

UNIVERSIDAD DE COSTA RICA

SISTEMA DE ESTUDIOS DE POSGRADO

MODELO BAYESIANO ESPACIO-TEMPORAL
PREDICTIVO SOBRE PERSONAS QUE FALLECERÍAN
DEBIDO AL COVID-19 EN CANTONES EN COSTA RICA
DURANTE EL MES DE MAYO DE 2021

Trabajo final de investigación aplicada sometido a la
consideración de la Comisión de Programa de Estudios de
Posgrado en Estadística para optar por el grado y título de
Maestría Profesional en Estadística

RICARDO ZAMORA MENNIGKE

Ciudad Universitaria Rodrigo Facio, Costa Rica
2022

DEDICATORIA

Han estado siempre ahí presentes,
Y mucho más cuando las he necesitado
Les agradezco la gran felicidad y emociones
que siempre traen a mi vida.
A lo más precioso de este mundo,
Mi familia

AGRADECIMIENTOS

Quiero agradecer primero a mi familia, por el apoyo incondicional brindado durante toda esta etapa de mi vida académica y por motivarme constantemente para seguir adelante. También deseo dejar patente de mi agradecimiento a todas las personas que de una u otra forma colaboraron conmigo a fin de que lograra la culminación de mis estudios en este programa. Mi agradecimiento al director de trabajo final, el Dr. Shu Wei Chou-Chen, y a los lectores de este estudio, el Dr. Gilbert Brenes Camacho y el MSc. Johnny Madrigal Pana. A todos muchas gracias.

"Este trabajo final de investigación aplicada fue aceptado por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar al grado de Maestría Profesional en Estadística y al título respectivo"

Shu Wei Chou Chen, Ph.D

Profesor Guía

Gilbert Brenes Camacho, Ph.D

Profesor Lector

Johnny Madrigal Pana, M.Sc.

Profesor Lector

Ricardo Zamora Mennigke

Sustentante

La persona Sustentante que consta en esta hoja de aprobación cuya rúbrica digital es omisa, se encuentra fuera de la República de Costa Rica, razón por la cual la grabación de la defensa de tesis, debidamente resguardada por la Universidad de Costa Rica, respalda la aprobación y participación de sus parte.

ÍNDICE

	<i>Página</i>
RESUMEN	vii
ABSTRACT	ix
ABREVIATURAS	xvi
CAPÍTULO 1	
INTRODUCCIÓN	1
1.1 Objetivos	6
1.1.1 Objetivo general	6
1.1.2 Objetivos específicos	6
1.2 Organización del trabajo	7
CAPÍTULO 2	
MARCO DE REFERENCIA	8
2.1 Estadística espacial	11
2.2 Manejo del proceso espacial	12
2.2.1 Distancia y vecindario	14
2.2.2 Índice I de Morán y C de Geary	15
2.2.3 Semi-variograma	16
2.3 Modelo condicional autorregresivo	17
2.4 Modelo simultáneo autorregresivo	19
2.5 Modelos de Poisson y modelos binomial negativos	20
2.6 Comparación de modelos	22
2.7 Riesgos relativos	23
CAPÍTULO 3	
MARCO METODOLÓGICO	26

3.1	Fuente de datos	26
3.2	Modelos propuestos	31
3.3	Softwares y códigos empleados	38
3.4	Algoritmo INLA	39
CAPÍTULO 4		
ANÁLISIS DE RESULTADOS		40
4.1	Covariable principal	41
4.2	Distribución del modelo	43
4.3	Análisis de covariables	43
4.4	Modelos según efectos espacio-temporales	47
4.5	Análisis del modelo final	48
4.6	Predicción de abril y mayo para 2021	53
CAPÍTULO 5		
CONCLUSIONES Y RECOMENDACIONES		56
 BIBLIOGRAFÍA		60
 APÉNDICES		76
A	Análisis inicial de datos exploratorios	76
ANEXOS		86
A	Acumulación de fallecidos por Covid-19 en Costa Rica	86
B	Predicciones de fallecidos por cantón para abril y mayo 2021	89

RESUMEN

La enfermedad respiratoria Covid-19, altamente contagiosa, ha causado tasas de mortalidad muy altas, así como un deterioro significativo en los pacientes que la contraen. El objetivo de este estudio ha sido el de proyectar, mediante modelos jerárquicos bayesianos espacio-temporales, la cantidad de personas que fallecerían por Covid-19 durante el mes de mayo de 2021, según los datos históricos y las covariables regionales con que se contaba (brindados oficialmente) sobre el Covid-19 en el periodo de abril 2020 a marzo 2021.

Aquí se analizan varios modelos bayesianos espacio-temporales con el fin de modelar variables de respuesta en epidemiología. Se establece que la variable de respuesta de fallecidos por la enfermedad es una variable con distribución binomial negativa. Además, se utiliza el paquete de R-INLA para realizar las estimaciones *a posteriori* de los modelos, debido a su facilidad y rapidez en comparación con los métodos de cadenas de Márkov.

En este estudio se analizan diferentes modelos según sus criterios de información (DIC, WAIC y CPO). Se sopesan las mejores combinaciones según las siguientes pautas: el tipo de interacción espacio-temporal, la distribución y las covariables más adecuadas para el modelo final.

La covariable más importante, de acuerdo con la literatura consultada, es la del *ratio* estándar de mortalidad (SMR por sus siglas en inglés) del mes anterior. Dentro de las covariables adicionales examinadas están: el índice de desarrollo humano, el porcentaje de adultos mayores y el porcentaje de vivienda urbana por cantones. Además, se incluyen covariables ambientales extraídas del espectrorradiómetro de imágenes de resolución moderada (MODIS). Al final la única covariable adicional del modelo final, además del SMR, es el porcentaje de vivienda urbana por cantones.

El modelo final que mejor ajusta los datos, incluye las covariables con una interacción de orden 2; es decir, estructurada en el tiempo, pero no en el espacio; y con un modelo aleatorio independiente (iid). El problema con este modelo se pone en evidencia al analizar los efectos, ya que en su SMR se presenta un efecto nulo, lo cual dificulta la estimación y predicción final. Los mapas de predicción, estimados por cantón, han sido empleados para generar predicciones futuras de la cantidad de personas que fallecerían en abril y en mayo de 2021 por esta enfermedad.

Los resultados sugieren que las covariables seleccionadas, no permiten mejorar el ajuste del modelo. Es importante señalar que las limitaciones en el acceso a ciertos datos, inhiben una conclusión precisa. Esto por cuanto, no se puede concluir con exactitud si existe asociación entre la cantidad de fallecidos y ciertas condiciones socioeconómicas o ambientales. En tal sentido los análisis espacio-temporales tienden a verse afectados por el modelo elegido y la poca cantidad de periodos disponibles para el análisis.

Para futuros estudios, puede ser relevante el análisis de las diferencias entre las cadenas de Márkov e INLA en modelos espacio-temporales, ya que algunos modelos en este estudio, son adaptados de referencias que simularon modelos con cadenas de Márkov. Se requiere investigar, más a fondo, si el uso de INLA, puede resultar inapropiado cuando se requiera un ajuste en los modelos espacio-temporales, con conteos bajos o limitaciones significativas en la cantidad de covariables y períodos comparado con las cadenas de Márkov.

ABSTRACT

The highly contagious respiratory disease Covid-19 has caused very high mortality rates, as well as significant deterioration in patients who contract it. The objective of this study has been to project, using hierarchical Bayesian space-time models, the number of people who would die from Covid-19 during the month of May 2021, according to historical data and the regional covariates available (officially provided) on Covid-19 in the period from April 2020 to March 2021.

Several Bayesian spatio-temporal models are analyzed in order to model response variables in epidemiology. It is established that the response variable of deaths from the disease is a variable with a negative binomial distribution. In addition, the R-INLA package is used to perform the posterior estimates of the models, due to its ease and speed compared to the Markov chain methods.

In this study, different models are analyzed according to their information criteria (DIC, WAIC and CPO). The best combinations are weighed according to the following guidelines: the type of spatio-temporal interaction, the distribution, and the most suitable covariates for the final model.

The most important covariate, according to the literature consulted, is the standard mortality ratio (SMR) of the previous month. Among the additional covariates examined are: the human development index, the percentage of older adults and the percentage of urban housing by cantons. In addition, environmental covariates extracted from the Moderate Resolution Imaging Spectroradiometer (MODIS) are included. In the end, the only additional covariate in the final model, besides the SMR, is the percentage of urban housing by canton.

The final model that best fits the data includes the covariates with an interaction of order 2; that is, structured in time, but not in space; and with an independent random model (iid). The problem with this model becomes evident when analyzing the effects, since in its SMR there is a null effect, which makes final estimation and prediction difficult. The prediction maps, estimated by canton, have been used to generate future predictions of the number of people who would die in April and May 2021 from this disease.

The results suggest that the selected covariates do not allow to improve the fit of the model. It is important to note that limitations in access to certain data inhibit an accurate conclusion. This is because it cannot be concluded exactly if there is an association between the number of deaths and certain socioeconomic or environmental conditions. In this sense, space-time analyzes tend to be affected by the chosen model and the small number of periods available for analysis.

For future studies, the analysis of the differences between Markov chains and INLA in space-time models may be relevant, since some models in this study are adapted from references that simulated models with Markov chains. It is necessary to inves-

investigate further whether the use of INLA may be inappropriate when an adjustment is required in space-time models, with low counts or significant limitations in the number of covariates and periods compared to Markov chains.

ÍNDICE DE CUADROS

Cuadro 3.1	Covariables analizadas para el modelo final	30
Cuadro 3.2	Covariables principales analizadas para estimar el número de fallecidos por Covid-19	33
Cuadro 3.3	Tipos de distribuciones de la variable de respuesta analizados en los modelos estadísticos analizados mediante R-INLA	33
Cuadro 3.4	Tipos de interacción espacio-temporal considerados en términos de la estructura del efecto espacial y temporal	34
Cuadro 3.5	Descripción de los 12 modelos principales que se consideraron para la comparación en cuanto a la especificación del logaritmo del riesgo relativo	36
Cuadro 4.6	Criterios de información para los modelos con distintas covariables	43
Cuadro 4.7	Criterios de información para los modelos con distribuciones y compensación (<i>offset</i>) diferentes	43
Cuadro 4.8	Criterios de información para los modelos con SMR del mes anterior y cada una de las covariables	46
Cuadro 4.9	Criterios de información para los modelos con interacción entre covariables y SMR del mes anterior	46
Cuadro 4.10	Criterios de información para los 12 modelos evaluados con diferentes tipos de efecto espacio-temporal	48
Cuadro 4.11	Resultados de efectos fijos para el modelo final	48
Cuadro 4.12	Resultados de precisión de hiperparámetros para el modelo final	50
Cuadro B.1	Comparación de valores predichos contra valores confirmados del mes de abril de 2021	104
Cuadro B.2	(Continuación) Comparación de valores predichos contra valores confirmados del mes de abril de 2021	105

Cuadro B.3	Comparación de valores predichos contra valores confirmados del mes de mayo de 2021	106
Cuadro B.4	(Continuación) Comparación de valores predichos contra valores confirmados del mes de mayo de 2021	107

ÍNDICE DE FIGURAS

Figura 4.1	Representación gráfica de fallecidos por Covid-19 en Costa Rica durante el período abril 2020 - marzo 2021	40
Figura 4.2	Representación gráfica del SMR por cantones en Costa Rica durante el mes de marzo de 2021	42
Figura 4.3	Representación gráfica de la correlación de las covariables predictoras	45
Figura 4.4	Representación gráfica de los marginales posteriores del intercepto (superior), el SMR del mes anterior (inferior izquierdo) y el porcentaje urbano (inferior derecho) en el modelo final	49
Figura 4.5	Representación gráfica de los valores aleatorios estructurados espacialmente y no estructurados	51
Figura 4.6	Representación gráfica de la media de riesgo relativo de coronavirus durante el período marzo 2021 por cantones en Costa Rica	52
Figura 4.7	Representación gráfica de la media, y sus límites (inferior y superior) de la estimación de fallecidos por coronavirus durante el período abril 2021 por cantones en Costa Rica	53
Figura 4.8	Representación gráfica de la media, y sus límites (inferior y superior) de la estimación de fallecidos por coronavirus durante el período Mayo 2021 por cantones en Costa Rica	54
Figura 9.9	Representación gráfica de fallecidos por Covid-19 por cantón en Costa Rica hasta marzo 2021	76
Figura 9.10	Representación gráfica de la distribución de fallecidos por Covid-19 a través del tiempo	77
Figura 9.11	Representación gráfica de rupturas personalizadas	78
Figura 9.12	Representación gráfica de rupturas naturales	79
Figura 9.13	Representación gráfica del Índice de Morán y sus p-valores	81
Figura 9.14	Representación gráfica del Índice de Morán y sus p-valores	82

Figura 9.15	Representación gráfica del diagrama de dispersión del Índice de Morán	83
Figura 9.16	Representación gráfica del diagrama de dispersión del Índice de Morán	84
Figura 9.17	Representación gráfica del diagrama de dispersión del Índice de Morán	85
Figura 10.18	Representación gráfica de la acumulación de fallecidos por Covid-19 en Costa Rica durante el período abril 2020 - marzo 2021 .	86
Figura 10.19	Representación gráfica de la acumulación de fallecidos por cantón en Costa Rica durante el período marzo 2020 - marzo 2021 .	87
Figura 10.20	Representación gráfica de la correlación de covariables predictoras (dendograma y mapa de calor)	87
Figura 10.21	Representación gráfica de la matriz de correlación de covariables predictoras	88
Figura 10.22	Representación gráfica del riesgo relativo de coronavirus durante el período abril 2020 - marzo 2021 por cantones en Costa Rica	90
Figura 10.23	Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Guanacaste	91
Figura 10.24	(Continuación) Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Guanacaste	92
Figura 10.25	Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de San José	93
Figura 10.26	(Continuación) Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de San José	94
Figura 10.27	(Continuación 2) Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de San José	95

Figura 10.28	Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Alajuela	96
Figura 10.29	(Continuación) Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Alajuela	97
Figura 10.30	Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Cartago	98
Figura 10.31	Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Heredia .	99
Figura 10.32	(Continuación) Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Heredia	100
Figura 10.33	Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Puntarenas	101
Figura 10.34	(Continuación) Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Puntarenas	102
Figura 10.35	Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Limón .	103

ABREVIATURAS

BYM	Modelo de Besag, York y Mollié
CAR	Distribución previa condicional autorregresiva
COVID-19	Enfermedad por coronavirus de 2019
CPO	Ordenada predictiva condicional
DIC	Criterio de información de desviación
EVI	Índice de vegetación mejorado
iid	Efectos aleatorios independientes
ICAR	Modelo condicional autorregresivo intrínseco
IDH	Índice de desarrollo humano
INLA	Aproximación de Laplace Integrada Anidada
LST	Índice temperatura de la superficie terrestre
MCMC	Métodos de cadenas de Markov Monte Carlo
MIR	Información radiométrica del infrarrojo medio
MSE	Error cuadrático medio
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Índice de vegetación de diferencia normalizada
Nbinomial	Distribución binomial negativa
NIR	Información radiométrica del infrarrojo cercano
OMS	Organización Mundial de la Salud
RMSE	Raíz del error cuadrático medio
SMR	Ratio estándar de mortalidad
WAIC	Información de Watanabe-Akaike



UNIVERSIDAD DE
COSTA RICA

SEP Sistema de
Estudios de Posgrado

Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.

Yo, Ricardo Zamora Mennigke, con cédula de identidad 115900325, en mi condición de autor del TFG titulado Modelo Bayesiano Espacio-temporal predictivo sobre personas que fallecerían debido al Covid-19 en cantones en Costa Rica durante el mes de mayo de 2021

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. SI NO *

*En caso de la negativa favor indicar el tiempo de restricción: _____ año (s).

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

FIRMA ESTUDIANTE

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

CAPÍTULO 1

INTRODUCCIÓN

El 8 de diciembre de 2019, el Comité de Salud y Salud Municipal de Wuhan informó sobre la presencia de una neumonía que, en ese momento, era inexplicable. Se trataba de una afección respiratoria muy contagiosa, que rápidamente causó altas tasas de mortalidad y un significativo deterioro en los pacientes que la contraían. La enfermedad fue llamada Covid-19. El agente infeccioso era un nuevo tipo de coronavirus, conocido como SARS-CoV-2, dada su semejanza con el virus responsable del Síndrome Respiratorio Agudo Grave (SARS) (Sah *et al.*, 2021).

Woo *et al.* (2012) explican que dicha enfermedad está estructuralmente relacionada con el Coronavirus (CoV), el cual pertenece a la familia *Coronaviridae* y al orden *Nidovirales*. A partir del análisis filogenético y genómico, esta familia se clasifica en cuatro géneros: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* y *Deltacoronavirus*. Las especies de *Alphacoronavirus* y *Betacoronavirus* infectan a los mamíferos, causan enfermedades respiratorias en los seres humanos y gastroenteritis en los animales, mientras que las especies de *Gammacoronavirus* y *Deltacoronavirus* infectan a las aves, pero algunas de ellas también pueden infectar a los mamíferos. Ge *et al.* (2013) agregan que, tanto las dos especies de virus del género *Betacoronavirus*, el síndrome respiratorio agudo severo (SARS-CoV) o el síndrome respiratorio de Oriente Medio (MERS-CoV), habían demostrado anteriormente que, los coronavirus pueden causar una amenaza pública significativa.

El Covid-19 se ha comportado como una pandemia única en este siglo. Afecta a adultos sanos y a personas mayores con algunos problemas de salud. La infección se produce según una tasa de aumento exponencial muy por encima del SARS o el MERS. Una de las grandes preocupaciones que se ha generado, desde el inicio de esta pandemia, radica en la posibilidad de que los centros de salud sean abrumados por la acumulación de pacientes. Según la OMS (2020), esta enfermedad supone un peligro inminente de

incremento acelerado en las solicitudes de atención hospitalaria. Por eso, es importante desarrollar herramientas que permitan proyectar factores como: características de las personas, duración de períodos hospitalarios requeridos y número de las personas diagnosticadas con Covid-19. Todo eso, con el fin de contar con la información pertinente para tomar decisiones adecuadas en el manejo de la pandemia.

Según Huang *et al.* (2020), el descubrimiento del Covid-19 se hizo efectivo en un grupo de pacientes con síndromes de dificultad respiratoria aguda en la ciudad de Wuhan en China. La diseminación de la enfermedad adoptó dimensiones globales en marzo de 2020. Lu *et al.* (2020) aclararon que, por el momento, la evidencia sugiere que las personas mayores y aquellas con problemas médicos subyacentes (como enfermedades cardiovasculares, diabetes, enfermedades respiratorias crónicas y cáncer) tienen más probabilidades de desarrollar casos graves a partir de esta afección. Existe una cantidad significativa de personas en las que la situación se vuelve crítica e incluso mortal.

De acuerdo con la OMS (2020), debe distinguirse terminológicamente la escala en que una enfermedad puede afectar la población. Un brote epidémico se produce tras la aparición repentina de una enfermedad infecciosa en lugares y momentos determinados. Para que la situación se considere pandémica, el brote debe afectar a más de un continente. Además, debe dar lugar a que los casos internos de cada país, más que importados, sean generados por transmisión comunitaria. Finalmente, una epidemia sucede cuando la enfermedad se propaga por un brote descontrolado y temporalmente continuo.

Según Cheng *et al.* (2019), las enfermedades infecciosas se adaptan principalmente a características antimicrobianas y de movilidad; luego adoptan la forma de pandemia o de epidemia. Entonces, las autoridades de salud pública se ven obligadas a comprender, tanto la virulencia de las enfermedades, como los factores demográficos y ambientales que, coadyuvan en el establecimiento de patrones de propagación en espacio y tiempo. Croner (2004) agrega que en ese proceso resultan relevantes los sistemas de información geográfica. Usualmente, tales sistemas proporcionan mapas dinámicos para compren-

der la distribución geográfica de las enfermedades. El mapeo colabora con el trabajo analítico sobre la frecuencia de casos, la ubicuidad y el grupo espacial de enfermedades, así como la asociación de enfermedades con factores ambientales, económicos, sociales, sanitarios, entre otros. Zhen *et al.* (2010) concluyen que, con tales visualizaciones y capacidades analíticas, se está llevando a cabo un crecimiento generalizado en los procesos de control y prevención de la salud pública.

Nykiforuk y Flaman (2009) indican que una de estas aplicaciones remite a la vigilancia de enfermedades infecciosas. Actualmente, esto resulta vital para analizar la dispersión y la magnitud que una enfermedad puede causar. De hecho, una de las grandes preocupaciones en otras pandemias (por ejemplo, la propagación mundial de la influenza aviar H5N1 en 2005-06) fue metodológica: cómo efectuar el rastreo epidemiológico. Rappole y Hubalek (2006) indican que, una de las principales razones para ello, fue la carencia de una herramienta de vigilancia de enfermedades en su etapa inicial. Eso causaba inaccesibilidad a las poblaciones en riesgo y generaba dificultades para implementar estrategias de inmunización a escala global.

De acuerdo con la Caja Costarricense de Seguro Social (2021), en Costa Rica, la mayoría de las víctimas del virus COVID-19 experimentarían una enfermedad respiratoria leve o moderada, y se recuperarían sin necesidad de un tratamiento especial. Para el 15 de marzo, Geovisión-UNED (2021), confirmó una cifra total de casos: 227.533. Para entonces, ya existían casos confirmados en 82 cantones de las 7 provincias; de ellos, 190.998 eran adultos, 17.231 eran adultos mayores y 19.193 eran menores de edad. De los casos confirmados, 112.698 eran mujeres y 114.835 eran hombres. Asimismo, 199.645 eran costarricenses y 27.888 eran extranjeros. En ese momento, 196.752 personas habían recuperado su salud y 3065 habían fallecido. En total, la cantidad de casos activos sumaba 27.716. Debe destacarse que el 86.47% de los casos confirmados se registraron como recuperados y la tasa de letalidad del virus en Costa Rica era de 1.34%. De los casos recuperados 97.228 eran mujeres y 99.524 eran hombres. Además, se estimaba que, de las 507 personas hospitalizadas, 236 estaban internadas en Unidades

de Cuidados Intensivos. Durante el mes de abril de 2021 se superó la barrera de los 3000 fallecidos en Costa Rica. El único cantón donde aún no registraban fallecidos el 17 de abril, era Turrubares. El estudio de esta enfermedad es muy importante, dada la cantidad de muertes y el cambio abrupto que ha provocado en la vida cotidiana, no solo en Costa Rica, sino en todo el orbe.

Por orden del Ministerio de Salud, se impusieron medidas preventivas en numerosas instituciones; también se implementaron numerosos procedimientos administrativos para contener la emergencia sanitaria (Rueda, 2021). A pesar de que en aquel momento Costa Rica tenía avanzado su programa de vacunación contra este virus, se pronosticaba que gran parte de la población (como la mundial) se mantendría, por meses o incluso por años, sin acceso a los programas de vacunación. Con la vacunación, se esperaba que los habitantes mejoraran su capacidad de resistencia y supervivencia a la enfermedad. También se esperaba que, conforme más personas recibieran la vacuna, se disminuyera la cantidad de muertes.

En Giuliani *et al.* (2020) se indica que la propagación de Covid-19 entre los países y dentro de ellos, no ha seguido un patrón homogéneo. Las causas de tal heterogeneidad aún no están claramente identificadas; pero Kandel *et al.* (2020) explican que los países tienen diferente manejo de la situación en función de sus habilidades de prevención, detección, estrategias de respuesta, función habilitadora y preparación operativa. Además, diferentes países han implementado diferentes medidas de cuarentena y control para prevenir y contener la epidemia, que afecta el movimiento de población, y por ende, el patrón de propagación del Covid-19. Dada la naturaleza altamente contagiosa del Covid-19, el patrón espacial de propagación de la enfermedad cambia rápidamente con el tiempo. Por lo tanto, comprender la dinámica espacio-temporal de la propagación del Covid-19 en diferentes países, resulta, sin duda, fundamental.

Esta pandemia se presenta como una prueba importante de la creciente demanda del sector de salud en cuanto al mapeo, análisis y visualización de datos de expansión de las enfermedades. Como respuesta al crecimiento exponencial de las tecnologías de

la información, ha habido un fuerte surgimiento de nuevos métodos de vigilancia de enfermedades transmisibles y la epidemiología en las últimas décadas (López y Blobel, 2007; Reeder *et al.*, 2012; Saran *et al.*, 2020). Según Carpenter (2011), los avances en la tecnología de la información han evidenciado la necesidad de contar con una tecnología geoespacial adecuada para el mapeo de enfermedades infecciosas con potencial epidémico. Asimismo, ha influido en la vigilancia, la predicción de brotes, la detección de agrupaciones y el análisis de patrones en el tiempo y en el espacio de la propagación.

La implementación de herramientas, técnicas y plataformas de sistemas de información geográfica brindan a la comunidad científica y a los profesionales una amplia gama de beneficios. Por ejemplo, en el caso del COVID-19 estos beneficios incluyen la visualización, el seguimiento de casos en tiempo real (de casos confirmados e informados), el seguimiento de contactos, la dirección de propagación y los puntos críticos (Dong *et al.*, 2020; Franch-Pardo *et al.*, 2020; Nøhr *et al.*, 2014). Probablemente, la aplicación más conocida de estos sistemas durante esta pandemia, es el panel de control Covid-19 en tiempo casi real; ha sido desarrollado por la Universidad Johns Hopkins (Dong *et al.*, 2020). La implementación de *software* para el análisis con sistemas de información geoespacial computarizado se remonta a mediados de la década de 1960.

Scala *et al.* (2020) indican que la propagación de Covid-19 ha influido en la introducción de una gran variedad de modelos epidémicos. Estos han sido generados con el propósito de identificar mecanismos específicos para el diseño de políticas públicas. Ballesteros *et al.* (2020) coinciden en que existe preocupación por la capacidad instalada en centros de salud para atender a las personas que deben ser hospitalizadas por Covid-19.

La dispersión de epidemias se da por muchos factores; algunos son complejos y difíciles de medir. Por ejemplo: el caso de los desplazamientos humanos por distintas regiones dentro y fuera de una ciudad. La incidencia de enfermedades infecciosas varía entre las ciudades. Algunos factores urbanos que inciden son: el tamaño de la población, las condiciones socioeconómicas imperantes, la estructura urbana y la conectividad con

otras ciudades (Dalziel *et al.*, 2018; Tian *et al.*, 2020). Niu *et al.* (2020), indican que los modelos predictivos clásicos de series de tiempo, no logran visualizar espacialmente los focos de Covid-19 en la población.

Es importante determinar números absolutos de esta pandemia para el territorio costarricense. Además, identificar las áreas geográficas donde se requieren mayores acciones preventivas. Por ello, se propone un estudio espacio-temporal para analizar las áreas donde se presentan más personas fallecerían debido a este virus. Este incluye generar mapeos predictivos con un modelo bayesiano jerárquico para estimar los meses de abril y mayo de 2021. Este tipo de análisis coadyuva en el proceso de toma de decisiones para focalizar zonas geográficas. De la misma manera, permite revisar para estos casos, los procedimientos estadísticos existentes.

1.1 Objetivos

1.1.1 Objetivo general

Predecir, mediante modelos jerárquicos bayesianos espacio-temporales que incluyen covariables regionales, la cantidad de personas que fallecerían por Covid-19 durante el mes de mayo de 2021 a partir de datos históricos por Covid-19 del período de abril 2020 a marzo 2021.

1.1.2 Objetivos específicos

1. Analizar las propiedades de los modelos Bayesianos espacio-temporales con el método de INLA aplicada a los datos de fallecimientos por COVID-19 en Costa Rica.
2. Comparar covariables deterministas para predecir los fallecimientos por Covid-19 en Costa Rica en distintos modelos jerárquicos bayesianos.

3. Aplicar los modelos bayesianos espacio-temporal para explicar las variaciones temporales y espaciales de las personas que fallecerían por Covid-19 según cantón en Costa Rica.
4. Proponer un modelo espacio-temporal que permite predecir los efectos temporales, espaciales y su interacción.

1.2 Organización del trabajo

Este trabajo se estructura en cinco capítulos. Tras el capítulo introductorio, en el segundo se trata la fundamentación teórica necesaria para comprender el análisis estadístico del trabajo. En esa sección, se incluyen conceptos sobre estadística espacial y riesgos relativos. El capítulo tres presenta el desarrollo metodológico llevado a cabo para completar los objetivos del estudio. En el capítulo cuatro se presentan y analizan los resultados de la simulación de modelos sobre las personas que, de acuerdo con los datos oficiales con que se cuenta, fallecerían. Finalmente, el último capítulo muestra las conclusiones y discusiones más relevantes del trabajo. Además de estos capítulos, se incluye la bibliografía consultada. Hay un apéndice y dos anexos con material relativo al estudio, aunque no esencial para el cumplimiento de los objetivos.

CAPÍTULO 2

MARCO DE REFERENCIA

En Desjardins *et al.* (2020), se indica que los estudios de patrones espaciales y temporales ayudan a dilucidar los mecanismos de propagación de una enfermedad en la población. Tales patrones, también sirven para identificar factores asociados con una distribución geográfica heterogénea. Gomes *et al.* (2020) estiman que, de manera similar, se requiere un análisis espacio-temporal prospectivo para monitorear los brotes. Dicho análisis permite la detección de conglomerados emergentes activos y del riesgo relativo para cada sitio afectado durante la epidemia. Por ello, considerando la situación actual de Covid-19, resulta importante analizar su tendencia y los conglomerados espacio-temporales de riesgo de mortalidad en Costa Rica. Esto, con el fin de definir áreas prioritarias para acciones de vigilancia y control más efectivas.

Casi todos los objetos, eventos o fenómenos pueden ser referenciados en el espacio y el tiempo. Por lo tanto, el análisis de datos espacio-temporales es de interés en muchos sectores industriales y de servicios. En consecuencia, durante los últimos años se ha incrementado la demanda de técnicas de análisis y modelado. Con Körner *et al.* (2012), se puede afirmar que las técnicas de análisis difieren de las técnicas clásicas de minería de datos, debido a las características de los datos y a su complejidad. Una de las principales diferencias entre los datos espaciales y no espaciales, es la autocorrelación; eso significa que, los objetos con características similares suelen agruparse en el espacio. Debe tomarse en cuenta que las características espaciales no necesariamente se distribuyen de manera uniforme en el espacio (Kargupta, 2009).

El análisis de tales datos es complejo, incluso si solo se consideran en relación con el comportamiento de los movimientos humanos. Ese fenómeno se ha estudiado por décadas. La primera clasificación formal de tales restricciones data de 1970 (Hägerstrand, 2005).

Según Körner *et al.* (2012), el estudio de datos espacio-temporales permite concentrarse en áreas específicas en determinados períodos. Por ejemplo, un análisis del riesgo de fatalidad permite obtener información valiosa. Se elabora a partir de los recuentos de casos observados dentro de pequeños distritos administrativos o regiones. Ha de obtenerse de esto, información potencialmente relevante de ciertos antecedentes como el número de individuos en riesgo. Probablemente sea necesario reconocer y evaluar algunas covariables, como la distribución regional por edades, las condiciones socioeconómicas de la muestra considerada y los niveles ambientales de contaminación. El problema fundamental para lograr este objetivo es la dificultad detallada por zonas. Debe tomarse en cuenta la necesidad de estudiar áreas pequeñas geográficamente; pero eso significa, que la muestra será reducida, lo cual causa un deterioro en la predicción estadística local. De esto se deriva la necesidad de establecer un balance adecuado entre el tamaño de las áreas administrativas estudiadas y los conteos de cada región (los cuales pueden ser bajos).

Las soluciones clásicas son poco factibles, ya que las muestras (datos) de cada región son insuficientes para alcanzar la precisión estadística requerida. El problema general consiste en que, los niveles de agrupación de los datos disponibles no coinciden con los objetivos del estudio. Por ejemplo, la disponibilidad de datos agrupados por áreas administrativas, no siempre permite evaluaciones en el nivel regional deseado. Por eso, los modelos espacio-temporales permiten un enfoque más adecuado para el estudio de áreas pequeñas. De esa forma se pueden mejorar las estimaciones locales, lo cual permite la *suavización* de las tasas de información requerida para tamaños de muestra locales pequeños. Fay y Herriot (1979) indican que dichos modelos espacio-temporales a menudo se expresan como modelos de efectos mixtos, y se remontan al trabajo de quienes propusieron el uso de intersecciones aleatorias para agrupar información y proporcionar tasas estimadas a nivel de los subgrupos.

La propuesta de Fay y Herriot (1979) sirve de base a una considerable literatura sobre estimación de áreas pequeñas; tales como, las formuladas por Ghosh y Rao (1994)

o las de Ghosh *et al.* (1998). Ambas son utilizadas en el análisis de encuestas estadísticas, incluidas las encuestas de salud, como la adaptada por Raghunathan *et al.* (2007). La mayoría de los enfoques tradicionales para la estimación de áreas pequeñas no son espaciales, dichos autores acometen el análisis de datos de subconjuntos con muestras pequeñas. Asimismo, los métodos espaciales suelen tomar información por igual en todas las áreas pequeñas, sin tener en cuenta sus ubicaciones espaciales relativas y suavizando las estimaciones hacia una media global. El mapeo de enfermedades se refiere a una colección de métodos que extiende la estimación de áreas pequeñas para utilizar directamente el entorno espacial. Se asume con ello una correlación espacial positiva entre observaciones. Esencialmente se trata de tomar prestada más la información de áreas vecinas, que de áreas lejanas, y se suavizan las tasas locales de acuerdo con los correspondientes valores locales o vecinos.

Clayton y Kaldor (1987) definieron ese tipo de mapeo como un conjunto de métodos empíricos bayesianos. Se contruyen a partir de la regresión de Poisson con intersecciones aleatorias definidas con correlación espacial. Se trata de un enfoque jerárquico, que proporciona un marco conceptual conveniente, el cual induce a una correlación espacial entre las tasas de enfermedad local estimadas mediante una distribución de efectos aleatorios condicionalmente autorregresiva (CAR). Besag (1974) asigna las intersecciones específicas del área. De hecho, Besag *et al.* (1991) ampliaron los modelos a un entorno completamente bayesiano, empleando en su momento los algoritmos de la cadena de Markov Monte Carlo (MCMC). El marco resultante es jerárquico y semi-personalizado para MCMC. Eso permite estimar los efectos de covariables, la predicción de datos faltantes y la estructura de covarianza espacio-temporal en los modelos.

El suavizamiento está determinado por los datos y la formulación del modelo, tanto en la configuración espacial, como en la no espacial. Este proceso permite visualizar el patrón geográfico subyacente de la enfermedad. Esto podría considerarse inapropiado si lo que se pretende es identificar límites o regiones cambiantes (Banerjee y Gelfand, 2006; Ma *et al.*, 2009).

2.1 Estadística espacial

Los datos espaciales no solo representan la ubicación geográfica, sino también el tamaño, la forma y los atributos de un objeto en el planeta; ya sea que se trate de un bosque, de un lago, de un camino, de una ciudad, etc. Según Khan (2018), la utilización de datos espaciales se ha incrementado significativamente en las últimas décadas en disciplinas, como la ciencia geológica, la climatología, las ciencias ambientales, las ciencias biológicas y la salud pública.

En cuenta a su utilidad y aplicación los datos espaciales son, normalmente, como otros datos. No obstante, Gelfand (2010) indica que hay algunos procesos que deben considerarse cuando se utilizan datos espaciales. Por ejemplo, los fenómenos espaciales pueden considerarse como objetos discretos con límites claros o como fenómenos continuos, que se observan en todas partes; pero no necesariamente tienen límites naturales. El concepto de *objetos espaciales discretos* denota ciertas realidades específicas, tales como un río, una carretera, un país, una ciudad, un sitio de investigación, etc. Los ejemplos de fenómenos continuos o *campos espaciales* incluyen la elevación, la temperatura y la calidad del aire. Los objetos espaciales generalmente se representan mediante datos vectoriales.

Los recientes avances en el mapeo de enfermedades incluyen la dimensión temporal. Eso favorece, en primer lugar, el análisis espacio-temporal de la variación del riesgo de enfermedad y, en segundo lugar, el análisis conjunto de enfermedades que comparten factores de riesgo comunes. En esa labor, se emplean modelos jerárquicos bayesianos que dividen el riesgo de enfermedad en dos componentes espacio-temporales. Usualmente, se considera un componente compartido y uno específico que calibra el diferencial, por ejemplo, entre las enfermedades. Richardson *et al.* (2006) advierten la necesidad de incluir las frecuencias absolutas de cada región en tiempos distintos. No obstante, la medición puede resultar poco eficaz; por eso en tales situaciones, se considera óptimo el uso de probabilidades.

El mapeo de enfermedades tiene como objetivo principal determinar el riesgo de enfermedad subyacente, el cual parte de los datos epidemiológicos dispersos y se representa, generalmente, en un mapa de color suavizado. Esta metodología suele basarse en la inferencia bayesiana y ser aplicada al estudio de las enfermedades no infecciosas de baja incidencia, cuya distribución de casos está estructurada espacialmente (y eventualmente, de manera temporal). Coly *et al.* (2021) señalan que, en las últimas décadas, el mapeo de enfermedades ha mejorado y ampliado su ámbito de acción en: la integración de la dimensión temporal y el tratamiento de los datos faltantes. Es preciso tomar en cuenta varios aspectos *a priori* (covariables ambientales y de población, supuestos sobre el reparto y la evolución del riesgo) y lidiar con la sobredispersión. En el contexto de una enfermedad contagiosa, el resultado de un caso primario puede generar además, efectos secundarios de la patología en una vecindad espacial y temporal cercana. Esto puede resultar en una sobredispersión local y en mayores dependencias espaciales y temporales debidas a la transmisión directa e indirecta.

El análisis bayesiano se basa en el cálculo de la distribución de probabilidad posterior de los parámetros del modelo. Obaromi (2019) indica que la distribución de probabilidad posterior es la distribución de probabilidad condicional de los parámetros desconocidos, dados los datos observados y ponderados por la información previa. El modelado bayesiano depende de la capacidad de calcular distribuciones posteriores, a fin de proporcionar estimaciones sobre todos los parámetros del modelo correspondiente. Las distribuciones posteriores son fáciles de calcular. Las distribuciones con un conjugado *a priori* suelen tener una distribución posterior que sigue una distribución estándar.

2.2 Manejo del proceso espacial

En el análisis estadístico espacial, el conocimiento de un proceso sirve para predecir los patrones espaciales que podrían surgir. La probabilidad de que cualquier patrón observado sea el resultado de este proceso, se identifica mediante un análisis de una o

más de sus componentes. En contraste, Unwin (1996) indica que el análisis exploratorio de datos espaciales sirve para examinar una distribución observada e intenta inferir el proceso que la produjo. El objetivo suele ser el de encontrar patrones en los datos que son significativos, en relación con el conocimiento de dominio existente de los investigadores. Por lo tanto, una región dada, D , se divide en S unidades de área, denominadas (W_1, W_2, \dots, W_s) y la respuesta resumida en la unidad W_s es y_s . Ahora se puede describir el modelo de regresión como:

$$y_s = X^T(s)\beta + e(s) \quad (1)$$

donde $X(s)$ es la variable predictora en la unidad W_s , β es el coeficiente de regresión y $e(s)$ es el residual en la s -ésima unidad de área.

A la ecuación 1 se le conoce como modelo de regresión lineal múltiple. Si se obtiene un valor alto de R^2 ajustado, se puede usar este modelo para hacer predicciones. Pero si el R^2 ajustado es bajo, primero se debe averiguar la razón y con base en esto realizar la modificación. Algunas posibilidades son: que la respuesta y las covariables no están fuertemente correlacionadas.

Este modelo requiere covariables adicionales o las mediciones de unidades de área adyacentes, que pueden tener alguna información, que no puede ser capturado por los regresores. En el caso de que se modifique el modelo, para el $\rho(s)$ se debe ir separando el residuo en términos espaciales, y efectos aleatorios no espaciales (s). Por lo tanto, se tendría un modelo conocido como modelo espacial a nivel de área:

$$y_s = XT(s)\beta + \rho(s) + e(s) \quad (2)$$

Debe tenerse en cuenta que los principales tipos de datos vectoriales son puntos, líneas y polígonos. En todos los casos, la geometría de estas estructuras de datos consta de conjuntos de pares de coordenadas (x, y) . Los puntos son el caso más simple; cada

punto tiene un par de coordenadas y n variables asociadas.

La geometría de las líneas es un poco más compleja, pues se refiere a un conjunto de una o más polilíneas (series conectadas de segmentos de línea). Las líneas se representan como conjuntos ordenados de coordenadas (nodos). Los segmentos de línea reales se pueden calcular conectando los puntos; es decir, la representación de una línea es muy similar a la de una estructura multipunto. La principal diferencia, radica en que el orden de los puntos es importante, ya que se debe saber qué puntos deben estar conectados. Un polígono se refiere a un conjunto de polilíneas cerradas. La geometría es muy similar a la de las líneas, pero para cerrar un polígono el último par de coordenadas coincide con el primer par. Varios polígonos se pueden considerar como una sola geometría.

2.2.1 Distancia y vecindario

Las etapas del análisis continúan con una definición y caracterización de la estructura espacial y temporal de una variable regional en estudio. Según Alvarado (2011), el objetivo debería ser el de estimar una función apropiada. Ha de servir para calcular la correlación espacio-temporal de la variable regionalizada, y luego comparada con la hipótesis de variabilidad. Respecto al análisis con datos coordinados, este se debe llevar a cabo con el índice del vecino más cercano: $NNI = d(NN)/d(ran)$.

El análisis del vecino más cercano mide la extensión o distribución de una variable en estudio en un espacio geográfico. El resultado de tal medición proporciona un valor numérico, el cual describe la medida en que un conjunto de puntos está agrupado o espaciado uniformemente. En investigaciones espaciales se emplea ese método para determinar si la frecuencia con la que una variable, observada espacialmente, es comparable con otras ubicaciones. Según Charlton (2009), así se puede proporcionar un valor numérico para la agrupación de un fenómeno geográfico. Eso permite que dicho valor sea comparado de forma más precisa con otros lugares. Esta estimación compara la distancia entre los puntos más cercanos y la distancia media aleatoria. Alvarado (2011) señala la importancia de notar que, si el resultado es igual a 1, la distribución se ajusta

a Poisson; si es mayor que 1, existe tendencia a la dispersión y si es menor que 1, existe tendencia a la agregación.

En el método del vecino más cercano promedio, si el índice es menor a 1, la distribución se agrupa; pero si el índice es mayor a 1, existe una tendencia a dispersarse. Según Aziz *et al.* (2012), el método del vecino más cercano medio es muy sensible, por lo que, un pequeño cambio en la distribución del uso del espacio puede provocar variaciones en los cálculos respectivos.

En cuanto a los datos espaciales, es de esperar que las mediciones de unidades de área que están cerca unas de otras, tiendan a ser más similares que las que están alejadas. Por eso, se requiere construir una matriz de vecindad, W , para definir la relación entre unidades de área. Algunos enfoques comunes para considerar las unidades de área como vecindario son:

- Límite compartido: dos unidades de área, digamos W_s y $W_{s'}$, se dice que son vecindarios si comparten bordes comunes, vértice o ambos.
- Distancias centroides: se fija el número de vecinos, k , o la distancia, d , más allá de la cual no hay influencia espacial directa entre unidades espaciales. Para k dado, todas las k unidades más cercanas, basadas en la distancia del centroide, de W_s se llaman vecinas de él. Por otro lado, si la distancia, d , es un criterio importante de influencia espacial, una celda, $W_{s'}$, se llama vecino de W_s si la distancia entre ellos es menor que d . Por lo general, las entradas de la matriz de vecindad, W , son binarias: $W(s, s_0) = 1$ si las celdas W_s y W_0 son vecinas y 0 en caso contrario.

2.2.2 Índice I de Morán y C de Geary

La correlación de datos se estima según varía en los ejes de coordenados X e Y , una región en intensidad y frecuencia según las variables en estudio. Por ello, la prueba I de Morán, según Anselin (2010), establece el tipo, la intensidad y el rango de patrón espacial, midiendo el grado de autocorrelación de una distribución. Este es un índice

general, que muestra si el patrón está agrupado, disperso o al azar. A cada punto agregado se le asigna una cantidad de intensidad y se necesita alguna variación de los valores para la medición de esta métrica. Ord y Getis (2010) indican que los puntos que tienen valores similares se expresan en los valores altos de Morán I, que son positivos o negativos. Los valores del índice de Morán varían entre +1 y -1. Los valores que se aproximan a +1 demuestran la agrupación y los valores que se aproximan a -1 demuestran la distribución aleatoria de los patrones de datos.

Antes de incorporar efectos espaciales aleatorios en el modelo, Banerjee *et al.* (2014), señalan que es necesario comprobar si existe alguna asociación espacial entre unidades de área de vecindario o no. El índice de Morán es una estadística popular que mide la fuerza de la asociación espacial entre unidades de área. De acuerdo con Getis y Ord (2010), el I de Morán se define como:

$$I = \frac{n \sum_s \sum_{s'} W(s, s') (y_s - \bar{y})(y_{s'} - \bar{y})}{(\sum_{s \neq s'} \sum_s (y_s - \bar{y}))} \quad (3)$$

donde $I \in (-1, 1)$, pero no estrictamente. Significativamente diferente de cero, los valores de I representan una fuerte asociación espacial y si I está cerca de cero, entonces no hay asociación espacial en los datos.

2.2.3 Semi-variograma

La separación entre los puntos de datos se especifica mediante la distancia de retraso. Los cálculos de semivariograma pueden comparar puntos de datos en todas las direcciones o en direcciones específicas. En este caso, se representa un vector con magnitud y orientación. La búsqueda del número $N(h)$ de puntos de datos en un conjunto fijo h se realiza especificando una vecindad alrededor de h . Fanchi (2010) agrega que la orientación del vector con magnitud h se especifica mediante el ángulo θ . La orientación del conjunto puede incluirse en el análisis dividiendo la distribución de áreas de las propiedades en un número finito de sectores, como cuatro cuadrantes u ocho

octantes.

2.3 Modelo condicional autorregresivo

Los modelos jerárquicos bayesianos se utilizan normalmente en el contexto del mapeo de enfermedades. Aquí se modela la superficie de riesgos mediante una combinación de datos de covariables disponibles y un conjunto de efectos espaciales aleatorios. De acuerdo con Lee (2011), estos efectos aleatorios se incluyen para modelar cualquier sobredispersión o correlación espacial en los datos de la enfermedad. Los efectos aleatorios se modelan típicamente mediante una distribución previa condicional autorregresiva (CAR), y se han propuesto varias especificaciones alternativas.

El mapeo de enfermedades también se puede utilizar para definir la disparidad geográfica de enfermedades, para detectar agrupaciones de enfermedades y producir mapas de enfermedades. Algunos autores, como Wakefield (2006), han realizado un buen número de revisiones estadísticas sobre el mapeo de enfermedades. El modelo principal y central empleado para un mapeo univariante de enfermedades es el modelo de Besag, York y Mollie (BYM) (Besag *et al.*, 1991). Ese modelo es un tipo de modelo de efectos mixtos lineales generalizados, que tiene dos efectos espaciales aleatorios; uno que está desestructurado espacialmente y modelado usando un *a priori* normal y otro, un efecto aleatorio estructurado espacialmente que es modelado por un *a priori* autorregresivo condicional intrínseco (ICAR). Un desafío notable en los análisis espaciales es el problema de la autocorrelación espacial. El modelado de las interfaces espaciales que surgen en los datos referenciados espacialmente, se lleva a cabo normalmente integrando la dependencia espacial en la estructura de covarianza, ya sea de forma explícita o indirecta a través de modelos autorregresivos.

Los dos modelos autorregresivos comúnmente utilizados son el CAR y el modelo autorregresivo simultáneo (SAR). De acuerdo con Clayton y Kaldor (1987), el CAR para estimaciones bayesianas, es una técnica utilizada en el mapeo de enfermedades

utilizado para suavizar el riesgo relativo. Dicho modelo proporciona cierta contracción y suavización espacial de la estimación bruta del riesgo relativo. Eso proporciona una estimación más estable del esquema del riesgo subyacente de enfermedad, que la proporcionada por las estimaciones brutas. Esta técnica toma prestada información de áreas vecinas que no sean áreas lejanas; también suaviza las tasas locales hacia valores vecinos locales. Eso disminuye la varianza en las estimaciones asociadas y permite el efecto espacial de las diferencias regionales en las poblaciones estatales. El CAR se utiliza regularmente para describir la variación espacial de cantidades de interés en forma de agregados sobre subregiones. Estos modelos se han utilizado para analizar datos en diversas disciplinas, como la demografía, la economía, la epidemiología y la geografía.

El objetivo general de tales modelos espaciales es mostrar y cuantificar las relaciones espaciales presentes entre los datos. Ha de hacerlo en términos específicos, determinar cómo las cantidades de interés difieren de las variables explicativas, así como identificar grupos de *puntos calientes*. Según Rue (2005), los modelos CAR son una clase de estimaciones aleatorias basadas en el teorema Gauss-Márkov. Los modelos CAR se han aplicado ampliamente en estadísticas espaciales para modelar datos observados; también para modelar variables latentes (no observadas) y efectos aleatorios que varían espacialmente (Cressie *et al.*, 2005; Pettitt *et al.*, 2002).

El modelo CAR fue presentado por Besag (1974) y el modelo jerárquico por Besag y Kooperberg (1995). Estos modelos CAR producen dependencia espacial en la estructura de covarianza, como una función de una matriz vecina W , y regularmente, un parámetro de correlación espacial desconocido fijo. Los CAR y los ICAR se utilizan ampliamente como distribuciones previas para los efectos espaciales aleatorios estructurados en modelos bayesianos. El modelo ICAR, el más simple de las CAR *a priori*, es el modelo intrínseco sugerido de BYM, que se da como:

$$\phi|\phi_i N\left(\frac{\rho \sum_{j=1}^n W_{ij}\phi_j}{\sum_{j=1}^n W_{ij}}, \frac{1}{\tau(\sum_{j=1}^n W_{ij})}\right) \quad (4)$$

aquí $i = 1, \dots, n$ donde el parámetro, τ , es la precisión condicional. La precisión es comparativa con el número de unidades vecinas, mientras que la expectativa condicional de θ_i es la media de los efectos aleatorios en las unidades de área vecinas. La formulación de precisión aquí es funcional, porque esperaríamos que la precisión fuera mayor cuando tiene más áreas vecinas y, por lo tanto, más información para estimar el valor de θ_i . Estos grupos de distribuciones condicionales concuerdan con la distribución normal multivariante, y con una media vectorial cero. La matriz de precisión impropia viene dada por $Q = \tau(\text{diag}(W_1) - W)$, con W_1 , como un vector que comprende el número de vecinos para cada una de las unidades de área.

Una limitación de este modelo es la falta de un parámetro para regular la fuerza de la autocorrelación espacial; si multiplica θ por 10, entonces la precisión τ , disminuiría, pero la estructura espacial no cambia. Según Obaromi (2019), eso implica que el modelo intrínseco solo es práctico en circunstancias donde la autocorrelación espacial en los datos es fuerte. Sin embargo, no es práctico para situaciones en las que existe una autocorrelación espacial débil o moderada en el área de estudio, ya que el modelo tendría una tendencia a producir una superficie de riesgo estimada, demasiado suave en estos casos.

2.4 Modelo simultáneo autorregresivo

El SAR es útil para ordenar varias clases de dependencia entre datos, que tienen soporte discreto en un espacio de interés. De acuerdo con Hooten *et al.* (2020), a menudo, estos modelos se especifican jerárquicamente como modelos de regresión de efectos mixtos, con una estructura de primer orden, que es controlada por un término de regresión lineal convencional, y por una estructura de segundo momento inducida por efectos aleatorios correlacionados.

En su forma general, los modelos SAR se asemejan a los CAR. Pueden incluso ser equivalentes; pero a menudo se parametrizan de manera diferente. Es importante

destacar que los modelos SAR pueden ser especificados mediante una regresión, que sería simultánea de un proceso espacial discreto. Por lo tanto, permiten construir modelos estadísticos para procesos con propiedades gráficas direccionales, que pertenecen a mecanismos de generación de datos. Más comúnmente, los modelos SAR se han utilizado para explicar la estructura entre los datos, con apoyo espacial en aplicaciones que involucran ecología, epidemiología, sociología y ciencias ambientales.

2.5 Modelos de Poisson y modelos binomial negativos

El uso de un modelo de Poisson, se justifica, usualmente, como una aproximación a un modelo binomial cuando la enfermedad es rara (es decir, la probabilidad binomial es pequeña). Wakefield (2004) señala que en caso de probabilidades pequeñas, el análisis debe concentrarse en los modelos de Poisson. Con ello se consigue determinar y analizar un enfoque binomial. Esto permite eludir la llamada *falacia ecológica*, es decir, la tendencia de las correlaciones obtenidas sobre ajuste agregado (por ejemplo, regional). Como se sabe, dicha falacia tiende a exagerar ciertas correlaciones, que se obtendrían si los datos permitieran el ajuste de modelos basados en casos individuales de riesgo. Waller y Carlin (2010) indican en este caso, que el modelo de Poisson en su forma más básica, con efectos aleatorios espaciales:

$$Y_i \sim \text{Poisson}(E_i\theta_i) \quad (5)$$

Este parámetro medio se puede vincular con los predictores:

$$\log(\theta_i) = \alpha + \beta x + u_i + v_i \quad (6)$$

Lo anterior sirve como base para la estimación del riesgo relativo, lo cual proporciona la tasa global de enfermedad observada por región. Tómese recuentos de casos de enfermedad como Y_i , para un conjunto de regiones $i = 1, \dots, I$, dividiendo nuestro

dominio de estudio D . Waller y Carlin (2010) agregan, que aún existe discusión sobre el resultado esperado de E_i con $i = 1, \dots, I$. Indican que se deben modelar los conteos como variables aleatorias de Poisson o binomiales en modelos lineales generalizados, utilizando una función de enlace log o logit, respectivamente. En algunos casos, también se puede haber observado valores de covariables x_i específicas de la región con los parámetros asociados β , y α es el intercepto. u_i indica el efecto espacial aleatorio estructurado capturado y v_i el efecto no estructurado.

Otros datos incluyen el número local de individuos en riesgo n_i , o un número local de casos esperados bajo algún modelo nulo de transmisión de la enfermedad (por ejemplo, riesgo constante para todos los individuos), también denominado E_i . Se asume que los valores de n_i (alternativamente, E_i) son fijos y conocidos.

El riesgo de referencia, estimado y definido anteriormente, es conocido como estandarización interna. Puede ser engañoso, ya que se sigue concibiendo en el E_i como conocido, aunque aun dependen de la estimación del riesgo relativo rr . Sin embargo, dado que el impacto de esta elección, se desvanece dentro de un número creciente de regiones I , y teniendo en cuenta que la definición de rr solo sirve para establecer la gran intersección β_0 relativamente poco significativa, esto parece una preocupación menor. Además, se puede estandarizar aún más los riesgos y expectativas para tener en cuenta la variación espacial en la distribución de factores de riesgo conocidos (como la edad), en lugar de ajustar dichos factores de riesgo en las covariables específicas de la región. Waller (2004) proporciona una descripción general de los mecanismos y argumentos a favor y en contra de la estandarización en la epidemiología espacial.

Por su parte, la distribución binomial negativa suele emplearse, también en datos de conteo sobredispersados. Esta distribución representa estrictamente el número de éxitos antes de un fracaso. Puede considerarse una distribución de beta-poisson, que es definida por dos parámetros (ubicación y escala). La distribución binomial negativa

aplicada en Besag y Kooperberg (1995) se define como:

$$Y_{it} \sim NegBin(E_i\theta_{it}) \quad (7)$$

Su similitud con Poisson radica en que, cuando el modelo tiene un parámetro de dispersión de 1, la distribución binomial negativa es equivalente a una distribución de Poisson. Estudios como Madden *et al.* (2021), y Rohleder y Bozorgmehr (2021) han empleado este tipo de distribución en el estudio de Covid-19 mediante el algoritmo de INLA, que se explica más adelante.

2.6 Comparación de modelos

El análisis de comparación, y la selección del mejor modelo para el estudio, resulta una tarea compleja. Una información valiosa, que debe ser considerada en la selección de modelos son las siguientes: el error cuadrático medio (MSE) de las estimaciones de los parámetros, los valores de y ajustados y el riesgo relativo (θ), la raíz del error cuadrático medio (RMSE), el número de parámetros efectivos (pD), la desviación media ($D(\theta)$) y el criterio de información de desviación (DIC).

El selección de los modelos utilizados en esta investigación ha sido realizada a partir de los criterios de información: el criterio de información de desviación (DIC), criterio de información de Watanabe-Akaike (WAIC) y la ordenada predictiva condicional (CPO). Spiegelhalter *et al.* (2002) definen el DIC como:

$$DIC = \overline{D(\theta)} + 2pD \quad (8)$$

El WAIC fue propuesto por Watanabe (2010). Tanto el DIC como el WAIC miden la bondad de ajuste de un modelo bayesiano al tiempo que tienen en cuenta su complejidad en términos del número de parámetros efectivos involucrados en el modelo.

Por regla general, el modelo con menor valor DIC y WAIC es el que presenta el mayor rendimiento; es decir, el mejor equilibrio entre desviación y complejidad. Gelman *et al.* (2013) explican que estos criterios analizan valores esperados y variabilidad respectivamente, comparados con la distribución posterior de los modelos. Como último criterio de selección, se emplea la CPO cuya aproximación se obtiene con el muestreo de la distribución posterior:

$$CPO_{it} = \frac{T}{\sum_{i=1}^T p(y_i|\theta^t)^{-1}} \quad (9)$$

donde T representa la cantidad de iteraciones en el muestreo de la distribución posterior. Otros estudios espacio-temporales (Jalilian y Mateu, 2021; Sutanto *et al.*, 2021) han considerado varios criterios de información para analizar modelos. Pettit (1990) indica que los valores pequeños de CPO_{it} indican posibles valores atípicos, observaciones influyentes y de alto apalancamiento.

2.7 Riesgos relativos

La función de riesgo relativo espacial es una herramienta útil para describir la variación geográfica en la incidencia de enfermedades (Hazelton, 2017). Siguiendo a Knorr-Held (2000), las razones del riesgo relativo se deben estimar en cada cantón, calculando los efectos espaciales aleatorios estructurados y no estructurados. Fekedulegn *et al.* (2010) indican que, en estudios epidemiológicos de esta índole estadística, se prefiere el uso de estimaciones de riesgos relativos por encima de otras razones de probabilidades. El fin, en estos casos, consiste en evaluar probabilidades posteriores, ya que se ocupa de la temporalidad, y también de una medida más intuitiva de la carga de mortalidad. Por lo tanto, los riesgos relativos estimados sirven para el análisis de las medias posteriores.

Las estimaciones de efectos espaciales y temporales aleatorios, así como su interacción, permiten asignar un riesgo relativo a cada unidad espacial, temporal o espacio-

temporal. Estos riesgos relativos se obtienen tomando, de manera exponencial, los parámetros espacio-temporales en cada uno de los modelos. Aquí se incorporan los casos estudiados para el i -ésimo cantón asumiendo una distribución binomial negativa con media $\mu_{i,t}$ ($i = 1, 2, \dots, n; t = 1, 2, \dots, t$). El índice i representa los diferentes cantones (espacio) y el índice t representa los diferentes meses/años (tiempo):

$$O_{i,t} = NegBin(\mu_{i,t}) \quad (10)$$

donde:

$$\mu_{i,t} = E_{i,t} \cdot \theta_{i,t} \quad (11)$$

donde $E_{i,t}$, son los casos esperados en el cantón i -ésimo y $\theta_{i,t}$ es el riesgo relativo (rr) logarítmico medio, como sigue:

$$rr_{i,t} = e^{\theta_{i,t}} \quad (12)$$

que se modela de la siguiente manera:

$$\log(\theta_{i,t}) = \alpha + U_i + S_i + G_t + T_t + \delta_{i,t} + \beta_j \cdot X_{i,t,j} + \Psi_{i,t} \quad (13)$$

Se calcula el modelo de estimación del efecto espacio-temporal: α es la intersección, U_i representa el efecto aleatorio no estructurado espacialmente y S_i representa el efecto aleatorio estructurado espacialmente. La fórmula final representa el modelo al que se agregó la variable explicativa, donde $X_{i,t,j}$ representa el valor de la variable j en el t -ésimo periodo del i -ésimo distrito; β_j es el coeficiente de regresión correspondiente a la variable j ; $\Psi_{i,t}$ representa el efecto aleatorio de acuerdo con He *et al.* (2020).

Para las estimaciones respectivas, se emplea la razón estándar de mortalidad (SMR) que funge como razón o porcentaje que aproxima el aumento o la disminución de la

mortalidad en este caso del coronavirus con respecto a la población general. Este índice se obtiene de la división de los casos observados $O_{i,t}$ entre los esperados $E_{i,t}$ (Moraga, 2018).

CAPÍTULO 3

MARCO METODOLÓGICO

El objetivo central de esta investigación es la modelación y predicción de la cantidad de personas que fallecerían por Covid-19 en cada cantón. Se realiza dicha labor desde una perspectiva espacial y temporal, que requiere aproximaciones bayesianas, debido a que los datos oficiales solo tienen conteos bajos e incluso nulos en algunos cantones.

Se procede con un análisis exploratorio de los datos espaciales. Posteriormente, se determinan las características de la estructura espacial y temporal con los modelos seleccionados para el estudio. Dichos modelos han sido escogidos a partir de un concienzudo estudio bibliográfico previo; y se ajustan de manera pertinente con los datos disponibles. Finalmente, se genera una predicción espacio-temporal empleando el modelo seleccionado.

3.1 Fuente de datos

Los datos para realizar el estudio se desprenden del programa de Geovision, que es una plataforma donde se presentan los datos sobre casos de Covid-19 en Costa Rica. Se encuentra administrada y facilitada conjuntamente por el Ministerio de Salud y la Universidad Estatal a Distancia (UNED), Geovisión-UNED (2021). De ahí se obtienen, por cantón y por distrito, datos sobre personas que han contraído el virus.

La variable de respuesta de los modelos es el resultado de la cantidad mensual de fallecidos en Costa Rica (según los datos oficiales con que se cuenta) por cantón hasta el mes de marzo de 2021. Las covariables predictoras incluidas de la base de Geovision son: la cantidad de personas positivas acumuladas hasta este mismo mes de marzo del 2021.

Los datos socioeconómicos que funcionan como covariables adicionales, fueron recolectadas de bases de datos facilitadas por el Centro Centroamericano de Población de la

Universidad de Costa Rica CCP (2021) y el Programa de las Naciones Unidas para el Desarrollo (PNUD) PNUD (2020). Dado que solo se cuenta con esos datos, en este estudio debe asumirse que las covariables socioeconómicas son constantes en el tiempo.

De la base de Proyecciones Distritales de Población de Costa Rica 2000-2050 se extrae la variable de cantidad de población por cantón según los grupos de edad proyectados para el 2021. La covariable para el modelo se estima a manera de proporción de las personas adultas mayores (65 o más años) respecto a la población total de cada cantón. Del Censo Nacional de Población y Vivienda de Costa Rica 2011 se extrae la proporción de población urbana por cantón en Costa Rica. Del Informe Atlas de Desarrollo Humano Cantonal de Costa Rica 2020 de la Universidad de Costa Rica y del Programa de las Naciones Unidas para el Desarrollo (PNUD), publicado en noviembre de 2020, se obtiene el Índice de Desarrollo Humano (IDH) correspondiente al 2018 y por cantones de Costa Rica.

El distrito antiguo de Río Cuarto se ha convertido, recientemente, en cantón. Al emplear los datos socioeconómicos, se imputan (por datos faltantes) los valores distritales al cantón de Río Cuarto. Además, se resta ese valor distrital al cantón de Grecia, ya que era el cantón al cual pertenecía Río Cuarto cuando era un distrito.

Desde el 2017 hay, oficialmente, 82 cantones en Costa Rica. Pero muchos archivos, datos e índices cantonales son presentados oficialmente como si todavía existieran solo 81 cantones; o con nombres antiguos. Por ello, en este trabajo ha sido necesario ajustar varios nombres cantonales en los datos empleados. Dado que el IDH de 2018 para el cantón de Río Cuarto en el informe de la PNUD no fue estimado, fue necesario imputar el mismo valor del IDH del cantón de Grecia al cantón de Río Cuarto.

Los datos disponibles sobre el número diario de casos nuevos y de muertes en diferentes regiones, ya han sido analizados en un número considerable de estudios. Lo mismo sucede con el análisis de la estructura de autocorrelación temporal y espacial subyacente en la propagación de Covid-19. En tal sentido, Kang *et al.* (2020) utilizan el I de Morán, en la estadística espacial, con varias estimaciones de vecinos más cercanos.

Determinan una asociación espacial significativa de Covid-19 en el número diario de casos nuevos en las provincias de China. Por su parte, resulta ejemplificador para este estudio, que Gayawan *et al.* (2020) utilicen un modelo de Poisson cero inflado para estudiar el número diario de nuevos casos de Covid-19 en el continente africano. Gross *et al.* (2020), también resultan ejemplificadores cuando estudian la propagación espaciotemporal de Covid-19 en China y la comparan con otras regiones del mundo. Concluyen que hay relación entre la movilidad y migración humana desde Hubei con la propagación de Covid-19. Danon *et al.* (2020) combinan los datos del censo de 2011, para capturar el tamaño y el movimiento de la población en Inglaterra y Gales. Sujath *et al.* (2020) modelan y pronostican la propagación de los casos de Covid-19 en India. Lo hacen mediante regresión lineal, con perceptrón multicapa y con autorregresión vectorial.

La selección de las covariables es muy problemática, por lo que la mayoría de estudios aplican variables distintas. Asimismo, las técnicas estadísticas varían de acuerdo con el estudio. Un factor determinante en la selección del proceso estadístico es la disponibilidad de información. A pesar de que en Alamo *et al.* (2020) se indica que hay muchas organizaciones nacionales e internacionales brindando datos abiertos sobre el número de casos confirmados y de muertes, debe tenerse presente que dicho acceso suele estar delimitado al país y a la región. En determinados casos puede estar restringido por razones de diversa índole: políticas, estructurales, institucionales, legales, etc.

El uso de ciertas covariables es incierto en numerosos estudios existentes. Una de las covariables más aceptadas en modelos epidemiológicos de incidencia de casos es la SIR (*Ratio* estándar de incidencia, por sus siglas en inglés). Pero el uso de esta covariable no es exclusivo, podrían utilizarse otras (Bastos y Cajueiro, 2020; Peng *et al.*, 2020a). El SMR, según Everitt y Skronidal (2015), es una covariable de estudio popular para modelos epidemiológicos de mortalidad. Es definida como la razón de mortalidad estándar. Una estimación simple del riesgo es el SMR, que aproxima al riesgo relativo igual al número observado de casos dividido por los casos esperados. Se implementa cual covariable principal en el estudio para analizar los distintos modelos, aunque en

un primer análisis se compara con otras posibles covariables. Es decir, se compara con un modelo nulo y otro modelo con los fallecidos del mes anterior a los estudiados. De esa forma se pueden comparar los criterios de información y establecer si resulta mejor el uso del SMR.

Varios estudios han implementado variables socioeconómicas como el IDH y el Índice Multidimensional de la Pobreza (de Souza *et al.*, 2021; Henao-Cespedes *et al.*, 2021; Liu *et al.*, 2020). En el caso de estudios sobre Covid-19, también se implementan variables ambientales, como es el caso de Briz-Redón (2021). Debido a eso, se emplean variables relacionadas como velocidad del viento, exposición solar y temperatura. Se emplean en el estudio los datos ambientales por cantón descargados de la plataforma de MODIS (*Moderate Resolution Imaging Spectroradiometer*) utilizando el paquete MODISTools y MODIS en el *software* estadístico de R.

En estudios previos dicha base de datos ha sido empleada para estudios bayesianos espacio-temporales, como es el caso de Militino *et al.* (2018) y Villandr e *et al.* (2021). Los datos de MODIS (2021) son extra idos de los sat elites TERRA y AQUA. TERRA MODIS y AQUA MODIS registran toda la superficie de la tierra aproximadamente cada 1 o 2 d as, para lo cual se adquieren datos en 36 bandas espectrales o grupos de longitudes de onda. En este trabajo tambi en se emplea el  ndice de Vegetaci n de Diferencia Normalizada (NDVI), el cual refleja el crecimiento de la vegetaci n y est  estrechamente relacionado con la cantidad de radiaci n activa absorbida fotosint ticamente (Weier y Herring, 2000).

Con Rouse (1974), se entiende que de los datos se pueden extraer informaci n sobre patrones de datos infrarrojos, como el infrarrojo cercano (NIR) y el infrarrojo medio (MIR). Seg n MODIS, el  ndice de vegetaci n mejorado (EVI), minimiza las variaciones del dosel y el suelo y mejora la sensibilidad en condiciones de vegetaci n densa. Finalmente, se ha empleado aqu  el  ndice de temperatura de la superficie terrestre (LST) y emisividad, los cuales se obtienen de MODIS por separado de d a y de noche. Dentro de las variables de inter s intr secas en este  ndice est  el vapor de

agua de la columna atmosférica y la temperatura de la superficie del aire en el límite inferior. Hernández-González (2019) emplea en su estudio espacial la covariable EVI extraída de MODIS, aunque esta no resultó significativa dentro de su modelo bayesiano. Esos índices sirven como covariables ambientales. Dada su disponibilidad mensual, son variables predictoras en el estudio que se logran implementar para el modelo en términos mensuales y espaciales.

Dentro de las limitaciones con las covariables socioeconómicas, se tiene que solo se lograron recuperar en términos espaciales (según cantón) y parcialmente en el caso temporal donde se tiene la información en términos anuales. Para los modelos, también se analizaron los casos positivos de cada mes como covariables. Un resumen de las covariables estudiadas se puede observar en el cuadro 3.1.

Cuadro 3.1. Covariables analizadas para el modelo final

Categoría	Covariable
Socioeconómica	Porcentaje adulto mayor
	Índice de Desarrollo Humano (2018)
	Porcentaje urbano
Ambiental	Índice de vegetación mejorado (EVI)
	Infrarrojo medio (MIR)
	Infrarrojo cercano (NIR)
	Índice de vegetación de diferencia normalizada (NDVI)
	Índice de temperatura de la superficie terrestre (LST) de día
Índice de temperatura de la superficie terrestre (LST) de noche	
Pandémica	SMR mes anterior
	Casos positivos del mes anterior
	Casos positivos de 2 meses anterior
	Fallecidos del mes anterior

Debe tenerse en cuenta que estos datos a menudo presentan información incompleta e inexacta; lo cual, como es sabido es una de las limitaciones considerables en cualquier análisis y modelado realizado sobre los datos disponibles (Langousis y Carsteanu, 2020). Este estudio no está exento de tales limitaciones ya que, con respecto al comportamiento del coronavirus y a sus efectos en las poblaciones humanas, la información continúa siendo limitada. Además, los datos sobre muchas variables

relevantes, como el movimiento y la interacción de la población, y el impacto de las políticas de cuarentena y distanciamiento social, no están disponibles ni se miden con precisión. Cualquier análisis como el presente estudio solo proporciona una descripción aproximada e imprecisa de la dinámica espacio-temporal subyacente de la pandemia.

3.2 Modelos propuestos

Inicialmente se consideró la distribución Poisson como la más adecuada basado en los datos de bajos conteos como lo más adecuada según los criterios de selección utilizados para el estudio. Investigaciones como Gayawan *et al.* (2020); Jalilian y Matteu (2021); Polo *et al.* (2020); Sartorius *et al.* (2021), han aplicado modelos espacio-temporales con distribución Poisson en sus estudios. Aunque en Costa Rica todavía no parecen existir estudios de esta índole, sí existen investigaciones con patologías similares que requirieron la implementación y comparación de modelos Poisson, empleando conteos bajos (Chou-Chen y Alvarado-Barrantes, 2018). Esos estudios utilizan la misma estructura de modelos jerárquicos, pero con diferentes métodos de estimación y estructuras de covariables.

Un análisis más exhaustivo lo realiza Briz-Redón (2021), al estudiar un total de 12 combinaciones con variables ambientales para determinar el impacto de elección de los resultados de los modelos. No obstante, al final del presente estudio, se emplea la distribución binomial negativa; apoyado en numerosos artículos, como los de Jaya y Folmer (2021) y Saavedra *et al.* (2021). Dichos autores han implementado esa distribución para el análisis espacio-temporal, comparando la distribución contra otras como la Poisson.

Cuando se dispone de información temporal, es posible construir modelos espacio-temporales que incluyen efectos aleatorios espaciales y temporales, así como efectos de interacción entre el espacio y el tiempo. Un modelo espacio-temporal resulta separable cuando la estructura de covarianza espacio-temporal puede descomponerse como un

término espacial y temporal. En Fuentes *et al.* (2008) se advierte que esto a menudo significa que la covarianza espacio-temporal puede escribirse como un producto de Kronecker de una covarianza espacial y temporal. WIKLE *et al.* (1998) indican que las estructuras de covarianza no separables no permiten un modelado separado de las covarianzas espaciales y temporales.

En este trabajo se asume que el número de nuevos casos diarios de Covid-19 observados en el cantón ($i = 1, \dots, 82$) del mes ($t = 1, \dots, 12$), denotado por O_{it} . Además si E_{it} denota el número de casos esperados en la región i en el mes t , y rr_{it} el riesgo relativo en lugar y tiempo con conteos bajos en numerosas regiones que en la mayoría de los casos tienen cero o pocos fallecidos. Para efectos de este estudio, se procede a adaptar parte de la metodología de análisis seguida por Rodrigues (2017) y Briz-Redón (2021), que son estudios que a su vez adoptan la metodología de estudios previos como Knorr-Held (2000) y Blangiardo *et al.* (2013).

Inicialmente se procedió a analizar si los modelos mantienen una mejor predicción con el término SMR del mes anterior como mejor covariable o el número de fallecidos del mes anterior. Además, se incluyó el término *offset* para modelar la variable dependiente como una variable discreta. Por último, se analizó también un modelo nulo para determinar si las covariables anteriores aportan mejora a las estimaciones (Cuadro 3.2). Todo lo anterior se realizó comparando tanto una distribución Poisson como la binomial negativa (BN). Todo esto se estimó ya con una estructura de efectos aleatorios, donde la parte estructurada se estima con el ICAR tipo "rw2" y asumiendo BYM. Por su parte, el efecto aleatorio no estructurado se especifica con el modelo "iid".

Luego se procedió a determinar la distribución que estimaba los criterios de información más bajos para el modelo final. En este caso se presentaba como principales posibilidades la Poisson, la Negativa Binomial y sus versiones de cero inflado. A pesar de que R-INLA permite una gran cantidad de distribuciones de probabilidad se probaron las más acordes al caso generando dos modelos separados por distribución empleando el SMR del mes anterior en uno (Cuadro 3.3). Para esta parte se determinó de acuerdo a

Cuadro 3.2. Covariables principales analizadas para estimar el número de fallecidos por Covid-19

Distribución	Covariable
Poisson	Nulo
BN	Nulo
Poisson	SMR (mes anterior)
BN	SMR (mes anterior)
Poisson	Fallecidos (mes anterior)
BN	Fallecidos (mes anterior)

los criterios de información proseguir el estudio con una distribución binomial negativa (nbinomial). La definición de la distribución y en los siguientes modelos se basó en un análisis de las estimaciones del DIC, el COP y el WAIC, aunque también se analizó para estos casos el RMSE.

Cuadro 3.3. Tipos de distribuciones de la variable de respuesta analizados en los modelos estadísticos analizados mediante R-INLA

Nombre en R-INLA	Distribución de verosimilitud
poisson	Poisson
nbinomial	Binomial Negativa
zeroinflatedpoisson1	Poisson cero-inflado
zeroinflatednbinomial1	Binomial Negativa cero-inflado

El objetivo del ajuste de estos modelos ha sido encontrar el modelo con los mejores rendimientos y predicciones. Como siguiente paso general para definir el modelo final, se ha procedido a seleccionar las covariables del modelo final. Las covariables analizadas son: SMR del mes anterior al periodo de fallecidos, casos positivos por Covid-19 del mes anterior al periodo de fallecidos, casos positivos por Covid-19 de 2 meses anterior al periodo de fallecidos, porcentaje de adultos mayores por cantón, IDH por cantón, porcentaje urbano por cantón, EVI mensual, MIR mensual, NIR mensual, NDVI mensual, LST de día mensual, LST de noche mensual.

Algunas de las covariables seleccionadas en este paso mejoran el ajuste del modelo final. Se ha evaluado la correlación entre covariables. Hay variables que, probablemente,

mantienen elevadas correlaciones con el SMR del mes anterior. Por eso, se ha analizado cada covariable por separado junto con el SMR del mes anterior. Una vez definidas las covariables que mejoraron el ajuste según criterio de información se ha procedido a agregarlas.

De las covariables, para adicionar al modelo final, solo resultó significativo el porcentaje urbano por cantón. Una vez definida esta parte resultó relevante determinar el tipo de interacción. Como muestra el Cuadro 3.4, es importante tener en cuenta que el término de interacción se puede definir de varias formas distintas. Para este caso, las matrices I_s e I_t representan la matriz de identidad del efecto no estructural espacial y temporal, respectivamente. Por su parte, las matrices R_s y R_t denotan la matriz de no identidad con un efecto espacial y temporal estructurado específico, respectivamente.

Cuadro 3.4. Tipos de interacción espacio-temporal considerados en términos de la estructura del efecto espacial y temporal

Tipo de interacción espacio-temporal	Matriz identidad
I	$I_s \otimes I_t$
II	$I_s \otimes R_t$
III	$R_s \otimes I_t$
IV	$R_s \otimes R_t$

Fuente: Knorr-Held (2000)

Respecto a las interacciones, se ha tomado el criterio de Knorr-Held (2000), según el cual es necesario indicar que la interacción de tipo I se puede considerar como covariables independientes no observadas para cada combinación de región y período (i, j) , y sin ninguna estructura. La contribución al riesgo en las regiones vecinas o en períodos de tiempo consecutivos aún puede ser altamente correlacionados, debido a los efectos principales. En la interacción de tipo II cada región tiene una estructura de evolución específica que es independiente de la de las regiones vecinas. La estructura de evolución de cada región puede tener tantas formas como el propio efecto principal temporal. De la misma forma que antes, esto no significa que cada región tenga una evolución independiente de las vecinas, ya que pueden compartir un efecto principal temporal común. La independencia solo afecta la desviación de la tendencia global.

La interacción de tipo III, de manera similar, supone que la interacción tiene una estructura espacial para cada período, independiente de los períodos adyacentes (sus vecinos en el tiempo). Al igual que el efecto de agrupamiento espacial, esto se estima típicamente con una distribución CAR para cada período. La inclusión de un término espacial de heterogeneidad adicional, como el de BYM, produce un término que tiende a la interacción de Tipo I. Se considera que cada región específica puede tener una ligera desviación de la tendencia global, pero que es probable que esta desviación sea similar a la de las regiones vecinas y, al mismo tiempo, independiente de la de la anterior o posterior período de tiempo.

La interacción de tipo IV surge, finalmente, cuando se supone que las desviaciones de las tendencias globales están correlacionadas con sus vecinos, tanto en el espacio como en el tiempo. Esto puede modelar factores ocultos, cuyos efectos exceden los límites de una o más regiones, y también persisten durante más de un período de tiempo. Los cuatro tipos de interacción son adecuados para modelar diferentes tipos de fenómenos, principalmente según su alcance. (Gilks, 1996; Knorr-Held, 2000)

Para formular los modelos se toman en cuenta los 12 modelos empleados por Briz-Redón (2021), y adaptados de Knorr-Held (2000). Esto se muestra en el cuadro 3.5. Donde se estiman dichos modelos con el SMR del mes anterior al periodo de fallecidos como covariable. En los modelos μ denota la intersección del modelo y x_j ($j = 1, \dots, 4$) representa las covariables, u_i y v_i representan el efecto espacial aleatorio estructurado y no estructurado del modelo. Por su parte, $\gamma_{w(t)}$ y $\phi_{w(t)}$ el efecto temporal aleatorio estructurado y no estructurado, y finalmente, $\delta_{i(t)}$ el efecto espacio-temporal aleatorio.

El primer modelo (M1) solo incluye el efecto fijo de cada una de las variables consideradas en el estudio. En M2 se procedió a incluir SMR. Estos primeros dos modelos asumen errores independientes que no poseen información espacio-temporal.

En M3 y M4 se implementa un modelo espacio-temporal sin interacción. Es decir, se estiman por separado los efectos regionales y temporales. También los efectos espaciales (u_i y v_i), que son modelados mediante BYM de Besag *et al.* (1991) y Besag y Kooperberg

Cuadro 3.5. Descripción de los 12 modelos principales que se consideraron para la comparación en cuanto a la especificación del logaritmo del riesgo relativo

Modelo	Descripción
M1	$\mu + \log(E_{it}) + \sum_{j=1}^4 \beta_j X_{jit}$
M2	$\mu + \log(E_{it}) + \sum_{j=1}^5 \beta_j X_{jit}$
M3	$\mu + \log(E_{it}) + \sum_{j=1}^5 \beta_j X_{jit} + u_i + v_i + \gamma_{w(t)} + \phi_{w(t)}$
M4	$\mu + \log(E_{it}) + \sum_{j=1}^5 \beta_j X_{jit} + u_i + v_i + \gamma_t + \phi_t$
M5	$\mu + \log(E_{it}) + \sum_{j=1}^5 \beta_j X_{jit} + u_i + v_i + \gamma_{w(t)} + \phi_{w(t)} + \delta_{iw(t)}$ (I)
M6	$\mu + \log(E_{it}) + \sum_{j=1}^5 \beta_j X_{jit} + u_i + v_i + \gamma_{w(t)} + \phi_{w(t)} + \delta_{iw(t)}$ (II)
M7	$\mu + \log(E_{it}) + \sum_{j=1}^5 \beta_j X_{jit} + u_i + v_i + \gamma_{w(t)} + \phi_{w(t)} + \delta_{iw(t)}$ (III)
M8	$\mu + \log(E_{it}) + \sum_{j=1}^5 \beta_j X_{jit} + u_i + v_i + \gamma_{w(t)} + \phi_{w(t)} + \delta_{iw(t)}$ (IV)
M9	$\mu + \log(E_{it}) + \sum_{j=1}^5 \beta_j X_{jit} + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$ (I)
M10	$\mu + \log(E_{it}) + \sum_{j=1}^5 \beta_j X_{jit} + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$ (II)
M11	$\mu + \log(E_{it}) + \sum_{j=1}^5 \beta_j X_{jit} + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$ (III)
M12	$\mu + \log(E_{it}) + \sum_{j=1}^5 \beta_j X_{jit} + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$ (IV)

Fuente: Briz-Redón (2021)

(1995), respectivamente. Los efectos temporales aleatorios se establecen mensualmente. Briz-Redón (2021) explica que este modelo asume que la distribución condicional del efecto estructurado espacialmente en la región (u_i) es equivalente a:

$$u_i | u_{j \neq i} \sim Normal\left(\frac{1}{N_i} \sum_{j \neq i=1}^n w_{ij} u_j, \frac{\sigma_u^2}{N_i}\right) \quad (14)$$

donde N_i es el número de vecinos que tiene la región i , w_{ij} es el elemento (i, j) de la matriz estandarizada por filas. En este caso, dado que se trabaja con 82 cantones, la dimensión de la matriz es de 82×82 , la cual representa la matriz de vecindad para las regiones. Debe tenerse en cuenta que si las regiones i y j son vecinos, entonces $w_{ij} = 1/N_i$; de no serlo $w_{ij} = 0$. Por último, σ_u^2 representa la varianza del efecto estructurado espacialmente. En el CAR, para el efecto espacialmente no estructurado sobre las regiones (v_i), se considera un *a priori* gaussiano independiente y con igual distribución. Aquí su σ_u^2 representa la varianza del efecto desestructurado espacialmente

del modelo:

$$v_i \sim Normal(0, \sigma_v^2) \quad (15)$$

Respecto a los dos efectos temporales, el efecto estructurado temporalmente (γ_t) se modeló mediante un recorrido aleatorio de segundo orden y con un componente de varianza σ_γ^2 .

$$\gamma_t | \gamma_{t-1}, \gamma_{t-2} \sim Normal(2\gamma_{t-1} + \gamma_{t-2}, \sigma_\gamma^2) \quad (16)$$

Se emplea, también un *a priori* gaussiano independiente y con distribución similar:

$$\phi_t \sim Normal(0, \sigma_\phi^2) \quad (17)$$

Luego, se ajustan modelos espacio-temporales, entre los cuales se contabiliza la interacción espacio-tiempo mediante BYM (Modelos 5 a 8) y de Besag (Modelos 9 a 12) (ICAR). En particular, se emplearon las cuatro estructuras espacio-temporales propuestas por Knorr-Held (2000). Cada una de estas estructuras consiste en especificar el término espacio-temporal no separable del modelo, según una combinación concreta de un efecto espacial estructurado y no estructurado con un efecto temporal estructurado y no estructurado. La combinación de estos efectos se realiza mediante el producto de Kronecker de las dos matrices, las cuales representan respectivamente el efecto espacial y el efecto temporal elegidos. La calidad del modelo se evalúa analizando los criterios de información a través del DIC, WAIC y el COP estimado, y revisando el RMSE.

Al final, se ofrece una predicción de los meses siguientes con el modelo final. Para la predicción, fue necesario modelar mes a mes. Primero se realiza la predicción de abril de 2021. Con los coeficientes estimados del mes de abril se determina el SMR para ese mismo mes. Con los datos estimados de fallecidos en abril de 2021 se procede al análisis de los fallecidos del mes de mayo de 2021 utilizando el modelo final. Para la

comprobación se adaptan pasos de validación de Moraga (2021), quien logró generar un estudio espacio-temporal mediante el cual determina la precisión y falta de covariables posibles.

3.3 Softwares y códigos empleados

Los estudios estadísticos se realizan mediante el *software* estadístico R. El paquete más relevante en este estudio fue R-INLA (Rue *et al.*, 2009). Los paquetes adicionales empleados son: maptools (Bivand y Lewin-Koh, 2021), tmaptools (Tennekes, 2021), broom (Robinson *et al.*, 2021) gpclib (Peng *et al.*, 2020b), raster (Hijmans, 2021), devtools (Wickham *et al.*, 2021b), mapproj (McIlroy, 2020), tidyr (Wickham, 2021b), plyr (Wickham, 2011), dplyr (Wickham *et al.*, 2021a), ggplot2 (Wickham, 2016), ggthemes (Arnold, 2021), forecats (Wickham, 2021a), readxl (Wickham y Bryan, 2019), sf (Pebesma, 2018), scales (Wickham y Seidel, 2020), rgdal (Bivand *et al.*, 2021), spdep (Bivand y Wong, 2018), tmap (Tennekes, 2018), tidyverse (Wickham *et al.*, 2019), magick (Ooms, 2021), plotly (Sievert, 2020), gridExtra (Auguie, 2017), RColorBrewer (Neuwirth, 2014), lattice (Sarkar, 2008), sqldf (Grothendieck, 2017), ggpubr (Kassambara, 2020), goftest (Faraway *et al.*, 2021), leaflet (Cheng *et al.*, 2021), xlsx (Dragulescu y Arendt, 2020), egg (Auguie, 2019), grid (R Core Team, 2021), doParallel (Corporation y Weston, 2020), furr (Vaughan y Dancho, 2021), MODISTools (Tuck *et al.*, 2014), corrplot (Wei y Simko, 2021), PerformanceAnalytics (Peterson y Carl, 2020), MLmetrics (Yan, 2016) y GGally (Schloerke *et al.*, 2021). Parte de la limpieza y confección de la base, también se ha realizado en Microsoft Excel. El programa R resulta necesario para llevar a cabo la mayor parte del análisis estadístico. El código empleado para el estudio se encuentra disponible en: <https://github.com/zamoraricardo15/Statistics-master-final-graduation-project>.

3.4 Algoritmo INLA

En este trabajo, la implementación de los modelos se realiza mediante el método de Aproximación de Laplace Integrada Anidada (INLA). Es un método computacional que permite estimar las distribuciones marginales posteriores de los parámetros involucrados en cada modelo. El algoritmo de INLA fue introducido por Rue *et al.* (2009) como un algoritmo determinista específicamente para inferencias bayesianas. Lindgren y Rue (2015) indican que INLA está enfocado al desarrollo de modelos gaussianos latentes, proporcionando resultados precisos para un tiempo de cálculo más eficiente que los modelos de Monte Carlo y las Cadenas de Markov.

El paquete R, recientemente desarrollado, también es denominado R-INLA. Se está convirtiendo en un paquete cada vez más utilizado para la inferencia bayesiana. Carroll *et al.* (2015), indica que el software INLA se ha promocionado como una alternativa rápida a MCMC. Se utiliza para aplicaciones de mapeo de enfermedades, y sirve para realizar análisis bayesianos, sin tener que utilizar métodos de muestreo posterior. A diferencia de los algoritmos MCMC, que se basan en la simulación de Monte Carlo, el paquete INLA permite análisis bayesianos mediante la estimación numérica, por lo que no requiere cálculo iterativo extenso. El marco bayesiano con estimación MCMC implica largas duraciones de cálculo en el contexto de estudios de simulación intensivos y requiere computación paralela para realizar los análisis. En la mayoría de los casos, la estimación bayesiana que utiliza la metodología INLA lleva mucho menos tiempo en comparación con la estimación que utiliza MCMC. (Martins *et al.*, 2013; Rue *et al.*, 2009)

CAPÍTULO 4

ANÁLISIS DE RESULTADOS

En esta sección se resumen los resultados proporcionados por cada uno de los modelos estadísticos ajustados en cada uno de las etapas antes mencionadas. Primero, se evalúa la calidad de los modelos para determinar el tipo de modelo, luego el tipo de interacción, y por último, las covariables.

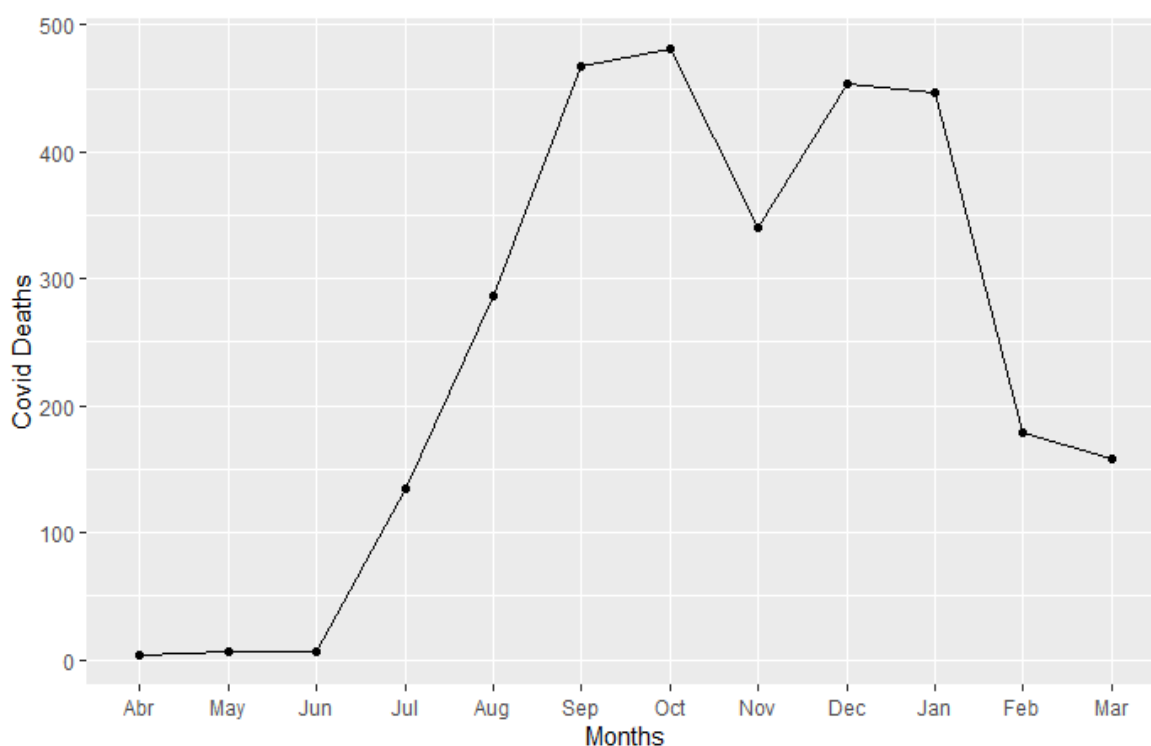


Figura 4.1. Representación gráfica de fallecidos por Covid-19 en Costa Rica durante el período abril 2020 - marzo 2021

La figura 4.1, muestra los casos de personas fallecidas por esta enfermedad. Se muestran varios picos en la tendencia especialmente en los meses de setiembre y octubre de 2020, y diciembre y enero. Dado que existen dos muertes registradas por coronavirus durante el mes de marzo de 2020. Cabe resaltar que, para efectos de estudio, ambos datos han sido imputados en el análisis. Durante la pandemia, se ha manifestado un

gran interés por analizar los datos relativos a ella de forma acumulativa. En este estudio, el análisis respectivo se ha hecho con datos separados de cada mes en estudio.

Debido a que la mayor parte de la población se concentra en la parte central del país, resulta esperable que los fallecidos se concentren en los cantones centrales. Un análisis estadístico descriptivo de ello se brinda en el Apéndice A y en el Anexo A de esta investigación. El cantón de Turrubares aun no presentaba fallecidos hasta el mes de marzo de 2021. En el mismo mes los cantones con mayores números de fallecidos son San José con 396 fallecidos registrados, seguido de Alajuela con 215 fallecidos y Desamparados con 181 muertes por esta enfermedad.

4.1 Covariable principal

Para el análisis de la covariable principal se emplean, en los modelos de prueba por separado: el SMR del mes anterior, su *offset*, los fallecidos del mes anterior y el modelo nulo. La figura 4.2 muestra el SMR calculado para el mes de marzo de 2021 en los distintos cantones. El SMR más elevado fue de 8.17 en el cantón de Dota, en la provincia de San José. El cantón de Turrubares al no contar con fallecidos tiene el SMR nulo. Existen numerosos cantones distribuidos por el país donde el SMR para el mes de marzo también es cero. Cabe resaltar que el criterio para estimar este valor fue determinado de manera mensual y cantonal. Por ello, se denota también que el SMR más elevado fue de 44.29 y ocurrió en el cantón de Guatuso durante el mes de junio de 2020. El SMR para Costa Rica durante el mes de marzo de 2021 está disperso por el país, lo que indica, a su vez, que el riesgo de mortalidad también está distribuido de manera heterogénea en el país.

El cuadro 4.6, muestra los valores DIC, WAIC y COP en distintos enfoques, para modelar el número total de fallecidos por coronavirus en Costa Rica en función del tiempo y del cantón. De acuerdo con los criterios de información, el modelo que mejor describe esta estructura es el que emplea el SMR del mes anterior bajo una distribución

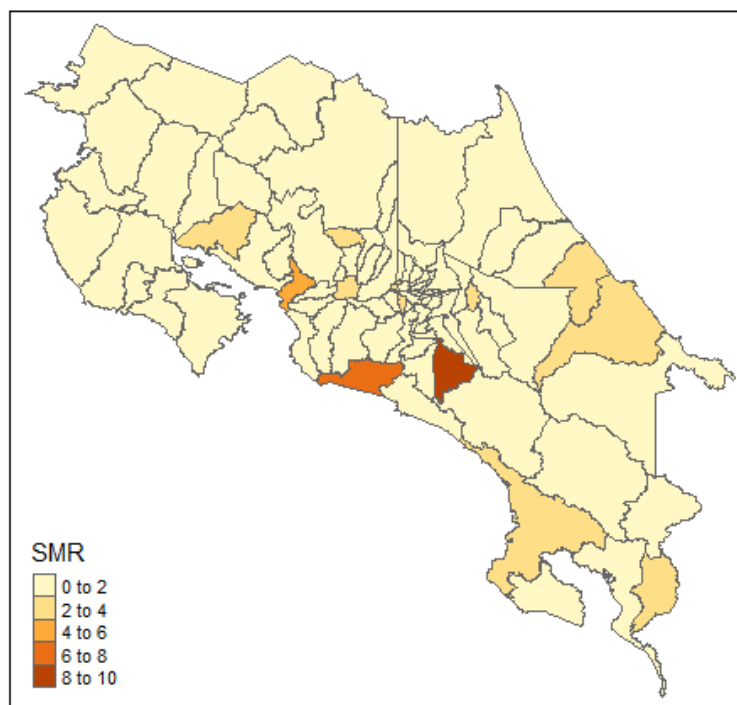


Figura 4.2. Representación gráfica del SMR por cantones en Costa Rica durante el mes de marzo de 2021

binomial negativa. Tiene un DIC de 2966.91, un WAIC de 2973.32 y un COP de 1487.82. Esto es relevante, ya que, al menos inicialmente, indicaría que existe una estructura autorregresiva. Además, se estimaron los coeficientes de raíz del error cuadrático medio (RMSE) para cada uno de los modelos, con el fin de analizar los ajustes en los mismos. El RMSE permite analizar comparativamente las diferencias entre los valores observados y los predichos. La diferencia entre el modelo nulo y el que emplea el número de fallecidos del mes anterior con distribución binomial negativa, no puede considerarse significativa. De hecho, este último modelo con fallecidos del mes anterior presenta un RMSE menor al de SMR del mes anterior. Para los siguientes modelos considerados más adelante, se emplea el SMR del mes anterior como covariable principal.

Cuadro 4.6. Criterios de información para los modelos con distintas covariables

Covariable	Distri.	DIC	WAIC	COP	RMSE
Sin covariable	Poisson	3200.33	3281.13	1643.57	6.94
Sin covariable	BN	2975.99	2978.96	1490.28	6.96
SMR mes anterior	Poisson	3161.70	3241.01	1626.28	6.98
SMR mes anterior	BN	2966.91	2973.32	1487.82	7.02
Fallecidos mes anterior	Poisson	3192.89	3284.28	1645.79	6.94
Fallecidos mes anterior	BN	2977.51	2982.18	1491.79	6.91

4.2 Distribución del modelo

Una vez definida la covariable principal del modelo, resulta importante definir la distribución (Cuadro 4.7). Según los criterios de información, la distribución que mejor describe la variabilidad de la variable de respuesta es la binomial negativa. El DIC de esta distribución es de 2966.91, un WAIC de 2973.32 y un COP de 1487.82. El RMSE es de 7.02, aunque no es el menor, muestra un valor bajo. La distribución de Poisson, así como los algoritmos de Poisson cero-inflado y los de binomial negativa cero-inflado tienen valores bajos, pero para los siguientes procedimientos se decide proseguir con la distribución binomial negativa.

Cuadro 4.7. Criterios de información para los modelos con distribuciones y compensación (*offset*) diferentes

Distribución	DIC	WAIC	COP	RMSE
poisson	3161.70	3241.01	1626.28	6.98
nbinomial	2966.91	2973.32	1487.82	7.02
zeroinflatedpoisson1	3142.55	3217.31	1614.65	6.98
zeroinflatednbinomial1	2969.26	2975.40	1488.98	7.02

4.3 Análisis de covariables

Varias relaciones entre las covariables presentes se pueden leer en la figura 4.3. Los valores numéricos de las correlaciones que se muestran en la gráfica revelan un nivel significativo y moderado de asociación positiva entre distintas variables, pero muestran

poca correlación con el SMR, lo que puede colaborar al final con la predicción del modelo. Para las covariables se denota una correlación y separación clara para denotar que las variables socioeconómicas, es decir: el IDH, el porcentaje de adultos mayores y el porcentaje urbano presentan fuertes correlaciones positivas entre sí. Por otro lado, se muestran las covariables ambientales con una situación similar, donde los coeficientes extraídos de MODIS presentan altas correlaciones. Por último, están el SMR, los fallecidos del mes anterior, los casos positivos para el mes anterior y el de dos meses anterior al período de fallecidos. Se muestra aquí una correlación alta, especialmente entre casos positivos de meses anteriores.

Las covariables permiten explicar gran parte de la variabilidad presente en un modelo espacio-temporal. El resumen de estos resultados se muestra en el cuadro 4.8. Los modelos fueron estimados utilizando la covariable específica, sumada al SMR del mes anterior. Las covariables se consideran significativas si con su adición al modelo logran mejorar significativamente el DIC de 2966.91, el WAIC de 2973.32 y el COP de 1487.82, estimados del modelo inicial que tiene solo el SMR del mes anterior.

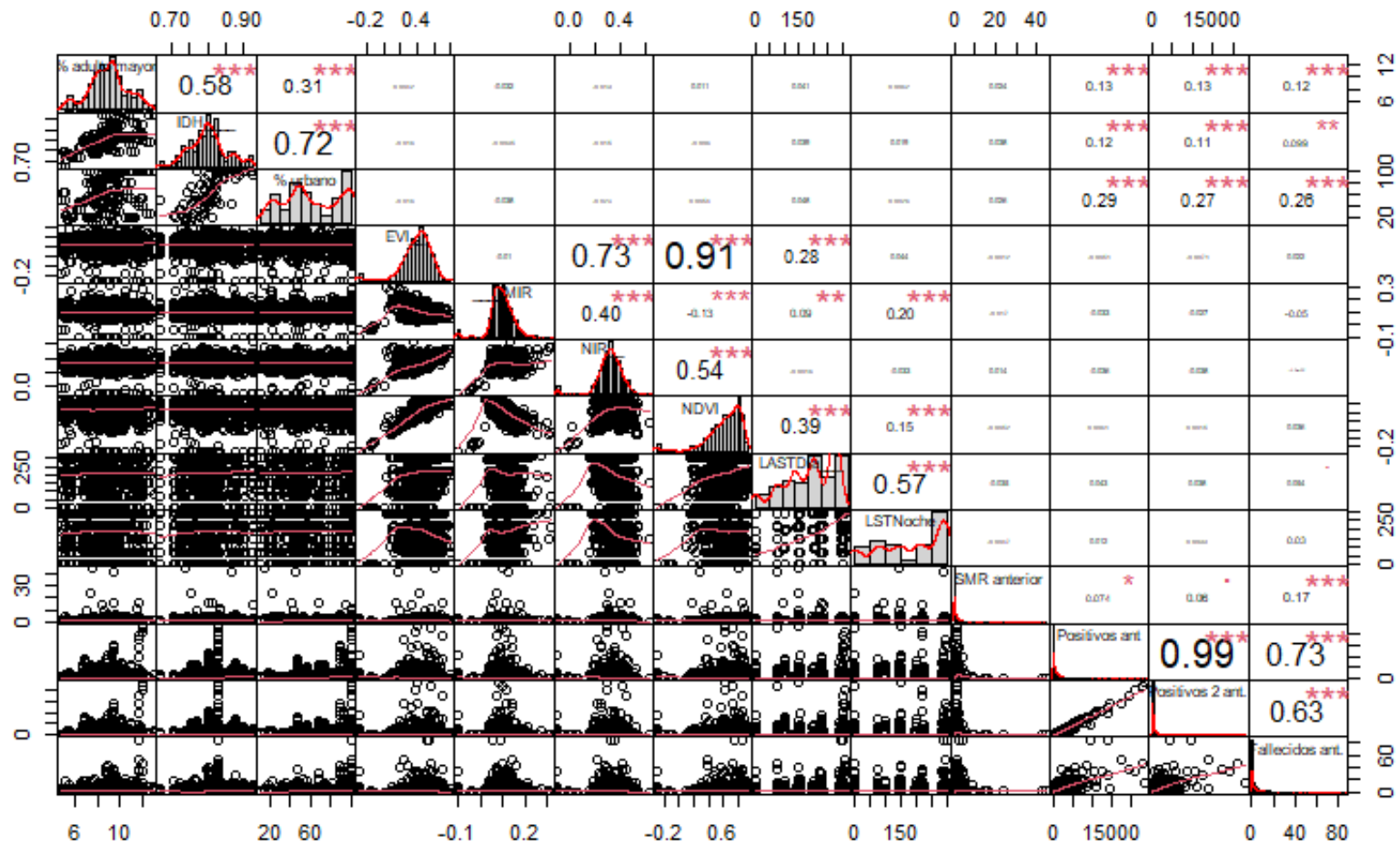


Figura 4.3. Representación gráfica de la correlación de las covariables predictoras

Cuadro 4.8. Criterios de información para los modelos con SMR del mes anterior y cada una de las covariables

Covariable	DIC	WAIC	COP	RMSE
Porcentaje adulto mayor	2967.17	2973.58	1487.94	7.02
IDH	2964.64	2970.97	1486.62	7.02
Porcentaje urbano	2960.62	2967.59	1484.72	7.00
EVI	2968.47	2975.34	1488.98	7.02
MIR	2968.49	2974.93	1488.71	7.02
NIR	2967.58	2974.39	1488.46	7.02
NDVI	2968.12	2975.41	1489.03	7.02
LST Día	2967.85	2974.05	1488.24	7.02
LST Noche	2966.83	2973.37	1487.88	7.02
Positivos mes anterior	2954.06	2960.33	1481.74	6.99
Positivos 2 meses anterior	2939.60	2945.84	1474.68	6.96

Primero, los términos ambientales extraídos del paquete MODIS, parecen entorpecer y disminuir la calidad del modelo según el DIC, WAIC y COP. En este caso, el modelo sin estas covariables resulta mejor. Por su parte, están las covariables socioeconómicas, en este caso, resulta peculiar notar que realmente existe ganancia en el IDH y el porcentaje urbano, tanto en el caso del DIC, WAIC y COP. Lo mismo ocurre con las covariables de casos positivos de ambos meses anteriores.

Cuadro 4.9. Criterios de información para los modelos con interacción entre covariables y SMR del mes anterior

Variable	DIC	WAIC	COP	RMSE
Positivos+Positivos 2	2928.88	2937.19	1470.06	6.83
Positivos+Positivos 2+% urbano+IDH	2921.89	2930.36	1466.44	6.83
Suma Positivos+% urbano+IDH	2936.81	2944.02	1473.48	6.96
% urbano+IDH	2961.69	2968.93	1485.38	7.00

En la figura 4.9 se agregan varias covariables al SMR del mes anterior. Uno de los problemas que presentan las covariables de casos positivos, es el hecho de que se encuentran altamente correlacionadas entre sí y con las covariables de fallecidos. Con respecto al efecto de cada covariable sobre la propagación de Covid-19, pueden incluirse los casos positivos del mes anterior; también los casos positivos de dos meses anterior al

periodo en estudio. Inclusive, puede incluirse la suma de ambas como una covariable. Sin embargo, la inclusión de esas covariables al modelo final, no resulta óptima, pues ambas producen un efecto fijo medio nulo sobre la variable de respuesta, que en este caso es el número de fallecidos. Al probar modelos con estas covariables sumadas al SMR, el efecto fijo sobre la predicción de fallecidos indica un efecto nulo. Por ello, no se consideran estas covariables en el modelo final. El IDH adicionado a porcentaje urbano parece no mejorar tampoco los criterios de información, por lo que, se decide incluir para el modelo final únicamente el porcentaje urbano y el SMR del mes anterior.

4.4 Modelos según efectos espacio-temporales

El cuadro 4.10 muestra que el modelo 6, donde se incluyen efectos temporales aleatorios de orden mensual, presenta los mejores ajustes según el DIC, WAIC y COP. En este modelo se presentan las interacciones de orden II (estructurada en el tiempo, pero no estructurada en el espacio). Nuevamente, los ajustes de otros modelos se presentan prometedores y con diferencias poco significativas con el modelo 6. Se considera este último como el adecuado para ajustar el modelo final. Algunos modelos con ajustes cercanos son el modelo 4, que también presenta efectos temporales mensuales, así como el modelo 3 y 5. Asimismo, el modelo 9, que presenta efectos temporales aleatorios a nivel anual y una interacción espacio-temporal de tipo I, también parece presentar resultados similares.

Se omiten los resultados del modelo 7 y 11 que presentaron una interacción de tipo III (estructurada en el espacio pero no estructurada en el tiempo). Esos modelos presentan dificultades en el proceso de convergencia. Los modelo 8 y 12 poseen una interacción espacio-tiempo tipo IV (estructurada tanto en el espacio como en el tiempo) y muestran valores WAIC y COP poco confiables (junto al RMSE); de hecho, el modelo 8 no converge. Se decide tomar como base el modelo 6 con interacción tipo II. Esto debido a que presenta mejores criterios de información: un DIC de 2925.58, un WAIC de 2934.76 y un COP de 1476.69.

Cuadro 4.10. Criterios de información para los 12 modelos evaluados con diferentes tipos de efecto espacio-temporal

Modelo	DIC	WAIC	COP	RMSE
M1	3067.02	3067.40	1533.70	7.05
M2	3030.80	3036.95	1518.71	8.69
M3	3006.31	3012.01	1506.23	7.45
M4	2960.11	2967.56	1484.70	7.00
M5	2960.36	2967.73	1484.79	7.00
M6	2925.58	2934.76	1476.69	6.91
M7	NA	NA	NA	NA
M8	NA	NA	NA	NA
M9	2969.92	2977.05	1489.40	7.02
M10	3853.66	3838.53	1977.16	5.24
M11	NA	NA	NA	NA
M12	4203.00	179043.08	29147.42	2.71e+18

4.5 Análisis del modelo final

La estimación mediante INLA devuelve un modelo ajustado. En él se evidencia la necesidad de analizar los márgenes posteriores de los parámetros. También los valores de intensidad del proceso espacial, presente en el efecto aleatorio espacialmente estructurado con ICAR y el efecto aleatorio no estructurado. Las estimaciones *a posteriori* de los efectos se presentan en el cuadro 4.11.

Cuadro 4.11. Resultados de efectos fijos para el modelo final

Efectos	media	sd	0.025q	0.5q	0.975q	moda	kld
Intercepto	-0.391	0.115	-0.521	-0.425	-0.082	-0.484	0.011
SMR anterior	0.000	0.022	-0.045	0.001	0.040	0.002	0.000
% urbano	-0.002	0.003	-0.006	-0.003	0.004	-0.005	0.004

El *intercepto* de -0.391 presenta un intervalo de credibilidad del 95% de (-0.521, -0.082). La covariable SMR del mes anterior presenta un efecto nulo en la cantidad de fallecidos con una media posterior, que tiende a cero; y un intervalo de credibilidad del 95% igual a (-0.045, 0.040). En el caso del porcentaje urbano por cantones, la media posterior es de -0.002 con un intervalo de (-0.006, 0.004). Eso indica un aporte marginal a la

predicción. El modelo final estimado, implica que las covariables no resultan adecuadas para predecir la variable respuesta. El modelo incluye efectos espacio-temporales de la interacción tipo II donde el efecto aleatorio del espacio y tiempo cobra mayor relevancia al ser las covariables poco útiles para predecir.

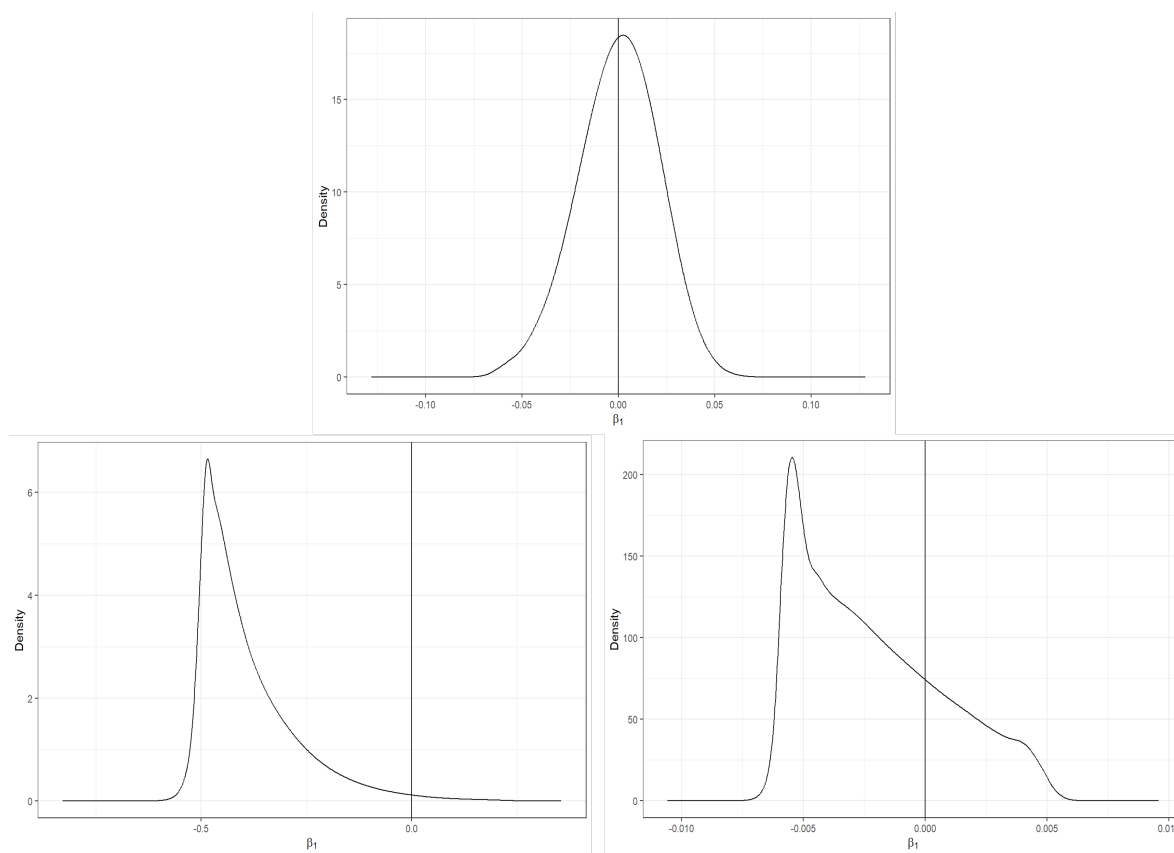


Figura 4.4. Representación gráfica de los marginales posteriores del intercepto (superior), el SMR del mes anterior (inferior izquierdo) y el porcentaje urbano (inferior derecho) en el modelo final

La figura 4.4 muestra las gráficas de la distribución posterior de los coeficientes del intercepto y las covariables. La distribución de los marginales posteriores, figuras 4.4, 4.5 y 4.6, muestra que las covariables tienen poca relación, dada la proximidad del pico de las distribuciones a cero. La excepción, en tal caso, sería la distribución de porcentaje urbano. Los gráficos muestran que el parámetro SMR (mes anterior) se ajusta, por ser simétrico (tiene media y moda similar).

Cuadro 4.12. Resultados de precisión de hiperparámetros para el modelo final

Precisión	media	sd	0.025q	0.5q	0.975q	moda
Observaciones BN	7.46	1.49	5.03	7.29	10.88	6.93
Cantón (iid)	2412.55	2557.53	210.53	1653.71	9205.86	582.10
Cantón (espacial)	3170.07	4072.09	320.74	1951.43	13454.47	820.12
Temporal estruct.	24375.47	27510.52	2169.17	16141.98	96241.52	5924.09
Temporal no estr.	20941.51	21317.80	1631.35	14651.14	77879.31	4542.65
Espacio-temporal	2783.30	2792.22	247.29	1966.71	10196.05	691.49

Las medidas de precisión de los hiperparámetros (cuadro 4.12) representan la inversa de la varianza del correspondiente efecto aleatorio. Es decir, un parámetro menor denota una mayor varianza. Al mismo tiempo refleja que dicho efecto tiene una mayor contribución a las variaciones relativas del riesgo. En el modelo se muestra la mayor contribución del efecto estructurado temporalmente, en relación con los riesgos relativos mensuales; posteriormente y en menor medida, contribuye al efecto espacial según el cantón.

La inclusión de efectos espacio-temporales ayuda a comprender cómo la enfermedad se ha extendido por el territorio en estudio. Específicamente, las estimaciones de efectos espaciales y temporales aleatorios y su interacción permiten asignar un riesgo relativo a cada unidad espacial, temporal o espacio-temporal bajo análisis. Estos riesgos relativos se obtienen exponenciando los parámetros espacio-temporales que describen la expresión $(r_i t)$ del modelo. La Figura 4.8 muestra los riesgos relativos a lo largo del tiempo en términos de los efectos temporales aleatorios estimados.

El riesgo relativo representado por el componente estructurado del efecto temporal aleatorio, captura la evolución de la pandemia en Costa Rica. El riesgo relativo era cercano a 0, a principios de abril de 2020, alcanzó un pico en marzo de 2021. Este hecho también se puede verificar comparando las estimaciones de los parámetros de precisión asociados a cada uno de los efectos aleatorios incluidos en los modelos, que se muestran en el cuadro 4.10.

La figura 4.5 muestra los mapas de los efectos aleatorios estructurados espacial-

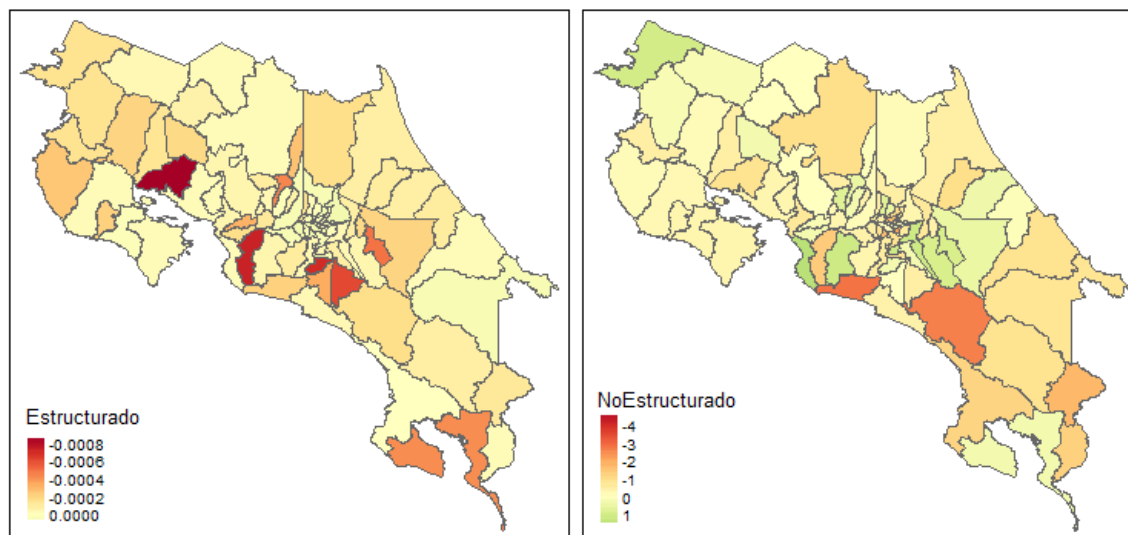


Figura 4.5. Representación gráfica de los valores aleatorios estructurados espacialmente y no estructurados

mente y no estructurados. Algo relevante que nos muestran ambos gráficos, es el hecho de que existe un patrón no constante del efecto aleatorio estructurado espacialmente. Esto permite concluir que la estimación del modelo no logra cubrir por completo estas intensidades. Lo cual apunta a que el modelo puede estar afectado por otros factores (covariables) espaciales que no han sido identificados e incluidos en el modelo. Además, el efecto aleatorio no estructurado muestra varias ubicaciones con alto valores que modifican la intensidad del proceso en ubicaciones individuales independientemente del resto. Los mapas muestran que existe variación residual estructurada espacialmente y no estructurada.

Conviene señalar en este punto que una falta de información, alguna omisión por parte de los organismos que deben brindarla y otros factores de tal índole, tendrán siempre un efecto negativo sobre los procesos investigativos, la aplicación de modelos y

la obtención de resultados reales.

Los mapas con riesgos relativos se muestran en la Figura 4.6 (marzo 2021) y el Anexo A (todos los meses estudiados en el modelo). La intensidad en la aparición de fallecidos por coronavirus es baja en la mayoría de los cantones. Existe una separación significativa entre cantones con grandes centros poblacionales y cantones menos poblados. Los riesgos relativos más bajos se encuentran en Abangares y Turrubares. Los más elevados han estado en el cantón de Mora y Santa Ana, es decir, hay cantones con riesgos relativos altos. La distribución espacial no se mantiene constante, ni homogénea.

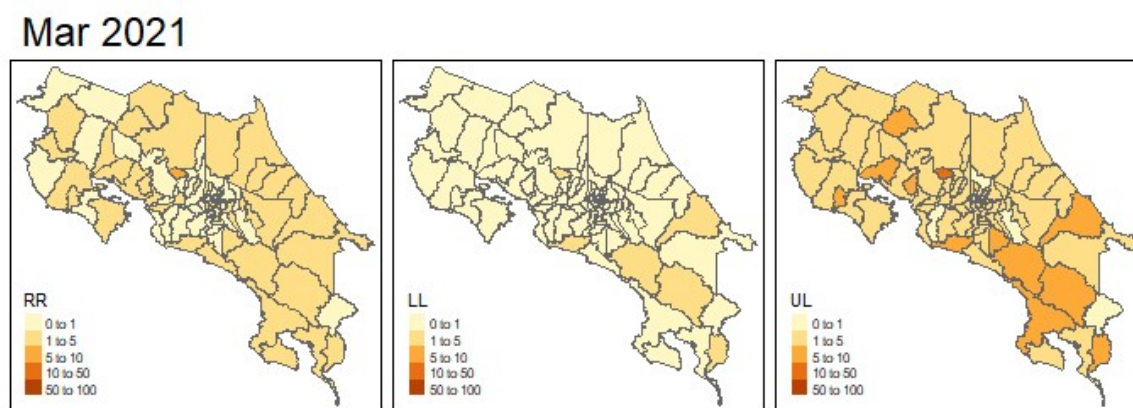


Figura 4.6. Representación gráfica de la media de riesgo relativo de coronavirus durante el período marzo 2021 por cantones en Costa Rica

Los límites de intervalos de credibilidad del 95% muestran la incertidumbre de estas predicciones. Debe tenerse en cuenta que, la formulación de riesgo relativo fue formulada con base en el modelo final. La inspección visual de las estimaciones de los riesgos relativos mensuales tiende a ser difícil. Un RR superior a 1 significa que la media posterior de los riesgos específicos del cantón, es mayor que el riesgo de incidencia promedio en los cantones de Costa Rica.

Abr 2021

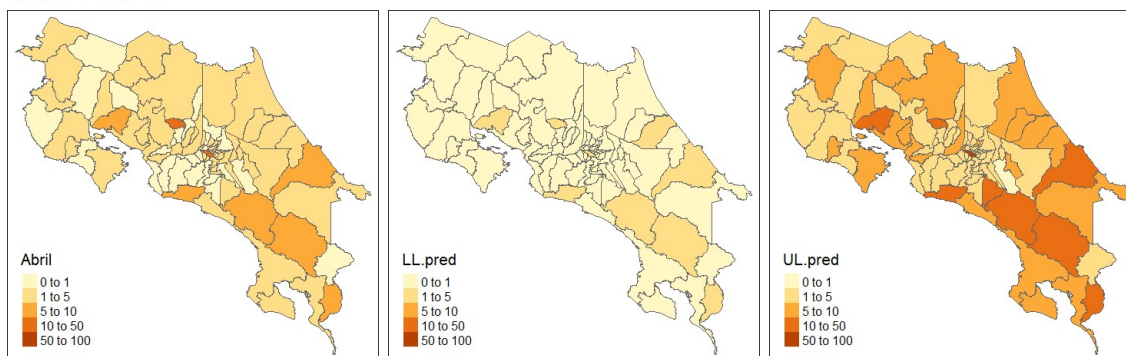


Figura 4.7. Representación gráfica de la media, y sus límites (inferior y superior) de la estimación de fallecidos por coronavirus durante el período abril 2021 por cantones en Costa Rica

4.6 Predicción de abril y mayo para 2021

Las figuras 4.7 y 4.8 muestran, respectivamente, la predicción para los meses de abril y mayo del 2021. Esta predicción (cuyos datos provienen de fuentes oficiales) indica que las medias con más fallecidos pronosticados, tienden a concentrarse en el centro del país; así también unos pocos cantones "periféricos". El Anexo B muestra gráficamente las predicciones por cantón según provincia, y los cuadros con las estimaciones predichas comparadas con los casos confirmados en Costa Rica, tanto para abril como para mayo de 2021. Dicho anexo permite determinar, además, que la predicción es acertada en 28 casos para el mes de abril comparando los intervalos del límite inferior y superior con el valor de ocurrencia real. En el caso de mayo, hay 17 predicciones correctas. Teniendo en cuenta que Costa Rica tiene 82 cantones, las predicciones podrían mejorarse (como

se explica más adelante).

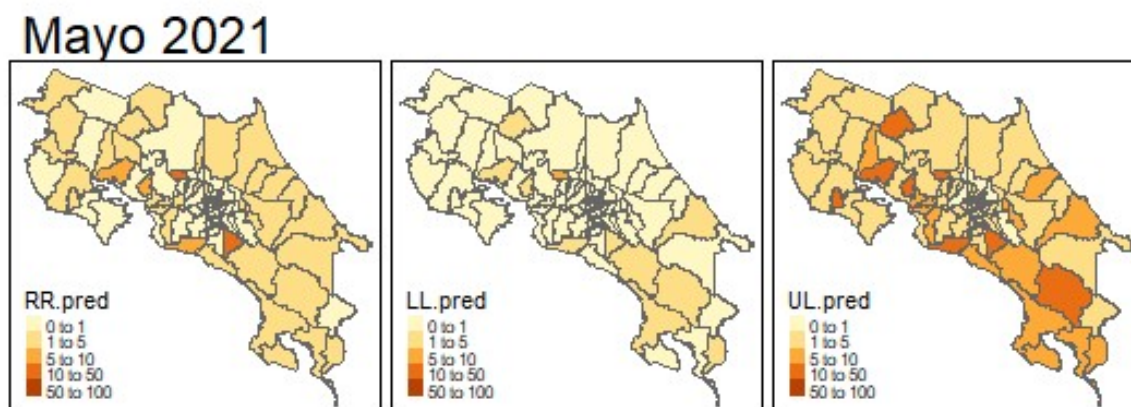


Figura 4.8. Representación gráfica de la media, y sus límites (inferior y superior) de la estimación de fallecidos por coronavirus durante el período Mayo 2021 por cantones en Costa Rica

Resulta complejo determinar las razones por las cuales las predicciones no coinciden con la realidad en numerosos casos. Una probable, ha de estar relacionada con la poca cantidad de periodos disponibles para ajustar el modelo. En este caso, se ha dispuesto de 12 periodos únicamente, a partir de los cuales se busca la predicción de 2 periodos futuros. A esto se debe agregar que en algunos cantones presentan casos nulos de fallecidos. Pese a que se trata de un aproximación con métodos bayesianos probabilísticos, el algoritmo cuenta con pocos periodos para pronosticar la información. El modelo presenta algunos problemas, especialmente al predecir casos nulos, así como los cantones mayor cantidad de fallecidos. La carencia de datos suficientes y precisos, así como la falta de una metodología estándar, inciden de forma relevante, en la acertabilidad del modelo. Además, pese a la inexistencia de una correlación alta poco evidente entre las

covariables en el modelo final, puede haber un problema de multicolinealidad.

Cabe señalar que los modelos bayesianos pueden generar conclusiones incongruentes, debido a su tratamiento de los parámetros como variables aleatorias y su uso de información subjetiva para establecer supuestos sobre estos parámetros. Otro problema con la serie, es la gran variabilidad de fallecidos entre meses. Esto complica la predicción, especialmente en los cantones con más fallecidos, pero que presentan periodos con volatilidad, variando de meses con casos nulos hasta meses con varias decenas de fallecidos.

CAPÍTULO 5

CONCLUSIONES Y RECOMENDACIONES

En esta última sección se presentan las conclusiones del trabajo. Se incluyen las reflexiones finales, se discuten brevemente los resultados y se sugieren lineamientos para futuras investigaciones del tema.

Ha sido posible generar aquí un modelo final de estimación bayesiana espacio-temporal, con el fin de abordar estadísticamente los casos de fallecimiento por Covid-19 en los cantones de Costa Rica. Un problema importante que se presenta en el modelo final, es que la covariable (SMR del mes anterior) presenta un efecto nulo sobre la estimación del modelo. Esto puede ser debido a multicolinealidad. Este modelo ha servido para mostrar que los niveles de asociación entre las covariables predictoras y la variable de respuesta son significativos. Aun así, dentro de las estimaciones se aprecia de manera significativa que existen covariables faltantes que expliquen la variabilidad de la variable de respuesta, lo cual afecta la exactitud de la predicción del modelo. En tal sentido, también ha debido insistirse a lo largo del trabajo, en la probable falta de exactitud y precisión de los datos, la cual incide, a todas luces en la adecuación de los cálculos con la realidad.

En estudios futuros, sería recomendable analizar las diferencias entre las cadenas de Márkov y el INLA en modelos espacio-temporales. Múltiples modelos en este estudio son adaptados de referencias que, en algunos casos, simularon modelos empleando cadenas de Márkov, no INLA. Un ejemplo relevante es el caso de Knorr-Held (2000), que propone varias formas de distribución previa en los modelos espacio-temporales.

Este estudio también adaptó modelos de estudios que emplearon INLA, como el de Briz-Redón (2021). Rodrigues (2017) también emplea INLA, así como distintas covariables para ajustar el modelo. En el presente trabajo se ha utilizado el SMR en forma analoga a lo que hace Moraga (2021), pues ha demostrado ser una covariable

adecuada en investigaciones de esta índole.

Se ha planteado una combinación de SMR, junto con algunas covariables presentes en diversos estudios a los que se ha hecho referencia. Más que predecir, en la mayoría de tales estudios se ha procurado ajustar los modelos, a fin de estimar riesgos relativos en el período de análisis.

En esta línea de investigación se requiere evaluar si el algoritmo de INLA propuesto por Rue *et al.* (2009) resulta apropiado o no, para los casos de ajuste de modelos espacio-temporales con conteos bajos o limitaciones en los datos. Probablemente, las estimaciones con este algoritmo resulten menos precisas que las obtenidas mediante el método de cadenas de Márkov.

El presente estudio ha enfrentado diversas limitaciones, entre las que se cuenta de forma primordial el acceso a información precisa. Probablemente, una eventual mejora en la disposición y acceso a los datos, una mayor cantidad de períodos disponibles, así como una mejoría en la definición metodológica establecida podrían, entre otros factores, resultar en una mejora del rendimiento predictivo de los modelos ajustados.

Cabe resaltar que el subregistro de casos asociados con el Covid-19, también es una situación que afecta la calidad de la información disponible. La forma en que se miden los casos positivos y los fallecidos por esta enfermedad ha sido objeto de discusión, ya que cabe establecer más certeza cuando una persona fallece por esta condición. La controversia sobre el conteo de fallecidos ha tenido una connotación global. Finalmente, la intensidad con que ha afectado la calidad de los datos disponibles al presente estudio resulta difícil de cuantificar, pero puede considerarse significativa. Este punto pudo haber afectado directamente los resultados que ha arrojado el modelo final. Esto también puede ser un punto por tener en cuenta en futuras investigaciones, ya que la calidad y veracidad de los datos puede incidir directamente en los resultados y conclusiones de un trabajo.

Fue difícil seleccionar, entre muchos, el método adecuado para hacer este trabajo.

La literatura existente es muy amplia. Ofrece diversas vías para investigar los procesos bayesianos espacio-temporales. Un aspecto importante que debería ser considerado en futuros estudios, es el que la tasa de detección de casos y fallecimientos se ha mantenido lejos de una detección completa desde el inicio de la pandemia Covid-19.

En Costa Rica, un problema importante es el acceso a la información veraz y debidamente actualizada; en ocasiones debe enfrentarse una notable falta de calidad en tal sentido. En este estudio, fue necesario ajustar numerosos datos, debido a que su registro se realiza por año. Algunas variables no se ajustaban por cantones, sino que obedecían a otro tipo de división administrativa. En el futuro, se pueden considerar otras covariables regionales, ambientales y socioeconómicas.

Al contar con poca variabilidad temporal, por falta de datos, la dificultad de convergencia de algunos modelos fue significativa. Eso podría tener un impacto sobre los resultados, de eventuales estudios que estimen coeficientes variabilidad espacial y temporal.

En esta investigación, se emplearon, modelos iniciales para determinar la mejor covariable principal. Para explicar las variaciones espaciales y temporales del riesgo relativo de los fallecidos por cantones en Costa Rica, también se consideraron: la mejor distribución de probabilidad, la inclusión de covariables adicionales y los 12 modelos espacio-temporales. Se emplearon criterios de información (DIC, WAIC y COP) para determinar los mejores modelos. A pesar de las limitaciones de los datos y la complejidad e incertidumbre en la propagación de Covid-19, los modelos logran captar tendencias temporales y espaciales en menor medida.

El modelo que mejor se ajustó a los requerimientos establecidos, fue el modelo con SMR del mes anterior y el porcentaje urbano por cantones (como covariables). Dicho modelo presenta, además, una estimación estructurada en el tiempo, pero no estructurada en el espacio; también ofrece un modelo aleatorio independiente (iid). El modelo final emplea una distribución binomial negativa. Al comprar las predicciones contra los valores reales para los meses de abril y mayo de 2021, el algoritmo logra

predecir correcta y respectivamente, 28 y 17 de los 82 cantones del país.

Respecto al porcentaje urbano, este trabajo resulta coincidente con otros estudios en cuanto a la inclusión de la covariable de marras. Covariables similares han resultado esenciales en otros modelos espacio-temporales bayesianos de ciertos estudios. Por ejemplo, Hu *et al.* (2018) emplearon promedio de residentes por comunidad, para analizar el crimen urbano en la ciudad de Wuhan en China. En el caso epidemiológico, por ejemplo, para Madden *et al.* (2021) y Jaya y Folmer (2021), los tamaños y densidades poblacionales resultaron esenciales en el estudio espacio-temporal del Covid-19.

Resulta peculiar y contradictorio, el hecho de que el modelo final tuviera un SMR del mes anterior con un valor que aproximara a cero; pero que según los criterios de información sería el modelo que mejor se ajusta a los datos. Ciertamente, la opción de emplear distintos modelos afecta el ajuste de los datos espacio-temporales. Puede evocarse, en tal sentido, a Briz-Redón (2021), quien analiza ciertas investigaciones con distintos procedimientos. En su conclusión advierte sobre ciertas contradicciones que se presentan al enfrentar estadísticamente esta temática, y sospecha de la metodología que suele emplearse en tal labor.

BIBLIOGRAFÍA

- Alamo, T., Reina, D. G., Mammarella, M., y Abella, A. (2020). Open data resources for fighting covid-19.
- Alvarado, W. R. (2011). Modelo geoestadístico espacio-temporal del crimen en el salvador: Análisis estructural y predictivo. *Revista de Matemática: Teoría y Aplicaciones*, 18(2):325.
- Anselin, L. (2010). Local indicators of spatial association-LISA. *Geographical Analysis*, 27(2):93–115.
- Arnold, J. B. (2021). *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*. R package version 4.2.4.
- Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3.
- Auguie, B. (2019). *egg: Extensions for 'ggplot2': Custom Geom, Custom Themes, Plot Alignment, Labelled Panels, Symmetric Scales, and Fixed Panel Size*. R package version 0.4.5.
- Aziz, S., Ngui, R., Lim, Y., Sholehah, I., Farhana, J., Abu samah, A., y Wan Sulaiman, W. (2012). Spatial pattern of 2009 dengue distribution in kuala lumpur using gis application. *Tropical biomedicine*, 29:113–20.
- Ballesteros, P., Salazar, E., Sánchez, D., y Bolanos, C. (2020). Spatial and spatiotemporal clustering of the COVID-19 pandemic in Ecuador. *Revista de la Facultad de Medicina*, 69(1).
- Banerjee, S., Carlin, B. P., y Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. CRC PR INC.

- Banerjee, S. y Gelfand, A. E. (2006). Bayesian wombling. *Journal of the American Statistical Association*, 101(476):1487–1501.
- Bastos, S. B. y Cajueiro, D. O. (2020). Modeling and forecasting the early evolution of the Covid-19 pandemic in Brazil. *Scientific Reports*, 10(1).
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.
- Besag, J. y Kooperberg, C. (1995). On conditional and intrinsic autoregression. *Biometrika*, 82(4):733.
- Besag, J., York, J., y Molli, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20.
- Bivand, R., Keitt, T., y Rowlingson, B. (2021). *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. R package version 1.5-27.
- Bivand, R. y Lewin-Koh, N. (2021). *maptools: Tools for Handling Spatial Objects*. R package version 1.1-2.
- Bivand, R. S. y Wong, D. W. S. (2018). Comparing implementations of global and local indicators of spatial association. 27(3):716–748.
- Blangiardo, M., Cameletti, M., Baio, G., y Rue, H. (2013). Spatial and spatio-temporal models with r-INLA. 4:33–49.
- Briz-Redón, Á. (2021). The impact of modelling choices on modelling outcomes: a spatio-temporal study of the association between COVID-19 spread and environmental conditions in Catalonia (Spain). *Stochastic Environmental Research and Risk Assessment*.
- Caja Costarricense de Seguro Social, C. (2021). Covid-19.

- Carpenter, T. (2011). The spatial epidemiologic (r)evolution: A look back in time and forward to the future. *Spatial and Spatio-temporal Epidemiology*, 2(3):119–124.
- Carroll, R., Lawson, A., Faes, C., Kirby, R., Aregay, M., y Watjou, K. (2015). Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. *Spatial and Spatio-temporal Epidemiology*, 14-15:45–54.
- CCP (2021). Centro centroamericano de población. bases de datos en línea.
- Charlton, M. (2009). Quantitative data. En *International Encyclopedia of Human Geography*, pp. 19–26. Elsevier.
- Cheng, J., Karambelkar, B., y Xie, Y. (2021). *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. R package version 2.0.4.1.
- Cheng, Y., Zhao, H., Song, P., Zhang, Z., Chen, J., y Zhou, Y.-H. (2019). Dynamic changes of lymphocyte counts in adult patients with severe pandemic h1n1 influenza a. *Journal of Infection and Public Health*, 12(6):878–883.
- Chou-Chen, S. W. y Alvarado-Barrantes, R. (2018). Un Modelo jerárquico Bayesiano espacio-temporal con variable de conteos: aplicación de VIH/SIDA en Costa Rica. *Comunicaciones en Estadística*, 11(1):9–35.
- Clayton, D. y Kaldor, J. (1987). Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43(3):671.
- Coly, S., Garrido, M., Abrial, D., y Yao, A.-F. (2021). Bayesian hierarchical models for disease mapping applied to contagious pathologies. *PLOS ONE*, 16(1):e0222898.
- Corporation, M. y Weston, S. (2020). *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. R package version 1.0.16.
- Cressie, N., Perrin, O., y Thomas-Agnan, C. (2005). Likelihood-based estimation for gaussian MRFs. *Statistical Methodology*, 2(1):1–16.

- Croner, C. M. (2004). Public health GIS and the internet. *Journal of Map & Geography Libraries*, 1(1):105–135.
- Dalziel, B. D., Kissler, S., Gog, J. R., Viboud, C., Bjørnstad, O. N., Metcalf, C. J. E., y Grenfell, B. T. (2018). Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. *Science*, 362(6410):75–79.
- Danon, L., Brooks-Pollock, E., Bailey, M., y Keeling, M. (2020). A spatial model of CoVID-19 transmission in England and Wales: early spread and peak timing.
- de Souza, A., Abreu, M. C., y de Oliveira-Júnior, J. F. (2021). Spatio-temporal analysis between the incidence of COVID-19 and human development in Mato Grosso do Sul, Brazil.
- Desjardins, M., Hohl, A., y Delmelle, E. (2020). Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied Geography*, 118:102202.
- Dong, E., Du, H., y Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5):533–534.
- Dragulescu, A. y Arendt, C. (2020). *xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files*. R package version 0.6.5.
- Everitt, B. S. y Skrondal, A. (2015). *The Cambridge Dictionary of Statistics*. Cambridge University Press.
- Fanchi, J. R. (2010). Reservoir characterization. En *Integrated Reservoir Asset Management*, pp. 187–204. Elsevier.
- Faraway, J., Marsaglia, G., Marsaglia, J., y Baddeley, A. (2021). *goftest: Classical Goodness-of-Fit Tests for Univariate Distributions*. R package version 1.2-3.

- Fay, R. E. y Herriot, R. A. (1979). Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, 74(366a):269–277.
- Fekedulegn, D., Andrew, M., Violanti, J., Hartley, T., Charles, L., y Burchfiel, C. (2010). Comparison of statistical approaches to evaluate factors associated with metabolic syndrome. 12(5):365–373.
- Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F., y Billa, L. (2020). Spatial analysis and GIS in the study of COVID-19. A review. *Science of The Total Environment*, 739:140033.
- Fuentes, M., Chen, L., y Davis, J. M. (2008). A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics*, 19(5):487–507.
- Gayawan, E., Awe, O. O., Oseni, B. M., Uzochukwu, I. C., Adekunle, A., Samuel, G., Eisen, D. P., y Adegboye, O. A. (2020). The spatio-temporal epidemic dynamics of COVID-19 outbreak in Africa. *Epidemiology and Infection*, 148.
- Ge, X.-Y., Li, J.-L., Yang, X.-L., Chmura, A. A., Zhu, G., Epstein, J. H., Mazet, J. K., Hu, B., Zhang, W., Peng, C., Zhang, Y.-J., Luo, C.-M., Tan, B., Wang, N., Zhu, Y., Cramer, G., Zhang, S.-Y., Wang, L.-F., Daszak, P., y Shi, Z.-L. (2013). Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*, 503(7477):535–538.
- Gelfand, A. (2010). *Handbook of spatial statistics*. CRC Press, Boca Raton.
- Gelman, A., Hwang, J., y Vehtari, A. (2013). Understanding predictive information criteria for bayesian models. *Statistics and Computing*, 24(6):997–1016.
- Geovisión-UNED (2021). Evolución de casos covid-19 costa rica.
- Getis, A. y Ord, J. K. (2010). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3):189–206.

- Ghosh, M., Natarajan, K., Stroud, T. W. F., y Carlin, B. P. (1998). Generalized linear models for small-area estimation. *Journal of the American Statistical Association*, 93(441):273–282.
- Ghosh, M. y Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9(1):55–76.
- Gilks, W. R. (1996). *Markov chain Monte Carlo in practice*. Chapman & Hall, London.
- Giuliani, D., Dickson, M. M., Espa, G., y Santi, F. (2020). Modelling and predicting the spatio-temporal spread of COVID-19 in Italy. *BMC Infectious Diseases*, 20(1).
- Gomes, D. S., Andrade, L. A., Ribeiro, C. J. N., Peixoto, M. V. S., Lima, S. V. M. A., Duque, A. M., Cirilo, T. M., Góes, M. A. O., Lima, A. G. C. F., Santos, M. B., Araújo, K. C. G. M., y Santos, A. D. (2020). Risk clusters of COVID-19 transmission in northeastern Brazil: prospective space–time modelling. *Epidemiology and Infection*, 148.
- Gross, B., Zheng, Z., Liu, S., Chen, X., Sela, A., Li, J., Li, D., y Havlin, S. (2020). Spatio-temporal propagation of COVID-19 pandemics. *EPL (Europhysics Letters)*, 131(5):58003.
- Grothendieck, G. (2017). *sqldf: Manipulate R Data Frames Using SQL*. R package version 0.4-11.
- Hazelton, M. L. (2017). Testing for changes in spatial relative risk. *Statistics in Medicine*, 36(17):2735–2749.
- He, X., Dong, S., Li, L., Liu, X., Wu, Y., Zhang, Z., y Mei, S. (2020). Using a Bayesian spatiotemporal model to identify the influencing factors and high-risk areas of hand, foot and mouth disease (HFMD) in Shenzhen. *PLOS Neglected Tropical Diseases*, 14(3):e0008085.

- Henao-Cespedes, V., Garcés-Gómez, Y. A., Ruggeri, S., y Henao-Cespedes, T. M. (2021). Relationship analysis between the spread of COVID-19 and the multidimensional poverty index in the city of Manizales, Colombia. *The Egyptian Journal of Remote Sensing and Space Science*.
- Hernández-González, G. (2019). Modelo espacial Bayesiano para la dinámica de transmisión de dengue en Puerto Rico para datos del 2014. *InterSedes*, 20:167 – 187.
- Hijmans, R. J. (2021). *raster: Geographic Data Analysis and Modeling*. R package version 3.4-13.
- Hooten, M. B., Pugh, S., y Roland, C. A. (2020). Geary's contiguity ratio (geary's c).
- Hu, T., Zhu, X., y Duan, L. (2018). Urban crime prediction based on spatio-temporal bayesian model. *PLOS ONE*, 13:e0206215.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., y Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395(10223):497–506.
- Hägerstrand, T. (2005). What about people in regional science? *Papers in Regional Science*, 24(1):7–24.
- Jalilian, A. y Mateu, J. (2021). A hierarchical spatio-temporal model to analyze relative risk variations of COVID-19: a focus on Spain, Italy and Germany. *Stochastic Environmental Research and Risk Assessment*, 35(4):797–812.
- Jaya, I. G. N. M. y Folmer, H. (2021). Bayesian spatiotemporal forecasting and mapping of COVID-19 risk with application to west java province, indonesia. *Journal of Regional Science*, 61(4):849–881.

- Kandel, N., Chungong, S., Omaar, A., y Xing, J. (2020). Health security capacities in the context of COVID-19 outbreak: an analysis of international health regulations annual report data from 182 countries. *The Lancet*, 395(10229):1047–1053.
- Kang, D., Choi, H., Kim, J.-H., y Choi, J. (2020). Spatial epidemic dynamics of the COVID-19 outbreak in China. *International Journal of Infectious Diseases*, 94:96–102.
- Kargupta, H. (2009). *Next generation of data mining*. CRC Press, Boca Raton.
- Kassambara, A. (2020). *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.4.0.
- Khan, K. (2018). Spatio-temporal reconstruction of remote sensing observations. mathesis, University of Arkansas. Theses and Dissertations.
- Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, 19(17-18):2555–2567.
- Körner, C., May, M., y Wrobel, S. (2012). Spatiotemporal modeling and analysis—introduction and overview. *KI - Künstliche Intelligenz*, 26(3):215–221.
- Langousis, A. y Carsteanu, A. A. (2020). Undersampling in action and at scale: application to the COVID-19 pandemic. *Stochastic Environmental Research and Risk Assessment*, 34(8):1281–1283.
- Lee, D. (2011). A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spatial and Spatio-temporal Epidemiology*, 2(2):79–89.
- Lindgren, F. y Rue, H. (2015). Bayesian spatial and spatiotemporal modelling with R-INLA. *J. Stat. Softw.*, 63.
- Liu, K., He, M., Zhuang, Z., He, D., y Li, H. (2020). Unexpected positive correlation between human development index and risk of infections and deaths of COVID-19 in Italy. *One Health*, 10:100174.

- López, D. y Blobel, B. (2007). Semantic interoperability between clinical and public health information systems for improving public health services. *Studies in health technology and informatics*, 127:256–67.
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J., Meng, Y., Wang, J., Lin, Y., Yuan, J., Xie, Z., Ma, J., Liu, W. J., Wang, D., Xu, W., Holmes, E. C., Gao, G. F., Wu, G., Chen, W., Shi, W., y Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, 395(10224):565–574.
- Ma, H., Carlin, B. P., y Banerjee, S. (2009). Hierarchical and joint site-edge methods for medicare hospice service region boundary analysis. *Biometrics*, 66(2):355–364.
- Madden, J. M., More, S., Teljeur, C., Gleeson, J., Walsh, C., y McGrath, G. (2021). Population mobility trends, deprivation index and the spatio-temporal spread of coronavirus disease 2019 in ireland. *International Journal of Environmental Research and Public Health*, 18(12):6285.
- Martins, T. G., Simpson, D., Lindgren, F., y Rue, H. (2013). Bayesian computing with INLA: New features. *Computational Statistics & Data Analysis*, 67:68–83.
- McIlroy, D. (2020). *mapproj: Map Projections*. R package version 1.2.7. Packaged for R by Ray Brownrigg and Thomas P Minka and transition to Plan 9 codebase by Roger Bivand.
- Militino, A. F., Ugarte, M. D., y Pérez-Goya, U. (2018). An introduction to the spatio-temporal analysis of satellite remote sensing data for geostatisticians. pp. 239–253. Springer International Publishing.
- MODIS (2021). Moderate resolution imaging spectroradiometer.
- Moraga, P. (2018). Small Area Disease Risk Estimation and Visualization Using R. *The R Journal*, 10(1):495–506.

- Moraga, P. (2021). Species Distribution Modeling using Spatial Point Processes: a Case Study of Sloth Occurrence in Costa Rica. *The R Journal*, 12(2):293–310.
- Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-2.
- Niu, X., Yue, Y., Zhou, X., y Zhang, X. (2020). How Urban Factors Affect the Spatiotemporal Distribution of Infectious Diseases in Addition to Intercity Population Movement in China. *ISPRS International Journal of Geo-Information*, 9(11):615.
- Nøhr, C., Sørensen, E. M., Gudes, O., Geraghty, E. M., Shaw, N. T., Bivona-Tellez, C., y and, A. K. L. (2014). A review and framework for categorizing current research and development in health related geographical information systems (GIS) studies. *Yearbook of Medical Informatics*, 23(01):110–124.
- Nykiforuk, C. I. J. y Flaman, L. M. (2009). Geographic information systems (GIS) for health promotion and public health: A review. *Health Promotion Practice*, 12(1):63–73.
- Obaromi, D. (2019). Spatial Modelling of Some Conditional Autoregressive Priors in A Disease Mapping Model: the Bayesian Approach. *Biomedical Journal of Scientific & Technical Research*, 14(3).
- OMS, O. (2020). Services, integrated health / who headquarters, (hq) / worldwide, w. h. o. maintaining essential health services:operational guidance for thecovid-19 context 2020-06.
- Ooms, J. (2021). *magick: Advanced Graphics and Image-Processing in R*. R package version 2.7.3.
- Ord, J. K. y Getis, A. (2010). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, 27(4):286–306.
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1):439–446.

- Peng, L., Yang, W., Zhang, D., Zhuge, C., y Hong, L. (2020a). Epidemic analysis of COVID-19 in China by dynamical modeling.
- Peng, R. D., Murdoch, D., Rowlingson, B., y Murta, A. (2020b). *gpclib: General Polygon Clipping Library for R*. R package version 1.5-6.
- Peterson, B. G. y Carl, P. (2020). *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis*. R package version 2.0.4.
- Pettit, L. I. (1990). The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1):175–184.
- Pettitt, A. N., Weir, I. S., y Hart, A. G. (2002). *Statistics and Computing*, 12(4):353–367.
- PNUD (2020). Programa de las Naciones Unidas para el Desarrollo. Índice de Desarrollo Humano (IDH).
- Polo, G., Acosta, C. M., Soler-Tovar, D., Villamil, J. F. P., Palencia, N. P., Penagos, M., Martínez, J. M., Bobadilla, J. N., Martín, L. V., Durán, S., Rodríguez, M., Carvajalino, C. M., Villamil, L. C., y Ortiz, E. B. (2020). Bayesian Spatio-Temporal Modeling of COVID-19: Inequalities on Case-Fatality Risk.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Raghunathan, T. E., Xie, D., Schenker, N., Parsons, V. L., Davis, W. W., Dodd, K. W., y Feuer, E. J. (2007). Combining information from two surveys to estimate county-level prevalence rates of cancer risk factors and screening. *Journal of the American Statistical Association*, 102(478):474–486.
- Rappole, J. y Hubalek, Z. (2006). Birds and Influenza H5N1 Virus Movement to and within North America. *Emerging Infectious Diseases*, 12(10):1486–1492.

- Reeder, B., Revere, D., Hills, R. A., Baseman, J. G., y Lober, W. B. (2012). Public health practice within a health information exchange: Information needs and barriers to disease surveillance. *Online Journal of Public Health Informatics*, 4(3).
- Richardson, S., Abellan, J. J., y Best, N. (2006). Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (UK). *Statistical Methods in Medical Research*, 15(4):385–407.
- Robinson, D., Hayes, A., y Couch, S. (2021). *broom: Convert Statistical Objects into Tidy Tibbles*. R package version 0.7.9.
- Rodrigues, . A. (2017). *Spatio-temporal modelling of tornados with R-INLA, at the county-level in Texas and Ocklahoma*. Tesis doctoral.
- Rohleder, S. y Bozorgmehr, K. (2021). Monitoring the spatiotemporal epidemiology of covid-19 incidence and mortality: A small-area analysis in germany. *Spatial and Spatio-temporal Epidemiology*, 38:100433.
- Rouse, J. W., J. H. R. H. D. D. W. (1974). Monitoring vegetation systems in the great plains with erts. *NASA. Goddard Space Flight Center 3d ERTS-1 Symp*, 1.
- Rue, H. (2005). *Gaussian Markov random fields : theory and applications*. Chapman & Hall/CRC, Boca Raton.
- Rue, H., Martino, S., y Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.
- Rueda, A. (2021). Costa Rica llama a democratizar la distribución de vacunas contra Covid-19.
- Saavedra, P., Santana, A., Bello, L., Pacheco, J.-M., y Sanjuán, E. (2021). A bayesian spatio-temporal analysis of mortality rates in spain: application to the COVID-19 2020 outbreak. *Population Health Metrics*, 19(1).

- Sah, P., Fitzpatrick, M. C., Zimmer, C. F., Abdollahi, E., Juden-Kelly, L., Moghadas, S. M., Singer, B. H., y Galvani, A. P. (2021). Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. 118(34):e2109229118.
- Saran, S., Singh, P., Kumar, V., y Chauhan, P. (2020). Review of geospatial technology for infectious disease surveillance: Use case on COVID-19. *Journal of the Indian Society of Remote Sensing*.
- Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5.
- Sartorius, B., Lawson, A. B., y Pullan, R. L. (2021). Modelling and predicting the spatio-temporal spread of COVID-19, associated deaths and impact of key risk factors in England. *Scientific Reports*, 11(1).
- Scala, A., Flori, A., Spelta, A., Brugnoli, E., Cinelli, M., Quattrocioni, W., y Pammolli, F. (2020). Time, space and social interactions: exit mechanisms for the covid-19 epidemics. *Scientific Reports*, 10(1).
- Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., y Crowley, J. (2021). *GGally: Extension to 'ggplot2'*. R package version 2.1.2.
- Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., y van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.
- Sujath, R., Chatterjee, J. M., y Hassanien, A. E. (2020). A machine learning forecasting model for COVID-19 pandemic in India. *Stochastic Environmental Research and Risk Assessment*, 34(7):959–972.

- Sutanto, H. T., Pramoedyo, H., Wardhani, W. S., y Astutik, S. (2021). The selection of bayesian polynomial regression with INLA by using DIC, WAIC and CPO. *Journal of Physics: Conference Series*, 1747(1):012029.
- Tennekes, M. (2018). tmap: Thematic maps in R. *Journal of Statistical Software*, 84(6):1–39.
- Tennekes, M. (2021). *tmertools: Thematic Map Tools*. R package version 3.1-1.
- Tian, H., Liu, Y., Li, Y., Wu, C.-H., Chen, B., Kraemer, M. U. G., Li, B., Cai, J., Xu, B., Yang, Q., Wang, B., Yang, P., Cui, Y., Song, Y., Zheng, P., Wang, Q., Bjornstad, O. N., Yang, R., Grenfell, B. T., Pybus, O. G., y Dye, C. (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*, 368(6491):638–642.
- Tuck, S. L., Phillips, H. R., Hintzen, R. E., Scharlemann, J. P., Purvis, A., y Hudson, L. N. (2014). MODISTools – downloading and processing MODIS remotely sensed data in R. *Ecology and Evolution*, 4(24):4658–4668.
- Unwin, D. J. (1996). GIS, spatial analysis and spatial statistics. *Progress in Human Geography*, 20(4):540–551.
- Vaughan, D. y Dancho, M. (2021). *furrr: Apply Mapping Functions in Parallel using Futures*. R package version 0.2.3.
- Villandr e, L., Plante, J.-F., Duchesne, T., y Brown, P. (2021). Fully Bayesian inference for spatiotemporal data with the multi-resolution approximation.
- Wakefield, J. (2004). A critique of statistical aspects of ecological studies in spatial epidemiology. *Environmental and Ecological Statistics*, 11(1):31–54.
- Wakefield, J. (2006). Disease mapping and spatial regression with count data. *Bio-statistics*, 8(2):158–183.
- Waller, G. (2004). *Applied Spatial Statistics*. John Wiley Sons.

- Waller, L. y Carlin, B. (2010). Disease mapping. En *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*, pp. 217–243. CRC Press.
- Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(116):3571–3594.
- Wei, T. y Simko, V. (2021). *R package 'corrplot': Visualization of a Correlation Matrix*. (Version 0.90).
- Weier, J. y Herring, D. (2000). Measuring vegetation (ndvi evi). *NASA Earth Observatory*.
- Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H. (2021a). *forcats: Tools for Working with Categorical Variables (Factors)*. R package version 0.5.1.
- Wickham, H. (2021b). *tidyr: Tidy Messy Data*. R package version 1.1.3.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., y Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.
- Wickham, H. y Bryan, J. (2019). *readxl: Read Excel Files*. R package version 1.3.1.
- Wickham, H., François, R., Henry, L., y Müller, K. (2021a). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.7.

- Wickham, H., Hester, J., y Chang, W. (2021b). *devtools: Tools to Make Developing R Packages Easier*. R package version 2.4.2.
- Wickham, H. y Seidel, D. (2020). *scales: Scale Functions for Visualization*. R package version 1.1.1.
- WIKLE, C. K., BERLINER, L. M., y CRESSIE, N. (1998). *Environmental and Ecological Statistics*, 5(2):117–154.
- Woo, P. C. Y., Lau, S. K. P., Lam, C. S. F., Lau, C. C. Y., Tsang, A. K. L., Lau, J. H. N., Bai, R., Teng, J. L. L., Tsang, C. C. C., Wang, M., Zheng, B.-J., Chan, K.-H., y Yuen, K.-Y. (2012). Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavirus. *Journal of Virology*, 86(7):3995–4008.
- Yan, Y. (2016). *MLmetrics: Machine Learning Evaluation Metrics*. R package version 1.1.1.
- Zhen, Z., Jing-min, J., y Liu, F. (2010). The application of geographic information system (GIS) in the field of public health. En *2010 Second IITA International Conference on Geoscience and Remote Sensing*. IEEE.

APÉNDICES

A Análisis inicial de datos exploratorios

El Covid-19 se ha extendido rápidamente por todo el mundo desde el 2019 y ha afectado la vida de muchas personas y naciones. Comprender la propagación de este virus en función de patrones espacio-temporales permitirá a las agencias planificar y utilizar mejor los recursos. La pandemia de Covid-19 llegó a Costa Rica en marzo de 2020, y durante abril del mismo año se empezaron a registrar los primeros fallecidos. El valle central, que es el núcleo y la zona más poblada del país, ha resultado ser el área más afectada, como se logra apreciar en la figura 9.9.

Fallecidos por COVID-19 en Costa Rica por cantón hasta marzo 2021

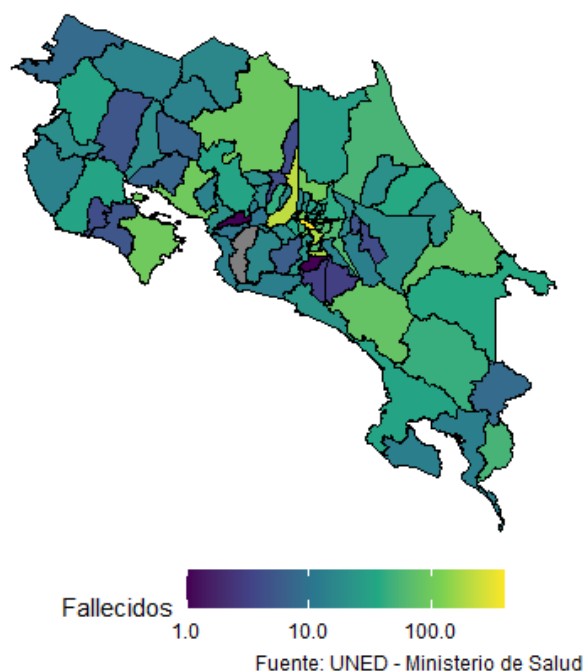


Figura 9.9. Representación gráfica de fallecidos por Covid-19 por cantón en Costa Rica hasta marzo 2021

Para este estudio los casos de Covid-19 se agregan según el mes epidemiológico, analizando los casos desde abril de 2020 hasta marzo de 2021. En la figura 9.9, se

puede ver la conformación cantonal del país. La figura 9.10, muestra que con el tiempo la cantidad de fallecidos por cantón debido al Covid-19 comienza a crecer. Sin embargo, necesitaremos utilizar un análisis geoespacial más localizado para comprender las relaciones temporales. Se puede observar que la distribución de las personas que perecieron por cantón de Covid-19 a lo largo de los meses está sesgada a la derecha. Inicialmente, en el mes de abril de 2020, la mayoría de los cantones tenían cero fallecidos, pero esto ha cambiado con los meses, y hasta marzo de 2021, únicamente el cantón de Turruabares aun no tenía fallecidos.

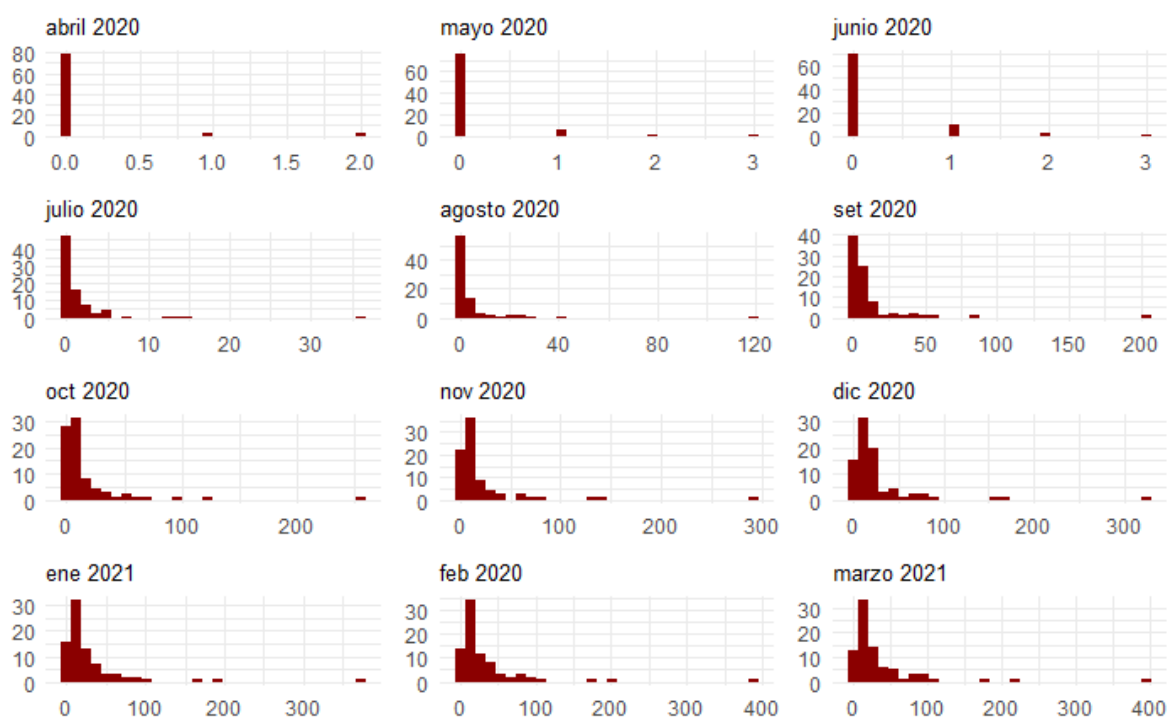


Figura 9.10. Representación gráfica de la distribución de fallecidos por Covid-19 a través del tiempo

Inicialmente, resulta adecuado analizar, mediante el mapa de coropletas, la distribución de las tasas de Covid-19 a nivel cantonal. Los mapas de coropletas, utilizados para visualizar la distribución espacial de los fallecidos de Covid-19 en los cantones de Costa Rica, se elaboran a lo largo de los meses. La figura 9.11, muestra la expansión de los fallecidos por el coronavirus; cómo empieza en el centro del país y en cantones en el

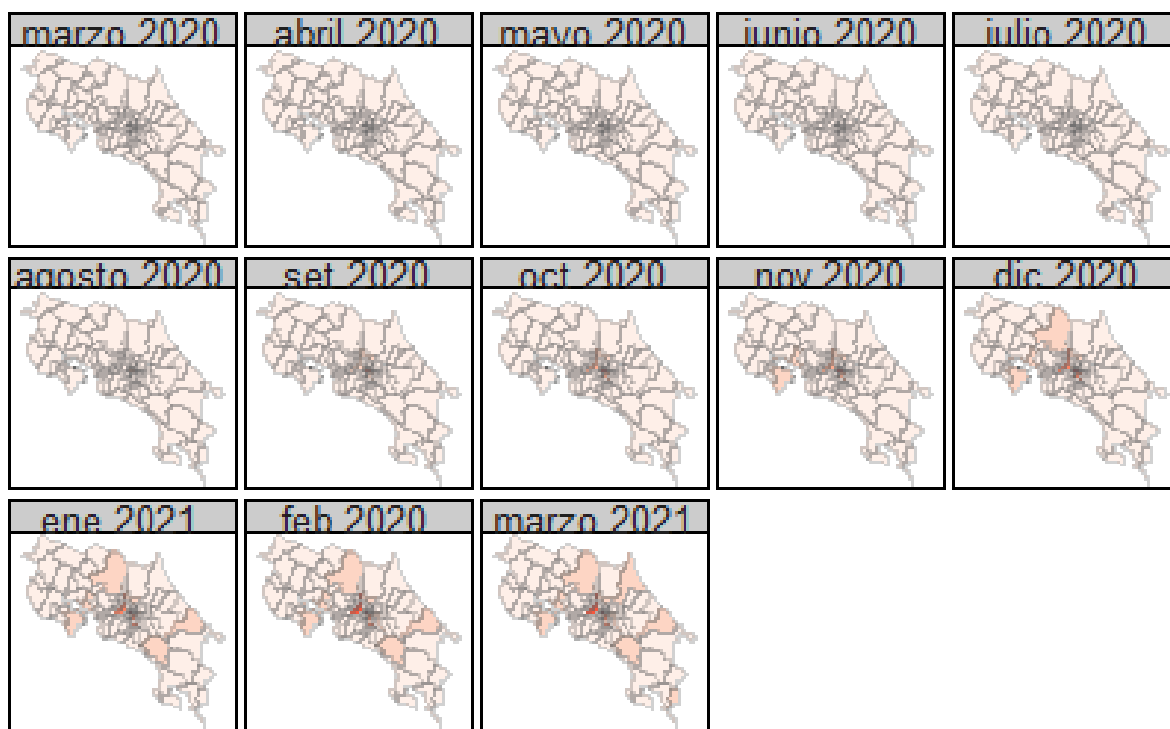


Figura 9.11. Representación gráfica de rupturas personalizadas

norte del país. Se nota que el virus se ha expandido, principalmente, cobrando víctimas que se ubican cada vez más en el centro y el norte del país. Paulatinamente, en julio de 2020 aparecen cantones de la zona sur y del Caribe del país con personas fallecidas. En este caso, para permitir una base común de comparación de las cantidades de fallecidos por Covid-19 a lo largo del tiempo, se debe utilizar un esquema de clasificación común para las tasas de Covid-19. La clasificación de mapas con los mismos intervalos de clase, permite comparaciones más fáciles, y con ello, la interpretación a largo plazo.

Para dar cuenta de estas rupturas personalizadas; es decir, para guiar la especificación de los puntos de corte, primero se han estudiado las estadísticas descriptivas de las tasas de Covid-19 a lo largo de los meses. De abril 2020 hasta marzo 2021, se puede apreciar que el intervalo es de 0 a 400 fallecidos. Un tono más oscuro de rojo indica una cantidad de fallecidos de Covid-19 más alta. Se puede observar que los fallecidos de Covid-19 parecen originarse en la parte central del país y en menor medida en el norte,

extendiéndose radialmente hacia afuera en todas direcciones a los cantones vecinos con el tiempo. Para marzo de 2021, los fallecidos por Covid-19 seguían encontrándose más concentrados en los cantones cercanos a la ubicación de origen.



Figura 9.12. Representación gráfica de rupturas naturales

Ahora se pueden observar patrones espaciales para todos los meses, gracias a la clasificación de rupturas naturales en la figura 9.12. Los casos de Covid-19 parecen extenderse radialmente desde los cantones ubicados en el centro de Costa Rica y el norte del país, utilizando un esquema de clasificación común como en la figura 9.11.

El análisis de datos preliminares mediante mapas de coropletas resulta deficiente; dadas las limitaciones de tales mapas para representar datos geoespaciales de una manera estadísticamente sólida. Por lo tanto, los métodos de análisis geoespacial localizado se utilizarán en la siguiente sección, solo para detectar estadísticamente conglomerados espaciales, valores atípicos, puntos calientes y puntos fríos relativos a los fallecidos por Covid-19 en Costa Rica.

Existen varios métodos para determinar la matriz de vecinos más cercanos. Para este estudio se evalúa el método de Queen y Rook. Basado en la matriz de vecinos de Queen, que hay 82 unidades de área, correspondientes a los cantones de Costa Rica y la unidad de área más conectada tiene 13 vecinos. Hay 5 unidades de área con solo dos vecinos y en promedio se tiene 5,097561 vecinos. Por su parte la matriz de Rook, identifica también 82 cantones, pero en este análisis, la unidad de área más conectada tiene 12 vecinos, 6 cantones con dos vecinos y una media de 4,829268 vecinos.

Un problema que presentan las matrices de ponderación basadas en la contigüidad, es que el número de vecinos en el conjunto de datos no es consistente. Eso podría deberse a las variaciones en los tamaños de los cantones en el país. Una matriz de peso más apropiada podría estar basada en la distancia. Hay dos matrices principales de tal índole: la fija y la adaptativa.

Son apropiadas las matrices adaptativas basadas en la distancia. Se ajustan a sí mismas en función de la densidad de datos. Podrían ser relevantes, porque la densidad poblacional de los cantones de Costa Rica varía en relación con las áreas respectivas.

En el caso de la distancia fija hay dos regiones sin vecinos. El número medio de enlaces que utilizan la medición de distancia fija es 17,53659, el cual resulta más alto que el de las mediciones de Rook y Queen. El cantón más conectado tiene 37 vecinos. Se puede observar en la parcela que, utilizando el método de distancia fija, las regiones con un mayor número de cantones más pequeños ubicados en el área tienen un mayor número de vecinos. La distancia mínima requerida para un municipio relativamente aislado, que tenga al menos un vecino, es mucho mayor que la distancia entre vecinos de un municipio ubicado en alguna región más densa.

Eso indica que los cantones de Costa Rica están espaciados de manera irregular, y por ello, varía el tamaño de los polígonos. En otras palabras, el uso de una distancia mínima puede no ser ideal, ya que da como resultado disparidades significativas en el número de vecinos. Esto puede producir variaciones significativas en la estimación de los datos cuando son escasos; inversamente puede enmascarar variaciones locales sutiles

donde los datos son densos.

Se estima aquí una matriz de peso estandarizada por filas. Está basada en los 17 vecinos más cercanos. Los pesos se asignan a cada cantón vecino. Se realiza la estandarización, porque el grado de conexión de un área (suma de pesos de vecinos) depende del número de sus vecinos si la matriz de ponderaciones no está estandarizada. Esto revela heterogeneidad entre cantones. Su análisis sirve posteriormente, para calcular la autocorrelación espacial global. Sus condiciones se establecen con una prueba de hipótesis:

- h_0 = Los patrones geográficos se distribuyen aleatoriamente
- h_1 = Los patrones geográficos no se distribuyen aleatoriamente

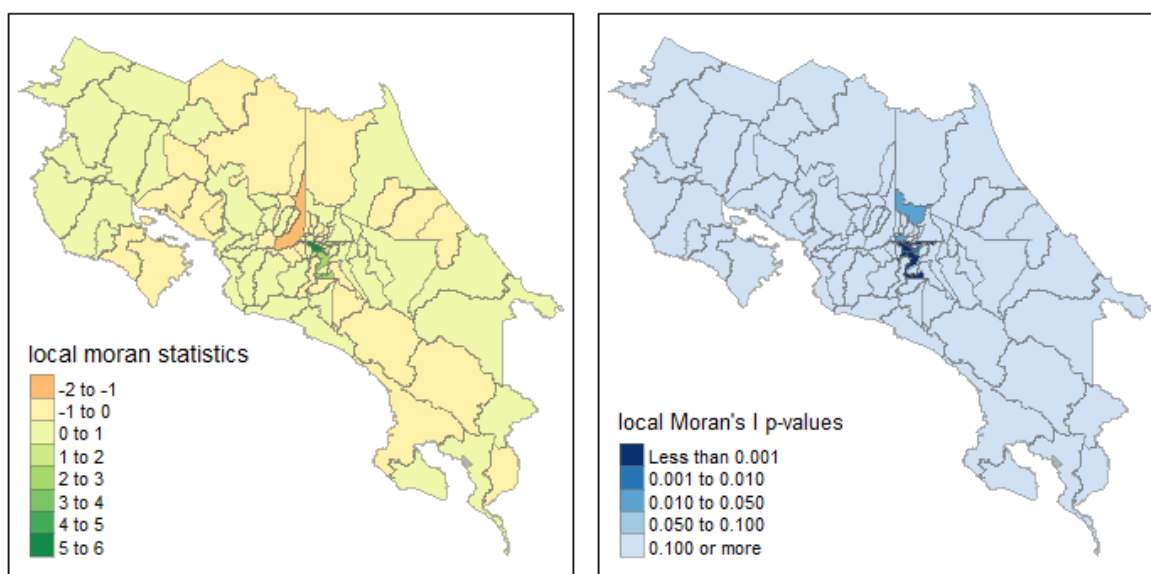


Figura 9.13. Representación gráfica del Índice de Morán y sus p-valores

Se emplea un valor crítico = 0.05 para analizar la hipótesis calculando las estadísticas I de Moran globales. Esta prueba ayuda a analizar si existe disociación o similitud

entre valores. Del mapa de la figura 9.13 se infiere que existe un agrupamiento espacial de los fallecidos de Covid-19 en el centro del país durante marzo de 2021. Para hacer esto, se emplea un análisis geoespacial localizado utilizando el análisis I de Moran local y el análisis G_i^* de Getis-Ord local. Se continúa utilizando la matriz de ponderación espacial adaptativa basada en la distancia.

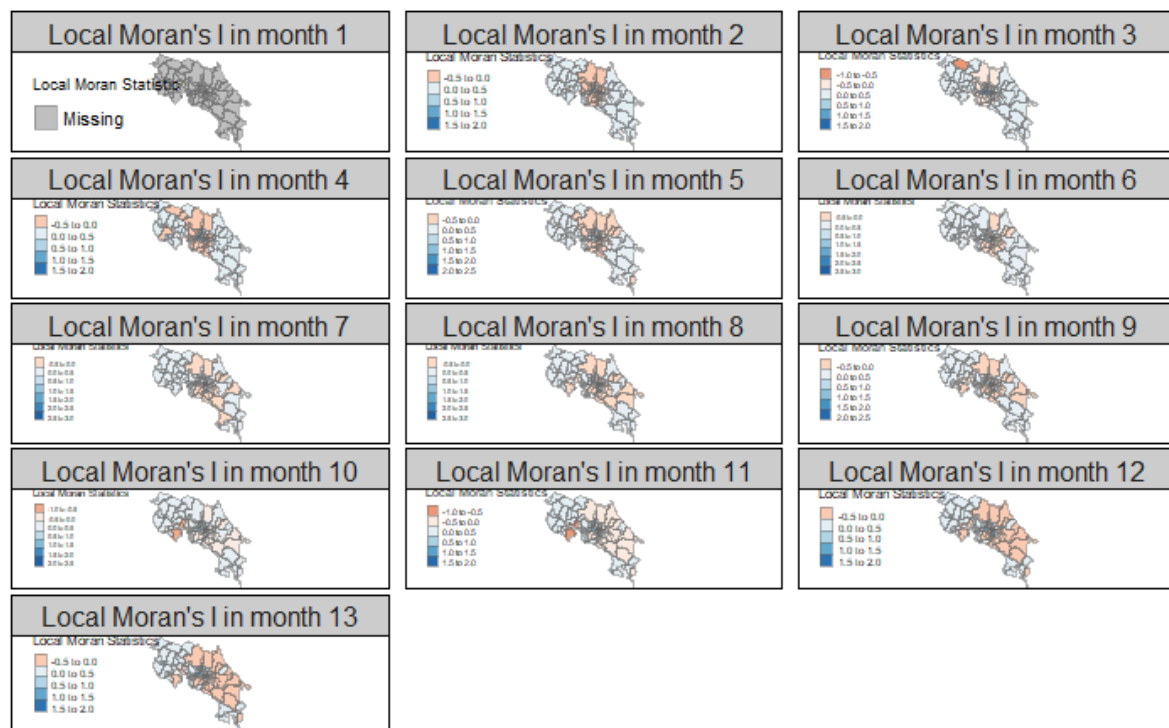


Figura 9.14. Representación gráfica del Índice de Morán y sus p-valores

En las figuras 9.13 y 9.14, las áreas azules (I local de Moran positivo) indican autocorrelación local espacial positiva. En cambio las áreas rojas (I de Moran local negativo) indican autocorrelación local espacial negativa. La figura 9.14 (coropleta local de Morán), muestra evidencia de valores I de Morán locales positivos y negativos al inicio y al final del período de tiempo epidemiológico que se está estudiando. No obstante, para interpretar si los valores tienen significancia estadística, es más útil considerar las estadísticas locales de Morán con sus valores p correspondientes.

Se puede observar que, desde los primeros meses, existen valores I de Morán locales

relativamente altos y positivos en los cantones centrales del país. Dichos valores son estadísticamente significativos. Se infiere, por lo tanto, la existencia de una autocorrelación local, la cual resulta significativa y positiva para la cantidad de fallecidos de Covid-19 en estas áreas, y sugiere una agrupación de cantones.

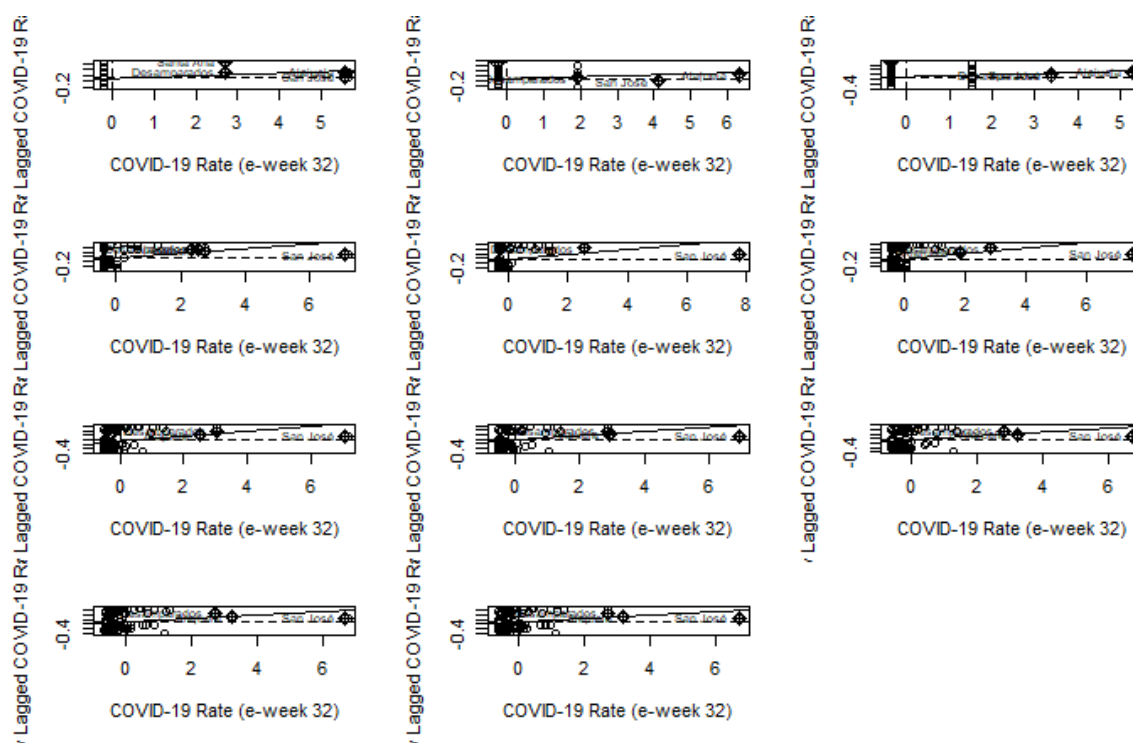


Figura 9.15. Representación gráfica del diagrama de dispersión del Índice de Morán

Cuando se estudia con el diagrama de *sacatterplot* de Morán (figura 9.15), se aprecian mejor los municipios que son estadísticamente significativos y su autocorrelación. Con base en este análisis desde el primer mes hasta el último, existe una gran agrupación de cantones; Desamparados es atípico. Hay un patrón de agrupamiento en términos de fallecimientos de Covid-19; y sugiere que estos cantones son puntos críticos potenciales para los casos de esta enfermedad.

Se puede observar, además, la dirección y la magnitud de la autocorrelación global a través de un diagrama de dispersión de Morán. También se ve la pendiente de la regresión lineal de los valores rezagados de las tasas de Covid-19, en contraste a las

frecuencias originales; eso resulta equivalente a la puntuación del I de Morán.

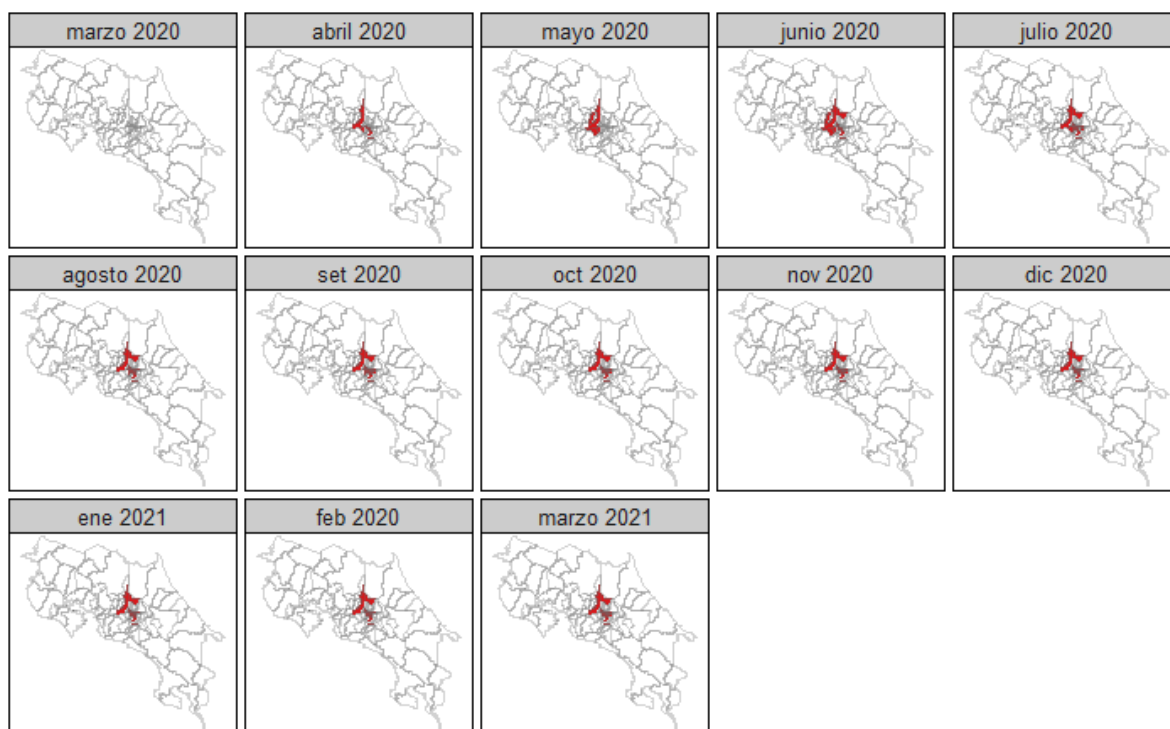


Figura 9.16. Representación gráfica del diagrama de dispersión del Índice de Morán

Mediante estos diagramas de dispersión, se muestran las autocorrelaciones negativas y positivas de cada municipio con la tasa de Covid-19. Pero se necesitan análisis complementarios para comprender su importancia estadística. No es fácil señalar las relaciones estadísticamente significativas de estos gráficos. La cartografía de conglomerados LISA (figura 9.16) podría ayudar a visualizar fácilmente todo esto en un mapa.

A partir de las gráficas de clúster de LISA, se visualiza la autocorrelación y si existe un clúster o un valor atípico. Los colores indican autocorrelaciones: alto-alto implica un clúster de autocorrelación positiva; bajo-bajo es un clúster de autocorrelación positiva; bajo-alto un valor atípico de autocorrelación negativa; y alto-bajo: un valor atípico de autocorrelación negativa.

El análisis G_i^* de Getis-Ord podría resultar más interesante que el análisis Morán I local, pues permite identificar áreas de puntos calientes y puntos fríos en términos de

fallecimientos por Covid-19. En este caso, vuelve a ser necesario el uso de la matriz de peso adaptativa basada en la distancia obtenida.

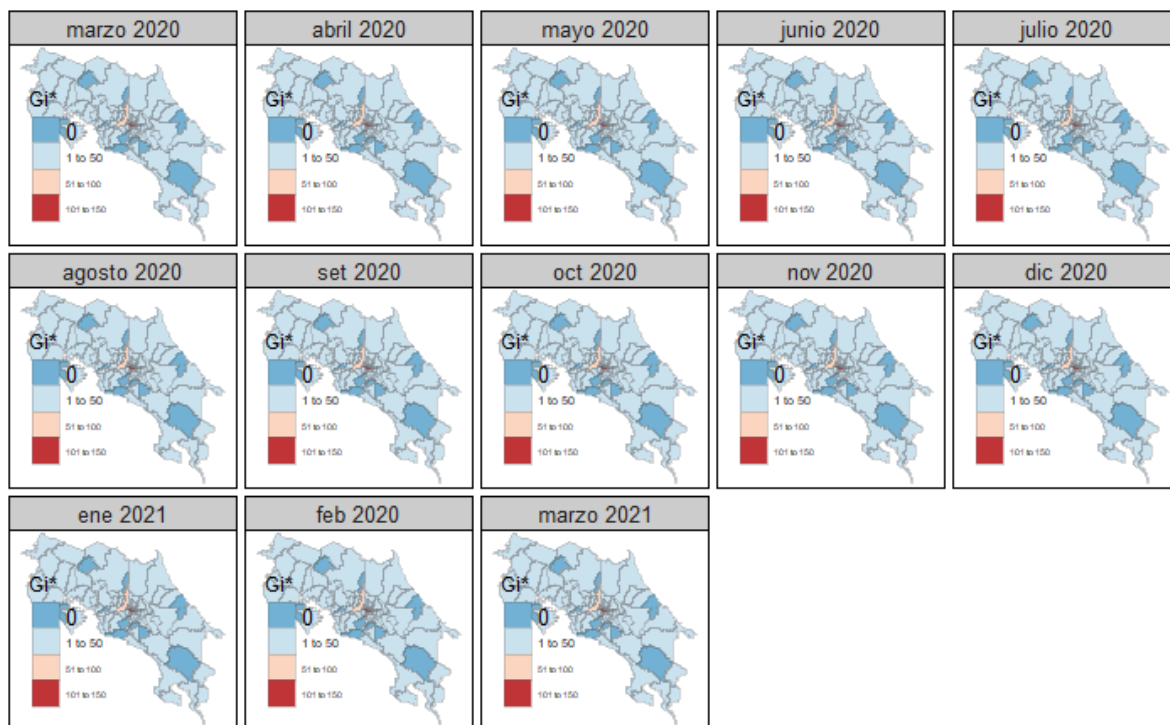


Figura 9.17. Representación gráfica del diagrama de dispersión del Índice de Morán

Los gráficos de la figura 9.17 muestran los puntos calientes y fríos de fallecidos por Covid-19 en el Costa Rica. Los cantones en rojo son las áreas de puntos calientes y los que están en azul son las áreas de puntos fríos. La oscuridad de los colores representa la intensidad de los valores G_i . Como tal, en el gráfico se muestra, de manera similar, que hay un punto caliente (rojo) en el centro del país. Los puntos fríos, en cambio, se encuentran fuera de la zona central del país. Con esto, se puede identificar la propagación espacio-temporal de los fallecidos por el virus a partir del centro del país. Además, existen signos de conglomerados espaciales en términos de fallecidos de Covid-19.

ANEXOS

En esta sección se presentan los resultados extraídos de los modelos corridos para determinar la estructura final del modelo. Además, se encuentra la matriz de correlación entre covariables y los gráficos de predicción.

A Acumulación de fallecidos por Covid-19 en Costa Rica

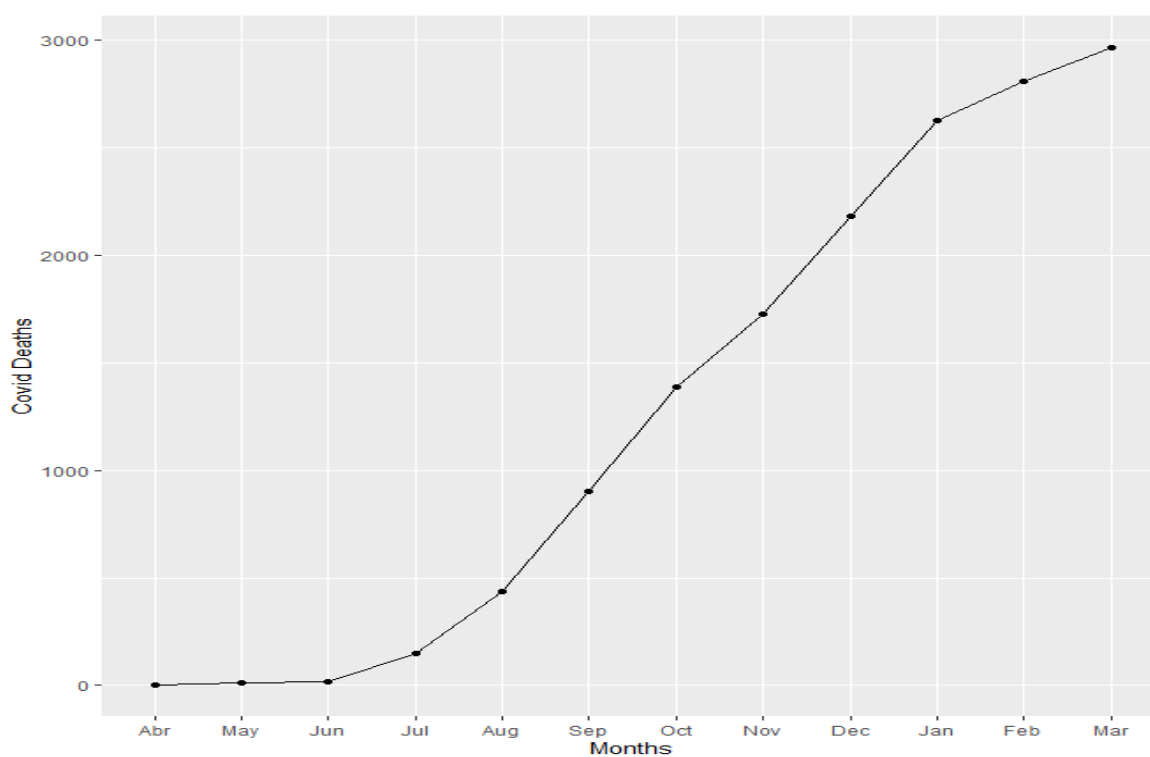


Figura 10.18. Representación gráfica de la acumulación de fallecidos por Covid-19 en Costa Rica durante el período abril 2020 - marzo 2021

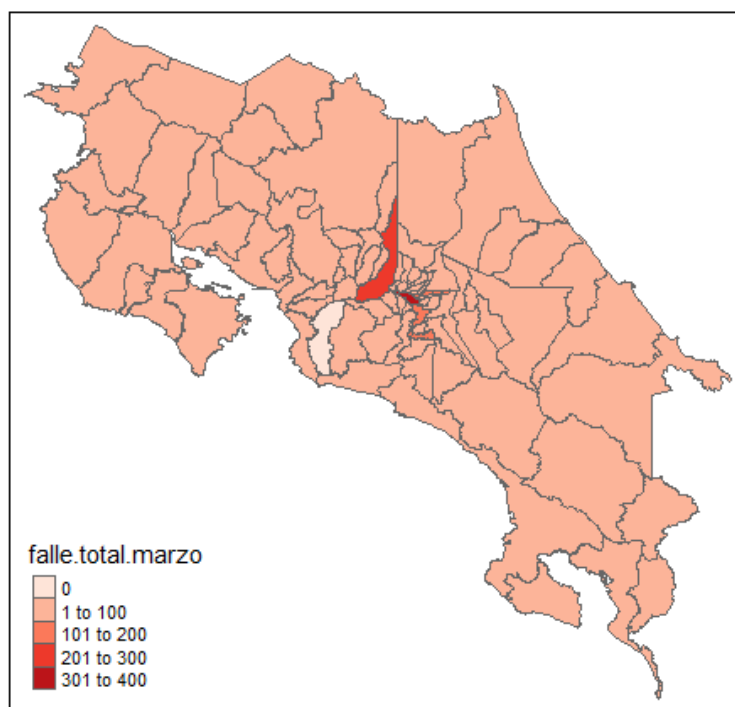


Figura 10.19. Representación gráfica de la acumulación de fallecidos por cantón en Costa Rica durante el período marzo 2020 - marzo 2021

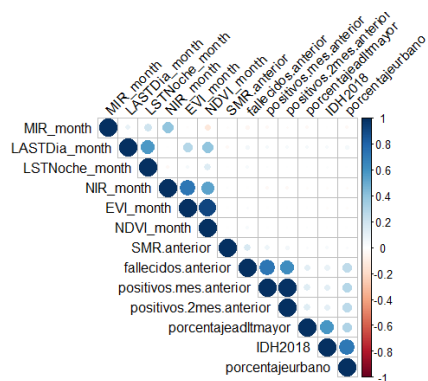


Figura 10.20. Representación gráfica de la correlación de covariables predictoras (dendograma y mapa de calor)

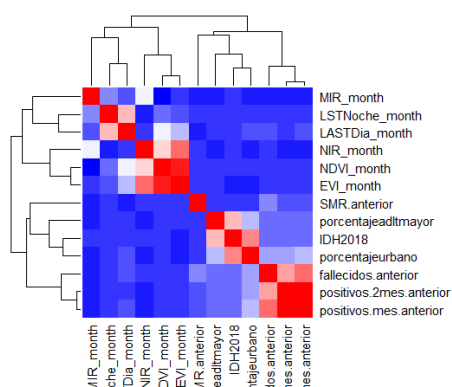


Figura 10.21. Representación gráfica de la matriz de correlación de covariables predictoras

B Predicciones de fallecidos por cantón para abril y mayo 2021

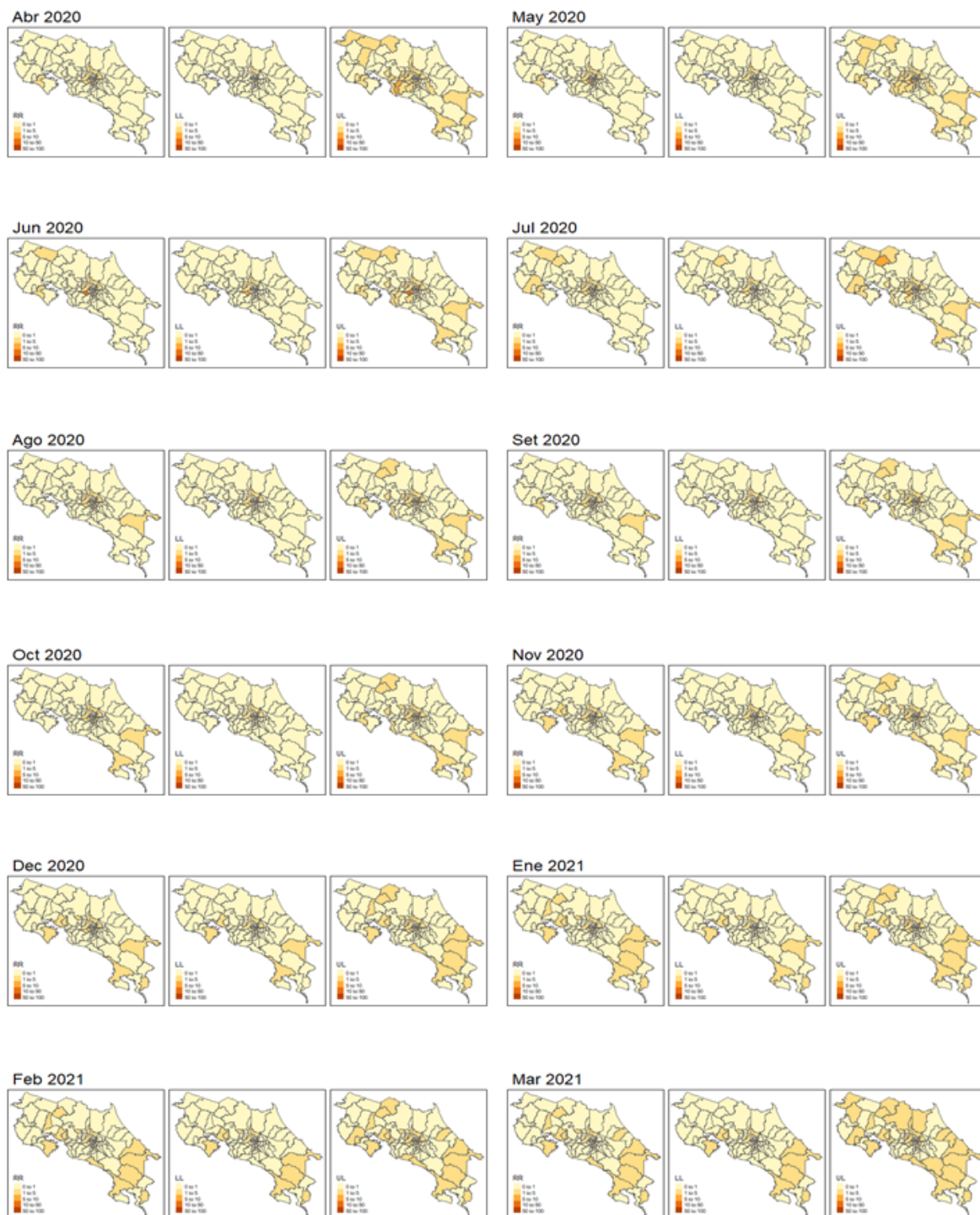


Figura 10.22. Representación gráfica del riesgo relativo de coronavirus durante el período abril 2020 - marzo 2021 por cantones en Costa Rica

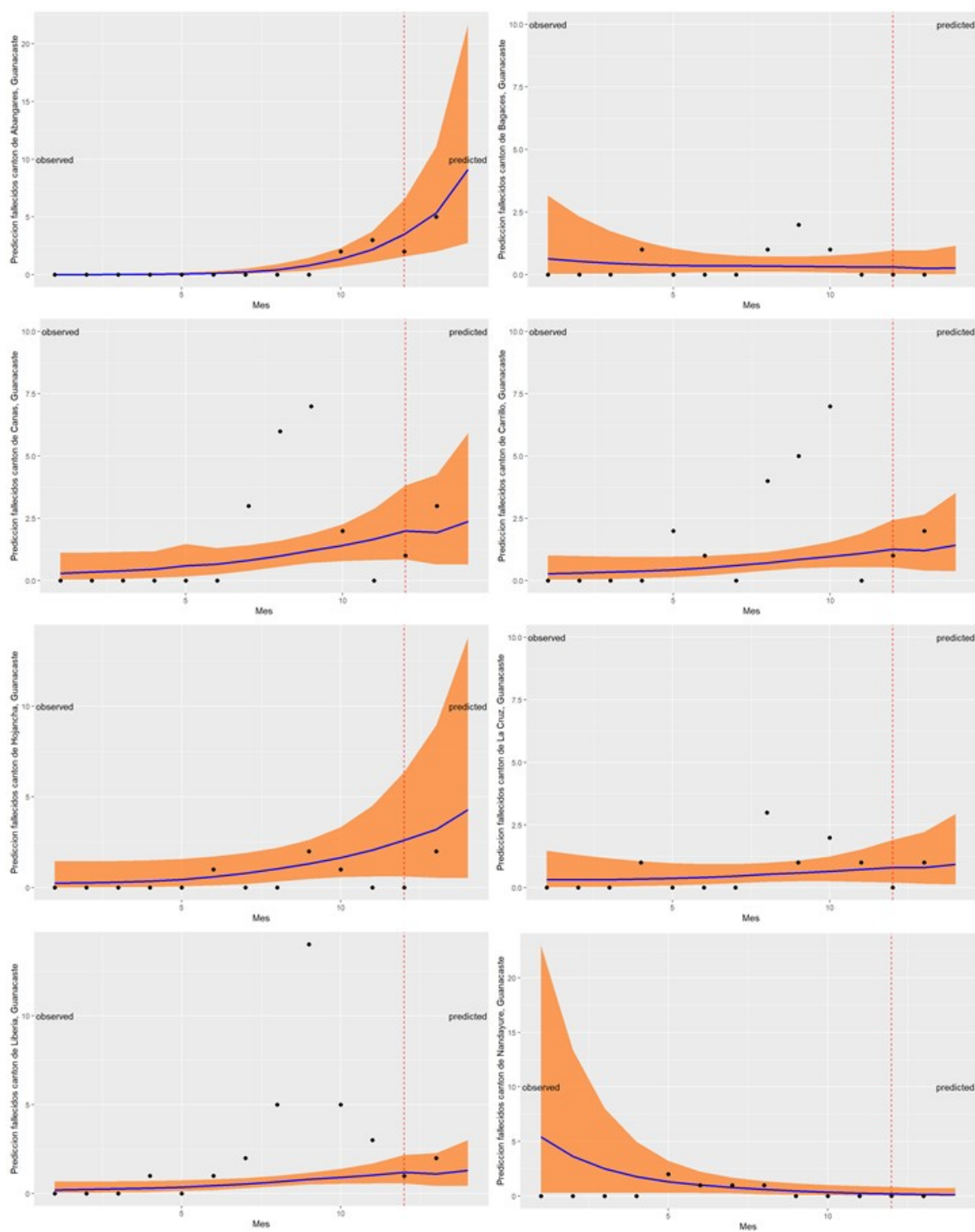


Figura 10.23. Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Guanacaste

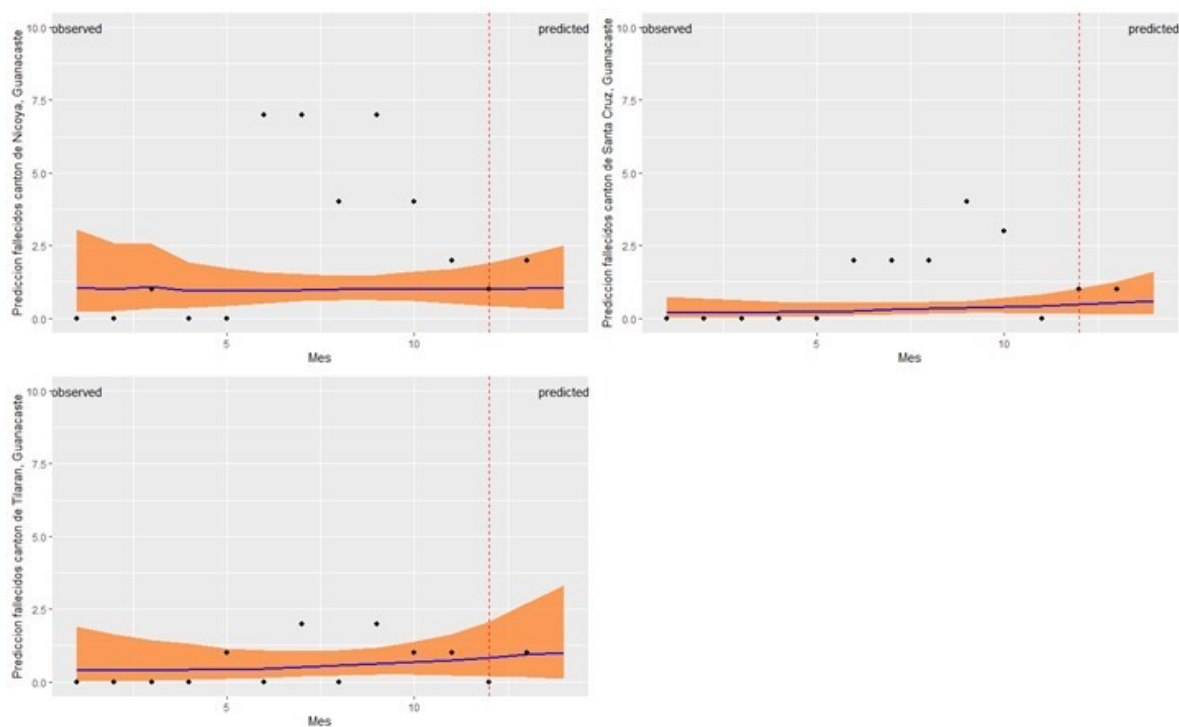


Figura 10.24. (Continuación) Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Guanacaste



Figura 10.25. Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de San José

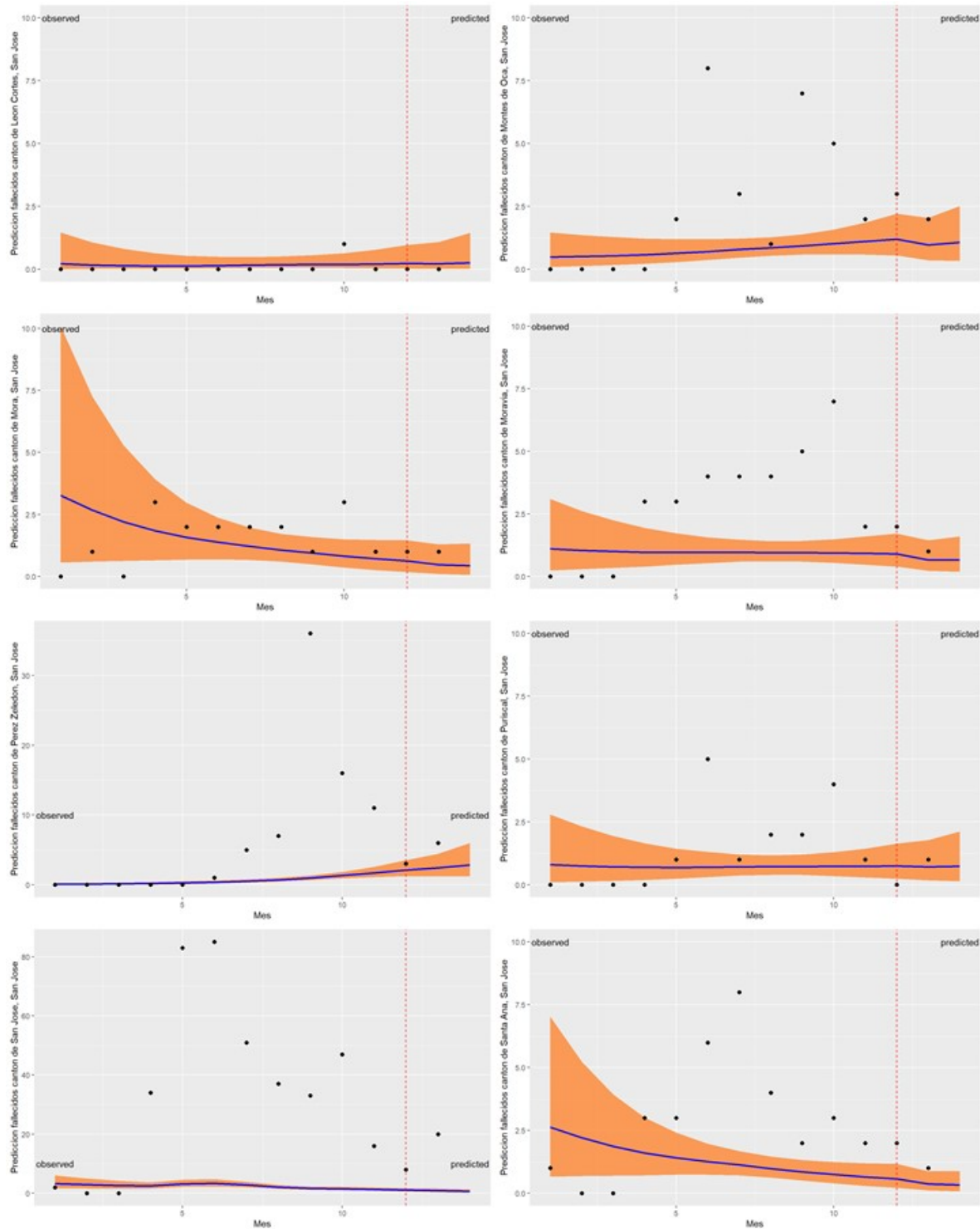


Figura 10.26. (Continuación) Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de San José

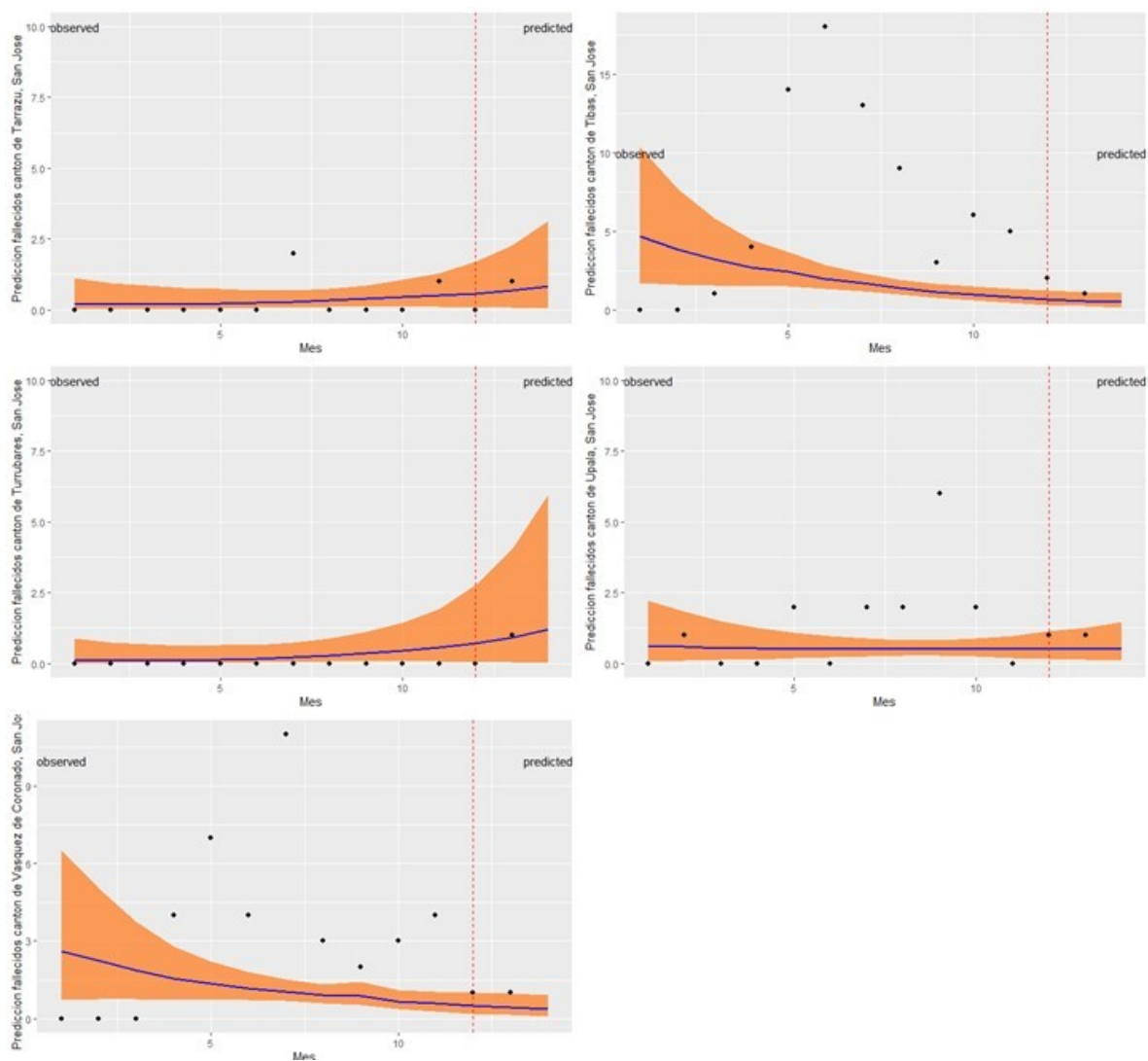


Figura 10.27. (Continuación 2) Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de San José

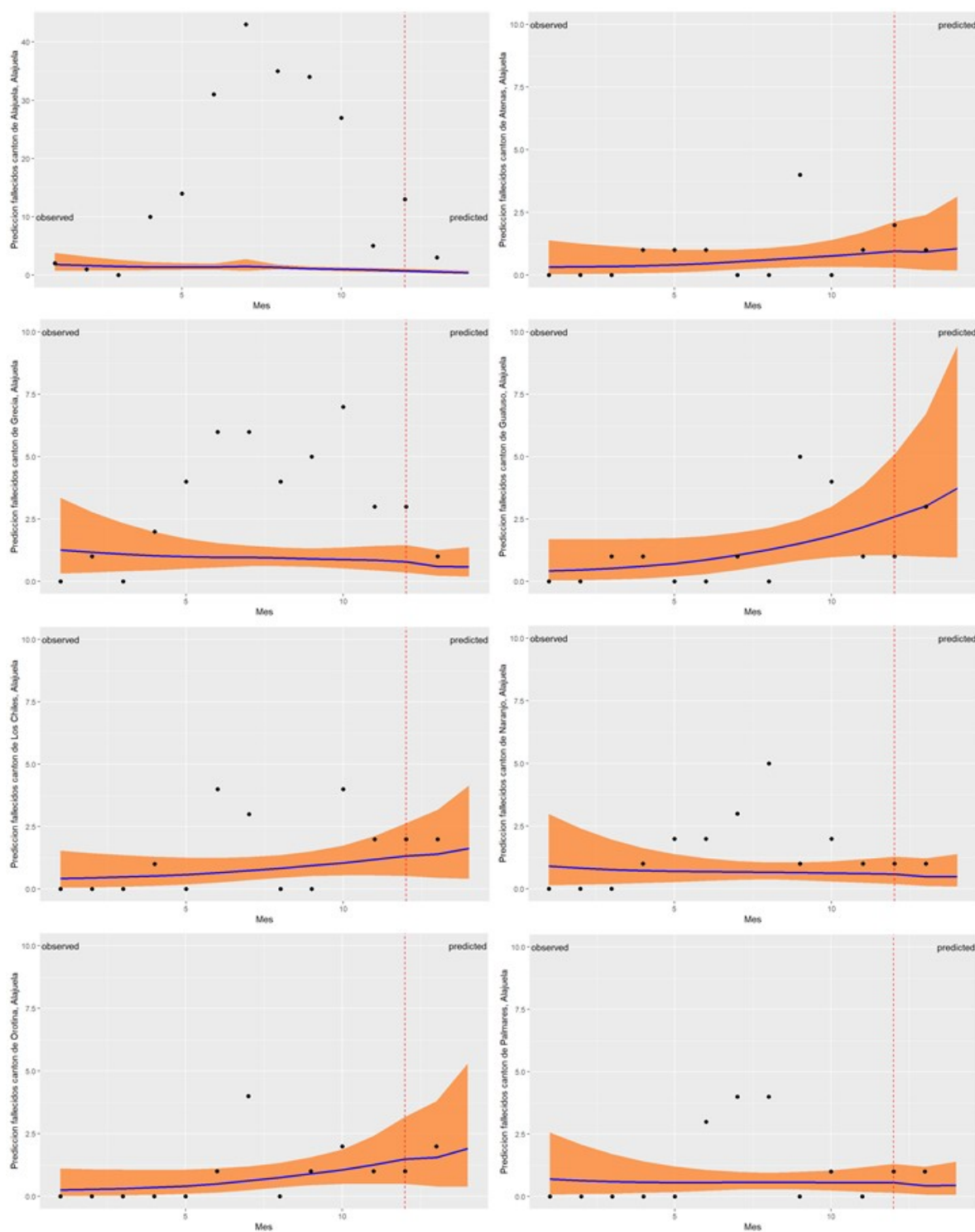


Figura 10.28. Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Alajuela

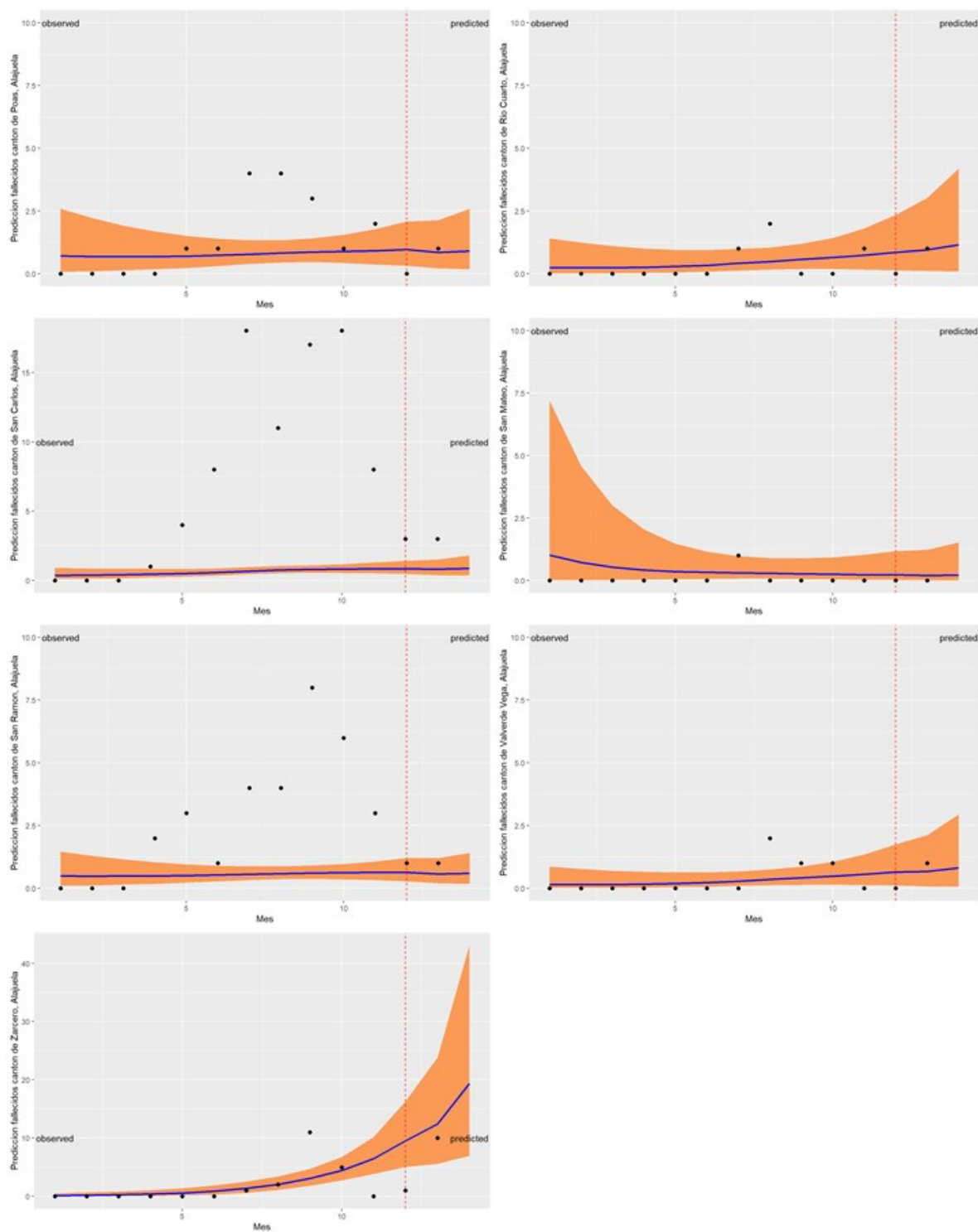


Figura 10.29. (Continuación) Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Alajuela

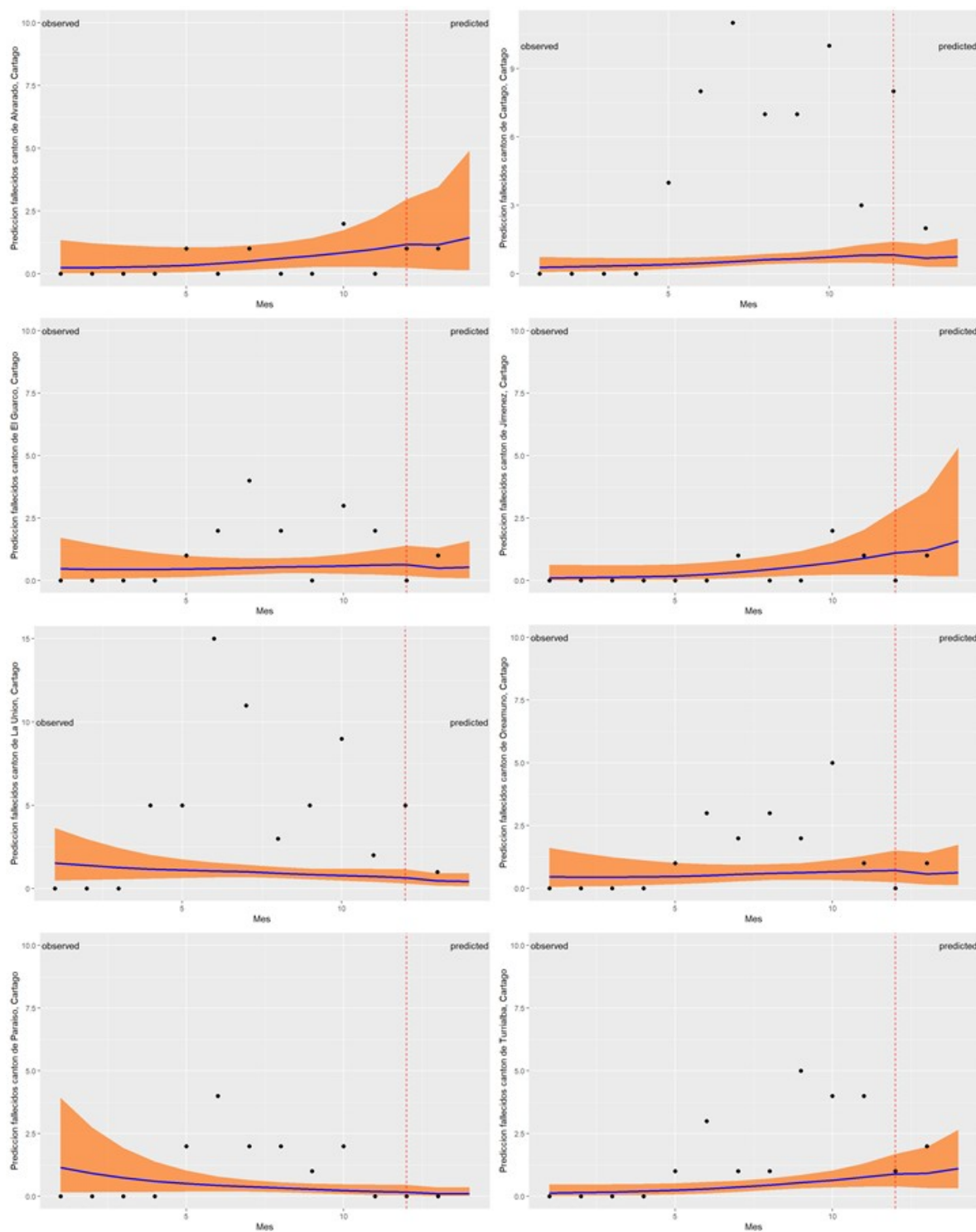


Figura 10.30. Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Cartago

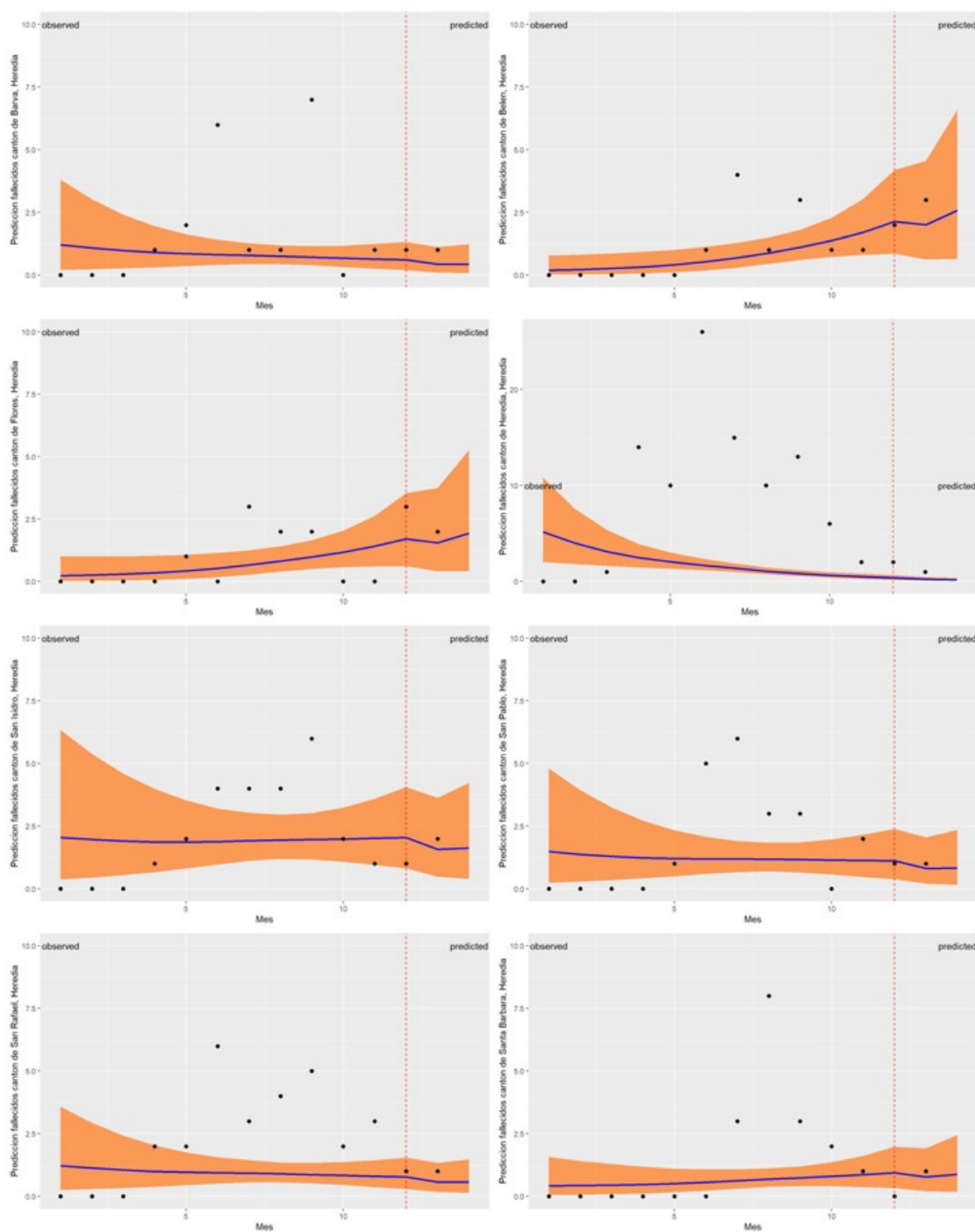


Figura 10.31. Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Heredia

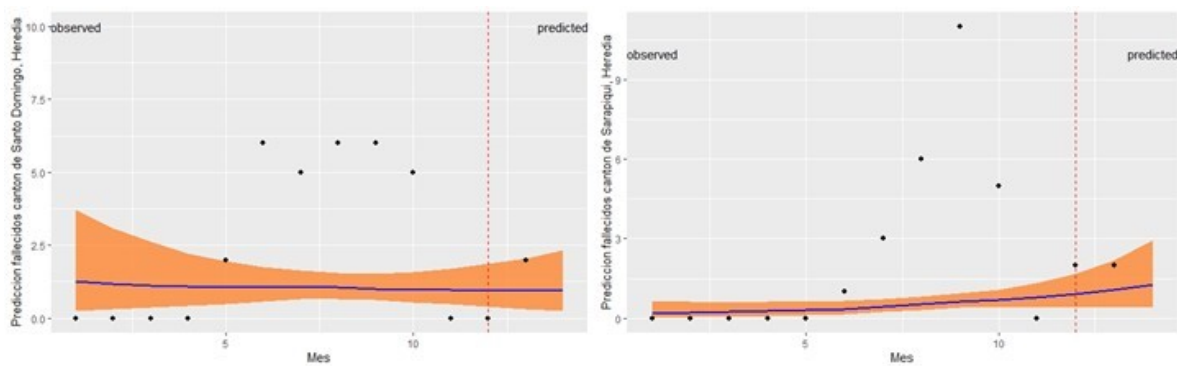


Figura 10.32. (Continuación) Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Heredia

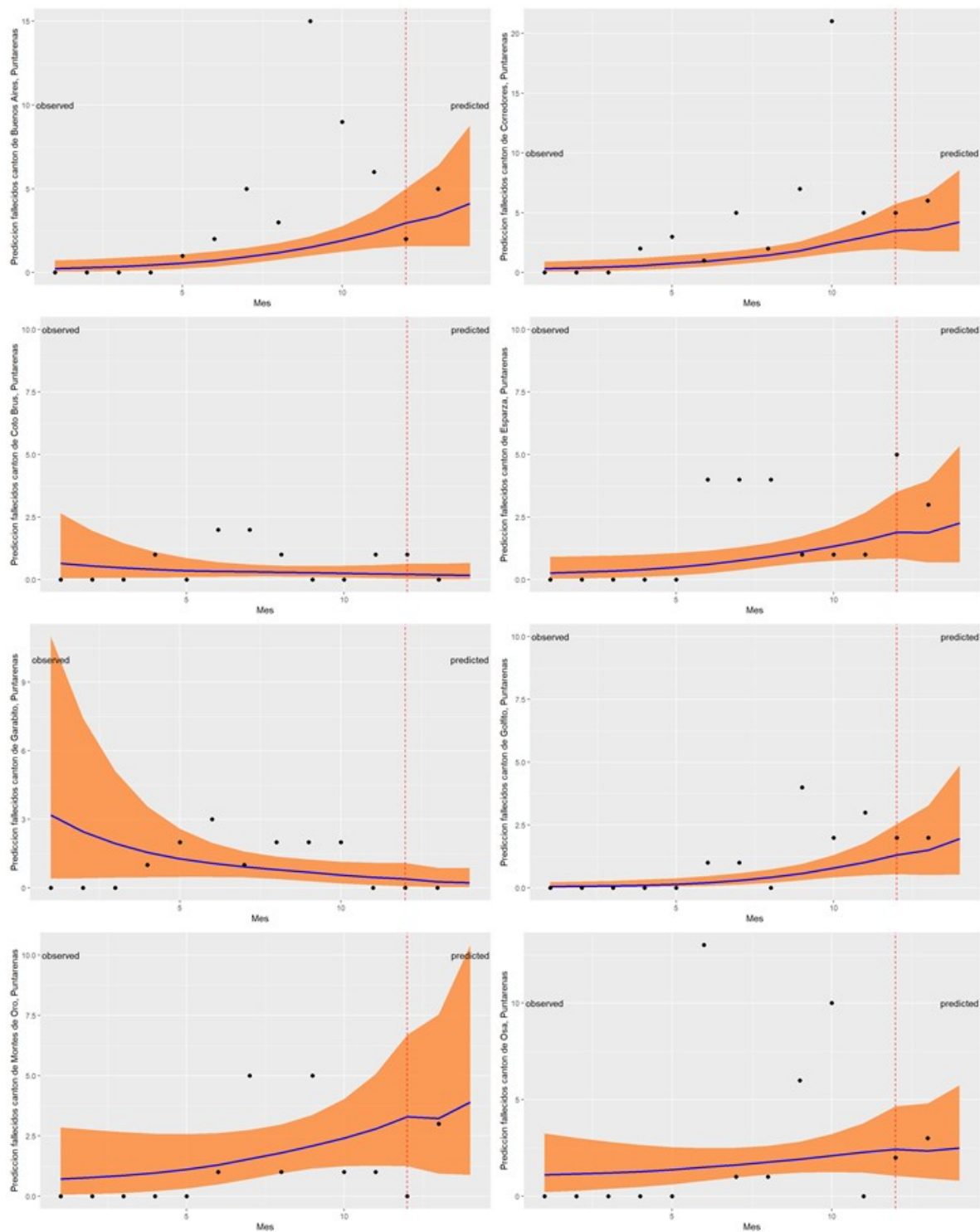


Figura 10.33. Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Puntarenas

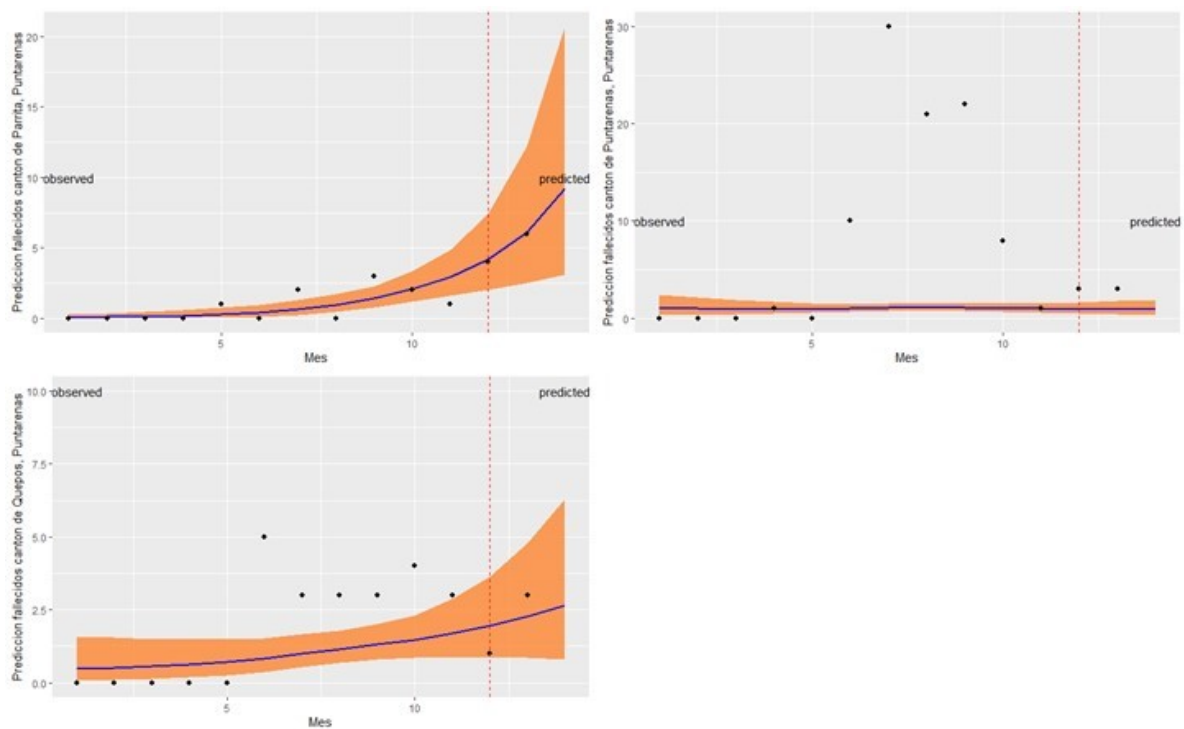


Figura 10.34. (Continuación) Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Puntarenas

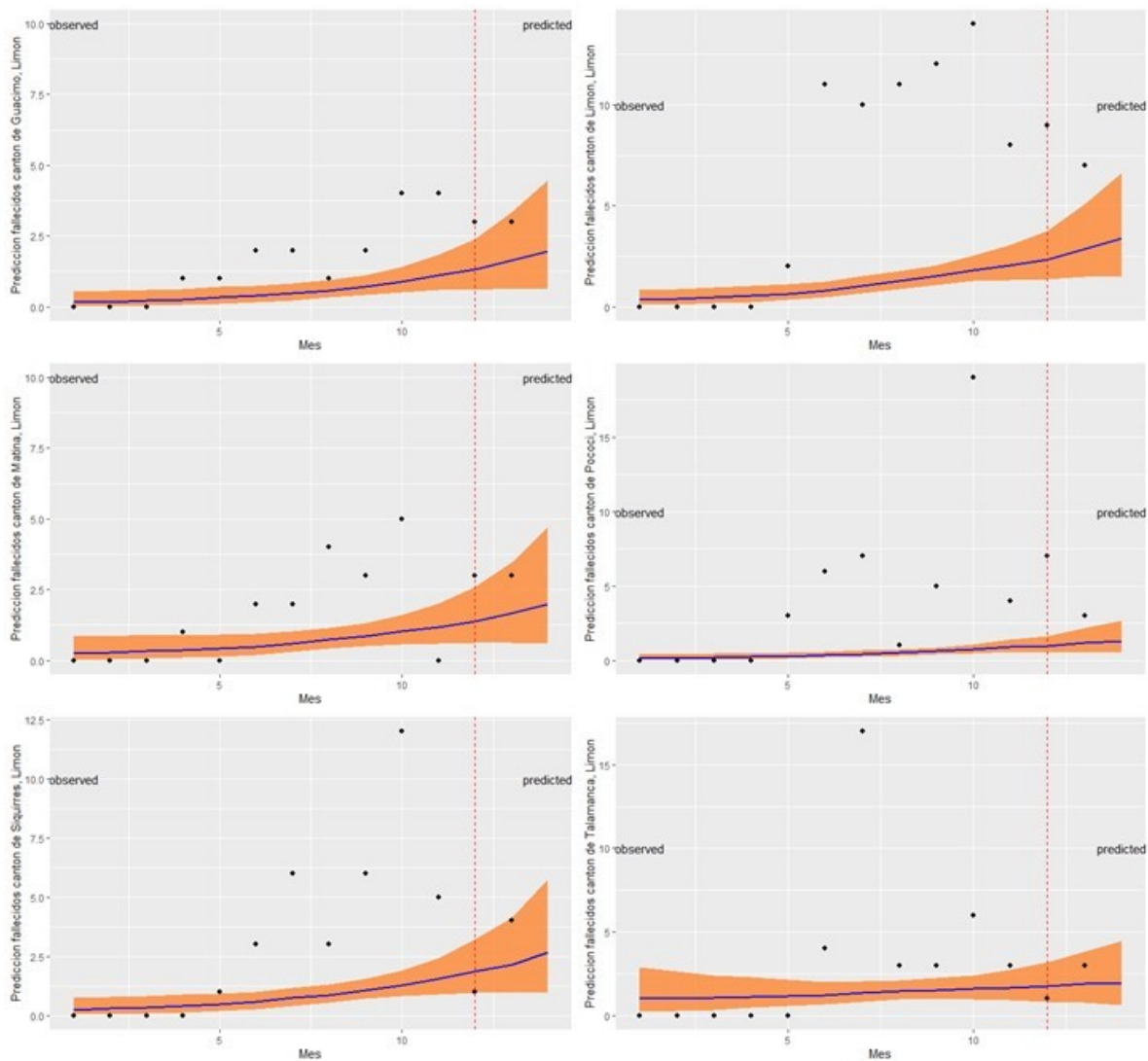


Figura 10.35. Predicción de fallecidos por coronavirus durante el período abril 2020 - mayo 2021 para los cantones de la provincia de Limón

Cuadro B.1. Comparación de valores predichos contra valores confirmados del mes de abril de 2021

Cantón	Fallecid. confir- mados	Predicc. abril 2021	Límite inferior	Límite superior	Intervalo
Abangares	0	6.03	2.36	12.41	
Acosta	1	0.50	0.06	1.66	Sí
Alajuela	25	0.55	0.29	0.95	
Alajuelita	4	0.72	0.32	1.38	
Alvarado	1	1.77	0.35	4.75	Sí
Aserrí	2	0.43	0.11	1.05	
Atenas	1	1.24	0.33	3.04	Sí
Bagaces	0	0.52	0.06	1.67	
Barva	4	0.45	0.11	1.16	
Belén	1	2.06	0.64	4.70	Sí
Buenos Aires	2	3.68	1.70	7.00	Sí
Cañas	0	1.98	0.67	4.39	
Carrillo	3	1.23	0.41	2.74	
Cartago	4	0.75	0.34	1.43	
Corredores	3	3.86	1.88	7.11	Sí
Coto Brus	2	0.34	0.05	1.01	
Curridabat	5	0.54	0.21	1.13	
Desamparados	13	0.32	0.15	0.59	
Dota	0	9.42	3.04	20.99	
El Guarco	0	0.52	0.12	1.36	
Escazú	0	0.77	0.30	1.58	
Esparza	1	1.93	0.71	4.11	Sí
Flores	0	1.96	0.58	4.55	
Garabito	0	0.46	0.08	1.37	
Goicoechea	9	0.64	0.29	1.20	
Golfito	2	1.82	0.65	3.91	Sí
Grecia	3	0.74	0.29	1.51	
Guácimo	5	1.52	0.58	3.14	
Guatuso	1	3.07	1.00	6.86	Sí
Heredia	5	0.23	0.09	0.48	
Hojancha	0	3.25	0.55	9.14	
Jiménez	2	1.83	0.36	4.93	Sí
La Cruz	1	0.82	0.15	2.27	Sí
La Unión	5	0.47	0.19	0.96	
León Cortés Castro	0	0.68	0.05	2.53	
Liberia	0	1.29	0.52	2.65	
Limón	7	2.39	1.19	4.36	
Los Chiles	0	1.44	0.45	3.27	
Matina	2	1.33	0.46	2.92	Sí
Montes de Oca	3	1.00	0.37	2.12	
Montes de Oro	0	3.97	1.26	8.95	

Cuadro B.2. (Continuación) Comparación de valores predichos contra valores confirmados del mes de abril de 2021

Cantón	Fallecid. confir- mados	Predicc. abril 2021	Límite inferior	Límite superior	Interva.
Mora	2	0.49	0.10	1.34	
Moravia	4	0.80	0.29	1.70	
Nandayure	0	0.48	0.04	1.92	
Naranjo	1	0.50	0.12	1.28	Sí
Nicoya	1	0.96	0.34	2.07	Sí
Oreamuno	0	0.60	0.15	1.48	
Orotina	2	1.59	0.41	3.93	Sí
Osa	2	2.41	0.94	4.96	Sí
Palmares	1	0.44	0.08	1.24	Sí
Paraíso	2	0.19	0.03	0.55	
Parrita	2	5.75	2.33	11.64	
Pérez Zeledón	20	2.65	1.28	5.02	
Poás	2	1.12	0.31	2.69	Sí
Pococí	5	1.04	0.48	1.97	
Puntarenas	6	0.81	0.36	1.59	
Puriscal	2	0.72	0.18	1.83	
Quepos	3	2.11	0.76	4.51	Sí
Río Cuarto	0	0.97	0.11	3.07	
San Carlos	11	0.83	0.39	1.54	
San Isidro	1	1.94	0.63	4.34	Sí
San José	24	0.46	0.25	0.80	
San Mateo	1	0.95	0.04	4.01	Sí
San Pablo	1	1.07	0.30	2.58	Sí
San Rafael	4	0.59	0.18	1.37	
San Ramón	3	0.59	0.21	1.25	
Santa Ana	3	0.39	0.11	0.92	
Santa Bárbara	2	1.01	0.29	2.39	Sí
Santa Cruz	5	0.49	0.13	1.20	
Santo Domingo	4	0.77	0.24	1.76	
Sarapiquí	2	1.06	0.40	2.22	Sí
Valverde Vega	1	0.69	0.09	2.17	Sí
Siquirres	5	2.03	0.88	3.99	
Talamanca	5	1.83	0.71	3.78	
Tarrazú	1	0.62	0.06	2.11	Sí
Tibás	12	0.46	0.17	0.95	
Tilarán	1	0.85	0.13	2.49	Sí
Turrialba	2	0.93	0.33	2.02	Sí
Turrubares	0	0.90	0.04	3.93	
Upala	2	0.49	0.12	1.22	
Vázquez de Coronado	3	0.37	0.11	0.85	
Zarcero	1	13.51	6.13	25.98	

Cuadro B.3. Comparación de valores predichos contra valores confirmados del mes de mayo de 2021

Cantón	Fallecid. confir- mados	Predicc. mayo 2021	Límite inferior	Límite superior	Intervalo
Abangares	1	10.62	3.36	25.00	
Acosta	1	0.53	0.04	2.03	Sí
Alajuela	56	0.50	0.23	0.95	
Alajuelita	9	0.65	0.25	1.36	
Alvarado	1	2.34	0.34	7.19	Sí
Aserrí	6	0.43	0.09	1.19	
Atenas	8	1.52	0.31	4.21	
Bagaces	1	0.60	0.05	2.18	Sí
Barva	10	0.44	0.08	1.29	
Belén	3	2.67	0.67	6.87	Sí
Buenos Aires	11	4.81	1.87	10.31	
Cañas	0	2.47	0.67	6.20	
Carrillo	5	1.47	0.40	3.69	
Cartago	17	0.83	0.32	1.77	
Corredores	5	4.89	2.03	10.03	Sí
Coto Brus	0	0.36	0.04	1.23	
Curridabat	22	0.49	0.16	1.14	
Desamparados	36	0.26	0.11	0.53	
Dota	1	16.64	4.22	42.41	
El Guarco	2	0.57	0.10	1.68	
Escazú	10	0.75	0.25	1.69	
Esparza	5	2.39	0.72	5.69	Sí
Flores	12	2.55	0.60	6.68	
Garabito	2	0.44	0.06	1.49	
Goicoechea	28	0.59	0.24	1.21	
Golfito	1	2.57	0.74	6.25	Sí
Grecia	15	0.74	0.24	1.69	
Guácimo	0	1.92	0.61	4.46	
Guatuso	3	3.87	0.99	9.87	Sí
Heredia	42	0.19	0.06	0.42	
Hojancha	0	4.44	0.54	14.26	
Jiménez	0	2.56	0.38	7.80	
La Cruz	2	0.96	0.13	3.03	Sí
La Unión	10	0.44	0.15	0.99	
León Cortés Castro	0	0.90	0.04	3.80	
Liberia	15	1.59	0.53	3.66	
Limón	7	2.87	1.24	5.78	
Los Chiles	10	1.67	0.41	4.28	
Matina	1	1.59	0.44	3.94	Sí
Montes de Oca	8	1.11	0.34	2.63	
Montes de Oro	1	5.07	1.27	13.03	

Cuadro B.4. (Continuación) Comparación de valores predichos contra valores confirmados del mes de mayo de 2021

Cantón	Fallecid. confir- mados	Predicc. mayo 2021	Límite inferior	Límite superior	Interva.
Mora	4	0.45	0.07	1.38	
Moravia	16	0.83	0.25	1.96	
Nandayure	1	0.47	0.02	2.20	Sí
Naranjo	13	0.51	0.10	1.46	
Nicoya	10	1.00	0.29	2.41	
Oreamuno	7	0.66	0.13	1.83	
Orotina	5	1.98	0.39	5.52	Sí
Osa	4	2.67	0.86	6.16	Sí
Palmares	15	0.47	0.07	1.47	
Paraíso	10	0.18	0.02	0.60	
Parrita	6	8.77	2.89	20.14	Sí
Pérez Zeledón	39	3.63	1.47	7.75	
Poás	10	1.27	0.28	3.45	
Pococí	23	1.26	0.50	2.64	
Puntarenas	9	0.81	0.30	1.76	
Puriscal	3	0.76	0.14	2.19	
Quepos	6	2.52	0.74	6.08	Sí
Río Cuarto	0	1.19	0.09	4.31	
San Carlos	33	0.89	0.36	1.84	
San Isidro	6	2.09	0.55	5.25	
San José	79	0.38	0.18	0.71	
San Mateo	0	1.22	0.03	5.84	
San Pablo	10	1.14	0.25	3.09	
San Rafael	10	0.59	0.15	1.53	
San Ramón	11	0.62	0.18	1.48	
Santa Ana	9	0.35	0.08	0.92	
Santa Bárbara	8	1.18	0.27	3.16	
Santa Cruz	19	0.56	0.12	1.56	
Santo Domingo	19	0.79	0.20	2.01	
Sarapiquí	7	1.26	0.39	3.00	
Valverde Vega	0	0.85	0.07	3.05	
Siquirres	5	2.53	0.92	5.57	Sí
Talamanca	4	1.99	0.64	4.62	Sí
Tarrazú	0	0.76	0.05	2.95	
Tibás	17	0.39	0.13	0.90	
Tilarán	0	0.99	0.11	3.31	
Turrialba	8	1.15	0.33	2.80	
Turrubares	0	1.22	0.03	6.01	
Upala	6	0.50	0.10	1.44	
Vázquez de Coronado	10	0.33	0.09	0.85	
Zarcero	0	21.69	7.86	48.43	