

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

AUTOMATIZACIÓN DE LA MEDICIÓN DEL TAMAÑO
FUNCIONAL DEL SOFTWARE PARA MODELOS FUNCIONALES
OBTENIDOS A PARTIR DEL ANÁLISIS DINÁMICO DEL
CÓDIGO FUENTE

Tesis sometida a la consideración de la Comisión del
Programa de Estudios de Posgrado en Computación e Informática
para optar al grado y título de
Doctorado Académico en Computación e Informática

CHRISTIAN QUESADA LÓPEZ

Ciudad Universitaria Rodrigo Facio, Costa Rica

2018

DEDICATORIA

A mi mamá por mostrarme el valor del estudio y a mi papá por enseñarme el valor del trabajo duro. A mi familia por estar ahí siempre sin importar la distancia.

AGRADECIMIENTOS

Mi agradecimiento a todos los que contribuyeron directa e indirectamente para la finalización de esta disertación.

Mi gratitud para mi supervisor y consejero Dr. Marcelo Jenkins Coronas por su dirección y guía. El profesor Jenkins ha sido mi mentor y un modelo a seguir en mi carrera académica. Les agradezco a los asesores de tesis: Dra. Gabriela Marín Raventós y Dr. Luis Guerrero Blanco, por su tiempo e interés para ayudarme a culminar este proceso con éxito. Un agradecimiento a la profesora Marín y a los profesores Dra. Ileana Castillo Arias, Dra. Gabriela Barrantes Sliesarieva, Dr. Vladimir Lara Villagrán y Dra. Alexandra Martínez Porras por su apoyo y consejo durante estos años como estudiante de doctorado.

Mi agradecimiento a los profesores Dr. Guilherme Horta Travassos de la Universidad Federal de Río de Janeiro (UFRJ), Dr. Oscar Pastor López de la de la Universidad Politécnica de Valencia (UPV) y Dra. Laurie Williams de la Universidad Estatal de Carolina del Norte (NCSU) por ofrecerme la oportunidad de tener experiencias de aprendizaje extraordinarias. Un agradecimiento al Dr. Thiago Silva de Souza de la UFRJ y a su familia por su apoyo incondicional. Asimismo, al Dr. Sergio España y Dr. Ignacio Panach de la UPV y a la Dra. Maria Riaz de NCSU por darme la oportunidad de aprender a su lado.

Le agradezco a los colegas y amigos del grupo de investigación por sus aportes y apoyo durante mi investigación. Un agradecimiento a mis compañeros de doctorado Juan Murillo, Abel Méndez y Luis Esquivel con los que compartí muchas de las experiencias que nos transformaron durante estos años. Igualmente, agradezco a Denisse Madrigal, Luis Carlos Salas, Melissa Jensen y Laura Solera por todos sus aportes al trabajo de investigación y finalmente, agradezco a mis compañeros de generación del Doctorado en Computación e Informática por impulsarme todos los días para lograr este objetivo.

El desarrollo de este trabajo fue parcialmente apoyado por el Ministerio de Ciencia, Tecnología y Telecomunicaciones de Costa Rica (MICITT), el Consejo Nacional para Investigaciones Científicas y Tecnológicas de Costa Rica (CONICIT), el Sistema de Estudios de Posgrado (SEP) de la Universidad de Costa Rica (UCR), la Oficina de Asuntos Internacionales y Cooperación Externa (OAICE), la Escuela de Computación e Informática (ECCI), el Centro de Investigaciones en Tecnologías de la Información y Comunicación (CITIC) y la

Escuela de Administración de Negocios (EAN) de la UCR. Asimismo, el *International Software Benchmarking Standards Group* (ISBSG) y el *Pan American Software Quality Institute* (PASQI) por medio del *Latin American and Caribbean Collaboration ICT Research* (LACCIR) y la *National Science Foundation* (NSF). Las opiniones expresadas en este material son las de los autores y no reflejan necesariamente las opiniones de las entidades mencionadas.

“Esta tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Computación e Informática de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Doctorado Académico en Computación e Informática.”

Dra. Ileana Castillo Arias

**Representante del Decano
Sistema de Estudios de Posgrado**

Dr. Marcelo Jenkins Coronas

Director de Tesis

Dra. Gabriela Marín Raventós

Asesora

Dr. Luis Guerrero Blanco

Asesor

Dr. Vladimir Lara Villagrán

**Representante de la Directora
Programa de Posgrado en Computación e Informática**

Christian Quesada López

Candidato

TABLA DE CONTENIDOS

DEDICATORIA	III
AGRADECIMIENTOS	V
TABLA DE CONTENIDOS	IX
RESUMEN.....	XIII
LISTA DE TABLAS.....	XV
LISTA DE FIGURAS.....	XVIII
LISTA DE ANEXOS	XXI
LISTA DE ABREVIATURAS	XXII
PRÓLOGO	XXIII
BIOGRAFÍA	XXIV
CAPÍTULO 1. INTRODUCCIÓN	1
1.1 Motivación.....	5
1.2 Definición del problema y objetivos de la investigación.....	7
1.3 Metodología de investigación	11
1.4 Contribuciones	18
1.5 Estructura del documento de tesis.....	20
CAPÍTULO 2. MARCO TEÓRICO.....	23
2.1 Medición del software	23
2.1.1. Medición del tamaño del software.....	24
2.1.2. Medición del tamaño funcional del software.....	26
2.1.3. Fundamentos de la medición del tamaño funcional del software.....	35
2.2 Metodologías de investigación para la ingeniería de software	45
2.1.4. Ciencia del diseño para la ingeniería del software	45
2.1.5. Ingeniería de software empírica.....	47
CAPÍTULO 3. TRABAJO RELACIONADO.....	57
3.1 Estructura y aplicabilidad de los métodos de medición del tamaño funcional IFPUG FPA y COSMIC FFP	57
3.2 Marcos de trabajo para la evaluación de modelos de estimación de esfuerzo ba- sados en el tamaño funcional	59
3.3 Modelos de representación funcional del software	63
3.4 Estudios secundarios relacionados con la formalización de la medición del ta- maño funcional	66
CAPÍTULO 4. PROCEDIMIENTOS DE MEDICIÓN DEL TAMAÑO FUNCIONAL: UN MAPEO Y REVISIÓN SISTEMÁTICA DE LITERATURA	75
4.1. Diseño del estudio	76
4.2. Resultados del estudio	83

4.2.1.	Estado actual de la investigación en el área y características de los procedimientos de medición del tamaño funcional	85
4.2.2.	Estrategias empíricas utilizadas para evaluar los procedimientos de medición del tamaño funcional	103
4.2.3.	Desafíos y oportunidades de los procedimientos de medición del tamaño funcional	110
4.3.	Resumen de resultados	117
CAPÍTULO 5. DISEÑO DE LA METODOLOGÍA PARA LA AUTOMATIZACIÓN DE LA MEDICIÓN DEL TAMAÑO FUNCIONAL DEL SOFTWARE		
119		
5.1.	Procedimiento de medición del tamaño funcional del software	119
5.1.1.	Diseño del procedimiento de medición.....	120
5.1.2.	Aplicación del procedimiento de medición.....	151
5.2.	Procedimiento de verificación de exactitud.....	171
5.2.1.	Diseño del protocolo de verificación.....	172
5.2.2.	Aplicación del protocolo de verificación.....	175
5.2.3.	Salidas del proceso de aplicación del protocolo.....	181
5.2.4.	Métricas de verificación de exactitud.....	183
5.3.	Procedimiento de evaluación de los resultados de medición.....	184
5.3.1.	Aplicación del procedimiento de evaluación.....	186
5.3.2.	Salidas del procedimiento de evaluación.....	192
5.3.3.	Métricas de evaluación de desempeño.....	193
5.4.	Herramientas prototipo para el soporte de la metodología de medición del tamaño funcional aFPA	197
5.4.1.	Módulo de medición del tamaño funcional	197
5.4.2.	Módulo de verificación de la exactitud.....	201
5.4.3.	Módulo de evaluación de los modelos de estimación.....	203
5.5.	Extensibilidad del procedimiento de medición en distintos contextos de desarrollo	205
5.5.1.	Etapas de análisis de requerimientos utilizando lenguaje natural bajo el estándar IEEE Std. 830-1998 y el estándar ISO/IEC/IEEE 29148:2011	207
5.5.2.	Etapas de diseño utilizando modelos conceptuales construidos bajo una herramienta de desarrollo dirigido por modelos	212
5.6.	Resumen.....	217
CAPÍTULO 6. EVALUACIÓN EMPÍRICA DE LOS PROCEDIMIENTOS DE MEDICIÓN Y VERIFICACIÓN DE LA METODOLOGÍA aFPA		
219		
6.1.	Evaluación de la efectividad del procedimiento de medición del tamaño funcional: un conjunto de casos de estudio.....	220
6.1.1.	Caso de estudio 01: Sistema de Universidad Contoso	222
6.1.2.	Caso de estudio 02: Sistema de registro C-Reg	240
6.1.3.	Caso de estudio 03: Sistema de proyectos PMS	247
6.1.4.	Caso de estudio 04: Un caso de estudio en la industria	255
6.1.5.	Limitaciones de los casos de estudio	263
6.1.6.	Discusión de los resultados de los casos de estudio	264
6.2.	Evaluación de la efectividad del procedimiento de medición del tamaño funcional: un cuasi experimento y una familia de replicaciones	264
6.2.1.	Descripción del estudio empírico	265
6.2.2.	Análisis e interpretación de los resultados	281
6.2.3.	Discusión de los resultados del estudio	302

6.3.	Resumen y discusión de resultados de las evaluaciones empíricas	304
CAPÍTULO 7. VALIDACIÓN EMPÍRICA DEL PROCEDIMIENTO DE EVALUACIÓN DE LA METODOLOGÍA		
AFPA	307	
7.1.	Evaluación empírica de la estructura interna y la aplicabilidad de puntos de función IFPUG FPA: Una replicación externa	309
7.1.1.	Descripción de los estudios originales.....	310
7.1.2.	Replicación externa	315
7.1.3.	Comparación de resultados	327
7.1.4.	Discusión de resultados de la replicación	341
7.2.	Evaluación empírica de los componentes funcionales básicos de puntos de función COSMIC FFP para la construcción de modelos de estimación de esfuerzo: Un estudio empírico	343
7.2.1.	Diseño del estudio.....	344
7.2.2.	Resultados del estudio.....	352
7.2.3.	Discusión de resultados del estudio	360
7.3.	Análisis de puntos de función para la creación de modelos de estimación de esfuerzo: Un estudio empírico.....	362
7.3.1.	Diseño del estudio.....	363
7.3.2.	Resultados del estudio.....	371
7.3.3.	Discusión de resultados del estudio	401
7.4.	Resumen y discusión de resultados de las evaluaciones empíricas	403
CAPÍTULO 8. CONCLUSIONES		
	405	
8.1.	Resumen de resultados	405
8.2.	Contribuciones	409
8.3.	Direcciones futuras.....	410
REFERENCIAS		413
ANEXOS.....		439

RESUMEN

Esta investigación propone una metodología para la automatización de la medición del tamaño funcional del software a partir del código fuente. En esta tesis se desarrolla **aFPA**, una metodología que permite realizar el conteo automático del tamaño funcional basada en los métodos de medición estandarizados IFPUG FPA y COSMIC FFP. La metodología **aFPA** toma como entrada el código fuente de la aplicación para construir un modelo funcional cuyos componentes son mapeados a los métodos de medición para producir la estimación del tamaño, manteniendo así una trazabilidad entre los artefactos de software y los resultados del conteo. Adicionalmente, **aFPA** es extendida para construir el modelo funcional a partir de los requerimientos funcionales bajo el estándar IEEE 830-1998 y los modelos conceptuales generados por una herramienta de desarrollo dirigido por modelos en un contexto de industria.

La metodología **aFPA** sigue la estructura de componentes de un modelo de procesos para métodos de medición del tamaño funcional, y se compone de tres procedimientos: el primero permite realizar la medición automática del tamaño funcional, el segundo verifica la exactitud de las mediciones obtenidas, y el tercero evalúa modelos de estimación de esfuerzo simplificados para mostrar un posible aprovechamiento de los resultados de la medición del tamaño funcional. Para probar la utilidad de la metodología, se implementa una herramienta prototipo para cada uno de estos tres procedimientos y se analizan los resultados de su utilización.

La metodología **aFPA** fue evaluada mediante un conjunto de estudios empíricos que muestran su factibilidad en procesos de desarrollo de software. La investigación fue estructurada de acuerdo al marco de trabajo Ciencia del Diseño para Sistemas de Información e Ingeniería del Software. Primero, se investigaron los desafíos y enfoques existentes para la automatización de la medición del tamaño funcional mediante un mapeo y revisión sistemática de literatura. Posteriormente, se diseñó la metodología **aFPA** mediante el modelo de procesos para métodos de medición del tamaño funcional y las herramientas prototipo mediante metodologías de desarrollo de software. Finalmente, se evaluó la exactitud de los resultados de medición mediante cuatro casos de estudio, la efectividad y las propiedades de adopción mediante cuatro estudios controlados cuasi experimentales orientados a personas y el aprovechamiento de los resultados de medición de tamaño funcional mediante tres replicaciones de estudios experimentales.

Los resultados indicaron que la metodología **aFPA** obtuvo una exactitud y reproducibilidad aceptable ($\pm 10\%$). Asimismo, la exactitud y reproducibilidad se mantuvo similar al medir artefactos de código fuente, requerimientos funcionales y modelos conceptuales. El procedimiento de medición de la metodología **aFPA** fue percibida por los profesionales como de igual facilidad de uso, utilidad e intención de uso que el método de medición IFPUG FPA. El procedimiento de verificación fue percibido como fácil de usar, útil y con intención de uso futuro. Finalmente, el procedimiento de evaluación mostró que los resultados del tamaño funcional podrían ser aprovechados para la construcción de modelos de estimación simplificados basados únicamente en conjuntos de los componentes funcionales básicos de los métodos de medición del tamaño funcional. Con el desarrollo de esta investigación contribuimos con una metodología validada empíricamente para la automatización de la medición del tamaño funcional.

LISTA DE TABLAS

Tabla 3-1 Estudios secundarios sobre medición del tamaño funcional.....	72
Tabla 4-1 Cantidad de artículos retornados por base de datos.....	83
Tabla 4-2 Publicaciones por año.....	86
Tabla 4-3 Foros de publicación por año.....	87
Tabla 4-4 Principales foros de publicación.	88
Tabla 4-5 Publicaciones por año y método de medición del tamaño funcional.	90
Tabla 4-6 Listado de artículos primarios.	93
Tabla 4-7 Artefactos por año y método de medición del tamaño funcional.	96
Tabla 4-8 Distribución de los estudios empíricos.....	104
Tabla 5-1 Puntos de función sin ajustar asignados por complejidad.....	149
Tabla 6-1 Estudios empíricos realizados para evaluar la metodología aFPA.....	220
Tabla 6-2 Resultados de medición del tamaño funcional.	236
Tabla 6-3 Exactitud de los resultados de medición del tamaño funcional.....	237
Tabla 6-4 Resultados de medición del tamaño funcional por requerimiento.	238
Tabla 6-5 Resultados de medición del tamaño funcional.	244
Tabla 6-6 Exactitud de los resultados de medición del tamaño funcional.....	244
Tabla 6-7 Resultados de medición del tamaño funcional por requerimiento.	246
Tabla 6-8 Resultados de medición del tamaño funcional.	251
Tabla 6-9 Exactitud de los resultados de medición del tamaño funcional.....	251
Tabla 6-10 Resultados de medición del tamaño funcional por requerimiento.	254
Tabla 6-11 Resultados de medición del tamaño funcional.	259
Tabla 6-12 Exactitud de los resultados de medición del tamaño funcional.....	259
Tabla 6-13 Resultados de medición del tamaño funcional por requerimiento.	261
Tabla 6-14 Diseño de los estudios empíricos.....	270
Tabla 6-15 Decisiones y configuración de los estudios empíricos.....	271
Tabla 6-16 Objetos de estudio.....	278
Tabla 6-17 Frecuencia de participantes por estudio.....	281
Tabla 6-18 Estadística descriptiva de los resultados de medición del tamaño funcional.....	282
Tabla 6-19 Exactitud (R) y reproducibilidad (R) de los resultados de medición.	284
Tabla 6-20 Exactitud y reproducibilidad de los resultados agregados.....	289
Tabla 6-21 Exactitud y reproducibilidad de los BFC.....	294

Tabla 6-22 Relaciones de los constructos con la intensidad de uso.....	300
Tabla 7-1 Resumen de objetivos y preguntas de investigación.	311
Tabla 7-2 Conjuntos de datos de los estudios originales.	312
Tabla 7-3 Listado de variables dependientes e independientes.	312
Tabla 7-4 Resumen de técnicas utilizadas en los estudios originales.....	314
Tabla 7-5 Criterios de selección de conjuntos de datos (DS).	319
Tabla 7-6 Detalle de variables seleccionadas en el estudio.....	321
Tabla 7-7 Conjuntos de datos utilizados en la replicación.	322
Tabla 7-8 Estadística descriptiva de los conjuntos de datos seleccionados.	323
Tabla 7-9 Tamaño funcional sin ajustar por variables de contexto.....	326
Tabla 7-10 Resultados de la relación de los UFP con el esfuerzo.	328
Tabla 7-11 Coeficientes de correlación <i>Kendall Tau</i> entre los BFC y los UFP.....	330
Tabla 7-12 Coeficientes de correlación entre los BFC, el UFP y el esfuerzo.....	333
Tabla 7-13 Modelos de estimación de esfuerzo basados en el UFP y los BFC.	336
Tabla 7-14 Evaluación de exactitud de los modelos de estimación de esfuerzo.	338
Tabla 7-15 Resultados de la prueba <i>Wilcoxon Signed Rank (W)</i> y <i>Sign test (S)</i>	341
Tabla 7-16 Criterios de selección del conjunto de datos (DS).....	345
Tabla 7-17 Conjuntos de datos utilizados en el estudio.	346
Tabla 7-18 Estadística descriptiva de los conjuntos de datos seleccionados.	349
Tabla 7-19 Tamaño funcional por variables de contexto.	352
Tabla 7-20 Coeficientes de correlación <i>Kendall Tau</i> entre los BFC y el CFP.	354
Tabla 7-21 Coeficientes de correlación entre los BFC, el CFP y el esfuerzo.....	355
Tabla 7-22 Modelos de estimación de esfuerzo basados en el CFP y los BFC.	358
Tabla 7-23 Resultados de la prueba <i>Wilcoxon Signed Rank (W)</i>	360
Tabla 7-24 Criterios de selección de conjuntos de datos (DS).	366
Tabla 7-25 Conjuntos de datos utilizados en el estudio.	367
Tabla 7-26 Grupos de datos utilizados en el estudio.	368
Tabla 7-27 Técnicas utilizadas en la construcción de los modelos de estimación.....	369
Tabla 7-28 Configuraciones de las técnicas utilizadas.....	371
Tabla 7-29 Estadística descriptiva de los resultados de evaluación de exactitud.	374
Tabla 7-30 Resultados de la repetibilidad para los modelos.	378
Tabla 7-31 Resultados de la comparación de los modelos para el método IFPUG FPA.....	384
Tabla 7-32 Comparación de los mejores modelos para el método IFPUG FPA.....	385

Tabla 7-33 Resultados de la comparación de los modelos para el método COSMIC FFP.	390
Tabla 7-34 Comparación de los mejores modelos para el método COSMIC FFP.	391
Tabla 7-35 Mejores modelos de estimación de esfuerzo por grupo.	395

LISTA DE FIGURAS

Figura 1-1 Metodología para la automatización de la medición del tamaño funcional del software aFPA.	10
Figura 1-2 Proyecto de Ciencia del Diseño para la automatización de la medición del tamaño funcional del software.....	12
Figura 1-3 Metodología de investigación.....	13
Figura 1-4 Metodología de investigación detallada.	22
Figura 2-1 Dominios de medición del software. Adaptado de (Abrahão, 2004).....	25
Figura 2-2 Estándares de medición del tamaño funcional. Adaptado de (ISO, 2007).	27
Figura 2-3 Evolución de los métodos de medición del tamaño funcional. Actualizado a partir de (Condori-Fernández, 2006; Abran, 2015).	29
Figura 2-4 Componentes del método IFPUG FPA. Adaptado de (Runeson, Borgquist, Landin & Bolanowski, 2000).	32
Figura 2-5 Componentes del método COSMIC FFP. Adaptado de (ISO, 2011).....	33
Figura 2-6 Fundamentos de medición y automatización. Adaptado de (Abran, 2010)...	37
Figura 2-9 Elementos de un procedimiento de medición (Abran, 2010).....	38
Figura 2-7 Pasos generales del modelo de procesos. Adaptado de (Abran, 2010).	40
Figura 2-8 Pasos detallados del modelo de procesos. Adaptado de (Abran, 2010).	44
Figura 2-10 Marco de trabajo de la metodología Ciencia del Diseño. Adaptado de (Wieringa, 2014).	46
Figura 2-11 Proceso de la revisión sistemática de literatura. Adaptado de (Kitchenham & Charters, 2007).	50
Figura 2-12 Proceso del mapeo sistemática de literatura. Adaptado de (Petersen, Feldt, Mujtaba & Mattsson, 2008).....	50
Figura 2-13 Proceso de experimentación. Adaptado de (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012).	52
Figura 2-14 Estructura del protocolo de casos de estudio en comparación con el proceso de experimentación empírica. Adaptado de (Runeson & Höst, 2009; Runeson, Host, Rainer & Regnell, 2012).....	54
Figura 4-1 Esquema de clasificación.	82
Figura 4-2 Artículos primarios en las bases de datos de búsqueda.....	83
Figura 4-3 Resultados del proceso de selección de estudios primarios.....	84
Figura 4-4 Calidad de los artículos primarios incluidos.	85
Figura 4-5 Listado de los artículos primarios seleccionados.	92

Figura 5-1 Metamodelo del método IFPUG FPA. Adaptado de (Abrahão, 2004; Hericko, Rozman & Zivkovic, 2006; Lavazza, 2015).	124
Figura 5-2 Metamodelo del método COSMIC FFP. Adaptado de (Hericko, Rozman & Zivkovic, 2006; Marín, 2011).	127
Figura 5-3 Metamodelo del procedimiento aFPA.	129
Figura 5-4 Modelo de clases para la representación del modelo funcional.	134
Figura 5-5 Mapeo con el metamodelo IFPUG FPA.	140
Figura 5-6 Mapeo con el metamodelo COSMIC FFP.	144
Figura 5-7 Pasos de la metodología para la automatización de la medición de tamaño funcional aFPA.	154
Figura 5-8 Niveles de automatización 0 y 1 del procedimiento de medición.	159
Figura 5-9 Niveles de automatización 2 y 3 del procedimiento de medición.	160
Figura 5-10 Modelo funcional de la aplicación <i>Customer</i>	166
Figura 5-11 Modelo de recolección de datos del protocolo.	174
Figura 5-12 Fases del protocolo de verificación de exactitud.	176
Figura 5-13 Niveles de automatización 0 y 1 del protocolo de verificación de exactitud.	181
Figura 5-14 Componentes de la estrategia de evaluación de modelos de estimación.	187
Figura 5-15 Proceso del procedimiento de aprovechamiento de resultados para la construcción de modelos de estimación.	192
Figura 5-16 Prototipos para el soporte de la metodología aFPA.	197
Figura 5-17 Componentes del módulo de medición.	198
Figura 5-18 Componentes del módulo de verificación.	202
Figura 5-19 Componentes del módulo de evaluación de modelos.	204
Figura 5-20 Metamodelo de la especificación de requerimientos.	209
Figura 5-21 Metamodelo de la plataforma dirigida por modelos.	213
Figura 6-1 Secuencia de casos de estudio realizados.	222
Figura 6-2 Lista de librerías configuradas en la aplicación.	229
Figura 6-3 Ejemplo del código instrumentado.	229
Figura 6-4 Ejemplo de trazo de la ejecución de la aplicación.	230
Figura 6-5 Ejemplo de archivo de interfaz XML.	230
Figura 6-6 Ejemplo de salidas de la simulación de la ejecución.	232
Figura 6-7 Ejemplo del modelo funcional en <i>Cypher</i>	232
Figura 6-8 Visualización del modelo funcional.	233
Figura 6-9 Librerías del componente <i>Constructor, Medidor y Reporteador</i>	234

Figura 6-10 Extracto de reporte de resultados del componente <i>Reporteador</i>	235
Figura 6-11 Cronología de la familia de replicaciones.....	266
Figura 6-12 Resultados de las mediciones de tamaño por estudio.....	283
Figura 6-13 Exactitud y reproducibilidad de los resultados de medición por estudio..	287
Figura 6-14 Exactitud y reproducibilidad de los resultados agregados.....	290
Figura 7-1 Tamaño funcional sin ajustar por variables de contexto.....	324
Figura 7-2 Tamaño funcional sin ajustar por variables de contexto.....	325
Figura 7-3 Relación de los UFP con el esfuerzo.....	328
Figura 7-4 Tamaño funcional por variables de contexto.....	350
Figura 7-5 Tamaño funcional por variables de contexto.....	351
Figura 7-6 Métricas de exactitud por conjuntos de datos.....	375
Figura 7-7 Métricas de exactitud para los mejores modelos del método IFPUG FPA por conjunto de datos.....	387
Figura 7-8 Métricas de exactitud para los mejores modelos del método COSMIC FFP por conjunto de datos.....	392
Figura 7-9 Mejores modelos (SA) para el conjunto de datos IFPUG FPA.....	399
Figura 7-10 Mejores modelos (SA) para el conjunto de datos COSMIC FFP.....	400

LISTA DE ANEXOS

Anexo 1 Diseño del estudio de mapeo y revisión sistemática de literatura.	439
Anexo 2 Formulario de extracción de los artículos relevantes del estudio secundario.	453
Anexo 3 Listado de artículos primarios del mapeo y revisión sistemática de literatura.	454
Anexo 4 Amenazas a la validez de los estudios empíricos orientados a personas.	478
Anexo 5 Tamaño funcional sin ajustar IFPUG FPA por variables de contexto.	482
Anexo 6 Tamaño funcional sin ajustar COSMIC FFP por variables de contexto.	485
Anexo 7 Estadística descriptiva de los resultados de la evaluación de exactitud de los modelos de estimación de esfuerzo.	487
Anexo 8 Mejores modelos de estimación de esfuerzo por grupo de análisis.	494
Anexo 9 Amenazas a la validez de los estudios empíricos orientados a personas.	496
Anexo 10 Artículos científicos publicados durante el proceso de la tesis.	499

LISTA DE ABREVIATURAS

- AFP: Puntos de función automatizados (*Automated Function Points*).
- AIAG: Grupo de acción de la industria automotriz (*Automotive Industry Action Group*).
- BFC: Componentes funcionales básicos (*Basic Functional Components*).
- CITIC: Centro de investigaciones en tecnologías de la información y comunicación.
- COSMIC: Consorcio internacional de medición de software (*Common Software Measurement International Consortium*).
- FPA: Análisis de puntos de función (*Function Point Analysis*).
- FSM: Medición del tamaño funcional (*Functional Size Measurement*).
- IEEE: Instituto de ingenieros eléctricos y electrónicos (*Institute of Electrical and Electronics Engineers*).
- IFPUG: Grupo internacional de usuarios de puntos de función (*International Function Points User Group*).
- LOC: Líneas de código (*Lines of Code*).
- Metrología: Ciencia que estudia los sistemas de medición.
- MIS: Sistemas de información gerencial (*Management Information Systems*).
- OMG: Grupo de administración de objetos (*Object Management Group*).
- SEI: Instituto de ingeniería del software (*Software Engineering Institute*).
- RTS: Sistemas en tiempo real (*Real Time Systems*).

PRÓLOGO

Este documento de tesis resume cinco años de investigación en el área de la medición del tamaño funcional del software y de la ingeniería del software empírica. En este proyecto, junto con el profesor Marcelo Jenkins, nos planteamos el reto de estudiar las metodologías para realizar el proceso de medición de tamaño funcional automáticamente a partir de los artefactos del código fuente de las aplicaciones. Producto de la investigación se desarrolló **aFPA**, una metodología para la automatización de la medición del tamaño funcional. Durante las evaluaciones empíricas realizadas en esta tesis se obtuvieron resultados prometedores y se determinaron oportunidades de mejora. Actualmente, se realizan investigaciones empíricas adicionales para permitir mejorar la metodología **aFPA** y para identificar nuevas oportunidades de investigación.

Con la realización de este doctorado, he crecido personal y profesionalmente. He disfrutado desarrollando esta investigación y he aprendido día con día. Entre muchas experiencias valiosas, he tenido la oportunidad de conocer el área de la ingeniería de software empírica (ESE). Esto me ha dado la oportunidad de estudiar las teorías y metodologías utilizadas por la comunidad de la ESE, aprender sobre la rigurosidad de las investigaciones que se desarrollan en el área y conocer cómo estas metodologías de investigación científica han ido evolucionando. En la ESE, he descubierto un campo fascinante y una oportunidad para mejorar todas mis habilidades como profesional. La ESE me ha permitido encontrar mi pasión como investigador, donde mi deseo es tener la oportunidad de seguir aprendiendo y aportando a la comunidad de la ESE mediante investigaciones empíricas. Me siento ilusionado en seguir trazando el camino que he elegido como investigador empírico y que inicia con la culminación de este proceso doctoral.

Durante el desarrollo de esta tesis, tuve la oportunidad de investigar un tema que me apasiona compartiendo con un gran equipo de profesionales en el Grupo de Investigación en Ingeniería del Software Empírica de la Escuela de Ciencias de la Computación e Informática de la Universidad de Costa Rica. Deseo agradecer a todos mis mentores, profesores y compañeros de investigación, sin ellos y ellas no hubiera sido posible finalizar esta tesis.

BIOGRAFÍA

Christian Quesada López completó su Bachillerato en Ingeniería en Computación del Instituto Tecnológico de Costa Rica (ITCR) en el 2002. En el 2009, realiza su Maestría en Computación e Informática en la Universidad de Costa Rica (UCR), la cual completa con honores en el 2013. Christian se unió a la UCR como docente e investigador desde el 2010, primero como docente en la Escuela de Administración de Negocios (EAN) y luego como investigador en el Centro de Investigaciones en Tecnologías de la Información y Comunicación (CITIC). Posteriormente, se incorpora como docente en el Programa de Maestría en Computación e Informática (PCI) y la Escuela de Ciencias de la Computación e Informática (ECCI) de la UCR. Después de una carrera profesional de 10 años en el sector público-privado y dado su fuerte inclinación hacia la investigación y la docencia universitaria, se incorpora al Programa de Doctorado en Computación e Informática de la UCR en el 2013. Sus intereses en investigación y docencia son los relacionados con la ingeniería de software (SE), SE empírica, metodologías de desarrollo, aplicaciones móviles, calidad de software e inteligencia artificial aplicada a la SE.

Capítulo 1. INTRODUCCIÓN

La medición del tamaño del software es fundamental en la disciplina de la Ingeniería del Software (Abran, 2010). El concepto de “medición” se define como “el proceso por el cual números o símbolos se asignan a los atributos de las entidades del mundo real para describirlas de acuerdo a un conjunto de reglas claramente definidas”, donde una medida es “el número o símbolo asignado a una entidad que caracteriza un atributo” (Fenton & Pfleeger, 1997). Según el Instituto de Ingeniería de Software (SEI), el proceso de medición implica la recolección y el análisis de datos cuantitativos acerca de los productos, procesos y proyectos de software para apoyar el proceso de toma de decisiones en una organización. La medición del software permite cuantificar los procesos de desarrollo del software y es esencial para el control de proyectos, productos y procesos. Asimismo, la medición es parte fundamental en la ejecución de estudios empíricos aplicados a la ingeniería de software (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012). Por ejemplo, los programas de medición de las organizaciones de desarrollo de software que comparan los datos de los procesos (*benchmarking*) y evalúan modelos de calidad, predicción de fallos, estimación de esfuerzo, productividad y costos (Abran, 2010).

El tamaño del software puede ser medido desde dos perspectivas, la perspectiva funcional o la perspectiva técnica. La perspectiva funcional es la del usuario final y la perspectiva técnica es la de los desarrolladores de software. El tamaño funcional es definido por el estándar ISO/IEC 14143-1 (2007) como “el tamaño del software derivado por la cuantificación de los requisitos funcionales de usuario (FUR)”. Las mediciones de tamaño funcional soportan la perspectiva del usuario final (Abrahão, 2004) y no deben verse afectados por aspectos de la implementación técnica y las tecnologías utilizadas. La serie de estándares ISO/IEC 14143 proveen un marco de trabajo para desarrollar, probar y refinar los métodos de medición del tamaño funcional. A partir de la aparición de este estándar, han surgido diferentes métodos para la medición funcional del software, en la actualidad, los principales métodos de medición del tamaño funcional son el estándar IFPUG FPA (ISO, 2009) y el estándar COSMIC FFP (ISO, 2011) los cuales son mantenidos por el *International Function*

Points User Group (IFPUG) y el *Common Software Measurement International Consortium* (COSMIC) respectivamente.

Durante el proceso del desarrollo del software, el tamaño funcional puede ser utilizado para la estimación del esfuerzo de proyectos de desarrollo, evaluación de calidad, evaluaciones comparativas mediante bases de datos de proyectos históricas y para la negociación de contratos. Asimismo, puede ser usado para presupuestar el desarrollo de proyectos, administrar y dar seguimiento a su ejecución, negociar modificaciones, determinar la cantidad de requerimientos entregados, estimar el activo total del software de una organización, administrar la productividad del desarrollo y/o mantenimiento y monitorear la calidad. Ha sido comprobado que la medición del tamaño funcional es uno de los principales predictores que influyen el esfuerzo de desarrollo y es clave para determinar el tiempo y el costo de los proyectos de software (Boehm, 1983; González-Ladrón-de-Guevara, Fernández-Diego & Lokan, 2016).

La medición del tamaño del software como entrada para los modelos de estimación del esfuerzo de desarrollo ha sido estudiada durante muchos años, sin embargo, en la industria no han sido adoptados como una práctica común. Muchas organizaciones utilizan aún el juicio experto, lo que produce estimaciones inexactas y atrasos en las fechas de finalización estimadas para los proyectos de software (Boehm, 1983; Molokken & Jorgensen, 2003). El uso correcto de los modelos de estimación es complejo y consume mucho tiempo, las organizaciones al tener que priorizar sus recursos limitados, comúnmente no los invierten en entrenamiento para el uso de modelos de estimación formales (Jorgensen, Boehm & Rifkin, 2009). Aún más, la construcción de un modelo de estimación de esfuerzo puede ser muy compleja y requerir una inversión de tiempo considerable (Jørgensen, 2007).

De acuerdo al reporte de *The Standish Group* (2015), solo el 29% de los proyectos de software son exitosos en cuanto al tiempo, el alcance y el costo estimados. En el 2013, este informe indicó que el 74% de los proyectos exceden el tiempo estimado, el 59% de los proyectos exceden los costos y en promedio solo el 69% de la funcionalidad es entregada a los usuarios finales. Este informe establece que uno de los factores de éxito es el tamaño de los

proyectos de desarrollo del software y que estos deben ser medidos mediante la utilización de metodologías fáciles y visuales para monitorear su progreso.

El modelo de madurez CMMI para la mejora de procesos (SEI, 2010), en su nivel 2, detalla las prácticas relacionadas con la administración de proyectos donde la medición del tamaño desempeña un rol sumamente importante (Condori-Fernández, 2006). Este modelo de madurez contiene los elementos esenciales para lograr un proceso de ingeniería de software efectivo, el cual debe ser orientado por la medición (Bégnocche, Abran & Buglione, 2007). La medición del tamaño del software es parte importante del proceso de desarrollo (Garmus & Herron, 2001) y es uno de los principales factores que potencialmente afecta el tiempo y costo de los proyectos de desarrollo de software (Boehm, 1983; Abrahão, 2004; González-Ladrón-de-Guevara, Fernández-Diego & Lokan, 2016).

Para que las mediciones de tamaño funcional sean consideradas una métrica, estas deben estar claramente definidas, tener propiedades matemáticas claras, y ser demostrablemente razonables en términos de precisión y exactitud, validez y confiabilidad (Abran, 2010). Para ello la medición del tamaño funcional debe realizarse mediante procesos sistemáticos. En los últimos años, la estructura y aplicabilidad de los métodos de medición IFPUG FPA y COSMIC FFP han sido investigadas para determinar distintos problemas que afectan la consistencia, confiabilidad y exactitud de los resultados de medición (De Freitas Junior, Fantinato & Sun, 2015; Silva, Pinheiro & Albuquerque, 2016).

El tamaño funcional es una de las técnicas más utilizadas para medir el tamaño del software (Bajwa, Gencel & Abrahamsson, 2014) pero su proceso de conteo es difícil de automatizar (Condori-Fernández, 2006; Huijgens, Bruntink, van Deursen, van der Storm & Vogelezang, 2016; Silva, Pinheiro & Albuquerque, 2016) lo que lo convierte en un proceso lento y costoso. En los últimos años, la investigación relacionada con la automatización de la medición del tamaño funcional se ha convertido en un tema cada vez más importante para la comunidad de la ingeniería del software (Bajwa, Gencel & Abrahamsson, 2014; Lavazza, 2015). Con la automatización del proceso de conteo de puntos de función, que se usa para medir el tamaño funcional del software, se alcanzan varios beneficios tales como la reducción del tiempo, esfuerzo y costo del conteo y la mejora en la confiabilidad, exactitud

y repetibilidad de los resultados de las mediciones (Lavazza, 2015). Además, un proceso de automatización de la medición del tamaño funcional podría convertirse en un componente estándar del proceso de desarrollo y mantenimiento del software para generar información histórica confiable de los proyectos para los procesos de *benchmarking* (Object Management Group, 2014).

Se hace necesario entonces realizar investigaciones que estudien la automatización del proceso de conteo de puntos de función para ayudar a entender los beneficios reales de la aplicación automatizada del proceso de conteo, comprobar su nivel de exactitud, confiabilidad, eficiencia y repetibilidad y determinar su factibilidad en contextos específicos de desarrollo de software. Además, la evidencia empírica generada a partir de estas investigaciones permitirá analizar y comparar las relaciones entre los diferentes procesos de medición (Lavazza, 2015). Es importante considerar que aunque el proceso de automatización de los métodos de medición del tamaño funcional es percibido como una herramienta importante para la toma de decisiones, aún presenta grandes variaciones en los resultados de medición que no son aceptables para los profesionales en medición (Heller, 2013; Lavazza, 2015; Huijgens, Bruntink, van Deursen, van der Storm & Vogelezang, 2016). Dado que la automatización del conteo de puntos de función es deseable, para demostrar la exactitud de los procesos de medición del tamaño funcional de forma independiente es necesario definir y utilizar protocolos de verificación de exactitud estandarizados. Estos deben asegurar que el proceso de medición produzca los resultados correctos (Soubra, Abran & Ramdane-Cherif, 2014).

En general, son requeridos procedimientos formales, automatizados y validados que puedan ser auditados para garantizar la calidad de los resultados del proceso de medición. En la actualidad existe poca investigación acerca de la verificación de los resultados de medición producidos por los diferentes procedimientos propuestos para la medición del tamaño funcional del software (Soubra, Abran & Ramdane-Cherif, 2014). Algunas de las dificultades para evaluar las nuevas propuestas de medición en un nivel práctico se dan por

la falta de rigurosidad en las validaciones empíricas y la ausencia de evaluaciones sistemáticas. Esto puede explicar la baja tasa de adopción de estas propuestas en la industria (Abrahão, Poels & Pastor, 2004).

Symons (2014) advierte que la información publicada acerca de las métricas de software debe ser interpretada con mucho cuidado. El autor indica que resultados de medición del tamaño funcional que presenten un nivel de exactitud muy alto, tienen que analizarse detalladamente, ya que todos los métodos de medición tienen sus limitaciones y muchos de los resultados obtenidos mediante procesos rápidos y automatizados, no cuentan con un respaldo científico comprobado.

1.1 MOTIVACIÓN

La medición del tamaño funcional del software es un proceso clave durante el desarrollo del software y provee información valiosa para la toma de decisiones de los administradores de proyectos y los ingenieros de software. Sin embargo, obtener el tamaño funcional del software es un proceso lento, costoso y requiere la intervención de expertos en medición. Asimismo, las herramientas para estimar el tamaño funcional a partir de artefactos de software tales como el código fuente son escasas, costosas y no han dado los resultados de exactitud esperados por la industria (Heller, 2013). Por lo anterior, se requieren metodologías para la automatización de la medición del tamaño funcional científicamente validadas que permitan a los profesionales mejorar la efectividad en el proceso de medición del tamaño funcional. Una metodología de medición automatizada del tamaño funcional del software puede permitir a la industria obtener mediciones consistentes de una forma eficiente, y así apoyar los procesos de toma de decisiones sobre sus proyectos de desarrollo de software (Object Management Group, 2014).

En los últimos años, se han presentado múltiples propuestas para la automatización de la medición del tamaño funcional donde la mayoría de ellas se basan en los métodos de medición IFPUG FPA y COSMIC FFP. La mayor motivación de este tipo de estudios es automatizar el proceso de la medición del tamaño funcional a partir de artefactos de software específicos generados durante el proceso de desarrollo de software y así reducir el tiempo

del proceso de conteo y mejorar la consistencia de los resultados de las mediciones obtenidas.

La medición representa un reto para las organizaciones de desarrollo de software. Sin embargo, si estas se conducen de forma correcta, representan una ventaja competitiva para las mismas (Abran, Meli & Symons, 2004; Huijgens, Bruntink, van Deursen, van der Storm & Vogelezang, 2016). Una de las principales preocupaciones en la industria del software es el desempeño de los procesos de medición para poder realizar evaluaciones comparativas del tamaño funcional. La mayoría de las herramientas actuales no automatizan totalmente el proceso de la medición del tamaño funcional, si no que más bien son herramientas de soporte al proceso de medición que realiza un experto (Stambollian & Abran, 2006; Quesada-López & Jenkins, 2017b). Asimismo, muchas de estas herramientas se basan en procesos de convertibilidad entre las líneas de código y el tamaño funcional (Huijgens, Bruntink, van Deursen, van der Storm & Vogelezang, 2016). Las herramientas para la automatización parcial o total del proceso de conteo del tamaño funcional deben desarrollarse tomando en cuenta la exactitud de los resultados de las mediciones a partir de su validación empírica rigurosa (Gencel & Demirors, 2008).

Para mejorar los procesos de medición del tamaño funcional, es necesaria la formalización de los artefactos de entrada del proceso de medición que representan los requerimientos funcionales. La formalización del proceso de medición, con reglas claramente definidas, habilita la posibilidad de automatización que permitirá reducir el esfuerzo y costo de proceso de conteo del tamaño funcional e incrementar la eficiencia del mismo (Marín, Giachetti & Pastor, 2008). La formalización de los métodos de medición, a partir de la aplicación de reglas específicas sobre especificaciones formales del software, incrementa las posibilidades de automatización. Dicha automatización, debe realizarse a partir de la interpretación consistente de las reglas de medición para así garantizar la consistencia y la repetibilidad de los resultados de medición (Özkan & Demirörs, 2010). Finalmente, la poca evidencia sobre la validación de las herramientas existentes que proponen la automatización de la medición del tamaño funcional, representa un vacío en el conocimiento, y por

tanto, una necesidad para el área de investigación de la medición funcional del software (Barkallah, Gherbi & Abran, 2011).

Por todo lo anterior, se requieren metodologías de medición sistematizados para contextos de desarrollo de software específicos, los cuales deben basarse en las mejores prácticas y las lecciones aprendidas de los estudios empíricos realizados en el área de la medición del tamaño funcional. Estas deben especificar el conjunto de artefactos de software de entrada y el nivel de detalle necesario para su aplicación en el proceso de medición que permitan mantener la exactitud en los resultados de las mediciones. Dichos resultados de medición, deben ser validados mediante la aplicación de protocolos de verificación sistematizados que evalúen rigurosamente la exactitud y la efectividad de los procesos de medición para garantizar la calidad de los resultados. Los procesos de medición deben realizarse de forma sistemática para que provean resultados consistentes. Estos deben mantener trazabilidad hacia los requerimientos funcionales de los usuarios finales y mostrar de una forma simple y visual los resultados de las mediciones para ser analizados por los tomadores de decisiones. Finalmente, los resultados de medición deben ser evaluados para determinar la manera en que pueden ser aprovechados para la construcción de distintos modelos de predicción tales como los modelos de estimación de esfuerzo (Abran, 2010).

1.2 DEFINICIÓN DEL PROBLEMA Y OBJETIVOS DE LA INVESTIGACIÓN

La automatización de la medición del tamaño funcional busca lograr la reducción en el tiempo y en el costo del proceso de conteo y la mejora en la confiabilidad, exactitud, y repetibilidad de los resultados de medición. La automatización a partir de artefactos de software tales como el código fuente es una tarea compleja, puesto que su proceso implica una transformación y mapeo desde el dominio técnico al dominio del problema dado que a partir de los artefactos del software se deben representar las funcionalidades del usuario final. Estas dos perspectivas no están necesariamente correlacionadas entre sí (Meli, 2000; Ellafi & Meli, 2006). Una de las principales características de las métricas funcionales es que se pueden obtener en las etapas iniciales del ciclo de vida del desarrollo del software. En el

caso de las métricas técnicas, estas se obtienen a partir de artefactos de software que se construyen en etapas avanzadas del desarrollo. La automatización de la medición del tamaño funcional en etapas avanzadas del desarrollo basado en, por ejemplo, el código fuente permite la comparación de los resultados en las distintas etapas del ciclo de vida del desarrollo. Sumado a esto, en el caso de los equipos de desarrollo utilizando metodologías ágiles, donde en la mayoría de los casos, solo se utilizan historias de usuario (Trudel, Desharnais & Cloutier, 2016) y en el caso de la industria de software costarricense la cual parece haber adoptado las metodologías ágiles para el desarrollo, el código fuente se convierte en uno de los pocos artefactos de software que se mantienen actualizados y con los cuales se puede calcular las distintas métricas necesarias para administrar los procesos (Quesada-López & Jenkins, 2017a; 2018). El reto es obtener una representación de un modelo funcional a partir de los artefactos técnicos y obtener el tamaño funcional desde la perspectiva del usuario final con una calidad verificable.

El problema a resolver consiste en realizar el proceso de conteo del tamaño funcional de aplicaciones automáticamente, a partir de los artefactos de software tales como los del código fuente. El proceso debe recibir como entrada los artefactos de software, los cuales analiza estática y dinámicamente para construir un modelo funcional basado en las transacciones las cuales son mapeadas a los componentes funcionales de los métodos de medición IFPUG FPA y COSMIC FFP. Basado en este mapeo, aplica las reglas de asignación numérica de tamaño funcional, verifica la confiabilidad de los resultados y evalúa su utilidad para la construcción de distintos modelos de estimación de esfuerzo. El proceso de medición debe mantener trazabilidad entre los requerimientos del usuario final y los artefactos de software de la aplicación, e idealmente, presentar visualmente el resultado del modelo para permitir analizar la evolución del tamaño del software durante el proceso de desarrollo y desde una perspectiva funcional.

En esta tesis se desarrolla **aFPA**, una metodología que permite realizar el conteo automático del tamaño funcional basado en los métodos de medición estandarizados IFPUG FPA y COSMIC FFP. La metodología **aFPA** toma como entrada el código fuente de la aplicación y construye un modelo funcional cuyos componentes son mapeados a los métodos

de medición IFPUG FPA y COSMIC FFP para producir la estimación del tamaño, manteniendo así una trazabilidad entre los artefactos y los resultados del conteo. Adicionalmente, **aFPA** es extendida para construir el modelo funcional a partir de los requerimientos funcionales bajo el estándar IEEE Std. 830-1998 y los modelos conceptuales generados por una herramienta de desarrollo dirigido por modelos. Como resultado de esta investigación se presenta una metodología que puede apoyar a los administradores de proyectos e ingenieros de software en sus procesos de medición del tamaño funcional del software mediante una metodología automatizada que pueda reducir el tiempo y costo del proceso de conteo y que provea resultados de medición con la exactitud y consistencia requerida por la industria.

El objetivo de esta investigación es *desarrollar una metodología para la automatización de la medición del tamaño funcional del software*.

La metodología sigue la estructura de componentes del modelo de procesos para métodos de medición del tamaño funcional (Jacquet & Abran, 1997; Habra, Abran, Lopez & Sellami, 2008; Abran, 2010), y se compone de tres procedimientos: el primero permite realizar la medición automática del tamaño funcional, el segundo verifica la exactitud de las mediciones obtenidas, y el tercero evalúa algunos modelos de estimación de esfuerzo para mostrar un posible aprovechamiento de los resultados de la medición del tamaño funcional. Para probar la utilidad de la metodología, se implementa una herramienta prototipo para cada uno de estos tres procedimientos y se analizan los resultados de su utilización. La Figura 1-1 detalla la estructura de la metodología para la automatización de la medición del tamaño funcional del software y los procedimientos que la componen. El procedimiento de medición recibe como entrada los artefactos de software de código fuente y obtiene el tamaño funcional, el procedimiento de verificación recibe como entrada los resultados del tamaño funcional y realiza la verificación de la exactitud, y el procedimiento de evaluación recibe como entrada las mediciones del tamaño funcional de proyectos y evalúa como estos pueden ser aprovechados para la construcción de modelos de estimación de esfuerzo.

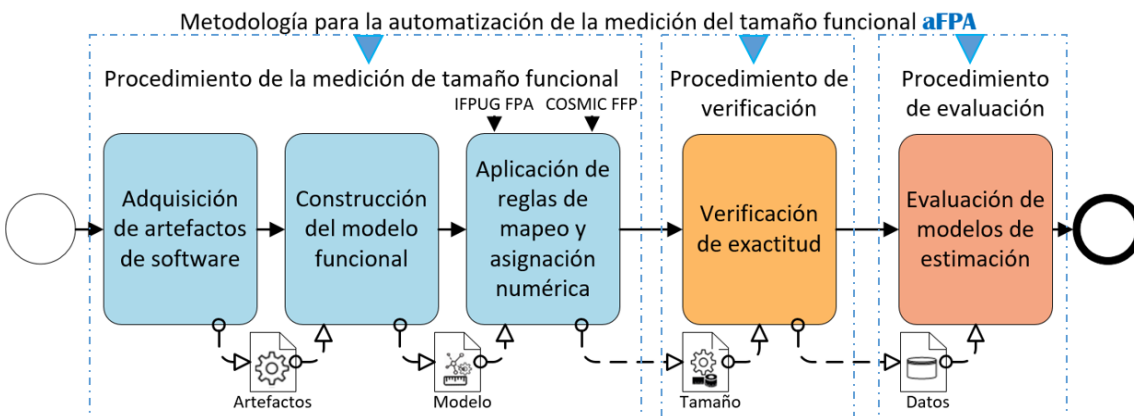


Figura 1-1 Metodología para la automatización de la medición del tamaño funcional del software **aFPA**.

Para lograr el objetivo general se plantean los siguientes objetivos específicos:

- OE1. Determinar las características, desafíos y oportunidades de los procedimientos existentes para la automatización de la medición del tamaño funcional del software.
- OE2. Diseñar una metodología para la automatización de la medición del tamaño funcional del software para modelos funcionales obtenidos a partir del análisis dinámico del código fuente.
- OE3. Evaluar la efectividad de la metodología diseñada para la automatización de la medición del tamaño funcional del software mediante estudios empíricos.

Este trabajo de investigación se contextualiza en el área de investigación de la medición de software, específicamente de la medición del tamaño funcional del software. El procedimiento se desarrolla para el dominio funcional de las aplicaciones transaccionales, también conocido como sistemas de información administrativos (MIS, por sus siglas en inglés) y se aplica sobre modelos funcionales obtenidos a partir del análisis dinámico de artefactos de software del código fuente. El procedimiento se basa en los principios de los métodos de medición IFPUG FPA y COSMIC FFP. Finalmente, todos los estudios realizados

se basan en las metodologías de la ingeniería del software empírica (ESE) y sus protocolos de experimentación.

1.3 METODOLOGÍA DE INVESTIGACIÓN

Para el desarrollo del proyecto de tesis se siguieron los lineamientos del marco de trabajo *Ciencia del Diseño* para Sistemas de Información e Ingeniería del Software (Wieringa, 2014) y las recomendaciones establecidas en los trabajos de (Hevner, March, Park & Ram, 2004; Peffers, Tuunanen, Rothenberger & Chatterjee, 2007). Este marco de trabajo permite el *diseño e investigación de artefactos en contexto* el cual se estructura en un conjunto de ciclos regulativos (de diseño y de investigación) para lograr la solución al problema. En nuestro proyecto, se diseña la metodología para la automatización de la medición del tamaño funcional (**aFPA**), la cual es el artefacto que interactúa con el contexto del problema y permite realizar mejoras en ese contexto. En este caso, **aFPA** se investiga en un contexto para la generación de conocimiento y el descubrimiento de nuevos retos de diseño.

Tal como se muestra en la Figura 1-2, el diseño de la investigación se compone de un conjunto de ciclos ingenieriles o de diseño (EC) y de investigación o empíricos (RC). Las actividades realizadas para cada uno de los ciclos permiten la definición de los problemas de diseño (DP) relacionados con el diseño de la metodología automatizada de medición y las preguntas del conocimiento (KQ) relacionadas con el conocimiento generado producto de la interacción entre la metodología y su contexto de aplicación. El marco de trabajo junto con las DP y las KQ enmarcan el objetivo y las preguntas de investigación de este trabajo.



Figura 1-2 Proyecto de Ciencia del Diseño para la automatización de la medición del tamaño funcional del software.

La metodología de automatización de la medición del tamaño funcional **aFPA** es el resultado de la conducción del proyecto de investigación bajo el marco de trabajo Ciencia del Diseño. El proyecto de investigación se realiza mediante la conducción de un ciclo ingenieril principal (EC1) que conduce el proceso para la investigación del problema (F1), el diseño e implementación de la solución (F2) y la evaluación de la implementación (F3), tal como se muestra en la Figura 1-3. Para cada una de las fases del ciclo ingenieril principal (EC1) se llevan a cabo múltiples ciclos de investigación (RC) anidados que permiten responder las preguntas de investigación para lograr el objetivo de investigación y múltiples ciclos ingenieriles (EC) anidados que permiten la solución de problemas de diseño asociados al problema general.

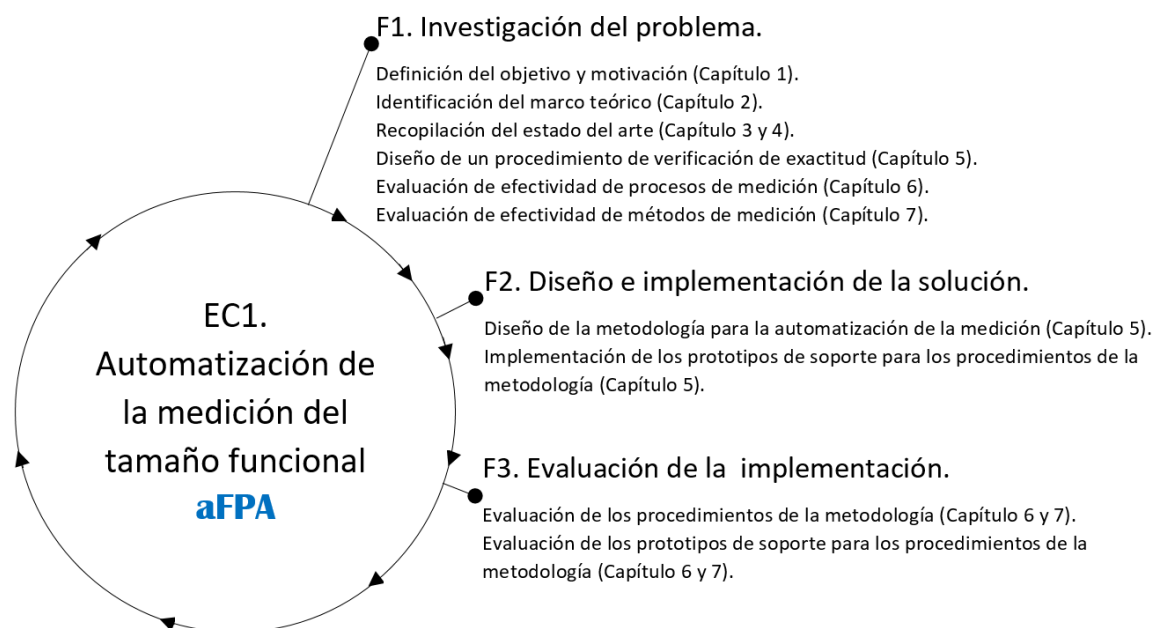


Figura 1-3 Metodología de investigación.

Las preguntas de investigación dirigen el proceso de investigación. De acuerdo al marco de trabajo Ciencia del Diseño, estas pueden clasificarse en preguntas del problema del diseño (DP) y preguntas del problema del conocimiento (KP). Por tanto, para cumplir el objetivo de investigación, se estudian las siguientes preguntas de investigación:

- RQ1. KP1: *¿Cuáles son las características, desafíos y oportunidades de los procedimientos existentes para la automatización de la medición del tamaño funcional del software?*
- RQ2. DP2: *¿Cómo diseñar una metodología para la automatización de la medición del tamaño funcional del software?*
- RQ3. KP3: *¿Cuál es la efectividad de la metodología para la automatización de la medición del tamaño funcional del software?*

La pregunta de investigación 1 (KP1) estudia los desafíos, oportunidades y características de las propuestas para la medición del tamaño funcional, identifica los requerimientos para la automatización del proceso de la medición del tamaño funcional, determina los factores de contexto donde estas propuestas son aplicables y analiza la evidencia empírica

existente sobre la efectividad de las propuestas para la automatización de la medición del tamaño funcional. Asimismo, analiza la relación entre los componentes funcionales de los métodos de la medición del tamaño funcional y los principales componentes funcionales que influyen la efectividad de los modelos de estimación de esfuerzo en proyectos de desarrollo de software. Finalmente, estudia las métricas cuantitativas y cualitativas y las pruebas que son utilizadas para la validación de la eficacia de los procedimientos y la evaluación de los modelos de estimación.

La pregunta de investigación 2 (DP2) identifica los lineamientos de diseño necesarios para guiar el desarrollo de la metodología para la automatización de la medición, determina los lineamientos metodológicos necesarios para facilitar la derivación de los modelos funcionales necesarios para el conteo de tamaño funcional y establece las reglas de mapeo y asignación numérica para la medición del tamaño funcional del software. Finalmente, determina los requerimientos y describe los lineamientos de diseño e implementación que debe contener una herramienta para la automatización de la medición del tamaño funcional, incluyendo la verificación de la calidad de los resultados y el aprovechamiento de los resultados de las mediciones del tamaño funcional.

Finalmente, la pregunta de investigación 3 (KP3) estudia la efectividad de la metodología para la automatización de la medición del tamaño funcional que incluye la verificación de la exactitud de los resultados de medición y la evaluación de los modelos de estimación para el aprovechamiento de los resultados de medición.

La Figura 1-4 detalla cada una de las fases y actividades de la metodología de investigación en orden secuencial. Cada fase y actividad se mapea con el objetivo específico y pregunta de investigación respectiva detallando el método de investigación utilizado y los productos esperados por cada una de las fases. La metodología se compone de tres fases principales, cada una de ellas con un subconjunto de ciclos ingenieriles y de investigación.

La primera fase (F1) permite la consecución del objetivo específico 1 (OE1) y obtener la respuesta a la pregunta de investigación 1 (KP1) mediante la investigación del problema. Para recopilar el estado del arte se lleva a cabo un ciclo de investigación (RC2) mediante el cual se realiza un *survey* exploratorio, un mapeo y una revisión sistemática de literatura.

Los métodos empíricos utilizados son el de mapeo (Petersen, Feldt, Mujtaba & Mattsson, 2008; Petersen, Vakkalanka & Kuzniarz, 2015) y el de revisión sistemática de literatura en la ingeniería del software (Biolchini, Mian, Natali & Travassos, 2005; Budgen & Brereton, 2006; Kitchenham, 2004; Kitchenham & Charters, 2007; Kitchenham, Budgen & Brereton, 2015). Esta actividad determina sistemáticamente los desafíos, oportunidades y características de los enfoques existentes para la automatización de la medición del tamaño funcional del software mediante una revisión de la literatura.

Para evaluar la efectividad de los métodos de medición IFPUG FPA y COSMIC FFP, y sus componentes funcionales básicos (BFC, por sus siglas en inglés) para la construcción de modelos de estimación de esfuerzo se lleva a cabo un ciclo de investigación (RC3). Para esto se realiza una familia de experimentos controlados y un conjunto de replicaciones de estudios previos en el área. Los métodos empíricos utilizados son el protocolo de experimentación (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012) y de replicación de estudios en la ingeniería del software (Shull, Carver, Vegas & Juristo, 2008; Juristo & Vegas, 2009; Carver, 2010; Carver, Juristo, Baldassarre & Vegas, 2014). Esta actividad evalúa empíricamente la efectividad de los métodos de medición del tamaño funcional para la estimación del esfuerzo del desarrollo del software y determina los principales componentes funcionales que deben extraer los procesos automatizados de medición para la creación de los modelos de estimación.

Para determinar la exactitud de la medición del tamaño funcional resultado de la medición de las aplicaciones se lleva a cabo un ciclo ingenieril (EC4) que permite diseñar el procedimiento de verificación de exactitud. El método utilizado es el modelo de procesos para métodos de medición del tamaño funcional (Jacquet & Abran, 1997; Habra, Abran, Lopez & Sellami, 2008; Abran, 2010). Esta actividad define el protocolo de verificación de exactitud que se aplica a los resultados de medición para determinar su exactitud.

Para evaluar la efectividad y determinar oportunidades de mejora de los procesos de medición del tamaño funcional existentes en la literatura se lleva a cabo un ciclo de investigación (RC5) que utiliza el protocolo de verificación diseñado. Para esto se lleva a

cabo un cuasi-experimento y una familia de replicaciones que utiliza el protocolo de experimentación en la ingeniería del software descrito en (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012). Esta actividad permite evaluar la efectividad de procesos de medición del tamaño funcional existentes mediante experimentos controlados. En este caso, se fijan las aplicaciones bajo medición y un conjunto de profesionales aplica distintos procedimientos de medición a partir de un conjunto de artefactos de software específicos. Posteriormente, se analizan las diferencias en los resultados de medición contra los valores reales para determinar oportunidades de mejora y las principales diferencias obtenidas. En este caso, los valores reales tomados como valores de control son obtenidos por expertos en medición. Este ciclo de investigación permite generar conocimiento sobre los procesos de medición y generar datos que luego pueden ser comparados con los obtenidos por la metodología de medición diseñada en la segunda fase del proyecto de investigación.

La segunda fase (F2) permite la consecución del objetivo específico 2 (OE2) y obtener la respuesta a la pregunta de investigación 2 (DP2) mediante el diseño e implementación de la solución que permite desarrollar la metodología para la automatización de la medición del tamaño funcional **aFPA**. Para construir cada uno de los tres procedimientos de la metodología (medición, verificación y evaluación) se lleva a cabo un ciclo ingenieril (EC6) el cual sigue el modelo de procesos para métodos de medición del tamaño funcional propuesto por (Abran, 2010; Habra, Abran, Lopez & Sellami, 2008; Jacquet & Abran, 1997). Esta actividad permite diseñar el procedimiento para la automatización de la medición, verificación y evaluación del tamaño funcional que incluye la definición de los objetivos de medición, el artefacto de entrada, los metamodelos, las reglas de mapeo, las reglas de asignación numérica, el proceso de adquisición de artefactos, las reglas de construcción del modelo funcional, las reglas de aplicación del procedimiento, el reporte de presentación de las mediciones, la verificación y auditoría de resultados y finalmente, la evaluación para el aprovechamiento de los resultados del tamaño funcional.

Para diseñar los prototipos de soporte de cada uno de los procedimientos se llevan a cabo múltiples ciclos ingenieriles (EC7). Para estos se sigue la metodología de desarrollo de software iterativa incremental (Pressman, 2005) y las recomendaciones del desarrollo

basado en las metodologías ágiles. En este ciclo se llevan a cabo todas las fases del ciclo de vida de desarrollo de software para construir las herramientas prototipo que apoyan el proceso automatizado de conteo del tamaño funcional, su verificación y su aprovechamiento de resultados. Esta actividad permite la implementación de los prototipos funcionales que soportan el proceso de medición y verificación de resultados de exactitud. Del mismo modo, se implementa el prototipo para la construcción y evaluación de los modelos de estimación de esfuerzo simplificados que permite el aprovechamiento de los resultados de medición.

La tercera fase (F3) permite la consecución del objetivo específico 3 (OE3) y obtener la respuesta a la pregunta de investigación 3 (KP3) mediante la evaluación de la metodología para la automatización de la medición del tamaño funcional **aFPA** y cada uno de los procedimientos que la componen. Para evaluar el procedimiento de medición y el prototipo de soporte se lleva a cabo un ciclo de investigación (RC8) mediante el cual se realizan múltiples casos de estudio. El método empírico utilizado es el protocolo de casos de estudio (Runeson & Höst, 2009; Runeson, Host, Rainer & Regnell, 2012) en la ingeniería del software. Esta actividad permite evaluar empíricamente la efectividad del procedimiento para la automatización de la medición del tamaño funcional mediante la ejecución de los casos de estudio en contexto. Cada caso de estudio selecciona aplicaciones específicas bajo medición para realizar la evaluación empírica en términos de la exactitud de las mediciones.

Del mismo modo, para evaluar la efectividad del procedimiento de medición del tamaño funcional y el procedimiento de verificación con profesionales se lleva a cabo un ciclo de investigación (RC9) mediante el cual se realiza un cuasi-experimento y tres repeticiones subsecuentes. El método empírico utilizado es el protocolo de experimentación en la ingeniería del software (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012). Esta actividad permite evaluar empíricamente la efectividad del procedimiento para la automatización de la medición del tamaño funcional mediante la ejecución de experimentos controlados. Además, permite comparar la efectividad de los resultados del procedimiento implementado con los resultados de medición de propuestas existentes para distintas aplica-

ciones bajo medición. Asimismo, permite evaluar el procedimiento de verificación de exactitud y su herramienta prototipo para determinar su efectividad para identificar errores de medición.

Finalmente, para evaluar la efectividad de los resultados de medición del tamaño funcional en la construcción de modelos de estimación de esfuerzo simplificados se lleva a cabo un ciclo de investigación (RC10) que utiliza el procedimiento de evaluación. El método empírico utilizado es el protocolo de experimentación (Carver, 2010; Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012; Carver, Juristo, Baldassarre & Vegas, 2014). Para esto se realiza una replicación de estudios previos y dos evaluaciones empíricas subsecuentes que comparan distintos modelos de estimación simplificados. Esta actividad evalúa mediante un conjunto de repeticiones la efectividad de los modelos y analiza los principales componentes funcionales que deben extraer los procedimientos automatizados de medición.

La descripción del marco de trabajo ciencia del diseño y las metodologías empíricas utilizadas se encuentran detallados en el Capítulo 2 de este documento de tesis. El diseño de cada uno de los ciclos ingenieriles y de investigación se detalla en cada uno de los capítulos correspondientes y finalmente, el diseño de cada uno de los estudios empíricos de los ciclos de investigación se detalla en el Capítulo 6 y el Capítulo 7.

1.4 CONTRIBUCIONES

Las contribuciones realizadas con este trabajo de investigación son las siguientes:

- Un mapeo y revisión sistemática de literatura que caracteriza los procedimientos de la medición del tamaño funcional existentes y presenta el estado del arte en el área de estudio.
- Una evaluación empírica de la estructura de los componentes funcionales básicos de los métodos de medición del tamaño funcional IFPUG FPA y COSMIC FFP para determinar

su impacto en la efectividad de los modelos de estimación del esfuerzo del desarrollo del software.

- Una evaluación empírica de la aplicabilidad de los componentes funcionales básicos de los métodos de medición del tamaño funcional IFPUG FPA y COSMIC FFP para la construcción de modelos simplificados de estimación.
- Una evaluación empírica sobre la efectividad de propuestas de automatización de la medición del tamaño funcional del software a partir de diferentes artefactos de software generados durante el ciclo de vida del desarrollo del software.
- Un protocolo sistemático para la verificación de la exactitud de los resultados de la medición del tamaño funcional que evalúa los resultados totales del conteo y de cada uno de los componentes funcionales básicos de un método de medición del tamaño funcional.
- Un proceso para obtener modelos funcionales de representación del software que permita realizar el proceso de conteo del tamaño funcional, mantener trazabilidad entre los requerimientos del usuario final y el código fuente de la aplicación y visualizar la funcionalidad del software desde la perspectiva del usuario final.
- Una metodología para la automatización de la medición del tamaño funcional que permite mediante un proceso sistematizado obtener el tamaño funcional basado en los métodos IFPUG FPA y COSMIC FFP, verificar los resultados de medición y evaluar algunos modelos de estimación de esfuerzo para mostrar un posible aprovechamiento de los resultados de la medición del tamaño funcional.
- Herramientas prototipo de soporte para la medición del tamaño funcional, la verificación de la exactitud y la evaluación de modelos de estimación de esfuerzo validadas empíricamente.
- Una evaluación empírica de la efectividad del procedimiento de medición de acuerdo con los métodos de medición IFPUG FPA y COSMIC FFP.
- Una evaluación empírica sobre la aplicabilidad del tamaño funcional para la construcción de modelos de estimación de esfuerzo utilizando los resultados de medición de los métodos IFPUG FPA y COSMIC FFP.

- Una evaluación empírica de la efectividad del procedimiento de medición de acuerdo con el método de medición IFPUG FPA en un contexto industrial.

1.5 ESTRUCTURA DEL DOCUMENTO DE TESIS

La estructura propuesta para la tesis es la siguiente:

- *Marco teórico*: El Capítulo 2 presenta los conceptos, métodos y teorías en el campo de la medición del tamaño del software y de las metodologías de la ingeniería del software empírica que son relevantes para nuestra investigación y que respaldan los diseños y estudios realizados en esta tesis.
- *Trabajo relacionado y estado del arte*: El Capítulo 3 presenta el trabajo relacionado a nuestra investigación que incluye estudios relevantes para la definición de cada uno de las partes la metodología para la automatización de la medición propuesta en este trabajo. El Capítulo 4 se documenta el mapeo y revisión sistemática de literatura sobre el estado actual de los procedimientos de medición del tamaño funcional del software.
- *Metodología para la automatización de la medición del tamaño funcional*: El Capítulo 5 presenta el diseño e implementación de la metodología para la automatización de la medición del tamaño funcional del software, sus procedimientos y las herramientas prototipo de soporte.
- *Evaluación*: El Capítulo 6 y el Capítulo 7 presenta la evidencia generada a partir de un conjunto de estudios empíricos para evaluar la efectividad de la metodología **aFPA**. En este caso, el procedimiento de medición y verificación propuesto se evalúa mediante múltiples casos de estudio y experimentos controlados cuasi experimentales presentados en el Capítulo 6. El Capítulo 7 presenta los experimentos realizados con el procedimiento de evaluación que estudian la estructura de los métodos de medición IFPUG FPA y COSMIC FFP y su efectividad para la construcción de modelos de estimación de esfuerzo del desarrollo de software para el aprovechamiento de los resultados de tamaño funcional.

- *Conclusiones y trabajo futuro*: El Capítulo 8 presenta las conclusiones y el trabajo futuro. Asimismo, detalla las principales contribuciones de la tesis, lecciones aprendidas durante su realización y las líneas de trabajo futuro.

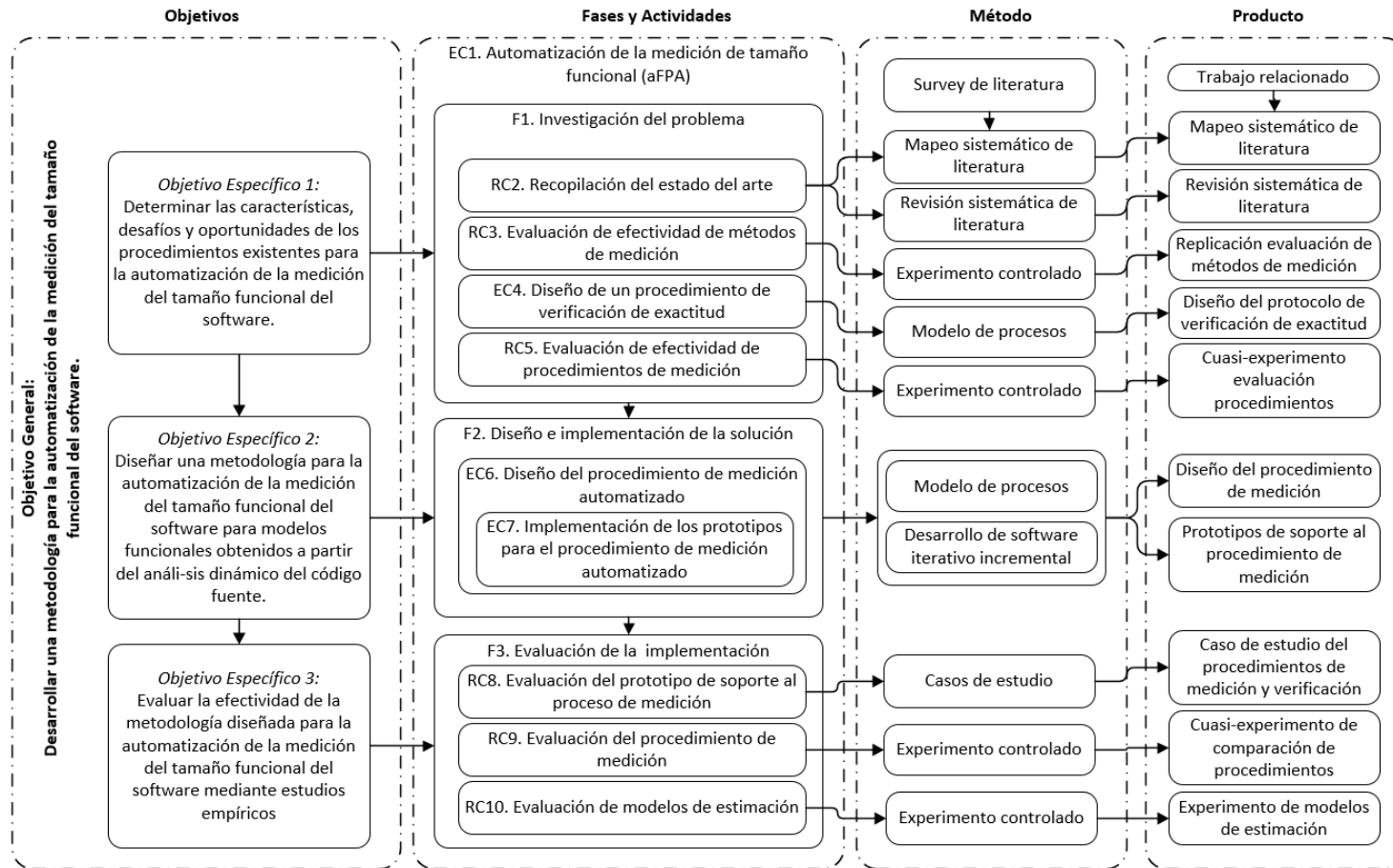


Figura 1-4 Metodología de investigación detallada.

Capítulo 2. MARCO TEÓRICO

Este capítulo presenta los conceptos, métodos y teorías en el campo de la medición del tamaño del software y de las metodologías empíricas utilizadas en la ingeniería de software que son relevantes para nuestra investigación.

2.1 MEDICIÓN DEL SOFTWARE

La medición es fundamental para la disciplina de la ingeniería del software (Abran, 2010). La *IEEE Computer Society* define la ingeniería de software como “La aplicación de un enfoque sistemático, disciplinado y cuantificable para el desarrollo, operación y mantenimiento del software”, esto es, la aplicación de la ingeniería al software. La medición del software es crucial para el control de proyectos, productos y procesos (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012) y en proyectos de software, no es posible controlar lo que no se puede medir (DeMarco, 1986).

El concepto de “medición” se define como “el proceso por el cual números o símbolos se asignan a los atributos de las entidades del mundo real para describirlas de acuerdo a un conjunto de reglas claramente definidas”, una medida es “el número o símbolo asignado a una entidad que caracteriza un atributo” (Fenton & Pfleeger, 1997).

La medición es parte fundamental en los estudios empíricos aplicados a la ingeniería del software. Para controlar los experimentos y describir las relaciones entre los objetos de estudio y las intervenciones, debemos medir las entradas y salidas del proceso y cuantificar el efecto causado. Sin métricas de software, no es posible controlar los experimentos y los estudios empíricos no pueden ser ejecutados (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012).

Las organizaciones de desarrollo de software han adoptado la medición como una práctica cotidiana en sus procesos de desarrollo (McGarry, 2002) y la medición del proceso de desarrollo del software ha permitido a las organizaciones administrar sus necesidades de información (Florac & Carleton, 1999). Para satisfacer estas necesidades de información,

las organizaciones deben determinar el conjunto de métricas que necesitan seleccionar, pero dada la cantidad de métricas que existen en la literatura, la elección de estas no es una tarea simple. Sumado a lo anterior, en la literatura de la ingeniería del software, las propuestas de definición de métricas para el software han proliferado en los últimos años, dando como resultado la baja adopción en la industria y la falta de estandarización de los procesos para validar cada una de las propuestas presentadas (Abran, 2010).

En años recientes, distintos grupos de profesionales y académicos han desarrollado un conjunto de mejores prácticas en medición de software desde diferentes perspectivas. Algunos de estos documentos que recopilan las buenas prácticas de medición son el estándar ISO 15939 (ISO, 2017) para procesos de medición del software y el modelo de madurez CMMI (SEI, 2010). Uno de los retos de las organizaciones de desarrollo de software es la integración de estas propuestas, para que mediante un modelo integrado, puedan dar soporte a sus programas de medición. En este caso, el estándar ISO 15939 define las actividades y tareas que son necesarias para implementar un proceso de medición y el modelo CMMI contiene los elementos esenciales para un proceso de ingeniería de software efectivo, que es orientado por la medición del proceso (Bégnoche, Abran & Buglione, 2007).

2.1.1. MEDICIÓN DEL TAMAÑO DEL SOFTWARE

El software puede ser medido desde dos perspectivas, la perspectiva funcional y la perspectiva técnica (Abrahão, 2004). Las métricas de tamaño funcional soportan la perspectiva del usuario final y se relacionan con lo que una aplicación de software hace por cada usuario y donde estas medidas funcionales se encuentran en el dominio del problema. Por otro lado, desde la perspectiva del desarrollador, un sistema se analiza desde las necesidades específicas para construir la aplicación donde las métricas técnicas representan estas necesidades del dominio del sistema, tal como se muestra en la Figura 2-1. Cuando se mide una aplicación de software, estas dos perspectivas no necesariamente están correlacionadas entre sí (Abrahão, 2004). La medición del software desde la perspectiva funcional

tiene un propósito distinto al de la perspectiva técnica por lo que extraer mediciones funcionales a partir de artefactos técnicos requiere procesos de transformación complejos.

Una de las principales diferencias es que las métricas funcionales se pueden obtener en las etapas iniciales del ciclo de vida del desarrollo del software y las métricas técnicas se obtienen a partir de artefactos de software que se construyen en etapas avanzadas del desarrollo. Las métricas técnicas cuantifican el producto y el proceso del software desde la perspectiva del desarrollador, mientras que las métricas funcionales lo hacen desde la perspectiva del usuario final (Meli, 2000).

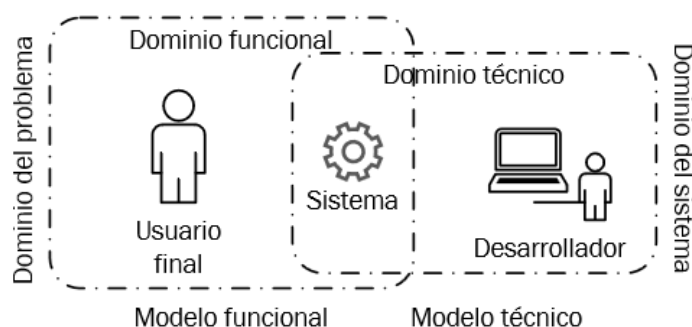


Figura 2-1 Dominios de medición del software. Adaptado de (Abrahão, 2004).

Las métricas técnicas miden las características físicas del software. Por ejemplo, a partir del código fuente podemos obtener las líneas de código, número de módulos, reportes, pantallas, métricas orientadas a objetos como clases y objetos y métricas de complejidad algorítmica, entre otras (Meli, 2000). Las principales críticas para las métricas técnicas son la dependencia con la tecnología y que estas se obtienen en etapas avanzadas del proceso de desarrollo. Además, dada la gran cantidad de lenguajes de programación, se dificulta la comparación de datos históricos de los proyectos (Jones, 2008). Finalmente, con el arribo de nuevas generaciones de lenguajes de programación y estilos de codificación, se presentan variaciones en la cantidad de líneas de código, clases y métodos necesarios para programar una funcionalidad específica del usuario (Abrahão, 2004). Por su lado, Albrecht (1979) indica que el concepto de líneas de código tiene un significado distinto dependiendo del lenguaje de programación.

Las métricas funcionales miden el software desde la perspectiva de las necesidades usuario final. Una de las principales ventajas de estas métricas, es que son independientes de la tecnología y de la implementación, y permiten medir el software en etapas tempranas de desarrollo (Abrahão, 2004). Las mediciones de puntos de función son métricas normalizadas que se han sido usadas con algún nivel de consistencia y un grado de exactitud aceptable en ambientes profesionales. En particular, las mediciones de puntos de función son independientes de la tecnología, no dependen de la plataforma de desarrollo o del lenguaje de programación y miden la funcionalidad entregada desde la perspectiva del usuario final. Las principales críticas a los métodos de medición del tamaño funcional son el esfuerzo requerido para realizar el conteo de una aplicación de software y las inconsistencias en los resultados que se pueden presentar entre el conteo realizado por distintos expertos debido a la interpretación de las reglas del método. Además, los resultados de las mediciones de tamaño funcional realizadas con diferentes métodos pueden variar significativamente y existen problemas con la adherencia de algunas de las definiciones de estos métodos con las teorías de medición (Runeson, Borgquist, Landin & Bolanowski, 2000).

2.1.2. MEDICIÓN DEL TAMAÑO FUNCIONAL DEL SOFTWARE

En la actualidad existen diferentes estándares y marcos de trabajo que definen los conceptos relacionados con la medición del tamaño funcional (FSM, por sus siglas en inglés). El marco de trabajo más importante para la medición del tamaño funcional es la serie de estándares ISO/IEC 14143 (ISO, 2007). Este provee los lineamientos para desarrollar, probar y refinar nuevos métodos de medición del tamaño funcional y se compone de seis partes principales, tal como se muestra en la Figura 2-2:

- Parte 1: Definición de conceptos, características y requerimientos necesarios para un método de medición del tamaño funcional (FSM).
- Parte 2: Evaluación de conformidad de un método de medición FSM con el estándar. Describe el proceso para la evaluación de la conformidad de un candidato a método de medición del tamaño funcional en relación con la Parte 1.

- Parte 3: Verificación del método de medición FSM. Establece un procedimiento para verificar los conceptos del método y para conducir pruebas que verifiquen las características de eficiencia: repetibilidad, reproductibilidad, exactitud y consistencia, convertibilidad y aplicabilidad en los dominios funcionales.
- Parte 4: Definición del modelo de referencia del método de medición FSM que presenta las entradas para la validación del método FSM.
- Parte 5: Determinación de los dominios funcionales para el uso del método de medición FSM. Describe las características de cómo los artefactos son utilizados de acuerdo al método más apropiado dependiendo del dominio funcional, sistemas de información gerencial (MIS), sistemas en tiempo real (RTS) y software científico (SS).
- Parte 6: Definición de una guía para el uso de la serie ISO/IEC 14143 en relación con otros estándares.

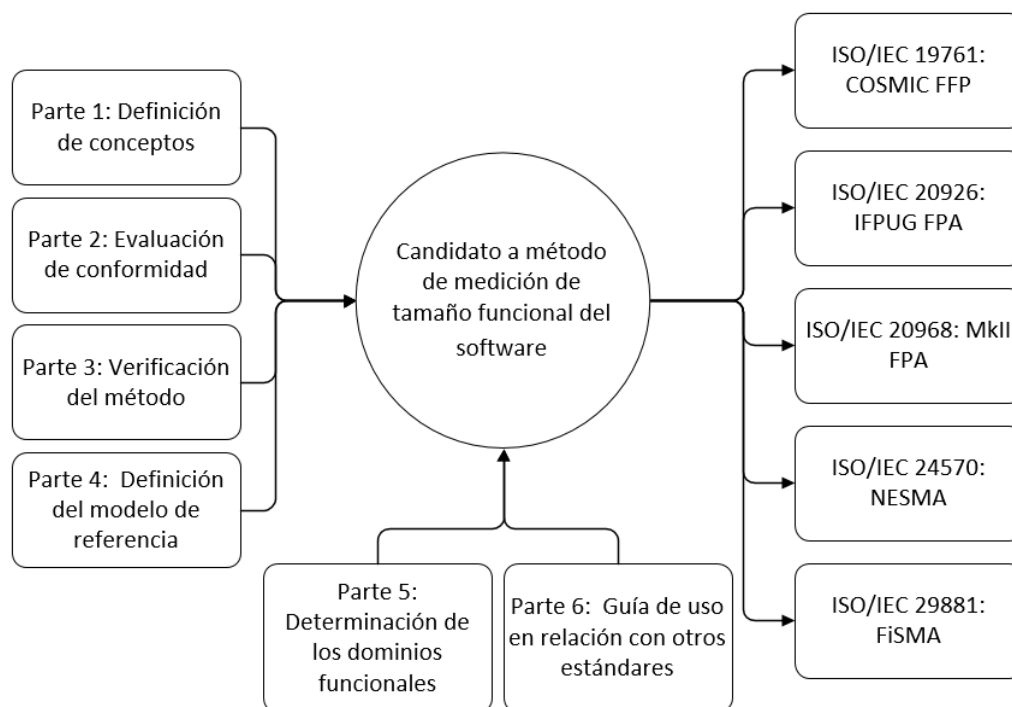


Figura 2-2 Estándares de medición del tamaño funcional. Adaptado de (ISO, 2007).

Este estándar permite la definición y aplicación de los métodos de medición del tamaño funcional. A partir de su aparición, se han estandarizado diferentes métodos FSM

para cuantificar el tamaño funcional del software: ISO/IEC 19761 (COSMIC FFP), ISO/IEC 20926 (IFPUG FPA), ISO/IEC 20968 (MkII), ISO/IEC 24570 (NESMA), y ISO/IEC 29881 (FiSMA). De los anteriores, el método IFPUG FPA es el más utilizado históricamente en la industria (Fingerman, 2011; Hill., 2010) y el método COSMIC FFP es el de mayor crecimiento en adopción en los últimos años (Dumke & Abran, 2011), esto en gran medida porque toda la documentación del método COSMIC FFP es de acceso libre para cualquier profesional.

Para el desarrollo del trabajo de investigación, dos métodos de medición son de interés. El primero es el método IFPUG FPA por ser el más utilizado en la industria y el segundo es el método COSMIC FFP por ser considerado la segunda generación de puntos de función y por ser el mayor crecimiento en adopción en los últimos años en la industria. Además, es de interés para nuestro estudio la especificación llamada *Automated Function Points* (AFP) porque propone una metodología para la medición automatizada del tamaño funcional a partir del código fuente de aplicaciones de software (Object Management Group, 2014). En las siguientes subsecciones se describen las características generales de cada uno de ellos.

Es importante mencionar que desde el año 1979 se ha estudiado el concepto del tamaño funcional iniciando con la propuesta de Albrecht (1979). Posterior a esta propuesta, múltiples métodos de medición del tamaño funcional se han desarrollado y han ido evolucionado, algunos de los cuales fueron estandarizados bajo los lineamientos de la serie de estándares ISO/IEC 14143, tal como se muestra en la Figura 2-3.

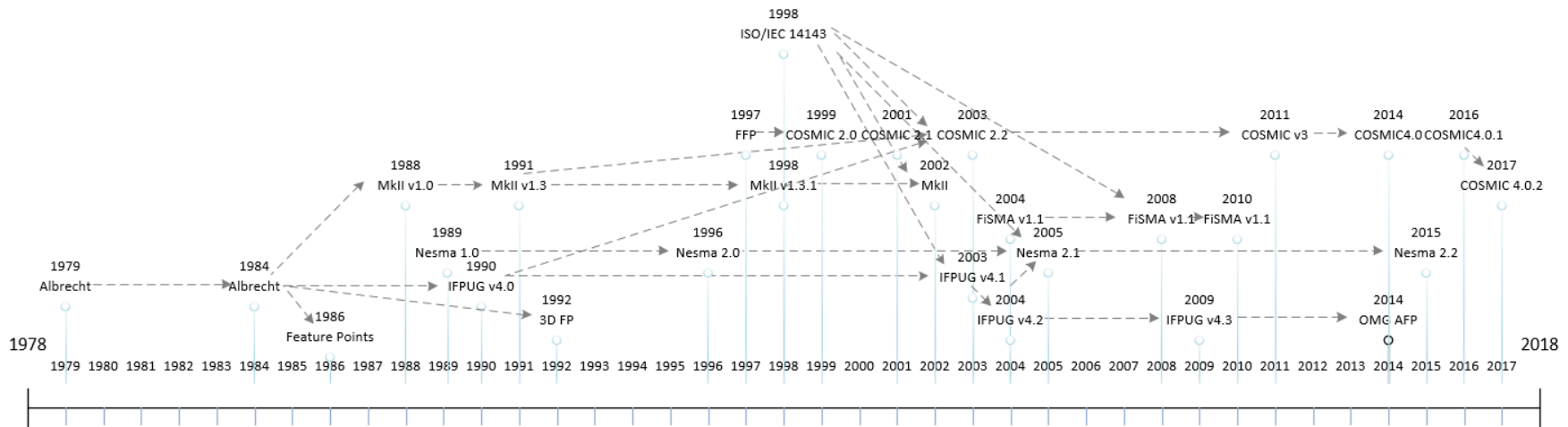


Figura 2-3 Evolución de los métodos de medición del tamaño funcional. Actualizado a partir de (Condori-Fernández, 2006; Abran, 2015).

2.1.2.1. IFPUG FPA

La medición de puntos de función IFPUG FPA (ISO, 2009) es una metodología utilizada para medir el tamaño funcional del software, la cual fue creada originalmente por Allan Albrecht en 1979 (Albrecht, 1979) y que actualmente es mantenida por el *International Function Point Users Group*. Este método pretende medir la funcionalidad del software entregada al usuario independientemente de la tecnología utilizada y su implementación. El conteo de puntos de función es una de las principales métricas de tamaño funcional en la industria y se estima que en los próximos años podría convertirse en una métrica de uso universal. Actualmente, gobiernos como Brasil exigen su utilización en sus contratos y otros como Corea del Sur e Italia podrían seguir esta misma tendencia (Jones, 2013).

Una aplicación de software representada mediante el método de medición IFPUG FPA se compone de cinco tipos de componentes funcionales básicos (BFC), tal como se muestra en la Figura 2-4 y los cuales se detallan a continuación:

- Funciones de datos (DF) los cuales se clasifican en:
 - Archivos lógicos internos: cualquier archivo o tabla interna de base de datos (*Internal logic files* [ILF]).
 - Interfaces externas: cualquier acceso de lectura a tablas o archivos de otros sistemas (*External interface files* [EIF]).
- Funciones transaccionales (TF) los cuales se clasifican en:
 - Entradas externas: cualquier entrada por pantallas que provea datos a la aplicación (*External Inputs* [EI]).
 - Salidas externas: cualquier salida en pantalla o reporte (*External Outputs* [EO]).
 - Consultas externas: cualquier consulta que genere una salida de información (*External Inquiries* [EQ]).

El tamaño funcional se determina calculando la complejidad de cada uno de los componentes funcionales básicos (BFC) a partir de los elementos de tipo dato (DET), elementos de tipo registro (RET) y tipos de archivo referenciado (FTR). La suma de estos valores corresponde al tamaño funcional sin ajustar (UFP). Los límites de la aplicación separan el software que se mide de su ambiente de operación (usuarios y otros sistemas). Los pasos para la aplicación de la técnica de puntos de función son los siguientes (Garmus & Herron, 2001):

1. Determinar el tipo de conteo y reunir la documentación disponible.
2. Identificar el alcance del conteo y los límites de la aplicación e identificar los requerimientos funcionales de los usuarios.
3. Identificar y contar las funciones de datos.
4. Identificar y contar las funciones transaccionales.
5. Realizar el conteo bruto de puntos función y calcular el tamaño funcional sin ajustar.
6. Determinar el factor de ajuste de acuerdo a las características generales del sistema.
7. Calcular el conteo del tamaño funcional ajustado.

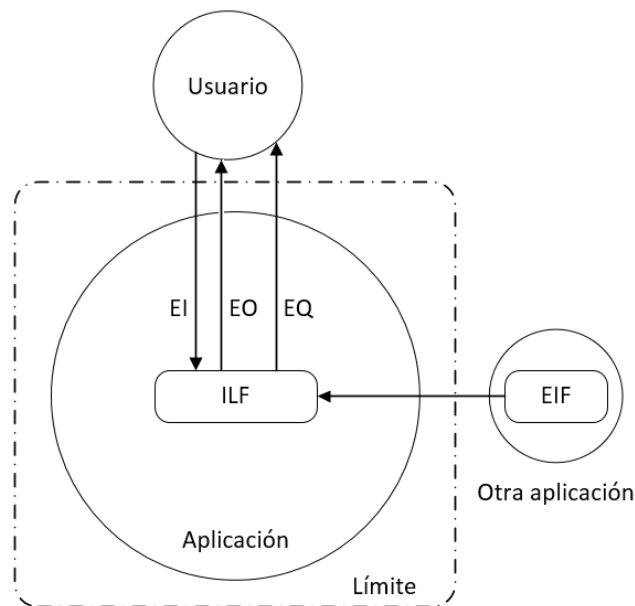


Figura 2-4 Componentes del método IFPUG FPA. Adaptado de (Runeson, Borgquist, Landin & Bolanowski, 2000).

2.1.2.2. COSMIC FFP

La medición de puntos de función COSMIC FFP (ISO, 2011) es una metodología utilizada para medir el tamaño funcional del software creada originalmente por St-Pierre en 1997 (St-Pierre, Maya, Abran, Desharnais & Bourque, 1997). Esta metodología pretende medir la funcionalidad del software identificando los movimientos de datos que ocurren entre los usuarios funcionales y los procesos funcionales. Una aplicación de software representada mediante el método de medición COSMIC FFP se compone de cuatro tipos de movimientos de datos (DM), tal como se muestra en la Figura 2-5 y los cuales se detallan a continuación:

- Movimientos de datos (DM) los cuales se clasifican en:
 - Entradas (E): entrada de datos a un proceso funcional.
 - Salidas (X): salida de datos desde un proceso funcional.
 - Lecturas (R): lectura de datos desde el almacenamiento persistente.

- Escrituras (W): escritura de datos en el almacenamiento persistente.

El tamaño funcional se determina calculando la cantidad de cada uno de los componentes funcionales básicos y la suma de estos valores corresponde al tamaño funcional. Los movimientos de datos se dan entre los usuarios funcionales y los procesos funcionales o su interacción con los grupos de datos.

Los pasos para la aplicación de la técnica de puntos de función son los siguientes:

1. Determinar la estrategia de medición.
2. Mapear los requerimientos del usuario (FUR) al modelo del software,
3. Medir todos los componentes de la fase anterior.

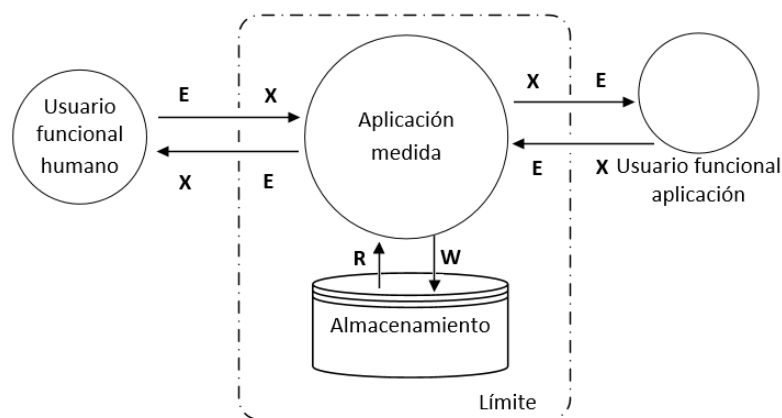


Figura 2-5 Componentes del método COSMIC FFP. Adaptado de (ISO, 2011).

2.1.2.3. OMG AFP

En el 2014, el *Object Management Group* (OMG) adopta la especificación *Automated Function Point* (AFP) la cual provee una guía para la automatización del conteo de puntos de función de acuerdo al método IFPUG FPA (Object Management Group, 2014). De acuerdo a OMG, AFP asegura la repetitividad y consistencia de la técnica de conteo y per-

mite medir software en desarrollo y en mantenimiento. También advierte que la especificación AFP es generalmente consistente con el IFPUG FPA versión 4.3, pero que difiere en los puntos en que la subjetividad debe usarse para realizar el conteo. Por tanto, esta especificación reemplaza algunas reglas para lograr la automatización. Se da prioridad a la repetitividad y consistencia sobre la consistencia con el método IFPUG FPA. AFP utiliza los estándares *Knowledge Discovery Meta-Model* (KDM) y el *Structure Metrics Meta-Model* (SMM) para especificar un modelo que se extrae mediante el análisis estático del código fuente del programa y que contiene los elementos básicos del conteo para calcular los puntos de función garantizando automatización, consistencia y verificabilidad.

Las entradas necesarias para utilizar esta especificación y aplicar el proceso de conteo son las siguientes:

- Entrada 1: Código fuente completo de la aplicación:
 - Código fuente de la aplicación (archivos con las unidades de interfaz, clases de la capa de negocios y scripts de base de datos).
 - Lista de archivos excluidos del código fuente.
 - Lista de librerías excluidas del código fuente.
 - Archivos de definición de datos (DDL).
 - Archivos planos mantenidos por la aplicación (CSV, XML, JSON, otros).
- Entrada 2: Lista de convención de nombres para la aplicación:
 - Bases de datos.
 - Archivos Planos.
 - Código fuente.
 - Patrones de la aplicación identificados como transacciones.

Además es requerido determinar para cada uno de los elementos del código fuente los límites de la aplicación y se deben determinar los siguientes aspectos para asegurar que el conteo refleje el tamaño del dominio de la aplicación que está siendo medida:

- Definir claramente los elementos dentro de los límites de la aplicación en el nivel de interfaces de usuario, servicios web y procesamiento de archivos por lotes, si aplica.
- Si se incluye código fuente externo a la aplicación debe ser marcado claramente como tal.

Para aplicar el método AFP se deben ejecutar los siguientes pasos y producir las siguientes salidas:

1. Obtener acceso al código fuente y determinar los límites de la aplicación.
2. Crear el modelo de la aplicación (KDM).
3. Determinar las funciones de datos.
4. Determinar las funciones transaccionales.
5. Determinar y distinguir entre archivos lógicos internos y externos.
6. Calcular el número total de puntos de función en la aplicación.
7. Generar las salidas detalladas del conteo.
8. Totalizar el número de funciones de datos de la aplicación.
9. Totalizar el número de funciones transaccionales de la aplicación.
10. Determinar el mapeo entre funciones transaccionales y de datos.
11. Totalizar el número de puntos de función de la aplicación.

El método AFP necesita para su primer paso la intervención manual de un experto en la aplicación a ser medida y todo el proceso de cálculo adicional debe implementarse de manera automática. Este mide cuatro componentes funcionales básicos del método IFPUG FPA: archivos lógicos internos (ILF), interfaces externas (EIF), entradas (EI) y salidas (EO).

2.1.3. FUNDAMENTOS DE LA MEDICIÓN DEL TAMAÑO FUNCIONAL DEL SOFTWARE

De acuerdo al estándar de términos de metrología (ISO, 2004) y basado en los fundamentos de las teorías de medición de software (Abran, 2010), la fase de diseño de un proceso de medición se divide en tres niveles secuenciales, tal como se muestra en la Figura

2-6. Primero, se realiza la definición del principio de medición¹ que incluye el conocimiento y entendimiento sobre el concepto² a ser medido y detalla una descripción precisa de lo que mide. Segundo, se define el método de medición basado en el principio de medición que da una descripción general del proceso. Tercero, se define el procedimiento de medición con la definición operacional del procedimiento que representa una implementación del método de medición en un contexto particular y presenta una descripción específica de cómo medir el software. El proceso de medición puede realizarse completamente automatizado, parcialmente automatizado o completamente manual. Un instrumento o herramienta de medición se construye para automatizar parcial o totalmente un procedimiento de medición (Quesada-López & Jenkins, 2017).

De acuerdo al estándar de términos de metrología, un procedimiento de medición del tamaño funcional presenta una descripción detallada y un conjunto de operaciones, descritas específicamente, para obtener una medición conforme a uno o más principios de medida y a un método de medición del tamaño funcional específico (Abran, 2010; Habra, Abran, Lopez & Sellami, 2008; ISO, 2004). Un método de medición en el cual se basa un procedimiento de medición presenta una secuencia lógica de operaciones, descritas de manera general, usadas para obtener una medición (ISO, 2004).

¹ Principio de medición: fenómeno que sirve como base de una medición

² Concepto a ser medido: la entidad y el atributo bajo consideración.

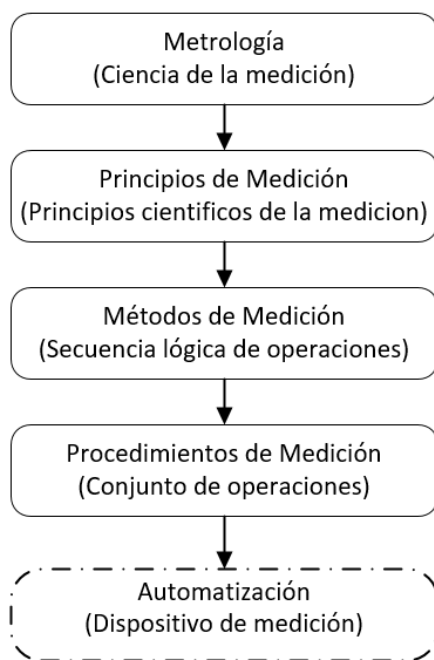


Figura 2-6 Fundamentos de medición y automatización. Adaptado de (Abran, 2010).

La aplicación de un procedimiento de medición requiere un número de elementos y procesos detallados para asegurar que el resultado de una medición sea de alta calidad y que se dé una correcta aplicación de las reglas de un procedimiento de medición. Para esto, se deben aplicar y verificar un conjunto de criterios bien definidos. En la Figura 2-9 se detalla el conjunto de elementos de un procedimiento de medición. Un procedimiento de medición tiene un conjunto de entradas, un conjunto de variables de control y una salida que representa el valor medido, en este caso el tamaño funcional. La instanciación del procedimiento de medición representa un modelo matemático que relaciona las cantidades envueltas en el proceso. Para ejecutar el proceso de medición el operador debe diseñar y seguir el procedimiento de medición. El proceso de medición puede verse influenciado por el operador, el principio de medición³, el método de medición seleccionado y las cantidades que influyen el proceso de medición.

³ Principio de medición: fenómeno que sirve como base de una medición

La calidad del resultado del proceso de medición se determina por los criterios de exactitud, valor verdadero⁴, repetibilidad y reproducibilidad, por su capacidad de calibración, verificación, validación y corrección, y finalmente por la trazabilidad y comparabilidad de los resultados de medición. Es deseable que el diseño y la construcción de un procedimiento de medición siga los pasos del modelo de procesos para métodos de medición del tamaño funcional propuesto por (Abran, 2010; Habra, Abran, Lopez & Sellami, 2008; Jacquet & Abran, 1997) y descrito en la Sección 2.1.3.1. Sin embargo, en la literatura son pocos los procedimientos de medición que han seguido estas recomendaciones de manera sistemática (Quesada-López & Jenkins, 2017). La metodología para la automatización de la medición del tamaño funcional propuesta en esta tesis se diseña basada en el modelo de procesos y se compone de tres procedimientos basados en los métodos IFPUG FPA y COSMIC FFP. Estos permiten realizar la medición automática de tamaño funcional, la verificación automática de la exactitud de las mediciones de tamaño funcional y la evaluación de modelos de estimación de esfuerzo.

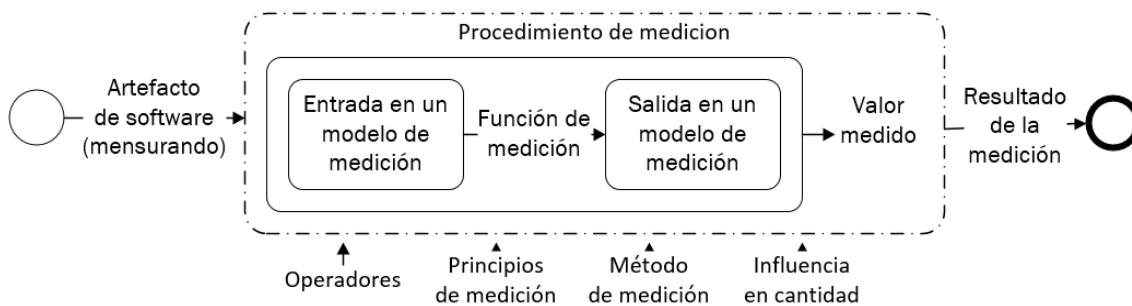


Figura 2-7 Elementos de un procedimiento de medición (Abran, 2010).

2.1.3.1. MODELO DE PROCESOS DE MEDICIÓN DEL TAMAÑO FUNCIONAL

Un procedimiento de medición del tamaño funcional sigue los pasos del modelo de procesos para métodos de medición del tamaño funcional propuesto por (Abran, 2010; Habra, Abran, Lopez & Sellami, 2008; Jacquet & Abran, 1997). En su versión inicial (Jacquet &

⁴ Valor verdadero: valor de referencia calculado por un experto

Abran, 1997), los autores propusieron cuatro pasos bien definidos: diseño y aplicación del método de medición, y verificación y aprovechamiento de los resultados. En la versión del 2010, las actividades de verificación se realizan durante el paso de aplicación. La metodología para la automatización de la medición del tamaño funcional propuesta en esta tesis se diseña basada en el modelo de procesos y se compone de tres procedimientos, los cuales permiten realizar la medición automática de tamaño funcional (que incluye el diseño y aplicación), la verificación de la exactitud de las mediciones de tamaño funcional y la evaluación de modelos de estimación de esfuerzo para el aprovechamiento de los resultados de la medición de tamaño funcional.

La Figura 2-7 presenta el modelo de procesos que plantea los pasos necesarios para el diseño de un método de medición. El modelo de procesos establece que para desarrollar una metodología de medición es necesario realizar un conjunto de pasos que incluyen la definición de un procedimiento de medición, la verificación de la calidad de los resultados y el aprovechamiento de los resultados de medición. Para cada uno de estos pasos se debe detallar el diseño y el proceso de aplicación.

El modelo de procesos (Abran, 2010; Habra, Abran, Lopez & Sellami, 2008) puede ser dividido en tres pasos bien definidos los cuales indican los pasos que se deben realizar. En el Paso 1 se realiza el diseño de la medición que incluye el principio de medición, el método de medición⁵ y el procedimiento de medición⁶. En el Paso 2 se realiza la aplicación del método de medición en un contexto determinado mediante el procedimiento de medición para producir un resultado del tamaño funcional. En el Paso 3 se realiza la explotación de los resultados de medición utilizando modelos de estimación en el contexto determinado. En la práctica, las actividades de verificación se deben realizar en cada una de las tres

⁵ Método de medición: descripción genérica de la secuencia lógica de operaciones utilizadas en una medición.

⁶ Procedimiento de medición: descripción detallada de una medición conforme a uno o más principios de medida y a un método de medición dado.

etapas descritas anteriormente para garantizar el suficiente grado de confianza en el método o procedimiento de medición. Estas actividades deben validar criterios de calidad tales como exactitud⁷, repetibilidad⁸ y reproducibilidad⁹.

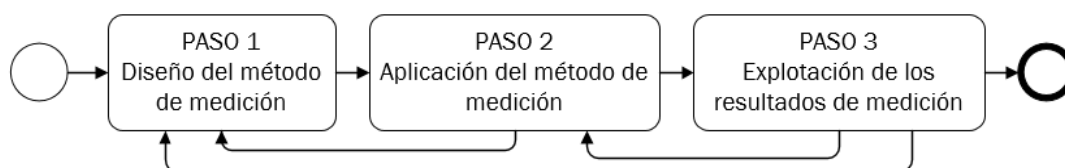


Figura 2-8 Pasos generales del modelo de procesos. Adaptado de (Abran, 2010).

La Figura 2-8 presenta los pasos detallados del modelo de procesos adaptados para el desarrollo de la metodología de medición propuesta en esta tesis. En este caso, los cuadros punteados representan los ajustes realizados que detallan la versión original. El modelo de procesos detalla la estructura una metodología de medición, por lo que la metodología propuesta en esta tesis representa una instanciación de este modelo para un contexto particular. Las sub etapas para cada uno de los pasos del proceso de medición del software son las siguientes:

1. Paso 1: El diseño del método de medición que comprende:
 - a. Definición de objetivos: definición de qué se desea medir, qué tipo de software se mide con el método, cuáles atributos, desde cuál punto de vista, y los usos que tiene el método.
 - b. Caracterización del concepto a ser medido: los conceptos a ser medidos deben ser claramente definidos. En este sub paso se detalla cómo un concepto es descompuesto en otros subconceptos hasta obtener su valor mediante las reglas numéricas. En el caso de los procedimientos de medición, los conceptos a ser

⁷ Exactitud: proximidad entre un valor medido y un valor verdadero.

⁸ Repetibilidad: precisión de medida bajo un conjunto de condiciones iguales. Proximidad entre las mediciones repetidas del mismo objeto.

⁹ Reproducibilidad: precisión de medida bajo un conjunto de condiciones diferentes (sujetos). Proximidad entre las mediciones repetidas del mismo objeto.

medidos se mapean con los conceptos del método de medición seleccionado para el procedimiento.

- c. Definición de los metamodelos: se detalla el grupo de características que representa los artefactos de entrada de software a ser medido y sus relaciones. En el caso de los procedimientos de medición, el metamodelo del procedimiento de medición es mapeado al metamodelo del método de medición seleccionado. El metamodelo representa el conjunto de objetos y sus relaciones que representan la perspectiva funcional del software (modelo funcional).
 - d. Definición de las reglas de mapeo: detalla cómo los componentes del metamodelo de representación funcional del software del procedimiento de medición se mapean con los componentes del metamodelo del método de medición seleccionado. Esto para determinar cada uno de los componentes del metamodelo del método de medición y aplicar las reglas de asignación numérica.
 - e. Definición de las reglas de asignación numérica: es la asignación numérica a cada uno de los elementos del metamodelo del método de medición seleccionado.
2. Paso 2: La aplicación de las reglas del método de medición sobre el artefacto de entrada comprende:
- a. Adquisición de los artefactos de software: comprende la recolección de los artefactos requeridos para la aplicación del método de medición y necesario para la construcción del modelo funcional.
 - b. Construcción del modelo funcional de software: con los artefactos de entrada recolectados, se construye el modelo funcional del software. El modelo describe los componentes medidos por el método y se basa en el metamodelo propuesto y el conjunto de reglas para la identificación de cada componente a partir de los artefactos de entrada. En el caso de procedimientos de medición este modelo funcional se construye con base en el metamodelo del procedimiento que luego

- es mapeado al metamodelo del método de medición. Cuando el modelo está construido se aplican las reglas de mapeo y de asignación numérica.
- c. Aplicación de las reglas de mapeo: en el caso de los procedimientos de medición, se aplican las reglas de mapeo para determinar cómo se asocian los componentes del metamodelo del procedimiento con los componentes del metamodelo del método de medición dando como resultado un modelo funcional a partir del metamodelo del método de medición. Las reglas de asignación numérica se aplican sobre este modelo funcional.
 - d. Aplicación de las reglas de asignación numérica: a partir del modelo funcional se aplican las reglas de asignación numérica para obtener el tamaño funcional del software.
 - e. El análisis de los resultados de la aplicación de las reglas de medición debe ser documentado y auditado. Este paso comprende las siguientes actividades:
 - Presentación de Resultados: todo el proceso de medición debe ser documentado de acuerdo al nivel de granularidad definido para la aplicación del método de medición. Se recomienda documentar el mayor nivel de granularidad reportando todo el detalle del proceso de la medición para mejorar el proceso de auditoría de resultados.
 - Auditoría de Resultados: los resultados deben ser revisados y validados para evaluar su calidad. Un protocolo de verificación de calidad debe definirse de forma clara y sistemática para aplicar los procesos de validación. Estos protocolos incluyen la verificación de todas las reglas aplicadas en el proceso, la calidad de los artefactos de entrada y la comparación con valores conocidos como verdaderos.
3. Paso 3: La explotación o aprovechamiento de los resultados del método de medición en modelos de estimación comprende la utilización del tamaño funcional resultado del proceso de medición y los factores del contexto como entrada de otros modelos tales

como: modelos de calidad (métricas para medir densidad de fallos o modelos de predicción de fallos), modelos de estimación de esfuerzo de desarrollo, productividad y costos.

- a. Esto incluye el análisis de los resultados cuantitativos de los modelos, usualmente utilizados para propósitos de evaluación, estadísticas, comparación, estimación y/o predicción.

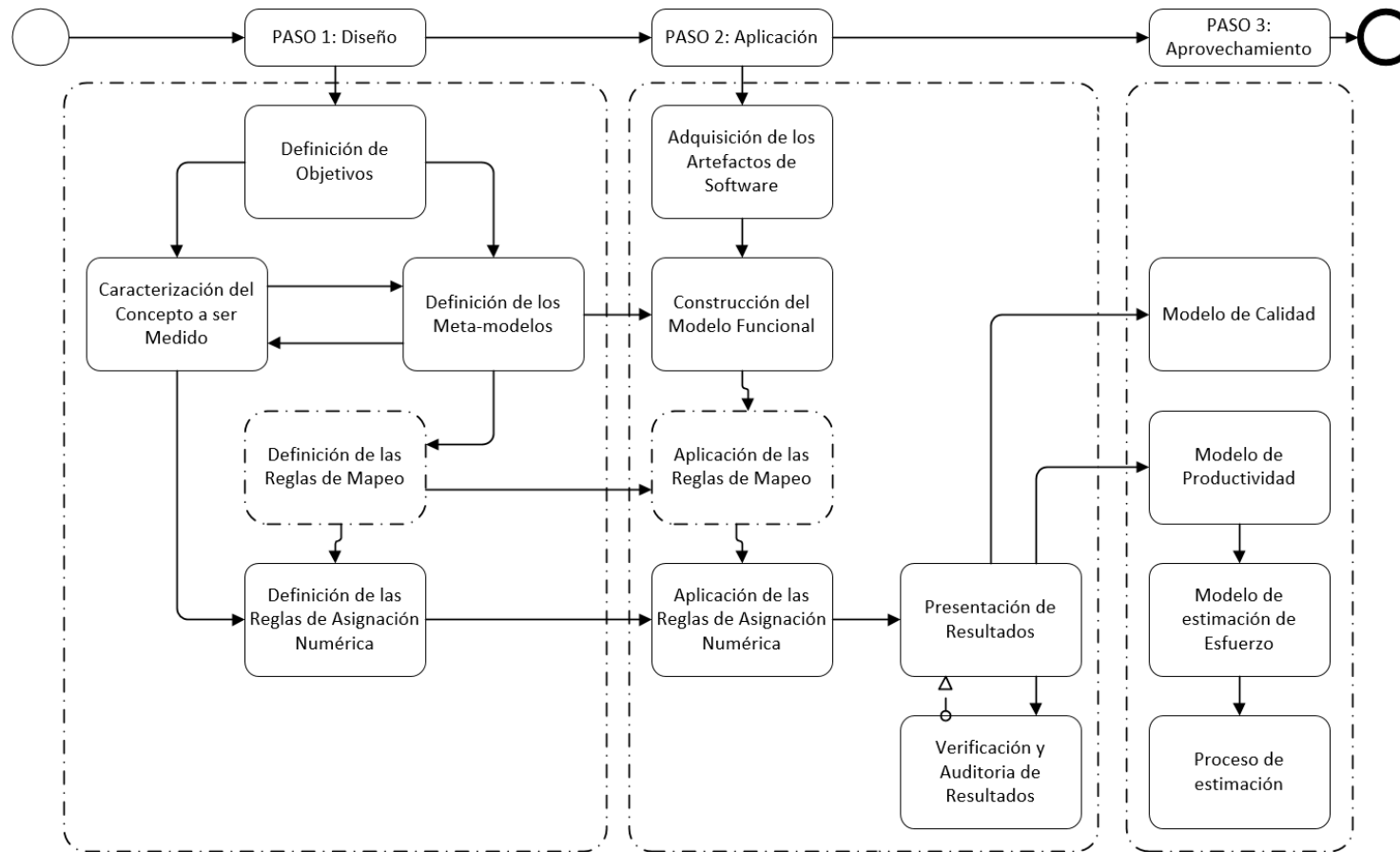


Figura 2-9 Pasos detallados del modelo de procesos. Adaptado de (Abran, 2010).

2.2 METODOLOGÍAS DE INVESTIGACIÓN PARA LA INGENIERÍA DE SOFTWARE

A continuación se detallan las metodologías utilizadas en esta investigación. Primero, presentamos la metodología de ciencia del diseño y posteriormente, detallamos las metodologías relacionadas con la ingeniería de software empírica.

2.1.4. CIENCIA DEL DISEÑO PARA LA INGENIERÍA DEL SOFTWARE

La metodología de investigación Ciencia del Diseño para Sistemas de Información e Ingeniería del Software propone un marco de trabajo para diseñar artefactos de software y obtener nuevo conocimiento en el área de investigación a partir de la investigación del problema y la evaluación de los artefactos donde todas las decisiones sobre el diseño se basan en la evidencia obtenida (Wieringa, 2014; Peffers, Tuunanen, Rothenberger & Chatterjee, 2007; Hevner, March, Park & Ram, 2004). Ciencia del diseño es el diseño e investigación de artefactos de software interactuando en un contexto particular (Wieringa, 2014).

Esta metodología establece que cuando nos enfrentamos a un problema de investigación en ingeniería de software, nos enfrentamos a un conjunto de problemas de diseño (dominio ingenieril) y un conjunto de preguntas de investigación (dominio del conocimiento). Por tanto, para resolver dicho problema de diseño y responder las preguntas de investigación, necesitamos ejecutar un conjunto de ciclos regulativos (ciclos ingenieriles y de investigación) que permiten obtener una solución. Es así como la metodología permite diseñar e investigar artefactos de software en un contexto particular donde los ciclos ingenieriles permiten diseñar el artefacto (en nuestro caso la metodología de medición y las herramientas prototipo) y los ciclos de investigación permiten generar nuevo conocimiento en el área de investigación (en nuestro caso las investigaciones para evaluar la efectividad de la metodología y sus procedimientos).

La Figura 2-10 detalla la estructura del marco de trabajo de la metodología Ciencia del Diseño. Los problemas de diseño son analizados en los ciclos ingenieriles (Parte II), las preguntas de investigación que requieren investigaciones empíricas, son analizadas en los

ciclos de investigación (Parte IV). Ambos ciclos son apoyados por teorías que permiten describir, explicar, y predecir los fenómenos bajo estudio (Parte III).

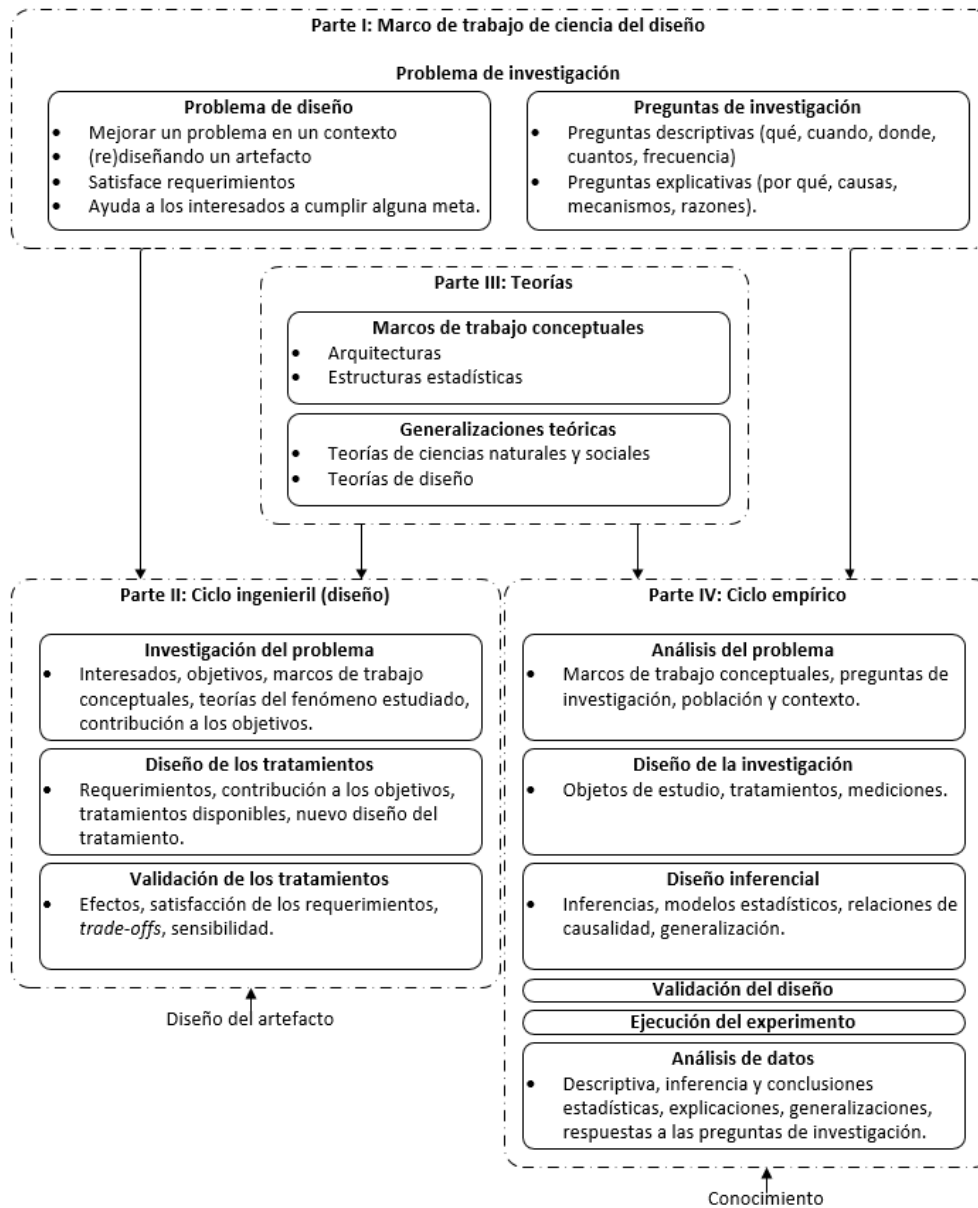


Figura 2-10 Marco de trabajo de la metodología Ciencia del Diseño. Adaptado de (Wieringa, 2014).

El producto de los ciclos ingenieriles son los artefactos validados. Estos ciclos incluyen la investigación del problema y el diseño y validación de la solución. Los ciclos de investigación incluyen el análisis del problema, el diseño de la investigación y la validación empírica mediante la ejecución de los estudios empíricos y el análisis de datos. En estos ciclos se conducen los experimentos a partir del diseño del artefacto y se analizan los resultados para contestar las preguntas de investigación.

En nuestro proyecto de investigación, mientras se diseña la metodología de medición, cada uno de sus procedimientos y prototipos, se realizan distintos estudios empíricos que permiten validar cada uno de ellos y generar conocimiento en el área de investigación.

2.1.5. INGENIERÍA DE SOFTWARE EMPÍRICA

La ingeniería de software empírica provee las metodologías para analizar y evaluar los fenómenos, métodos, metodologías y procesos de la ingeniería de software (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012). Los estudios empíricos proveen enfoques sistemáticos y científicos para recolectar evidencia que permite probar hipótesis y responder preguntas de investigación específicas. La ingeniería de software basada en evidencia (Kitchenham, Dyba & Jorgensen, 2004) es un área que se ha desarrollado en los últimos años y continua en constante evolución. Los estudios empíricos juegan un rol fundamental en la evolución de la disciplina de la ingeniería de software y permiten la creación de un cuerpo de conocimiento a partir de la observación y evidencia empírica (Basili, 2007).

Los principales tipos de estudios empíricos son las metodologías para el mapeo sistemático (Petersen, Feldt, Mujtaba & Mattsson, 2008) y la revisión sistemática de literatura (Biolchini, Mian, Natali & Travassos, 2005; Budgen & Brereton, 2006; Kitchenham, 2004), casos de estudio (Runeson & Höst, 2009; Yin, 2009), experimentos controlados (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012) y las replicaciones de estudios empíricos (Carver, 2010; Juristo & Vegas, 2009; Shull, Carver, Vegas & Juristo, 2008). Los estudios empíricos pueden contribuir para mejorar la investigación y las prácticas de la ingeniería

del software. Cada vez más se requiere una mayor formalización de estos estudios que incrementen su calidad, con un enfoque en la síntesis de la evidencia y la creación de teorías que generen valor agregado a la academia y a la industria (Sjoberg, Dyba & Jorgensen, 2007).

2.1.5.1. ESTUDIOS SECUNDARIOS

Las metodologías de investigación para realizar estudios secundarios permiten, a partir de un proceso sistemático, identificar, clasificar, analizar y caracterizar investigaciones relevantes relacionadas con un área de investigación y con el alcance delimitado por una pregunta de investigación (Kitchenham & Charters, 2007). Un estudio secundario sistemático minimiza los posibles sesgos cuando se agrega evidencia de múltiples estudios primarios. El diseño de estos estudios define rigurosamente cada uno de los pasos que se ejecutan para obtener de manera sistemática la evidencia relacionada con un tema de interés. Con un mapeo sistemático del área se delimita el alcance de la revisión sistemática de literatura (Petersen, Feldt, Mujtaba & Mattsson., 2008), luego, se analizan las preguntas de investigación de la revisión sistemática. Realizar un proceso riguroso y sistemático que sigue un protocolo de mapeo y revisión sistemática de literatura reduce el sesgo de selección de artículos, permite mayor confiabilidad en los resultados, y es un proceso bien definido y documentado que posibilita ser replicado. Además, es una metodología probada y adoptada por el área de la ingeniería del software experimental (Budgen & Brereton, 2006).

Los fundamentos teóricos para la ejecución de estos estudios secundarios sistemáticos se describen en (Badampudi, Wohlin & Petersen, 2015; Biolchini, Mian, Natali & Travassos, 2005; Brereton, Kitchenham & Budgen, 2007; Kitchenham & Charters, 2007; Budgen & Brereton, 2006; Kitchenham, Budgen & Brereton, 2011; Kitchenham, 2004; Petersen, Feldt, Mujtaba & Mattsson, 2008; Petersen, Vakkalanka & Kuzniarz, 2015; Wohlin, 2014). Está comprobado que un estudio sistemático de mapeo y revisión de literatura apoya el proceso de investigación en la ingeniería del software. En particular, los estudios de mapeo mejoran el proceso de identificación de líneas de investigación y proveen las líneas base

para establecer nuevos esfuerzos en investigación siempre que sean de alta calidad en términos de rigurosidad y completitud, lo que permite que sean una base confiable en la definición de nuevas investigaciones (Kitchenham, Budgen & Brereton, 2011).

Esta metodología detalla los pasos necesarios para el diseño y la ejecución de cada una de las etapas del proceso de revisión tal y como se muestra en la Figura 2-11. Estas etapas comprenden: (1) la definición del objetivo de investigación y las preguntas de investigación, (2) la definición de la estrategia de búsqueda, (3) el proceso de selección de estudios primarios basado en criterios específicos de inclusión y exclusión, (4) la evaluación de la calidad y del rigor de los artículos seleccionados, (5) la extracción de los datos, (6) el proceso de análisis y clasificación de la evidencia identificada. A partir de los resultados del proceso se responden las preguntas de investigación de acuerdo a la evidencia analizada. La Figura 2-12 detalla las particularidades para la conducción de un mapeo sistemático de revisión de literatura. Para la estrategia de búsqueda, el proceso de “bola de nieve” (Badampudi, Wohlin & Petersen, 2015; Wohlin, 2014), hacia adelante y hacia atrás, permite mejorar la confiabilidad de la búsqueda y la identificación de los artículos relevantes para la investigación.

Está comprobado que un estudio sistemático de revisión de literatura apoya el proceso de investigación en la ingeniería del software. En particular, los estudios de mapeo mejoran el proceso de identificación de líneas de investigación y proveen las líneas base para establecer nuevos esfuerzos en investigación siempre que sean de alta calidad en términos de rigurosidad y completitud, lo que permite que sean una base confiable en la definición de nuevas investigaciones (Kitchenham, Budgen & Brereton, 2011).

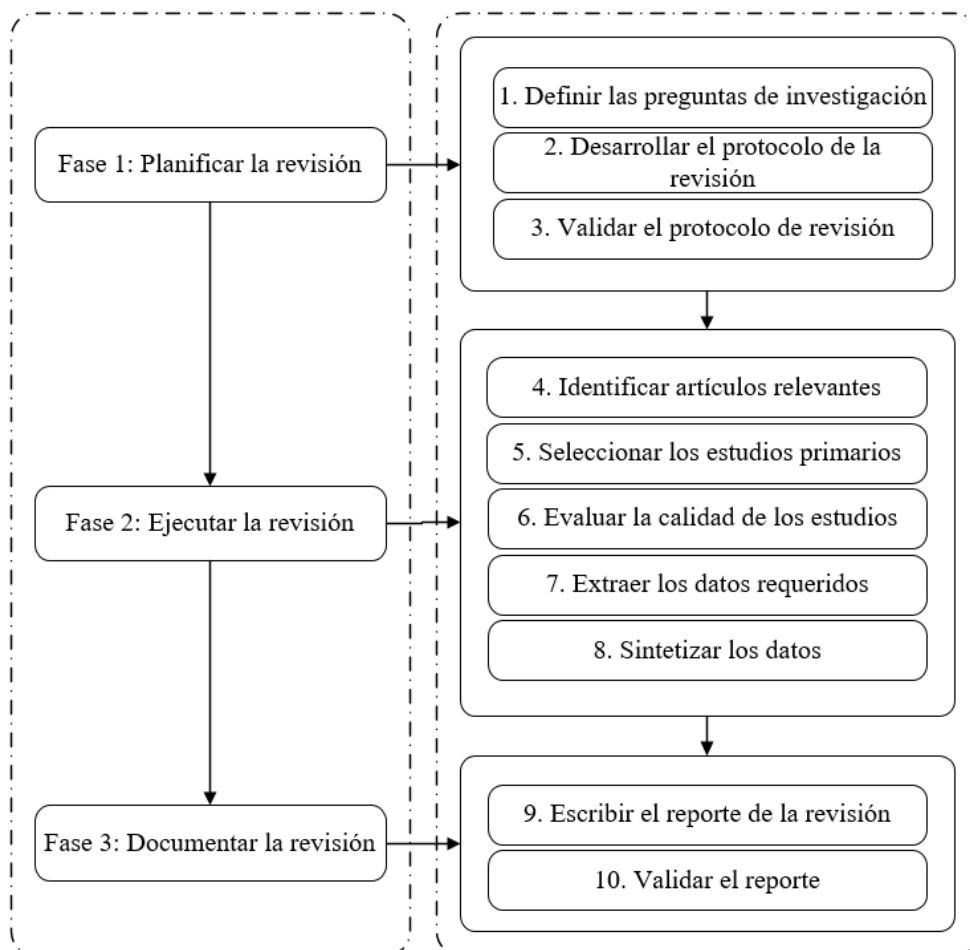


Figura 2-11 Proceso de la revisión sistemática de literatura. Adaptado de (Kitchenham & Charters, 2007).

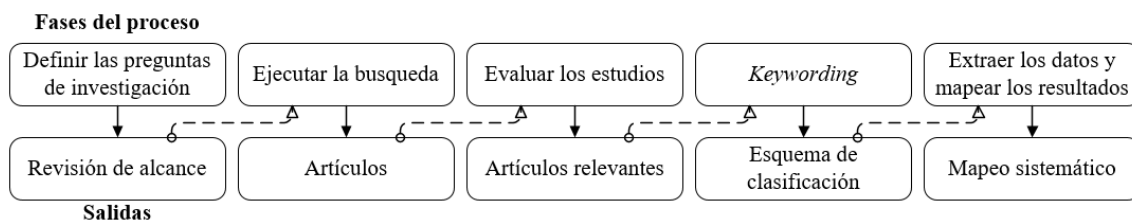


Figura 2-12 Proceso del mapeo sistemática de literatura. Adaptado de (Petersen, Feldt, Mujtaba & Mattsson, 2008).

2.1.5.2. EXPERIMENTOS CONTROLADOS Y ESTUDIOS CUASI EXPERIMENTALES

Los experimentos controlados son útiles para estudiar la causalidad o efecto de un tratamiento (variables independientes) sobre resultado (variable dependientes) en un contexto controlado. Los fundamentos teóricos y los detalles de cómo ejecutar los experimentos se describen en (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012).

El protocolo metodológico detalla los pasos necesarios para la ejecución de cada una de las etapas del proceso de experimentación tal y como se muestra en la Figura 2-13. Estas etapas comprenden: (1) La definición del alcance y el objetivo de la investigación. (2) El planteamiento del problema que incluye la selección del contexto, la formulación de las hipótesis, la selección de variables, objetos y sujetos, el diseño experimental, la instrumentación y la evaluación de la validez. (3) La operación del experimento que incluye la preparación, ejecución del experimento y la validación de los datos. (4) El análisis e interpretación de los resultados que incluye la estadística descriptiva, la reducción de los conjuntos de datos si es necesario y las pruebas de hipótesis. Finalmente, (5) la presentación del reporte de los resultados y paquete del experimento.

Los principales aspectos a considerar en los experimentos es mantener la validez de los estudios durante el diseño, la ejecución y el análisis para lograr la confiabilidad de los resultados (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012) y reportar de manera detallada los experimentos para soportar la replicación de los estudios (Carver, 2010). Es importante considerar que dadas las características particulares de los datos de la ingeniería del software, la selección de métodos robustos para el análisis de las métricas recolectadas en los procesos experimentales representa un factor clave para poder obtener conclusiones válidas en los estudios realizados en el área (Kitchenham, Madeyski, Budgen, Keung, Brereton, Charters & Pohthong, 2017).

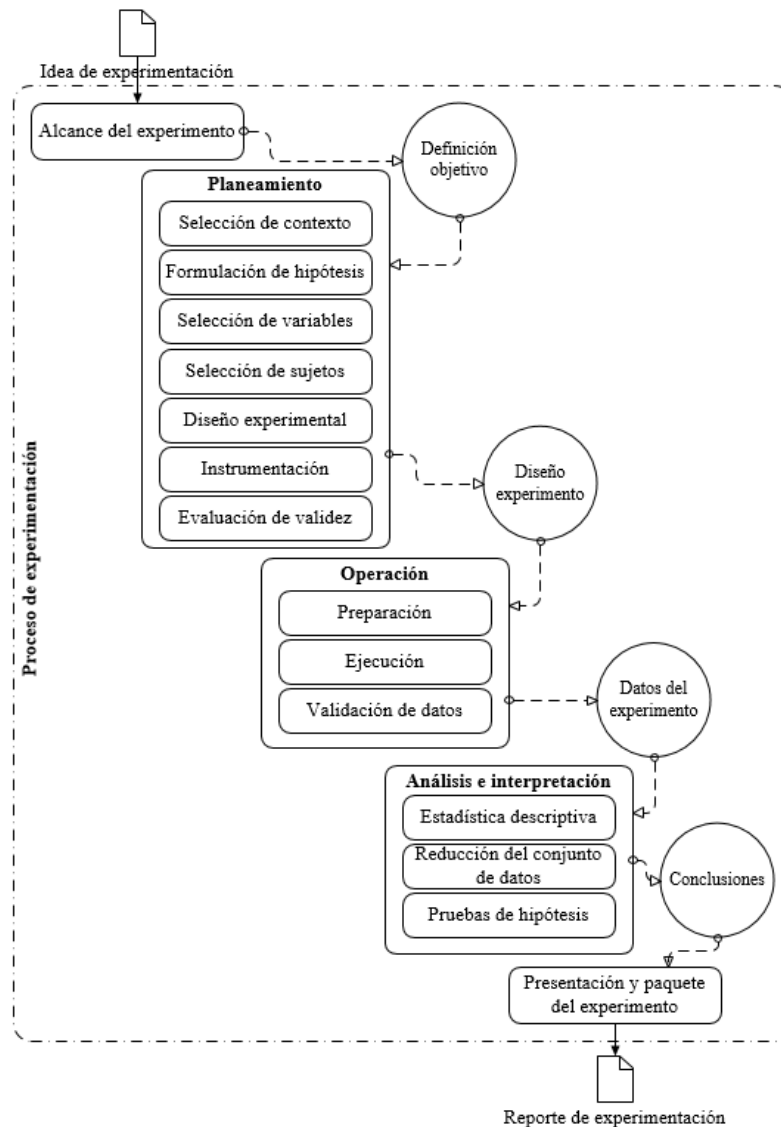


Figura 2-13 Proceso de experimentación. Adaptado de (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012).

2.1.5.3. CASOS DE ESTUDIO

Los casos de estudios permiten estudiar los fenómenos en su contexto natural (Runeson & Höst, 2009; Runeson, Host, Rainer & Regnell, 2012). Esto implica un análisis profundo de uno o más casos (personas, grupos, métodos, artefactos o sistemas) en un contexto específico. Los casos de estudio son útiles cuando se estudian preguntas de investigación

descriptivas y explicativas (Yin, 2009). El protocolo metodológico detalla los pasos necesarios para la ejecución de cada una de las etapas del proceso de conducción de un caso de estudio, tal y como se muestra en la Figura 2-14. Estas etapas comprenden: (1) El diseño del caso de estudio que incluye la definición de las preguntas de investigación, la selección del caso, la selección de los sujetos, el procedimiento de recolección de datos, el procedimiento de análisis y el procedimiento de validez. (2) El análisis de resultados que incluye la descripción del caso y de los sujetos, los aspectos de análisis e interpretación y la evaluación de validez. En la Figura 2-14 se detallan las principales diferencias de un proceso experimental y un caso de estudio los cuales se consideran para la construcción de los estudios empíricos. Estos representan la estructura de un protocolo y un reporte de experimentación.

2.1.5.4. REPLICACIÓN DE ESTUDIOS EMPÍRICOS

Las replications de experimentos juegan un rol fundamental en la comunidad de la ingeniería del software porque permiten construir conocimiento sobre la evidencia obtenida y bajo qué condiciones fue obtenida dicha evidencia (Shull, Carver, Vegas & Juristo, 2008). Estas implican la conducción de los estudios empíricos bajo condiciones similares con variaciones particulares para evaluar sus efectos, esto con el fin de evaluar la confiabilidad y la generalización de los resultados encontrados en los experimentos realizados (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012). Uno de los principales beneficios de las replications es la posibilidad de adquirir conocimientos sobre un tema de investigación particular a partir de las diferencias encontradas en los distintos estudios experimentales replicados (Juristo & Vegas, 2009). Los estudios experimentales requieren ser replicados en distintos contextos y bajo diferentes condiciones para producir resultados que puedan ser generalizados (Carver, Juristo, Baldassarre & Vegas, 2014).

Experimento		Caso de estudio	
Sección	Sub sección	Sección	Sub sección
Motivación		Introducción	
	Problema		Problema
	Objetivos		Objetivos
	Contexto		Contexto
Trabajo relacionado		Trabajo relacionado	
			Estudios previos
			Teorías
Diseño experimental		Diseño del caso de estudio	
	Objetivos		Preguntas de investigación
	Hipótesis		Selección de caso
	VARIABLES		Selección de sujetos
	Diseño		Procedimiento de recolección de datos
	Sujetos		Procedimiento de análisis
	Objetos		Procedimiento de validez
	Instrumentación		
	Procedimiento de recolección de datos		
	Procedimiento de análisis		
	Evaluación de validez		
Ejecución		Resultados	
	Muestra		Descripción del caso y sujetos
	Preparación		Aspectos de análisis e interpretación
	Recolección de datos		Evaluación de validez
	Procedimiento de validez		
Análisis			
	Estadística descriptiva		
	Reducción de conjunto de datos		
	Prueba de hipótesis		
Interpretación			
	Evaluación de resultados		
	Interpretación de resultados		
Conclusiones y trabajo futuro		Conclusiones y trabajo futuro	
	Relación con la evidencia existente		Resumen de conclusiones
	Impacto de los resultados		Relación con la evidencia existente
	Limitaciones		Impacto e implicaciones de los resultados
	Trabajo futuro		Limitaciones
			Trabajo futuro

Figura 2-14 Estructura del protocolo de casos de estudio en comparación con el proceso de experimentación empírica. Adaptado de (Runeson & Höst, 2009; Runeson, Host, Rainer & Regnell, 2012).

El protocolo metodológico para reportar los estudios de replicación comprende el detalle de un conjunto de secciones que permiten identificar las similitudes y diferencias entre los estudios originales con respecto a las repeticiones realizadas (Carver, 2010; Carver, Juristo, Baldassarre & Vegas, 2014). Estos estudios deben incluir al menos el reporte de

la siguiente información: (1) La información relacionada con los estudios originales que incluya las preguntas de investigación, los participantes, el diseño experimental, la instrumentación, las variables de contexto y el resumen de los resultados. (2) La información de la replicación que incluya la motivación para realizar el estudio, el nivel de interacción con los investigadores originales y los cambios al experimento original. (3) La comparación de los resultados del estudio original con los de la replicación que incluya los resultados consistentes y las principales diferencias. Finalmente, (4) las conclusiones a través de los estudios incluyendo el análisis de los resultados que son obtenidos al analizar el conjunto de experimentos. En el caso de familias de replications, se deben proporcionar detalles de los resultados de todos los estudios relacionados.

2.1.5.5. VALIDEZ DE LOS ESTUDIOS EMPÍRICOS

La validez de los resultados de los estudios empíricos debe ser analizada cuando se realiza la planificación de los estudios y debe ser considerada para determinar la confiabilidad de los resultados obtenidos y que tan generalizables son estos resultados. Primero, los resultados deben ser válidos para la población de la que se extraen las muestras, y segundo debemos analizar si los resultados se pueden generalizar a una población más amplia (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012). Las amenazas a la validez deben reportarse en los estudios empíricos ya que representan potenciales limitaciones sobre los resultados obtenidos, de igual manera, los investigadores deben reportar las acciones tomadas para mitigar estas amenazas a la validez. Existen diferentes clasificaciones para las amenazas a la validez, en el contexto de los estudios empíricos de este trabajo de investigación, se utiliza la categorización de amenazas a la validez descrita en (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012) la cual detallamos a continuación:

- *Validez de la conclusión:* esta se refiere a la relación entre el tratamiento y los resultados obtenidos. La validez de la conclusión representa el grado en que los investigadores extraen conclusiones estadísticamente significativas de los datos y con suficiente poder

estadístico. La validez de la conclusión se refiere muchas veces como la validez de la conclusión estadística.

- *Validez interna*: cuando se encuentra una relación entre el tratamiento y los resultados debemos verificar si existe una relación causal y que no es resultado de un factor del cual no tenemos control. La validez interna representa el grado en que podemos asegurar que el tratamiento es el que causa los efectos. En este caso, los procedimientos, tratamientos y participantes pueden amenazar la validez interna. Las amenazas a la validez interna se preocupan por las relaciones causales que no existen.
- *Validez del constructo*: esta se refiere a la relación entre las teorías base y la observación. Si existe una relación de causa y efecto debemos asegurarnos que el tratamiento refleja el constructo de la causa y que los resultados reflejan el constructo del efecto. La validez del constructo representa el grado en que el experimento está midiendo lo que pretende medir, donde los constructos deben presentar una definición única que no se preste a interpretaciones. Esta analiza hasta qué punto la configuración del experimento realmente refleja el constructo bajo estudio.
- *Validez externa*: esta se refiere a la generalización de los resultados fuera del contexto y la configuración del experimento. Si existe una relación causal entre el constructo y el efecto debemos determinar si es posible generalizar los resultados fuera del alcance de nuestro estudio. La validez externa representa el grado en que los resultados son generalizables a otros sujetos, situaciones y contextos.

Capítulo 3. TRABAJO RELACIONADO

Este capítulo presenta el trabajo relacionado con la investigación que incluye estudios relevantes para la definición de cada uno de los procedimientos de la metodología de medición propuesta en esta tesis. El análisis se realiza para obtener el conocimiento necesario que apoye el proceso del diseño del procedimiento de medición de la metodología propuesta en el Capítulo 5 y la definición de los estudios empíricos que lo evalúan en el Capítulo 6 y el Capítulo 7.

3.1 ESTRUCTURA Y APLICABILIDAD DE LOS MÉTODOS DE MEDICIÓN DEL TAMAÑO FUNCIONAL IFPUG FPA Y COSMIC FFP

Esta sección discute aspectos relacionados con la estructura y aplicabilidad de los métodos de medición IFPUG FPA y COSMIC FFP. El Paso 3 del modelo de procesos para métodos de medición del tamaño funcional (Abran, 2010) indica que con los resultados de las mediciones se realiza el aprovechamiento de los resultados, por ejemplo, en modelos de estimación de esfuerzo de desarrollo para contextos específicos. La construcción de modelos de estimación simplificados podría basarse en conjuntos de los componentes funcionales básicos (BFC) que representarían una oportunidad para reducir el costo y tiempo que se requiere para la medición del tamaño funcional requerida para la implementación de estos modelos.

Los estudios que se describen a continuación fueron identificados realizando una búsqueda exploratoria que permite obtener estudios relevantes sobre la evaluación de la estructura y aplicabilidad de los métodos de medición del tamaño funcional estandarizados relevantes para nuestra investigación. La cadena de búsqueda utilizada para su identificación en las bases de datos automatizadas se compone de los siguientes términos: ("*Function* size* measure**" OR "*function point**" OR "*COSMIC*" OR "*IFPUG*") AND ("*BFC*" OR "*basic functional component**"). A partir de la identificación de los artículos se seleccionan los estudios relevantes para el análisis basado en un conjunto de criterios de inclusión y exclusión que

permiten seleccionar estudios sobre la evaluación de la estructura y aplicabilidad de los métodos IFPUG FPA y COSMIC FFP. La biblioteca utilizada para la búsqueda fue la base de datos referencial Scopus. A continuación presentamos los principales hallazgos.

Las mediciones basadas en puntos de función han sido sometidas a investigaciones para evaluar su estructura interna y aplicabilidad para la construcción de modelos de estimación de esfuerzo del desarrollo de software. Existe evidencia empírica que indica relaciones entre los diferentes componentes funcionales básicos (BFC) y entre los BFC y el esfuerzo del desarrollo del software (Silva, Pinheiro & Albuquerque, 2016). Esto implica dos aspectos a considerar, el primero, la correlación entre BFC puede indicar que algunos de los componentes se están midiendo dos veces, lo que representa doble esfuerzo en el conteo, y segundo, medir de manera duplicada un BFC puede afectar la confiabilidad del método de medición (Kitchenham & Kansala, 1993; Lavazza, Morasca & Robiolo, 2013). Asimismo, esto sugiere que un método simplificado de medición puede obtener resultados similares que la aplicación del método completo para la construcción de estos modelos de estimación de esfuerzo (Kitchenham & Kansala, 1993; Lavazza, Morasca & Robiolo, 2013; Lokan, 1999). Por tanto, la construcción y evaluación de modelos simplificados de estimación de esfuerzo en contextos específicos puede suponer un beneficio para la práctica profesional en la que los procedimientos de medición puedan llevarse a cabo de manera más simple, rápida y manteniendo la exactitud de las estimaciones. Estos estudios sugieren que el uso de los BFC para la construcción de modelos simplificados de estimación puede igualar el rendimiento de las estimaciones de los modelos completos (Buglione & Gencel, 2008; Gencel & Buglione, 2008; Gencel, 2008; Amasaki & Yokogawa, 2012; Lavazza, Morasca & Robiolo, 2013).

En la literatura se han analizado, principalmente, los métodos de medición IFPUG FPA y COSMIC FFP. Por ejemplo, Kitchenham & Kansala (1993) analizan la consistencia interna del método FPA y el uso del tamaño funcional para predecir el esfuerzo de desarrollo. Por su parte, Jeffery, Low & Barnes (1993) investigaron la correlación de los componentes funcionales básicos del método FPA. Jeffery & Stathis (1996) evalúan empíricamente la independencia de los BFC y la relación de estos BFC con el esfuerzo de desarrollo. Lokan

(1999) estudia las correlaciones de los BFC y analiza los factores de influencia entre las relaciones de los BFC. Amasaki & Yokogawa (2012) analizan la efectividad de modelos de predicción de esfuerzo basados en los BFC para el método de medición IFPUG FPA. Finalmente, Lavazza, Morasca & Robiolo (2013) analizan la relación de los BFC y analizan la posibilidad de construir modelos simplificados basados en el tamaño funcional para la estimación de esfuerzo. Finalmente, Gencel, (2008), Buglione & Gencel (2008) y Gencel & Buglione (2008) investigan modelos de predicción de esfuerzo basados en los BFC del método COSMIC FFP y cómo estos tienen un impacto en la confiabilidad de los resultados.

En el caso del diseño de la metodología de medición automatizada propuesta en esta tesis, el procedimiento de evaluación de modelos podría basarse en conjuntos de los BFC que simplificarían el proceso de medición y permitan construir modelos de estimación de esfuerzo simplificados. Para comprobar si esto es posible, se deben realizar evaluaciones empíricas que corroboren si los resultados de los modelos simplificados de estimación de esfuerzo mantienen un desempeño similar a los modelos basados en el total de los puntos de función.

3.2 MARCOS DE TRABAJO PARA LA EVALUACIÓN DE MODELOS DE ESTIMACIÓN DE ESFUERZO BASADOS EN EL TAMAÑO FUNCIONAL

Esta sección presenta aspectos relacionados con la evaluación de los modelos de estimación de esfuerzo de desarrollo de software basados en el tamaño funcional. En el Paso 3 del modelo de procesos para métodos de medición del tamaño funcional (Abran, 2010) que se desarrolla en este proyecto de investigación, debemos aprovechar los resultados de las mediciones, por ejemplo, para construir modelos de estimación de esfuerzo en el contexto en que se desarrollan las aplicaciones de software que medimos. Asimismo, es de interés evaluar la efectividad de modelos simplificados de estimación construidos a partir de conjuntos de los componentes funcionales básicos (BFC). Para esto se debe definir un procedimiento riguroso que permitan la selección de los conjuntos de los BFC y evaluar la efectividad de estos modelos simplificados.

Los estudios analizados son identificados realizando una búsqueda exploratoria que permite obtener estudios relevantes sobre marcos de trabajo para la evaluación de los modelos de estimación de esfuerzo de desarrollo de software basado en el tamaño funcional. La cadena de búsqueda utilizada para la identificación en las bases de datos automatizadas se compone de los siguientes términos: ("*framework*") AND ("*model*") AND ("*effort estimation*" OR "*effort prediction*"). A partir de la identificación de los artículos se seleccionan los relevantes para el análisis basado en un conjunto de criterios de inclusión y exclusión que permiten seleccionar estudios sobre la evaluación de los modelos de estimación de esfuerzo utilizando como métricas predictoras las mediciones de los métodos IFPUG FPA y COSMIC FFP. La biblioteca utilizada fue la base de datos Scopus. A continuación se presentan los principales hallazgos.

La investigación sobre la evaluación de modelos de estimación de esfuerzo de desarrollo de software a partir del tamaño funcional presenta aun grandes retos para la comunidad de la ingeniería del software (Jorgensen & Shepperd, 2007; Gencel, 2008; Dejaeger, Verbeke, Martens & Baesens, 2012). Uno de los principales aspectos a considerar es el cómo se realiza la selección de datos a partir de repositorios históricos para obtener estimaciones consistentes (Gencel & Buglione, 2008; Kitchenham & Mendes, 2009) y el cómo se puede realizar una comparación justa entre modelos que garantice que no existan sesgos en la selección de los casos para la construcción de cada uno de los modelos bajo comparación (Kitchenham & Mendes, 2009). Se ha determinado que en estimación de esfuerzo, el cómo se realiza la selección de los datos, es aún más importante que los algoritmos de construcción de modelos que se aplican a los datos (Menzies, Mathew, Boehm & Hihn, 2016). Por lo anterior, dado que pueden existir diferentes modelos que presentan distintos resultados dependientes de los conjuntos de datos con los que se construyen los modelos, una evaluación completa que permita una selección acertada de cada modelo en un contexto determinado es clave para tomar decisiones acerca de su efectividad (Shepperd, 2007).

Dada esta problemática, se han desarrollado estudios que plantean enfoques para la evaluación comparativa de modelos de estimación de esfuerzo y definen un conjunto de métricas de evaluación y modelos de línea base que permiten realizar estas comparaciones

entre modelos con el menor sesgo posible (Whigham, Owen & Macdonell, 2015). Estos se dirigen, principalmente, a lograr mantener la estabilidad de los resultados de evaluación a través de diferentes conjuntos de datos. Dejaeger, Verbeke, Martens & Baesens (2012) establecen que la selección y combinación de diferentes técnicas de minería de datos e inteligencia artificial permiten mejorar los resultados de la evaluación de los métodos de predicción para conjuntos de datos con características específicas. La motivación detrás del uso exhaustivo de estas combinaciones es depender lo mínimo posible de las características de los datos cuando se realiza el entrenamiento de los modelos de predicción. Harman & Jones (2001) plantean que las técnicas de búsqueda deben apoyar la resolución de los problemas de la ingeniería del software, ya que en su mayoría, estos son problemas de optimización. Por lo tanto, el área de la ingeniería del software basada en búsqueda (SBSE) enfoca los problemas del área como problemas de búsqueda, en un espacio de búsqueda específico y con un conjunto de soluciones posibles (Harman, Burke, Clark & Yao, 2012). Los modelos de estimación dependen directamente de las características de los conjuntos de datos y no existen reglas únicas para decidir cuál es la mejor combinación y configuración de las técnicas para la construcción de modelos de estimación de esfuerzo. Los enfoques de SBSE pueden apoyar el proceso de selección de técnicas y parámetros que puedan mejorar los resultados de estimación (Ferrucci, Harman & Sarro, 2014).

Durante los últimos años, las evaluaciones de modelos se han desarrollado mediante estudios comparativos (*benchmarking*) utilizando múltiples conjuntos de datos y técnicas de inteligencia artificial y minería de datos. Por ejemplo, Dejaeger, Verbeke, Martens & Baesens (2012) realizaron una comparación empírica a gran escala usando diferentes tipos de técnicas de inteligencia artificial y minería de datos y analizaron los aspectos relacionados con la selección de las características de los conjuntos de datos. Oliveira, Braga, Lima & Cornélio (2010) realizaron investigaciones comparativas para evaluar el efecto de la configuración de parámetros en el desempeño de los modelos de predicción de esfuerzo construidos mediante técnicas de regresión. Del mismo modo, Song, Minku & Yao (2013) realizaron comparaciones para evaluar el efecto de la configuración de parámetros en el desempeño de los modelos construidos a partir de técnicas de aprendizaje máquina. Por su parte,

Keung, Kocaguneli & Menzies (2013) evalúan noventa predictores con veinte conjuntos de datos y siete métricas de evaluación para determinar la estabilidad según la selección de los predictores. Finalmente, Huang, Li & Xie (2015) evalúan un conjunto de técnicas de pre procesamiento y de aprendizaje máquina en modelos de predicción de esfuerzo y analizan las características de los conjuntos de datos y algoritmos de acuerdo a su efectividad.

Con el fin de establecer marcos de trabajo para la evaluación comparativa de modelos de estimación de esfuerzo, múltiples estudios en el área han propuesto enfoques para mejorar los procesos de comparación entre modelos. Por ejemplo, Shepperd & MacDonell (2012) proponen un marco de trabajo para evaluación de modelos de predicción que reduce la inconsistencia a través de resultados comparativos. Este contempla el uso de métricas que disminuyen el sesgo y que pueden permitir la agregación de resultados entre estudios y futuros meta análisis. Menzies & Shepperd (2012) y Keung, Kocaguneli & Menzies (2013) discuten el problema de la inestabilidad de los sistemas de predicción y proponen marcos de trabajo para el área. Estos estudios analizan las fuentes de inestabilidad y proveen recomendaciones para disminuir estos problemas. Los estudios además plantean que la evaluación de combinaciones de técnicas para distintos conjuntos de datos es un campo de investigación abierto. Mittas & Angelis (2013) proponen un marco de trabajo estadístico basado en un algoritmo de múltiples comparaciones para clasificar la efectividad de los modelos de predicción de esfuerzo. Por su parte, Whigham, Owen & Macdonell (2015) presentaron un método para establecer una línea base de comparación para los modelos de predicción de esfuerzo evaluados basado en regresiones lineales múltiples. Dolado, Rodriguez, Harman, Langdon & Sarro (2016) proponen una medida para comparar modelos de estimación basada en un enfoque de pruebas del mínimo intervalo de equivalencia para el error absoluto y una estimación aleatoria como línea base de referencia y Langdon, Dolado, Sarro & Harman (2016) proponen un método para establecer líneas base de selección aleatoria (*random guessing*) para la comparación de los modelos de predicción de esfuerzo evaluados. Finalmente, Lavazza & Morasca (2017) proponen un marco de trabajo para la construcción de indicadores de exactitud sobre modelos de predicción de esfuerzo. Estos indicadores son

basados en los cuadrados y los valores absolutos de los residuos de los resultados obtenidos por los modelos.

En el caso de la metodología desarrollada en esta investigación, construimos un procedimiento sistemático y riguroso de evaluación de modelos de estimación de esfuerzo que permite realizar una comparación reduciendo el sesgo entre los distintos modelos simplificados basados en los BFC de los métodos IFPUG FPA y COSMIC FFP.

3.3 MODELOS DE REPRESENTACIÓN FUNCIONAL DEL SOFTWARE

Esta sección discute aspectos relacionados con la construcción de modelos funcionales de representación del software a partir del comportamiento obtenido mediante la ejecución de las funcionalidades del sistema. El Paso 2 del modelo de procesos para métodos de medición del tamaño funcional (Abran, 2010) requiere la construcción de un modelo funcional del software que se genera, en nuestro caso, a partir de los artefactos de código fuente de la aplicación. A partir de la adquisición de los artefactos de software, se construye el modelo funcional que se mapea a los componentes de un método de medición mediante las reglas de mapeo. Una vez mapeados se realiza la asignación numérica para obtener el tamaño funcional de la aplicación. Este modelo funcional debe describir la estructura de una aplicación de software detallando sus funcionalidades. Un modelo funcional describe cómo una aplicación de software cambia de un estado a otro dependiendo de las acciones que son ejecutadas, lo que permite representar las funciones y los flujos de datos. Este modela objetos o entidades y las relaciones entre sí que tienen un objetivo definido. En el caso del software, la especificación funcional de una aplicación es independiente de su tecnología de implementación (Mayer, Painter & de Witte, 1994; Menzel & Mayer, 1998).

Para identificar literatura relevante, se realiza una búsqueda exploratoria que permite obtener literatura sobre modelos de representación del comportamiento del software extraídos a partir del código fuente de la aplicación. La cadena de búsqueda utilizada en las bases de datos automatizadas se compone de los siguientes términos: ("*software*") AND

("graph" OR "tree") AND ("source code") AND ("model") AND ("metric*"). A partir de la identificación de los artículos se seleccionan los relevantes para el análisis basado en un conjunto de criterios de inclusión y exclusión que permiten seleccionar estudios sobre representación en grafos de modelos de comportamiento de aplicaciones de software basado en el análisis dinámico del código fuente. La biblioteca utilizada fue la base de datos referencial Scopus. A continuación se presentan los resultados.

Las técnicas de extracción intentan obtener los modelos a partir del código existente, sin embargo, esta no es una tarea sencilla. Con la combinación del análisis estático y dinámico del código fuente se pueden obtener mejores modelos de representación del software para ser analizados desde diferentes perspectivas (Duarte, Kramer & Uchitel, 2017). En la mayoría de los casos, la construcción de los modelos funcionales a partir del código fuente solo se realiza mediante el análisis estático del código fuente, lo que limita la construcción de los modelos para la representación del software (Duarte, Kramer & Uchitel, 2017). Los modelos de comportamiento extraídos a partir del análisis estático y dinámico del código fuente pueden representar con mayor exactitud una aplicación de software, la información recolectada a partir del seguimiento de las ejecuciones puede detallar la funcionalidad del sistema que genera conocimiento sobre el comportamiento de las aplicaciones representando el software mediante, por ejemplo, máquinas de estados finitos (Duarte, Kramer & Uchitel, 2017).

Duarte, Kramer & Uchitel (2006) introducen el concepto de “contextos” que representan un estado abstracto de una aplicación de software. Cada contexto describe una combinación de ejecuciones en un grafo de control de flujo, un conjunto de valores para el estado del sistema y una pila de llamadas. Los modelos de comportamiento son útiles para el análisis, verificación y pruebas de las aplicaciones de software (Lorenzoli, Mariani & Pezzè, 2008). En nuestro caso, consideramos que estos modelos son un enfoque viable que podemos utilizar para realizar el proceso de medición del tamaño funcional. Por ejemplo, Ferreira & Barros (2011) exploraron la posibilidad de establecer trazabilidad entre los puntos de función de una aplicación y su código fuente. En la misma línea el *Object Manage-*

ment Group (2014) con la especificación de la especificación *Automated Function Points* provee una guía en la que establecen las relaciones que pueden existir entre el código fuente y los componentes del conteo del tamaño funcional.

Estudios realizados han propuesto que la extracción de los modelos se puede realizar por medio de enfoques de descubrimiento de procesos basados en eventos (Cook & Wolf, 1998), análisis mecánicos a partir del código fuente que puedan representar con exactitud el comportamiento del software (Holzmann & Smith, 1999), generación de modelos de relaciones a partir del seguimiento de las iteraciones de los componentes y los datos (Lorenzoli, Mariani & Pezzè, 2008) y finalmente, a partir del seguimiento de la ejecución de los sistemas de software (Bodhuin, Pagnozzi, Santone, Tortorella & Villani, 2009). Todo lo anterior, para lograr la implementación de análisis basados en contexto donde existe una trazabilidad entre la información recopilada en tiempo de ejecución y el código fuente que mantiene las conexiones implícitas entre la información que se representa en el modelo (Cito, Oliveira, Leitner, Nagpurkar & Gall, 2017).

En el caso de nuestra investigación, construimos un modelo funcional de representación del software. A partir de un modelo obtenido desde una perspectiva dinámica del software, el procedimiento de medición de la metodología propuesta realiza el proceso de conteo de tamaño funcional, manteniendo trazabilidad entre los requerimientos del usuario final y el código fuente de la aplicación. Además de poder visualizar los modelos funcionales desde la perspectiva del usuario final. Este modelo basado en teorías de grafos y máquinas de estado finito representa los flujos de información transaccional de las aplicaciones. El enfoque para la construcción de los modelos funcionales en nuestra propuesta es el del seguimiento y rastreo de ejecuciones de la aplicación de software las cuales se pueden representar mediante grafos dirigidos (Walkinshaw, Taylor & Derrick, 2016). Según Frappier (1999), los lenguajes de máquinas de estados y las especificaciones de trazabilidad (*trace-based languages*) son los más apropiados para el conteo de puntos de función.

3.4 ESTUDIOS SECUNDARIOS RELACIONADOS CON LA FORMALIZACIÓN DE LA MEDICIÓN DEL TAMAÑO FUNCIONAL

Esta sección discute distintos estudios secundarios que han analizado métodos, procedimientos y herramientas para apoyar el proceso de medición del tamaño funcional del software. El análisis de estos estudios secundarios complementa los estudios primarios del mapeo y la revisión de literatura que se presenta en el Capítulo 4 de este documento de tesis. Estos estudios se analizan con el objetivo de identificar oportunidades de mejora en el área de la medición, que incluyen: la formalización de las reglas de mapeo y la medición de diferentes artefactos de software, la posibilidad de automatizar parcial o totalmente los procedimientos de medición en diferentes dominios funcionales y diferentes contextos de desarrollo de software. También, se discuten las posibilidades de mejora en la productividad de la medición y la exactitud, así como en la repetibilidad de los resultados de las mediciones.

Los estudios secundarios que se describen son identificados realizando una búsqueda exploratoria que permite obtener artículos relevantes sobre la formalización de la medición del tamaño funcional. La cadena de búsqueda utilizada se compone de los términos siguientes: (*"software"* OR *"web"* OR *"system"* OR *"application"*) AND (*"function point"* OR *"functional size"* OR *"function * measurement"*) AND (*"automat"* OR *"systematic"* OR *"procedure"* OR *"tool"* OR *"mapping"* OR *"rule"* OR *"formal * representation"*). A partir de la identificación de los artículos se seleccionan los relevantes mediante la aplicación de un conjunto de criterios de inclusión y exclusión seleccionando estudios secundarios sobre la formalización y automatización de la medición del tamaño funcional. La búsqueda se realizó en la base de datos Scopus. Posteriormente, se realiza un proceso de búsqueda en bola de nieve para identificar estudios relevantes adicionales. La Tabla 3-1 lista el autor, el año de publicación, el enfoque del estudio y el tipo de estudio en orden cronológico. A continuación se describen los estudios secundarios en orden cronológico.

Frappier (1999) analiza diferentes lenguajes de formalización de sistemas de software para evaluar qué tan adecuados son para la especificación de los sistemas de información y para la automatización de la medición del tamaño funcional. El autor determina

que para la formalización de puntos de función, los lenguajes de máquinas de estados son los más apropiados. Además, si se considera la representación de los sistemas de información, las especificaciones de trazabilidad (*trace-based languages*) son buenas candidatas para la definición y el conteo de puntos de función. Finalmente, recomienda formalizar el método de puntos de función usando lenguajes de máquinas de estados y definir reglas para estimar el tamaño funcional a partir de las especificaciones de trazabilidad.

Abran, Meli & Symons (2004) presentan el estado del arte con respecto al método de medición del tamaño funcional COSMIC ISO/IEC 19761, y sus perspectivas futuras. Los autores determinan que algunas de las principales preocupaciones en la industria del software son el desempeño de la medición y la evaluación comparativa (*benchmarking*) de las mediciones. También, establecen que la estimación aún es un reto para las organizaciones de software, e indican que si la estimación se realiza con éxito, representa una ventaja competitiva para las organizaciones de desarrollo de software. Finalmente, alertan que muy pocas empresas han adoptado programas de medición y son muchas menos las que comparten estándares de medición. Bajo este panorama, la recolección de datos es costosa y se realiza de forma manual.

Stambollian & Abran (2006) presentan un marco de trabajo de referencia con el conjunto básico de las funciones de interés para los profesionales que implementan estándares de medición del tamaño funcional. Además, incluyen una recopilación de las herramientas de medición basadas en el método COSMIC FFP, que están disponibles en la industria y en la comunidad académica. Finalmente, analizan las oportunidades de mejora para estas herramientas y determinan que ninguna de las herramientas automatiza totalmente la medición del tamaño funcional. Los autores establecen que se deben realizar evaluaciones de las herramientas para establecer su efectividad y determinan que la automatización de la medición, debe considerar diferentes artefactos de entrada, sin la necesidad de la intervención de un experto para realizar la medición automatizada.

Jorgensen & Shepperd (2007) realizan una revisión sistemática de literatura en el área de la estimación de software que incluye un análisis de los métodos de estimación de tamaño funcional y de su efectividad para la estimación de esfuerzo del desarrollo. Los

autores determinan que muchos de los estudios basados en métodos de medición del tamaño funcional se evalúan analizando su efectividad para la construcción de los modelos de estimación de esfuerzo. También sugieren la necesidad de más estudios empíricos sobre la efectividad de los métodos de estimación, donde los investigadores detallen el contexto y las características de las bases de datos de proyectos históricos que los modelos utilizan y cómo estas características impactan los resultados de la evaluación. Los autores establecen que los métodos formales de estimación aún no superan la efectividad de los métodos de estimación basados en el juicio experto.

Marín, Giachetti & Pastor (2008) realizan un análisis de los métodos de estimación de tamaño funcional que utilizan modelos conceptuales como artefactos de entrada para la medición con el método COSMIC FFP. Los autores identifican lecciones aprendidas y realizan una clasificación de los procedimientos de medición. Los autores identifican la necesidad de que los procedimientos de medición se basen en las últimas versiones de los métodos y determinan que se requiere la formalización de los artefactos de entrada para representar los requerimientos funcionales. Además, indican que los procedimientos de medición deben establecer claramente las fases de la estrategia de medición, las reglas de mapeo entre los conceptos del método de medición y los artefactos de entrada, y las reglas de identificación de los elementos funcionales del método de medición. Asimismo, establecen que la automatización es necesaria para reducir el costo del conteo del tamaño funcional e incrementar la eficiencia del proceso de medición. Finalmente, indican que se deben realizar validaciones empíricas de los procedimientos de medición, para asegurar la calidad de los resultados obtenidos.

Gencel & Demirors (2008) evalúan e identifican los métodos de medición del tamaño funcional y sus oportunidades de mejora. Los autores evalúan los estudios empíricos en medición y realizan una comparación con datos de la industria publicados por el ISBSG (*The International Software Benchmarking Standards Group*). Los autores plantean que la investigación en el área se debe direccionar hacia la evaluación de la significancia de las reglas del proceso de medición, el análisis del nivel de abstracción de las entidades de medición y la determinación de los efectos de los niveles de granularidad de los atributos

medidos. Además, se deben analizar los diferentes métodos de medición que existen y su convertibilidad, para determinar cuáles son mejores y su aplicabilidad en los diferentes dominios funcionales y contextos de desarrollo. Los autores identifican la necesidad de desarrollar herramientas para la automatización parcial o total del proceso de conteo de tamaño funcional donde la precisión y la exactitud de las mediciones deben ser empíricamente validadas.

Ozkan & Demirors (2009; 2010) analizan propuestas de formalización para la medición del tamaño funcional del software. Determinan que estas propuestas buscan resolver las ambigüedades de los métodos de medición existentes, sus conceptos y sus reglas de aplicación, e intentan reducir la variabilidad que genera la interpretación de las reglas generales de estos métodos de medición. Los autores determinan que dada la formalización de los métodos de medición a partir de la aplicación de reglas detalladas a especificaciones formales del software, es posible explorar las posibilidades de automatización. La automatización de los procesos de los métodos de medición puede ser parcial o total, y se debe realizar a partir de la interpretación consistente de las reglas de medición, para garantizar la consistencia de los resultados de medición. Los autores establecen que las definiciones formales de los métodos de medición pueden permitir comparar, transparentemente, las mediciones del tamaño funcional entre los diferentes métodos.

Barkallah, Gherbi & Abran (2011) proponen un marco de trabajo para la automatización de los procedimientos de medición basados en COSMIC FFP, y que utilizan modelos UML como artefactos de entrada para los procedimientos. Los autores determinan que existe poca evidencia sobre la validación de las herramientas existentes que proponen la automatización de la medición del tamaño funcional, y que aún se requieren herramientas para la automatización de la medición. Los autores determinan la necesidad de definir un perfil de modelos UML que formalice la automatización de la medición del tamaño funcional basados en el estándar de medición COSMIC FFP. Este trabajo es una continuación del trabajo realizado por Stambollian & Abran (2006). Por su parte Fehlmann & Kranich (2011) recopilan el trabajo realizado sobre las experiencias realizadas con el método COSMIC calculando el tamaño funcional a partir de diagramas de secuencia UML. Vogelesang, Symons,

Lesterhuis, Meli & Daneva (2013) y Vogelesang (2005) presentan los lineamientos para realizar aproximaciones de la medición del tamaño funcional basado en el método COSMIC. Los autores detallan distintos métodos de aproximación, sus ventajas y desventajas, y las recomendaciones para su aplicación. La guía contempla los métodos de aproximación: *average functional process approximation*, *fixed size classification approximation*, *equal size bands approximation*, *average use case approximation*, y *early & quick COSMIC approximation*. El objetivo principal de estos métodos es obtener el tamaño funcional en etapas tempranas del desarrollo del software mediante una aproximación al tamaño real de las aplicaciones y reduciendo el tiempo y el esfuerzo del conteo.

Bajwa, Gencel & Abrahamsson (2014) realizan un mapeo sistemático para identificar los métodos de medición del tamaño del software incluyendo los de tamaño funcional. Los autores determinan las entidades de medición, el tipo de atributo medido, las medidas de tamaño utilizadas, la validación teórica y empírica realizada, y la aplicabilidad para la metodología de desarrollo utilizada. En su estudio, determinan que la mayoría de los estudios proponen métodos para la medición del tamaño funcional y pocos estudian las medidas físicas del software. También, identifican métodos de medición ajustados a diferentes contextos de desarrollo basados en los métodos más utilizados en la industria, y nuevas propuestas de métodos de medición. Según los autores, la mayoría de estudios analizan métodos existentes de tamaño funcional, lo cual corrobora la aceptación de estas métricas de tamaño en la comunidad académica.

Por su parte, De Souza & De Aquino (2014) discuten la aplicabilidad de los métodos formales de medición de software para la estimación del esfuerzo de desarrollo de aplicaciones móviles. Abdullah & Rusli (2015b) identifican y discuten literatura relacionada con métodos formales de estimación de tamaño funcional a partir del modelado UML para representar los requerimientos de aplicaciones móviles. Este estudio analiza la medición del tamaño funcional en términos del proceso de medición y las reglas de mapeo. Para ello analizan literatura sobre medición funcional para aplicaciones móviles y para modelos UML con el fin de proveer lineamientos para la medición del tamaño funcional de estas aplicaciones utilizando modelos UML.

Finalmente, el autor, el director de esta tesis y un asesor externo, realizan un mapeo sistemático de literatura para identificar y caracterizar los procedimientos de medición del tamaño funcional basados en los métodos estandarizados de medición, así como analizar sus limitaciones (Quesada-López & Jenkins, 2017b; Quesada-López, Jenkins & Travassos, 2018). El detalle de este estudio se presenta en el Capítulo 4 de este documento.

Todo lo anterior corrobora la necesidad de la formalización y automatización de los procesos de medición del tamaño funcional a partir de modelos de representación funcional de software bien definidos para los principales métodos de medición del tamaño funcional IFPUG FPA y COSMIC FFP. Los procedimientos de automatización de la medición del tamaño funcional deben apoyar a los administradores de proyectos e ingenieros de software para obtener el tamaño funcional permitiendo una mejora en el tiempo y el costo del proceso de conteo y proveyendo resultados de medición con la exactitud y consistencia requerida por la industria. Para esto es necesario estudiar la factibilidad de implementar estos procedimientos en contextos específicos de desarrollo en las organizaciones de software.

Tabla 3-1 Estudios secundarios sobre medición del tamaño funcional.

Referencia	Enfoque	Tipo
(Frappier, 1999)	Analiza los lenguajes de formalización para evaluar qué tan adecuados son para la especificación de sistemas de información y la automatización de la medición del tamaño funcional.	SVY
(Abran, Meli & Symons, 2004)	Presentan el estado del arte del método de medición del tamaño funcional COSMIC ISO/IEC 19761 y definen las perspectivas futuras.	SVY
(Stambollian & Abran, 2006)	Presentan un marco de trabajo con el conjunto de funciones de interés para implementar estándares de medición del tamaño funcional e incluyen una recopilación de herramientas COSMIC.	SVY
(Jorgensen & Shepperd, 2007)	Analizan los métodos de estimación del tamaño funcional y su efectividad para la estimación del esfuerzo de desarrollo del software. Realizan un análisis de la evidencia empírica existente para definir las líneas de investigación en el área.	SLR
(Marín, Giachetti & Pastor, 2008)	Realizan un análisis de los métodos de estimación del tamaño funcional que se basan en COSMIC FFP y que utilizan modelos conceptuales como artefactos de entrada.	SVY
(Gencel & Demirors, 2008)	Evalúan e identifican métodos de medición del tamaño funcional y las oportunidades de mejora en el área.	SVY
(Ozkan & Demirors, 2009; 2010)	Analizan propuestas de formalización para la medición del tamaño funcional del software y analizan las posibilidades de automatización de los procedimientos de medición y las oportunidades de mejora.	SVY
(Barkallah, Gherbi & Abran, 2011)	Analizan procedimientos de medición del tamaño funcional que se basa en el método COSMIC FFP y utiliza modelos UML como entrada para el proceso de medición.	SVY
(Fehlmann & Kranich, 2011)	Recopilan el trabajo realizado sobre las experiencias realizadas con el método COSMIC calculando el tamaño funcional a partir de diagramas de secuencia UML.	SVY
(Vogelezang, Symons, Meli & Daneva, 2013; Vogelezang, 2005)	Presentan los lineamientos para realizar aproximaciones de la medición del tamaño funcional basado en el método COSMIC. Los autores detallan distintos métodos de aproximación, sus ventajas, desventajas y las recomendaciones para su aplicación.	SVY
(Bajwa, Gencel & Abrahamsson, 2014)	Identifican métodos de medición del tamaño de software, las entidades medidas y la validación empírica.	SMR

(continúa)

Referencia	Enfoque	Tipo
(De Souza & de Aquino, 2014)	Discuten la aplicabilidad de los métodos formales de medición de software para la estimación del esfuerzo de desarrollo de aplicaciones móviles.	SVY
(Abdullah & Rusli, 2015b)	Identifican y discuten literatura relacionada con métodos formales de estimación del tamaño funcional a partir del modelado UML para representar los requerimientos de aplicaciones móviles.	SVY
(Quesada-López & Jenkins, 2017b)	Identifican y caracterizan procedimientos de medición del tamaño funcional basados en los métodos estandarizados de medición y analizan sus limitaciones. Analizan los procesos de mapeo de componentes hacia los métodos de medición, discuten las características y el nivel de automatización de las herramientas de medición del tamaño funcional e identifican la investigación empírica realizada en el área.	SMR
(Quesada-López, Jenkins & Travassos, 2018)	Analizan los procesos de mapeo de componentes hacia los métodos de medición, discuten las características y el nivel de automatización de las herramientas de medición del tamaño funcional e identifican la investigación empírica realizada en el área.	SLR

Capítulo 4. PROCEDIMIENTOS DE MEDICIÓN DEL TAMAÑO FUNCIONAL: UN MAPEO Y REVISIÓN SISTEMÁTICA DE LITERATURA

En este capítulo, se agrega conocimiento para la consecución del objetivo específico 1 (OE1) que permite responder parte de la primera pregunta de investigación (KP1) de la tesis.

El estudio de la literatura científica permite recolectar el conocimiento existente sobre el tema de investigación seleccionado, demostrar la importancia del área de investigación que le compete y justificar la relevancia del problema de investigación.

Este estudio representa el producto del ciclo de investigación 2 (RC2).

Este capítulo presenta el estado del arte en el área de investigación. Para esto se realiza un mapeo y revisión sistemática de la literatura que permite caracterizar los enfoques de medición del tamaño funcional existentes. El análisis se realiza para obtener el conocimiento necesario que apoye el proceso del diseño del procedimiento de medición y verificación de la metodología propuesta en el Capítulo 5 y la definición de los estudios empíricos que lo evalúan en el Capítulo 6.

En nuestro caso, el mapeo y la revisión sistemática de literatura permiten identificar, clasificar, caracterizar y analizar los procedimientos de medición del tamaño funcional del software y determinar el nivel de automatización que han alcanzado estas propuestas. Asimismo, se determinan los desafíos, oportunidades, fortalezas y debilidades de la automatización de los procedimientos de medición del tamaño funcional y se obtienen conocimientos acerca de las mejores prácticas y los aspectos a considerar en la evaluación empírica de los procedimientos de medición.

El objetivo de este estudio es identificar y caracterizar literatura existente sobre los procedimientos de medición del tamaño funcional. El análisis incluye consideraciones para mitigar algunas de las limitaciones de la investigación en el área y la síntesis de los principales resultados obtenidos a partir de la evidencia existente (Quesada-López & Jenkins, 2017b; Quesada-López, Jenkins & Travassos, 2018). Como resultado se obtiene una caracterización de los estudios del área, se identifican tendencias de investigación, se determina el nivel de automatización que han alcanzado los procedimientos de medición del tamaño funcional y se analiza la evidencia empírica sobre la evaluación de los procedimientos. Los estudios se mapean principalmente de acuerdo con el método de medición, los artefactos de entrada, procedimiento de medición, su nivel de automatización, el contexto, el método de investigación, la validación empírica y las amenazas a la validez reportadas.

A continuación se describe el diseño, ejecución y resultados de un mapeo y revisión sistemática de literatura. La Sección 4.1 presenta el resumen del diseño del mapeo y revisión sistemática de literatura y la Sección 4.2 discute los resultados del estudio.

4.1. DISEÑO DEL ESTUDIO

Para llevar a cabo el estudio, primero se realiza el mapeo sistemático que caracteriza la literatura y luego se analizan las preguntas de investigación de la revisión sistemática. El mapeo y revisión sistemática de literatura siguen los lineamientos para la ejecución de estudios secundarios propuestos por (Kitchenham, 2004; Biolchini, Mian, Natali & Travassos, 2005; Kitchenham & Charters, 2007; Kitchenham, Budgen & Brereton, 2015; Petersen, Vakkalanka & Kuzniarz, 2015) y consideran las recomendaciones descritas en (Petersen, Feldt, Mujtaba & Mattsson, 2008; Kitchenham, Budgen & Brereton, 2011; Wohlin, 2014; Badampudi, Wohlin & Petersen, 2015; Mourão, Kalinowski, Murta, Mendes & Wohlin, 2017).

El diseño del estudio describe los pasos necesarios para la ejecución de cada una de las etapas del proceso de mapeo y revisión sistemática las cuales comprenden: la definición del objetivo y las preguntas de investigación, la definición de la estrategia de búsqueda, el

proceso de selección de estudios primarios y sus criterios de inclusión y exclusión, la evaluación de la calidad de los artículos seleccionados, la extracción de los datos y el proceso de clasificación y análisis de la evidencia. En el Anexo 1 se detalla el protocolo del diseño de este estudio el cual resumimos a continuación.

Se establecen las preguntas de investigación para guiar diseño y el análisis durante la conducción del estudio. Las preguntas de mapeo se responden a través de la clasificación de los estudios primarios seleccionados, mientras que las preguntas de revisión se responden mediante el análisis y síntesis de la evidencia. Las preguntas de investigación son las siguientes:

RQ1. ¿Cuál es el estado actual de la investigación en el área y las características de los procedimientos de medición del tamaño funcional del software que se encuentran en la literatura científica?

Respondiendo a esta pregunta de mapeo es posible estructurar el conocimiento en el área identificando: (1) la frecuencia de los trabajos publicados a través del tiempo que muestra las tendencias de publicación de acuerdo a las diferentes categorías de clasificación de los artículos seleccionados, (2) los principales foros de publicación donde es posible encontrar publicaciones relevantes en el área y en las cuales se pueden publicar este tipo de trabajos, (3) los investigadores más activos y sus intereses de investigación de acuerdo a las diferentes categorías de clasificación de los artículos seleccionados y (4) los principales enfoques de la investigación realizada en el área.

RQ2. ¿Cómo han sido evaluados los procedimientos de medición del tamaño funcional?

Respondiendo a esta pregunta de mapeo e investigación es posible clasificar y analizar: (1) los tipos de preguntas de investigación e hipótesis que han sido investigadas empíricamente, (2) el tipo de evaluación realizada para los procedimientos de medición del

tamaño funcional y los diseños empíricos para evaluar los procedimientos, (3) los constructos, métricas y mediciones que han sido utilizadas para evaluar los procedimientos de medición del tamaño funcional que permitan demostrar su efectividad y que habiliten la posibilidad de comparación de resultados en contextos similares, (4) el contexto de la investigación empírica realizada para validar los procedimientos de medición del tamaño funcional tales como los participantes, los objetos de estudio y los entornos de desarrollo, (5) las amenazas a la validez que se han reportado en el diseño y evaluación empíricos de los procedimientos y (6) los aspectos que se deben considerar para realizar replicaciones para la evaluación procedimientos de medición.

RQ3. ¿Cuáles son los desafíos y oportunidades de los procedimientos existentes para la medición del tamaño funcional?

Respondiendo a esta pregunta de investigación es posible determinar: (1) los principales desafíos y oportunidades en el área de los procedimientos de medición del tamaño funcional, (2) las posibles líneas de investigación futura y (3) las recomendaciones que se pueden realizar con respecto a la investigación que se realiza en el área de los procedimientos de medición con el fin de mejorar cómo se reportan los resultados de estos estudios.

La aplicación de la estrategia de búsqueda, el proceso de selección de estudios primarios y sus criterios de inclusión y exclusión, la evaluación de la calidad de los artículos seleccionados, la extracción de los datos y el proceso de clasificación y análisis de la evidencia se realiza de acuerdo al diseño del protocolo presentado en el Anexo 1. La información extraída para cada artículo primario es la siguiente:

- *Caracterización (RQ1)*: extrae la información para el mapeo de los estudios y su caracterización que incluye el identificador, título, año, autores, foro de publicación, enfoque, método, artefacto de entrada, etapa del ciclo de vida, tipo de artículo, procedimiento de

medición, reglas de mapeo, reglas de asignación numérica, marco de trabajo utilizado, las herramientas que lo soportan y nivel de automatización.

- *Evaluación (RQ2)*: extrae la información sobre la evaluación de los procedimientos que incluye los objetivos, tipos de preguntas de investigación e hipótesis, el método de investigación empírica, tipo de evaluación, participantes, tamaño de la muestra, instrumentación, los constructos, métricas y mediciones, factores de contexto y dominio funcional, las amenazas a la validez interna y externa, constructo, conclusión y tipo de replicación.
- *Desafíos y oportunidades (RQ3)*: extrae la información de los desafíos, oportunidades, líneas de investigación futura y recomendaciones.
- *Calidad*: detalla la evaluación de la calidad de los artículos.

La Figura 4-1 detalla el esquema de clasificación de artículos primarios que considera las categorías utilizadas para este estudio. Todos los artículos primarios con clasificados de acuerdo a este esquema y posteriormente analizados y sintetizados para responder cada uno de las preguntas de investigación. A continuación describimos las categorías del esquema de clasificación:

- *Método de medición del tamaño funcional*: la propuesta se basa en los métodos de medición IFPUG FPA, COSMIC FFP, MkII FPA, NESMA y FiSMA.
- *Tipo de medición*: la propuesta presenta un procedimiento para la aproximación, estimación o medición del tamaño funcional.
- *Dominio funcional*: el procedimiento aplica para sistemas de información, sistemas en tiempo real o software científico.
- *Atributo medido*: mide el producto, proceso o proyecto.
- *Etapa del ciclo de vida*: los artefactos utilizados como entrada para el procedimiento se generan en la etapa del ciclo de vida de desarrollo requerimientos, diseño, codificación, pruebas o producción/mantenimiento.

- *Artefacto de entrada*: el tipo de artefacto de entrada que se mide durante la aplicación del procedimiento es una especificación de requerimientos en lenguaje natural, modelo de negocios, modelo de requerimientos, lenguajes formales de especificación de requerimientos, diseño de interfaz de usuario, modelos de objetos, diagramas ER, diagramas UML, modelos de diseño no UML, modelos conceptuales específicos, código fuente, casos de pruebas, sistema de software en ejecución, entre otros.
- *Reglas del procedimiento de medición*: se verifica si se realiza la especificación de las reglas de construcción del metamodelo, las reglas del mapeo entre el artefacto de entrada y los componentes del método de medición y las reglas de asignación numérica.
- *Análisis*: el análisis sobre los artefactos de entrada al procedimiento se realiza estático, dinámico o ambas.
- *Evaluación*: el procedimiento presenta una evaluación teórica y/o empírica.
- *Tipo de artículo*: se clasifica el estudio primario como una propuesta, modelo, herramienta o estudio empírico.
- *Estudio empírico*: un estudio empírico se clasifica como análisis crítico, reporte de experiencia, prueba de concepto, encuesta, caso de estudio, cuasi-experimento o experimento.
- *Validez del estudio*: se identifica si los estudios empíricos detallan las amenazas a la validez interna, del constructo, conclusión y externa.
- *Contexto experimental*: se identifica si el estudio utiliza objetos de la industria o de ambientes controlados en laboratorio. Asimismo, si los participantes del estudio son estudiantes y/o profesionales.
- *Ambiente*: el estudio se realiza en un ambiente académico o industrial.
- *Métricas de evaluación*: se clasifican las métricas de efectividad (exactitud, repetibilidad, reproducibilidad, productividad, defectos) y las métricas de adopción (facilidad de uso percibida, utilidad percibida, intención de uso futuro, satisfacción).
- *Factores de contexto*: se identifica si el estudio se realiza para ambientes y software orientados a objetos, aplicaciones web, aplicaciones móviles, bajo metodologías ágiles, desarrollo dirigido por modelos u otros.

- *Aprovechamiento en modelos de predicción:* se determina el uso de los resultados de medición del tamaño funcional en modelos de predicción de esfuerzo, de tamaño del software, de memoria utilizada por el software, calidad, otros.
- *Nivel de automatización:* se determina si el procedimiento se aplica de forma manual, si existe una intención de automatización, si existe una herramienta de soporte para el proceso de conteo. Asimismo, se determina si es una herramienta básica de soporte al conteo (por ejemplo, una herramienta de registro de datos), si es una herramienta que permite hacer una automatización parcial o si es totalmente automatizada.

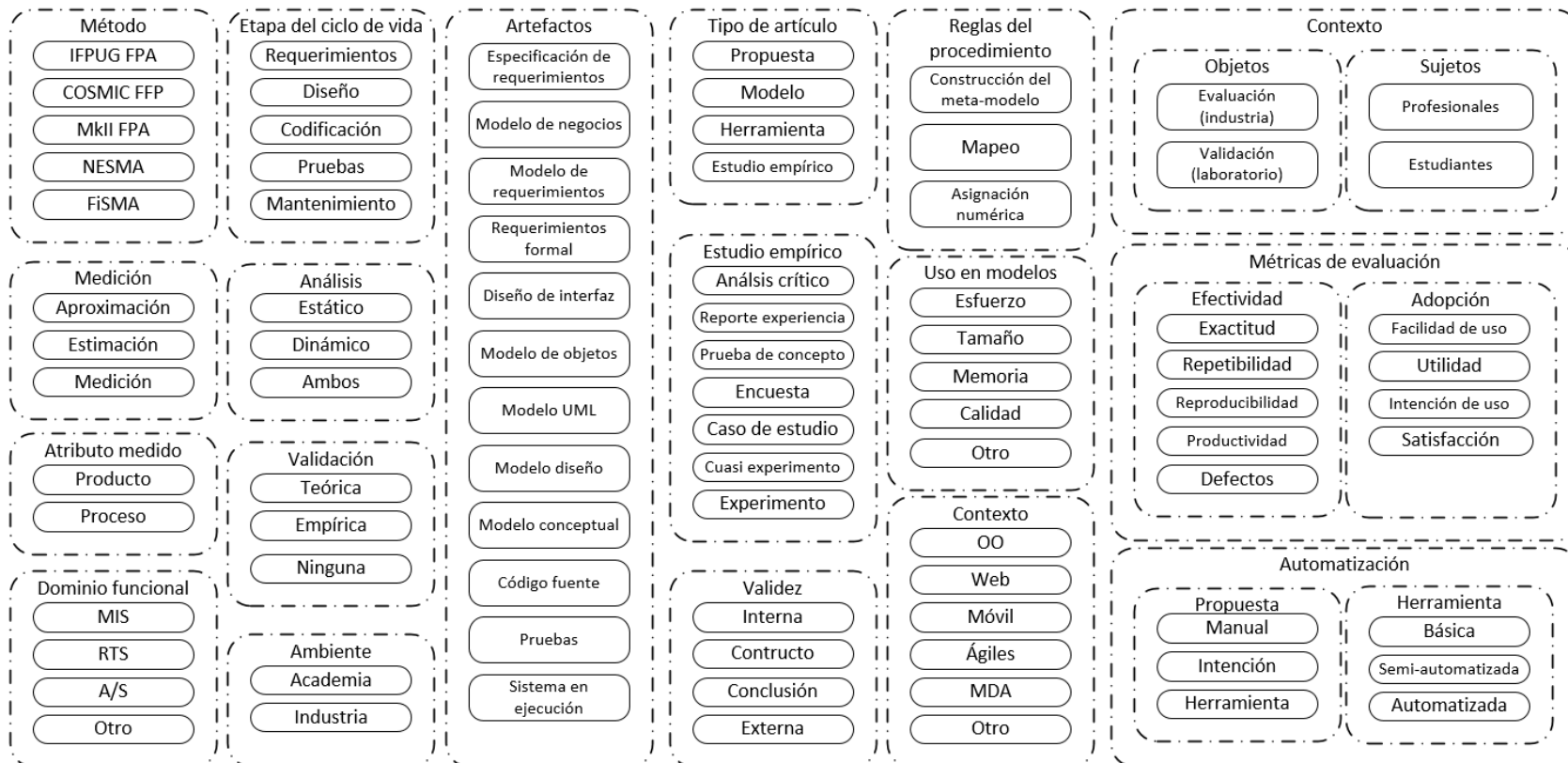


Figura 4-1 Esquema de clasificación.

4.2. RESULTADOS DEL ESTUDIO

En esta sección se presentan los resultados del mapeo y de la revisión sistemática de literatura que responden cada pregunta de investigación (RQ). Todos los artículos identificados en este estudio son contabilizados por separado. La Tabla 4-1 muestra el número de resultados obtenidos en cada base de datos (n), los estudios incluidos después de realizar el proceso de inclusión y exclusión (I/E) y finalmente, los artículos de control identificados en cada base de datos. La mayoría de los artículos seleccionados durante la etapa de I/E están indexados en *SCOPUS* (114), seguido de *Engineering Village* (53), *IEEE Xplore* (34) y *Web of Science* (24). La Figura 4-2 muestra el total de artículos encontrados en las bases de datos automatizadas y la distribución por cada una de ellas. Algunos de los estudios fueron encontrados en más de una base de datos.

Tabla 4-1 Cantidad de artículos retornados por base de datos.

Base de datos	n	I/E	Control
<i>SCOPUS</i>	657	110	11
<i>IEEE Xplore</i>	92	32	3
<i>ISI Web of Science</i>	196	23	4
<i>Engineering Village Compendex</i>	128	53	8

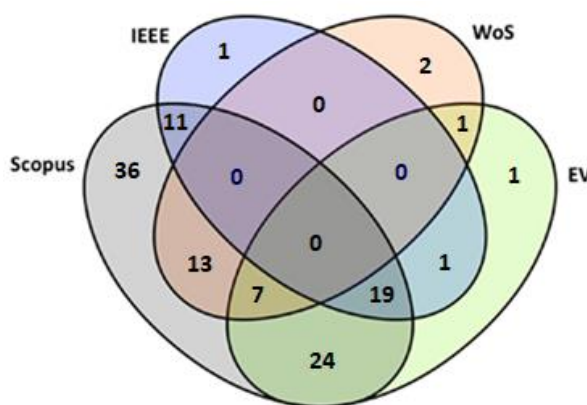


Figura 4-2 Artículos primarios en las bases de datos de búsqueda.

En total, 180 estudios primarios relevantes son identificados y analizados. En la búsqueda en bases de datos automatizadas se identifican 116 artículos que incluyen los 11 artículos de control. Como resultado del proceso de bola de nieve se identifican 64 artículos relevantes: 37 hacia atrás y 27 hacia adelante.

La Figura 4-3 muestra los resultados de cada etapa del proceso de identificación y de selección de estudios. A partir de los estudios identificados, cada pregunta de investigación se analiza en las siguientes subsecciones.

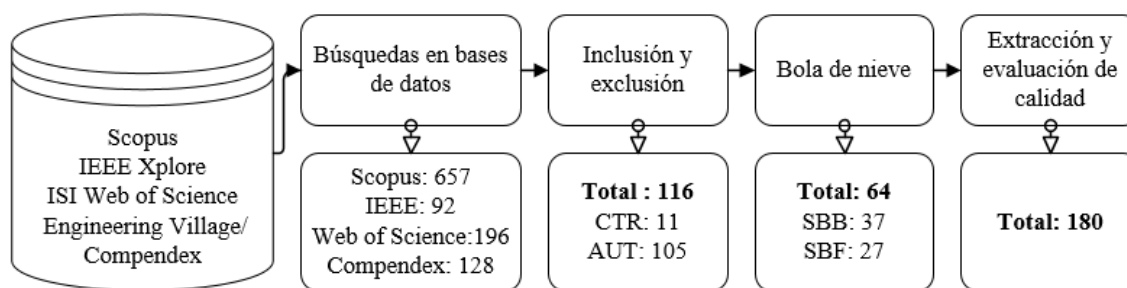


Figura 4-3 Resultados del proceso de selección de estudios primarios.

Se utiliza Mendeley como herramienta de administración de referencias y MS Excel para los procesos de selección, evaluación y extracción de datos.

El listado completo de publicaciones y su clasificación se encuentra en orden cronológico en el Anexo 2. A cada artículo se le asigna un código de referencia única para el análisis. A los 180 estudios primarios se les asigna el consecutivo a partir de las tres primeras letras del primer apellido del autor y dos dígitos del año, en el caso de duplicados se le asignan una letra adicional. Estos códigos se utilizan a lo largo de este capítulo para referir cada uno de los artículos. El Anexo 2 muestra cada una de las referencias de los artículos, su código único y el detalle de la evaluación de calidad.

La Figura 4-4 muestra los valores de evaluación de calidad para los artículos incluidos, la cual varía entre 0 y 17. El promedio de calidad es de un 10.64 (63%), con un 75% de los estudios con una calidad superior al 53% y el 25% de los estudios con una calidad superior al 76%. Los estudios presentan menores calificaciones de calidad en los ítems relacionados con el detalle de las amenazas a la validez, el diseño experimental y el análisis de

resultados. En el caso de las publicaciones en revistas, el promedio de calidad es de un 12.58 (74%) y en el caso de las publicaciones en conferencias o talleres es del 10.18 (60%). Un puntaje de calidad mayor indica que el reporte provee un mayor nivel de detalle y no debe asociarse con la calidad del estudio, ni la calidad del foro de publicación.

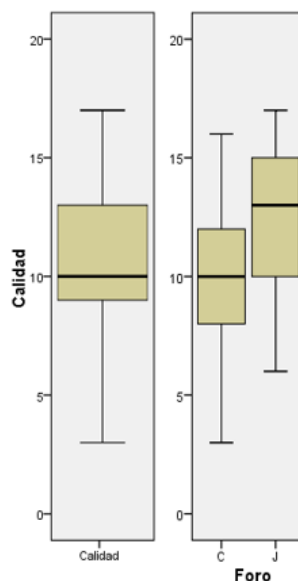


Figura 4-4 Calidad de los artículos primarios incluidos.

4.2.1. ESTADO ACTUAL DE LA INVESTIGACIÓN EN EL ÁREA Y CARACTERÍSTICAS DE LOS PROCEDIMIENTOS DE MEDICIÓN DEL TAMAÑO FUNCIONAL

RQ1. *¿Cuál es el estado actual de la investigación en el área y las características de los procedimientos de medición del tamaño funcional del software que se encuentran en la literatura científica?*

En esta sección se estructura el conocimiento en el área identificando la frecuencia de los trabajos publicados a través del tiempo. Se identifican los principales foros de publicación donde es posible encontrar publicaciones relevantes y en las cuales se pueden publicar este tipo de trabajos, los investigadores más activos y sus intereses de investigación,

las tendencias de publicación de acuerdo a las diferentes categorías de clasificación de los artículos seleccionados y los principales enfoques de la investigación realizada en el área.

4.2.1.1. FRECUENCIA DE LAS PUBLICACIONES

La Tabla 4-2 muestra el número de estudios entre los años 1991 y 2017. Los resultados presentan la tendencia de los estudios en el área de la medición del tamaño funcional, específicamente, los relacionados con los procedimientos de medición del tamaño funcional. En total se mapearon 180 artículos. Entre los años 2008 y 2015 se publicaron la mayor cantidad de artículos con 95 (53%). Del año 2003 al 2016 se acumula la mayor cantidad de artículos con 146 (81%). La tendencia de las publicaciones muestra un incremento sostenido a través de los años en los temas estudiados.

Tabla 4-2 Publicaciones por año.

1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total
1	0	1	0	2	0	5	3	5	2	4	4	8	8	8	10	8	14	10	11	11	9	13	13	14	9	7	180

4.2.1.2. FOROS DE PUBLICACIÓN

La Tabla 4-3 muestra la distribución de estudios por foro de publicación a través de los años. Un mayor número de artículos han sido publicados en conferencias o talleres 140 (78%) y en menor medida en revistas 37 (21%). En los años 2006 y 2015 se publicaron la mayor cantidad de artículos de los procedimientos de medición en conferencias (89), y del año 2009 al 2013 se publicaron la mayor cantidad de artículos en revistas (17).

La Tabla 4-4 muestra los principales foros de publicación, para las revistas se muestran las 3 principales y para las conferencias las 5 principales. Los foros se ordenan según el ranking (R) basado en la cantidad de publicaciones. Para cada artículo se muestra el nombre y las siglas del foro, la cantidad de artículos publicados (n) y la cantidad de artículos publicados por método de medición para COSMIC FFP (C), IFPUG FPA (F) y MkII FPA (M).

Las principales conferencias son *Software Measurement and Software Process and Product Measurement* (IWSM-MENSURA, 37 artículos), *Software Metrics y Empirical Software Engineering and Measurement* (ESEM [METRICS], 10 artículos), *Software Measurement European Forum* (SMEF, 6 artículos), *Software Engineering and Advanced Applications* (SEAA), *Quality Software* (QSIC) e *Ibero-American Conference on Software Engineering* (CibSE) con 5 artículos cada uno; finalmente, *Product-Focused Software Process Improvement* (PROFES) con 4 artículos.

Las revistas principales fueron *Information and Software Technology* (IST) con 7 artículos, *Journal of Systems and Software* (JSS) con 3 artículos y *Journal of Data & Knowledge Engineering* (DKE), *Empirical Software Engineering* (ESE), *ACM Transactions on Software Engineering and Methodology* (TOSEM) con 2 artículos cada una.

Estas conferencias y revistas son algunas de las más reconocidas en el área de la ingeniería del software. Por ejemplo, la revista *Information and Software Technology* (Q1) tiene un factor de impacto SJR: 1.391, JCR: 1.046, y 1.612 (5 años). El *Journal of Systems and Software* (Q1) tiene un SJR: 1.381, JCR: 1.352, y 1.485 (5 años); finalmente, el *Empirical Software Engineering* (Q1) tiene un SJR: 1.566, JCR: 2.161, y 2.585 (5 años). Estos resultados indican que los artículos pueden ser publicados en foros de calidad.

Tabla 4-3 Foros de publicación por año.

Foro	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total
J	0	0	1	0	0	0	0	0	1	0	1	0	2	1	3	2	2	0	4	3	3	4	3	1	2	2	2	37
C	0	0	0	0	1	0	4	3	4	2	3	4	6	7	5	8	6	14	6	8	8	5	10	12	12	7	5	140
R	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
T	1	0	1	0	2	0	5	3	5	2	4	4	8	8	8	10	8	14	10	11	11	9	13	13	14	9	7	180

J=Revista, C=Conferencia o taller, R=Reporte técnico

Tabla 4-4 Principales foros de publicación.

R	Revista	n	C	I	M	R	Conferencia	n	C	I	M
1	Journal of Information and Software Technology (IST)	7	5	3	0	1	Software Measurement and Software Process and Product Measurement (IWSM-MENSURA)	37	29	9	0
2	Journal of Systems and Software (JSS)	3	1	3	1	2	Software Metrics (METRICS) & Empirical Software Engineering and Measurement (ESEM)	8	1	7	0
3	Journal of Data & Knowledge Engineering (DKE)	2	2	0	0	3	Software Measurement European Forum (SMEF)	6	2	5	0
3	Empirical Software Engineering (ESE)	2	0	2	0	4	Quality Software (QSIC)	5	4	1	0
3	ACM Transactions on Software Engineering and Methodology (TOSEM)	2	1	1	0	4	Software Engineering and Advanced Applications (SEAA)	5	4	1	0
						5	Ibero-American Conference on Software Engineering (CibSE)	5	1	3	1
						5	Product-Focused Software Process Improvement (PROFES)	4	3	1	0

El Anexo 2 detalla la lista completa de todos los foros de publicación para los artículos seleccionados.

4.2.1.3. INVESTIGADORES EN EL ÁREA

Los autores más activos en el área de investigación son: A. Abran (20 artículos, 15 para COSMIC FFP y 5 para IFPUG FPA), O. Pastor (19 artículos, 15 para COSMIC FFP y 4 para IFPUG FPA), S. Abrahao (13 artículos, 3 para COSMIC FFP y 10 para IFPUG FPA), L. Lavazza (11 artículos, 5 para COSMIC FFP y 6 para IFPUG FPA), N. Condori-Fernandez (10 artículos, 9 para COSMIC FFP y 1 para IFPUG FPA), G. Poels (10 artículos, 4 para COSMIC FFP y 6 para IFPUG FPA) y B. Marín (10 artículos, 9 para COSMIC FFP y 1 para IFPUG FPA). Las frecuencias de publicación se totalizan independientemente de la cantidad y el orden de los autores en un artículo.

Investigadores como Oscar Pastor, Luigi Lavazza y Geert Poels han trabajado con los métodos IFPUG FPA y COSMIC FFP equitativamente, Alain Abran, Nelly Condori-Fernandez y Beatriz Marín principalmente con el método COSMIC FFP pero también con IFPUG FPA y finalmente, Silvia Abrahao principalmente en el método IFPUG FPA pero

también ha trabajado con el método COSMIC FFP. Las principales colaboraciones son internas, en el grupo de investigación de Oscar Pastor, con Silvia Abrahao, y Nelly Condory-Fernández. Además se dan colaboraciones entre este grupo de investigación, Alain Abran y Geert Poels.

4.2.1.4. CARACTERÍSTICAS DE LOS PROCEDIMIENTOS DE MEDICIÓN

La Tabla 4-5 presenta la distribución de los métodos de medición para los estudios identificados. Del total de artículos, 92 (51%) utilizan el método COSMIC FFP, 81 (45%) el método IFPUG FPA, y solo 2 (1%) utilizan el método MkII FPA. En este caso, del total de 180 artículos, dos estudios fueron publicados analizando los métodos IFPUG FPA, COSMIC FFP y MkII FPA simultáneamente y tres artículos para IFPUG FPA y COSMIC FFP.

Según la tendencia mostrada, el área de investigación se encuentra en crecimiento principalmente para el método de medición COSMIC FFP y más constante para el método IFPUG FPA. Los datos muestran que, hasta el año 2007, la mayor cantidad de estudios de los procedimientos de medición se realizaban basados en el método IFPUG FPA, y a partir del 2008, se ha ido incrementado la cantidad de estudios basados en el método COSMIC FFP. Aunque en los últimos años, COSMIC FFP acapara la mayoría de la investigación en el área, estudios basados en el método IFPUG FPA se han continuado realizando de manera constante. La investigación de los procedimientos de medición basados en el método COSMIC FFP ha mantenido una tendencia en crecimiento desde el año 2000. Por su parte, la investigación basada en el método IFPUG FPA se ha mantenido constante hasta el año 2010 donde disminuyó. Para el año 2015 presenta un pequeño repunte de estos estudios.

Tabla 4-5 Publicaciones por año y método de medición del tamaño funcional.

Método	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total
C	0	0	0	0	0	0	0	0	1	0	2	1	5	2	3	5	3	7	5	6	9	7	11	10	9	7	4	97
I	1	0	1	0	2	0	5	3	4	2	2	3	3	5	6	6	5	7	5	6	2	3	2	2	5	3	3	86
M	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	4
T	1	0	1	0	2	0	5	3	5	2	4	4	8	8	8	10	8	14	10	11	11	9	13	13	14	9	7	180

C=COSMIC FFP, I=IFPUG FPA, M=MkII

La Figura 4-5 presenta la línea del tiempo de los estudios primarios de los procedimientos de medición del tamaño funcional. Los estudios que se encuentran sobre la línea de tiempo principal muestran los procedimientos que utilizaron el método de medición del tamaño funcional IFPUG FPA y los estudios bajo la línea de tiempo principal presentan los procedimientos que utilizaron el método COSMIC FFP. Los estudios [BER04, JIN14] utilizaron MkII FPA y los estudios [JBER05, HER06, JIC10, SAN12, OCH16a] utilizaron más de un método de medición del tamaño funcional.

La Tabla 4-6 lista un extracto de los estudios primarios incluidos en este estudio. La tabla con el listado completo con los 180 estudios primarios ordenados cronológicamente se detalla en el Anexo 3. Para cada uno de los estudios se muestran las siguientes columnas:

- *Secuencia (#)*: secuencia del estudio primario.
- *Código de referencia (CR)*: identificador único del estudio primario.
- *Referencia (Ref.)*: referencia bibliográfica del estudio primario.
- *Año (AN)*: año de publicación del estudio primario.
- *Grupo (GR)*: grupo de artículos relacionados que estudiaron propuestas de medición similares.
- *Enfoque (E)*: descripción del enfoque principal del artículo.
- *Origen del artículo (OR)*: indica dónde se identifica el artículo durante el proceso de búsqueda y selección. CTR=artículos de control, AUT=búsqueda automática, SBB=proceso de bola de nieve hacia atrás, SBF=proceso de bola de nieve hacia adelante.

- *Tipo de foro* (TF): indica donde el estudio fue publicado. J=Revista, C=Conferencia, B=Capítulo de libro, R=Reporte técnico.
- *Google hits* (GH): detalla la cantidad de *hits* de cada artículo según *Google Scholar*.
- *Método* (ME): indica cual método de tamaño funcional se utiliza. I=IFPUG FPA, C=COSMIC FFP, M=MKII FPA.
- *Tipo de medición* (TM): tipo de medición que plantea el artículo. A=Aproximación de tamaño (*early estimation*), E=Estimación de tamaño, M=Medición del tamaño.
- *Atributo* (AT): indica el atributo medido. P=Producto, R=Proceso.
- *Artefacto* (Artef.): se identifica el artefacto de entrada utilizado para realizar el proceso de medición.
- *Etapa* (EC): indica la etapa de ciclo de vida de desarrollo donde puede ser aplicado el procedimiento. R=Requerimientos, D=Diseño, C=Codificación, M=Mantenimiento.
- *Tipo de artículo* (TP): indica el tipo de artículo. P=Propuesta, H=Herramienta, E=Empírico.
- *Tipo empírico* (TE): indica el tipo del estudio si es empírico. AC=Análisis crítico, RE=Reporte de experiencia, PR=Propuesta, PC=Prueba de concepto, CS=Caso de estudio, EC=Cuasi experimento o experimento controlado, EX=Experimento.
- *Dominio* (DF): indica el dominio del estudio. MIS=Sistemas de información, RTS= sistemas en tiempo real.
- *Herramienta* (H): indica si el estudio menciona una herramienta o no.
- *Calidad* (Q): indica la calificación de la evaluación de calidad.

El listado completo de publicaciones y su clasificación resultado del proceso de extracción de datos se encuentra disponible en el Anexo 2.

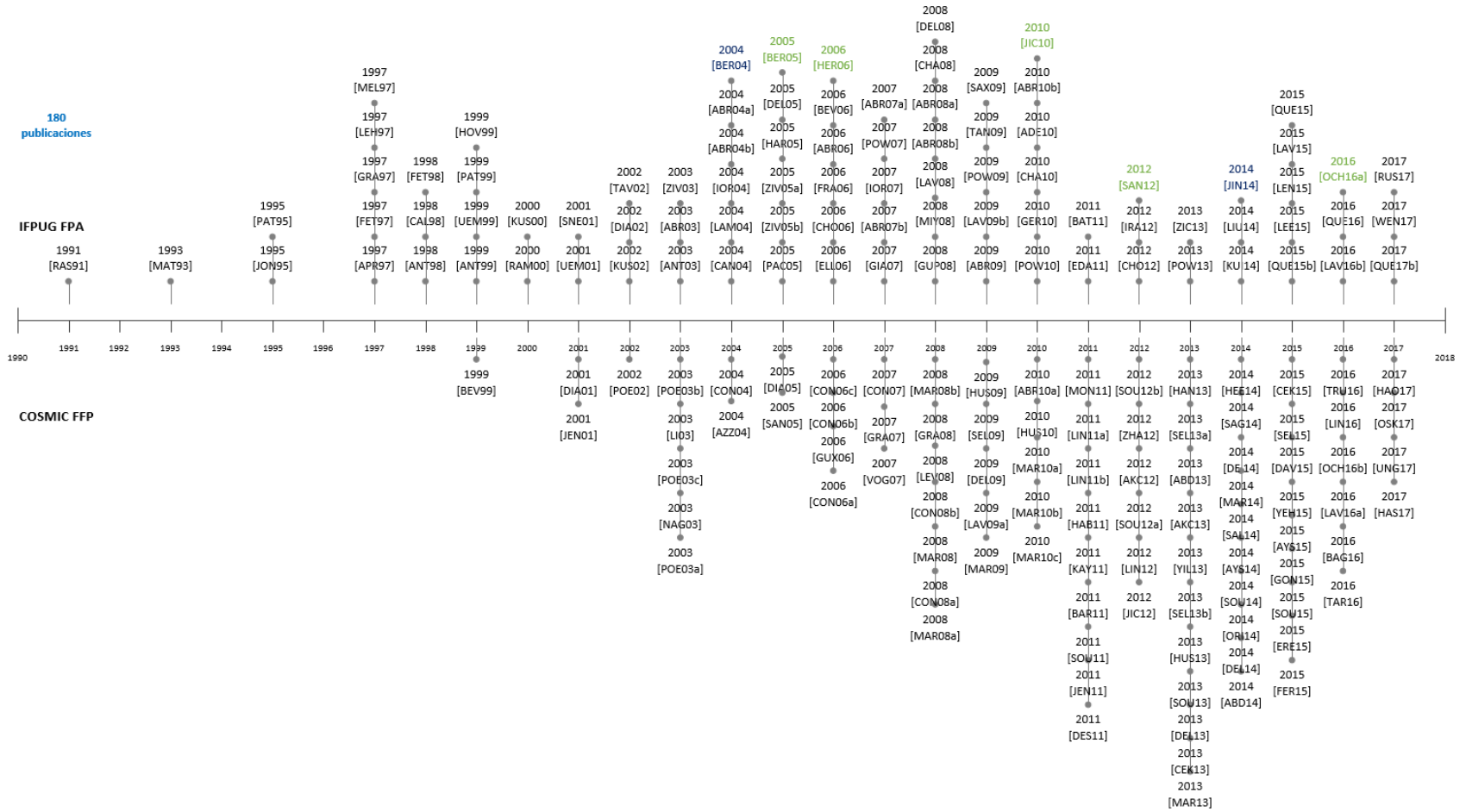


Figura 4-5 Listado de los artículos primarios seleccionados.

Tabla 4-6 Listado de artículos primarios.

#	CR	Ref.	AN	GR	Enfoque (E)	OR	TF	GH	ME	TM	AT	Artef.	EC	TP	TE	DF	H	Q
1	RAS91	(Rask, 1991)	1991		Propone las reglas para contar los puntos de función IFPUG a partir de diagramas de flujos de datos y modelos entidad relación.	SBB	R	11	I	M	P	Modelos conceptuales	R D	P	PR	MIS		8
2	MAT93	(Matson & Mellichamp, 1993)	1993		Describe una herramienta para el conteo de puntos de función en aplicaciones orientadas a objetos.	AUT	J	10	I	E	P	Modelos conceptuales	D	PH	CS	MIS	Y	9
3	JON95	(Jones, 1995)	1995		Propone un procedimiento de conteo ajustando el método IFPUG para contar puntos de función a partir de las entidades de datos automáticamente.	SBB	C	4	I	M	P	Modelos de diseño	D	PH	PC	MIS	Y	7
4	PAT95	(Paton & Abran, 1995)	1995		Propone una especificación formal para el método de medición IFPUG y detalla el conjunto de reglas que pueden ser automatizadas para el conteo de tamaño funcional.	SBB	R	10	I	M	P	Modelos conceptuales	R D	P	PR	MIS		8
...
176	QUE17b	(Quesada-López, Jenkins, Salas & Gómez, 2017b)	2017	G001	Propone una herramienta de medición para modelos de diseño.	SBF	C	0	I	E	P	Modelos de diseño	D	PHE	CS	MIS	Y	15
177	RUS17	(Rusli & Abdullah, 2017)	2017	G020	Presenta un procedimiento para medir aplicaciones móviles de juegos mediante el método IFPUG basadas en diagramas UML.	AUT	J	0	I	E	P	Modelos UML	R D	P	PC	RTS		9
178	HAS17	(Soubra & Abran, 2017)	2017		Propone los lineamientos para medir aplicaciones del dominio IoT	SBF	C	0	C	M	P	Código fuente	C	P	PC	RTS		8
179	UNG17	(Ungan, Trudel & Poulin, 2017)	2017	G032	Propone el uso de patrones de medición del tamaño funcional para facilitar la medición de requerimientos de seguridad	SBF	C	0	C	E	P	Modelos de diseño	R	PE	CS	MIS		11
180	WEN17	(Wentzlaff, 2017)	2017	G032	Propone el uso del tamaño funcional como punto de referencia para clasificar los requerimientos y estimar el tamaño funcional de manera reproducible. Plantea la medición utilizando <i>problem frames</i> asociando los BFC a cada patrón.	SBF	C	0	I	E	P	Especificación de requerimientos	R D	P	PC	MIS		4

(continúa)

#: Secuencia, CR: Código de referencia, Ref.: referencia bibliográfica, AN: año, GR: grupo, OR: origen, TF: tipo de foro, G: Google hits, ME: método de medición, TM: tipo de medición, AT: atributo medido, Artef.: artefacto de entrada, EC: etapa del ciclo de vida de desarrollo, TP: tipo de propuesta, TE: tipo empírico, DF: dominio funcional, H: Herramienta, Q: suma total de calidad.

OR: CTR=artículos de control, AUT=búsqueda automática, SBB=proceso de bola de nieve hacia atrás, SBF=proceso de bola de nieve hacia adelante.

TF: J=Revista, C=Conferencia, B=Capítulo de libro, R=Reporte técnico.

ME: I=IFPUG FPA, C=COSMIC FFP, M=MKII FPA.

TM: A=Aproximación de tamaño (*early estimation*), E=Estimación de tamaño, M=Medición del tamaño.

AT: P=Producto, R=Proceso

EC: R=Requerimientos, D=Diseño, C=Codificación, M=Mantenimiento.

TP: P=Propuesta, H=Herramienta, E=Empírico.

TE: AC=Análisis crítico, RE=Reporte de experiencia, PR=Propuesta, PC=Prueba de concepto, CS=Caso de estudio, EC=Cuasi experimento o experimento controlado, EX=Experimento.

DF: MIS=Sistemas de información, RTS= sistemas en tiempo real.

Los artículos realizan procesos de medición para la aproximación, estimación o medición del tamaño funcional a partir de los artefactos de software. Ozkan & Demirors (2016) establecen que la estimación y la medición del tamaño funcional del software no son lo mismo. Las técnicas de estimación o aproximación son utilizadas para obtener el tamaño de los requerimientos funcionales cuando están incompletos o sin el suficiente detalle, principalmente en las etapas tempranas del desarrollo. Por otro lado, la medición del tamaño funcional sigue un procedimiento de medición que cumple con los conceptos y reglas de un método de medición del tamaño funcional. Haoues, Sellami & Ben-Abdallah (2017) indican que al inicio del proceso de desarrollo los requerimientos no se encuentran bien detallados, por lo cual se utilizan técnicas de aproximación para estimar el tamaño funcional en etapas tempranas. Finalmente, indica que cuando se cuenta con los requerimientos detallados se puede realizar la medición mediante el proceso completo establecido por el método o mediciones rápidas si el tiempo para realizarlo es limitado. Los métodos rápidos de medición pueden ser utilizados para producir las mediciones rápidamente y con una aproximación confiable al tamaño funcional real (Vogelezang, Symons, Lesterhuis, Meli & Deneva, 2013).

Del total de artículos, 96 (53%) afirman realizar la medición del tamaño funcional y 84 (47%) afirman realizar la estimación o aproximación del tamaño funcional. Los artículos de medición se distribuyen en 60 para COSMIC FFP y 36 para IFPUG FPA. Los artículos de estimación se distribuyen en 37 para COSMIC FFP y para 50 IFPUG FPA. Para el método MkII FPA se encontró un artículo para medición y un artículo para estimación. Todos los artículos que mencionaron utilizar una técnica de aproximación se desarrollaron para el método COSMIC FFP (11 artículos). Los totales presentados incluyen artículos que presentan propuestas para más de un método, 3 artículos realizan la propuesta para IFPUG FPA y COSMIC FFP, todos ellos para estimación. Además, 2 artículos realizan la propuesta para IFPUG FPA, COSMIC FFP y MkII FPA, uno de ellos para estimación del tamaño y el otro para medición.

Los procedimientos pueden aplicar el conjunto de reglas de mapeo entre componentes de los métodos de medición y los conceptos de los artefactos de software utilizados

para la medición. La Tabla 4-7 detalla la tendencia de uso de los artefactos en las publicaciones de los procedimientos. En total, 157 artículos utilizan artefactos desarrollados en las (E) etapas tempranas del desarrollo del software (87%), 23 artículos utilizan artefactos (D) del desarrollo tales como el código fuente (13%). De los artículos que utilizan artefactos de etapas tempranas, 87 lo hacen con el método COSMIC FFP y 73 con el método IFPUG FPA. Para los que utilizan artefactos de etapas de desarrollo, 10 lo hacen con el método COSMIC FFP y 13 con el método IFPUG FPA. Con el método MkII FPA solo se utilizan artefactos de etapas tempranas de desarrollo.

En etapas tempranas el método COSMIC FFP presenta la mayor cantidad de publicaciones en el 2013 con 10, el método IFPUG FPA se mantuvo estable entre el 2004-2010 con entre 5 y 6 publicaciones en promedio. De los artículos que utilizan artefactos de la etapa del desarrollo como el código fuente, el método COSMIC FFP presenta la mayor cantidad de publicaciones del 2015-2017 con 2, el método IFPUG FPA lo hace en el 2015 con 4, principalmente relacionados con el análisis de OMG AFP (Object Management Group, 2014).

Tabla 4-7 Artefactos por año y método de medición del tamaño funcional.

Método	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total
Artefactos en etapas tempranas del desarrollo (E)																												
C	0	0	0	0	0	0	0	0	1	0	2	1	5	2	3	4	3	7	5	6	9	6	10	9	7	5	2	87
I	1	0	1	0	2	0	4	3	2	2	1	2	3	5	6	5	5	6	5	6	1	3	2	2	1	2	3	73
M	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	4
T	1	0	1	0	2	0	4	3	3	2	3	3	8	8	8	8	8	13	10	11	10	8	12	12	8	6	5	157
Artefactos en etapas del desarrollo (D)																												
C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	1	2	2	2	10
I	0	0	0	0	0	0	1	0	2	0	1	1	0	0	0	1	0	1	0	0	1	0	0	0	0	4	1	13
M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	0	0	1	0	2	0	1	1	0	0	0	2	0	1	0	0	1	1	1	1	6	3	2	23

C=COSMIC FFP, I=IFPUG FPA, M=MkII

Durante el análisis de las publicaciones se deduce la o las etapas del ciclo de vida del desarrollo del software donde es posible aplicar los procedimientos de medición del

tamaño funcional a partir de los artefactos de entrada. En el caso de algunas propuestas, estos se puedan aplicar en diferentes etapas. La mayor cantidad de procedimientos de medición se han desarrollado para la etapa de diseño (102 artículos), tanto para el método IFPUG FPA (54) como para el método COSMIC FFP (45). Adicionalmente, se presenta un procedimiento que contempla los métodos IFPUG FPA y COSMIC FFP, un procedimiento que contempla los métodos IFPUG FPA, COSMIC FFP y MkII FPA, y un procedimiento para el método MkII FPA.

La segunda etapa es la de requerimientos (101 artículos), tanto para el método IFPUG FPA (38) como para el método COSMIC FFP (57). Adicionalmente, se presentan 3 procedimientos que contempla los métodos IFPUG FPA y COSMIC FFP, un procedimiento que contempla los métodos IFPUG FPA, COSMIC FFP y MkII FPA, y 2 procedimientos para el método MkII FPA.

Muchas de estas propuestas (45) se pueden aplicar en la etapa de requerimientos si existe un modelo formal o una especificación formal de requerimientos y dependiendo del contexto de operación de desarrollo de aplicaciones de software. En la etapa de codificación (23) existen propuestas tanto para COSMIC FFP (10) como para IFPUG FPA (13). De los procedimientos que se aplican en esta etapa, el proceso de análisis se realiza mediante un análisis de las aplicaciones en tiempo de ejecución o un análisis estático sobre el código fuente. En la aplicación de un procedimiento de medición, en la mayoría de los casos, se utiliza el código fuente para hacer un análisis de los flujos de datos que representan los componentes funcionales.

El 36% de las publicaciones utilizan artefactos basados en modelos UML (65) [34 IFPUG, 28 COSMIC, 1 Mark II, 1 que contempla IFPUG, COSMIC y MkII, y 1 que contempla IFPUG y COSMIC] tales como: diagramas de clase, objetos, secuencia, actividad, casos de uso y componentes; 31% modelos conceptuales (55) [21 IFPUG, 33 COSMIC y 1 que contempla IFPUG, COSMIC y MkII] que utilizan modelos de clases, dinámicos, flujo de datos, funcionales y de presentación; 13% especificaciones de requerimientos (23) [6 IFPUG, 15 COSMIC y 2 que contempla IFPUG y COSMIC] como requerimientos funcionales, historias de usuarios y ontologías de requerimientos, 3% lenguajes formales de especificación (6) [5

IFPUG, 1 Mark II] como metas y escenarios, notaciones matemáticas y especificaciones VDM-SL, 2% modelos de negocio (4) [4 COSMIC] como BPMN y 2% otros modelos de diseño particulares (4) tales como modelos propietarios, de datos y de interfaz. El 13% de las publicaciones utilizan artefactos basados en el código fuente (23) [13 IFPUG, 10 COSMIC]. Las líneas de investigación que se han mantenido activas a través de los años son las que utilizan modelos UML y modelos conceptuales. Las propuestas para calcular el tamaño funcional a partir de modelos de negocio (BPM) solo se han realizado para COSMIC, entre el 2011-2014. El detalle de los artefactos de entrada para todos los procedimientos de medición se encuentra detallado en el Anexo 2.

Las publicaciones plantean propuestas para el proceso de medición, desarrollan herramientas para automatizarlo y/o validan empíricamente los resultados de su aplicación en contextos específicos. Un total de 73 artículos realizan una propuesta, una herramienta y su validación (41%) [34 IFPUG FPA, 35 COSMIC FFP, 4 otras], 28 una propuesta y herramienta (16%) [14 IFPUG FPA, 14 COSMIC FFP], 40 una propuesta y validación (22%) [15 IFPUG FPA, 23 COSMIC FFP, 2 otras], 25 solo una propuesta (14%) [10 IFPUG FPA, 14 COSMIC FFP, 1 otras], 8 una herramienta y validación (4%) [5 IFPUG FPA y 3 COSMIC FFP] y 6 realizan solo una validación empírica (3%) [3 IFPUG FPA, 3 COSMIC FFP] basada en un trabajo previo en el cual se utiliza una herramienta de soporte al proceso.

Todos los artículos miden la entidad producto. Un total de 98 (55%) publicaciones dicen ser para ambientes y metodologías orientadas a objetos (OO), 38 (21%) para desarrollo dirigido por modelos (MDA), 25 para aplicaciones web (14%), 7 publicaciones para aplicaciones móviles (3%) y se identifican 5 publicaciones para ambientes ágiles (3%). Un total de 29 (16%) publicaciones son para el dominio funcional de sistemas de tiempo real (RTS) y 152 (84%) para el dominio de sistemas de información (MIS). IFPUG FPA es utilizado en MIS (45%) y en RTS (1%), COSMIC en MIS (40%) y en RTS (15%). Esto es, para el dominio MIS, 82 de las propuestas usan el método de medición IFPUG FPA y 73 utilizan el método COSMIC FFP. Sin embargo, para el dominio RTS, 26 de las propuestas usan el método de medición COSMIC FFP y solo 3 utilizan el método IFPUG FPA.

Un procedimiento de medición del tamaño funcional sigue los pasos del modelo de procesos para métodos de medición del tamaño funcional propuesto por (Abran, 2010; Habra, Abran, Lopez & Sellami, 2008; Jacquet & Abran, 1997). Esto incluye la validación teórica y que se realice la definición del procedimiento y las reglas de aplicación. Esto incluye la especificación de las reglas de construcción del metamodelo funcional, las reglas de mapeo entre el artefacto de entrada y los componentes del método de medición, y las reglas de asignación numérica. Del total de artículos primarios, 25 realizan o mencionan una validación teórica del procedimiento de medición del tamaño funcional. En total, 70 artículos definen las características del procedimiento de medición y el proceso de aplicación del procedimiento de medición. En la mayoría de los casos, los estudios se centran en la propuesta del procedimiento de medición. Sin embargo, 27 artículos primarios también detallan el modelo de predicción que utiliza como base la métrica de tamaño funcional. Para los modelos de predicción, en 11 artículos se mencionan modelos de esfuerzo de desarrollo, y en 13 artículos modelos de predicción de tamaño físico de software (líneas de código, tamaño de memoria).

Las publicaciones mencionan herramientas que pueden apoyar la automatización del proceso de conteo en diferentes niveles. El 78% (141) de las publicaciones mencionan la posibilidad de automatización a partir del procedimiento de medición y el 60% (109) dicen utilizar una herramienta. La mayoría de las herramientas realizan análisis estático sobre los artefactos de software. El análisis dinámico se realiza sobre artefactos de software como el código fuente y las aplicaciones corriendo en ambientes controlados. El 30% utilizan herramientas para IFPUG FPA y el 31% para COSMIC FFP. Es importante notar que muchas de las propuestas de las herramientas planteadas en las investigaciones no automatizan totalmente el proceso de medición, sino que representan un soporte para mejorar la productividad de los procesos de medición del tamaño funcional.

Los artefactos de software más utilizados como entrada para herramientas en el proceso de medición son los diferentes diagramas de los modelos UML, modelos de requerimientos y modelos conceptuales. El código fuente de las aplicaciones de software también se ha utilizado para realizar el proceso de medición automatizado. Las etapas del ciclo de

desarrollo donde se han aplicado herramientas son las de diseño, la de requerimientos y la etapa de codificación.

4.2.1.5. ENFOQUES DE INVESTIGACIÓN EN EL ÁREA

En la Tabla 4-6 se detallan los grupos de estudios relacionados. Estos grupos de artículos estudian propuestas de medición similares.

Los artículos del grupo G000 [HEE14, DAV15, FER15, HAO17] aplican el método COSMIC FFP para medir aplicaciones móviles. Las mediciones las realizan sobre artefactos de las etapas tempranas de desarrollo tales como especificación de requerimientos funcionales y casos de uso. [DAV15, FER15] utilizan una estrategia de identificación de patrones del software para realizar la medición (M). En el caso de [HEE14, HAO17] proponen técnicas de estimación (E) del tamaño funcional. El grupo G020 [ABD13, ABD14, RUS17] realiza la medición de aplicaciones móviles para juegos utilizando modelos UML. En el caso de [ABD13, ABD14] lo hacen para COSMIC FFP y [RUS17] estima el tamaño para el método IFPUG FPA. El grupo G032 [TRU16, UNG17, WEN17] propone la estimación del tamaño funcional a partir de patrones de medición para el método COSMIC FFP sobre la especificación de requerimientos o modelos de diseño.

El grupo G030 [HUS09, HUS10, HUS13, OCH16a, OCH16b] realiza y evalúa una propuesta de aproximación del tamaño funcional COSMIC FFP e IFPUG FPA a partir de los requerimientos textuales como historias de usuario o casos de uso. Los autores aproximan el tamaño a partir de la aplicación de técnicas de minería de texto. El grupo G003 [CHO06, ADE10, CHO12] propone un procedimiento de estimación de tamaño basado en reglas aplicadas a requerimientos textuales basado en el modelo de metas y escenarios para IFPUG FPA. El grupo G008 [MEL97, SAN05, IOR07, VOG07, SAN12, DEV14, DEL14, LAV16a] propone y evalúa técnicas rápidas y tempranas de estimación de tamaño funcional principalmente para el método COSMIC FFP y para IFPUG FPA. Los artículos del grupo G028 [BEV06, BAG16] proponen la medición del tamaño funcional basados en una ontología de requerimientos.

El grupo G002 [SEL09, SEL13a, SEL13b, SEL15] propone y analiza procedimientos de la medición del tamaño funcional basado en el método COSMIC FFP que usa como entrada diagramas UML tales como casos de uso, secuencia, actividad y componentes. El grupo G004 [LIN11a, LIN11b, LIN12] desarrolla un método basado en perfiles UML, para aplicaciones embebidas con el método COSMIC FFP. El grupo G011 [CAN04, PAC05] analiza la medición en modelos UML para el método IFPUG FPA. El grupo G013 [ANT98, CAL98, ANT99, ANT03] analiza aplicaciones orientadas a objetos basado en modelos UML para el método IFPUG FPA al igual que el grupo G015 [UEM99, KUS00, UEM01] y G027 [CHA08, CHA10]. El grupo G022 [BEV99, LEV08] lo hace para el método COSMIC FFP. El grupo G017 [POW09, POW10, POW13] presenta un procedimiento para identificar los componentes del método IFPUG FPA a partir de diagramas de clase UML. El grupo G018 [ZIV03, ZIV05a, ZIV05b, HER06] presenta un mapeo entre modelos UML y el método IFPUG FPA para software orientado a objetos. Discuten estrategias de mapeo existentes y define la estrategia de mapeo. Finalmente, propone una representación de métodos de tamaño funcional y los mapea en el sistema GASS para el cálculo de puntos de función IFPUG FPA, COSMIC FFP, MkII FPA.

El grupo G021 [JIC10, JIC12] propone un procedimiento para medir puntos de función COSMIC FFP, IFPUG FPA y MarkII FPA a partir de diagramas UML. El grupo G023 [JEN01, JEN11] analiza los problemas de granularidad al realizar la medición del tamaño funcional COSMIC FFP sobre modelos UML. El grupo G025 [DEL05, DEL08, LAV08, DEL09, LAV09a, LAV09b, DEL13] propone un procedimiento para la construcción de modelos UML para aplicar las reglas de los métodos de medición y provee un conjunto de mediciones para distintos casos de estudio.

El grupo G006 [MAR08b, MAR08a, MAR08, MAR09, ABR10a, MAR10a, MAR10b, MAR10c, MAR13] mide aplicaciones desarrolladas bajo el método *OO-Method* para COSMIC FPA, de igual forma el grupo G007 [ABR03, ABR04a, ABR04b, ABR06, ABR07b, ABR07a, GIA07, ABR08b, ABR09, ABR10b] lo hace para IFPUG FPA. El grupo G014 [POE02, POE03a, POE03b, POE03c] analiza la medición sobre modelos conceptuales para COSMIC FFP. El grupo G009 [CON04, CON06a, CON06b, CON06c, CON07, CON08a, CON08b] mide

aplicaciones desarrolladas bajo el método *OO-Method* para COSMIC FFP a partir de la especificación de requerimientos.

El grupo G010 [GRA07, GRA08] mide puntos de función COSMIC FFP bajo el modelo PRIM. El grupo G016 [AYS14, AYS15] estima el tamaño funcional COSMIC FFP bajo la notación UPROM. El grupo G029 [DIA01, DIA05] proponen la medición para COSMIC FFP basado en las especificaciones ROOM y modelos generados mediante la herramienta *Rational Rose* para sistemas de tiempo real. El grupo G019 [KAY11, MON11, MAR14, SAL14] propone la medición del método COSMIC FFP a partir modelos de negocio BPMN.

El grupo G005 [KUS02, GUP08, EDA11] realiza el conteo del método IFPUG FPA a partir del código fuente de aplicaciones Java. El grupo G026 [SOU11, SOU12a, SOU12b, SOU13, ORI14, SOU14, SOU15] utiliza el lenguaje Simulink y modelos AUTOSAR como entrada para medir el tamaño funcional COSMIC FFP para aplicaciones en tiempo real. Además, propone un protocolo de verificación de exactitud para el método COSMIC FFP para demostrar la exactitud de las mediciones automáticas de tamaño funcional.

El grupo G024 [AKC12, AKC13, SAG14, GON15, TAR16, OSK17] realiza el conteo del método COSMIC FFP a partir del código fuente de las aplicaciones analizando los flujos de datos y procesos. El grupo G031 [YEH15, LIN16] presenta una herramienta de ingeniería inversa para obtener a partir de código fuente diagramas de secuencia y calcular el tamaño funcional del software COSMIC FFP. El grupo G012 [GRA97, LAM04] analiza el conteo a partir de diagramas de flujos de datos y modelos entidad relación para el método IFPUG FPA.

El grupo G001 realiza una evaluación empírica [QUE15, QUE15b, QUE16] y un análisis crítico [LAV15] de puntos de función automatizados OMG AFP basado en el método IFPUG FPA. A partir de la evaluación de distintos métodos de medición automatizada [QUE16] realiza una propuesta de medición del tamaño funcional para el método IFPUG FPA [QUE17b].

4.2.2. ESTRATEGIAS EMPÍRICAS UTILIZADAS PARA EVALUAR LOS PROCEDIMIENTOS DE MEDICIÓN DEL TAMAÑO FUNCIONAL

RQ2. ¿Cómo han sido evaluados los procedimientos de medición del tamaño funcional?

En esta sección se identifican los tipos de evaluación empírica y cómo han sido evaluados los procedimientos de medición del tamaño funcional, las preguntas de investigación que han sido investigadas empíricamente, los constructos, métricas y mediciones que han sido utilizadas para evaluar los procedimientos, las amenazas a la validez que se han reportado en el diseño y evaluación empíricos de los procedimientos y los aspectos que se deben considerar para realizar repeticiones para la evaluación de procedimientos de medición.

4.2.2.1. TIPOS DE EVALUACIÓN EMPÍRICA

La Tabla 4-8 presenta la distribución de los estudios empíricos para los artículos identificados. Los estudios que presentan procedimientos para dos o más procedimientos son contados por cada uno de los métodos de medición. El 87% de las publicaciones (156) realizan algún tipo de validación para los procedimientos. El 13% (24) realizan procesos experimentales (EC y EX), el 48% (87) casos de estudio (CS) y el 25% (45) pruebas de concepto o estudios piloto (PC). Dos publicaciones realizan un análisis crítico (AC) y las demás solo realizan una propuesta (PR). El método COSMIC FFP ha sido evaluado principalmente mediante casos de estudio (49 estudios) y pruebas de concepto (28 estudios). El método IFPUG FPA mediante casos de estudio (40) y pruebas de concepto (17 estudios). El método MkII FPA se ha evaluado mediante 2 casos de estudio.

En total, 91 publicaciones realizan validaciones en ambientes académicos (58%) y 68 publicaciones realizan evaluaciones en la industria (44%). Las validaciones en la industria varían entre la industria automotriz, financiera, gobierno, telecomunicaciones, juegos, entre otras. Las evaluaciones realizadas en la industria para el método COSMIC FFP repre-

sentan el 56% y las validaciones en la academia el 44%. El mismo comportamiento se presenta para IFPUG FPA, donde las evaluaciones realizadas en la industria representan el 58% y las validaciones en la academia el 42%.

Tabla 4-8 Distribución de los estudios empíricos.

Tipo	COSMIC FFP			IFPUG FPA			TG
	Validación (Academia)	Evaluación (Industria)	T	Validación (Academia)	Evaluación (Industria)	T	
Análisis crítico (AC)	1	0	1	2	0	2	2
Propuesta (PR)	0	0	12	3	0	10	22
Prueba de concepto (PC)	20	8	28	14	3	17	45
Caso de estudio (CS)	23	26	49	15	25	40	87
Estudio cuasi-experimental (EC)	4	2	7	12	2	14	21
Experimento (EX)	0	0	0	1	2	3	3
Total	48	36	97	47	32	86	

T=Total, TG=Total General

En las evaluaciones en la industria (68), el método principal es el de caso de estudio, con un 75% de las veces. En las validaciones en la academia (91), el 37% son pruebas de concepto, 39% caso de estudio y el 17% procesos experimentales. Los estudios experimentales regularmente realizan procesos con estudiantes para las validaciones (65%) pero también se han utilizado en casos de estudio (30%) y pruebas de concepto (5%). Solo seis publicaciones son reportadas explícitamente como replicaciones de estudios anteriores [CON06c, ABR08b, ABR09, FER15, OCH16b, QUE16].

Del total de estudios empíricos, 15 complementaron el estudio al realizar encuestas de percepción (en la mayoría de los casos, con estudiantes y/o con profesionales en ambientes académicos). Las tendencias de estudios empíricos por año muestran que para los casos de estudio se ha presentado un incremento constante; sin embargo, la cantidad de publicaciones que realizan solo una prueba de concepto se mantiene igualmente creciente.

La tendencia de las evaluaciones muestra la necesidad de investigación empírica rigurosa en el campo de los procedimientos de medición del tamaño funcional. No basta con realizar pruebas de concepto sobre las propuestas de los procedimientos de medición del tamaño funcional. Es necesario realizar los casos de estudio en contextos reales, con

aplicaciones de software reales, así como experimentos de laboratorio que permitan generalizar los resultados sobre la efectividad y la adopción de los procesos de medición.

4.2.2.2. PREGUNTAS DE INVESTIGACIÓN

Las preguntas de investigación identifican los aspectos que los autores investigan en el área de los procedimientos de medición del tamaño funcional. El principal objetivo de los estudios primarios seleccionados ha sido investigar la exactitud de los resultados del proceso de medición del tamaño funcional. Estos estudios analizan cómo el nivel de detalle de los artefactos de entrada utilizados durante la aplicación del proceso de medición impacta en la exactitud de los resultados de medición. Asimismo, un tema recurrente en el análisis es el nivel de detalle y la efectividad necesaria para que los modelos de representación funcional del software sean efectivos en la aplicación de los procedimientos de medición.

La formalización, automatización y aplicación de las reglas de mapeo son temas de análisis en los estudios. Estos estudios intentan sistematizar los procesos de medición proveyendo un conjunto de reglas de aplicación específicas para el contexto donde se realizan las mediciones.

El estudio de la confiabilidad, consistencia y repetibilidad de los resultados de medición cuando se aplican procedimiento de medición son de interés, así como las propiedades de adopción de los procesos de medición.

La calibración de los procedimientos de medición para obtener resultados aceptables de medición se discute en los estudios en el área. Asimismo, otro aspecto de estudio es el esfuerzo requerido para la aplicación de los procedimientos de medición y cómo este impacta en las propiedades de adopción cuando es utilizado por profesionales.

4.2.2.3. CONSTRUCTOS, MÉTRICAS Y MEDICIONES PARA LA EVALUACIÓN

La estandarización de los constructos, métricas y mediciones habilita la posibilidad de comparar los resultados de los procedimientos de medición en contextos similares o

distintos para evaluar su eficacia. Los procedimientos de medición evalúan las diferentes métricas que se pueden clasificar como: métricas de eficacia (exactitud, repetibilidad y reproducibilidad), eficiencia (tiempo y productividad) y métricas de adopción (facilidad de uso percibida, utilidad percibida e intención de uso futuro).

Los estudios validan las mediciones comparando la exactitud entre el valor obtenido por el procedimiento y un valor de referencia logrado mediante la medición de uno o varios expertos, validando la repetibilidad o reproducibilidad, propiedades de adopción como facilidad de uso, utilidad e intención de uso futuro y productividad.

Del total de 180 estudios primarios, el 61% evalúa la exactitud de los resultados de medición, además, un 4% evalúa la repetibilidad y un 7% evalúa la reproducibilidad. El detalle de los procesos de evaluación de repetibilidad y reproducibilidad se puede encontrar en los artículos [ABR04a, ABR04b, CON06a, CON06b, CON07, ABR08b, DEL08, MAR08b, MAR10c, EDA11, QUE15, QUE15b, QUE16]. De estos estudios, 13 artículos primarios evalúan la productividad en la aplicación de los procedimientos de medición. Para la evaluación de efectividad, solo 4 estudios primarios reportan explícitamente la utilización de un protocolo sistemático de evaluación de los resultados de la medición: el [MAR08b, YIL13, SOU14] lo utiliza para COSMIC FFP y el [QUE15b] para IFPUG FPA, todos basados en [SOU14]. Solo 7 publicaciones mencionan explícitamente el uso de un protocolo de verificación de exactitud.

Con respecto a las propiedades de adopción, 15 artículos primarios evalúan la facilidad de uso percibida, 10 la utilidad percibida, 10 la intención de uso futuro y 1 la satisfacción. El detalle de cada uno de los procesos de evaluación de las propiedades de adopción se puede encontrar en los artículos [ABR04b, ABR06, CON06c, ABR07b, ABR08b, CON08a, ABR09, QUE15, QUE15b, QUE16]. Estas, en su mayoría, se basan en la propuesta original de (Davis, 1989) y sus adaptaciones para métodos de medición (Abraham, 2004).

Los resultados de la exactitud de las mediciones varían dependiendo de cada estudio y su contexto. En el caso de procedimientos basados en el método de medición COSMIC FFP, aunque la mayoría de estudios recientes han reportado mediciones dentro del $\pm 10\%$ de variación, otros indican diferencias de hasta un 53%, para casos específicos de medición.

De la misma manera, para el método IFPUG FPA algunos estudios han reportado mediciones dentro del $\pm 10\%$ de variación, así como diferencias de hasta un 40% para procesos específicos de medición. En el caso de la adopción de los procedimientos de medición, pocos estudios reportan estas propiedades y en ninguno se halló información sobre la adopción en la industria. Sin embargo, se puede observar que a mayor sistematización y automatización de los procedimientos de medición que permitan realizar un proceso de medición más eficiente, consistente y sin perder exactitud, mayor es la intención de uso de estos procedimientos.

4.2.2.4. AMENAZAS A LA VALIDEZ

Para el análisis se establecieron las categorías establecidas en (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012) y se extrajo la información de cada estudio primario cuando la amenaza fue reportada explícitamente por los investigadores originales. Del total de estudios primarios, 47 discuten las amenazas a la validez y de estos, 23 detallan las amenazas por una o más categorías de acuerdo con la clasificación establecida.

4.2.2.4.1. VALIDEZ INTERNA

La utilización de personas (estudiantes, profesionales y expertos) presenta una variación natural en la aplicación de los procesos de medición. Los experimentos en la academia con estudiantes seleccionados a conveniencia se reportan como una amenaza, pero se reporta que el nivel de conocimiento y experiencia similar permite comparar los rendimientos obtenidos. Se recomienda validar la confiabilidad en la recolección de los datos y mantener la motivación de los participantes; además, seleccionar aleatoriamente a los participantes. El uso de un solo experto para determinar las medidas de referencia es una amenaza a la validez; algunos estudios indican la participación de varios expertos para determinar ese valor de referencia y, en algunos casos, los valores de referencia no son calculados por contadores certificados.

El tamaño y la cantidad de las aplicaciones y artefactos utilizados en los estudios representan una amenaza, el contexto y nivel de detalle de los artefactos de software para la aplicación de los procedimientos son un factor de influencia en los resultados. Los estudios que comparan procedimientos fijan este factor y utilizan artefactos del mismo dominio funcional y con características funcionales similares. Se recomienda el uso de artefactos estandarizados de acuerdo a especificaciones reconocidas y la utilización de material revisado y validado en estudios anteriores.

4.2.2.4.2. VALIDEZ DEL CONSTRUCTO

Los estudios reportan métricas de evaluación ampliamente aceptadas en el campo de medición del tamaño funcional. En el caso de propiedades de efectividad se utilizan métricas reconocidas, tales como exactitud, repetibilidad, reproducibilidad y productividad. Respecto a las propiedades de adopción, se utilizan modelos de adopción de tecnologías basados en (Davis, 1989) y sus adaptados para métodos de medición. Algunos investigadores advierten que las herramientas automatizadas pueden no seguir completamente las reglas establecidas por los estándares de medición.

4.2.2.4.3. VALIDEZ DE LAS CONCLUSIONES

Los autores reportan la heterogeneidad de los participantes, aplicaciones y artefactos utilizados como una amenaza; además, de la cantidad y tamaño limitado de los artefactos utilizados para la experimentación.

4.2.2.4.4. VALIDEZ EXTERNA

El uso de aplicaciones pequeñas en contextos académicos es una limitación. Se deben utilizar aplicaciones similares a las que se encuentran en la industria o aplicaciones reales de la industria. La generalización de los resultados se ve limitada porque los procesos de conteo se aplican en contextos específicos de desarrollo, bajo un dominio funcional particular y con artefactos de software específicos. Los estudios reportan que la participación

de sujetos que no son regularmente usuarios de los procedimientos de medición, representan una amenaza para generalizar los resultados sobre el uso y la adopción.

4.2.2.5. REPLICACIÓN DE ESTUDIOS

Del total de estudios primarios, sólo 6 estudios son explícitamente reportados como un estudio de replicación por los investigadores originales: [CON06c, ABR08b, ABR09, FER15, OCH16b, QUE16]:

- En [CON06c] se evalúa un procedimiento de medición para sistemas orientados a objetos que utiliza el método COSMIC FFP a partir de las especificaciones de requerimientos del OO-Method. Este artículo es una replicación interna de los autores originales del estudio [CON06a].
- En [ABR08b] se evalúa el procedimiento de medición OOmFP para aplicaciones web que utiliza modelos conceptuales como entrada para el procedimiento de medición. Este artículo es una replicación interna estricta de los autores originales del estudio [ABR03, ABR04b].
- En [ABR09] se realiza una familia de experimentos para la validación del procedimiento de medición OOmFPWeb utilizando el modelo de aceptación de tecnología (TAM). Este artículo es una familia de experimentos de los autores originales del estudio [ABR03, ABR04b, ABR08b].
- En [FER15] se aplica el procedimiento de medición para contar el tamaño funcional COSMIC FFP de aplicaciones móviles en el que se utiliza una estrategia de medición de patrones del software. Este artículo es una replicación interna de los autores originales del estudio [HEE14, FER15a].
- En [OCH16b] se valida un procedimiento de aproximación del tamaño funcional COSMIC FFP a partir de requerimientos textuales. Este artículo es una replicación externa del estudio [HUS13].

- En [QUE16] se realiza una familia de experimentos para la evaluación empírica de varias propuestas de medición de puntos de función basadas en IFPUG FPA. Este artículo es una familia de experimentos de los autores originales del estudio [QUE15, QUE15b].

Asimismo, existen estudios relacionados entre sí que evalúan propuestas de medición similares. Los grupos de estudios se encuentran detallados en la Sección 4.2.1.5. En términos generales, son muy pocos estudios los que utilizan un protocolo de experimentación y un protocolo de replicación reconocido en el área de ingeniería de software experimental para reportar los estudios y las replications. Muchos de estos estudios no incluyen el nivel de detalle necesario para conducir replications que permitan generalizar resultados de los procedimientos de medición. Algunos de los estudios que reportan protocolos de experimentación detallados y que pueden ser una guía para futuros estudios son: [ABR03, ABR04a, CON04, ABR04b, DIA05, CON06c, ABR06, CON07, ABR07a, ABR07b, ABR08b, ABR08a, ABR09, LAV08, LAV09b, ABR10a, MAR10c, EDA11, DEL14, FER15, SEL15, OCH16a]. Es necesario considerar que una descripción clara de los protocolos de los estudios empíricos y del contexto de operación es requisito clave para la replicación y agregación de resultados (Sjoberg, Dyba & Jorgensen, 2007).

4.2.3. DESAFÍOS Y OPORTUNIDADES DE LOS PROCEDIMIENTOS DE MEDICIÓN DEL TAMAÑO FUNCIONAL

***RQ3.** ¿Cuáles son los desafíos y oportunidades de los procedimientos existentes para la medición del tamaño funcional?*

En esta sección se identifican los principales desafíos y oportunidades en el área de los procedimientos de medición del tamaño funcional, las recomendaciones para la experimentación, los problemas abiertos y las posibles líneas de investigación futura.

4.2.3.1. DESAFÍOS, OPORTUNIDADES Y LIMITACIONES

A partir de las limitaciones reportadas se identifican las siguientes oportunidades de mejora y desafíos para la investigación en el área:

- Dado que los procedimientos de medición dependen de contextos específicos de desarrollo donde son necesarios artefactos particulares para la aplicación del procedimiento de medición del tamaño funcional, es necesario determinar las mejores prácticas de los procedimientos existentes para cada contexto y unificar las reglas de medición (mapeo y aplicación). Asimismo, se debe definir el conjunto de las características mínimas necesarias de los artefactos de software para realizar el proceso del conteo del tamaño funcional. Cuando se construyen los artefactos de software utilizados como entrada para los procedimientos de medición, se debe pensar en el nivel de detalle necesario para obtener los resultados de medición esperados en cuanto a los resultados de exactitud.
- A partir de estos procedimientos de medición, la comunidad requiere realizar replicasiones que permitan generalizar los resultados e impulsar la transferencia tecnológica a la industria.
- Es necesario poner a disposición de las comunidades académica y profesional herramientas que apoyen el proceso de medición para los distintos contextos de aplicación y que permitan potenciar los estudios de replicación.
- La comunidad debe definir conjuntos de aplicaciones de prueba estándar que puedan ser usadas para las validaciones preliminares de las propuestas de los procedimientos de medición. Ejemplo de estas aplicaciones son el caso *Rice Cooker* (Lavazza & Del Bianco, 2009), el portafolio de aplicaciones de (Fetcke, 1999), entre otras.
- Se requiere la construcción de modelos funcionales simples que permitan mapear sus componentes, de manera natural, con los componentes funcionales de los métodos de medición del tamaño funcional. Un modelo de representación funcional de software

sencillo permite realizar el mapeo de componentes funcionales con uno o varios métodos de medición y, a partir de la recolección de datos para varios métodos, realizar comparaciones sobre las relaciones de los componentes funcionales de cada uno de los métodos evaluados.

- Es necesario definir y utilizar protocolos sistematizados para evaluar los resultados de medición del tamaño funcional. Estos no solamente deben evaluar los resultados generales de tamaño funcional, sino también los resultados específicos por componente funcional básico. Ejemplos de estos protocolos de verificación de exactitud son los propuestos en (Marín, Condori-Fernández & Pastor, 2008; Yilmaz, Tunalilar & Demirors, 2013; Soubra, Abran & Ramdane-Cherif, 2014; Quesada-López & Jenkins, 2015b).
- Gran cantidad de estudios han planteado la posibilidad de automatización del proceso de medición del tamaño funcional y han evaluado diferentes herramientas de soporte al proceso de la medición. La automatización total mediante una herramienta de medición no es trivial, por tanto, cualquier herramienta que apoye el proceso de automatización de la medición del tamaño funcional debe mantener la trazabilidad hacia los artefactos de entrada de donde se obtienen las mediciones. Además, debe permitir la interacción con los expertos de medición en el caso de que deban realizar ajustes en el proceso de la medición.

Por tanto, las principales limitaciones reportadas para los procedimientos de medición del tamaño funcional son las siguientes:

- Dependen de contextos específicos de desarrollo donde son necesarios artefactos particulares para la aplicación del procedimiento de medición.
- La construcción de los artefactos de entrada y el nivel de detalle en la construcción de estos artefactos afectan la calidad de los resultados de medición del tamaño funcional.
- En algunos casos, la construcción de los modelos funcionales para la aplicación de las reglas de mapeo se considera una tarea extensa en tiempo.

- Muchos de los procedimientos de medición del tamaño funcional se evalúan con aplicaciones pequeñas en ambientes académicos, mientras otros solo realizan pruebas de concepto con aplicaciones específicas. Esto dificulta la comparación de resultados entre los estudios y los procedimientos.
- En algunos casos, los participantes de las evaluaciones son estudiantes o profesionales en ambientes académicos.
- La mayoría de los procedimientos se concentran en el producto y no en el proceso de desarrollo.
- En la mayoría de los casos, el procedimiento de medición se limita para un solo método, lo cual no permite la comparación de métricas de tamaño funcional entre los diferentes métodos.
- La mayoría de los procedimientos de medición no utilizan un protocolo sistematizado para evaluar los resultados de la medición del tamaño funcional.
- Existen pocas replicaciones para las diferentes propuestas de los procedimientos que permitan validar y generalizar los resultados.
- No existen herramientas académicas disponibles para realizar experimentación en el campo de los procedimientos de medición del tamaño funcional.

4.2.3.2. RECOMENDACIONES PARA LA EXPERIMENTACIÓN

En cuanto a la experimentación es necesario incrementar la rigurosidad de los estudios empíricos con el fin de reportar los resultados de manera más clara para que esta información luego pueda ser utilizada para la agregación de evidencia.

Es necesario el uso de protocolos de experimentación reconocidos en el área de ingeniería de software experimental, así como el uso de métricas comunes que permitan la comparación de los resultados de los procedimientos de medición. Para habilitar la comparación de los resultados de medición, los estudios que realicen procesos de medición del tamaño funcional, deben usar aplicaciones de software similares a las utilizadas en otros estudios. Más aún, el uso de protocolos sistemáticos de verificación de la exactitud permite

estandarizar el proceso de comparación entre los resultados de los procedimientos. Estos no solo deben realizar una comparación general del total del tamaño funcional, sino también una comparación detallada de todos los componentes funcionales básicos del método de medición con el fin de mejorar la calibración de los procedimientos y las herramientas de soporte en el proceso de medición.

Además, se deben identificar las mejores prácticas para la aplicación de los procedimientos de medición del tamaño funcional que son particulares de cada contexto de desarrollo, y unificar las reglas de medición que permitan que la comunidad académica realice repeticiones. Lo anterior para mejorar la generalización de resultados e impulsar la transferencia tecnológica de la academia a la industria. Ejemplos de este tipo de trabajos son (Lavazza, del Bianco & Garavaglia, 2008; Lavazza & Garavaglia, 2009) y el estudio secundario descrito en (Barkallah , Gherbi & Abran, 2011).

En general, existen pocas repeticiones para las diferentes propuestas de los procedimientos que permitan validar y generalizar los resultados. En particular, no se encontraron herramientas académicas disponibles en línea para realizar experimentación sobre procedimientos de medición específicos, lo que limita la posibilidad de realizar repeticiones. Algunos estudios han generado instrumentos y artefactos de referencia para utilizar en las validaciones de las nuevas propuestas de procedimientos de medición los cuales los investigadores deben aprovechar.

Los principales pasos que pueden ser desarrollados para la definición y la evaluación de los procedimientos de medición del tamaño funcional son los siguientes:

1. Definición del método de medición de acuerdo con el modelo de procesos de (Jacquet & Abran, 1997; Abran, 2010).
2. Validación del diseño y validación teórica del procedimiento de acuerdo con las teorías de medición (Abran, 2010) y los métodos de medición específicos.
3. Validación y calibración del procedimiento mediante pruebas de concepto y utilización de protocolos sistemáticos de verificación. Si es posible, el proceso de medición

y validación se debe apoyar con herramientas de soporte para la automatización del proceso de medición.

4. Validación empírica a partir de aplicaciones reconocidas en el campo de investigación que han sido medidas y reportadas previamente.
5. Automatización del procedimiento de medición mediante herramientas de soporte para el proceso de medición.
6. Validación empírica mediante experimentos controlados para evaluar propiedades de efectividad y adopción.
7. Validación empírica mediante la conducción de casos de estudio con aplicaciones reales y en ambiente reales en la industria para la adaptación al contexto particular de operación.
8. Creación y evaluación de modelos de estimación o predicción para el aprovechamiento de los resultados de medición.
9. Replicación de estudios empíricos que permitan la generalización de los resultados.

4.2.3.3. PROBLEMAS ABIERTOS Y LÍNEAS DE INVESTIGACIÓN

A continuación se detallan las posibles líneas de investigación y problemas abiertos en el área:

- La formalización de los procedimientos de medición puede potenciar la automatización, mejorar la productividad, reducir los costos y mejorar la consistencia de los resultados. La evidencia empírica sobre cada una de las propuestas es esencial para entender su nivel de madurez.
- Definir procedimientos de medición estandarizados para contextos específicos de desarrollo de software, basados en las mejores prácticas y las lecciones aprendidas de los estudios empíricos realizados en el campo de la medición del tamaño funcional.

- Definir el conjunto de artefactos de software y su nivel de detalle necesario para aplicar el procedimiento de medición eficientemente y sin perder la exactitud de las mediciones.
- Definir modelos, lo más simple posibles, para la representación funcional del software que permitan aplicar sistemática y eficientemente las reglas de mapeo y de asignación numérica. Los modelos de representación del software deben reducir la brecha entre los modelos de abstracción y los metamodelos de los métodos de medición del tamaño funcional.
- Implementar herramientas de soporte para automatizar la medición del tamaño funcional, que sean lo suficientemente flexibles para permitir la interacción con los expertos de medición. Estas herramientas deben realizar los procesos de conteo eficientemente y sin perder la efectividad en los resultados de medición. Asimismo, deben permitir mantener la trazabilidad hacia los artefactos de entrada de donde se obtienen las mediciones e identificar posibilidades de visualización de los modelos funcionales del software desde la perspectiva del usuario final.
- Definir y utilizar protocolos sistematizados para evaluar los resultados de medición del tamaño funcional para cada método. Estos no solo deben evaluar los resultados totales del tamaño funcional, sino también los resultados específicos para cada uno de los componentes funcionales básicos de cada uno de los métodos de medición.
- Obtener, a partir de los modelos de representación funcional del software, el tamaño funcional de los diferentes métodos de medición, que permita realizar comparaciones e investigar las relaciones entre los componentes funcionales básicos de los diferentes métodos.
- Conducir casos de estudio con aplicaciones reales de la industria y replicaciones para los procedimientos de medición propuestos. Obtener evidencia empírica, a partir de procesos de experimentación rigurosos, sobre la efectividad de los procedimientos de medición del tamaño funcional y la relación entre los resultados de medición funcional de los diferentes métodos.

- Identificar las posibilidades de trazabilidad entre los requerimientos funcionales y los artefactos de software utilizados como entrada de los procedimientos de medición para determinar el origen del tamaño funcional reportado. Este proceso de trazabilidad debe posibilitar el seguimiento del crecimiento del software a través de las diferentes etapas del ciclo de desarrollo del software.
- Determinar patrones de medición para distintos modelos de representación del software que permita obtener el tamaño funcional de las aplicaciones de manera más rápida y manteniendo la confiabilidad de los resultados de las mediciones.
- La medición del tamaño funcional sobre modelos conceptuales y UML son las más estudiadas, existe evidencia empírica sobre su aplicabilidad y se han propuesto herramientas para su automatización. Sin embargo, el nivel de granularidad con que se detallan los artefactos de entrada impacta la exactitud de las mediciones.
- Las herramientas de automatización de medición deben ser validadas mediante protocolos sistemáticos para garantizar la confiabilidad de los resultados. No solo es necesario evaluar los resultados totales de medición, sino también los resultados obtenidos para cada uno de los componentes funcionales. Se deben formalizar marcos de trabajo para la automatización del conteo de puntos de función y validarlos mediante el uso de protocolos de verificación.

4.3. RESUMEN DE RESULTADOS

En resumen, existe una necesidad de evidencia empírica sobre la automatización de procedimientos de medición del tamaño funcional. Esto incluye el uso estandarizado de protocolos de evaluación y un mayor nivel de detalle en los reportes para facilitar las comparaciones y replicaciones de estudios. La formalización de los artefactos de entrada, así como reglas de mapeo permite un mayor nivel de automatización; sin embargo, las diferencias de los objetos utilizados en las evaluaciones empíricas limitan la comparación de los resultados. Aunque la información sobre la eficacia y eficiencia de las propuestas se

encuentra en la mayoría de los estudios, es difícil determinar todos los factores de influencia para realizar una comparación entre ellos. Existe una necesidad de procedimientos automatizados para diferentes dominios funcionales y diferentes modelos de desarrollo de software. La formalización en la definición de los artefactos de entrada y la rigurosa definición de las reglas de mapeo puede mejorar el nivel de automatización, lo que a su vez puede incrementar la productividad del proceso de medición y reducir los costos, y mejorar la consistencia.

Es necesario analizar la relación entre los artefactos de software reportados y la confiabilidad y exactitud de las mediciones. Asimismo, estudiar cuáles son los factores que influyen los resultados de medición, la efectividad de las herramientas automatizadas, el nivel de automatización sin la intervención de un experto y el cumplimiento de cada una de las fases que los métodos estandarizados de medición establecen. A partir de los resultados se debe analizar la posibilidad de establecer marcos de trabajo y de validación empírica. Finalmente, es de interés analizar los diseños experimentales y la confiabilidad de los resultados alcanzados, se debe analizar la posibilidad de realizar la agregación de los resultados a partir de los diferentes estudios que reportan sobre una misma propuesta de medición.

Los resultados del mapeo y revisión de literatura guiaron el proceso para el diseño de la metodología propuesta en el Capítulo 5 y la conducción de los estudios empíricos para la evaluación de los procedimientos de medición y verificación los cuales se describen en el Capítulo 6.

Capítulo 5. DISEÑO DE LA METODOLOGÍA PARA LA AUTOMATIZACIÓN DE LA MEDICIÓN DEL TAMAÑO FUNCIONAL DEL SOFTWARE

En este capítulo, se aborda el proceso del diseño de la metodología de medición para la consecución del objetivo específico 2 (OE2) que permite responder a la segunda pregunta de investigación (DP2) de la tesis.

Este capítulo presenta la metodología **aFPA** para la automatización de la medición del tamaño funcional del software a partir del código fuente basado en los métodos de medición estandarizados IFPUG FPA y COSMIC FFP. En las siguientes secciones se describe el diseño de la metodología la cual se lleva a cabo siguiendo las recomendaciones del modelo de procesos propuesto por (Abran, 2010; Habra, Abran, Lopez & Sellami, 2008; Jacquet & Abran, 1997). Asimismo, se presentan los procedimientos que permiten realizar la medición automática del tamaño funcional, la verificación de la exactitud de las mediciones obtenidas y la evaluación de algunos modelos de estimación de esfuerzo para mostrar un posible aprovechamiento de los resultados de la medición del tamaño funcional. Finalmente, se presenta el diseño de las herramientas prototipo que soportan cada uno de los procedimientos de la metodología **aFPA** y se discute la extensibilidad del procedimiento de medición para distintos contextos de desarrollo.

5.1. PROCEDIMIENTO DE MEDICIÓN DEL TAMAÑO FUNCIONAL DEL SOFTWARE

Este procedimiento representa el producto del ciclo ingenieril 6 (EC6).

El procedimiento de medición de la metodología **aFPA** permite la automatización del proceso de medición para modelos funcionales obtenidos a partir del análisis dinámico

del código fuente (Quesada-López, Jenkins, Salas & Gómez, 2017a). El procedimiento describe los lineamientos de diseño necesarios para desarrollar un proceso automatizado de medición, derivar los modelos funcionales a partir del análisis dinámico del código fuente y aplicar las reglas de mapeo y de asignación numérica para la de medición del tamaño funcional. El procedimiento se diseña en dos etapas:

- 1) El diseño de la medición que incluye la definición del principio de medición, el método de medición y el procedimiento de medición.
- 2) La aplicación del proceso de medición en un contexto determinado mediante el procedimiento para obtener el resultado del tamaño funcional.

En las siguientes secciones se detallan los principales aspectos de cada una de las etapas del diseño del procedimiento.

5.1.1. DISEÑO DEL PROCEDIMIENTO DE MEDICIÓN

El diseño del procedimiento de medición incluye la definición de los objetivos, la caracterización del concepto a ser medido, la definición de los metamodelos que representan el conjunto de objetos y relaciones de los componentes, la definición de las reglas de mapeo entre los componentes del metamodelo y los métodos de medición y la definición de las reglas de asignación numérica para la cuantificación del tamaño funcional. El proceso de diseño del procedimiento corresponde al Paso 1 del modelo de procesos de (Abran, 2010; Habra, Abran, Lopez & Sellami, 2008; Jacquet & Abran, 1997). La estrategia de medición se construye siguiendo las recomendaciones de trabajos relacionados en el área de procedimientos de medición tales como (Abrahamo, 2004; Hericko, Rozman & Zivkovic, 2006; Edagawa, Akaike, Higo, Kusumoto, Hanabusa & Shibamoto, 2011; Marín, 2011; Object Management Group, 2014; Gonultas & Tarhan, 2015; Tarhan, Ozkan & Içoz, 2016).

5.1.1.1. OBJETIVO DEL PROCEDIMIENTO

El objetivo es *diseñar un procedimiento de medición con el propósito de medir el tamaño funcional de aplicaciones de software a partir del análisis de los modelos funcionales obtenidos a partir del análisis dinámico del código fuente.*

El propósito del procedimiento es medir el tamaño funcional del software desde una perspectiva de un usuario final y se ajusta a las reglas de los métodos de medición IFPUG FPA y COSMIC FFP. El procedimiento puede medir el tamaño de la funcionalidad a medida que evoluciona durante el desarrollo del software y los resultados del proceso de medición pueden ser utilizados como entrada para un procedimiento de estimación del esfuerzo del desarrollo del software.

Los modelos funcionales de representación del software se generan a partir del análisis estático y dinámico del código fuente. El procedimiento determina los eventos del usuario final y rastrea las transacciones de datos durante la ejecución de las funcionalidades del software. Por tanto, el alcance de la medición es el modelo funcional obtenido a partir del análisis del análisis dinámico del código fuente. En nuestro caso, este modelo funcional delimita la frontera o los límites de la aplicación a ser medida. El modelo se construye a partir de la unión de los sub modelos de eventos, transacciones y datos, los cuales representan una arquitectura de tres capas. El modelo de eventos representa la capa de interfaz de usuario, el modelo de datos representa la capa de almacenamiento persistente y el modelo de transacciones representa la capa de negocios. Por tanto, estas capas representan las piezas de software a ser medidas las cuales se miden como una única unidad por el procedimiento de medición.

El nivel de granularidad de los artefactos se encuentra a un bajo nivel dado que estos detallan todas las funciones que se proveen al usuario final. Las mediciones se realizan a partir del modelo funcional que se extrae del código fuente de la aplicación, por lo tanto, el nivel de granularidad para el método COSMIC FFP es el Nivel 3 (sub funciones detalladas) y el nivel de conteo para el método IFPUG FPA es el Nivel 1 (trazable y etiquetado hacia

los artefactos medidos), ambos que son los niveles más detallados de conteo. Estos niveles de granularidad permiten que el conteo pueda ser auditable, con resultados confiables, bien documentados y fáciles de mantener. El tipo de conteo para el método IFPUG FPA es el conteo de puntos de función de un nuevo desarrollo de aplicación de software (DFP).

Los usuarios funcionales que interactúan con la aplicación de software a ser medida son los que envían o reciben datos durante la ejecución de las funcionalidades del mismo, quienes se identifican a partir del análisis dinámico del código fuente. Dado que el procedimiento realiza un rastreo de las transacciones de datos, un evento de un usuario funcional podría ser realizado por humanos u otras aplicaciones o piezas de software (fuera de los límites) que le solicitan o envían información a la aplicación medida. Por tanto, los usuarios funcionales que se identifican durante la medición pueden ser ya sea usuarios humanos o piezas de software mediante interfaces externas.

5.1.1.2. CARACTERIZACIÓN DEL CONCEPTO A SER MEDIDO

El procedimiento de medición se aplica sobre el modelo funcional que representa las funcionalidades que la aplicación le ofrece al usuario final. Este modelo funcional es construido con base en los eventos del usuario final y generado a partir del análisis estático y dinámico del código fuente y el rastreo de las transacciones que representan la funcionalidad. El procedimiento recibe como entrada los artefactos del código fuente de la aplicación de software. Esto incluye los archivos fuente de la aplicación y la aplicación compilada y configurada para ser ejecutada en un ambiente controlado de operación. El resultado del tamaño funcional obtenido representa la cantidad de funcionalidad representada por el modelo funcional.

Por tanto, la entidad a ser medida es *modelo funcional* y el atributo a ser medido es el *tamaño funcional* definido como “*el tamaño del software derivado de la cuantificación de los requerimientos funcionales del usuario*” (ISO, 2007). El tamaño funcional se calcula de acuerdo al método de medición IFPUG FPA versión 4.3 y COSMIC FFP versión 4.01. Este procedimiento excluye requerimientos no funcionales y requerimientos técnicos.

5.1.1.3. SELECCIÓN Y DEFINICIÓN DE LOS METAMODELOS

Un metamodelo se define como el conjunto de componentes usados para representar el software y sus relaciones (Abrahão, 2004). En el procedimiento de medición del tamaño funcional se utilizan tres metamodelos: (1) el metamodelo del método de medición IFPUG FPA, (2) el metamodelo del método de medición COSMIC FFP y (3) el metamodelo del procedimiento **aFPA** que define el modelo funcional con el cual se realiza el proceso de medición. Durante el proceso de medición, el modelo funcional de la aplicación es mapeado a cada uno de los metamodelos de los métodos de medición. En este caso, primero se construye el modelo funcional de la aplicación basada en el metamodelo del procedimiento, luego se aplican las reglas de mapeo para identificar las relaciones con los componentes de los métodos de medición, y finalmente, se aplican las reglas de asignación numérica.

5.1.1.3.1. METAMODELO IFPUG FPA

La Figura 5-1 muestra el metamodelo para el método de medición IFPUG FPA. Este detalla la información necesaria que se debe recolectar para construir la representación de la aplicación de software y para la ejecución del proceso de medición. En este caso, el procedimiento construye este metamodelo automáticamente a partir de la instanciación de su metamodelo para la aplicación de las reglas de asignación numérica.

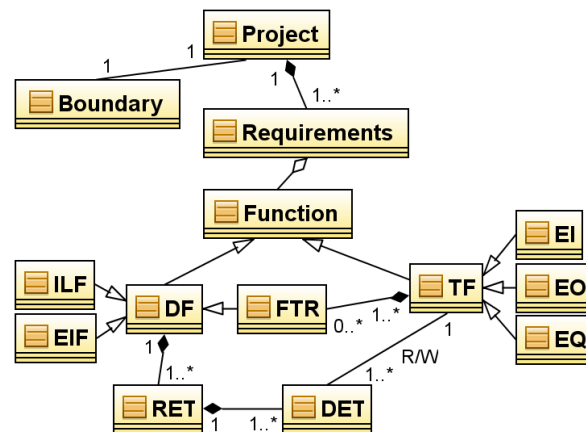


Figura 5-1 Metamodelo del método IFPUG FPA. Adaptado de (Abrahão, 2004; Hericko, Rozman & Zivkovic, 2006; Lavazza, 2015).

El método de medición IFPUG FPA requiere la identificación de los siguientes componentes:

- **Función (*Function*):** una función es un proceso elemental o grupo de datos que representa un componente funcional básico de la aplicación de software. Una función se puede clasificar como una función de datos (DF) o una función transaccional (TF).
- **Función de datos (*Data Function*):** una función de datos (DF) es un grupo de datos o información de control lógicamente relacionada e identificable por el usuario. Existen dos tipos de funciones de datos: archivos lógicos internos (ILF) e archivos de interfaces externas (EIF).
- **Función transaccional (*Transactional Function*):** una función transaccional (TF) es un proceso elemental de la aplicación que procesa datos o información de control. Existen tres tipos de funciones transaccionales: entradas externas (EI), salidas externas (EO) y consultas externas (EQ).
- **Archivos lógicos internos (*Internal Logical File*):** un archivo lógico interno (ILF) es un grupo de datos o información de control lógicamente relacionada e identificable por el usuario que se mantiene dentro de los límites de la aplicación medida. Un ILF es mantenido por uno o más procesos elementales de la aplicación medida. La complejidad se

determina por el número de elementos de tipo registro (RET) y elementos de tipo dato (DET).

- Interfaces externas (*External Interface File*): un archivo de interfaz externa (EIF) es un grupo de datos o información de control lógicamente relacionada e identificable por el usuario que se mantiene fuera de los límites de la aplicación medida. Una EIF es accedida por uno o más procesos elementales, pero no mantenida. La complejidad se determina por el número de elementos de tipo registro (RET) y elementos de tipo dato (DET).
- Entradas (*External Input*): una entrada externa (EI) es un proceso elemental de la aplicación que procesa datos o información de control que viene desde afuera de los límites de la aplicación medida. Un EI mantiene uno o más ILF y/o altera el comportamiento de la aplicación. La complejidad se determina por el número de elementos de tipo dato (DET) y tipos de archivo referenciado (FTR).
- Salidas (*External Outputs*): una salida externa (EO) es un proceso elemental que envía datos o información de control fuera de los límites de la aplicación medida. Presenta información al usuario a partir de la recuperación y procesamiento de información. La complejidad se determina por el número de elementos de tipo dato (DET) y tipos de archivo referenciado (FTR). Una EO procesa información aplicando cálculos matemáticos, mantiene al menos un ILF y altera el comportamiento de la aplicación.
- Consultas (*External Inquiry*): una consulta externa (EQ) es un proceso elemental que envía datos o información de control afuera de los límites de la aplicación medida. Presenta información al usuario a partir de la recuperación de información. La complejidad de una EQ se determina por el número de elementos de tipo dato (DET) y tipos de archivo referenciado (FTR). Una EQ no procesa información mediante la aplicación de cálculos matemáticos y no altera el comportamiento de la aplicación.
- Elemento de tipo dato (*Data Element Type*): un elemento de tipo dato (DET) es un campo o atributo único identificable por el usuario mantenido por un ILF o un EIF.
- Elemento de tipo registro (*Record Element Type*): un elemento de tipo registro (RET) es un subgrupo de datos reconocible por el usuario mantenido por un ILF o un EIF.

- Tipo de archivo referenciado (*File Types Referenced*): un tipo de archivo referenciado (FTR) es un conjunto de ILF y EIF mantenidos, leídos o referenciados por cada una de las funciones transaccionales EI, EO y EQ.

La técnica de medición de puntos de función IFPUG FPA cuenta los componentes que implementan la funcionalidad de la aplicación de software. Una aplicación de software representada mediante el método de medición IFPUG FPA se mide a partir de los cinco componentes funcionales básicos (BFC): archivos lógicos internos (ILF) e interfaces externas (EIF) y entradas (EI), salidas (EO) y consultas (EQ). El tamaño funcional se determina calculando la complejidad de cada uno de los BFC determinada por la cantidad de elementos identificados en el conteo (DET, RET, FTR). La suma de estos valores corresponde al tamaño funcional sin ajustar (UFP).

5.1.1.3.2. METAMODELO COSMIC FFP

La Figura 5-2 muestra el metamodelo para el método de medición COSMIC FFP. Este detalla la información necesaria que se debe recolectar para construir la representación de la aplicación de software y para la ejecución del proceso de medición. De igual forma que con el método IFPUG FPA, el procedimiento construye este metamodelo automáticamente a partir de la instanciación de su metamodelo para la aplicación de las reglas de asignación numérica.

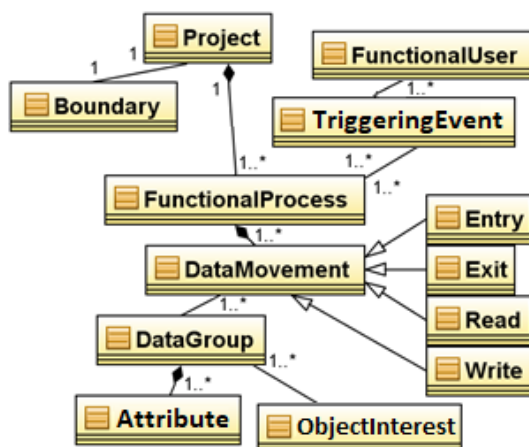


Figura 5-2 Metamodelo del método COSMIC FFP. Adaptado de (Hericko, Rozman & Zivkovic, 2006; Marín, 2011).

El método de medición COSMIC FFP requiere la identificación de los siguientes elementos:

- Evento desencadenante (*Triggering event*): es un evento conocido en los requerimientos funcionales de la aplicación medida. Cada proceso funcional es disparado por un evento desencadenante cuando es llevado a cabo por un usuario funcional.
- Proceso funcional (*Functional Process*): un proceso funcional es un conjunto de movimientos de datos que representan una parte elemental y única de los requerimientos funcionales. Cada proceso funcional es ocasionado por un evento desencadenador. Los procesos funcionales se componen de dos tipos de sub procesos que mueven o manipulan datos. Los procesos de manipulación de datos no se reconocen por separado porque se asocian a los movimientos de datos con los que están asociados.
- Movimiento de datos (*Data Movement*): un movimiento de datos es un componente funcional básico que mueve un único tipo de grupo de datos. Existen cuatro tipos de movimientos de datos: las entradas y salidas mueven un grupo de datos hacia adentro

y afuera de un proceso funcional, y de y hacia un usuario funcional. La lectura y escritura, los cuales mueven un grupo de datos entre el proceso funcional y el almacenamiento persistente.

- Entrada (*Entry*): una entrada es un movimiento de datos que mueve un grupo de datos desde un usuario funcional hacia el proceso funcional y a través de los límites de la aplicación medida.
- Salida (*Exit*): una salida es un movimiento de datos que mueve un grupo de datos desde el proceso funcional hacia un usuario funcional que los requiere y a través de los límites de la aplicación medida.
- Lectura (*Read*): una lectura es un movimiento de datos que mueve un grupo de datos desde el almacenamiento persistente hacia el proceso funcional que los requiere.
- Escritura (*Write*): una escritura es un movimiento de datos que mueve un grupo de datos que se encuentra en un proceso funcional hacia el almacenamiento persistente.
- Grupo de datos (*Data Group*): un grupo de datos es un conjunto, no vacío, no ordenado y no redundante de atributos asociados que describen el objeto de interés de un usuario funcional. Cada grupo de datos es único y distinguible y participa en uno o más movimientos de datos.
- Atributo (*Attribute*): un atributo de datos es la pieza más pequeña de información que tiene significado desde la perspectiva de los requerimientos funcionales.
- Objeto de interés (*Object of Interest*): cualquier elemento o grupo de elementos que tienen significado para el usuario funcional y se identifican en los requerimientos funcionales y sobre los que se requiere que la aplicación procese y/o almacene datos.

La técnica de medición COSMIC FFP cuenta los componentes que implementan la funcionalidad de la aplicación de software. Una aplicación de software representada mediante el método de medición COSMIC FFP se mide a partir de los cuatro tipos de movimientos de datos (DM): entradas (E) y salidas (X), y lecturas (R) y escrituras (W). El tamaño funcional se determina calculando la cantidad de cada uno de los componentes funcionales básicos. La suma de estos valores corresponde al tamaño funcional. Los movimientos de

datos se dan entre los usuarios funcionales y los procesos funcionales o su interacción con los grupos de datos.

5.1.1.3.3. METAMODELO DEL PROCEDIMIENTO DE MEDICIÓN **aFPA**

La Figura 5-3 detalla el metamodelo funcional de representación del software requerido para realizar el conteo del tamaño funcional con el procedimiento de medición **aFPA**. Este detalla la información necesaria que se debe recolectar desde la aplicación para construir el modelo funcional que permite la aplicación de las reglas de mapeo y de asignación numérica durante el proceso de medición. El modelo funcional se construye a partir del análisis de los eventos y transacciones obtenidas durante la simulación de la ejecución de los procesos funcionales. La aplicación de software en ejecución permite la construcción del modelo funcional orientado a eventos y transacciones.

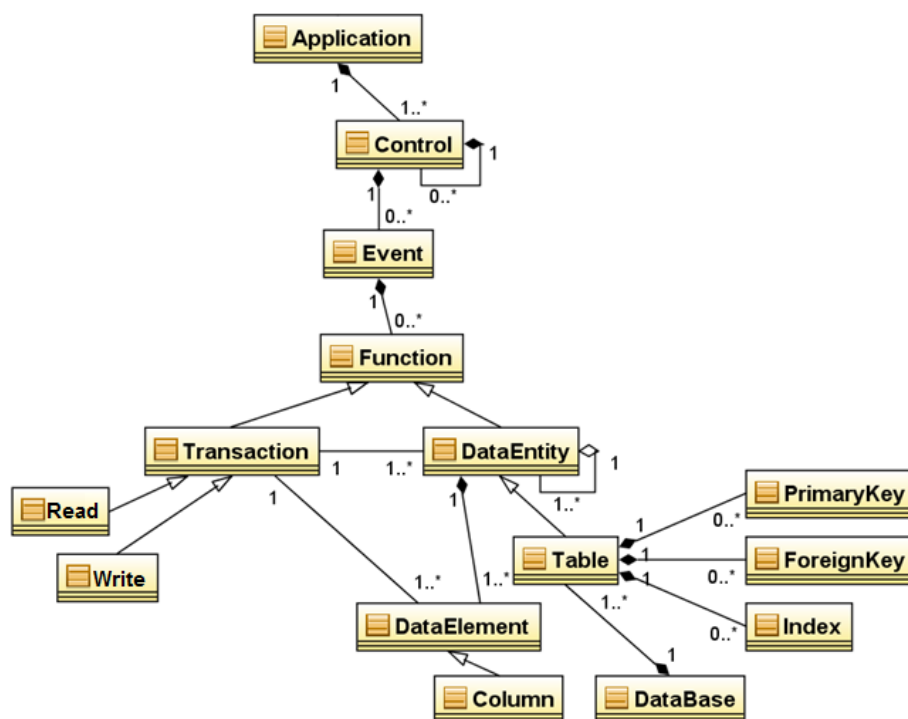


Figura 5-3 Metamodelo del procedimiento **aFPA**.

El modelo funcional del procedimiento **aFPA** requiere la identificación de los siguientes elementos:

- *Aplicación (Application)*: representa la aplicación de software a ser medida. Los artefactos de entrada del código fuente determinan los límites de la aplicación. Una aplicación se compone de uno o más controles.
- *Control (Control)*: es un componente de interfaz de usuario y puede contener otros controles. El usuario funcional accede los procesos funcionalidades mediante la ejecución de los eventos asociados a los controles. Un control puede tener asociado ningún evento o uno o más eventos. Un componente de interfaz de usuario puede representar una interfaz humano computador o una interfaz con otras aplicaciones de software. Ejemplos: ventanas, botones, interfaces, campos de edición, tablas de datos, entre otros.
- *Evento (Event)*: permite la ejecución de una opción que la aplicación de software medida proporciona a un usuario. Un evento puede ejecutar ninguna o muchas funciones transaccionales, por tanto puede representar un evento de navegación o de función y es clasificado como tipo función o navegación. La navegación permite ir a otro control de interfaz de usuario sin realizar ninguna transacción de datos y una función realiza accesos a las entidades de datos por medio de transacciones de lectura o escritura.
- *Función (Function)*: una función puede representar un proceso funcional, una función transaccional, una función de datos o un grupo de datos. Se clasificada como una transacción de lectura [seleccionar, listar] o escritura [insertar, actualizar, borrar] o una entidad de datos. Una transacción de datos accede una entidad de datos de la aplicación mediante la ejecución de una funcionalidad del usuario. Ambos tipos de función son identificados en el proceso de construcción del modelo funcional.
- *Transacción (Transaction)*: es un conjunto de flujos de datos hacia la base de datos realizados como una sola unidad y en orden. Esto se da desde que el usuario funcional dispara un evento en la interfaz hasta que la aplicación devuelve el control al usuario funcional. Una transacción de datos accede una o varias entidades de datos de la aplicación medida por medio de operaciones de lectura (*read*) y/o escritura (*write*).

- Entidad de datos (*DataEntity*): una entidad de datos es una tabla de la base de datos, archivo o un grupo de tablas o archivos (almacenamiento persistente) que representan un agrupamiento lógico. Un agrupamiento lógico es el conjunto de todas las entidades de datos que se relacionan entre sí y que son vistas como una única unidad desde la perspectiva del usuario funcional. Cada entidad de datos se compone de elementos de datos. Las relaciones entre grupos de entidades se referencian por medio de llaves foráneas y/o accesos simultáneos a las entidades de datos.
- Elemento de datos (*DataElement*): es un atributo o columna de un archivo o tabla de base de datos que puede representar un campo físico o calculado y que tiene significado para el usuario funcional.
- Base de datos (*DataBase*): representa la o las bases de datos (almacenamiento persistente) con las cuales la aplicación interactúa durante la ejecución de las funcionalidades del usuario funcional.

5.1.1.3.3.1 ESPECIFICACIÓN DEL MODELO FUNCIONAL

La extracción y generación del modelo funcional a partir del código fuente de la aplicación medida es una de las tareas esenciales para obtener mediciones confiables y consistentes del tamaño funcional a partir del procedimiento de medición **aFPA**. Una especificación detallada del proceso de construcción del modelo es necesaria, ya que dependiendo de la complejidad de la aplicación de software esta puede ser una tarea compleja de administrar. A continuación se especifica el modelo funcional que permite la identificación de los elementos a partir del código fuente y la construcción de la representación funcional de la aplicación medida. El modelo funcional detalla todos los componentes que forman parte de los procesos funcionales y que representan el comportamiento de la aplicación. El modelo se construye a partir del análisis de los eventos y las transacciones que conforman la

funcionalidad de la aplicación de software manteniendo la trazabilidad de la secuencia de interacciones realizadas por los usuarios funcionales.

El modelo funcional se construye a partir de tres sub modelos.

- 1) El *modelo de eventos* representa la capa de interfaz de usuario. En la construcción de este modelo se identifican todos los controles que representan las interfaces con los usuarios funcionales. Para cada control, se identifican todos los componentes asociados que cuenten con eventos desencadenadores en la aplicación. Para cada evento se identifican y registran todas las entradas requeridas para la ejecución del evento. Si el evento desencadena una transacción de datos, se registra la transacción como un evento funcional y se detallan los accesos realizados en el almacenamiento persistente. Si no, se registra la transacción como un evento de navegación. Para cada evento se registra el control de interfaz y el control específico que generó el evento, el nombre del evento y la lista de funciones que representan las operaciones transaccionales que intenta realizar la aplicación hacia y desde el almacenamiento persistente. El modelo de eventos representa todas las funcionalidades que un usuario funcional puede realizar desde la interfaz de la aplicación medida. El modelo de eventos detalla el mapa de la interfaz de la aplicación con cada uno de los componentes y los eventos asociados a ellos.
- 2) El *modelo de datos* representa la capa de almacenamiento persistente. En la construcción de este modelo se identifican todas las entidades de datos o grupos de datos en almacenamiento persistente (bases de datos y otras estructuras de datos) que son accedidas por los procesos funcionales. Para cada entidad de datos, se identifican los elementos de datos o atributos relacionados, los identificadores únicos y las relaciones entre grupos de datos. El modelo de datos asocia cada entidad con sus elementos de datos y relaciona las entidades de datos entre sí. El modelo de datos representa las relaciones entre entidades de datos y elementos de datos. Una vez construido el modelo de eventos y el modelo de datos se procede con la construcción del modelo de transacciones que permite obtener el modelo funcional.

- 3) El *modelo de transacciones* representa la capa de negocios que contiene las transacciones realizadas por la aplicación medida. En la construcción de este modelo se identifican los procesos funcionales (transacciones) que dispara un componente de interfaz mediante un evento en las entidades de datos y cada uno de sus atributos. Si un mismo evento realiza múltiples accesos a una entidad o elemento de datos se eliminan los duplicados y se incrementa el peso para este acceso. Cada uno de los accesos a los datos se marca como una transacción de lectura o escritura de acuerdo a las funcionalidades realizadas por la aplicación medida. El modelo de transacciones registra todas las transacciones realizadas sobre cada elemento de datos desde la interfaz de usuario. Con base en la información generada a partir de los tres modelos se construye el modelo funcional de la aplicación medida.

La instanciación del metamodelo del procedimiento de medición **aFPA** se realiza mediante una estructura de grafo dirigido¹⁰, etiquetado¹¹ y ponderado¹² donde sus nodos representan cada uno de los elementos funcionales identificados en la aplicación y las aristas representan las relaciones entre cada uno de ellos. El grafo G se define como un par (N, E) , donde N es un conjunto finito de elementos no vacío y $E \subseteq N \times N$. El par (N, E) es un grafo dirigido sobre N , donde N es el conjunto de nodos (vértices) y E es su conjunto de aristas que contienen pares ordenados de elementos de N . Escribimos $G = (N, E)$ para denotar el grafo que representa el modelo funcional y el conjunto de elementos N y sus relaciones E . Donde:

$$N \in T_v$$

$$T_v = \{Application, Control, Event, Function, Read, Write, Table, Column, PrimaryKey, ForeignKey, Index, DataBase\}.$$

$$E \in T_e$$

¹⁰ La arista entre dos nodos tiene dirección.

¹¹ Cada arista y/o nodo tiene una etiqueta o valor.

¹² Cada arista tiene un peso.

$$T_e = \{Father, Relationship\}.$$

La Figura 5-4 muestra el modelo de clases para representar un grafo con el modelo funcional del procedimiento de medición **aFPA**. En este grafo cada nodo es etiquetado con un tipo de elemento: *Application*, *Control*, *Event*, *Function*, *Read*, *Write*, *Table*, *Column*, *PrimaryKey*, *ForeignKey*, *Index*, *DataBase*, y cada arista asocia un par de nodos con una relación de tipo: *Father*, *Relationship*. El modelo funcional representa una abstracción de estados concretos de la funcionalidad de la aplicación de software que combina información estática y dinámica a partir del código fuente para modelar el comportamiento funcional.

Cada uno de los nodos del grafo almacena las propiedades que describen el componente funcional y que son actualizadas en el proceso de aplicación de las reglas de mapeo y de asignación numérica. Las propiedades que se administran para cada nodo son las siguientes:

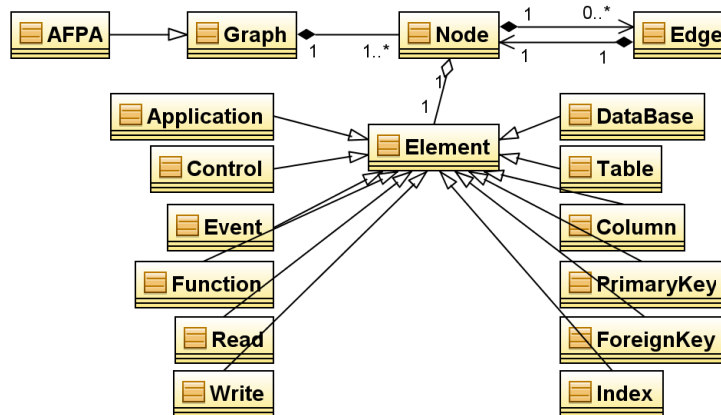


Figura 5-4 Modelo de clases para la representación del modelo funcional.

- *Application*: identificador y nombre del nodo en relación al componente funcional, nombre de la aplicación medida y un número *hash* como identificador único dentro del grafo.
- *Control*: identificador y nombre del nodo en relación al componente funcional, nombre del control y un número *hash* como identificador único dentro del grafo. Un nodo tipo *UserInterface* es un nodo *Control* que representa una ventana de la aplicación.

- *Event*: identificador y nombre del nodo en relación al componente funcional, nombre y descripción del evento, comentario sobre el requerimiento y un número *hash* como identificador único dentro del grafo.
- *Function*: identificador y nombre del nodo en relación al componente funcional y un número *hash* como identificador único dentro del grafo. La función se clasifica con el tipo *Select*, *Insert*, *Update*, *Delete* que describe las transacciones de lectura y escritura que realiza con los grupos de datos.
- *Read*: identificador y nombre del nodo en relación al componente funcional, nombre y esquema de la base de datos, nombre de la operación de lectura, tipo de transacción, tipo interno o externo, tabla técnica y un número *hash* como identificador único dentro del grafo.
- *Write*: identificador y nombre del nodo en relación al componente funcional, nombre de la operación de escritura, tipo de transacción y un número *hash* como identificador único dentro del grafo.
- *Table*: identificador y nombre del nodo en relación al componente funcional, nombre y esquema de la base de datos, nombre de la tabla, tipo interno o externo, grupo de datos, tabla técnica y un número *hash* como identificador único dentro del grafo.
- *Column*: identificador y nombre del nodo en relación al componente funcional, nombre, esquema, bases de datos, tabla, nombre de la columna, tipo de dato, largo, tipo campo o calculado y un número *hash* como identificador único dentro del grafo.
- *PrimaryKey*: identificador y nombre del nodo en relación al componente funcional, nombre, esquema, bases de datos, tabla, columna, nombre de la llave y un número *hash* como identificador único dentro del grafo.
- *ForeignKey*: identificador y nombre del nodo en relación al componente funcional, nombre, esquema, bases de datos, tabla, columna, nombre de la llave, tabla y columna foránea y un número *hash* como identificador único dentro del grafo.
- *Index*: identificador y nombre del nodo en relación al componente funcional, nombre, esquema, bases de datos, tabla, columna, nombre del índice, si es llave primaria, si es único y un número *hash* como identificador único dentro del grafo.

- *DataBase*: identificador y nombre del nodo en relación al componente funcional, nombre de la base de datos, tipo local o externa y un número *hash* como identificador único dentro del grafo.

El modelo funcional permite un conjunto específico de relaciones entre los componentes representados por los nodos en el grafo. Las aristas entre tipos de nodo son las siguientes:

- *Application*<-*Control*: el nodo aplicación tiene una arista *Father* con uno o más nodos de interfaz de usuario que se representan por los nodos control. Un nodo aplicación se asocia con uno o más nodos control que tienen asociados los nodos evento.
- *Control*<-*Event*, *Event*<-*Control*: el nodo control tiene una arista *Father* con los nodos evento. Un nodo control puede tener asociado ningún nodo evento o uno o más nodos evento que representa la funcionalidad que se desencadena con la ejecución de un evento del control. Del mismo modo, un nodo evento puede tener asociado ningún nodo control o uno o más nodos control que representa una asociación de navegación hacia otro componente de interfaz en un evento de navegación.
- *Event*<-*Function*: el nodo evento tiene una arista *Father* con los nodos función. Un nodo evento puede tener asociado ningún nodo función o uno o más nodos función.
- *Function*<-*Read*, *Function*<-*Write*: el nodo función tiene una arista *Father* con los nodos lectura y escritura. Los nodos de transacciones lectura y escritura realizan accesos a los grupos de datos como una sola unidad los cuales asocian uno o más nodos tabla y columna. El nodo función se clasifica con el tipo *Select*, *Insert*, *Update*, *Delete* a partir del análisis de las relaciones con los nodos transacciones de lectura y escritura.
- *Read*<-*Table*, *Read*<-*Column*, *Write*<-*Table*, *Write*<-*Column*: los nodos de transacciones lectura y escritura tiene una arista *Father* con los nodos tabla y columna. Los nodos de transacciones lectura y escritura realizan accesos a uno o más nodos tabla y columna.
- *Table*<-*Column*, *Table*<-*PrimaryKey*, *Table*<-*ForeignKey*, *Table*<-*Index*: el nodo tabla tiene una arista *Father* con los nodos columna, llave e índice. Un nodo tabla se asocia

con uno o más nodos columna y con ningún nodo llave e índice o uno o más nodos llave e índice.

- *Column* \leftarrow *ForeignKey*, *ForeignKey* \leftarrow *Column*: el nodo columna tiene una arista *Relationship* con los nodos llave foránea y viceversa. Esta relación denota una asociación de entidades que en algunos casos representa un grupo lógico de datos.
- *DataBase* \leftarrow *Table*: el nodo base de datos tiene una arista *Father* con los nodos tabla. Un nodo base de datos se asocia con uno o más nodos tabla.

Un ejemplo de la construcción del grafo del modelo funcional se presenta en la Sección 5.1.2.3. El modelo funcional del procedimiento **aFPA** es un grafo G que representa la aplicación de software a ser medida A y se define como $G_A = (N, E)$. Un grupo de datos (función de datos) D se define como un conjunto finito de componentes no vacío y sus relaciones (C, R) donde $R \subseteq C \times C$, $C \in \{Table, Column, PrimaryKey, ForeignKey, Index, DataBase\}$ y $R \in \{Father, Relationship\}$. El grupo de datos D_i se compone de un conjunto de componentes que cumplen con las siguientes restricciones de asociación de nodos:

$$D_i = \{(DataBase_{i1}, Table_{ij}), (Table_{i1}, Column_{ij}), (Table_{i1}, PrimaryKey_{ij}), (Table_{i1}, ForeignKey_{ij}), (Table_{i1}, Index_{ij}), (Column_{ij}, ForeignKey_{ij})\}.$$

Un proceso funcional (función transaccional) F se define como un conjunto finito de componentes no vacío y sus relaciones (C, R) donde $R \subseteq C \times C$ y $C \in \{Application, Control, Event, Function, Read, Write, Table, Column\}$ y $R \in \{Father\}$. El proceso funcional F_i es un conjunto de componentes que cumple con las siguientes restricciones de asociación de nodos:

$$F_i = \{(Application_{i1}, Control_{ij}), (Control_{i1}, Event_{ij}), (Event_{i1}, Function_{ij}), (Function_{i1}, Read_{ij}), (Function_{i1}, Write_{ij}), (Read_{i1}, Table_{ij}), (Write_{i1}, Table_{ij}), (Read_{i1}, Column_{ij}), (Write_{i1}, Column_{ij})\}.$$

5.1.1.4. REGLAS DE MAPEO ENTRE METAMODELOS

A partir de la construcción de los metamodelos del procedimiento de medición, el proceso de aplicación de las reglas de mapeo determina los candidatos de cada uno de los componentes funcionales del método de medición correspondiente. Las reglas de mapeo determinan la asociación entre los elementos de los metamodelos. Para realizar la medición del modelo funcional basado en los métodos de medición IFPUG FPA y COSMIC FFP, el metamodelo del procedimiento **aFPA** es mapeado al metamodelo de cada uno de los métodos de medición. El mapeo se realiza sobre un subgrafo que representa los procesos funcionales de acuerdo a las reglas de medición de cada uno de los métodos. El grafo permite mantener la trazabilidad entre el conteo funcional, cada uno de los componentes de datos y los componentes de la interfaz de la aplicación que representan los requerimientos del usuario funcional. En este caso, el procedimiento de medición se encarga de instanciar cada metamodelo de un método de medición automáticamente a partir del grafo del modelo funcional. Posterior al proceso de mapeo se realiza la aplicación de las reglas de asignación numérica para cada uno de los métodos de medición que permite obtener el tamaño funcional.

El proceso de mapeo es el encargado de determinar los candidatos de los componentes funcionales que son analizados por el procedimiento de medición para identificar si cumplen con las características de un elemento funcional desde la perspectiva del usuario funcional. Posterior a esto, cada elemento es cuantificado de acuerdo a las reglas específicas de cada método de medición. El objetivo de las reglas de mapeo es establecer las relaciones entre los conceptos de cada uno de los metamodelos lo cual permite identificar los componentes funcionales básicos para el conteo del tamaño funcional.

5.1.1.4.1. MAPEO CON EL MÉTODO IFPUG FPA

La Figura 5-5 muestra el mapeo de componentes entre el metamodelo del procedimiento de medición **aFPA** y el metamodelo del método de medición IFPUG FPA. Para el procedimiento de medición, el alcance de la medición es el modelo funcional que determina

los límites de la aplicación medida. El tipo de conteo aplicado es el de un nuevo desarrollo de aplicación de software (DFP). El nivel de granularidad es el Nivel 1 (trazable y etiquetado hacia los artefactos medios).

Los usuarios funcionales identificados a partir de las interacciones realizadas en el modelo funcional son el usuario humano y el usuario aplicación externa. Un usuario humano interactúa con la aplicación de software por medio de las interfaces de usuario y el usuario aplicación externa por medio de cualquier otra interfaz. Las reglas de mapeo para la identificación de los usuarios son las siguientes:

- R1. Se identifica un *usuario humano* cuando un nodo tipo *event* se desencadena por medio de un nodo *control* tipo interfaz de usuario.
- R2. Se identifica un *usuario aplicación externa* cuando un nodo tipo *event* se desencadena por medio de un nodo *control* tipo interfaz.
- R3. Se identifica como el límite de la aplicación el que se encuentra entre un nodo *control* tipo interfaz de usuario y un nodo tipo *event*.
- R4. Se identifica como el límite de la aplicación el que se encuentra entre un nodo *control* tipo interfaz y un nodo tipo *event*.

las DF y sus RET, las entidades de datos se clasifican como candidatas a archivos lógicos internos (ILF) o interfaces externas (EIF).

En el modelo **aFPA**, un elemento *function* se mapea con el concepto función transaccional (TF), un elemento *transaction read* se mapea con una salida (EO) o una consulta (EQ). Un elemento *transaction write* se mapea con una entrada (EI), un elemento *DataEntity* con un archivo lógico interno (ILF), interfaz externa (EIF), función de datos (DF), elemento de tipo registro (RET) y/o elemento de tipo archivo referenciado (FTR). Finalmente, un *DataElement* se relaciona con el concepto de elemento de tipo dato referenciado (DET).

Los candidatos son analizados por el procedimiento de medición para determinar si cumplen las características de un elemento funcional. Las reglas generales de mapeo entre los componentes del metamodelo del procedimiento y del método de medición IFPUG FPA son las siguientes:

- R5. Se identifica un candidato de tipo función transaccional (TF) para cada nodo tipo *event* si este tiene asociado uno o varios nodos tipo *function*. En caso contrario se identifica como un nodo *event* de tipo navegación.
- R6. Se identifica, por cada nodo tipo *function*, un candidato de tipo dato referenciado (DET) para cada nodo tipo *column* asociado.
- R7. Se identifica, por cada nodo tipo *function*, un candidato de tipo archivo referenciado (FTR) para cada nodo tipo *table* asociado.
- R8. Se identifica, por cada nodo tipo *function*, los nodos tipo *read* y *write* asociados. Por cada nodo tipo *function* los nodos tipo *column* asociados. Si existe al menos una asociación con un nodo tipo *write* se identifica como un candidato de tipo operación de entrada (EI). Si existen solo asociaciones con nodos de tipo *read* y se identifican asociaciones solo con nodos *column* tipo *field* se identifica como un candidato de tipo de operación de consulta (EQ), si no se identifica como un candidato de tipo de operación de salida (EO). En el caso de los tipos EI, no se consideran como elementos DET los nodos tipo *column* asociados que han sido marcados como acceso de solo lectura.

- R9. Se identifica, por cada nodo tipo *read* o *write*, los nodos tipo *table* asociados. Si la cantidad de nodos tipo *table* asociados es mayor que uno, entonces se identifica si pertenecen a un mismo grupo lógico de datos. En el caso de que el almacenamiento persistente sea una base de datos relacional, se identifica un candidato a grupo de datos lógico si: (a) existe una asociación maestro-detalle entre los nodos de tipo *table* que se determinan por los nodos tipo *foreign key* que se asocian con un nodo tipo *primary key* del nodo *table* y (b) existe una asociación dada por múltiples nodos tipo *primary key* que se relacionan con nodos tipo *foreign key* de todos los nodos tipo *column* de un nodo tipo *table* distinto. Si se identifica una asociación entre nodos tipo *table*, estos pertenecen a un agrupamiento de datos lógico y se asocian en un único grupo lógico candidato de tipo archivo referenciado (FTR). Si solo existe un nodo tipo *table* asociado entonces identificamos que no pertenece a un grupo lógico existente, si es así, se identifica como un nuevo grupo lógico candidato de tipo archivo referenciado (FTR). Esta asociación sustituye las existentes para cualquier candidato de tipo archivo referenciado (FTR) identificado en reglas anteriores.
- R10. Se identifica un candidato de función de datos (DF) para cada grupo lógico único identificado y se mapea de la siguiente manera: (a) se identifica un candidato de tipo archivo lógico interno (ILF) si al menos tiene una asociación con un nodo tipo *write* y (b) se identifica un candidato de tipo interfaz externa (EIF) si solo se asocia con nodos tipo *read*. Todos los nodos de tipo *table* que no se encuentran asociados con un nodo tipo *function* no se consideran parte de la funcionalidad y son marcados como componentes técnicos que no se consideran en el conteo del tamaño funcional.
- R11. Solo se identifican como un candidato de tipo de operación de consulta (EQ) las que al menos cuenten con un mínimo de un FTR.

Después de aplicar las reglas de mapeo se reportan todos los resultados, las decisiones tomadas en el proceso, los componentes no considerados y la trazabilidad a cada uno de los componentes del modelo funcional. Del mismo modo se identifican y reportan los

procesos funcionales identificados como posibles duplicados a partir de la comparación de los caminos de cada una de las funcionalidades.

5.1.1.4.2. MAPEO CON EL MÉTODO COSMIC FFP

La Figura 5-6 muestra el mapeo de componentes entre el metamodelo del procedimiento de medición **aFPA** y el metamodelo del método de medición COSMIC FFP. El modelo se construye a partir de la unión de los sub modelos de eventos, transacciones y datos los cuales representan una arquitectura de tres capas para las piezas de software a ser medidas por el procedimiento. En este caso, el procedimiento realiza la medición de la funcionalidad de la aplicación sin detallar el tamaño de cada una de las capas de la aplicación por separado, sino como una única pieza de software.

El nivel de granularidad para la aplicación del método COSMIC FFP es el Nivel 3 (sub funciones detalladas) que representa el nivel más bajo de granularidad. El alcance de la medición es el modelo funcional que determina los límites de la aplicación medida. El procedimiento realiza la medición utilizando como entrada el modelo funcional donde se identifican los procesos funcionales que se desencadenan a partir de los eventos desde los componentes de interfaz de cada uno de los usuarios funcionales.

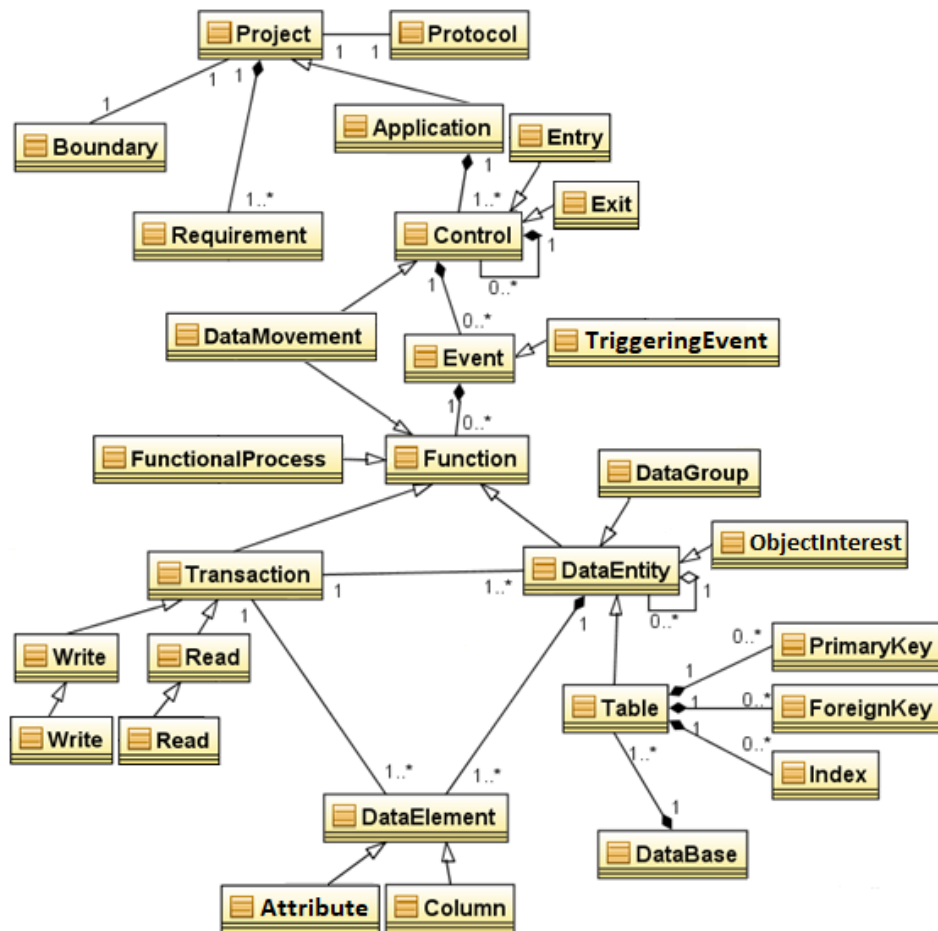


Figura 5-6 Mapeo con el metamodelo COSMIC FFP.

Los usuarios funcionales identificados a partir de las interacciones realizadas en el modelo funcional son el usuario humano y el usuario aplicación externa. Un usuario humano interactúa con la aplicación de software por medio de las interfaces humano computador y el usuario aplicación externa por medio de cualquier otra interfaz. Las reglas de mapeo para la identificación de los usuarios son las siguientes:

- R1. Se identifica un *usuario humano* cuando un nodo tipo *event* se desencadena por medio de un nodo *control* tipo interfaz de usuario.
- R2. Se identifica un *usuario aplicación externa* cuando un nodo tipo *event* se desencadena por medio de un nodo *control* tipo interfaz.

- R3. Se identifica como el límite de la aplicación el que se encuentra entre un nodo *control* tipo interfaz de usuario y un nodo tipo *event*.
- R4. Se identifica como el límite de la aplicación el que se encuentra entre un nodo *control* tipo interfaz y un tipo *event*.

Con base en el modelo funcional se identifican los eventos desencadenantes y se determinan los candidatos a procesos funcionales. Posteriormente, se clasifican los candidatos a cada una de los movimientos de datos por cada uno de los procesos funcionales de acuerdo al flujo de datos y los tipos de acceso realizado sobre los objetos de interés. De acuerdo a los accesos simultáneos realizados en los grupos de datos se reportan los procesos funcionales únicos.

En el modelo **aFPA**, un elemento *event* se mapea con el concepto de evento desencadenante y un elemento *function* con un proceso funcional. Un elemento *transaction read* se mapea con un movimiento de datos de lectura (R), un elemento *transaction write* con un movimiento de datos de escritura (W), un elemento *control* de interfaz de usuario se mapea con un movimiento de datos de entrada (E) o salida (X) a partir del flujo de datos y el tipo de accesos que realiza la aplicación en los objetos de interés. Finalmente, un elemento *DataEntity* se mapea con un grupo de datos y un *DataElement* con un atributo.

Los candidatos de componentes son analizados por el procedimiento de medición para determinar si cumplen las características de un componente funcional y un proceso funcional. Las reglas generales de mapeo entre los componentes del metamodelo del procedimiento y del método de medición COSMIC FFP son las siguientes:

- R5. Se identifica un candidato de un proceso funcional para cada nodo tipo *function* si este se encuentra asociado a un nodo tipo *event*.
- R6. Se identifica cada nodo tipo *event* como candidato de evento desencadenante.
- R7. Se identifica un candidato de un movimiento de datos para cada transacción. Se identifica un candidato de lectura (R) para cada nodo tipo *read* y un candidato de escritura (W) para cada nodo tipo *write*.

- R8. Se identifica un candidato de un movimiento de entrada (E) para cada nodo *control* de tipo interfaz y un movimiento de salida (X) para cada nodo *control* de tipo interfaz asociado a un proceso funcional.
- R9. Se identifica por cada nodo tipo *function* un candidato de grupo de datos para cada nodo tipo *table* asociado. Los candidatos a grupo de datos lógico se agrupan si: (a) existe una asociación maestro-detalle entre los nodos de tipo *table* que se determinan por los nodos tipo *foreign key* que se asocia con un nodo tipo *primary key* del nodo *table* y (b) existe una asociación dada por múltiples nodos tipo *primary key* que se relacionan con nodos tipo *foreign key* de todos los nodos tipo *column* de un nodo tipo *table* distinto. Si se identifica una asociación entre nodos tipo *table*, estas pertenecen a un agrupamiento de datos lógico y se asocian en un único grupo lógico candidato. Si solo existe un nodo tipo *table* asociado entonces identificamos que no pertenece a un grupo lógico existente, si es así, se identifica como un nuevo grupo lógico candidato.
- R10. Se identifica por cada nodo tipo *function* un candidato de atributo para cada nodo tipo *column* asociado al nodo *function* y a un nodo tipo *table*.
- R11. Se reclasifica la identificación de un candidato de movimiento de lectura (R) para cada nodo tipo *read*, el cual se encuentra asociado a un candidato de grupo lógico de datos que solo sea utilizado con acceso de lectura, como un candidato de movimiento de salida (X). El grupo lógico de datos se considera una interfaz fuera de los límites de la aplicación.
- R12. Solo se identifican como procesos funcionales candidatos los flujos de datos con un mínimo de dos movimientos de datos que incluyan al menos un movimiento de entrada (E) y un movimiento de escritura (W) o un movimiento de salida (X).

Las reglas de mapeo del procedimiento de medición pueden ser aplicadas automáticamente por la herramienta prototipo descrita en la Sección 5.2.1. Esta herramienta, después de aplicar las reglas de mapeo, reporta todos los resultados, las decisiones tomadas en el proceso, los componentes no considerados y la trazabilidad a cada uno de los componentes del modelo funcional. Del mismo modo es posible identificar y reportar los procesos

funcionales reconocidos como posibles duplicados a partir de la comparación de los caminos de cada una de las funcionalidades.

5.1.1.5. REGLAS DE ASIGNACIÓN NUMÉRICA

A partir del mapeo e identificación de cada uno de los componentes funcionales para cada uno de los métodos de medición, se definen las reglas de asignación numérica para establecer un valor numérico al modelo funcional. Este valor numérico representa el tamaño funcional de la aplicación medida. Para obtener el tamaño del modelo funcional del procedimiento **aFPA**, se definen las reglas de medición para los métodos IFPUG FPA y COSMIC FFP de acuerdo a las reglas estandarizadas para cada uno de estos métodos. Dado que el procedimiento de medición del tamaño funcional se basa en los métodos IFPUG FPA y COSMIC FFP, existen correspondencias que permiten la aplicación de sus reglas de medición sobre los componentes funcionales identificados durante la aplicación de las reglas de mapeo.

5.1.1.5.1. ASIGNACIÓN NUMÉRICA PARA EL MÉTODO IFPUG FPA

Las reglas de asignación numérica se definen de conformidad con las establecidas por el método de medición IFPUG FPA. Por tanto, el tamaño funcional de la aplicación medida es igual al *número de puntos de función sin ajustar (UFP)*. El procedimiento de medición asigna los valores de los UFP para cada una de las funciones transaccionales y de datos identificadas y el total del tamaño funcional de la aplicación es la sumatoria de cada una de ellas.

La complejidad de una entrada (EI) se calcula por el número de tipos de archivo referenciados (FTR) y elementos de tipo dato (DET) asociados. Con el número de componentes FTR y DET identificados en la aplicación a partir de las reglas de mapeo, la entrada se clasifica como una función transaccional de complejidad baja, media o alta de acuerdo a las siguientes reglas:

$$Complejidad_{EI} = \left\{ \begin{array}{l} \text{Baja:} \\ ((0 \leq FTR \leq 1) \wedge (1 \leq DET \leq 15)) \vee \\ ((FTR = 2) \wedge (1 \leq DET \leq 4)) \\ \text{Media:} \\ ((0 \leq FTR \leq 1) \wedge (DET \geq 16)) \vee \\ ((FTR = 2) \wedge (5 \leq DET \leq 15)) \vee \\ ((FTR \geq 3) \wedge (1 \leq DET \leq 4)) \\ \text{Alta:} \\ ((FTR = 2) \wedge (DET \geq 16)) \vee \\ ((FTR \geq 3) \wedge (DET \geq 5)) \end{array} \right.$$

La complejidad de una salida (EO) y de una consulta (EQ) se calcula por el número de tipos de archivo referenciados (FTR) y elementos de tipo dato (DET) asociados. Con el número de componentes FTR y DET identificados en la aplicación a partir de las reglas de mapeo, la salida y consulta se clasifica como una función transaccional de complejidad baja, media o alta de acuerdo a las siguientes reglas:

$$Complejidad_{EO,EQ} = \left\{ \begin{array}{l} \text{Baja:} \\ ((0 \leq FTR \leq 1) \wedge (1 \leq DET \leq 19)) \vee \\ ((2 \leq FTR \leq 3) \wedge (1 \leq DET \leq 5)) \\ \text{Media:} \\ ((0 \leq FTR \leq 1) \wedge (DET \geq 20)) \vee \\ ((2 \leq FTR \leq 3) \wedge (6 \leq DET \leq 19)) \vee \\ ((FTR \geq 4) \wedge (1 \leq DET \leq 5)) \\ \text{Alta:} \\ ((2 \leq FTR \leq 3) \wedge (DET \geq 20)) \vee \\ ((FTR \geq 4) \wedge (DET \geq 6)) \end{array} \right.$$

La complejidad de un archivo lógico interno (ILF) y de un archivo de interfaz externa (EIF) se calcula por el número de elementos de tipo registro (RET) y elementos de tipo dato (DET) asociados. Con el número de componentes RET y DET identificados en la aplicación a partir de las reglas de mapeo, los archivos lógicos internos y las interfaces externas se clasifican como una función de datos de complejidad baja, media o alta de acuerdo a las siguientes reglas:

$$Complejidad_{ILF,EIF} = \left\{ \begin{array}{l} \text{Baja:} \\ ((RET = 1) \wedge (1 \leq DET \leq 50)) \vee \\ ((2 \leq RET \leq 5) \wedge (1 \leq DET \leq 19)) \\ \text{Media:} \\ ((RET = 1) \wedge (DET \geq 51)) \vee \\ ((2 \leq RET \leq 5) \wedge (20 \leq DET \leq 50)) \vee \\ ((RET \geq 6) \wedge (1 \leq DET \leq 19)) \\ \text{Alta:} \\ ((2 \leq RET \leq 5) \wedge (DET \geq 51)) \vee \\ ((RET \geq 6) \wedge (DET \geq 20)) \end{array} \right.$$

Tal como se muestra en la Tabla 5-1, el tamaño en puntos de función sin ajustar se calcula de acuerdo a la complejidad de cada uno de los componentes funcionales básicos de la siguiente manera: un EI y EQ de complejidad baja agrega 3 puntos de función, de complejidad media agrega 4 puntos de función y de complejidad alta agrega 6 puntos de función. Un EO de complejidad baja agrega 4 puntos de función, de complejidad media agrega 5 puntos de función y de complejidad alta agrega 7 puntos de función. Un ILF de complejidad baja agrega 7 puntos de función, de complejidad media agrega 10 puntos de función y de complejidad alta agrega 15 puntos de función. Finalmente, una EIF de complejidad baja agrega 5 puntos de función, de complejidad media agrega 7 puntos de función y de complejidad alta agrega 10 puntos de función.

Tabla 5-1 Puntos de función sin ajustar asignados por complejidad.

BFC	Baja	Media	Alta
EI	3	4	6
EO	4	5	7
EQ	3	4	6
ILF	7	10	15
EIF	5	7	10

La suma total de los puntos de función sin ajustar representa el tamaño funcional de la aplicación medida. Esta se obtiene de la sumatoria de los puntos de función de cada uno de los componentes funcionales básicos:

$$\begin{aligned} \text{Tamaño}_{UFP}(\text{Función Transaccional}_i) \\ = \sum \text{tamaño}(EI_i) + \sum \text{tamaño}(EO_i) + \sum \text{tamaño}(EQ_i) \end{aligned}$$

$$\text{Tamaño}_{UFP}(\text{Función de Datos}_i) = \sum \text{tamaño}(ILF_i) + \sum \text{tamaño}(EIF_i)$$

$$\begin{aligned} \text{Tamaño}_{UFP}(\text{Modelo funcional}) \\ = \sum \text{tamaño}(\text{Función de Datos}_i) + \sum \text{tamaño}(\text{Función Transaccional}_i) \end{aligned}$$

5.1.1.5.2. ASIGNACIÓN NUMÉRICA PARA EL MÉTODO COSMIC FFP

Las reglas de asignación numérica se definen en conformidad con las establecidas por el método de medición COSMIC FFP. Por tanto, el tamaño funcional de la aplicación medida es igual al *número de los movimientos de datos*. Esto es, 1 CFP (punto de función COSMIC) se define como el tamaño de un movimiento de datos. El procedimiento de medición asigna 1 CFP por cada uno de los movimientos de datos identificados en un proceso funcional y el total del tamaño funcional de la aplicación es la sumatoria de los tamaños de los procesos funcionales.

La suma total de los movimientos de datos representa el tamaño funcional de la aplicación medida:

$$\begin{aligned} \text{Tamaño}(\text{proceso funcional}_i) \\ = \sum \text{tamaño}(\text{Entradas}_i) + \sum \text{tamaño}(\text{Salidas}_i) \\ + \sum \text{tamaño}(\text{Lecturas}_i) + \sum \text{tamaño}(\text{Escrituras}_i) \end{aligned}$$

$$\text{Tamaño (Modelo funcional)} = \sum \text{Tamaño}(\text{proceso funcional}_i)$$

5.1.2. APLICACIÓN DEL PROCEDIMIENTO DE MEDICIÓN

La Figura 5-7 provee una descripción de los pasos para aplicar la metodología de medición de tamaño funcional **aFPA**. La metodología recibe como entrada los artefactos de software y construye el modelo funcional con los componentes funcionales que son mapeados a los métodos de medición del tamaño funcional IFPUG FPA y COSMIC FFP para obtener el tamaño funcional. Asimismo, verifica la exactitud de las mediciones y realiza la evaluación de los resultados de medición para ser aprovechados en la construcción de modelos de estimación de esfuerzo. La Figura 5-7 (a) describe los pasos para el procedimiento de medición, la Figura 5-7 (b) describe los pasos para el procedimiento de verificación y la Figura 5-7 (c) describe los pasos para el procedimiento de evaluación. En esta sección se detalla el proceso de aplicación del procedimiento de medición para obtener el tamaño funcional a partir de los artefactos del código fuente. El proceso de aplicación del procedimiento de verificación de la exactitud de los resultados que se realiza posterior a la aplicación del procedimiento de medición se detalla en la Sección 5.2 y el proceso de aplicación del procedimiento de evaluación de modelos para el aprovechamiento de resultados de medición se detalla en la Sección 5.3.

Los pasos para la aplicación del proceso del procedimiento de medición de la metodología **aFPA** incluyen la adquisición de los artefactos del código fuente de la aplicación en un ambiente de operación controlado, la construcción del modelo funcional a partir del análisis estático y dinámico del código fuente de la aplicación, la medición mediante la aplicación de las reglas de mapeo entre metamodelos y las reglas de asignación numérica. El proceso de aplicación del procedimiento de medición corresponde al Paso 2 del modelo

de procesos de (Abran, 2010). La descripción de los pasos para la aplicación del procedimiento de medición de tamaño funcional son las siguientes:

- *Adquisición de artefactos de software*: el procedimiento de medición recibe como entrada los artefactos del código fuente de la aplicación y la aplicación instalada funcionando en un ambiente de operación controlado. El proceso de adquisición de artefactos de software recopila todos los artefactos necesarios para realizar el proceso de medición.
- *Construcción del modelo funcional*: a partir del análisis estático y dinámico del código fuente de la aplicación que realiza el rastreo de transacciones se construye el modelo funcional que representa los componentes funcionales y sus relaciones. En el proceso de análisis se determinan los eventos del usuario final y se rastrean las transacciones hacia las entidades de datos obteniendo un modelo funcional orientado a eventos y transacciones que representa los procesos funcionales.
- *Aplicación de reglas de mapeo y asignación numérica*: el procedimiento de medición obtiene el tamaño funcional realizando el mapeo entre los elementos del modelo funcional y los de los métodos de medición para lo cual identifica los procesos funcionales únicos e independientes y aplica las reglas de mapeo y de asignación numérica. La etapa de mapeo entre elementos de la aplicación, los elementos del procedimiento de medición y los elementos de los métodos de medición permite mantener la trazabilidad entre los resultados de medición, los accesos y movimientos de datos, el modelo funcional, los elementos del modelo de la aplicación y los requerimientos funcionales. Las reglas de mapeo y asignación numérica con las cuales el procedimiento obtiene el tamaño funcional se realiza de acuerdo a los métodos de medición de tamaño funcional IFPUG FPA y COSMIC FFP.
- *Verificación de exactitud*: el procedimiento de medición reporta el detalle de los resultados de medición y el procedimiento de verificación lleva a cabo el proceso de la verificación de exactitud de los resultados de las mediciones. Para esto se realiza la aplicación

del protocolo de verificación de exactitud que permite validar la calidad de las mediciones obtenidas por el procedimiento de medición. El procedimiento de verificación de la exactitud de los resultados se detalla en la Sección 5.2.

- *Evaluación de modelos de estimación:* el procedimiento de evaluación estudia algunos modelos de estimación de esfuerzo para mostrar un posible aprovechamiento de los resultados de la medición del tamaño funcional. El procedimiento de evaluación de modelos para el aprovechamiento de resultados de medición se detalla en la Sección 5.3.

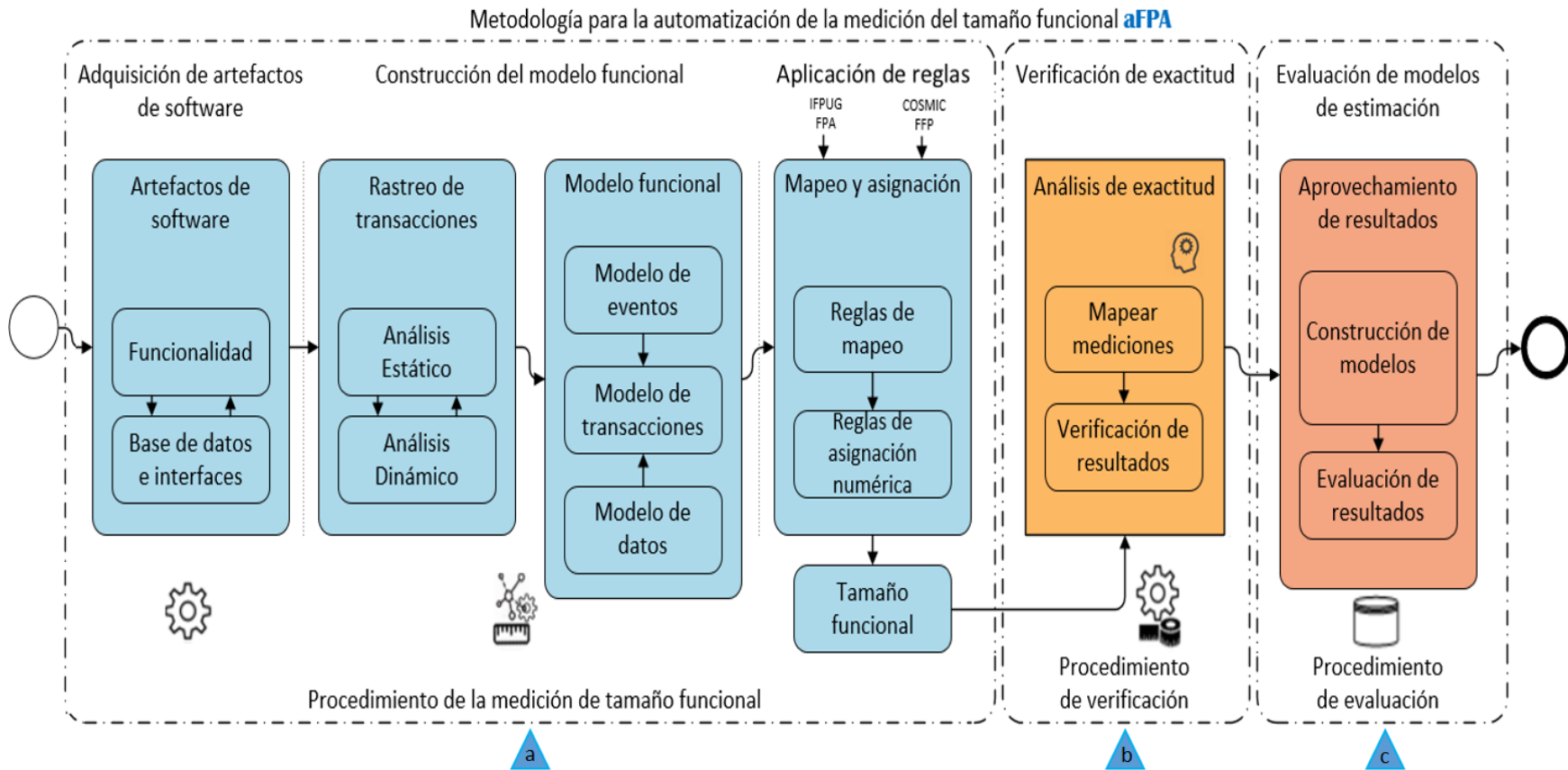


Figura 5-7 Pasos de la metodología para la automatización de la medición de tamaño funcional aFPA.

En la etapa de adquisición de artefactos de software el procedimiento de medición recibe como entrada los artefactos del código fuente de la aplicación medida. Esto incluye todos los artefactos requeridos para poder instalar y correr la funcionalidad en un ambiente de operación controlado. Con la aplicación instalada y en operación se realizan las tareas de configuración para el proceso de medición que incluye la agregación de las librerías que permiten realizar la identificación de los eventos del usuario funcional, el rastreo de las transacciones hacia las entidades de datos y el cálculo del conteo de tamaño funcional. En la práctica, el ambiente de medición se configuraría, por ejemplo, en un servidor de medición o de pruebas separado del ambiente de desarrollo. En la etapa de construcción del modelo funcional se crea el modelo con los componentes funcionales del procedimiento y sus relaciones que representa la funcionalidad ofrecida por la aplicación medida a los usuarios funcionales. El modelo funcional es orientado a eventos y transacciones. A este modelo en la etapa de medición, se le aplican las reglas de mapeo para identificar las relaciones entre los componentes del modelo funcional y los componentes del método de medición utilizado. Basado en la identificación de cada uno de los elementos que componen los procesos funcionales, se aplican las reglas de asignación numérica que permiten calcular el tamaño funcional.

A continuación se detallan los aspectos generales del proceso de aplicación del procedimiento de medición para la obtención del tamaño funcional del modelo funcional. Primero, se describen los niveles de automatización con los cuales el procedimiento puede ser aplicado. Segundo, se detalla el proceso de adquisición de los artefactos de entrada. Tercero, se describe el proceso de construcción del modelo funcional. Cuarto, se detalla el proceso de aplicación de las reglas de mapeo y las reglas de asignación numérica y por último, se listan las salidas del procedimiento.

5.1.2.1. NIVEL DE AUTOMATIZACIÓN DEL PROCEDIMIENTO DE MEDICIÓN

El nivel de automatización con el que es posible aplicar el procedimiento de medición depende del nivel de control que el profesional tenga sobre la aplicación medida y su

ambiente de operación. Se definen cuatro niveles de automatización que van desde la aplicación manual del procedimiento hasta la aplicación completamente automatizada. Estos niveles permiten que las reglas del procedimiento de medición puedan ser aplicadas en distintos ambientes de operación y en distintas etapas del ciclo de vida de desarrollo. Tal como se muestra en la Figura 5-8 y en la Figura 5-9, los cuatro niveles de automatización en los que el procedimiento de medición puede ser aplicado son:

N0. En el *nivel 0* de automatización se realiza la identificación de los componentes del procedimiento, la construcción del modelo funcional, la aplicación de las reglas de mapeo y las reglas de asignación numérica manualmente a partir del diseño y los lineamientos del procedimiento, lo cual no requiere ninguna adaptación del código fuente de la aplicación. En este caso, el profesional controla el proceso de medición y el procedimiento no requiere la configuración de ninguna librería de acceso a la aplicación medida. Asimismo, la aplicación medida no requiere la instrumentación del código fuente mediante los procesos de etiquetado o la inclusión de las librerías de automatización específicas para la plataforma tecnológica. Finalmente, el procedimiento no requiere el acceso a la aplicación medida en un ambiente controlado de operación y no es necesaria la configuración e inclusión de las librerías de medición.

N1. En el *nivel 1* de automatización se realiza la identificación de los componentes del procedimiento y la construcción del modelo funcional manualmente a partir del diseño y los lineamientos del procedimiento, lo cual no requiere ninguna adaptación del código fuente de la aplicación. En este caso, el profesional controla el proceso inicial de medición y el procedimiento no necesita la configuración de ningún acceso a la aplicación medida. Para aplicar el proceso del procedimiento en el *nivel 1*, este provee una interfaz para recibir como entrada el modelo generado manualmente. El procedimiento procesa el modelo y lo mide automáticamente mediante la aplicación de las reglas de mapeo y las reglas de asignación numérica. En este caso, la aplicación medida no requiere la instrumentación del código fuente mediante procesos de instrumentación o la inclusión de librerías de automatización específicas para la plataforma tecnológica. Además, el

procedimiento no requiere el acceso a la aplicación medida en un ambiente controlado de operación y no es necesaria la configuración e inclusión de las librerías de medición.

N2. En el *nivel 2* de automatización, el profesional a partir del diseño y los lineamientos del procedimiento, realiza la identificación de los componentes manual o automáticamente con el apoyo de sus herramientas, lo cual no requiere ninguna adaptación del código fuente de la aplicación. El procedimiento provee una interfaz que recibe como entrada un archivo XML que contiene el mapa de la interfaz de usuario y las estructuras de almacenamiento persistente. En un ambiente controlado de operación, el procedimiento automáticamente configura lo requerido para realizar la construcción del modelo funcional a partir del modelo de eventos, datos y transacciones mediante la simulación de la ejecución de la funcionalidad. Asimismo, realiza la medición del tamaño funcional mediante la aplicación de las reglas de mapeo y las reglas de asignación numérica. En este caso, el profesional controla el inicio del proceso y construye el archivo de interfaz XML. El procedimiento requiere la configuración del ambiente de operación de la aplicación medida pero no el acceso mismo a la aplicación. En este caso, la aplicación medida no requiere la instrumentación del código fuente mediante procesos de etiquetado o la inclusión de librerías de automatización específicas para la plataforma tecnológica, el procedimiento no requiere el acceso a la aplicación medida en un ambiente controlado de operación y no es necesaria la configuración e inclusión de las librerías de medición. Sin embargo, sí es requerido un ambiente de operación en el que el prototipo de soporte a la medición pueda operar para realizar el proceso de conteo.

N3. En el *nivel 3* de automatización se realiza la identificación de los componentes del procedimiento, la construcción del modelo funcional y la medición del tamaño funcional mediante la aplicación de las reglas de mapeo y las reglas de asignación numérica automáticamente; por ejemplo, por la herramienta prototipo propuesta en esta tesis. En este caso, el profesional configura el ambiente de operación al inicio del proceso y el procedimiento realiza todos los pasos para la medición funcional. La aplicación medida requiere la instrumentación del código fuente mediante un proceso de etiquetado o la inclusión de librerías de automatización específicas para la plataforma tecnológica, el

procedimiento requiere el acceso a la aplicación medida en un ambiente controlado de operación y es necesaria la configuración e inclusión de las librerías de medición para obtener el tamaño funcional. En este nivel de automatización, todas las librerías del procedimiento son incluidas en el código fuente de la aplicación medida.

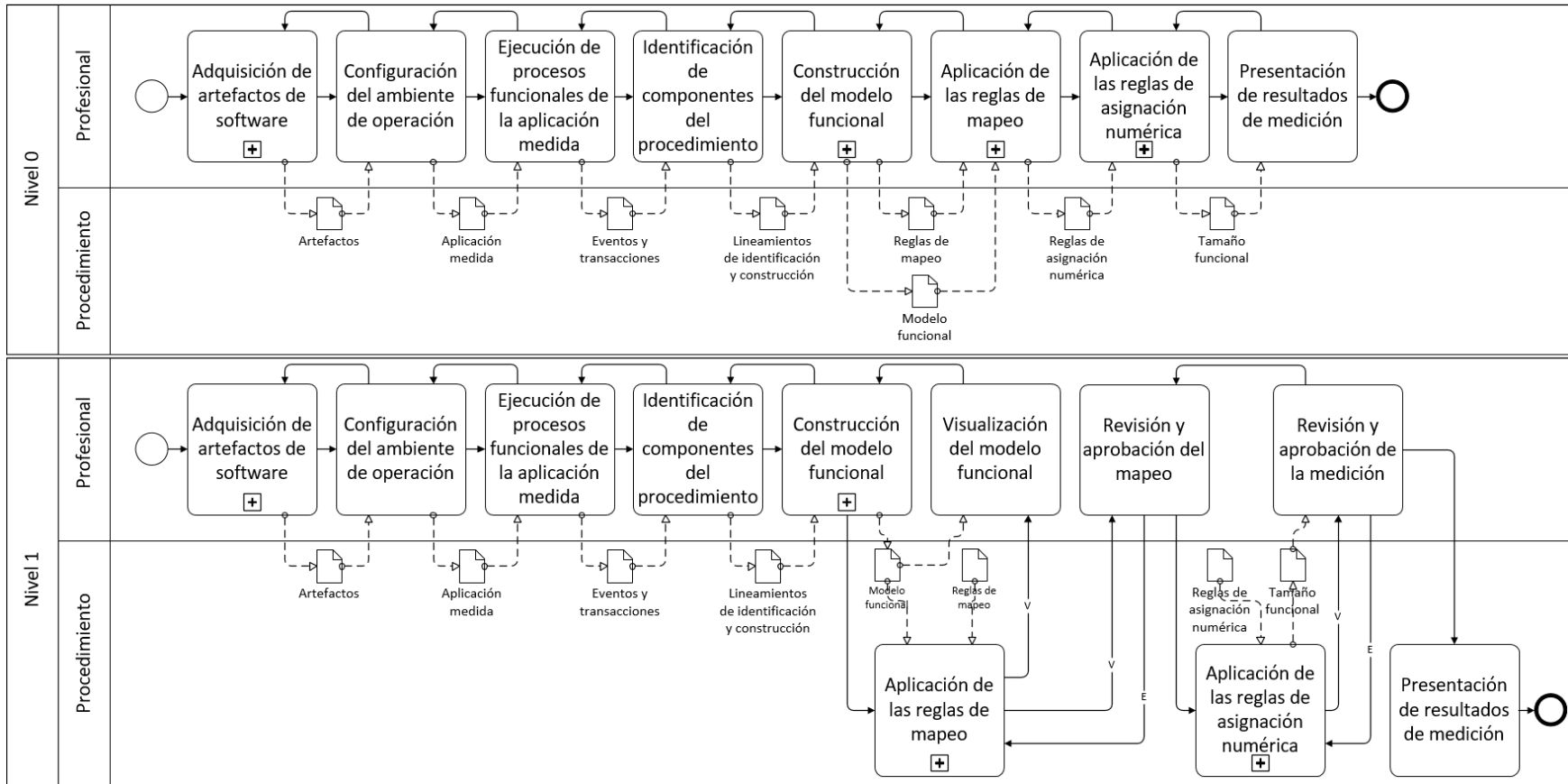


Figura 5-8 Niveles de automatización 0 y 1 del procedimiento de medición.

Es recomendable que el profesional o los profesionales que llevan a cabo el proceso de medición y que aplican el procedimiento cuenten con conocimientos sobre los objetivos y estrategia establecidos para obtener el tamaño funcional de la aplicación de software bajo medición. Por lo tanto, un profesional en el rol de analista de medición debe: (a) conocer el dominio del problema de negocio que resuelve la aplicación medida mediante sus requerimientos funcionales, (b) contar con la experticia técnica para realizar la configuración del ambiente controlado de ejecución, el proceso de instrumentación del código fuente y la inclusión de las librerías de medición, según sea el caso. Finalmente, (c) conocer el programa de medición de la organización de desarrollo de software para identificar los objetivos de la aplicación del procedimiento de medición para que se alineen con los procesos de la toma de decisiones. Tal como se recomienda en (Tarhan, Ozkan & Içoz, 2016), este tipo de procesos requieren la participación de uno o varios profesionales con perfiles relacionados con la administración de proyectos, análisis, medición, desarrollador y/o pruebas.

En la Sección 5.4 se detalla la propuesta de la herramienta prototipo de medición y cómo este puede ser utilizado en los distintos niveles de automatización del procedimiento. El detalle de cómo aplicar cada una de las reglas de mapeo y de asignación numérica en una aplicación de software se describe en los casos de estudio presentados en el Capítulo 6.

Durante la aplicación del procedimiento, el proceso de instrumentación del código fuente de la aplicación medida es dependiente de la tecnología utilizada en el ambiente de operación. No obstante, una vez que se identifican los componentes funcionales del procedimiento, la aplicación del proceso de medición es independiente de la plataforma tecnológica. Esta característica del procedimiento de medición permite extender su aplicación no solo a artefactos de la etapa de desarrollo del código fuente, sino, a etapas tempranas en el ciclo de vida, las cuales se discuten en la Sección 5.5.

5.1.2.2. ADQUISICIÓN DE LOS ARTEFACTOS DE SOFTWARE

Los artefactos de software utilizados para la aplicación del procedimiento de medición incluyen todos los archivos del código fuente de la aplicación medida. Esto incluye las últimas versiones liberadas de los archivos del código fuente, desde la capa de interfaz de usuario hasta la capa de bases de datos. Dependiendo de nivel de automatización con el que se aplica el procedimiento de medición, estos archivos pueden incluir todos los artefactos requeridos para la instalación y la ejecución de la funcionalidad en un ambiente de operación controlado. En este caso, la aplicación del procedimiento requiere contar con la aplicación de software instalada y funcionando en este ambiente de operación, donde el profesional tiene acceso a todo el sistema y su código fuente en modo administrador. En la práctica, se recomienda que este ambiente de operación controlado sea configurado sobre, por ejemplo, los ambientes de pruebas pero no los de desarrollo. Si existen interfaces, bases de datos y/o servicios externos, estos deben ser configurados para permitir la ejecución de todas las funcionalidades de la aplicación medida desde la perspectiva del usuario funcional. Los artefactos de entrada mínimos requeridos para realizar el proceso de conteo bajo el procedimiento de medición son los siguientes:

- Versiones actuales de todos los archivos del código fuente que permitan la ejecución de la aplicación, desde la capa de interfaz de usuario hasta la capa de bases de datos.
- Versiones actuales de los archivos de definición de estructuras de almacenamiento de datos tales como los *scripts* de las bases de datos.
- Versiones actuales de los archivos de interfaces externas de la aplicación tales como estructuras XML, JSON u otros.
- Versiones actuales de los archivos de configuración para la instalación de la aplicación.
- Cualquier otro artefacto requerido para la instalación y la ejecución de la funcionalidad de la aplicación medida en un ambiente de operación controlado.

El nivel de control que debe tener el profesional que realiza el proceso de medición depende del nivel de automatización con el que se aplica el procedimiento. El profesional que realiza el proceso de medición sobre la aplicación de software debe tener acceso a los artefactos del código fuente que le permitan identificar los componentes funcionales del procedimiento de medición. El procedimiento se puede aplicar de manera manual o automatizada, donde mientras más alto sea el nivel de automatización, más control se requiere sobre la aplicación. En el nivel 0, el nivel 1 y el nivel 2, el proceso de medición no requiere ninguna adaptación del código fuente de la aplicación de software. El procedimiento recibe las entradas requeridas para su ejecución a partir de las interfaces provistas para ese fin. En este caso, el procedimiento funciona como un módulo independiente de la aplicación medida. Por otro lado, en el nivel 3 con la aplicación configurada en un ambiente de operación controlado, se realiza la instrumentación del código fuente mediante un proceso de etiquetado o la inclusión de librerías de automatización específicas para la plataforma tecnológica y finalmente, se realiza la configuración e inclusión de las librerías de medición.

Las librerías provistas por el procedimiento de medición permiten analizar el rastreo de las transacciones y construir los sub modelos de eventos, datos y transacciones necesarios para construir el modelo funcional. Asimismo, dependiendo del enfoque de automatización en el contexto de operación determinado, el proceso de instrumentación del código fuente permite realizar el conteo sin necesidad de la intervención de un profesional una vez que se hayan realizado las configuraciones iniciales. En la Sección 5.4 se detalla la propuesta de prototipo de medición para la plataforma utilizada en los casos de estudio realizados en esta investigación. Los casos de estudio describen los procesos de instrumentación del código fuente y la inclusión de las librerías de automatización específicas que dependen de la plataforma tecnológica específica.

En el caso de que el profesional no cuente con las librerías de medición para su plataforma, puede implementarlas él mismo o realizar el proceso de identificación de los componentes del modelo manualmente. En este caso, el profesional debe determinar cómo los distintos flujos de información mueven los datos a través de la aplicación medida hacia y desde el almacenamiento persistente, las interfaces externas y las interfaces de usuario.

Una de las ventajas del procedimiento es que sus reglas de mapeo y de asignación numérica pueden ser aplicadas de manera manual o automatizada dependiendo de la disponibilidad de las librerías de medición en una plataforma tecnológica específica.

5.1.2.3. CONSTRUCCIÓN DEL MODELO FUNCIONAL

El modelo funcional del procedimiento de medición se construye a partir de los artefactos de código fuente recopilados en la fase de adquisición. El metamodelo del procedimiento de medición presentado en la Figura 5-4 detalla cada uno de los componentes y las relaciones a partir de los cuales se construye el grafo que representa la funcionalidad de la aplicación medida. El modelo funcional se construye a partir del análisis estático y dinámico del código fuente de la aplicación configurada en un ambiente de operación controlado. Para extraer los componentes del procedimiento se deben ejecutar los requerimientos funcionales que desencadenan los flujos de información desde la capa de interfaz hasta la capa de datos. El modelo funcional representa la instanciación del metamodelo del procedimiento.

Para la identificación de los componentes del procedimiento se realiza el rastreo de las funcionalidades que permite la construcción del modelo de la siguiente manera:

- *Modelo de eventos*: este modelo representa el mapa de navegación de la interfaz de los usuarios funcionales. Se identifican todos los controles de interfaz y para cada uno de ellos, cada uno de los eventos asociados a una llamada de una función transaccional o una operación de navegación. Asimismo, para cada componente de interfaz se identifican los controles asociados a la interfaz y los eventos asociados a cada uno de estos controles. En este caso, la interfaz de usuario principal de la aplicación medida se identifica como el componente tipo *Application* y como el primer componente de tipo *Control* tipo interfaz de usuario. Cada control adicional se identifica como tipo *Control* y cada evento asociado a los controles se identifican como tipo *Event*. Cada evento se

identifica como un tipo *Function* cuando se encuentra asociada a una función transaccional que involucra una transacción de información desde o hacia datos internos o externos. La aplicación se explora para encontrar los componentes de interfaz asociados a los eventos de navegación y aplicar el mismo proceso hasta que ya no se encuentren más eventos de navegación o las llamadas ya se encuentren asociadas a componentes de interfaz ya explorados.

- *Modelo de datos*: este modelo representa la capa de almacenamiento persistente interno o externo. Se identifican cada una de las bases de datos asociadas a la aplicación como un componente tipo *DataBase* y para cada una de ellas todas las tablas y columnas asociadas como tipo *Table* y *Column* respectivamente. Asimismo, se identifican todos los componentes de llaves e índices clasificándolos como tipo *PrimaryKey*, *ForeignKey* e *Index* según corresponda. Finalmente, se identifican las entidades de almacenamiento externo como tipo *Table* y sus datos asociados como tipo *Column*.
- *Modelo de transacciones*: este modelo representa los flujos de información de los procesos funcionales. Para cada uno de los componentes *Function* se identifican un componente tipo *Transaction Read* y *Transaction Write* a partir de los tipos de acceso que realiza en los componentes tipo *Table* y *Column* que representan las entidades de almacenamiento persistente.

La instanciación del metamodelo del procedimiento de medición **aFPA** crea la estructura de grafo la cual se utiliza en los procesos de mapeo de componentes con los métodos de medición IFPUG FPA y COSMIC FFP. El proceso de análisis instancia el modelo a partir del rastreo de las transacciones desde los eventos del usuario hasta las entidades de datos.

En el caso de la aplicación del procedimiento en el Nivel 3 de automatización, la aplicación medida configurada en un ambiente de operación controlado requiere la instrumentación del código fuente mediante un proceso de etiquetado o la inclusión de librerías de automatización específicas para la plataforma tecnológica y la configuración e inclusión de las librerías de medición. Este proceso se detalla en la Sección 5.2, donde se describe la

propuesta del prototipo de medición para la plataforma seleccionada en los casos de estudio.

La Figura 5-10 muestra un ejemplo de la instanciación del metamodelo para obtener el modelo funcional que representa los requerimientos de la aplicación bajo medición llamada *Customer*. Esta aplicación administra los datos de un cliente (nombre, dirección, ciudad, estado, código postal, teléfono, fax, correo) y permite realizar las operaciones de mantenimiento básicas sobre esta entidad (insertar, leer, actualizar, borrar y listar).

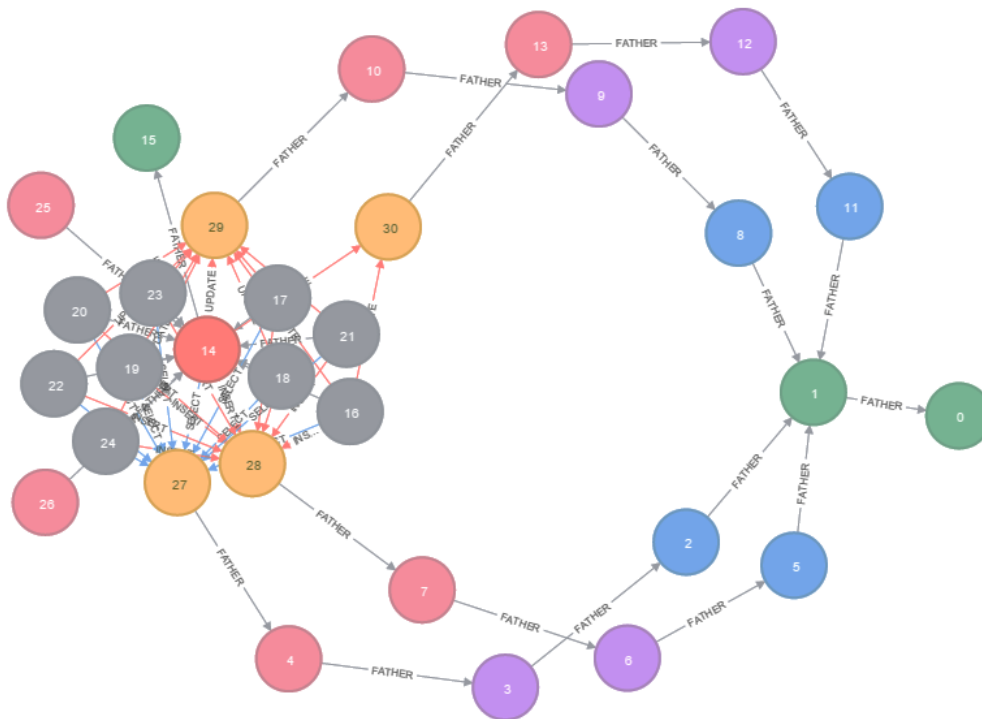


Figura 5-10 Modelo funcional de la aplicación *Customer*.

El proceso para la construcción del modelo funcional se describe a continuación:

P1. *Modelo de eventos*: La aplicación *Customer* {nodo 0=*Application*} presenta una ventana principal *MainForm* {nodo 1=*Control*} arista {(0, 1)} que a su vez es un control {nodo 2=*Control*} arista {(1, 2)} con un evento asociado *Load* {nodo 3=*Event*} arista {(2, 3)}. El

control *MainForm* tiene asociados tres botones *buttonInsert*, *buttonUpdate*, *button* {nodos 5, 8, 11=*Control*} aristas {(1, 5), (1, 8), (1, 11)} los cuales tienen un evento asociado *Click* cada uno {nodos 6, 9, 12=*Event*} aristas {(5, 6), (8, 9), (11, 12)}. Cada evento tiene asociada una función {nodos 4, 7, 10, 13=*Function*} aristas {(3, 4), (6, 7), (9, 10), (12, 13)} donde una función es de lectura {nodo 27=*Read*} y tres de escritura {nodos 28, 29, 30=*Write*} aristas {(4, 27), (7, 28), (10, 29), (13, 30)}.

- P2. *Modelo de datos*: La aplicación accede a una base de datos *CustomerDB* {nodo 15=*Database*} que tiene una tabla *Customer* {nodo 14=*Table*} arista {(15, 14)}. La tabla tiene un conjunto de atributos *Identifier*, *Name*, *Address*, *City*, *State*, *ZipCode*, *Telephone*, *Fax*, *Email* {nodos 16, 17, 18, 19, 20, 21, 22, 23, 24=*Column*} aristas {(14, 16), (14, 17), (14, 18), (14, 19), (14, 20), (14, 21), (14, 22), (14, 23), (14, 24)}, una llave primaria *PkCustomer* {nodo 25=*PrimaryKey*} arista {(14, 25)} y un índice *PkCustomerIdx* {nodo 26=*Index*} arista {(14, 26)}.
- P3. *Modelo de transacciones*: La función de lectura {nodo 27=*Read*} y las funciones de escritura {nodos 28, 29, 30=*Write*} acceden a la tabla *Customer* {nodo 14=*Table*} aristas {(27, 14), (28, 14), (29, 14), (30, 14)}. Esto es, la arista {(27,14)} es de lectura (*Select*), la arista {(28,14)} es de escritura (*Insert*), la arista {(29,14)} es de escritura (*Update*) y la arista {(30,14)} es de escritura (*Delete*). La función de lectura {nodo 27=*Read*} (*Select*) accede a las columnas *Identifier*, *Name*, *Address*, *City*, *State*, *ZipCode*, *Telephone*, *Fax*, *Email* {16, 17, 18, 19, 20, 21, 22, 23, 24=*Column*} aristas {(27, 16), (27, 17), (27, 18), (27, 19), (27, 20), (27, 21), (27, 22), (27, 23), (27, 24)}. La función de escritura {nodo 28=*Write*} (*Insert*) accede a las columnas *Identifier*, *Name*, *Address*, *City*, *State*, *ZipCode*, *Telephone*, *Fax*, *Email* {16, 17, 18, 19, 20, 21, 22, 23, 24=*Column*} aristas {(28, 16), (28, 17), (28, 18), (28, 19), (28, 20), (28, 21), (28, 22), (28, 23), (28, 24)}. La función de escritura {nodo 29=*Write*} (*Update*) accede a las columnas *Identifier*, *Name*, *Address*, *City*, *State*, *ZipCode*, *Telephone*, *Fax*, *Email* {16, 17, 18, 19, 20, 21, 22, 23, 24=*Column*} aristas {(29, 16), (29, 17), (29, 18), (29, 19), (29, 20), (29, 21), (29, 22), (29, 23), (29, 24)}. La función de escritura {nodo 30=*Write*} (*Delete*) accede a las columnas *Identifier* {16=*Column*} arista {(30, 16)}.

El modelo funcional contiene la información de cada uno de los procesos funcionales asociados a cada componente de interfaz que puede ser extraído del seguimiento de las transacciones.

A partir del modelo funcional se realiza el proceso de aplicación de las reglas de mapeo para la identificación de los componentes funcionales de acuerdo a cada uno de los métodos de medición y se aplican las reglas de asignación numérica para obtener el tamaño funcional. En las siguientes secciones se describe el proceso de aplicación de las reglas de mapeo y de asignación numérica para los métodos de medición.

5.1.2.4. APLICACIÓN DE LAS REGLAS DE MAPEO

La aplicación de las reglas de mapeo se realiza sobre el modelo funcional y de acuerdo al método de medición seleccionado. Mediante la aplicación de las reglas de mapeo se identifican los procesos funcionales únicos e independientes, los candidatos a componentes funcionales y sus relaciones. Las reglas específicas de mapeo para el método de medición IFPUG FPA y COSMIC FFP se describen en la Sección 5.1.1.4. La identificación de los componentes de cada uno de los metamodelos de los métodos de medición permite realizar el conteo del tamaño funcional de la aplicación medida. El proceso de mapeo permite determinar la asociación de los componentes del metamodelo del procedimiento **aFPA** con los componentes del metamodelo de cada método de medición dando como resultado la instanciación del metamodelo de cada método lo que permite la aplicación de las reglas de asignación numérica. En este proceso, a partir del modelo funcional construido, se analizan cada uno de los trazos de ejecución que representan los flujos de datos y transacciones y se realiza el mapeo de componentes funcionales al método de medición particular.

En el caso de la aplicación del procedimiento en el Nivel 3 de automatización, la aplicación medida funcionando en un ambiente de operación controlado requiere la configuración e inclusión de las librerías de medición. Este proceso de configuración se detalla en la Sección 5.4 donde se describe la propuesta de prototipo de medición.

5.1.2.5. APLICACIÓN DE LAS REGLAS DE ASIGNACIÓN NUMÉRICA

La aplicación de las reglas de asignación numérica se realiza sobre la instanciación del metamodelo de cada uno de los métodos de medición en los que se basa el procedimiento. Las reglas de asignación numérica para el método de medición IFPUG FPA y COSMIC FFP se describen en la Sección 5.1.1.5. Con la aplicación de las reglas de asignación numérica, el procedimiento de medición obtiene el tamaño funcional de la aplicación bajo medición.

Del mismo modo que con la aplicación de las reglas de mapeo, en el caso de la aplicación del procedimiento en el Nivel 3 de automatización, la aplicación medida funcionando en un ambiente de operación controlado requiere la configuración e inclusión de las librerías de medición cuyo proceso se detalla en la descripción del prototipo de medición que se encuentra en la Sección 5.4.

Después de que el procedimiento obtiene el tamaño funcional de la aplicación bajo medición se realiza el proceso de verificación de la exactitud de los resultados obtenidos para validar su confiabilidad y calibrar el procedimiento. Posterior a esto, los resultados obtenidos por los procesos de medición del tamaño funcional son evaluados para determinar cómo pueden ser aprovechados para la construcción de modelos de estimación. En la Sección 5.2 se detalla el procedimiento de verificación de exactitud de los resultados de medición del tamaño funcional y en la Sección 5.3 se describe el procedimiento de evaluación para la explotación o aprovechamiento de las mediciones de tamaño funcional en la construcción de los modelos de estimación de esfuerzo de desarrollo de software.

5.1.2.6. SALIDAS DEL PROCESO DE APLICACIÓN DEL PROCEDIMIENTO

Como resultado del proceso de aplicación del procedimiento de medición se reporta el detalle de los resultados del tamaño funcional que incluye las siguientes salidas:

- Nombre, versión y módulos de la aplicación medida.

- Lista de los artefactos de entrada utilizados para la medición con su nombre, versión, ubicación.
- Descripción y dominio funcional de la aplicación medida.
- Objetivo, alcance y nivel de granularidad del proceso de medición.
- Fecha de medición, nombre, perfil, cualidades y certificaciones del profesional.
- Nombre y versión del procedimiento de medición.
- Etapa del ciclo de vida de desarrollo donde se realiza el conteo.
- Lista de usuarios funcionales.
- Lista de procesos funcionales.
- Trazabilidad entre los artefactos de entrada y los procesos funcionales.
- Modelo funcional de la aplicación medida.
- Trazabilidad entre cada uno de los componentes del modelo funcional y los artefactos de entrada.
- Trazabilidad entre cada uno de los componentes del modelo funcional y los procesos funcionales.
- Detalle del mapeo de los componentes funcionales del modelo funcional y el metamodelo de cada uno de los métodos de medición utilizados en el proceso de conteo.
- Detalle de la asignación numérica para cada uno de los componentes funcionales y procesos funcionales indicando el número y tamaño funcional correspondiente con el método de medición utilizado en el proceso de conteo.
- Tamaño funcional, número total y clasificación de funciones o movimiento de datos.
- Trazabilidad entre las funciones o movimiento de datos y cada uno de los requerimientos y los artefactos de entrada.
- Trazabilidad con todos los elementos asociados a cada una de las funciones o movimiento de datos.
- Tamaño funcional de la aplicación medida en puntos de función de acuerdo al método de medición.

El primer caso de estudio de la Sección 6.1 ejemplifica cada una de las entradas y salidas para cada uno de los pasos del proceso de aplicación del procedimiento de medición de la metodología **aFPA**.

5.2. PROCEDIMIENTO DE VERIFICACIÓN DE EXACTITUD

Este procedimiento representa el producto del ciclo ingenieril 4 (EC4).

Como resultado de la aplicación del procedimiento de medición se obtiene el tamaño funcional de la aplicación bajo medición. El procedimiento de medición reporta todos los detalles de los resultados de las mediciones de tamaño funcional que permite llevar a cabo la ejecución del procedimiento de verificación de exactitud para validar la calidad de las mediciones obtenidas. En esta etapa del proceso, los resultados de medición son verificados, documentados y auditados mediante la utilización del protocolo de verificación. La Figura 5-7 (b) describe los pasos para el procedimiento de verificación. Desde la perspectiva de la metodología de medición **aFPA**, el objetivo del protocolo de verificación de exactitud es realizar un proceso sistemático que permita validar y calibrar las reglas de mapeo y de asignación numérica, lo que permite identificar y entender las fuentes de inexactitud en el proceso de medición. En los procesos de calibración de los procedimientos y herramientas de medición, el protocolo busca determinar la consistencia con las mediciones reportadas por un experto en medición que son tomadas como valores de control. El protocolo incluye no solo la verificación de las reglas de mapeo y de asignación numérica aplicadas por el procedimiento, sino también la verificación de la calidad de los artefactos de entrada y la variación encontrada producto de la comparación con los valores verdaderos establecidos como valores de control. El procedimiento de medición presenta los resultados de todo el proceso de medición de acuerdo al nivel de granularidad definido. El nivel de granularidad de los resultados reportados determina el nivel de detalle con que puede ser aplicado el procedimiento de verificación.

5.2.1. DISEÑO DEL PROTOCOLO DE VERIFICACIÓN

El protocolo de verificación de exactitud propone una evaluación de arriba hacia abajo (*top-down*) de los resultados de medición para identificar diferencias durante la aplicación del proceso de medición en comparación con los valores verdaderos (*true value*) obtenidos a partir de la medición de uno o varios expertos. El protocolo verifica el tamaño de cada uno de los procesos y componentes funcionales resultado del proceso de medición conduciendo las siguientes fases: (F1) Comparación de los resultados totales de medición, (F2) Comparación detallada de la exactitud de los resultados de cada uno de los componentes funcionales básicos e (F3) Identificación y recuperación de errores (Quesada-López, Jenkins, 2015b). El proceso de verificación mantiene trazabilidad para cada uno de los resultados de medición de los componentes funcionales del método, los componentes del procedimiento y el artefacto del código fuente, esto habilita la posibilidad de realizar una auditoría de los resultados. La estrategia de verificación para el método de medición IFPUG FPA se construye siguiendo las recomendaciones de trabajos relacionados en el área de verificación de la medición funcional del método COSMIC FFP (Yilmaz, Tunalilar & Demirors, 2013; Soubra, Abran & Ramdane-Cherif, 2014).

Desde la perspectiva de la confiabilidad de los resultados del proceso de medición, el protocolo de verificación permite realizar un análisis sistemático y detallado de los resultados de medición obtenidos por un procedimiento de medición. El protocolo realiza la comparación de resultados de medición obtenidos por diferentes procedimientos de medición, que puede incluir el uso de herramientas y artefactos en distintas fases del ciclo de vida del desarrollo del software y comparar las diferencias obtenidas por cada uno de ellos. El objetivo del protocolo es reportar las diferencias de los resultados de medición obtenidos en distintos procesos de conteo e indicar los puntos específicos en el proceso donde las diferencias son identificadas. Las actividades de verificación deben garantizar el suficiente

grado de confianza en el procedimiento de medición y validar criterios de calidad tales como exactitud¹³, repetibilidad¹⁴ y reproducibilidad¹⁵.

El protocolo requiere como entrada para la aplicación del proceso de verificación los resultados de medición obtenidos por cada uno de los procedimientos de medición. Estos permiten identificar y analizar los componentes del modelo de recolección de datos que se presentan en la Figura 5-11. El modelo determina el nivel de detalle de los datos que deben ser recolectados en un proceso de conteo para la aplicación del protocolo de verificación. El protocolo se clasifica como nivel de detalle 1¹⁶, trazable y etiquetado hacia los artefactos medidos. El modelo almacena todos los niveles específicos de datos extraídos a partir del proceso de conteo de tamaño funcional bajo el método de medición. Para esto, todos los procesos funcionales son identificados, asociados a un requerimiento específico, y relacionados con los distintos componentes funcionales básicos. El protocolo mantiene trazabilidad para todos los resultados de medición hacia los requerimientos de la aplicación bajo medición.

Para realizar un proyecto (*Project*) de verificación de exactitud, el protocolo (*Protocol*) compara los resultados de uno o varios procedimientos de medición (*Procedure*). El alcance del proyecto de verificación se determina por los límites de la aplicación medida (*Boundary*) y el nivel de granularidad (*CountLevel*) de los resultados obtenidos por los procedimientos de medición. El procedimiento de medición detalla los procesos funcionales (*Function*) identificados con los cuales se mantiene trazabilidad (*Traceability*) hacia los requerimientos (*Requirements*) y los artefactos de entrada (*Inputs*). Las funciones (*Function*) clasificadas como transacciones (*Transaction*, *ExternalInput*, *ExternalOutput*, *ExternalQuery*) y los grupos de archivos lógicos (*LogicalFile*, *InternalLogicalFile*, *ExternalLogicalFile*) son relacionados por medio de las referencias realizadas por las transacciones hacia los

¹³ Exactitud: proximidad entre un valor medido y un valor verdadero.

¹⁴ Repetibilidad: precisión de medida bajo un conjunto de condiciones iguales. Proximidad entre las mediciones repetidas del mismo objeto.

¹⁵ Reproducibilidad: precisión de medida bajo un conjunto de condiciones diferentes (sujetos). Proximidad entre las mediciones repetidas del mismo objeto.

¹⁶ "Detailed Linked and Flagged Count". *Levels of Function Point Counting*. Disponible en: <http://www.totalmetrics.com/function-point-resources/downloads/Levels-of-Function-Point-Counting.pdf>

archivos (*FileTypeReferenced*). Los elementos de datos (*DataElement*) son agrupados por los registros de elementos (*RecordElement*) los cuales son asociados a los archivos lógicos. De igual manera, para cada uno de estos componentes se mantiene la trazabilidad hacia los requerimientos y los artefactos de entrada. Todos los componentes corresponden a cada uno de los elementos funcionales básicos del método de medición IFPUG FPA por lo que el procedimiento de medición los reporta a partir del proceso de mapeo y de asignación numérica.

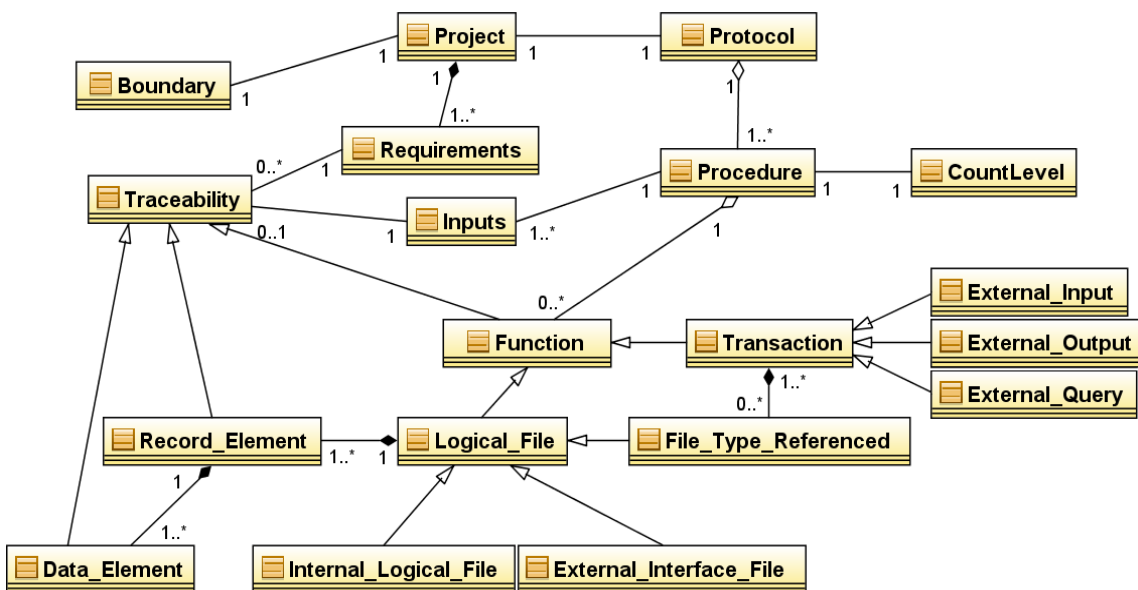


Figura 5-11 Modelo de recolección de datos del protocolo.

Dado que el protocolo de verificación se utiliza para la calibración del procedimiento de medición, este requiere la recolección de los valores verdaderos utilizados como valores de control en la comparación de los resultados. Por ejemplo, los artefactos de entrada deben ser medidos por uno o varios expertos a partir de las reglas del método correspondiente en paralelo y manteniendo el nivel de trazabilidad requerido para realizar la comparación.

5.2.2. APLICACIÓN DEL PROTOCOLO DE VERIFICACIÓN

La Figura 5-12 describe las tres fases principales del protocolo de verificación de exactitud que se aplican a las mediciones recolectadas por los procedimientos de medición. Cada una de las fases del protocolo se describe a continuación:

- F1. *Comparación de la exactitud de los totales de medición:* en la primera fase, los totales del proceso de medición para cada uno de los conteos de tamaño funcional pueden ser comparados entre sí y contra el valor verdadero (*true value*) proporcionado por un conteo experto. En esta fase solo los resultados totales del conteo son verificados, en este caso los puntos de función sin ajustar (UFP). Si los resultados coinciden, no existe una diferencia entre los totales reportados por los procedimientos de medición. Sin embargo, la coincidencia de resultados en esta etapa no significa que los procesos y componentes identificados por las mediciones en el proceso para obtener el total del tamaño funcional sean los mismos. Si el proceso de verificación se detiene en esta fase solo los resultados generales son comparados.
- F2. *Comparación detallada de la exactitud de los resultados de medición:* en la segunda fase, todos los pasos del proceso de medición son comparados contra los valores verdaderos y entre los procedimientos de medición. Esto incluye la comparación del número y tamaño funcional de cada uno de los componentes funcionales básicos y los elementos que los componen, y el número y tamaño funcional de cada uno de los requerimientos identificados. Esta fase representa una verificación detallada del proceso de medición para cada uno de los procedimientos que permite identificar los lugares específicos en el proceso de medición donde existen inconsistencias de número o tamaño. Cada vez que un error en esta fase es identificado, la fase 3 se dispara. Al final de esta fase, cualquier causa asignable responsable por un error es identificada, separada, asociada y reportada basada en la respuesta obtenida de la fase 3. Para la conducción de la segunda fase se realizan las siguientes actividades:

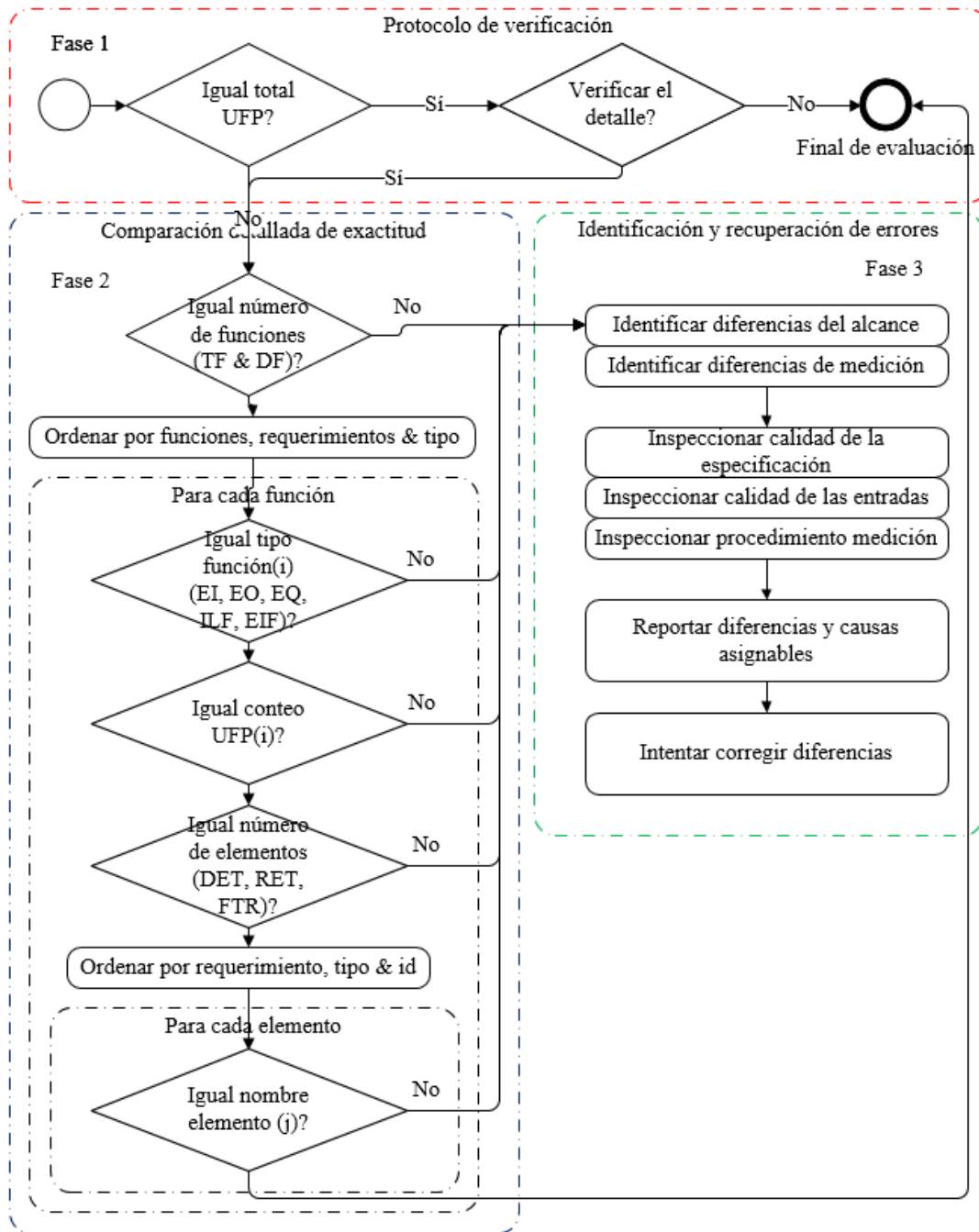


Figura 5-12 Fases del protocolo de verificación de exactitud.

A1. *Comparación del número de cada una de las funciones transaccionales y de datos: si existen diferencias en el número de funciones por tipo se dispara la fase*

3 y el error es reportado. Una inconsistencia en el número de funciones transaccionales o de datos significa que la funcionalidad identificada por los distintos procedimientos puede no ser la misma. Por ejemplo, esto puede presentarse por inconsistencias entre los artefactos de entrada utilizados por los procedimientos de medición. En el caso de mediciones sobre aplicaciones de software ya desarrolladas, pueden existir inconsistencias tales como requerimientos documentados y no implementados o funcionalidad implementada que no se encuentra documentada. Asimismo, puede presentarse por la duplicación de la funcionalidad de la aplicación bajo medición.

A2. *Comparación del tamaño funcional de los requerimientos y tipos de funciones:* para habilitar esta comparación, todas las funciones se ordenan y agrupan por los requerimientos del usuario y el tipo de función. En esta actividad las diferencias en el número y tipo de las funciones transaccionales (TF) y de datos (DF) entre los procedimientos de medición son identificadas. Todas las funciones son verificadas contra el valor verdadero de referencia para determinar los casos donde existe una diferencia en los identificadores de los requerimientos que alteren la clasificación. Cuando se presenta una diferencia, se dispara la fase 3, la inconsistencia es reportada y posibles causas asignadas son identificadas. Por ejemplo, se reportan las inconsistencias basadas en los identificadores de los requerimientos y las recomendaciones de reclasificación que se determinan por la comparación del número y del tamaño de los procesos funcionales.

A3. *Comparación del número y tamaño funcional de los componentes funcionales básicos de los requerimientos y tipos de funciones:* para cada una de las funciones que coincida entre los procedimientos de medición a partir del identificador de requerimiento funcional se realiza una comparación de los cinco tipos de componentes básicos, su número y tamaño funcional. Finalmente, se compara el número y cantidad de elementos que determinan la complejidad de los componentes. Primero, se verifica si los tipos de componentes funcionales básicos (EI,

EO, EQ, ILF, EIF) coinciden entre los resultados de los procedimientos de medición. Si existe una diferencia en la clasificación, se dispara la fase 3 y el error es reportado ya que las funcionales identificadas no son iguales. Por ejemplo, estas inconsistencias pueden presentarse por diferencias en cómo los procedimientos interpretan como los artefactos de entrada realizan las operaciones de mantenimiento en las funciones de datos. Segundo, se verifica si el tamaño funcional de cada uno de los tipos de componentes funcionales básicos (EI, EO, EQ, ILF, EIF) coincide entre los resultados de los procedimientos de medición. Por ejemplo, una inconsistencia de este tipo puede presentarse por diferencias en el número de elementos DET, RET y FTR identificados para cada función lo que afecta el cálculo del tamaño funcional total por componente. Tercero, se verifica el número de los elementos DET, RET y FTR para cada una de las funciones. Este proceso verifica las diferencias en el número de elementos tipo DET y RET para las funciones de datos y el número de DET y FTR para las funciones transaccionales. En paralelo, se verifican los cálculos realizados para determinar la complejidad de las funciones. Para realizar las comparaciones, los elementos de cada una de las funciones deben ser ordenados y agrupados para cada uno de los requerimientos de usuario, tipo de función y la identificación de cada elemento. Cuarto, para cada uno de los elementos DET, RET y FTR de una función se compara el identificador que determina si los elementos identificados por cada uno de los procedimientos para una función coinciden. Cualquier diferencia identificada durante la ejecución de las verificaciones anteriores dispara la fase 3, las inconsistencias son reportadas y las posibles causas asignadas son identificadas.

A4. *Reporte de errores y diferencias*: al final de la fase dos, cada una de las causas asignables responsables de un error o diferencia son aisladas y reportadas por el protocolo manteniendo trazabilidad entre cada uno de los procedimientos de medición, requerimientos y artefactos de entrada. En este punto, el proceso de

verificación de los resultados de la aplicación de software se ha completado dado que la fase tres se fue disparando durante esta fase.

F3. *Identificación y recuperación de errores*: en la tercera fase se determinan las causas asignables por las cuales se presenta una diferencia en alguno de los resultados del proceso de medición. Esta fase se ejecuta en paralelo cada vez que en la segunda fase se identifica una inconsistencia. Para determinar una causa asignable se realiza una inspección sobre la calidad de los requerimientos, los artefactos de entrada, y el proceso de conteo de cada uno de los procedimientos para identificar el origen del error. Las diferencias encontradas entre las mediciones son almacenadas y reportadas para su análisis y corrección. Si el error puede ser reparado, las acciones correspondientes son realizadas, los resultados de la medición son actualizados y el protocolo es ejecutado nuevamente para verificar los resultados. En esta fase se realizan inspecciones paralelas de cada uno de los procedimientos de medición para determinar los pasos responsables de las diferencias y de los artefactos de entrada para identificar posibles inconsistencias. En el momento que el error es detectado se analizan las causas asignables y si es posible se toman las acciones correspondientes:

- A1. *Errores causados por el procedimiento de medición*: si el error es causado por el procedimiento, el error es reportado y el proceso de medición es revisado. Con la identificación de la etapa y el paso específico donde se produce la inconsistencia, la corrección es realizada y los resultados de medición son actualizados. El proceso de verificación es aplicado nuevamente para asegurar que los nuevos cálculos son correctos y que no existen inconsistencias.
- A2. *Errores causados por los artefactos de entrada*: si el error está relacionado con la calidad o completitud de los artefactos de entrada, el error es reportado y el artefacto es revisado. Si el artefacto puede ser corregido, la corrección es realizada y los resultados de medición son actualizados para cada procedimiento de medición. El proceso de verificación es aplicado nuevamente para asegurar que los nuevos cálculos son correctos y que no existen inconsistencias.

La Figura 5-13 presenta los dos niveles de automatización con los cuales es posible aplicar el protocolo de verificación de exactitud. Este proceso de verificación puede ser aplicado de forma manual o automatizada con el soporte de una herramienta de comparación tal como se describe a continuación:

N0. En el *nivel 0* de automatización se realiza el proceso de verificación para cada uno de los resultados de los procedimientos de medición contra los valores verdaderos manualmente siguiendo los lineamientos del protocolo. En este caso, el profesional es el encargado de recolectar las mediciones y aplicar cada una de las fases y actividades del protocolo y los valores verdaderos son reportados por expertos en medición.

N1. En el *nivel 1* de automatización se realiza el proceso de verificación para cada uno de los resultados de los procedimientos de medición contra los valores verdaderos automáticamente por el protocolo. En este caso, la herramienta protocolo de soporte del proceso de verificación reporta los resultados de comparación con cada una de las diferencias encontradas y las causas asignables. El profesional es el encargado de recolectar las mediciones para la aplicación de cada una de las fases y actividades de verificación con las que el protocolo obtiene y presenta los resultados. El profesional revisa y aprueba los resultados del proceso de verificación y el protocolo genera el reporte de resultados.

El profesional que interactúa con el protocolo de verificación de exactitud debe: (a) conocer el dominio del problema, (b) entender los artefactos de entrada utilizados por los procedimientos de medición y (c) conocer el programa de medición de la organización de desarrollo de software que le permita la toma de decisiones durante el proceso de verificación.

El prototipo de automatización para apoyar el proceso de aplicación del protocolo de verificación se describe en la Sección 5.2 y el detalle de cómo aplicar cada una de las etapas y actividades del protocolo para distintos procedimientos de medición se describe en los casos de estudio y estudios cuasi experimentales presentados en el Capítulo 6.

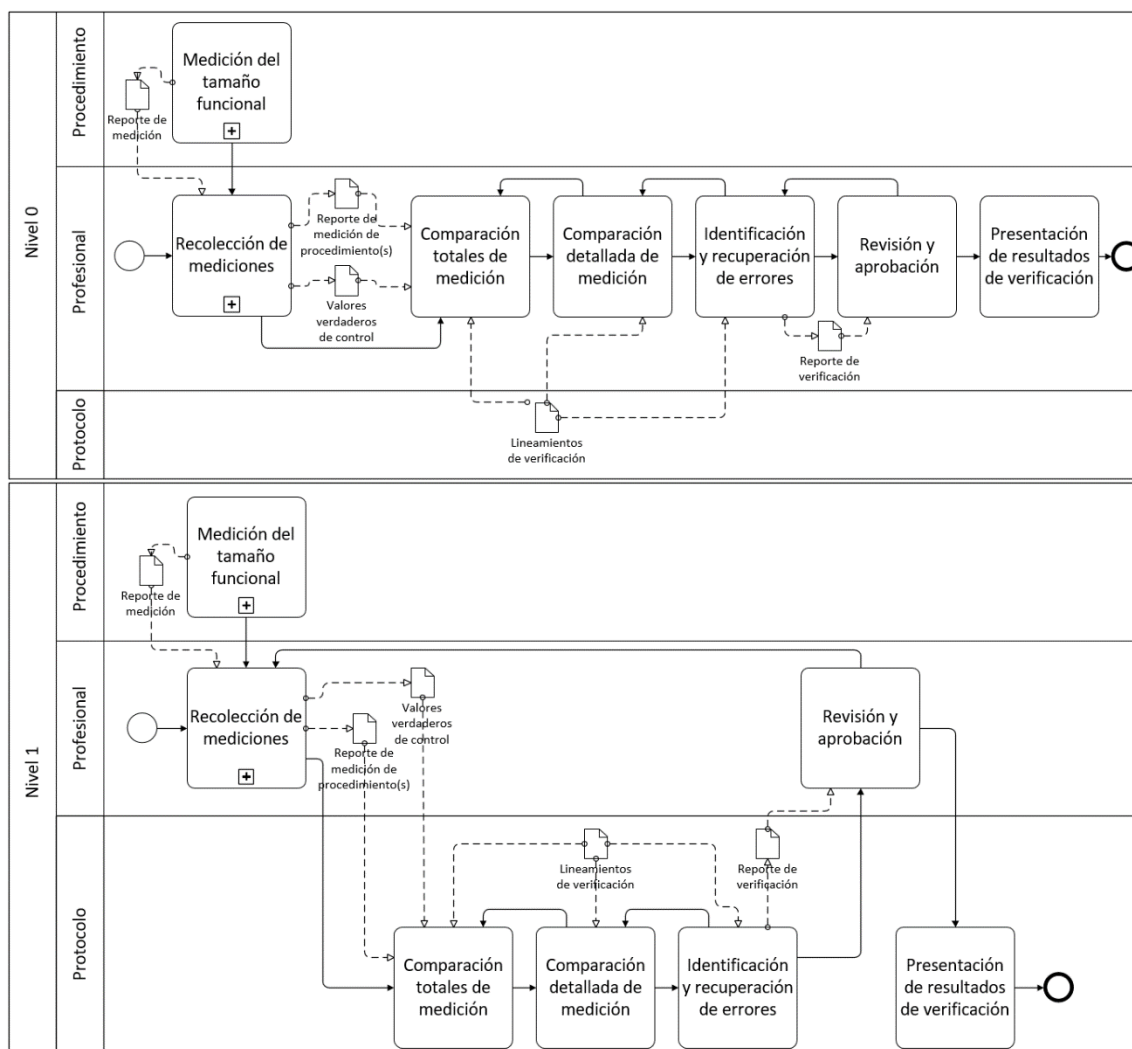


Figura 5-13 Niveles de automatización 0 y 1 del protocolo de verificación de exactitud.

5.2.3. SALIDAS DEL PROCESO DE APLICACIÓN DEL PROTOCOLO

El protocolo de verificación reporta la comparación de los resultados de medición por cada uno de los procedimientos incluyendo las siguientes salidas:

- Descripción de los artefactos de entrada utilizados en cada uno de los procesos de medición.

- Etapa del ciclo de vida de desarrollo donde se realiza el conteo de cada uno de los procesos de medición.
- Fecha de verificación, nombre, perfil, cualidades y certificaciones del profesional.
- Requerimientos funcionales de la aplicación medida.
- Lista de los artefactos de entrada por proceso de medición.
- Trazabilidad entre los artefactos de entrada y los requerimientos del usuario.
- Tamaño funcional de la aplicación medida en puntos de función sin ajustar.
- Tamaño funcional y número total de funciones de datos.
- Tamaño funcional y número total de funciones transaccionales.
- Funciones de datos y funciones transaccionales clasificadas de acuerdo a su tipo de componente funcional básico: EI, EO, EQ, ILF, EIF.
- Lista de funciones mapeadas a cada uno de los requerimientos y los artefactos de entrada.
- Notas explicativas asociadas a cada una de las funciones que presenten los resultados de la verificación y la identificación de causas asignables. Esto incluye el detalle de las métricas del proceso de verificación.
- Para cada una de las funciones transaccionales, el tamaño funcional y número total de elementos de los tipos DET y FTR asociados. La complejidad de la función y la trazabilidad entre cada uno de los elementos, los artefactos de entrada y los requerimientos. Asimismo, todas las entidades de datos y atributos de datos asociados a cada función transaccional.
- Para cada una de las funciones de datos, el tamaño funcional y número total de elementos de tipo DET y RET asociados. La complejidad de la función y la trazabilidad entre cada uno de los elementos, los artefactos de entrada y los requerimientos. Asimismo, todas las entidades de datos y atributos de datos asociados a cada función de datos.

El reporte de los resultados de la verificación de las mediciones de tamaño funcional presenta un reporte comparativo con las inconsistencias entre cada uno de los procedimientos de medición y el valor verdadero. Asimismo, el detalle del análisis de posibles causas asignables a las diferencias de los procesos de medición y de los artefactos de entrada. En el caso de la verificación de las mediciones del método COSMIC FFP se aplica el protocolo de verificación propuesto en (Soubra, Abran & Ramdane-Cherif, 2014) y se reporta el mismo nivel de detalle sugerido anteriormente.

5.2.4. MÉTRICAS DE VERIFICACIÓN DE EXACTITUD

El protocolo reporta los resultados de la verificación de exactitud de las mediciones de acuerdo a las métricas comúnmente utilizadas en el área de medición para comparar los valores verdaderos contra los valores calculados mediante la aplicación de un procedimiento. Estas incluyen: la diferencia reportada por el error absoluto (MAE), la magnitud del error relativo (MRE), la magnitud relativa del error (MER), el error relativo balanceado (BRE), el error relativo balanceado invertido (IBRE) y la reproducibilidad (REP) del número y tamaño funcional para cada uno de los requerimientos, funciones, componentes funcionales básicos y elementos de complejidad. La descripción de las funciones de medición para cada una de las métricas mencionadas es el siguiente:

Magnitud del error absoluto (MAE):

$$MAE = |e_i - \hat{e}_i|$$

Donde e_i es valor real y \hat{e}_i es la observación con el valor estimado.

Magnitud del error relativo (MRE):

$$MRE = \frac{|e_i - \hat{e}_i|}{e_i}$$

Donde e_i es valor real y \hat{e}_i es la observación con el valor estimado.

Magnitud relativa del error (MER):

$$MER = \frac{|e_i - \hat{e}_i|}{\hat{e}_i}$$

Donde e_i es valor real y \hat{e}_i es la observación con el valor estimado.

Error relativo balanceado (BRE):

$$BRE = \frac{|e_i - \hat{e}_i|}{\min(e_i, \hat{e}_i)}$$

Donde e_i es valor real y \hat{e}_i es la observación con el valor estimado.

Error relativo balanceado invertido (IBRE):

$$IBRE = \frac{|e_i - \hat{e}_i|}{\max(e_i, \hat{e}_i)}$$

Donde e_i es valor real y \hat{e}_i es la observación con el valor estimado.

Reproducibilidad (REP):

$$REP_i = \frac{|\text{promedio}(\hat{e}_{\text{otros}}) - \hat{e}_i|}{\text{promedio}(\hat{e}_{\text{otros}})}$$

Donde \hat{e}_i es una observación con el valor estimado y REP_i es la reproducibilidad del sujeto i .

El uso, aplicación e interpretación de las métricas para el análisis de los resultados de cada proceso de verificación se detalla en los estudios empíricos presentados en el Capítulo 6.

5.3. PROCEDIMIENTO DE EVALUACIÓN DE LOS RESULTADOS DE MEDICIÓN

La aplicación del procedimiento de evaluación permite determinar cómo los resultados del tamaño funcional obtenidos a partir de la aplicación del procedimiento de medición pueden ser aprovechados o explotados para la construcción de modelos de estimación

tales como los modelos de estimación de esfuerzo. El proceso de aprovechamiento de los resultados corresponde al Paso 3 del modelo de procesos propuesto por (Abran, 2010; Habra, Abran, Lopez & Sellami, 2008; Jacquet & Abran, 1997). El aprovechamiento de los resultados del procedimiento de medición en modelos de estimación de esfuerzo de desarrollo de proyectos de software comprende la utilización del tamaño funcional y los distintos factores del contexto como entradas para la construcción de los modelos y su evaluación mediante el cálculo de las métricas de desempeño. El procedimiento para la evaluación construye los modelos de estimación de esfuerzo basados en la medición del tamaño funcional del software y los evalúa para determinar cuál combinación de factores genera los mejores resultados en el contexto de desarrollo (Quesada-López, Murillo-Morera, Castro-Herrera & Jenkins, 2016; Murillo-Morera, Quesada-López, Castro-Herrera & Jenkins, 2017). Este proceso detalla una estrategia de evaluación de modelos de estimación del esfuerzo de desarrollo en el contexto de aplicaciones orientadas a datos, utilizando las mediciones de tamaño funcional y los distintos factores de contexto como predictores y obteniendo como salida la estimación del esfuerzo de desarrollo de un proyecto de software.

La estrategia de evaluación se basa principalmente en las recomendaciones del marco de trabajo propuesto por (Song, Jia, Shepperd, Ying & Liu, 2011) y en trabajos relacionados en el área de modelos de estimación de esfuerzo tales como (Kitchenham & Mendes, 2009; Shepperd & MacDonell, 2012; Dejaeger, Verbeke, Martens & Baesens, 2012; Langdon, Dolado, Sarro & Harman, 2016; Menzies, Yang, Mathew, Boehm & Hihn, 2016; Lavazza & Morasca, 2017).

La Figura 5-7 (c) describe los pasos para el procedimiento de evaluación. Desde la perspectiva de la metodología de medición **aFPA**, el procedimiento de evaluación de los modelos de estimación utiliza los resultados del proceso de medición del tamaño funcional los cuales se almacenan en los repositorios de datos históricos de proyectos de desarrollo de software. El objetivo de la estrategia de evaluación es permitir una comparación exhaustiva y sin sesgo entre los modelos de estimación de esfuerzo para determinar el que mejor desempeño produce en el contexto de evaluación.

5.3.1. APLICACIÓN DEL PROCEDIMIENTO DE EVALUACIÓN

La Figura 5-14 presenta los componentes del marco de trabajo que permite la ejecución de la estrategia de evaluación de modelos. Primero, el componente de evaluación permite analizar el desempeño de los modelos a partir de los datos históricos de proyectos y segundo, el componente de estimación de esfuerzo de desarrollo construye los modelos a partir de las técnicas que mejor resultado presentaron en el contexto y realiza la estimación de esfuerzo a partir de los nuevos datos. En el proceso de construcción de los modelos se combinan distintas técnicas para el pre procesamiento de datos, la selección de atributos o características y los algoritmos de aprendizaje. Cada combinación de los tres tipos de técnicas representa un esquema de aprendizaje que es comparado durante la aplicación de la estrategia de evaluación.

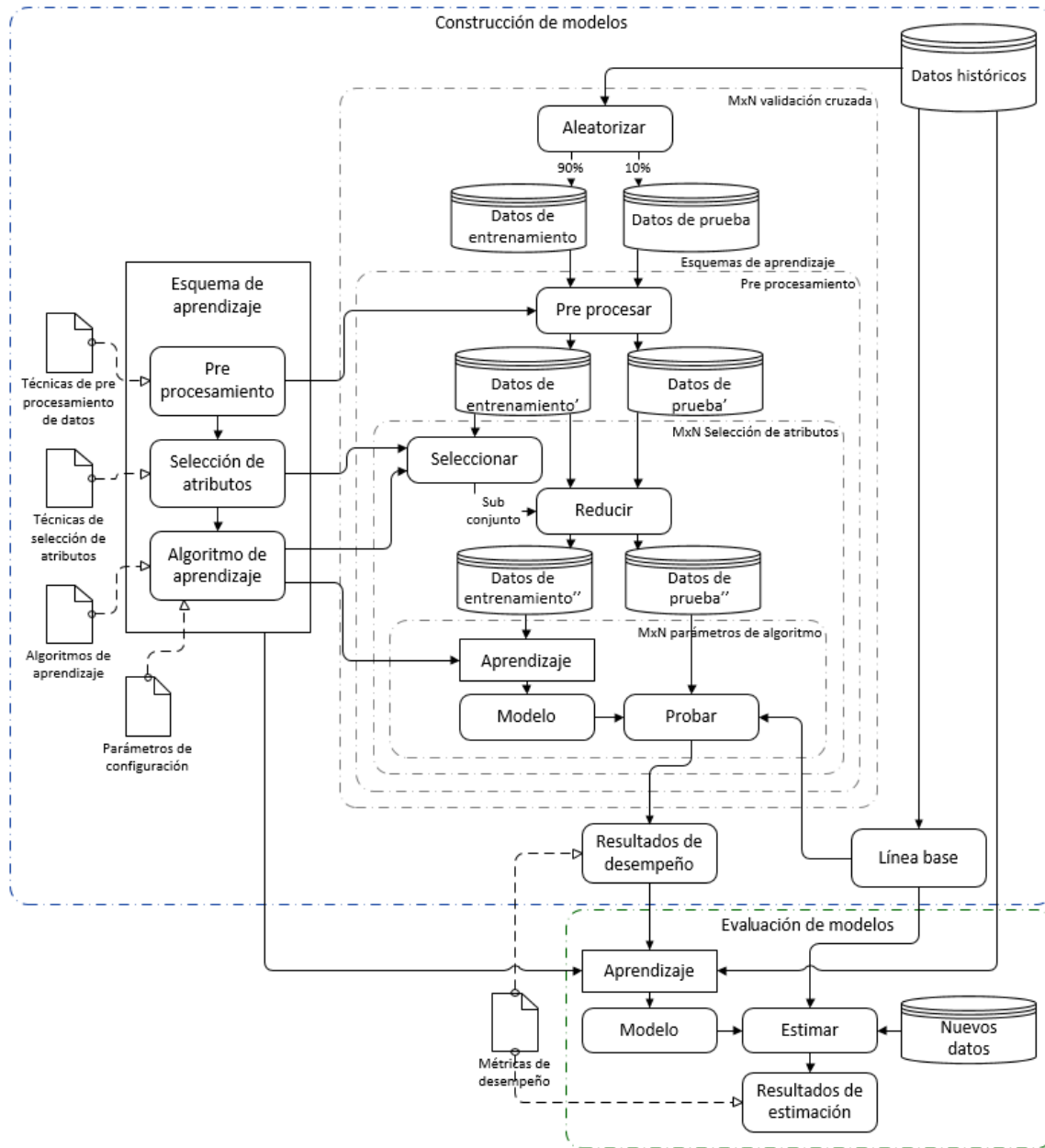


Figura 5-14 Componentes de la estrategia de evaluación de modelos de estimación.

La descripción de las fases de la estrategia de evaluación son las siguientes:

- F1. *Construcción de los modelos de estimación de esfuerzo:* en esta fase se analizan los modelos de estimación construidos a partir de cada uno de los esquemas de aprendizaje para

determinar su desempeño en la predicción del esfuerzo de desarrollo. Los modelos se construyen para cada uno de los esquemas de aprendizaje y los datos históricos de los proyectos desarrollados que funcionan como los conjuntos de entrenamiento. Cada esquema de aprendizaje se construye a partir de la selección de una técnica de pre procesamiento de datos, una de selección de atributos y un algoritmo de aprendizaje. Las técnicas de pre procesamiento se utilizan, principalmente, para la normalización o transformación de los datos, la eliminación de valores atípicos, la creación de variables ficticias y el tratamiento de los valores faltantes, las técnicas de selección de atributos permiten la elección de un subconjunto con los atributos más relevantes para la estimación del esfuerzo las cuales se realizan a partir de distintos métodos de filtrado o envoltura; finalmente, las técnicas o algoritmos de aprendizaje permiten realizar la predicción del esfuerzo mediante, principalmente, algoritmos de aprendizaje de máquina y técnicas de regresión que han probado ser de utilidad en este contexto. Los parámetros de las técnicas o algoritmos de aprendizaje son configurados para determinar cuáles ofrecen un mejor resultado en las métricas de desempeño. El proceso de construcción contempla los siguientes aspectos:

- A1. El desempeño de los esquemas de aprendizaje es evaluado utilizando datos históricos de proyectos para determinar las capacidades de predicción del esfuerzo de desarrollo para nuevos proyectos.
- A2. Los datos históricos se dividen en dos conjuntos de datos: un conjunto de datos de entrenamiento (90%) que permite la construcción de los modelos a partir de los distintos esquemas de aprendizaje y un conjunto de datos de prueba (10%) que permite evaluar el desempeño de los modelos.
- A3. Para la separación de los conjuntos de datos se realiza un proceso de una ronda de validación cruzada $M \times N$ que permite la partición del conjunto de datos históricos en dos subconjuntos complementarios. Se realizan M repeticiones donde se aplica una aleatorización al conjunto de datos históricos y se generan N particiones de las cuales se asigna un 10% de las instancias para el conjunto

de datos de prueba y un 90% de las instancias para el conjunto de datos de entrenamiento. En este caso, la selección de datos se realiza aleatoriamente donde primero se hace la separación de N partes, el modelo es construido a partir de las $N - 1$ partes y la parte restante es utilizada para las pruebas. Esto se repite para las N rondas lo que permite que cada parte sea utilizada para el entrenamiento y las pruebas lo que minimiza el sesgo de la selección. El proceso se repite M veces aplicando una aleatorización para lograr la reducción del efecto del ordenamiento.

- A4. El conjunto de datos de prueba es independiente del proceso de construcción de los modelos y es utilizado para evaluar la generalización de los resultados a partir de los esquemas de aprendizaje seleccionados.
- A5. El proceso de validación cruzada permite la construcción y evaluación de $M \times N$ modelos por cada conjunto de datos y para cada uno de los esquemas de aprendizaje. Cada una de las evaluaciones realizadas reporta los resultados de las métricas de desempeño las cuales son analizadas. Esta estrategia de múltiples rondas de validación cruzada para cada uno de los conjuntos de datos se realiza para reducir la variabilidad de los resultados del desempeño de cada uno de los modelos construidos. Los resultados de cada una de las rondas son promediados para obtener el valor total del desempeño que se reporta. En este caso, se utiliza la validación cruzada de $M \times N$ rondas para cada uno de los modelos de estimación construidos.
- A6. Los modelos se construyen y evalúan para cada uno de los esquemas de aprendizaje a partir del conjunto de datos de entrenamiento. La cantidad de esquemas de aprendizaje se obtiene de la combinación única de la cantidad de técnicas de pre procesamiento, técnicas de selección de atributos y algoritmos de aprendizaje.
- A7. El proceso de aplicación de cada uno de los esquemas de aprendizaje para la generación de los modelos se lleva a cabo en el conjunto de datos de entrenamiento realizando tres pasos: Primero, se aplica la técnica de pre procesamiento

de datos para la transformación, la creación de variables ficticias en escala [0, 1] de los predictores en estala categórica, eliminación de valores atípicos y el tratamiento de los valores faltantes. Segundo, se realiza la selección de atributos para la reducción de la dimensión del conjunto de datos aplicando métodos de filtrado o envoltura y evaluando el desempeño de cada uno de los subconjuntos seleccionados mediante una validación cruzada que permite determinar el subconjunto de atributos que muestra mejor desempeño. Tercero, se realiza la construcción del modelo utilizando las técnicas de aprendizaje de máquina y de regresión a partir del conjunto de datos reducido. En este caso se evalúa el desempeño de distintas configuraciones de parámetros mediante una validación cruzada que permite determinar la configuración con mejor desempeño.

A8. Para realizar el proceso de validación, el conjunto de datos de prueba es tratado del mismo modo que el conjunto de entrenamiento y los valores reales y de predicción son calculados a partir de las métricas de desempeño. Cada uno de los resultados de evaluación se almacenan para calcular los promedios y varianzas durante cada uno de los procesos de validación cruzada de $M \times N$ rondas.

A9. El proceso de evaluación de esquemas de aprendizaje permite construir los modelos y determinar los mejores esquemas de aprendizaje que pueden ser utilizados para la estimación del esfuerzo de desarrollo para los nuevos datos.

F2. *Evaluación de modelos de estimación de esfuerzo*: en la fase de evaluación y a partir de los resultados de la fase de construcción de modelos se selecciona el o los esquemas de aprendizaje para construir los modelos que permiten realizar las estimaciones de esfuerzo a partir de los nuevos datos y evaluar su desempeño. El proceso de evaluación contempla los siguientes aspectos:

A1. Para realizar la estimación de esfuerzo se seleccionan los mejores esquemas de aprendizaje y se construyen los modelos a partir de todo el conjunto de datos históricos.

A2. El proceso de creación de los modelos de estimación se realiza de acuerdo a la estrategia descrita en la primera fase. Se pre procesan los datos, se realiza la

selección de los mejores atributos para los nuevos datos y se selecciona la mejor configuración de parámetros de acuerdo a los resultados de los modelos analizados en la fase anterior.

A3. El proceso obtiene la estimación de esfuerzo para los nuevos datos y reporta los resultados de estimación y desempeño para los nuevos datos.

Las técnicas específicas para cada uno de los componentes de los esquemas de aprendizaje se detallan en cada una de las configuraciones de los experimentos del Capítulo 7.

Todas las fases realizadas en este proceso permiten mejorar el proceso de evaluación de los modelos de estimación que se construyen a partir de los conjuntos de datos históricos. La estrategia permite la estabilización de los modelos y reducir el riesgo de que los mismos presenten problemas de sobre ajuste o sub ajuste que puede producir resultados no confiables sobre el desempeño. Asimismo, mediante la implementación de las estrategias de modelos de línea base, bajo el mismo conjunto de datos, es posible realizar comparaciones más objetivas sobre el desempeño de los distintos modelos de estimación. Esto es posible porque la estrategia de evaluación plantea la comparación de los resultados de los modelos construidos para cada uno de los esquemas de aprendizaje contra los modelos construidos a partir de la línea base, ambos a partir de los mismos datos en el contexto evaluado. En este caso, se comparan los modelos generados por la estrategia exhaustiva contra los modelos basados en una estrategia de selección aleatoria de acuerdo al marco de trabajo de línea base propuesto por (Langdon, Dolado, Sarro & Harman, 2016). Como se menciona anteriormente, este proceso permite una comparación sin sesgo a partir de las métricas de desempeño calculadas con el mismo conjunto de datos con el que se construyen los modelos de estimación basados en los distintos esquemas de aprendizaje.

La Figura 5-15 presenta el proceso de aplicación del procedimiento de evaluación. Este proceso se aplica de forma automatizada con el soporte de una herramienta de evaluación de modelos que permite realizar la estimación del esfuerzo de desarrollo del software. En este caso, el profesional provee los datos históricos de proyectos para el entrena-

miento de los modelos, determina bajo qué conjunto de técnicas quiere evaluar los esquemas de aprendizaje, proporciona los nuevos datos y revisa y aprueba los resultados de evaluación del desempeño generados por la herramienta.

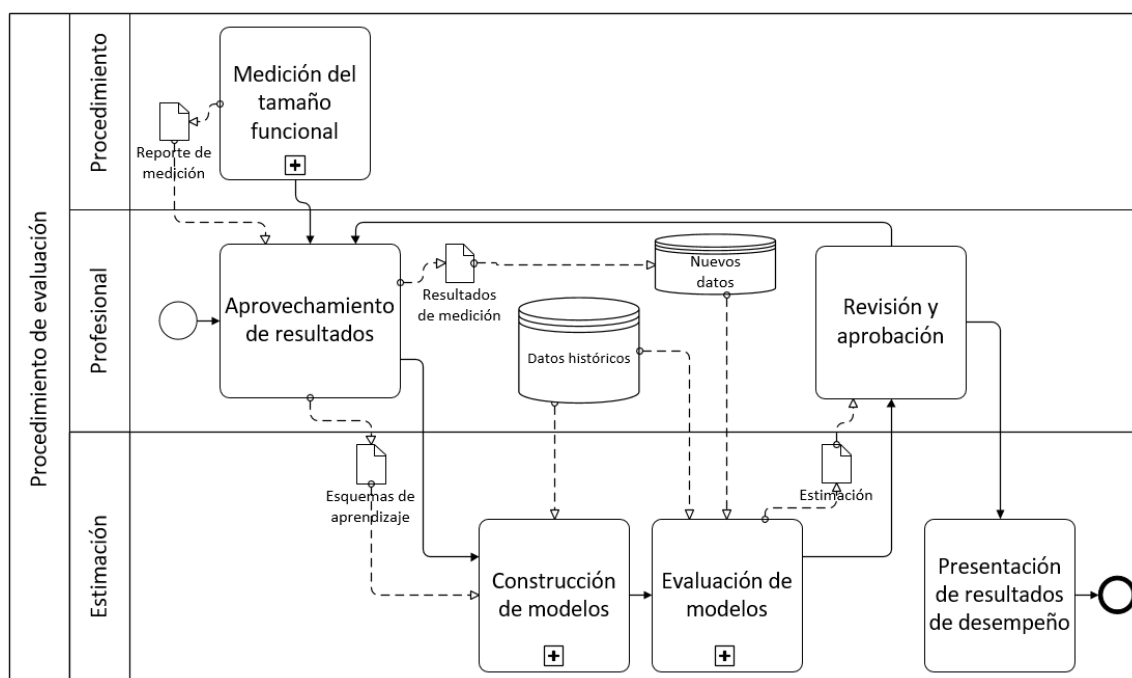


Figura 5-15 Proceso del procedimiento de aprovechamiento de resultados para la construcción de modelos de estimación.

Los estudios que listan las configuraciones de cada una de los componentes de la estrategia de evaluación de modelos y muestran los resultados del desempeño se detallan en los experimentos presentados en el Capítulo 7.

5.3.2. SALIDAS DEL PROCEDIMIENTO DE EVALUACIÓN

El procedimiento de evaluación reporta el desempeño de los modelos y el resultado de la estimación de esfuerzo de desarrollo para los nuevos datos con los siguientes detalles:

- Fecha de evaluación.

- Descripción del conjunto de datos históricos.
- Descripción de los nuevos datos.
- Lista de las técnicas de pre procesamiento, selección de atributos y algoritmos de aprendizaje.
- Lista de esquemas de aprendizaje evaluados con el detalle de las métricas de desempeño, parámetros de configuración y subconjunto de atributos seleccionados.
- Estimación del esfuerzo de desarrollo basado en el tamaño funcional para los nuevos datos.

El reporte de los resultados de la evaluación presenta el detalle de las comparaciones del desempeño de cada uno de los esquemas de aprendizaje, los parámetros de configuración y los subconjuntos de datos seleccionados para realizar la estimación lo que permite evaluar el impacto de cada una de las configuraciones evaluadas. El prototipo de automatización para apoyar el proceso de construcción y evaluación de los modelos de estimación de esfuerzo se describe en la Sección 5.4.

5.3.3. MÉTRICAS DE EVALUACIÓN DE DESEMPEÑO

El desempeño de los modelos se evalúa de acuerdo a las métricas comúnmente utilizadas en el área de modelos de estimación de esfuerzo de desarrollo del software. Estas comparan las mediciones estimadas contra los valores reales y calculan las variaciones encontradas para determinar el desempeño de cada modelo. Estas variaciones deben ser lo más pequeñas posibles dado que las desviaciones grandes pueden tener un impacto en el costo estimado para el desarrollo del software. Las métricas aplicadas para la evaluación de los modelos son: el coeficiente de correlación de Spearman (SC), la magnitud del error relativo (MRE), la magnitud relativa del error (MER), el error relativo balanceado (BRE), el error relativo balanceado invertido (IBRE) y el número de predicciones con un porcentaje (%) de los valores reales ($Pred_n$). Las mediciones individuales de los MRE promediadas generan la magnitud del error relativo promedio (MMRE) y la mediana de las mediciones

individuales de los MRE generen la magnitud del error relativo medio (MdmRE) que permite reducir el impacto de los valores atípicos. Para el MER y el BRE también se calculan los promedios y las medianas. Asimismo, se aplican las métricas del promedio de los errores absolutos (MMAE), el error de cuadrados medio (RMSE) y la exactitud estandarizada (SA) calculadas a partir de la aplicación de la estrategia de selección aleatoria de acuerdo al marco de trabajo de línea base propuesto por (Langdon, Dolado, Sarro & Harman, 2016). La descripción de las funciones de medición para cada una de las métricas es la siguiente:

Coefficiente de correlación de Spearman (SC):

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Donde d_i representa la diferencia entre los rangos asignados a cada una de las observaciones y n el número de observaciones. El coeficiente se utiliza como medida de la relación entre los valores estimados y reales.

Magnitud del error relativo (MRE):

$$MRE = \frac{|e_i - \hat{e}_i|}{e_i}$$

Donde e_i es valor real y \hat{e}_i es la observación con el valor estimado.

Magnitud relativa del error (MER):

$$MER = \frac{|e_i - \hat{e}_i|}{\hat{e}_i}$$

Donde e_i es valor real y \hat{e}_i es la observación con el valor estimado.

Error relativo balanceado (BRE):

$$BRE = \frac{|e_i - \hat{e}_i|}{\min(e_i, \hat{e}_i)}$$

Donde e_i es valor real y \hat{e}_i es la observación con el valor estimado.

Error relativo balanceado invertido (IBRE):

$$IBRE = \frac{|e_i - \hat{e}_i|}{\max(e_i, \hat{e}_i)}$$

Donde e_i es valor real y \hat{e}_i es la observación con el valor estimado.

Promedio de la magnitud del error relativo promedio (MMRE):

$$MMRE = \frac{100}{n} \sum_{i=1}^n \frac{|e_i - \hat{e}_i|}{e_i}$$

Donde e_i es valor real, \hat{e}_i es la observación con el valor estimado y n el número de observaciones.

Mediana de la magnitud del error relativo promedio (MdMRE):

$$MdMRE = \text{mediana}(MRE_i)$$

Donde se calcula la mediana de los resultados del MRE_i para el número de observaciones.

De la misma manera se calculan los promedios y medianas para las métricas MER, BRE, IBRE:

$MMER = \text{promedio}(MER_i)$, $MBRE = \text{promedio}(BRE_i)$, $MIBRE = \text{promedio}(IBRE_i)$

$MdMER = \text{mediana}(MER_i)$, $MdBRE = \text{mediana}(BRE_i)$, $MdIBRE = \text{mediana}(IBRE_i)$

Número de predicciones de los valores reales (Pred):

$$Pred(x) = \frac{100}{n} \sum_{i=1}^n \begin{cases} 1 & \text{si } MRE \leq x/100 \\ 0 & \text{en otro caso} \end{cases}$$

Donde x es el porcentaje de variación y n el número de observaciones.

Promedio de los errores absolutos (MMAE):

$$MMAE = \frac{100}{n} \sum_{i=1}^n |e_i - \hat{e}_i|$$

Donde e_i es valor real, \hat{e}_i es la observación con el valor estimado y n el número de observaciones.

Error de cuadrados medio (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \hat{e}_i)^2}$$

Donde e_i es valor real, \hat{e}_i es la observación con el valor estimado y n el número de observaciones.

Exactitud estandarizada (SA):

$$SA = 1 - \frac{MAR}{MAR_{P_0}} * 100$$

Donde MAR es el promedio de los residuos absolutos que es calculado como:

$$MAR = \frac{1}{N} \sum_{i=1}^N |e_i - \hat{e}_i|$$

Donde MAR_{P_0} es la línea base sin sesgo del MAR recomendado para conjuntos pequeños que es calculado como:

$$MAR_{P_0} = \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^{j<i} |e_i - e_j|$$

El uso, aplicación e interpretación de cada una de las métricas para el análisis de los resultados de desempeño de cada evaluación se detalla en los experimentos presentados en el Capítulo 7.

5.4. HERRAMIENTAS PROTOTIPO PARA EL SOPORTE DE LA METODOLOGÍA DE MEDICIÓN DEL TAMAÑO FUNCIONAL **aFPA**

El diseño del conjunto de herramientas de soporte a la metodología representa el producto del ciclo ingenieril (EC7).

Para el soporte de la ejecución de cada una de las etapas de los procedimientos de la metodología de medición **aFPA** se diseñan tres herramientas prototipo. Tal como se muestra en la Figura 5-16, el primer prototipo apoya el procedimiento de medición del tamaño funcional, el segundo prototipo apoya el procedimiento de aplicación del protocolo de verificación de exactitud y finalmente, el tercer prototipo apoya el procedimiento de construcción y evaluación de los modelos de estimación de esfuerzo.

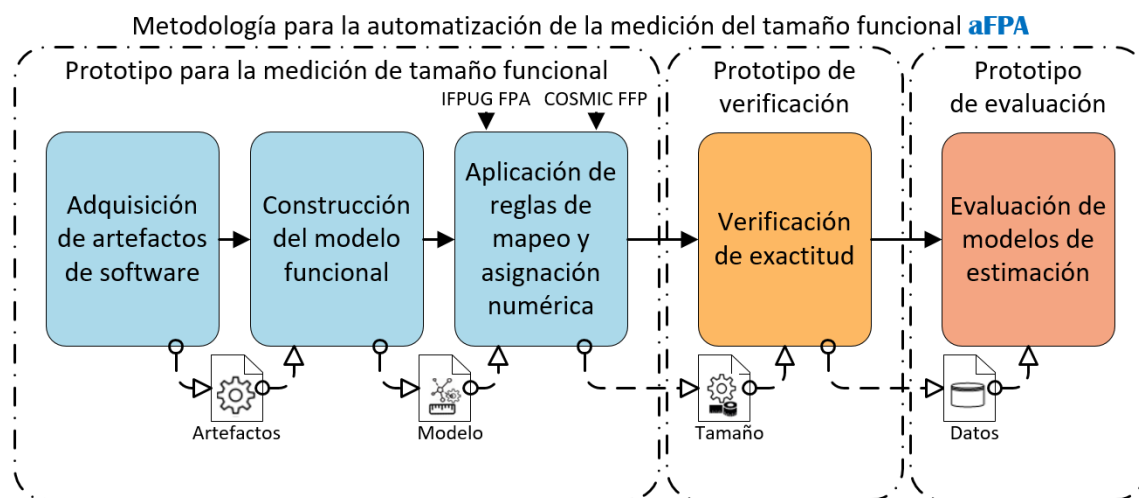


Figura 5-16 Prototipos para el soporte de la metodología **aFPA**.

5.4.1. MÓDULO DE MEDICIÓN DEL TAMAÑO FUNCIONAL

La Figura 5-17 muestra los componentes del módulo de medición del tamaño funcional que incluye el componente generador, constructor, medidor y reporteador (Quesada-López, Jenkins, Salas & Gómez, 2017b).

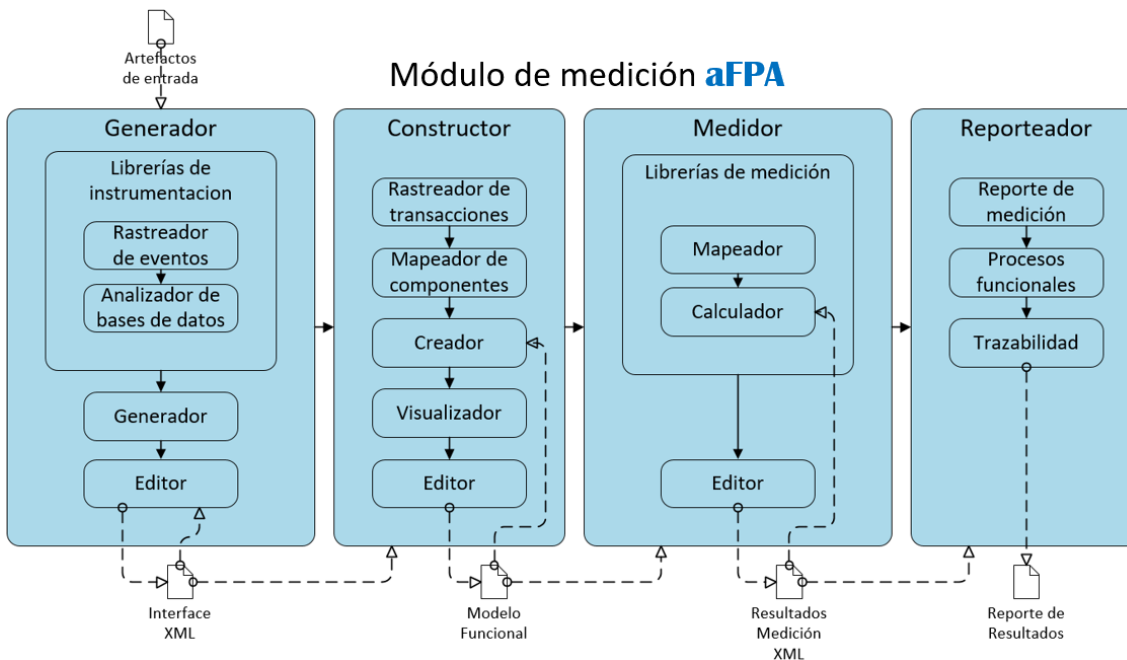


Figura 5-17 Componentes del módulo de medición.

Este módulo permite la medición del tamaño funcional a partir de los artefactos de entrada de acuerdo a los distintos niveles de automatización con los que se aplica el procedimiento e incluye los siguientes componentes:

- *Generador*: el componente de generación se encarga de crear un archivo de interfaz XML a partir de los artefactos de entrada que detalla toda la información requerida para la construcción del modelo funcional. El archivo de interfaz es generado a partir de la instancia de los objetos de la aplicación mediante la implementación de un conjunto de funcionalidades que automatizan la identificación de los elementos de la interfaz, los eventos desencadenadores, los elementos de almacenamiento persistente y los accesos desde la interfaz hasta el almacenamiento persistente que representan los procesos funcionales de la aplicación bajo medición. Este componente incluye las librerías de instrumentación del código fuente que permite el proceso de etiquetado y automatización para la identificación de los elementos de la aplicación medida en la plataforma tecnológica utilizada. La instrumentación permite realizar el proceso de rastreo de eventos y análisis de las estructuras de almacenamiento persistente. Asimismo, incluye

las librerías de generación de los archivos de interfaz XML que detallan los elementos identificados y la interfaz de edición que permite que el profesional puede revisar, y si es necesario editar los resultados.

- *Constructor*: el componente de construcción se encarga de construir el modelo funcional que incluye la información del modelo de eventos, modelo de datos y modelo de transacciones basado en la información provista en el archivo de interfaz XML. Este componente incluye las librerías de simulación de la ejecución de las transacciones mediante el rastreador de transacciones que genera una bitácora de los flujos de información los cuales son utilizados para crear el grafo que representa el modelo funcional de la aplicación bajo medición, las librerías de visualización del modelo funcional y la interfaz de edición que permite que el profesional pueda revisar, y si es necesario ajustar el modelo funcional.
- *Medidor*: el componente de medición se encarga de realizar el proceso de aplicación de las reglas de mapeo entre los elementos del modelo funcional y los elementos funcionales de los métodos de medición del tamaño funcional IFPUG FPA y COSMIC FFP, determinar los procesos funcionales, funciones transaccionales y funciones de datos únicos, cohesivos e independientes y calcular el tamaño funcional de acuerdo a un método de medición. Este componente incluye las librerías de medición que permiten el mapeo de los componentes funcionales mediante la aplicación de las reglas de mapeo y el cálculo del tamaño funcional de acuerdo a las reglas de asignación numérica y la interfaz de edición que permite que el profesional pueda revisar, y si es necesario ajustar los elementos de influencia del conteo del modelo funcional.
- *Reporteador*: el componente de reporte se encarga de realizar el proceso de presentación de los resultados de medición. Este componente incluye el reporte de los resultados de medición total, el detalle de todas las etapas del proceso y el desglose de las mediciones manteniendo la trazabilidad con el modelo funcional y sus componentes. Los resultados obtenidos por este reporte permiten realizar la aplicación del protocolo de verificación de exactitud de los resultados de medición.

Para validar el modelo de la herramienta propuesta se desarrolla la primera versión del prototipo de medición **aFPA** en el ambiente de operación de la plataforma de desarrollo *Visual Studio .Net* y el *API* del *Framework .Net*. En este caso, el objetivo del prototipo es aplicar el Nivel 3 de automatización del procedimiento de medición del tamaño funcional. Para la implementación del componente *Generador*, se desarrollan las librerías de generación del archivo de interfaz mediante las librerías *System XML*, las librerías de instrumentación que permite el proceso de etiquetado y automatización para la identificación de los elementos de la aplicación medida mediante las librerías *System.Data.Entity.DbContext*, *System.Data.Entity.Infrastructure.Interception.DbCommandInterceptor*, *System.Reflection* y *System.Diagnostics*. Otras opciones para realizar esta implementación son las librerías del *Windows Automation API System, UI automation Microsoft VisualStudio TestTools UITest Common UIMap*. El proceso de rastreo de eventos y análisis de las estructuras de almacenamiento persistente fue implementado mediante las librerías *Microsoft ScriptDom*. Para la implementación del componente *Constructor*, se desarrollan las librerías de lectura del archivo de interfaz mediante las librerías *System XML*, las librerías de simulación de la ejecución de las transacciones con el rastreador de transacciones mediante las librerías *Microsoft ScriptDom*, las librerías y la interfaz de visualización del modelo funcional mediante las librerías del *Neo4j API, Neo4j for .Net Developers* y el lenguaje *Neo4j Graph Query Language Cypher*. Para la implementación del componente *Medidor*, se desarrollan las librerías de mapeo y medición mediante librerías propias que utilizan las librerías del *Neo4j API, Neo4j for .Net Developers* y el lenguaje *Neo4j Graph Query Language Cypher*. Los componentes de edición y las aplicaciones que soportan la interacción con el profesional se desarrollaron utilizando las tecnologías *Asp.Net MVC* y las interfaces de archivos XML mediante las librerías *System XML*.

El detalle de cómo se utiliza cada uno de los componentes y de las librerías del módulo de medición se describe en los casos de estudio presentados en el Capítulo 6 Sección 6.1. Los casos de estudio detallan los procesos de instrumentación del código fuente y la inclusión de las librerías de automatización específicas que dependen de la plataforma tecnológica seleccionada en cada uno de ellos.

5.4.2. MÓDULO DE VERIFICACIÓN DE LA EXACTITUD

La Figura 5-18 muestra los componentes del módulo de verificación de la exactitud (Madrigal-Sánchez, Quesada-López & Jenkins, 2018). Este módulo permite la verificación de los resultados del tamaño funcional a partir de la comparación de las mediciones con los valores de control e incluye los siguientes componentes:

- *Mapeador*: el componente de mapeo se encarga de procesar los archivos de interfaz XML que contienen los resultados de las mediciones de tamaño funcional y de la instanciación de los modelos requeridos para realizar el proceso de comparación de mediciones. En este caso, el prototipo construye un modelo por cada uno de los resultados de medición provistos por el profesional para procesar la comparación. Este componente incluye las librerías de lectura de los archivos de interfaz XML y las librerías de generación de los modelos. Los modelos se construyen como una representación de nodos para cada uno de los componentes funcionales del método de medición.
- *Verificador*: el componente de verificación realiza los procesos de comparación de los modelos construidos para cada una de las mediciones provistas. El proceso de comparación se realiza contra el modelo construido a partir de los valores de control. El proceso se implementa mediante un recorrido en profundidad que identifica un pareo de nodos entre las mediciones. Si todos los nodos de un camino son pareados entre grafos entonces se identifica el mismo proceso funcional en ambos modelos. Si se encuentra una diferencia, entonces se reportan como rutas inválidas y se detalla el error manteniendo trazabilidad entre los componentes funcionales que pudieron ser pareados y los que no. A partir de la identificación de las rutas inválidas se realiza el proceso de análisis de similitudes para sugerir posibles causas asignables. Los detalles están descritos en (Madrigal-Sánchez, Quesada-López & Jenkins, 2018).
- *Reporteador*: el componente de reporte se encarga de realizar el proceso de presentación de los resultados de verificación. Este componente incluye el reporte de los resultados

totales, el detalle de la comparación de todas las etapas del proceso y el desglose entre las mediciones manteniendo la trazabilidad con el modelo y sus componentes. Los resultados obtenidos por este reporte identifican posibles causas asignables que son notificadas al profesional.

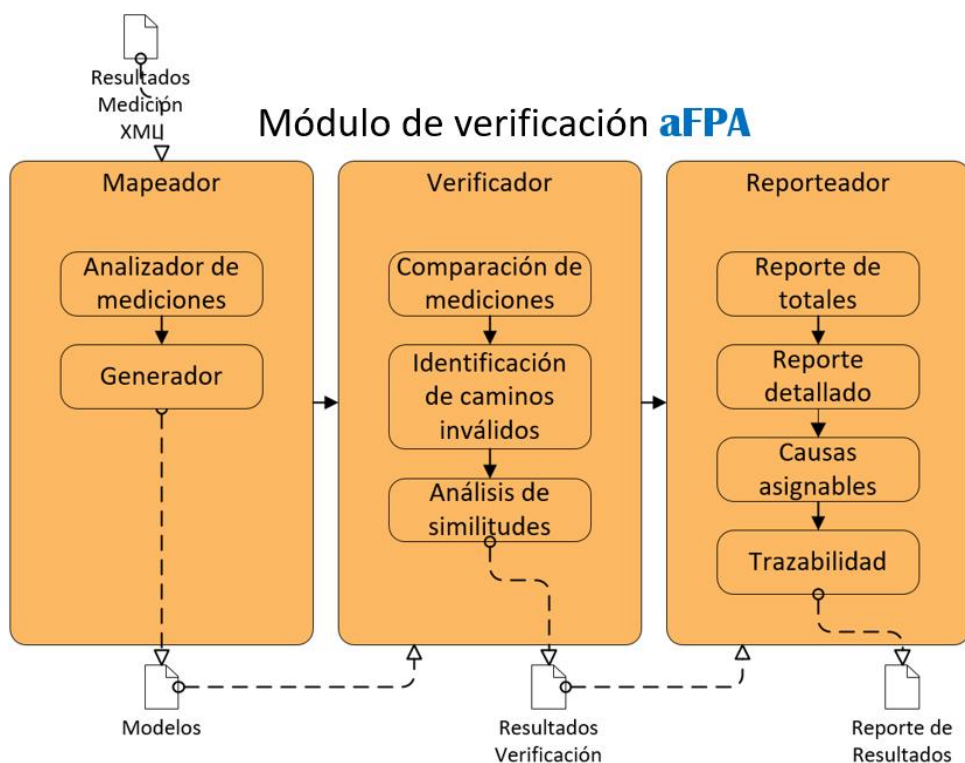


Figura 5-18 Componentes del módulo de verificación.

Para validar el modelo de la herramienta propuesta se desarrolla la primera versión del prototipo de verificación **aFPA** en el ambiente de operación de la plataforma de desarrollo *Java* y el *API* del *Framework JDK*. En este caso, el objetivo del prototipo es aplicar el Nivel 1 de automatización del protocolo de verificación de la exactitud de los resultados del tamaño funcional. Para la implementación del componente de mapeo, se desarrollan las librerías de procesamiento de los archivos de interfaz mediante las librerías de análisis *Java XML* y *W3C Dom*, las librerías de generación de los modelos mediante las librerías del *Neo4j API*, *Neo4j for Java Spring Users*, *Neo4j for Java Developers* y el lenguaje *Neo4j Graph*

Query Language Cypher. Del mismo modo, para la implementación del componente de verificación se desarrollan las librerías de comparación de los modelos mediante el lenguaje *Java* y *Neo4j Graph Query Language Cypher*. Los componentes que soportan la interacción con el profesional se desarrollaron utilizando las tecnologías del *Framework Spring-Boot* y *Neo4j for Java Spring Users*.

El detalle de cómo se utiliza el prototipo y sus componentes se detalla en los casos de estudio y los cuasi experimentos presentados en el Capítulo 6 Sección 6.1.

5.4.3. MÓDULO DE EVALUACIÓN DE LOS MODELOS DE ESTIMACIÓN

La Figura 5-19 muestra los componentes del módulo de evaluación de los modelos de estimación de esfuerzo para el aprovechamiento de los resultados de medición. Este módulo permite la evaluación de los modelos de estimación generados a partir de las mediciones del tamaño funcional en contexto e incluye los siguientes componentes:

- *Constructor*: el componente de construcción de modelos se encarga de construir y evaluar el desempeño de los modelos creados a partir de cada uno de los esquemas de aprendizaje y los datos históricos. Este componente incluye el pre procesamiento de datos, la selección de atributos y la configuración de los parámetros de los algoritmos de aprendizaje. El proceso de evaluación construye los modelos y determina los mejores esquemas de aprendizaje que son utilizados para la estimación de los nuevos datos.
- *Evaluador*: el componente de evaluación permite aplicar los mejores esquemas de aprendizaje para la construcción de los modelos utilizados en la estimación del esfuerzo de desarrollo del software. Este componente incluye las mejores técnicas de pre procesamiento, selección de atributos, configuración de atributos y algoritmo de aprendizaje que son aplicados en la estimación del esfuerzo de los nuevos datos.
- *Reporteador*: el componente de reporte se encarga de realizar el proceso de presentación de los resultados de evaluación de los modelos y de la estimación del esfuerzo de desarrollo. Este componente incluye el reporte de los resultados de desempeño para cada

uno de los esquemas de aprendizaje y el resultado del proceso de estimación. Los resultados obtenidos por este componente son reportados al profesional.

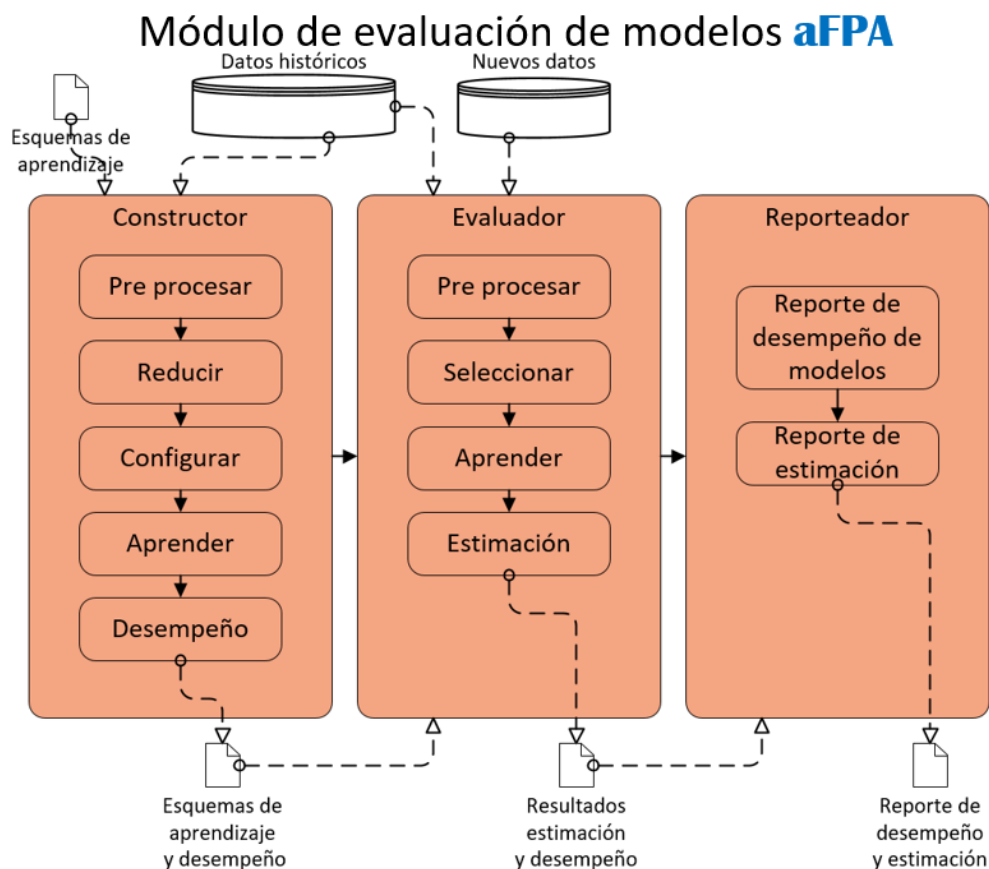


Figura 5-19 Componentes del módulo de evaluación de modelos.

Para validar el modelo de la herramienta propuesta se desarrolla la primera versión del prototipo de evaluación **aFPA** en el ambiente de operación de la plataforma de desarrollo *Java* y el *API* del *Framework JDK*. En este caso, el objetivo del prototipo es evaluar los resultados del tamaño funcional para la construcción de modelos de estimación de esfuerzo. Para la implementación del componente de construcción y evaluación se desarrollan las librerías de procesamiento de los archivos de interfaz ARFF mediante las librerías de *Waikato Environment for Knowledge Analysis (WEKA)* para *Java* (Witten, Frank, Hall & Pal, 2016). Del mismo modo se utilizan estas librerías de *WEKA* para la implementación de cada

uno de los pasos de construcción y evaluación, estas librerías proveen la implementación de las técnicas de pre procesamiento, selección de atributos y algoritmos de aprendizaje. Para el cálculo de las métricas de desempeño y la estrategia de selección aleatoria de línea base se utilizaron las librerías *Commons Math3 Stat* de *Apache*. Los componentes que soportan la interacción con el profesional se desarrollaron utilizando las tecnologías del *Framework NetBeans IDE*.

El detalle de cómo se utiliza el prototipo y sus componentes, y las configuraciones particulares utilizadas en la estrategia de evaluación se presentan en los estudios empíricos del Capítulo 6 Sección 6.3.

5.5. EXTENSIBILIDAD DEL PROCEDIMIENTO DE MEDICIÓN EN DISTINTOS CONTEXTOS DE DESARROLLO

Tal como se ha descrito anteriormente, el procedimiento para la automatización de la medición del tamaño funcional se lleva a cabo en tres grandes fases: 1) el diseño y aplicación del procedimiento para obtener el tamaño funcional; 2) el protocolo de verificación de los resultados de la exactitud de las mediciones; 3) la evaluación de los resultados de medición para aprovecharlos en la construcción de modelos de estimación de esfuerzo. El protocolo de verificación y la evaluación de modelos de estimación de esfuerzo pueden ser utilizados en conjunto o independientemente de la primera fase, lo que permite que puedan ser utilizados en distintos contextos de desarrollo sin realizar ninguna adaptación.

Por otro lado, el diseño y aplicación del procedimiento para obtener el tamaño funcional se ha desarrollado para modelos funcionales obtenidos a partir del análisis dinámico del código fuente, el cual toma como entrada dichos artefactos de software del código fuente y construye el modelo funcional que es mapeado a los métodos de medición del tamaño funcional IFPUG FPA y COSMIC FFP para obtener el tamaño funcional de la aplicación medida. En este caso, el procedimiento de medición **aFPA** se puede llevar a cabo en distintos niveles de automatización dependiendo del control del profesional sobre el contexto de desarrollo y la aplicación bajo medición. Dado que existe la posibilidad de aplicar

el procedimiento en distintos niveles de automatización, es de interés determinar las posibilidades de utilizar las reglas del procedimiento bajo distintos contextos de desarrollo que incluyen el uso de diferentes artefactos de entrada recolectados en las distintas etapas del ciclo de vida de desarrollo.

La posibilidad de extender el procedimiento para varias etapas de desarrollo permite a los profesionales aplicar el mismo procedimiento en las distintas fases del ciclo de vida para realizar comparaciones. Por ejemplo, el proceso de medición podría realizarse durante el análisis de requerimientos, el diseño, la codificación, las pruebas y el mantenimiento, tal como se discute en (Garmus & Herron, 2001).

En las etapas tempranas del desarrollo, análisis de requerimientos y diseño, el paso clave para habilitar la integración del procedimiento **aFPA** a los procesos ingenieriles que realiza un equipo de desarrollo es el de la identificación de los componentes funcionales del procedimiento de medición. Es así como el procedimiento podría ser utilizado en el nivel 0, nivel 1 y nivel 2 de automatización para cualquier etapa temprana del desarrollo. Sin embargo, el mayor aprovechamiento del procedimiento puede obtenerse en el nivel 2 de automatización, en el cual a partir de la identificación de los componentes funcionales se utiliza la interfaz que recibe como entrada un archivo XML con el mapa de la interfaz de usuario y las estructuras de almacenamiento persistente. En este caso, el procedimiento automáticamente realiza la construcción del modelo funcional y la medición del tamaño funcional mediante la aplicación de las reglas de mapeo y las reglas de asignación numérica. En el caso de las etapas posteriores a la codificación y dependiendo del contexto de desarrollo y el nivel de control del profesional sobre la aplicación medida, se deben analizar las posibilidades de utilizar el procedimiento en los distintos niveles de automatización para realizar la medición.

A continuación se proponen dos extensiones del procedimiento de medición **aFPA** para que pueda ser aplicado en la etapa de análisis de requerimientos utilizando artefactos de requerimientos en lenguaje natural bajo el estándar IEEE Std. 830-1998 y el ISO/IEC/IEEE 29148:2011 y en la etapa de diseño utilizando modelos conceptuales construidos bajo una herramienta de desarrollo dirigido por modelos.

5.5.1. ETAPA DE ANÁLISIS DE REQUERIMIENTOS UTILIZANDO LENGUAJE NATURAL BAJO EL ESTÁNDAR IEEE STD. 830-1998 Y EL ESTÁNDAR ISO/IEC/IEEE 29148:2011

Para extender el procedimiento de medición **aFPA** para que pueda ser aplicado en la etapa de análisis de requerimientos utilizando requerimientos en lenguaje natural bajo los estándares IEEE Std. 830-1998 e ISO/IEC/IEEE 29148:2011, se requiere identificar los componentes funcionales del procedimiento de medición a partir de la especificación de requerimientos. Para esto, primero se determina el metamodelo de los requerimientos y luego se definen las reglas de mapeo entre el metamodelo de los artefactos de requerimientos y el metamodelo del procedimiento de medición. Posterior a que se realiza el mapeo de los componentes que requiere el procedimiento de medición, este automáticamente realiza la construcción del modelo funcional y la medición del tamaño funcional mediante la aplicación de las reglas de mapeo y las reglas de asignación numérica descritas en la Sección 5.1.1.4.

En las siguientes subsecciones se detallan los principales aspectos necesarios para extender el diseño del procedimiento. Los aspectos del diseño del procedimiento necesarios para ser aplicado en estos artefactos, solo describen las principales extensiones realizadas al procedimiento original descrito en la Sección 5.1, lo que significa que las partes no descritas en esta sección no son modificadas con respecto a las originales.

5.5.1.1. EXTENSIÓN AL PROCEDIMIENTO

El objetivo es *diseñar un procedimiento de medición con el propósito de estimar el tamaño funcional de aplicaciones de software a partir del análisis de la especificación de los requerimientos descritos bajo el estándar IEEE Std. 830-1998.*

Los modelos funcionales de representación del software se generan a partir del análisis estático del documento de requerimientos realizado en la etapa de especificación de requerimientos. Por tanto, el alcance de la medición es el modelo funcional obtenido a partir de la especificación de requerimientos funcionales.

5.5.1.2. METAMODELO DE LOS REQUERIMIENTOS FUNCIONALES DESCRITOS BAJO EL ESTÁNDAR IEEE STD. 830-1998

Los estándares de requerimientos IEEE Std. 830-1998 e ISO/IEC/IEEE 29148:2011 detallan los procesos y productos relacionados con la ingeniería de requerimientos. Dentro de los productos, detallan las partes que debe contener un documento de especificación de requerimientos de software. En particular, un documento de requerimientos lista las funciones del producto y los requisitos funcionales específicos que detallan las funciones que la aplicación de software requiere implementar, las interfaces externas y los requisitos del almacenamiento persistente. El estándar ISO/IEC/IEEE 29148:2011 reemplaza, entre otros, el estándar IEEE Std. 830-1998 y adapta cómo se reporta un documento de especificación de requerimientos. El estándar ISO/IEC/IEEE 29148:2011 contiene las disposiciones para los procesos y los productos relacionados con la ingeniería de requerimientos para sistemas y servicios de software durante el ciclo de vida de desarrollo, en este caso es de interés el documento de especificación de requerimientos.

Para realizar la aplicación del proceso de medición se requiere contar con el documento de requerimientos funcionales al cual se le aplica un proceso de etiquetado que busca identificar para cada uno de los elementos que son mapeados a los componentes del procedimiento de medición (Quesada-López, Madrigal & Jenkins, 2016). La Figura 5-20 detalla el metamodelo que describe el documento de especificación de requerimientos de una aplicación bajo medición. Primero se realiza la especificación de requerimientos y luego se lleva a cabo el proceso de etiquetado sobre el documento de requerimientos identificando dentro de las funciones del producto y los requerimientos funcionales las entidades, atributos, eventos, mensajes y funciones únicas que se describen. En este caso el documento queda etiquetado para realizar la identificación de cada uno de los elementos del metamodelo del procedimiento. El metamodelo de requerimientos requiere la identificación de los siguientes elementos:

- Aplicación (*Application*): una aplicación se compone de uno o más requerimientos. La aplicación contiene todos los elementos identificados en el documento de requerimientos.
- Requerimiento (*Requirements*): un requerimiento detalla la funcionalidad requerida por un usuario de la aplicación. Un requerimiento puede ser de tipo requerimiento base o requerimiento agregado que se compone de un conjunto de requerimientos base.
- Control (*Control*): un control permite a un usuario disparar un evento o visualizar datos de salida de los procesos funcionales. Este puede ser de tipo evento o mensaje.
- Función (*Function*): una función corresponde a una operación de lectura o escritura de datos que se realiza en respuesta a un requerimiento mediante un evento realizado por el usuario. Una función permite la ejecución de transacciones que realizan operaciones sobre las entidades y los atributos.
- Entidad (*Entity*): una entidad representa un grupo lógico de datos donde se almacena la información de la aplicación. Una entidad puede ser interna o externa dependiendo si pertenece al almacenamiento persistente de la aplicación o a una interfaz. Un grupo lógico puede estar compuesto por uno o varias entidades de datos.
- Atributo (*Attribute*): un atributo es un elemento de datos que pertenece a una entidad, puede representar un campo físico o calculado y tiene significado para el usuario funcional.

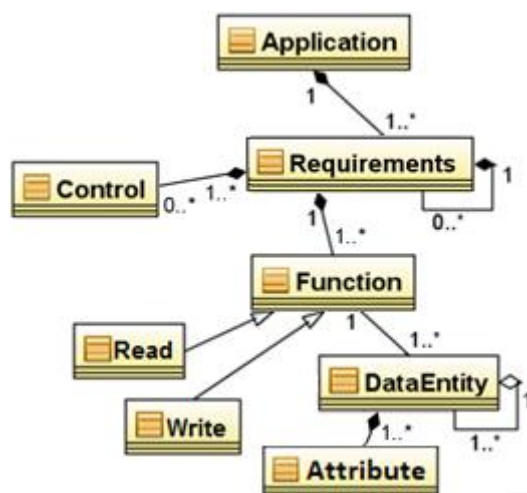


Figura 5-20 Metamodelo de la especificación de requerimientos.

El modelo describe todas las funcionalidades del usuario final que permite generar una aplicación funcional. En este caso, el modelo de la aplicación contiene la información para la construcción del modelo funcional del procedimiento **aFPA** utilizado para obtener el tamaño funcional mediante la aplicación de las reglas de mapeo y asignación numérica.

5.5.1.3. REGLAS DE MAPEO ENTRE METAMODELOS

A partir de la instanciación del metamodelo, el proceso de mapeo determina los candidatos de cada uno de los componentes funcionales del procedimiento de medición. El objetivo del mapeo es establecer las relaciones entre los conceptos de cada uno de los metamodelos. Aplicando las reglas de mapeo se obtienen los componentes funcionales básicos para el conteo del tamaño funcional. Las reglas de mapeo entre el metamodelo de la especificación de requerimientos funcionales y el modelo funcional del procedimiento son las siguientes:

- R1. La aplicación genera un elemento *Application*. Los límites de la aplicación se determinan por el documento de requerimientos.
- R2. Cada requerimiento genera un elemento de tipo *Control* de interfaz y el elemento tipo *Function* correspondiente. Los requerimientos tipo base son asociados con sus respectivos requerimiento tipo agregado asociados con un elemento de navegación.
- R3. Cada control genera un elemento tipo *Control* asociado a un control de interfaz.
- R4. Cada función genera un elemento tipo *Function* y un elemento tipo transacción cuando realiza operaciones sobre un elemento tipo *Entity*. Cada una de las funciones se asocian con sus respectivas entidades y atributos de datos (*Attribute*). El elemento transacción es clasificado como operación de lectura y/o escritura.
- R5. Cada entidad genera un elemento *Entity* que representa una tabla de la base de datos, archivo o un grupo de tablas o archivos (almacenamiento persistente) que constituye un agrupamiento lógico.

- R6. Cada elemento *Attribute* genera un elemento de datos que puede representar un campo físico o calculado y que se encuentra asociado a una entidad de datos.
- R7. La aplicación genera un elemento *DataBase* que representa la o las bases de datos (almacenamiento persistente) con las cuales la aplicación interactúa durante la ejecución de las funcionalidades del usuario funcional.

El mapeo de los componentes funcionales del procedimiento de medición, la construcción del modelo funcional, las reglas de mapeo para cada uno de los métodos de medición, las reglas de asignación numérica, el proceso de aplicación, la verificación y la evaluación de los resultados se realiza de acuerdo a las reglas y procesos originales descritos en la Sección 5.1.

Para la aplicación del procedimiento utilizando el documento de especificación de requerimientos en la actividad de adquisición de los artefactos de software, el artefacto de software utilizado para la aplicación del procedimiento de medición es el documento de análisis de requerimientos etiquetado. El resultado de aplicar el procedimiento de medición es la estimación del tamaño funcional que depende del nivel de granularidad de los requerimientos. El profesional es el encargado de definir, en la estrategia de medición, los niveles de granularidad utilizados. Esta flexibilidad permite que el procedimiento pueda medir el tamaño de la funcionalidad conforme evoluciona el desarrollo y se formalizan los requerimientos. En este caso el profesional administra las expectativas de exactitud esperada para procesos de estimación temprana versus las realizadas en etapas donde los artefactos tienen un nivel de granularidad superior.

El detalle de cómo se utiliza el procedimiento de medición con los documentos de especificación de requerimientos se describe en el caso de estudio presentado en el Capítulo 6 Sección 6.1 y Sección 6.2.

5.5.2. ETAPA DE DISEÑO UTILIZANDO MODELOS CONCEPTUALES CONSTRUIDOS BAJO UNA HERRAMIENTA DE DESARROLLO DIRIGIDO POR MODELOS

Para extender el procedimiento de medición del tamaño funcional para que pueda ser aplicado en la etapa de diseño utilizando modelos conceptuales generados por una herramienta de desarrollo dirigido por modelos, se requiere identificar los componentes funcionales del procedimiento **aFPA** a partir de los componentes del modelo conceptual. Para esto, primero se determina el metamodelo de los modelos conceptuales y luego se definen las reglas de mapeo entre el metamodelo de los artefactos del modelo conceptual y el metamodelo del procedimiento de medición. Posterior a que se realiza el mapeo de los componentes, este automáticamente realiza la construcción del modelo funcional y la medición del tamaño funcional mediante la aplicación de las reglas de mapeo y las reglas de asignación numérica descritas en la Sección 5.1.1.4.

En las siguientes subsecciones se detallan los principales aspectos necesarios para extender el diseño del procedimiento para los modelos conceptuales generados por la herramienta de desarrollo dirigido por modelos. Los aspectos del diseño del procedimiento solo describen las principales extensiones realizadas al procedimiento original descrito en la Sección 5.1, lo que significa que las partes no descritas en esta sección no son modificadas con respecto a las originales. En este caso, sólo se describen los cambios necesarios para explicar cómo es aplicado en este contexto de desarrollo y con estos artefactos de software.

5.5.2.1. EXTENSIÓN AL PROCEDIMIENTO

El objetivo es diseñar un procedimiento de medición con el propósito de estimar el tamaño funcional de aplicaciones de software a partir del análisis de los modelos conceptuales generados por una herramienta de desarrollo dirigido por modelos.

Los modelos funcionales de representación del software se generan a partir del análisis estático y dinámico del modelo conceptual realizado en la plataforma de desarrollo dirigido por modelos. Por tanto, el alcance de la medición es el modelo funcional obtenido a partir del modelo conceptual.

5.5.2.2. METAMODELO DE LOS MODELOS CONCEPTUALES GENERADOS EN UNA HERRAMIENTA DE DESARROLLO DIRIGIDA POR MODELOS

FastWorks es una plataforma de desarrollo dirigida por modelos que consiste en una herramienta de especificación y generación de código que permite el modelado y desarrollo de aplicaciones transaccionales suministrando una arquitectura estandarizada que soporta la implementación de los módulos y sus funcionalidades. Esta ha sido desarrollada y mantenida por la organización Grupo Asesor en Informática S.A. durante más de 15 años, obteniendo gran cantidad de casos de éxito en el desarrollo de aplicaciones para sus clientes, en su mayoría en el sector de Gobierno. La plataforma es utilizada para el desarrollo de sistemas de información basados en la plataforma *Microsoft .Net* (Quesada-López, Jenkins, Salas & Gómez, 2017b).

Para la producción del software se modelan en FastWorks un conjunto de elementos y sus relaciones que representan una aplicación de software. La Figura 5-21 detalla el metamodelo para la construcción de aplicaciones bajo esta herramienta. En las primeras fases del proceso de desarrollo se especifica el modelo de la aplicación que incluye los módulos, la base de datos y todos los elementos que describen la aplicación.

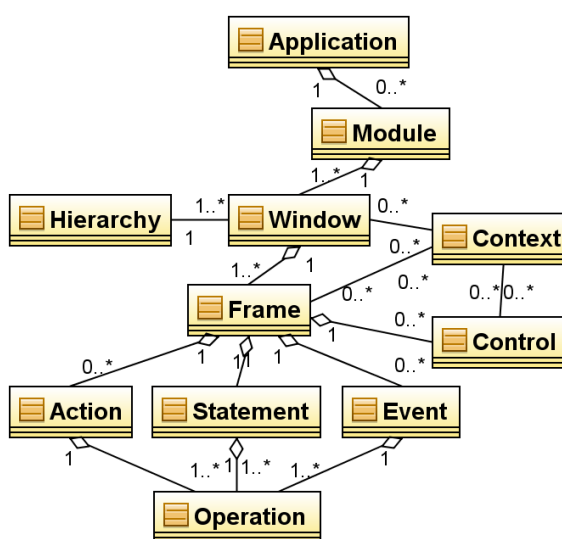


Figura 5-21 Metamodelo de la plataforma dirigida por modelos.

El modelo funcional requiere la identificación de los siguientes elementos del modelo conceptual:

- **Módulo (*Module*):** una aplicación se compone de uno o más módulos. Un módulo se compone de una o más ventanas y donde se define una ventana principal para cada módulo.
- **Ventana (*Window*):** es una pantalla en el módulo que puede contener uno o más marcos, con interacción entre ellos y una organización específica en la ventana.
- **Marco (*Frame*):** contiene una o más sentencias que permiten la administración de datos en el marco, uno o más controles, acciones y eventos. Los marcos se clasifican en tipo Listado (LL) que contienen un único control de tipo tabla o tipo Formulario (FF) y contiene múltiples controles.
- **Base de datos (*DataBase*):** esta detalla las entidades de datos de la aplicación y todos sus atributos.
- **Sentencia (*Statements*):** corresponde a una sentencia de programación para la administración de los datos de un marco. Esta puede incluir sentencias SQL para las bases de datos, además pueden representar llamadas a vistas, procedimientos y funciones almacenadas. Existen sentencias de marco y de acción.
- **Operación (*Operation*):** corresponde a una operación de programación para la administración de los datos que se realiza en respuesta a una solicitud de datos o un evento realizado por el usuario.
- **Control (*Control*):** permite a un usuario la visualización, entrada o salida de los datos en un marco. Un control puede contener otros controles y puede tener relacionados uno o varios eventos.
- **Acción (*Action*):** permite la ejecución de transacciones en el marco al usuario basado en uno o más sentencias. Una transacción permite realizar operaciones sobre otros controles o sobre las bases de datos.

- Evento (*Event*): permite la ejecución de transacciones como respuesta a eventos realizados por el usuario en el marco. Puede realizar operaciones sobre otros controles o sobre la base de datos mediante una acción.
- Jerarquías (*Hierarchy*): permiten organizar las ventanas posibilitando ventanas padre y una o varias ventanas hijas.
- Contexto (*Context*): representa la información global de la aplicación. Habilita un diccionario disponible para que las ventanas, marcos y controles puedan leer o escribir valores para establecer relaciones o interacciones entre ellos.

El modelo describe todas las funcionalidades del usuario final que permite generar una aplicación completamente funcional. En este caso, el modelo de la aplicación contiene toda la información necesaria para la construcción del modelo funcional del procedimiento utilizado para obtener el tamaño funcional mediante la aplicación de las reglas de mapeo y de asignación numérica.

5.5.2.3. REGLAS DE MAPEO ENTRE METAMODELOS

A partir de la instanciación del metamodelo, el proceso de mapeo determina los candidatos de cada uno de los componentes funcionales del procedimiento de medición. El objetivo del mapeo es establecer las relaciones entre los conceptos de cada uno de los metamodelos. Aplicando las reglas de mapeo se obtienen los componentes funcionales básicos para el conteo del tamaño funcional. Las reglas de mapeo entre el metamodelo de la plataforma de desarrollo dirigida por modelos y del modelo funcional del procedimiento son las siguientes:

- R1. Un módulo genera un elemento aplicación. Los límites de la aplicación se determinan por el modelo que representa un módulo de la aplicación.
- R2. Cada ventana, marco y control genera un control. Para todo control, si el control tiene una sentencia relacionada, se define un evento.

- R3. Cada sentencia, acción y evento genera un evento. En este evento la sentencia de programación, acción y evento se trata como una función
- R4. Cada operación genera una función con sus respectivas transacciones. Cada sentencia de programación de base de datos se mapea a una función del evento mediante una o más transacciones.
- R5. La base de datos genera las entidades y elementos de datos y un elemento de base de datos.

Al igual que con la extensión explicada en la sección anterior, el mapeo de los componentes funcionales del procedimiento de medición, la construcción del modelo funcional, las reglas de mapeo para cada uno de los métodos de medición, las reglas de asignación numérica, el proceso de aplicación, la verificación y la evaluación de los resultados se realiza de acuerdo a las reglas y procesos originales descritos en la Sección 5.1.

Para la aplicación del procedimiento utilizando los modelos conceptuales, en la actividad de adquisición de los artefactos de software, el artefacto de software utilizado para la aplicación del procedimiento de medición es el modelo conceptual que es provisto por la plataforma de desarrollo dirigida por modelos. La plataforma permite la construcción de modelos tempranos simplificados para estimación y posteriormente los modelos completos que se utilizan para la generación de código fuente. El procedimiento de medición puede ser aplicado con ambos modelos conceptuales. En este caso, los modelos completos usados para generar el código fuente de la aplicación se encuentran a un bajo nivel de granularidad detallando todas las funciones que se proveen al usuario final. Por lo tanto, el nivel de granularidad para el método COSMIC FFP es el Nivel 3 (sub funciones detalladas) y el nivel de conteo para el método IFPUG FPA es el Nivel 1 (trazable y etiquetado hacia los artefactos medidos), ambos los niveles más detallados de conteo. Esto permite mejorar los resultados en la exactitud de los conteos. En el caso de los modelos tempranos simplificados, el nivel de granularidad puede variar por lo que el profesional es el encargado de definir en su estrategia de medición los niveles de granularidad utilizados y las expectativas de exactitud

esperadas. Esta flexibilidad permite que el procedimiento pueda medir el tamaño de la funcionalidad conforme evoluciona el desarrollo dirigido por modelos y de acuerdo a las expectativas de exactitud esperadas por el profesional.

El detalle de cómo se utiliza el procedimiento de medición bajo la herramienta de desarrollo dirigida por modelos se describe en el caso de estudio de industria presentado en el Capítulo 6 Sección 6.1.5.

5.6. RESUMEN

Este capítulo ha presentado la metodología para la automatización de la medición del tamaño funcional del software **aFPA**. La metodología **aFPA** detalla los lineamientos para obtener el tamaño funcional a partir de modelos funcionales obtenidos del análisis dinámico del código fuente y permite la medición funcional de los artefactos de entrada de acuerdo a los métodos de medición IFPUG FPA y COSMIC FFP. Las secciones del capítulo han presentado el diseño y el proceso de aplicación del procedimiento de medición, el protocolo del procedimiento de verificación de la exactitud y el procedimiento de evaluación para el aprovechamiento de los resultados de medición utilizando modelos de estimación de esfuerzo de desarrollo. Asimismo, se han descrito los prototipos que soportan la metodología de medición mediante la implementación de tres módulos funcionales para la medición, verificación y evaluación. Finalmente, se ha descrito una propuesta para la extensión del procedimiento en la medición de artefactos de software bajo distintos contextos de desarrollo para las etapas del ciclo de vida de desarrollo del análisis de requerimientos y del diseño. Las evaluaciones realizadas para cada una de las propuestas descritas en este capítulo son presentadas en el Capítulo 6 y el Capítulo 7 de este documento de tesis.

Capítulo 6. EVALUACIÓN EMPÍRICA DE LOS PROCEDIMIENTOS DE MEDICIÓN Y VERIFICACIÓN DE LA METODOLOGÍA **aFPA**

Este capítulo agrega conocimiento para la consecución del objetivo específico 2 (OE1) que permite responder parte de la primera pregunta de investigación (KP1) y genera evidencia que permite responder la tercera pregunta de investigación (KP3) del objetivo específico 3 (OE3) sobre la efectividad de la metodología para la automatización de la medición del tamaño funcional del software.

En este capítulo se presentan los estudios empíricos realizados para la evaluación de la metodología de medición de tamaño funcional **aFPA**. En este capítulo se evalúa el procedimiento de medición de tamaño funcional y el procedimiento de verificación de exactitud de los resultados de medición. La evaluación del procedimiento de evaluación para el aprovechamiento de los resultados de medición mediante la creación de modelos de estimación de esfuerzo se presenta en el Capítulo 7. La metodología **aFPA** es evaluada mediante un conjunto de estudios empíricos que estudian la efectividad del procedimiento de medición, las herramientas prototipo de soporte a la medición y las extensiones propuestas para su aplicación utilizando los artefactos de la especificación de requerimientos y de los modelos conceptuales en un ambiente de desarrollo dirigido por modelos.

En las siguientes secciones se describe el diseño y el análisis de los resultados de cada uno de los estudios empíricos que presentan la evidencia generada sobre la evaluación de la efectividad de la metodología **aFPA** y sus procedimientos de medición y verificación. La evaluación empírica se realiza mediante múltiples casos de estudio y estudios cuasi experimentales orientados a personas. Finalmente, la evaluación de los modelos de estimación de esfuerzo presentado en el Capítulo 7 se realiza mediante experimentos controlados. La Tabla 6-1 presenta una matriz con el mapeo de cada una de las evaluaciones empíricas realizadas y las partes de la metodología **aFPA** que evalúan. Los objetivos específicos de

cada uno de los estudios empíricos se detallan en las secciones subsecuentes. El procedimiento de medición de tamaño funcional así como las extensiones propuestas, el procedimiento de verificación y las herramientas de soporte son evaluadas en los casos de estudio CS-01, CS-02, CS-03 y CS-04. Del mismo modo, el procedimiento de medición, el procedimiento de verificación y la herramienta de soporte, así como la extensión propuesta para requerimientos es evaluado con profesionales en los estudios empíricos E2-14, E1-15, E2-15 y E1-18. Finalmente, la estrategia de aprovechamiento de los resultados de medición del tamaño funcional para la construcción de los modelos de estimación de esfuerzo y su herramienta de soporte es evaluada en los estudios experimentales RE-01, CE-02 y CE-03 presentados en el Capítulo 7.

Tabla 6-1 Estudios empíricos realizados para evaluar la metodología **aFPA**.

Sec- ción	Propuestas asociadas al procedimiento de medición	Estudios empíricos											
		Sección 6.1				Sección 6.2				Capítulo 7			
		CS-01	CS-02	CS-03	CS-04	E2-14	E1-15	E2-15	E1-18	RE-01	CE-02	CE-03	
5.1	Procedimiento de medición de tamaño funcional.	×	×	×	×			×	×				
5.2	Procedimiento de verificación de la exactitud.	×	×	×	×	×	×	×	×				
5.3	Procedimiento de evaluación de los resultados de medición.									×	×	×	
5.4.1	Módulo para la medición del tamaño funcional.	×	×		×								
5.4.2	Módulo para la verificación de exactitud.					×	×	×					
5.4.3	Módulo para la evaluación de los modelos de estimación.									×	×	×	
5.5.1	Extensión para la aplicación con requerimientos.			×					×				
5.5.2	Extensión para la aplicación con modelos conceptuales.				×								

6.1. EVALUACIÓN DE LA EFECTIVIDAD DEL PROCEDIMIENTO DE MEDICIÓN DEL TAMAÑO FUNCIONAL: UN CONJUNTO DE CASOS DE ESTUDIO

Estos estudios representan el producto obtenido del ciclo de investigación 8 (RC8).

Esta sección presenta un conjunto de casos de estudio para la evaluación del procedimiento de medición del tamaño funcional. Los casos de estudio investigan la efectividad del procedimiento para obtener el tamaño funcional basado en los métodos IFPUG FPA y

COSMIC FFP. El detalle del diseño y el proceso de aplicación se detallan en la Sección 5.1. El prototipo de soporte al proceso de medición utilizado se describe en la Sección 5.4.1.

El objetivo de los casos de estudio es determinar cómo los resultados de medición del tamaño funcional obtenidos a partir de la aplicación del procedimiento de medición se comparan con los obtenidos por mediciones expertas tomadas como valores de control. Para la comparación de las mediciones se utiliza el protocolo de verificación descrito en la Sección 5.2. La estructura de cada uno de los reportes de los estudios se realiza de acuerdo a los lineamientos del protocolo de casos de estudio descrito en (Runeson & Höst, 2009; Runeson, Host, Rainer & Regnell, 2012; Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012). Para cada caso de estudio se selecciona una aplicación específica para realizar la evaluación empírica en términos de la exactitud de las mediciones bajo un contexto de operación particular.

En el primer caso de estudio se aplica el procedimiento para obtener los puntos de función sin ajustar IFPUG FPA utilizando como entrada los artefactos del código fuente de una aplicación desarrollada en Asp.Net C# bajo la arquitectura MVC (CS-01). Este estudio presenta el ejemplo detallado del proceso de medición utilizando el prototipo de soporte. El segundo caso de estudio realiza el proceso del primer caso de estudio para obtener los puntos de función COSMIC FFP con una aplicación distinta (CS-02). El tercer caso de estudio lleva a cabo el conteo del tamaño funcional utilizando la extensión del procedimiento para una especificación de requerimientos bajo el estándar IEEE Std. 830-1998 (CS-03). En este caso se obtienen los puntos de función sin ajustar IFPUG FPA. Finalmente, el cuarto caso de estudio realiza el conteo del tamaño funcional utilizando el prototipo de soporte en el nivel 2 de automatización para obtener los puntos de función sin ajustar (UFP) utilizando modelos conceptuales en el contexto del desarrollo dirigido por modelos (CS-04). En la Figura 6-1 se detalla la secuencia de los múltiples casos de estudio realizados. En las siguientes secciones se presenta cada uno de los casos de estudio.

(CS-01) IFPUG FPA Código fuente	(CS-02) COSMIC FFP Código fuente	(CS-03) IFPUG FPA Requerimientos IEEE 830	(CS-04) IFPUG FPA Modelos conceptuales
---------------------------------------	--	---	--

Figura 6-1 Secuencia de casos de estudio realizados.

6.1.1. CASO DE ESTUDIO 01: SISTEMA DE UNIVERSIDAD CONTOSO

En este caso de estudio se lleva a cabo el conteo del tamaño funcional de una aplicación transaccional utilizando el prototipo de soporte del procedimiento de medición y se comparan los resultados obtenidos con los valores de control reportados por un experto. En este caso, se aplica el procedimiento de medición de la metodología **aFPA** para obtener los puntos de función sin ajustar IFPUG FPA utilizando como entrada del proceso de medición los artefactos del código fuente de una aplicación desarrollada en *Asp.Net C#* bajo la arquitectura *MVC*. A continuación se describe el diseño del estudio, se detalla el ejemplo de medición utilizando el prototipo y se analizan los resultados.

6.1.1.1. DESCRIPCIÓN DEL ESTUDIO

El objetivo del estudio es determinar cuál es la efectividad de las mediciones obtenidas por el procedimiento de medición. La efectividad del procedimiento de medición se evalúa en términos de la exactitud en comparación con los valores de control. A partir de las mediciones recolectadas se realizan los análisis y se comparan los resultados. Con base en el objetivo de la investigación, se plantean las siguientes preguntas de investigación:

RQ1. *¿Cuál es la exactitud de la medición del tamaño en puntos de función sin ajustar obtenido por el procedimiento a partir del código fuente en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

RQ2. *¿Cuál es la exactitud de la medición del tamaño de los componentes funcionales básicos obtenido por el procedimiento a partir del código fuente en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

6.1.1.1.1. OBJETOS DE ESTUDIO E INSTRUMENTOS

El detalle de los objetos de estudio y los instrumentos utilizados en este caso de estudio es el siguiente:

- **Objetos experimentales:** los objetos utilizados son los artefactos de software de una aplicación transaccional: Universidad Contoso (Quesada-López, Madrigal & Jenkins, 2016). Esta aplicación tiene un tamaño de 144 UFP lo que la clasifica como una aplicación pequeña (S). El artefacto de software utilizado en el proceso de medición del experto es la especificación de requerimientos bajo el estándar IEEE 830. En el caso de la aplicación del procedimiento de medición de la metodología **aFPA** se utilizan los artefactos del código fuente desarrollado en *Asp.Net C#* bajo la arquitectura *MVC*. La aplicación transaccional medida implementa funcionalidades con las operaciones básicas de creación, lectura, actualización, borrado y listado de información (CRUDL).
- **Instrumentos de recolección de datos:** para cada una de las corridas se recolectan los resultados en los formularios de recolección de datos provistos.

6.1.1.1.2. PROCEDIMIENTO DE RECOLECCIÓN DE DATOS

El proceso realizado para la medición del tamaño funcional y la recolección de los datos contempla las siguientes actividades:

- **Se preparan y validan los objetos de estudio:** en este caso se utiliza la especificación de requerimientos para la medición con el método IFPUG FPA y el código fuente de la

aplicación para la medición con el soporte del prototipo del procedimiento de medición de la metodología **aFPA**.

- Se obtiene la medición del tamaño funcional IFPUG FPA: el experto realiza el proceso de conteo y de validación de las mediciones del tamaño funcional de la aplicación y las reporta en el formulario de recolección de datos. Estas mediciones son utilizadas como valores de control en la verificación de resultados que determina la exactitud de los resultados reportados por el procedimiento. Los expertos que miden la aplicación son el director y el autor de esta tesis. Los resultados del conteo son obtenidos por separado y validados para consensuar el tamaño final. Asimismo, el modelo de requerimientos de la aplicación es medida con una herramienta de conteo que valida los resultados.
- Se obtiene la medición del tamaño funcional mediante la aplicación del procedimiento de medición de medición de la metodología **aFPA**: se realiza el proceso de aplicación del proceso de medición. Este incluye los pasos de adquisición de los artefactos de software, la construcción del modelo funcional y la aplicación de las reglas de mapeo y de asignación numérica. El detalle del proceso de medición se ejemplifica en la Sección 6.1.1.2. El proceso de medición se aplica en el nivel 3 de automatización del procedimiento de medición, que se describe en la Figura 5-9.
- Se aplica el protocolo de verificación de exactitud: se aplica manualmente el proceso del procedimiento de verificación de exactitud de las mediciones obtenidas por el procedimiento de medición y se comparan contra las mediciones obtenidas por el experto para determinar la efectividad.

6.1.1.1.3. PROCEDIMIENTO DE ANÁLISIS

La variable medida es la relacionada con la efectividad. Para esto se recolectan las mediciones del tamaño funcional y se calcula la exactitud. Las siguientes métricas son recolectadas:

- Tamaño funcional (cuantitativa, escala intervalo): el tamaño funcional obtenido al aplicar el proceso de medición al objeto. Este recolecta los puntos de función sin ajustar (UFP) y el detalle de los puntos de función para cada uno de los componentes funcionales básicos (BFC). Esta variable es recolectada durante la medición del experto y durante la aplicación del procedimiento de medición.
- Exactitud: (cuantitativa, escala razón): la proximidad de los resultados entre un valor medido y el valor verdadero (utilizado como medida de control). Para evaluar la exactitud se utiliza la métrica MRE que se calcula como la diferencia de los valores absolutos entre el conteo producido por el procedimiento y el valor verdadero, relativo al valor verdadero. Esta métrica se calcula a partir del tamaño funcional reportado por el procedimiento de medición de la metodología **aFPA** y el tamaño funcional de control. La ecuación de la métrica de exactitud se detalla en la Sección 5.2.4. Esta se basa en las recomendaciones presentadas en (Abrahamo, 2004; Marín, Condori-Fernández & Pastor, 2008).

6.1.1.2. APLICACIÓN DEL PROCESO DE MEDICIÓN

El prototipo descrito en la Sección 5.4.1 soporta el proceso de automatización de la medición del tamaño funcional en el ambiente de operación de la plataforma de desarrollo *Visual Studio .Net* y el *API Framework .Net*. Para ejemplificar el proceso de aplicación del procedimiento **aFPA** se realiza la medición del tamaño funcional de la aplicación Universidad Contoso. En las siguientes secciones se muestra cada uno de los pasos llevados a cabo para obtener los resultados de medición. Esta sección detalla cómo se utiliza cada uno de los componentes y las librerías del módulo de medición. En este caso, el procedimiento se aplica en el nivel 3 de automatización, tal como se describe en la Figura 5-9. Para realizar el proceso de medición, la aplicación transaccional es instalada y ejecutada en un ambiente controlado de operación, el código fuente es instrumentado y las librerías de medición son incluidas y configuradas. En este nivel de automatización, todas las librerías del procedimiento de medición son agregadas al proyecto de desarrollo de la aplicación medida. A

continuación se describen los pasos llevados a cabo para obtener el tamaño funcional utilizando el procedimiento **aFPA** y el prototipo de soporte a la medición.

6.1.1.2.1. ADQUISICIÓN DE LOS ARTEFACTOS DE SOFTWARE

En el paso de adquisición de los artefactos de software, todos los artefactos del código fuente son recolectados y utilizados como entradas del procedimiento de medición. En esta etapa se realiza la configuración de la aplicación medida en el ambiente de operación controlado. Esto incluye la configuración de las librerías para el proceso de instrumentación del código fuente, la identificación de los componentes de interfaz de usuario y la generación de los rastros de las transacciones de datos. En este primer paso se configura el componente *Generador* del módulo de soporte para la medición. A continuación se detallan aspectos técnicos de la aplicación Universidad Contoso y los artefactos recolectados en esta etapa:

- La aplicación Universidad Contoso se encuentra desarrollada en *Asp.Net Web Application* en el lenguaje C# bajo el modelo MVC.
- El entorno de desarrollo integrado (IDE) que se utiliza en el proceso de medición es Microsoft Visual Studio Professional 2013.
- El servidor de bases de datos utilizado es Microsoft SQL Server 2012.
- Los artefactos del código fuente son los de la aplicación *Asp.Net Web Application*. Los artefactos incluyen lo necesario para la ejecución de la aplicación que contiene los archivos de configuración de la aplicación, la configuración de la conexión con la base de datos, los archivos de las vistas, los controladores y los modelos que implementan las funcionalidades (*cshhtml*, *cs.*, *css*), entre otros.

Las configuraciones de las librerías en esta etapa es la siguiente:

- Configuración de la librería de contexto de la base de datos. Esta clase coordina la funcionalidad del *Entity Framework* para un modelo de datos. Esta clase se desarrolla personalizando la clase *System.Data.Entity.DbContext*. Esto implica especificar las entidades del modelo de datos de la aplicación. La clase configurada en la aplicación es: *SchoolContext.cs*.
- Configuración de las librerías de registro de ejecución del código. Estas clases permiten rastrear la ejecución del código fuente y las llamadas a servicios externos como las de las bases de datos SQL. En este caso, se proveen funcionalidades provistas por las clases *System.Diagnostics.Trace* para mostrar los mensajes de los trazos de la ejecución de las funcionalidades y las transacciones de datos mediante el acceso a los servicios tales como las bases de datos SQL. La clase configurada en la aplicación es: *ILogger.cs* y *Logger.cs*.
- Configuración de las librerías de intercepción de transacciones. Las clases que interceptan las transacciones se implementan utilizando la clase *System.Data.Entity.Infrastructure.Interception.DbCommandInterceptor* y permiten obtener los rastros de los accesos a la base de datos mediante la comunicación con la librería del contexto de la bases de datos. La clase configurada en la aplicación es: *SchoolInterceptorLogging.cs*.
- Configuración de las librerías de intercepción de eventos. En este estudio se implementa una librería particular que permite realizar el proceso de instrumentación del código fuente para rastrear los eventos asociados a las transacciones. Esta permite, para las clases controladores relacionadas con la interfaz del usuario, agregar el código necesario para el rastreo de la ejecución en los constructores y los métodos públicos. Esto incluye el llamado a la clase de registro de ejecución del código, el uso de la clase *System.Reflection* para obtener los componentes del espacio de nombres y la clase *System.Diagnostics* que permite acceder la pila de llamadas entre clases.

Es importante indicar que la incorporación de la mayoría de las librerías descritas anteriormente, forman parte del proceso de desarrollo común en este tipo de aplicaciones transaccionales en un ambiente de operación *Asp.Net Web Application*.

La Figura 6-2 lista las librerías configuradas en la aplicación. La Figura 6-3 muestra un extracto del código instrumentado para las clases controladores. La Figura 6-4 muestra un extracto de un trazo de ejecución de la aplicación que detalla los componentes de eventos y las transacciones realizadas. A partir de los trazos de ejecución se construye el archivo de interfaz XML que detalla toda la información requerida para la construcción del modelo funcional. En este estudio se implementa una librería particular que permite realizar el proceso de análisis (*parsing*) del archivo de trazos para construir la interfaz XML. El proceso de análisis procesa el archivo de trazos para extraer los componentes de interfaz de usuario y las transacciones asociadas. Para esto, identifica los rastros marcados como “*Component:aFPA Generator --->>> UI*” y “*Component:aFPA Generator --->>> DB*” y para cada una de ellos, las etiquetas *<event>* que contienen la información del componente de interfaz de usuario, los eventos ejecutados y las etiquetas *<function>* que contienen la información de las transacciones asociadas a los eventos. La Figura 6-5 muestra un extracto del archivo de interfaz XML generado a partir de los trazos de ejecución de la aplicación. Este archivo de interfaz XML es la entrada del componente *Constructor* del módulo de soporte para la medición.

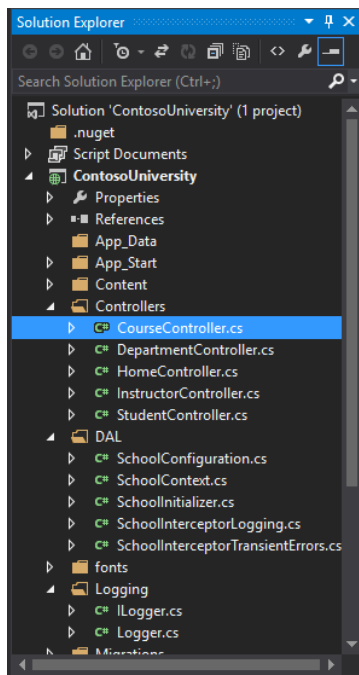


Figura 6-2 Lista de librerías configuradas en la aplicación.

```

Package Manager Console Error List Output Find Results 1 CourseController.cs x Logger.cs
ContosoUniversity.Controllers.CourseController Details(int? id)
/* aFPA Instrumentacion */
using ContosoUniversity.Logging;
using System.Diagnostics;
using System.Reflection;

namespace ContosoUniversity.Controllers
{
    public class CourseController : Controller
    {
        private SchoolContext db = new SchoolContext();

        /* aFPA Instrumentacion */
        private ILogger _logger = new Logger();

        public CourseController()
        {
            /* aFPA Instrumentacion */
            _logger.TraceUI("aFPA Generator ---->>> UI <event>",
                "<user_interface_name>" + this.GetType().Name + "</user_interface_name>",
                "<control_name>" + new StackTrace(true).GetFrame(1).GetMethod().Name + "</control_name>",
                "<event_name>" + MethodBase.GetCurrentMethod().Name + "</event_name>",
                "<file_name>" + new StackTrace(true).GetFrame(0).GetFileName() + "</file_name>",
                "<description>Constructor: Entrada a la clase</description><funtions>");
        }

        // GET: Course/Details/5
        public ActionResult Details(int? id)
        {
            /* aFPA Instrumentacion */
            _logger.TraceUI("aFPA Generator ---->>> UI ",
                "<user_interface_name>" + this.GetType().Name + "</user_interface_name>",
                "<control_name>" + new StackTrace(true).GetFrame(1).GetMethod().Name + "</control_name>",
                "<event_name>" + MethodBase.GetCurrentMethod().Name + "</event_name>",
                "<file_name>" + new StackTrace(true).GetFrame(0).GetFileName() + "</file_name>",
                "<description>Evento ejecutado</description><funtions>");

            if (id == null)
            {
                return new HttpStatusCodeResult(HttpStatusCode.BadRequest);
            }
        }
    }
}

```

Figura 6-3 Ejemplo del código instrumentado.

```

Output
Show output from: Debug
iisexpress.exe Information: 0 : Component: aFPA Generator --->> UI <event><user_interface_name>CourseController</
user_interface_name><control_name>CreateInstanceDefaultctor</control_name><event_name>.ctor</event_name><file_name>e:\Desarrollo
\Net\ContosoUniversity_v2\C#\ContosoUniversity\Controllers\CourseController.cs</file_name><description>Constructor: Entrada a la
Clase</description><funtions>
iisexpress.exe Information: 0 : Component: aFPA Generator --->> UI <user_interface_name>CourseController</
user_interface_name><control_name>lambda_method</control_name><event_name>Details</event_name><file_name>e:\Desarrollo\Net
\ContosoUniversity_v2\C#\ContosoUniversity\Controllers\CourseController.cs</file_name><description>Evento ejecutado</
description><funtions>
iisexpress.exe Information: 0 : Component: aFPA Generator --->> DB ReaderExecuted Timespan:00:00:00.0011534<funtion><sql>SELECT TOP
(2)
[Extent1].[CourseID] AS [CourseID],
[Extent1].[Title] AS [Title],
[Extent1].[Credits] AS [Credits],
[Extent1].[DepartmentID] AS [DepartmentID]
FROM [dbo].[Course] AS [Extent1]
WHERE [Extent1].[CourseID] = @p0</sql></funtion>
iisexpress.exe Information: 0 : Component: aFPA Generator --->> DB ReaderExecuted Timespan:00:00:00.0009938<funtion><sql>SELECT
[Extent1].[DepartmentID] AS [DepartmentID],
[Extent1].[Name] AS [Name],
[Extent1].[Budget] AS [Budget],
[Extent1].[StartDate] AS [StartDate],
[Extent1].[InstructorID] AS [InstructorID],
[Extent1].[RowVersion] AS [RowVersion]
FROM [dbo].[Department] AS [Extent1]
WHERE [Extent1].[DepartmentID] = @EntityKeyValue1</sql></funtion>
Error List Output Locals Watch 1 Call Stack Immediate Window

```

Figura 6-4 Ejemplo de trazo de la ejecución de la aplicación.

```

aFPA_Requirements...tosoUniversity.xml
<?xml version="1.0"?>
<event_model>
  <application_id>ContosoUniversity</application_id>
  <application_name>ContosoUniversity</application_name>
  <events>
    <event>
      <user_interface_name>CourseController</user_interface_name>
      <control_name>lambda_method</control_name>
      <event_name>Details</event_name>
      <file_name>e:\Desarrollo\Net\ContosoUniversity_v2\C#\ContosoUniversity\Controllers\CourseController.cs</
      <position></position>
      <requirement></requirement>
      <description>Evento ejecutado</description>
      <funtions>
        <funtion>
          <sql>
            SELECT TOP (2)
            [Extent1].[CourseID] AS [CourseID],
            [Extent1].[Title] AS [Title],
            [Extent1].[Credits] AS [Credits],
            [Extent1].[DepartmentID] AS [DepartmentID]
            FROM [dbo].[Course] AS [Extent1]
            WHERE [Extent1].[CourseID] = = @p0
          </sql>
        </funtion>
        <funtion>
          <sql>
            SELECT
            [Extent1].[DepartmentID] AS [DepartmentID],
            [Extent1].[Name] AS [Name],
            [Extent1].[Budget] AS [Budget],
            [Extent1].[StartDate] AS [StartDate],
            [Extent1].[InstructorID] AS [InstructorID],
            [Extent1].[RowVersion] AS [RowVersion]
            FROM [dbo].[Department] AS [Extent1]
            WHERE [Extent1].[DepartmentID] = @EntityKeyValue1
          </sql>
        </funtion>
      </funtions>
    </event>
  </events>
</event_model>
100%

```

Figura 6-5 Ejemplo de archivo de interfaz XML.

6.1.1.2.2. CONSTRUCCIÓN DEL MODELO FUNCIONAL

En el paso de la construcción del modelo funcional, el componente *Constructor* del módulo de soporte para la medición procesa el archivo de interfaz XML entregado por el

componente *Generador*. Para realizar la construcción del modelo, el componente *Constructor* debe estar configurado con los accesos a las bases de datos de la aplicación. El componente provee un conjunto de librerías que permiten la simulación de la ejecución de las transacciones que genera una bitácora de flujos de datos que se utilizan para la construcción del grafo. En el proceso de construcción, se proveen librerías para el análisis del archivo XML que permite construir el modelo de eventos, para la conexión y análisis de las bases de datos que permiten obtener todos los objetos de datos y sus atributos para construir el modelo de datos. Finalmente, se proveen las librerías que se encargan de asociar los elementos del modelo de eventos a los elementos del modelo de datos mediante la identificación de los trazos de ejecución. En el análisis del archivo XML, las funciones son pre procesadas para eliminar elementos que no son necesarios en la construcción del modelo funcional. Por ejemplo, se eliminan parámetros y cláusulas de orden. Asimismo, pueden presentarse duplicidades de funcionalidades que el proceso de construcción del modelo intenta identificar para notificar al profesional. En este caso de estudio, los modelos fueron revisados y depurados para disminuir la funcionalidad duplicada en el modelo utilizado para realizar la medición. Los algoritmos que construyen el modelo funcional implementan el proceso detallado en la Sección 5.1.

El procesamiento del archivo XML se realiza utilizando las librerías *System XML*, la simulación de la ejecución de las transacciones con el rastreador de transacciones se realiza utilizando las librerías *Microsoft ScriptDom* y la interfaz de visualización del modelo funcional mediante las librerías del *Neo4j API*, *Neo4j for .Net Developers* y el lenguaje *Neo4j Graph Query Language Cypher*. Con la construcción del modelo funcional, el grafo es explorado para generar los nodos y aristas que permiten la visualización del modelo y que son las entradas para el componente *Medidor* junto con los trazos de ejecución. La Figura 6-6 muestra un extracto de procesamiento realizado al archivo XML y de las salidas de la simulación de la ejecución de las transacciones. La Figura 6-7 muestra un extracto del modelo generado en lenguaje *Cypher*. Finalmente, la Figura 6-8 muestra la visualización del modelo funcional el cual es medido por el componente *Medidor* para determinar su tamaño funcional. Las librerías construidas para el componente *Generador* se listan en la Figura 6-

9 y se explican en la siguiente sección. El componente *Generador* y el componente *Medidor* se encuentran implementados en el mismo paquete como una aplicación de apoyo a la medición.

```

Output
Show output from: Debug
Application: CONTOSO CONTOSO
Node: CourseController e:\Desarrollo\Net\ContosoUniversity_v2\C#\ContosoUniversity\Controllers\CourseController.cs lambda_method Details_lambda_method
sql: [
    SELECT TOP (2)
    [Extent1].[CourseID] AS [CourseID],
    [Extent1].[Title] AS [Title],
    [Extent1].[Credits] AS [Credits],
    [Extent1].[DepartmentID] AS [DepartmentID]
    FROM [dbo].[Course] AS [Extent1]
    WHERE [Extent1].[CourseID] = ''
]
SQL Parser: Microsoft.SqlServer.TransactSql.ScriptDom.SelectStatement
Token: TokenType [WhiteSpace] Text [
]
Token: TokenType [WhiteSpace] Text [
]
Token: TokenType [Select] Text [SELECT]
Token: TokenType [WhiteSpace] Text [ ]
Token: TokenType [Top] Text [TOP]
Token: TokenType [WhiteSpace] Text [ ]
Token: TokenType [LeftParenthesis] Text [(]
Token: TokenType [Integer] Text [2]
Token: TokenType [RightParenthesis] Text [)]
Token: TokenType [WhiteSpace] Text [
]
-----> Imprimir Grafo <-----
[Node]: NodeType: APPLICATION Element: 66337667 Name: CONTOSO Count: 1 Description: [ApplicationModel
id: CONTOSO
name: CONTOSO]
--> [Edge]: EdgeType: FATHER [ [Node]: NodeType: USER_INTERFACE Element: 35489797 Name: CourseController Count: 1 ]
[Node]: NodeType: USER_INTERFACE Element: 35489797 Name: CourseController Count: 1 Description: [Control
id: CourseController
name: CourseController]
--> [Edge]: EdgeType: FATHER [ [Node]: NodeType: CONTROL Element: 45795543 Name: lambda_method Count: 2 ]
[Node]: NodeType: CONTROL Element: 45795543 Name: lambda_method Count: 2 Description: [Control
id: lambda_method
name: lambda_method]

```

Figura 6-6 Ejemplo de salidas de la simulación de la ejecución.

```

Output
Show output from: Debug
-----> Script Graph Neo4j <-----
CREATE (h66337667:APPLICATION {name:'CONTOSO', hash:66337667, id:'CONTOSO', AppName:'CONTOSO'})
CREATE (h35489797:USER_INTERFACE {name:'CourseController', hash:35489797, id:'CourseController', uiName:'CourseController', fileName:'e:\Desarrollo\Net\ContosoUniversity_v2\C#\ContosoUniversity\Controllers\CourseController.cs'})
CREATE (h45795543:CONTROL {name:'lambda_method', hash:45795543, id:'lambda_method', controlName:'lambda_method'})
CREATE (h65677972:EVENT {name:'Details', hash:65677972, id:'Details', eventName:'Details', position:'', requirement:'', description:'Evento ejecutado Details'})
CREATE (h10589941:FUNCTION {name:'UCRQLAFP_TransactionalFuntion_1_1_1_Select', hash:10589941, id:'Select'})
CREATE (h31534420:FUNCTION {name:'UCRQLAFP_TransactionalFuntion_1_1_2_Select', hash:31534420, id:'Select'})
CREATE (h8442299:EVENT {name:'Delete', hash:8442299, id:'Delete', eventName:'Delete', position:'', requirement:'', description:'Evento ejecutado Delete'})

CREATE (h35489797)-[:FATHER]->(h66337667)
CREATE (h45795543)-[:FATHER]->(h35489797)
CREATE (h65677972)-[:FATHER]->(h45795543)
CREATE (h22583577)-[:FATHER]->(h53683562)
CREATE (h53511455)-[:FATHER]->(h53683562)

```

Figura 6-7 Ejemplo del modelo funcional en *Cypher*.

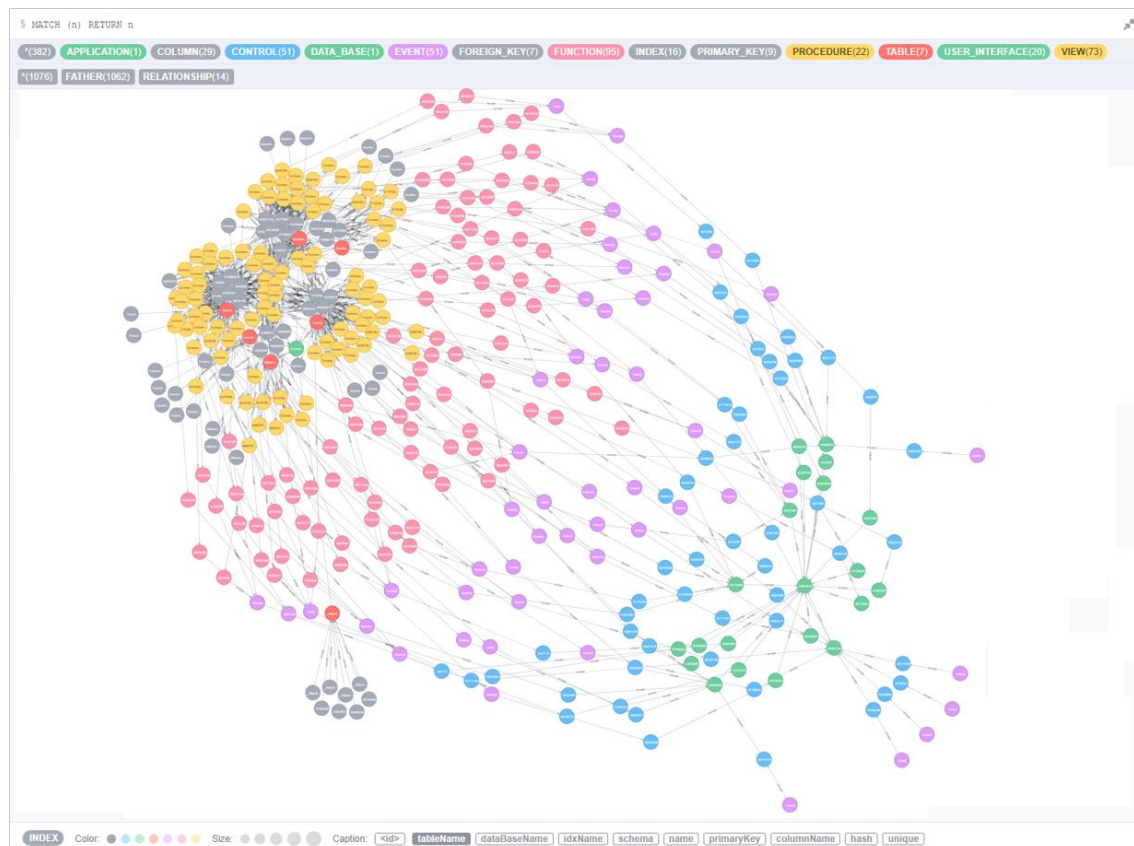


Figura 6-8 Visualización del modelo funcional.

6.1.1.2.3. APLICACIÓN DE LAS REGLAS DE MAPEO Y DE ASIGNACIÓN NUMÉRICA

En este paso, el componente *Medidor* del módulo de soporte para el conteo del tamaño funcional aplica las reglas de mapeo para asociar los elementos del modelo funcional del procedimiento **aFPA** a los elementos del método de medición IFPUG FPA. Asimismo, con los elementos identificados se aplican las reglas de asignación numérica para obtener el tamaño funcional en puntos de función sin ajustar (UFP). Por su parte, el componente *Reporteador* presenta los resultados de medición.

El componente *Medidor* procesa el modelo funcional entregado por el componente *Constructor*. Para realizar el análisis y la medición del tamaño funcional, el prototipo provee las librerías de medición que implementan una estructura de grafos y los algoritmos de mapeo y conteo. Los algoritmos implementan el proceso detallado en la Sección 5.1. En este

proceso, a partir del modelo funcional construido, se analizan cada uno de los trazos de ejecución que representan los flujos de datos y transacciones, se realiza el mapeo de componentes funcionales al método de medición IFPUG FPA y se asigna el valor del tamaño funcional.

La Figura 6-9 lista las librerías construidas para el componente *Constructor*, *Medidor* y *Reporteador*. Estas incluyen las clases que implementan cada uno de los elementos del modelo funcional: *ApplicationModel.cs*, *UserInterface.cs*, *Control.cs*, *Event.cs*, *Function.cs*, *View.cs*, *Procedure.cs*, *DataBase.cs*, *Table.cs*, *Column.cs*, *PrimaryKey.cs*, *ForeignKey.cs* e *Index.cs*. Las clases que implementan el grafo: *Graph.cs*, *Node.cs*, *Element.cs* y *Edge.cs*. La clase de análisis del archivo XML: *XMLParser.cs*. Las clases de apoyo para la simulación y análisis de transacciones: *SQLparserResult.cs* y *SQLParser.cs*. Las clases para administrar la interacción con las librerías del *Neo4j API*, *Neo4j for .Net Developers* y el lenguaje *Neo4j Graph Query Language Cypher: Neo4jScript.cs*. Las clases de configuración y de llamado desde línea de comandos: *Connection.cs* y *Program.cs*. Finalmente, la clase que implementa los algoritmos de construcción del modelo, mapeo de elementos y medición del tamaño funcional: *UCRCQLAFP.cs*. La Figura 6-10 muestra un extracto del reporte de resultados entregado por el componente *Reporteador* con las mediciones del tamaño funcional.

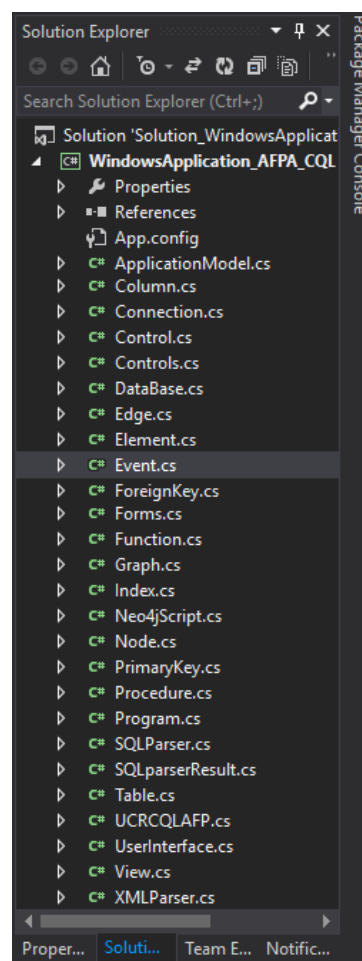


Figura 6-9 Librerías del componente *Constructor*, *Medidor* y *Reporteador*.

```

aFPA Reporter
-----
App: ContosoUniversity          Version: 3.0.0
DateTime: 13/05/2018 04:20:00 p.m.  Counted by: aFPA Procedure 2.1.0
Method: IFPUG FPA 4.3
-----
EVENT  FUNCTION  TABLE          COLUMN          FC  T  RW  FTR  RET  DET  BFC  Complex  UFP
-----
Node:  EVENT  Event: Details  UI: CourseController  Control: lambda_method
File:  e:\Desarrollo\Net\ContosoUniversity_v2\C#\ContosoUniversity\Controllers\CourseController.cs
Description: Evento ejecutado      Requirement: NR

Details VIEW  TABLE: Course  COLUMN: CourseID  (F)( )(R)      1      1  EQ  A      4
              COLUMN: Title    (F)( )( )
              COLUMN: Credits  (F)( )( )
              COLUMN: DepartmentID (F)(T)( )

              TABLE: Department  COLUMN: DepartmentID (F)(T)( )  1
              COLUMN: Name      (F)( )( )
              COLUMN: Budget    (F)( )( )
              COLUMN: StartDate (F)( )( )
              COLUMN: InstructorID (F)(T)( )
              COLUMN: RowVersion (F)(T)( )

              Button          1
              ErrorMessage    1
              ---
              2                8

Node:  EVENT  Event: Delete  UI: CourseController  Control: lambda_method
File:  e:\Desarrollo\Net\ContosoUniversity_v2\C#\ContosoUniversity\Controllers\CourseController.cs
Description: Evento ejecutado      Requirement: NR

```

Figura 6-10 Extracto de reporte de resultados del componente *Reporteador*.

6.1.1.3. ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

En esta sección se presenta la comparación de los resultados de la medición del tamaño funcional y se responden las preguntas de investigación. La Tabla 6-2 presenta el detalle de los resultados de medición en puntos de función IFPUG FPA. En esta se detallan los resultados del tamaño funcional (UFP) para cada uno de los procesos de medición (experto y procedimiento) y para cada uno de los componentes funcionales básicos (BFC). Asimismo, se presentan los resultados del tamaño para las funciones transaccionales (TF), las funciones de datos (DF), las entradas (EI), las salidas (EO), las consultas (EQ), los archivos internos (ILF) y las interfaces externas (EIF). Para todos los anteriores, se detalla la cantidad de componentes identificados.

La Tabla 6-3 presenta la exactitud obtenida a partir de los resultados de medición. Se presentan los porcentajes de las métricas MRE, MER, BRE e IBRE para cada uno de los

BFC y para el total de puntos de función sin ajustar (UFP). Las métricas se detallan para las TF y las DF. Asimismo, se detallan los porcentajes obtenidos para las EI, las EO, las EQ, los ILF y las EIF. Además, para cada uno de ellos se detalla la diferencia absoluta del tamaño funcional. A continuación se analiza cada una de las preguntas de investigación a partir de los resultados obtenidos en el proceso de medición.

6.1.1.3.1. EXACTITUD DE LOS RESULTADOS DE MEDICIÓN DEL TAMAÑO FUNCIONAL

RQ1. *¿Cuál es la exactitud de la medición del tamaño en puntos de función sin ajustar obtenido por el procedimiento a partir del código fuente en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

El conteo manual realizado por el experto sobre la especificación de requerimientos IEEE obtiene un total de 144 puntos de función sin ajustar (UFP). En el caso de la aplicación del proceso automatizado de medición se obtienen un total de 141 UFP. En este caso, los resultados obtienen una variación en la exactitud (MRE) del 2.1% para el total de puntos de UFP. El conteo genera una diferencia de 3 UFP.

Tabla 6-2 Resultados de medición del tamaño funcional.

Proceso	Var	TF				DF			
		EI	EO	EQ	Total	ILF	EIF	Total	Total
Experto	UFP	71	19	26	116	28	0	28	144
	Cantidad	(17)	(3)	(7)	(27)	(4)	(0)	(4)	(31)
aFPA	UFP	70	17	26	113	28	0	28	141
	Cantidad	(17)	(3)	(7)	(27)	(4)	(0)	(4)	(31)

Tabla 6-3 Exactitud de los resultados de medición del tamaño funcional.

	TF				DF			
	EI	EO	EQ	Total	ILF	EIF	Total	Total
MRE	1.4%	10.5%	0.0%	2.6%	0.0%	-	0.0%	2.1%
MER	1.4%	11.8%	0.0%	2.7%	0.0%	-	0.0%	2.1%
BRE	1.4%	11.8%	0.0%	2.7%	0.0%	-	0.0%	2.1%
IBRE	1.4%	10.5%	0.0%	2.6%	0.0%	-	0.0%	2.1%
Diff	1	2	0	3	0	0	0	3

Seguidamente, se realiza un análisis de los valores de puntos de función para cada uno de los requerimientos. En la Tabla 6-4 se presenta el conteo de ambos procesos de medición (experto y procedimiento de medición de la metodología **aFPA**) y para cada uno de los componentes funcionales básicos (BFC). Asimismo, se detalla la cantidad de los elementos de tipo dato (DET), los tipos de archivo referenciado (FTR), los elementos de tipo registro (RET) y finalmente, la complejidad (C). Para cada uno de los requerimientos se detalla la exactitud (MRE) y la diferencia entre el valor de control y el valor medido por el procedimiento (Dif). La aplicación medida cuenta con 31 requerimientos relacionados con las operaciones CRUDL.

Tabla 6-4 Resultados de medición del tamaño funcional por requerimiento.

RQ	aPPA					Experto					MRE	Dif
	BCF	FTR/RET	DET	C	UFP	BCF	FTR/RET	DET	C	UFP		
(R1.1)	EI	2	7	A	4	EI	2	8	A	4	0%	-
(R1.2)	EI	2	4	L	3	EI	2	8	A	4	25%	+1
(R1.3)	EI	2	5	A	4	EI	2	7	A	4	0%	-
(R1.4)	EQ	2	8	A	4	EQ	2	8	A	4	0%	-
(R1.5)	EQ	2	7	A	4	EQ	2	8	A	4	0%	-
(R1.6)	EI	2	6	A	4	EI	2	6	A	4	0%	-
(R1.7)	EQ	2	8	A	4	EQ	2	8	A	4	0%	-
(R1.8)	EO	4	14	H	7	EO	4	19	H	7	0%	-
(R2.1)	EI	2	8	A	4	EI	1	8	L	3	33%	-1
(R2.2)	EI	2	8	A	4	EI	1	8	L	3	33%	-1
(R2.3)	EI	3	5	H	6	EI	1	8	L	3	100%	-3
(R2.4)	EQ	3	12	A	4	EQ	2	8	A	4	0%	-
(R2.5)	EI	2	5	A	4	EI	2	5	A	4	0%	-
(R2.6)	EO	3	13	A	5	EO	3	14	A	5	0%	-
(R3.1)	EI	2	8	A	4	EI	2	7	A	4	0%	-
(R3.2)	EI	2	10	A	4	EI	2	8	A	4	0%	-
(R3.3)	EI	2	6	A	4	EI	3	7	H	6	33%	+2
(R3.4)	EQ	2	9	A	4	EQ	2	8	A	4	0%	-
(R3.5)	EI	2	5	A	4	EI	2	6	A	4	0%	-
(R4.1)	EI	1	7	L	3	EI	1	7	L	3	0%	-
(R4.2)	EI	1	6	L	3	EI	1	7	L	3	0%	-
(R4.3)	EI	1	5	L	3	EI	3	8	H	6	50%	+3
(R4.4)	EQ	1	6	L	3	EQ	1	7	L	3	0%	-
(R4.5)	EQ	1	6	L	3	EQ	1	7	L	3	0%	-
(R4.6)	EI	3	8	H	6	EI	3	5	H	6	0%	-
(R4.7)	EO	3	9	A	5	EO	3	17	H	7	29%	+2
(R4.8)	EI	3	6	H	6	EI	3	6	H	6	0%	-
(R1.9)	ILF	1	4	L	7	ILF	1	5	L	7	0%	-
(R3.9)	ILF	1	5	L	7	ILF	1	5	L	7	0%	-
(R2.9)	ILF	2	7	L	7	ILF	1	6	L	7	0%	-
-	ILF	1	2	L	7	-	-	-	-	-	-	-7
(R4.9)	-	-	-	-	-	ILF	1	6	L	7	100%	+7

En este caso, para realizar la comparación se analizan las funciones generadas en el reporte de medición del procedimiento sin considerar las posibles funciones duplicadas y las funciones adicionales. Para esto se realiza una validación contra los requerimientos que se encuentran en la especificación medida por el experto. Por ejemplo, en el proceso de validación se identifican casos en que la forma en que es desarrollada la aplicación agrega

funcionalidad tal como cuando ofrece filtros o listados adicionales que no se encuentren en la especificación de requerimientos funcionales. Después de filtrar las funcionalidades, se encuentra que las principales diferencias en el conteo se presentan por la identificación de campos técnicos y entidades de datos. Es importante notar que algunas de las variaciones en las funciones de datos se balancean con otras variaciones. Por lo tanto, se deben revisar las reglas de mapeo para reducir estas diferencias.

6.1.1.3.2. EXACTITUD DE LOS RESULTADOS DE MEDICIÓN DE LOS COMPONENTES FUNCIONALES BÁSICOS

RQ2. *¿Cuál es la exactitud de la medición del tamaño de los componentes funcionales básicos obtenido por el procedimiento a partir del código fuente en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

Al realizar el análisis para cada uno de los componentes funcionales básicos (BFC) se encuentran diferencias entre las mediciones reportadas por el experto y las obtenidas por el procedimiento de medición de la metodología **aFPA**. En total se cuentan 113 puntos de función transaccionales y 28 puntos de función de datos. De las 27 funciones transaccionales, 17 son entradas (EI), 3 son salidas (EO) y 7 son consultas (EQ). Cuatro funciones de datos son clasificadas como archivos internos (ILF) y no se reportan interfaces externas (EIF). En términos de exactitud, los resultados de la métrica MRE alcanzan un 2.6% para las TF y un 0% para las DF. En este caso, las EI obtienen un resultado del 1.4% y las EQ del 0%; sin embargo, las EO presentan una variabilidad del 10.5%. La identificación de las salidas externas (EO) presenta diferencias en los resultados esperados dados los elementos mapeados para las funciones. Sin embargo, los resultados se consideran aceptables para un sistema de medición. Al analizar los requerimientos particulares se encuentran algunas interpretaciones que se realizaron en la implementación que variaron los resultados con respecto a la especificación de requerimientos.

6.1.1.4. RESUMEN DE RESULTADOS

El prototipo del procedimiento de medición presenta resultados de exactitud aceptables. La métrica MRE obtiene un 2.1% para el total de puntos de función sin ajustar (UFP), un 2.6% para las funciones transaccionales (TF) y un 0% para las funciones de datos (DF). Los análisis de la exactitud para cada uno de los componentes funcionales básicos (BFC), muestra que las entradas (EI), las consultas (EQ) y las salidas (EO) obtienen igualmente resultados aceptables. En el caso de los EO se presentan algunas desviaciones que deben ser mejoradas. El procedimiento identifica funciones que reporta como duplicadas. Al analizar el código fuente de la aplicación que implementa los requerimientos funcionales, se encuentran trazos de ejecución duplicados que requieren ser analizados y validados. Estas funciones son validadas para seleccionar el conjunto de funciones que deben formar parte del alcance del conteo. El procedimiento reporta los casos al profesional para que este tome las respectivas decisiones del alcance de la medición. Finalmente, la toma de decisiones automáticas sobre los flujos de datos duplicados requiere más análisis y se contempla como parte del trabajo futuro de esta propuesta.

6.1.2. CASO DE ESTUDIO 02: SISTEMA DE REGISTRO C-REG

En este caso de estudio se lleva a cabo el conteo del tamaño funcional de una aplicación transaccional utilizando el prototipo del procedimiento de medición de la metodología **aFPA**. Asimismo, se comparan los resultados de medición contra los valores de control reportados por el conteo experto. En este caso, se aplica el procedimiento de medición para obtener los puntos de función COSMIC FFP utilizando como entrada los artefactos del có-

digo fuente de una aplicación desarrollada en *Asp.Net C#* bajo la arquitectura *MVC*. El proceso realizado es similar al descrito en la Sección 6.1.1. A continuación se describe el diseño del estudio y se analizan los resultados.

6.1.2.1. DESCRIPCIÓN DEL ESTUDIO

El objetivo del estudio es determinar cuál es la efectividad de las mediciones obtenidas por el procedimiento de medición de la metodología **aFPA** basado en el método COSMIC FFP. La efectividad del procedimiento de medición se evalúa en términos de la exactitud. Para esto se realizan los análisis a partir de las mediciones recolectadas y se comparan los resultados contra los reportados por un experto. Basado en el objetivo de la investigación se definen las siguientes preguntas de investigación:

- RQ1. *¿Cuál es la exactitud de la medición del tamaño en puntos de función obtenido por el procedimiento a partir del código fuente en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*
- RQ2. *¿Cuál es la exactitud de la medición del tamaño de los componentes funcionales obtenido por el procedimiento a partir del código fuente en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

6.1.2.1.1. OBJETOS DE ESTUDIO E INSTRUMENTOS

El detalle de los objetos de estudio y los instrumentos utilizados en el caso de estudio es el siguiente:

- **Objetos experimentales:** los objetos utilizados son los artefactos de software de una aplicación transaccional: Registro (Bundschuh & Dekkers, 2008; COSMIC, 2015). Esta aplicación tiene un tamaño de 118 CFP lo que la clasifica como una aplicación pequeña

(S). El artefacto de software utilizado en el proceso de medición del experto es la especificación de requerimientos bajo el estándar IEEE 830. En el caso de la aplicación del procedimiento de medición de la metodología **aFPA** se utilizan los artefactos del código fuente del sistema desarrollado en *Asp.Net C#* bajo la arquitectura *MVC*. La aplicación transaccional implementa funcionalidades con las operaciones básicas de creación, lectura, actualización, borrado y listado de información (CRUDL). Durante la implementación de la aplicación, en el caso en que las funcionalidades requieran realizar llamadas a interfaces externas, estas son implementadas mediante *mocking*.

- Instrumentos de recolección de datos: para cada una de las corridas se recolectan los resultados en los formularios de recolección de datos previamente definidos.

6.1.2.1.2. PROCEDIMIENTO DE RECOLECCIÓN DE DATOS

El proceso realizado para la medición de la aplicación y la recolección de los datos contempla las siguientes actividades:

- Se preparan y validan los objetos de estudio: en este caso se utiliza la especificación de requerimientos para la medición del método COSMIC FFP realizada por el experto y el código fuente de la aplicación para la medición con el soporte del prototipo del procedimiento de medición de la metodología **aFPA**.
- Se obtiene la medición del tamaño funcional COSMIC FFP: el experto valida el conteo del tamaño funcional de la aplicación y reporta los detalles en el formulario de recolección de datos. Estas mediciones son utilizadas como valores de control en la verificación de resultados que determina la exactitud de los resultados del procedimiento. Los valores de medición son los reportados en (COSMIC, 2015). Los expertos que adaptan y validan el conteo de la aplicación son el director y el autor de esta tesis.
- Se obtiene la medición del tamaño funcional mediante la aplicación del procedimiento de medición de la metodología **aFPA**: se realiza el proceso de aplicación del proceso de

medición. Este incluye los pasos de adquisición de los artefactos de software, la construcción del modelo funcional y la aplicación de las reglas de mapeo y de asignación numérica. Este proceso es el que se ejemplifica en la Sección 6.1.1.2. El proceso de medición aplica el nivel 3 de automatización del procedimiento de medición descrito en la Figura 5-9.

- Se aplica el protocolo de verificación de exactitud: se aplica manualmente el proceso de verificación de las mediciones obtenidas por el procedimiento de medición y se comparan contra las mediciones expertas para determinar la efectividad.

Durante el análisis se recolectan las mediciones de tamaño funcional y se calcula la métrica de la exactitud. El procedimiento de análisis es el descrito en la Sección 6.1.1.1.3 del caso de estudio 01.

6.1.2.2. ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

La Tabla 6-5 presenta el detalle de los resultados de medición del tamaño funcional en puntos de función COSMIC FFP. En esta se detallan los resultados del tamaño funcional (CFP) para cada uno de los procesos de medición realizados (experto y procedimiento) y para cada uno de los componentes funcionales (BFC). En este caso se presentan los resultados para las funciones de entrada (E), las salidas (X), y las funciones de lectura (R) y escritura (W) en el almacenamiento persistente. La Tabla 6-6 presenta la exactitud obtenida a partir de los resultados de medición. Se presentan los porcentajes de las métricas MRE, MER, BRE e IBRE para cada uno de los BFC y para el total de puntos de función (CFP). Las métricas se detallan para las E, las X, las R y las W. Además, para cada uno de ellos se detalla la diferencia absoluta del tamaño funcional. A continuación se analizan los resultados obtenidos en el proceso de medición para responder cada una de las preguntas de investigación.

6.1.2.2.1. EXACTITUD DE LOS RESULTADOS DE MEDICIÓN DEL TAMAÑO FUNCIONAL

RQ1. *¿Cuál es la exactitud de la medición del tamaño en puntos de función obtenido por el procedimiento a partir del código fuente en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

El conteo manual realizado por el experto sobre la especificación de requerimientos IEEE obtiene un total de 118 puntos de función (CFP). En el caso de la aplicación del proceso automatizado de medición se obtienen un total de 119 CFP. En este caso, los resultados obtienen una variación en la exactitud (MRE) del 0.8%. El conteo automatizado genera una diferencia de un punto de función.

Tabla 6-5 Resultados de medición del tamaño funcional.

Proceso	E	X	R	W	Total
Experto	35	53	19	11	118
aFPA	36	53	19	11	119

Tabla 6-6 Exactitud de los resultados de medición del tamaño funcional.

	E	X	R	W	Total
MRE	2.9%	0.0%	0.0%	0.0%	0.8%
MER	2.8%	0.0%	0.0%	0.0%	0.8%
BRE	2.9%	0.0%	0.0%	0.0%	0.8%
IBRE	2.8%	0.0%	0.0%	0.0%	0.8%
Diff	1	0	0	0	1

Se realiza un análisis de los valores de puntos de función para cada uno de los requerimientos. La Tabla 6-7 detalla los resultados de medición del tamaño funcional para cada uno de estos requerimientos. En esta se detalla el conteo para cada uno de los procesos de medición realizados (experto y procedimiento de medición de la metodología **aFPA**) y para cada uno de los componentes funcionales básicos (BFC). En cada uno de los requerimientos se detalla la exactitud (MRE) y la diferencia entre el valor de control y el valor

medido (Dif). La aplicación medida cuenta con 21 requerimientos relacionados con las operaciones CRUDL.

En este caso, para realizar la comparación se analizan las funciones generadas en el reporte de medición del procedimiento sin considerar las notificadas como posibles funciones duplicadas. Estas fueron validadas contra los requerimientos para determinar el alcance de la medición. Adicionalmente, se detectan funcionalidades que el experto notifica en una sola función y que el procedimiento las reporta en varias funciones. Esto se da porque el procedimiento contempla una función como una única interacción con el usuario funcional persona. Para efectos de la comparación, estos resultados fueron unificados para compararlos con los reportados por el experto. Después de filtrar las funcionalidades, las principales diferencias en el conteo se presentan por la identificación de las salidas, las cuales son identificadas por cada elemento control sin considerar la cantidad de objetos de interés utilizados. Es importante notar, que algunas de las variaciones en el conteo de las funciones se balancean con otras variaciones, lo que hace que mejore la métrica de exactitud final. Por lo tanto, se deben revisar las reglas de mapeo para reducir estas diferencias.

Tabla 6-7 Resultados de medición del tamaño funcional por requerimiento.

RQ	aFPA					Experto					MRE	Dif
	E	X	R	W	CFP	E	X	R	W	CFP		
(R1.1)	1	1	1	0	3	1	1	1	0	3	0%	-
(R2.1)	1	1	1	1	4	1	1	1	1	4	0%	-
(R2.2)	1	2	1	0	4	1	2	1	0	4	0%	-
(R2.3)	1	1	0	1	3	1	1	0	1	3	0%	-
(R2.4)	2	2	0	1	5	2	2	0	1	5	0%	-
(R3.1)	3	3	2	0	8	2	3	2	0	7	14%	-1
(R3.2)	2	3	0	0	5	3	3	0	0	6	17%	1
(R3.3)	3	4	0	0	7	3	4	0	0	7	0%	-
(R3.4)	2	3	0	0	5	1	3	0	0	4	25%	-1
(R4.1)	1	2	0	1	4	1	2	0	1	4	0%	-
(R4.2)	1	2	1	0	4	1	2	1	0	4	0%	-
(R4.3)	1	2	0	1	4	1	1	0	1	3	33%	-1
(R4.4)	1	2	1	1	5	1	1	1	1	4	25%	-1
(R5.1)	2	3	0	0	5	2	3	0	0	5	0%	-
(R5.2)	2	3	1	1	7	2	2	1	1	6	17%	-1
(R5.3)	2	4	2	1	9	2	3	2	1	8	13%	-1
(R5.4)	2	3	1	1	7	1	3	1	1	6	17%	-1
(R6.1)	2	4	1	0	7	2	4	1	0	7	0%	-
(R6.2)	1	2	2	0	5	1	2	2	0	5	0%	-
(R6.3)	2	3	2	1	8	2	5	2	1	10	20%	+2
(R6.4)	3	3	3	1	10	4	5	3	1	13	23%	+3

6.1.2.2.2. EXACTITUD DE LOS RESULTADOS DE MEDICIÓN DE LOS COMPONENTES FUNCIONALES BÁSICOS

RQ2. *¿Cuál es la exactitud de la medición del tamaño de los componentes funcionales obtenido por el procedimiento a partir del código fuente en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

Al realizar el análisis para cada uno de los componentes funcionales básicos (BFC) se encuentran diferencias entre las mediciones reportadas por el experto y las obtenidas por el procedimiento, principalmente para las entradas (E). En total se cuentan 119 puntos de función de los cuales se identifican 26 entradas (E), 53 salidas (X), 19 accesos de lectura

(R) y 11 accesos de escritura (W). En términos de exactitud, los resultados de la métrica MRE alcanzan un 2.9% para las entradas (E) y un 0% para las salidas (X), lecturas (R) y escrituras (W). Los resultados se consideran aceptables. Al analizar los requerimientos particulares se encuentran algunas interpretaciones que se realizaron en la implementación que variaron levemente la funcionalidad con respecto a la especificación de requerimientos.

6.1.2.3. RESUMEN DE RESULTADOS

El prototipo del procedimiento de medición de la metodología **aFPA** presenta resultados de exactitud aceptables. La métrica MRE obtiene un 0.8% para el total de puntos de función (CFP), un 2.9% para las entradas (E) y un 0% para las demás funciones (X, R, W). El procedimiento identifica funciones que reporta como duplicadas. Asimismo, para la realización de la comparación con el conteo experto se unifican algunas de las funciones. En la práctica, estas decisiones deben ser validadas por el profesional para seleccionar el conjunto de funciones que deben formar parte del alcance del conteo. La toma de decisiones automáticas sobre los flujos de datos duplicados o sobre la unificación de flujos requiere más investigación y se contemplan como parte del trabajo futuro de esta propuesta.

6.1.3. CASO DE ESTUDIO 03: SISTEMA DE PROYECTOS PMS

En este caso de estudio se lleva a cabo el proceso de conteo del tamaño funcional de una aplicación transaccional utilizando la extensión del procedimiento de medición de la metodología **aFPA** propuesta en la Sección 5.5.1. Esta extensión permite aplicar el procedimiento utilizando la especificación de requerimientos bajo el estándar IEEE Std. 830-1998. Los resultados de medición obtenidos por el procedimiento son comparados con los valores de control reportados por un experto. El procedimiento de medición de la metodología **aFPA** se aplica para obtener los puntos de función sin ajustar IFPUG FPA utilizando como entrada la especificación de requerimientos funcionales. A continuación se describe el diseño del estudio y se analizan los resultados.

6.1.3.1. DESCRIPCIÓN DEL ESTUDIO EMPÍRICO

El objetivo del estudio es determinar cuál es la efectividad de las mediciones obtenidas por el procedimiento de medición utilizando como entrada la especificación de requerimientos funcionales. La efectividad del procedimiento de medición se evalúa en términos de la exactitud. A partir de las mediciones recolectadas se realizan los análisis y se comparan los resultados contra las mediciones validadas por un experto. Basado en el objetivo de la investigación se plantean las siguientes preguntas de investigación:

RQ1. *¿Cuál es la exactitud de la medición del tamaño en puntos de función sin ajustar obtenido por el procedimiento a partir de la especificación de requerimientos en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

RQ2. *¿Cuál es la exactitud de la medición del tamaño de los componentes funcionales básicos obtenido por el procedimiento a partir de la especificación de requerimientos en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

6.1.3.1.1. OBJETOS DE ESTUDIO E INSTRUMENTOS

El detalle de los objetos de estudio y los instrumentos utilizados en el caso de estudio es el siguiente:

- **Objetos experimentales:** los objetos utilizados son los artefactos de software de una aplicación transaccional: Proyectos PMS (Abrahamo, 2004). Esta aplicación tiene un tamaño de 152 UFP lo que la clasifica como una aplicación pequeña (S). El artefacto de software utilizado en el proceso de medición del experto es la especificación de requerimientos bajo el estándar IEEE 830. Asimismo, esta especificación es utilizada para el conteo mediante el procedimiento de medición. La aplicación se compone de 36 reque-

rimientos funcionales. Esta es una aplicación transaccional que implementa funcionalidades con las operaciones básicas de creación, lectura, actualización, borrado y listado de información (CRUDL).

- Instrumentos de recolección de datos: para cada una de las corridas se recolectan los resultados en los formularios de recolección de datos provistos.

6.1.3.1.2. PROCEDIMIENTO DE RECOLECCIÓN DE DATOS

El proceso realizado para la medición de la aplicación y la recolección de datos contempla las siguientes actividades:

- Se preparan y validan los objetos de estudio: en este caso se utiliza la especificación de requerimientos para la medición del método IFPUG FPA realizada por el experto. Esta misma especificación de requerimientos se utiliza para que el investigador identifique los componentes funcionales del procedimiento y realice el proceso de medición.
- Se obtiene la medición del tamaño funcional IFPUG FPA: el experto valida las mediciones del tamaño funcional de la aplicación y las reporta en el formulario de recolección de datos. Los valores de medición son los reportados en (Abrahamo, 2004). Los expertos que adaptan y validan el conteo de la aplicación son el director y el autor de esta tesis. Estas mediciones son utilizadas como valores de control en la verificación de resultados que determina la exactitud de los resultados del procedimiento.
- Se obtiene la medición del tamaño funcional mediante la aplicación del procedimiento de medición de la metodología **aFPA**: se realiza el proceso de aplicación del proceso de medición. A partir de la especificación de los requerimientos se identifican los componentes funcionales del procedimiento de medición. Para esto, primero se determina el metamodelo de los requerimientos y luego se aplican las reglas de mapeo entre el metamodelo de los artefactos de requerimientos y el metamodelo del procedimiento de

medición. Posterior a que se realiza el mapeo de los componentes, se realiza la construcción del modelo funcional y la medición del tamaño funcional mediante la aplicación de las reglas de mapeo y de asignación numérica.

- Se aplica el protocolo de verificación de exactitud: se aplica manualmente el proceso de verificación de las mediciones obtenidas por el procedimiento de medición y se comparan contra las mediciones reportadas por el experto para determinar la efectividad.

Durante el análisis se recolectan las mediciones de tamaño funcional y se calcula la métrica de la exactitud. El procedimiento de análisis es el descrito en la Sección 6.1.1.1.3 del caso de estudio 01.

6.1.3.2. ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

La Tabla 6-8 presenta el detalle de los resultados de medición del tamaño funcional en puntos de función IFPUG FPA. En esta se detallan los resultados del tamaño funcional (UFP) para cada uno de los procesos de medición (experto y procedimiento) y para cada uno de los componentes funcionales básicos (BFC). En este caso se presentan los resultados para las funciones transaccionales (TF), las funciones de datos (DF), las entradas (EI), las salidas (EO), las consultas (EQ), los archivos internos (ILF) y las interfaces externas (EIF). Asimismo, la cantidad de componentes identificados.

La Tabla 6-9 presenta la exactitud obtenida a partir de los resultados de medición. Se presentan los porcentajes de las métricas MRE, MER, BRE e IBRE para cada uno de los BFC y para el total de puntos de función sin ajustar (UFP). Las métricas se detallan para las TF y las DF. Asimismo, se detallan los porcentajes obtenidos para las EI, las EO, las EQ, los ILF y las EIF. Para cada uno de los componentes se detalla la diferencia absoluta del tamaño funcional. A continuación se analizan los resultados para responder cada una de las preguntas de investigación.

6.1.3.2.1. EXACTITUD DE LOS RESULTADOS DE MEDICIÓN DEL TAMAÑO FUNCIONAL

RQ1. *¿Cuál es la exactitud de la medición del tamaño en puntos de función sin ajustar obtenido por el procedimiento a partir de la especificación de requerimientos en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

El conteo manual realizado por el experto sobre la especificación de requerimientos IEEE obtiene un total de 152 puntos de función sin ajustar (UFP). En el caso de la aplicación del proceso de medición se obtienen un total de 151 UFP. En este caso, los resultados obtienen una variación en la exactitud (MRE) del 0.7%. Los resultados indican una diferencia de un punto de función sin ajustar (UFP).

Tabla 6-8 Resultados de medición del tamaño funcional.

Proceso	Var	TF				DF			
		EI	EO	EQ	Total	ILF	EIF	Total	Total
Experto	UFP	90	10	24	124	28	0	28	152
	Cantidad	(27)	(2)	(7)	(36)	(4)	(0)	(4)	(40)
aFPA	UFP	89	10	24	123	28	0	28	151
	Cantidad	(27)	(2)	(7)	(36)	(4)	(0)	(4)	(40)

Tabla 6-9 Exactitud de los resultados de medición del tamaño funcional.

	TF				DF			
	EI	EO	EQ	Total	ILF	EIF	Total	Total
MRE	1.1%	0.0%	0.0%	0.8%	0.0%	-	0.0%	0.7%
MER	1.1%	0.0%	0.0%	0.8%	0.0%	-	0.0%	0.7%
BRE	1.1%	0.0%	0.0%	0.8%	0.0%	-	0.0%	0.7%
IBRE	1.1%	0.0%	0.0%	0.8%	0.0%	-	0.0%	0.7%
Diff	1	0	0	1	0	0	0	1

Se realiza el análisis de los valores de puntos de función para cada uno de los requerimientos. La Tabla 6-10 detalla los resultados de la medición del tamaño funcional para cada uno de los requerimientos. En esta se detalla el conteo para cada uno de los procesos de medición (experto y procedimiento de medición de la metodología **aFPA**) y para cada

uno de los componentes funcionales básicos (BFC). Asimismo, se detallan las cantidades de los elementos de tipo dato (DET), los tipos de archivo referenciado (FTR), los elementos de tipo registro (RET) y de la complejidad (C). Para cada uno de los requerimientos se detalla la exactitud (MRE) y la diferencia entre el valor de control y el valor medido (Dif). La aplicación medida cuenta con 36 requerimientos relacionados con las operaciones CRUDL.

Los resultados de la comparación del tamaño de las funciones identificadas indican que existe poca variabilidad en los resultados en comparación con los valores de control. Es importante notar, que algunas de las funciones no presentan variaciones en el conteo, pero presentan distintas cantidades de DETs. En este caso, los intervalos de las reglas de asignación numérica hacen que el total de tamaño funcional no se vea afectado. El requerimiento que presenta la diferencia de un punto de función sí se ve afectado por el número de DETs identificados. Por lo tanto, se deben revisar y aclarar las reglas de mapeo para reducir estas diferencias. Sin embargo, en este proceso de medición, los resultados presentan una variación en la exactitud cercana a cero.

6.1.3.2.2. EXACTITUD DE LOS RESULTADOS DE MEDICIÓN DE LOS COMPONENTES FUNCIONALES BÁSICOS

RQ2. ¿Cuál es la exactitud de la medición del tamaño de los componentes funcionales básicos obtenido por el procedimiento a partir de la especificación de requerimientos en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?

Al realizar el análisis para cada uno de los componentes funcionales básicos (BFC) se encuentran pequeñas diferencias entre las mediciones reportadas por el experto y las obtenidas por el procedimiento. En total se cuentan 124 puntos de función transaccionales (TF) y 28 puntos de función de datos (DF). De las 36 funciones transaccionales, 27 son entradas (EI), 2 son salidas (EO) y 7 son consultas (EQ). Cuatro funciones de datos son clasificadas como archivos internos (ILF) y no se reportan interfaces externas (EIF). En términos de exactitud, los resultados de la métrica MRE alcanzan un 0.8% para las TF y un 0% para

las DF. En este caso, las EI obtienen un resultado del 1.1%, las EQ del 0% y las EO del 0.0%. Los resultados se consideran aceptables.

Tabla 6-10 Resultados de medición del tamaño funcional por requerimiento.

RQ	aPPA					Experto					MRE	Dif
	BCF	FTR/RET	DET	C	UFP	BCF	FTR/RET	DET	C	UFP		
(R0.1)	EQ	1	5	L	3	EQ	1	5	L	3	0%	-
(R0.2)	EI	1	5	L	3	EI	1	5	L	3	0%	-
(R1.1)	EI	1	16	A	4	EI	1	16	A	4	0%	-
(R1.2)	EI	2	6	A	4	EI	2	5	A	4	0%	-
(R1.3)	EI	2	8	A	4	EI	2	7	A	4	0%	-
(R1.4)	EI	2	4	L	3	EI	2	6	A	4	25%	+1
(R1.5)	EI	1	5	L	3	EI	1	5	L	3	0%	-
(R1.6)	EI	1	5	L	3	EI	1	5	L	3	0%	-
(R1.7)	EO	3	15	A	5	EO	3	16	A	5	0%	-
(R1.8)	EI	1	5	L	3	EI	1	5	L	3	0%	-
(R1.9)	EI	1	6	L	3	EI	1	5	L	3	0%	-
(R2.1)	EI	1	12	L	3	EI	1	12	L	3	0%	-
(R2.2)	EI	2	6	A	4	EI	2	7	A	4	0%	-
(R2.3)	EI	2	7	A	4	EI	2	8	A	4	0%	-
(R2.4)	EI	2	8	A	4	EI	2	8	A	4	0%	-
(R2.5)	EI	2	9	A	4	EI	2	8	A	4	0%	-
(R2.6)	EQ	2	13	A	4	EQ	2	12	A	4	0%	-
(R3.1)	EI	1	7	L	3	EI	1	7	L	3	0%	-
(R3.2)	EI	2	4	L	3	EI	2	4	L	3	0%	-
(R3.3)	EI	1	4	L	3	EI	1	4	L	3	0%	-
(R3.4)	EI	1	4	L	3	EI	1	4	L	3	0%	--
(R3.5)	EI	1	4	L	3	EI	1	4	L	3	0%	-
(R3.6)	EQ	1	4	L	3	EQ	1	8	L	3	0%	-
(R3.7)	EQ	1	11	L	3	EQ	1	11	L	3	0%	-
(R3.8)	EQ	2	12	A	4	EQ	2	12	A	4	0%	-
(R3.9)	EQ	2	11	A	4	EQ	2	12	A	4	0%	-
(R4.1)	EI	1	4	L	3	EI	1	4	L	3	0%	-
(R4.2)	EI	1	3	L	3	EI	1	3	L	3	0%	-
(R4.3)	EI	1	4	L	3	EI	1	4	L	3	0%	-
(R4.4)	EO	3	14	A	5	EO	3	15	A	5	0%	-
(R4.5)	EI	2	4	L	3	EI	2	4	L	3	0%	-
(R5.1)	EI	1	6	L	3	EI	1	6	L	3	0%	-
(R5.2)	EI	1	3	L	3	EI	1	3	L	3	0%	-
(R5.3)	EI	1	4	L	3	EI	1	4	L	3	0%	-
(R5.4)	EQ	2	5	L	3	EQ	2	5	L	3	0%	-
(R5.5)	EI	2	5	A	4	EI	2	5	A	4	0%	-
(R1,R2)	ILF	2	22	L	7	ILF	2	22	L	7	0%	-
(R3)	ILF	1	6	L	7	ILF	1	6	L	7	0%	-
(R4)	ILF	1	2	L	7	ILF	1	2	L	7	0%	-
(R5)	ILF	1	4	L	7	ILF	1	4	L	7	0%	-

6.1.3.3. RESUMEN DE RESULTADOS

La extensión del procedimiento de medición de la metodología **aFPA** para medir documentos de especificación de requerimientos provee resultados de exactitud aceptables. La métrica MRE obtiene un 0.7% para el total de puntos de función sin ajustar (UFP), un 1.1% para las funciones transaccionales (TF) y un 0% para las funciones de datos (DF). Los análisis de la exactitud para cada uno de los componentes funcionales básicos (BFC), muestra que las entradas (EI), las consultas (EQ) y las salidas (EO) obtienen resultados aceptables. Es necesario analizar las diferencias en la identificación de los elementos de tipo dato (DET) que eventualmente podrían afectar los resultados de la medición del tamaño funcional.

6.1.4. CASO DE ESTUDIO 04: UN CASO DE ESTUDIO EN LA INDUSTRIA

En este caso de estudio se lleva a cabo el conteo del tamaño funcional de una aplicación transaccional utilizando el prototipo de soporte del procedimiento de medición de la metodología **aFPA** en el nivel 2 de automatización. Para esto se realiza el proceso de medición utilizando modelos conceptuales construidos bajo una herramienta de desarrollo dirigido por modelos (Quesada-López, Salas, Gómez & Jenkins, 2017a; 2017b). El procedimiento de medición de la metodología **aFPA** se aplica para obtener los puntos de función sin ajustar IFPUG FPA utilizando como entrada el modelo conceptual de una aplicación desarrollada utilizando la plataforma de desarrollo FastWorks. Los resultados de medición obtenidos por el procedimiento son comparados con los valores de control obtenidos por un experto para determinar su exactitud. Este caso de estudio evalúa la propuesta presentada en la Sección 5.5.2. A continuación se describe el diseño del estudio y se analizan los resultados.

6.1.4.1. DESCRIPCIÓN DEL ESTUDIO

El objetivo del estudio es determinar cuál es la efectividad de las mediciones obtenidas por el procedimiento de medición de la metodología **aFPA** utilizado la plataforma de

desarrollo FastWorks. La efectividad del procedimiento de medición se evalúa en términos de la exactitud. A partir de las mediciones recolectadas se realizan los análisis y se comparan los resultados contra los reportados por un experto. Basado en el objetivo de la investigación se plantean las siguientes preguntas de investigación:

RQ1. *¿Cuál es la exactitud de la medición del tamaño en puntos de función sin ajustar obtenido por el procedimiento a partir del modelo conceptual en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

RQ2. *¿Cuál es la exactitud de la medición del tamaño de los componentes funcionales básicos obtenido por el procedimiento a partir del modelo conceptual en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

6.1.4.1.1. OBJETOS DE ESTUDIO E INSTRUMENTOS

El detalle de los objetos de estudio y los instrumentos utilizados en el caso de estudio es el siguiente:

- **Objetos experimentales:** los objetos utilizados son los artefactos de software de una aplicación transaccional: Universidad Contoso (Quesada-López, Madrigal & Jenkins, 2016). Esta aplicación tiene un tamaño de 164 UFP lo que la clasifica como una aplicación pequeña (S). El artefacto de software utilizado en el proceso de medición experta es la especificación de requerimientos bajo el estándar IEEE 830. En el caso de la medición mediante FastWorks FPA se utiliza el modelo conceptual que describe las funcionalidades. La aplicación consta de 32 requerimientos funcionales. Esta es una aplicación transaccional que implementa funcionalidades con las operaciones básicas de creación, lectura, actualización, borrado y listado de información (CRUDL).
- **Instrumentos de recolección de datos:** para cada una de las corridas se recolectan los resultados en los formularios de recolección de datos provistos.

6.1.4.1.2. PROCEDIMIENTO DE RECOLECCIÓN DE DATOS

El proceso realizado para la medición de la aplicación y la recolección de datos contempla las siguientes actividades:

- Se preparan y validan los objetos de estudio: en este caso se utiliza la especificación de requerimientos para la medición del método IFPUG FPA por parte del experto. Esta misma especificación de requerimientos se utiliza para que el profesional modele la aplicación en la plataforma de desarrollo FastWorks. El modelo conceptual de la aplicación se utiliza para la medición con el soporte del prototipo FastWorks FPA (Quesada-López, Salas, Gómez & Jenkins, 2017b).
- Se obtiene la medición del tamaño funcional IFPUG FPA: el experto realiza el proceso de conteo y de validación de las mediciones del tamaño funcional de la aplicación y las reporta en el formulario de recolección de datos. Estas mediciones son utilizadas como valores de control en la verificación de resultados que determina la exactitud de los resultados del procedimiento.
- Se obtiene la medición del tamaño funcional mediante la aplicación del nivel 2 de automatización del procedimiento de medición de la metodología **aFPA**: se realiza el proceso de aplicación del proceso de medición. En el nivel 2 de automatización, el profesional, a partir del diseño y los lineamientos del procedimiento, realiza la identificación de los componentes funcionales por medio del análisis del modelo conceptual, lo cual no requiere ninguna adaptación del código fuente de la aplicación. En este caso, la herramienta FastWorks FPA es calibrada mediante la ejecución de múltiples corridas piloto. Posteriormente, el procedimiento de medición recibe el archivo de interfaz XML como entrada para procesarlo. A partir de este punto, el procedimiento construye el modelo funcional y realiza el proceso de medición tal como se detalla en la Sección 6.1.1.2. El proceso de medición aplica el nivel 2 de automatización del procedimiento de medición de la metodología **aFPA** el cual se describe en la Figura 5-9. El proceso detallado de los pasos para realizar la medición bajo este contexto de desarrollo dirigido por

modelos se detalla en el estudio presentado en (Quesada-López, Jenkins, Salas & Gómez, 2017b).

- Se aplica el protocolo de verificación de exactitud: se aplica manualmente el proceso de verificación de las mediciones obtenidas por el procedimiento de medición y se comparan contra las mediciones expertas para determinar la efectividad.

Durante el análisis se recolectan las mediciones de tamaño funcional y se calcula la métrica de la exactitud. El procedimiento de análisis es el descrito en la Sección 6.1.1.1.3 del caso de estudio 01.

6.1.4.2. ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

La Tabla 6-11 presenta el detalle de los resultados de medición del tamaño funcional en puntos de función IFPUG FPA. En esta se detallan los resultados del tamaño funcional (UFP) para cada uno de los procesos de medición (experto y procedimiento) y para cada uno de los componentes funcionales básicos (BFC). En este caso se presentan los resultados para las funciones transaccionales (TF), las funciones de datos (DF), las entradas (EI), las salidas (EO), las consultas (EQ), los archivos internos (ILF) y las interfaces externas (EIF). Además, para cada uno de ellos se detalla la cantidad de componentes identificados.

La Tabla 6-12 presenta la exactitud obtenida a partir de los resultados de medición. Se presentan los porcentajes de las métricas MRE, MER, BRE e IBRE para cada uno de los BFC y para el total de puntos de función sin ajustar (UFP). Las métricas se detallan para las TF y las DF. Asimismo, se detallan los porcentajes para las EI, las EO, las EQ, los ILF y las EIF. Para cada uno de los componentes se detalla la diferencia absoluta del tamaño funcional.

Tabla 6-11 Resultados de medición del tamaño funcional.

Proceso	Var	TF				DF			
		EI	EO	EQ	Total	ILF	EIF	Total	Total
Experto	UFP	58	21	50	129	35	0	35	164
	Cantidad	(15)	(4)	(13)	(32)	(5)	(0)	(5)	(37)
aFPA	UFP	61	9	50	120	35	0	35	155
	Cantidad	(15)	(2)	(13)	(30)	(5)	(0)	(5)	(35)

Tabla 6-12 Exactitud de los resultados de medición del tamaño funcional.

	TF				DF			
	EI	EO	EQ	Total	ILF	EIF	Total	Total
MRE	5.2%	57.1%	0.0%	7.0%	0.0%	-	0.0%	5.5%
MER	4.9%	133.3%	0.0%	7.5%	0.0%	-	0.0%	5.8%
BRE	5.2%	133.3%	0.0%	7.5%	0.0%	-	0.0%	5.8%
IBRE	4.9%	57.1%	0.0%	7.0%	0.0%	-	0.0%	5.5%
Diff	3	12	0	9	0	0	0	9

A continuación se analizan los resultados obtenidos en el proceso de medición y se responden las preguntas de investigación.

6.1.4.2.1. EXACTITUD DE LOS RESULTADOS DE MEDICIÓN DEL TAMAÑO FUNCIONAL

RQ1. *¿Cuál es la exactitud de la medición del tamaño en puntos de función sin ajustar obtenido por el procedimiento a partir del modelo conceptual en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

El conteo manual realizado por el experto sobre la especificación de requerimientos IEEE obtiene un total de 164 puntos de función sin ajustar (UFP). En el caso de la aplicación del proceso automatizado de medición se obtienen un total de 155 UFP. En este caso, los resultados obtienen una variación en la exactitud (MRE) del 5.5%. El procedimiento de medición identifica 35 de las 37 funciones las cuales generan una diferencia total de 9 UFP.

Durante el modelado de la aplicación en FastWorks se registran los requerimientos a los que responden cada uno de los elementos del modelo. La Tabla 6-13 detalla los resultados de la medición del tamaño funcional por cada uno de los requerimientos. En esta se detalla el conteo para cada uno de los procesos de medición (experto y procedimiento de medición de la metodología **aFPA**) y para cada uno de los componentes funcionales básicos (BFC). Asimismo, se detalla la cantidad de los elementos de tipo dato (DET), los tipos de archivo referenciado (FTR), los elementos de tipo registro (RET) y de la complejidad (C). Para cada uno de los requerimientos se detalla la exactitud (MRE) y la diferencia entre el valor de control y el valor medido (Dif). La aplicación medida cuenta con 32 requerimientos relacionados con las operaciones CRUDL. Las principales diferencias se presentan con funcionalidades de filtrado que no fueron identificadas en el conteo automático. En términos generales, las diferencias en los componentes se dan en la identificación de campos técnicos y entidades de datos. Asimismo, la identificación de las salidas externas (EO) no presentan los resultados esperados ya que el procedimiento las clasifica en la mayoría de los casos como consultas externas (EQ).

Tabla 6-13 Resultados de medición del tamaño funcional por requerimiento.

RQ	aFPA					Experto					MRE	Dif
	BCF	FIR/RET	DET	C	UFP	BCF	FIR/RET	DET	C	UFP		
(R1.1)	EI	2	7	A	4	EI	2	8	A	4	0.0%	-
(R1.2)	EI	2	8	A	4	EI	2	8	A	4	0.0%	-
(R1.3)	EI	2	7	A	4	EI	2	7	A	4	0.0%	-
(R1.4)	EQ	2	6	A	4	EQ	2	8	A	4	0.0%	-
-	EQ	1	2	L	3	-	-	-	-	0	-	-3
(R1.5)	-	-	-	-	0	EQ	2	8	A	4	100.0%	+4
(R1.6)	-	-	-	-	0	EQ	2	6	A	4	100.0%	+4
(R1.7)	EQ	2	7	A	4	EQ	2	8	A	4	0.0%	-
(R1.8)	EQ	5	20	H	6	EO	5	19	H	7	14.3%	+1
(R1.9)	EI	1	2	L	3	EI	1	2	L	3	0.0%	-
(R2.1)	EI	2	11	A	4	EI	1	8	L	3	33.3%	-1
(R2.2)	EI	2	12	A	4	EI	1	8	L	3	33.3%	-1
(R2.3)	EI	2	10	A	4	EI	1	8	L	3	33.3%	-1
(R2.4)	EQ	2	9	A	4	EQ	2	8	A	4	0.0%	-
(R2.5)	EQ	2	6	A	4	EQ	2	4	L	3	33.3%	-1
(R2.6)	EQ	3	5	L	3	EQ	4	7	H	6	50.0%	+3
(R2.7)	EQ	1	6	L	3	EQ	1	6	L	3	0.0%	-
(R2.8)	EQ	3	14	A	4	EO	3	14	A	5	20.0%	+1
(R3.1)	EI	2	9	A	4	EI	2	7	A	4	0.0%	-
(R3.2)	EI	2	10	A	4	EI	2	8	A	4	0.0%	-
(R3.3)	EI	2	10	A	4	EI	2	7	A	4	0.0%	-
(R3.4)	EQ	2	9	A	4	EQ	2	8	A	4	0.0%	-
(R3.5)	EQ	2	8	A	4	EQ	2	8	A	4	0.0%	-
(R4.1)	EI	1	7	L	3	EI	1	7	L	3	0.0%	-
(R4.2)	EI	1	7	L	3	EI	1	7	L	3	0.0%	-
(R4.3)	EI	2	7	A	4	EI	2	8	A	4	0.0%	-
(R4.4)	EQ	1	7	L	3	EQ	1	7	L	3	0.0%	-
(R4.5)	-	-	-	-	0	EQ	1	7	L	3	100.0%	+3
(R4.6)	EI	3	16	H	6	EI	4	5	H	6	0.0%	-
(R4.7)	EO	3	13	A	5	EO	3	17	A	5	0.0%	-
(R4.8)	EI	3	12	H	6	EI	4	6	H	6	0.0%	-
(R4.9)	EQ	3	13	A	4	EQ	3	7	A	4	0.0%	-
(R4.10)	EO	0	1	L	4	EO	3	4	L	4	0.0%	-

6.1.4.2.2. EXACTITUD DE LOS RESULTADOS DE MEDICIÓN DE LOS COMPONENTES FUNCIONALES BÁSICOS

RQ2. *¿Cuál es la exactitud de la medición del tamaño de los componentes funcionales básicos obtenido por el procedimiento a partir del modelo conceptual en comparación con los resultados obtenidos por un experto a partir de la especificación de requerimientos?*

Al realizar el análisis para cada uno de los componentes funcionales básicos (BFC) se encuentran diferencias entre las mediciones reportadas por el experto y las obtenidas por el procedimiento. El conteo manual identifica 32 funciones transaccionales (TF) y 5 funciones de datos (DF). En total se cuentan 129 puntos de función transaccionales y 35 puntos de función de datos. Las 32 funciones transaccionales consisten de 15 entradas (EI), 4 salidas (EO) y 13 consultas (EQ). Cinco funciones de datos son clasificadas como archivos internos (ILF) y no se reportan interfaces externas (EIF). El procedimiento reporta las mismas funciones con excepción de las salidas (EO) donde identifica solo dos funciones. En términos de exactitud, los resultados de la métrica MRE alcanzan un 7% para las TF y un 0% para las DF. En este caso, las EI obtienen un resultado del 5% y las EQ del 0%; sin embargo, las EO presentan una variabilidad del 57%. La identificación de las salidas externas (EO) no presenta los resultados esperados ya que el procedimiento las clasifica en muchos casos como consultas externas (EQ). En este caso, el procedimiento no identifica campos calculados durante el análisis del modelo funcional que permitan la clasificación de este tipo de funciones.

6.1.4.3. RESUMEN DE RESULTADOS

El procedimiento de medición de tamaño funcional de la metodología **aFPA** aplicado mediante la plataforma de desarrollo FastWorks presenta resultados de exactitud aceptables. La métrica MRE obtiene un 5.5% para el total de puntos de función sin ajustar (UFP), un 7% para las funciones transaccionales (TF) y un 0% para las funciones de datos (DF). Los

análisis de la exactitud para cada uno de los componentes funcionales básicos (BFC), muestra que, aunque las entradas (EI) y las consultas (EQ) obtienen resultados aceptables, las salidas (EO) presentan desviaciones importantes con respecto a los valores de control (57%). Es necesario implementar acciones para corregir estos problemas. Específicamente, el procedimiento no clasifica algunos de los EQ como EO porque no puede identificar los campos calculados; por lo tanto, el proceso de generación de este tipo de información desde la plataforma debe ser adaptada e implementada. Las mejoras en el proceso de medición mediante la plataforma FastWorks FPA se considera como trabajo futuro de este estudio.

6.1.5. LIMITACIONES DE LOS CASOS DE ESTUDIO

A continuación se analizan las principales limitaciones a considerar a partir de los estudios realizados. Para cada uno de los casos de estudio presentados, el procedimiento de medición de tamaño funcional de la metodología **aFPA** obtiene resultados de exactitud aceptables, lo que indica que puede estimar el tamaño funcional de las aplicaciones transaccionales bajo estudio. Sin embargo, estos resultados solo pueden ser generalizables a los objetos de estudio y los artefactos utilizados en el proceso de medición. Los casos de estudio presentados se implementaron para tecnologías *Asp.Net C#* bajo la arquitectura *MVC* pero pueden ser adaptables a distintos ambientes de desarrollo o inclusive ser utilizados como servicios de medición externos bajo distintos niveles de automatización. El caso de estudio de la Sección 6.1.4 detalla un ejemplo de cómo el procedimiento de medición puede ser utilizado en un contexto de desarrollo particular y calibrado a las necesidades específicas de una organización bajo una plataforma específica.

Para cada una de los procesos de medición realizados, se ejecutaron múltiples corridas de calibración que permitieron obtener los resultados indicados. Por lo tanto, al utilizar el procedimiento de medición en contextos de desarrollo particulares es necesario realizar procesos de calibración similares. Es importante mencionar que, en algunos casos, las reglas de mapeo y de medición fueron calibradas de acuerdo a como las aplicaciones transaccionales fueron construidas. Esto indica que es recomendable que las reglas sean

parametrizadas a partir de las prácticas de desarrollo específicas de cada organización durante el proceso de calibración. Los resultados obtenidos son prometedores; sin embargo, adaptar el proceso de medición a los procesos de desarrollo de software de las organizaciones requiere un plan de adaptación del procedimiento para ajustarse a las necesidades particulares de una organización de desarrollo.

6.1.6. DISCUSIÓN DE LOS RESULTADOS DE LOS CASOS DE ESTUDIO

En esta sección se han presentado los resultados de un conjunto de casos de estudio que evaluaron la exactitud del procedimiento de medición del tamaño funcional de la metodología **aFPA**. Asimismo se ha ejemplificado el proceso de medición utilizando el procedimiento propuesto en distintos contextos de desarrollo. Los casos de estudio miden aplicaciones transaccionales para obtener el tamaño funcional basado en los métodos IFPUG FPA y COSMIC FFP utilizando como entrada distintos artefactos de software. Los resultados de exactitud obtenidos son aceptables y el MRE varía entre un 0.8% y un 5.5% para el total de puntos de función. Los resultados indican que, para las aplicaciones medidas, los resultados de tamaño funcional obtienen una variación en la exactitud menor al $\pm 10\%$, las cuales son consideradas satisfactorias para un sistema de medición (AIAG, 2002; Abran, 2015).

6.2. EVALUACIÓN DE LA EFECTIVIDAD DEL PROCEDIMIENTO DE MEDICIÓN DEL TAMAÑO FUNCIONAL: UN CUASI EXPERIMENTO Y UNA FAMILIA DE REPLICACIONES

Este estudio representa el producto obtenido del ciclo de investigación 5 y del ciclo de investigación 9 (RC5, RC9).
--

En esta sección se describen los estudios empíricos orientados a personas que permiten evaluar la efectividad y adopción de cuatro procesos de medición del tamaño funcional. Para esto se seleccionan tres procesos de medición existentes (IFPUG FPA, COSMIC

FFP, OMG AFP) y el procedimiento de medición de la metodología **aFPA** propuesto en el Capítulo 5. Específicamente, este conjunto de replications evalúa los resultados de la aplicación de los procesos de medición para determinar cómo las mediciones obtenidas por los profesionales se comparan con las reportadas por un experto. En total, participan en estos cuatro estudios empíricos 58 profesionales de la ingeniería del software que tomaron cursos de maestría en computación e informática relacionados con temas de métricas del software. Para la comparación de las mediciones se utiliza el protocolo de verificación descrito en la Sección 5.2. Asimismo, se realiza una validación del módulo propuesto en la Sección 5.4.2.

El objetivo de los estudios es determinar cuál es la efectividad de las mediciones obtenidas por los participantes, así como las propiedades de adopción de los procesos de medición (Quesada-López & Jenkins, 2015a; 2015b; Quesada-López, Madrigal & Jenkins, 2016; Madrigal-Sánchez, Quesada-López & Jenkins, 2018). Para esto se lleva a cabo un estudio cuasi experimental y un conjunto de replications diferenciadas. La familia de replications ejecuta cuatro estudios empíricos orientados a personas. Los sujetos aplican los procesos de medición a los artefactos de software asignados para obtener el tamaño funcional. A partir de los datos recolectados se realizan los análisis y se comparan los resultados. La estructura del reporte de la investigación se realiza de acuerdo a los lineamientos para estudios experimentales propuesta por (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012). En este caso, primero se describe el diseño del estudio y segundo se presentan los análisis de resultados. Asimismo, se consideran las recomendaciones para reportar familias de experimentos descritas en (Carver, 2010).

6.2.1. DESCRIPCIÓN DEL ESTUDIO EMPÍRICO

El estudio ejecuta una familia de replications diferenciadas donde se evalúa la efectividad obtenida durante la aplicación de procesos de medición sobre aplicaciones de software transaccionales. Los procesos de medición del tamaño funcional evaluados son el método de medición IFPUG FPA, el método de medición COSMIC FFP, la especificación de medición OMG AFP y el procedimiento de medición de la metodología **aFPA**. Primero, se

ejecuta un estudio controlado cuasi experimental con 14 participantes para comparar los procesos de medición entre el método IFPUG FPA y la especificación OMG AFP (E2-14). En las subsecuentes tres replicaciones se comparan los procesos de medición entre el método IFPUG FPA y la especificación OMG AFP con 12 profesionales (E1-15) y entre los procesos de medición de la especificación OMG AFP y el procedimiento de medición de la metodología **aFPA** con 12 profesionales (E2-15). Finalmente, se analizan los procesos de medición utilizando el método IFPUG FPA, el método COSMIC FFP y el procedimiento de medición de la metodología **aFPA** con 20 profesionales (E1-18), tal como se muestra en la Figura 6-11. La motivación para realizar las diferentes replicaciones es determinar la efectividad de distintos procesos de medición bajo un conjunto de artefactos de software y analizar las distintas condiciones que pueden afectar la efectividad de los resultados obtenidos. Asimismo, es de interés analizar el nivel de adopción percibido por los profesionales para cada uno de los procesos de medición.

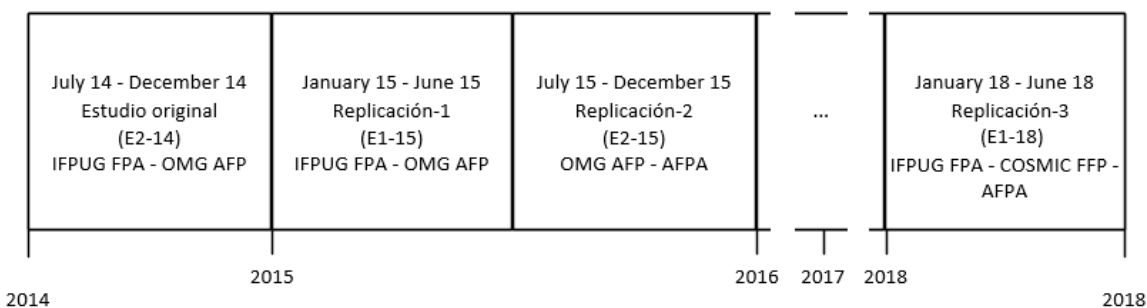


Figura 6-11 Cronología de la familia de replicaciones.

Los detalles del proceso experimental y la configuración de los estudios empíricos, las hipótesis, la instrumentación y los objetos experimentales, y la selección de los sujetos se detalla en las secciones posteriores. Todas las replicaciones internas fueron conducidas por los investigadores originales. Estas se ejecutaron como actividades en línea durante distintos cursos de maestría donde los profesionales se encontraban matriculados. Las decisiones para cada uno de los estudios se toman de acuerdo a las restricciones de los cursos

en los cuales fueron ejecutados. Los investigadores configuran los estudios para que las actividades en línea no alteren los objetivos de los cursos.

Basado en los objetivos de la investigación se analizan las siguientes preguntas de investigación:

RQ1. *¿Cuál es la efectividad de las mediciones realizadas por los profesionales aplicando los procesos de medición?*

RQ2. *¿Cuáles son las propiedades de adopción de los procesos de medición percibidas por los profesionales?*

6.2.1.1. PROCEDIMIENTO EXPERIMENTAL

En términos generales, el proceso realizado en cada uno de los estudios contempla las siguientes actividades:

- **Actividades de capacitación:** los participantes asisten a las sesiones de capacitación de cada uno de los procesos de medición. En cada sesión los conceptos teóricos, las reglas de medición y los ejemplos de medición son presentados. Asimismo, las tareas de medición y los sistemas bajo medición son presentados.
- **Actividades de medición:** los sujetos son divididos en dos grupos y asignados al grupo de tratamiento o al grupo de control de manera aleatoria. Cada uno de los grupos tiene asignado un tratamiento diferente. En la medida de lo posible se mantienen los grupos balanceados. Cada uno de los participantes aplica las reglas del proceso de medición asignado a los artefactos de software y recolecta las mediciones de tamaño funcional en la hoja de recolección de datos suministrada. Se solicita a los participantes detallar todo el proceso de medición y detallar las mediciones para cada uno de los componentes funcionales básicos. Finalmente, los datos recolectados son procesados para determinar la efectividad.

- Actividades de verificación: cada uno de los participantes aplica el protocolo de verificación para determinar las diferencias de exactitud en comparación con los valores verdaderos tomados como valores de control.
- Encuesta de salida: cada uno de los participantes completa una encuesta que recolecta los datos demográficos, la percepción sobre el proceso de medición y el proceso de verificación para determinar las propiedades de adopción. Se indica explícitamente que la asignación de la nota para la encuesta de salida se asigna por completitud de llenado y no por el contenido, por lo que se les invita a ofrecer respuestas honestas con respecto a los procesos de medición.

En total se ejecutan cuatro estudios que combinan diseños entre-sujetos e intra-sujetos. Un diseño entre-sujetos es aquel en el que cada sujeto recibe un único tratamiento y un diseño intra-sujetos es aquel en el que cada sujeto recibe todos los tratamientos (Dolado & Fernandez, 2000; Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012; Genero, Cruz-Lemus & Piattini, 2014). Los pros y contras de cada uno de los diseños se discuten en las amenazas a la validez de los estudios en la Sección 6.2.2.1.5. Las decisiones de cada uno de los diseños se tomaron de acuerdo a las restricciones del curso en el cual fueron ejecutados. El diseño de cada uno de los estudios empíricos y el detalle de las corridas realizadas se especifica a continuación:

- Estudio original (E2-14): diseño simple entre-sujetos con aleatorización. Se asignan catorce sujetos aleatoriamente a dos grupos (G1, G2). Se utilizan dos tratamientos (FPA, AFP) con dos objetos de estudio de complejidad similar (A, B). En este estudio se realiza una corrida donde el grupo G1 mide el objeto A con el proceso de medición FPA y el grupo G2 mide el objeto B con el proceso de medición AFP. En este caso, durante el curso cada participante solo aplica uno de los dos procesos de medición. Los objetos experimentales se detallan en las secciones posteriores.

- Replicación-1 (E1-15): diseño simple intra-sujetos. Se cuenta con doce sujetos los cuales reciben todos los tratamientos (G1). Se asignan los tratamientos FPA y AFP a dos objetos de estudio con complejidad similar (B, C). En este estudio, los tratamientos son asignados a los sujetos en orden y el proceso se realiza en dos corridas. En la corrida 1 el grupo G1 mide el objeto C con el proceso de medición FPA. En la corrida 2 el grupo G1 mide el objeto B con el proceso de medición AFP. En este caso, la restricción del curso es que los participantes apliquen los dos procesos de medición, primero el proceso FPA y segundo el proceso AFP que permite medir el código fuente.
- Replicación-2 (E2-15): diseño simple intra-sujetos. Se cuenta con doce sujetos quienes reciben todos los tratamientos (G1). Se asignan los tratamientos AFP y **aFPA** a dos objetos de estudio con complejidad similar (D, E). Los tratamientos son asignados a los sujetos en orden y el proceso se realiza en dos ejecuciones. En este estudio se realizan dos corridas, en la corrida 1 el grupo G1 mide el objeto D con el proceso de medición AFP. En la corrida 2 el grupo G1 mide el objeto E con el proceso de medición de la metodología **aFPA**. En este caso, la restricción del curso es que los participantes apliquen los dos procesos de medición, primero AFP como un ejemplo de análisis estático de código fuente y luego **aFPA** como un ejemplo de análisis estático y dinámico sobre el código fuente de la aplicación ejecutándose en un entorno controlado.
- Replicación-3 (E1-18): diseño intra-sujetos con confusión parcial. Se utilizan dos objetos que en combinación con los tratamientos permite la asignación de los sujetos a cuatro tratamientos distintos. Se asignan trece sujetos aleatoriamente a dos grupos (G1, G2). Se asignan los tratamientos FPA y FFP a dos objetos de estudio con complejidad y tamaño similar (F, G). Se realizan dos corridas donde los tratamientos son asignados a los sujetos cambiando el orden de los dos objetos (F, G). El estudio se complementa con un diseño simple entre-sujetos sin aleatorización para el tratamiento (**aFPA**) y el objeto (H). Para esto siete sujetos son asignados a un tercer grupo (G3). En este caso, se comparan los dos tratamientos (FPA, **aFPA**) asignados a dos objetos de estudio (F, H). En total, se analizan tres tratamientos (FPA, FFP, **aFPA**) con tres objetos de estudio de complejidad similar (F, G, H). En este estudio se realizan tres corridas. En la corrida 1 el

grupo G1 mide el objeto F y el grupo G2 mide el objeto G, ambos con el proceso de medición FPA. En la corrida 2 el grupo G1 mide el objeto G y el grupo G2 mide el objeto F, ambos con el proceso de medición FFP. Finalmente, en la corrida 3 el grupo G3 mide el objeto H con el proceso de medición de la metodología **aFPA**. En este caso, la restricción del curso es que los participantes aplicaran los dos procesos de medición, primero FPA como primera generación y luego FFP como segunda generación de los métodos de medición. El proceso de medición de la metodología **aFPA** es ejecutado por un grupo independiente sin conocimientos de los otros procesos de medición utilizado para la comparación.

La Tabla 6-14 lista las corridas realizadas por estudio empírico detallando la asignación de los tratamientos y objetos a los grupos. La Tabla 6-15 resume las decisiones y las configuraciones específicas de cada uno de los estudios indicando los participantes, el periodo de la corrida, el diseño, la asignación de objetos a sujetos, los procesos de medición utilizados, la duración de las actividades, las métricas recolectadas, los artefactos y los objetos de estudio, las aplicaciones bajo medición y los materiales y manuales. El detalle de los sujetos que participaron en el estudio se presenta en la Sección 6.2.2.1.4 y los objetos de estudio son descritos en la Sección 6.2.2.1.3.

Tabla 6-14 Diseño de los estudios empíricos.

Estudio	Diseño	Tratamiento	Corrida	Grupo	n	Objeto
E2-14	Entre-sujetos	FPA	1	G1	7	A
		AFP	1	G2	7	B
E1-15	Intra-sujetos	FPA	1	G1	12	C
		AFP	2	G1	12	B
E2-15	Intra-sujetos	AFP	1	G1	12	D
		aFPA	2	G1	12	E
E1-18	Intra-sujetos	FPA	1	G1	6	F
		FPA	1	G2	7	G
		FFP	2	G1	6	G
		FFP	2	G2	7	F
		Entre-sujetos	aFPA	3	G3	6

Tabla 6-15 Decisiones y configuración de los estudios empíricos.

	Estudio original (E2-14)	Replicación-1 (E1-15)	Replicación-2 (E2-15)	Replicación-3 (E1-18)
Participantes	14 profesionales	12 profesionales	12 profesionales	20 profesionales
Periodo	II-2014	I-2015	II-2015	I-2018
Diseño	Entre-sujetos (un factor con dos tratamientos).	Intra-sujetos (un factor con dos tratamientos).	Intra-sujetos (un factor con dos tratamientos).	Entre-sujetos (un factor dos tratamientos) e Intra-sujetos (un factor tres tratamientos).
Objetos a sujetos	Asignación aleatoria	Asignación aleatoria	Asignación aleatoria	Asignación aleatoria
Proceso de medición	IFPUG FPA (n=7), OMG AFP (n=7).	IFPUG FPA (n=12), OMG AFP (n=12).	OMG AFP (n=9), aFPA (n=12).	IFPUG (n=13), COSMIC (n=13), aFPA (n=6).
Duración	Actividad en casa (2 semanas cada uno, 20% de la nota).	Actividad en casa (1 semana cada uno, 20% de la nota).	Actividad en casa (2 semanas cada uno, 20% de la nota).	Actividad en casa (1 semana cada uno, FPA 20%, FFP 20%, aFPA 10%).
Métricas	Reproducibilidad, Exactitud, Tiempo, Facilidad de uso percibida, Utilidad percibida, Intención de uso	Reproducibilidad, Exactitud, Tiempo, Facilidad de uso percibida, Utilidad percibida, Intención de uso	Reproducibilidad, Exactitud, Tiempo, Facilidad de uso percibida, Utilidad percibida, Intención de uso	Reproducibilidad, Exactitud, Tiempo, Facilidad de uso percibida, Utilidad percibida, Intención de uso
Dominio de la aplicación	Transaccional (FPA [U: 101 UFP], AFP [U: 115 UFP]).	Transaccional (FPA [U: 144 UFP], AFP [U: 115 UFP]).	Transaccional (AFP [U: 164 UFP], aFPA [P: 152 UFP]).	Transaccional (FPA [R: 146, I: 118 UFP], FFP [R: 118, I: 111 FP] aFPA [152 UFP]).
Capacitación	Dos sesiones por método (3 horas cada una).	Una sesión por método (1 hora cada una).	Una sesión por método (2 horas AFP y 1 hora aFPA).	Dos sesiones método FPA (3 horas cada una). Una sesión FFP (3 horas). No se provee capacitación para aFPA.
Materiales y guías	Manual FPA, Guía AFP, Guía rápida FPA, Formularios de recolección de datos.	Manual FPA, Manual AFP, Guía rápida y manual FPA, Formularios de recolección de datos.	Manual AFP, Procedimiento aFPA, Formularios de recolección de datos.	Manual FPA, Manual FFP, Guía rápida FPA, FFP, Procedimiento aFPA, Formularios de recolección de datos.
Aplica verificación	Investigador (n=14).	Participantes (actividad en casa, n=24, 5% de la nota).	Participantes (actividad en casa, n=21, 5% de la nota).	Participantes (actividad en casa, n=13, 10% de la nota), Investigador (n=6).
Durante el experimento.	Presentación de ejercicio por método (15 minutos cada una).	Presentación de ejercicio por método (15 minutos cada una).	Presentación de ejercicio por método (15 minutos cada una).	Presentación de ejercicio por método (15 minutos cada una).
Nivel participantes	Estudiantes de maestría (ESE).	Estudiantes de maestría (Métricas).	Estudiantes de maestría (ESE).	Estudiantes de maestría (SEE, ESE).

6.2.1.2. HIPÓTESIS Y SELECCIÓN DE VARIABLES

La variable independiente es el proceso de medición (FPA, FFP, AFP, **aFPA**) utilizado por los profesionales para medir los artefactos de software de las aplicaciones transaccionales. Las variables dependientes son las medidas relacionadas con la efectividad (tamaño funcional, exactitud, reproducibilidad, tiempo de medición) y las relacionadas con la adopción percibida (facilidad de uso, utilidad, intención de uso). En cada corrida, las siguientes métricas son recolectadas:

- Tamaño funcional (cuantitativa, escala intervalo): el tamaño funcional obtenido al aplicar el proceso de medición al objeto. Este recolecta los puntos de función totales y el detalle de los puntos de función para cada uno de los componentes funcionales básicos.
- Exactitud: (cuantitativa, escala razón): la proximidad de los resultados entre un valor medido y el valor verdadero (utilizado como medida de control). Para evaluar la exactitud se utiliza la métrica MRE que se calcula como la diferencia de los valores absolutos entre el conteo producido por un sujeto y el valor verdadero, relativo al valor verdadero. Esta métrica se calcula a partir del tamaño funcional reportado por el sujeto y el tamaño funcional de control.
- Reproducibilidad (cuantitativa, razón): la proximidad de los resultados entre las mediciones sucesivas del mismo objeto por distintos sujetos. Para evaluar el grado de variación entre sujetos se utiliza la métrica de reproducibilidad que se calcula como las diferencias de los valores absolutos entre los conteos producidos por un sujeto y el promedio de los conteos producidos por los otros sujetos de la muestra, relativo al promedio del conteo de los otros sujetos. Esta métrica se calcula a partir del tamaño funcional reportado por los sujetos.
- Tiempo de medición (cuantitativa, escala razón): el esfuerzo real en horas y minutos que le toma a un participante realizar el proceso de medición. Esta medición es reportada por los participantes.

- Facilidad de uso percibida (cualitativa, escala ordinal): el grado en que un sujeto percibe que usar el proceso de medición particular no es una tarea compleja. Este constructo mide la percepción sobre el esfuerzo requerido para aprender y usar el proceso de medición. Los ítems fueron formulados como una escala de Likert de cinco puntos usando un formato con preguntas de declaraciones opuestas. Este dato es recolectado mediante un instrumento de encuesta que los participantes completan.
- Utilidad percibida (cualitativa, escala ordinal): el grado en que un sujeto percibe que usar el proceso de medición particular es efectivo para lograr el objetivo para el cual fue creado. Este constructo mide la percepción sobre la utilidad del proceso de medición. Los ítems fueron formulados como una escala de Likert de cinco puntos usando un formato con preguntas de declaraciones opuestas. Este dato es recolectado mediante un instrumento de encuesta que los participantes completan.
- Intención de uso futuro (cualitativa, escala ordinal): el grado en que una persona tiene la intención de utilizar un proceso de medición particular. Este constructo mide la intención de uso a partir de la percepción sobre el desempeño del proceso de medición. Los ítems fueron formulados como una escala de Likert de cinco puntos usando un formato con preguntas de declaraciones opuestas. Este dato es recolectado mediante un instrumento de encuesta que los participantes completan.

Las ecuaciones de las métricas de efectividad se detallan en la Sección 5.2.4. Estas se basan en las recomendaciones presentadas en (Abrahamo, 2004; Marín, Condori-Fernández & Pastor, 2008). Asimismo, el modelo de adopción utilizado se fundamenta en los estudios propuestos por (Abrahamo, 2004; Abrahamo, Poels & Pastor, 2004a) que se encuentran basados en los modelos de aceptación de tecnologías (Davis, 1989; Moody, 2001).

Para comparar los procesos de medición se plantea un conjunto de hipótesis que evalúan la efectividad y las propiedades de adopción de los procesos de medición. Asimismo, analizan la relación entre los constructos. A continuación se detalla cada una de las preguntas de investigación y sus respectivas hipótesis.

Para la pregunta de investigación 1 se desea probar si existe diferencia entre la efectividad de las mediciones de los procesos de medición (P_i), de la siguiente manera:

RQ1. *¿Cuál es la efectividad de las mediciones realizadas por los profesionales aplicando los procesos de medición?*

H_{0A} : *no existe diferencia entre la exactitud de las mediciones del P_i y el P_j*

H_{0R} : *no existe diferencia entre la reproducibilidad de las mediciones del P_i y el P_j*

Donde los procesos de medición P_i y P_j toman los valores de una de las variables independientes (FPA, FFP, AFP, **aFPA**). La exactitud y la reproducibilidad se comparan tanto para los valores totales de los puntos de función como para los conteos de los componentes funcionales básicos.

Para la pregunta de investigación 2 se desea probar si existe diferencia entre las propiedades de adopción que perciben los profesionales al aplicar los procesos de medición (P_i). Asimismo, se analiza la relación entre los constructos (C_i), de la siguiente manera:

RQ2. *¿Cuáles son las propiedades de adopción de los procesos de medición percibidas por los profesionales?*

H_{0F} : *no existe diferencia entre la facilidad de uso percibida del P_i y el P_j*

H_{0U} : *no existe diferencia entre la utilidad percibida del P_i y el P_j*

H_{0I} : *no existe diferencia entre la intención de uso del P_i y el P_j*

H_{0C} : *el C_i no está relacionado con el C_j*

Donde los procesos de medición P_i y P_j toman los valores de una de las variables independientes (FPA, FFP, AFP, **aFPA**). Los constructos C_i y C_j toman los valores de las propiedades de adopción para analizar sus relaciones.

Los procedimientos de análisis y las pruebas realizadas para contestar cada una de las preguntas de investigación y sus respectivas hipótesis se detallan en la Sección 6.2.2.2 de análisis e interpretación de los resultados.

6.2.1.3. INSTRUMENTACIÓN

Los instrumentos utilizados en cada uno de los estudios fueron preparados previamente y validados mediante corridas piloto. La instrumentación incluye los objetos experimentales, los materiales de las capacitaciones y la encuesta de salida. El detalle de los instrumentos utilizados es la siguiente:

- **Objetos experimentales:** los objetos utilizados son los distintos artefactos de software de cuatro aplicaciones transaccionales: Universidad, Proyectos, Registro e Inventario. Estas aplicaciones tienen tamaños que varían entre 101 y 164 puntos de función lo que las clasifica como aplicaciones pequeñas (S). La complejidad de las funcionalidades de las aplicaciones es similar. Los artefactos de software son la especificación de requerimientos bajo el estándar IEEE 830 y el código fuente del sistema desarrollado en *.Net C#* bajo la arquitectura *MVC*. Cada uno de los artefactos utilizados en los estudios se describe posteriormente.
- **Material de capacitación:** el material de capacitación incluye las presentaciones sobre los procesos de medición y los ejemplos de medición utilizados en las sesiones de entrenamiento. Asimismo, los ejercicios básicos para resolver como ejemplos en la capacitación. Para cada uno de los procesos de medición se proveen las guías y manuales de medición. En el caso del método de medición IFPUG FPA se provee el manual oficial de conteo versión 4.3, el material de capacitación de *Software Metrics*¹⁷ y la guía rápida de *Total Metrics*¹⁸. Para el método de medición COSMIC FFP se provee el manual oficial de

¹⁷ *Software Metrics, Function Points Analysis Training Course*

¹⁸ *Total Metrics, Function Point Basics IFPUG CPM 4.3*

conteo versión 4.0.2, los lineamientos para el conteo de aplicaciones de negocios versión 1.3.1 y la guía rápida para aplicaciones de negocios 1.0.1. En el caso de AFP se provee el manual de conteo versión 1.0. Finalmente, para el procedimiento de medición de la metodología **aFPA** se proveen los lineamientos con las reglas de aplicación y de asignación numérica.

- Instrumentos de recolección de datos: para cada una de las corridas se proveen los formularios de recolección de datos de las mediciones y la encuesta de salida para recolectar los datos de percepción de los participantes. La encuesta de salida se compone de las preguntas de datos demográficos, las 14 preguntas cerradas de adopción ordenadas aleatoriamente donde la mitad de ellas son la negación de otra pregunta para reducir el riesgo de obtener preguntas monótonas. El instrumento de encuesta se describe posteriormente.

Los objetos de estudio utilizados para medir cada una de las aplicaciones transaccionales (Universidad, Proyectos, Registro e Inventario) son los artefactos de software requeridos como entrada para realizar cada uno de los procesos de medición. Las aplicaciones transaccionales medidas implementan funcionalidades de las operaciones básicas de creación, lectura, actualización, borrado y listado de información (CRUDL). Los objetos de estudio utilizados han sido utilizados en la literatura para ejemplificar procesos de medición: Universidad fue utilizado en (Quesada-López, Madrigal & Jenkins, 2016), Proyectos fue detallado en (Abrahamo, 2004), Registro fue descrito en (Bundschuh & Dekkers, 2008; COSMIC, 2015) y finalmente, Inventario fue desarrollado en (Fetcke, 1999). Las aplicaciones tienen un tamaño en puntos de función que varía entre 101 y 164 UFP y entre 111 y 118 FP. La Tabla 6-16 detalla los objetos experimentales y cada uno de los artefactos utilizados en cada una de las corridas de los estudios empíricos:

Tabla 6-16 Objetos de estudio.

Aplicación	Objeto	Descripción	Tamaño	Medición
Universidad	A	Requerimientos IEEE 830. Modelo de clases y ER.	101 UFP	FPA
Universidad	B	Código fuente. <i>Net C# MVC</i> .	115 UFP	AFP
Universidad	C	Requerimientos IEEE 830.	144 UFP	FPA
Universidad	D	Código fuente. <i>Net C# MVC</i> . Requerimientos IEEE 830.	164 UFP	AFP
Proyectos	E	Código fuente. <i>Net C# MVC</i> . Requerimientos IEEE 830.	152 UFP	aFPA
Registro	F	Requerimientos IEEE 830.	146 UFP, 118 FP	FPA, FFP
Inventario	G	Requerimientos IEEE 830.	118 UFP, 111 FP	FPA, FFP
Proyectos	H	Requerimientos IEEE 830.	152 UFP	aFPA

El instrumento de la encuesta de salida tiene la siguiente estructura:

- Datos demográficos: se consultan los datos de la experiencia profesional en computación e informática y los conocimientos en procesos de medición del tamaño funcional. Los participantes seleccionan entre las categorías de experiencia >10 años, de 6-9 años, de 2-5 años, <2 años. El conocimiento en los métodos de medición es seleccionado entre las categorías de conocimiento (N) no lo conocía, (T) conocía la teoría, (A) lo ha aplicado en la universidad, (P) lo ha aplicado en pilotos en su organización y (I) se encuentra implementado en su organización.
- Preguntas cerradas de adopción: catorce preguntas¹⁹ para medir la facilidad de uso (preguntas 1, 2, 4, 6, 9), utilidad percibida (preguntas 3, 5, 8, 10, 11) y la intención de uso (preguntas 7, 12, 13, 14).
- Preguntas abiertas sobre el proceso de medición: las preguntas abiertas se relacionan con las preguntas de adopción para conocer la opinión y sugerencias de los participantes con respecto a los procesos de medición.

¹⁹ <http://bit.ly/2LmlynA>

6.2.1.4. SELECCIÓN DE SUJETOS DE ESTUDIO

Los participantes fueron seleccionados a conveniencia. Estos son profesionales de la industria de software costarricense tomando cursos de Maestría en Computación e Informática de la UCR. Los participantes fueron informados que los datos recolectados serían utilizados para propósitos de la investigación y tratados de manera confidencial. Se mantiene la privacidad de los participantes. La selección de los sujetos se da por su accesibilidad durante cursos de la maestría. Las asignaciones fueron obligatorias como parte de las actividades evaluadas del curso. Cada una de las actividades de medición representa entre un 10% y un 20% del total de la nota del curso. Las actividades se realizaron al inicio del curso después de recibir los conceptos de medición. Las asignaciones fueron realizadas en casa con una duración que varía entre una y dos semanas, en el caso de la Replicación-3 la actividad se inicia en clase y finaliza en casa. Los participantes no conocen las preguntas de investigación ni los objetivos del estudio.

Los participantes para cada uno de los estudios son los siguientes:

- Estudio original (E2-14): catorce estudiantes de maestría matriculados en un curso de Ingeniería de Software Experimental (ESE) durante el segundo periodo del 2014.
- Replicación-1 (E1-15): catorce estudiantes de maestría matriculados en un curso de Métricas de software durante el primer periodo del 2015.
- Replicación-2 (E2-15): doce estudiantes de maestría matriculados en un curso de Ingeniería de Software Experimental (ESE) durante el segundo periodo del 2015.
- Replicación-3 (E1-18): trece estudiantes de maestría matriculados en un curso de Estimación de Software (SEE) y siete estudiantes de maestría matriculados en un curso de Ingeniería de Software Experimental (ESE) durante el primer periodo del 2018.

Al completar las actividades de medición, los participantes completaron una encuesta de opinión en la que indican sus datos demográficos relacionados con su experiencia académica y profesional. Los participantes no son expertos en métodos de medición del

tamaño funcional pero cuentan con experiencia en las prácticas de la ingeniería del software. La Tabla 6-17 resume los datos demográficos de los profesionales que participaron en el estudio original y en cada una de las replicaciones. Se detalla la cantidad de participantes para cada uno de los estudios indicando su experiencia y conocimiento de los métodos de medición. Se resaltan las frecuencias máximas por grupo. Los participantes reportan que cuentan con una experiencia profesional (EP) y conocimiento de los métodos de medición del tamaño funcional (CMF) similar para cada uno de los estudios.

El Anexo 4 detalla el análisis de las amenazas a la validez de los estudios descritos en esta evaluación. Las amenazas a la validez son las relacionadas con los diseños experimentales, los experimentos orientados a personas, los sujetos y los objetos de estudio.

Tabla 6-17 Frecuencia de participantes por estudio.

Estudio	G	N	M	EP	n	CMF	n	G	N	M	EP	n	CMF	n	
E2-14	G1	7	FPA	>10	1	P	1	G2	7	AFP	>10	0	P	0	
				6-9	0	A	1				6-9	0	A	0	
				2-5	0	T	3				2-5	0	T	1	
				<2	6	N	2				<2	7	N	6	
				NR	0	NR	0				NR	0	NR	0	
E1-15	G1	12	FPA	>10	0	P	0	G2	12	AFP	>10	0	P	0	
				6-9	0	A	3				6-9	0	A	3	
				2-5	2	T	2				2-5	2	T	2	
				<2	10	N	6				<2	10	N	7	
				NR	0	NR	1				NR	0	NR	0	
E2-15	G1	12	AFP	>10	1	P	0	G2	12	aFPA	>10	1	P	0	
				6-9	0	A	2				6-9	0	A	2	
				2-5	2	T	1				2-5	2	T	4	
				<2	7	N	7				<2	7	N	4	
				NR	2	NR	2				NR	2	NR	2	
E1-18	G1	13	FPA	>10	0	P	0	G2	13	FFP	>10	0	P	0	
				6-9	1	A	1				6-9	1	A	0	
				2-5	7	T	0				2-5	7	T	1	
				<2	5	N	11				<2	5	N	11	
				NR	0	NR	1				NR	0	NR	1	
	G3	7	aFPA	>10	0	P	0								
				6-9	0	A	0								
				2-5	2	T	1								
				<2	4	N	6								
				NR	1	NR	0								

Grupo (G), Método (M), Experiencia profesional en años (EP), Conocimiento medición funcional (CMF). NR=No reporta, N=No lo conocía, T=conocía la teoría, A=lo ha aplicado en la universidad, P= lo ha aplicado en pilotos en la organización. N=participantes, n=participantes por categoría.

6.2.2. ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

En esta sección se presentan los resultados de este conjunto de estudios empíricos. La Tabla 6-18 presenta el detalle de las estadísticas descriptivas de los resultados de medición del tamaño funcional por cada una de las corridas indicando los procesos de medición utilizados (Tratamiento), la cantidad de sujetos (N) y los valores verdaderos de referencia (Control). Después de validar los resultados reportados por cada participante, para la Replicación-2 (E2-15) y el proceso de medición AFP no se consideran dos de las mediciones

por estar incompletas. Del mismo modo, en la Replicación-3 (E1-18) y el proceso de medición aFPA no se considera una de las mediciones por estar incompleta.

Los datos de las corridas muestran que los conjuntos de datos tienen un *skewness* que se encuentra en el rango de ± 2 y en el caso del *kurtosis* este se encuentra en el rango de ± 3 . Las pruebas de normalidad *Kolmogorov-Smirnov* y *Shapiro-Wilk* indican que las mediciones en las distintas corridas pertenecen a una distribución normal. La excepción se presenta en la corrida E1-18-FFP-7 donde $D(7) = 0.312, p < 0.038$ lo que indica que las mediciones se desvían de una distribución normal. La Figura 6-12 presenta los diagramas de cajas con los datos del tamaño funcional para cada uno de los estudios y sus respectivas corridas. En este se pudo observar la variación obtenida en cada corrida. Asimismo, se indica una marca con el valor verdadero (de control) a la izquierda de cada gráfico para referencia. Los resultados de medición varían dependiendo de las características específicas de cada corrida, los procesos de medición utilizados y los objetos experimentales medidos. En las secciones siguientes se analizan cada uno de estos factores y se responden las preguntas de investigación. Asimismo, se procesan y analizan los datos para realizar las pruebas planteadas en el diseño.

Tabla 6-18 Estadística descriptiva de los resultados de medición del tamaño funcional.

Estudio	Tratamiento	N	Control	Prom	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rango	Std.D.	Skew	Kurt
E2-14	FPA	7	101	95	85	85	85	86	96	101	-	-	104	19	7.35	-0.63	-1.15
E2-14	AFP	7	115	108	91	91	91	102	112	118	-	-	119	28	10.37	-0.72	-0.62
E1-15	FPA	12	144	137	93	93	93	99	141	165	185	-	188	95	34.11	-0.07	-1.50
E1-15	AFP	12	115	107	78	78	82	92	104	118	144	-	149	71	19.49	0.81	0.77
E2-15	AFP	9	164	143	113	113	113	138	146	150	-	-	164	51	13.95	-1.10	2.73
E2-15	aFPA	12	152	165	153	153	155	162	165	169	176	-	177	24	6.33	0.03	0.71
E1-18	FPA	6	146	142	121	121	121	133	139	155	-	-	165	44	14.99	0.32	0.48
E1-18	FPA	7	118	117	111	111	111	114	118	120	-	-	122	11	3.73	-0.69	-0.01
E1-18	FFP	6	111	102	98	98	98	98	101	107	-	-	111	13	5.01	1.18	0.91
E1-18	FFP	7	118	113	107	107	107	108	110	116	-	-	130	23	8.04	2.02	4.18
E1-18	aFPA	6	152	155	140	140	140	147	154	168	-	-	169	29	11.36	0.01	-1.55

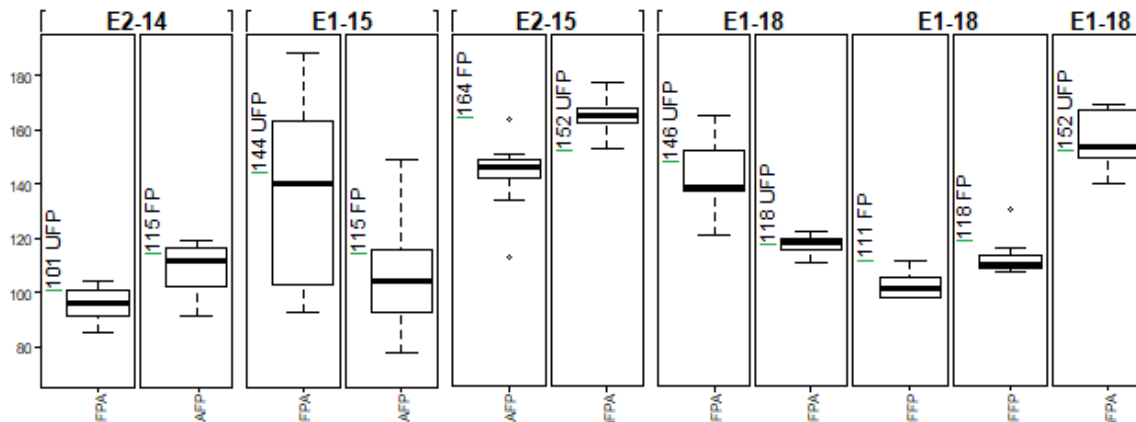


Figura 6-12 Resultados de las mediciones de tamaño por estudio.

6.2.2.1. EFECTIVIDAD DE LOS RESULTADOS DE LOS PROCESOS DE MEDICIÓN

RQ1. *¿Cuál es la efectividad de las mediciones realizadas por los profesionales aplicando los procesos de medición?*

El objetivo del análisis es estudiar la efectividad, en términos de la exactitud y reproducibilidad, de los resultados de medición obtenidos por los profesionales. La Tabla 6-19 detalla los resultados de las métricas de exactitud y reproducibilidad para cada una de las corridas indicando los procesos de medición utilizados (Tratamiento), la métrica (Var) y la cantidad de sujetos (N). En general, los mejores valores de las métricas de exactitud y reproducibilidad son las que tienden a cero.

La Figura 6-13 detalla los resultados de las métricas de exactitud y reproducibilidad mediante los histogramas con la distribución de cada uno de las corridas realizadas en los estudios. Cada uno de los histogramas se presenta por proceso de medición aplicado. La línea punteada muestra la mediana y la línea continua muestra el valor verdadero de control. La distancia entre la media y el valor de control indica la exactitud de las estimaciones de tamaño. Si los valores de una corrida se encuentran cercanos a la media, indica que existe una precisión en las mediciones realizadas por los participantes (reproducibilidad).

Los resultados muestran algunas variaciones dependiendo de la configuración de cada una de las corridas, las cuales se discuten a continuación.

Tabla 6-19 Exactitud (R) y reproducibilidad (R) de los resultados de medición.

Estudio	Tratamiento	Var	N	Prom	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rango	Std.D.	Skew	Kurt
E2-14	FPA	E	7	6%	0%	0%	0%	1%	5%	15%	-	-	16%	16%	6%	0.89	-1.02
		R		7%	1%	1%	1%	1%	7%	11%	-	-	13%	12%	5%	-0.31	-1.62
E2-14	AFP	E	7	7%	0%	0%	0%	3%	3%	11%	-	-	21%	21%	7%	1.13	0.76
		R		9%	4%	4%	4%	7%	7%	12%	-	-	18%	14%	5%	1.19	1.52
E1-15	FPA	E	12	21%	3%	3%	5%	11%	19%	34%	-	-	35%	32%	11%	0.09	-1.43
		R		23%	3%	3%	5%	14%	22%	33%	40%	-	-	42%	39%	12%	-0.15
E1-15	AFP	E	12	15%	3%	3%	3%	5%	14%	22%	31%	-	32%	29%	10%	0.48	-0.86
		R		15%	2%	2%	2%	4%	12%	24%	40%	-	44%	42%	13%	1.17	1.00
E2-15	AFP	E	9	13%	0%	0%	0%	9%	11%	16%	-	-	31%	31%	8%	1.15	2.92
		R		7%	1%	1%	1%	2%	5%	12%	-	-	23%	22%	8%	1.46	1.20
E2-15	aFPA	E	12	9%	1%	1%	2%	7%	9%	11%	-	-	16%	15%	4%	-0.08	0.68
		R		3%	0%	0%	0%	1%	2%	5%	8%	-	8%	8%	3%	0.87	-0.52
E1-18	FPA	E	6	8%	4%	4%	4%	4%	6%	14%	-	-	17%	13%	5%	1.08	-0.55
		R		9%	2%	2%	2%	4%	7%	18%	-	-	20%	18%	8%	0.68	-1.69
E1-18	FPA	E	7	2%	0%	0%	0%	1%	2%	3%	-	-	6%	6%	2%	1.08	1.44
		R		3%	0%	0%	0%	1%	3%	5%	-	-	6%	6%	2%	0.26	-0.79
E1-18	FFP	E	6	8%	0%	0%	0%	4%	9%	12%	-	-	12%	12%	5%	-1.03	0.33
		R		4%	0%	0%	0%	2%	4%	6%	-	-	10%	10%	3%	0.79	1.70
E1-18	FFP	E	7	7%	2%	2%	2%	6%	8%	9%	-	-	10%	8%	3%	-1.40	2.46
		R		6%	2%	2%	2%	3%	4%	6%	-	-	18%	16%	6%	2.33	5.71
E1-18	aFPA	E	6	6%	2%	2%	2%	2%	6%	10%	-	-	11%	9%	4%	0.09	-2.51
		R		7%	2%	2%	2%	4%	7%	11%	-	-	12%	10%	4%	-0.11	-1.77

E=Exactitud (MRE). R=Reproducibilidad.

En el estudio E2-14, el promedio de la exactitud es de un 6% con una reproducibilidad del 7% utilizando IFPUG FPA con una aplicación de 101 UFP (n=7). En el caso de la medición utilizando OMG AFP con una aplicación de 115 FP (n=7), la exactitud es de un 7% con una reproducibilidad del 9%. Los resultados indican que la efectividad de las mediciones de los participantes en el estudio E2-14 utilizando los dos procesos de medición sobre los artefactos de requerimientos y el código fuente de la aplicación es similar.

En el estudio E1-15, el promedio de la exactitud es de un 21% con una reproducibilidad del 23% utilizando IFPUG FPA con una aplicación de 144 UFP (n=12). En el caso de la medición utilizando OMG AFP con una aplicación de 115 FP (n=12), la exactitud es de un 15% con una reproducibilidad del 15%. En el estudio E2-14, los resultados indican que la efectividad de los participantes utilizando el proceso de OMG AFP presenta diferencias de 7% de exactitud y de 9% de reproducibilidad. Un factor a considerar que puede influenciar los resultados es la diferencia en el tamaño de las aplicaciones medidas. Del mismo modo, entre el estudio E2-14 y E1-15, las variaciones entre los grupos pueden presentarse por el tamaño de la aplicación utilizada para IFPUG FPA (de 101 a 144 UFP), los artefactos utilizados para la medición que incluyen o no modelos de la aplicación, las variaciones en los procesos de capacitación (seis horas por proceso de medición en el estudio E2-14, una hora por proceso en el estudio E1-15) y la cantidad de sujetos.

Para el estudio E2-15, el promedio de la exactitud es de un 13% con una reproducibilidad del 7% utilizando OMG AFP con una aplicación de 164 UFP (n=9). En el caso de la medición utilizando el procedimiento de medición de la metodología **aFPA** con una aplicación de 152 UFP (n=12), la exactitud es de un 9% con una reproducibilidad del 3%. En este caso, las mediciones OMG AFP mantienen resultados similares a los reportados por los estudios anteriores. En este proceso de medición, el tamaño de la aplicación se incrementa (115 UFP a 164 UFP) y se incrementa la duración de las actividades de capacitación (dos horas para OMG AFP). En el caso del procedimiento de medición de la metodología **aFPA** se alcanzan buenos resultados proporcionando una hora de capacitación y midiendo una la aplicación de 152 UFP. En este caso, la efectividad puede explicarse por las reglas del procedimiento que toman decisiones para garantizar consistencia (Reproducibilidad=3%) lo que afecta el resultado de la exactitud contra los valores verdaderos (Exactitud=9%).

En el estudio E1-18, el promedio de la exactitud es de un 8% con una reproducibilidad del 9% utilizando IFPUG FPA con una aplicación de 146 UFP (n=6). En el caso de la medición de una aplicación de 118 UFP (n=7), la exactitud es de un 2% con una reproducibilidad del 3%. En este caso, las mediciones presentan mejoras para la aplicación de 118 UFP. En el caso del método COSMIC FFP, el promedio de la exactitud es de un 8% con una

reproducibilidad del 4% con una aplicación de 111 FP (n=6) y en el caso de la medición de una aplicación de 118 FP (n=7), la exactitud es de un 7% con una reproducibilidad del 6%. Los procesos de capacitación realizados son de seis horas para IFPUG FPA y de tres horas para COSMIC FFP. Finalmente, en el caso de la medición utilizando el procedimiento de medición de la metodología **aFPA** con una aplicación de 152 UFP (n=6), la exactitud es de un 6% con una reproducibilidad del 7%. Estos resultados son similares a los reportados en el estudio E2-15, las variaciones en la reproductividad podrían estar asociadas a la utilización de la extensión del procedimiento para documentos de requerimientos y a las diferencias en la cantidad de sujetos. En el caso del procedimiento de medición de la metodología **aFPA**, las sesiones de capacitación fueron introductorias o del todo omitidas (menores a 15 minutos). Los resultados de efectividad indican que bajo las condiciones descritas, es posible obtener mediciones aceptables utilizando el proceso de medición de la metodología **aFPA** con el beneficio de que este procedimiento puede ser automatizado.

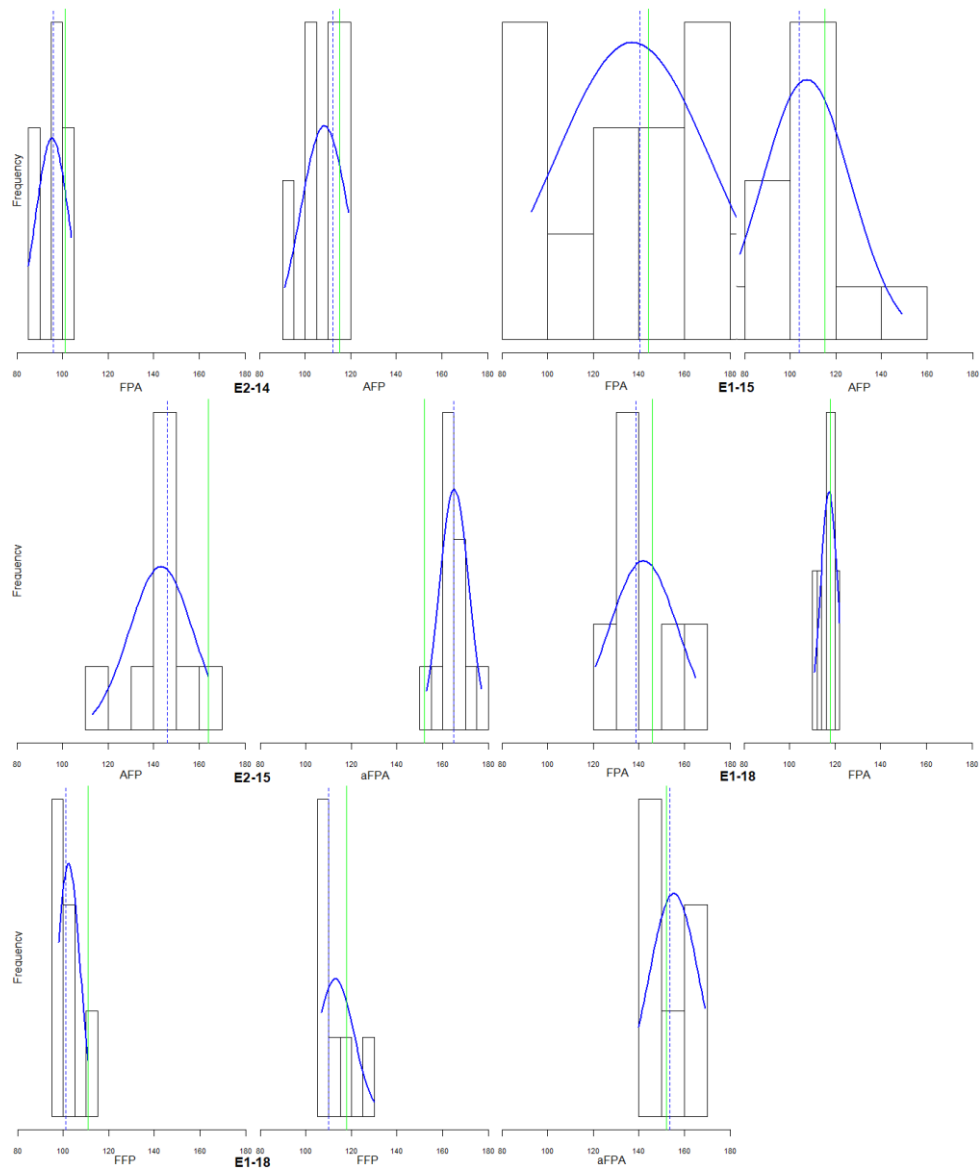


Figura 6-13 Exactitud y reproducibilidad de los resultados de medición por estudio.

Los resultados obtenidos muestran que, en promedio y bajo las condiciones descritas, los profesionales sin experiencia previa en medición pueden obtener resultados de tamaño funcional con una exactitud y una reproducibilidad menor al $\pm 10\%$, que son consideradas satisfactorias para un sistema de medición (AIAG, 2002; Abran, 2015). En algunos

casos, las mediciones se encuentran entre el 10-30% lo cual indica que el sistema de medición podría ser utilizado dependiendo de las necesidades de la organización y el costo de obtener y mantener las mediciones.

A continuación se realiza el análisis de los resultados agregados obtenidos para cada uno de los procesos de medición IFPUG FPA, OMG AFP y **aFPA**. Asimismo, se analizan los resultados del proceso de medición utilizando COSMIC FFP. Es de interés realizar el análisis de los resultados agregados para estudiar las diferencias obtenidas entre los procesos de medición para la métrica de exactitud y reproducibilidad.

Las pruebas de normalidad *Kolmogorov-Smirnov* y *Shapiro-Wilk* indican que la mayoría de las mediciones agregadas por proceso de medición pertenecen a una distribución normal incluyendo las del procedimiento de medición de la metodología **aFPA**. Sin embargo, los resultados del proceso de medición IFPUG FPA, $D(32) = 0.214, p < 0.001$ se desvían de una distribución normal para la exactitud. Lo mismo se presenta para la métrica de reproducibilidad en el caso del proceso de medición OMG AFP, $D(28) = 0.225, p < 0.001$ y COSMIC FFP donde $D(13) = 0.283, p < 0.005$. Considerando las diferencias presentadas y dependiendo de las configuraciones de los distintos estudios, los análisis de los resultados agregados deben ser interpretados cuidadosamente.

La Tabla 6-20 detalla los resultados de las métricas de exactitud y reproducibilidad para cada una de los procesos de medición utilizados (Tratamiento), las métricas (Var) y la cantidad de sujetos (N). La Figura 6-14 detalla los resultados de las métricas de exactitud y reproducibilidad para cada uno de los procesos de medición. El promedio de la exactitud es de un 11% con una reproducibilidad del 13% utilizando IFPUG FPA con aplicaciones de 101 UFP a 146 UFP (n=32). En el caso de la medición utilizando OMG AFP con aplicaciones de 115 UFP a 164 UFP (n=28), la exactitud es de un 12% con una reproducibilidad del 11%. El promedio de la exactitud es de un 8% con una reproducibilidad del 4% utilizando el procedimiento de medición de la metodología **aFPA** con aplicaciones de 152 UFP (n=18). Finalmente, en el caso de la medición utilizando COSMIC FFP con aplicaciones de 111 FP a 118 FP (n=13), la exactitud es de un 7% con una reproducibilidad del 5%. En este caso el procedimiento de medición de la metodología **aFPA** cuenta con soporte para la medición de los

puntos de función COSMIC FFP por lo que como trabajo futuro se desea evaluar con profesionales la efectividad de su aplicación para obtener este tipo de mediciones.

En cuanto a la productividad de los procesos de conteo, el método IFPUG FPA reporta en promedio 26 puntos de función por hora persona (UFP/hp), con una mediana (Md) de 23 UFP/hp y una desviación estándar (SD) de 23 UFP/hp (n=26). Los procesos de conteo de OMG AFP reportan en promedio 20 UFP/hp (Md=17, SD=15, n=23). En el caso de los procesos de conteo del procedimiento **aFPA**, estos reportan en promedio 36 UFP/hp (Md=25, SD=28, n=8). Finalmente, los procesos de conteo del método COSMIC FFP reportan en promedio 30 FP/hp (Md=30, SD=8, n=11). Los datos de esfuerzo en tiempo son recolectados por los participantes durante la aplicación manual de cada uno de los procesos de medición. Como valor de referencia se reporta que un experto en puntos de función puede contar hasta 50 FP/hp (Jones, 2008). Los tiempos reportados indican la importancia de contar con soporte para la automatización de las mediciones que pueda significar mejoras en la productividad.

Tabla 6-20 Exactitud y reproducibilidad de los resultados agregados.

Tratamiento	Var	N	Prom	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rango	Std.D.	Skew	Kurt
FPA	E	32	11%	0%	0%	1%	3%	6%	16%	34%	35%	35%	35%	11%	1.12	0.25
	R		13%	0%	1%	1%	3%	9%	20%	33%	37%	42%	42%	12%	1.00	0.02
AFP	E	28	12%	0%	0%	3%	4%	11%	18%	30%	32%	32%	32%	9%	0.76	-0.08
	R		11%	1%	1%	2%	4%	7%	17%	26%	37%	44%	43%	10%	1.66	3.15
aFPA	E	18	8%	1%	1%	2%	4%	9%	10%	14%	-	16%	15%	4%	-0.01	-0.28
	R		4%	0%	0%	0%	1%	4%	8%	11%	-	12%	12%	4%	0.68	-0.66
FFP	E	13	7%	0%	0%	1%	6%	8%	10%	12%	-	12%	12%	4%	-0.82	0.39
	R		5%	0%	0%	1%	3%	4%	6%	15%	-	18%	18%	5%	2.15	5.50

E=Exactitud (MRE). R=Reproducibilidad.

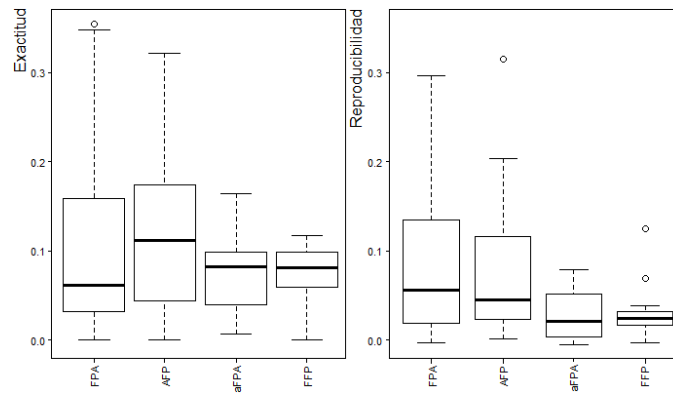


Figura 6-14 Exactitud y reproducibilidad de los resultados agregados.

Para comparar los procesos de medición se plantean dos hipótesis que evalúan la efectividad de los procesos de medición. Se desea inicialmente probar si existe diferencia entre la efectividad de las mediciones de los procesos de medición IFPUG FPA, OMG AFP y **aFPA**. Posteriormente, se comparan con el método COSMIC FFP. La prueba estadística utilizada es *Kruskal-Wallis H*, que es la alternativa no paramétrica a la prueba de ANOVA. Esta prueba permite determinar si existen diferencias significativas entre dos o más grupos para variables continuas u ordinales. Esta es la extensión de la prueba *Mann-Whitney U* para más de dos grupos. La primera hipótesis compara los resultados de la exactitud entre los procesos de medición:

$$H_{0A}: \text{no existe diferencia entre la exactitud de las mediciones del } P_i \text{ y el } P_j$$

Los resultados de la prueba de hipótesis *Kruskal-Wallis H* para los datos agregados indican que no existe una diferencia significativa en los resultados de exactitud de los procesos de medición $\chi^2(3) = 3.138, p = 0.371$, con rangos medios de 45.05 para IFPUG FPA, 52.82 para OMG AFP, 41.50 para el procedimiento de medición de la metodología **aFPA** y 39.88 para COSMIC FFP. Al realizar la prueba *Mann-Whitney U* para comparar cada par de los procesos de medición relacionados con el método IFPUG FPA se obtiene que no existe diferencia significativa en la exactitud de los procesos de medición IFPUG FPA y OMG AFP ($U = 396.5, p = 0.444$) y los procesos de medición IFPUG FPA y **aFPA** ($U = 277.5, p =$

0.832). En el caso de los procesos de medición OMG AFP (Prom=12%, Md=11%, SD=9%, n=28) y **aFPA** (Prom=8%, Md=9%, SD=4%, n=18), aunque no se encuentra una diferencia significativa ($U = 175.0, p = 0.082$) la suma de los rangos es de 735 y 346 en favor del procedimiento de medición de la metodología **aFPA**. Estos resultados indican que es posible obtener resultados de exactitud aceptables utilizando procedimientos de medición que podrían ser automatizados.

La segunda hipótesis compara los resultados de la reproducibilidad entre los procesos de medición:

H_{0R} : *no existe diferencia entre la reproducibilidad de las mediciones del P_i y el P_j*

Los resultados de la prueba de hipótesis *Kruskal-Wallis H* para los datos agregados indican que existe una diferencia significativa en los resultados de reproducibilidad de los procesos de medición $\chi^2(3) = 10.954, p = 0.012$, con rangos medios de 52.88 para IFPUG FPA, 52.13 para OMG AFP, 31.83 para el procedimiento de medición de la metodología **aFPA** y 35.50 para COSMIC FFP. Al realizar la prueba *Mann-Whitney U* para comparar cada par de los procesos de medición relacionados con el método IFPUG FPA se obtiene que no existe diferencia significativa en la reproducibilidad de los procesos de medición IFPUG FPA y OMG AFP ($U = 431.5, p = 0.807$). Para los procesos de medición IFPUG FPA y **aFPA** ($U = 160.0, p = 0.010$) y OMG AFP y **aFPA** ($U = 138.5, p = 0.010$) se encuentran diferencias significativas que indican que el procedimiento de medición de la metodología **aFPA** presenta mejores resultados en cuanto a la reproducibilidad. Los resultados indican que es posible obtener consistencia en las mediciones utilizando el procedimiento de medición de la metodología **aFPA**.

6.2.2.1.1. VALIDACIÓN DEL PROCEDIMIENTO AFPA

Durante los estudios se realizan dos validaciones para el procedimiento de medición de la metodología **aFPA**. Estas validan la aplicación del procedimiento para medir los artefactos de código fuente y los artefactos de requerimientos mediante la versión extendida del procedimiento presentado en la Sección 5.5.1. La primera validación se realiza en la Replicación-2 (E2-15) donde 12 sujetos miden el código fuente de la aplicación (**aFPA_c**) y la segunda validación se realiza en la Replicación-3 (E1-18) donde 6 sujetos miden la especificación de los requerimientos de la aplicación (**aFPA_r**). Para comparar las dos versiones de los procedimientos se plantea la hipótesis que compara la efectividad de los dos procesos de medición. La prueba estadística utilizada es *Mann-Whitney U* entre los dos grupos. Estas comparan los resultados de la exactitud y la reproducibilidad entre los procesos realizados:

H_{0_A} : no existe diferencia entre los resultados de exactitud del P_i y el P_j

H_{0_R} : no existe diferencia entre los resultados de reproducibilidad del P_i y el P_j

Los resultados de la prueba *Mann-Whitney U* indican que no existe diferencia significativa en la exactitud de los procesos de medición **aFPA_c** y **aFPA_r** ($U = 26.0, p = 0.385$). Caso contrario, la comparación de reproducibilidad indica que existe una diferencia significativa en la reproducibilidad de los procesos de medición **aFPA_c** y **aFPA_r** ($U = 12.0, p = 0.024$) donde el procedimiento aplicado a los artefactos de código fuente (Prom=3%, Md=2%, SD=2%, n=12) presenta mejores resultados que la aplicación en los artefactos de especificación de requerimientos (Prom=7%, Md=7%, SD=4%, n=6). Los resultados indican que es posible obtener resultados similares en la exactitud utilizando ambas versiones del procedimiento. En el caso de la reproducibilidad es posible que las reglas de identificación de transacciones sobre el código fuente sean más específicas que la identificación de elementos en un documento de requerimientos. Por lo tanto, se requiere hacer revisiones sobre las reglas de mapeo.

Es importante considerar algunos de los factores de contexto que variaron en ambas corridas y que podrían haber influido en los resultados. Primero, la medición se realiza sobre artefactos de software distintos. Sumado a esto, los participantes del estudio E2-15 ya contaban con experiencia en procesos de medición con el proceso OMG AFP. Adicionalmente, en la replicación E2-15 se provee una sesión de capacitación de una hora y en la replicación E1-18 no se provee sesión de capacitación. Finalmente, el proceso de medición de la replicación E2-15 se realizó a mediados del curso y el proceso de medición del E1-18 se realizó al final del curso junto con otros proyectos por entregar.

6.2.2.1.2. EFECTIVIDAD DE LOS RESULTADOS DE LOS COMPONENTES FUNCIONALES BÁSICOS

El protocolo de verificación (Quesada-López & Jenkins, 2015b; Quesada-López, Margiral & Jenkins, 2016) indica que no solo se deben analizar los resultados de los totales de puntos de función, sino también los resultados de los componentes funcionales básicos (BFC). En este caso, aunque los resultados de los totales del conteo de tamaño funcional presentan resultados promedio de exactitud y reproducibilidad de menos del $\pm 10\%$, en el caso de los BFC las variaciones presentadas son mayores. La Tabla 6-21 detalla los resultados de la exactitud y reproducibilidad de las funciones transaccionales (TF) y las funciones de datos (DF).

Tabla 6-21 Exactitud y reproducibilidad de los BFC.

Estudio	Tratamiento	BFC	Var	N	Prom	Min	Max	Var	N	Prom	Min	Max
E2-14	FPA	TF	E	7	8%	0%	22%	R	7	11%	1%	19%
		DF	E	7	4%	0%	25%	R	7	7%	4%	25%
E2-14	AFP	TF	E	7	11%	0%	33%	R	7	11%	1%	28%
		DF	E	7	27%	0%	103%	R	7	35%	10%	116%
E1-15	FPA	TF	E	12	14%	1%	28%	R	12	13%	1%	36%
		DF	E	12	65%	0%	100%	R	12	59%	10%	100%
E1-15	AFP	TF	E	12	13%	0%	32%	R	12	15%	0%	34%
		DF	E	12	34%	10%	100%	R	12	42%	5%	100%
E2-15	AFP	TF	E	8	13%	0%	19%	R	8	5%	0%	17%
		DF	E	8	4%	0%	17%	R	8	7%	5%	15%
E2-15	aFPA	TF	E	12	6%	0%	13%	R	12	4%	0%	7%
		DF	E	12	22%	0%	50%	R	12	15%	3%	27%
E1-18	FPA	TF	E	6	11%	1%	22%	R	6	9%	2%	16%
		DF	E	6	15%	6%	38%	R	6	11%	1%	27%
E1-18	FPA	TF	E	6	3%	0%	9%	R	6	4%	1%	9%
		DF	E	6	9%	0%	51%	R	6	16%	9%	51%
E1-18	aFPA	TF	E	6	8%	1%	13%	R	6	10%	4%	13%
		DF	E	6	28%	14%	43%	R	6	6%	1%	14%

Los resultados indican que dependiendo del tamaño de la aplicación los valores calculados para las TF y las DF pueden variar más allá del $\pm 10\%$ esperado. En el caso de los procesos de medición IFPUG FPA, los valores promedio de la exactitud de las TF pueden variar entre 3% y 14%. Para las DF estos valores pueden variar entre 4% y 65%. Para el proceso de medición OMG AFP, los valores para TF pueden variar entre 11% y 13%, y los valores DF entre 4% y 34%. En el caso del procedimiento de medición de la metodología **aFPA** los valores para TF pueden variar entre 6% y 8% y los valores DF pueden variar entre 22% y 28%. Los resultados de reproducibilidad indican que para los procesos de medición IFPUG FPA las TF pueden variar entre 4% y 13%. Para las DF estos valores pueden variar entre 7% y 59%. Para el proceso de medición OMG AFP, los valores para TF pueden variar entre 5% y 15%, y los valores DF entre 7% y 42%. En el caso del procedimiento de medición de la metodología **aFPA** los valores para TF pueden variar entre 4% y 10% y los valores DF pueden variar entre 6% y 15%.

Al realizar el análisis para cada uno de los BFC: entradas (EI), salidas (EO), consultas (EQ), archivos internos (ILF) e interfaces externas (EIF), los resultados de la exactitud y la reproducibilidad presentan variaciones no aceptables. Estos resultados indican la importancia de la verificación detallada de los resultados de las mediciones de tamaño funcional. En el caso de los procesos de medición IFPUG FPA, los valores de la exactitud de las EI pueden variar entre 0% y 47% con un promedio de 13%. Para las EO estos valores pueden variar entre 0% y 195% con un promedio de 71%. Para las EQ los valores pueden variar entre 0% y 131% con un promedio de 44%. Los ILF pueden variar entre 0% y 100% con un promedio de 28% y los EIF entre 0% y 105% con un promedio de 31%. Para el proceso de medición OMG AFP, los valores de la exactitud de las EI pueden variar entre 0% y 69% con un promedio de 15%. Para las EO estos valores pueden variar entre 0% y 100% con un promedio de 26%. Los ILF pueden variar entre 0% y 100% con un promedio de 23%. En el caso del procedimiento de medición de la metodología **aFPA**, los valores para EI pueden variar entre 0% y 23% con un promedio de 10%. Para las EO estos valores pueden variar entre 0% y 120% con un promedio de 37%. Para las EQ los valores pueden variar entre 0% y 46% con un promedio de 13%. Los ILF pueden variar entre 0% y 75% con un promedio de 31%.

El análisis detallado de las diferencias en los resultados de cada uno de los componentes funcionales básicos (BFC) y los tipos de causas asignables que producen estas diferencias se detallan en los estudios presentados en (Quesada-López & Jenkins, 2015b; Quesada-López, Madrigal & Jenkins, 2016; Madrigal-Sánchez, Quesada-López & Jenkins, 2018). Los resultados reportados en estos artículos consideran los datos recolectados en los estudios E2-14, E1-15 y E2-15. Las principales causas asignables se relacionan con la identificación y clasificación de las funciones de datos, la identificación de los grupos lógicos desde la perspectiva del usuario final, la clasificación de las entradas externas (EO) y su relación con las consultas externas (EQ) y finalmente, la identificación de los elementos de tipo dato (DET) desde la perspectiva del usuario final.

En particular, el estudio reportado en (Madrigal-Sánchez, Quesada-López & Jenkins, 2018) realiza una validación del prototipo del protocolo de verificación detallado en la Sección 5.4.3, el cual alcanza una efectividad del 93.92% de exactitud, el 98.69% de precisión y

el 94.94% de *recall* encontrando diferencias en las mediciones. El valor del área bajo la curva es de 0.82, lo que indica que la efectividad de la herramienta es buena. Actualmente se están realizando las corridas para validar la herramienta incorporando los datos del estudio E1-18.

6.2.2.2. PROPIEDADES DE ADOPCIÓN DE LOS PROCESOS DE MEDICIÓN

RQ2. *¿Cuáles son las propiedades de adopción de los procesos de medición percibidas por los profesionales?*

El objetivo del análisis es estudiar la percepción de adopción de los profesionales al aplicar los procesos de medición en términos de la facilidad de uso percibida, la utilidad percibida y la intención de uso. Se realiza la prueba estadística *Kruskal-Wallis H* para determinar si existen diferencias significativas entre dos o más grupos de variables ordinales. La primera hipótesis compara los resultados de la facilidad de uso percibida entre los procesos de medición:

H_{0F} : *no existe diferencia entre la facilidad de uso percibida del P_i y el P_j*

Los resultados de la prueba de hipótesis para los datos agregados indican que existe una diferencia significativa en los resultados de facilidad de uso percibida de los procesos de medición $\chi^2(2) = 11.351, p = 0.003$, con rangos medios de 205.99 para IFPUG FPA, 165.54 para OMG AFP y 179.63 para el procedimiento de medición de la metodología **aFPA**. Al realizar la prueba *Mann-Whitney U* para comparar cada par de los procesos de medición relacionados con el método IFPUG FPA se obtiene que existe diferencia significativa entre los procesos de medición IFPUG FPA y OMG AFP ($U = 8210.5, p = 0.001$). En este caso, el método IFPUG FPA (Md=3, Prom=3.05, SD=1.11, n=150) es percibido como más fácil de usar que el proceso de medición de AFP OMG (Md=2, Prom=2.59, SD=1.15, n=140). La comparación de los procesos de medición IFPUG FPA y **aFPA** ($U = 5065.5, p = 0.062$) indica

que no existe diferencia significativa. En este caso el método IFPUG FPA (Md=3, Prom=3.05, SD=1.11, n=150) es percibido igual de fácil de usar que el proceso de medición de **aFPA** (Md=3, Prom=2.73, SD=1.09, n=79). Del mismo modo, la comparación de los procesos de medición OMG AFP y **aFPA** ($U = 5094.5, p = 0.318$) indica que no existe diferencia significativa en la facilidad de uso percibida. Los resultados indican que los profesionales perciben al procedimiento de medición de la metodología **aFPA** igual de fácil de usar que los otros procesos de medición; sin embargo, perciben el proceso OMG AFP más difícil de usar que el método IFPUG FPA.

Al realizar la prueba *Mann-Whitney U* con la comparación entre los procesos de medición IFPUG FPA y COSMIC FFP ($U = 3159.5, p = 0.001$) los resultados indican que existe diferencia significativa entre ellos. En este caso el método COSMIC FFP (Md=4, Prom=3.63, SD=1.3, n=60) es percibido como más fácil de usar que el proceso de medición de IFPUG FPA (Md=3, Prom=3.05, SD=1.11, n=150). Para los resultados recolectados, el método COSMIC FFP es percibido como más fácil de usar que los demás procesos de medición.

La segunda hipótesis compara los resultados de la utilidad percibida entre los procesos de medición:

$$H_{0U}: \text{no existe diferencia entre la utilidad percibida del } P_i \text{ y el } P_j$$

Los resultados de la prueba de hipótesis para los datos agregados indican que existe una diferencia significativa en los resultados de utilidad percibida de los procesos de medición $\chi^2(2) = 15.493, p = 0.000$, con rangos medios de 207.85 para IFPUG FPA, 160.13 para OMG AFP y 187.99 para el procedimiento de medición de la metodología **aFPA**.

Al realizar la prueba *Mann-Whitney U* para comparar cada par de los procesos de medición relacionados con el método IFPUG FPA se obtiene que existe diferencia significativa entre los procesos de medición IFPUG FPA y OMG AFP ($U = 7811.5, p = 0.000$). En este caso el método IFPUG FPA (Md=4, Prom=3.47, SD=1.15, n=150) es percibido más útil que el proceso de medición de AFP OMG (Md=3, Prom=2.97, SD=1.12, n=140). La comparación de los procesos de medición IFPUG FPA y **aFPA** ($U = 5335.5, p = 0.151$) indica

que no existe diferencia significativa. En este caso, el método IFPUG FPA (Md=4, Prom=3.47, SD=1.15, n=150) es percibido igual de útil que el proceso de medición de **aFPA** (Md=3, Prom=3.28, SD=1.09, n=80). Para la comparación de los procesos de medición OMG AFP y **aFPA** ($U = 4736.5, p = 0.049$) se encuentra que existe una diferencia significativa en la utilidad percibida. Los resultados indican que los profesionales perciben al procedimiento de medición de la metodología **aFPA** como más útil que OMG AFP. Al realizar la prueba *Mann-Whitney U* para la comparación entre los procesos de medición IFPUG FPA y COSMIC FFP ($U = 4375.5, p = 0.743$) los resultados indican que no existe diferencia significativa entre ellos. Esto de acuerdo a los valores que presenta el método COSMIC FFP (Md=4, Prom=3.55, SD=1.05, n=60). Del mismo modo, la comparación entre los procesos de medición COSMIC FFP y **aFPA** ($U = 2059.0, p = 0.135$) indica que no existe diferencia significativa en la utilidad percibida.

La tercera hipótesis compara los resultados de la intención de uso entre los procesos de medición:

$$H_{0j}: \text{no existe diferencia entre la intención de uso del } P_i \text{ y el } P_j$$

Los resultados de la prueba de hipótesis para los datos agregados indican que existe una diferencia significativa en los resultados de la intención de uso de los procesos de medición $\chi^2(2) = 7.152, p = 0.028$, con rangos medios de 120.17 para IFPUG FPA, 97.08 para OMG AFP y 120.48 para el procedimiento de medición de la metodología **aFPA**.

Al realizar la prueba *Mann-Whitney U* para comparar cada par de los procesos de medición relacionados con el método IFPUG FPA se obtiene que existe diferencia significativa entre los procesos de medición IFPUG FPA y OMG AFP ($U = 3006.5, p = 0.017$). En este caso el método IFPUG FPA (Md=3, Prom=3.27, SD=1.31, n=90) tiene una intención de uso mayor que el proceso de medición de AFP OMG (Md=3, Prom=2.80, SD=1.25, n=84). La comparación de los procesos de medición IFPUG FPA y **aFPA** ($U = 2153.0, p = 0.974$) indica que no existe diferencia significativa. En este caso, el método IFPUG FPA (Md=3, Prom=3.27, SD=1.31, n=90) tiene una intención de uso similar que el proceso de medición

de **aFPA** (Md=3, Prom=3.29, SD=1.15, n=48). En el caso de la comparación de los procesos de medición OMG AFP y **aFPA** ($U = 1578.0, p = 0.033$) se encuentra que existe diferencia significativa en la utilidad percibida. Los resultados indican que los profesionales tienen una intención de uso mayor para el procedimiento de medición de la metodología **aFPA**.

Al realizar la prueba *Mann-Whitney U* con la comparación entre los procesos de medición IFPUG FPA y COSMIC FFP ($U = 1446.0, p = 0.335$) los resultados indican que no existe diferencia significativa entre ellos. Esto de acuerdo a los valores que presenta el método COSMIC FFP (Md=4, Prom=3.53, SD=1.23, n=36). Del mismo modo, la comparación entre los procesos de medición COSMIC FFP y **aFPA** ($U = 761.0, p = 0.338$) indica que no existe diferencia significativa en la intención de uso.

Finalmente, para evaluar la aceptación del procedimiento de medición de la metodología **aFPA** se comparan los valores de cada uno de los constructos del modelo de aceptación de tecnología verificando si los valores son significativamente mayores que los valores medios, en este caso comparando con el valor neutral 3 (Abrahamo, 2004). Al realizar la prueba *Mann-Whitney U* ($U = 2646.5, p = 0.51$), los resultados indican que no existe diferencia significativa entre los valores de la facilidad de uso percibida y los valores neutros. Aunque el procedimiento no presenta una tendencia negativa, no es posible concluir que el procedimiento es percibido como fácil de usar. En el caso de la utilidad percibida los resultados ($U = 2488.5, p = 0.006$) indican que el procedimiento de medición de la metodología **aFPA** es percibido como útil. Por último, al analizar los resultados de la prueba para verificar la intención de uso ($U = 960.0, p = 0.104$), los resultados indican que no existe diferencia significativa entre los valores de la intención de uso y los valores neutros por lo que no es posible concluir que los participantes tienen la intención de utilizar el procedimiento.

6.2.2.2.1. RELACIONES ENTRE LOS CONSTRUCTOS DEL MODELO DE ADOPCIÓN

La cuarta hipótesis compara las relaciones entre los constructos utilizados para medir las propiedades de adopción entre los procesos de medición:

$$H_{0c}: \text{el } C_i \text{ no está relacionado con el } C_j$$

Para analizar las relaciones se estudian los constructos C_i y C_j que toman los valores de las propiedades de adopción. Para cada uno de los constructos se presentan los resultados del coeficiente de correlación de *Kendall Tau* y *Spearman Rho*. La Tabla 6-22 detalla los resultados de los coeficientes de correlación entre la intención de uso, la facilidad de uso percibida y la utilidad percibida.

De acuerdo al modelo de adopción de tecnología (TAM) existen relaciones entre los constructos del modelo. En este caso, en el modelo de adopción se esperan relaciones entre los constructos de la facilidad de uso percibida y la intención de uso, y entre los constructos de la utilidad percibida y la intención de uso. Los resultados soportan el planteamiento del modelo TAM, existe una relación entre la facilidad de uso percibida y la intención de uso (*Kendall s* = 0.375, $p < 0.000$) y la utilidad percibida con la intención de uso (*Kendall s* = 0.540, $p < 0.000$). Además, se encuentra una relación entre la facilidad de uso percibida y la utilidad percibida (*Kendall s* = 0.320, $p < 0.000$).

Tabla 6-22 Relaciones de los constructos con la intención de uso.

Constructo	Kendall Tau	Spearman
	Intención de uso	
Facilidad de uso percibida	0.375 $p < 0.000$	0.486 $p < 0.000$
Utilidad percibida	0.540 $p < 0.000$	0.671 $p < 0.000$

6.2.2.2.2. PROPIEDADES DE ADOPCIÓN DEL PROTOCOLO DE VERIFICACIÓN

El protocolo de verificación de exactitud (Quesada-López & Jenkins, 2015b) es aplicado por los participantes de la Replicación-1 (E1-15), Replicación-2 (E2-15) y Replicación-

3 (E1-18) manualmente. Después de su aplicación se solicita llenar una encuesta de salida para evaluar las propiedades de adopción. En total se obtienen 20 respuestas de los participantes. El objetivo del análisis es estudiar la percepción de adopción de los profesionales al aplicar el protocolo de verificación de exactitud en términos de la facilidad de uso percibida, la utilidad percibida y la intención de uso. Para evaluar la aceptación del protocolo de verificación se comparan los valores de cada uno de los constructos del modelo de aceptación de tecnología verificando si los valores son significativamente mayores que los valores neutros, en este caso comparando con el valor 3 de la escala de Likert utilizada en los ítems de la encuesta de percepción (Abrahamo, 2004).

Al realizar la prueba *Mann-Whitney* ($U = 3250.0, p = 0.000$) los resultados indican que existe diferencia significativa entre los valores de la facilidad de uso percibida y los valores neutros. En este caso el protocolo de verificación es percibido como fácil de usar (Md=4, Prom=3.56, SD=1.12, n=100). Al analizar la utilidad percibida, los resultados ($U = 3050.0, p = 0.000$) indican que el protocolo es percibido como útil (Md=4, Prom=3.58, SD=1.04, n=100). Finalmente, al analizar los resultados de la prueba para verificar la intención de uso ($U = 1410.0, p = 0.018$), los resultados indican que existe diferencia significativa entre los valores de la intención de uso y los valores neutros. Los resultados indican que los participantes tienen la intención de utilizar el protocolo (Md=3, Prom=3.37, SD=1.08, n=60).

Los participantes fueron consultados sobre su percepción acerca de cómo el protocolo de verificación les puede ofrecer beneficios en los procesos de medición del tamaño funcional. Para esto se plantean cuatro preguntas sobre la utilidad y la intención de usarlo, la cual responden 13 participantes. El 70% de los participantes que responden tienen una opinión favorable acerca de la utilidad del protocolo de verificación. Los participantes están de acuerdo en que el protocolo permite identificar las diferencias en las mediciones y asociar causas asignables a esas diferencias. Las principales preocupaciones se relacionan con la medición de control (valor verdadero) y con el tiempo necesario para la aplicación del proceso de verificación. Finalmente, el 92.3% de los participantes indican que la aplicación del protocolo les permitió mejorar el entendimiento de los procesos de medición realizados

y que su aplicación es importante en los procesos de enseñanza y aprendizaje de los métodos de medición del tamaño funcional. El 61.5% de los participantes indican que usarían un protocolo de verificación en el futuro si sus organizaciones adoptaran procesos de medición del tamaño funcional. La mayoría de los participantes (92.3%) recomendaría el uso del protocolo. Finalmente, la mitad de los participantes indican la necesidad de ofrecer herramientas de soporte para la automatización del proceso de verificación de exactitud.

6.2.3. DISCUSIÓN DE LOS RESULTADOS DEL ESTUDIO

En esta sección se presentaron un conjunto de estudios empíricos que compararon la efectividad y las propiedades de adopción de los procesos de medición IFPUG FPA, OMG AFP, **aFPA** y COSMIC FFP. Para esto se llevó a cabo un estudio cuasi experimental y tres repeticiones diferenciadas sucesivas. Con respecto a la comparación entre la efectividad de los procesos de medición, los resultados muestran que, en promedio y bajo las condiciones de contexto, los profesionales sin experiencia previa en medición pueden obtener resultados de tamaño funcional con una exactitud y una reproducibilidad de al menos el 10% para los totales de medición. En algunos casos, las mediciones pueden variar entre el 10% y el 30% lo cual indica que el sistema de medición debe ser revisado. En el caso de los resultados obtenidos por el procedimiento de medición de la metodología **aFPA**, la exactitud presentada varió en promedio entre un 6% y 9% con una reproducibilidad entre el 3% y el 7% (n=18).

Los resultados de las pruebas de hipótesis indican que no existe una diferencia significativa en los resultados de exactitud de los procesos de medición IFPUG FPA, OMG AFP y **aFPA**. Los resultados indican que es posible obtener resultados similares en la exactitud utilizando el procedimiento de medición automatizado. Por otro lado, los resultados de las pruebas de hipótesis indican que existe una diferencia significativa en la reproducibilidad de los procesos de medición IFPUG FPA, OMG AFP y **aFPA**. En este caso, el procedimiento

de medición de la metodología **aFPA** presenta mejores resultados en cuanto a la reproducibilidad. Los resultados indican que es posible obtener consistencia en las mediciones utilizando el procedimiento de medición propuesto.

Asimismo, el procedimiento de medición de la metodología **aFPA** puede ser aplicado utilizando como entrada el código fuente de la aplicación o la especificación de los requerimientos funcionales. Los resultados indican que no existe diferencia significativa en la exactitud de la aplicación del procedimiento con el código fuente o la especificación de los requerimientos funcionales. La implicación de estos resultados es que el proceso de medición podría ser aplicado durante la fase de requerimientos y posteriormente, durante el desarrollo para medir el código fuente de la aplicación obteniendo resultados de exactitud similares. En este caso, las mediciones realizadas sobre el código fuente de la aplicación presentan mejores resultados de reproducibilidad.

Los resultados de exactitud y reproducibilidad de las mediciones reportadas para los componentes funcionales básicos (BFC) presentan variaciones mayores que las reportadas para los totales de los puntos de función. En el caso particular del procedimiento de medición de la metodología **aFPA** los valores promedio de la exactitud de las TF pueden variar entre 6% y 8%; sin embargo, los valores para las DF pueden variar entre 22% y 28%. Esto implica que se deben revisar los lineamientos que explican las reglas de mapeo para garantizar que los valores no se desvíen más del 10%. Para la reproducibilidad, los valores para las TF pueden variar entre 4% y 10% y los valores DF pueden variar entre 6% y 15%. Finalmente, al realizar el análisis para cada uno de los BFC: entradas (EI), salidas (EO), consultas (EQ), archivos internos (ILF) e interfaces externas (EIF), los resultados de la exactitud y la reproducibilidad pueden presentar variaciones no aceptables. Estos resultados indican la importancia de la verificación detallada de los resultados de las mediciones de tamaño funcional.

Con respecto a la comparación entre las propiedades de adopción de los procesos de medición, los resultados de las pruebas de hipótesis indican que existe una diferencia significativa en los resultados de facilidad de uso percibida de los procesos de medición IFPUG FPA, OMG AFP y **aFPA**. En este caso, el método IFPUG FPA es percibido más fácil

de usar que el proceso de medición de OMG AFP. La comparación de los procesos de medición IFPUG FPA y **aFPA** indican que el método IFPUG FPA es percibido como de igual facilidad de uso que el proceso de medición de **aFPA**. Asimismo, existe una diferencia significativa en los resultados de utilidad percibida de los procesos de medición IFPUG FPA, OMG AFP y **aFPA**. En este caso, el método IFPUG FPA es percibido más útil que el proceso de medición de AFP OMG, el proceso de medición IFPUG FPA es percibido igual de útil que el proceso de medición de **aFPA** y finalmente, el proceso **aFPA** es percibido más útil que OMG AFP. Por último, los resultados indican que existe una diferencia significativa en los resultados de la intención de uso de los procesos de IFPUG FPA, OMG AFP y **aFPA**. El procedimiento de medición de la metodología **aFPA** tiene una intención de uso similar que el proceso de medición IFPUG FPA. El proceso de medición OMG AFP presenta una intención de uso menor que el IFPUG FPA y **aFPA**.

Con respecto al procedimiento de medición de la metodología **aFPA**, en comparación con los valores neutros de adopción, no es posible concluir que es percibido como fácil de usar y que existe la intención de uso; sin embargo, sí es posible verificar que es percibido como útil. En cuanto al procedimiento de verificación de exactitud de la metodología, los resultados indican que este protocolo es: percibido como fácil de usar, percibido como útil y que existe una intención de uso por parte de los participantes.

6.3. RESUMEN Y DISCUSIÓN DE RESULTADOS DE LAS EVALUACIONES EMPÍRICAS

En este capítulo se presentaron los estudios empíricos para la evaluación de la metodología para la automatización de la medición del tamaño funcional **aFPA**. La evaluación de la metodología y su procedimiento de medición se realizó mediante la conducción de múltiples casos de estudio para determinar su efectividad, la comparación del procedimiento de medición con otras propuestas de medición y la evaluación del procedimiento de verificación se realizó mediante estudios cuasi experimentales.

Se realizaron cuatro casos de estudio para determinar la exactitud de los resultados de medición del tamaño funcional. En el primer caso de estudio se analizaron los resultados

de la exactitud del procedimiento para obtener los puntos de función sin ajustar IFPUG FPA utilizando el código fuente de una aplicación. El procedimiento obtiene una variación en la exactitud (MRE) de un 2.1% para el total de puntos de función sin ajustar (UFP). En el segundo caso de estudio se analizaron los resultados de la exactitud para los puntos de función COSMIC FFP utilizando el código fuente de una aplicación. El procedimiento obtiene una variación en la exactitud (MRE) de un 0.8% para el total de puntos de función (CFP). En el tercer caso de estudio se analizaron los resultados de la exactitud de la medición funcional IFPUG FPA utilizando la extensión del procedimiento para una especificación de requerimientos IEEE Std 830-1998. En este caso, el procedimiento obtiene una variación en la exactitud (MRE) de un 0.7% para el total de puntos de función sin ajustar (UFP). Finalmente, en el cuarto caso de estudio se analizaron los resultados de la exactitud de la medición IFPUG FPA utilizando modelos conceptuales en el contexto del desarrollo dirigido por modelos. El procedimiento obtiene una variación en la exactitud (MRE) de un 5.5% para el total de puntos de función sin ajustar (UFP). Los resultados son alentadores, pero son necesarios más estudios empíricos que permitan replicar los resultados obtenidos en distintos ambientes de desarrollo. La evidencia muestra que el procedimiento de medición de la metodología **aFPA** puede obtener resultados aceptables de exactitud. Estos resultados pueden alcanzarse mediante la calibración del procedimiento en el contexto particular de operación.

Se realizaron tres replicaciones diferenciadas de un estudio controlado cuasi experimental orientados a personas para evaluar la efectividad y las propiedades de adopción de cuatro procesos de medición del tamaño funcional: IFPUG FPA, OMG AFP, **aFPA** y COSMIC FFP. Los resultados de los procesos de medición indican que, en promedio, los profesionales obtienen resultados de tamaño funcional con una variación en la exactitud y en la reproducibilidad menor al 10% para los totales de medición. Sin embargo, en algunos casos, las mediciones pueden variar entre el 10% y el 30% lo cual indica que el sistema de medición debe ser revisado. Para el caso del procedimiento de medición de la metodología **aFPA**, los resultados presentan una variación promedio en la exactitud entre un 6% y un 9%

con una reproducibilidad entre un 3% y un 7% (n=18). Asimismo, el procedimiento de medición de la metodología **aFPA** presenta mejores resultados en la reproducibilidad que los procesos de medición IFPUG FPA y OMG AFP.

Los resultados indican que el procedimiento de medición de la metodología **aFPA** obtiene resultados similares cuando se aplica utilizando como entrada el código fuente de la aplicación o la especificación de los requerimientos funcionales. Sin embargo, las mediciones realizadas sobre el código fuente de la aplicación presentan mejores resultados de reproducibilidad. Los resultados de exactitud y reproducibilidad presentados en las mediciones de los componentes funcionales básicos (BFC) deben ser analizados para mejorar su efectividad. Estos resultados muestran la importancia de la verificación detallada de los resultados de las mediciones de tamaño funcional. Los resultados de las propiedades de adopción indican que los procesos de medición IFPUG FPA y **aFPA** son percibidos de igual facilidad de uso. Del mismo modo, el proceso de medición IFPUG FPA es percibido igual de útil que el proceso de medición **aFPA** y el proceso **aFPA** es percibido más útil que el proceso de medición OMG AFP. Finalmente, el procedimiento de la metodología **aFPA** tiene una intención de uso similar que IFPUG FPA y el proceso de medición OMG AFP presenta una intención de uso menor que el IFPUG FPA y el **aFPA**. En cuanto al protocolo de verificación de exactitud, los resultados indican que este protocolo es percibido como fácil de usar, percibido como útil y que existe una intención de uso por parte de los participantes.

Capítulo 7. VALIDACIÓN EMPÍRICA DEL PROCEDIMIENTO DE EVALUACIÓN DE LA METODOLOGÍA **aFPA**

Este capítulo agrega conocimiento para la consecución del objetivo específico 2 (OE1) que permite responder parte de la primera pregunta de investigación (KP1) y genera evidencia que permite responder la tercera pregunta de investigación (KP3) del objetivo específico 3 (OE3) sobre la efectividad de la metodología para la automatización de la medición del tamaño funcional del software.

En este capítulo se presenta los estudios empíricos realizados para la evaluación de la metodología de medición de tamaño funcional **aFPA** con respecto al procedimiento de evaluación para el aprovechamiento de los resultados de medición mediante la creación de modelos de estimación de esfuerzo. El procedimiento de evaluación de los modelos de estimación propuesto en la Sección 5.3 utiliza como entrada los resultados del proceso de medición del tamaño funcional **aFPA**. Esta sección presenta una replicación externa y una familia de experimentos para la evaluación de modelos de estimación de esfuerzo de desarrollo basados en las mediciones de los métodos de tamaño funcional IFPUG FPA y COSMIC FFP. El objetivo es realizar comparaciones exhaustivas entre distintos modelos de estimación de esfuerzo que permitan determinar el que mejor desempeño produce en el contexto de evaluación. Asimismo, identificar si existen modelos simplificados basados en los componentes funcionales básicos que produzcan resultados similares a los modelos basados en el total de los puntos de función. El proceso de aprovechamiento de los resultados corresponde al Paso 3 del modelo de procesos propuesto por (Abran, 2010; Habra, Abran, Lopez & Sellami, 2008; Jacquet & Abran, 1997) y presenta el detalle de cómo explotar los resultados de medición de tamaño funcional obtenidos durante el procedimiento de medición de tamaño funcional de la metodología **aFPA**. Los estudios realizados investigan la

efectividad del procedimiento de evaluación para determinar cómo las mediciones del tamaño funcional pueden ser aprovechadas en la construcción de modelos de estimación de esfuerzo.

En la Sección 7.1 se presenta una replicación que evalúa la estructura interna y la aplicabilidad del método de medición IFPUG FPA. Esta replicación analiza la relación entre los componentes funcionales básicos (BFC), los principales BFC que influyen la exactitud de los modelos de estimación y los modelos de estimación de esfuerzo simplificados construidos a partir de conjuntos de los BFC para determinar su nivel de exactitud en comparación con modelos basados en el conteo total del tamaño funcional. En la Sección 7.2 se evalúa la estructura interna y la aplicabilidad del método de medición COSMIC FFP empíricamente. Del mismo modo, se analiza la relación entre los componentes funcionales básicos (BFC), los principales BFC que influyen la exactitud de los modelos de estimación y los modelos simplificados construidos a partir de conjuntos de los BFC. Finalmente, en la Sección 7.3 se valida el procedimiento para la evaluación de los resultados de medición en la construcción de modelos de estimación de esfuerzo de desarrollo. Este estudio analiza la estrategia de evaluación de modelos de estimación del esfuerzo de desarrollo en el contexto de aplicaciones orientadas a datos, utilizando las mediciones de tamaño funcional y los distintos factores de contexto como predictores y obteniendo como salida la estimación del esfuerzo de desarrollo de un proyecto de software. Este estudio utiliza el módulo descrito en la Sección 5.4.3 que permite realizar un análisis comparativo exhaustivo que contempla la combinación de distintas técnicas de pre procesamiento de datos, selección de atributos y algoritmos de aprendizaje bajo distintos conjuntos de datos. El módulo de evaluación de modelos apoya el proceso de construcción de modelos específicos para aprovechar los resultados de medición de tamaño funcional a partir de conjuntos de datos específicos.

7.1. EVALUACIÓN EMPÍRICA DE LA ESTRUCTURA INTERNA Y LA APLICABILIDAD DE PUNTOS DE FUNCIÓN IFPUG FPA: UNA REPLICACIÓN EXTERNA

Este estudio representa un producto obtenido del ciclo de investigación 3 (RC3).

En esta sección se complementa el análisis de la primera pregunta de investigación (KP1) donde se estudian los desafíos y oportunidades de los métodos de medición del tamaño funcional. Para eso se lleva a cabo una replicación externa que evalúa la efectividad del método de medición IFPUG FPA y sus componentes funcionales básicos (BFC) para la construcción de modelos de estimación de esfuerzo. En esta replicación se verifican los resultados de un conjunto de estudios base para lo cual se realiza una repetición deliberada del procedimiento empírico en nuevas corridas con el fin de determinar si es posible reproducir los resultados anteriores (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012).

Este estudio empírico reporta una replicación llevada a cabo con subconjuntos de proyectos del repositorio del *International Software Benchmarking Standards Group dataset (ISBSG)* para evaluar la estructura interna y la aplicabilidad de la metodología de puntos de función IFPUG FPA. El estudio es una replicación externa sin interacción con los investigadores originales (Carver, 2010). El objetivo de esta familia de replications (Quesada-López & Jenkins, 2014; Quesada-López & Jenkins, 2015c; Quesada-López & Jenkins, 2016) es agregar evidencia acerca de los aspectos internos de la métrica de puntos de función IFPUG FPA. Asimismo, los resultados agregados son comparados con los reportados en los estudios originales descritos en (Kitchenham & Kansala, 1993; Jeffery, Low & Barnes, 1993; Jeffery & Stathis, 1996; Lavazza, Morasca & Robiolo, 2013).

La estructura de la investigación se reporta de acuerdo a los lineamientos para replications experimentales propuestas por (Carver, 2010). Para esto, se describen los estudios originales, se detalla el diseño de la replicación actual y se comparan los resultados de la replicación con los estudios originales.

7.1.1. DESCRIPCIÓN DE LOS ESTUDIOS ORIGINALES

Los estudios originales (Kitchenham & Kansala, 1993; Jeffery, Low & Barnes, 1993; Jeffery & Stathis, 1996; Lokan, 1999; Lavazza, Morasca & Robiolo, 2013) realizaron evaluaciones sobre la estructura y aplicabilidad de la metodología de puntos de función IFPUG FPA como métrica de tamaño funcional. Las correlaciones entre los componentes funcionales básicos (BFC) implican que aspectos del tamaño se miden más de una vez ya que algunos BFC pueden haber sido medidos por otros BFC. Los estudios examinan el conteo de puntos de función para determinar cuáles BFC son independientes de otros BFC y por tanto apropiados para el modelo aditivo de tamaño. Asimismo, investigan la relación entre el tamaño funcional y el esfuerzo de desarrollo que incluye el análisis de los puntos de función sin ajustar (UFP), puntos de función ajustados (AFP) y los componentes funcionales básicos (BFC).

7.1.1.1. OBJETIVOS Y PREGUNTAS DE INVESTIGACIÓN

Kitchenham & Kansala (1993) analizan la consistencia interna del método IFPUG FPA y el uso del tamaño funcional para estimar el esfuerzo de desarrollo. Jeffery, Low & Barnes (1993) investigan los ajustes de complejidad del método y la correlación entre los componentes funcionales básicos (BFC) para analizar métodos simplificados de medición. Jeffery & Stathis (1996) analizan empíricamente los BFC de los puntos de función sin ajustar (UFP) y cuándo cada uno de los BFC son estadísticamente independientes entre sí. Asimismo, investigan la relación entre el esfuerzo de desarrollo, los BFC, los UFP y los AFP. Lokan (1999) estudia las correlaciones entre los BFC y analiza cuáles son los factores que influyen el balance entre los BFC. Lavazza, Morasca & Robiolo (2013) investigan las correlaciones entre los BFC para evaluar la posibilidad de una definición simplificada de puntos de función.

En términos generales, el método de medición de puntos de función IFPUG FPA fue analizado para determinar en qué medida los componentes funcionales básicos (BFC) del

método son independientes entre sí y por lo tanto apropiados para un modelo aditivo de tamaño. El problema consiste en que (a) desde un punto de vista práctico, la correlación entre los BFC implica que algunos aspectos se miden dos veces, lo que representa un desperdicio del esfuerzo de medición y (b) desde el punto de vista teórico, medir los BFC que ya han sido medidos por otro BFC podría afectar la confiabilidad del método de medición (Kitchenham & Kansala, 1993; Lavazza, Morasca & Robiolo, 2013).

En la práctica, los profesionales podrían utilizar las relaciones BFC para la estimación del tamaño funcional a partir de elementos individuales sin aplicar todo el método de conteo (Lokan, 1999). Asimismo, las correlaciones entre esfuerzo, los componentes funcionales básicos (BFC) y los puntos de función sin ajustar (UFP) podrían ser utilizados para determinar si una métrica de tamaño simplificada puede ser apropiada para simplificar la construcción de modelos de estimación del esfuerzo de desarrollo.

Los objetivos y preguntas de investigación de los estudios originales relacionados con la replicación se listan en la Tabla 7-1.

Tabla 7-1 Resumen de objetivos y preguntas de investigación.

Autores	Objetivos y preguntas de investigación
Kitchenham & Kansala (1993)	(1) Determinar cuándo todos los componentes funcionales básicos son requeridos para proporcionar una medida de tamaño válida. (2) Determinar cuándo la suma de todos los componentes funcionales básicos es un mejor estimador de esfuerzo que los elementos que la componen.
Jeffery, Low & Barnes (1993)	(1) Determinar si un conteo de puntos de función simplificado puede obtener resultados similares a los obtenidos utilizando el método completo.
Jeffery & Stathis (1996)	(1) Determinar en qué medida los componentes funcionales básicos de puntos de función son independientes entre sí y por lo tanto apropiados para un modelo aditivo de tamaño. (2) Investigar la relación entre el esfuerzo, los componentes funcionales básicos y los puntos de función sin ajustar.
Lokan (1999)	(1) Identificar las correlaciones entre los componentes funcionales básicos de puntos de función de acuerdo al tipo de desarrollo, tipo de lenguaje y lenguajes de programación.
Lavazza, Morasca & Robiolo (2013)	(1) Investigar la construcción de modelos de estimación de esfuerzo utilizando subconjuntos de BFC que presenten al menos el mismo rendimiento que el método completo. (2) Analizar las correlaciones entre los puntos de función sin ajustar, BFC y esfuerzo.

7.1.1.2. CONTEXTO Y VARIABLES

Los estudios originales realizaron su análisis basado en conjuntos de datos de proyectos reales de distintas organizaciones de desarrollo de software donde el principal tipo de aplicaciones pertenecen al dominio de aplicaciones de negocio (MIS). La Tabla 7-2 resume la información acerca del contexto de los conjuntos de datos utilizados por cada estudio. La Tabla 7-3 lista las variables dependientes e independientes analizadas en cada estudio. En particular, en nuestro estudio es de interés realizar el análisis para el dominio de aplicaciones de negocio (MIS).

Tabla 7-2 Conjuntos de datos de los estudios originales.

Autores	Conjunto de datos	Tipo	Dominio
Kitchenham & Kansala (1993)	40 proyectos de 9 organizaciones de desarrollo de software.	Multicompañía	MIS
Jeffery, Low & Barnes (1993)	64 proyectos de 1 organización de desarrollo de software.	Intra-compañía	MIS
Jeffery & Stathis (1996)	17 proyectos de 1 organización de desarrollo de software.	Intra-compañía	MIS
Lokan (1999)	269 proyectos del repositorio ISBSG R4.	Multicompañía	MIS, DSS
Lavazza, Morasca & Robiolo (2013)	Más de 600 proyectos del repositorio ISBSG R11.	Multicompañía	MIS, DSS

Tabla 7-3 Listado de variables dependientes e independientes.

Independientes		Dependientes
Nombre	Descripción	
Tamaño de BFC (UUF and UFP)	Conteo de puntos de función para las entradas, salidas, consultas, archivos internos e interfaces externas.	Esfuerzo de desarrollo
Tamaño en UUF	Puntos de función sin ponderar y sin ajustar.	
Tamaño en UFP	Puntos de función sin ajustar.	
Tamaño en AFP	Puntos de función ajustados.	
Variables de contexto	Tipo de desarrollo, lenguaje de programación, tipo de lenguaje, grupo de aplicación, tamaño del equipo.	

7.1.1.3. RESUMEN DE TÉCNICAS DE ANÁLISIS DE DATOS

La Tabla 7-4 resume las métricas de evaluación de resultados utilizadas en cada estudio y las técnicas utilizadas para la construcción de los modelos de estimación de esfuerzo.

Para la construcción de los modelos de estimación de esfuerzo basados en puntos de función, distintas técnicas de regresión fueron aplicadas, tanto con transformación de datos como sin transformación. Las principales técnicas utilizadas son las de regresión lineal simple, *stepwise multivariate regression*, *ordinary least square regression*, *least median of squares regression* y estimación por analogía. La correlación entre los componentes funcionales básicos (BFC) se ha analizado utilizando pruebas paramétricas como el coeficiente de correlación de *Pearson*, R^2 y pruebas no paramétricas tales como el coeficiente de correlación de *Kendall* y de *Spearman*. Estos coeficientes de correlación indican la asociación entre variables, en este caso el tamaño funcional y el esfuerzo de desarrollo. Para utilizar la prueba de *Pearson* ambas variables deben ser continuas y estar normalmente distribuidas, en el caso de las pruebas no paramétricas de *Spearman* y de *Kendall* no son necesarias las suposiciones sobre la distribución. Estas pruebas pueden ser usadas tanto para variables ordinales como continuas. Lavazza, Morasca & Robiolo (2013) indican que la correlación entre variables no necesariamente implica la obtención de modelos de estimación de esfuerzo exactos. Por tanto, otros indicadores fueron utilizados para evaluar la exactitud de los modelos de estimación tales como, la magnitud del error relativo (MRE) y el número de predicciones con un % de los valores reales ($Pred_n$), los cuales se describen en la Sección 5.3.3.

Tabla 7-4 Resumen de técnicas utilizadas en los estudios originales.

Autores	Técnicas	Evaluación
Kitchenham & Kansala (1993)	Regresión lineal simple, <i>stepwise multivariate regression</i> .	R^2 , coeficiente de correlación de Pearson y de Kendall.
Jeffery, Low & Barnes (1993), Jeffery & Stathis (1996)	Regresión lineal simple.	R^2 , coeficiente de correlación de Pearson y de Kendall.
Lokan (1999)	Regresión lineal simple.	Coeficiente de correlación de Kendall.
Lavazza, Morasca & Robiolo (2013)	Regresión lineal simple, transformaciones logarítmicas, <i>ordinary least square regression, least median of squares regression, analogy criteria estimation</i> .	Distancia de Cook, coeficiente de correlación de Kendall y de Spearman. R^2 , MRE , $Pred(25)$.

7.1.1.4. RESUMEN DE RESULTADOS

Kitchenham & Kansala (1993) reportan correlaciones entre los componentes funcionales básicos (BFC) lo que implica que los BFC no son independientes. Esto indica que el método de medición IFPUG FPA no tiene las características de un método aditivo dado que algunos componentes parecen ser medidos más de una vez en el método original. Asimismo, algunos de los BFC se relacionan con el esfuerzo de desarrollo, lo que indica que algunos modelos de estimación de esfuerzo; por ejemplo, los basados en un subconjunto de los BFC tales como EI y EO, pueden ser tan buenos como los basados en el total de los puntos de función. Los autores plantean que un proceso simplificado de conteo puede reducir la variabilidad del proceso de medición.

Del mismo modo, Jeffery, Low & Barnes (1993) encuentran que los BFC no son independientes. Los autores concluyen que los ajustes de complejidad no tienen un efecto en el desempeño de los modelos de estimación de esfuerzo. Jeffery & Stathis (1996) encuentran correlaciones significativas entre los UFP y los EI, EQ, ILF y entre los BFC y el esfuerzo de desarrollo. Asimismo, determinan que los valores de ajuste no mejoran el poder de la métrica de puntos de función en los modelos de estimación de esfuerzo. Además, sugieren que una métrica simplificada puede ser apropiada para tales efectos.

Por su parte, Lokan (1999) reporta evidencia de la inter-correlación entre los EI y los ILF y donde los EIF rara vez se correlaciona con otros BFC. Los autores determinan la influencia de factores de contexto específicos tales como el tipo de desarrollo y el tipo de

lenguaje en el balance de los BFC. Lavazza, Morasca & Robiolo (2013) determinan correlaciones entre los BFC y evalúan modelos simplificados de estimación de esfuerzo obteniendo resultados alentadores. Los autores proponen la creación de métodos simplificados de puntos de función IFPUG FPA basados en subconjuntos de los BFC.

Los resultados muestran que los BFC se pueden correlacionar entre sí y con el esfuerzo de desarrollo. Esto sugiere que un método simplificado del método de medición de puntos de función IFPUG FPA; por ejemplo, basado en subconjuntos de los BFC, puede ser aplicable en distintos dominios. Se espera que un conteo simplificado pueda reducir la variabilidad de los resultados de medición y el esfuerzo para realizar el conteo.

7.1.2. REPLICACIÓN EXTERNA

Esta sección detalla el diseño del estudio de replicación. El objetivo de esta familia de replications (Quesada-López & Jenkins, 2014; Quesada-López & Jenkins, 2015c; Quesada-López & Jenkins, 2016) es agregar evidencia sobre la estructura interna y aplicabilidad de la metodología de puntos de función IFPUG FPA. El objetivo del estudio es agregar evidencia acerca de la estructura interna y aplicabilidad de la métrica de puntos de función. Para esto se analizan los componentes funcionales básicos (BFC) del método y se evalúan modelos de estimación de esfuerzo de desarrollo basados en subconjuntos de estos componentes. En particular, es de interés en el procedimiento para la automatización de la medición del tamaño funcional del software determinar los componentes que tienen un impacto mayor en la exactitud de los modelos de estimación basados en el tamaño funcional.

7.1.2.1. MOTIVACIÓN

Múltiples modelos de estimación de esfuerzo han sido propuestos en la literatura durante los últimos años; sin embargo, no se ha llegado a conclusiones sobre cuál técnica de estimación de esfuerzo es la mejor ya que los resultados pueden depender de los conjuntos de datos analizados y la combinación de técnicas utilizadas (Shepperd, 2007; Dejaeger, Verbeke, Martens & Baesens, 2012; Shepperd & MacDonell, 2012).

La investigación sobre la evaluación de modelos de estimación de esfuerzo de desarrollo de software a partir del tamaño funcional presenta aun grandes retos para la comunidad de la ingeniería del software (Jorgensen & Shepperd, 2007; Gencel, 2008; Dejaeger, Verbeke, Martens & Baesens, 2012). Uno de los principales aspectos a considerar es el cómo se realiza la selección de datos a partir de repositorios históricos para obtener estimaciones consistentes (Gencel & Buglione, 2008; Kitchenham & Mendes, 2009) y el cómo se puede realizar una comparación justa entre modelos que garantice que no existan sesgos en la selección de los casos para la construcción de cada uno de los modelos bajo comparación (Kitchenham & Mendes, 2009).

El procedimiento de evaluación de los resultados de medición propuesto en la Sección 5.3 propone un proceso para la construcción de modelos de estimación de esfuerzo. Esta estrategia de evaluación permite realizar una comparación sin sesgo entre modelos de estimación para determinar el que mejor desempeño produce en el contexto de evaluación. La motivación de esta replicación es aplicar parte del procedimiento de evaluación para analizar conjuntos de datos en el contexto de las aplicaciones de negocios (MIS) y comparar los resultados obtenidos en los estudios originales. El análisis contempla el uso de métricas que disminuyen el sesgo y que pueden permitir la agregación de resultados entre estudios.

La combinación de los resultados de una familia de replicaciones es interesante ya que se analizan preguntas de investigación relacionadas en distintos contextos. La agregación de resultados por replicaciones es útil en el área de la ingeniería del software para consolidar evidencia (Carver, 2010). En esta replicación se analiza la estructura interna del método IFPUG FPA y la posibilidad de construir modelos de estimación de esfuerzo simplificados basados en subconjuntos de los BFC. Los resultados son comparados cuando el estudio original reporta los datos necesarios para permitir dicha comparación.

Finalmente, los estudios originales reportaron distintos conjuntos de BFC que presentan correlaciones a partir de los datos de análisis. Asimismo, no hay consenso en cuales de estos componentes funcionales son requeridos para crear los modelos de estimación de esfuerzo simplificados. Esta replicación busca agregar conocimiento utilizando un conjunto de proyectos diferenciado y una estrategia de evaluación que permita obtener resultados

sin sesgo para la comparación de los modelos. El conjunto de datos ha sido seleccionado de acuerdo a un conjunto de criterios específicos del dominio de aplicaciones de negocio (MIS) que permite realizar el análisis de las coincidencias y las diferencias encontradas en comparación con los resultados previos.

7.1.2.2. NIVEL DE INTERACCIÓN CON LOS AUTORES ORIGINALES

El estudio realiza una replicación externa sin interacción con los autores originales. Para la replicación no se cuenta con los conjuntos de datos de los paquetes experimentales de los estudios originales y se consideran solo los resultados reportados en los artículos. Este aspecto limita la comparación de los resultados o la posibilidad de realizar un estudio de metaanálisis.

7.1.2.3. CAMBIOS REALIZADOS CON RESPECTO A LOS ESTUDIOS ORIGINALES

El estudio fue diseñado para mantener los análisis de los estudios originales con el objetivo de realizar la comparación de los resultados. Sin embargo, un conjunto de cambios fue realizado con respecto a los análisis originales. Tres tipos de cambios fueron realizados intencionalmente: (1) la selección del conjunto de datos y su contexto, (b) la selección de las variables independientes y (b) las métricas y las técnicas estadísticas utilizadas para el análisis de los datos. Cada uno de los cambios realizados se detalla en las secciones siguientes.

En este estudio se realiza análisis empírico donde (a) se investigan las correlaciones entre los puntos de función sin ajustar (UFP) y cada uno de los componentes funcionales básicos (BFC) que lo componen (EI, EO, EQ, ILF, EIF); asimismo, la inter correlación entre los BFC, (b) se analizan las correlaciones entre los UFP y los BFC con el esfuerzo de desarrollo y finalmente, (c) se estudian los modelos simplificados de estimación basados en los BFC y se comparan con los modelos construidos a partir de los UFP. Basado en los objetivos del estudio se analizan las siguientes preguntas de investigación:

- RQ1. *¿Existen correlaciones entre los puntos de función sin ajustar (UFP) y cada uno de los componentes funcionales básicos (BFC) que lo componen? ¿Existen correlaciones entre los BFC?*
- RQ2. *¿Existen correlaciones entre los puntos de función sin ajustar (UFP) y los componentes funcionales básicos (BFC) con el esfuerzo de desarrollo?*
- RQ3. *¿Cuál es la exactitud de los modelos de estimación simplificados en comparación con los modelos construidos a partir de los puntos de función sin ajustar (UFP)?*

7.1.2.3.1. DESCRIPCIÓN DEL PROCESO

La conducción del proceso general del estudio conlleva las siguientes fases:

- Selección de datos y pre procesamiento: a partir del repositorio de datos ISBSG se realiza el proceso de selección y pre procesamiento para obtener los conjuntos de datos y variables utilizados en el experimento.
- Evaluación de los UFP y los BFC: se analizan los puntos de función sin ajustar (UFP) y los componentes funcionales básicos (BFC) para determinar la independencia entre sí y las relaciones con el esfuerzo de desarrollo.
- Evaluación de los modelos de estimación: se evalúa la aplicabilidad del método de tamaño funcional en la creación de modelos de estimación simplificados y se determina el nivel de exactitud de las estimaciones.

7.1.2.3.2. SELECCIÓN DE LOS CONJUNTOS DE DATOS

Se seleccionan tres conjuntos de datos (DS) para el estudio. Los conjuntos de datos seleccionan aplicaciones de negocio del dominio de sistemas de información administrativo (MIS) para el método de medición IFPUG FPA. La selección se realiza a partir de las bases de datos de proyectos del ISBSG versión R11 y R12. Las instancias seleccionadas para cada

uno de los conjuntos de datos contienen información de proyectos de desarrollo realizados en múltiples compañías.

El repositorio de proyectos ISBSG provee un conjunto de datos de proyectos de desarrollo de software. Este repositorio contiene proyectos de múltiples compañías, países, industrias y múltiples ambientes de desarrollo. En su última versión (R12), el repositorio contiene 6000 proyectos con 105 características entre las cuales se encuentra el tamaño funcional, el esfuerzo de desarrollo y múltiples factores de contexto. El ISBSG recomienda que para utilizar los datos de este repositorio es necesario realizar una selección de proyectos y atributos de acuerdo al uso específico para el cual se requiere utilizar (Fingerman, 2011).

La selección de cada uno de los conjuntos de datos (DS) se realiza bajo los criterios de la calidad de los datos, la calificación del conteo, el método de conteo y su versión, el año de desarrollo de los proyectos, el grupo de aplicaciones y el nivel de los recursos en el registro del esfuerzo de desarrollo, tal como se muestra en la Tabla 7-5. La motivación para los criterios de selección de proyectos son los siguientes: (a) selección de los proyectos medidos con los últimos manuales de conteo del método de puntos de función IFPUG FPA versión 4+, (b) solo los datos de proyectos con un nivel alto de calidad e integridad, (c) solo los conteos con un nivel alto de calidad, (d) solo proyectos recientes que utilizan nuevas tecnologías, (e) del grupo de aplicaciones de negocios y (f) donde solo se reporta el esfuerzo del equipo de desarrollo.

Tabla 7-5 Criterios de selección de conjuntos de datos (DS).

DS	Método	Calidad datos	Calidad conteo	Periodo	Grupo	Nivel recursos	Lenguaje
DS12	IFPUG 4+	A	A	≥ 2008	BA	1	C#
DS72	IFPUG 4+	A	A	≥ 2008	BA	1	-
DS202	IFPUG 4+	A	A B	≥ 2005	BA	1	-

La Tabla 7-6 lista las variables seleccionadas para el estudio. Dado que el repositorio de proyectos ISBSG presenta proyectos con valores faltantes, se seleccionan los proyectos que reportan los puntos de función sin ajustar (UFP) y cada uno de los cinco componentes funcionales básicos (BFC) completos, así como el esfuerzo de desarrollo de los equipos. El

conjunto de variables de contexto utilizado en cada una de las evaluaciones se selecciona independientemente de los valores faltantes y de acuerdo a las recomendaciones de (González-Ladrón-de-Guevara & Fernández-Diego, 2014; González-Ladrón-de-Guevara, Fernández-Diego & Lokan, 2016).

Como resultado del proceso de selección inicial se obtienen tres subconjuntos de datos, tal como se muestra en la Tabla 7-7. En total, el primer subconjunto selecciona 14 proyectos, el segundo subconjunto selecciona 72 proyectos y el tercer subconjunto selecciona 204 proyectos. Para reducir los conjuntos de datos se realiza la prueba de distancia de Cook (1977; 2000) que permite determinar si existen valores atípicos que puedan influenciar la efectividad de los modelos basados en regresiones lineales. Después del análisis, se encuentran dos valores atípicos en el primer y tercer subconjunto de datos los cuales fueron removidos. Estos proyectos representan proyectos muy grandes que varían entre 1,659 y 2,250 UFP con una tasa de productividad diferente a la media para proyectos grandes (entre 1.5 y 13.7 HP/FP).

Tabla 7-6 Detalle de variables seleccionadas en el estudio.

Variable	Escala	Descripción
Conteo de entradas	Razón	Puntos de función sin ajustar (UFP) de las entradas externas (EI).
Conteo de salidas	Razón	Puntos de función sin ajustar (UFP) de las entradas externas (EO).
Conteo de consultas	Razón	Puntos de función sin ajustar (UFP) de las consultas externas (EQ).
Conteo de archivos internos	Razón	Puntos de función sin ajustar (UFP) de los archivos lógicos internos (ILF).
Conteo de interfaces externas	Razón	Puntos de función sin ajustar (UFP) de las interfaces externas (EIF).
Tamaño funcional	Razón	Puntos de función sin ajustar (UFP) de la aplicación.
Esfuerzo de desarrollo (<i>Normalized Level 1</i>)	Razón	El esfuerzo de desarrollo del equipo en horas persona recolectadas a lo largo del proyecto.
Productividad (<i>Normalized Level 1</i>)	Razón	Tasa de entrega de funcionalidad en horas y por unidad de tamaño en puntos de función sin ajustar.
Tipo de desarrollo	Nominal	Proyecto de mejora E= <i>Enhancement</i> , nuevo desarrollo N= <i>New Development</i> , R= <i>Re-development</i> .
Tamaño relativo	Ordinal	1. XXS, 2. XS, 3. S, 4. M1, 5. M2, 6. L, 7. XL, 8. XXL, 9. XXXL.
Tamaño equipo de desarrollo	Ordinal	2, 3-4, 5-8, 9-14, 15-20, 21-30, 31-40, 61-70. Tamaño del equipo, tamaño máximo del equipo y tamaño promedio del equipo de desarrollo.
Año	Nominal	Año del proyecto derivado de la entrega.
Plataforma de desarrollo	Nominal	MP= <i>Multi-Platform</i> , MF= <i>Main Frame</i> , PC=PC, MR= <i>Mid-Range</i> .
Arquitectura	Nominal	CS=Cliente servidor, SA= <i>Stand alone</i> , MC=Multicapa, MW=Multicapa Web, ND = No definido.
Tipo de leguaje	Nominal	3GL, ApG, 4GL
Lenguaje de programación	Nominal	Lenguaje principal: JV = Java, COO = COOLGen, ASN = ASPNet, C# = C#, JS = JavaScript, ABA = ABAP, PL = PLI, VB = VisualBasic, PB = PowerBuilder, ASP = ASP, SQL = SQL, NET = VisualStudioNet, DST = Datastage, NET = Net, IBM = IBMWTX, XML = XML, COB = COBOL, ABI = ABINITIO, AG = AG, C++ = C++.
Método de desarrollo	Nominal	Método de desarrollo utilizado en el proyecto.
Estándar de medición	Nominal	Versión de estándar de medición del tamaño funcional utilizado.
Conteo manual	Nominal	Sí/No.

Finalmente, el primer subconjunto selecciona 12 proyectos (DS12), el segundo subconjunto selecciona 72 proyectos (DS72) y el tercer subconjunto selecciona 202 proyectos (DS202) para el análisis. La Tabla 7-8 presenta el detalle de las estadísticas descriptivas de cada uno de los sub conjuntos de datos seleccionados para las variables: puntos de función sin ajustar (UFP), entradas (EI), salidas (EO), consultas (EQ), archivos internos (ILF), interfaces externas (EIF) y esfuerzo (Eff). Para la mayoría de las variables los datos muestran que los conjuntos de datos tienen un *skewness* con simetría positiva que indica que la cantidad de proyectos pequeños y medianos es mayor que la cantidad de proyectos grandes.

En el caso del *kurtosis* las variables muestran valores negativos para el DS12 y valores positivos para el DS72 y DS202. Las pruebas de normalidad *Kolmogorov-Smirnov* y *Shapiro-Wilk* indican que solo las variables del conjunto de datos DS12 (EI, EO, EQ, ILF, UFP, Esfuerzo) pertenecen a una distribución normal. Por otro lado, todas las demás variables en los distintos conjuntos de datos no pertenecen a una distribución normal. La transformación logarítmica de las variables permite normalizar los datos donde la prueba de normalidad indica que solo la variable EIF del conjunto de datos DS202 no pertenece a una distribución normal. La descripción detallada de los conjuntos de datos es necesaria dado que los resultados del estudio tienen que ser analizados con respecto a los conjuntos de datos seleccionados.

Tabla 7-7 Conjuntos de datos utilizados en la replicación.

DS	Método	N	Repositorio	Conjunto de datos	Periodo	Tipo	Dominio
DS12	IFPUG FPA	12	ISBSG R11	12 proyectos del ISBSG R11	2008-2011	Multicompañía	MIS
DS72	IFPUG FPA	72	ISBSG R12	72 proyectos del ISBSG R12	2008-2011	Multicompañía	MIS
DS202	IFPUG FPA	202	ISBSG R12	202 proyectos del ISBSG R12	2005-2011	Multicompañía	MIS

La Figura 7-1 y la Figura 7-2 muestran la distribución del tamaño funcional de cada uno de los subconjuntos de datos agrupados por las variables de contexto. Se detallan las distribuciones para el tipo de lenguaje (TL), el lenguaje de programación (LP), el tipo de desarrollo (TD), la plataforma de desarrollo (PD), el tamaño del equipo de desarrollo (TE), el año de liberación (AN), la arquitectura (AR) y el tamaño relativo (TR). La Tabla 7-9 lista las estadísticas descriptivas de un conjunto de las variables de contexto de los subconjuntos de datos. En el Anexo 5 se detallan las estadísticas para cada una de las variables de contexto. Es importante conocer las variables de contexto utilizadas en la evaluación ya que estas pueden influenciar el esfuerzo de desarrollo que toma desarrollar los proyectos a partir de su tamaño funcional.

Tabla 7-8 Estadística descriptiva de los conjuntos de datos seleccionados.

DS	N	Var	Prom	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rango	Std.D.	Skew	Kurt
DS12	12	UFP	157.75	40	40.00	42.10	76.50	133.00	237.00	308.50	-	313	273	97.38	0.38	-1.41
	12	EI	46.83	14	14.00	14.30	17.75	54.50	71.50	78.40	-	79	65	25.88	-0.15	-1.86
	12	EO	28.42	0	0.00	1.20	6.00	25.50	46.25	71.60	-	77	77	24.14	0.72	-0.26
	12	EQ	44.42	0	0.00	0.00	10.25	42.50	68.00	107.00	-	116	116	37.07	0.46	-0.66
	12	ILF	27.92	0	0.00	2.10	8.75	28.00	41.50	68.60	-	77	77	21.43	0.99	1.29
	12	EIF	10.17	0	0.00	0.00	0.00	5.00	16.50	37.00	-	40	40	13.39	1.29	0.93
	12	Eff	3,240.42	81	81.00	164.70	982.50	2,065.00	4,979.75	11,259.40	-	13,489	13,408	3,725.39	2.17	5.26
DS72	72	UFP	240.42	24	54.00	70.90	93.25	183.50	323.50	483.40	600.15	1,337	1,313	202.30	2.66	11.27
	72	EI	88.78	0	1.95	6.60	26.25	71.50	108.75	179.60	320.55	551	551	95.26	2.64	8.94
	72	EO	46.40	0	0.00	0.00	7.00	26.00	62.00	130.90	172.00	287	287	59.70	2.12	5.04
	72	EQ	58.72	0	0.00	6.00	18.25	43.50	77.75	125.80	195.40	275	275	57.00	1.76	3.57
	72	ILF	39.71	0	0.00	7.00	7.00	26.00	51.25	95.80	137.75	252	252	48.18	2.50	7.58
	72	EIF	6.81	0	0.00	0.00	0.00	0.00	10.00	29.10	36.75	54	54	12.62	2.16	4.30
	72	Eff	6,134.29	167	837.80	1,150.70	2,006.75	3,347.50	7,386.25	12,850.40	19,641.00	71,729	71,562	9,135.85	5.52	37.96
DS202	202	UFP	246.90	6	38.45	53.30	90.00	183.50	312.25	552.10	731.80	1,337	1,331	224.02	2.00	4.67
	202	EI	96.99	0	4.30	8.30	26.75	66.00	131.25	235.60	363.05	551	551	100.02	1.90	3.91
	202	EO	37.56	0	0.00	0.00	5.00	21.00	49.00	94.00	137.25	287	287	50.46	2.64	8.20
	202	EQ	65.08	0	0.00	4.00	14.00	42.50	90.25	148.90	220.10	419	419	70.87	1.99	5.02
	202	ILF	37.07	0	0.00	0.00	7.00	21.00	51.25	84.00	123.55	403	403	52.75	3.65	18.31
	202	EIF	10.21	0	0.00	0.00	0.00	0.00	10.00	30.00	53.85	261	261	26.84	5.43	40.83
	202	Eff	4,726.56	167	600.70	773.80	1,288.75	2,752.00	4,999.75	9,478.10	13,299.75	71,729	71,562	7,921.06	5.74	40.35

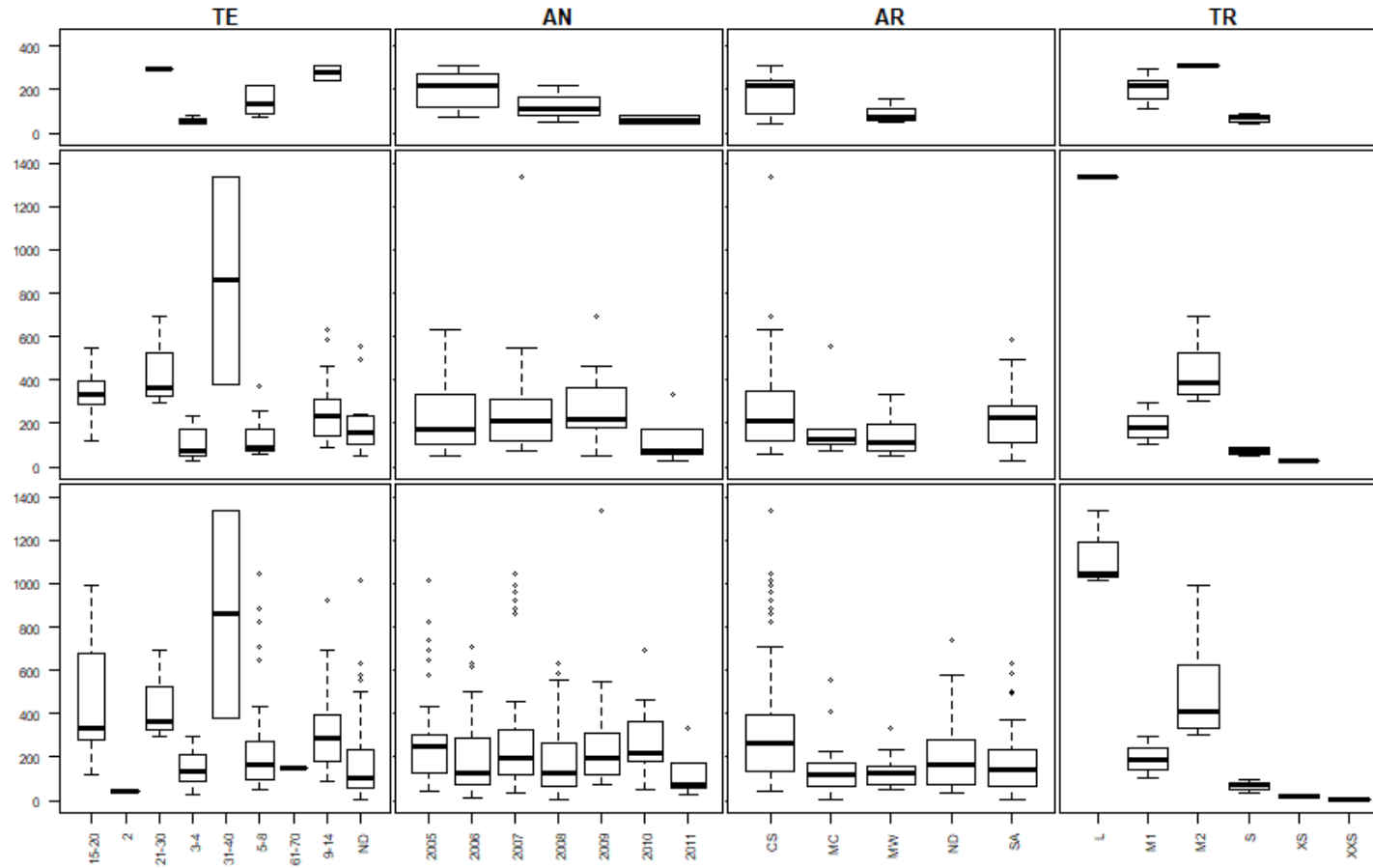


Figura 7-2 Tamaño funcional sin ajustar por variables de contexto.

Tabla 7-9 Tamaño funcional sin ajustar por variables de contexto.

DS	Vr	Niv	N	%	Mean	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rng	SD	Skw	Krt		
DS12	TL	3GL	12	1.0	158	40	40	42	77	133	237	309	-	313	273	97	0.38	-1.41		
	LP	C#	12	1.0	158	40	40	42	77	133	237	309	-	313	273	97	0.38	-1.41		
	TD	E	7	0.6	120	40	40	40	47	81	219	-	-	222	182	78	0.51	-1.86		
		N	4	0.3	236	90	90	90	128	270	309	-	-	313	223	102	-1.51	2.04		
	R(N=1, 0.1%, 112 UFP)																			
DS72	TL	3GL	46	0.6	253	24	52	73	90	198	319	562	673	1,337	1,313	235	2.54	9.14		
		4GL	16	0.2	230	59	59	69	157	191	331	426	-	461	402	118	0.59	-0.55		
		ApG	9	0.1	181	54	54	54	69	155	238	-	-	493	439	136	1.63	3.30		
		ND(N=1, 0.0%, 372 UFP)																		
	LP	PL	18	0.3	226	61	61	73	117	224	305	465	-	550	489	135	0.88	0.53		
		C#	16	0.2	283	47	47	67	90	187	309	888	-	1,337	1,290	325	2.63	7.62		
		COO	9	0.1	181	54	54	54	69	155	238	-	-	493	439	136	1.63	3.30		
		ABA	8	0.1	284	180	180	180	187	269	369	-	-	461	281	106	0.55	-1.15		
		JV	8	0.1	236	24	24	24	75	116	499	-	-	583	559	225	0.90	-1.10		
		ASN	3	0.0	122	73	73	73	73	139	-	-	-	153	80	43	-1.53	-		
		JS	2	0.0	244	89	89	89	89	244	-	-	-	399	310	219	-	-		
		NET(N=1, 0.0%, 236 UFP) AG(N=1, 0.0%, 84 UFP) ABI(N=1, 0.0%, 59 UFP) COB(N=1, 0.0%, 632 UFP)																		
		IBM(N=1, 0.0%, 372 UFP) SQL(N=1, 0.0%, 169 UFP) VB(N=1, 0.0%, 361 UFP) XML(N=1, 0.0%, 169 UFP)																		
TD	E	57	0.8	220	47	54	68	90	179	303	504	588	695	648	159	1.31	1.14			
	N	14	0.2	332	24	24	57	118	301	402	899	-	1,337	1,313	319	2.61	8.38			
	R(N=1, 0.0%, 112 UFP)																			
DS202	TL	3GL	112	0.6	266	8	47	71	102	199	328	605	861	1,337	1,329	241	1.97	4.38		
		ApG	52	0.3	155	6	15	32	54	99	239	305	495	635	629	135	1.52	2.64		
		4GL	36	0.2	315	59	71	115	163	232	392	754	908	1,048	989	240	1.68	2.30		
		ND	2	0.0	361	349	349	349	349	361	-	-	-	372	23	16	-	-		
	LP	COO	52	0.3	155	6	15	32	54	99	239	305	495	635	629	135	1.52	2.64		
		C#	38	0.2	319	47	74	88	127	262	335	837	979	1,337	1,290	279	2.09	4.58		
		PL	36	0.2	262	49	59	74	118	199	310	606	997	1,018	969	235	2.11	4.39		
		JV	29	0.1	217	8	11	24	64	128	374	583	632	649	641	207	1.04	-0.38		
		ABA	17	0.1	367	122	122	168	205	279	429	865	-	883	761	234	1.40	0.94		
		DST	4	0.0	461	117	117	117	171	340	873	-	-	1,048	931	405	1.58	2.96		
		JS	4	0.0	232	89	89	89	100	220	376	-	-	399	310	146	0.25	-3.80		
		VB	4	0.0	213	144	144	144	148	174	317	-	-	361	217	100	1.83	3.40		
		ASN	3	0.0	122	73	73	73	73	139	-	-	-	153	80	43	-1.53	-		
		PB	3	0.0	461	285	285	285	285	401	-	-	-	696	411	212	1.17	-		
		COB	2	0.0	361	90	90	90	90	361	-	-	-	632	542	383	-	-		
		SQL	2	0.0	187	169	169	169	169	187	-	-	-	204	35	25	-	-		
		NET(N=1, 0.0%, 236 UFP) AG(N=1, 0.0%, 84 UFP) ABI(N=1, 0.0%, 59 UFP) ASP(N=1, 0.0%, 109 UFP)																		
	C++(N=1, 0.0%, 146 UFP) IBM(N=1, 0.0%, 372 UFP) XML(N=1, 0.0%, 169 UFP) ND(N=1, 0.0%, 119 UFP)																			
TD	E	161	0.8	210	6	34	50	84	154	271	447	635	960	954	180	1.82	3.65			
	N	39	0.2	394	24	51	90	228	303	430	993	1,048	1,337	1,313	311	1.46	1.58			
	R	2	0.0	364	112	112	112	112	364	-	-	-	615	503	356	-	-			

TL=Tipo de lenguaje (3GL, ApG, 4GL). LP=Lenguaje programación (JV=Java, COO=COOLGen, ASN=ASPNet, C#=C#, JS=JavaScript, ABA=ABAP, PL= PLI, VB=VisualBasic, PB=PowerBuilder, ASP=ASP, SQL=SQL, NET=VisualStudioNet, DST=Datastage, NET=Net, IBM=IBMWTX, XML=XML, COB=COBOL, ABI=ABINITIO, AG=AG, C++=C++). TD=Tipo de desarrollo (E=Enhancement, N=New Development, R= Re-development).

7.1.3. COMPARACIÓN DE RESULTADOS

A continuación se realizan los análisis y se comparan los resultados con los estudios originales de la replicación.

7.1.3.1. ANÁLISIS DE DATOS

La Figura 7-3 muestra los gráficos de dispersión entre el esfuerzo de desarrollo y los puntos de función sin ajustar (UFP). Se encuentra una relación positiva entre variables para cada uno de los conjuntos de datos. Para el conjunto de datos DS12 la regresión lineal simple del esfuerzo basado en el tamaño encuentra una regresión significativa $F(1, 10) = 16.024, p < 0.003$ con un $R^2 = 0.577$ donde la ecuación de regresión lineal es $Eff = -1495 + 30.01 \times UFP$. Para el conjunto de datos DS72 la regresión lineal simple del esfuerzo basado en el tamaño encuentra una regresión significativa $F(1, 70) = 148.970, p < 0.000$ con un $R^2 = 0.676$ donde la ecuación de regresión lineal es $Eff = -2820 + 37.24 \times UFP$. Para el conjunto de datos DS202 la regresión lineal simple del esfuerzo basado en el tamaño encuentra una regresión significativa $F(2, 200) = 175.625, p < 0.000$ con un $R^2 = 0.465$ donde la ecuación de regresión lineal es $Eff = -1242 + 24.17 \times UFP$. En este caso los residuos para cada uno de los conjuntos de datos se encuentran normalmente distribuidos. Para cada uno de los análisis todas las suposiciones son revisadas. En el caso de ser necesario se realizan las transformadas de los datos para cumplir con las suposiciones de normalidad.

Es importante notar que mientras el conjunto de datos crece el coeficiente de correlación disminuye por la variabilidad de los datos. Del mismo modo, los residuos para los conjuntos de datos pequeños se ajustan mejor a la distribución normal. Los coeficientes de determinación R^2 son los comúnmente reportados en estudios de estimación de esfuerzo utilizando repositorios multicompañía. Asimismo, modelos de regresión no lineales pueden mejorar los resultados del coeficiente de determinación. Por ejemplo, para el DS12 el mo-

delo lineal obtuvo un $R^2 = 0.577$, sin embargo un modelo exponencial $R^2 = 0.681$, polinomial $R^2 = 0.673$ o de potencia $R^2 = 0.743$ obtiene un mejor resultado. En el caso del DS72 el modelo lineal obtuvo un $R^2 = 0.676$, sin embargo un modelo polinomial $R^2 = 0.826$ obtiene un mejor resultado. En el caso del DS202 el modelo lineal obtuvo un $R^2 = 0.465$, sin embargo un modelo exponencial $R^2 = 0.491$, polinomial $R^2 = 0.575$ o de potencia $R^2 = 0.527$ obtiene un mejor resultado. Todos los resultados descritos muestran evidencia de la relación entre los UFP y el esfuerzo de desarrollo.

La comparación con los estudios previos se presenta en la Tabla 7-10. Los resultados muestran la sensibilidad de los modelos de regresión dependiendo de la selección de los datos.

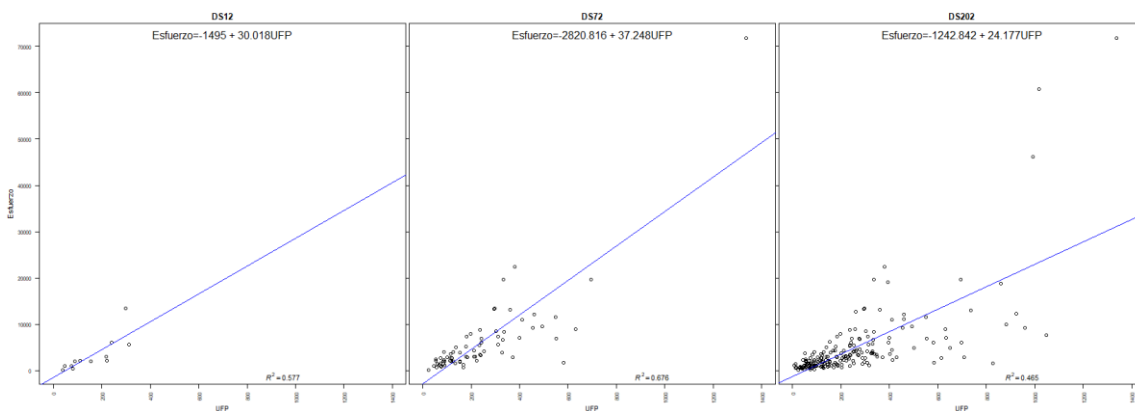


Figura 7-3 Relación de los UFP con el esfuerzo.

Tabla 7-10 Resultados de la relación de los UFP con el esfuerzo.

Autores	n	R^2	p
Albrecht & Gaffney (1983)	24	0.90	< 0.001
Kemerer (1987)	15	0.54	< 0.001
Kitchenham & Kansala (1993)	40	0.41	< 0.010
Jeffery, Low & Barnes (1993)	64	0.36	< 0.001
Jeffery & Stathis (1996)	17	0.95	< 0.001
Jeffery & Stathis (1996)	14	0.58	< 0.001
Este estudio (DS12)	12	0.58	< 0.003
Este estudio (DS12)	14	0.94	< 0.000
Este estudio (DS72)	72	0.68	< 0.000
Este estudio (DS202)	202	0.46	< 0.000

7.1.3.2. ESTRUCTURA INTERNA DE LOS PUNTOS DE FUNCIÓN IFPUG FPA

RQ1. ¿Existen correlaciones entre los puntos de función sin ajustar (UFP) y cada uno de los componentes funcionales básicos (BFC) que lo componen? ¿Existen correlaciones entre los BFC?

La Tabla 7-11 muestra los resultados del coeficiente de *Kendall Tau* entre los pares de componentes funcionales básicos (BFC) y los UFP para cada uno de los conjuntos de datos. Asimismo, se presentan los resultados de los estudios previos. Los datos indican que los BFC no son independientes entre sí y muestran que existe una relación de algunos de ellos con los UFP. Los resultados del análisis confirman la relación reportada en estudios previos entre los EI y los EQ, y los EQ y los ILF. Con respecto a la relación entre los UFP y los BFC, los resultados muestran que, en todos los estudios, los EI, EQ y los ILF tienen una correlación significativa con los UFP. Los resultados además indican una relación entre las funciones transaccionales ($TF = EI + EO + EQ$) y los UFP. Para el conjunto de datos DS12 ($Kendall\ s = 0.727, p < 0.001$), DS72 ($s = 0.860, p < 0.000$) y DS202 ($s = 0.868, p < 0.000$). Del mismo modo, una relación entre las funciones de datos ($DF = ILF + EIF$) y los UFP. Para el conjunto de datos DS12 ($s = 0.595, p < 0.007$), DS72 ($s = 0.460, p < 0.000$) y DS202 ($s = 0.544, p < 0.000$). Finalmente, se encuentran relaciones entre los TF y los DF para los conjuntos de datos DS72 ($s = 0.313, p < 0.000$) y DS202 ($s = 0.407, p < 0.000$). Esta relación es reportada también en Lavazza, Morasca & Robiolo (2013) donde indican que las TF se encuentran relacionadas con los UFP y las TF con las DF.

Los estudios previos reportan similitudes y diferencias. Kitchenham & Kansala (1993), Lokan (1999), Lavazza, Morasca & Robiolo (2013) encuentran correlaciones entre los EO y los EI tal como las reportadas este estudio (DS12, DS202), las cuales no se encuentran en Jeffery & Stathis (1996). Asimismo, encuentran una correlación entre los EO y los EQ y entre los EO y los ILF que no se encuentran en Jeffery & Stathis, ni en este estudio. Jeffery & Stathis confirma las correlaciones entre los EI y los EQ, los EI y los ILF, y los EQ y los ILF, los cuales se reportan en Kitchenham & Kansala, Lokan, Lavazza, Morasca & Robiolo y este estudio. En términos generales, los resultados muestran evidencia de la relación entre los UFP y los BFC.

Los resultados obtenidos presentan consistencias con los estudios previos; sin embargo, se encuentran diferencias reportadas que pueden ser causadas por las características de los conjuntos de datos seleccionados. Estas características pueden incluir los tipos de aplicación, lenguajes de programación y otras variables de contexto. En este caso, el análisis se limita al conjunto de variables del tamaño funcional puesto que no se cuenta con los datos originales de los estudios previos para realizar un análisis detallado.

Tabla 7-11 Coeficientes de correlación *Kendall Tau* entre los BFC y los UFP.

Autores	BFC	UFP	EI	EO	EQ	EIF
Kitchenham & Kansala (1993)	EI	0.670 p<0.001				
Jeffery & Stathis (1996)		0.540 p<0.010				
Lokan (1999)		(n.r.)				
Lavazza, Morasca & Robiolo (2013)		0.658 (n.r.)				
Este estudio (DS12)		0.740 p<0.000				
Este estudio (DS72)		0.640 p<0.000				
Este estudio (DS202)		0.650 p<0.000				
Kitchenham & Kansala (1993)	EO	0.530 p<0.001	0.470 p<0.001			
Jeffery & Stathis (1996)		0.270 (n.s.)	0.030 (n.s.)			
Lokan (1999)		(n.r.)	0.370 p<0.001			
Lavazza, Morasca & Robiolo (2013)		0.597 (n.r.)	0.438 (n.r.)			
Este estudio (DS12)		0.450 p<0.040	0.550 p<0.010			
Este estudio (DS72)		0.340 p<0.000	0.190 p<0.190			
Este estudio (DS202)		0.360 p<0.000	0.247 p<0.000			
Kitchenham & Kansala (1993)	EQ	0.470 p<0.001	0.470 p<0.001	0.320 p<0.010		
Jeffery & Stathis (1996)		0.680 p<0.001	0.720 p<0.001	-0.060 (n.s.)		
Lokan (1999)		(n.r.)	0.480 p<0.001	0.290 p<0.001		
Lavazza, Morasca & Robiolo (2013)		0.528 (n.r.)	0.448 (n.r.)	0.288 (n.r.)		
Este estudio (DS12)		0.800 p<0.000	0.610 p<0.000	0.250 p<0.270		
Este estudio (DS72)		0.540 p<0.000	0.380 p<0.000	0.030 p<0.660		
Este estudio (DS202)		0.575 p<0.000	0.384 p<0.000	0.086 p<0.760		
Kitchenham & Kansala (1993)	EIF	0.320 p<0.010	0.140 (n.s.)	0.310 p<0.010	0.600 (n.s.)	
Jeffery & Stathis (1996)		-0.370 (n.s.)	-0.560 p<0.050	0.030 (n.s.)	-0.530 p<0.050	
Lokan (1999)		(n.r.)	-0.020 (n.s.)	0.100 (n.s.)	0.000 (n.s.)	
Lavazza, Morasca & Robiolo (2013)		0.264 (n.r.)	0.072 (n.r.)	0.194 (n.r.)	0.097 (n.r.)	
Este estudio (DS12)		0.420 p<0.070	0.160 p<0.500	0.000 p<1.000	0.410 p<0.080	
Este estudio (DS72)		-0.040 p<0.690	-0.150 p<0.110	-0.270 p<0.770	-0.020 p<0.800	
Este estudio (DS202)		0.572 p<0.000	0.387 p<0.000	0.183 p<0.000	0.427 p<0.000	
Kitchenham & Kansala (1993)	ILF	0.600 p<0.001	0.510 p<0.001	0.300 p<0.010	0.310 p<0.010	0.170 (n.s.)
Jeffery & Stathis (1996)		0.730 p<0.001	0.440 p<0.050	0.110 (n.s.)	0.650 p<0.001	-0.390 (n.s.)
Lokan (1999)		(n.r.)	0.480 p<0.001	0.330 p<0.001	0.410 p<0.001	0.080 p<0.020
Lavazza, Morasca & Robiolo (2013)		0.619 (n.r.)	0.449 (n.r.)	0.417 (n.r.)	0.327 (n.r.)	0.195 (n.r.)
Este estudio (DS12)		0.660 p<0.000	0.440 p<0.050	0.190 p<0.400	0.510 p<0.020	0.560 p<0.020
Este estudio (DS72)		0.580 p<0.000	0.380 p<0.000	0.110 p<0.210	0.410 p<0.000	0.600 p<0.520
Este estudio (DS202)		0.225 p<0.000	0.040 p<0.478	0.104 p<0.060	0.150 p<0.005	0.240 p<0.000

(n.s.) no significativo. (n.r.) no reportado.

7.1.3.3. APLICABILIDAD DE LOS COMPONENTES FUNCIONALES BÁSICOS (BFC) PARA LA CREACIÓN DE MODELOS DE ESTIMACIÓN DE ESFUERZO

RQ2. ¿Existen correlaciones entre los puntos de función sin ajustar (UFP) y los componentes funcionales básicos (BFC) con el esfuerzo de desarrollo?

La Tabla 7-12 muestra los resultados de los coeficientes de correlación entre los componentes funcionales básicos (BFC), los puntos de función sin ajustar (UFP) y el esfuerzo de desarrollo. Dada la correlación entre los UFP y el esfuerzo de desarrollo y los BFC y los UFP es de interés investigar la relación que existe entre los BFC y el esfuerzo de desarrollo. Esto para evaluar la posibilidad de construir modelos de estimación de esfuerzo simplificados basados en subconjuntos de los BFC que presenten resultados de desempeño al menos como los modelos construidos a partir de los UFP.

Para cada uno de los conjuntos de datos se presentan los resultados del coeficiente de correlación de *Pearson*, *Kendall Tau* y *Spearman Rho*. Se presentan los distintos coeficientes para permitir la comparación con los estudios originales y para permitir futuras comparaciones. *Pearson* es una prueba paramétrica para la cual se requieren un conjunto de suposiciones las cuales los conjuntos de datos de este estudio no cumplen. En el caso de *Kendall Tau* y *Spearman Rho* representan una prueba paramétrica donde *Spearman Rho* puede verse afectada por valores atípicos. Asimismo, *Kendall Tau* es utilizada para muestras pequeñas. Los análisis presentados para los conjuntos de datos de este estudio se presentan basados en el coeficiente de *Kendall Tau*.

Los resultados muestran que algunos de los BFC están correlacionados con el esfuerzo. Asimismo, dependiendo de las características de los conjuntos de datos, los BFC se correlacionan de manera similar que los UFP con el esfuerzo. Los resultados del análisis confirman los resultados reportados en estudios previos, existen correlaciones similares entre el esfuerzo y los UFP, EQ, EI y los ILF. Asimismo, los resultados de esta replicación indican una relación entre las funciones transaccionales (TF), funciones de datos (DF) y el esfuerzo. En el caso de las TF, estas presentan una relación con el esfuerzo en el conjunto

de datos DS12 (*Kendall* $s = 0.667, p < 0.003$), DS72 ($s = 0.607, p < 0.000$) y DS202 ($s = 0.520, p < 0.000$). En el caso de las DF, el conjunto de datos DS202 presenta una relación con el esfuerzo ($s = 0.452, p < 0.000$).

Kitchenham & Kansala (1993) reportan las correlaciones de los EI y los EO con el esfuerzo y determinan que la combinación de los EI y los EO pueden ofrecer una mejor correlación con el esfuerzo que los UFP. Del mismo modo, Jeffery & Stathis (1996) reportan correlaciones de los EQ y los ILF con el esfuerzo. Lavazza, Morasca & Robiolo (2013) reportan que un modelo de estimación de esfuerzo basado en los EI, EO y los TF pueden ofrecer modelos con resultados similares a los basados en los UFP.

Los resultados de los estudios muestran evidencia de la relación entre los BFC y el esfuerzo de desarrollo. Estos proveen evidencia de que un modelo basado en un subconjunto de componentes funcionales básicos (BFC) puede ofrecer un resultado de estimación de esfuerzo al menos tan bueno como los modelos basados en el total de puntos de función. Dado que la efectividad de los modelos basados en los subconjuntos de los BFC es afectada por la selección de datos, es necesario construir los modelos a partir de las características específicas de los conjuntos de datos bajo evaluación.

Tabla 7-12 Coeficientes de correlación entre los BFC, el UFP y el esfuerzo.

Autores	BFC	Pearson	Kendall Tau	Spearman
Kitchenham & Kansala (1993)	UFP	0.650 p<0.001	(n.r.)	(n.r.)
Jeffery & Stathis (1996)		0.580 p<0.010	(n.r.)	(n.r.)
Este estudio (DS12)		0.785 p<0.003	0.758 p<0.001	0.916 p<0.000
Este estudio (DS72)		0.825 p<0.000	0.607 p<0.000	0.793 p<0.000
Este estudio (DS202)		0.684 p<0.000	0.550 p<0.000	0.749 p<0.000
Kitchenham & Kansala (1993)	EI	0.600 p<0.001	(n.r.)	(n.r.)
Jeffery & Stathis (1996)		0.370 p<0.001	(n.r.)	(n.r.)
Este estudio (DS12)		0.531 p<0.076	0.615 p<0.006	0.796 p<0.002
Este estudio (DS72)		0.720 p<0.000	0.484 p<0.000	0.667 p<0.000
Este estudio (DS202)		0.582 p<0.000	0.417 p<0.000	0.590 p<0.000
Kitchenham & Kansala (1993)	ILF	0.440 p<0.010	(n.r.)	(n.r.)
Jeffery & Stathis (1996)		0.730 p<0.001	(n.r.)	(n.r.)
Este estudio (DS12)		0.588 p<0.044	0.406 p<0.071	0.610 p<0.035
Este estudio (DS72)		0.622 p<0.000	0.456 p<0.000	0.613 p<0.000
Este estudio (DS202)		0.530 p<0.000	0.443 p<0.000	0.612 p<0.000
Kitchenham & Kansala (1993)	EQ	0.280 (n.s)	(n.r.)	(n.r.)
Jeffery & Stathis (1996)		0.630 p<0.001	(n.r.)	(n.r.)
Este estudio (DS12)		0.861 p<0.001	0.800 p<0.000	0.898 p<0.000
Este estudio (DS72)		0.596 p<0.000	0.416 p<0.000	0.561 p<0.000
Este estudio (DS202)		0.516 p<0.000	0.400 p<0.000	0.552 p<0.000
Kitchenham & Kansala (1993)	EO	0.660 p<0.001	(n.r.)	(n.r.)
Jeffery & Stathis (1996)		0.030 (n.s)	(n.r.)	(n.r.)
Este estudio (DS12)		0.277 p<0.383	0.394 p<0.075	0.476 p<0.118
Este estudio (DS72)		0.525 p<0.000	0.320 p<0.000	0.431 p<0.000
Este estudio (DS202)		0.469 p<0.000	0.291 p<0.000	0.406 p<0.000
Kitchenham & Kansala (1993)	EIF	0.310 (n.s)	(n.r.)	(n.r.)
Jeffery & Stathis (1996)		0.005 (n.s)	(n.r.)	(n.r.)
Este estudio (DS12)		0.857 p<0.000	0.453 p<0.055	0.602 p<0.038
Este estudio (DS72)		0.233 p<0.049	0.040 p<0.659	0.057 p<0.632
Este estudio (DS202)		0.254 p<0.000	0.262 p<0.000	0.343 p<0.000

(n.s.) no significativo. (n.r.) no reportado.

7.1.3.3.1. VALIDACIÓN DE LOS MODELOS DE ESTIMACIÓN SIMPLIFICADOS

RQ3. ¿Cuál es la exactitud de los modelos de estimación simplificados en comparación con los modelos construidos a partir de los puntos de función sin ajustar (UFP)?

Para determinar la aplicabilidad de los puntos de función se realiza la evaluación de la exactitud de los modelos de estimación de esfuerzo basados en el tamaño funcional y sus componentes funcionales básicos (BFC). Como una primera evaluación se construye un conjunto de modelos de estimación simplificados basados en subconjuntos de los BFC y se comparan con los modelos basados en los UFP aplicando parte del procedimiento de evaluación descrito en la Sección 5.3. En este caso, se seleccionan dos de los conjuntos de datos

(DS72 y DS202) y se fijan los parámetros de evaluación del procedimiento. La validación del procedimiento de evaluación de modelos y la evaluación exhaustiva de los modelos de estimación de esfuerzo se presenta en la Sección 7.3.

La pregunta que se investiga es si pueden existir modelos de estimación de esfuerzo basados en subconjuntos de los BFC del tamaño funcional con un desempeño similar a los basados en el total del tamaño funcional (UFP). El análisis de correlación indica que elementos tales como los EQ, EI, ILF y los TF se encuentran significativamente relacionados con el esfuerzo de desarrollo. Sin embargo, las características específicas de un conjunto de datos deben ser analizadas para determinar cuál subconjunto de los BFC puede ser utilizado para la construcción de los modelos simplificados. Por ejemplo, en el caso del conjunto de datos DS202, las correlaciones con el esfuerzo presentadas fueron: UFP ($s = 0.550, p < 0.000$), TF ($s = 0.520, p < 0.000$), DF ($s = 0.452, p < 0.000$), ILF ($s = 0.443, p < 0.000$) y EI ($s = 0.417, p < 0.000$). Por lo que un modelo simplificado debe construirse con base a la combinación de algunos de estos componentes.

Asimismo, es de interés para el estudio la selección de las variables independientes que expliquen la variación del esfuerzo de desarrollo. En este caso, primero se aplican las técnicas de selección de atributos provistas por las técnicas de regresión paso a paso (*stepwise*) que permite realizar la selección de las variables del contexto y segundo se realiza un análisis de comunalidades del análisis de componentes principales (PCA) para corroborar los principales componentes funcionales. Estas técnicas de reducción, permiten seleccionar un subconjunto de variables que explican la mayor parte de la variación del conjunto original (Fodor, 2002).

Basado en la selección de atributos, las variables de contexto utilizadas para crear los modelos de estimación son el tipo de desarrollo (TD), el tipo de lenguaje (TL), el lenguaje programación (LP), el tamaño relativo (TR), la plataforma de desarrollo (PD), la arquitectura (AR) y el tamaño del equipo de desarrollo (TE). Las variables de contexto nominales son pre procesadas para codificarlas con valores entre 0 y 1 (*dummy coding*). Por ejemplo, el TD cuenta con tres valores que son reemplazados por una variable ficticia (*dummy*) cada uno.

Basado en los resultados de las comunalidades del PCA, las variables de tamaño más fuertes fueron seleccionadas. Las comunalidades representan la cantidad de varianza que recogen los componentes principales seleccionados y permite examinar cómo las variables individuales reflejan las fuentes de variabilidad. Las variables con una comunalidad de al menos 6.0 representa de buena manera la variabilidad. Los resultados del análisis obtienen como mejores valores de las comunalidades los siguientes (DS202): TF ($c = 0.988$), DF ($c = 0.964$), EI ($c = 0.876$), ILF ($c = 0.796$), EIF ($c = 0.68$), EQ ($c = 0.684$) y EO ($c = 0.414$). El análisis confirma que algunos de los BFC pueden ser utilizados como predictores de los modelos de estimación de esfuerzo sin utilizar la totalidad de los UFP del conteo funcional.

Para simplificar los modelos de estimación de esfuerzo, distintos conjuntos de los componentes funcionales básicos (BFC) son utilizados para la construcción de los modelos utilizando distintas técnicas de regresión. La Tabla 7-13 lista un conjunto de modelos de estimación basados en los BFC y las variables de contexto. En este caso, se utilizan distintas combinaciones de técnicas para la creación de los modelos basado en los estudios originales de esta replicación. Primero, utilizamos solo las mediciones de tamaño funcional (UFP y BFC) y luego, las variables nominales de contexto son agregadas para analizar el ajuste de los modelos. La descripción de las técnicas utilizadas en la construcción de los modelos de estimación se realiza en la Sección 7.3. Las etapas realizadas para la construcción de los modelos son las siguientes:

- Pre procesamiento: las variables nominales son pre procesadas mediante *dummy coding* y los datos son o no transformados mediante la función logarítmica (Dejaeger, Verbeke, Martens & Baesens, 2012).
- Selección de atributos: para la selección de variables se utiliza la técnica de *forward selection* y *backward elimination* en el caso de los modelos construidos con *Stepwise regression*.
- Técnicas para la creación de los modelos: para la creación de los modelos se utilizan las técnicas *stepwise regression* (SR), regresión lineal simple (LR), regresión no lineal de potencia (NLR) y *least median of squares regression* (LMSR).

Los resultados del coeficiente de determinación de la Tabla 7-13 muestran la relación entre los puntos de función sin ajustar (UFP) y el esfuerzo de desarrollo y entre los componentes funcionales básicos (BFC) y el esfuerzo de desarrollo. Los resultados respaldan los hallazgos de los estudios originales, la evidencia sugiere que un subconjunto de BFC puede ofrecer un modelo de estimación de esfuerzo al menos con el mismo desempeño que un modelo basado en los UFP. Del mismo modo, las variables de contexto mejoran el coeficiente de determinación de los modelos de estimación. Los resultados muestran que el uso de estos atributos de contexto permite ajustar los modelos.

Por ejemplo, para el conjunto de datos DS72, un modelo de regresión basado en los UFP presenta un coeficiente de determinación de $R^2 = 0.68$ (1, a), un modelo basado en las TF presenta un $R^2 = 0.63$ (1, d) y un modelo basado en los EI, EO e ILF presenta un $R^2 = 0.65$ (1, e). Por otro lado, un modelo basado en los UFP y las variables de contexto (TD, LP, AR, TE) presenta un $R^2 = 0.87$ (2, a) y un modelo basado en los EI, EO, ILF y las variables de contexto (TL, LP, PD, AR, TE) presenta un $R^2 = 0.89$ (2, b). En el caso del conjunto de datos DS202, un modelo basado en los UFP presenta un coeficiente de determinación de $R^2 = 0.47$ (1, a), un modelo basado en las TF presenta un $R^2 = 0.42$ (1, b) y un modelo basado en los EI, EO, EQ, ILF presenta un $R^2 = 0.47$ (1, c). Por otro lado, un modelo basado en los UFP y las variables de contexto (TR, TE, LP, TD, PD) presenta un $R^2 = 0.81$ (4, a), un modelo basado en los TF y las variables de contexto (TR, TE, LP, TD) presenta un $R^2 = 0.79$ (4, b) y un modelo basado en los EI y las variables de contexto (TR, TE, TD, LP, PD) presenta un $R^2 = 0.79$ (4, c).

Para cada uno de los conjuntos de datos, los residuos se encuentran normalmente distribuidos. Asimismo, se validan las suposiciones de normalidad, multicolinealidad, linealidad y homocedasticidad según corresponda con cada técnica de creación de modelos de estimación.

Tabla 7-13 Modelos de estimación de esfuerzo basados en el UFP y los BFC.

Autores	Id	Modelo basado en:	R²	Técnica	
Kitchenham & Kansala (1993)		UFP	0.42	SR	
		EI, EO	0.50	SR	
Jeffery & Stathis (1996)		UFP	0.58	SR	
		EI, EO	(n.s)	SR	
Lavazza, Morasca & Robiolo (2013)		TF	0.74	LMSR	
		EI	0.41	LMSR	
Este estudio (DS72)	(1)	(a) UFP	0.68	SR	
		(b) EI, EO	0.56	SR	
		(c) EI	0.52	SR	
		(d) TF	0.63	SR	
		(e) EI, EO, ILF	0.65	SR	
	(2)	(a) UFP, TD, LP, AR, TE	0.87	SR	
		(b) EI, EO, ILF, TL, LP, PD, AR, TE	0.89	SR	
	(3)	(a) UFP	0.68	LMSR	
		(b) TF	0.63	LMSR	
		(c) EI	0.52	LMSR	
	Este estudio (DS202)	(1)	(a) UFP	0.47	LR
			(b) TF	0.42	LR
			(c) EI, EO, EQ, ILF	0.47	SR
			(d) EI, EO, ILF	0.44	SR
			(e) EI, ILF	0.42	SR
(f) EI			0.34	SR	
(2)		(a) UFP	0.47	LMSR	
		(b) TF	0.42	LMSR	
		(c) EI	0.34	LMSR	
(3)		(a) UFP	0.53	NLR	
		(b) TF	0.49	NLR	
		(c) EI	0.34	LMSR	
(4)		(a) UFP, TR, TE, LP, TD, PD	0.81	SR	
		(b) TF, TR, TE, LP, TD	0.79	SR	
		(c) EI, TR, TE, TD, LP, PD	0.79	SR	

EI=entradas, EO= salidas, EQ= consultas, ILF=archivos internos, EIF= interfaces externas, UFP= puntos de función sin ajustar, TF=funciones transaccionales (EI+EO+EQ),

TD=Tipo de desarrollo, LP=Lenguaje programación, AR=Arquitectura, TE=Tamaño del equipo de desarrollo, TL=Tipo de lenguaje, PD=Plataforma de desarrollo, TR=Tamaño relativo. SR: Stepwise regression, LMS: Least Median of Squares, LMSR: LMS Regression Log transformation, LR: Linear Regression, NLR: Non Linear regression (power). (n.s.) no significativo.

Para evaluar la exactitud de los modelos de estimación de esfuerzo se calculan las métricas de exactitud: magnitud del error relativo (MRE), magnitud relativa del error (MER), error relativo balanceado (BRE) y número de predicciones de los valores reales (Pred₂₅). Eso basado en los datos del conjunto DS202. Estas métricas realizan la comparación entre los residuos de los valores estimados y los valores reales para determinar el nivel de exactitud de las estimaciones. La evaluación con estas métricas se realiza dado que la presencia de una correlación no implica necesariamente que un modelo presente una exactitud aceptable en las estimaciones.

Los resultados de las métricas de exactitud se resumen en la Tabla 7-14 para un subconjunto de los modelos del conjunto de datos DS202. Los resultados confirman hallazgos previos presentados por Lavazza, Morasca & Robiolo (2013), los modelos de estimación de esfuerzo basados en subconjuntos de los BFC presentan relaciones similares a los basados en los UFP. En la práctica, se espera que los resultados para las métricas MRE, MER, BRE tiendan a cero con valores ideales cercanos a 25% y para el $Pred_{25}$ se espera que tienda a uno con valores cercanos al 75%. Tal como se reporta en Dejaeger, Verbeke, Martens & Baesens, (2012) los valores típicos para el MdmRE en estudios similares a esta replicación varían entre un 30% y 100% y para el $Pred_{25}$ entre un 10 y 60%. En esta replicación los valores de las métricas se encuentran lejanos a los valores esperados por la industria; sin embargo, son similares a los reportados por estudios previos (Lokan & Mendes, 2006; Kitchenham, Mendes & Travassos, 2006; Dejaeger, Verbeke, Martens & Baesens, 2012; Lavazza, Morasca & Robiolo, 2013).

Tabla 7-14 Evaluación de exactitud de los modelos de estimación de esfuerzo.

DS	Id	Modelo basado en:	MMRE	MdMRE	MMER	MdMER	MBRE	MdBRE	Pred ₂₅	Pred ₅₀
DS202	(1)	(a) UFP	0.82	0.55	3.56	0.51	3.94	0.74	0.25	0.47
		(b) TF	0.89	0.55	5.00	0.54	5.45	0.78	0.26	0.48
		(c) EI, EO, EQ and ILF	0.85	0.52	1.85	0.52	2.25	0.81	0.23	0.48
		(d) EI, EO and ILF	0.88	0.54	1.14	0.52	1.57	0.86	0.23	0.48
		(e) EI and ILF	0.94	0.61	2.00	0.56	2.47	0.82	0.19	0.41
		(f) EI	1.07	0.59	0.98	0.54	1.59	0.90	0.20	0.42
	(3)	(a) UFP	0.63	0.43	0.64	0.41	0.90	0.55	0.33	0.59
		(b) TF	0.67	0.44	0.67	0.44	0.96	0.55	0.29	0.58
	(4)	(a) UFP, TR, TE, LP, TD, PD	0.79	0.46	1.45	0.38	1.85	0.53	0.28	0.55
		(b) TF, TR, TE, LP, TD	0.91	0.46	0.94	0.38	1.47	0.54	0.28	0.54
		(c) EI, TR, TE, TD, LP, PD	0.89	0.55	1.17	0.47	1.63	0.75	0.25	0.47

Los resultados presentan evidencia de que por ejemplo, para el conjunto de datos DS202, un modelo de regresión basado en los UFP presenta un $MdMRE = 0.43$, $Pred_{25} = 0.33$ (3, a) y un modelo basado en los TF presenta un $MdMRE = 0.44$, $Pred_{25} = 0.29$ (3, b). Por otro lado, un modelo basado en los UFP y las variables de contexto (TR, TE, LP, TD, PD) presenta un $MdMRE = 0.46$, $Pred_{25} = 0.28$ (4, a) y un modelo basado en los TF y las variables de contexto (TR, TE, LP, TD) presenta un $MdMRE = 0.46$, $Pred_{25} = 0.28$ (4, b).

Para evaluar la exactitud entre los modelos de estimación basados en los UFP y los BFC se comparan los resultados de las métricas de exactitud. En estudios previos, la equivalencia de modelos se ha evaluado aplicando las pruebas *t*, *Wilcoxon signed rank* y *Mann-Whitney* (Kitchenham, Pfleeger, McColl & Eagan, 2002; Lavazza, Morasca & Robiolo, 2013). En este caso, se aplican las pruebas no paramétricas *Wilcoxon signed Rank* y *paired-samples sign* dado que los resultados de las métricas MRE, MER, BRE no pertenecen a una distribución normal. La Tabla 7-15 muestra los resultados de la aplicación de las pruebas que comparan los pares de modelos presentados en la Tabla 7-14, en este caso los construidos a partir de los UFP y los construidos a partir de los BFC. La hipótesis que se desea probar es si existe diferencia entre la exactitud de los modelos construidos a partir de los UFP y los BFC, de la siguiente manera:

H_0 : no existe diferencia entre la exactitud del Modelo_{*i*} y el Modelo_{*j*}

Donde el Modelo_{*i*} es un modelo construido a partir de los UFP y el Modelo_{*j*} es un modelo construido a partir de un subconjunto de los BFC. La exactitud se mide a partir de los resultados de las métricas MRE, MER, BRE.

La Tabla 7-15 presenta los resultados de las pruebas de hipótesis que indican que, para cada comparación de los modelos construidos a partir de los UFP y los subconjuntos de los BFC, con excepción de los modelos (1, a, e), (3, a, b) y (4, a, c), no existe una diferencia significativa para los resultados de las métricas de exactitud MRE, MER y BRE. Con estos resultados se confirma que existen modelos de estimación de esfuerzo basados en subconjuntos de los BFC que pueden tener el mismo nivel de exactitud que los modelos construidos a partir de los UFP. Por ejemplo, no se encuentran diferencias significativas en la comparación de los resultados de exactitud de modelos de regresión lineal basados en los UFP y los TF (1, a, b), los UFP y los EI, EO, EQ, ILF (1, a, c) y finalmente, los UFP y los EI, EO, ILF (1, a, d). De igual manera, no se encuentran diferencias significativas en la comparación de los resultados de exactitud de modelos de *stepwise regression* basados en los UFP y los TF con las variables de contexto (4, a, b).

En el caso de la comparación de los modelos (1, a, e), la prueba indica que la exactitud del modelo basado en los UFP es significativamente diferente al modelo basado en los EI y los ILF donde los resultados de la prueba para la métrica MRE son $Z = -2.748, p < 0.006$, para MER $Z = -2.154, p < 0.031$ y para BRE $Z = -2.383, p < 0.017$. En este caso, las medianas de los resultados de las métricas de exactitud del modelo basado en los UFP fueron mejores a las del modelo simplificado basado en los EI y los ILF. El mismo caso se presenta con la comparación de modelos no lineales (3, a, b) donde los resultados de las métricas de exactitud del modelo basado en los UFP fueron mejores a los del modelo simplificado basado en los TF. Por su parte, en la comparación de los modelos (4, a, c), la prueba indica que la exactitud del modelo basado en los UFP y las variables de contexto es significativamente diferente al modelo basado en los EI y las variables de contexto donde los resultados de la prueba para la métrica MRE son $Z = -3.006, p < 0.003$, para MER $Z = -2.261, p < 0.024$ y para BRE $Z = -2.602, p < 0.009$. En este caso, las medianas de los resultados de las métricas de exactitud del modelo basado en los UFP fueron mejores a las del modelo simplificado basado en los EI.

Tabla 7-15 Resultados de la prueba *Wilcoxon Signed Rank (W)* y *Sign test (S)*.

DS	Modelos			MRE		MER		BRE	
	Id	Par	Prueba	Z	p	Z	p	Z	p
DS202	(1)	(a)(b)	W	-2.022	<0.053	-0.183	<0.855	-1.087	<0.277
			S	-1.643	<0.100	-1.500	<0.134	-1.626	<0.104
	(a)(c)	W	-0.493	<0.622	-0.353	<0.724	-0.351	<0.725	
		S	-0.285	<0.776	-0.285	<0.776	-0.354	<0.724	
	(a)(d)	W	-1.114	<0.265	-0.339	<0.734	-0.334	<0.739	
		S	-0.352	<0.725	-0.282	<0.778	-0.633	<0.527	
	(a)(e)	W	-2.748	<0.006	-2.154	<0.031	-2.383	<0.017	
		S	-2.836	<0.005	-3.041	<0.002	-2.899	<0.004	
	(a)(f)	W	-2.739	<0.006	-0.901	<0.368	-1.842	<0.065	
		S	-1.843	<0.650	-0.570	<0.569	-1.129	<0.259	
	(3)	(a)(b)	W	-2.879	<0.004	-2.263	<0.024	-3.312	<0.001
			S	-1.500	<0.134	-1.872	<0.061	-1.701	<0.089
	(4)	(a)(b)	W	-0.900	<0.368	-0.452	<0.652	-0.510	<0.610
			S	-1.282	<0.200	-1.357	<0.175	-1.418	<0.156
		(a)(c)	W	-3.006	<0.003	-2.261	<0.024	-2.602	<0.009
			S	-2.203	<0.028	-2.578	<0.010	-2.333	<0.020

En la práctica, los resultados indican que es posible la construcción de algunos modelos de estimación de esfuerzo basados en subconjuntos de los componentes funcionales básicos (BFC) con resultados de exactitud similares a los obtenidos por los modelos construidos a partir de los puntos de función sin ajustar (UFP).

7.1.4. DISCUSIÓN DE RESULTADOS DE LA REPLICACIÓN

Los resultados de la replicación confirman algunos de los resultados de los estudios originales y agregan evidencia sobre la exactitud de los modelos de estimación simplificados basados en el tamaño funcional IFPUG FPA. Sin embargo, dados los distintos conjuntos de datos y las diferentes configuraciones de los estudios, no es posible generalizar los resultados totalmente. Los resultados confirman que, para los conjuntos de datos analizados, algunos de los componentes funcionales básicos (BFC) del método de medición IFPUG FPA se encuentran correlacionados. Asimismo, existe una relación entre los BFC y el esfuerzo de desarrollo de software. La evidencia sugiere que es posible construir modelos de estimación de esfuerzo basados únicamente en las funciones transaccionales (TF) o en las entradas internas (EI), las salidas externas (EO) y los archivos lógicos internos (ILF) que pueden desempeñarse al menos con el mismo rendimiento que un modelo basado en la totalidad

de los puntos de función sin ajustar (UFP). Asimismo, la utilización de las variables de contexto tales como el tipo de desarrollo (TD), la plataforma de desarrollo (PD), el lenguaje de programación (LP), el tamaño del equipo de desarrollo (TE) y el tamaño relativo (TR) pueden representar una mejora para la exactitud de los modelos.

Los resultados obtenidos podrían incidir en una reducción en el tiempo del proceso de medición del tamaño funcional. Simplificando el proceso de medición basado en el conteo de subconjuntos de los BFC es posible realizar ahorros en el esfuerzo de medición, preservando la exactitud de los modelos de estimación de esfuerzo. Para tales efectos, se recomienda que los modelos de estimación de esfuerzo simplificados sean construidos de acuerdo a las características de los proyectos históricos y el contexto específico de cada organización de desarrollo de software para lograr un mejor ajuste de los modelos. Esta investigación representa un punto de partida para la evaluación de modelos de estimación simplificados en contextos reales.

Dado los resultados obtenidos y para mejorar la generalización de los mismos, es necesario realizar estudios comparativos comprensivos para la comparación de modelos de estimación simplificados donde se puedan fijar las características de los conjuntos de datos, las configuraciones de las técnicas utilizadas, el proceso de construcción de los modelos y las métricas de comparación contra modelos de línea base. En este caso, la naturaleza de la replicación externa no permite conocer detalles de la configuración de los estudios originales. Por este mismo motivo, este estudio no considera la realización de un procedimiento de metaanálisis para la agregación de los resultados. En la Sección 7.3 se presenta una evaluación empírica exhaustiva que contempla todas las características anteriormente descritas para permitir la ejecución de un procedimiento sistemático y riguroso de comparación de modelos que permita reducir el sesgo de la selección de los datos y la configuración del procedimiento de evaluación.

7.2. EVALUACIÓN EMPÍRICA DE LOS COMPONENTES FUNCIONALES BÁSICOS DE PUNTOS DE FUNCIÓN COSMIC FFP PARA LA CONSTRUCCIÓN DE MODELOS DE ESTIMACIÓN DE ESFUERZO: UN ESTUDIO EMPÍRICO

Este estudio representa un producto obtenido del ciclo de investigación 3 (RC3).

En la replicación reportada en la Sección 7.1 se analiza la estructura interna y la aplicabilidad de puntos de función IFPUG FPA encontrando hallazgos sobre la correlación entre sus componentes funcionales básicos (BFC) y la posibilidad de construir modelos de estimación de esfuerzo simplificados basados en subconjuntos de los BFC. Dado que el método COSMIC FFP es considerado la segunda generación de los métodos de medición del tamaño funcional (Abran, 2015) es de interés evaluar la estructura interna y la aplicabilidad del método de medición COSMIC FFP empíricamente. Estudios previos, Gencel (2008), Buglione & Gencel (2008) y Gencel & Buglione (2008) investigaron la construcción de modelos de estimación de esfuerzo basados en subconjuntos de los BFC del método COSMIC FFP y cómo estos tienen un impacto en los resultados. Este estudio pretende generar evidencia acerca de la estructura interna del método de medición y la aplicabilidad de los modelos de estimación de esfuerzo construidos a partir de subconjuntos de los BFC.

Este estudio complementa el estudio reportado en la Sección 7.1 que permite el análisis de la primera pregunta de investigación (KP1) donde se estudian los desafíos y oportunidades de los métodos de medición del tamaño funcional. El objetivo del estudio (Quesada-López, Murillo-Morera, Castro-Herrera & Jenkins, 2016) es evaluar la estructura interna y aplicabilidad del método COSMIC FFP para la construcción de modelos simplificados de estimación en el contexto de aplicaciones de negocios (MIS). Para realizar el análisis empírico se seleccionan subconjuntos del repositorio de proyectos del *International Software Benchmarking Standards Group dataset (ISBSG)*. La estructura del reporte de la investigación se realiza de acuerdo a los lineamientos para estudios experimentales propuesta por (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012). En este caso, primero se describe el diseño del estudio y segundo se presenta el análisis de los resultados.

7.2.1. DISEÑO DEL ESTUDIO

El diseño del estudio y los procedimientos de análisis de datos se basan en los presentados en la Sección 7.1. Por tanto, para este estudio reportamos los aspectos más relevantes del diseño y los resultados.

En este estudio se realiza análisis empírico donde: (a) se investigan las correlaciones entre los puntos de función (CFP) y cada uno de los componentes funcionales básicos (BFC) que lo componen (E, X, R, W); asimismo, la inter correlación entre los BFC, (b) se analizan las correlaciones entre los CFP y los BFC con el esfuerzo de desarrollo y finalmente, (c) se estudian los modelos simplificados de estimación basados en los BFC y se comparan con los modelos construidos a partir de los CFP. Basado en los objetivos del estudio se analizan las siguientes preguntas de investigación:

- RQ1. *¿Existen correlaciones entre los puntos de función (CFP) y cada uno de los componentes funcionales básicos (BFC) que lo componen? ¿Existen correlaciones entre los BFC?*
- RQ2. *¿Existen correlaciones entre los puntos de función (CFP) y los componentes funcionales básicos (BFC) con el esfuerzo de desarrollo?*
- RQ3. *¿Cuál es la exactitud de los modelos de estimación simplificados en comparación con los modelos construidos a partir de los puntos de función (CFP)?*

7.2.1.1. PROCEDIMIENTO EXPERIMENTAL

El proceso general del estudio se lleva a cabo realizando las siguientes fases:

- Selección de datos y pre procesamiento: a partir del repositorio de datos ISBSG se realiza el proceso de selección y pre procesamiento para obtener los conjuntos de datos y variables utilizados en el experimento.

- Evaluación de los CFP y los BFC: se analizan los puntos de función (CFP) y los componentes funcionales básicos (BFC) para determinar la independencia entre sí y las relaciones con el esfuerzo de desarrollo.
- Evaluación de los modelos de estimación de esfuerzo: se evalúa la aplicabilidad del método de tamaño funcional en la creación de modelos de estimación simplificados y se determina el nivel de exactitud de las estimaciones.

7.2.1.2. SELECCIÓN DEL CONJUNTO DE DATOS

El conjunto de datos (DS) del estudio incluye la selección de aplicaciones de negocio del dominio de sistemas de información administrativo (MIS) medidas bajo el método de medición COSMIC FFP. Los proyectos son seleccionados de la base de datos de proyectos del ISBSG versión R12. Las instancias seleccionadas para el conjunto de datos contienen información de proyectos de desarrollo realizados en múltiples compañías. La selección del conjunto de datos (DS) se realiza bajo los criterios de calidad, el método de conteo, el año de desarrollo de los proyectos, el grupo de aplicaciones y el nivel de los recursos, tal como se muestra en la Tabla 7-16. La motivación para los criterios de selección de proyectos es: (a) selección de los proyectos medidos con el método COSMIC FFP, (b) solo los datos de proyectos con un nivel aceptable de calidad, (c) solo proyectos recientes que utilizan nuevas tecnologías, (d) del grupo de aplicaciones de negocios y (f) donde solo se reporta el esfuerzo del equipo de desarrollo.

Tabla 7-16 Criterios de selección del conjunto de datos (DS).

Método	Calidad datos	Calidad conteo	Periodo	Grupo	Nivel recursos
COSMIC	A B	A	≥ 2008	BA	1

Dado que el repositorio de proyectos ISBSG presenta proyectos con valores faltantes, se seleccionan los proyectos que reportan los puntos de función COSMIC FFP y cada uno de los cuatro componentes funcionales básicos (BFC), así como el esfuerzo de los equi-

pos de desarrollo. Las variables de tamaño funcional seleccionadas son las del tamaño funcional en puntos de función COSMIC FFP (CFP), las entradas (E), salidas (X), lecturas (R) y escrituras (W). Las variables de contexto seleccionadas para el estudio incluyen el tipo de lenguaje (TL), el lenguaje de programación (LP), el tipo de desarrollo (TD), la plataforma de desarrollo (PD), el tamaño del equipo de desarrollo (TE), la arquitectura (AR) y el tamaño relativo (TR).

Como resultado del proceso de selección inicial se obtiene un subconjunto de datos con 63 proyectos (DS63) desarrollados entre el 2008 y el 2011, tal como se muestra en la Tabla 7-17. A partir de este conjunto se realiza una validación para determinar si existen valores atípicos que puedan influenciar la efectividad de los modelos basados en regresiones lineales mediante la prueba de distancia de Cook (1977; 2000). Después del análisis, se encuentran diez valores atípicos en el subconjunto de datos los cuales fueron removidos. Como resultado obtenemos un segundo subconjunto de datos DS63 y se crea un segundo conjunto de datos que selecciona 53 proyectos (DS53). En este estudio se utilizan ambos conjuntos de datos para realizar los análisis. Es de interés evaluar cómo los valores atípicos pueden impactar los resultados. Asimismo, en el área de estudio esto es una situación común en los repositorios de proyectos que debe ser considerada para la construcción de los modelos de estimación.

Tabla 7-17 Conjuntos de datos utilizados en el estudio.

DS	Método	N	Repositorio	Conjunto de datos	Periodo	Tipo	Dom.
DS63	COSMIC FFP	63	ISBSG R12	63 proyectos del ISBSG R12	2008-2012	Multicompañía	MIS
DS53	COSMIC FFP	53	ISBSG R12	53 proyectos del ISBSG R12	2008-2012	Multicompañía	MIS

La Tabla 7-18 presenta el detalle de las estadísticas descriptivas de los sub conjuntos de datos seleccionados para las variables: puntos de función (CFP), entradas (E), salidas (X), lecturas (R), escrituras (W) y esfuerzo (Eff). Para la mayoría de las variables los datos muestran que los conjuntos de datos tienen un *skewness* con simetría positiva que indica que la cantidad de proyectos pequeños y medianos es mayor que la cantidad de proyectos gran-

des. En el caso del *kurtosis* las variables muestran valores positivos. Las pruebas de normalidad *Kolmogorov-Smirnov* y *Shapiro-Wilk* indican que las variables en los distintos conjuntos de datos no pertenecen a una distribución normal. La descripción detallada de los conjuntos de datos es necesaria dado que los resultados del estudio tienen que ser analizados con respecto a los conjuntos de datos seleccionados.

La Figura 7-4 y la Figura 7-5 muestran la distribución del tamaño funcional de los subconjuntos de datos agrupados por las variables de contexto. La Figura 7-4 muestra las distribuciones del tamaño funcional para el conjunto de datos DS63 en la sección superior y para el conjunto de datos DS53 en la sección inferior para las variables TL=Tipo de lenguaje (3GL, ApG, 4GL), LP=Lenguaje programación (JV=Java, COO=COOLGen, ASN=ASPNet, C#=C#, JS=JavaScript, ABA=ABAP, PL=PLI, VB=VisualBasic, PB=PowerBuilder, ASP=ASP, SQL=SQL, NET=VisualStudioNet, DST=Datastage, NET=Net, IBM=IBMWTX, XML=XML, COB=COBOL, ABI=ABINITIO, AG=AG, C++=C++), TD=Tipo de desarrollo (E=Enhancement, N=New Development, R= Re-development) y PD=Plataforma de desarrollo (MP=Multi-Platform, MF=Main Frame, PC=PC, MR=Mid-Range). Por su parte, la Figura 7-5 muestra las distribuciones del tamaño funcional para las variables TE=Tamaño del equipo de desarrollo (2, 3-4, 5-8, 9-14, 15-20, 21-30, 31-40, 61-70), AN=Año (2005, 2006, 2007, 2008, 2009, 2010, 2011), AR=Arquitectura (CS=Cliente servidor, SA=Stand alone, MC=Multicapa, MW=Multicapa Web, ND=No definido) y TR=Tamaño relativo (1.XXS, 2.XS, 3.S, 4.M1, 5.M2, 6.L).

La Tabla 7-19 lista las estadísticas descriptivas de un conjunto de las variables de contexto del conjunto de datos DS63 y DS53. En el Anexo 6 se detallan las estadísticas para cada una de las variables de contexto. Las variables de contexto utilizadas en la evaluación se deben conocer ya que estas pueden influenciar el esfuerzo de desarrollo requerido para desarrollar los proyectos a partir de su tamaño funcional. Asimismo, las características de los conjuntos de datos influyen los resultados de la evaluación de los modelos de estimación de esfuerzo. Para cada uno de los conjuntos de datos se detalla la variable de contexto (Vr), el nivel de la variable (Niv), la cantidad de proyectos (n), el mínimo (Min), el

máximo (Max), el promedio (Mean), la mediana (Med), el rango (Rng), la desviación estándar (SD), los percentiles (Q), la *kurtosis* (Krt) y el *Skewness* (Skw).

Tabla 7-18 Estadística descriptiva de los conjuntos de datos seleccionados.

DS	N	Var	Prom	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rango	Std.D.	Skew	Kurt
DS63	63	CFP	151.57	2	2.20	3.00	6.00	74.00	187.00	476.60	748.60	853	851	214.15	1.89	3.02
	63	E	31.48	1	1.00	1.00	2.00	13.00	39.00	76.40	189.40	255	254	51.96	2.99	9.55
	63	X	46.27	-	-	1.00	1.00	14.00	61.00	154.40	221.80	287	287	69.29	1.89	3.08
	63	R	50.11	-	-	1.00	2.00	20.00	64.00	187.40	197.80	304	304	71.40	1.75	2.31
	63	W	23.71	-	-	-	-	6.00	33.00	66.80	109.80	193	193	36.76	2.49	7.46
	63	Eff	3,050.03	24	27.20	40.00	123.00	721.00	2,929.00	9,256.00	18,698.40	29,246	29,222	5,713.07	2.84	8.63
DS53	53	CFP	97.34	2	2.00	3.00	5.50	43.00	123.00	293.60	481.90	579	577	140.07	2.07	4.05
	53	E	17.19	1	1.00	1.00	1.50	6.00	23.50	53.20	64.70	96	95	22.16	1.72	2.65
	53	X	28.26	-	-	0.40	1.00	3.00	31.00	107.60	134.90	231	231	47.30	2.43	6.42
	53	R	33.81	-	-	1.00	1.50	16.00	46.00	94.20	177.80	198	198	49.38	2.03	3.77
	53	W	18.08	-	-	-	-	4.00	31.00	45.00	65.90	193	193	31.11	3.73	18.87
	53	Eff	2,400.06	24	24.00	40.00	77.00	348.00	1,420.50	8,874.60	17,179.40	21,700	21,676	4,761.95	2.83	7.91

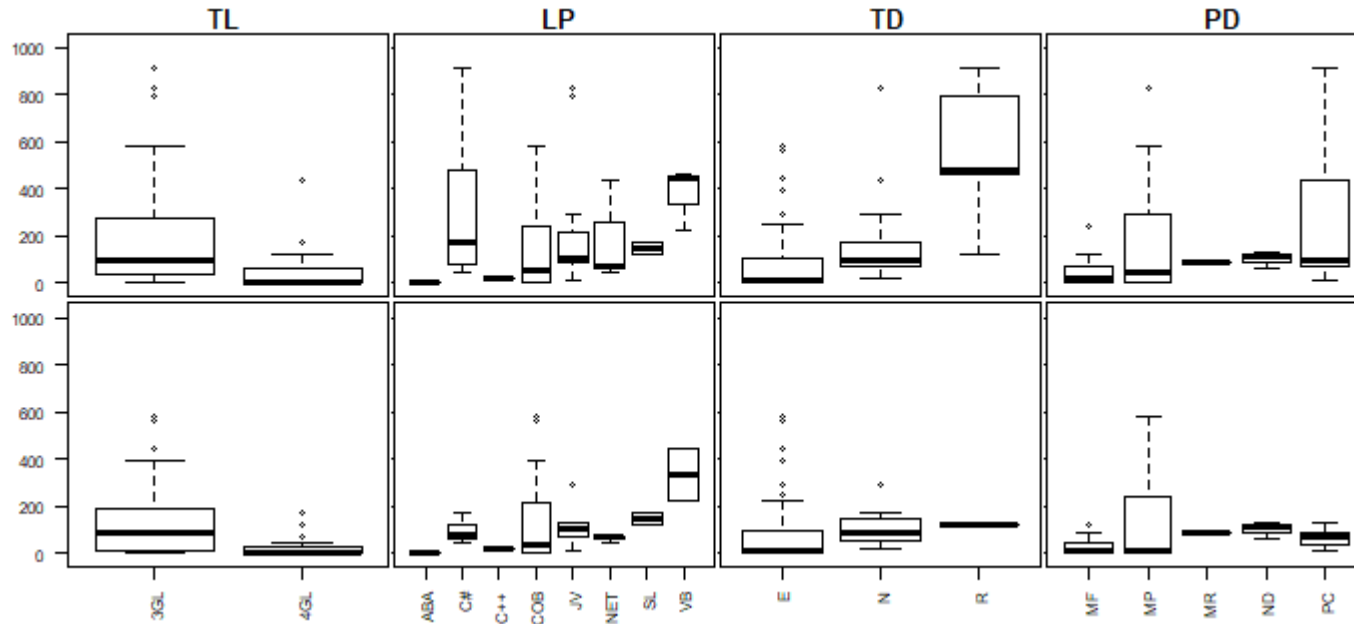


Figura 7-4 Tamaño funcional por variables de contexto.

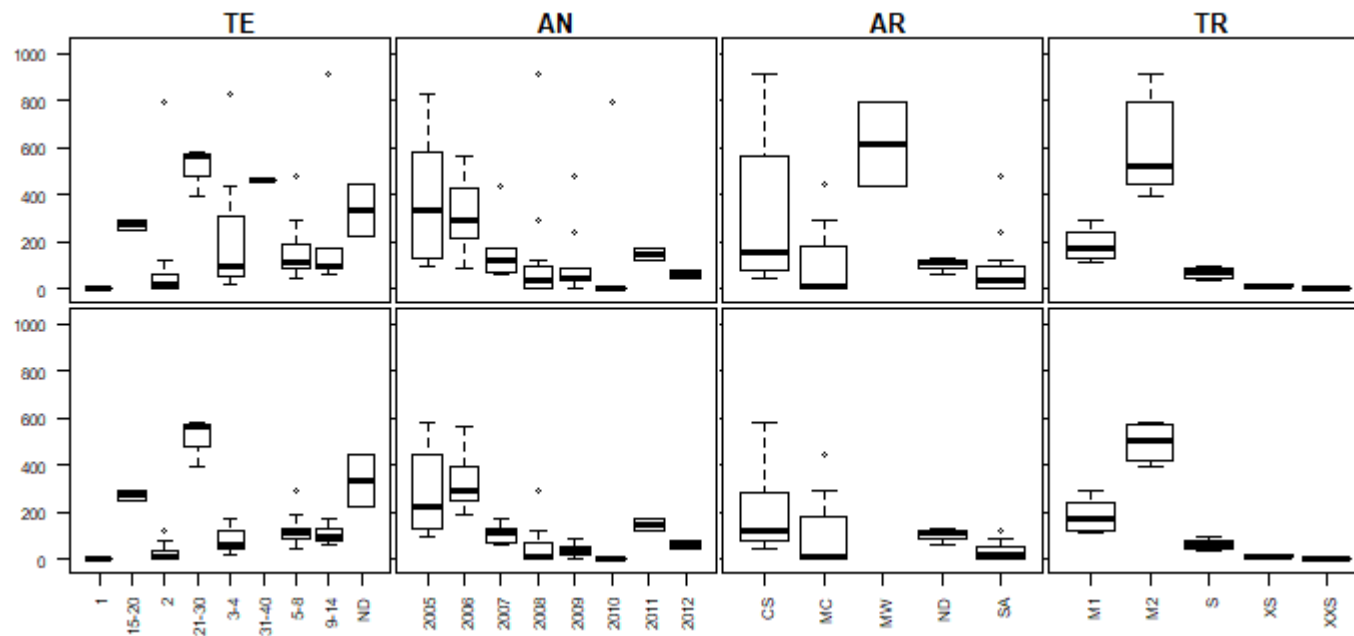


Figura 7-5 Tamaño funcional por variables de contexto.

Tabla 7-19 Tamaño funcional por variables de contexto.

DS	Vr	Niv	N	%	Mean	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rng	SD	Skw	Krt	
DS63	TL	3GL	43	0.7	200	3	3	5	33	98	293	572	819	912	909	240	1.58	1.79	
		4GL	20	0.3	50	2	2	2	3	6	68	169	428	441	439	103	3.22	11.46	
	LP	COB	22	0.3	137	3	3	3	6	50	246	512	576	579	576	179	1.50	1.35	
		ABA	14	0.2	5	2	2	2	3	4	6	9	-	9	7	2	0.66	-0.76	
	C#	JV	12	0.2	228	14	14	28	89	105	253	816	-	826	812	279	1.83	2.02	
		C#	5	0.1	338	45	45	45	64	169	698	-	912	867	364	1.25	0.63		
		NET	4	0.1	160	48	48	48	55	75	350	-	-	441	393	188	1.97	3.92	
		VB	3	0.0	380	226	226	226	226	448	-	-	-	467	241	134	-1.69	-	
		SL	2	0.0	147	118	118	118	118	147	-	-	-	175	57	40	-	-	
	C++(N=1, 0.0%, 23 FP)																		
	TD	E	43	0.7	97	2	2	3	5	11	118	356	538	579	577	153	1.98	3.28	
		N	15	0.2	178	23	23	35	62	98	175	595	-	826	803	209	2.49	6.65	
		R	5	0.1	554	118	118	118	293	483	852	-	-	912	794	311	-0.35	-0.55	
DS53	TL	3GL	34	0.6	135	3	3	5	13	84	197	423	566	579	576	160	1.57	1.77	
		4GL	19	0.4	29	2	2	2	3	6	48	118	-	175	173	48	2.07	3.91	
	LP	COB	20	0.4	134	3	3	3	6	39	234	545	578	579	576	186	1.53	1.27	
		ABA	14	0.3	5	2	2	2	3	4	6	9	-	9	7	2	0.66	-0.76	
	C#	JV	8	0.2	116	14	14	14	68	105	130	-	-	294	280	82	1.57	3.85	
		C#	3	0.1	99	45	45	45	45	82	-	-	-	169	124	64	1.10	-	
		NET	3	0.1	66	48	48	48	48	74	-	-	-	75	27	15	-1.72	-	
		VB	2	0.0	337	226	226	226	226	337	-	-	-	448	222	157	-	-	
		SL	2	0.0	147	118	118	118	118	147	-	-	-	175	57	40	-	-	
	C++(N=1, 0.0%, 23 FP)																		
	TD	E	41	0.8	93	2	2	3	5	9	109	377	550	579	577	155	2.05	3.47	
		N	11	0.2	110	23	23	27	48	86	169	270	-	294	271	79	1.33	1.91	
		R(N=1, 0.0%, 118 FP)																	

(Continúa)

TL=Tipo de lenguaje (3GL, ApG, 4GL). LP=Lenguaje programación (JV=Java, COO=COOLGen, ASN=ASPNet, C#=C#, JS=JavaScript, ABA=ABAP, PL=PLI, VB=VisualBasic, PB=PowerBuilder, ASP=ASP, SQL=SQL, NET=VisualStudioNet, DST=Datastage, NET=Net, IBM=IBMWTX, XML=XML, COB=COBOL, ABI=ABINITIO, AG=AG, C++=C++). TD=Tipo de desarrollo (E=Enhancement, N=New Development, R= Re-development).

7.2.2. RESULTADOS DEL ESTUDIO

A continuación se presentan los principales resultados del estudio empírico.

7.2.2.1. ANÁLISIS DE DATOS

Los conjuntos de datos DS63 y DS53 muestran una relación entre el esfuerzo de desarrollo y el tamaño funcional. Para el conjunto de datos DS53, la regresión lineal simple del esfuerzo basado en el tamaño funcional encuentra una regresión significativa $F(1, 51) = 296.188, p < 0.000$ con un $R^2 = 0.853$ donde la ecuación de regresión lineal es $Eff = -656.48 + 31.40 \times FP$. Para el conjunto de datos DS63, la regresión lineal simple del

esfuerzo basado en el tamaño encuentra una regresión significativa $F(1, 61) = 21.237, p < 0.000$ con un $R^2 = 0.258$ donde la ecuación de regresión lineal es $Eff = 995.20 + 13.56 \times FP$. En este caso, los residuos para cada uno de los conjuntos de datos se acercan a una distribución normalmente distribuida. Es importante notar que mientras el conjunto de datos crece, el coeficiente de correlación disminuye por la variabilidad de los datos.

Es posible observar que existen modelos de regresión no lineales que pueden mejorar los resultados de coeficiente de determinación. Por ejemplo, para el conjunto DS53 el modelo lineal obtuvo un $R^2 = 0.853$, sin embargo un modelo de potencia $R^2 = 0.862$ obtiene un mejor resultado. En el caso del DS63, el modelo lineal obtuvo un $R^2 = 0.258$ y un modelo de potencia $R^2 = 0.744$ obtiene un mejor resultado. Los resultados obtenidos muestran evidencia de la relación entre los CFP y el esfuerzo de desarrollo.

7.2.2.2. CONSISTENCIA INTERNA DE LOS PUNTOS DE FUNCIÓN COSMIC FFP

RQ1. ¿Existen correlaciones entre los puntos de función (CFP) y cada uno de los componentes funcionales básicos (BFC) que lo componen? ¿Existen correlaciones entre los BFC?

La Tabla 7-20 muestra los coeficientes de *Kendall Tau* entre cada uno de los pares de los componentes funcionales básicos (BFC) y entre los BFC y los CFP. Los resultados indican que los BFC no son independientes entre sí. Los resultados confirman la relación entre todos los BFC (E, X, R y W). Del mismo modo, se encuentra una relación entre los CFP y los BFC. Los resultados muestran que las correlaciones más fuertes se presentan entre los CFP y las entradas (E), las salidas (X) y las escrituras (X). Asimismo, entre las lecturas (R) y las escrituras (X). Aunque se encuentran relaciones entre los BFC y los CFP, estos resultados deben ser analizados a partir de las características de cada uno de los conjuntos de datos que incluyen los tipos de aplicación, lenguajes de programación y otras variables de contexto. Por tanto, la reducción del método de medición debe limitarse a los conjuntos de datos para los cuales se realiza el análisis.

Tabla 7-20 Coeficientes de correlación Kendall Tau entre los BFC y el CFP.

Autores	BFC	CFP	E	X	R
Este estudio (DS63)	E	0.843 p<0.000			
	X	0.804 p<0.000	0.739 p<0.000		
	R	0.853 p<0.000	0.710 p<0.000	0.698 p<0.000	
	W	0.764 p<0.000	0.694 p<0.000	0.583 p<0.000	0.759 p<0.000
Este estudio (DS53)	E	0.830 p<0.000			
	X	0.770 p<0.000	0.680 p<0.000		
	R	0.863 p<0.000	0.709 p<0.000	0.678 p<0.000	
	W	0.774 p<0.000	0.703 p<0.000	0.550 p<0.000	0.759 p<0.000

7.2.2.3. APLICABILIDAD DE LOS COMPONENTES FUNCIONALES BÁSICOS (BFC) PARA LA CREACIÓN DE MODELOS DE ESTIMACIÓN DE ESFUERZO

RQ2. ¿Existen correlaciones entre los puntos de función (CFP) y los componentes funcionales básicos (BFC) con el esfuerzo de desarrollo?

La Tabla 7-21 muestra los resultados de los coeficientes de correlación entre los componentes funcionales básicos (BFC), los puntos de función (CFP) y el esfuerzo de desarrollo. Dada la correlación existente entre los CFP y el esfuerzo de desarrollo y entre los BFC y los CFP es de interés investigar la relación que existe entre los BFC y el esfuerzo de desarrollo. Esto con el fin de evaluar la posibilidad de la construcción de modelos de estimación de esfuerzo simplificados basados en subconjuntos de los BFC que puedan presentar resultados de exactitud similares a los modelos construidos a partir del total de los CFP.

Los resultados muestran que los BFC se encuentran correlacionados con el esfuerzo de desarrollo. Asimismo, algunos de los BFC se correlacionan con el esfuerzo con resultados similares a la correlación entre los CFP y el esfuerzo. Por ejemplo, el DS63 presenta relaciones similares entre los BFC y el esfuerzo. En el caso del conjunto de datos DS53, los W y los R presentan relaciones muy similares a las presentadas por los CFP con el esfuerzo. Los resultados confirman los datos reportados en estudios previos que indican que existen relaciones entre el esfuerzo de desarrollo y los BFC (Buglione & Gencel, 2008; Gencel & Buglione, 2008). La evidencia sugiere la posibilidad de construir modelos de estimación de esfuerzo basados en subconjuntos de los BFC donde estos puedan ofrecer resultados similares a los basados en el conteo total de los CFP. En la práctica, dado que la efectividad de

los modelos basados en los subconjuntos de los BFC es afectada por la selección de datos, es necesario construirlos a partir de las características específicas de los conjuntos de datos bajo evaluación.

Tabla 7-21 Coeficientes de correlación entre los BFC, el CFP y el esfuerzo.

Autores	BFC	Kendall Tau	Spearman
Este estudio (DS63)	CFP	0.666 p<0.000	0.847 p<0.000
	E	0.629 p<0.000	0.824 p<0.000
	X	0.600 p<0.000	0.761 p<0.000
	R	0.625 p<0.000	0.809 p<0.000
	W	0.641 p<0.000	0.824 p<0.000
Este estudio (DS53)	CFP	0.758 p<0.000	0.911 p<0.000
	E	0.685 p<0.000	0.862 p<0.000
	X	0.646 p<0.000	0.769 p<0.000
	R	0.728 p<0.000	0.888 p<0.000
	W	0.726 p<0.000	0.882 p<0.000

7.2.2.3.1. VALIDACIÓN DE LOS MODELOS DE ESTIMACIÓN SIMPLIFICADOS

RQ3. ¿Cuál es la exactitud de los modelos de estimación simplificados en comparación con los modelos construidos a partir de los puntos de función (CFP)?

Para determinar la aplicabilidad de los puntos de función se realiza la evaluación de la exactitud de los modelos de estimación de esfuerzo basados en el tamaño funcional y sus componentes funcionales básicos (BFC). Como una primera evaluación se construye un conjunto de modelos de estimación simplificados basados en subconjuntos de los BFC y se comparan con los modelos basados en el total de los CFP aplicando el procedimiento descrito en la Sección 5.3. En este caso, para los conjuntos de datos (DS53 y DS63) se seleccionan las variables de análisis y se fijan los parámetros de evaluación del procedimiento. La validación del procedimiento de evaluación de modelos y la evaluación exhaustiva de los modelos de estimación de esfuerzo se presenta en la Sección 7.3.

La pregunta que se investiga es si existen modelos de estimación de esfuerzo basados en subconjuntos de los BFC del tamaño funcional que puedan presentar un desempeño similar a los modelos basados en el total del tamaño funcional (CFP). Dados los resultados previos en que todos los BFC presentan una correlación con el esfuerzo de desarrollo, se

evalúan distintas combinaciones posibles de modelos de estimación simplificados extraídos a partir de los subconjuntos de datos. Asimismo, es de interés para el estudio la selección de las variables independientes que expliquen la variación del esfuerzo de desarrollo. En este caso, se aplican las técnicas de selección de atributos provistas por las técnicas de regresión paso a paso (*stepwise*) que permiten realizar la selección de las variables del contexto. Basado en la selección de atributos, las variables de contexto utilizadas para crear los modelos de estimación son el tipo de desarrollo (TD), el tipo de lenguaje (TL), el lenguaje programación (LP), el tamaño relativo (TR), la plataforma de desarrollo (PD), la arquitectura (AR) y el tamaño del equipo de desarrollo (TE). Las variables de contexto nominales son pre procesadas para codificarlas con valores entre 0 y 1 (*dummy coding*).

Para simplificar los modelos de estimación de esfuerzo, distintos conjuntos de los componentes funcionales básicos (BFC) son usados para la construcción de modelos utilizando distintas técnicas de regresión. La Tabla 7-22 lista un conjunto de modelos de estimación basados en los BFC y las variables de contexto. En este caso, se utilizan distintas combinaciones de técnicas para la creación de los modelos. Primero, se analizan modelos basados solo en las mediciones de tamaño funcional (CFP y BFC) y luego se analizan modelos que incorporan las variables nominales de contexto para evaluar el ajuste de los modelos. La descripción de las técnicas utilizadas en la construcción de los modelos de estimación se realiza en la Sección 7.3. Las etapas realizadas para la construcción de los modelos son las siguientes:

- Pre procesamiento: las variables nominales son pre procesadas mediante *dummy coding* y los datos son o no transformados mediante la función logarítmica.
- Selección de atributos: para la selección de variables se utilizan las técnicas de *forward selection* y *backward elimination* en el caso de los modelos construidos con *Stepwise regression*
- Técnicas para la creación de los modelos: para la creación de los modelos se utilizan las técnicas *stepwise regression* (SR), regresión lineal simple (LR), regresión no lineal de potencia (NLR) y *least median of squares regression* (LMSR).

Los resultados de la Tabla 7-22 muestran la relación entre los puntos de función (CFP) y el esfuerzo de desarrollo y entre los componentes funcionales básicos (BFC) y el esfuerzo de desarrollo. La evidencia sugiere que modelos basados en un subconjunto de los BFC pueden ofrecer resultados de exactitud al menos similares en comparación con un modelo basado en los CFP. Por ejemplo, para el conjunto de datos DS63, un modelo de regresión basado en los CFP presenta un coeficiente de determinación de $R^2 = 0.258$ (1, a), un modelo basado en los BFC presenta un $R^2 = 0.471$ (1, b) y un modelo basado en los W presenta un $R^2 = 0.338$ (1, d). En el caso del conjunto de datos DS53, un modelo basado en los CFP presenta un coeficiente de determinación de $R^2 = 0.853$ (1, a), un modelo basado en los BFC presenta un $R^2 = 0.862$ (1, b) y un modelo basado en los E, W presenta un $R^2 = 0.856$ (1, g). En este caso, los residuos de los conjuntos de datos se acercan a una distribución normalmente distribuida y para cada una de las técnicas se chequean las suposiciones de normalidad, multicolinealidad, linealidad y homocedasticidad según corresponda a cada técnica de creación de modelos de estimación.

Tabla 7-22 Modelos de estimación de esfuerzo basados en el CFP y los BFC.

Autores	Id	Modelo basado en:	R ²	Técnica
Este estudio (DS63)	(1)	(a) CFP	0.258	LR
		(b) E, X, R, W	0.471	SR
		(c) E, X, W	0.471	SR
		(d) W	0.338	SR
	(2)	(a) CFP	0.744	NLR
		(3) (a) CFP, TR, TE, LP, TD, PD, AR	0.552	SR
Este estudio (DS53)	(1)	(b) E, X, R, W, TR, TE, LP, TD, PD, AR	0.396	SR
		(a) CFP	0.853	LR
		(b) E, X, R, W	0.862	SR
		(c) E, X, W	0.861	SR
		(d) E, R, W	0.858	SR
		(e) E, R	0.841	SR
		(f) X, W	0.856	SR
		(g) E, W	0.856	SR
	(2)	(a) CFP	0.825	SR
		(3) (a) CFP, TR, TE, LP, TD, PD, AR	0.862	NLR
(3)	(a) CFP, TR, TE, LP, TD, PD, AR	0.836	SR	
	(b) E, X, R, W, TR, TE, LP, TD, PD, AR	0.645	SR	

E=entradas, X= salidas, R= lecturas, W= escrituras, CFP= puntos de función.

SR: Stepwise regression, LR: Linear Regression, NLR: Non Linear regression (power).

Para evaluar la exactitud de los modelos de estimación de esfuerzo, se analizan las métricas de exactitud: magnitud del error relativo (MRE), coeficiente de correlación de Spearman (SP), exactitud estandarizada (SA) y número de predicciones de los valores reales (Pred₂₅) basado en los datos del conjunto DS53. Basado en los resultados de estas métricas se realizan las comparaciones entre los residuos de los valores estimados y los valores reales de los modelos para determinar el nivel de exactitud de las estimaciones. En este caso, se aplican las pruebas no paramétricas *Wilcoxon signed Rank* para las comparaciones dado que los resultados de las métricas no pertenecen a una distribución normal.

La Tabla 7-23 muestra los resultados de la aplicación de las pruebas estadísticas que comparan un subconjunto de los modelos presentados en la Tabla 7-22. La hipótesis que se desea probar es si existe diferencia entre la exactitud de los modelos construidos a partir de los CFP y los modelos construidos a partir de los BFC, de la siguiente manera:

H_0 : no existe diferencia entre la exactitud del Modelo_i y el Modelo_j

Donde el $Modelo_i$ es un modelo construido a partir del total de los puntos de función (CFP) y el $Modelo_j$ es un modelo construido a partir de los componentes funcionales básicos (BFC). La exactitud se mide a partir de los resultados de las métricas MRE, SP, $Pred_{25}$, SA. La exactitud estandarizada (SA) es una métrica recomendada para la comparación de los modelos de estimación que permite obtener valores sin sesgo y a partir del conjunto de datos bajo evaluación (Shepperd & MacDonell, 2012; Whigham, Owen & Macdonell, 2015; Dolado, Rodriguez, Harman, Langdon & Sarro, 2016). Esta se calcula comparando los valores estimados con los valores obtenidos por la línea base y representa el promedio de los residuos absolutos contra la línea base de selección aleatoria (*random guessing*) (Dolado, Rodriguez, Harman, Langdon & Sarro, 2016). Un valor cercano a cero no es útil puesto que estima similar a una adivinación aleatoria. Los modelos de estimación deben maximizar el valor de la métrica SA (Shepperd & MacDonell, 2012).

En la Tabla 7-23 se listan los resultados de las pruebas de hipótesis con la comparación de modelos construidos a partir de los CFP y subconjuntos de los BFC. Los resultados indican que no se encuentra una diferencia significativa entre los modelos comparados para la métrica de exactitud MRE. Estos resultados indican que existen modelos de estimación de esfuerzo basados en subconjuntos de los BFC que pueden tener un nivel de exactitud similar a los modelos construidos a partir de los CFP. Por ejemplo, no se encuentran diferencias significativas en la comparación de los resultados de exactitud de modelos de regresión lineal basados en los CFP y los BFC (1, a, b), los CFP y los X, W (1, a, f) y finalmente, los CFP y los E, W (1, a, g). De igual manera, no se encuentran diferencias significativas en la comparación de la métrica SA para el modelo basado en los CFP y los X, W (1, a, f). Sin embargo, los resultados de las demás métricas para los modelos (1, a, b), (1, a, f) y (1, a, g) presentan diferencias significativas. En este caso, las medianas de los resultados de las métricas de exactitud del modelo basado en los CFP fueron mejores a las de los modelos simplificados.

En el caso de los modelos basados en los CFP y BFC que utilizan las variables de contexto como predictores, no se encuentran diferencias significativas en la comparación de los resultados de exactitud de modelos de regresión lineal (3, a, b). Por lo que, en este

caso, no se encuentra evidencia de que las variables de contexto mejoran el desempeño de los modelos de estimación. En el caso del modelo basado en los CFP que utiliza o no las variables de contexto, los resultados de la prueba para la métrica MRE son $Z = -3.086, p < 0.700$. Para el modelo basado en los BFC que utiliza o no las variables de contexto los resultados de la prueba para la métrica MRE son $Z = -1.943, p < 0.052$.

Tabla 7-23 Resultados de la prueba *Wilcoxon Signed Rank (W)*.

DS	Modelos		MRE		SP		Pred ₂₅		SA	
	Id	Par	Z	p	Z	p	Z	p	Z	p
DS53	(1)	(a)(b)	-0.972	0.331	-4.295	0.000	-2.147	0.032	-3.086	0.002
		(a)(f)	-0.514	0.607	-3.671	0.000	-2.149	0.032	-1.629	0.103
		(a)(g)	-0.198	0.843	-4.214	0.000	-2.188	0.029	-2.857	0.004
	(3)	(a)(b)	-0.543	0.587	-0.888	0.375	-0.476	0.634	-1.657	0.097

En la práctica, los resultados indican que es posible explorar la utilización de modelos de estimación de esfuerzo basados en subconjuntos de los componentes funcionales básicos (BFC) con resultados de exactitud similares (MRE) a los obtenidos por los modelos construidos a partir de los puntos de función (CFP). Sin embargo, la simplificación se debe realizar con cautela donde los modelos sean construidos con datos específicos de las organizaciones que permitan evaluar el ajuste y los resultados de exactitud.

7.2.3. DISCUSIÓN DE RESULTADOS DEL ESTUDIO

Los resultados del estudio agregan evidencia sobre la exactitud de los modelos de estimación simplificados basados en el tamaño funcional COSMIC FFP. La evidencia indica que los componentes funcionales básicos (BFC) del método de medición COSMIC FFP se encuentran correlacionados. Asimismo, existe una relación entre los BFC y el esfuerzo de desarrollo de software. Sumado a esto, los resultados del coeficiente de determinación indican que los modelos de estimación basados en los BFC podrían tener un desempeño similar a los basados en los CFP. La evidencia sugiere la posibilidad de construir modelos de

estimación de esfuerzo basados únicamente en subconjuntos de los componentes funcionales básicos (BFC). Por ejemplo, un modelo de regresión basado en las salidas (X) y las escrituras (W) podría desempeñarse al menos con la misma exactitud que un modelo basado en la totalidad de los puntos de función (CFP). De la misma forma, un modelo basado en los CFP y los BFC utilizando o no variables de contexto tales como el tipo de desarrollo (TD), la plataforma de desarrollo (PD), el lenguaje de programación (LP), el tamaño del equipo de desarrollo (TE) y el tamaño relativo (TR) parece presentar el mismo nivel de exactitud. Esto es, para el conjunto de datos analizado, no se evidencia que las variables de contexto mejoran el desempeño del modelo de estimación.

Por otro lado, los resultados obtenidos presentan variaciones dependiendo de las métricas de exactitud comparadas, los conjuntos de datos seleccionados y las configuraciones aplicadas para la construcción de los modelos. Por lo tanto, no es posible generalizar los resultados. En la práctica, la simplificación del proceso de medición podría hacer posible un ahorro en el esfuerzo de medición preservando la exactitud de los modelos. Sin embargo, para realizar la simplificación y ajustar los modelos, las organizaciones deben construirlos de acuerdo a las características de sus proyectos históricos y en el contexto específico de sus proyectos de desarrollo.

Esta investigación representa un punto de partida para la evaluación de modelos de estimación simplificados en el contexto de COSMIC FFP. Para analizar los factores que pueden afectar el ajuste de los modelos simplificados es necesario realizar estudios comparativos comprensivos donde se puedan fijar las características de los conjuntos de datos, las configuraciones de las técnicas utilizadas, el proceso de construcción de los modelos y las métricas de comparación contra modelos de línea base. En la Sección 7.3 se presenta una evaluación empírica exhaustiva para la comparación de modelos que permite reducir el sesgo de la selección de los datos y la configuración del procedimiento de evaluación.

7.3. ANÁLISIS DE PUNTOS DE FUNCIÓN PARA LA CREACIÓN DE MODELOS DE ESTIMACIÓN DE ESFUERZO: UN ESTUDIO EMPÍRICO

Este estudio representa el producto obtenido del ciclo de investigación 3 (RC10).

En los últimos años, una gran cantidad de técnicas han sido utilizadas para la construcción de modelos de estimación del esfuerzo de desarrollo del software (Dejaeger, Verbeke, Martens & Baesens, 2012; Huang, Li & Xie, 2015). Este estudio analiza la efectividad del procedimiento de automatización para la evaluación de los resultados de medición en la construcción de los modelos de estimación propuesto en la Sección 5.3. Para ejecutar la recolección de datos se utiliza el módulo descrito en la Sección 5.4.3 que permite realizar un análisis comparativo exhaustivo que contempla la combinación de distintas técnicas de pre procesamiento de datos (DP), selección de atributos (AS) y algoritmos de aprendizaje (LA) bajo distintos conjuntos de datos. Esta sección complementa el análisis de la tercera pregunta de investigación (KP3) donde se estudia la efectividad de la metodología para la automatización. Asimismo, complementa los estudios reportados en la Sección 7.1 y la Sección 7.2 realizando una evaluación exhaustiva de los modelos de estimación de esfuerzo. Este estudio realiza una validación del procedimiento para la evaluación de los resultados de medición en la construcción de los modelos de estimación que permite realizar el aprovechamiento de los resultados de medición de tamaño funcional.

El objetivo de este estudio (Murillo, J., Quesada-López, C., Castro-Herrera, C. & Jenkins, 2016; 2017) es validar el procedimiento para la evaluación de los resultados de medición en la construcción de modelos de estimación de esfuerzo de desarrollo. El procedimiento de evaluación recibe como entrada los resultados del tamaño funcional y los factores de contexto de los proyectos de desarrollo. Por ejemplo, las mediciones podrían ser recolectadas por el procedimiento de medición de la metodología **aFPA** durante la aplicación del proceso de medición y ser almacenadas en repositorios de datos históricos con las métricas de los proyectos de desarrollo de software. En este caso, para realizar el análisis se seleccionan subconjuntos de datos del repositorio de proyectos del *International Software*

Benchmarking Standards Group dataset (ISBSG) y se presenta una evaluación empírica exhaustiva para la comparación de modelos de estimación que busca determinar su desempeño en el contexto de aplicaciones de negocios (MIS). La estructura del reporte de la investigación se realiza de acuerdo a los lineamientos para estudios experimentales propuesta por (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012). En este caso, primero se describe el diseño del estudio y segundo se presentan los análisis de los resultados.

7.3.1. DISEÑO DEL ESTUDIO

Para realizar el estudio experimental se aplica el proceso descrito por el procedimiento de evaluación de modelos detallado en la Sección 5.3. Por tanto, para este estudio reportamos los aspectos más relevantes del diseño y las decisiones tomadas en la configuración para la evaluación de modelos. El estudio realiza análisis empírico exhaustivo donde (a) se valida la estabilidad de los resultados del procedimiento para la construcción de modelos de estimación de esfuerzo basados en el tamaño funcional, (b) se estudian los modelos simplificados de estimación de esfuerzo y (c) se reportan las técnicas que presentan los mejores resultados para los conjuntos de datos estudiados. Basado en los objetivos del estudio se analizan las siguientes preguntas de investigación:

- RQ1. *¿El procedimiento de evaluación de modelos con la configuración seleccionada permite obtener resultados de exactitud repetibles²⁰ a través de múltiples corridas?*
- RQ2. *¿Cuál es la exactitud²¹ de los modelos de estimación simplificados construidos por el procedimiento de evaluación?*
- RQ3. *¿Cuáles son las técnicas que obtienen los mejores resultados de exactitud para los modelos de estimación de esfuerzo construidos por el procedimiento de evaluación?*

²⁰ Repetibilidad: precisión de medida bajo un conjunto de condiciones iguales. Proximidad entre las mediciones repetidas del mismo objeto.

²¹ Exactitud: proximidad entre un valor medido y un valor verdadero.

Las respuestas a las preguntas de investigación se analizan en la sección de resultados.

7.3.1.1. PROCEDIMIENTO EXPERIMENTAL

El proceso general del estudio se lleva a cabo realizando las siguientes fases:

- Selección de datos y pre procesamiento: a partir del repositorio de proyectos ISBSG R12 se realiza el proceso de selección de datos y el pre procesamiento para obtener los subconjuntos de datos y variables utilizadas en el experimento.
- Evaluación de los modelos de estimación de esfuerzo: se evalúa la aplicabilidad del método de tamaño funcional IFPUG FPA y COSMIC FFP en la creación de modelos de estimación simplificados y se determina el nivel de exactitud de las estimaciones.

Para la evaluación de la exactitud se construye un conjunto de modelos de estimación aplicando el procedimiento descrito en la Sección 5.3. En este caso, la configuración particular ejecutada es la siguiente:

- La selección del conjunto de datos se realiza a partir del procedimiento descrito en los estudios de la Sección 7.1 y de la Sección 7.2. Cada uno de los conjuntos de datos (DS202 y DS53) es dividido manualmente para obtener distintos grupos que contengan subconjuntos de los componentes funcionales básicos (BFC) y variables de contexto que permitan realizar las comparaciones de efectividad de los modelos simplificados. Por ejemplo, conjuntos de datos que solo contengan el tamaño funcional total, subconjuntos de BFC y/o variables de contexto.
- Las variables de contexto (atributos nominales) son pre procesadas para codificarlas con valores entre 0 y 1 (*dummy coding*). No se aplica ningún tratamiento para los valores faltantes de las variables de contexto. Se aplican técnicas de distancia para la identificación de los valores atípicos.

- En total se utilizan 2 técnicas de pre procesamiento de datos (DP), 5 técnicas de selección de atributos (AS) y 15 métodos de algoritmos de aprendizaje (LA) que son combinadas para la construcción de los modelos de estimación ($DP \times AS \times LA$). La selección de las técnicas utilizadas se detalla en la Sección 7.3.1.3. En el caso de la técnica de transformación de *BoxCox* se utilizan 6 parámetros para el λ .
- Cada sub conjunto de datos se evalúa por separado. Un sub conjunto de datos se divide aleatoriamente en dos conjuntos, un conjunto de datos de entrenamiento (90%) y un conjunto de datos de prueba (10%). Para la división de los datos se realiza un proceso de una ronda de validación cruzada $M \times N$ (10×10).
- Se aplica el enfoque de *filter* para la reducción de las variables de los conjuntos de datos. Se selecciona este enfoque dado que, aunque no ofrece resultados de exactitud tan buenos como un enfoque de *wrapper*, es menos costoso computacionalmente. En este caso, para la comparación de los modelos solo es necesario fijar este factor.
- Por cada uno de los sub conjuntos de datos se construye un modelo para cada uno de los esquemas de aprendizaje ($DP \times AS \times LA$). La configuración de cada una de las técnicas utilizadas se detalla en la Sección 7.3.1.3 las cuales se mantienen fijas.
- En esta evaluación, para cada uno de los sub conjuntos de datos se ejecuta un $N_PASS = 10$ donde para cada pasada de la evaluación se calculan y promedian las métricas de exactitud. En total se realizan mil ejecuciones ($N_PASS = 10$, $M \times N$ donde $M = 10$ y $N = 10$) por esquema de aprendizaje ($DP \times AS \times LA$).
- La exactitud de los modelos de estimación de esfuerzo se analiza basada en las métricas de exactitud: magnitud del error relativo (MRE), coeficiente de correlación de Spearman (SP), exactitud estandarizada (SA) y número de predicciones de los valores reales ($Pred_{25}$).

7.3.1.2. SELECCIÓN DEL CONJUNTO DE DATOS

La selección de los conjuntos de datos es uno de los aspectos más importantes en los estudios de estimación. Estos pueden influir directamente en la exactitud de las estimaciones dependiendo de la representatividad de los proyectos históricos, la calidad de los datos, el porcentaje de los valores faltantes, entre otros (Dejaeger, Verbeke, Martens & Baesens, 2012; Huang, Li & Xie, 2015). Por tal motivo, los modelos de estimación deben ser analizados con respecto a las características específicas de los conjuntos de datos seleccionados (Dejaeger, Verbeke, Martens & Baesens, 2012). Para mayor detalle sobre el repositorio de proyectos utilizado, la descripción general del ISBSG se encuentra en (ISBSG, 2013).

Los conjuntos de datos (DS) utilizados en los análisis de este estudio (DS202 y DS53) son los seleccionados en la Sección 7.1 y la Sección 7.2. Los proyectos seleccionados son aplicaciones de negocio del dominio de sistemas de información administrativo (MIS) medidas bajo los métodos de medición IFPUG FPA y COSMIC FFP. La Tabla 7-24 muestra los parámetros de selección utilizados. Se seleccionan los proyectos con calidad A o B, del método de medición IFPUG FPA y COSMIC FFP y del año 2005 o superior. La Tabla 7-25 lista los conjuntos de datos seleccionados como resultado del proceso. Los criterios de selección específicos, las características de los proyectos, las estadísticas descriptivas y el análisis de los conjuntos de datos se detallan en la Sección 7.1 y la Sección 7.2. Este estudio incorpora las variables de contexto de los proyectos tales como el tipo de desarrollo (TD), el tipo de lenguaje (TL), el lenguaje programación (LP), la plataforma de desarrollo (PD), la arquitectura (AR) y el tamaño del equipo de desarrollo (TE).

Tabla 7-24 Criterios de selección de conjuntos de datos (DS).

DS	Método	Calidad datos	Calidad conteo	Periodo	Grupo	Nivel recursos
DS202	IFPUG FPA 4+	A	A B	≥ 2005	BA	1
DS53	COSMIC FFP 2+, 3+	A B	A B	≥ 2005	BA	1

Tabla 7-25 Conjuntos de datos utilizados en el estudio.

DS	Método	N	Repositorio	Conjunto de datos	Periodo	Tipo	Dom.
DS202	IFPUG FPA	202	ISBSG R12	202 proyectos del ISBSG R12	2005-2011	Multicompañía	MIS
DS53	COSMIC FFP	53	ISBSG R12	53 proyectos del ISBSG R12	2008-2012	Multicompañía	MIS

Cada uno de los conjuntos de datos (DS202 y DS53) es dividido manualmente para obtener subconjuntos de datos que contengan distintos grupos de variables de interés para los análisis específicos. Por ejemplo, combinaciones de conjuntos de datos que contengan el total del tamaño de los puntos de función, el tamaño de los sub conjuntos de los componentes funcionales básicos (BFC), el tamaño relativo (TR) y/o las variables de contexto (TD, TL, LP, PD, AR, TE). La división manual se realiza para obtener distintos grupos que contengan subconjuntos de los componentes funcionales básicos (BFC) y variables de contexto que permitan realizar las comparaciones de efectividad de los modelos simplificados.

La división de los datos permite evaluar cada uno de los modelos simplificados por separado y realizar las comparaciones de efectividad en comparación con los modelos basados en el total de puntos de función. Los conjuntos de datos del método de medición IFPUG FPA (DS202) se dividen en 21 grupos y los conjuntos de datos del método de medición COSMIC FFP (DS53) se dividen en 9 grupos, tal como se lista en la Tabla 7-26. En el caso de los conjuntos de datos IFPUG FPA se utiliza el tamaño de los puntos de función sin ajustar (UFP) y el tamaño de los sub conjuntos de los BFC (EI, EO, EQ, ILF, EIF). En el caso de los conjuntos de datos COSMIC FFP se utiliza el tamaño de los puntos de función (CFP) y el tamaño de los sub conjuntos de los BFC (E, X, R, W). La selección de las variables predictoras de tamaño funcional se realiza con base en los resultados presentados en los estudios descritos en la Sección 7.1 y la Sección 7.2 que indican que modelos de estimación simplificados basados en conjuntos de los BFC del tamaño funcional podrían ofrecer un desempeño en la exactitud similar a los modelos de estimación de esfuerzo basados en el total de los puntos de función.

Tabla 7-26 Grupos de datos utilizados en el estudio.

		DS202				DS53	
Grupo	VARIABLES	Grupo	VARIABLES	Grupo	VARIABLES	Grupo	VARIABLES
1	UFP-Eff	8	BFC,DF,TF-Eff	15	EI,EO,EQ,ILF-Eff	1	FP,TR,Cxt-Eff
2	UFP,TR,Cxt-Eff	9	BFC-Eff	16	EI,EO,ILF-Eff	2	FP,Cxt-Eff
3	UFP,Cxt-Eff	10	DF,TF,TR,Cxt-Eff	17	EI,ILF-Eff	3	FP-Eff
4	BFC,TR,DF,TF,Cxt-Eff	11	DF,TF,Cxt-Eff	18	EI,EO-Eff	4	BFC,TR,Cxt-Eff
5	BFC,DF,TF,Cxt-Eff	12	DF,TF-Eff	19	EI-Eff	5	BFC,Cxt-Eff
6	BFC,TR,Cxt-Eff	13	TF-Eff	20	TR,Cxt-Eff	6	BFC-Eff
7	BFC,Cxt,Eff	14	TF,ILF-Eff	21	TR-Eff	7	E,W-Eff
						8	X,W-Eff
						9	W-Eff

7.3.1.3. TÉCNICAS UTILIZADAS PARA LA CONSTRUCCIÓN DE MODELOS

La Tabla 7-27 presenta la descripción general de cada una de las técnicas utilizadas en la construcción de los modelos durante la aplicación de la estrategia de evaluación. Las técnicas de pre procesamiento de datos (DP) permiten limpiar el ruido, eliminar datos faltantes y valores atípicos, transformar las variables, entre otros. Las técnicas de selección de atributos (AS) permiten seleccionar los atributos predictores más apropiados que explican las variaciones para predecir la variable de salida. Finalmente, los algoritmos de aprendizaje (LA) son los modelos que permiten realizar las estimaciones. Las principales categorías de algoritmos utilizados son las funciones y regresiones, bayesianos, árboles, reglas, redes, entre otros. Los algoritmos utilizados en la evaluación son los implementados por las librerías de *WEKA* para *Java* (Witten, Frank, Hall & Pal, 2016). Del mismo modo, los detalles de las técnicas utilizadas pueden ser encontrados en (Witten, Frank, Hall & Pal, 2016).

Tabla 7-27 Técnicas utilizadas en la construcción de los modelos de estimación.

Técnica	Descripción
<i>Pre procesamiento de datos (DP)</i>	
1. Sin procesamiento (<i>None</i>)	Los conjuntos de datos se utilizan sin ninguna transformación.
2. Logarítmica (<i>Log</i>)	Los valores numéricos son reemplazados por la transformada logarítmica. Es una técnica de transformación de potencia para reducir anomalías como la no normalidad y heterocedasticidad.
3. <i>Box-Cox</i> (<i>BC</i>)	Los valores numéricos son reemplazados por la transformada de <i>Box-Cox</i> ($\lambda=-2, \lambda=-1, \lambda=-0.5, \lambda=0.5, \lambda=1, \lambda=2$). Es una técnica de transformación de potencia para reducir anomalías como la no normalidad y heterocedasticidad.
<i>Selección de atributos (AS)</i>	
1. <i>GeneticSearch</i> (<i>GS</i>)	Selecciona los atributos basado en un algoritmo genético simple. Incluye los parámetros del tamaño de la población, número de generaciones, <i>crossover</i> y mutación.
2. <i>BestFirst</i> (<i>BF</i>)	Es una estrategia de selección de atributos que minimiza el costo para alcanzar el objetivo. Empieza con una selección vacía y busca hacia adelante o empieza con el conjunto de atributos y busca hacia atrás. Es posible especificar la cantidad de nodos consecutivos que no mejoran antes de hacer el retroceso.
3. <i>LinearForwardSelection</i> (<i>LFS</i>)	Presenta una extensión de <i>BestFirst</i> que considera un número restringido de atributos restantes. La técnica selecciona los mejores <i>k</i> atributos.
4. <i>BackwardElimination</i> (<i>BE</i>)	Esta técnica selecciona los atributos más relevantes con un esquema de eliminación hacia atrás. Empieza con el conjunto de variables completo y va eliminando los atributos por iteración mientras esto represente un mejor rendimiento.
5. <i>ForwardSelection</i> (<i>FS</i>)	Esta técnica empieza con un conjunto vacío de variables y solo los atributos que incrementan el rendimiento son agregados. Primero selecciona el atributo que represente un mejor rendimiento e itera con los atributos restantes hasta encontrar el mejor conjunto.
<i>Algoritmos de aprendizaje (LA)</i>	
1. <i>GaussianProcesses</i> (<i>GP</i>)	Este algoritmo implementa un proceso de redes Bayesianas/Gaussianas para regresiones no lineales. El método es equivalente al <i>kernel ridge regression</i> . Este aplica procesos de normalización y de reemplazo de valores faltantes.
2. <i>LeastMedSq</i> (<i>LMS</i>)	Esta técnica representa una regresión lineal robusta que es menos vulnerable a los valores atípicos. El procedimiento minimiza las medias de los cuadrados que divergen de la línea de regresión aplicando repetidamente la regresión lineal a subconjuntos de datos para obtener el menor error cuadrado medio.
3. <i>LinearRegression</i> (<i>LR</i>)	Esta técnica ejecuta la clásica <i>Ordinary Least Squares multiple regression</i> y opcionalmente puede realizar el proceso de selección de atributos hacia adelante o hacia atrás.
4. <i>MultilayerPerceptron</i> (<i>MP</i>)	Esta técnica implementa una red neuronal con propagación hacia atrás para clasificar las instancias. Los nodos de salida se convierten en unidades lineales sin umbral.
5. <i>RBFNetwork</i> (<i>RBFN</i>)	Este algoritmo implementa una red Gaussiana de función lineal normalizada. Esta usa algoritmos de <i>k-means clustering</i> y regresión logística donde realiza la clasificación midiendo la similitud de las entradas.
6. <i>SMOreg</i> (<i>SMO</i>)	Esta técnica implementa máquinas de soporte vectorial para la regresión donde los parámetros se pueden aprender a partir de varios algoritmos. El método implementa un algoritmo secuencial de optimización mínima para el modelo de regresión vectorial.

(continúa)

Técnica	Descripción
7. <i>AdditiveRegression</i> (AR)	Este algoritmo busca mejorar el rendimiento de la regresión lineal donde en cada iteración ajusta el modelo a los residuos de la iteración previa. La estimación se realiza agregando las predicciones de cada uno de los modelos construidos.
8. <i>Bagging</i> (BGN)	Algoritmo para reducir la variación y evitar el sobre ajuste mediante la evaluación de conjuntos de entrenamiento. Puede realizar clasificaciones o regresión dependiendo del algoritmo base.
9. <i>ConjunctiveRule</i> (CR)	El algoritmo implementa la regla conjuntiva para predicción de clases numéricas. La regla consiste en el conjunto de antecedentes y el consecuente valor de clase para la clasificación o la regresión. Para la regresión se utiliza la información del promedio ponderado de los errores de los cuadrados medios cubiertos y no cubiertos por la regla.
10. <i>DecisionTable</i> (DT)	Este algoritmo permite utilizar tablas de decisión como clasificador. En este caso, consiste de una tabla jerárquica en la que cada entrada de una tabla de alto nivel se desglosa por los valores de un par de atributos para formar otra tabla.
11. <i>M5Rules</i> (M5R)	Este algoritmo genera una lista de decisión para problemas de regresión bajo el enfoque de separar y conquistar. En cada iteración se construye un árbol utilizando M5 y convierte la mejor hoja en una regla.
12. <i>ZeroR</i> (ZR)	Es un algoritmo que predice la categoría basado en la moda o el promedio dependiendo de la clase del atributo.
13. <i>DecisionStump</i> (DS)	Es un algoritmo que se utiliza con el método <i>boosting</i> . Construye un árbol de decisión de un nivel. Realiza la decisión (<i>decision stump</i>) basado en una sola característica de entrada.
14. <i>M5P</i> (M5P)	Esta técnica combina árboles de clasificación con funciones de regresión lineal en los nodos. Primero se construye el árbol de clasificación y las hojas se ajustan a una regresión lineal basados en el tratamiento de la desviación estándar. Este es una extensión de <i>Classification And Regression Trees</i> (CART) para valores continuos.
15. <i>REPTree</i> (RT)	Este algoritmo construye un árbol de decisión/regresión usando la información de la varianza y lo poda utilizando el criterio de reducción del error. Los valores faltantes se los trata dividiendo las instancias en segmentos.

Las configuraciones de parámetros específicas utilizadas para cada una de las técnicas implementadas por los algoritmos de aprendizaje (LA) se presentan en la Tabla 7-28. Estas se seleccionan a partir de las recomendaciones presentadas por la documentación de las librerías de *WEKA* para *Java* (Witten, Frank, Hall & Pal, 2016) las cuales fueron utilizadas para la implementación del procedimiento de evaluación.

Tabla 7-28 Configuraciones de las técnicas utilizadas.

Id	Configuración
GP	<i>Level of Gaussian Noise=1, normalize/standardize/neither=normalize, Kernel=PolyKernel [Exponent=1.0, lower-order terms=no, size of the cache=250007], Random number seed=1, decimal places=2.</i>
LMS	<i>Sample size=4, seed used to generate samples=0.</i>
LR	<i>Attribute selection=None, ridge parameter=1.0e-8, output additional statistics.</i>
MP	<i>Learning Rate for the backpropagation algorithm=0.3, Momentum Rate for the backpropagation algorithm=0.3, Number of epochs to train=500, Percentage size of validation set to use to terminate training=0, value used to seed the random number=0, consecutive number of errors allowed for validation before the network terminates=20, NominalToBinary.</i>
RBFN	<i>Number of clusters=2, random seed to be used by K-means=1, ridge for linear regression, maximum number of iterations for the logistic regression=-1, minimum standard deviation for the clusters=0.1.</i>
SMO	<i>Complexity=1, 0=normalize/1=standardize/2=neither=0, Optimizer=RegSMOImproved [tolerance=0.001, variant=1, round-off error epsilon=1.0e-12, epsilon-insensitive loss function=1.0e-3, random number seed=1], Kernel=PolyKernel [debugging=off, checks=on, cache=250007, Exponent=1.0, lower-order terms=no].</i>
AR	<i>Shrinkage rate=1.0, iterations=10, Minimize absolute error instead of squared error, classifier=DecisionStump.</i>
BGN	<i>Size of each bag=100, Random number seed=1, Number of execution slots=1, Number of iterations=10, classifier=REPTree [Number of instances=2, minimum variance for split=1e-3, number of folds=3, seed=1, Maximum tree depth=-1].</i>
CR	<i>Number of folds=3, randomization, exclusive expressions for nominal attributes=false, minimal weights of instances within a split=2.0, number of antecedents=-1, seed=1.</i>
DT	<i>Search method=BestFirst [Direction of search=forward], number of folds=1, evaluation measure for selecting attributes=RMSE.</i>
M5R	<i>Unpruned tree/rules, unsmoothed predictions, regression tree/rule, minimum number of instances per leaf=4.</i>
ZR	<i>Debug mode.</i>
DS	<i>Classifier=DecisionStump, debug mode.</i>
M5P	<i>Unpruned tree/rules, unsmoothed predictions, regression tree/rule, minimum number of instances per leaf=4, save instances at the nodes in the tree.</i>
RT	<i>Number of instances=2, minimum variance for split=1e-3, number of folds=3, seed=1, Maximum tree depth=-1.</i>

7.3.2. RESULTADOS DEL ESTUDIO

A continuación se presentan los principales resultados del estudio empírico.

7.3.2.1. ANÁLISIS DE DATOS

Las estadísticas descriptivas de los resultados de la exactitud de los modelos se presentan para mostrar las tendencias de los datos, para señalar aspectos importantes sobre su comportamiento que determinan las decisiones sobre los análisis realizados y para apoyar la interpretación de los resultados presentados en las secciones posteriores. La Tabla 7-

29 presenta el detalle de las estadísticas descriptivas de los resultados de la exactitud estandarizada (SA) de los modelos para cada uno de los sub conjuntos de datos (DS) y los grupos evaluados (Grupo). En el Anexo 7 se presentan las métricas de exactitud de la magnitud del error relativo (MRE), el coeficiente de correlación de Spearman (SP), el número de predicciones de los valores reales (Pred₂₅) y la exactitud estandarizada (SA). Para realizar los cálculos se consideran todas las combinaciones de los modelos construidos en la evaluación. Para cada uno de los grupos de datos se detalla la métrica de exactitud (SP, MRE, MdMRE, SA, Pred₂₅), la cantidad de modelos construidos (n) y el mínimo (Min), el máximo (Max), el promedio (Mean), la mediana (Med), el rango (Rng), la desviación estándar (SD), los percentiles (Q), la *kurtosis* (Krt) y el *Skewness* (Skw) de cada una de las métricas.

Estudios previos indican que se espera que los resultados de las métricas de exactitud obtenidos por la evaluación de los modelos tiendan a cero para la MRE con valores cercanos a 25% y que oscilen entre el 30% y 100% para sus medias. En el caso del SP se espera que los valores tiendan a uno con valores cercanos al 70% o superior. Para el Pred₂₅ se espera que tienda a uno con valores cercanos al 75% y que oscilen entre un 10% y 60%. Finalmente, para la SA se espera maximizar su valor, ya que valores cercanos a cero significan que el modelo estima similar a una adivinación aleatoria. En estudios previos, la SA ha presentado valores que oscilan entre el 30% y 60% con valores que han alcanzado hasta el 80% lo que significa que el modelo podría superar la línea base de selección aleatoria hasta por un 80%.

Tal como se visualiza en la Figura 7-6, los datos recolectados presentan una alta variabilidad de los resultados de las métricas de exactitud dependiendo de los conjuntos de datos y los grupos de variables utilizados en la construcción de los modelos. En esta figura se muestran los resultados de los modelos (DS) basados en el método IFPUG FPA en la parte superior y los modelos basados en el método COSMIC FFP en la parte inferior de la figura para cada uno de los grupos detallados en la Tabla 7-26.

En el caso de la MRE y media del MRE existen modelos que han alcanzado valores cercanos a cero, para el SP existen valores que han alcanzado el ~86%, en el caso de la SA

se han alcanzado valores cercanos al 80% y finalmente, en el caso del Pred₂₅ se han alcanzado valores del 38%. Sin embargo, también existen modelos con una exactitud deficiente, por lo que es de interés encontrar los modelos que permiten los mejores resultados por cada uno de los conjuntos de datos. El análisis de los mejores modelos por conjunto de datos se realiza en las secciones posteriores. Del mismo modo, tal como se muestra en la Figura 7-6, el análisis de las métricas de exactitud para las técnicas de construcción de modelos (LA) presentan variaciones dependiendo de las combinaciones utilizadas $DP \times AS \times LA$ y entre los conjuntos de datos evaluados. Las tendencias de las métricas presentadas indican que los resultados de exactitud de una combinación de técnicas de $DP \times AS \times LA$ depende del conjunto de datos y los grupos de variables con los cuales se construyen y evalúan los modelos de estimación. Para cada una de las técnicas (LA) se muestran los resultados de los modelos basados en el método IFPUG FPA en la parte superior y los resultados de los modelos basados en el método COSMIC FFP en la parte inferior de la figura.

En este estudio se comparan los modelos de estimación obtenidos por conjunto de datos utilizando una estrategia de evaluación sin sesgo. Por lo tanto, la optimización y configuración de parámetros de las técnicas utilizadas para la construcción de los modelos es considerado como trabajo futuro. En este estudio, aunque los valores de las métricas se encuentran lejanos a los valores esperados por la industria; los modelos con mejor desempeño presentan valores similares a los reportados por estudios previos (Lokan & Mendes, 2006; Kitchenham, Mendes & Travassos, 2006; Dejaeger, Verbeke, Martens & Baesens, 2012; Lavazza, Morasca & Robiolo, 2013; Minku & Yao, 2013; Dolado, Rodriguez, Harman, Langdon & Sarro, 2016).

Tabla 7-29 Estadística descriptiva de los resultados de evaluación de exactitud.

DS	Grupo	Métrica	N	Prom	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rango	Std.D.	Skew	Kurt
DS202	1	SA	600	41.189	13.556	27.532	29.132	34.663	42.515	48.244	51.274	53.415	58.794	45.237	8.740	-0.401	-0.636
	2	SA	600	43.601	3.605	29.112	32.016	35.583	44.695	51.778	54.990	56.681	61.469	57.864	9.605	-0.613	0.559
	3	SA	600	43.601	3.496	29.094	31.698	35.361	44.150	51.845	55.141	57.083	63.043	59.547	9.873	-0.616	0.624
	4	SA	600	40.089	-79.680	24.640	27.812	31.610	41.295	48.476	53.761	55.602	61.568	141.249	11.850	-2.159	17.119
	5	SA	600	40.013	1.935	23.883	27.265	32.068	41.176	48.149	52.912	56.126	60.945	59.010	10.645	-0.542	0.358
	6	SA	600	39.594	-1.815	25.340	27.993	32.251	41.166	47.209	51.604	53.461	61.020	62.835	10.077	-0.780	1.049
	7	SA	600	39.203	-22.451	25.488	28.058	31.764	40.343	46.302	51.779	54.567	60.606	83.058	10.112	-0.827	2.419
	8	SA	600	37.978	-15.459	23.508	26.831	30.771	39.420	44.654	48.924	53.082	56.429	71.888	9.446	-0.596	1.227
	9	SA	600	37.832	4.841	23.934	27.745	30.793	39.422	44.325	48.186	52.824	57.472	52.632	8.926	-0.329	0.040
	10	SA	600	41.488	-37.818	24.933	27.896	33.000	42.731	50.243	54.470	56.748	61.110	98.928	11.011	-0.957	3.738
	11	SA	600	41.106	-58.731	24.370	27.672	33.118	42.181	50.022	54.080	56.355	64.054	122.785	11.203	-1.488	9.736
	12	SA	600	39.868	-15.307	25.419	27.733	31.453	41.497	47.740	51.626	54.279	57.973	73.281	10.042	-0.666	1.177
	13	SA	600	39.245	13.472	27.193	29.018	32.203	40.096	46.160	48.630	51.212	55.104	41.632	8.219	-0.293	-0.749
	14	SA	600	39.646	11.350	26.496	28.275	31.498	41.074	46.726	50.166	53.096	57.097	45.746	8.884	-0.264	-0.851
	15	SA	600	37.755	-99.458	24.930	27.973	30.813	39.169	44.305	48.516	52.452	55.692	155.150	10.338	-4.043	50.763
	16	SA	600	37.983	12.065	25.020	28.309	31.593	39.155	43.870	47.295	50.218	53.761	41.696	7.922	-0.304	-0.444
	17	SA	600	37.026	11.608	24.782	26.749	30.360	38.667	43.155	46.814	49.611	54.533	42.925	8.009	-0.312	-0.515
	18	SA	600	36.355	7.233	24.401	28.415	31.560	36.592	41.484	44.856	48.067	54.009	46.776	7.228	-0.437	0.630
	19	SA	600	34.681	11.045	21.440	27.799	30.508	34.790	38.935	43.271	46.791	49.938	38.893	6.829	-0.390	0.641
	20	SA	600	40.285	3.488	27.372	29.562	35.198	41.153	46.665	50.273	52.034	57.394	53.906	8.315	-0.758	1.049
	21	SA	600	15.486	-491.632	-22.836	14.626	27.153	31.127	42.249	45.993	46.920	50.208	541.840	92.451	-5.048	24.163
DS53	1	SA	600	52.970	2.495	19.862	25.817	43.339	54.694	67.757	75.316	77.242	79.693	77.198	17.673	-0.596	-0.192
	2	SA	600	52.918	2.983	20.063	25.612	42.741	55.090	67.143	75.113	76.829	79.855	76.872	17.647	-0.579	-0.284
	3	SA	600	55.161	8.257	20.996	25.155	47.007	59.657	69.093	72.361	74.758	80.302	72.045	16.965	-0.825	-0.334
	4	SA	600	36.118	-3284.809	9.079	21.087	39.838	54.145	65.767	72.376	74.840	77.432	3362.241	165.835	-15.780	287.899
	5	SA	600	33.145	-5238.018	3.976	20.575	39.364	54.671	65.927	72.648	74.637	77.562	5315.580	230.881	-20.356	456.773
	6	SA	600	8.813	-3999.033	0.350	20.175	39.744	52.285	61.877	66.067	68.718	72.652	4071.685	333.673	-10.334	111.343
	7	SA	600	48.583	-21.653	20.032	25.287	41.315	49.910	61.272	65.859	67.730	75.800	97.452	15.712	-1.014	1.553
	8	SA	600	-7.014	-6442.474	-3.116	19.007	39.319	49.374	60.423	64.818	68.836	74.307	6516.781	495.527	-11.351	133.734
	9	SA	600	37.811	-59.110	16.143	20.009	28.614	38.209	48.287	57.119	61.311	67.309	126.419	15.083	-0.931	3.930

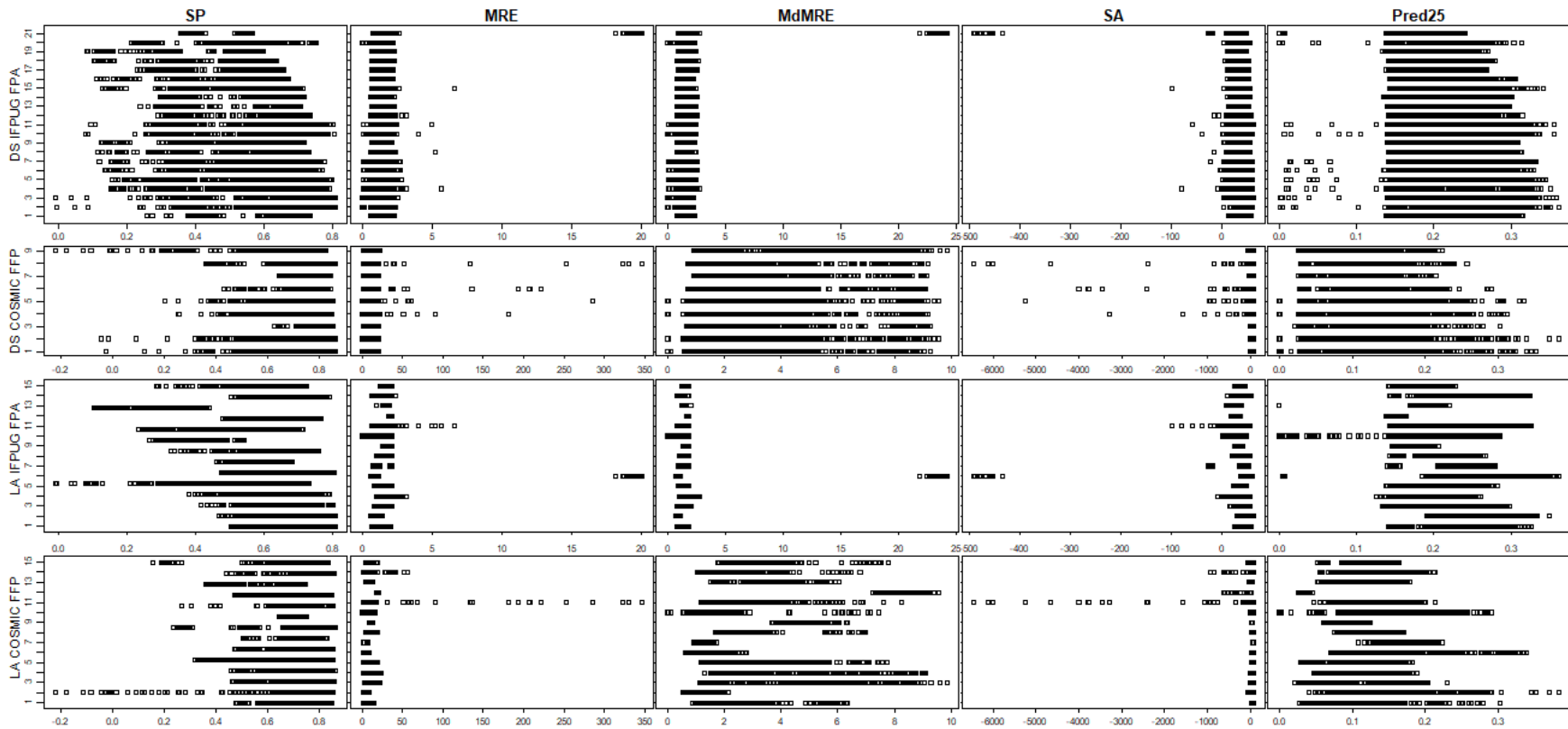


Figura 7-6 Métricas de exactitud por conjuntos de datos.

7.3.2.2. REPETIBILIDAD DEL PROCEDIMIENTO DE EVALUACIÓN DE MODELOS DE ESTIMACIÓN DE ESFUERZO

RQ1. ¿El procedimiento de evaluación de modelos con la configuración seleccionada permite obtener resultados de exactitud repetibles a través de múltiples corridas?

El objetivo del análisis es determinar la estabilidad de los resultados obtenidos por el procedimiento para la construcción de modelos de estimación de esfuerzo basados en el tamaño funcional. Para evaluar la repetibilidad de los modelos de estimación de esfuerzo se analizan las métricas de exactitud: la magnitud del error relativo (MRE), el coeficiente de correlación de Spearman entre las predicciones y los valores reales (SP), la exactitud estandarizada (SA) y el número de predicciones de los valores reales (Pred₂₅). Los análisis se realizan para el conjunto de datos del método de medición COSMIC FFP y el conjunto del método de medición IFPUG FPA. En este caso, se realizan dos ejecuciones aplicando la estrategia de evaluación para cada uno de los 30 grupos detallados en la Tabla 7-26 y se comparan los resultados de las métricas. Para cada sub conjunto de datos y sus respectivos grupos se construyen un total de 600 modelos a partir de la combinación de las técnicas de DP (n=8), las técnicas de AS (n=5) y los LA (n=15). Una vez que se recolectan las métricas de exactitud de las dos corridas por combinación, se aplica la prueba *Wilcoxon signed Rank* para determinar la consistencia entre ellas. Los resultados de las métricas seleccionadas no pertenecen a una distribución normal, por lo que se aplica esta prueba no paramétrica que permite realizar una prueba pareada entre los resultados de las dos corridas.

La Tabla 7-30 presenta los resultados de la aplicación de las pruebas estadísticas que comparan las métricas de exactitud para las dos corridas realizadas para el conjunto de datos del método de medición IFPUG FPA y del método COSMIC FFP. La hipótesis que se desea probar es si existe una diferencia entre los resultados de exactitud obtenidos en cada una de las corridas realizadas para la construcción de los modelos de estimación. La prueba se plantea de la siguiente manera:

H₀: no existe diferencia entre los resultados de la exactitud obtenidos en la

Corrida_i y la Corrida_j para el Modelo_k

Donde el *Modelo_k* es un modelo construido a partir de la combinación de las técnicas de $DP \times AS \times LA$. La comparación se realiza a partir de los resultados de exactitud del *Modelo_k* obtenidos de la *Corrida_i* y *Corrida_j*. La exactitud se mide a partir de los resultados de las métricas MRE, SP, Pred₂₅ y SA. Los modelos se construyen por cada uno de los grupos de cada conjunto de datos.

Los análisis indican que no se encuentra una diferencia significativa entre los resultados de exactitud entre las dos corridas realizadas con la estrategia de evaluación para los modelos de estimación. Esto implica que la estrategia de evaluación de modelos basada en múltiples validaciones cruzadas y la selección aleatoria no afectan la obtención de resultados consistentes. En la práctica, la consistencia de la estrategia de evaluación permite realizar comparaciones exhaustivas y sin sesgo entre los modelos de estimación de esfuerzo para determinar el modelo que mejor desempeño produce en el contexto de evaluación. Del mismo modo, al permitir aplicar una estrategia sin sesgo de selección de proyectos es posible mejorar la confiabilidad en cuanto al sobre ajuste en los resultados de los modelos. Por tanto, a partir de los resultados obtenidos en la evaluación de los modelos, es posible analizar la exactitud de los modelos simplificados, las técnicas que obtienen los mejores resultados de exactitud y los aspectos que pueden impactar el desempeño. Estos análisis se presentan en las siguientes secciones.

Tabla 7-30 Resultados de la repetibilidad para los modelos.

DS	Método	Mo- delo	SP		MRE		SA		Pred ₂₅		
			Z	p	Z	p	Z	p	Z	P	
DS202	IFPUG	1	-2.132	0.033	-1.247	0.212	-	-	-1.334	0.182	
		FPA	2	-1.953	0.051	-0.301	0.763	-	-	-1.210	0.226
			3	0.000	1.000	-1.826	0.068	-1.000	0.317	-1.604	0.109
			4	-1.096	0.273	-0.879	0.380	-	-	-0.180	0.857
			5	0.000	1.000	-1.826	0.068	-1.000	0.317	0.000	1.000
			6	-0.125	0.901	-0.472	0.637	-	-	-0.193	0.847
			7	-0.539	0.590	-0.305	0.760	-0.079	0.937	-1.133	0.257
			8	-0.194	0.846	-1.275	0.202	-	-	-0.145	0.885
			9	-0.463	0.643	-0.707	0.479	-	-	-0.383	0.701
			10	-0.150	0.881	-0.040	0.968	-	-	-0.866	0.386
			11	-1.000	0.317	-1.826	0.068	-1.000	0.317	0.000	1.000
			12	-0.511	0.609	-0.965	0.335	-	-	-1.362	0.173
			13	-1.826	0.068	-1.826	0.068	-1.000	0.317	-1.604	0.109
			14	-1.000	0.317	0.000	1.000	0.000	1.000	0.000	1.000
			15	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
			16	-1.000	0.317	-1.826	0.068	-1.000	0.317	-1.604	0.109
			17	-1.000	0.317	-1.826	0.068	-1.000	0.317	0.000	1.000
			18	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
			19	-1.000	0.317	-1.826	0.068	-1.000	0.317	-1.604	0.109
			20	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
			21	-0.472	0.637	-0.325	0.745	-0.740	0.459	-0.587	0.557
DS53	COSMIC	1	-1.000	0.317	-1.000	0.317	-1.000	0.317	-1.000	0.317	
		FFP	2	-1.000	0.317	-1.000	0.317	-1.342	0.180	-1.000	0.317
			3	-1.342	0.180	-1.342	0.180	-1.000	0.317	-1.000	0.317
			4	-1.000	0.317	0.000	1.000	-1.000	0.317	-1.000	0.317
			5	-1.000	0.317	-1.826	0.068	-1.000	0.317	-1.000	0.317
			6	-1.000	0.317	-1.604	0.109	-1.342	0.180	-1.000	0.317
			7	-1.342	0.180	-1.342	0.180	-1.000	0.317	0.000	1.000
			8	-1.342	0.180	-1.826	0.068	-1.604	0.109	-1.000	0.317
			9	-1.342	0.180	-1.604	0.109	-1.000	0.317	-1.000	0.317

7.3.2.3. COMPARACIÓN DE LA EXACTITUD DE LOS MODELOS SIMPLIFICADOS DE ESTIMACIÓN DE ESFUERZO

RQ2. ¿Cuál es la exactitud de los modelos de estimación simplificados construidos por el procedimiento de evaluación?

El objetivo del análisis es determinar la exactitud de los modelos simplificados de estimación construidos a partir de los componentes funcionales básicos (BFC) en comparación con los basados en el total del tamaño funcional. Este estudio investiga si existen modelos de estimación de esfuerzo basados en subconjuntos de los BFC del tamaño funcional que puedan presentar un desempeño similar a los modelos basados en el total del tamaño funcional que podrían simplificar el proceso de medición del tamaño funcional. En este estudio se comparan los dos conjuntos de datos (DS202 y DS53) para cada uno de los grupos detallados en la Tabla 7-26. En total se comparan 21 grupos para el subconjunto del método de medición IFPUG FPA (DS202) y 9 grupos para el subconjunto del método de medición COSMIC FFP (DS53). En el caso del método IFPUG FPA se comparan los pares de grupos basados en el total de puntos de función y en subconjuntos de los componentes funcionales básicos (1, 8), (1, 9), (1, 12), (1, 13), (1, 14), (1, 15), (1, 16), (1, 17), (1, 18), (1, 19), basados en el total de puntos de función y en subconjuntos de los componentes funcionales básicos incorporando las variables de contexto y el tamaño relativo (2, 4), (2, 6), (2, 10), basados en el total de puntos de función y en subconjuntos de los componentes funcionales básicos incorporando las variables de contexto (3, 5), (3, 7), (3, 11), basados en el total de puntos de función y en subconjuntos de los componentes funcionales con los que solo utilizan el tamaño relativo incorporando o no las variables de contexto (1, 21), (3, 20) y finalmente, los basados en el total de puntos de función y en subconjuntos de los componentes funcionales incorporando o no las variables de contexto y el tamaño relativo (1, 2), (9, 6).

Para el método COSMIC FFP se comparan los grupos basados en el total de puntos de función y en subconjuntos de los componentes funcionales básicos incorporando las variables de contexto y el tamaño relativo (1, 4), basados en el total de puntos de función y en subconjuntos de los componentes funcionales básicos incorporando las variables de contexto (2, 5), basados en el total de puntos de función y en subconjuntos de los componentes funcionales básicos (3, 6), (3, 7), (3, 8) y (3, 9) y finalmente, los basados en el total de puntos de función y en subconjuntos de los componentes funcionales incorporando o no las variables de contexto y el tamaño relativo (1, 3), (6, 4).

Basado en los resultados de las métricas de exactitud se realizan las comparaciones entre los residuos de los valores estimados de los modelos y los valores reales de esfuerzo para determinar el nivel de exactitud de las estimaciones. Se procesan los datos de la ejecución de la estrategia de evaluación para cada uno de los 30 grupos. Para cada grupo de datos se construyen un total de 600 modelos a partir de la combinación de las técnicas $DP \times AS \times LA$. En este análisis, se aplica la prueba no paramétrica *Wilcoxon signed Rank* para la comparación de los modelos dado que los resultados de las métricas no pertenecen a una distribución normal. La exactitud de los modelos de estimación se analiza a partir de las métricas de exactitud: magnitud del error relativo (MRE), coeficiente de correlación de Spearman (SP), exactitud estandarizada (SA) y número de predicciones de los valores reales ($Pred_{25}$).

La hipótesis que se desea probar es si existe diferencia entre la exactitud de los modelos construidos a partir de los FP y los modelos construidos a partir de los BFC, de la siguiente manera:

H₀: no existe diferencia entre la exactitud del Modelo_i y el Modelo_j

Donde el *Modelo_i* es un modelo construido a partir del total de los puntos de función (FP) y el *Modelo_j* es un modelo construido a partir de los componentes funcionales básicos (BFC). La exactitud se mide a partir de los resultados de las métricas MRE, SP, $Pred_{25}$, SA.

Primero se presentan los análisis para el conjunto de datos basado en el método de medición IFPUG FPA (Sección 7.3.2.3.1) y posteriormente los análisis para el sub conjunto de datos basado en el método de medición COSMIC FFP (Sección 7.3.2.3.2).

7.3.2.3.1. MODELOS SIMPLIFICADOS BASADOS EN EL MÉTODO IFPUG FPA

En la primera prueba se consideran los 600 modelos construidos que comparan los pares de grupos de los conjuntos de datos para cada una de la combinación de técnicas

utilizadas $DP \times AS \times LA$. Al realizar la prueba estadística pareada *Wilcoxon signed Rank* para todos los modelos, estos presentan diferencias significativas para cada una de las comparaciones. Esto es, los resultados de la prueba por cada uno de los modelos construidos a partir de la combinación de las técnicas presentan distinto desempeño entre los grupos basados en el total de puntos de función y los modelos que se basan en subconjuntos de los componentes funcionales básicos. En este caso, no es posible concluir que existen técnicas que presenten el mismo desempeño para los distintos grupos de los conjuntos de datos comparados.

Tal como se obtiene en los resultados de los estudios descritos en la Sección 7.1 y la Sección 7.2, la variabilidad de los resultados de las métricas de exactitud que se presentan dependiendo de la combinación de las diferentes técnicas de construcción de modelos y de los conjuntos de datos evaluados puede explicar que no se encuentren resultados similares. Esto se da, por ejemplo, cuando un modelo puede presentar un buen desempeño para ciertos grupos de un conjunto de datos específico; sin embargo, presenta un desempeño distinto para otros grupos de un conjunto de datos diferente. Asimismo, es posible encontrar diferentes combinaciones de técnicas de construcción de modelos ($DP \times AS \times LA$) que presentan distintos resultados para un mismo conjunto de datos y una combinación de técnicas de construcción de modelos que puede presentar diferencias en el desempeño dependiendo de los conjuntos de datos evaluados.

Dado lo anterior, es de interés determinar si existe alguna combinación de técnicas de construcción de modelos que permitan obtener resultados similares entre los modelos de los grupos para cada uno de los conjuntos de datos basados en el total de puntos de función y los modelos que se basan en subconjuntos de los componentes funcionales básicos. En la Sección 7.1, se presentaron evidencias que indican que si la construcción de los modelos se basa en técnicas de regresión es posible obtener resultados similares en la exactitud. Esto es, las técnicas de regresión parecen mantener consistencia a través de distintos conjuntos de datos que contienen métricas del tamaño funcional. Por tanto, se realiza la comparación de los modelos construidos a partir de las técnicas de pre procesamiento de

datos logarítmica y sin pre procesamiento (DP=1, 2), de selección de atributos *BackwardElimination* y *ForwardSelection* (AS=4, 5) y de regresión *LeastMedSq*, *LinearRegression*, *SMOreg* y *AdditiveRegression* (LA=2, 3, 6, 7) para los 21 subconjuntos de datos. Estas técnicas son seleccionadas porque mostraron estabilidad entre los grupos de los distintos conjuntos de datos en las evaluaciones previas.

Al realizar la prueba pareada para los modelos seleccionados entre los grupos de los conjuntos de datos para el método IFPUG FPA es posible encontrar modelos simplificados que presentan un desempeño similar en la exactitud de las estimaciones a los modelos basados en el total de puntos de función. La Tabla 7-31 presenta los resultados de las pruebas de hipótesis con la comparación de modelos construidos a partir de los UFP y los conjuntos de los BFC. En la tabla únicamente se presentan los resultados de las comparaciones de los modelos de los grupos de los conjuntos de datos que no presentaron diferencias en la exactitud de las estimaciones. Los resultados indican que existen modelos simplificados que no encuentran una diferencia significativa en comparación con los modelos basados en los puntos de función.

Por ejemplo, no se encuentran diferencias significativas en la comparación de los resultados de exactitud (SA) de los modelos basados en los UFP y los BFC, TF y DF (1, 8), los UFP y los TF y DF (1, 12), los UFP y los TF e ILF (1, 14), los modelos basados en los UFP, tamaño relativo, variables de contexto y los TF, DF, tamaño relativo, variables de contexto (2, 10), al igual que los basados en los UFP, variables de contexto y los basados en el tamaño relativo, variables de contexto (3, 20). En el caso del modelo basado en los UFP y el basado en EI, EO, EQ, ILF (1, 15) presenta resultados similares para la MRE y en el caso del modelo basado en los UFP, variables de contexto y el basado en los BFC y contexto (3, 7) presenta resultados similares para la MRE y el Pred₂₅.

En el caso de las pruebas para los modelos basados en los UFP o los BFC que utilizan o no las variables de contexto, los resultados utilizando las 600 combinaciones indicaron que existe diferencia significativa para la comparación de los modelos (1, 2) y (9, 6) donde, los modelos que incorporan las variables de contexto obtienen mejores resultados de exactitud basado en las métricas de la SA, MRE y Pred₂₅. Al realizar las pruebas con los modelos

construidos a partir de las técnicas de pre procesamiento de datos logarítmica y sin pre procesamiento DP (1, 2), de selección de atributos *BackwardElimination* y *ForwardSelection* AS (4, 5) y de regresión *LeastMedSq*, *LinearRegression*, *SMOreg* y *AdditiveRegression* LA (2, 3, 6, 7) para los 21 grupos del conjunto de datos, los resultados indican que existe diferencia significativa para la comparación de los modelos (1, 2) donde para la SA el resultado fue $Z = -2.275, p < 0.023$) y (9, 6) donde para la SA el resultado fue $Z = -2.534, p < 0.011$). Esto comprueba la evidencia que indica que los modelos que incorporan las variables de contexto obtienen resultados de exactitud significativamente mejores para las métricas SA y $Pred_{25}$.

Los resultados muestran que pueden existir modelos de estimación de esfuerzo basados en conjuntos de los BFC que pueden tener un nivel de exactitud similar a los modelos construidos a partir de los UFP totales. Aún más, en algunos casos podrían construirse modelos basados en el tamaño relativo de la aplicación, más las variables de contexto, que puedan ofrecer resultados similares a los basados en los puntos de función sin ajustar. Por otro lado, aunque es posible encontrar resultados contradictorios dependiendo de los grupos seleccionados, existen técnicas como las de regresión, que pueden presentar un comportamiento estable entre modelos. Sin embargo, para realizar la evaluación de exactitud de los modelos de estimación es requerido aplicar las técnicas con los conjuntos de datos específicos de una organización para poder realizar la selección. Esto se da porque pueden existir combinaciones de técnicas de construcción de modelos que pueden funcionar con valores aceptables para conjuntos de proyectos con características específicas pero con resultados no aceptables para otros conjuntos de proyectos con otras características. Es importante recalcar que, para un contexto de industria, algunos de los resultados entre modelos que obtuvieron diferencia significativa por las pruebas estadísticas podrían seleccionarse, ya que en la práctica no representan variaciones importantes en la exactitud comparadas con los beneficios en el tiempo y el costo requerido para realizar las mediciones.

Tabla 7-31 Resultados de la comparación de los modelos para el método IFPUG FPA.

DS	Mode- los	SP		MRE		SA		Pred ₂₅	
		Z	p	Z	p	Z	p	Z	P
DS202	(1)(8)	-3.361	0.001	-2.068	0.039	-1.603	0.109	-2.637	0.008
	(1)(12)	-1.603	0.109	-1.138	0.255	-1.913	0.056	-0.465	0.642
	(1)(14)	-2.947	0.003	-1.551	0.121	-0.155	0.877	-3.051	0.002
	(1)(15)	-3.516	0.000	-1.862	0.063	-2.120	0.034	-0.879	0.379
	(2)(10)	-2.379	0.017	-2.223	0.026	-1.086	0.278	-2.534	0.011
	(3)(7)	-3.154	0.002	-1.396	0.163	-2.120	0.034	-2.585	0.010
	(3)(20)	-3.103	0.002	-1.706	0.088	-1.758	0.079	-2.689	0.007

Finalmente, dado que los resultados de las técnicas de construcción de modelos de estimación podrían depender de los grupos de variables y los conjuntos de datos utilizados, es de interés evaluar si los mejores resultados de exactitud obtenidos en la evaluación para los modelos simplificados pueden desempeñarse de manera similar a los obtenidos por los mejores modelos basados en los puntos de función totales. Por tanto, en este análisis se consideran solo los modelos que proveen los mejores resultados de exactitud para la métrica SA. Para cada uno de los 21 grupos del conjunto de datos se construyeron 600 modelos de los cuales son seleccionados los mejores 60, que representan el 90% percentil. En este caso, se aplica la prueba *Mann-Whitney U* que permite la comparación de los resultados independientes de los mejores modelos. Se utiliza esta prueba no paramétrica dado que los resultados de las métricas no pertenecen a una distribución normal.

Al realizar la prueba estadística para los mejores modelos entre los grupos del conjunto de datos del método IFPUG FPA se encuentran modelos simplificados que presentan un desempeño similar en la exactitud de las estimaciones a los modelos basados en el total de los puntos de función. La Tabla 7-32 presenta los resultados de las pruebas de hipótesis con la comparación de modelos construidos a partir de los UFP y los conjuntos de los BFC. En esta tabla, solo se presentan las comparaciones de los modelos de los grupos del conjunto de datos que no presentaron diferencias en la exactitud de las estimaciones y los que muestran resultados distintos a los de los análisis anteriores. Asimismo, la Figura 7-7 muestra un diagrama de cajas con los resultados de las métricas de exactitud para los 21 grupos del conjunto de datos utilizados para el método IFPUG FPA. Los resultados indican que

existen modelos simplificados que no encuentran una diferencia significativa en los resultados de exactitud en comparación con los modelos basados en los totales de los puntos de función.

Tabla 7-32 Comparación de los mejores modelos para el método IFPUG FPA.

DS	Mode- los	SP		MRE		SA		Pred ₂₅	
		Z	p	Z	p	Z	p	Z	P
DS202	(1)(8)	-6.067	0.000	-1.039	0.299	-1.811	0.070	-1.624	0.104
	(1)(9)	-9.138	0.000	-1.764	0.078	-2.288	0.022	-1.522	0.128
	(1)(12)	-0.661	0.508	-0.745	0.456	-2.383	0.017	-1.842	0.065
	(1)(14)	-8.576	0.000	-1.317	0.188	-2.089	0.037	-2.819	0.005
	(1)(15)	-9.406	0.000	-2.189	0.029	-3.228	0.001	-0.635	0.525
	(2)(6)	-8.104	0.000	-1.947	0.052	-7.658	0.000	-5.511	0.000
	(2)(4)	-2.861	0.00	-1.333	0.182	-3.238	0.001	-1.302	0.193
				4					
	(2)(10)	-1.806	0.071	-1.821	0.069	-0.787	0.431	-0.976	0.329
	(3)(5)	-2.855	0.00	-0.546	0.585	-3.286	0.001	-0.541	0.589
				4					
	(3)(7)	-7.269	0.000	-1.013	0.311	-6.729	0.000	-3.049	0.002
	(3)(11)	-7.269	0.000	-1.013	0.311	-6.729	0.000	-3.049	0.002
	(3)(20)	-9.148	0.000	-5.868	0.000	-9.143	0.000	-8.540	0.000
	(1)(2)	-8.865	0.000	-0.509	0.611	-8.051	0.000	-5.915	0.000
	(9)(6)	-4.467	0.000	-1.417	0.156	-3.034	0.002	-0.635	0.525

Similar al análisis anterior, no se encuentran diferencias significativas en la comparación de los resultados de exactitud para la métrica SA de los modelos basados en los UFP y los BFC, TF y DF (1, 8) y los TF e ILF (1, 14), los modelos basados en los UFP, tamaño relativo, variables de contexto y los TF, DF, tamaño relativo, variables de contexto (2, 10). En el caso de los modelos basados en los UFP y los TF y DF (1, 12) no presentan diferencia para las métricas MRE y Pred₂₅. Para los modelos basados en los UFP y el basado en EI, EO, EQ, ILF (1, 15) presenta resultados similares para la Pred₂₅ y en el caso del modelo basado en los UFP, variables de contexto y el basado en los BFC y contexto (3, 7) presenta resultados similares para la MRE. Los modelos basados en UFP, variables de contexto y los basados en los BFC, DF, TF, contexto (3, 5) y los basados en los BFC y los construidos a partir de los BFC, tamaño relativo y contexto (9, 6) no presentan diferencias significativas para las métricas MRE y Pred₂₅. En este caso, si una organización no cuenta con valores de contexto para sus proyectos podría explorar la posibilidad de utilizar los BFC en lugar de los UFP. En esta evaluación, los modelos basados en los UFP, variables de contexto y los basados en

el tamaño relativo, variables de contexto (3, 20) sí presentan diferencias significativas en favor de los que utilizan los UFP totales.

Basado en la comparación de los mejores modelos, los resultados obtenidos ofrecen evidencia que confirma que pueden existir modelos de estimación de esfuerzo basados en los BFC que pueden tener un nivel de exactitud similar a los modelos construidos a partir de los UFP. Aún más, en algunos casos podría decidirse utilizar o no las variables de contexto dependiendo de los datos de proyectos con los que cuente la organización. La evidencia indica que la construcción de los modelos de estimación se debe realizar para las características específicas de los conjuntos de datos para los cuales van a ser utilizados.

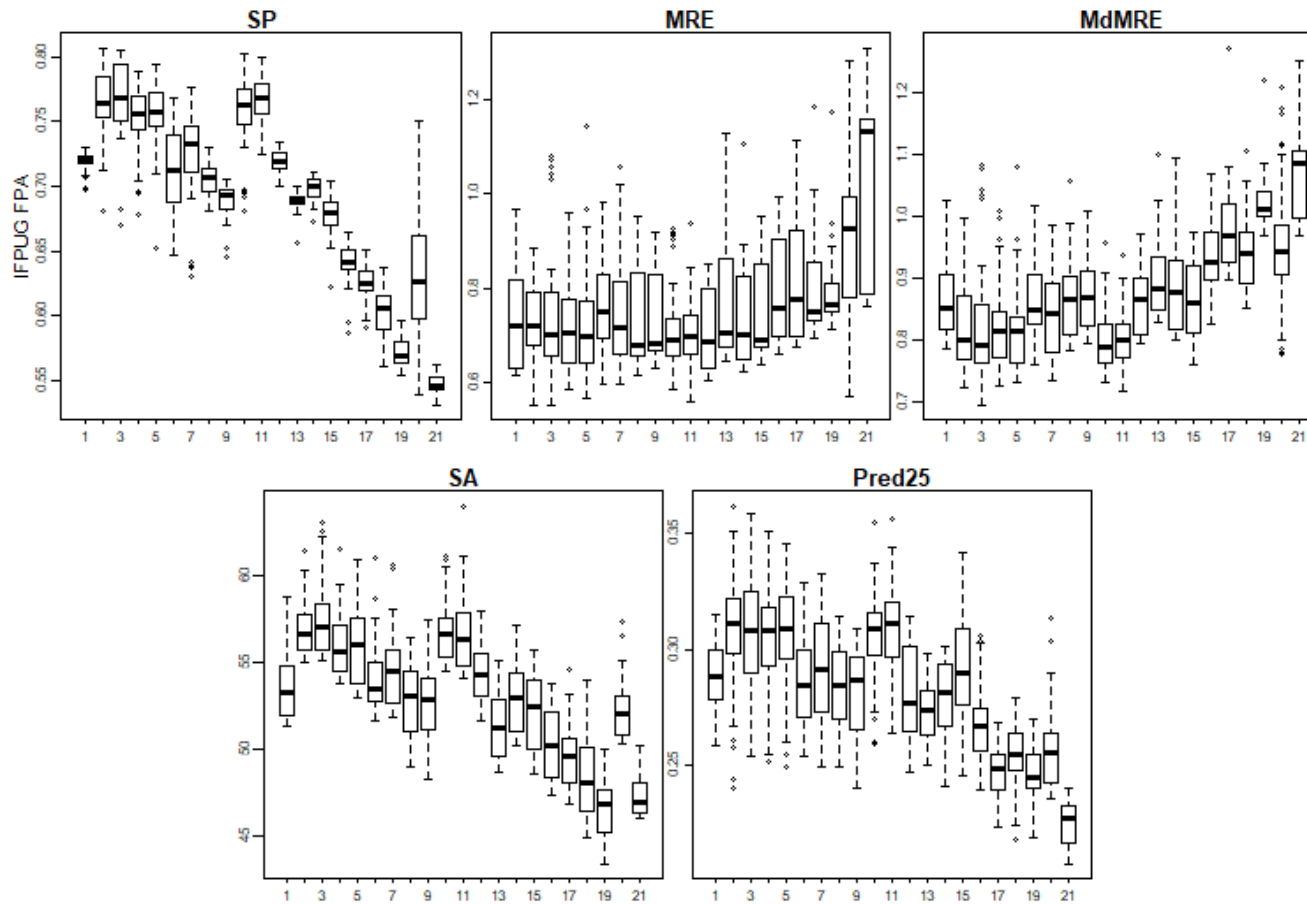


Figura 7-7 Métricas de exactitud para los mejores modelos del método IFPUG FPA por conjunto de datos.

7.3.2.3.2. MODELOS SIMPLIFICADOS BASADOS EN EL MÉTODO COSMIC FFP

Para realizar la evaluación de los modelos de los grupos del conjunto de datos del método COSMIC FFP se consideran 600 combinaciones de técnicas $DP \times AS \times LA$ y se comparan los pares de los grupos. Se realiza la prueba estadística pareada *Wilcoxon signed Rank* para todos los modelos comparando los modelos construidos a partir de los CFP y los BFC. A diferencia de los modelos basados en los grupos del conjunto de datos del método IFPUG FPA, para el método COSMIC FFP sí se encuentran modelos con un desempeño similar en la exactitud de las estimaciones para los distintos grupos comparados. En este caso, los modelos basados en el total de puntos de función y en conjuntos de los componentes funcionales básicos (BFC) incorporando las variables de contexto y el tamaño relativo (1, 4) no presentan diferencias significativas para las métricas MRE donde $Z = -1.163, p < 0.245$ y $Pred_{25}$ donde $Z = -1.169, p < 0.242$. De igual forma, los modelos basados en el total de puntos de función y en conjuntos de los BFC incorporando las variables de contexto (2, 5) no presentan diferencias significativas para la métrica $Pred_{25}$ donde $Z = -0.41, p < 0.968$. Finalmente, los modelos basados en el total de puntos de función y en conjuntos de los BFC incorporando o no las variables de contexto y el tamaño relativo (1, 3) (6, 4) no presentan diferencias significativas para la métrica $Pred_{25}$ donde $Z = -1.119, p < 0.263$ y para la métrica MRE donde $Z = -0.610, p < 0.542$ respectivamente. Por otro lado, la comparación de los modelos basados en el total de puntos de función y en conjuntos de los componentes funcionales básicos (3, 6), (3, 7), (3, 8) y (3, 9) presentan diferencias significativas en los resultados de las métricas de exactitud.

Similar al comportamiento de los modelos del conjunto de datos IFPUG FPA, los resultados de las métricas de exactitud para los grupos del conjunto de datos del método COSMIC FFP pueden presentar variabilidad dependiendo de la combinación de las diferentes técnicas de construcción de modelos y de los conjuntos de datos evaluados. Dados los resultados obtenidos en la Sección 7.2, se analizan los modelos construidos a partir de la combinación de técnicas de regresión *LeastMedSq*, *LinearRegression*, *SMOreg* y *AdditiveRegression* (LA=2, 3, 6, 7), de selección de atributos *BackwardElimination* y *ForwardSelection*

(AS=4, 5) y de pre procesamiento de datos logarítmica y sin pre procesamiento (DP=1, 2). Al realizar la prueba pareada para los modelos seleccionados entre los grupos del conjunto de datos para el método COSMIC FFP es posible encontrar modelos simplificados que presentan un desempeño similar en la exactitud de las estimaciones con respecto a los modelos construidos basados en el total de los puntos de función.

La Tabla 7-33 presenta los resultados de las pruebas de hipótesis con la comparación de modelos construidos a partir de los FP y los conjuntos de los BFC. Únicamente se presentan las comparaciones de los modelos de los grupos que no presentaron diferencias en la exactitud de las estimaciones. Los resultados indican que existen modelos simplificados que no encuentran una diferencia significativa en comparación con los modelos basados en los puntos de función.

No se encuentran diferencias significativas entre los resultados de exactitud de los modelos basados en el total de puntos de función y en conjuntos de los componentes funcionales básicos (BFC) incorporando las variables de contexto y el tamaño relativo (1, 4), los basados en el total de puntos de función y en conjuntos de los BFC incorporando las variables de contexto (2, 5) y los basados en el total de puntos de función y en los conjuntos de los componentes funcionales básicos X y W (3, 8). Los anteriores para las métricas MRE, SA y Pred₂₅. Del mismo modo, para los modelos basados en el total de puntos de función y en los subconjuntos de los componentes funcionales básicos E, X, R y W (3, 6) para las métricas SA y Pred₂₅, en los E y W para la métrica MRE (3, 7) y en el W para la métrica Pred₂₅ (3, 9) no se encuentran diferencias significativas. En el caso de las pruebas para los modelos basados en los CFP que utilizan o no las variables de contexto, los resultados indican que no existe diferencia significativa para la comparación de los modelos (1, 3) para las métricas MRE, SA y Pred₂₅. En el caso de los modelos basados en los BFC (6, 4) no existe diferencia significativa para las métricas MRE y SA pero sí para el Pred₂₅.

Los resultados indican que pueden existir modelos de estimación de esfuerzo basados en conjuntos de los BFC con un nivel de exactitud similar a los modelos construidos a partir de los FP y que, para el conjunto de datos bajo análisis, la incorporación de las características de contexto de los proyectos específicos no representa una mejora en los modelos

de estimación. En la práctica, la utilización de los modelos simplificados podría ayudar a reducir el tiempo necesario para la recolección de las métricas de tamaño funcional y para la construcción de modelos de estimación, en este tipo de proyectos, parece no ser requerido recolectar las variables de contexto.

Tabla 7-33 Resultados de la comparación de los modelos para el método COSMIC FFP.

DS	Mo- delos	SP		MRE		SA		Pred ₂₅	
		Z	p	Z	p	Z	p	Z	P
DS53	(1)(4)	-1.086	0.278	-0.465	0.642	-0.465	0.642	-0.310	0.756
	(2)(5)	-0.621	0.535	-0.155	0.877	-0.103	0.918	-0.827	0.408
	(3)(6)	-3.516	0.000	-1.965	0.049	-1.551	0.121	-0.827	0.408
	(3)(7)	-3.464	0.001	-1.758	0.079	-2.689	0.007	-2.637	0.008
	(3)(8)	-3.258	0.001	-0.362	0.717	-0.982	0.326	-1.189	0.234
	(3)(9)	-3.464	0.001	-2.223	0.026	-3.516	0.000	-0.879	0.379
	(1)(3)	-3.309	0.001	-0.155	0.877	-0.465	0.642	-1.551	0.121
	(6)(4)	-3.361	0.001	-1.551	0.121	-1.189	0.234	-2.741	0.006

Finalmente, dado que los resultados de las técnicas de construcción de modelos de estimación podrían depender de los grupos de variables y los conjuntos de datos utilizados, es de interés evaluar si los mejores resultados de exactitud obtenidos en la evaluación para los modelos simplificados pueden desempeñarse de manera similar a los mejores obtenidos por los modelos basados en los puntos de función totales. Por tanto, para este análisis se consideran solo los modelos que proveen los mejores resultados de exactitud para la métrica SA. Para cada uno de los 9 grupos del conjunto de datos se construyeron 600 modelos de los cuales son seleccionados los mejores 60, que representan el 90% percentil. En este caso, se aplica la prueba no paramétrica *Mann-Whitney U* que permite la comparación de los resultados independientes de los mejores modelos. Se utiliza esta prueba dado que los resultados de las métricas no pertenecen a una distribución normal.

La Tabla 7-34 presenta los resultados de las pruebas de hipótesis con la comparación de modelos construidos a partir de los FP y los conjuntos de los BFC. Solo se presentan las comparaciones de los modelos de los grupos que no presentaron diferencias en la exactitud de las estimaciones. Asimismo, la Figura 7-8 muestra los resultados de las métricas de exactitud para los 9 grupos del conjunto de datos utilizados para el método COSMIC FFP. Los

resultados indican que existen modelos simplificados que no presentan una diferencia significativa en los resultados de exactitud en comparación con los modelos basados en los totales de los puntos de función. No se encuentran diferencias significativas en la comparación de los resultados de exactitud de los modelos basados en el total de puntos de función (FP) y en conjuntos de los componentes funcionales básicos (BFC) incorporando las variables de contexto y el tamaño relativo (1, 4), los basados en los FP y en conjuntos de los BFC incorporando las variables de contexto (2, 5) y los basados en el total de FP y el componente funcional básico W (3, 9). Esto para las métricas MRE y Pred₂₅. De igual manera, para los modelos basados en el total de CFP y en los conjuntos de los componentes funcionales básicos E, X, R y W (3, 6) y los basados en los X y W (3, 8) no se entran diferencias para la métrica MRE. Finalmente, en el caso de las pruebas para los modelos basados en los CFP y BFC que utilizan o no las variables de contexto, los resultados indican que no existe diferencia significativa para la comparación de los modelos (1, 3) (6, 4) para la métrica MRE. Estos resultados confirman que los mejores modelos de estimación de esfuerzo basados en los BFC pueden obtener valores de exactitud con un nivel similar a los mejores modelos construidos a partir de los CFP.

En la práctica, los resultados indican que es posible explorar la utilización de modelos de estimación de esfuerzo basados en los componentes funcionales básicos (BFC) con resultados de exactitud similares a los obtenidos por los modelos construidos a partir de los puntos de función (CFP).

Tabla 7-34 Comparación de los mejores modelos para el método COSMIC FFP.

DS	Mode- los	SP		MRE		SA		Pred ₂₅	
		Z	p	Z	p	Z	p	Z	P
DS202	(1)(4)	-1.459	0.145	-1.359	0.174	-8.135	0.000	-1.798	0.072
	(2)(5)	-1.018	0.309	-1.622	0.105	-7.773	0.000	-1.073	0.283
	(3)(6)	-6.823	0.000	-1.690	0.091	-9.432	0.000	-2.703	0.007
	(3)(7)	-7.799	0.000	-4.272	0.000	-8.151	0.000	-7.705	0.000
	(3)(8)	-1.774	0.076	-0.740	0.459	-8.886	0.000	-5.511	0.000
	(3)(9)	-7.285	0.000	-0.399	0.690	-9.448	0.000	-0.081	0.935
	(1)(3)	-7.925	0.000	-1.454	0.146	-6.451	0.000	-2.472	0.013
	(6)(4)	-5.926	0.000	-1.517	0.129	-9.432	0.000	-3.648	0.000

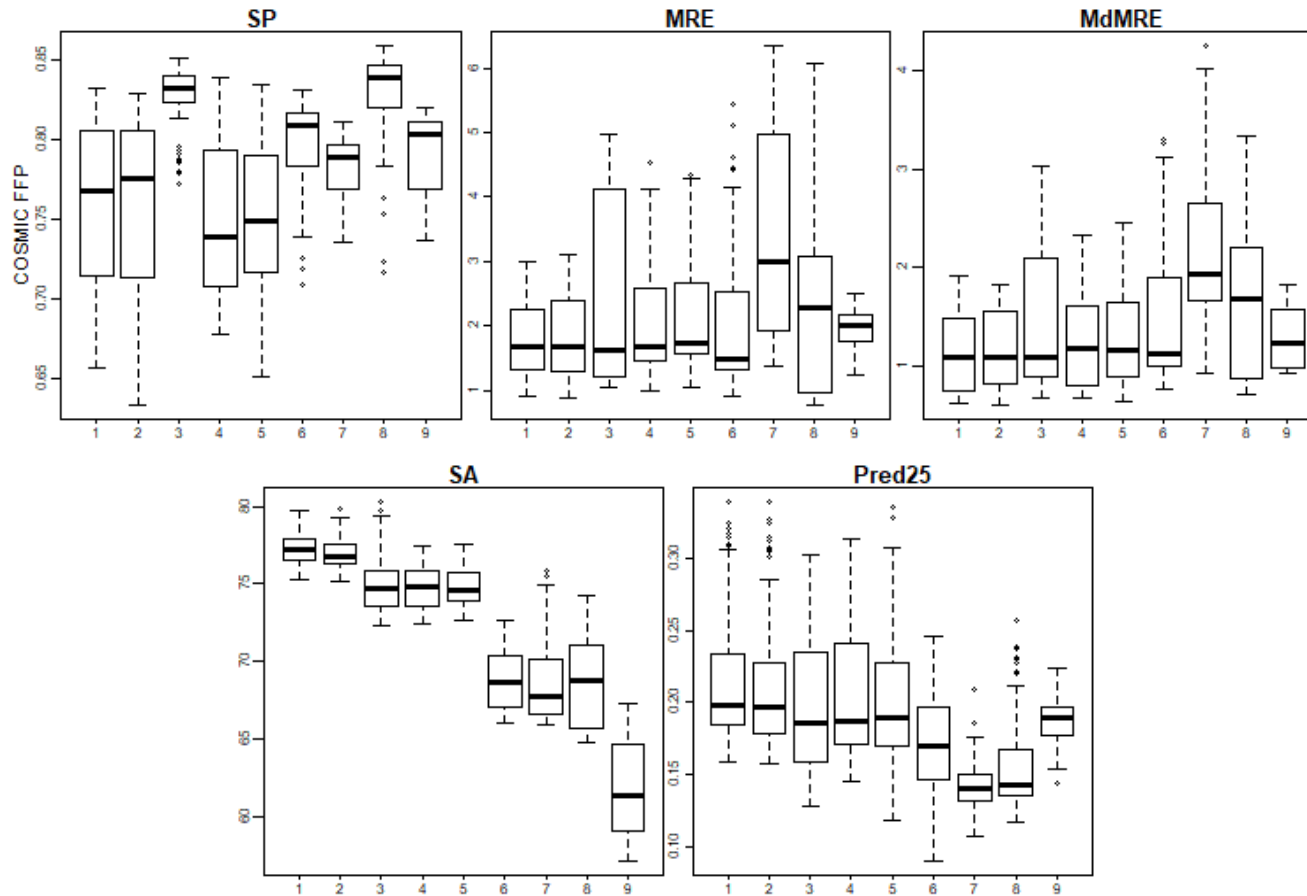


Figura 7-8 Métricas de exactitud para los mejores modelos del método COSMIC FFP por conjunto de datos.

7.3.2.4. EXACTITUD DE LAS TÉCNICAS DE CONSTRUCCIÓN DE LOS MODELOS DE ESTIMACIÓN DE ESFUERZO

RQ3. ¿Cuáles son las técnicas que obtienen los mejores resultados de exactitud para los modelos de estimación de esfuerzo contruidos por el procedimiento de evaluación?

En esta sección se reportan las técnicas de construcción de modelos de estimación de esfuerzo que presentan mejores resultados de exactitud para los grupos de los conjuntos de datos estudiados. Para determinar los mejores modelos se utiliza la métrica de la exactitud estandarizada (SA) y se reportan las de la magnitud del error relativo (MRE), el coeficiente de correlación de Spearman (SP) y el número de predicciones de los valores reales (Pred₂₅). Los resultados de exactitud son realizados para el método de medición IFPUG FPA y COSMIC FFP. Se presentan los resultados para cada uno de los 30 grupos de los dos conjuntos de datos analizados, 21 para los basados en el método de medición IFPUG FPA y 9 para los basados en el método de medición COSMIC FFP. Se seleccionan los modelos con mejores resultados de exactitud de un total de 600 modelos por grupo producto de la ejecución de la estrategia de evaluación con la combinación de las técnicas de $DP \times AS \times LA$.

En la Tabla 7-35 se lista, para cada uno de los grupos de los conjuntos de datos, los modelos que presentan los mejores resultados en la exactitud de la estimación con base en la métrica SA. En el Anexo 8 se detallan los seis modelos que presentan los mejores resultados en la exactitud de la estimación con base en la métrica SA. Estos representan el 1% superior de los modelos para los resultados de exactitud. Asimismo, para cada uno de los modelos se presentan las métricas asociadas al modelo: la MRE, MdMRE y Pred₂₅. En el caso de la exactitud estandarizada (SA), los mejores resultados de los grupos para el método de medición IFPUG FPA oscilan entre el 49.94% y el 64.05% y para el método de medición COSMIC FFP oscilan entre el 67.31% y el 80.30%. Esto significa que existen modelos basados en los proyectos del método IFPUG FPA que podrían superar la línea base de selección aleatoria hasta por un 54.05% y hasta un 80.30% en el caso del método COSMIC FFP.

Considerando los modelos con los mejores resultados (1% superior), la mayor cantidad de técnicas de construcción (LA) utilizadas para el conjunto de datos del método

IFPUG FPA son las de la regresión (2) *Least LeastMedSq* que representa un 54.0% (n=68) del total y (6) *SMOreg* que representa un 41.3% (n=52) del total. En el caso de los conjuntos de datos del método COSMIC FFP son las de (6) *SMOreg* que representa un 29.6% (n=16) del total, (14) *MP5* que representa un 29.6% (n=16) del total y (7) *AdditiveRegression* que representa un 24.1% (n=13) del total. Es importante notar que todas estas técnicas (LA) incorporan estrategias de regresión para su implementación. Los modelos con mejor desempeño presentan valores similares a los reportados por estudios previos (Lokan & Mendes, 2006; Kitchenham, Mendes & Travassos, 2006; Dejaeger, Verbeke, Martens & Baesens, 2012; Lavazza, Morasca & Robiolo, 2013; Minku & Yao, 2013; Dolado, Rodriguez, Harman, Langdon & Sarro, 2016).

Las combinaciones de técnicas de $DP \times AS \times LA$ más utilizadas para los grupos del conjunto del método de medición IFPUG FPA son las de $BC \times BF \times LMS$ (3x2x2) que representa un 10.3% (n=13) del total, $BC \times BE \times SMO$ (3x4x6) que representa un 9.5% (n=12) del total, $BC \times BE \times LMS$ (3x4x2) que representa un 8.7% (n=11) del total y $BC \times LFS \times LMS$ (3x3x2) que representa un 7.9% (n=10) del total. En el caso del método de medición COSMIC FFP fueron las de $BC \times LFS \times SMO$ (3x3x6) que representa un 7.4% (n=4) del total, $BC \times FS \times AR$ (3x5x7) que representa un 7.4% (n=4) del total y $None \times BF \times M5P$ (1x2x14) que representa un 5.6% (n=3) del total.

Tabla 7-35 Mejores modelos de estimación de esfuerzo por grupo.

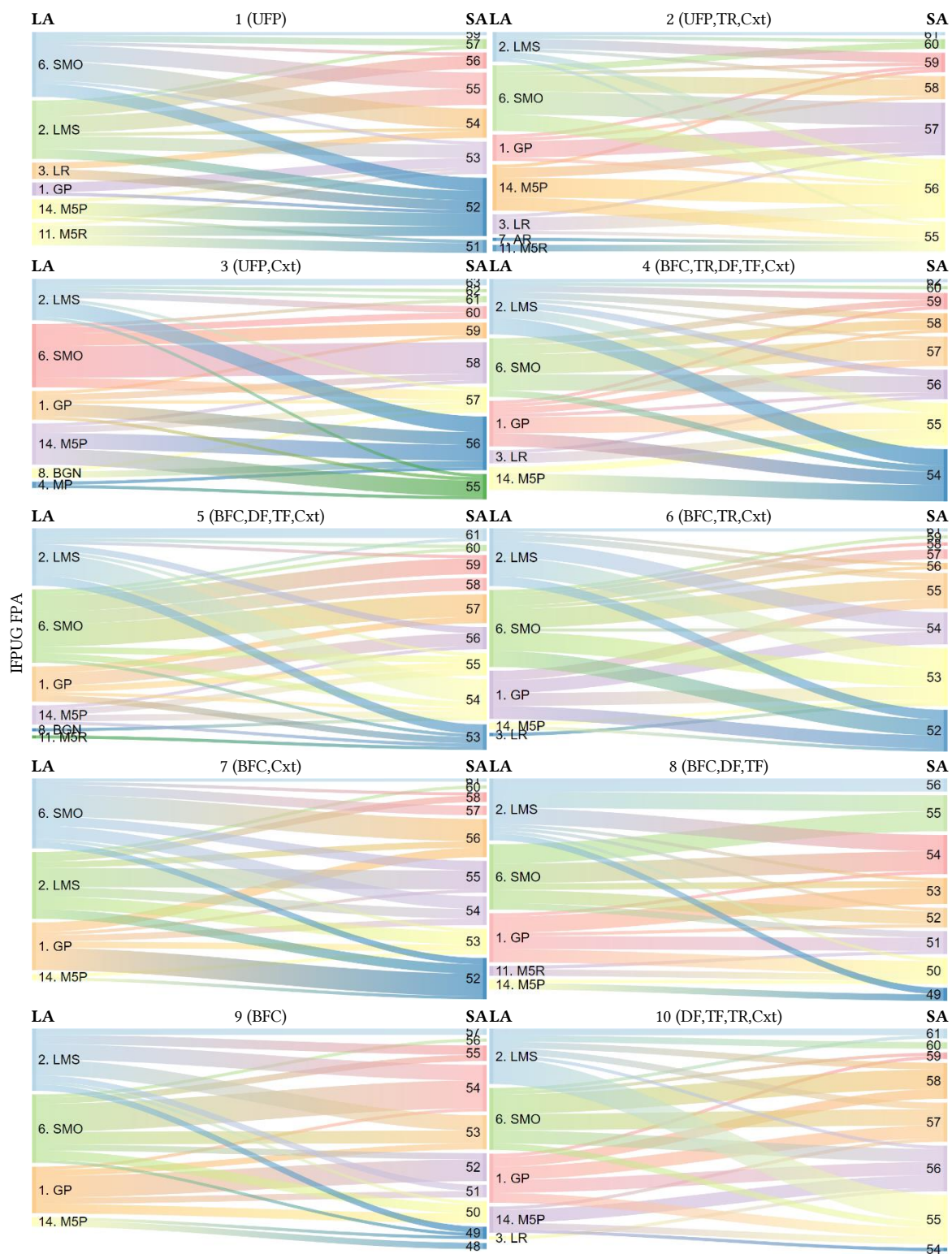
DS	DPxASxLA	SP	MRE	MdMRE	SA	Pred	DS	DPxASxLA	SP	MRE	MdMRE	SA	Pred		
DS	1	3x5x6	0.722	0.616	0.785	58.79	0.315	DS	12	1x5x6	0.716	0.621	0.799	57.97	0.301
202	2	3x5x2	0.788	0.557	0.736	61.47	0.314	202	13	3x3x2	0.695	0.649	0.857	55.10	0.285
	3	3x3x2	0.778	0.557	0.727	63.04	0.321		14	1x4x6	0.705	0.650	0.803	57.10	0.290
	4	3x4x2	0.749	0.620	0.772	61.57	0.303		15	1x2x2	0.670	0.685	0.824	55.69	0.294
	5	3x4x2	0.762	0.601	0.751	60.94	0.312		16	1x3x6	0.649	0.687	0.840	53.76	0.297
	6	3x4x2	0.747	0.597	0.768	61.02	0.301		17	1x4x2	0.617	0.684	0.924	54.53	0.260
	7	3x4x6	0.765	0.646	0.767	60.61	0.330		18	1x1x2	0.615	0.693	0.868	54.01	0.265
	8	3x2x2	0.702	0.649	0.809	56.43	0.300		19	1x4x2	0.563	0.752	1.002	49.94	0.243
	9	3x5x2	0.683	0.667	0.832	57.47	0.285		20	1x4x2	0.750	0.569	0.785	57.39	0.290
	10	3x4x6	0.803	0.618	0.734	61.11	0.355		21	3x1x6	0.554	0.820	0.973	50.21	0.232
	11	3x2x2	0.766	0.599	0.745	64.05	0.312		-	-	-	-	-	-	
DS	1	1x4x7	0.724	1.659	1.038	79.69	0.214	DS	6	1x2x14	0.809	2.452	1.906	72.65	0.155
53	2	1x2x7	0.721	1.478	1.007	79.86	0.201	53	7	3x3x14	0.799	1.937	1.524	75.80	0.176
	3	3x3x6	0.826	1.217	0.889	80.30	0.196		8	1x1x6	0.850	0.905	0.714	74.31	0.257
	4	3x3x6	0.795	1.658	0.838	77.43	0.256		9	3x4x6	0.808	1.244	0.943	67.31	0.170
	5	1x5x14	0.795	2.395	1.646	77.56	0.177								

La Figura 7-9 muestra los resultados de la métrica de la exactitud estandarizada (SA) para los modelos construidos a partir de las mediciones del método IFPUG FPA. Se detalla para cada grupo del conjunto de datos IFPUG FPA los resultados de exactitud obtenidos por cada técnica de construcción de los modelos (LA). La Figura 7-10 muestra los resultados para los modelos construidos a partir de las mediciones del método COSMIC FFP. Del mismo modo, se detalla para cada grupo del conjunto de datos COSMIC FFP los resultados de exactitud obtenidos por cada técnica de construcción de los modelos (LA). Los datos mostrados se obtienen de la selección de los mejores 60 modelos que representan el 90% percentil. Para cada uno de los grupos de los conjuntos de datos se detalla cómo una técnica (LA) está relacionada con la exactitud obtenida durante la evaluación (SA). Los mejores resultados de exactitud se encuentran ordenados del valor más alto al más bajo para los modelos del mejor 10% de los resultados.

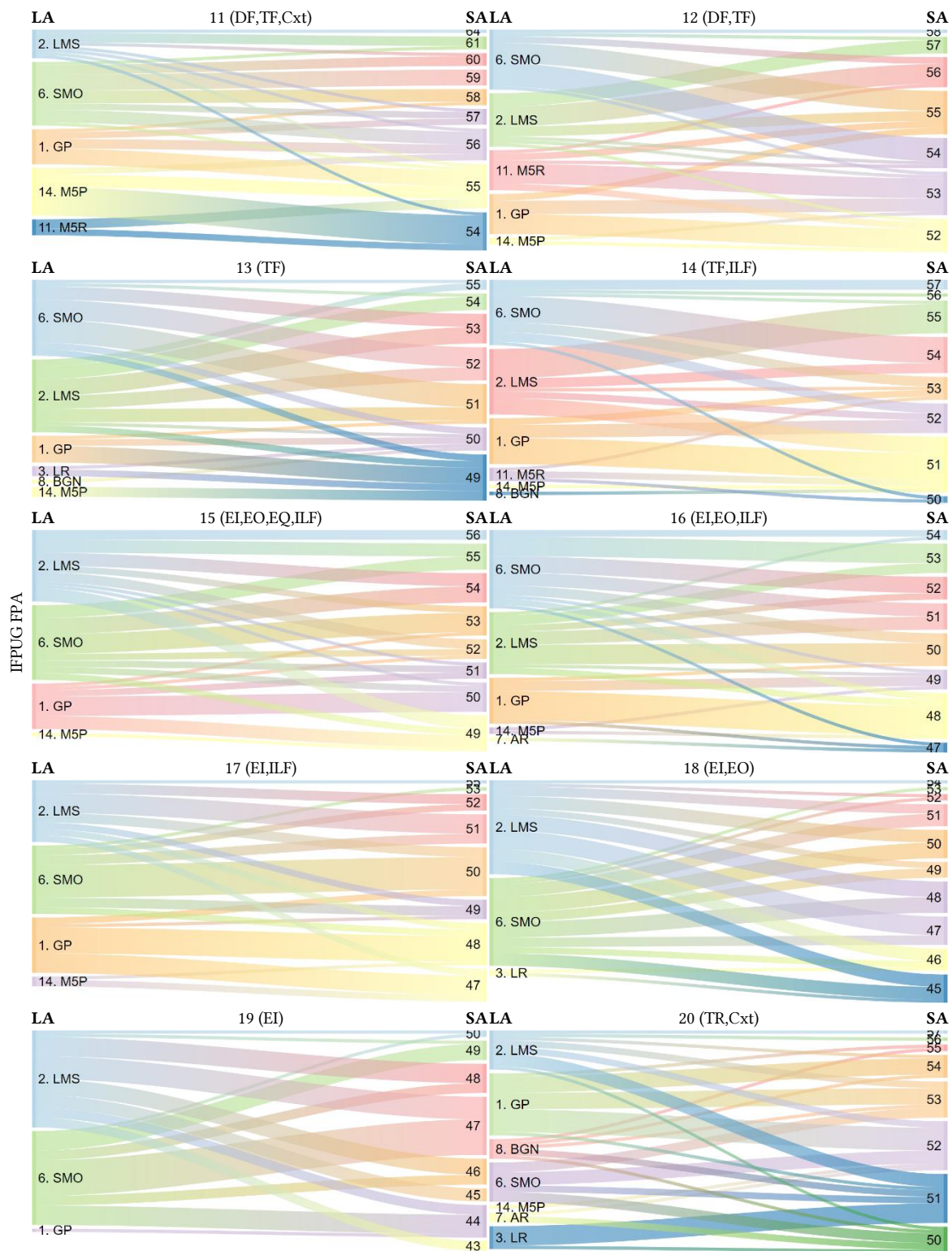
Considerando el mejor 10% de los resultados de la exactitud (SA) de los modelos, la mayor cantidad de técnicas de construcción (LA) para los grupos de los conjuntos de datos del método IFPUG FPA son las de la regresión (6) *SMOreg* que representa un 34.6% (n=436) del total, la regresión (2) *LeastMedSq* que representa un 30.7% (n=387) del total y la técnica (1) *GaussianProcesses* que implementa procesos gaussianos para la regresión no lineal con un 18.6% (n=234) del total. En el caso de los grupos del conjunto de datos del método

COSMIC FFP son las de (6) *SMOreg* que representa un 20.7% (n=112) del total, (7) *Additive-Regression* que representa un 20.4% (n=110) del total y (14) *MP5* que representa un 19.3% (n=104) del total.

Las combinaciones de técnicas de $DP \times AS \times LA$ más utilizadas para los grupos del conjunto de dato del método de medición IFPUG FPA son las de $BC \times BE \times SMO$ (3x4x6) que representa un 4.2% (n=53) del total, $BC \times FS \times SMO$ (3x5x6) que representa un 4.1% (n=51) del total, $BC \times GS \times SMO$ (3x1x6) y $BC \times BF \times SMO$ (3x2x6) que representa un 4% (n=50) del total, y $BC \times GS \times LMS$ (3x1x2) y $BC \times LFS \times SMO$ (3x3x6) que representa un 3.9% (n=49) del total. En el caso del método de medición COSMIC FFP son las de $BC \times BF \times AR$ (3x2x7) que representa un 3.7% (n=20) del total, $BC \times LFS \times AR$ (3x3x7) que representa un 3.5% (n=19) del total, y $BC \times BF \times SMO$ (3x2x6) y $BC \times LFS \times SMO$ (3x3x6) que representa un 2.8% (n=15) del total.



(Continúa)



(Continúa)

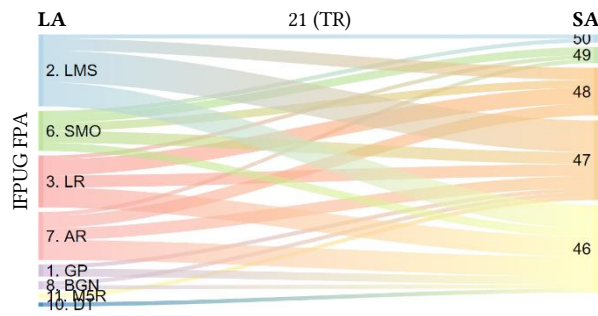


Figura 7-9 Mejores modelos (SA) para el conjunto de datos IFPUG FPA.

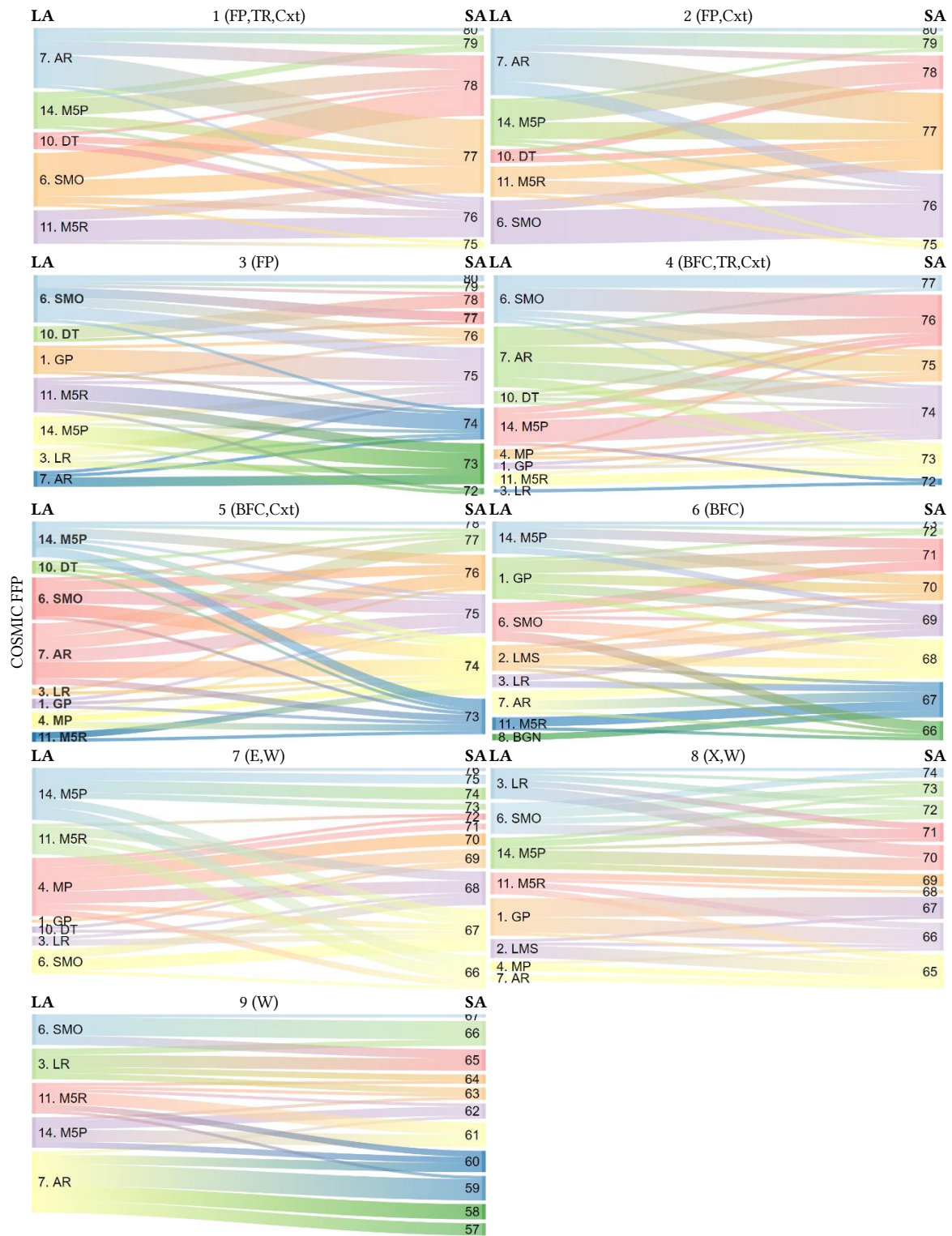


Figura 7-10 Mejores modelos (SA) para el conjunto de datos COSMIC FFP.

El Anexo 9 detalla el análisis de las amenazas a la validez de los estudios descritos en esta evaluación. Las amenazas a la validez del estudio son las relacionadas al repositorio de proyectos ISBSG, estudios de correlación y análisis de modelos de estimación de esfuerzo de desarrollo.

7.3.3. DISCUSIÓN DE RESULTADOS DEL ESTUDIO

Los resultados del estudio indican que el procedimiento de evaluación de modelos de estimación de esfuerzo presenta una repetibilidad aceptable en los resultados a través de múltiples corridas. Los análisis no encuentran diferencias significativas entre los resultados de exactitud de dos corridas realizadas. Esto implica que la estrategia basada en múltiples validaciones cruzadas y en selección aleatorias no afecta la obtención de resultados consistentes. Por tanto, el procedimiento de evaluación permite realizar comparaciones exhaustivas y sin sesgo entre los modelos de estimación. Con el procedimiento de evaluación es posible analizar la exactitud de los modelos de estimación, las técnicas que obtienen los mejores resultados de exactitud y los aspectos que pueden impactar en el desempeño.

Basado en la comparación de los mejores modelos, los resultados obtenidos ofrecen evidencia que confirma que pueden existir modelos de estimación de esfuerzo basados en los BFC que pueden tener un nivel de exactitud similar a los modelos construidos a partir de los FP. La evidencia indica que la construcción de los modelos de estimación se debe realizar para las características específicas de los conjuntos de datos para los cuales van a ser utilizados.

En la práctica, los resultados indican que es posible explorar la utilización de modelos de estimación de esfuerzo basados en los componentes funcionales básicos (BFC) con resultados de exactitud similares a los obtenidos por los modelos construidos a partir de los puntos de función. Simplificando el proceso de medición basado en el conteo de conjuntos de los BFC podría ser posible realizar ahorros en el esfuerzo de medición, preservando la exactitud de los modelos de estimación de esfuerzo.

La exactitud estandarizada (SA) permite la comparación de los modelos de estimación a partir del conjunto de datos bajo evaluación (Shepperd & MacDonell, 2012; Whigham, Owen & Macdonell, 2015; Dolado, Rodriguez, Harman, Langdon & Sarro, 2016). Los modelos de estimación deben maximizar el valor de la métrica SA (Shepperd & MacDonell, 2012). En estudios previos, la SA ha presentado valores que oscilan entre el 30% y 60% con valores que han alcanzado hasta el 80%.

Los modelos de estimación evaluados alcanzaron valores entre el 49.94% y 64.05% para el método de medición IFPUG FPA y entre el 80.30% y el 67.31% para el método de medición COSMIC FFP con base en la exactitud estandarizada (SA). Esto significa que la exactitud de los modelos puede superar la línea base de selección aleatoria hasta por un 54.05% en el caso de IFPUG FPA y hasta por un 80.30% en el caso de COSMIC FFP. Las técnicas de construcción de modelos (LA) que mejores resultados obtuvieron para los conjuntos de datos del método IFPUG FPA fueron: *LeastLeastMedSq*, *SMOreg* y *GaussianProcesses*. En el caso de los conjuntos de datos del método COSMIC FFP fueron las técnicas: *SMOreg*, *MP5* y *AdditiveRegression*. Los resultados indican que las técnicas que incorporan estrategias de regresión para su implementación son las que mejores resultados obtienen para estos conjuntos de datos.

Las combinaciones de técnicas de $DP \times AS \times LA$ que mejores resultados obtienen para el método de medición IFPUG FPA son las de $BC \times BF \times LMS$, $BC \times BE \times SMO$, $BC \times BE \times LMS$, y $BC \times LFS \times LMS$. En el caso del método de medición COSMIC FFP son las de $BC \times LFS \times SMO$, $BC \times FS \times AR$ y $None \times BF \times M5P$. Los resultados indican el impacto de las técnicas de $SP \times AS$ en las técnicas de construcción de modelos (LA) por lo que se requiere realizar análisis adicionales en esta línea.

La selección de las técnicas de construcción de modelos de estimación de esfuerzo depende de los conjuntos de datos para los cuales son creadas, por lo que el procedimiento de evaluación de la Sección 5.3 puede permitir a los profesionales identificar los modelos que mejor se ajustan a sus proyectos particulares. Asimismo, en un proceso de mejora continua, este procedimiento podría ejecutarse cada vez que existen nuevos conjuntos de datos

para ofrecer modelos de estimación actualizados. Una herramienta exhaustiva permite evaluar los modelos de estimación para determinar cuál combinación de técnicas se ajusta mejor al subconjunto de datos bajo evaluación. La estrategia de evaluación de modelos de estimación se propone como una herramienta de apoyo y esta no puede reemplazar los profesionales que administran los procesos de estimación. Finalmente, el objetivo de la estrategia de evaluación de modelos es servir como herramienta de apoyo a la toma de decisiones para los profesionales

7.4. RESUMEN Y DISCUSIÓN DE RESULTADOS DE LAS EVALUACIONES EMPÍRICAS

En este capítulo se presentaron las evaluaciones empíricas realizadas para la validación del procedimiento de evaluación de la metodología propuesta para la automatización de la medición del tamaño funcional **aFPA**. La evaluación de los modelos simplificados de estimación de esfuerzo basados en el tamaño funcional se realizó mediante experimentos controlados. Se realizaron tres experimentos para la evaluación de los modelos de estimación de esfuerzo de desarrollo basados en los métodos de medición del tamaño funcional IFPUG FPA y COSMIC FFP con subconjuntos de proyectos del repositorio ISBSG R12. Primero, se reporta una replicación externa llevada a cabo para evaluar el método de medición IFPUG FPA. Los resultados agregan evidencia sobre la exactitud de los modelos de estimación simplificados basados en el tamaño funcional IFPUG FPA. Algunos de los componentes funcionales básicos (BFC) del método de medición se encuentran correlacionados. Asimismo, existe una relación entre los BFC y el esfuerzo de desarrollo de software. La evidencia sugiere que es posible construir modelos de estimación de esfuerzo basados únicamente en las funciones transaccionales (TF) o en las entradas externas (EI), las salidas externas (EO) y los archivos lógicos internos (ILF) que pueden desempeñarse al menos con el mismo rendimiento que un modelo basado en la totalidad de los puntos de función sin ajustar (UFP).

Segundo, se reporta un estudio controlado experimental para evaluar el método de medición COSMIC FFP. La evidencia indica que los componentes funcionales básicos (BFC)

del método de medición COSMIC FFP se encuentran correlacionados. Asimismo, existe una relación entre los BFC y el esfuerzo de desarrollo de software. Los resultados indican que los modelos de estimación basados en los BFC podrían tener un desempeño similar a los basados en los FP. Por ejemplo, un modelo de regresión basado en las salidas (X) y las escrituras (W) podría desempeñarse al menos con la misma exactitud que un modelo basado en la totalidad de los puntos de función (FP).

Finalmente, se realiza un estudio comparativo exhaustivo que realiza una evaluación de los modelos de estimación de esfuerzo. Los resultados obtenidos ofrecen evidencia que confirma que pueden existir modelos de estimación de esfuerzo basados en los componentes funcionales básicos que pueden tener un nivel de exactitud similar a los modelos construidos a partir de los puntos de función.

La evidencia indica que la construcción de los modelos de estimación de esfuerzo se debe realizar tomando en cuentas las características específicas de los conjuntos de datos para los cuales van a ser utilizados. Por lo tanto, una herramienta de evaluación de modelos puede apoyar el proceso de construcción de modelos específicos para una organización de desarrollo. Aún más, en la práctica, los resultados indican que es posible explorar la utilización de modelos de estimación de esfuerzo simplificados lo que puede ayudar a reducir el esfuerzo de medición. Los modelos de estimación de esfuerzo simplificados deben ser construidos de acuerdo a las características de los proyectos históricos y el contexto específico de cada organización de desarrollo de software para lograr un mejor desempeño en los resultados.

Capítulo 8. CONCLUSIONES

En esta tesis se ha desarrollado una metodología para la automatización de la medición del tamaño funcional del software (**aFPA**). La metodología **aFPA** permite realizar el conteo automático del tamaño funcional utilizando los métodos de medición estandarizados IFPUG FPA y COSMIC FFP. Esta toma como entrada el código fuente de la aplicación para construir un modelo funcional cuyos componentes son mapeados a los métodos de medición IFPUG FPA y COSMIC FFP para producir la estimación del tamaño, manteniendo así una trazabilidad entre los artefactos y los resultados del conteo. Adicionalmente, **aFPA** es extendida para construir el modelo funcional utilizando como entrada los requerimientos funcionales bajo el estándar IEEE Std. 830-1998 y los modelos conceptuales generados por una herramienta de desarrollo dirigido por modelos.

La metodología **aFPA** se compone de tres procedimientos: el primero permite realizar la medición automática del tamaño funcional, el segundo verifica la exactitud de las mediciones obtenidas, y el tercero evalúa algunos modelos de estimación de esfuerzo para mostrar un posible aprovechamiento de los resultados de la medición del tamaño funcional. Se implementó una herramienta prototipo para cada uno de los tres procedimientos y se analizaron los resultados de su utilización. La metodología **aFPA** fue evaluada mediante un conjunto de casos de estudio, cuasi-experimentos y replicaciones que muestran la factibilidad de su uso en procesos de desarrollo de software. En las secciones siguientes se resumen los principales hallazgos de la investigación, se presentan las contribuciones de la tesis, y finalmente se discuten las direcciones futuras de investigación.

8.1. RESUMEN DE RESULTADOS

En esta tesis se ha estudiado la aplicación de la metodología **aFPA** que permite medir artefactos de software, tales como el código fuente de las aplicaciones, con el fin de obtener la medición de tamaño funcional con una exactitud aceptable para un sistema de medición en la industria (AIAG, 2002; Abran, 2015). Para lograr este objetivo, se ha realizado un

conjunto de ciclos de diseño y de investigación utilizando el marco de trabajo *Ciencia del Diseño* para Sistemas de Información e Ingeniería del Software. Esta metodología ha permitido guiar el diseño de la metodología **aFPA**, los procedimientos que la componen, los prototipos de soporte y los de estudios empíricos que la evalúan para generar conocimiento. La metodología de medición desarrollada detalla los lineamientos para obtener el tamaño funcional a partir de modelos funcionales obtenidos del análisis dinámico del código fuente y permite la medición funcional de los artefactos de acuerdo a los métodos de medición IFPUG FPA y COSMIC FFP. Los prototipos desarrollados apoyan la metodología de medición mediante la implementación de un módulo de soporte a la medición del tamaño funcional, un módulo de soporte para la verificación de exactitud y un módulo de soporte para la evaluación de modelos de estimación de esfuerzo.

Durante el desarrollo de la tesis se realizaron múltiples estudios empíricos. Se realizó un mapeo y una revisión sistemática de literatura para determinar los desafíos, oportunidades y enfoques existentes de los procedimientos para la automatización de la medición del tamaño funcional del software. Asimismo, se realizaron múltiples casos de estudio, un cuasi-experimento y una familia de replicaciones para determinar la exactitud de la medición del tamaño funcional resultado de la aplicación de los procesos de medición. Finalmente, se realizó una familia de experimentos controlados y un conjunto de replicaciones de estudios para evaluar la efectividad de los métodos de medición y sus componentes funcionales básicos en los modelos de estimación de esfuerzo.

Durante la evaluación de la metodología para la automatización de medición del tamaño funcional **aFPA**, se condujeron los estudios empíricos para determinar la efectividad de los procedimientos propuestos. Primero, se realizaron cuatro casos de estudio para determinar la exactitud de los resultados del procedimiento de medición. Los resultados de la exactitud del procedimiento para obtener los puntos de función sin ajustar IFPUG FPA presentaron una variación (MRE) de un 2.1%. Los resultados de la exactitud del procedimiento para obtener los puntos de función COSMIC FFP presentan una variación (MRE) de un 0.8%. La extensión del procedimiento para medir la especificación de requerimientos

IEEE Std 830-1998 presentaron una variación (MRE) de un 0.7% en los resultados de la medición funcional IFPUG FPA. Finalmente, los resultados de la exactitud del procedimiento para obtener los puntos de función IFPUG FPA utilizando modelos conceptuales obtuvieron una variación en la exactitud (MRE) de un 5.5%. Los resultados evidencian que el procedimiento de medición de la metodología **aFPA** puede obtener resultados aceptables de exactitud (AIAG, 2002; Abran, 2015).

La evaluación realizada para determinar la efectividad y las propiedades de adopción del procedimiento de medición y del procedimiento de verificación de la metodología **aFPA** se realizó mediante tres replicaciones diferenciadas de un estudio controlado cuasi experimental con 58 profesionales que tomaron cursos de maestría en computación e informática relacionados con temas de métricas de software. Los resultados de medición indicaron que, en promedio, los profesionales obtienen resultados de tamaño funcional con una variación en la exactitud y en la reproducibilidad menor al 10%. Los resultados del procedimiento de medición presentaron una variación promedio en la exactitud entre un 6% y un 9% con una reproducibilidad entre un 3% y un 7% (n=18). Además, el procedimiento de medición presentó mejores resultados en la reproducibilidad que los procesos de medición IFPUG FPA y OMG AFP. Los resultados indicaron que el procedimiento de medición obtuvo una exactitud similar cuando se aplica utilizando como entrada el código fuente de la aplicación o la especificación de los requerimientos. Las mediciones realizadas sobre el código fuente presentaron mejores resultados de reproducibilidad. Para las propiedades de adopción, los resultados indicaron que los procesos de medición IFPUG FPA y **aFPA** fueron percibidos de igual facilidad de uso y utilidad. El proceso de medición de **aFPA** fue percibido más útil que el proceso de medición OMG AFP. El procedimiento de medición de la metodología **aFPA** presentó una intención de uso similar que IFPUG FPA. El procedimiento de verificación de exactitud de la metodología **aFPA** obtuvo resultados que indican que fue percibido como fácil de usar, fue percibido como útil y que existe una intención de uso por parte de los participantes del estudio.

El análisis realizado para la evaluación de los modelos de estimación de esfuerzo basados en los métodos IFPUG FPA y COSMIC FFP se efectuó mediante tres experimentos. El procedimiento de evaluación de modelos de estimación de esfuerzo de la metodología **aFPA** fue validado para determinar su consistencia. Las pruebas realizadas obtuvieron resultados repetibles a través de múltiples corridas. Los análisis de la estructura interna de los métodos de medición encontraron correlaciones entre algunos de los componentes funcionales básicos (BFC) y además se reportaron relaciones entre los BFC y el esfuerzo de desarrollo de software. A partir de estos resultados se construyeron modelos de estimación de esfuerzo basados únicamente en subconjuntos de los BFC para los cuales, algunos de ellos, obtuvieron un desempeño de al menos el mismo rendimiento que un modelo basado en la totalidad de los puntos de función. Los modelos de estimación evaluados alcanzaron valores entre el 49.94% y el 64.05% para el método de medición IFPUG FPA y entre el 67.31% y el 80.30% para el método de medición COSMIC FFP con base en la exactitud estandarizada (SA). Los resultados obtenidos son similares a los reportados en estudios previos en el área que han alcanzado valores que oscilan entre el 30% y 60% con valores de hasta el 80% (Shepherd & MacDonell, 2012; Whigham, Owen & Macdonell, 2015; Dolado, Rodriguez, Harman, Langdon & Sarro, 2016). Las técnicas de construcción de modelos (LA) que mejores resultados obtuvieron para los conjuntos de datos del método IFPUG FPA fueron: *LeastMedSq*, *SMOreg* y *GaussianProcesses*. En el caso de los conjuntos de datos del método COSMIC FFP los mejores resultados los obtuvieron las técnicas: *SMOreg*, *MP5* y *AdditiveRegression*. Los resultados mostraron que las técnicas que incorporaron estrategias de regresión para su implementación son las que mejores resultados obtuvieron para los conjuntos de datos evaluados.

La aplicación del marco de trabajo *Ciencia del Diseño* para Sistemas de Información e Ingeniería del Software permitió determinar los desafíos, oportunidades y características de los enfoques existentes para la automatización de la medición del tamaño funcional, diseñar la metodología para la automatización de la medición del tamaño funcional (**aFPA**) y evaluar la efectividad de cada uno de los procedimientos diseñados mediante estudios empíricos. Por todo lo anterior, se cumplen los objetivos planteados para la investigación.

8.2. CONTRIBUCIONES

Las principales contribuciones de la tesis son las relacionadas con el diseño de la metodología para la medición automatizada del tamaño funcional y la evidencia empírica generada por los estudios realizados. Las principales contribuciones de la tesis son las siguientes:

- Una metodología para la automatización de la medición del tamaño funcional (**aFPA**) que permite mediante un proceso sistematizado obtener el tamaño funcional basado en los métodos IFPUG FPA y COSMIC FFP, verificar los resultados de medición y evaluar algunos modelos de estimación de esfuerzo para mostrar un posible aprovechamiento de los resultados de la medición del tamaño funcional.
- Un proceso sistemático para obtener modelos funcionales de representación del software que permita realizar el proceso de conteo del tamaño funcional, mantener trazabilidad entre los requerimientos del usuario final y el código fuente de la aplicación y visualizar la funcionalidad del software desde la perspectiva del usuario final.
- Un protocolo sistemático para la verificación de la exactitud de los resultados de la medición del tamaño funcional que evalúa los resultados totales del conteo y de cada uno de los componentes funcionales básicos de un método de medición del tamaño funcional.
- Tres herramientas prototipo de soporte para: (a) la medición del tamaño funcional, (b) la verificación de la exactitud y (c) la evaluación de modelos de estimación de esfuerzo basado en el código fuente y otros artefactos de diseño.
- Un mapeo y revisión sistemática de literatura que caracteriza los procedimientos de la medición del tamaño funcional existentes y presenta el estado del arte en el área de estudio.
- Una evaluación empírica de la estructura de los componentes funcionales básicos de los métodos de medición del tamaño funcional IFPUG FPA y COSMIC FFP para determinar

su impacto en la efectividad de los modelos de estimación del esfuerzo de desarrollo del software.

- Una evaluación empírica de la aplicabilidad de los componentes funcionales básicos de los métodos de medición del tamaño funcional IFPUG FPA y COSMIC FFP para la construcción de modelos simplificados de estimación.
- Una evaluación empírica sobre la efectividad de propuestas de medición del tamaño funcional del software a partir de diferentes artefactos de software generados durante el ciclo de vida del desarrollo del software.
- Una evaluación empírica de la efectividad de la metodología de medición **aFPA** de acuerdo con los métodos de medición IFPUG FPA y COSMIC FFP.
- Una evaluación empírica sobre la aplicabilidad del tamaño funcional para la construcción de modelos de estimación de esfuerzo simplificados utilizando los resultados de medición de los métodos IFPUG FPA y COSMIC FFP.
- Dos extensiones del procedimiento de medición de la metodología de medición **aFPA** para construir el modelo funcional utilizando como entrada los requerimientos funcionales bajo el estándar IEEE Std. 830-1998 y los modelos conceptuales generados por una herramienta de desarrollo dirigido por modelos.
- Una evaluación empírica de la efectividad del procedimiento de medición de la metodología de medición **aFPA** en un contexto industrial de acuerdo con el método de medición IFPUG FPA.

8.3. DIRECCIONES FUTURAS

En esta tesis se ha presentado una metodología para la automatización de la medición del tamaño funcional del software (**aFPA**). Asimismo, ha desarrollado los prototipos para proveer el soporte a los procedimientos de medición, verificación y evaluación. Como trabajo futuro, se podrían desarrollar nuevas funcionalidades de automatización para varios de los pasos del proceso de medición de tamaño funcional. Por ejemplo, el proceso de exploración de la aplicación medida podría realizarse automáticamente para obtener los

rastros de ejecución sin la intervención del profesional. Una vez obtenidos los trazos de ejecución por medio de la exploración automática, el profesional determinaría los que desea incluir en el alcance de la medición. Asimismo, es posible mejorar el proceso de identificación de funcionalidades duplicadas. En este caso, es de interés investigar cómo el prototipo puede, automáticamente, tomar las decisiones sobre si debe o no incluir un conjunto de funciones identificadas por el proceso de exploración. Actualmente, el procedimiento identifica los candidatos a funcionalidades duplicadas y la reporta; sin embargo, es el profesional el que toma las decisiones de cuáles funciones incluir en el alcance de la medición.

En esta línea de investigación, es necesario realizar más estudios empíricos para analizar la exactitud del procedimiento con aplicaciones transaccionales de un nivel de complejidad superior. Para esto, actualmente se está realizando un proceso de experimentación con aplicaciones reales en la industria. Asimismo, es de interés comparar el procedimiento de medición de la metodología **aFPA** con otras propuestas de automatización tales como el procedimiento OOmFP en su versión para IFPUG FPA y COSMIC FFP (Abrahamo, 2004; Abrahamo, Poels & Pastor, 2004a; Abrahamo, Poels & Insfran, 2008; Marín, Pastor & Giachetti, 2008). En particular, estos procedimientos obtienen el tamaño funcional a partir de modelos conceptuales.

Con respecto a los modelos funcionales, es de interés analizar si es posible la simplificación de sus componentes sin perder exactitud y manteniendo la semántica asociada a la trazabilidad del proceso de medición. Asimismo, es interesante explorar la posibilidad de identificar patrones de medición para los trazos de ejecución que se analizan contra el grafo del modelo. Esto es, analizar la posibilidad de calcular el tamaño funcional a partir de los patrones de los caminos del trazo de ejecución en el modelo funcional con la información de los tipos de nodos y etiquetas. Adicionalmente, es posible explorar distintas métricas obtenidas a partir del grafo del modelo funcional y estudiar su relación con el tamaño funcional.

Finalmente, en el procedimiento de evaluación para el aprovechamiento de los resultados de medición en la construcción de modelos de estimación de esfuerzo, es posible realizar múltiples análisis que permitan obtener conocimiento de cómo mejorar el proceso

de toma de decisiones sobre la selección de las técnicas utilizadas. Como trabajo futuro, se desea determinar el impacto de las distintas técnicas de pre procesamiento de datos, selección de atributos y algoritmos de aprendizaje en los resultados de exactitud de los modelos. Asimismo, es de interés evaluar las técnicas *filter* y *wrapper* para determinar en qué casos el costo computacional de un *wrapper* justifica su utilización. Además, se desea analizar el impacto de los parámetros de las técnicas utilizadas para determinar los valores óptimos que permitan mejorar los resultados de exactitud de los modelos. Finalmente, se desean analizar los conjuntos de variables seleccionadas en los modelos de predicción para explorar cuál es el conjunto mínimo de variables predictoras necesarias para mantener los resultados de exactitud esperados. Los análisis descritos anteriormente forman parte del trabajo en progreso que se está desarrollando. Asimismo, es de interés replicar los estudios utilizando el repositorio ISBSG R16.

Otras posibles líneas de investigación relacionadas con esta tesis son: 1) la convertibilidad de las mediciones del tamaño funcional desde el método IFPUG FPA al método COSMIC FFP y viceversa, 2) los procedimientos de aproximación del tamaño funcional en etapas tempranas de desarrollo y 3) la aplicación de procesos de medición de tamaño funcional en ambientes de desarrollo ágil.

REFERENCIAS

- Abdullah, N. A. S., Rusli, N. I. A., & Ibrahim, M. F. (2013). A case study in COSMIC functional size measurement: Angry bird mobile application. En 2013 IEEE Conference on Open Systems, ICOS 2013 (pp. 139-144). <http://doi.org/10.1109/ICOS.2013.6735063>
- Abdullah, N. A. S., Rusli, N. I. A., & Ibrahim, M. F. (2015). Mobile game size estimation: COSMIC FSM rules, UML mapping model and Unity3D game engine. En *ICOS 2014 - 2014 IEEE Conference on Open Systems* (pp. 42-47). <http://doi.org/10.1109/ICOS.2014.7042407>
- Abdullah, N. A. S., & Rusli, N. I. A. (2015b). Reviews on functional size measurement in mobile application and UML model. In *Proceedings of the 5th International Conference on Computing and Informatics* (pp. 353-358).
- Abrahao, S., & Pastor, O. (2003). Measuring the functional size of web applications. *International Journal of Web Engineering and Technology*, 1(1), 5. <http://doi.org/10.1504/IJWET.2003.003265>
- Abrahão, S. M. (2004). *On the Functional Size Measurement of Object-Oriented Conceptual Schemas: Design and Evaluation Issues*. Universidad Politecnica de Valencia (Spain).
- Abrahao, S., Poels, G., & Pastor, O. (2004a). Evaluating a functional size measurement method for Web applications: an empirical analysis. En *Software Metrics, 2004. Proceedings. 10th International Symposium on* (pp. 358-369).
- Abrahão, S., Poels, G., & Pastor, O. (2004). Assessing the reproducibility and accuracy of functional size measurement methods through experimentation. En *Proceedings - 2004 International Symposium on Empirical Software Engineering, ISESE 2004* (pp. 189-198). <http://doi.org/10.1109/ISESE.2004.1334906>
- Abrahão, S., & Poels, G. (2006). Functional Size Measurement Method for Object-Oriented Conceptual Schemas: Design and Evaluation Issues A Functional Size Measurement Method for Object-Oriented Conceptual Schemas: Design and Evaluation Issues. *Software & Systems Modeling*, 5(January), 48-71. Recuperado a partir de <http://link.springer.com/article/10.1007/s10270-005-0098-x>
- Abrahão, S., Mendes, E., Gomez, J., & Insfran, E. (2007). A model-driven measurement procedure for sizing Web applications: design, automation and validation. In *MODELS* (pp. 467-481).
- Abrahão, S., & Poels, G. (2007). Experimental evaluation of an object-oriented function point measurement procedure. *Information and Software Technology*, 49(4), 366-380. <http://doi.org/10.1016/j.infsof.2006.06.001>
- Abrahão, S., & Insfran, E. (2008). A metamodeling approach to estimate software size from requirements specifications. En *EUROMICRO 2008 - Proceedings of the 34th EUROMICRO Conference on Software Engineering and Advanced Applications, SEAA 2008* (pp. 465-475). <http://doi.org/10.1109/SEAA.2008.53>
- Abrahao, S., Poels, G., & Insfran, E. (2008). A replicated study on the evaluation of a size measurement procedure for web applications. En *Web Engineering, 2008. ICWE'08. Eighth International Conference on* (pp. 217-223).

- Abrahão, S., & Poels, G. (2009). The Journal of Systems and Software A family of experiments to evaluate a functional size measurement procedure for Web applications q. *The Journal of Systems & Software*, 82(2), 253-269. <http://doi.org/10.1016/j.jss.2008.06.031>
- Abrahão, S., De Marco, L., Ferrucci, F., Gravino, C., & Sarro, F. (2010). A COSMIC measurement procedure for sizing web applications developed using the OO-H method. En *Proceedings of the Workshop on Advances in Functional Size Measurement and Effort Estimation - FSM '10* (pp. 1-8). <http://doi.org/10.1145/1921705.1921707>
- Abrahão, S., Gómez, J., & Insfran, E. (2010). Validating a Functional Size Measure for Effort Estimation in Model- Driven Web Development 1 Introduction. *Information Sciences*, 180(20), 3932-3954.
- Abran, A., Meli, R., & Symons, C. (2004). COSMIC-FFP (ISO 19761) Software size measurement: State of the art 2004. En *Software Measurement European* Recuperado a partir de <http://publicationslist.org/data/a.abran/ref-2375/789.pdf>
- Abran, A. (2010). *Software Metrics and Software Metrology. Software Metrics and Software Metrology*. John Wiley & Sons. <http://doi.org/10.1002/9780470606834>
- Abran, A. (2015). *Software project estimation: the fundamentals for providing high quality information to decision makers*. John Wiley & Sons.
- Adem, N. A. Z., & Kasirun, Z. M. (2010). Automating function points analysis based on functional and non functional requirements text. En *2010 The 2nd International Conference on Computer and Automation Engineering, ICCAE 2010* (Vol. 5, pp. 664-669). <http://doi.org/10.1109/ICCAE.2010.5451322>
- Akça, A. A. (2012). Run-Time Measurement of Cosmic Functional Size for Java Business Applications. En *Software Measurement and the 2012 Seventh International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2012 Joint Conference of the 22nd International Workshop on* (pp. 226-231).
- Akca, A. A., & Tarhan, A. (2013). Run-time measurement of COSMIC functional size for Java business applications: Is it worth the cost? En *Proceedings - Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement, IWSM-MENSURA 2013* (pp. 54-59). <http://doi.org/10.1109/IWSM-Mensura.2013.19>
- Al-Batran, B., Schätz, B., & Hummel, B. (2011). Model Driven Engineering Languages and Systems. En *MoDELS* (Vol. 6981, pp. 258-272). Springer. <http://doi.org/10.1007/978-3-642-24485-8>
- Albrecht, A. J. (1979). Measuring application development productivity. En *Proceedings of the Joint SHARE/GUIDE/IBM Application Development Symposium* (Vol. 10, pp. 83-92). Recuperado a partir de <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Measuring+Application+Development+Productivity#0>
- Amasaki, S., & Yokogawa, T. (2012). A study on predictive performance of regression-based effort estimation models using base functional components. *Product-Focused Software Process Improvement*, 350-354.
- Antoniol, G., Calzolari, F., Cristoforetti, L., Fiutem, R., & Caldiera, G. (1998). Adapting function points to object oriented information systems. En *Advanced Information*

- Systems Engineering* (pp. 59-76).
- Antoniol, G., Fiutem, R., & Lokan, C. (2003). Object-oriented function points: An empirical validation. *Empirical Software Engineering*, 8(3), 225-254. <http://doi.org/10.1023/A:1024472727275>
- Antoniol, G., Lokan, C., Caldiera, G., & Fiutem, R. (1999). A function point-like measure for object-oriented software. *Empirical Software Engineering*, 4(3), 263-287.
- April, A., Merlo, E., & Abran, A. (1997). A reverse engineering approach to evaluate function point rules. En *Reverse Engineering, 1997. Proceedings of the Fourth Working Conference on* (pp. 236-245).
- AIAG, ASQ. (2002). Measurement Systems Analysis-Reference Manual. The Automotive Industries Action Group, Troy, MI.
- Aysolmaz, B., & Demirors, O. (2014). Automated functional size estimation using business process models with UPROM method. En *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2014 Joint Conference of the International Workshop on* (pp. 114-124).
- Aysolmaz, B., & Demirörs, O. (2015). Unified process modeling with UPROM tool. Lecture Notes in Business Information Processing, 204, 250-266. http://doi.org/10.1007/978-3-319-19270-3_16
- Azzouz, S., & Abran, A. (2003). A Proposed Measurement Role in the Rational Unified Process and its Implementation with ISO 19761: COSMIC-FFP. En *Software Measurement European Forum, Rome, Italy* (p. 12).
- Badampudi, D., Wohlin, C., & Petersen, K. (2015). Experiences from using snowballing and database searches in systematic literature studies. En *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering - EASE '15* (pp. 1-10). <http://doi.org/10.1145/2745802.2745818>
- Bagriyanik, S. & Karahoca, A. (2016). Automated COSMIC Function Point measurement using a requirements engineering ontology. *Information and Software Technology*, 72, 189-203. <http://doi.org/10.1016/j.infsof.2015.12.011>
- Bajwa, S. S., Gencel, C., & Abrahamsson, P. (2014). *Software Product Size Measurement Methods: A Systematic Mapping Study*. 2014 Joint Conference of the International Workshop on Software Measurement and the International Conference on Software Process and Product Measurement, 176-190. <http://doi.org/10.1109/IWSM.Mensura.2014.24>
- Barkallah, S., Gherbi, A., & Abran, A. (2011). COSMIC functional size measurement using UML models. En *Software Engineering, Business Continuity, and Education* (pp. 137-146). Springer.
- Basili, V. R. (1992). *Software modeling and measurement: the Goal/Question/Metric paradigm*.
- Basili, V. (2007). The role of controlled experiments in software engineering research. In *Empirical Software Engineering Issues. Critical Assessment and Future Directions* (pp. 33-37). Springer Berlin Heidelberg.
- Batista, V. A., Peixoto, D. C. C., Borges, E. P., Pádua, W., Resende, R. F., & Pádua, C. I. P. S. (2011). ReMoFP: A Tool for Counting Function Points from UML Requirement Models. *Advances in Software Engineering, 2011*, 1-7. <http://doi.org/10.1155/2011/495232>

- Bégnoche, L., Abran, A., & Buglione, L. (2007). A measurement approach integrating iso 15939, cmmi and the isbsg. En *Proc. of Software Measurement European Forum (SMEF) Conference, Rome, Italy* (pp. 111-130).
- Berg, K., Dekkers, T., & Oudshoorn, R. (2005). Functional size measurement applied to UML-based user requirements. *2nd Software Measurement European Forum, SMEF 2005*. Recuperado a partir de <http://doc.utwente.nl/64099/>
- Bertolami, M., & Oliveros, A. (2004). Functionality measurement process on the requirements elicitation. En *Proc. Workshop Ibero American Requirements Engineering and Software Environment, Arequipa-Per u* (pp. 91-102).
- Bevo, V., Levesque, G., & Abran, A. (1999). UML notation for functional size measurement method. En *Proc. 9th International Workshop on Software Measurement, Canada* (pp. 230-242).
- Bevo, E. V., & Lévesque, G. (2006). Development of computer aided measurement tools for software functional size measurement: An ontology driven approach. En *Proceedings of the 10th IASTED International Conference on Software Engineering and Applications, SEA 2006* (pp. 288-293).
- Biolchini, J., Mian, P. G., Natali, A. C. C., & Travassos, G. H. (2005). Systematic Review in Software Engineering. *Technical Report ES 679/05, 679*(May), 45.
- Bodhuin, T., Pagnozzi, F., Santone, A., Tortorella, M., & Villani, M. L. (2009). Abstracting models from execution traces for performing formal verification. *ASEA*, 143-150.
- Boehm, B. W. (1983). *Software Engineering Economics. Software Engineering, IEEE Transactions* (Vol. SE-10). Prentice-hall Englewood Cliffs (NJ).
<http://doi.org/10.1109/TSE.1984.5010193>
- Brereton, P., Kitchenham, B., & Budgen, D. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of systems and ...*, 80(4), 571-583. Recuperado a partir de <http://www.sciencedirect.com/science/article/pii/S016412120600197X>
- Budgen, D., & Brereton, P. (2006). Performing systematic literature reviews in software engineering. En *Int. Conf. Soft. Engin.* (Vol. 2, p. 1051).
<http://doi.org/10.1145/1134285.1134500>
- Buglione, L., & Gencel, C. (2008). Impact of base functional component types on software functional size based effort estimation. En *International Conference on Product Focused Software Process Improvement* (pp. 75-89).
- Bundschuh, M., & Dekkers, C. (2008). *The IT measurement compendium: estimating and benchmarking success with functional size measurement*. Springer Science & Business Media.
- Caldiera, G., Antoniol, G., Fiutem, R., & Lokan, C. (1998). Definition and experimental evaluation of function points for object-oriented systems. En *Software Metrics Symposium, 1998. Metrics 1998. Proceedings. Fifth International* (pp. 167-178).
- Cantone, G., Pace, D., & Calavaro, G. (2004). Applying function point to unified modeling language: Conversion model and pilot study. En *Proceedings - International Software Metrics Symposium* (pp. 280-291). <http://doi.org/10.1109/METRIC.2004.1357912>
- Carver, J. (2010). Towards reporting guidelines for experimental replications: a proposal. En *1st International Workshop on Replication in Empirical ...* (pp. 2-5). Recuperado a

- partir de http://carver.cs.ua.edu/Papers/Conference/2010/2010_RESER.pdf
- Carver, J., Juristo, N., Baldassarre, M. & Vegas, S. (2014). Replications of *software engineering* experiments. *Empirical Software Engineering*. (pp. 267--276).
- Ceke, D., Durek, M., & Kasapovic, S. (2013). Web application functional size estimation based on COSMIC method and UWE approach. En *Information & Communication Technology Electronics & Microelectronics (MIPRO), 2013 36th International Convention on* (pp. 396-403).
- Ceke, D., & Milasinovic, B. (2015). Automated web application functional size estimation based on a conceptual model. En *Software, Telecommunications and Computer Networks (SoftCOM), 2015 23rd International Conference on* (pp. 234-241).
- Chamundeswari, A., & Babu, C. (2008). An Extended Function Point Approach for Size Estimation of Object-Oriented *Software*. En *ICSOFT 2008 - Proceedings of the 3rd International Conference on Software and Data Technologies* (Vol. SE, pp. 139-145). Recuperado a partir de <http://www.scopus.com/inward/record.url?eid=2-s2.0-57649155775&partnerID=tZOtx3y1>
- Chamundeswari, A., & Babu, C. (2010). An extended function point approach for size estimation of object-oriented *software*. In *Proceedings of 3rd India Software Engineering Conference (ISECESE), India, Pebruary*.
- Chen, Y., Li, Q., & Gu, X. (2010). Estimating the size of applications in MDA environments. En *2nd International Conference on Information Science and Engineering, ICISE2010 - Proceedings* (pp. 4526-4529). <http://doi.org/10.1109/ICISE.2010.5689284>
- Choi, S., Park, S., & Sugumaran, V. (2006). Function Point Extraction Method from Goal and Scenario Based Requirements Text*. En *Natural Language Processing and Information Systems Lecture Notes in Computer Science* (pp. 12-24). Springer. Recuperado a partir de http://link.springer.com/chapter/10.1007/11765448_2
- Choi, S., Park, S., & Sugumaran, V. (2012a). A rule-based approach for estimating *software* development cost using function point and goal and scenario based requirements. *Expert Systems with Applications*, 39(1), 406-418.
- Choi, S., Park, S., & Sugumaran, V. (2012b). Expert Systems with Applications A rule-based approach for estimating *software* development cost using function point and goal and scenario based requirements. *Expert Systems With Applications*, 39(1), 406-418. <http://doi.org/10.1016/j.eswa.2011.07.029>
- Cito, J., Oliveira, F., Leitner, P., Nagpurkar, P., & Gall, H. C. (2017). Context-based analytics: establishing explicit links between runtime traces and source code. In *Proceedings of the 39th International Conference on Software Engineering: Software Engineering in Practice Track* (pp. 193-202). IEEE Press.
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15-18.
- Cook, R. D. (2000). Detection of influential observation in linear regression. *Technometrics*, 42(1), 65-68.
- Cook, J. E., & Wolf, A. L. (1998). Discovering models of *software* processes from event-based data. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 7(3), 215-249.
- Condori-Fernandez, N., Abrahao, S., & Pastor, O. (2004). Towards a functional size

- measure for object-oriented systems from requirements specifications. En *Fourth International Conference on Quality Software, 2004. QSIC 2004. Proceedings.* (pp. 94-101). <http://doi.org/10.1109/QSIC.2004.1357949>
- Condori-Fernández, N. (2006). Un Procedimiento de Medición de Tamaño Funcional para Especificaciones de Requisitos. Universidad Politecnica de Valencia (Spain).
- Condori-fernández, N., & Pastor, O. (2006a). An Empirical Study on the Likelihood of Adoption in Practice of a Size Measurement Procedure for Requirements Specification 2 . The Requirements Model and the RmFFP measurement procedure. En *2006 Sixth International Conference on Quality Software (QSIC'06)* (pp. 133-140).
- Condori-Fernández N., P. Ó. (2006b). Evaluating the productivity and reproducibility of a measurement procedure. En *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 4231 LNCS, pp. 352-361). Springer. Recuperado a partir de <http://www.scopus.com/inward/record.url?eid=2-s2.0-33845186373&partnerID=40&md5=e71a0aa07b69b29a41d60fccebc4dec>
- Condori-Fernandez, N., & Pastor, O. (2006c). Re-assessing the intention to use a measurement procedure based on COSMIC-FFP. *Mensura*, 63.
- Condori-Fernandez, N. ., Pastor, O. ., Daneva, M. ., Abran, A. ., & Castro, J. . (2008). Quantifying functional reuse from object oriented requirements specifications. *11th Workshop on Requirements Engineering, WER 2008 - Proceedings*, 59-69. Recuperado a partir de <http://www.scopus.com/inward/record.url?eid=2-s2.0-84870406531&partnerID=40&md5=d9af57fcc04713ab23d3edf2fbef7136>
- Condori-Fernandez, N., Abrahao, S., & Pastor, O. (2007). On the Estimation of Software Functional Size from Requirements Specifications. *Journal of Computer Science and Technology (JCST), Special Issue on Advances in Software Metrics and Software Processes. Springer (accepted for publication)*.
- Condori-Fernández, N., & Pastor, O. (2008). Analyzing the influence of certain factors on the acceptance of a model-based measurement procedure in practice: An empirical study. *CEUR Workshop Proceedings*, 392, 61-70.
- COSMIC – Common Software Measurement International Consortium. (2015). The COSMIC Functional Size Measurement Method Versión 4.0.1 Course Registration ('C-REG') System Case Study. Version 2.0.
- D'Avanzo, L., Ferrucci, F., Gravino, C., & Salza, P. (2015). Cosmic functional measurement of mobile applications and code size estimation. En *Proceedings of the 30th Annual ACM Symposium on Applied Computing* (pp. 1631-1636).
- Davis, F. D. (1989). Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319-340. <http://doi.org/10.2307/249008>
- De Freitas Junior, M., Fantinato, M., & Sun, V. (2015). Improvements to the Function Point Analysis Method: A Systematic Literature Review. *Engineering Management, IEEE Transactions on*, 62(4), 495-506.
- De Vito, G., & Ferrucci, F. (2014). Approximate COSMIC size: The quick/early method. En *Proceedings - 40th Euromicro Conference Series on Software Engineering and Advanced Applications, SEAA 2014* (pp. 69-76). <http://doi.org/10.1109/SEAA.2014.30>

- Del Bianco, V., & Lavazza, L. (2005). An empirical assessment of function point-like object-oriented metrics. En *Proceedings - International Software Metrics Symposium* (Vol. 2005, pp. 371-380). <http://doi.org/10.1109/METRICS.2005.9>
- Del Bianco, V., Gentile, C., & Lavazza, L. (2008). An Evaluation of Function Point Counting Based on Measurement-Oriented Models. In EASE.
- Del Bianco, V., & Lavazza, L. (2009). Applying the COSMIC functional size measurement method to problem frames. En *Proceedings of the IEEE International Conference on Engineering of Complex Computer Systems, ICECCS* (pp. 282-290). <http://doi.org/10.1109/ICECCS.2009.25>
- Del Bianco, V., Lavazza, L., Liu, G., Morasca, S., & Abualkishik, A. Z. (2013). Model-based Simplified Functional Size Measurement-an Experimental Evaluation with COSMIC Function Points. In EESSMOD@ MoDELS (pp. 13-22).
- Del Bianco, V., Lavazza, L., Liu, G., Morasca, S., & Abualkishik, A. Z. (2014). Model-based early and rapid estimation of COSMIC functional size-An experimental evaluation. *Information and Software Technology*, 56(10), 1253-1267.
- DeMarco, T. (1986). *Controlling Software Projects: Management, Measurement, and Estimates*. Prentice Hall PTR.
- Demirel, H. & Özkan, B. (2017). Automated Functional Size Measurement For Multi-tier Object Relational Mapping Architectures. *Collegium of Economic Analysis Annals*, Warsaw School of Economics, Collegium of Economic Analysis, issue 43, pages 51-68.
- Dejaeger, K., Verbeke, W., Martens, D., & Baesens, B. (2012). Data mining techniques for software effort estimation: a comparative study. *IEEE transactions on software engineering*, 38(2), 375-397.
- Desharnais, J.-M. ., Buglione, L. ., & Kocatürk, B. . (2011). Using the COSMIC method to estimate Agile User Stories. En *ACM International Conference Proceeding Series* (pp. 68-73). <http://doi.org/10.1145/2181101.2181117>
- De Souza, L. S., & de Aquino Jr, G. S. (2014). Estimating the Effort of Mobile Application Development. In *Proceedings of Second International Conference on Computational Science and Engineering* (pp. 45-63).
- Diab, H., Frappier, M., & Denis, R. St. (2001). Formalizing COSMIC-FFP using ROOM. En *Proceedings ACS/IEEE International Conference on Computer Systems and Applications* (pp. 312-318). <http://doi.org/10.1109/AICCSA.2001.934002>
- Diab, H., Frappier, M., & St-Denis, R. (2002). A Formal Definition of Function Points for Automated Measurement of B Specifications. En *Icfem 2002* (pp. 483-494). Springer.
- Diab, H., Koukane, F., Frappier, M., & St-Denis, R. (2005). cROSE: Automated measurement of COSMIC-FFP for Rational Rose RealTime. *Information and Software Technology*, 47(3), 151-166. <http://doi.org/10.1016/j.infsof.2004.06.007>
- Dixon-woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. (2016). Synthesising qualitative and quantitative evidence: a review of possible methods. *Journal of health services research & policy*, 10(1), 45-53B.
- Dolado, J. J., Rodriguez, D., Harman, M., Langdon, W. B., & Sarro, F. (2016). Evaluation of estimation models using the Minimum Interval of Equivalence. *Applied Soft Computing*, 49, 956-967.

- Dolado, J. J., & Fernandez, L. (2000). *Medición para la gestión en la ingeniería del software*. Libro Editorial RA-MA, ISBN, 84-7897.
- Duarte, L. M., Kramer, J., & Uchitel, S. (2006). Model extraction using context information. *In International Conference on Model Driven Engineering Languages and Systems* (pp. 380-394). Springer, Berlin, Heidelberg.
- Duarte, L. M., Kramer, J., & Uchitel, S. (2017). Using contexts to extract models from code. *Software & Systems Modeling*, 16(2), 523-557.
- Dumke, R., & Abran, A. (2011). *COSMIC Function Points: Theory and Advanced Practices*. CRC Press.
- Easterbrook, S., Singer, J., Storey, M.-A., & Damian, D. (2008). Selecting Empirical Methods for *Software Engineering Research*. In *Guide to Advanced Empirical Software Engineering* (pp. 285-311). Springer. http://doi.org/10.1007/978-1-84800-044-5_11
- Edagawa, T., Akaike, T., Higo, Y., Kusumoto, S., Hanabusa, S., & Shibamoto, T. (2011). Function point measurement from Web application source code based on screen transitions and database accesses. *Journal of Systems and Software*, 84(6), 976-984. <http://doi.org/10.1016/j.jss.2011.01.029>
- Ellafi, R., & Meli, R. (2006). A Source Code Analysis-based Function Point Estimation Method integrated with a Logic Driven Estimation Method. *Smef*.
- Eren, O., Ozkan, B., & Demirors, O. (2015). PL FSM: An Approach and a Tool for the Application of FSM in SPL Environments. En *Software Engineering and Advanced Applications (SEAA), 2015 41st Euromicro Conference on* (pp. 119-126).
- Fetcke, T., Abran, A., & Nguyen, T. H. (1997). Mapping the OO-Jacobson approach into function point analysis. In *Technology of Object-Oriented Languages and Systems, 1997. TOOLS 23. Proceedings* (pp. 192-202). IEEE.
- Fenton, N. E., & Pfleeger, S. L. (1997). *Software Metrics: A Rigorous and Practical Approach. It Professional* (Vol. 2). CRC Press. <http://doi.org/10.1201/b17461>
- Ferrucci, F., Harman, M., & Sarro, F. (2014). Search-based *software* project management. In *Software Project Management in a Changing World* (pp. 373-399). Springer Berlin Heidelberg.
- Ferrucci, F., Gravino, C., Salza, P., & Sarro, F. (2015). Investigating Functional and Code Size Measures for Mobile Applications. En *Product-Focused Software Process Improvement* (pp. 271-287). Springer.
- Fetcke, T., Abran, A., & Nguyen, T.-H. (1997). Mapping the OO-Jacobsen Approach into Function Point Analysis. En *Tools'USA* (pp. 192-202).
- Fetcke, T., Abran, A., & Nguyen, T.-H. (1998). Function point analysis for the OO-Jacobson method: a mapping approach. En *Proceedings of the FESMA* (Vol. 98, pp. 395-410).
- Fetcke, T. (1999). *The Warehouse Software Portfolio A Case Study in Functional Size Measurement*.
- Fehlmann, T. M., & Kranich, E. (2011). COSMIC functional sizing based on UML sequence diagrams. MetriKon, Kaiserslautern.
- Fingerman, S. (2011). *Practical software project estimation; a toolkit for estimating software development effort & duration*. *Sci-Tech News* (Vol. 65). McGraw Hill Professional.
- Florac, W. A., & Carleton, A. D. (1999). *Measuring the software process: statistical process*

- control for software process improvement*. Addison-Wesley Professional.
- Fodor, I. K. (2002). A survey of dimension reduction techniques (No. UCRL-ID-148494). Lawrence Livermore National Lab., CA (US).
- Frappier, M. (1999). An Overview of Formal Specification Languages and their Adequacy for Formalizing the Definition of Function Points. *Project FP-Formal Technical Report*, 1-23.
- Fraternali, P., Tisi, M., & Bongio, A. (2006). Automating function point analysis with model driven development. En *Proceedings of the 2006 conference of the Center for Advanced Studies on Collaborative research* (p. 18).
- Garmus, D., & Herron, D. (2001). *Function point analysis: measurement practices for successful software projects*. Addison-Wesley Longman Publishing Co., Inc.
- Gencel, C. (2008). How to Use COSMIC Functional Size in Effort Estimation Models? En *Software Process and Product Measurement* (pp. 196-207). Springer.
- Gencel, C., & Buglione, L. (2008). Do base functional component types affect the relationship between *software* functional size and effort? En *Software Process and Product Measurement* (pp. 72-85). Springer.
- Gencel, C., & Demirors, O. (2008). Functional size measurement revisited. *ACM Transactions on Software Engineering and Methodology*, 17(3), 1-36.
<http://doi.org/10.1145/1363102.1363106>
- Genero, M., Cruz-Lemus, J. A., & Piattini, M. (2014). Métodos de investigación en ingeniería del *software*. Editorial RA-MA: Madrid, Spain, 171-199.
- Gerlec, Č., Krajnc, A., & Goljat, U. (2010). A method for evaluating the actual functional size in model driven *software* development. In *Software Technology and Engineering (ICSTE), 2010 2nd International Conference on* (Vol. 1, pp. V1-185). IEEE.
- Giachetti, G., Marin, B., Condori-Fernandez, N., & Molina, J. C. (2007). Updating oo-method function points. En *Quality of Information and Communications Technology, 2007. QUATIC 2007. 6th International Conference on the* (pp. 55-64).
- Giachetti, G., & Pastor, O. (2008). Measurement of Functional Size in Conceptual Models: A Survey of Measurement Procedures. En *Software Process and Product Measurement* (pp. 170-183). Springer.
- Gonultas, R., & Tarhan, A. (2015). Run-Time Calculation of COSMIC Functional Size via Automatic Installment of Measurement Code into Java Business Applications. En *Software Engineering and Advanced Applications (SEAA), 2015 41st Euromicro Conference on* (pp. 112-118).
- González-Ladrón-de-Guevara, F., & Fernández-Diego, M. (2014). ISBSG variables most frequently used for *software* effort estimation: a mapping review. In *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement* (p. 42). ACM.
- González-Ladrón-de-Guevara, F., Fernández-Diego, M., & Lokan, C. (2016). The usage of ISBSG data fields in *software* effort estimation: A systematic mapping study. *Journal of Systems and Software*, 113, 188-215.
- Gramantieri, F., Lamma, E., Mello, P., & Riguzzi, F. (1997). A system for measuring function points from specifications. DEIS--Universita Di Bologna, Bologna and Dipartimento Di Ingegneria, Ferrara, Tech. Rep DEIS-LIA-97-006.

- Grau, G., & Franch, X. (2007). Using the PRiM method to Evaluate Requirements Model with COSMIC-FFP. En *International Conference on Software Process and Product Measurement (IWSM-MENSURA)* (pp. 110-120).
- Grau, G. (2008). Adapting the COSMIC method for evaluating the functional size in PRiM. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4895 LNCS, 139-153. http://doi.org/10.1007/978-3-540-85553-8_12
- Gu, X., Song, G., & Li, Q. (2006). An improved FSM method for web-based applications. En *CIMCA 2006: International Conference on Computational Intelligence for Modelling, Control and Automation, Jointly with IAWTIC 2006: International Conference on Intelligent Agents Web Technologies ...* (p. 21). <http://doi.org/10.1109/CIMCA.2006.41>
- Gupta, D. ., Kaushal, S. J. ., & Sadiq, M. . (2008). *Software estimation tool based on three - layer model for software engineering metrics*. En *Proceedings of the 4th IEEE International Conference on Management of Innovation and Technology, ICMIT* (pp. 623-628). <http://doi.org/10.1109/ICMIT.2008.4654437>
- Habela, P., Glowacki, E., Serafinski, T., & Subieta, K. (2011). 3.4 Adapting the Use Case Model for COSMIC FFP-Based Measurement. *COSMIC Function Points: Theory and Advanced Practices*, 204. <http://doi.org/10.1017/CBO9781107415324.004>
- Habra, N., Abran, A., Lopez, M., & Sellami, A. (2008). A framework for the design and verification of *software measurement methods*. *Journal of Systems and Software*, 81(5), 633-648. <http://doi.org/10.1016/j.jss.2007.07.038>
- Han, W. M., & Chen, W. T. (2013). Sizing cloud applications with ISO/IEC 19761: A case study. En *Lecture Notes in Electrical Engineering* (Vol. 234 LNEE, pp. 403-409). Springer. http://doi.org/10.1007/978-1-4614-6747-2_48
- Haoues, M., Sellami, A., & Ben-Abdallah, H. (2017). A Rapid Measurement Procedure for SizingWeb and Mobile Applications based on COSMIC FSM Method. In *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA)*, 2017 Joint Conference of the International Workshop on (pp. 130-138). ACM.
- Harman M, Jones BF (2001) Search-based *software engineering*. *Information in Software Technology*. 43(14):833–839.
- Harman M, Burke E, Clark JA, Yao X (2012) Dynamic adaptive search based *software engineering*. In: *Empirical Software Engineering and Measurement (ESEM)*, 2012 ACM-IEEE International Symposium On. IEEE, Lund Sweden. pp 1–8.
- Harput, V., Kaindl, H., & Kramer, S. (2005). Extending function point analysis to object-oriented requirements specifications. En *Proceedings - International Software Metrics Symposium* (Vol. 2005, pp. 361-370). <http://doi.org/10.1109/METRICS.2005.25>
- Heeringen, H. Van, & Gorp, E. Van. (2014). Measure the functional size of a mobile app: Using the cosmic functional size measurement method. En *Proceedings - 2014 Joint Conference of the International Workshop on Software Measurement, IWSM 2014 and the International Conference on Software Process and Product Measurement, Mensura 2014* (pp. 11-16). <http://doi.org/10.1109/IWSM.Mensura.2014.8>
- Heller, R. (2013). Automated function point counting--a fact based analysis. *Q/P Management Group, Inc. www.qpmg.com/Library/confirm_download.php*.

- Hericko, M., Rozman, I., & Zivkovic, A. (2006). A formal representation of functional size measurement methods. *Journal of Systems and Software*, 79(9), 1341-1358. Recuperado a partir de <http://www.sciencedirect.com/science/article/B6V0N-4J0XS1M-1/2/d4c2a5d67e028451056977b972749dcc>
- Hevner, von A., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 28(1), 75-105.
- Hill, P. (2010). *Practical software project estimation: a toolkit for estimating software development effort & duration*. McGraw Hill Professional.
- Ho, V. T., & Abran, A. (1999). A Framework for Automatic Function Point Counting from Source Code. En *International Workshop on Software Measurement* (pp. 248-255).
- Holzmann, G. J., & Smith, M. H. (1999). A practical method for verifying event-driven software. In Proceedings of the 21st international conference on *Software engineering* (pp. 597-607). ACM.
- Huang, J., Li, Y. F., & Xie, M. (2015). An empirical analysis of data preprocessing for machine learning-based software cost estimation. *Information and software Technology*, 67, 108-127.
- Huijgens, H., Bruntink, M., van Deursen, A., van der Storm, T., & Vogelezang, F. (2016). An exploratory study on functional size measurement based on code. En *Proceedings of the International Workshop on Software and Systems Process* (pp. 56-65).
- Hussain, I., Ormandjieva, O., & Kosseim, L. (2009). Mining and Clustering Textual Requirements to Measure Functional Size of Software with COSMIC. In *Software Engineering Research and Practice* (pp. 599-605).
- Hussain, I., Kosseim, L., & Ormandjieva, O. (2010). Towards approximating COSMIC functional size from user requirements in agile development processes using text mining. *Natural Language Processing and Information Systems*, 80-91.
- Hussain, I., Kosseim, L., & Ormandjieva, O. (2013). Approximation of COSMIC functional size to support early effort estimation in Agile. *Data & Knowledge Engineering*, 85, 2-14.
- Iorio, T. (2004). IFPUG Function Point analysis in a UML framework. *Proceedings of Software Measurement European Forum*. Recuperado a partir de <http://www.dpo.it/resources/papers/2004-smef-iorio-en.pdf>
- Iorio, T., Meli, R., & Perna, F. (2007). Early & Quick Function Points® v3. 0: enhancements for a Publicly Available Method. In Proceedings *Software Measurement European Forum (SMEF)* (pp. 179-198).
- Irawati, A. R., & Mustofa, K. (2012). Measuring Software Functionality Using Function Point Method Based On Design Documentation. *International Journal of Computer Science Issues*, 9(3/1), 124-130.
- ISO. (2004). *International vocabulary of basic and general terms in metrology (VIM)*. Geneva, CH.
- ISO. (2007). *Information technology – Software measurement – Functional size measurement. Part 1: Definition of concepts. International Standard ISO/IEC 14143-1* (Vol. 2007). Geneva, CH.
- ISO. (2009). *ISO/IEC 20926:2009 Software and systems engineering - Software measurement - IFPUG functional size measurement method 2009* (Vol. 2009). Geneva, CH. Recuperado

- a partir de http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=51717
- ISO. (2011). *ISO/IEC 19761:2011 Software engineering -- COSMIC: a functional size measurement method 2011* (Vol. 2011). Geneva, CH. Recuperado a partir de http://www.iso.org/iso/catalogue_detail.htm?csnumber=54849
- ISO. (2017). *ISO/IEC/IEEE 15939:2017 Systems and software engineering -- Measurement process*. Geneva, CH. Recuperado a partir de <https://www.iso.org/standard/71197.html>
- ISBSG. (2013). The ISBSG Development & Enhancement project data. ISBSG, Ed., R 12 ed.
- Jacquet, J. P., & Abran, A. (1997). From *software* metrics to *software* measurement methods: A process model. En *Software Engineering Standards Symposium and Forum, 1997. 'Emerging International Standards'. ISESS 97., Third IEEE International* (Vol. 3000, pp. 128–135). <http://doi.org/10.1109/SESS.1997.595954>
- Jeffery, D. R., Low, G. C., & Barnes, M. (1993). A comparison of function point counting techniques. *IEEE Transactions on Software Engineering*, 19(5), 529-532.
- Jeffery, R., & Stathis, J. (1996). Function point sizing: structure, validity and applicability. *Empirical Software Engineering*, 1(1), 11-30.
- Jenner, M. S. (2001). COSMIC-FFP and UML: Estimation of the Size of a System Specified in UML--Problems of Granularity. En *Proc. the Fourth European Conference on Software Measurement and ICT Control* (pp. 173-184).
- Jenner, S. (2011). Automation of Counting of Functional Size Using COSMIC FFP in UML. In *COSMIC Function Points: Theory and Advanced Practices*, Editors: Dumke, R. and Abran, A., CRC Press. pp. (5.1) 276 – 283
- Ji, C., Yan, S., Ma, X., & Song, G. (2010). Unified Model of Functional Size Measurement. In *E-Product E-Service and E-Entertainment (ICEEE), 2010 International Conference on* (pp. 1-5). IEEE. <http://doi.org/10.1109/ICEEE.2010.5661067>
- Ji, C., & Yan, S. (2012). UML-based full function point measurement method. *Applied Mechanics and Materials*, 220-223, 2879-2885. <http://doi.org/10.4028/www.scientific.net/AMM.220-223.2879>
- Jing, F., Zeng, Y., Liu, H., Cheng, L., & Lu, J. (2014). A metric method for object-oriented systems based on MarkII method. En *2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference, ITAIC 2014* (pp. 564-570). <http://doi.org/10.1109/ITAIC.2014.7065114>
- Jones, E. L. (1995). Automated calculation of function points. *Inquiry*, 4, x6.
- Jones, C. (2008). *Applied software measurement: global analysis of productivity and quality*. McGraw-Hill Education Group.
- Jones, C. (2013). Function points as a universal *software* metric. *ACM SIGSOFT Software Engineering Notes*, 38(4), 1. <http://doi.org/10.1145/2492248.2492268>
- Jørgensen, M. (2007). Forecasting of *Software* Development Work Effort : Evidence on Expert Judgment and Formal Models 2 . *Software Development Effort Estimation. International Journal of Forecasting*, 23(3), 449-462.
- Jorgensen, M., & Shepperd, M. (2007). A Systematic Review of *Software* Development Cost Estimation Studies. *IEEE Transactions on Software Engineering*, 33(1), 33-53. <http://doi.org/10.1109/TSE.2007.256943>
- Jorgensen, M., Boehm, B., & Rifkin, S. (2009). *Software* Development Effort Estimation:

- Formal Models or Expert Judgment? *IEEE Computer Society*, 26(02), 14-19.
<http://doi.org/10.1109/MS.2009.47>
- Juristo, N., & Vegas, S. (2009). Using differences among replications of *software engineering* experiments to gain knowledge. En *Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement* (pp. 356-366).
- Kaya, M., & Demirors, O. (2011). E-Cosmic: A Business Process Model Based Functional Size Estimation Approach. En *2011 37th EUROMICRO Conference on Software Engineering and Advanced Applications* (pp. 404-410).
<http://doi.org/10.1109/SEAA.2011.60>
- Keung, J., Kocaguneli, E., & Menzies, T. (2013). Finding conclusion stability for selecting the best effort predictor in *software* effort estimation. *Automated Software Engineering*, 20(4), 543-567.
- Kitchenham, B., & Kansala, K. (1993). Inter-item correlations among function points. En *Proceedings of 1993 15th International Conference on Software Engineering* (pp. 477-480). <http://doi.org/10.1109/ICSE.1993.346018>
- Kitchenham, B., Pfleeger, S. L., McColl, B., & Eagan, S. (2002). An empirical study of maintenance and development estimation accuracy. *Journal of systems and software*, 64(1), 57-77.
- Kitchenham, B. A. (2002). Principles of Survey Research Part 2: Designing a Survey Barbara A. Kitchenham Department of Computer Science Keele University, Staffordshire, UK. *ACM SIGSOFT Software Engineering Notes*, 27(1), 18.
- Kitchenham, B. A., & Pfleeger, S. L. (2002). Principles of survey research: part 3: constructing a survey instrument. *ACM SIGSOFT Software Engineering Notes*, 27(2), 20-24.
- Kitchenham, B., & Pfleeger, S. L. (2002a). Principles of survey research part 4: questionnaire evaluation. *ACM SIGSOFT Software Engineering Notes*, 27(3), 20-23.
- Kitchenham, B., & Pfleeger, S. L. (2002b). Principles of survey research: part 5: populations and samples. *ACM SIGSOFT Software Engineering Notes*, 27(5), 17-20.
- Kitchenham, B., & Pfleeger, S. L. (2003). Principles of survey research part 6: data analysis. *ACM SIGSOFT Software Engineering Notes*, 28(2), 24-27.
- Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(TR/SE-0401), 28. <http://doi.org/10.1.1.122.3308>
- Kitchenham, B., Dyba, T., & Jorgensen, M. (2004). Evidence-based *software engineering*. In *Proceedings of the 26th international conference on software engineering* (pp. 273-281). IEEE Computer Society.
- Kitchenham, B., Mendes, E., & Travassos, G. (2006). A systematic review of cross-vs. within-company cost estimation studies. En *Proceedings of Empirical ...* (pp. 81-90). Recuperado a partir de http://bcs.org/upload/pdf/ewic_ea06_paper10.pdf
- Kitchenham, B., & Charters, S. (2007). Guidelines for Performing Systematic Literature Reviews in *Software Engineering*. Technical Report EBSE-2007-01 School of Computer Science and Mathematics, Keele University.
- Kitchenham, B., & Mendes, E. (2009). Why comparative effort prediction studies may be invalid. In *Proceedings of the 5th international Conference on Predictor Models in Software Engineering* (p. 4). ACM.
- Kitchenham, B. (2010). What's up with *software* metrics?—A preliminary mapping study.

- Journal of Systems and Software*, 83(1), 37–51. <http://doi.org/10.1016/j.jss.2009.06.041>
- Kitchenham, B., Pretorius, R., Budgen, D., Brereton, O. P., Turner, M., Niazi, M., & Linkman, S. (2010). Systematic literature reviews in *software engineering*—a tertiary study. *Information and Software Technology*, 52(8), 792-805.
- Kitchenham, B. A., Budgen, D., & Pearl Brereton, O. (2011). Using mapping studies as the basis for further research—A participant-observer case study. *Information and Software Technology*, 53(6), 638–651. <http://doi.org/10.1016/j.infsof.2010.12.011>
- Kitchenham, B., Sjöberg, D., Dyba, T., Brereton, P., Budgen, D., Host, M., & Runeson, P. (2013). Trends in the Quality of Human-Intensive *Software Engineering Experiments: A Quasi-Experiment*. *Software Engineering, IEEE Transactions on*, PP(99), 1. <http://doi.org/10.1109/TSE.2012.76>
- Kitchenham, B. A., Budgen, D., & Brereton, P. (2015). Evidence-based *software engineering and systematic reviews* (Vol. 4). CRC Press.
- Kitchenham, B., Madeyski, L., Budgen, D., Keung, J., Brereton, P., Charters, S. & Pohthong, A. (2017). Robust statistical methods for empirical *software engineering*. *Empirical Software Engineering*, 22(2), 579-630.
- Kitchenham, B., Madeyski, L., & Curtin, F. (2017). Corrections to effect size variances for continuous outcomes of cross-over clinical trials. *Statistics in Medicine* <https://doi.org/10.1002/sim.7379>.
- Kuijpers, C. (2014). Automated FPA (eFPA) in SAP environment - Visions and experiences of automated function point analysis. En Proceedings - 2014 Joint Conference of the International Workshop on *Software Measurement*, IWSM 2014 and the International Conference on *Software Process and Product Measurement*, Mensura 2014 (pp. 72-78). <http://doi.org/10.1109/IWSM.Mensura.2014.35>
- Kusumoto, S., Inoue, K., Kasimoto, T., Suzuki, A., Yuura, K., & Tsuda, M. (2000). Function point measurement for object-oriented requirements specification. En *Proceedings - IEEE Computer Society's International Computer Software and Applications Conference* (pp. 543-548). Recuperado a partir de <http://www.scopus.com/inward/record.url?eid=2-s2.0-0034507803&partnerID=tZOtx3y1>
- Kusumoto, S., Imagawa, M., Inoue, K., Morimoto, S., Matsusita, K., & Tsuda, M. (2004). Function point measurement from Java programs. En *Natural Language Processing and Information Systems* (pp. 159-170). <http://doi.org/10.1145/581410.581412>
- Lamma, E., Mello, P., & Riguzzi, F. (2004). A system for measuring function points from an ER-DFD specification. *Computer Journal*, 47(3), 358-372. <http://doi.org/10.1093/comjnl/47.3.358>
- Langdon, W. B., Dolado, J., Sarro, F., & Harman, M. (2016). Exact mean absolute error of baseline predictor, MARP0. *Information and Software Technology*, 73, 16-18.
- Lavazza, L. A., del Bianco, V., & Garavaglia, C. (2008). Model-based Functional Size Measurement. En *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement* (pp. 100–109). <http://doi.org/10.1145/1414004.1414021>
- Lavazza, L., & Del Bianco, V. (2009). A case study in COSMIC functional size measurement: The rice cooker revisited. En *Lecture Notes in Computer Science (including sub-series Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol.

- 5891 LNCS, pp. 101-121). Springer. http://doi.org/10.1007/978-3-642-05415-0_8
- Lavazza, L., & Garavaglia, C. (2009). Using Function Points to measure and estimate real-time and embedded *software*: Experiences and guidelines. En *2009 3rd International Symposium on Empirical Software Engineering and Measurement, ESEM 2009* (pp. 100-110). <http://doi.org/10.1109/ESEM.2009.5316018>
- Lavazza, L., Morasca, S., & Robiolo, G. (2013). Towards a simplified definition of Function Points. *Information and Software Technology*, 55(10), 1796-1809. <http://doi.org/10.1016/j.infsof.2013.04.003>
- Lavazza, L. (2015). Automated function points: Critical evaluation and discussion. En *International Workshop on Emerging Trends in Software Metrics, WETSoM* (Vol. 2015-Augus, pp. 35-43). <http://doi.org/10.1109/WETSoM.2015.13>
- Lavazza, L., & Morasca, S. (2016). An Empirical Evaluation of Two COSMIC Early Estimation Methods. In *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2016 Joint Conference of the International Workshop on* (pp. 65-74). IEEE.
- Lavazza, L., Lenarduzzi, V., & Taibi, D. (2016). Towards Component-Aware Function Point Measurement. In *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2016 Joint Conference of the International Workshop on* (pp. 35-44). IEEE.
- Lavazza, L., & Morasca, S. (2017). On the Evaluation of Effort Estimation Models. In *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering* (pp. 41-50). ACM.
- Lee, H. K. (2015). Automatic Measurement of Function Points from Java Applications. En *The International Conference on Software Engineering, Mobile Computing and Media Informatics (SEMCMi2015)* (p. 59).
- Lehne, A. (1997). Experience report: function points counting of object oriented analysis and design based on the OOram method. En *Proceedings of the Conference on Object-Oriented Programming Systems, Languages, and Applications (OOPSLA'97)*.
- Lent, J. S., & Qu, Y. (2015). On XML Based Automated Function Point Analysis: An Effective Method to Assess Developer Productivity. *Lecture Notes on Software Engineering*, 3(4), 245-250. <http://doi.org/10.7763/LNSE.2015.V3.199>
- Levesque, G., Bevo, V., & Cao, D. T. (2008). Estimating *software* size with UML models. En *Proceedings of the 2008 C3S2E conference on - C3S2E '08* (p. 81). <http://doi.org/10.1145/1370256.1370268>
- Li, Z., Nonaka, M., Kakurai, a., & Azuma, M. (2003). Measuring functional size of interactive *software*: a support system based on XForms-format user interface specifications. En *Third International Conference on Quality Software, 2003. Proceedings.* (pp. 368-373). <http://doi.org/10.1109/QSIC.2003.1319123>
- Linaker, J., Sulaman, S. M., de Mello, R., & Hst, M. (2015). Guidelines for Conducting Surveys in *Software Engineering*.
- Lin, C. J., & Yeh, D. M. (2016). A *Software* Maintenance Project Size Estimation Tool Based On Cosmic Full Function Point. In *Computer Symposium (ICS), 2016 International* (pp. 555-560). IEEE.
- Lind, K., & Heldal, R. (2008). Estimation of real-time *software* component size. *Nordic*

- Journal of Computing*, 14(4), 282-300.
- Lind, K., & Heldal, R. (2009). Estimation of real-time *software* code size using COSMIC FSM. En *Proceedings of the 2009 IEEE International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing, ISORC 2009* (pp. 244-248). <http://doi.org/10.1109/ISORC.2009.15>
- Lind, K., & Heldal, R. (2011). A model-based and automated approach to size estimation of embedded *software* components. En *Model Driven Engineering Languages and Systems* (pp. 334-348). Springer. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-642-24485-8_24
- Lind, K., Heldal, R., Harutyunyan, T., & Heimdahl, T. (2011). CompSize: Automated size estimation of embedded *software* components. En *Proceedings - Joint Conference of the 21st International Workshop on Software Measurement, IWSM 2011 and the 6th International Conference on Software Process and Product Measurement, MENSURA 2011* (pp. 86-95). <http://doi.org/10.1109/IWSM-MENSURA.2011.49>
- Lind, K. ., & Heldal, R. . (2012). Comp size: A model-based and automated approach to size estimation of embedded *software* components. *IEICE Transactions on Information and Systems*, E95-D(9), 2183-2192. <http://doi.org/10.1587/transinf.E95.D.2183>
- Liu, L. . c, Yang, A. ., Tao, Q. ., Zhu, L. ., & Wu, D. (2014). Study of the *software* size estimation model based on UML. En *Conference Proceedings - 2014 International Conference on System Science and Engineering, ICSSE 2014* (pp. 138-142). <http://doi.org/10.1109/ICSSE.2014.6887921>
- Lokan, C. J. (1999). An empirical study of the correlations between function point elements [*software* metrics]. En *Proceedings Sixth International Software Metrics Symposium (Cat. No.PR00403)* (pp. 200-206). <http://doi.org/10.1109/METRIC.1999.809741>
- Lokan, C., & Mendes, E. (2006). Cross-company and single-company effort models using the ISBSG database: a further replicated study. In *Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering* (pp. 75-84). ACM.
- Lorenzoli, D., Mariani, L., & Pezzè, M. (2008). Automatic generation of *software* behavioral models. In *Proceedings of the 30th international conference on Software engineering* (pp. 501-510). ACM.
- Lother, M., & Dumke, R. (2001). Points metrics-comparison and analysis. En *International Workshop on Software Measurement (IWSM 2001), Montr{é}al, Qu{é}bec* (pp. 155-172).
- Madrigal-Sánchez, D., Quesada-López, C. & Jenkins, M. (2018). Towards the automation of a defect detection protocol for functional size measurements. *Proceedings of the XXI Ibero-american conference on software engineering (CibSE-ESELAW)*. Bogota, Colombia, 23-27 April 2018.
- Marín, B., Condori-Fernández, N., & Pastor, O. (2008). Towards a Method for Evaluating the Precision of *Software* Measures (Short Paper). In *2008 The Eighth International Conference on Quality Software* (pp. 305-310). IEEE.
- Marín, B., Giachetti, G., & Pastor, O. (2008). Measurement of Functional Size in Conceptual Models: A Survey of Measurement Procedures Based on COSMIC. *Software Process and Product Measurement SE - 15*, 5338, 170-183. http://doi.org/10.1007/978-3-540-89403-2_15
- Marín, B., Pastor, O., & Giachetti, G. (2008). Automating the measurement of functional

- size of conceptual models in an MDA environment. En *Product-Focused Software Process Improvement* (pp. 215-229). Springer.
- Marín, B., Giachetti, G., & Pastor, O. (2009). Applying a functional size measurement procedure for defect detection in MDD environments. *Communications in Computer and Information Science*, 42, 57-68. http://doi.org/10.1007/978-3-642-04133-4_5
- Marín, B., Pastor, O., & Abran, A. (2010). Towards an accurate functional size measurement procedure for conceptual models in an MDA environment. *Data & Knowledge Engineering*, 69(5), 472-490. <http://doi.org/10.1016/j.datak.2010.01.001>
- Marín, B., Giachetti, G., Pastor, O., & Vos, T. E. (2010). A tool for automatic defect detection in models used in model-driven engineering. In *Quality of Information and Communications Technology (QUATIC), 2010 Seventh International Conference on the* (pp. 242-247). IEEE.
- Marín, B., Giachetti, G., Pastor, O., Vos, T. E., & Abran, A. (2010). Evaluating the usefulness of a functional size measurement procedure to detect defects in MDD models. In *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement* (p. 49). ACM.
- Marín, B. M. (2011) *Functional Size Measurement and Model Verification for Software Model-Driven Developments: A COSMIC-based Approach*. Universidad Politecnica de Valencia (Spain).
- Marín, B., Giachetti, G., Pastor, O., Vos, T. E., & Abran, A. (2013). Using a functional size measurement procedure to evaluate the quality of models in MDD environments. *ACM Transactions on Software Engineering and Methodology*, 22(3). <http://doi.org/10.1145/2491509.2491520>
- Marín, B., & Quinteros, J. (2014). A COSMIC measurement procedure for BPMN diagrams. En *Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE (Vol. 2014-Janua, pp. 408-411)*.
- Matson, J. E., & Mellichamp, J. M. (1993). Object oriented tool for function point analysis. *Expert Systems*, 10(1), 3-14.
- Mayer, R. J., Painter, M. K., & de Witte, P. S. (1994). *IDEF family of methods for concurrent engineering and business re-engineering applications*. Knowledge Based Systems College Station, Tex, USA.
- McGarry, J. (2002). *Practical software measurement: objective information for decision makers*. Addison-Wesley Professional.
- Meli, R. (1997). Early and Extended Function Point: a new method for Function Points estimation. In: *Proceedings of the IFPUG Fall Conference, IFPUG, Scottsdale, Arizona*.
- Meli, R. (2000). Functional and technical software measurement: conflict or integration. En *3rd. European Software Measurement Conference (FESMA'00), Madrid, Spain*.
- Méndez-Porras, A., Quesada-López, C., y Jenkins, M. (2015). Automated Testing of Mobile Applications: A Systematic Map and Review. J. Araujo y cols. (Eds.), *Proceedings of the XVIII ibero-american conference on software engineering (CIbSE)*. Lima, Perú, 22-24 April 2015.
- Menzel, C., & Mayer, R. J. (1998). The IDEF family of languages. En *Handbook on architectures of information systems* (pp. 209-241). Springer.

- Menzies, T., & Shepperd, M. (2012). Special issue on repeatable results in *software engineering prediction*. *Empirical Software Engineering*, 17(1), 1-17.
- Menzies, T., Yang, Y., Mathew, G., Boehm, B., & Hihn, J. (2016). Negative results for *software effort estimation*. *Empirical Software Engineering*, 1-26.
- Minku, L. L., & Yao, X. (2013). Ensembles and locality: Insight on improving *software effort estimation*. *Information and Software Technology*, 55(8), 1512-1528.
- Mittas, N., & Angelis, L. (2013). Ranking and clustering *software cost estimation models* through a multiple comparisons algorithm. *IEEE Transactions on software engineering*, 39(4), 537-551.
- Miyawaki, T., Iijima, J., & Ho, S. (2008). Measuring function points from VDM-SL specifications. En *5th International Conference Service Systems and Service Management - Exploring Service Dynamics with Science and Innovative Technology, ICSSSM'08* (pp. 1-6). <http://doi.org/10.1109/ICSSSM.2008.4598523>
- Moody, D. L. (2001). Dealing with complexity: a practical method for representing large entity relationship models (Doctoral dissertation, University of Melbourne, Department of Information Systems).
- Molokken, K., & Jorgensen, M. (2003). A review of *software surveys on software effort estimation*. En *Empirical Software Engineering, 2003. ISESE 2003. Proceedings. 2003 International Symposium on* (pp. 223-230). <http://doi.org/10.1109/ISESE.2003.1237981>
- Monsalve, C., Abran, A., & April, A. (2011). Measuring *Software Functional Size From Business Process Models*. *International Journal of Software Engineering and Knowledge Engineering*, 21(3), 311-338. <http://doi.org/10.1142/S0218194011005359>
- Mourão, E., Kalinowski, M., Murta, L., Mendes, E., & Wohlin, C. (2017). Investigating the Use of a Hybrid Search Strategy for Systematic Reviews. En *2017 11th International Symposium on Empirical Software Engineering and Measurement, ESEM 2017* (pp.0-0).
- Murillo, J., Quesada-López, C., Castro, C. y Jenkins, M. (2015). *Software Fault Prediction: A Systematic Map and Review*. J. Araujo y cols. (Eds.), *Proceedings of the XVIII ibero-american conference on software engineering (CibSE-ESELAW)*. Lima, Perú, 22-24 April 2015.
- Murillo, J., Quesada-López, C., Castro, C. y Jenkins, M. (2016). An empirical validation of an automated genetic *software effort prediction framework* using the ISBSG Dataset. *Proceedings of the XIX Ibero-American Conference on Software Engineering (CibSE/ESELAW)*. Quito, Ecuador, 27-29 April 2016.
- Murillo-Morera, J., Quesada-López, C., Castro-Herrera, C., & Jenkins, M. (2017). A genetic algorithm based framework for *software effort prediction*. *Journal of Software Engineering Research and Development*, 5(1), 4. <https://rdcu.be/QRwx>
- Nagano, S., & Ajisaka, T. (2003). Functional metrics using COSMIC-FFP for object-oriented real-time systems. En *13th International Workshop on Software Measurement (IWSM)* (pp. 1-7).
- Object Management Group. (2014). *Automated Function Points (AFP) Version 1.0, OMG Document Number: formal/2014-01-03* (Vol. 2014). <http://www.omg.org/spec/AFP>.
- Ochodek, M. (2016a). Functional size approximation based on use-case names. *Information and Software Technology*, 80, 73-88.

- Ochodek, M. (2016b). Approximation of COSMIC functional size of scenario-based requirements in Agile based on syntactic linguistic features—a replication study. In *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2016 Joint Conference of the International Workshop on* (pp. 201-211). IEEE.
- Oliveira, A. L., Braga, P. L., Lima, R. M., & Cornélio, M. L. (2010). GA-based method for feature selection and parameters optimization for machine learning regression applied to *software* effort estimation. *Information and Software Technology*, 52(11), 1155-1166.
- Oriou, A., Bronca, E., Bouzid, B., Guetta, O., & Guillard, K. (2014). Manage the Automotive Embedded *Software* Development Cost & Productivity with the Automation of a Functional Size Measurement Method (COSMIC). En *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2014 Joint Conference of the International Workshop on* (pp. 1-4).
- Ozkan, B., & Demirors, O. (2009). Formalization Studies in Functional Size Measurement: How Do They Help? En *Software Process and Product Measurement* (pp. 197-211). Springer.
- Özkan, B., & Demirörs, O. (2010). Formalization studies in functional size measurement. *Modern Software Engineering Concepts and Practices: Advanced Approaches*, 242-262. Recuperado a partir de <http://www.scopus.com/inward/record.url?eid=2-s2.0-84898558226&partnerID=40&md5=160e69581c7b693c5e234ebbd2e98246>
- Özkan, B., & Demirörs, O. (2010). Formalization Studies in Functional Size Measurement. *Modern Software Engineering Concepts and Practices: Advanced Approaches: Advanced Approaches*, 242.
- Özkan, B., & Demirors, O. (2016). On the Seven Misconceptions about Functional Size Measurement. In *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2016 Joint Conference of the International Workshop on* (pp. 45-52). IEEE.
- Pace, D., Cantone, G., & Calavaro, G. (2005). A tool for counting Function Points of UML *software*. *Smef 2005*, 81.
- Pastor, O., & Abran, A. (2008). Measuring the Functional Size of Conceptual Models in an MDA Environment 1. En *Information Systems* (pp. 33-36).
- Paton, K., & Abran, A. (1995). A formal notation for the rules of function point analysis. *Software Engineering Measurement Research Laboratory, Université du Québec a Montreal (UQAM) Canada*.
- Paton, K., & Lambert, S. (1999). Automatic Function Point Counting Using Static and Dynamic Code Analysis. En *International Workshop on Software Measurement (IWSM)* (p. 6).
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3), 45-77.
- Petersen, K. (2011). Measuring and predicting *software* productivity: A systematic map and review. *Information and Software Technology*, 53(4), 317-343. <http://doi.org/10.1016/j.infsof.2010.12.001>

- Petersen, K., Feldt, R., Mujtaba, S., & Mattsson, M. (2008). Systematic Mapping Studies in Software Engineering. En *Ease* (Vol. 17, pp. 68-77).
- Petersen, K., & Wohlin, C. (2009). Context in industrial software engineering research. En *2009 3rd International Symposium on Empirical Software Engineering and Measurement, ESEM 2009* (pp. 401-404). <http://doi.org/10.1109/ESEM.2009.5316010>
- Petersen, K., & Gencel, C. (2013). Worldviews, research methods, and their relationship to validity in empirical software engineering research. En *Proceedings - Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement, IWSM-MENSURA 2013* (pp. 81-89). <http://doi.org/10.1109/IWSM-Mensura.2013.22>
- Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64, 1-18.
- Pfleeger, S. L., & Kitchenham, B. A. (2001). Principles of survey research: part 1: turning lemons into lemonade. *ACM SIGSOFT Software Engineering Notes*, 26(6), 16-18.
- Poels, G. (2002). A Functional Size Measurement Method for Event-Based Object oriented Enterprise Models. In *International Conference On Enterprise Information Systems* (pp. 667-675).
- Poels, G. (2003a). Definition and validation of a COSMIC-FFP functional size measure for object-oriented systems. In *Proc. 7th Int. ECOOP Workshop Quantitative Approaches OO Software Eng.* Darmstadt.
- Poels, G. (2003b). Functional size measurement of multi-layer object-oriented conceptual models. En *Object-Oriented Information Systems* (pp. 334-345). Springer.
- Poels, G. . b. (2003c). Functional size measurement of layered conceptual models. In *ICEIS 2003 - Proceedings of the 5th International Conference on Enterprise Information Systems* (Vol. 3, pp. 411-416).
- Pow-Sang, J. A., & Imbert, R. (2007). Including the composition relationship among classes to improve Function Points Analysis. *Proceeding VI Jornadas Peruanas de Computación-JPC*, 7.
- Pow-Sang, J. A., Gasco, L., & Nakasone, A. (2009). A Function Point Logic File Identification Technique Using UML Analysis Class Diagrams. In *International Conference on Advanced Software Engineering and Its Applications* (pp. 160-167). Springer Berlin Heidelberg.
- Pow-Sang, J. A., Gasco, L., & Nakasone, A. (2010). Evaluating the Applicability of a Function Point Logic File Identification Technique through Controlled Experiments. *International Journal of Software Engineering and Its Applications*, 4(3).
- Pow-Sang, J. A. ., Villanueva, D. ., Flores, L. ., & Rusu, C. . (2013). A conversion model and a tool to identify Function Point logic files using UML analysis class diagrams. In *Proceedings - Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement, IWSM-MENSURA 2013* (pp. 126-134). <http://doi.org/10.1109/IWSM-Mensura.2013.28>
- Pressman, R. S. (2005). *Software engineering: a practitioner's approach*. Palgrave Macmillan.

- Quesada-López, C., y Jenkins, M. (2014). Function point structure and applicability validation using the ISBSG Dataset: A replicated study. *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM). Poster extended abstract doi: 10.1145/2652524.2652595. Torino, Italia, 15-19 September 2014.*
- Quesada-López, C., Jenkins, M., Travassos G. (2014). Automatización de la medición del tamaño funcional del *software*: Resultados preliminares de un mapeo de literatura. Reporte técnico proyecto No. 834-B5- A18, Universidad de Costa Rica.
- Quesada-López, C., & Jenkins, M. (2015a). An evaluation of functional size measurement methods. En *CIBSE 2015 - XVIII Ibero-American Conference on Software Engineering* (pp. 151-165). Recuperado a partir de <http://www.scopus.com/inward/record.url?eid=2-s2.0-84936105768&partnerID=40&md5=dc4c79ca3a170de1e01ad3b8fc241491>
- Quesada-López, C., & Jenkins, M. (2015b). Applying a verification protocol to evaluate the accuracy of functional size measurement procedures: An empirical approach. En *International Conference on Product-Focused Software Process Improvement. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9459, pp. 243-250). Springer. http://doi.org/10.1007/978-3-319-26844-6_18
- Quesada-López, C., y Jenkins, M. (2015c). An empirical validation of function point structure and applicability: A replication study. J. Araujo y cols. (Eds.), *Proceedings of the XVIII ibero-american conference on software engineering (CibSE-ESELAW). Lima, Perú, 22-24 April 2015.*
- Quesada-López, C., Madrigal, D., & Jenkins, M. (2016). An Empirical Evaluation of Automated Function Points. En *CIBSE 2016 - XIX Ibero-American Conference on Software Engineering* (pp. 151-165).
- Quesada-López, C., & Jenkins, M. (2016). Function Point Structure and Applicability: A Replicated Study. *Journal of Object Technology*, 15(3), 2-1.
- Quesada-López, C., Murillo-Morera, J., Castro-Herrera, C., & Jenkins, M. (2016). Cosmic base functional components in functional size based effort estimation models. In *Central American and Panama Convention (CONCAPAN XXXVI), 2016 IEEE 36th* (pp. 1-6). IEEE.
- Quesada-López, C. & Jenkins, M. (2017a). Un estudio sobre las prácticas de la ingeniería del *software* en Costa Rica: Resultados preliminares. *Proceedings of the XX Ibero-American Conference on Software Engineering (CibSE 2017). Buenos Aires, Argentina, 22-23 May 2017.*
- Quesada-López, C. & Jenkins, M. (2017b). Procedimientos de medición del tamaño funcional: un mapeo sistemático de literatura. In *Proceedings of the XX Ibero-American Conference on Software Engineering (CibSE 2017). Buenos Aires, Argentina, 22-23 May 2017.*
- Quesada-López, C., Jenkins, M., Salas, L. & Gómez, J. (2017a). Towards an automated functional size measurement procedure: An industrial case study. In *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2017 Joint Conference of the International Workshop on* (pp. 139-

- 145). ACM.
- Quesada-López, C., Salas, L., Gómez, J., Jenkins, M. (2017b). FastWorks FPA: Una herramienta para automatizar la medición del tamaño funcional. *Proceedings of XLIII Conferencia Latinoamericana de Informática y 46 Jornadas Argentinas de Informática*. Córdoba, Argentina, 4-8 Setiembre del 2017. IEEE.
- Quesada-López, C. & Jenkins, M. (2018). Factores asociados a prácticas de desarrollo y pruebas de software en Costa Rica: Un estudio exploratorio. *Proceedings of the XXI Ibero-american conference on software engineering (CibSE-ESELAW)*. Bogota, Colombia, 23-27 April 2018.
- Quesada-López, C., Jenkins, M., Travassos G. (2018). Automatización de la medición del tamaño funcional del *software*: Un mapeo sistemático de literatura. Reporte técnico proyecto No. 834-B5- A18, Universidad de Costa Rica.
- Ram, D. J., & Raju, S. (2000). Object oriented design function points. En *Quality Software, 2000. Proceedings. First Asia-Pacific Conference on* (pp. 121-126).
- Rask, R. (1991). *Algorithms for Counting Unadjusted Function Points from Dataflow Diagrams* Raimo Rask. Science. Citeseer.
- Riaz, M., Breaux, T., & Williams, L. (2015). How Have We Evaluated *Software* Pattern Application? A Systematic Mapping Study of Research Design Practices. *Information and Software Technology*, 65, 14-38. <http://doi.org/10.1016/j.infsof.2015.04.002>
- Runeson, P., Borgquist, N., Landin, M., & Bolanowski, W. (2000). An evaluation of functional size methods and a bespoke estimation method for real-time systems. En *Product Focused Software Process Improvement* (pp. 339–352). Springer. Recuperado a partir de http://link.springer.com/chapter/10.1007/978-3-540-45051-1_30
http://pdf.aminer.org/000/552/856/an_evaluation_of_functional_size_methods_and_a_bespoke_estimation.pdf
- Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in *software engineering*. *Empirical software engineering*, 14(2), 131-164.
- Runeson, P., Host, M., Rainer, A., & Regnell, B. (2012). Case study research in *software engineering*: Guidelines and examples. John Wiley & Sons.
- Rusli, N. & Abdullah, N. (2017). UML point for mobile game A measurement method for sizing mobile game design. *Journal of Engineering and Applied Sciences*, 12 (3), pp. 481-487.
- Sag, M. A., & Tarhan, A. (2014). Measuring COSMIC *software* size from functional execution traces of java business applications. En *Proceedings - 2014 Joint Conference of the International Workshop on Software Measurement, IWSM 2014 and the International Conference on Software Process and Product Measurement, Mensura 2014* (pp. 272-281). <http://doi.org/10.1109/IWSM.Mensura.2014.29>
- Salmanouglu, M., Demirors, O., & Türetken, O. (2014). Exploration of a Method for COSMIC Size Estimation from S-BPM. In *S-BPM ONE-Application Studies and Work in Progress* (pp. 55–66). Springer.
- Saxena, V., & Shrivastava, M. (2009). Performance of Function Point Analysis Through UML Modeling. *SIGSOFT Softw. Eng. Notes*, 34(2), 1-4. <http://doi.org/10.1145/1507195.1507214>
- Santillo, L., Conte, M., & Meli, R. (2005). Early & Quick function point: sizing more with

- less. In *Software Metrics*, 2005. 11th IEEE International Symposium (pp. 41-41). IEEE.
- Santillo, L. (2012). Easy Function Points--'Smart' Approximation Technique for the IFPUG and COSMIC Methods. In *Software Measurement and the 2012 Seventh International Conference on Software Process and Product Measurement (IWSM-MENSURA)*, 2012 Joint Conference of the 22nd International Workshop on (pp. 137-142). IEEE.
- SEI. (2010). Capability Maturity Model Integration (CMMI) for Development. Version 1.3. Carnegie Mellon, USA. Recuperado a partir de https://resources.sei.cmu.edu/asset_files/TechnicalReport/2010_005_001_15287.pdf
- Sellami, A., & Ben-Abdallah, H. (2009). Functional size of use case diagrams: A fine-grain measurement. En *4th International Conference on Software Engineering Advances, ICSEA 2009, Includes SEDES 2009: Simposio para Estudantes de Doutorado em Engenharia de Software* (pp. 282-288). <http://doi.org/10.1109/ICSEA.2009.96>
- Sellami, A. ., Haoues, M. ., & Ben-Abdallah, H. . (2013a). Automated COSMIC-Based Analysis and Consistency Verification of UML Activity and Component Diagrams. *Communications in Computer and Information Science*, 417 CCIS, 48-63. http://doi.org/10.1007/978-3-642-54092-9_4
- Sellami, A., Haoues, M., & Ben-Abdallah, H. (2013b). Analyzing UML activity and component diagrams: An approach based on cosmic functional size measurement. En *ENASE 2013 - Proceedings of the 8th International Conference on Evaluation of Novel Approaches to Software Engineering* (pp. 36-44).
- Sellami, A., Hakim, H., Abran, A., & Ben-Abdallah, H. (2015). A measurement method for sizing the structure of UML sequence diagrams. *Information and Software Technology*, 59, 222-232. <http://doi.org/10.1016/j.infsof.2014.11.002>
- Seo, Y., Bae, D. & Jeffery, R. (2013). AREION: *Software effort estimation based on multiple regressions with adaptive recursive data partitioning*. *Information and Software technology*, 55(10), 1710-1725.
- Shepperd, M. (2007). *Software project economics: a roadmap*. In *2007 Future of Software Engineering* (pp. 304-315). IEEE Computer Society.
- Shepperd, M., & MacDonell, S. (2012). Evaluating prediction systems in *software project estimation*. *Information and Software Technology*, 54(8), 820-827.
- Shepperd, M., Ajenka, N., & Counsell, S. (2018). The role and value of replication in empirical *software engineering results*. *Information and Software Technology*.
- Shull, F. J., Carver, J. C., Vegas, S., & Juristo, N. (2008). The role of replications in Empirical *Software Engineering*. *Empirical Software Engineering*, 13(2), 211-218. <http://doi.org/10.1007/s10664-008-9060-1>
- Silva, A., Pinheiro, P., & Albuquerque, A. (2016). A Brief Analysis of Reported Problems in the Use of Function Points. En *Software Engineering Perspectives and Application in Intelligent Systems* (pp. 117-126). Springer.
- Sjoberg, D., Dyba, T. & Jorgensen, M. (2007). The future of empirical methods in *software engineering research*. In *Future of Software Engineering, 2007. FOSE'07* (pp. 358-378). IEEE.
- Sneed, H. M. (2001). Extraction of Function-Points from source-code. En *New Approaches in Software Measurement* (pp. 135-146). Springer.
- Song, L., Minku, L. L., & Yao, X. (2013). The impact of parameter tuning on *software effort*

- estimation using learning machines. In *Proceedings of the 9th international conference on predictive models in software engineering* (p. 9). ACM.
- Soubra, H., Abran, A., Stern, S., & Ramdane-Cherif, A. (2011). Design of a Functional Size Measurement Procedure for Real-Time Embedded Software Requirements Expressed using the Simulink Model. En *Software Measurement, 2011 Joint Conference of the 21st Int'l Workshop on and 6th Int'l Conference on Software Process and Product Measurement (IWSM-MENSURA)* (pp. 76-85).
- Soubra, H., & Chaaban, K. (2012). Functional Size Measurement of Electronic Control Units Software Designed Following the AUTOSAR Standard: A Measurement Guideline Based on the COSMIC ISO 19761 Standard. In *Software Measurement and the 2012 Seventh International Conference on Software Process and Product Measurement (IWSM-MENSURA)*, 2012 Joint Conference of the 22nd International Workshop on (pp. 78-84).
- Soubra, H., Abran, A., & Ramdane-Cherif, A. (2012). A Refined Functional Size Measurement Procedure for Real-Time Embedded Software Requirements Expressed Using the Simulink Model. En *Software Measurement and the 2012 Seventh International Conference on Software Process and Product Measurement (IWSM-MENSURA)*, 2012 Joint Conference of the 22nd International Workshop on (pp. 70-77).
- Soubra, H. (2013). Fast functional size measurement with synchronous languages: An approach based on LUSTRE and on the COSMIC ISO 19761 standard. En *Proceedings - Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement, IWSM-MENSURA 2013* (pp. 3-8). <http://doi.org/10.1109/IWSM-Mensura.2013.11>
- Soubra, H., Abran, A., & Ramdane-Cherif, A. (2014). Verifying the accuracy of automation tools for the measurement of software with COSMIC - ISO 19761 including an AUTOSAR-based example and a case study. En *Proceedings - 2014 Joint Conference of the International Workshop on Software Measurement, IWSM 2014 and the International Conference on Software Process and Product Measurement, Mensura 2014* (pp. 23-31). <http://doi.org/10.1109/IWSM.Mensura.2014.26>
- Soubra, H., Abran, A., & Sehit, M. (2015). Functional size measurement for processor load estimation in AUTOSAR. In *International Workshop on Software Measurement* (pp. 114-129). Springer International Publishing.
- Soubra, H. & Abran, A. (2017). Functional Size Measurement for the Internet of Things (IoT): An example using COSMIC and the Arduino open-source platform. In *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA)*, 2017 Joint Conference of the International Workshop on (pp. 123-129). ACM.
- Stambollian, A., & Abran, A. (2006). Survey of Automation Tools Supporting COSMIC-FFP - ISO 19761. *IWSM Mensura - MetriKon*.
- St-Pierre, D., Maya, M., Abran, A., Desharnais, J.-M., & Bourque, P. (1997). Full function points: Counting practices manual. *Software Engineering Management Research Laboratory and Software Engineering Laboratory in Applied Metrics*.
- Symons, C. (2014). Lies, damned lies and software metrics. En *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-*

- MENSURA), 2014 Joint Conference of the International Workshop on (pp. 174-175).
- Tan, H. B. K., Zhao, Y., & Zhang, H. (2009). Conceptual data model-based software size estimation for information systems. *ACM Transactions on Software Engineering and Methodology*, 19(2), 1-37. <http://doi.org/10.1145/1571629.1571630>
- Tavares, H., Carvalho, A., & Castro, J. (2002). Function points measurement from requirement specification. En *Proc. 5th Workshop Engineering Requirements, Valency, Spain* (pp. 278-298).
- Tarhan, A., Özkan, B., & İçöz, G. C. (2016). A Proposal on Requirements for COSMIC FSM Automation from Source Code. In *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2016 Joint Conference of the International Workshop on* (pp. 195-200). IEEE.
- The Standish Group. (2015). *CHAOS MANIFESTO 2013, Think Big, Act Small. CHAOS MANIFESTO 2015*.
- Trudel, S., Desharnais, J. M., & Cloutier, J. (2016). Functional size measurement patterns: A proposed approach. In *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2016 Joint Conference of the International Workshop on* (pp. 23-34). IEEE.
- Uemura, T., Kusumoto, S., & Inoue, K. (1999). Function point measurement tool for UML design specification. En *Software Metrics Symposium, 1999. Proceedings. Sixth International* (pp. 62-69). <http://doi.org/10.1109/METRIC.1999.809727>
- Uemura, T., Kusumoto, S., & Inoue, K. (2001). Function-point analysis using design specifications based on the unified modelling language. *Journal of Software Maintenance*, 13(4), 223-243.
- Ungan, E., Trudel, S. & Poulin, L. (2017). Using FSM Patterns to Size Security Non Functional Requirements with COSMIC. In *Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2017 Joint Conference of the International Workshop on* (pp. 65-77). ACM.
- Vianna Ferreira, P. J. A., & Barros, M. D. O. (2011). Traceability between function point and source code. In *Proceedings of the 6th International Workshop on Traceability in Emerging Forms of Software Engineering* (pp. 10-16). ACM.
- Villalobos-Arias, L., Quesada-López, C., Martinez, A. & Jenkins, M. (2018) A tertiary study on model-based testing areas, tools and challenges: Preliminary results. *Proceedings of the XXI Ibero-american conference on software engineering (CibSE-ESELAW)*. Bogota, Colombia, 23-27 April 2018.
- Vogelezang, F. (2005). Early estimating using COSMIC-FFP. In *SMEF Software Measurement European Forum 2005* (pp. 227-234).
- Vogelezang, F., Prins, T., & BV, S. N. (2007). Approximate size measurement with the COSMIC method Factors of influence. In *Software Measurement European Forum SMEF 2007* (pp. 167-178).
- Vogelezang, F., Symons, C., Lesterhuis, A., Meli, R., & Daneva, M. (2013). Approximate COSMIC Functional Size--Guideline for Approximate COSMIC Functional Size Measurement. In *Software Measurement and the 2013 Eighth International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2013 Joint Conference of the 23rd International Workshop on* (pp. 27-32). IEEE.

- Walkinshaw, N., Taylor, R., & Derrick, J. (2016). Inferring extended finite state machine models from *software* executions. *Empirical Software Engineering*, 21(3), 811-853.
- Whigham PA, Owen CA, Macdonell SG (2015) A baseline model for *software* effort estimation. *ACM Trans Softw Eng Methodol* 24(3):20:1–20:11
- Wentzlaff, I. (2017). Establishing a Requirements Baseline by Functional Size Measurement Patterns. In REFSQ Workshops.
- Wieringa, R. J. (2014). *Design Science Methodology for Information Systems and Software Engineering. the 32nd ACM/IEEE International Conference*. Springer.
<http://doi.org/10.1007/978-3-662-43839-8>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in Software Engineering Hardcover*. Springer Science & Business Media.
[http://doi.org/10.1016/S0065-2458\(08\)60338-1](http://doi.org/10.1016/S0065-2458(08)60338-1)
- Wohlin, C. (2014). Guidelines for Snowballing in Systematic Literature Studies and a Replication in *Software Engineering*. En *18th International Conference on Evaluation and Assessment in Software Engineering (EASE 2014)* (pp. 1-10).
<http://doi.org/10.1145/2601248.2601268>
- Yeh, D., Chen, Y.-H., & Yang, C.-Y. (2015). Applying Reverse Engineering Techniques to Verify the Estimation of *Software* Code Size using COSMIC Full Function Point. *Recent Researches in Applied Informatics*.
- Yilmaz, G., Tunalilar, S., & Demirors, O. (2013). Towards the Development of a Defect Detection Tool for COSMIC Functional Size Measurement. In *Software Measurement and the 2013 Eighth International Conference on Software Process and Product Measurement (IWSM-MENSURA), 2013 Joint Conference of the 23rd International Workshop on* (pp. 9-16). IEEE.
- Yin, R. (2009). *Case Study Research: Design and Methods*. 4th ed. Sage Publications.
- Zhang, X., & Ren, A. (2012). Functional size measurement using use case: From the viewpoint of flow of event. En *Lecture Notes in Electrical Engineering* (Vol. 125 LNEE, pp. 205-210). Springer. http://doi.org/10.1007/978-3-642-25789-6_30
- Zickert, F., & Beck, R. (2013). A mapping model for assessing project effort from requirements. *Information Systems and e-Business Management*, 11(3), 377-401.
<http://doi.org/10.1007/s10257-012-0195-7>
- Zivkovic, A., Heričko, M., & Rozman, I. (2003). Function points in object oriented analysis and design. En *IASTED International Multi-Conference on Applied Informatics* (Vol. 21, pp. 986-990).
- Živkovič, A., Heričko, M., Brumen, B., Beloglavec, S., & Rozman, I. (2005). The impact of details in the class diagram on *software* size estimation. *Informatica*, 16(2), 295-312. Recuperado a partir de <http://www.scopus.com/inward/record.url?eid=2-s2.0-21244473252&partnerID=tZOtx3y1>
- Živkovič, A., Rozman, I., & Heričko, M. (2005). Automated *software* size estimation based on function points using UML models. *Information and Software Technology*, 47(13), 881-890. Recuperado a partir de <http://www.sciencedirect.com/science/article/pii/S0950584905000303>

ANEXOS

Anexo 1 Diseño del estudio de mapeo y revisión sistemática de literatura.

El Anexo 1 detalla el diseño del estudio de mapeo y revisión sistemática de literatura para la identificación y análisis de los procedimientos de medición del tamaño funcional. El protocolo base para este mapeo y revisión fue desarrollado durante el año 2014 (Quesada-López, Jenkins & Travassos, 2014) y actualizado mediante corridas piloto para este estudio (Quesada-López & Jenkins, 2017b; Quesada-López, Jenkins & Travassos, 2018).

Diseño del estudio

El diseño del estudio describe los pasos necesarios para la ejecución de cada una de las etapas del proceso de mapeo y revisión sistemática las cuales comprenden: la definición del objetivo y las preguntas de investigación, la definición de la estrategia de búsqueda, el proceso de selección de estudios primarios y sus criterios de inclusión y exclusión, la evaluación de la calidad de los artículos seleccionados, la extracción de los datos y el proceso de clasificación y análisis de la evidencia.

Preguntas de investigación

Las preguntas de investigación de un mapeo y de una revisión sistemática tienen diferentes objetivos (Petersen, Feldt, Mujtaba & Mattsson, 2008). Las preguntas del mapeo estructuran el conocimiento actual en el área y muestran las distribuciones y tendencias de los artículos relevantes. Por su parte, las preguntas de investigación de la revisión analizan y sintetizan la evidencia existente y proveen recomendaciones basadas en la evidencia. El estudio sobre los procedimientos de medición del tamaño funcional puede permitir establecer la factibilidad de las propuestas de medición en distintos contextos de operación y sus limitaciones. Las preguntas de mapeo se responden a través de la clasificación de los

estudios primarios seleccionados, mientras que las preguntas de revisión se responden mediante el análisis y síntesis de la evidencia. Se establecen las siguientes preguntas de investigación para guiar el análisis durante la conducción del estudio:

- RQ1. *¿Cuál es el estado actual de la investigación en el área y las características de los procedimientos de medición del tamaño funcional del software que se encuentran en la literatura científica?*
- RQ2. *¿Cómo han sido evaluados los procedimientos de medición del tamaño funcional?*
- RQ3. *¿Cuáles son los desafíos y oportunidades de los procedimientos existentes para la medición del tamaño funcional?*

Estrategia de búsqueda

La estrategia de búsqueda conlleva realizar un conjunto de actividades que permiten seleccionar los artículos relevantes para responder a las preguntas de investigación planteadas. Esta estrategia se construye a partir de las preguntas del mapeo y de la revisión. El objetivo en esta etapa del proceso de revisión es establecer los lineamientos para identificar todos los estudios primarios relacionados con estas preguntas de investigación.

El protocolo base para este mapeo y revisión fue desarrollado durante el año 2014 (Quesada-López, Jenkins & Travassos, 2014) y actualizado mediante corridas piloto para este estudio. Se realiza una búsqueda exploratoria para identificar estudios relevantes que son usados como artículos de control. Esta búsqueda se basa en las preguntas de investigación, en términos utilizados en los estudios secundarios relacionados con el área (Jorgensen & Shepperd, 2007; Marín, Giachetti & Pastor, 2008; Ozkan & Demirors, 2010; Barkallah, Gherbi & Abran, 2011; Bajwa, Gencel & Abrahamsson, 2014) y en experiencias y lecciones aprendidas durante la conducción de distintos estudios secundarios (Quesada-López, Jenkins & Travassos, 2014; Méndez-Porras, Quesada-López & Jenkins, 2015; Murillo, Quesada-López, Castro & Jenkins, 2015; Quesada-López & Jenkins, 2017b; Villalobos-Arias, Quesada-López, Martínez & Jenkins, 2018). La búsqueda exploratoria utiliza una combinación de las palabras “*software*”, “*tamaño funcional*”, “*procedimiento*”, “*herramienta*” y “*automatización*”.

A partir de este proceso se seleccionan 11 artículos de control y se determinan los términos de búsqueda para el estudio. La cadena de búsqueda se construye a partir de la extracción de términos clave del título y del resumen del conjunto de artículos de control y se basa en las preguntas de investigación de este estudio. Los artículos de control a partir de los cuales se construye la cadena de búsqueda son: (Uemura, Kusumoto & Inoue, 2001; Abrahao, Poels & Pastor, 2004a; Abrahao, Poels & Pastor, 2004; Zivkovic, Rozman & Hericko, 2005; Hericko, Rozman & Zivkovic, 2006; Abrahao, Mendes, Gomez & Insfran, 2007; Marín, Pastor & Giachetti, 2008; Edagawa, Akaike, Higo, Kusumoto, Hanabusa & Shibamoto, 2011; Soubra, Abran, Stern & Ramdane-Cherifet, 2011; Aysolmaz & Demirors, 2014; Soubra, Abran & Ramdane-Cherif, 2014).

Se desarrolla el modelo PICO (Población, Intervención, Comparación, Salidas, por sus siglas en inglés) sugerido por (Budgen & Brereton, 2006). La cadena de búsqueda utilizada es la siguiente: *Población* (aplicaciones de software y proyectos de software. Ejemplo: “sistema”, “software”, “aplicación”), *Intervención* (medición del tamaño funcional. Ejemplo: “puntos de función”, “tamaño funcional”, “medición”), *Comparación* (no fue considerada) y *Salidas* (procedimientos, reglas de mapeo, formalización o automatización de la medición del tamaño funcional. Ejemplo: “automatización”, “reglas mapeo”, “sistemático”, “herramienta”, “representación formal”).

Los términos de búsqueda fueron identificados según la estrategia utilizada por (Kitchenham, Mendes & Travassos, 2006) que propone los siguientes pasos: (1) derivar los términos principales de las preguntas de investigación e identificar los componentes del modelo PICO, (2) identificar los sinónimos y los descriptores alternativos con base en el conjunto de artículos de control, (3) incluir descriptores alternativos a partir del conocimiento experto de los investigadores, (4) construir la cadena de búsqueda usando el operador booleano OR para los términos de un mismo componente del modelo PICO y el operador booleano AND para unir los diferentes grupos del modelo PICO, (5) pilotear la cadena de búsqueda mediante un conjunto de pruebas piloto para reducir el ruido. Para este estudio, también se verificó que los artículos de control de la lista de referencia fueran retornados por las bases de datos automatizadas de búsqueda.

Con la cadena refinada se realizan las búsquedas en las bases de datos digitales las cuales fueron seleccionadas de acuerdo con las recomendaciones de estudios relacionados en el área de interés e incluyen *SCOPUS*²², *IEEE Xplore*²³, *ISI Web of Science*²⁴ y *Engineering Village-Inspeccompendex*²⁵. Estas bases de datos han sido utilizadas en estudios secundarios previos por su estabilidad e interoperabilidad con múltiples bases de datos académicas. Por ejemplo, *SCOPUS* indexa artículos de *ACM*, *IEEE* y *Elsevier*. Estas bases de datos garantizan un proceso sistemático y repetible y permiten el uso de opciones para la búsqueda avanzada de artículos científicos mediante cadenas de búsqueda complejas.

La cadena de búsqueda base que fue utilizada en las bases de datos automatizadas es la siguiente:

- ("software*" OR "web*" OR "system*" OR "application*") AND ("function* point*" OR "functional size*" OR "function* measurement") AND ("automat*" OR "systematic" OR "procedure*" OR "tool*" OR "mapping*" OR "rule*" OR "formal * representation")

La búsqueda automatizada se realiza en agosto del 2017 y los estudios se analizan durante el segundo semestre del 2017. La búsqueda considera estudios primarios recuperados hasta el año 2017. Los ajustes particulares en la cadena de búsqueda para cada una de las bases de datos son los siguientes: *SCOPUS* (TITLE-ABS-KEY ((())) AND (LIMIT-TO (SUBJAREA, "COMP") OR LIMIT-TO (SUBJAREA, "ENGI"))), *Web of Science* (TS=()) y *Engineering Village-Inspeccompendex* ((() WN KY)).

Inclusión y exclusión de estudios

Los criterios de inclusión y de exclusión (I/E) determinan los estudios primarios que son relevantes para esta investigación. El proceso de inclusión y exclusión evalúa cada uno

²² <http://www.scopus.com/>

²³ <http://ieeexplore.ieee.org/>

²⁴ <http://www.isiknowledge.com/>

²⁵ <http://www.engineeringvillage.com/>

de los artículos identificados en las bases de datos con base en su título, palabras clave y resumen. En los casos en que el título, las palabras clave y el resumen no proporcionan información suficiente para tomar la decisión de inclusión y de exclusión, esta se hace realizando la lectura completa del artículo.

En nuestro estudio excluimos publicaciones que no cumplen con la fórmula (E1 OR E2 OR E3 OR E4 OR E5) donde:

- E1. Estudios secundarios o terciarios.
- E2. Estudios en idiomas diferentes al inglés con texto completo.
- E3. Estudios relacionados con la estimación del tamaño funcional a partir de métodos de conversión. Por ejemplo, a partir de las líneas de código fuente.
- E4. Estudios que traten únicamente el tema de modelos de estimación basados en el tamaño funcional del software sin considerar el proceso de medición.
- E5. Literatura gris que no esté referenciada por los estudios seleccionados para el proceso de bola de nieve.

Para esta revisión, no son de interés los artículos que estudian la convertibilidad a partir de las líneas de código fuente, pues el tema de *backfiring* no es de interés para esta investigación.

Los estudios que no fueron excluidos y que cumplen con la fórmula (I1 AND I2 AND (I3 OR I4)) son incluidos de acuerdo con:

- I1. Estudios relacionados con el área de medición del tamaño funcional del software en el campo de la ingeniería del software.
- I2. Estudios primarios relacionados con un procedimiento de medición del tamaño funcional basado en los métodos de tamaño funcional estandarizados IFPUG FPA, COSMIC FFP, MARK II, FiSMA y/o NESMA.
- I3. Estudios que proveen alguna descripción del procedimiento de medición del tamaño funcional.

- I4. Estudios primarios que proveen algún detalle de la evaluación de los procedimientos de medición del tamaño funcional.

Para esta revisión, son de interés los artículos que proponen y evalúan los procedimientos de medición del tamaño funcional basándose en métodos de medición estandarizados: IFPUG FPA, COSMIC FFP, MARK II, FiSMA y NESMA. El criterio inclusivo de nuestro estudio permite seleccionar procedimientos de medición que realizan modificaciones a los métodos estandarizados y procedimientos de estimación de tamaño funcional. Los criterios de inclusión y de exclusión fueron definidos, evaluados y ajustados durante varias pruebas piloto para garantizar la reducción del ruido en la selección de los artículos relevantes.

Procedimiento de selección de estudios primarios

Con el resultado obtenido a partir de la búsqueda de los artículos en las bases de datos, se lleva a cabo el proceso de selección. Los criterios de inclusión y de exclusión para la identificación de artículos relevantes son aplicados después de que el investigador principal realiza la búsqueda en las bases de datos y elimina los duplicados.

El proceso se realiza de la siguiente manera: (1) el investigador principal realiza la lectura e incluye o excluye los artículos basado en el título, resumen, palabras clave o lectura completa, (2) un segundo investigador, que actúa como consultor, evalúa una muestra aleatoria de los artículos y valida los resultados. Si se presentan inconsistencias en la selección de estudios primarios, el proceso se ejecuta individualmente y de forma completa por cada investigador de acuerdo a lo recomendado por (Petersen, Vakkalanka & Kuzniarz, 2015). Para este proceso se definen criterios específicos de inclusión y exclusión. En el caso de este mapeo y revisión, los resultados del proceso de auditoría confirman la consistencia en el proceso de inclusión de los artículos. En este proceso se aplica una validación *test-retest* (Kitchenham, 2010) para mejorar los resultados.

Procedimiento de bola de nieve

Para mejorar el alcance de nuestra búsqueda, aplicamos una estrategia de búsqueda híbrida que realiza un proceso de búsqueda de bola de nieve (*backward and forward snowball sampling*) a partir de las referencias de los artículos relevantes seleccionados (Wohlin, 2014; Petersen, Vakkalanka & Kuzniarz, 2015; Badampudi, Wohlin & Petersen, 2015; Mourão, Kalinowski, Murta, Mendes & Wohlin, 2017). Los criterios de inclusión y de exclusión son aplicados para la selección de estos artículos adicionales. El proceso de *snowballing forward* se apoya en la herramienta *Google Scholar*²⁶.

El proceso de bola de nieve mejora la confiabilidad de la búsqueda y la identificación de artículos relevantes. Por ejemplo, en los procedimientos de medición del tamaño funcional, un proyecto de investigación podría ser dividido en varios artículos: uno para la propuesta y varios artículos para la validación empírica. El proceso de “bola de nieve” identifica los artículos relacionados. Por otra parte, favorece la identificación de las propuestas que no son definidas como procedimientos de medición.

Evaluación de calidad

La evaluación de la calidad refleja el nivel de detalle ofrecido para los aspectos más relevantes del análisis. El objetivo de los criterios de calidad fue evaluar la completitud del procedimiento de medición y la rigurosidad de su validación empírica. La evaluación de calidad tiene por objetivo identificar y destacar los artículos que son más relevantes para responder las preguntas de investigación. Los artículos seleccionados son evaluados de acuerdo al siguiente conjunto de criterios de calidad:

- Q1. ¿El artículo menciona claramente el método de medición utilizado para el procedimiento de medición? (1 punto).
- Q2. ¿El artículo describe claramente o referencia los artefactos de entrada utilizados para el procedimiento de medición? (3 puntos) [3-Lista y describe los artefactos.

²⁶ <https://scholar.google.com/>

- 2-Lista y menciona los artefactos. 1-Menciona tipos de artefactos de manera general].
- Q3. ¿El artículo explica o referencia las reglas de mapeo entre los artefactos de entrada y los componentes del método de medición del tamaño funcional? (3 puntos) [3-Explica las reglas de mapeo entre los artefactos de entrada y los componentes del método y su aplicación. 2-Menciona las reglas, pero no describe cómo aplicarlas. 1-Resume un conjunto de reglas generales para el mapeo].
- Q4. ¿El artículo menciona el desarrollo de una herramienta de automatización para el procedimiento de medición del tamaño funcional? (1 punto).
- Q5. ¿El artículo describe el proceso experimental para la validación del procedimiento de medición del tamaño funcional de manera completa? (3 puntos) [3-Describe la pregunta de investigación o hipótesis, el procedimiento de recolección y análisis de datos y las características de los objetos y participantes detalladamente. 2-Menciona brevemente cada una de las anteriores. 1-Resume el proceso de manera general].
- Q6. ¿El artículo describe las amenazas a la validez y las acciones para minimizarlas? (3 puntos) [3-Las clasifica y las describe detalladamente. 2-Las clasifica y menciona de manera resumida. 1-Las describe en términos generales].
- Q7. ¿El artículo presenta los resultados de la validación empírica del procedimiento de medición del tamaño funcional? (3 puntos) [3-Menciona las métricas y aplica un procedimiento de validación. 2-Menciona las métricas y los resultados. 1-Solo menciona resultados generales].

El puntaje se asigna de acuerdo a la escala de 0-3 (0 = “No en lo absoluto”; 1 = Algo; 2 = Parcialmente; 3 = Totalmente), tal como se recomienda en (Kitchenham, Pretorius, Budgen, Brereton, Turner, Niazi & Linkman, 2010). La evaluación de la calidad es aplicada en su totalidad por un investigador y validada por un segundo investigador (consultor) sobre una muestra aleatoria de los artículos.

Los valores de la evaluación varían entre 0 y 17. Un puntaje de calidad mayor indica que el reporte provee un mayor nivel de detalle y no debe asociarse con la calidad del estudio, ni la calidad del foro de publicación.

Extracción de datos

Para los artículos seleccionados se extrae la información relevante para el análisis y los componentes del formulario de extracción se encuentran asociados a cada pregunta de investigación (RQ). La extracción es realizada por un investigador, y luego validada por un segundo investigador (consultor) a partir de una muestra aleatoria de los artículos extraídos. Siguiendo las recomendaciones de (Petersen, Vakkalanka & Kuzniarz, 2015), el consultor revisa los datos extraídos y a partir de la lectura del artículo correspondiente, determina si los datos extraídos son correctos. Durante el proceso de extracción, los datos son tabulados en tablas previamente definidas y luego unificados en un único archivo que consolida la totalidad de los datos. El formulario de extracción permite para cada artículo extraer la información relevante que responde cada una de las preguntas del mapeo y de la revisión.

El formulario de extracción cuenta con las siguientes secciones y detalla los elementos para la recolección de datos:

- *Caracterización (RQ1)*: extrae la información para el mapeo de los estudios y su caracterización que incluye el identificador, título, año, autores, foro de publicación, enfoque, método, artefacto de entrada, etapa del ciclo de vida, tipo de artículo, procedimiento de medición, reglas de mapeo, reglas de asignación numérica, marco de trabajo utilizado, las herramientas que lo soportan y nivel de automatización.
- *Evaluación (RQ2)*: extrae la información sobre la evaluación de los procedimientos que incluye los objetivos, tipos de preguntas de investigación e hipótesis, el método de investigación empírica, tipo de evaluación, participantes, tamaño de la muestra, instru-

mentación, los constructos, métricas y mediciones, factores de contexto y dominio funcional, las amenazas a la validez interna y externa, constructo, conclusión y tipo de replicación.

- *Desafíos y oportunidades* (RQ3): extrae la información de los desafíos, oportunidades, líneas de investigación futura y recomendaciones.
- *Calidad*: detalla la evaluación de la calidad de los artículos.

Análisis y clasificación

Para el análisis y síntesis de la información se aplican las siguientes estrategias. La información de cada elemento extraído se tabula y grafica de acuerdo a cada una de las preguntas del mapeo. Asimismo, se realiza un análisis de los datos extraídos para responder a cada una de las preguntas de la investigación.

Todos los artículos se organizan de acuerdo a las diferentes categorías determinadas, las cuales se validan con la literatura y las clasificaciones existentes. Por ejemplo, en el caso de las amenazas a la validez se utilizan las categorías específicas descritas en (Petersen & Gencel, 2013). Los artículos relacionados a una sola propuesta son contabilizados por separado.

Las categorías de la clasificación de los artículos primarios se determinan utilizando la técnica para generar esquemas de clasificación llamada *keywording* (Petersen, Feldt, Mujtaba & Mattsson, 2008). El *keywording* se realiza a partir de los resúmenes de los estudios primarios, lo que permite que las categorías establecidas surjan a partir de los datos y sean justificadas por los estudios previos. Esta técnica complementa las categorías que se determinan a partir de la literatura seleccionada para este estudio secundario. A través del análisis de los estudios primarios, las categorías son refinadas y ajustadas hasta lograr la categorización completa para los estudios seleccionados, tal como se describe en (Riaz, Breaux & Williams, 2015). Las categorías establecidas se utilizan para clasificar los datos extraídos durante el análisis, son determinadas por un investigador y discutidas y aprobadas conjun-

tamente con el segundo investigador. Estas se complementan con las distintas clasificaciones definidas en los estudios relacionados (Lothar & Dumke, 2001; Abrahão, 2004; Jorgensen & Shepperd, 2007; Marín, Giachetti & Pastor, 2008; Petersen & Wohlin, 2009; Ozkan & Demirors, 2010; Barkallah, Gherbi & Abran, 2011; Petersen, 2011; Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012; Bajwa, Gencel & Abrahamsson, 2014; Riaz, Breaux & Williams, 2015; Ozkan & Demirors, 2016).

Para sintetizar la evidencia recolectada se realiza un análisis de contenidos y se resumen los hallazgos que posibilitan contestar a las preguntas de investigación. Estos enfoques para el análisis de información cualitativa y cuantitativa se describen en (Dixon-woods, Agarwal, Jones, Young & Sutton, 2016). El análisis de contenidos facilita categorizar los datos y analizar las frecuencias de los temas recurrentes, lo cual permite convertir la evidencia cualitativa en datos cuantitativos; para simplificar el análisis de los resultados se realiza la tabulación de resultados. Por su parte, el resumen de contenidos posibilita recapitular los hallazgos y describirlos en conjunto.

En esta investigación también se aplican los principios de la *Grounded Theory* que permiten analizar los datos cualitativos para generar teorías. Este proceso permite un análisis de datos sistemático para minimizar el sesgo que produce la interpretación de los investigadores durante el análisis de la información (Easterbrook, Singer, Storey & Damian, 2008). Es importante recalcar que la técnica de síntesis de datos meta-análisis no se utiliza en este estudio debido a la heterogeneidad de los estudios primarios (Wohlin, Runeson, Höst, Ohlsson, Regnell & Wesslén, 2012).

Esquema de clasificación

El esquema de clasificación es la estructura de los aspectos de interés en el campo de estudio el cual los grupos de clasificación que permiten categorizar y cuantificar los resultados. Lothar & Dumke (2001) proponen un conjunto de criterios para delimitar el contexto donde los métodos de medición son aplicables. Los criterios necesarios para definir una propuesta de medición son: el método de medición utilizado, el dominio funcional,

el tipo de medición, los artefactos de entrada, la etapa del ciclo de desarrollo donde se realiza la medición y el instrumento de medición. Asimismo, para realizar la validación de las mediciones se deben considerar criterios como: el contexto de la evaluación empírica, la estrategia empírica utilizada, y las métricas de validación evaluadas. La Figura 4-1 detalla el esquema de clasificación que considera las categorías utilizadas para este estudio.

Amenazas a la validez del mapeo y de la revisión

Las amenazas a la validez determinan las fortalezas y las limitaciones del estudio con respecto a la validez de los resultados obtenidos, y deben detallarse claramente. A continuación se discuten las amenazas a la validez del mapeo y revisión sistemática de literatura.

Selección de la cadena de búsqueda y las bibliotecas digitales. Con respecto a la selección de las palabras clave para la cadena de búsqueda, estas fueron definidas a partir de un conjunto de artículos de control, y piloteada en varias pruebas para reducir el ruido. Por su parte, las bases de datos seleccionadas son reconocidas por tener gran cobertura en el área de la ingeniería de software: *SCOPUS*, *IEEEExplore*, *ISI Web of Science*, y *Engineering Village*. Asimismo, se revisaron las referencias de los artículos identificados para localizar artículos relevantes que no se identificaron durante las búsquedas automáticas mediante el proceso de bola de nieve.

Identificación de estudios primarios. Debido a que el proceso de inclusión y exclusión, el proceso de extracción de datos y de la evaluación de la calidad es realizado por un solo investigador, podría existir un sesgo en la selección y extracción de los artículos primarios. Para reducir la posible amenaza, se aplicaron las siguientes acciones: el protocolo del mapeo y la revisión sistemática de literatura fue revisado por un par (colega) y piloteado en varias corridas de prueba, lo que asegura la claridad de los criterios de inclusión y exclusión, así como de los datos del formulario de extracción. Con respecto a la validez del proceso de inclusión y exclusión y la evaluación de calidad, a partir de una muestra de artículos seleccionados aleatoriamente, los resultados de la revisión de un segundo investigador mostraron un alto grado de acuerdo entre los investigadores. Aquellos casos en donde no

existió acuerdo, se revisaron en detalle mediante la lectura completa. Los estudios secundarios identificados durante el proceso validan la selección de artículos primarios. Durante la evaluación de calidad, la mayoría de las preguntas fueron booleanas y, para las que no lo son, se estableció una escala de clasificación previamente definida. Para minimizar el riesgo de excluir literatura relevante se adoptó un criterio inclusivo durante la evaluación de la inclusión y de la exclusión. Cualquier artículo primario que genera dudas sobre su relevancia se lee de forma completa para determinar si debe ser incluido. La evaluación de calidad no fue un criterio para la exclusión de ningún artículo, pues existen artículos que representan propuestas relevantes de los procedimientos de medición del tamaño funcional. Para la selección de artículos no se consideró una ventana de tiempo específico, lo que permitió recolectar la totalidad de la literatura disponible en los medios electrónicos, sin importar el año de publicación. Tampoco se consideró la literatura gris, salvo en casos específicos de documentos técnicos de autores reconocidos que fueron referenciados por artículos primarios seleccionados.

Extracción y clasificación. La extracción se basa en un esquema de clasificación específico. El proceso de clasificación y extracción es realizado por un solo investigador y validado por un segundo investigador, además se ejecutan revisiones ciegas aleatorias en un subconjunto de artículos. Las ambigüedades se resuelven durante el proceso. La interpretación de los resultados por parte de los investigadores es una amenaza a la validez. Después de la extracción de los datos, el investigador revisó el formulario de extracción para determinar su completitud, la existencia de inconsistencias, y finalmente, verificar y depurar cualquier información extraída que se considerara necesaria. Cada vez que se detectó una inconsistencia, el investigador revisó el artículo en detalle para determinar la *correctitud* de la información extraída. Este enfoque de doble revisión, llamado *test-retest*, ha sido exitosamente aplicado en otros estudios secundarios en el área de la ingeniería de software (Kitchenham, 2010; Petersen, 2011). Los artículos son clasificados de acuerdo con lo reportado por los autores originales y en caso de no ser reportados explícitamente los investigadores de este estudio, si es posible, realizan la clasificación. Se realiza un proceso sistemático para la clasificación de los artículos. Se diseña un formulario de extracción para

la recolección de datos que guía el proceso y que puede ser revisado. La aplicación de los criterios de calidad es realizada solo por un investigador, lo que representa una amenaza a la validez.

Generalización de los resultados. La generalización de resultados se limita a los estudios incluidos en el mapeo. Los componentes de contexto se extrajeron de cada artículo primario para ser tomados en cuenta cuando se determina el grado de generalización. Los investigadores tienen experiencia en el área de medición, lo que es una ventaja, pero a su vez, cabe la posibilidad de que sus intereses y sus opiniones afecten las recomendaciones que se realizan. Durante todo el proceso, se aplican lineamientos en el área de la SE experimental que detallan los protocolos y las mejores prácticas para ejecutar estudios secundarios. Se reporta detalladamente el proceso para facilitar el análisis y replicación.

Anexo 2 Formulario de extracción de los artículos relevantes del estudio secundario.

El Anexo 2 de este documento de tesis se encuentra en el dispositivo de almacenamiento adjunto.

Anexo 3 Listado de artículos primarios del mapeo y revisión sistemática de literatura.

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
1	RAS91	(Rask, 1991)	1991		Propone las reglas para contar los puntos de función IFPUG a partir de diagramas de flujos de datos y modelos entidad relación.	SBB	R	11	I	M	P	Modelos conceptuales	R D	P	PR	MIS		8
2	MAT93	(Matson & Mellichamp, 1993)	1993		Describe una herramienta para el conteo de puntos de función en aplicaciones orientadas a objetos.	AUT	J	10	I	E	P	Modelos conceptuales	D	PH	CS	MIS	Y	9
3	JON95	(Jones, 1995)	1995		Propone un procedimiento de conteo ajustando el método IFPUG para contar puntos de función a partir de las entidades de datos automáticamente.	SBB	C	4	I	M	P	Modelos de diseño	D	PH	PC	MIS	Y	7
4	PAT95	(Paton & Abran, 1995)	1995		Propone una especificación formal para el método de medición IFPUG y detalla el conjunto de reglas que pueden ser automatizadas para el conteo de tamaño funcional.	SBB	R	10	I	M	P	Modelos conceptuales	R D	P	PR	MIS		8
5	APR97	(April, Merlo & Abran, 1997)	1997		Propone las reglas para extraer los puntos de función de forma automática a partir del código fuente.	AUT	C	7	I	M	P	Código fuente	C	PH	PR	MIS	Y	10
6	FET97	(Fetcke, Abran & Nguyen, 1997)	1997		Aplica IFPUG en el software desarrollado con el método OO-Jacobson en la fase de análisis de requerimientos.	AUT	C	131	I	M	P	Modelos UML	R D	P	CS	MIS		11
7	GRA97	(Gramantieri, Lamma, Mello & Riguzzi, 1997)	1997	G012	Propone las reglas para contar los puntos de función IFPUG a partir de diagramas de flujos de datos y modelos entidad relación. Propone un grafo de representación a partir de los diagramas.	SBB	R	12	I	M	P	Modelos conceptuales	D	PH	PC	MIS	Y	9
8	LEH97	(Lehne, 1997)	1997		Propone las reglas para medir aplicaciones orientadas a objetos modelados con el método OORAM basado en roles.	SBB	C	13	I	M	P	Modelos conceptuales	D	P	PR	MIS		0

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
9	MEL97	(Meli, 1997)	1997	G008	Propone un enfoque para estimar el tamaño funcional del método IFPUG	SBB	C	0	I	E	P	Especificación de requerimientos	R	PE	CS	MIS		7
10	ANT98	(Antoniol, Calzolari, Cristoforetti, Fiutem & Caldiera, 1998)	1998	G013	Propone una adaptación de puntos de función para medir aplicaciones orientadas a objetos, basado en el modelo de objetos para el método IFPUG.	AUT	C	21	I	E	P	Modelos UML	R D	PH	PC	MIS	Y	14
11	CAL98	(Caldiera, Antoniol, Fiutem & Lokan, 1998)	1998	G013	Presenta un procedimiento para medir el tamaño funcional de aplicaciones de software orientadas a objetos con el método IFPUG.	AUT	C	71	I	E	P	Modelos UML	R D	PH	CS	MIS	Y	14
12	FET98	(Fetcke, Abran & Nguyen, 1998)	1998		Aplica IFPUG en aplicaciones orientadas a objetos representados con casos de uso. Utiliza el enfoque OO-Jacobson.	SBB	C	11	I	M	P	Modelos UML	R D	P	CS	MIS		9
13	ANT99	(Antoniol, Lokan, Caldiera & Fiutem, 1999)	1999	G013	Propone un procedimiento para medir aplicaciones orientadas a objetos con el método IFPUG. Adapta los componentes de medición.	AUT	J	69	I	E	P	Modelos UML	R D	PHE	CS	MIS	Y	14
14	BEV99	(Bevo, Levesque & Abran, 1999)	1999	G022	Mapea diagramas UML para medición mediante el método COSMIC.	SBB	C	12	C	M	P	Modelos UML	R D	PHE	CS	MIS	Y	0
15	HOV99	(Ho & Abran, 1999)	1999		Propone un marco de trabajo que permite la construcción de un modelo para el conteo automático de puntos de función a partir del código fuente de COBOL mediante la técnica <i>program slicing</i> .	SBB	C	13	I	M	P	Código fuente	C	PHE	PC	MIS	Y	8
16	PAT99	(Paton & Lambert, 1999)	1999		Propone una representación de grafo para el flujo de datos de la aplicación derivado mediante <i>tracing</i> para realizar el conteo de puntos de función.	SBB	C	13	I	M	P	Código fuente	C	P	PR	MIS		8

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
17	UEM99	(Uemura, Kusumoto & Inoue, 1999)	1999	G015	Propone un procedimiento de medición del tamaño funcional para el método IFPUG basado en especificaciones de diseño UML.	AUT	C	93	I	M	P	Modelos UML	D	PHE	CS	MIS	Y	12
18	KUS00	(Kusumoto, Inoue, Kasimoto, Suzuki, Yuura & Tsuda, 2000)	2000	G015	Evalúa un procedimiento de medición del tamaño funcional para el método IFPUG basado en especificaciones de diseño UML y evalúa una herramienta de apoyo.	AUT	C	11	I	M	P	Modelos conceptuales	R	PHE	CS	MIS	Y	12
19	RAM00	(Ram & Raju, 2000)	2000		Propone un procedimiento de medición de la funcionalidad de aplicaciones orientadas a objetos.	AUT	C	28	I	M	P	Modelos conceptuales	D	PE	CS	MIS		7
20	DIA01	(Diab, Frappier & Denis, 2001)	2001	G029	Propone la formalización del método COSMIC para automatizar la medición de las especificaciones ROOM.	AUT	C	22	C	M	P	Modelos conceptuales	D	PHE	PC	RTS	Y	8
21	JEN01	(Jenner, 2001)	2001	G023	Analiza aspectos de granularidad del conteo de tamaño funcional de casos de uso.	SBB	C	50	C	E	P	Modelos UML	R D	PHE	CS	MIS	Y	9
22	SNE01	(Sneed, 2001)	2001		Propone una herramienta para contar el tamaño funcional de aplicaciones basado en el código fuente.	AUT	C	6	I	E	P	Código fuente	C	PHE	CS	MIS	Y	9
23	UEM01	(Uemura, Kusumoto & Inoue, 2001)	2001	G015	Propone las reglas detalladas para medir IFPUG sobre especificaciones UML. Describe una herramienta de medición que utiliza los modelos UML generados por <i>Rational Rose</i> como entradas del proceso de medición.	CTR	J	58	I	M	P	Modelos UML	D	PHE	CS	MIS	Y	12
24	DIA02	(Diab, Frappier & St-Denis, 2002)	2002		Propone la formalización del método IFPUG para automatizar la medición de las especificaciones B.	AUT	C	20	I	M	P	Lenguajes formales de especificación	R	PHE	CS	MIS	Y	10

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
25	KUS02	(Kusumoto, Imagawa, Inoue, Morimoto, Matsusita & Tsuda, 2002)	2002	G005	Propone un procedimiento de medición del tamaño funcional para el método IFPUG para aplicaciones orientadas a objetos a partir del código fuente.	AUT	C	32	I	E	P	Código fuente	C	PH	CS	MIS	Y	11
26	POE02	(Poels, 2002)	2002	G014	Realiza una medición sobre modelos basados en eventos (MERODE) en aplicaciones de múltiples capas basados en el método COSMIC.	AUT	C	15	C	M	P	Modelos conceptuales	D	PH	PR	MIS	Y	10
27	TAV02	(Tavares, Carvalho & Castro, 2002)	2002		Define reglas de representación entre los conceptos del método de tamaño funcional IFPUG y las primitivas del lenguaje de modelado unificado (UML)	SBB	C	6	I	E	P	Modelos UML	R D	PE	CS	MIS		8
28	ABR03	(Abrahamo & Pastor, 2003)	2003	G007	Propone un procedimiento basado en IFPUG para medir aplicaciones web que usan el método OO-H.	AUT	J	38	I	M	P	Modelos conceptuales	D	PHE	EC	MIS	Y	9
29	ANT03	(Antoniol, Fiutem & Lokan, 2003)	2003	G013	Evalúa un procedimiento para medir aplicaciones orientadas a objetos con el método IFPUG.	SBB	J	45	I	E	P	Modelos UML	R D	PHE	EX	MIS	Y	14
30	LI03	(Li, Nonaka, Kukurai & Azuma, 2003)	2003		Reporta una aplicación de soporte a la medición del tamaño funcional COSMIC basado en especificaciones de la interfaz de usuario X-Forms.	AUT	C	3	C	M	P	Modelos de diseño	D	PH	PC	RTS	Y	10
31	NAG03	(Nagano & Ajisaka, 2003)	2003		Propone un procedimiento de medición del tamaño funcional COSMIC a partir de UML para aplicaciones orientadas a objetos en tiempo real.	SBB	C	22	C	M	P	Modelos UML	R	PE	PC	RTS		8
32	POE03a	(Poels, 2003a)	2003	G014	Presenta una validación teórica para el método COSMIC.	SBB	C	0	C	M	P	Modelos conceptuales	D	P	PR	RTS		7
33	POE03b	(Poels, 2003b)	2003	G014	Realiza una medición sobre modelos de negocios y servicios (MERODE) para aplicaciones de múltiples capas basados en el método COSMIC.	AUT	C	24	C	M	P	Modelos conceptuales	D	P	PR	MIS		10

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
34	POE03c	(Poels, 2003c)	2003	G014	Realiza una medición de aplicaciones de múltiples capas basados en el método COSMIC.	AUT	C	2	C	M	P	Modelos conceptuales	D	PH	PR	MIS	Y	10
35	ZIV03	(Zivkovic, Hericko & Rozman, 2003)	2003	G018	Presenta un mapeo entre modelos UML y puntos de función para software orientado a objetos.	AUT	C	0	I	E	P	Modelos UML	D	PH	PC	MIS	Y	9
36	ABR04a	(Abrahao, Poels & Pastor, 2004)	2004	G007	Evalúa un procedimiento de medición basado en el método IFPUG que utilizan el método de desarrollo <i>OO-Method</i> .	CTR	C	20	I	M	P	Modelos conceptuales	D	HE	EC	MIS	Y	16
37	ABR04b	(Abrahao, Poels & Pastor, 2004)	2004	G007	Evalúa el procedimiento de medición OOmFP para aplicaciones web, el cual utiliza modelos conceptuales como entrada para el procedimiento de medición.	CTR	C	30	I	M	P	Modelos conceptuales	D	HE	EC	MIS	Y	16
38	AZZ04	(Azzouz & Abran, 2004)	2004		Presenta un procedimiento para la automatización del tamaño funcional COSMIC para el proceso de desarrollo RUP donde los conceptos UML son mapeados a los componentes COSMIC.	SBB	C	49	C	M	P	Modelos UML	R D	PHE	CS	MIS	Y	10
39	BER04	(Bertolami & Oliveros, 2004)	2004		Propone un procedimiento de medición a partir de escenarios en la etapa de análisis de requerimientos.	SBB	C	5	M	E	P	Lenguajes formales de especificación	R	PE	CS	MIS		8
40	CAN04	(Cantone, Pace & Calavaro, 2004)	2004	G011	Propone un mapeo entre los componentes de modelos UML y las entidades del método de puntos de función IFPUG.	AUT	C	33	I	E	P	Modelos UML	R D	PH	PC	MIS	Y	13
41	CON04	(Condori-Fernandez, Abrahao & Pastor, 2004)	2004	G009	Presenta un procedimiento de medición para sistemas orientados a objetos. Utiliza el método COSMIC sobre las especificaciones de requerimientos del método <i>OO-Method</i> .	AUT	C	28	C	M	P	Modelos conceptuales	R	PE	CS	MIS		9

#	CR	Ref.	A	G	Enfoque (E)	O	T	G	M	T	A	Artef.	E	T	T	D	H	Q
			N	R		R	F	H	E	M	T		C	P	E	F		
42	IOR04	(Iorio, 2004)	2004		Propone lineamientos para aplicar el método de tamaño funcional IFPUG sobre documentación UML.	SBB	C	13	I	E	P	Modelos UML	R D	P	PR	MIS		8
43	LAM04	(Lamma, Mello & Riguzzi, 2004)	2004	G012	Propone las reglas para contar los puntos de función IFPUG a partir de diagramas de flujos de datos y modelos entidad relación. Propone un grafo de representación a partir de los diagramas.	AUT	J	23	I	M	P	Modelos conceptuales	D	PHE	CS	MIS	Y	12
44	BER05	(Berg, Dekkers & Oudshoorn, 2005)	2005		Propone una estrategia de mapeo entre requerimientos expresados en diagramas UML y el método de medición IFPUG y COSMIC.	SBB	C	25	I, C	E	P	Modelos UML	R D	PE	CS	MIS		11
45	DEL05	(Del Bianco & Lavazza, 2005)	2005	G025	Evalúa un conjunto de métricas orientadas a objetos basadas en puntos de función para determinar su eficacia en modelos de predicción de esfuerzo.	AUT	C	11	I	E	P	Modelos UML	D	E	CS	MIS		0
46	DIA05	(Diab, Koukane, Frappier & St-Denis, 2005)	2005	G029	Presenta una herramienta CASE que automáticamente cuenta el tamaño funcional de modelos de <i>Rational Rose</i> para sistemas de tiempo real con el método COSMIC.	AUT	J	46	C	M	P	Modelos UML	D	PHE	CS	RTS	Y	13
47	HAR05	(Harput, Kaindl & Kramer, 2005)	2005		Aplica el método de medición IFPUG sobre modelos UML orientados a objetos. Propone las reglas para la transformación semiautomática de los requerimientos orientados a objetos al modelo IFPUG.	AUT	C	26	I	E	P	Modelos UML	R D	PHE	CS	MIS	Y	15
48	PAC05	(Pace, Cantone & Calavaro, 2005)	2005	G011	Propone un mapeo entre los componentes de modelos UML y las entidades del método de puntos de función IFPUG e implementa una herramienta de soporte.	SBF	C	1	I	E	P	Modelos UML	R D	PH	PC	MIS	Y	10
49	SAN05	(Santillo, Conte & Meli, 2005)	2005	G008	Propone un enfoque para estimar el tamaño funcional del método COSMIC	SBB	C	0	C	E	P	Especificación de requerimientos	R	PE	CS	MIS	Y	4

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
50	ZIV05a	(Zivkovic, Hericko, Brumen, Beloglavec & Rozman, 2005)	2005	G018	Discute estrategias de mapeo existentes y define una estrategia de mapeo entre modelos UML y el método IFPUG.	AUT	J	15	I	E	P	Modelos UML	D	PHE	CS	MIS	Y	13
51	ZIV05b	(Zivkovic, Rozman & Hericko, 2005)	2005	G018	Presenta un mapeo entre modelos UML para software orientado a objetos y el método de medición IFPUG.	CTR	J	43	I	E	P	Modelos UML	D	PHE	EX	MIS	Y	14
52	ABR06	(Abraham, Poels & Pastor, 2006)	2006	G007	Propone las reglas de un procedimiento de medición OOmFP para aplicaciones web. Utiliza modelos conceptuales como entrada para el procedimiento de medición.	SBB	J	39	I	M	P	Modelos conceptuales	D	PHE	EC	MIS	Y	17
53	BEV06	(Bevo & Levesque, 2006)	2006	G028	Propone una herramienta de medición del tamaño funcional basada en una ontología.	AUT	C	0	I	E	P	Especificación de requerimientos	R	PH	PR		Y	0
54	CHO06	(Choi, Park & Sugumar, 2006)	2006	G003	Discute un procedimiento para extraer puntos de función a partir de requerimientos textuales. Utiliza el método de metas y escenarios.	AUT	C	6	I	E	P	Lenguajes formales de especificación	R	PHE	PC	MIS		9
55	CON06a	(Condori-fernández & Pastor, 2006a)	2006	G009	Evalúa un procedimiento de medición para sistemas orientados a objetos usando el método COSMIC a partir de las especificaciones de requerimientos basadas en el método <i>OO-Method</i> .	AUT	C	10	C	M	P	Modelos conceptuales	R	PHE	EC	MIS	Y	15
56	CON06b	(Condori-Fernández & Pastor, 2006b)	2006	G009	Evalúa el procedimiento RmFFP para determinar su productividad y reproducibilidad al medir puntos de función COSMIC.	AUT	C	14	C	M	P	Modelos conceptuales	R	PHE	EC	MIS	Y	15
57	CON06c	(Condori-Fernández & Pastor, 2006c)	2006	G009	Evalúa un procedimiento de medición para sistemas orientados a objetos. Utiliza el método COSMIC a partir de las especificaciones de requerimientos del método <i>OO-Method</i> .	SBF	C	4	C	M	P	Modelos conceptuales	R	PHE	EC	MIS	Y	13

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
58	ELL06	(Ellafi & Meli, 2006)	2006		Propone un análisis para calcular automáticamente el tamaño funcional IFPUG mediante modelos funcionales contruidos a partir del código fuente.	SBB	C	3	I	E	P	Código fuente	C	PH	PC	MIS	Y	11
59	FRA06	(Fraternali, Tisi & Bongio, 2006)	2006		Presenta un procedimiento automatizado para medir modelos conceptuales en ambientes MDA.	AUT	C	30	I	E	P	Modelos UML	D	PHE	CS	MIS	Y	12
60	GUX06	(Gu, Song & Li, 2006)	2006		Propone un método para medir puntos de función COSMIC para aplicaciones web a partir del código fuente de la interfaz de usuario.	SBB	C	6	C	M	P	Código fuente	C	PE	CS	MIS		6
61	HER06	(Hericko, Rozman & Zivkovic, 2006)	2006	G018	Propone una representación de métodos de tamaño funcional y los mapea en el sistema GASS para el cálculo de puntos de función IFPUG, COSMIC, MkII.	CTR	J	13	I, C, M	E	P	Modelos UML	D	PH	CS	MIS	Y	13
62	ABR07b	(Abrahão & Poels, 2007)	2007	G007	Evalúa el procedimiento de medición OOmFP para aplicaciones web que utiliza modelos conceptuales como entrada para el procedimiento de medición	AUT	J	29	I	M	P	Modelos conceptuales	D	HE	EC	MIS	Y	14
63	ABR07a	(Abrahao, Mendes, Gomez & Insfran, 2007)	2007	G007	Propone un procedimiento basado en IFPUG para medir aplicaciones web que usan el método OO-H para modelos de predicción de esfuerzo.	CTR	C	22	I	M	P	Modelos conceptuales	D	PH	CS	MIS	Y	12
64	CON07	(Condori-Fernandez, Abrahao & Pastor, 2007)	2007	G009	Presenta un procedimiento de medición para sistemas orientados a objetos que utiliza el método COSMIC a partir de las especificaciones de requerimientos del método <i>OO-Method</i> .	AUT	J	2	C	M	P	Modelos conceptuales	R	PE	EC	MIS		16
65	GIA07	(Giachetti, Marin, Condori-Fernandez & Molina, 2007)	2007	G007	Presenta las reglas de medición para los elementos del modelo conceptual del método <i>OO-Method</i> para el método IFPUG.	AUT	C	14	I	M	P	Modelos conceptuales	D	PH	CS	MIS	Y	12
66	GRA07	(Grau & Franch, 2007)	2007	G010	Propone un método para medir puntos de función COSMIC bajo el modelo PRIM.	SBB	C	14	C	M	P	Modelos conceptuales	R	PHE	CS	MIS	Y	10

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
67	IOR07	(Iorio, Meli & Perna, 2007)	2007	G008	Describe el método para medir el tamaño funcional bajo <i>E&QFP</i> 3.0.	SBB	C	0	I	E	P	Especificación de requerimientos	R	P	PR	MIS		7
68	POW07	(Pow-Sang & Imbert, 2007)	2007		Presenta la reglas para el mapeo de los componentes de IFPUG y diagramas de clase	SBB	C	0	I	E	P	Modelos UML	R D	PE	EC	MIS		13
69	VOG07	(Vogelezang & Prins, 2007)	2007	G008	Propone un procedimiento para la estimación temprana del tamaño funcional COSMIC	SBF	C	0	C	A	P	Especificación de requerimientos	R	PE	EC	MIS		7
70	ABR08a	(Abrahão & Insfran, 2008)	2008		Introduce un procedimiento de medición para aplicaciones orientadas a objetos a partir de la especificación de requerimientos de la herramienta REST.	AUT	C	6	I	E	P	Modelos conceptuales	R	PHE	CS	MIS	Y	13
71	ABR08b	(Abraham, Poels & Insfran, 2008)	2008	G007	Evalúa el procedimiento de medición OomFP para aplicaciones web que utiliza modelos conceptuales como entrada para el procedimiento de medición.	AUT	C	1	I	M	P	Modelos conceptuales	D	HE	EC	MIS	Y	13
72	CHA08	(Chamundeswari & Babu, 2008)	2008	G027	Propone un procedimiento para medir puntos de función a partir de diagramas de casos de uso.	AUT	C	4	I	E	P	Modelos UML	R D	PE	PC	MIS		9
73	CON08a	(Condori-Fernández & Pastor, 2008)	2008	G009	Evalúa un procedimiento de medición automatizado para sistemas orientados a objetos usando el método COSMIC a partir de las especificaciones de requerimientos basadas en el método <i>OO-Method</i> .	SBF	C	3	C	M	P	Modelos conceptuales	R	HE	CS	MIS	Y	13
74	CON08b	(Condori-Fernández, Pastor, Dagneva, Abran & Castro, 2008)	2008	G009	Presenta un procedimiento de medición de reusabilidad para sistemas orientados a objetos. Utiliza el método COSMIC que se aplica en las especificaciones de requerimientos del método <i>OO-Method</i> .	AUT	C	0	C	M	P	Modelos conceptuales	R	PH	PC	MIS	Y	12

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
75	DEL08	(Del Bianco, Gentile & Lavazza, 2008)	2008	G025	Propone un procedimiento para la construcción de modelos UML para aplicar las reglas del método de medición IFPUG.	SBB	C	12	I	E	P	Modelos UML	R D	PE	EC	MIS		11
76	GRA08	(Grau, 2008)	2008	G010	Propone un método para medir puntos de función COSMIC bajo el modelo PRIM.	AUT	C	1	C	M	P	Modelos conceptuales	R	PHE	CS	MIS	Y	10
77	GUP08	(Gupta, Kaushal & Sadiq, 2008)	2008	G005	Propone un procedimiento para calcular puntos de función en aplicaciones de tres capas.	AUT	C	6	I	E	P	Código fuente	C M	PH	PR	MIS	Y	10
78	LAV08	(Lavazza, del Bianco & Garavaglia, 2008)	2008	G025	Propone reglas para la construcción de modelos UML orientados a puntos de función para facilitar su medición.	SBF	C	35	I	E	P	Modelos UML	D	PHE	CS	MIS	Y	12
79	LEV08	(Levesque, Bevo & Cao, 2008)	2008	G022	Propone un procedimiento para medir el tamaño funcional COSMIC a partir de modelos UML.	AUT	C	43	C	M	P	Modelos UML	R D	PE	CS	MIS		11
80	MAR08b	(Marín, Condori-Fernández & Pastor, 2008)	2008	G006	Presenta una propuesta para evaluar la precisión de las mediciones de acuerdo al estándar ISO 5725	AUT	C	6	C	M	P	Modelos conceptuales	D	PHE	PC	MIS	Y	6
81	MAR08a	(Marin, Condori-Fernandez, Pastor & Abran, 2008)	2008	G006	Presenta el procedimiento de medición OOmFP para medir el tamaño funcional de aplicaciones desarrolladas con el método <i>OO-Method</i> para el método COSMIC.	AUT	C	17	C	M	P	Modelos conceptuales	D	P	PR	MIS		3
82	MAR08	(Marín, Pastor & Giachetti, 2008)	2008	G006	Presenta el procedimiento de medición OOmFP para medir automáticamente el tamaño funcional de aplicaciones desarrolladas con el método <i>OO-Method</i> para el método COSMIC.	CTR	C	32	C	M	P	Modelos conceptuales	D	PHE	CS	MIS	Y	12

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
83	MIY08	(Miyawaki, Iijima & Ho, 2008)	2008		Propone un procedimiento de medición basado en el método IFPUG aplicado sobre las especificaciones formales del lenguaje VDM-SL (<i>Vienna Development Method Specification Language</i>).	AUT	C	5	I	M	P	Lenguajes formales de especificación	R	PH	PC	RTS	Y	11
84	ABR09	(Abrahão & Poels, 2009)	2009	G007	Realiza una validación del procedimiento de medición OOmFPWeb. Utiliza el modelo de aceptación de tecnología (TAM).	AUT	J	21	I	M	P	Modelos conceptuales	D	HE	EC	MIS	Y	13
85	DEL09	(Del Bianco & Lavazza, 2009)	2009	G025	Explora la posibilidad de aplicar el método de medición COSMIC a los modelos del método <i>Problem Frames</i> (UML y nativo).	AUT	C	4	C	E	P	Modelos UML	R D	P	PC	MIS		9
86	HUS09	(Hussain, Ormandjieva & Kosseim, 2009)	2009	G030	Realiza una propuesta de aproximación del tamaño funcional COSMIC a partir de requerimientos textuales	SBB	C	0	C	A	P	Especificación de requerimientos	R	P	PR	MIS	Y	6
87	LAV09a	(Lavazza & Del Bianco, 2009)	2009	G025	Propone un procedimiento para la construcción de modelos UML para aplicar las reglas del método de medición COSMIC de forma sencilla y consistente.	AUT	C	31	C	E	P	Modelos UML	R D	PE	CS	RTS		10
88	LAV09b	(Lavazza & Garavaglia, 2009)	2009	G025	Detalla un procedimiento para la aplicación del método de medición IFPUG en aplicaciones en tiempo real.	AUT	C	18	I	E	P	Modelos UML	R D	PE	CS	RTS		9
89	MAR09	(Marin, Giachetti & Pastor, 2009)	2009	G006	Utiliza un procedimiento de medición del tamaño funcional para encontrar defectos en modelos conceptuales. Utiliza el método COSMIC.	AUT	C	9	C	M	P	Modelos conceptuales	D	PHE	CS	MIS	Y	8
90	POW09	(Pow-Sang, Gasco & Nakasone, 2009)	2009	G017	Presenta la reglas para el mapeo de los componentes de IFPUG y diagramas de clase	AUT	J	1	I	E	P	Modelos UML	R D	PE	EC	MIS		16
91	SAX09	(Saxena & Shrivastava, 2009)	2009		Evalúa el proceso de medición del tamaño funcional IFPUG sobre modelos UML.	SBB	J	3	I	E	P	Modelos UML	D	PE	PC	MIS		6

#	CR	Ref.	A	G	Enfoque (E)	O	T	G	M	T	A	Artef.	E	T	T	D	H	Q
			N	R		R	F	H	E	M	T		C	P	E	F		
92	SEL09	(Sellami & Ben-Abdallah, 2009)	2009	G002	Analiza la medición del tamaño funcional basado en COSMIC. Utiliza como entrada diagramas de secuencia UML.	SBF	C	17	C	E	P	Modelos UML	R D	PE	PC	MIS		9
93	TAN09	(Tan, Zhao & Zhang, 2009)	2009		Propone una estimación de tamaño en LOC basado en métricas de los modelos conceptuales.	SBB	J	25	I	E	P	Modelos UML	D	PE	CS	MIS		0
95	ABR10b	(Abrahão, Gómez & Insfran, 2010)	2010	G007	Valida el procedimiento de medición OO-HFP y compara los modelos de predicción de esfuerzo.	SBF	J	10	I	M	P	Modelos conceptuales	D	PHE	EC	MIS	Y	13
96	ADE10	(Adem & Kasirun, 2010)	2010	G003	Presenta un procedimiento de medición del tamaño funcional IFPUG sobre una especificación de requerimientos que se basa en metas y escenarios.	AUT	C	8	I	E	P	Lenguajes formales de especificación	R	PH	PC	MIS	Y	11
97	CHA10	(Chamundeswari & Babu, 2010)	2010	G027	Presenta un procedimiento para medir el tamaño funcional de las aplicaciones de software orientadas a objetos con el método IFPUG a partir de modelos UML.	SBF	C	5	I	E	P	Modelos UML	R D	PE	CS	MIS		9
98	GER10	(Gerlec, Krajnc & Goljat, 2010)	2010		Presenta una herramienta para medir el tamaño funcional IFPUG FPA en ambientes MDA	AUT	C	1	I	E	P	Modelos conceptuales	D	PHE	CS	MIS	Y	8
99	HUS10	(Hussain, Kosseim & Ormandjieva, 2010)	2010	G030	Realiza una propuesta de aproximación del tamaño funcional COSMIC a partir de requerimientos textuales	AUT	C	9	C	A	P	Especificación de requerimientos	R	PE	PC	MIS		9
100	JIC10	(Ji, Yan, Ma & Song, 2010)	2010	G021	Propone un procedimiento para medir puntos de función COSMIC, IFPUG y MarkII a partir de diagramas UML	AUT	C	0	I, C, M	M	P	Modelos conceptuales	R	PHE	CS	MIS	Y	8
101	MAR10a	(Marin, Giachetti, Pastor & Vos, 2010)	2010	G006	Presenta una herramienta de detección de errores en modelos por medio de un procedimiento de medición del tamaño funcional	AUT	C	8	C	M	P	Modelos conceptuales	D	HE	PC	MIS	Y	7

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
102	MAR10b	(Marín, Giachetti, Pastor, Vos & Abran, 2010)	2010	G006	Utiliza un procedimiento de medición del tamaño funcional para encontrar defectos en modelos conceptuales. Utiliza el método COSMIC.	AUT	C	1	C	M	P	Modelos conceptuales	D	PHE	CS	MIS	Y	13
103	MAR10c	(Marín, Pastor & Abran, 2010)	2010	G006	Presenta un procedimiento de medición para medir aplicaciones orientadas a objetos a partir de los modelos conceptuales en UML. Utiliza el método COSMIC.	AUT	J	24	C	M	P	Modelos conceptuales	D	PH	CS	MIS	Y	15
104	POW10	(Pow-Sang, Gasco & Nakasone, 2010)	2010	G017	Propone un procedimiento para identificar los componentes del método de conteo IFPUG a partir del análisis de diagramas de clases UML. Evalúa las diferentes representaciones de asociaciones y detalla las reglas de mapeo.	SBF	J	3	I	E	P	Modelos UML	R D	PH	PC	MIS	Y	17
105	BAR11	(Barkallah, Gherbi & Abran, 2011)	2011		Introduce un marco de trabajo para la automatización de procedimientos FSM COSMIC basados en artefactos UML. Detalla el mapeo entre UML y COSMIC.	AUT	C	4	C	M	P	Modelos UML	R D	P	PR	MIS		8
106	BAT11	(Batista, Peixoto, Borges, Pádua, Resende & Pádua, 2011)	2011		Describe un procedimiento y una herramienta para contar el tamaño funcional IFPUG a partir de modelos de requerimientos expresados en UML.	SBF	J	4	I	M	P	Modelos UML	R	PHE	CS	MIS	Y	9
107	DES11	(Desharnais, Buglione & Kocaturk, 2011)	2011		Propone un procedimiento para medir puntos de función COSMIC a partir de historias de usuario.	AUT	C	9	C	E	P	Especificación de requerimientos	R	P	PR	MIS		6
108	EDA11	(Edagawa, Akaike, Higo, Kusumoto, Hanabusa & Shibamoto, 2011)	2011	G005	Examina la posibilidad de automatizar la medición del tamaño funcional IFPUG a partir del código fuente por medio de análisis estático.	CTR	J	3	I	E	P	Código fuente	C M	PHE	CS	MIS	Y	16
109	HAB11	(Habela, Glowacki, Serafinski & Subieta, 2011)	2011		Propone un procedimiento de medición para modelos de casos de uso. Utiliza el método COSMIC.	SBB	B	32	C	M	P	Modelos UML	D	P	PR	MIS		7

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
110	JEN11	(Jenner, 2011)	2011	G023	Propone la automatización del conteo de puntos de función COSMIC a partir de diagramas UML.	SBB	B	28	C	E	P	Modelos UML	R D	PH	PR	MIS	Y	9
111	KAY11	(Kaya & Demirors, 2011)	2011	G019	Propone un mapeo entre el método de medición COSMIC y modelos de procesos de negocios.	AUT	C	11	C	E	P	Modelo de negocios	R	PHE	CS	MIS		13
112	LIN11a	(Lind & Heldal, 2011)	2011	G004	Propone un procedimiento basado en modelos UML para la medición del tamaño funcional. Utiliza COSMIC para aplicaciones embebidas.	AUT	C	17	C	E	P	Modelos UML	R	PHE	CS	RTS	Y	15
113	LIN11b	(Lind, Heldal, Harutyunyan & Heimdahl, 2011)	2011	G004	Desarrolla una herramienta para automatizar la medición del tamaño funcional basado en perfiles UML.	AUT	C	5	C	E	P	Modelos UML	R	PHE	CS	RTS	Y	11
114	MON11	(Monsalve, Abran & April, 2011)	2011	G019	Analiza el uso de modelos BPMN para medir aplicaciones transaccionales y en tiempo real. Utiliza el método COSMIC.	AUT	J	16	C	M	P	Modelo de negocios	R	PE	PC	RTS		10
115	SOU11	(Soubra, Abran, Stern & Ramdane-Cherifet, 2011)	2011	G026	Utiliza el lenguaje Simulink como entrada para medir el tamaño funcional. Utiliza COSMIC para aplicaciones en tiempo real.	CTR	C	9	C	M	P	Modelos conceptuales	D	PH	PC	RTS	Y	11
116	AKC12	(Akca & Tarhan, 2012)	2012	G024	Desarrolla una librería que obtiene los procesos funcionales de las aplicaciones en tiempo de ejecución mediante la interfaz de usuario y mide el tamaño funcional con el método COSMIC.	AUT	C	4	C	M	P	Código fuente	C M	PH	PC	MIS	Y	9
117	CHO12	(Choi, Park & Sugumaran, 2012a)	2012	G003	Propone un procedimiento basado en reglas para obtener el tamaño funcional a partir de requerimientos textuales. El documento de requerimientos se basa en el modelo de metas y escenarios.	AUT	J	5	I	E	P	Lenguajes formales de especificación	R	PH	CS	MIS	Y	13

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
118	IRA12	(Irawati & Mustofa, 2012)	2012		Propone una herramienta para ejecutar el análisis de puntos de función IFPUG de forma eficiente a partir de archivos XML resultante de la documentación basada en modelos UML.	SBF	J	4	I	E	P	Modelos UML	D	PH	CS	MIS	Y	9
119	JIC12	(Ji & Yan, 2012)	2012	G021	Propone un procedimiento para medir puntos de función COSMIC a partir de diagramas UML	AUT	J	1	C	M	P	Modelos UML	R	PE	CS	MIS		9
120	LIN12	(Lind & Haldal, 2012)	2012	G004	Propone una herramienta para la estimación del tamaño del código basado en COSMIC basado en modelos UML.	AUT	J	17	C	E	P	Modelos UML	R	PHE	CS	RTS	Y	16
121	SAN12	(Santillo, 2012)	2012	G008	Realiza una comparación entre técnicas para la aproximación del tamaño funcional en IFPUG y COSMIC	SBB	C	0	I, C	A	P	Especificación de requerimientos	R	P	AC	MIS		5
122	SOU12a	(Soubra & Chaaban, 2012)	2012	G026	Presenta los lineamientos para medir unidades de control electrónico de aplicaciones de software para automóviles. Utiliza el método COSMIC.	SBF	C	4	C	M	P	Modelos conceptuales	D	PH	PC	RTS	Y	13
123	SOU12b	(Soubra, Abran & Ramdane-Cherif, 2012)	2012	G026	Utiliza el lenguaje Simulink como entrada para medir el tamaño funcional. Utiliza COSMIC para aplicaciones en tiempo real.	AUT	C	5	C	M	P	Modelos conceptuales	D	PH	PC	RTS	Y	10
124	ZHA12	(Zhang & Ren, 2012)	2012		Propone un procedimiento de mapeo entre casos de uso y COSMIC. Se enfoca en los flujos de los eventos de los casos de uso para refinar los resultados de medición.	SBF	C	0	C	E	P	Modelos UML	D	P	PR	MIS		8
125	ABD13	(Abdullah, Rusli & Ibrahim, 2013)	2013	G020	Presenta un procedimiento para medir aplicaciones móviles de juegos mediante el método COSMIC basadas en diagramas UML.	AUT	C	2	C	M	P	Modelos UML	R D	P	CS	RTS		7
126	AKC13	(Akca & Tarhan, 2013)	2013	G024	Mide procesos funcionales disparados desde la interfaz de usuario de forma semiautomática para aplicaciones java de tres capas.	AUT	C	11	C	M	P	Código fuente	C M	PH	CS	MIS	Y	16

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
127	CEK13	(Ceke, Durek & Kasapovic, 2013)	2013		Presenta un procedimiento de medición que se aplica en modelos UML basados en ingeniería Web (UWE). Calcula el tamaño en puntos de función COSMIC.	SBF	C	6	C	E	P	Modelos UML	R D	PH	PC	MIS	Y	10
128	DEL13	(Del Bianco, Lavazza, Liu, Morasca & Abualkishik, 2013)	2013	G025	Propone medir el tamaño funcional a partir de modelos UML simplificados	SBB	C	0	C	E	P	Modelos UML	R D	PE	CS	MIS		12
129	HAN13	(Han & Chen, 2013)	2013		Propone los lineamientos para medir el tamaño funcional de servicios de software en la nube con el método COSMIC.	AUT	C	0	C	M	P	Especificación de requerimientos	R	E	CS	RTS		5
130	HUS13	(Hussain, Kosseim, Ormandjieva, 2013)	2013	G030	Proponen un proceso para aproximar los puntos de función COSMIC a partir de los requerimientos textuales en ágiles.	AUT	J	25	C	A	P	Especificación de requerimientos	R	PHE	CS	MIS	Y	11
131	MAR13	(Marin, Giachetti, Pastor, Vos, Abran, 2013)	2013	G006	Utiliza un procedimiento de medición del tamaño funcional para encontrar defectos en modelos conceptuales. Utiliza el método COSMIC.	AUT	J	8	C	M	P	Modelos conceptuales	D	PHE	CS	MIS	Y	13
132	POW13	(Pow-Sang, Villanueva, Flores & Rusu, 2013)	2013	G017	Presentan las reglas para identificar puntos de función a partir de diagramas de clase UML.	AUT	C	5	I	E	P	Modelos UML	R D	PHE	CS	MIS	Y	12
133	SEL13a	(Sellami, Haoues & Ben-Abdallah, 2013)	2013	G002	Presenta una propuesta para evaluar la consistencia entre los diagramas de actividad y componentes UML basado en un procedimiento de medición basado en COSMIC.	AUT	C	5	C	E	P	Modelos UML	R D	PHE	PC	MIS	Y	10
134	SEL13b	(Sellami, Haoues & Ben-Abdallah, 2013)	2013	G002	Presenta una propuesta para evaluar la consistencia entre los diagramas de actividad y componentes UML basado en un procedimiento de medición basado en COSMIC.	AUT	C	2	C	E	P	Modelos UML	R D	PE	PC	MIS		9

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
135	SOU13	(Soubra, 2013)	2013	G026	Propone un procedimiento de medición para lenguajes sincrónicos (LUSTRE) bajo el método de medición COSMIC.	AUT	C	2	C	M	P	Modelos conceptuales	D	PH	PC	RTS	Y	13
136	YIL13	(Yilmaz, Tunalilar & Demirors, 2013)	2013		Proponen una herramienta para la detección automatizada de defectos en mediciones COSMIC	AUT	C	3	C	M	P	Especificación de requerimientos	R	PHE	CS	MIS	Y	7
137	ZIC13	(Zickert & Beck, 2013)	2013		Desarrolla un modelo de mapeo sobre los requerimientos (MMAPER) usando el método IFPUG. Aplica COCOMO II para la estimación de esfuerzo.	AUT	J	3	I	E	P	Modelos conceptuales	R	PH	CS	MIS	Y	10
138	ABD14	(Abdullah, Rusli & Ibrahim, 2014)	2014	G020	Presenta un procedimiento para medir aplicaciones móviles de juegos mediante el método COSMIC.	AUT	C	6	C	M	P	Modelos UML	R D	P	PC	RTS		7
139	AYS14	(Aysolmaz & Demirors, 2014)	2014	G016	Propone un procedimiento para automatizar la medición del tamaño funcional a partir de modelos de procesos de negocio. Utiliza el procedimiento UPROM para COSMIC.	CTR	C	4	C	E	P	Modelos conceptuales	R	PH	CS	MIS	Y	14
140	DEV14	(De Vito & Ferrucci, 2014)	2014	G008	Presenta un procedimiento de medición que se aplica a modelos de casos de uso para mejorar la productividad del conteo. Propone un conjunto de plantillas para la identificación de los componentes funcionales.	SBF	C	7	C	E	P	Modelos UML	D	P	CS	MIS		10
141	DEL14	(Del Bianco, Lavazza, Liu, Morasca & Abualkishik, 2014)	2014	G008	Propone medir el tamaño funcional a partir de modelos UML y evalúa como modelos más detallados proporcionan mayor exactitud en las mediciones.	SBB	J	7	C	E	P	Modelos UML	R D	PE	EC	MIS		10
142	JIN14	(Jing, Zeng, Liu, Cheng & Lu, 2014)	2014		Propone un procedimiento para medir aplicaciones orientadas a objetos con el método MarkII usando el modelo UML.	AUT	C	0	M	M	P	Modelos UML	R D	PHE	CS	MIS	Y	11

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
143	KUI14	(Kuijpers, 2014)	2014		Describe una solución automatizada de conteo de puntos de función mediante un diccionario de modelos.	AUT	C	2	I	E	P	Especificación de requerimientos	R	PH	CS	MIS	Y	13
144	LIU14	(Liu, Yang, Tao, Zhu & Wu, 2014)	2014		Mide diagramas UML con el método IFPUG.	AUT	C	1	I	E	P	Modelos UML	RD	PH	PC	MIS	Y	7
145	MAR14	(Marin & Quinteros, 2014)	2014	G019	Propone un procedimiento de medición basado en COSMIC que mide diagramas BPMN.	AUT	C	1	C	M	P	Modelo de negocios	R	PH	PC	MIS	Y	11
146	ORI14	(Oriou, Bronca, Bouzid, Guetta & Guillard, 2014)	2014	G026	Propone una herramienta de medición basado en COSMIC para obtener el conteo de aplicaciones representadas con la especificación de sistemas de la herramienta Simulink.	AUT	C	3	C	M	P	Modelos conceptuales	D	PH	CS	RTS	Y	9
147	SAG14	(Sag & Tarhan, 2014)	2014	G024	Desarrolla una herramienta prototipo de automatización de la medición del tamaño COSMIC de aplicaciones Java. El prototipo utiliza diagramas de secuencia UML generados a partir de la ejecución del programa (código fuente y código binario) y realiza la medición del tamaño.	AUT	C	4	C	M	P	Código fuente	CM	PH	PC	MIS	Y	9
148	SAL14	(Salmanouglu, Demirors & Turetken, 2014)	2014	G019	Realiza la medición del tamaño funcional COSMIC usando diagramas S-BPM. Puede aplicarse en etapas tempranas del proceso de desarrollo.	SBF	C	0	C	M	P	Modelo de negocios	R	P	CS	MIS		12
149	SOU14	(Soubra, Abran & Ramdane-Cherif, 2014)	2014	G026	Propone un protocolo de verificación de exactitud para el método COSMIC para demostrar la exactitud de las mediciones automáticas de tamaño funcional COSMIC por medio de un proceso sistemático. Valida el protocolo aplicando una herramienta de conteo en dos casos de estudio.	CTR	C	6	C	M	P	Modelos conceptuales	D	PHE	CS	RTS	Y	14

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
150	HEE14	(Heeringen & Gorp, 2014)	2014	G000	Describe un procedimiento para la medición de aplicaciones móviles con el método COSMIC.	SBF	C	6	C	A	P	Especificación de requerimientos	R	P	PC	MIS		9
151	AYS15	(Aysolmaz & Demirors, 2015)	2015	G016	Propone un procedimiento para automatizar la medición del tamaño funcional a partir de modelos de procesos de negocio. Utiliza el procedimiento UPROM para COSMIC.	AUT	C	0	C	E	P	Modelos conceptuales	R	PH	CS	MIS	Y	10
152	CEK15	(Ceke & Milasinovic, 2015)	2015		Propone las reglas de mapeo entre el modelo conceptual (UWE) y el método COSMIC, para aplicaciones web. Presenta una herramienta para la automatización del conteo.	AUT	C	1	C	E	P	Modelos conceptuales	R	PHE	CS	MIS	Y	9
153	DAV15	(DAvanzo, Ferrucci, Gravino & Salza, 2015)	2015	G000	Presenta las reglas para la aplicación del método COSMIC en ambientes móviles. El procedimiento identifica múltiples patrones del software para la medición del tamaño funcional.	SBF	C	5	C	M	P	Especificación de requerimientos	R	PE	CS	MIS		12
154	ERE15	(Eren, Ozkan & Demirors, 2015)	2015		Presenta las reglas de mapeo entre el método de medición COSMIC y los elementos conceptuales de modelos UML.	AUT	C	0	C	M	P	Modelos UML	R D	PHE	CS	RTS	Y	14
155	FER15	(Ferrucci, Gravino, Salza & Sarro, 2015)	2015	G000	Aplica el procedimiento de medición para contar el tamaño funcional COSMIC de aplicaciones móviles. Utiliza una estrategia de medición de patrones del software.	AUT	C	2	C	M	P	Especificación de requerimientos	R	PE	EC	MIS		13
156	GON15	(Gonultas & Tarhan, 2015)	2015	G024	Propone un procedimiento para automatizar la medición del tamaño funcional COSMIC durante la ejecución de escenarios en tiempo de ejecución.	AUT	C	2	C	M	P	Código fuente	C M	PH	PC	MIS	Y	11
157	LAV15	(Lavazza, 2015)	2015	G001	Realiza un análisis crítico de la propuesta automatizada de puntos de función (AFP). Propone posibles usos de la especificación AFP.	AUT	C	5	I	M	P	Código fuente	C	P	AC	MIS		11

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
158	LEE15	(Lee, 2015)	2015		Propone un mapeo entre modelos UML y puntos de función IFPUG para automatizar el conteo basado en los diagramas de clase y secuencia.	SBF	C	0	I	M	P	Modelos UML	D	P	PR	MIS		7
159	LEN15	(Lent & Qu, 2015)	2015		Propone un procedimiento de medición del tamaño funcional IFPUG para automatizar el conteo basado en coincidencia de patrones sobre archivos XML de representación de la interfaz del usuario.	SBB	J	0	I	M	P	Código fuente	C	PH	EX	MIS	Y	11
160	QUE15	(Quesada-López & Jenkins, 2015a)	2015	G001	Presenta una evaluación empírica con profesionales de la propuesta automatizada de puntos de función (AFP).	AUT	C	3	I	M	P	Código fuente	C	E	EC	MIS		13
161	QUE15b	(Quesada-López & Jenkins, 2015b)	2015	G001	Propone un protocolo de verificación de exactitud para el método IFPUG para evaluar la exactitud de las mediciones de tamaño funcional IFPUG por medio de un proceso sistemático.	AUT	C	1	I	M	P	Código fuente	C	PE	EC	MIS		11
162	SEL15	(Sellami, Hakim, Abran & Ben-Abdallah, 2015)	2015	G002	Propone un procedimiento de medición del tamaño funcional basado en COSMIC que usa como entrada diagramas de secuencia UML.	SBF	J	3	C	M	P	Modelos UML	R D	PE	PC	MIS		10
163	SOU15	(Soubra, Abran & Sehit, 2015)	2015	G026	Evalúa una herramienta de medición basado en COSMIC para la estimación.	AUT	C	0	C	M	P	Modelos conceptuales	D	HE	CS	RTS	Y	14
164	YEH15	(Yeh, Chen & Yang, 2015)	2015	G031	Desarrolla una herramienta de ingeniería inversa para obtener a partir de código fuente diagramas de secuencia y calcular el tamaño funcional del software.	AUT	C	0	C	E	P	Código fuente	C	PHE	PC	MIS	Y	8
165	BAG16	(Bagriyanik & Karahoca, 2016)	2016	G028	Define una ontología para requerimientos que utiliza como modelo para la automatización del conteo de puntos de función COSMIC.	AUT	J	3	C	M	P	Especificación de requerimientos	R	PHE	CS	RTS	Y	15

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
166	LAV16a	(Lavazza & Morasca, 2016)	2016	G008	Realiza una evaluación de métodos de estimación de tamaño rápidos para COSMIC	SBF	C	0	C	A	P	Especificación de requerimientos	R	E	CS	MIS		11
167	LAV16b	(Lavazza, Lenarduzzi & Taibi, 2016)	2016		Propone una modificación al método IFPUG para contemplar los componentes reutilizables	SBF	C	0	I	M	P	Especificación de requerimientos	R	PE	CS	MIS		10
168	LIN16	(Lin & Yeh, 2016)	2016	G031	Desarrolla una herramienta de ingeniería inversa para obtener a partir de código fuente diagramas de secuencia y calcular el tamaño funcional del software.	AUT	C	0	C	E	P	Código fuente	C	PHE	CS	MIS	Y	6
169	OCH16a	(Ochodek, 2016a)	2016	G030	Propone un procedimiento de aproximación del tamaño funcional COSMIC a partir de casos de uso	AUT	J	2	I, C	A	P	Especificación de requerimientos	R	PHE	CS	MIS	Y	17
170	OCH16b	(Ochodek, 2016b)	2016	G030	Validar un procedimiento de aproximación del tamaño funcional COSMIC a partir de requerimientos textuales	AUT	C	0	C	A	P	Especificación de requerimientos	R	E	CS	MIS		16
171	QUE16	(Quesada-López, Madrigal & Jenkins, 2016)	2016	G001	Presenta una evaluación empírica con profesionales de la propuesta automatizada de puntos de función (AFP).	AUT	C	0	I	M	P	Código fuente	C	E	EC	MIS		14
172	TAR16	(Tarhan, Ozkan & Icoz, 2016)	2016	G024	Propone los requerimientos para el desarrollo de una herramienta de automatización de la medición del tamaño COSMIC de aplicaciones a partir del código fuente.	AUT	C	1	C	M	P	Código fuente	C M	P	PR	MIS		8
173	TRU16	(Trudel, Desharnais & Cloutier, 2016)	2016	G032	Propone un conjunto de patrones de medición para obtener el tamaño funcional	AUT	C	1	C	E	P	Especificación de requerimientos	R	P	PC	RTS		11
174	OSK17	(Demirel & Özkan, 2015)	2017	G024	Propone la medición del tamaño funcional COSMIC para aplicaciones en la arquitectura <i>Three-Tier Object Relational Mapping architecture</i> .	SBB	J	0	C	M	P	Código fuente	C M	PH	CS	MIS	Y	15

#	CR	Ref.	A N	G R	Enfoque (E)	O R	T F	G H	M E	T M	A T	Artef.	E C	T P	T E	D F	H	Q
175	HAO17	(Haoues, Sellami & Ben-Abdallah, 2017)	2017	G000	Propone un procedimiento de medición para el cálculo rápido del tamaño funcional	SBF	C	0	C	A	P	Modelos UML	R D	PE	CS	MIS		11
176	QUE17b	(Quesada-López, Jenkins, Salas & Gómez, 2017b)	2017	G001	Propone una herramienta de medición para modelos de diseño.	SBF	C	0	I	E	P	Modelos de diseño	D	PHE	CS	MIS	Y	15
177	RUS17	(Rusli & Abdullah, 2017)	2017	G020	Presenta un procedimiento para medir aplicaciones móviles de juegos mediante el método IFPUG basadas en diagramas UML.	AUT	J	0	I	E	P	Modelos UML	R D	P	PC	RTS		9
178	HAS17	(Soubra & Abran, 2017)	2017		Propone los lineamientos para medir aplicaciones del dominio IoT	SBF	C	0	C	M	P	Código fuente	C	P	PC	RTS		8
179	UNG17	(Ungan, Trudel & Poulin, 2017)	2017	G032	Propone el uso de patrones de medición del tamaño funcional para facilitar la medición de requerimientos de seguridad	SBF	C	0	C	E	P	Modelos de diseño	R	PE	CS	MIS		11
180	WEN17	(Wentzlaff, 2017)	2017	G032	Propone el uso del tamaño funcional como punto de referencia para clasificar los requerimientos y estimar el tamaño funcional de manera reproducible. Plantea la medición utilizando <i>problem frames</i> asociando los BFC a cada patrón.	SBF	C	0	I	E	P	Especificación de requerimientos	R D	P	PC	MIS		4

#	CR	Ref.	A	G	Enfoque (E)	O	T	G	M	T	A	Artef.	E	T	T	D	H	Q
			N	R		R	F	H	E	M	T		C	P	E	F		

#: Secuencia, CR: Código de referencia, Ref.: referencia bibliográfica, AN: año, GR: grupo, OR: origen, TF: tipo de foro, G: Google hits, ME: método de medición, TM: tipo de medición, AT: atributo medido, Artef.: artefacto de entrada, EC: etapa del ciclo de vida de desarrollo, TP: tipo de propuesta, TE: tipo empírico, DF: dominio funcional, H: Herramienta, Q: suma total de calidad.

OR: CTR=artículos de control, AUT=búsqueda automática, SBB=proceso de bola de nieve hacia atrás, SBF=proceso de bola de nieve hacia adelante.

TF: J=Revista, C=Conferencia, B=Capítulo de libro, R=Reporte técnico.

ME: I=IFPUG FPA, C=COSMIC FFP, M=MKII FPA.

TM: A=Aproximación de tamaño (*early estimation*), E=Estimación de tamaño, M=Medición del tamaño.

AT: P=Producto, R=Proceso

EC: R=Requerimientos, D=Diseño, C=Codificación, M=Mantenimiento.

TP: P=Propuesta, H=Herramienta, E=Empírico.

TE: AC=Análisis crítico, RE=Reporte de experiencia, PR=Propuesta, PC=Prueba de concepto, CS=Caso de estudio, EC=Cuasi experimento o experimento controlado, EX=Experimento.

DF: MIS=Sistemas de información, RTS= sistemas en tiempo real.

Anexo 4 Amenazas a la validez de los estudios empíricos orientados a personas.

El Anexo 4 detalla el análisis de las amenazas a la validez de los estudios descritos en la Sección 6.2.

Amenazas a la validez

A continuación se analizan las amenazas a la validez y las acciones realizadas para mitigarlas. Las amenazas a la validez son las relacionadas con los diseños experimentales, los experimentos orientados a personas, los sujetos y los objetos de estudio.

Validez de la conclusión

El tamaño de la muestra puede afectar el poder de las pruebas estadísticas utilizadas. Asimismo, la experiencia y conocimiento de los participantes pueden afectar los resultados obtenidos. En este caso los participantes del estudio tienen características similares lo que disminuye la afectación de este factor. Las tareas de medición fueron realizadas en casa, lo que podría permitir que los participantes compartan información. Los tiempos asignados a cada tarea y los materiales utilizados en las capacitaciones pueden influir en los resultados del estudio. Se utilizaron materiales de capacitación y ejemplos similares a los utilizados en procesos de capacitación en medición del tamaño funcional. Las condiciones y artefactos específicos utilizados en cada replicación pueden influir en los resultados que se utilizan para realizar las comparaciones.

Validez interna

La validez interna se relaciona con el diseño de los experimentos y la selección de los sujetos. Los sujetos mantienen un nivel de experiencia profesional y conocimiento de los métodos de medición homogénea; sin embargo, la variación natural de los procesos realizados por los participantes puede afectar las salidas. Todos los instrumentos fueron

validados mediante estudios piloto. Existe la amenaza de que el tamaño de los distintos artefactos pueda influenciar los resultados. Otra amenaza es que los participantes puedan compartir información. Los resultados se chequean para validar que no existan resultados duplicados; sin embargo, cabe la posibilidad de que en algunos casos esta situación se pueda presentar. Del mismo modo, la motivación de los participantes puede afectar la calidad en la ejecución de los procesos de medición. En este caso, las asignaciones representan un importante porcentaje de su nota del curso.

En total se ejecutan cuatro estudios que combinan diseños entre-sujetos y diseños intra-sujetos. Las decisiones de cada uno de los diseños se tomaron de acuerdo a restricciones particulares de los cursos en los cuales fueron ejecutados. En el estudio original (E2-14) se selecciona un diseño simple entre-sujetos con aleatorización por restricciones de tiempo. En este caso la experiencia de los sujetos puede afectar los resultados del estudio; sin embargo, los participantes cuentan con experiencia profesional similar y conocimientos sobre medición comparables.

En la replicación-1 (E1-15) y la replicación-2 (E2-15) se selecciona un diseño simple intra-sujetos. En este caso los tratamientos son asignados a los sujetos en orden y el proceso se realiza en dos ejecuciones. El efecto del orden puede afectar los resultados del estudio ya que los participantes miden primero con el proceso de FPA y luego con el de AFP en la replicación-1, y miden primero con el proceso de AFP y luego con el de **aFPA** en la replicación-2. Asimismo, el efecto de aprendizaje puede afectar los resultados del estudio. Para mitigar este efecto los artefactos utilizados en cada uno de las corridas varían según las entradas requeridas por el proceso de medición. Además, para la replicación-2 se varía la aplicación medida en cada corrida. Los objetos de estudio utilizados son de complejidad similar; sin embargo, el tamaño podría influenciar los resultados.

En la replicación-3 (E1-18) se selecciona un diseño intra-sujetos con confusión parcial. El efecto de aprendizaje se mitiga utilizando distintos artefactos de distintos sistemas bajo medición en cada uno de las corridas. Los objetos de estudio utilizados son de complejidad similar. El efecto de aprendizaje también se contrarresta tomando en cuenta la secuencia de las corridas como factor. En este estudio se puede presentar el efecto fatiga dado

las corridas realizadas en clase. Para este efecto, las actividades son iniciadas en clase y terminadas en casa. Asimismo, se ejecuta una corrida con un diseño simple entre-sujetos sin aleatorización por restricciones de tiempo. Aunque la experiencia de los sujetos puede afectar los resultados, en este caso los participantes cuentan con similar experiencia profesional y conocimientos sobre medición.

En el caso de las mediciones verdaderas tomadas como valores de control, las aplicaciones utilizadas fueron medidas y reportadas en estudios previos y reportes técnicos. En algunos casos, las mediciones fueron validadas por dos expertos en medición no certificados para obtener las mediciones detalladas por cada uno de los componentes funcionales básicos. Para la encuesta de percepción, se indica explícitamente que la asignación de la nota se asigna por completitud de llenado y no por el contenido, por lo que se les invita a ofrecer respuestas honestas con respecto a los procesos de medición.

Validez del constructo

La validez al constructo se ve influenciada por las mediciones recolectadas. Todas las medidas utilizadas son las recomendadas en el área de medición. En el caso de los cuestionarios de percepción fueron elaborados con base en estudios previos en la literatura que utilizan el modelo de aceptación de tecnología (TAM). Las mediciones del tiempo de esfuerzo real de cada tarea de medición fueron tomadas y reportadas por cada uno de los participantes lo que puede afectar los datos recolectados.

Validez externa

Los cuatro estudios son realizados con estudiantes de maestría en computación e informática, quienes son profesionales de la ingeniería de software que, aunque no son representativos de profesionales de medición, sí están familiarizados con las prácticas y procesos de desarrollo de software. Las tareas de medición se presentan con un alto nivel de detalle, lo cual permite su ejecución sin contar con experiencia previa en el campo de

medición. El conocimiento previo de los profesionales matriculados en el curso es homogéneo, lo que presenta una ventaja en cada uno de los estudios. Por otro lado, los artefactos de software utilizados, aunque mantienen una complejidad similar, pueden ser considerados como aplicaciones pequeñas con variaciones en el tamaño. Los resultados con sistemas más grandes y complejos podrían presentar variaciones que deben ser estudiadas. Las aplicaciones bajo medición son las comúnmente utilizadas en este tipo de estudios dado que su nivel de complejidad permite asignarse como trabajos en clase. Asimismo, las metodologías de especificación de requerimientos, modelado y desarrollo son comunes en la práctica profesional. Los profesionales son, en su mayoría, ingenieros de software por lo que la generalización no se puede realizar más allá de esta población. Todas las actividades realizadas se enmarcan dentro del contexto de la medición funcional con aplicaciones transaccionales de negocios.

Anexo 5 Tamaño funcional sin ajustar IFPUG FPA por variables de contexto.

El Anexo 5 lista el detalle de las estadísticas descriptivas para cada una de las variables de contexto de los subconjuntos de datos.

DS	Vr	Niv	N	%	Mean	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rng	SD	Skw	Krt	
DS12	TL	3GL	12	1.0	158	40	40	42	77	133	237	309	-	313	273	97	0.38	-1.41	
	LP	C#	12	1.0	158	40	40	42	77	133	237	309	-	313	273	97	0.38	-1.41	
	TD	E	7	0.6	120	40	40	40	47	81	219	-	-	222	182	78	0.51	-1.86	
		N	4	0.3	236	90	90	90	128	270	309	-	-	313	223	102	-1.51	2.04	
	R(N=1, 0.1%, 112 UFP)																		
	PD	MP	12	1.0	158	40	40	42	77	133	237	309	-	313	273	97	0.38	-1.41	
	TE	5-8	6	0.5	145	75	75	75	86	133	220	-	-	222	147	64	0.34	-2.12	
		3-4	3	0.3	56	40	40	40	40	47	-	-	-	81	41	22	1.54	-	
		9-14	2	0.2	278	242	242	242	242	278	-	-	-	313	71	50	-	-	
		21-30	1	0.1	-	298	-	-	-	-	-	-	-	298	-	-	-	-	
	AN	2009	7	0.6	199	75	75	75	90	222	298	-	-	313	238	95	-0.21	-1.71	
		2010	3	0.3	126	47	47	47	47	112	-	-	-	219	172	87	0.71	-	
		2011	2	0.2	61	40	40	40	40	61	-	-	-	81	41	29	-	-	
	AR	CS	9	0.8	180	40	40	40	86	219	270	-	-	313	273	101	-0.05	-1.70	
		MW	3	0.3	92	47	47	47	47	75	-	-	-	154	107	55	1.25	-	
	TR	M1	6	0.5	208	112	112	112	144	221	256	-	-	298	186	66	-0.27	-0.30	
		S	5	0.4	67	40	40	40	44	75	86	-	-	90	50	22	-0.40	-2.58	
	M2(N=1, 0.1%, 313 UFP)																		
	DS72	TL	3GL	46	0.6	253	24	52	73	90	198	319	562	673	1,337	1,313	235	2.54	9.14
			4GL	16	0.2	230	59	59	69	157	191	331	426	-	461	402	118	0.59	-0.55
		ApG	9	0.1	181	54	54	54	69	155	238	-	-	493	439	136	1.63	3.30	
ND(N=1, 0.0%, 372 UFP)																			
LP		PL	18	0.3	226	61	61	73	117	224	305	465	-	550	489	135	0.88	0.53	
		C#	16	0.2	283	47	47	67	90	187	309	888	-	1,337	1,290	325	2.63	7.62	
		COO	9	0.1	181	54	54	54	69	155	238	-	-	493	439	136	1.63	3.30	
		ABA	8	0.1	284	180	180	180	187	269	369	-	-	461	281	106	0.55	-1.15	
		JV	8	0.1	236	24	24	24	75	116	499	-	-	583	559	225	0.90	-1.10	
		ASN	3	0.0	122	73	73	73	73	139	-	-	-	153	80	43	-1.53	-	
		JS	2	0.0	244	89	89	89	89	244	-	-	-	399	310	219	-	-	
NET(N=1, 0.0%, 236 UFP) AG(N=1, 0.0%, 84 UFP) ABI(N=1, 0.0%, 59 UFP) COB(N=1, 0.0%, 632 UFP)																			
IBM(N=1, 0.0%, 372 UFP) SQL(N=1, 0.0%, 169 UFP) VB(N=1, 0.0%, 361 UFP) XML(N=1, 0.0%, 169 UFP)																			
TD		E	57	0.8	220	47	54	68	90	179	303	504	588	695	648	159	1.31	1.14	
		N	14	0.2	332	24	24	57	118	301	402	899	-	1,337	1,313	319	2.61	8.38	
R(N=1, 0.0%, 112 UFP)																			
PD		MP	46	0.6	268	47	65	75	116	212	343	551	656	1,337	1,290	226	2.63	10.30	
		MF	19	0.3	229	54	54	54	117	235	303	493	-	632	578	150	1.24	1.81	
		PC	6	0.1	77	24	24	24	50	80	99	-	-	127	103	35	-0.14	0.64	
MR(N=1, 0.0%, 169 UFP)																			
TE	9-14	21	0.3	270	89	92	117	133	235	356	559	627	632	543	154	1.01	0.40		
	5-8	19	0.3	138	61	61	70	75	90	184	254	-	372	311	83	1.47	2.16		
	ND	13	0.2	202	54	54	54	94	155	238	529	-	553	499	156	1.52	1.63		
	15-20	8	0.1	338	121	121	121	265	334	426	-	-	550	429	128	0.02	0.85		
	3-4	6	0.1	104	24	24	24	41	75	186	-	-	236	212	82	0.97	-0.39		
	21-30	3	0.0	451	298	298	298	298	361	-	-	-	695	397	213	1.56	-		
	31-40	2	0.0	859	381	381	381	381	859	-	-	-	1,337	956	676	-	-		
AN	2008	29	0.4	234	54	54	61	97	169	352	553	608	632	578	173	1.01	-0.14		

DS	Vr	Niv	N	%	Mean	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rng	SD	Skw	Krt	
			2009	25	0.3	255	73	74	76	104	211	313	449	1,101	1,337	1,264	255	3.45	14.28
			2010	13	0.2	269	47	47	64	146	219	380	601	-	695	648	177	1.15	1.57
			2011	5	0.1	132	24	24	24	42	70	253	-	-	337	313	127	1.40	1.47
	AR	CS		39	0.5	275	61	74	84	117	212	361	550	695	1,337	1,276	237	2.63	9.77
		SA		20	0.3	221	24	26	54	97	223	291	481	579	583	559	145	0.92	0.95
		MW		8	0.1	142	47	47	47	74	109	216	-	-	337	290	99	1.30	1.04
		MC		5	0.1	205	70	70	70	87	128	361	-	-	553	483	198	2.05	4.34
			AR(CS=Cliente servidor, SA=Stand alone, MC=Multicapa, MW=Multicapa Web, ND=No definido)																
	TR	M1		33	0.5	184	103	109	117	131	180	236	250	295	298	195	55	0.31	-0.87
		M2		20	0.3	430	303	303	312	333	390	536	627	692	695	392	118	0.88	-0.29
		S		17	0.2	74	47	47	53	60	75	89	90	-	90	43	15	-0.50	-1.07
			XS(N=1, 0.0%, 24 UFP) L(N=1, 0.0%, 1,337 UFP)																
	TL	3GL		112	0.6	266	8	47	71	102	199	328	605	861	1,337	1,329	241	1.97	4.38
		ApG		52	0.3	155	6	15	32	54	99	239	305	495	635	629	135	1.52	2.64
		4GL		36	0.2	315	59	71	115	163	232	392	754	908	1,048	989	240	1.68	2.30
		ND		2	0.0	361	349	349	349	349	361	-	-	-	372	23	16	-	-
	LP	COO		52	0.3	155	6	15	32	54	99	239	305	495	635	629	135	1.52	2.64
		C#		38	0.2	319	47	74	88	127	262	335	837	979	1,337	1,290	279	2.09	4.58
		PL		36	0.2	262	49	59	74	118	199	310	606	997	1,018	969	235	2.11	4.39
		JV		29	0.1	217	8	11	24	64	128	374	583	632	649	641	207	1.04	-0.38
		ABA		17	0.1	367	122	122	168	205	279	429	865	-	883	761	234	1.40	0.94
		DST		4	0.0	461	117	117	117	171	340	873	-	-	1,048	931	405	1.58	2.96
		JS		4	0.0	232	89	89	89	100	220	376	-	-	399	310	146	0.25	-3.80
		VB		4	0.0	213	144	144	144	148	174	317	-	-	361	217	100	1.83	3.40
		ASN		3	0.0	122	73	73	73	73	139	-	-	-	153	80	43	-1.53	-
		PB		3	0.0	461	285	285	285	285	401	-	-	-	696	411	212	1.17	-
		COB		2	0.0	361	90	90	90	90	361	-	-	-	632	542	383	-	-
		SQL		2	0.0	187	169	169	169	169	187	-	-	-	204	35	25	-	-
			NET(N=1, 0.0%, 236 UFP) AG(N=1, 0.0%, 84 UFP) ABI(N=1, 0.0%, 59 UFP) ASP(N=1, 0.0%, 109 UFP)																
			C++(N=1, 0.0%, 146 UFP) IBM(N=1, 0.0%, 372 UFP) XML(N=1, 0.0%, 169 UFP) ND(N=1, 0.0%, 119 UFP)																
	TD	E		161	0.8	210	6	34	50	84	154	271	447	635	960	954	180	1.82	3.65
		N		39	0.2	394	24	51	90	228	303	430	993	1,048	1,337	1,313	311	1.46	1.58
		R		2	0.0	364	112	112	112	112	364	-	-	-	615	503	356	-	-
	PD	MP		110	0.5	296	8	49	75	119	235	352	690	918	1,337	1,329	252	1.80	3.40
		MF		76	0.4	181	6	26	40	69	145	252	349	520	736	730	151	1.58	2.85
		PC		14	0.1	234	24	24	42	85	136	314	769	-	923	899	254	1.96	3.57
		MR		2	0.0	161	153	153	153	153	161	-	-	-	169	16	11	0.60	1.15
	TE	5-8		56	0.3	236	51	69	74	91	162	275	496	835	1,048	997	214	2.20	4.88
		9-14		36	0.2	309	89	104	117	181	287	395	598	730	923	834	182	1.46	2.77
		3-4		18	0.1	149	24	24	45	82	138	218	267	-	297	273	80	0.20	-0.89
		15-20		16	0.1	480	121	121	207	277	336	706	970	-	993	872	274	0.81	-0.63
		21-30		3	0.0	451	298	298	298	298	361	-	-	-	695	397	213	1.56	-
		31-40		2	0.0	859	381	381	381	381	859	-	-	-	1,337	956	676	-	-
		ND		69	0.3	173	6	14	32	55	103	238	410	567	1,018	1,012	177	2.33	7.25
			2(N=1, 0.0%, 41 UFP) 61-70(N=1, 0.0%, 153 UFP)																
	AN	2007		46	0.2	287	33	43	68	113	193	330	895	981	1,048	1,015	278	1.74	1.96
		2005		43	0.2	283	41	43	55	122	252	307	677	809	1,018	977	221	1.60	2.55
		2008		35	0.2	199	6	8	28	61	128	294	517	593	632	626	176	1.12	0.24
		2006		34	0.2	203	13	15	30	73	126	293	558	653	708	695	187	1.37	1.09
		2009		26	0.1	251	73	74	76	110	198	312	432	1,062	1,337	1,264	250	3.53	14.88
		2010		13	0.1	269	47	47	64	146	219	380	601	-	695	648	177	1.15	1.57
		2011		5	0.0	132	24	24	24	42	70	253	-	-	337	313	127	1.40	1.47
	AR	CS		95	0.5	332	41	76	90	132	264	399	756	967	1,337	1,296	264	1.61	2.39
		SA		58	0.3	167	6	16	33	65	139	236	321	504	635	629	139	1.55	2.60
		ND		24	0.1	202	32	34	45	75	163	285	460	697	736	704	172	1.71	3.38
		MC		16	0.1	153	8	8	12	60	116	174	453	-	553	545	144	1.83	3.43

DS202

DS	Vr	Niv	N	%	Mean	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rng	SD	Skw	Krt
	MW		9	0.0	144	47	47	47	74	127	195	-	-	337	290	92	1.28	1.36
TR	M1		92	0.5	192	102	110	118	139	185	241	276	291	299	197	59	0.20	-1.23
	M2		51	0.3	497	303	308	313	335	411	632	853	938	993	690	196	1.08	0.11
	S		49	0.2	68	32	36	41	53	72	85	90	97	99	67	19	-0.20	-1.11
	XS		5	0.0	19	13	13	13	14	16	26	-	-	28	15	7	0.69	-2.07
	L		3	0.0	1,134	1,018	1,018	1,018	1,018	1,048	-	-	-	1,337	319	176	1.68	-
	XXS		2	0.0	7	6	6	6	6	7	-	-	-	8	2	1	-	-

TL=Tipo de lenguaje (3GL, ApG, 4GL). LP=Lenguaje programación (JV=Java, COO=COOLGen, ASN=ASPNet, C#=C#, JS=JavaScript, ABA=ABAP, PL= PLI, VB=VisualBasic, PB=PowerBuilder, ASP=ASP, SQL=SQL, NET=VisualStudioNet, DST=Datastage, NET=Net, IBM=IBMWTX, XML=XML, COB=COBOL, ABI=ABINITIO, AG=AG, C++=C++). TD=Tipo de desarrollo (E=Enhancement, N=New Development, R= Re-development). PD=Plataforma de desarrollo (MP=Multi-Platform, MF=Main Frame, PC=PC, MR=Mid-Range). TE=Tamaño del equipo de desarrollo (2, 3-4, 5-8, 9-14, 15-20, 21-30, 31-40, 61-70). AN=Año (2005, 2006, 2007, 2008, 2009, 2010, 2011). AR=Arquitectura (CS=Cliente servidor, SA=Stand alone, MC=Multicapa, MW=Multicapa Web, ND=No definido). TR=Tamaño relativo (1.XXS, 2.XS, 3.S, 4.M1, 5.M2, 6.L).

Anexo 6 Tamaño funcional sin ajustar COSMIC FFP por variables de contexto.

El Anexo 6 lista el detalle de las estadísticas descriptivas para cada una de las variables de contexto de los subconjuntos de datos.

DS	Vr	Niv	N	%	Mean	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rng	SD	Skw	Krt		
DS63	TL	3GL	43	0.7	200	3	3	5	33	98	293	572	819	912	909	240	1.58	1.79		
		4GL	20	0.3	50	2	2	2	3	6	68	169	428	441	439	103	3.22	11.46		
	LP	COB	22	0.3	137	3	3	3	6	50	246	512	576	579	576	179	1.50	1.35		
		ABA	14	0.2	5	2	2	2	3	4	6	9	-	9	7	2	0.66	-0.76		
	JV	12	0.2	228	14	14	28	89	105	253	816	-	826	812	279	1.83	2.02			
	C#	5	0.1	338	45	45	45	64	169	698	-	-	912	867	364	1.25	0.63			
	NET	4	0.1	160	48	48	48	55	75	350	-	-	441	393	188	1.97	3.92			
	VB	3	0.0	380	226	226	226	226	448	-	-	-	467	241	134	-1.69	-			
	SL	2	0.0	147	118	118	118	118	147	-	-	-	175	57	40	-	-			
	C++(N=1, 0.0%, 23 FP)																			
	TD	E	43	0.7	97	2	2	3	5	11	118	356	538	579	577	153	1.98	3.28		
		N	15	0.2	178	23	23	35	62	98	175	595	-	826	803	209	2.49	6.65		
		R	5	0.1	554	118	118	118	293	483	852	-	-	912	794	311	-0.35	-0.55		
	PD	MP	29	0.5	176	2	2	2	4	48	294	561	703	826	824	224	1.31	1.11		
		MF	16	0.3	47	3	3	3	5	22	80	156	-	245	242	64	2.20	5.58		
		PC	14	0.2	240	14	14	19	67	98	452	852	-	912	898	296	1.51	1.07		
		ND	3	0.0	100	62	62	62	62	110	-	-	-	128	66	34	-1.21	-		
	MR(N=1, 0.0%, 86 FP)																			
	TE	1	17	0.3	5	2	2	2	3	5	6	9	-	9	7	2	0.54	-0.72		
		5-8	13	0.2	159	43	43	55	88	110	216	407	-	483	440	120	1.90	3.81		
		2	12	0.2	96	3	3	3	6	24	73	589	-	791	788	222	3.31	11.21		
		3-4	8	0.1	220	23	23	23	50	97	375	-	-	826	803	279	1.86	3.07		
		9-14	5	0.1	266	62	62	62	75	99	541	-	-	912	850	363	2.17	4.76		
21-30		3	0.0	513	398	398	398	398	561	-	-	-	579	181	100	-1.67	-			
15-20		2	0.0	272	250	250	250	250	272	-	-	-	294	44	31	-	-			
ND	2	0.0	337	226	226	226	226	337	-	-	-	448	222	-	-	-				
31-40(N=1, 0.0%, 467 FP)																				
AN	2008	16	0.3	112	3	3	4	6	33	98	479	-	912	909	226	3.35	11.99			
	2010	15	0.2	57	2	2	2	3	5	7	322	-	-	789	203	3.87	14.99			
	2009	9	0.1	111	3	3	3	24	48	166	-	-	483	480	157	2.08	4.14			
	2006	7	0.1	321	88	88	88	187	293	467	-	-	561	473	165	0.11	-0.90			
	2005	6	0.1	385	99	99	99	123	337	641	-	-	826	727	286	0.64	-0.90			
	2007	6	0.1	164	62	62	62	71	119	237	-	-	441	379	141	2.06	4.49			
	2011	2	0.0	147	118	118	118	118	147	-	-	-	175	57	40	-	-			
	2012	2	0.0	64	45	45	45	45	64	-	-	-	82	37	26	-	-			
AR	MC	24	0.4	87	2	2	2	3	8	183	294	410	448	446	129	1.46	1.23			
	SA	20	0.3	73	3	3	3	6	34	96	232	471	483	480	113	2.87	9.29			
	CS	14	0.2	324	45	45	60	80	153	566	869	-	912	867	298	0.89	-0.54			
	ND	3	0.0	100	62	62	62	62	110	-	-	-	128	66	-	-1.21	-			
MW	2	0.0	616	441	441	441	441	616	-	-	-	791	350	247	-	-				
TR	XXS	21	0.3	5	2	2	2	3	5	6	9	9	9	7	2	0.54	-0.74			
	S	16	0.3	70	33	33	34	46	75	90	98	-	99	66	24	-0.27	-1.46			
	M1	13	0.2	188	110	110	113	123	175	248	294	-	294	184	67	0.41	-1.32			
	M2	10	0.2	591	398	398	402	446	522	800	903	-	912	514	185	0.83	-0.95			
	XS	3	0.0	16	11	11	11	11	14	-	-	-	23	12	6	1.29	-			
TL	3GL	34	0.6	135	3	3	5	13	84	197	423	566	579	576	160	1.57	1.77			

DS	Vr	Niv	N	%	Mean	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rng	SD	Skw	Krt
		4GL	19	0.4	29	2	2	2	3	6	48	118	-	175	173	48	2.07	3.91
LP		COB	20	0.4	134	3	3	3	6	39	234	545	578	579	576	186	1.53	1.27
		ABA	14	0.3	5	2	2	2	3	4	6	9	-	9	7	2	0.66	-0.76
		JV	8	0.2	116	14	14	14	68	105	130	-	-	294	280	82	1.57	3.85
		C#	3	0.1	99	45	45	45	45	82	-	-	-	169	124	64	1.10	-
		NET	3	0.1	66	48	48	48	48	74	-	-	-	75	27	15	-1.72	-
		VB	2	0.0	337	226	226	226	226	337	-	-	-	448	222	157	-	-
		SL	2	0.0	147	118	118	118	118	147	-	-	-	175	57	40	-	-
		C++(N=1, 0.0%, 23 FP)																
TD		E	41	0.8	93	2	2	3	5	9	109	377	550	579	577	155	2.05	3.47
		N	11	0.2	110	23	23	27	48	86	169	270	-	294	271	79	1.33	1.91
		R(N=1, 0.0%, 118 FP)																
PD		MP	27	0.5	141	2	2	2	3	9	250	471	572	579	577	183	1.20	0.42
		MF	14	0.3	30	3	3	3	5	10	47	104	-	118	115	36	1.51	1.60
		PC	8	0.2	68	14	14	14	29	75	95	-	-	131	117	39	0.08	-0.54
		ND	3	0.1	100	62	62	62	62	110	-	-	-	128	66	34	-1.21	-
		MR(N=1, 0.0%, 86 FP)																
TE		1	17	0.3	5	2	2	2	3	5	6	9	-	9	7	2	0.54	-0.72
		5-8	9	0.2	127	43	43	43	80	110	159	-	-	293	250	74	1.53	2.71
		2	11	0.2	32	3	3	3	5	14	45	111	-	118	115	37	1.55	1.86
		3-4	6	0.1	83	23	23	23	42	66	132	-	-	175	152	55	1.00	0.43
		9-14	3	0.1	110	62	62	62	62	99	-	-	-	169	107	54	0.87	-
		21-30	3	0.1	513	398	398	398	398	561	-	-	-	579	181	100	-1.67	-
		15-20	2	0.0	272	250	250	250	250	272	-	-	-	294	44	31	-	-
		ND	2	0.0	337	226	226	226	226	337	-	-	-	448	222	-	-	-
AN		2008	13	0.2	53	3	3	3	6	11	83	224	-	294	291	82	2.44	6.60
		2010	14	0.3	5	2	2	2	3	4	6	9	-	9	7	2	0.66	-0.76
		2009	7	0.1	39	3	3	3	14	34	57	-	-	86	83	28	0.48	0.18
		2006	5	0.1	338	187	187	187	219	293	480	-	-	561	374	147	0.95	0.34
		2005	5	0.1	297	99	99	99	115	226	514	-	-	579	480	209	0.61	-1.94
		2007	5	0.1	109	62	62	62	68	110	149	-	-	169	107	43	0.45	-0.83
		2011	2	0.0	147	118	118	118	118	147	-	-	-	175	57	40	-	-
		2012	2	0.0	64	45	45	45	45	64	-	-	-	82	37	26	-	-
AR		MC	23	0.4	87	2	2	2	3	7	187	294	417	448	446	131	1.43	1.07
		SA	16	0.3	33	3	3	3	5	17	54	98	-	118	115	37	1.23	0.49
		CS	11	0.2	212	45	45	51	75	118	398	575	-	579	534	201	1.25	-0.10
		ND	3	0.1	100	62	62	62	62	110	-	-	-	128	66	-	-1.21	-
TR		XXS	21	0.4	5	2	2	2	3	5	6	9	9	9	7	2	0.54	-0.74
		S	13	0.2	64	33	33	33	44	62	84	95	-	99	66	22	0.07	-1.39
		M1	12	0.2	183	110	110	112	121	172	244	294	-	294	184	68	0.63	-1.04
		M2	4	0.1	497	398	398	398	411	505	575	-	-	579	181	88	-0.24	-4.26
		XS	3	0.1	16	11	11	11	11	14	-	-	-	23	12	6	1.29	-

TL=Tipo de lenguaje (3GL, ApG, 4GL). LP=Lenguaje programación (JV=Java, COO=COOLGen, ASN=ASPNet, C#=C#, JS=JavaScript, ABA=ABAP, PL=PLI, VB=VisualBasic, PB=PowerBuilder, ASP=ASP, SQL=SQL, NET=VisualStudioNet, DST=Datastage, NET=Net, IBM=IBMWTX, XML=XML, COB=COBOL, ABI=ABINITIO, AG=AG, C++=C++). TD=Tipo de desarrollo (E=Enhancement, N=New Development, R= Re-development). PD=Plataforma de desarrollo (MP=Multi-Platform, MF=Main Frame, PC=PC, MR=Mid-Range). TE=Tamaño del equipo de desarrollo (2, 3-4, 5-8, 9-14, 15-20, 21-30, 31-40, 61-70). AN=Año (2005, 2006, 2007, 2008, 2009, 2010, 2011). AR=Arquitectura (CS=Cliente servidor, SA=Stand alone, MC=Multicapa, MW=Multicapa Web, ND=No definido). TR=Tamaño relativo (1.XXS, 2.XS, 3.S, 4.M1, 5.M2, 6.L).

Anexo 7 Estadística descriptiva de los resultados de la evaluación de exactitud de los modelos de estimación de esfuerzo.

El Anexo 7 lista el detalle de las estadísticas descriptivas de los resultados de la exactitud de los modelos para cada uno de los sub conjuntos de datos (DS) y los grupos evaluados (Grupo). Las métricas de exactitud presentadas son la magnitud del error relativo (MRE), el coeficiente de correlación de Spearman (SP), el número de predicciones de los valores reales (Pred₂₅) y la exactitud estandarizada (SA).

DS	Grupo	Métrica	N	Prom	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rango	Std.D.	Skew	Kurt
DS202	1	SP	554	0.646	0.256	0.391	0.418	0.640	0.704	0.720	0.723	0.725	0.732	0.476	0.116	-1.509	0.711
		MRE	600	1.292	0.616	0.669	0.786	0.906	1.164	1.657	1.978	2.036	2.294	1.678	0.436	0.292	-1.174
		MdMRE	600	1.279	0.785	0.850	0.899	0.982	1.168	1.516	1.788	1.899	2.410	1.625	0.356	0.733	-0.237
		SA	600	41.189	13.556	27.532	29.132	34.663	42.515	48.244	51.274	53.415	58.794	45.237	8.740	-0.401	-0.636
		Pred ₂₅	600	0.221	0.139	0.149	0.154	0.181	0.223	0.261	0.283	0.293	0.315	0.176	0.047	-0.009	-1.172
	2	SP	557	0.625	-0.008	0.349	0.371	0.520	0.691	0.746	0.768	0.785	0.807	0.815	0.155	-1.010	0.059
		MRE	600	1.238	0.006	0.677	0.724	0.846	1.140	1.627	1.885	2.017	2.358	2.353	0.438	0.278	-0.781
		MdMRE	600	1.191	0.006	0.771	0.810	0.905	1.121	1.446	1.708	1.789	2.258	2.252	0.344	0.326	-0.013
		SA	600	43.601	3.605	29.112	32.016	35.583	44.695	51.778	54.990	56.681	61.469	57.864	9.605	-0.613	0.559
		Pred ₂₅	600	0.234	0.001	0.149	0.172	0.189	0.236	0.274	0.305	0.316	0.362	0.361	0.054	-0.356	0.557
	3	SP	556	0.626	-0.010	0.338	0.362	0.550	0.689	0.746	0.769	0.784	0.806	0.815	0.157	-1.082	0.250
		MRE	600	1.242	0.014	0.677	0.712	0.844	1.149	1.633	1.890	2.021	2.530	2.516	0.446	0.257	-0.755
		MdMRE	600	1.196	0.014	0.764	0.814	0.910	1.122	1.453	1.697	1.794	2.234	2.220	0.349	0.240	0.009
		SA	600	43.601	3.496	29.094	31.698	35.361	44.150	51.845	55.141	57.083	63.043	59.547	9.873	-0.616	0.624
		Pred ₂₅	600	0.233	0.002	0.152	0.174	0.188	0.233	0.272	0.305	0.318	0.359	0.356	0.055	-0.343	0.729
	4	SP	557	0.602	0.151	0.298	0.348	0.518	0.659	0.719	0.752	0.762	0.789	0.637	0.156	-1.077	0.056
		MRE	600	1.296	0.055	0.669	0.720	0.896	1.212	1.712	1.888	2.019	5.662	5.607	0.501	1.344	8.665
		MdMRE	600	1.223	0.057	0.779	0.821	0.960	1.149	1.497	1.700	1.790	2.773	2.716	0.359	0.490	0.734
		SA	600	40.089	-79.680	24.640	27.812	31.610	41.295	48.476	53.761	55.602	61.568	141.249	11.850	-2.159	17.119
		Pred ₂₅	600	0.226	0.010	0.150	0.166	0.183	0.233	0.258	0.302	0.313	0.351	0.341	0.053	-0.228	0.349
5	SP	556	0.602	0.155	0.277	0.335	0.536	0.654	0.721	0.752	0.764	0.795	0.640	0.155	-1.069	0.011	
	MRE	600	1.294	0.061	0.671	0.721	0.904	1.212	1.723	1.906	2.015	2.779	2.718	0.461	0.198	-0.867	
	MdMRE	600	1.228	0.059	0.773	0.822	0.974	1.163	1.508	1.695	1.766	2.493	2.434	0.348	0.288	0.172	
	SA	600	40.013	1.935	23.883	27.265	32.068	41.176	48.149	52.912	56.126	60.945	59.010	10.645	-0.542	0.358	
	Pred ₂₅	600	0.225	0.010	0.151	0.168	0.183	0.230	0.256	0.300	0.314	0.345	0.336	0.052	-0.142	0.440	
6	SP	556	0.574	0.135	0.309	0.353	0.490	0.621	0.680	0.717	0.733	0.769	0.634	0.145	-1.111	0.507	
	MRE	600	1.311	0.039	0.712	0.782	0.911	1.238	1.721	1.892	2.023	2.735	2.695	0.451	0.194	-0.781	
	MdMRE	600	1.252	0.036	0.829	0.871	0.980	1.194	1.531	1.678	1.774	2.573	2.537	0.342	0.353	0.678	
	SA	600	39.594	-1.815	25.340	27.993	32.251	41.166	47.209	51.604	53.461	61.020	62.835	10.077	-0.780	1.049	

DS	Grupo	Métrica	N	Prom	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rango	Std.D.	Skew	Kurt
7	Pred ₂₅		600	0.220	0.009	0.151	0.168	0.181	0.225	0.254	0.279	0.293	0.329	0.320	0.047	-0.355	0.922
		SP	557	0.574	0.118	0.313	0.347	0.490	0.619	0.684	0.722	0.741	0.776	0.658	0.147	-1.038	0.295
	MRE	600	1.317	0.068	0.695	0.766	0.917	1.239	1.742	1.892	2.025	2.743	2.675	0.453	0.169	-0.840	
	MdMRE	600	1.261	0.065	0.806	0.877	1.001	1.216	1.539	1.695	1.772	2.603	2.538	0.340	0.294	0.622	
8	SA		600	39.203	-22.451	25.488	28.058	31.764	40.343	46.302	51.779	54.567	60.606	83.058	10.112	-0.827	2.419
		Pred ₂₅	600	0.218	0.014	0.150	0.165	0.181	0.223	0.247	0.283	0.294	0.332	0.319	0.047	-0.172	0.929
	SP	560	0.584	0.114	0.280	0.353	0.531	0.632	0.681	0.705	0.712	0.730	0.616	0.139	-1.434	1.199	
	MRE	600	1.338	0.616	0.677	0.825	0.961	1.293	1.735	1.925	2.034	5.225	4.609	0.456	1.163	7.292	
9	MdMRE		600	1.278	0.782	0.851	0.900	1.008	1.212	1.516	1.764	1.818	2.467	1.685	0.333	0.673	-0.057
		SA	600	37.978	-15.459	23.508	26.831	30.771	39.420	44.654	48.924	53.082	56.429	71.888	9.446	-0.596	1.227
	Pred ₂₅	600	0.218	0.141	0.149	0.164	0.183	0.221	0.247	0.276	0.287	0.314	0.173	0.042	0.117	-0.866	
	SP	560	0.560	0.124	0.321	0.354	0.486	0.594	0.669	0.691	0.700	0.715	0.592	0.133	-1.180	0.877	
10	MRE		600	1.341	0.630	0.681	0.844	0.965	1.324	1.735	1.928	2.031	2.125	1.494	0.422	0.127	-1.271
		MdMRE	600	1.295	0.794	0.847	0.907	1.028	1.231	1.558	1.752	1.815	2.374	1.580	0.327	0.475	-0.589
	SA	600	37.832	4.841	23.934	27.745	30.793	39.422	44.325	48.186	52.824	57.472	52.632	8.926	-0.329	0.040	
	Pred ₂₅	600	0.215	0.139	0.151	0.162	0.177	0.214	0.244	0.275	0.288	0.309	0.170	0.042	0.234	-0.897	
11	SP		556	0.613	0.077	0.311	0.344	0.534	0.671	0.731	0.758	0.769	0.803	0.727	0.156	-1.036	-0.186
		MRE	600	1.287	0.031	0.678	0.709	0.910	1.197	1.717	1.909	2.020	4.016	3.985	0.469	0.464	0.676
	MdMRE	600	1.209	0.027	0.775	0.812	0.942	1.140	1.474	1.703	1.794	2.488	2.461	0.351	0.375	0.033	
	SA	600	41.488	-37.818	24.933	27.896	33.000	42.731	50.243	54.470	56.748	61.110	98.928	11.011	-0.957	3.738	
12	Pred ₂₅		600	0.230	0.007	0.151	0.169	0.189	0.233	0.268	0.305	0.313	0.355	0.349	0.052	-0.174	0.149
		SP	556	0.614	0.093	0.310	0.341	0.547	0.669	0.732	0.761	0.772	0.800	0.707	0.156	-1.053	-0.151
	MRE	600	1.299	0.043	0.674	0.711	0.922	1.195	1.732	1.906	2.017	4.989	4.945	0.476	0.896	4.651	
	MdMRE	600	1.217	0.037	0.773	0.816	0.952	1.137	1.477	1.670	1.782	2.477	2.440	0.341	0.364	0.043	
12	SA		600	41.106	-58.731	24.370	27.672	33.118	42.181	50.022	54.080	56.355	64.054	122.785	11.203	-1.488	9.736
		Pred ₂₅	600	0.228	0.010	0.151	0.170	0.188	0.227	0.262	0.303	0.313	0.356	0.346	0.051	0.005	0.119
	SP	560	0.623	0.288	0.357	0.369	0.618	0.664	0.702	0.720	0.725	0.734	0.446	0.120	-1.481	0.805	
	MRE	600	1.322	0.605	0.682	0.801	0.932	1.250	1.713	1.939	2.030	3.144	2.539	0.451	0.327	-0.760	
MdMRE	600	1.267	0.794	0.866	0.900	0.974	1.158	1.523	1.762	1.842	2.563	1.769	0.348	0.770	-0.015		

DS	Grupo	Métrica	N	Prom	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rango	Std.D.	Skew	Kurt
		SA	600	39.868	-15.307	25.419	27.733	31.453	41.497	47.740	51.626	54.279	57.973	73.281	10.042	-0.666	1.177
		Pred ₂₅	600	0.221	0.141	0.154	0.167	0.183	0.226	0.259	0.274	0.284	0.315	0.174	0.042	0.020	-1.040
	13	SP	560	0.612	0.237	0.322	0.366	0.601	0.675	0.689	0.693	0.695	0.705	0.467	0.126	-1.566	0.844
		MRE	600	1.341	0.646	0.698	0.858	0.989	1.231	1.717	1.953	2.023	2.322	1.676	0.421	0.188	-1.294
		MdMRE	600	1.305	0.830	0.879	0.930	1.022	1.188	1.545	1.778	1.908	2.492	1.662	0.340	0.707	-0.215
		SA	600	39.245	13.472	27.193	29.018	32.203	40.096	46.160	48.630	51.212	55.104	41.632	8.219	-0.293	-0.749
		Pred ₂₅	600	0.217	0.140	0.148	0.152	0.181	0.229	0.250	0.271	0.280	0.298	0.158	0.043	-0.113	-1.201
	14	SP	560	0.603	0.296	0.346	0.368	0.578	0.655	0.687	0.702	0.707	0.716	0.421	0.119	-1.316	0.289
		MRE	600	1.316	0.624	0.694	0.823	0.946	1.243	1.681	1.932	2.016	2.280	1.657	0.423	0.200	-1.253
		MdMRE	600	1.278	0.802	0.874	0.924	1.004	1.172	1.504	1.786	1.952	2.564	1.762	0.350	0.815	-0.080
		SA	600	39.646	11.350	26.496	28.275	31.498	41.074	46.726	50.166	53.096	57.097	45.746	8.884	-0.264	-0.851
		Pred ₂₅	600	0.218	0.135	0.151	0.161	0.183	0.225	0.249	0.269	0.286	0.302	0.167	0.041	-0.055	-1.018
	15	SP	560	0.567	0.126	0.324	0.360	0.525	0.605	0.660	0.683	0.690	0.712	0.586	0.129	-1.462	1.654
		MRE	600	1.359	0.638	0.691	0.852	0.962	1.347	1.732	1.935	2.025	6.600	5.962	0.475	2.306	23.214
		MdMRE	600	1.294	0.762	0.836	0.920	1.017	1.241	1.566	1.746	1.801	2.475	1.713	0.328	0.458	-0.593
		SA	600	37.755	-99.458	24.930	27.973	30.813	39.169	44.305	48.516	52.452	55.692	155.150	10.338	-4.043	50.763
		Pred ₂₅	600	0.218	0.143	0.154	0.166	0.179	0.218	0.249	0.276	0.293	0.342	0.199	0.043	0.310	-0.729
	16	SP	560	0.545	0.109	0.330	0.361	0.512	0.585	0.630	0.644	0.651	0.670	0.561	0.117	-1.521	1.801
		MRE	600	1.360	0.661	0.719	0.875	0.989	1.340	1.728	1.922	2.024	2.225	1.563	0.405	0.115	-1.237
		MdMRE	600	1.308	0.826	0.925	0.954	1.040	1.236	1.561	1.759	1.815	2.236	1.410	0.315	0.561	-0.639
		SA	600	37.983	12.065	25.020	28.309	31.593	39.155	43.870	47.295	50.218	53.761	41.696	7.922	-0.304	-0.444
		Pred ₂₅	600	0.214	0.143	0.153	0.167	0.179	0.215	0.247	0.267	0.273	0.306	0.162	0.039	0.104	-1.092
	17	SP	560	0.542	0.228	0.305	0.342	0.521	0.585	0.616	0.629	0.635	0.657	0.429	0.106	-1.436	0.852
		MRE	600	1.376	0.674	0.720	0.864	1.036	1.322	1.739	1.948	2.033	2.231	1.557	0.409	0.101	-1.241
		MdMRE	600	1.344	0.897	0.968	1.005	1.063	1.251	1.579	1.802	1.920	2.539	1.641	0.330	0.812	0.029
		SA	600	37.026	11.608	24.782	26.749	30.360	38.667	43.155	46.814	49.611	54.533	42.925	8.009	-0.312	-0.515
		Pred ₂₅	600	0.206	0.138	0.151	0.159	0.178	0.211	0.235	0.246	0.252	0.269	0.130	0.033	-0.215	-1.177
	18	SP	560	0.516	0.102	0.297	0.338	0.494	0.567	0.602	0.615	0.621	0.636	0.534	0.119	-1.548	1.860
		MRE	600	1.437	0.693	0.749	0.883	1.139	1.420	1.754	1.977	2.040	2.258	1.564	0.386	-0.096	-1.002

DS	Grupo	Métrica	N	Prom	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rango	Std.D.	Skew	Kurt
		MdMRE	600	1.379	0.851	0.940	0.999	1.153	1.347	1.563	1.776	1.973	2.657	1.806	0.311	0.818	0.918
		SA	600	36.355	7.233	24.401	28.415	31.560	36.592	41.484	44.856	48.067	54.009	46.776	7.228	-0.437	0.630
		Pred ₂₅	600	0.203	0.142	0.152	0.165	0.179	0.200	0.226	0.248	0.260	0.279	0.137	0.031	0.212	-0.679
	19	SP	560	0.502	0.080	0.208	0.298	0.514	0.555	0.570	0.580	0.584	0.596	0.515	0.121	-1.941	2.611
		MRE	600	1.475	0.712	0.763	0.882	1.214	1.442	1.818	1.973	2.039	2.286	1.573	0.387	-0.206	-0.915
		MdMRE	600	1.441	0.969	1.012	1.058	1.223	1.395	1.610	1.794	2.061	2.483	1.514	0.299	0.800	0.609
		SA	600	34.681	11.045	21.440	27.799	30.508	34.790	38.935	43.271	46.791	49.938	38.893	6.829	-0.390	0.641
		Pred ₂₅	600	0.187	0.134	0.148	0.153	0.167	0.180	0.206	0.234	0.246	0.270	0.136	0.029	0.740	-0.189
	20	SP	557	0.529	0.212	0.270	0.427	0.478	0.556	0.588	0.625	0.654	0.750	0.538	0.102	-1.253	1.819
		MRE	600	1.357	0.002	0.761	0.823	1.001	1.361	1.670	1.925	2.037	2.249	2.247	0.404	-0.038	-0.617
		MdMRE	600	1.310	0.003	0.906	0.944	1.038	1.324	1.513	1.740	1.814	2.332	2.329	0.324	0.188	0.842
		SA	600	40.285	3.488	27.372	29.562	35.198	41.153	46.665	50.273	52.034	57.394	53.906	8.315	-0.758	1.049
		Pred ₂₅	600	0.211	0.000	0.150	0.170	0.188	0.211	0.239	0.255	0.264	0.313	0.313	0.036	-0.922	3.729
	21	SP	297	0.525	0.354	0.373	0.404	0.538	0.545	0.549	0.555	0.557	0.566	0.212	0.055	-2.277	3.517
		MRE	600	2.173	0.761	0.846	0.940	1.174	1.672	2.025	2.071	2.160	20.019	19.259	3.235	5.063	24.302
		MdMRE	600	2.248	0.969	1.008	1.104	1.169	1.705	1.784	2.000	2.249	24.205	23.236	3.958	5.157	24.928
		SA	600	15.486	-491.632	-22.836	14.626	27.153	31.127	42.249	45.993	46.920	50.208	541.840	92.451	-5.048	24.163
		Pred ₂₅	600	0.173	0.000	0.005	0.144	0.149	0.170	0.222	0.232	0.235	0.241	0.241	0.057	-1.535	2.784
DS53	1	SP	546	0.689	-0.026	0.481	0.510	0.632	0.717	0.769	0.817	0.834	0.863	0.889	0.120	-1.418	3.522
		MRE	600	6.736	0.000	1.002	1.522	2.531	5.325	9.557	14.029	18.807	20.797	20.797	5.174	0.987	0.184
		MdMRE	600	2.948	0.000	0.691	0.943	1.415	2.323	3.856	6.376	8.037	9.262	9.262	2.100	1.229	0.893
		SA	600	52.970	2.495	19.862	25.817	43.339	54.694	67.757	75.316	77.242	79.693	77.198	17.673	-0.596	-0.192
		Pred ₂₅	600	0.133	0.000	0.034	0.043	0.097	0.129	0.166	0.204	0.257	0.352	0.352	0.062	0.581	0.974
	2	SP	546	0.686	-0.044	0.454	0.501	0.636	0.721	0.770	0.818	0.833	0.862	0.906	0.127	-1.604	4.326
		MRE	600	6.821	0.000	1.052	1.553	2.551	5.274	9.834	14.591	19.261	20.815	20.815	5.275	1.008	0.185
		MdMRE	600	2.981	0.000	0.706	0.926	1.368	2.311	3.906	6.186	8.126	9.561	9.561	2.140	1.246	0.938
		SA	600	52.918	2.983	20.063	25.612	42.741	55.090	67.143	75.113	76.829	79.855	76.872	17.647	-0.579	-0.284
		Pred ₂₅	600	0.133	0.000	0.034	0.046	0.095	0.131	0.165	0.205	0.255	0.383	0.383	0.062	0.626	1.050
	3	SP	560	0.806	0.627	0.726	0.735	0.786	0.827	0.836	0.843	0.846	0.855	0.228	0.043	-1.219	0.664

DS	Grupo	Métrica	N	Prom	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rango	Std.D.	Skew	Kurt
		MRE	600	6.196	0.776	1.073	1.487	2.811	4.665	8.209	12.772	18.863	20.485	19.709	4.879	1.408	1.408
		MdMRE	600	3.083	0.685	0.845	1.074	1.692	2.415	3.702	6.733	8.357	9.249	8.564	2.110	1.409	1.305
		SA	600	55.161	8.257	20.996	25.155	47.007	59.657	69.093	72.361	74.758	80.302	72.045	16.965	-0.825	-0.334
		Pred ₂₅	600	0.130	0.022	0.034	0.045	0.103	0.139	0.158	0.186	0.204	0.302	0.280	0.051	-0.112	0.254
4	SP		552	0.696	0.256	0.535	0.567	0.652	0.713	0.757	0.794	0.811	0.850	0.595	0.091	-1.251	2.542
		MRE	600	7.213	0.000	1.210	1.530	2.437	4.836	9.500	14.330	18.934	180.925	180.925	10.011	10.380	161.179
		MdMRE	600	2.913	0.000	0.767	0.941	1.395	2.283	3.740	5.888	8.230	9.157	9.157	2.084	1.328	1.166
		SA	600	36.118	-3284.809	9.079	21.087	39.838	54.145	65.767	72.376	74.840	77.432	3362.241	165.835	-15.780	287.899
		Pred ₂₅	600	0.133	0.000	0.035	0.053	0.100	0.133	0.165	0.200	0.243	0.313	0.313	0.057	0.313	0.531
5	SP		552	0.695	0.201	0.511	0.564	0.655	0.717	0.754	0.798	0.813	0.854	0.652	0.098	-1.500	3.525
		MRE	600	7.386	0.000	1.281	1.589	2.433	4.928	9.617	15.268	19.387	286.385	286.385	13.178	16.265	337.098
		MdMRE	600	2.927	0.000	0.778	0.991	1.386	2.216	3.784	5.665	8.296	9.570	9.570	2.121	1.368	1.297
		SA	600	33.145	-5238.018	3.976	20.575	39.364	54.671	65.927	72.648	74.637	77.562	5315.580	230.881	-20.356	456.773
		Pred ₂₅	600	0.133	0.000	0.034	0.054	0.097	0.133	0.163	0.202	0.236	0.335	0.335	0.057	0.279	0.631
6	SP		560	0.749	0.430	0.648	0.684	0.718	0.758	0.795	0.813	0.820	0.843	0.413	0.064	-1.828	5.163
		MRE	600	8.275	0.824	1.284	1.498	2.578	5.354	8.601	13.694	19.447	221.521	220.697	18.166	9.542	100.050
		MdMRE	600	3.100	0.719	0.969	1.141	1.593	2.670	3.710	6.363	8.173	9.101	8.383	1.994	1.408	1.510
		SA	600	8.813	-3999.033	0.350	20.175	39.744	52.285	61.877	66.067	68.718	72.652	4071.685	333.673	-10.334	111.343
		Pred ₂₅	600	0.122	0.027	0.037	0.070	0.095	0.124	0.143	0.173	0.197	0.290	0.263	0.042	0.257	0.983
7	SP		560	0.759	0.643	0.683	0.700	0.730	0.769	0.789	0.800	0.811	0.846	0.203	0.041	-0.707	-0.084
		MRE	600	6.428	0.843	1.320	1.563	2.294	5.160	9.534	12.428	18.490	20.585	19.742	4.829	1.174	0.846
		MdMRE	600	3.150	0.904	1.094	1.224	1.626	2.676	4.111	5.791	8.076	9.140	8.236	1.964	1.296	1.138
		SA	600	48.583	-21.653	20.032	25.287	41.315	49.910	61.272	65.859	67.730	75.800	97.452	15.712	-1.014	1.553
		Pred ₂₅	600	0.113	0.026	0.036	0.059	0.096	0.117	0.135	0.153	0.166	0.214	0.188	0.036	-0.391	0.482
8	SP		560	0.746	0.361	0.490	0.658	0.719	0.757	0.812	0.840	0.847	0.862	0.500	0.095	-1.954	4.672
		MRE	600	9.073	0.784	1.233	1.939	2.558	5.657	8.332	13.169	19.277	346.988	346.204	26.266	11.029	128.814
		MdMRE	600	3.115	0.714	1.051	1.208	1.670	2.617	3.618	6.483	7.979	9.171	8.457	1.964	1.371	1.278
		SA	600	-7.014	-6442.474	-3.116	19.007	39.319	49.374	60.423	64.818	68.836	74.307	6516.781	495.527	-11.351	133.734

DS	Grupo	Métrica	N	Prom	Min	5Q	10Q	25Q	Med	75Q	90Q	95Q	Max	Rango	Std.D.	Skew	Kurt
		Pred ₂₅	600	0.118	0.029	0.036	0.070	0.093	0.122	0.142	0.161	0.179	0.257	0.228	0.040	0.043	0.800
	9	SP	560	0.665	-0.221	0.236	0.420	0.617	0.735	0.793	0.809	0.815	0.827	1.047	0.180	-1.978	4.123
		MRE	600	9.107	0.922	1.720	1.914	3.358	8.780	13.322	18.381	19.336	21.730	20.808	5.837	0.347	-1.007
		MdMRE	600	4.222	0.930	1.047	1.195	1.750	4.301	5.901	7.671	8.391	9.858	8.928	2.369	0.230	-1.029
		SA	600	37.811	-59.110	16.143	20.009	28.614	38.209	48.287	57.119	61.311	67.309	126.419	15.083	-0.931	3.930
		Pred ₂₅	600	0.101	0.027	0.034	0.042	0.065	0.094	0.130	0.181	0.194	0.224	0.197	0.047	0.603	-0.421

Anexo 8 Mejores modelos de estimación de esfuerzo por grupo de análisis.

El Anexo 8 se lista, para cada uno de los grupos de los conjuntos de datos, los seis modelos que presentan los mejores resultados en la exactitud de la estimación con base en la métrica SA. Asimismo, para cada uno de los modelos se presentan las métricas asociadas al modelo: la MRE, MdMRE y Pred₂₅.

DS	DPxASxLA	SP	MRE	MdMRE	SA	Pred	DS	DPxASxLA	SP	MRE	MdMRE	SA	Pred
DS 202	1 3x5x6	0.722	0.616	0.785	58.79	0.315	DS 202	12 1x5x6	0.716	0.621	0.799	57.97	0.301
	1x4x6	0.715	0.636	0.800	56.82	0.300		3x3x2	0.719	0.622	0.805	57.23	0.309
	1x1x6	0.724	0.631	0.798	56.68	0.304		1x4x2	0.724	0.631	0.819	57.19	0.307
	3x2x2	0.718	0.624	0.809	56.64	0.293		1x3x6	0.721	0.629	0.798	57.14	0.300
	1x4x2	0.725	0.622	0.815	56.37	0.293		3x2x2	0.717	0.628	0.812	56.74	0.306
	1x1x2	0.714	0.635	0.825	55.97	0.291		1x3x2	0.715	0.627	0.825	56.73	0.301
	2 3x5x2	0.788	0.557	0.736	61.47	0.314		13 3x3x2	0.695	0.649	0.857	55.10	0.285
3x4x2	0.776	0.551	0.722	60.32	0.328	1x3x6	0.692	0.675	0.837	55.05	0.266		
2x2x6	0.794	0.795	0.793	59.63	0.328	3x4x2	0.685	0.678	0.854	55.04	0.276		
3x4x6	0.791	0.616	0.739	59.53	0.346	3x1x2	0.684	0.668	0.849	54.32	0.280		
3x3x2	0.773	0.578	0.755	59.30	0.304	3x3x2	0.694	0.659	0.851	54.04	0.279		
3x2x2	0.782	0.558	0.728	59.26	0.318	3x5x6	0.690	0.698	0.830	53.78	0.297		
3 3x3x2	0.778	0.557	0.727	63.04	0.321	14 1x4x6	0.705	0.650	0.803	57.10	0.290		
3x2x2	0.770	0.554	0.730	62.60	0.318	3x3x6	0.704	0.645	0.803	56.76	0.294		
3x5x2	0.783	0.574	0.727	62.31	0.319	1x1x6	0.701	0.654	0.805	56.58	0.293		
3x4x2	0.787	0.551	0.694	61.44	0.349	3x4x6	0.700	0.654	0.807	56.14	0.298		
3x4x6	0.801	0.600	0.738	61.34	0.345	1x3x2	0.693	0.649	0.822	55.24	0.293		
3x1x2	0.761	0.605	0.750	60.42	0.307	1x2x6	0.697	0.657	0.810	55.19	0.297		
4 3x4x2	0.749	0.620	0.772	61.57	0.303	15 1x2x2	0.670	0.685	0.824	55.69	0.294		
3x2x2	0.753	0.601	0.758	59.51	0.309	1x1x2	0.678	0.658	0.800	55.67	0.308		
3x5x2	0.772	0.602	0.749	59.47	0.308	1x4x2	0.674	0.664	0.800	55.59	0.304		
3x3x2	0.763	0.587	0.755	59.29	0.309	1x5x2	0.680	0.671	0.831	55.43	0.295		
3x1x6	0.789	0.663	0.728	59.13	0.351	3x4x6	0.689	0.690	0.787	55.32	0.333		
2x3x6	0.767	0.779	0.843	58.99	0.323	3x2x2	0.684	0.673	0.811	54.89	0.298		
5 3x4x2	0.762	0.601	0.751	60.94	0.312	16 1x3x6	0.649	0.687	0.840	53.76	0.297		
3x2x2	0.775	0.587	0.733	60.85	0.318	3x5x2	0.652	0.672	0.895	53.69	0.273		
3x5x2	0.772	0.565	0.737	60.83	0.314	3x4x6	0.648	0.755	0.912	53.68	0.269		
3x2x6	0.794	0.613	0.758	60.65	0.341	3x1x6	0.649	0.698	0.840	53.39	0.298		
3x4x6	0.781	0.627	0.747	59.95	0.337	1x3x2	0.636	0.701	0.903	53.38	0.265		
3x3x2	0.762	0.577	0.754	59.52	0.307	3x2x6	0.651	0.697	0.833	53.33	0.296		
6 3x4x2	0.747	0.597	0.768	61.02	0.301	17 1x4x2	0.617	0.684	0.924	54.53	0.260		
2x3x6	0.750	0.789	0.893	58.65	0.294	1x4x6	0.629	0.693	0.912	53.11	0.252		
2x4x6	0.758	0.777	0.869	57.59	0.309	3x5x2	0.634	0.674	0.924	52.43	0.264		
3x1x6	0.756	0.746	0.793	57.46	0.322	3x3x2	0.611	0.687	0.940	52.31	0.259		
3x5x6	0.715	0.733	0.815	57.08	0.312	3x1x2	0.616	0.677	0.927	52.21	0.268		
3x4x6	0.769	0.665	0.769	56.96	0.329	3x5x6	0.621	0.693	0.897	51.87	0.266		
7 3x4x6	0.765	0.646	0.767	60.61	0.330	18 1x1x2	0.615	0.693	0.868	54.01	0.265		
3x4x2	0.762	0.596	0.746	60.38	0.312	3x4x6	0.589	0.740	0.873	52.65	0.268		
3x1x2	0.760	0.632	0.734	58.03	0.332	1x3x2	0.600	0.707	0.892	51.93	0.261		
2x4x6	0.751	0.779	0.873	57.67	0.308	1x2x6	0.589	0.735	0.859	51.58	0.279		
3x2x2	0.726	0.650	0.768	57.60	0.311	1x1x6	0.599	0.732	0.873	51.44	0.268		
1x3x6	0.751	0.683	0.784	57.27	0.323	3x5x6	0.592	0.738	0.878	51.39	0.268		

DS	DPxASxLA	SP	MRE	MdMRE	SA	Pred	DS	DPxASxLA	SP	MRE	MdMRE	SA	Pred
8	3x2x2	0.702	0.649	0.809	56.43	0.300	19	1x4x2	0.563	0.752	1.002	49.94	0.243
	1x2x2	0.695	0.638	0.824	56.24	0.290		1x1x6	0.569	0.738	0.982	49.75	0.255
	3x1x2	0.709	0.651	0.823	55.81	0.294		3x4x6	0.563	0.748	0.991	49.26	0.267
	1x3x2	0.702	0.645	0.820	55.58	0.293		1x2x6	0.565	0.742	0.979	49.21	0.256
	3x1x6	0.705	0.655	0.795	55.38	0.307		3x1x6	0.573	0.744	0.981	49.19	0.265
	1x5x2	0.703	0.643	0.807	55.24	0.299		3x3x6	0.569	0.746	0.998	48.96	0.255
9	3x5x2	0.683	0.667	0.832	57.47	0.285	20	1x4x2	0.750	0.569	0.785	57.39	0.290
	3x2x2	0.691	0.636	0.836	57.06	0.291		3x4x2	0.741	0.595	0.801	56.49	0.286
	1x5x6	0.690	0.673	0.796	55.57	0.301		3x1x8	0.707	1.098	1.116	55.12	0.258
	3x2x2	0.687	0.666	0.821	55.30	0.291		3x5x1	0.564	0.980	0.969	54.73	0.255
	3x3x6	0.700	0.679	0.826	55.25	0.290		3x3x1	0.624	0.914	0.932	54.48	0.259
	3x4x6	0.698	0.660	0.795	55.15	0.308		3x4x2	0.589	0.717	0.878	54.41	0.257
10	3x4x6	0.803	0.618	0.734	61.11	0.355	21	3x1x6	0.554	0.820	0.973	50.21	0.232
	3x4x2	0.779	0.584	0.733	60.98	0.322		3x1x2	0.550	0.764	0.980	50.04	0.222
	3x3x2	0.758	0.588	0.760	60.52	0.306		3x5x7	0.553	1.131	1.099	48.81	0.238
	3x5x2	0.772	0.586	0.751	60.01	0.307		3x2x6	0.547	0.829	1.005	48.76	0.218
	3x2x2	0.757	0.593	0.759	59.72	0.303		1x4x6	0.558	0.834	1.008	48.69	0.227
	3x5x1	0.681	0.918	0.891	59.13	0.292		3x1x3	0.550	1.157	1.107	48.56	0.234
11	3x2x2	0.766	0.599	0.745	64.05	0.312	-	-	-	-	-	-	-
	3x4x2	0.784	0.612	0.717	61.12	0.331	-	-	-	-	-	-	-
	3x1x2	0.775	0.647	0.739	60.86	0.312	-	-	-	-	-	-	-
	3x5x2	0.779	0.589	0.734	60.65	0.313	-	-	-	-	-	-	-
	3x2x6	0.791	0.618	0.741	60.60	0.344	-	-	-	-	-	-	-
	3x3x2	0.782	0.558	0.733	59.72	0.317	-	-	-	-	-	-	-
DS 1	1x4x7	0.724	1.659	1.038	79.69	0.214	DS 6	1x2x14	0.809	2.452	1.906	72.65	0.155
53	3x4x7	0.719	1.471	1.033	79.14	0.200	53	3x3x1	0.820	1.320	1.003	72.40	0.216
	3x3x7	0.712	1.446	1.088	78.81	0.198		1x4x14	0.810	2.700	1.948	71.72	0.162
	1x3x14	0.807	1.977	1.344	78.77	0.184		3x2x14	0.818	2.393	1.628	71.37	0.145
	3x5x7	0.658	1.807	1.181	78.65	0.207		1x3x1	0.804	1.363	0.998	71.28	0.233
	1x2x14	0.820	2.154	1.549	78.51	0.180		3x5x14	0.813	2.477	1.940	71.19	0.134
2	1x2x7	0.721	1.478	1.007	79.86	0.201	7	3x3x14	0.799	1.937	1.524	75.80	0.176
	3x1x7	0.713	1.424	0.996	79.28	0.208		3x2x14	0.798	1.899	1.640	75.49	0.154
	3x1x14	0.784	2.475	1.279	79.23	0.205		1x4x14	0.793	2.085	1.863	74.91	0.149
	3x5x7	0.723	1.529	1.059	78.84	0.199		1x1x14	0.792	2.013	1.716	74.61	0.162
	3x3x7	0.690	1.638	1.003	78.84	0.207		3x1x14	0.792	1.980	1.669	74.38	0.154
	3x2x7	0.679	1.547	1.065	78.69	0.201		3x4x14	0.811	2.027	1.676	74.16	0.153
3	3x3x6	0.826	1.217	0.889	80.30	0.196	8	1x1x6	0.850	0.905	0.714	74.31	0.257
	3x1x6	0.830	1.199	0.914	79.74	0.179		3x5x3	0.845	3.399	2.189	74.22	0.151
	3x2x6	0.840	1.209	0.844	79.36	0.197		3x2x6	0.829	0.927	0.758	74.11	0.239
	2x3x10	0.779	1.121	0.695	78.26	0.271		1x4x6	0.833	0.932	0.756	73.45	0.237
	3x4x6	0.832	1.211	0.833	78.24	0.197		1x2x14	0.839	2.199	1.704	73.40	0.162
	2x1x10	0.786	1.376	0.695	78.08	0.250		3x3x6	0.837	0.954	0.796	73.31	0.230
4	3x3x6	0.795	1.658	0.838	77.43	0.256	9	3x4x6	0.808	1.244	0.943	67.31	0.170
	3x3x7	0.738	2.122	1.307	77.26	0.180		1x2x3	0.817	2.133	1.650	66.25	0.192
	1x3x6	0.686	1.210	0.707	76.99	0.306		1x1x6	0.811	1.245	0.964	66.14	0.173
	3x5x6	0.839	1.327	0.763	76.96	0.255		3x1x3	0.810	1.977	1.400	66.13	0.196
	1x5x6	0.730	1.181	0.695	76.86	0.296		3x3x6	0.807	1.263	0.978	66.06	0.164
	2x2x10	0.793	1.383	0.693	76.47	0.244		3x5x6	0.807	1.265	0.988	65.97	0.162
5	1x5x14	0.795	2.395	1.646	77.56	0.177	-	-	-	-	-	-	-
	2x4x10	0.790	1.276	0.745	77.11	0.240	-	-	-	-	-	-	-
	3x5x14	0.804	2.726	1.791	76.94	0.189	-	-	-	-	-	-	-
	3x2x7	0.737	1.908	1.142	76.85	0.189	-	-	-	-	-	-	-
	3x5x7	0.734	1.775	1.131	76.82	0.186	-	-	-	-	-	-	-
	3x5x7	0.717	1.583	1.115	76.65	0.205	-	-	-	-	-	-	-

Anexo 9 Amenazas a la validez de los estudios empíricos orientados a personas.

El Anexo 9 detalla el análisis de las amenazas a la validez de los estudios descritos en el Capítulo 7.

Amenazas a la validez de los estudios empíricos

A continuación se analizan las amenazas a la validez de los estudios presentados y las acciones realizadas para mitigarlas. Las amenazas a la validez del estudio son las relacionadas al repositorio de proyectos ISBSG, estudios de correlación y análisis de modelos de estimación de esfuerzo de desarrollo.

Validez de la conclusión

La cantidad y la naturaleza de los datos pueden afectar el poder de las pruebas estadísticas utilizadas para la evaluación. La selección y las características de los conjuntos de datos, las configuraciones de las técnicas utilizadas y las decisiones en el proceso de construcción de los modelos pueden afectar los resultados obtenidos. En algunos casos, para poder realizar la comparación de los resultados con estudios previos se reportan resultados aun cuando las pruebas estadísticas paramétricas no cumplieran las suposiciones. Este estudio fue conservador para eliminar valores atípicos de los conjuntos por lo que algunos valores extremos podrían influenciar los análisis de regresión utilizadas en los modelos de estimación de esfuerzo. Esta decisión se toma para que los modelos representen de mejor manera la realidad de proyectos de desarrollo de software de las organizaciones. Adicionalmente, se reportan los resultados de las pruebas no paramétricas correspondientes que permiten mejorar el análisis. Los datos reportados deben utilizarse con cautela dadas las características de los datos. Los lectores deben analizar los resultados en contexto y cada uno de los análisis presentados puede requerir adaptaciones específicas a las necesidades particulares de la organización.

Validez interna

La limitada cantidad de proyectos y la calidad de los conjuntos de datos seleccionados a partir del repositorio ISBSG pueden influenciar los resultados. Cada uno de los conjuntos de datos fue filtrado para garantizar que solo se seleccionara información con un nivel de calidad alto. Los conjuntos de datos fueron validados para eliminar posibles valores atípicos que pudieran influenciar los resultados, además de permitir la selección de muestras homogéneas dentro, de lo posible, de acuerdo a las características de interés para la investigación.

Validez del constructo

Las medidas utilizadas para la evaluación son las recomendadas en el área de medición. Se utilizaron múltiples métricas de comparación que permiten complementar los análisis para asegurar que las mediciones se realizan de manera adecuada y disminuyendo el sesgo. Para realizar los análisis se utilizaron técnicas robustas y aceptadas en la comunidad de la ingeniería del software de acuerdo a cada una de las evaluaciones. Cuando no se cumplían las suposiciones para las pruebas estadísticas paramétricas, se utilizaron las pruebas no paramétricas correspondientes. Asimismo, para realizar las evaluaciones de los modelos de estimación de esfuerzo se aplicaron las estrategias recomendadas en el área y de acuerdo a una estrategia de evaluación robusta y aleatoria. En los estudios se reportan distintas métricas de exactitud, algunos de los análisis son realizados solo para una métrica específica y las demás solo se detallan como valores de referencia.

Validez externa

Las características específicas de los proyectos de los conjuntos de datos seleccionados a partir del repositorio ISBSG pueden afectar los resultados. Las características de los datos se seleccionaron de acuerdo a las recomendaciones de los estudios previos y siguiendo criterios de inclusión específicos. El tamaño y las características de algunos de los

conjuntos de datos pueden limitar la generalización de los resultados. Los resultados son válidos para proyectos del dominio de aplicaciones de negocios y de acuerdo al método de medición evaluado. Se intentan seleccionar proyectos recientes lo que puede mejorar la aplicabilidad de los resultados en proyectos actuales. El análisis y comparación de los resultados se limita al conjunto de variables y métricas reportadas en los artículos ya que no se cuenta con los datos originales de los estudios previos.

Anexo 10 Artículos científicos publicados durante el proceso de la tesis.

El Anexo 10 de este documento de tesis se encuentra en el dispositivo de almacenamiento adjunto.