

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

ANÁLISIS PROBABILÍSTICO DE SEQUÍA METEOROLÓGICA
MEDIANTE CADENAS DE MARKOV Y REDES BAYESIANAS;
CUENCA DEL RÍO TEMPISQUE, GUANACASTE

Tesis sometida a la consideración de la Comisión del Programa de
Estudios de Posgrado en Geografía para optar al grado y título de
Maestría Académica en Geografía

Kenneth Gutiérrez García

Ciudad Universitaria Rodrigo Facio, Costa Rica

2021

Agradecimientos

Primeramente, quiero agradecer a mis padres Xinia García y Lidier Gutiérrez por su increíble amor y apoyo incondicional que siempre me han brindado.

Agradezco también a mi abuela Nidia, a mi tía Lilliana, a mi hermana Melissa, a mi hermano Erick a mi primo Gabriel, por su acompañamiento en todas las etapas de mi vida.

A mis amigos de universidad y de la vida que me han motivado y apoyado, especialmente a Valeria Meoño, Melisa Mena, Nelson Venegas y Marcos Moreno.

A la *Deutscher Akademischer Austausch Dienst* (DAAD) por el financiamiento brindado al proyecto TropiSeca, el cual me brindo la posibilidad de una pasantía en el Departamento de Recursos Hídricos y Ciencias Ambientales de la Universidad de Cuenca, a los cuales también agradezco su hospitalidad, especialmente al Dr. Rolando Célleri. También agradezco la ayuda brindada por la Dr. Alexandra Nauditt y M. Sc Justyna Syz.

A los compañeros y amigos investigadores que forman parte del Observatorio del Agua y Cambio Global, a la Escuela de Geografía y a la Universidad de Costa Rica.

Finalmente quiero agradecer profundamente a los doctores Rafael Arce, Christian Birkel y Alex Avilés por la gran oportunidad que me dieron de realizar esta tesis, les agradezco sus valiosos aportes en mi formación académica, pero aún más sus consejos de motivación los cuales fueron muy importantes para mí.

Esta Tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Geografía de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Geografía

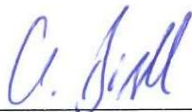
Jonathan Reyes Ch.

M.Sc. Jonathan Reyes Chaves
**Representante del Decano
Sistema de Estudios de Posgrado**



Dr. Rafael Arce Mesén
Profesor Guía

Dr. Alex Avilés Añazco
Lector



Dr. Christian Birkel
Lector



Dr. Adolfo Quesada Román
**Representante del Director del
Posgrado en Geografía**

Kenneth Gutiérrez G.

Kenneth Gutiérrez García
Sustentante

Nota aclaratoria: Por encontrarse fuera de la República de Costa Rica, el Dr. Alex Avilés Añazco omite estampar su rúbrica en esta hoja de aprobación, quedando de respaldo para la Universidad de Costa Rica, la grabación de la defensa de la Tesis.

Tabla de Contenidos

Agradecimientos	ii
Hoja de aprobación	iii
Resumen	vi
Abstract	vii
Lista de tablas	viii
Lista de figuras	ix
Lista de abreviaturas	xi
Capítulo 1. Introducción al problema, al área de estudio, a los objetivos, conceptos y a la metodología	1
1.1 El problema y su importancia	1
1.2 Área de estudio	2
1.3 Antecedentes	4
1.4 Objetivo general	8
1.5 Objetivos específicos	8
1.6 Marco teórico.....	8
1.6.1 Periodos secos en el Pacífico Norte	8
1.6.2 El concepto de sequía.....	9
1.6.3 Modelo probabilístico	11
1.7.1 Datos	12
1.7.2 Categorías de sequía	13
1.7.3 Cadenas de Markov	13
1.7.4 Redes Bayesianas	15
1.7.5 Verificación de los modelos	18
Capítulo 2. Aplicación de las Cadenas de Markov de primer y segundo orden	20
2.1 Datos de la Estación Aeropuerto, Liberia	20

2.1.1 Índice Estandarizado de Precipitación	21
2.1.2 Clasificación de los estados de la sequía	23
2.2 Cadenas de Markov de primer orden	25
2.3 Cadenas de Markov de segundo orden	27
Capítulo 3. Modelos probabilísticos basados en Redes Bayesianas	32
3.1 Datos de entrada para los modelos de Redes Bayesianas	32
3.2 Redes Bayesianas de primer orden	33
3.2.1 Ajuste de cópulas de dos dimensiones	33
3.2.2 Probabilidades del modelo de Redes Bayesianas de primer orden	34
3.3 Redes Bayesianas de segundo orden	36
3.3.1 Ajuste de cópulas de tres dimensiones	36
3.3.2 Probabilidades del modelo de Redes Bayesianas de segundo orden	37
Capítulo 4. Evaluación de los modelos probabilísticos	39
4.1 Ranked Probability Score para todas las categorías	39
4.2 Ranked Probability Score únicamente para los estados de sequía	40
4.3 Ranked Probability Skill Score para todas las categorías	41
4.4 Ranked Probability Skill Score únicamente para los estados de sequía	43
Capítulo 5. Aplicación del modelo de Redes Bayesianas de primer orden en las estaciones Meteorológicas Cañas Dulces y Talolinga	44
5.1 Datos de precipitación de las estaciones Cañas Dulces y Talolinga	44
5.2 Estados de sequía para las estaciones Cañas Dulces y Talolinga	46
5.3 Ajuste de cópulas para las estaciones Cañas Dulces y Talolinga	47
Conclusiones y recomendaciones	55
Referencias bibliográficas	57
Anexos	62

Resumen

La sequía es un proceso atmosférico difícil de detectar, pues sus impactos socio-económicos son solo apreciables meses después de su inicio. Un ejemplo de esto es el caso de la sequía que ocurrió a Centro América en entre el 2014 y 2015, la cual afectó a 3.5 millones de personas y según la FAO para el 2016 se había perdido el 90% de las cosechas en el Corredor Seco Centroamericano. Costa Rica no escapa a estos impactos, pues históricamente la Región Pacifico Norte se ha visto afectada por eventos de sequía, ya que esta región presenta una estación seca de al menos cuatro meses, la cual se puede extender si se presentan sequías en los meses de lluvia. Debido a esto es necesario buscar formas de pronosticar eventos de sequía, para así poder establecer mecanismos de adaptación que minimicen los impactos en las comunidades. En esta tesis, se estudiaron modelos estadísticos basados en Cadenas de Markov (de primer y segundo orden) y Redes Bayesianas (de primer y segundo orden) con el objetivo de generar pronósticos de eventos de sequía meteorológica en tres estaciones meteorológicas ubicadas dentro de la cuenca del Río Tempisque. Primero se utilizaron los datos de la estación Aeropuerto para aplicar los modelos en una serie temporal de precipitación de 82 años y se le calculó el Índice Estandarizado de Precipitación a escala mensual (*SPI-1*), posteriormente se reclasificaron los valores en cuatro estados (No Sequía, Sequía Moderada, Sequía Severa y Sequía Extrema). Los modelos fueron aplicados a un periodo de calibración (1937-1999) y su rendimiento fue probado en un periodo de validación (2000-2020). Para seleccionar el modelo de mejor rendimiento se utilizaron los métodos de *Ranked Probability Score (RPS)* y de *Ranked Probability Skill Score (RPSS)*. Una vez seleccionado el modelo con mejor rendimiento se aplicó el mismo en las estaciones meteorológicas de Cañas Dulces y Talolinga. Las Cadenas de Markov lograron un mejor rendimiento en la época seca, pues pronosticaban que el estado de No Sequía era el de mayor probabilidad, sin embargo, esto era esperable ya que estos meses secos son una característica del clima de la zona y no debe confundirse con eventos de sequía. Por otro lado, las Redes Bayesianas superaron los resultados de las Cadenas de Markov, pues en sus distribuciones de probabilidad en la época lluviosa mostraron un aumento en probabilidad de sequía cuando se observaron eventos de sequías en la serie temporal. De las validaciones se comprobó que las Redes Bayesianas de primer orden tienen el mejor rendimiento en la época lluviosa, que es cuando históricamente se han observado sequías meteorológicas. Este fue el modelo seleccionado para aplicarlo en las otras estaciones. Los resultados fueron similares a los de la estación Aeropuerto, pues el modelo logró diferenciar entre la estación seca y lluviosa, además las probabilidades de que se presentara algún estado de sequía aumentaban cuando se observaban eventos de sequía en la serie temporal. Así se concluye que el modelo de Redes Bayesianas de primer orden es un modelo apropiado para aplicar en la región, además esta metodología puede ser utilizada como un sistema de alerta temprana para otros tipos de sequía.

Abstract

Drought is an atmospheric process that is difficult to monitor, as its socio-economic impacts are only perceptible months later. An example of this is the case of the drought that affected Central America between 2014 and 2015, which affected 3.5 million people and according to FAO by 2016 90% of the crops in the Central American Dry Corridor were lost. Costa Rica is no exception to these impacts, as the North Pacific Region historically has been affected by drought events, as this region has a dry season of at least four months, which can be extended if droughts occur during the rainy season. Therefore, it is necessary to develop ways of forecasting drought events in order to establish adaptation mechanisms that minimize the impacts on the communities. In this thesis, statistical models based on Markov Chains (first and second order) and Bayesian Networks (first and second order) were studied with the objective of generating forecasts of meteorological drought events in three meteorological stations located within the Tempisque River Basin. First, data from the Airport Station were used to apply the models to an 82-year precipitation time series where the Standardized Precipitation Index on a monthly scale (SPI-1) was calculated, and then these values were reclassified into four states (No Drought, Moderate Drought, Severe Drought and Extreme Drought). The models were applied to a calibration period (1937-1999) and their performance was evaluated in a validation period (2000-2020). The models were applied to a calibration period (1937-1999) and their performance was evaluated in a validation period (2000-2020). The Ranked Probability Score (RPS) and Ranked Probability Skill Score (RPSS) methods were used to select the best performing model. Once the best performing model was selected, we applied it to the Cañas Dulces and Talolinga weather stations. The Markov Chains obtained a better performance in the dry season, predicting that the No Drought state was the most likely, however, this was expected as these dry months are a characteristic of the climate of the region and should not be confused with drought events. The Bayesian Networks outperformed the results of the Markov Chains, since their probability distributions in the rainy season showed an increase in the probability of drought when drought events were observed in the time series. The validations showed that the first-order Bayesian Networks have the best performance in the rainy season, which is when meteorological droughts historically have been observed. For this reason, this was the model selected to be applied to the other meteorological stations. The results of the model were similar to those of the Airport station, since the model was able to identify the difference between the dry and rainy seasons, and the probabilities of a drought state increased when drought events were observed in the time series. It is concluded that the first-order Bayesian Networks model is an appropriate model to apply in the region, and that this methodology can be used as an early warning system for other types of droughts.

Lista de tablas

Tabla 1. Clasificación de las categorías de sequía, según valores de SPI y CDF.	13
Tabla 2. Frecuencia de los eventos de sequía en la Estación Aeropuerto (1937-2020).....	24
Tabla 3. Frecuencia de los eventos de sequía para el periodo de calibración (1936-1999), en la Estación Aeropuerto.....	25
Tabla 4. Frecuencia de los eventos de sequía para el periodo de validación (2000-2020), en la Estación Aeropuerto.....	25
Tabla 5. Probabilidades de transición de los estados del modelo MC1, en la época lluviosa para la Estación Aeropuerto.....	26
Tabla 6. Probabilidades de transición de los estados del modelo MC2, en la época lluviosa para la Estación Aeropuerto.....	28
Tabla 7. Resultado del ajuste mensual de cópulas de dos dimensiones.....	33
Tabla 8. Correlaciones de las pseudo-observaciones para cada mes en la Estación Aeropuerto.	36
Tabla 9. Resultado del ajuste de cópulas de tres dimensiones.	36
Tabla 10. Umbrales de la distribución acumulada de las estaciones Cañas Dulces y Talolinga según cada categoría de sequía.	48
Tabla 11. Resultado del ajuste mensual de cópulas de dos dimensiones para la Estación de Cañas Dulces.	49
Tabla 12. Resultado del ajuste mensual de cópulas de dos dimensiones para la Estación de Talolinga.....	49

Lista de figuras

Figura 1. Localización de la cuenca del Río Tempisque.	3
Figura 2. Componentes para el pronóstico de sequías. (Mishra y Singh, 2011). ...	5
Figura 3. Clasificación de la sequía y su desarrollo. Adaptado Wilhite & Glantz (1985) y Van Loon (2015).	10
Figura 4. Precipitación promedio mensual en la Estación Aeropuerto (1937- junio, 2020).	20
Figura 5. Mapa de calor que representa la precipitación promedio mensual (1937- junio, 2020) para la Estación Aeropuerto. El tono blanco representa menor precipitación, mientras que los azules una mayor cantidad de lluvia.	21
Figura 6. Índice estandarizado de precipitación para la Estación Aeropuerto (1937- junio, 2020). Con la letra A, se señala la sequía extrema de los meses julio-agosto de 1939. La letra B, marca las sequías extremas de setiembre de 1996 y 1997, mientras que la letra C, resalta la de julio del 2014.	22
Figura 7. Mapa de calor que representa los estados de sequía por mes de toda la serie temporal para la Estación Aeropuerto.	23
Figura 8. Distribución Acumulada de la serie temporal del Índice Estandarizado de Precipitación, para la Estación Aeropuerto.	32
Figura 9. Probabilidades por mes del modelo de Redes Bayesianas de primer orden, en la época lluviosa del 2000 al 2020.	35
Figura 10. Probabilidades por mes del modelo de Redes Bayesianas de segundo orden, en la época lluviosa del 2000 al 2020.	38
Figura 11. Ranked Probability Score (RPS) de los modelos de Cadenas de Markov y Redes Bayesianas de primer y segundo orden, para las cuatro categorías.	39
Figura 12. Ranked Probability Score (RPS) de los modelos de Cadenas de Markov y Redes Bayesianas de primer y segundo orden, únicamente para los estados de sequía.	41
Figura 13. Ranked Probability Skill Score (RPSS) de los modelos de Cadenas de Markov y Redes Bayesianas de primer y segundo orden, para las cuatro categorías.	42

Figura 14. Ranked Probability Skill Score (RPSS) de los modelos de Cadenas de Markov y Redes Bayesianas de primer y segundo orden, únicamente para los estados de sequía.....	43
Figura 15. Precipitación promedio mensual para las estaciones Cañas Dulces (A) y Talolinga (B). Ambas con un periodo de 1950 a junio, 2020.....	44
Figura 16. Mapa de calor que representa la precipitación promedio mensual (1950-junio, 2020) para las estaciones Cañas Dulces (A) y Talolinga (B). El tono blanco representa menor precipitación, mientras que los azules una mayor cantidad de lluvia.....	45
Figura 17. Mapa de calor que representa los estados de sequía por mes de toda la serie temporal para las estaciones Cañas Dulces (A) y Talolinga (B).....	47
Figura 18. Probabilidades por mes del modelo de Redes Bayesianas de primer orden, en la época lluviosa del 2000 al 2020, en la Estación Cañas Dulces.	51
Figura 19. Probabilidades por mes del modelo de Redes Bayesianas de primer orden, en la época lluviosa del 2000 al 2020, en la Estación Talolinga.	52
Figura 20. Ranked Probability Score (RPS) del modelo Redes Bayesianas de primer, únicamente para los estados de sequía en las estaciones de Cañas Dulces (A) y Talolinga (B).	53
Figura 21. Ranked Probability Skill Score (RPSS) del modelo Redes Bayesianas de primer, únicamente para los estados de sequía en las estaciones de Cañas Dulces (A) y Talolinga (B).	54

Lista de abreviaturas

<i>BN1</i>	<i>Bayesian Network</i> de primer orden
<i>BN2</i>	<i>Bayesian Network</i> de segundo orden
<i>CDF</i>	<i>Cumulative Distribution Function</i>
<i>CHIRPS</i>	<i>Climate Hazards Group InfraRed Precipitation with Station</i>
<i>DRAT</i>	Distrito de Riego Arenal - Tempisque
<i>DI</i>	<i>Drought Index</i>
<i>ENOS</i>	El Niño – Oscilación del Sur
<i>FAO</i>	<i>Food and Agriculture Organization</i>
<i>IMN</i>	Instituto Meteorológico Nacional
<i>MAG</i>	Ministerio de Agricultura y Ganadería
<i>MC1</i>	<i>Markov Chain</i> de primer orden
<i>MC2</i>	<i>Markov Chain</i> de segundo orden
<i>NASA</i>	<i>National Aeronautics and Space Administration</i>
<i>PIAAG</i>	Programa Integral de Abastecimiento de Agua para Guanacaste
<i>RPS</i>	<i>Ranked Probability Score</i>
<i>RPSS</i>	<i>Ranked Probability Skill Score</i>
<i>SPI</i>	<i>Standardized Precipitation Index</i>
<i>WMO</i>	<i>World Meteorological Organization</i>
<i>ZCIT</i>	Zona de Convergencia Intertropical



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, Kenneth Jesús Gutiérrez García, con cédula de identidad 1-1495-0693, en mi condición de autor del TFG titulado 'Análisis probabilístico de sequía meteorológica mediante Cadenas de Markov y Redes Bayesianas; Cuenca del Río Tempisque'

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.

Kenneth Gutiérrez G.

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 al problema, al área de estudio, a los objetivos, conceptos y a la metodología

1.1 El problema y su importancia

La sequía es un proceso atmosférico que afecta a todas las zonas climáticas del planeta, y su impacto se ve reflejado en costos sociales, políticos y económicos; Ban Ki-moon (2013) afirmó en su mensaje para el “Día Mundial de Lucha contra la Desertificación y la Sequía” que: para el 2011, la sequía en el Cuerno de África afectó a casi 13 millones de personas, además “en 2012, los Estados Unidos de América padecieron su peor sequía desde los años cincuenta, que afectó al 80% de las tierras de cultivo”. Según Andreu (2015) las proyecciones del cambio climático a nivel mundial indican que es probable que las sequías sean más frecuentes y graves debido a la mayor probabilidad de inviernos cálidos y veranos más calientes, junto con disminuciones en la precipitación y aumentos en la evaporación y la transpiración. En 2016, América Central vivió una de las sequías más graves en los últimos 10 años, pues más de 3,5 millones de personas necesitaron asistencia; y los mayores impactos se reflejaron en la seguridad alimentaria, pues los cultivos no tuvieron una buena cosecha, lo cual resultó en grandes pérdidas económicas (FAO, 2016).

Costa Rica no es la excepción, la región Pacífico Norte históricamente, ha sido la más afectada por estos eventos. En el 2015 se presentó una de las sequías más intensas en la provincia de Guanacaste, en este caso El Niño contribuyó a agravar la situación, pues hubo déficit en las precipitaciones medias de hasta un -65% en la Estación Meteorológica Liberia (IMN, 2015). Sin embargo, esta región se caracteriza por tener una vegetación adaptada a sobrevivir periodos de poca precipitación, ya que la época seca es muy marcada e incluso hay meses en los que la precipitación registrada es cero. Esto quiere decir que la sequía no es el problema, pues la región naturalmente está adaptada. El problema se da cuando la sociedad es afectada. Según Mishra y Desai (2006) existe una deficiencia en los esfuerzos para mitigar los efectos de la sequía, y la incapacidad para pronosticar las condiciones secas con una anticipación de, al menos, algunos meses es la causa principal. En el caso de Costa Rica, Birkel (2006) sustenta que se han realizado

“pocos esfuerzos para analizar el comportamiento de sequías hidrológicas y, en particular, las tendencias en las series temporales en Costa Rica”.

De estos argumentos se infiere la importancia de pronosticar la sequía, y principalmente la sequía meteorológica, pues la disminución de precipitaciones es el inicio de varios efectos como la disminución en la escorrentía de los ríos, en la humedad del suelo y en el nivel de los acuíferos, con el consiguiente impacto en la producción agrícola y el bienestar de las poblaciones. Como existe un desfase temporal entre la disminución de precipitación y sus efectos, un pronóstico de sequía meteorológica conlleva inherentemente la anticipación de sus impactos, tanto naturales como sociales. Con un pronóstico probabilístico de sequía meteorológica se pueden tomar decisiones a tiempo con el fin de minimizar los impactos económicos, sociales y ambientales; y de esta manera contribuir a optimizar la gestión del riesgo y el manejo de los recursos hídricos.

1.2 Área de estudio

La cuenca del río Tempisque (Figura 1) se encuentra en la región climática denominada Pacífico Norte, según el Instituto Meteorológico Nacional (IMN) (2008) esta región es una de las zonas con menor riqueza hídrica del país, pues los caudales disminuyen notablemente durante la estación seca, que se prolonga de cinco a seis meses. En promedio, precipita anualmente 1700 mm, con temperaturas de 32°C en el día y 22°C en la noche. Presenta un clima tropical seco, con una temporada seca que inicia en diciembre y finaliza en abril, mientras que la temporada lluviosa se extiende de mayo a noviembre; sin embargo, es normal que en el mes de julio haya una disminución en la precipitación, a la cual se le conoce como Veranillo de San Juan o canícula, esto por un período de 2 a 3 semanas (Birkel, *et ál.*, 2017). Las máximas precipitaciones ocurren en octubre y se relacionan con la posición de la Zona de Convergencia Intertropical (ZCIT). La variación de las estaciones se muestra también en los caudales de los ríos, pues desde 1973 al 2003 la estación de medición Guardia, en el Tempisque, mostró promedios de caudales de 24.6 m³/s; sin embargo, se han medido caudales de 2.6 m³/s en época seca (Birkel *et ál.*, 2017).

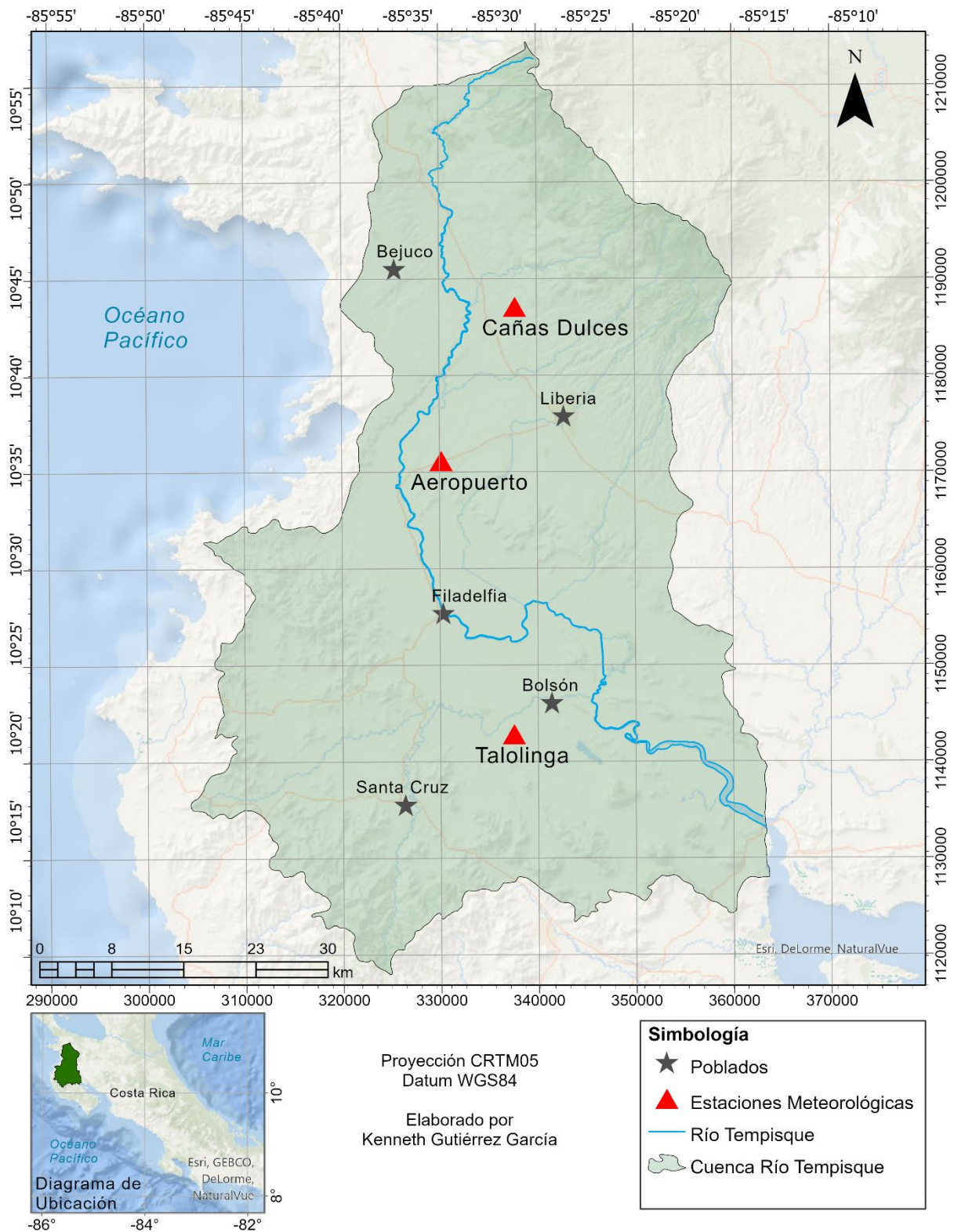


Figura 1. Localización de la cuenca del Río Tempisque.

La región es una de las principales productoras de arroz y melón, genera casi el 45% de azúcar nacional y procesa más del 50% de la caña (PIAAG, s.f.). El área utilizada para la agricultura aumentó de un 0% en 1950 a un 27.8% en el 2010, esto se debió al cambio de modelo agrícola que se dio en la década de los 80, en donde antes había un modelo orientado a la subsistencia, se impuso un modelo industrial orientado a la exportación (Birkel *et ál*, 2017).

Estas actividades se realizan de manera intensiva y utilizan al río Tempisque como fuente principal para sustentar sus necesidades hídricas, lo cual realizan por medio de concesiones para la extracción de agua. En todo el río existen 16 concesiones con una cantidad asignada cada una, “pero no hay forma real de confirmar y controlar que sea ese volumen el que finalmente se extrae” (Mora, 2016). Según Mena (2019) casi el 65% de la población económicamente activa trabaja en el sector terciario, un 21% en el sector primario y un 14% en el sector secundario. En 1984 se estableció el Distrito de Riego Arenal-Tempisque (DRAT) (Decreto Ejecutivo N°15321-MAG, 1984) para aprovechar las aguas procedentes del Proyecto Hidroeléctrico del Arenal mediante canales de riego para la agricultura. De esta manera se ha logrado mantener la actividad agrícola en una parte de la cuenca sin presionar al río Tempisque.

1.3 Antecedentes

Alrededor del mundo ha habido una creciente necesidad por conocer eventos de sequía al futuro, para poder tomar las mejores decisiones posibles y así minimizar los impactos de muchas amenazas. Esto se ha logrado en el caso de muchos eventos climáticos, como por ejemplo los huracanes, los cuales por su comportamiento ha sido posible identificar las probabilidades de su recorrido e impacto; sin embargo, el caso de la sequía es más complejo, ya que puede variar en duración e intensidad, e incluso, puede haber impactos en zonas en donde no se registró un déficit de agua.

Por ejemplo, si la sequía sucede en la parte alta de una cuenca, puede que los caudales de los ríos disminuyan, lo cual afectaría a las zonas media y baja de la cuenca. Además, la severidad del impacto depende las necesidades hídricas que existan en un determinado sistema. Como consecuencia de esto, ha habido esfuerzos de varios

investigadores para plantear índices, modelos y metodologías que traten de explicar, predecir y delimitar la sequía.

En el 2011, Mishra y Singh, publicaron un artículo titulado *Drought modeling – A review*, en el cual hacen una revisión de diferentes metodologías que han sido utilizadas para el modelamiento de sequía. Este artículo es importante, porque presenta un “resumen” de lo que se necesita para poder modelar eventos de sequía. En la explicación que hacen de cada metodología presentan los resultados de los trabajos de otros investigadores que han aplicado las técnicas y, además, muestran los diferentes componentes necesarios para el pronóstico por medio de un esquema (Figura 2) que sirve de guía metodológica y que resume las diferentes variables hidrometeorológicas, los índices de sequía, los índices climáticos y las metodologías de modelamiento de sequía más utilizadas.

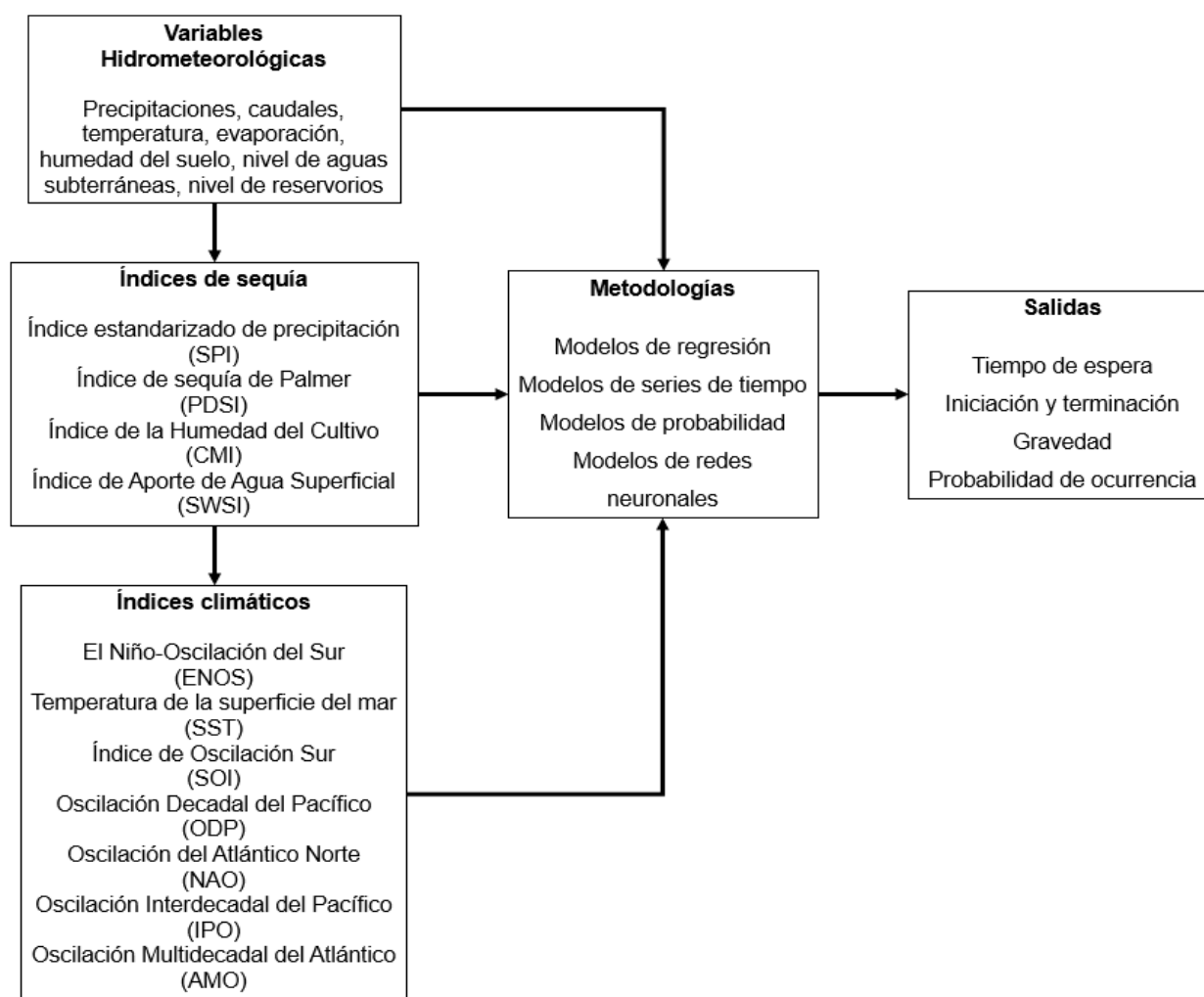


Figura 2. Componentes para el pronóstico de sequías. (Mishra y Singh, 2011).

En el artículo titulado *Probabilistic Drought Analysis of Weekly Rainfall Data using Markov Chain Model*, los autores Alam, N., et al. (2015); investigaron el patrón de ocurrencia de semanas húmedas y secas en áreas de la India, propensas a la sequía. Ellos sugieren que los resultados del artículo pueden ser útiles para los planificadores agrícolas, en el sentido que se pueden identificar las áreas en donde el desarrollo agrícola debería enfocarse a estrategias de mitigación a corto y largo plazo. Con esta investigación se comprueba que las Cadenas de Markov pueden ser utilizadas para analizar las frecuencias de eventos meteorológicos, en este caso las sequías y comprueba su funcionalidad para la toma de decisiones.

La tesis doctoral de Avilés (2017), titulada *Pronóstico Probabilístico de Eventos de Sequía y Evaluación del Riesgo en la Gestión de Sistemas de Recursos Hídricos. Caso de Estudio en una Cuenca Andina Regulada*, “evalúa el riesgo de fallo en el suministro de agua mediante la introducción de pronósticos probabilísticos de eventos de sequía, como soporte para la toma de decisiones con respecto a la satisfacción de las demandas” (Avilés, 2017). Esta tesis utilizó modelos estocásticos basados en Cadenas de Markov y Redes Bayesianas para predecir eventos de sequía; además, relacionó estos pronósticos con una metodología de evaluación para conocer el riesgo de fallo en los sistemas de recursos hídricos. Esta tesis se utilizó como fuente para la metodología de esta investigación.

Para el caso de Costa Rica, el pronóstico de sequía no ha sido investigado con mucha profundidad. Sin embargo, las implicaciones de la sequía sí se han estudiado en diferentes trabajos. La mayor cantidad de información se puede encontrar en los boletines que emite el Instituto Meteorológico Nacional (IMN), no obstante, se limitan a presentar los datos de meses anteriores y las proyecciones se basan en el comportamiento de El Niño. Sin embargo, sí ha habido varios trabajos que han tratado de explicar, más allá de los números, el comportamiento de la sequía en Costa Rica.

Birkel (2005), en su tesis de maestría titulada *Temporal and Spatial Variability of Drought Indices in Costa Rica* trata de determinar el comportamiento temporal y espacial de la sequía en Costa Rica. También vinculó los patrones de circulación atmosférica con las sequías regionales, para determinar sus causas. Además, el autor cree que el estudio

puede alentar al desarrollo de modelos que pronostiquen las sequías a largo plazo, con el fin de proveer información importante para realizar pasos preventivos en los planes de acción para un manejo sostenible de los recursos hídricos.

En el artículo *Recurso Hídrico Superficial en la cuenca del río Tempisque, Costa Rica: Escenarios de Disponibilidad y Demanda*, Guzmán-Arias (2013), proyecta al 2030 la disponibilidad y demanda del recurso hídrico en la parte alta y media de la cuenca. En este trabajo, la autora incorpora transformaciones en los flujos de agua por cambios en el uso de la tierra, así como por el cambio climático. Un resultado de las proyecciones para el 2030 es que el balance de disponibilidad y demanda para la parte alta de la cuenca, es de números negativos. Este balance se realizó considerando la oferta, demanda, oferta real y la demanda ambiental, es decir, que ya desde la parte alta hay un déficit en la disponibilidad de agua. Esto significa que es posible que en la parte media y baja ya no haya un caudal que logre mantener la demanda del agua. Guzmán-Arias cree que la única alternativa es trabajar en un modelo de planificación, por eso es importante conocer los eventos de sequía y así se tendrán proyecciones acerca de la disponibilidad de agua en las diferentes zonas de la cuenca y en diferentes épocas del año.

Durham *et al.* (2016); realizaron un trabajo titulado *Monitoring Drought and Water Balance in the Guanacaste Province to Enhance Decision Making and Response Planning in Costa Rica*, en el cual utilizaron información de varios satélites y de estaciones *in situ* para analizar y monitorear el estado de la sequía meteorológica y agrícola en el 2016. Para ello, utilizaron dos índices con el objeto de evaluar los dos tipos de sequía por separado y, además, crearon información para la evaluación de un balance hídrico en la cuenca Arenal-Tempisque. Los autores afirman que los resultados permiten hacer una planificación de la gestión hídrica más eficiente.

Steyn *et al.* (2016); elaboraron un informe para el proyecto *FuturAgua*, llamado *Temporal and Spatial Variability of Annual Rainfall Patterns in Guanacaste, Costa Rica*. En este trabajo analizaron 38 años de datos de precipitación en Guanacaste, con el propósito de descubrir la variabilidad espacial y temporal de la lluvia; para esto aplicaron análisis estadísticos y técnicas de modelamiento y determinaron que las precipitaciones en períodos menores a dos semanas son dominadas por una fuerte variabilidad estocástica,

mientras que los períodos más largos tienen una variación sistemática. La importancia del trabajo es que utilizaron un análisis Bayesiano para confirmar la conveniencia del Modelo Gaussiano utilizado para determinar la variabilidad de la lluvia, esto quiere decir que en Guanacaste ya se logró implementar eficientemente un modelo probabilístico como el Bayesiano.

1.4 Objetivo general

Analizar el rendimiento de los modelos probabilísticos, Cadenas de Markov y Redes Bayesianas, con eventos de sequías meteorológicas históricas desde 1937 hasta 2020, en tres estaciones meteorológicas ubicadas dentro de la cuenca del río Tempisque, para que funcionen como una herramienta para el pronóstico de eventos de sequía meteorológica.

1.5 Objetivos específicos

- Desarrollar y comparar los modelos probabilísticos de Cadenas de Markov y Redes Bayesianas de primer y segundo orden con registros a largo plazo de precipitación de la Estación Aeropuerto, usando los criterios *RPS* y *RPSS* para evaluar su rendimiento.
- Validar el modelo de mejor rendimiento con datos de precipitación de las estaciones Cañas Dulces y Talolinga, esto para comprobar o no, la aplicabilidad del modelo con diferentes series de precipitación.

1.6 Marco teórico

1.6.1 Periodos secos en el Pacífico Norte

Según Patterson (1992), la sequía en Guanacaste es un fenómeno natural que se presenta cada cierto tiempo, y que se genera por la combinación de varios factores atmosféricos, económicos, sociales y geológicos. Sin embargo, es importante realizar una diferenciación entre periodos secos y sequías, pues son eventos que se presentan de manera diferente y tienen un origen diferenciado.

El Pacífico Norte es la región con menos precipitación del país (1500 mm), sin embargo, si se compara con otras zonas del planeta, por ejemplo, el Mediterráneo, en donde la precipitación promedio anual esta entre 400 y 800 mm, se podría pensar que en el

Pacífico Norte no debería tener problemas de abastecimiento de agua, pero, según Vargas (2001) el balance hídrico se vuelve crítico entre noviembre y abril. Durante estos meses las condiciones edáficas de suelos que son muy deficientes en la retención del agua, así como la fuerza de los vientos del este y noreste; provocan que los periodos secos varíen en su intensidad y recurrencia. A estos factores meteorológicos se le debe agregar un factor geológico, el cual afecta la región en de la depresión del río Tempisque, ya que existe una formación de tobas e ignimbritas, las cuales, por su porosidad, favorece la infiltración rápida y profunda del agua, generando problemas con la retención superficial de agua para las plantas.

Un factor a tomar en cuenta es la oscilación *ENOS*, concretamente en su fase El Niño, la cual se caracteriza por el calentamiento de las aguas del Océano Pacífico Oriental, frente a las costas de Perú y Ecuador, esto coincide con un debilitamiento de los vientos alisios y la disminución de la presión atmosférica en esta zona, llevando a un calentamiento de las aguas oceánicas que provoca, en la vertiente Pacífica de Costa Rica, una estación seca más cálida y árida de lo normal, mientras que las precipitaciones tienden a disminuir durante la estación lluviosa, situación que se acentúa en la provincia de Guanacaste donde las lluvias comienzan a finales de mayo, sufren un descenso en junio y se restablecen con regularidad hasta setiembre (Díaz y Alfaro, 2016).

En un estudio realizado por Patterson (1992) se identificó que, durante 32 años, hubo 63 meses de sequías, que se intensificaban durante los meses de mayo, julio y agosto, se puede decir que los eventos están asociados con la disminución en la fuerza de los alisios del sur (periodo del Veranillo de San Juan), con esto se podría asumir que los vientos del sur son los que aportan mayor cantidad de lluvia a la región. Sin embargo, el mismo autor afirma que la sequía en la zona no solo obedece a una disminución en la cantidad de precipitación sino también a una distribución irregular tanto en el tiempo como en el espacio.

1.6.2 El concepto de sequía

Una de las deficiencias en la mitigación de la sequía, es definir universalmente este concepto, pues a diferencia de otros eventos hidrometeorológicos, como huracanes, tormentas o inundaciones, la sequía no cuenta con una definición que aplique en

cualquier parte del mundo. Para Birkel (2005) la sequía al igual que las inundaciones son extremos hidrológicos que forman parte de la variabilidad natural del clima. Van Loon (2015) en su artículo titulado *Hydrological Drought explained* define la sequía como un periodo sostenido de baja disponibilidad de agua con respecto a lo normal, con características espaciales y temporales que varían de una región a otra.

Wilhite y Glantz (1985) clasificaron la sequía en cuatro términos básicos para aproximar su medición, estas divisiones son: la sequía meteorológica, hidrológica, agrícola y socioeconómica, visualizadas en el la Figura 3. Las tres primeras tratan de aspectos físicos de la sequía, mientras que la última está planteada en términos de oferta y demanda y como la sequía puede afectar el sistema socioeconómico.

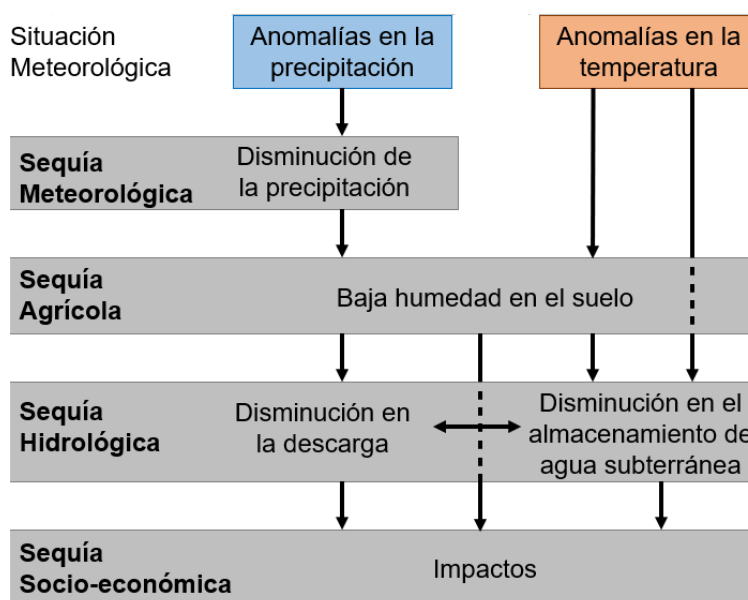


Figura 3. Clasificación de la sequía y su desarrollo. Adaptado Wilhite & Glantz (1985) y Van Loon (2015).

La sequía meteorológica usualmente es definida como el grado de sequedad, en comparación con la cantidad promedio, y en la duración de un periodo seco. En ese sentido esta definición depende de la región en la que se aplique, pues los promedios de precipitación varían de un lugar a otro y con esto el grado de sequedad que pueda soportar el sistema, así como el tiempo de duración del evento. En el caso de la sequía hidrológica se asocia con los efectos que puede provocar la disminución de la precipitación en el sistema hidrológico, usualmente existe un desfase temporal entre este

tipo de sequía y la meteorológica, ya que toma cierto tiempo para que se evidencie la disminución de precipitación en una disminución en la humedad del suelo, en los niveles del agua subterránea, en el caudal de los ríos y en los reservorios.

Por otro lado, la sequía agrícola se refiere al impacto de la sequía meteorológica en la productividad de los cultivos, el cual varía según el requerimiento hídrico de las plantas que depende de las condiciones meteorológicas, las características biológicas de las especies cultivadas, el estado de crecimiento y de las propiedades físicas y biológicas del suelo (Wilhite & Glantz, 1985). Como se mencionó brevemente, la sequía socioeconómica normalmente se plantea en términos de oferta y demanda del recurso hídrico que genera un impacto al sistema económico y social. Hoyt (citado por Wilhite & Glantz, 1985), en 1936 propuso una definición bastante simple, pero comprensible, según él, este tipo de sequía ocurre cuando la precipitación no es suficiente para satisfacer las necesidades establecidas por las actividades humanas, en donde el desarrollo económico de una región genera una mayor demanda de agua de la que normalmente está disponible.

Como se evidencia, el concepto de sequía depende de varios factores y es adaptable según el objetivo de estudio o de la región. En ese sentido para esta tesis se propone determinar sequía desde la perspectiva meteorológica, pues se trabajó solo con datos de precipitación; además, teniendo en cuenta que la cuenca del Tempisque se caracteriza por tener una estación seca bien marcada de diciembre a abril, en la que hay una disminución de precipitación que se da de manera natural todos los años, la sequía meteorológica en la región solo puede ocurrir en los meses de estación lluviosa.

1.6.3 Modelo probabilístico

Según Mora (2010) el objetivo de los modelos probabilísticos es establecer procedimientos teóricos de evaluación o valoración del grado de certeza con que se espera que ocurra una de las eventualidades posibles de un fenómeno aleatorio. Con esta definición básica se puede decir que con un modelo probabilístico se obtiene la certidumbre de que un evento suceda en el futuro y al mismo tiempo la incertidumbre de que no ocurra.

En el artículo de Mishra y Singh (2011) se hace una revisión de los diferentes métodos para el pronóstico y modelamiento de las sequías. Uno de ellos es el análisis de regresión, que asume que el predictor y la predicción son lineales, lo que impide un adecuado pronóstico a largo plazo (Avilés, 2016). Otro método son los modelos de series temporales, que a pesar de que identifican, estiman y diagnostican los eventos, también tienen un comportamiento lineal (Mishra y Desai, 2006). Las redes neuronales son otra de las metodologías utilizadas, son modelos no lineales que a partir de los datos puede caracterizar los patrones y estimar su relación, sin embargo, tienen la desventaja de ser una caja negra, tener un desarrollo empírico, además de ser una gran carga computacional (Mishra y Singh, 2011).

La ventaja de los modelos probabilísticos yace en la capacidad de considerar la incertidumbre asociada con las variables, algo que los modelos citados en el párrafo anterior no logran ya que principalmente son determinísticos. Otra ventaja, es que estos modelos calculan la probabilidad condicional, lo cual permite pronosticar futuros eventos de sequía en función de las condiciones actuales (Avilés, 2016).

1.7 Metodología

1.7.1 Datos

Para el análisis y modelamiento de la sequía se cuenta con una serie temporal de 70 años (1950-2020) en dos estaciones meteorológicas, las cuales son, la estación Cañas Dulces y la estación Talolinga. Se utilizó una tercera estación (Aeropuerto) la cual tiene un registro desde 1937 hasta junio 2020. Los datos se obtuvieron del IMN, sin embargo, estas series de tiempo tenían vacíos de información que fueron rellenos por medio del sensor remoto *Terrestrial Air Temperature and Precipitation: V4.01* de la NASA (Willmott y Matsuura, 2001).

Adicionalmente para completar los últimos 5 años se utilizaron datos del *Climate Hazards Group Infrared Precipitation with Stations (CHIRPS)* (Funk et ál, 2014), todos estos datos fueron corregidos de manera mensual y fueron calibrados para su ajuste con los datos observados *in situ*, debido a esto es esperable que haya diferencias entre los datos observados y los datos *CHIRPS*, pues son productos globales y calculados a partir de información satelital. La serie temporal se dividió en un periodo de calibración, desde

1937 hasta 1999 para la estación del Aeropuerto y desde 1950 hasta 1999 para Cañas Dulces y Talolinga; y en un periodo de validación desde el 2000 hasta junio 2020, para todas las estaciones.

1.7.2 Categorías de sequía

Con los datos del periodo de calibración se construyó el Índice Estandarizado de Precipitación (*SPI*) de manera mensual, que posteriormente se reclasificó en cuatro estados para tener así un Índice de Sequía (*DI*), que permite caracterizar los períodos de sequía, ya que el *DI* representa una serie de tiempo con la cual se calcula la frecuencia, severidad y probabilidad de ocurrencia de los estados de la sequía. El *DI* básicamente es una reclasificación de los valores del *SPI*, para esta tesis se clasificó en cuatro categorías los eventos de sequía (Tabla 1).

Tabla 1. Clasificación de las categorías de sequía, según valores de *SPI* y CDF.

Categoría (Estado)	<i>SPI</i>	Umbral CDF
No sequía (0)	Mayor a -0.99	Mayor a 0.158
Sequía moderada (1)	-1.0 a -1.49	0.158 a 0.066
Sequía severa (2)	-1.5 a -1.99	0.065 a 0.021
Sequía Extrema (3)	Menor a -2.0	Menor a 0.02

En esta clasificación, se agrupan las categorías de moderado, muy húmedo, y extremadamente húmedo en una sola categoría denominada “No sequía”, con esto se entiende que la condición de humedad no es relevante como sí lo es la severidad de la sequía, lo cual es útil en el caso de que solo se cuente con datos de precipitación o caudales. Con esto se creó una matriz de probabilidades de transición de primer y segundo orden que se utilizó en el modelo basado en cadenas de Markov. Mientras que se calculó la función de distribución normal acumulada (CDF) de los valores del *SPI*, necesaria para los modelos basados en la red Bayesiana.

1.7.3 Cadenas de Markov

Los modelos probabilísticos basados en Cadenas de Markov, son los que normalmente se utilizan para la predicción de la sequía (Mishra y Singh, 2010; Banimah y Khalili, 2013; Avilés, 2016), estos modelos se componen de “un conjunto de matrices de probabilidades de transición que indican las probabilidades de ocurrencia de los estados de un sistema

para un intervalo de tiempo del futuro a partir de la información del estado actual y/o estados de los intervalos pasados, dependiendo de la orden del modelo” (Avilés, 2017).

Estas probabilidades de transición son las que rigen el comportamiento de las Cadenas de Markov y esto las convierte en probabilidades condicionales (Wilks, 2006). A partir de estas matrices de transición es que se puede pronosticar la categoría de sequía del siguiente mes, basado en la observación del mes actual para Cadenas de Markov de primer orden (MC1). En el caso de Cadenas de Markov de segundo orden (MC2), es el mismo principio, pero, con la diferencia que el pronóstico del siguiente mes, se basa en la observación de dos meses previos. Como en esta tesis se trabajó con series temporales continuas, se utilizó Cadenas de Markov no homogéneas, así cada probabilidad de transición varía mes a mes.

1.7.3.1 Cadenas de Markov de primer orden

Wilks (2006) explica que las Cadenas de Markov de primer orden (MC1) tienen la característica de que las probabilidades de transición de cada observación en la serie temporal dependen sólo del valor del miembro anterior de la serie temporal. Es decir, la probabilidad de transición de un mes se rige por la categoría del mes anterior. Según Wilks (2006) esta característica es conocida como la Propiedad de Markov, que se expresa de la siguiente manera:

$$p_{ij} = Pr\{X_{t+1} = j | X_t = i\} \quad (1)$$

Donde p_{ij} representa la probabilidad de transición en la que, X_{t+1} es igual a la categoría j , dado que X_t es igual a la categoría i . Para calcular la probabilidad de transición (\hat{p}_{ij}) se debe estimar las frecuencias relativas que son condicionales de las transiciones (m_{ij}), lo cual se puede expresar como:

$$\hat{p}_{ij} = \frac{m_{ij}}{\sum_j m_{ij}} \quad i, j = 1 \quad (2)$$

El numerador representa la cantidad de transiciones de la categoría i a la categoría j , mientras que el denominador es la suma de la cantidad de transiciones de la categoría i a cualquier otra categoría.

1.7.3.2 Cadenas de Markov de segundo orden

Las Cadenas de Markov de segundo orden (MC2) siguen el mismo principio de la Propiedad de Markov, con la diferencia de que se consideran en las matrices de transición dependencias de más de dos periodos de tiempo (Wilks, 2006). A diferencia de los modelos de MC1, los MC2 requieren de tres subíndices para el cálculo de las probabilidades de transición. El primer subíndice es el estado o categoría del tiempo $t - 1$, el segundo es la categoría del tiempo t , y el tercero es el estado en el tiempo futuro $t + 1$. Según Wilks (2006), esto se expresa así:

$$p_{hij} = Pr\{X_{t+1} = j | X_t = i, X_{t-1} = h\} \quad (3)$$

En este caso, la probabilidad de transición p_{hij} , en la que X_{t+1} sea igual al estado j , es dada por X_t igual a la categoría i , y X_{t-1} igual a la categoría h . Las probabilidades de transición \hat{p}_{hij} se estiman con:

$$\hat{p}_{hij} = \frac{m_{hij}}{\sum_j m_{hij}} \quad h, i, j = 1 \quad (4)$$

Para los modelos MC2, el numerador es la cantidad de transiciones de las categorías h, i a la categoría j y el denominador es igual a la suma de la cantidad de transiciones de las categorías h, i , a cualquier otra categoría. Es importante tener en cuenta que este modelo, a diferencia del MC1, toma en cuenta valores más atrás en el tiempo, por lo que el número de transiciones posibles aumenta exponencialmente, de esta manera se obtiene matrices más extensas.

1.7.4 Redes Bayesianas

Las Redes Bayesianas consisten, según Avilés (2016), en una estructura de redes que abarcan un conjunto de afirmaciones de independencia condicional con ciertas variables, así como una serie de distribuciones de probabilidad asociadas a una variable. Para Heckerman (1995), estos modelos usualmente representan la incertidumbre entre la relación de las variables de un dominio. Así dado un dominio de variables $U = \{x_1, \dots, x_n\}$, la probabilidad de la distribución conjunta para U es una distribución de probabilidad sobre los estados de U .

Según Heckerman (1995), el primer paso para construir una Red Bayesiana es decidir las variables y los estados que se modelaran, en este caso son las categorías representadas en la Tabla 1. El segundo paso consiste en construir un gráfico acíclico dirigido que codifica las afirmaciones de independencia condicional, a este gráfico se le denomina como la estructura de la Red Bayesiana, la cual, de manera general, se expresa de la siguiente forma:

$$P(x_{tn} | x_{t1}, \dots, x_{tn-1}) = \frac{P(x_{t1}, \dots, x_{tn})}{P(x_{t1}, \dots, x_{tn-1})} \quad (5)$$

En donde x_{tn} representa las probabilidades del pronóstico, que se calculan a partir de la división de las variables predictoras, es decir las distribuciones conjuntas que se asocian a las variables.

1.7.4.1 Cópulas

Según Madadgar y Moradkhani (2013), se requiere un esfuerzo analítico intenso para modelar directamente el comportamiento de las variables y sus distribuciones, para lograrlo se utilizaron funciones de cópulas. Según Schmidt (2006) las cópulas son una herramienta para modelar la dependencia de variables aleatorias. El término proviene de la palabra latina *copulare*, que significa conectar o unir, y fue usado por primera vez por Sklar en 1959. Para Avilés (2016), las cópulas son una función de distribución multivariada cuyos márgenes unidimensionales son uniformes en el intervalo [0,1].

De acuerdo con la formulación de Yan (2006), si se considera un vector al azar $(U_1, \dots, U_n)^T$, en donde cada marginal $U_i, i = 1, \dots, n$, es una variable aleatoria uniforme en el intervalo [0,1]. Y suponiendo que la función de distribución acumulada (CDF) de $(U_1, \dots, U_n)^T$, es:

$$C(U_1, \dots, U_n) = P(U_1 \leq u_1, \dots, U_n \leq u_n) \quad (6)$$

Se tiene que la función C es una copula n-dimensional y la CDF especifica las probabilidades en las que un valor aleatorio X no exceda umbrales particulares (Avilés, 2016). También se debe tomar en cuenta el Teorema de Sklar, el cual describe cómo las cópulas interactúan entre las funciones de distribución multivariada y su distribución

marginal. El teorema establece que una distribución multivariada $F(X_1, \dots, X_n)$, para todas las X en el dominio F , puede expresarse de la siguiente manera (Avilés, 2016):

$$F(X_1, \dots, X_n) = C\{U_1(X_1), \dots, U_n(X_n)\} \quad (7)$$

En donde la distribución marginal en el intervalo $[0,1]$ se representa por $U_1(X_1)$, mientras que C es la distribución acumulada de la copula. Con esto en cuenta la Ecuación (5), se puede reformular como:

$$P(x_{tn} | x_{t1}, \dots, x_{tn-1}) = \frac{C(U_{t1}, \dots, U_{tn})}{C(U_{t1}, \dots, U_{tn-1})} \quad (8)$$

A partir de la Ecuación (8), se construyen los modelos de Redes Bayesianas. Para los de primer orden (BN1), es necesario que $n = 2$; mientras que si $n = 3$, tenemos los modelos de Redes Bayesianas de segundo orden (BN2), los cuales se expresan así respectivamente:

$$P(X_{t2} | X_{t1}) = \frac{C(U_{t1}, U_{t2})}{U_{t1}} \quad (9)$$

$$P(X_{t3} | X_{t1}, X_{t2}) = \frac{C(U_{t1}, U_{t2}, U_{t3})}{C(U_{t1}, U_{t2})} \quad (10)$$

Siguiendo la metodología de Avilés (2016), las Ecuaciones (9) y (10), se deben reescribir para que las probabilidades de cada categoría no excedan los umbrales de la CDF establecidos en la Tabla 1. Considerando que xd_s representa el umbral de la CDF, las ecuaciones se definen como:

$$P(X_{t2} \leq xd_s | X_{t1}) = \frac{C[U(X_{t1}), U(X_{t2} \leq xd_s)]}{U(X_{t1})} \quad (11)$$

$$P(X_{t3} \leq xd_s | X_{t1}, X_{t2}) = \frac{C[U(X_{t1}), U(X_{t2}), U(X_{t3} \leq xd_s)]}{C[U(X_{t1}), U(X_{t2})]} \quad (12)$$

1.7.4.2 Ajuste de Cópulas

De acuerdo a Avilés (2016), el siguiente paso es encontrar estructuras de dependencia adecuadas y formar distribuciones conjuntas. Para esto se modeló la dependencia temporal realizando un ajuste de cópulas de acuerdo a las distribuciones marginales. La función de distribución acumulada se utilizó para convertir las observaciones en pseudo-observaciones con valores en el intervalo $[0,1]$, que son la distribución marginal utilizada en el ajuste de las cópulas.

Se trabajó con cuatro tipos de cópulas, dos elípticas (Normal y *Student*) y dos arquimedianas (Clayton y Frank). Las de mejor ajuste se seleccionaron según dos criterios. El primero, que cumplieran con un *p-value* mayor a 0.05; y el segundo es que dentro de las cópulas que pasaran el primer criterio, se seleccionó la copula con el valor estadístico (S) más bajo. Posteriormente se aplicaron estas cópulas en las Ecuaciones (11) y (12) para formar las redes bayesianas, tanto de primer orden como de segundo orden.

1.7.5 Verificación de los modelos

De los pasos anteriores se obtienen cuatro modelos que fueron verificados estadísticamente con el objetivo de seleccionar el más apto para pronosticar eventos de sequía. La validación se determinó mediante dos formas, la primera por medio de un puntaje ponderado de probabilidad, conocido en inglés como *Ranked Probability Score (RPS)*, el cual es expresado como:

$$RPS = \sum_{m=1}^s (Y_m - O_m)^2 \quad (13)$$

Este método muestra los errores del pronóstico en términos de la probabilidad asignada a los eventos, además, incluye una penalización por los pronósticos que están más lejanos de las observaciones (Avilés, 2017). El resultado de la Ecuación (13) es la diferencia entre las probabilidades pronosticadas (Y_m) y las probabilidades observadas (O_m). Un pronóstico perfecto deberá tener un RPS igual a cero, y entre más se aleje de este valor mayor diferencia habrá entre los pronosticado y lo observado.

Para la segunda forma de validar los resultados se utilizó un puntaje de habilidad llamado *Ranked Probability Skill Score (RPSS)*, que indica la precisión relativa de un conjunto de pronósticos modelados, con respecto a un conjunto de pronósticos de referencia ($RPSS_{clim}$), la ecuación se expresa como:

$$RPSS = 1 - \frac{\langle RPS \rangle}{\langle RPS_{clim} \rangle} \quad (14)$$

La Ecuación (14) se puede interpretar como el porcentaje de mejora sobre el pronóstico de referencia (Wilks, 2006). Siguiendo con la metodología de Avilés (2016), un $RPSS$ igual a 1 significa una mejora del 100% sobre el pronóstico de referencia, pero si él $RPSS$ es igual o menor a cero, quiere decir que el pronóstico modelado no supera al de referencia.

$$z = \frac{x - \bar{a}}{s} \quad (15)$$

El pronóstico de referencia se calcula con la Ecuación 15, donde x es el valor de precipitación; \bar{a} es la media de precipitación para dado mes y s representa la desviación estándar para cada mes. A partir de esta ecuación se obtiene una variable estandarizada que posteriormente es reclasificada de la misma manera que con los valores del SPI de la Tabla 1, así se crea un pronóstico de referencia que indica los estados sequía esperables de acuerdo a la reclasificación.

Capítulo 2. Aplicación de las Cadenas de Markov de primer y segundo orden

2.1 Datos de la Estación Aeropuerto, Liberia

La Estación Meteorológica Aeropuerto Internacional Daniel Oduber Quirós ubicada en Liberia de Guanacaste, cuenta con una serie temporal de aproximadamente 83 años (1937-2020). En la Figura 4 se muestra el gráfico de precipitación promedio por mes en milímetros para la serie completa.

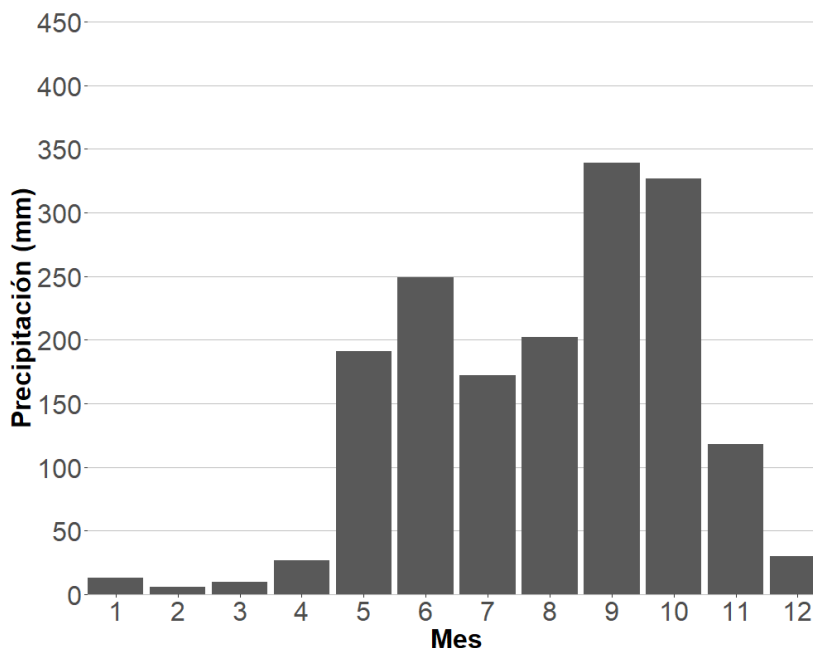


Figura 4. Precipitación promedio mensual en la Estación Aeropuerto (1937- junio, 2020).

En la Figura 4 se observa la marcada estacionalidad entre la época seca y la lluviosa. De diciembre a abril, se presentan meses secos con precipitaciones en promedio menores a 50 mm; y la estación lluviosa de mayo a noviembre. Entre abril y mayo hay una diferencia de aproximadamente 165 mm; mientras que de noviembre a diciembre el cambio de la precipitación promedio es de casi 90 mm. Con esto se puede decir que no hay una época de transición entre las estaciones, pues el aumento en las lluvias en mayo es casi inmediato, así como la disminución de la precipitación en noviembre.

En la Figura 5, la estacionalidad se puede apreciar con los valores promedio mensuales según cada año de la serie. Los valores con menor precipitación tienden al blanco, mientras que los de mayor precipitación pasan al color azul. Así, se logra apreciar que la

época seca, marcada entre los meses de diciembre y abril, es un periodo recurrente de menor, y casi nula, precipitación.

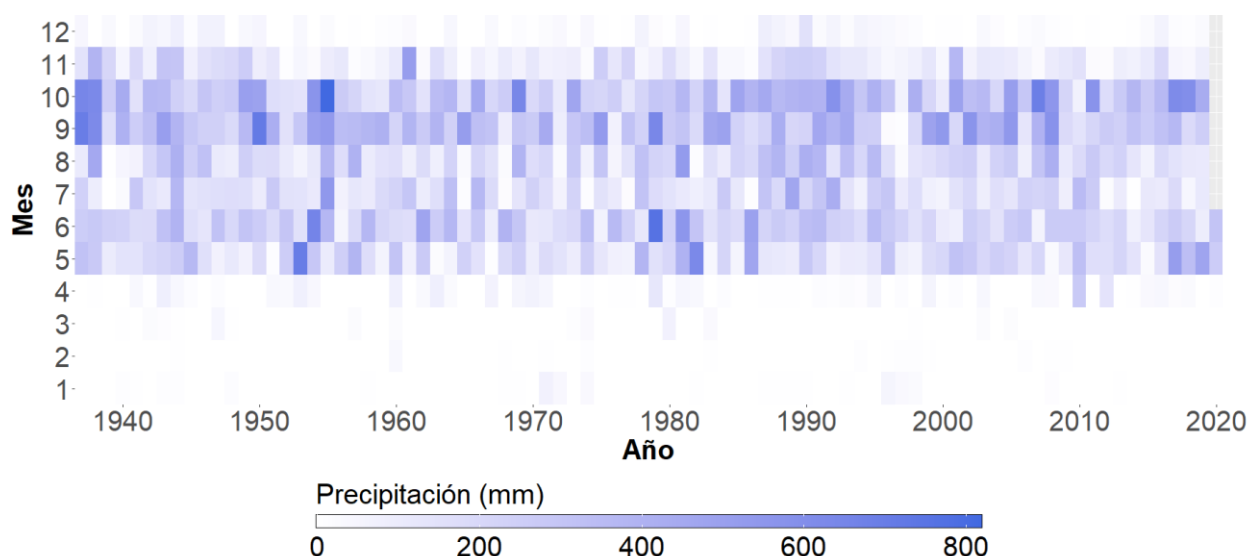


Figura 5. Mapa de calor que representa la precipitación promedio mensual (1937- junio, 2020) para la Estación Aeropuerto. El tono blanco representa menor precipitación, mientras que los azules una mayor cantidad de lluvia.

Es importante tener esto en cuenta para la construcción y análisis de los modelos, ya que este periodo seco no debe confundirse con sequía meteorológica, puesto que es una época característica del clima tropical seco de la zona de estudio. De igual forma se debe tener cuidado al momento de interpretar las probabilidades de transición entre abril y mayo, pues es un cambio brusco en la cantidad de precipitación.

2.1.1 Índice Estandarizado de Precipitación

A partir de los datos de precipitación se calculó el Índice Estandarizado de Precipitación (*SPI*) a escala mensual (Figura 6). Este índice ajusta la serie temporal a una distribución de probabilidad, que luego se transforma a una distribución normal. Así el valor cero representa el promedio del *SPI* para el lugar y periodo calculado (WMO, 2012). Los valores positivos indican una precipitación mayor al promedio, mientras que los negativos indican una precipitación menor al promedio.

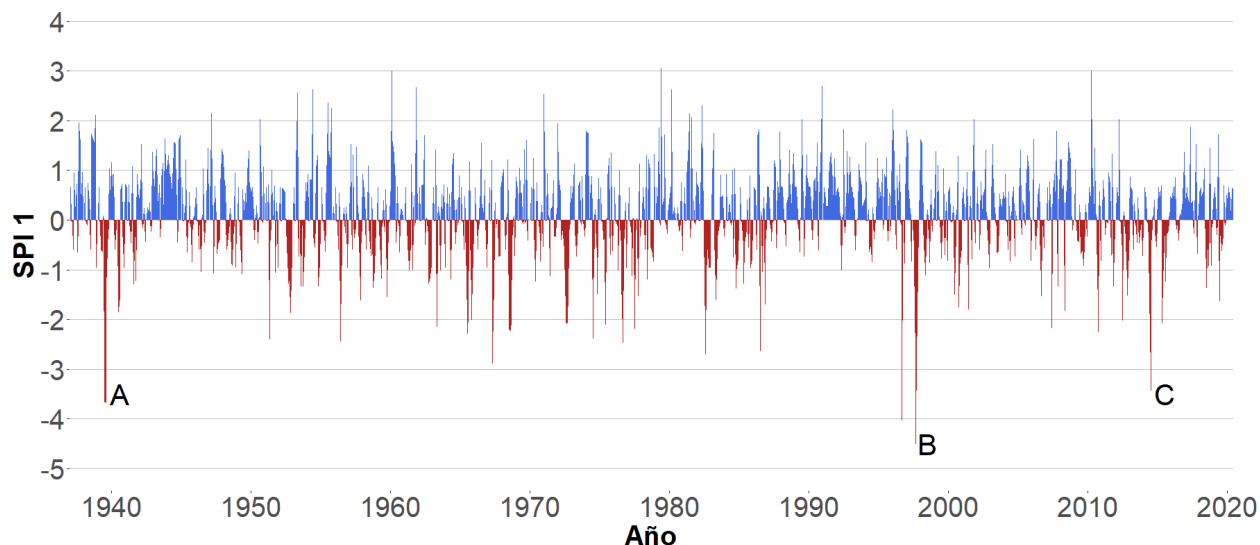


Figura 6. Índice estandarizado de precipitación para la Estación Aeropuerto (1937- junio, 2020). Con la letra A, se señala la sequía extrema de los meses julio-agosto de 1939. La letra B, marca las sequías extremas de setiembre de 1996 y 1997, mientras que la letra C, resalta la de julio del 2014.

Con el *SPI* a escala mensual se tiene un panorama más detallado del comportamiento de la precipitación. Tanto para los meses con lluvias por encima del promedio, como por debajo del mismo, es decir los meses con sequía. En la Figura 6 se muestra la serie temporal del *SPI* mensual (*SPI* 1), en la que se muestra como los eventos de sequía se han presentado históricamente. Estos eventos son los que están en colores rojos y con valores menores a -1.

En la Tabla 1 se presentó la clasificación de los eventos de sequías en cuatro categorías, siendo las sequías extremas con valores menores a -2. Si bien en la Figura 6 se presenta que ese estado se ha alcanzado en varias ocasiones, hay cinco eventos que sobresalen por sus valores extremos de sequedad. El primer y segundo evento de sequía que se registró se produjo en julio y agosto de 1939. Este evento alcanzó valores de *SPI* de -3.7 en los dos meses que se mantuvo.

Durante 57 años ningún mes presentó valores menores a -3, sino hasta setiembre de 1996 cuando hubo una sequía con un valor de -4; un año después en setiembre de 1997, se presentó el valor de *SPI* más negativo de la serie con -4.5. En julio del 2014, ocurrió una sequía con un valor *SPI* de -3.4. De esta manera, se tiene que en la serie temporal ha ocurrido cinco eventos de sequía extrema con valores *SPI* menores a -3, dos de esos

eventos fueron en meses consecutivos en 1939 y los únicos durante 57 años. Los otros tres eventos han ocurrido en 1996, 1997 y 2014.

2.1.2 Clasificación de los estados de la sequía

Se reclasificó los valores de *SPI* de acuerdo con la Tabla 1, para obtener un índice de sequía (Figura 7) que indica la severidad de la sequía para los meses en que se presentó algún evento. En la Figura 7, los estados de no sequía están en color blanco. La categoría de sequía moderada está en amarillo, el estado de sequía severa en anaranjado, mientras que los eventos de sequía extrema están en rojo oscuro.

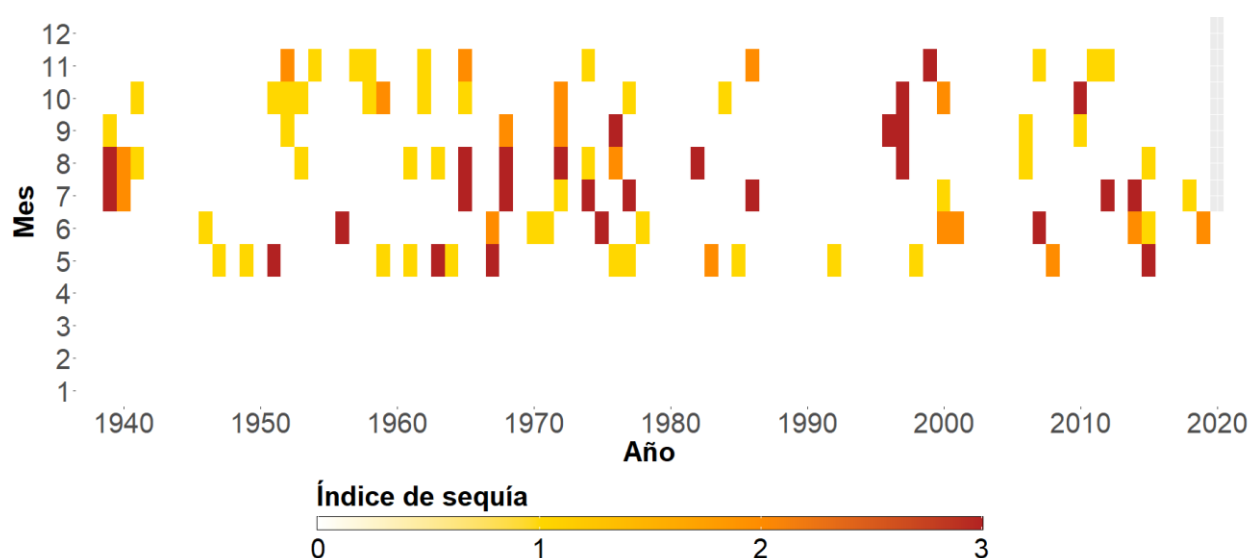


Figura 7. Mapa de calor que representa los estados de sequía por mes de toda la serie temporal para la Estación Aeropuerto.

En la Estación Aeropuerto, los eventos de sequía meteorológica han ocurrido históricamente entre mayo y noviembre, es decir, en la época lluviosa. Esto tiene sentido ya que en los meses secos la precipitación es escasa e incluso nula, esto significa que es imposible una disminución de precipitación con respecto al promedio de lluvia para los meses de la estación seca.

Popularmente se tiene la idea que las sequías en Guanacaste se deben a la influencia de El Niño, sin embargo, en los periodos 1982-1983, 1986-1987 y 1992-1993, cuando El Niño presentó una fuerte señal (Wolter y Timlin, 1998), no se refleja en la cantidad de meses con algún estado de sequía ya que entre 1982-1983 hubo tres meses no consecutivos con estado de sequía. Para el periodo 1986-1987, también solo tres meses

presentaron eventos, pero solo en 1986, al año siguiente no se registraron sequías. Entre 1992 y 1993, solo mayo del 1992 presentó sequía, esta fue moderada con un valor SPI de -1, es decir, su señal no fue tan fuerte.

No obstante, otros periodos clasificados con un ENOS fuerte, que ocurrieron en los años de 1972-1973, 1997-1998 (Wolter y Timlin, 1998), sí se observan en la serie temporal. Ya que, en 1972, hubo cuatro meses consecutivos con algún estado de sequía, esto entre julio y octubre; en 1973 no se presentó alguna sequía, pero para la época lluviosa de ese año, El Niño ya se había debilitado. Entre 1997 y 1998, se registró un El Niño bastante fuerte que sí se registró en la región, pues para 1998, desde agosto a octubre la categoría de sequía fue de extrema para esos tres meses.

Así, tenemos que se han registrado al menos seis ENOS con una señal muy fuerte, de estos, tres se reflejan con un aumento de eventos de sequía o con categorías más extremas. Pero, las otras tres no se manifestaron de una manera fuerte en la cuenca. Esto quiere decir que El Niño no es necesariamente la única causa de las sequías.

De los 1002 meses de la serie temporal, 92 estuvieron en alguna categoría de sequía, lo que representa al 9% de los datos (Tabla 2). La categoría de sequía moderada es la que se ha presentado con mayor frecuencia, seguida por la sequía extrema y por último la sequía severa. Sin embargo, estas últimas unidas representan el 52% de los meses con algún grado de sequía, mientras que la categoría de sequía modera representa el 48% de los meses con sequía.

Tabla 2. Frecuencia de los eventos de sequía en la Estación Aeropuerto (1937-2020).

Categoría	Cantidad de meses	Porcentaje de los datos
No sequía	910	0.91
Sequía moderada	44	0.04
Sequía severa	20	0.02
Sequía extrema	28	0.03
Total	1002	1

La frecuencia de la sequía con respecto al total de los datos es baja, no obstante, se muestra que es recurrente, ya que puede presentarse entre una y tres veces en un año, e incluso cuatro veces como en 1965 y 1972. Entre 1960 y finales de los setentas, hubo

una mayor actividad de eventos de sequías, mientras que la década de los noventas fue en la que se presentó menor cantidad de sequías.

El periodo en estudio, se dividió en dos secciones. La primera en el periodo comprendido desde 1937 hasta 1999 (Tabla 3), esto para la calibración de los modelos; lo cual corresponde a, aproximadamente, el 75% de los datos, es decir 756 meses. Mientras que la segunda sección (Tabla 4) son los años restantes (2000-junio 2020), este es el periodo de validación de los modelos, lo que representa al 25% de los datos (246 meses).

Como se observa en las Tablas 3 y 4, los valores porcentuales de ambos periodos son similares entre ellos mismos y también con el periodo total de la serie (Tabla 2). Esto quiere decir que los periodos seleccionados representan el comportamiento y frecuencia de los eventos de sequía en la Estación Aeropuerto.

Tabla 3. Frecuencia de los eventos de sequía para el periodo de calibración (1936-1999), en la Estación Aeropuerto.

Categoría	Cantidad de meses	% de los datos
No sequía	686	0.91
Sequía moderada	35	0.04
Sequía severa	12	0.02
Sequía extrema	23	0.03
Total	756	1

Tabla 4. Frecuencia de los eventos de sequía para el periodo de validación (2000-2020), en la Estación Aeropuerto.

Categoría	Cantidad de meses	% de los datos
No sequía	224	0.91
Sequía moderada	9	0.04
Sequía severa	8	0.03
Sequía extrema	5	0.02
Total	246	1

2.2 Cadenas de Markov de primer orden

Una vez establecido el periodo de validación, se procedió a aplicar la Ecuación 2, con la cual se calcula las frecuencias relativas de las transiciones entre los meses. Así se obtuvo doce matrices de transición, una para cada mes. Estas matrices establecen las

probabilidades de transición para el mes actual dado el mes anterior (Ecuación 1). Los resultados para la época lluviosa se muestran en la Tabla 5, mientras que en el Anexo 1, están los resultados para los doce meses.

Tabla 5. Probabilidades de transición de los estados del modelo MC1, en la época lluviosa para la Estación Aeropuerto.

Estado mes anterior i	Estado mes actual j	Probabilidades del mes actual j						
		MAY	JUN	JUL	AGO	SET	OCT	NOV
0	0	0.78	0.88	0.86	0.87	0.96	0.84	0.90
	1	0.16	0.08	0.02	0.07	0.02	0.14	0.06
	2	0.02	0	0.02	0.02	0	0.02	0.02
	3	0.05	0.04	0.11	0.04	0.02	0	0.02
1	0	1	1	1	0	1	0.50	0.56
	1	0	0	0	0	0	0.50	0.22
	2	0	0	0	0	0	0	0.22
	3	0	0	0	1	0	0	0
2	0	1	1	1	0	0.50	0.50	1
	1	0	0	0	0	0	0	0
	2	0	0	0	1	0	0.50	0
	3	0	0	0	0	0.50	0	0
3	0	1	0.67	1	0.33	0.33	0.67	1
	1	0	0	0	0.17	0.17	0	0
	2	0	0.33	0	0	0.33	0	0
	3	0	0	0	0.50	0.17	0.33	0

Se decide mostrar solo los meses de lluvia, ya que es en este periodo cuando históricamente han ocurrido las sequías meteorológicas. Además, los resultados para la época seca establecen que hay una probabilidad de transición a la categoría de No sequía del 100% para todos los meses secos. Es decir, según este modelo, no hay probabilidad de sequía para la época seca.

Para interpretar la Tabla 5 se debe seleccionar el mes actual j y leer las probabilidades de ocurrencia que tienen los estados para ese mes, con respecto al estado del mes anterior i . Por ejemplo, si abril presenta la categoría de No sequía (estado 0), mayo tiene una probabilidad del 78% de mantenerse en esa misma categoría; un 16% de pasar al

estado 1 (Sequía Moderada); un 2% de que se presente el estado 2 (Sequía Severa); y hay un 5% de probabilidades de un evento de Sequía Extrema (estado 3).

Siguiendo con el ejemplo de mayo, se nota que hay un 100% de probabilidades de No sequía cuando procede de un abril de estados 1, 2 y 3; esto se debe a que históricamente abril nunca ha tenido una sequía meteorológica, por lo tanto, es en mayo cuando se empiezan las transiciones a alguna categoría de sequía. Así es como se observa que cuando mayo presenta un estado 3, junio tiene una probabilidad del 67% de cambiar a un estado 0, y un 33% de probabilidades de bajar al estado 2. Agosto, setiembre y octubre son los meses que presentan mayores probabilidades de transición al estado 3, es decir, es en ese trimestre que podrían ocurrir las sequías extremas, además de junio.

Es importante mencionar la situación que se da cuando setiembre procede de un agosto estado 2, ya que hay un 50% de probabilidad de que no ocurra sequía, pero también hay un 50% de posibilidades de que ocurra una sequía extrema. Una situación similar ocurre para octubre cuando setiembre presenta el estado 1 y 2, pues hay posibilidades del 50% de que no ocurra sequía y del otro 50% de que se presente el estado 1 o 2, respectivamente. Esos casos presentan un inconveniente para establecer con alguna seguridad la ocurrencia o no de la sequía, por lo que la decisión dependería de un análisis de la situación de la cuenca y de otras variables como la influencia de ENOS, o la fuerza de los vientos.

2.3 Cadenas de Markov de segundo orden

Con la Ecuación 4 se calculó las frecuencias de las transiciones de los estados de sequía para las Cadenas de Markov de segundo orden. Estas transiciones toman en cuenta los estados de dos meses anteriores al mes que se quiere pronosticar. Debido a esto las posibles combinaciones aumentan, de 16 posibilidades que se tienen en el modelo MC1 a 64 en el MC2 (Tabla 6).

Tabla 6. Probabilidades de transición de los estados del modelo MC2, en la época lluviosa para la Estación Aeropuerto.

Estados meses anteriores i	Estado mes actual j	Probabilidades del mes actual j						
		MAY	JUN	JUL	AGO	SET	OCT	NOV
0-0	0	0.77	0.88	0.84	0.86	0.96	0.88	0.90
	1	0.16	0.08	0.02	0.08	0.02	0.10	0.06
	2	0.02	0	0.02	0.02	0	0.02	0.02
	3	0.05	0.04	0.12	0.04	0.02	0	0.02
0-1	0	-	1	1	0	1	0	0.62
	1	-	0	0	0	0	1	0.25
	2	-	0	0	0	0	0	0.13
	3	-	0	0	1	0	0	0
0-2	0	-	1	-	0	0	-	1
	1	-	0	-	0	0	-	0
	2	-	0	-	1	0	-	0
	3	-	0	-	0	1	-	0
0-3	0	-	0.67	1	0.33	0.50	1	-
	1	-	0	0	0.17	0	0	-
	2	-	0.33	0	0	0	0	-
	3	-	0	0	0.50	0.50	0	-
1-0	0	-	-	0.90	1	-	0.60	1
	1	-	-	0	0	-	0.40	0
	2	-	-	0	0	-	0	0
	3	-	-	0.10	0	-	0	0
1-1	0	-	-	-	-	-	-	0
	1	-	-	-	-	-	-	0
	2	-	-	-	-	-	-	1
	3	-	-	-	-	-	-	0
1-2	0	-	-	-	-	-	-	-
	1	-	-	-	-	-	-	-

	2	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-
1-3	0	-	-	-	-	0	-	-
	1	-	-	-	-	0	-	-
	2	-	-	-	-	1	-	-
	3	-	-	-	-	0	-	-
2-0	0	-	-	1	1	-	1	1
	1	-	-	0	0	-	0	0
	2	-	-	0	0	-	0	0
	3	-	-	0	0	-	0	0
2-1	0	-	-	-	-	-	-	-
	1	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-
2-2	0	-	-	-	-	1	-	1
	1	-	-	-	-	0	-	0
	2	-	-	-	-	0	-	0
	3	-	-	-	-	0	-	0
2-3	0	-	-	-	-	-	1	-
	1	-	-	-	-	-	0	-
	2	-	-	-	-	-	0	-
	3	-	-	-	-	-	0	-
3-0	0	-	-	1	1	1	0.50	1
	1	-	-	0	0	0	0.50	0
	2	-	-	0	0	0	0	0
	3	-	-	0	0	0	0	0
3-1	0	-	-	-	-	1	1	-
	1	-	-	-	-	0	0	-
	2	-	-	-	-	0	0	-
	3	-	-	-	-	0	0	-

	0	-	-	1	-	-	0.50	-
3-2	1	-	-	0	-	-	0	-
	2	-	-	0	-	-	0.50	-
	3	-	-	0	-	-	0	-
	0	-	-	-	-	0.34	0	1
3-3	1	-	-	-	-	0.33	0	0
	2	-	-	-	-	0.33	0	0
	3	-	-	-	-	0	1	0
	0	-	-	-	-	0	1	0

En la Tabla 6, se presenta las probabilidades de las posibles combinaciones de transición de estados para los meses de la estación lluviosa. Históricamente ha habido combinaciones que no se han registrado; la cuales se presentan en la tabla con un guion para una mejor visualización de las combinaciones que sí han ocurrido y que, por lo tanto, tienen una probabilidad calculable. Sin embargo, el hecho de que algunas combinaciones no se hayan observado, no significa que no puedan ocurrir. Por eso lo correcto sería asignarles alguna probabilidad a todas; en este caso se decidió distribuir un 25% entre los cuatro estados, ya que la suma de las probabilidades debe ser igual a 100%. La tabla con esos valores asignados, además de la inclusión de los meses secos, se presenta en el Anexo 2.

La interpretación de la Tabla 6 es similar a la del modelo MC1, con la diferencia que el modelo MC2 toma en cuenta dos estados previos. De esta forma se obtuvo que, mayo tiene una probabilidad de 16% de presentar sequía moderada (estado 1), cuando marzo y abril estuvieron en el estado 0. En otro ejemplo, junio tiene una probabilidad del 33% de estar en el estado 2, cuando abril fue estado 0 y mayo estado 3. Se observa que algunas matrices presentan una probabilidad del 100% de que se presente una sola transición, esto se debe a que ese determinado cambio de estado ocurrió solo una vez en el periodo de calibración; sin embargo, este resultado no debe ser confiable, pues podría ser que en un periodo anterior a 1937, sí se presentaron diversos cambios; o bien podrían presentarse en el futuro.

También es importante anotar que el problema que muestra el MC1, en donde un mes tiene probabilidades del 50% de que se presenten dos estados opuestos, sigue ocurriendo en el modelo MC2. Como se observa en la Tabla 6, esta situación se presenta en los dos modelos en los meses de setiembre y octubre. Lo que quiere decir que aun considerando dos meses previos los modelos de Cadenas de Markov tienen el inconveniente de otorgar la misma probabilidad para dos estados opuestos, lo que representan un problema al momento de decidir si habrá o no algún estado de sequía.

Capítulo 3. Modelos probabilísticos basados en Redes Bayesianas

3.1 Datos de entrada para los modelos de Redes Bayesianas

A diferencia de los modelos de Cadenas de Markov, que utilizaron una reclasificación del Índice Estandarizado de Precipitación (*SPI*) para establecer los cambios de los estados, las Redes Bayesianas necesitan la distribución de la probabilidad de las categorías de sequía. Para conocer esto, se le calculó al *SPI* de la Estación Aeropuerto, la función de la distribución acumulada (*CDF*). La Figura 8, muestra el grafico de la distribución acumulada para los datos de la Estación Aeropuerto.

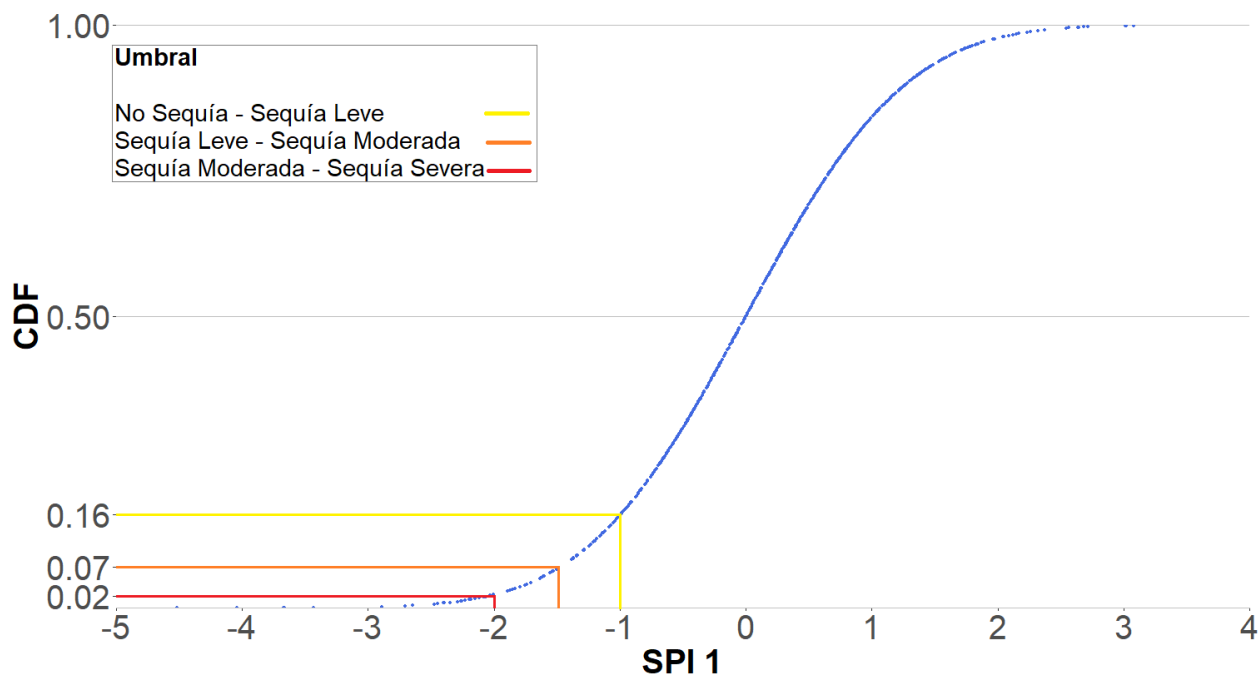


Figura 8. Distribución Acumulada de la serie temporal del Índice Estandarizado de Precipitación, para la Estación Aeropuerto.

Con la *CDF* se establecen los umbrales de los estados de sequía, pues como se aprecia en la Figura 8, se puede conocer la posición de los datos con respecto a su distribución acumulada. Así, sabemos que los valores de *SPI* menores a -1, tienen una probabilidad menor a 0.16 de presentarse, es decir existe un 16% de probabilidades de que se presente algún estado de sequía.

3.2 Redes Bayesianas de primer orden

3.2.1 Ajuste de cópulas de dos dimensiones

De acuerdo al Teorema de Sklar, las cópulas son funciones que se caracterizan por tener las marginales uniformes. Para lograr esto, se calcularon pseudo-observaciones a partir de los valores de la *CDF* de cada mes en el periodo de calibración. Una vez realizado este paso se procedió a realizar el ajuste de las cópulas, cuyo resultado se muestra en el Anexo 3, mientras que la Tabla 11 es el resumen con las cópulas que cumplieron con los criterios de selección por mes y sus respectivos parámetros.

Tabla 7. Resultado del ajuste mensual de cópulas de dos dimensiones.

Mes	Copula	Parámetro
Enero	Frank	6.776
Febrero	Frank	4.401
Marzo	Normal	0.058
Abril	T	0.016
Mayo	T	0.226
Junio	Clayton	-0.049
Julio	Clayton	0.674
Agosto	Clayton	0.381
Setiembre	Normal	0.201
Octubre	Clayton	0.372
Noviembre	T	0.399
Diciembre	Frank	3.917

Como se evidencia en la Tabla 7, las distribuciones mensuales se ajustan a diferentes tipos de cópulas y aun cuando algunos meses comparten el mismo tipo, el parámetro es distinto. También se puede observar que de manera general en los meses de época seca la copula que mejor se ajusta es la de tipo Frank, mientras que en la época lluviosa es la copula Clayton la que predomina.

Esto quiere decir que en los meses de diciembre a febrero los datos se distribuyen de una manera amplia y uniforme con una pendiente positiva, por lo que su correlación es débil y no tendrían cola en ninguno de los extremos. En el caso de abril, mayo y noviembre, mostraron un mejor ajuste con la copula T, lo que implica que los datos de

estos meses tampoco poseen una cola y su distribución es mucho más amplia y sin una pendiente visible.

Por otro lado, las cópulas Clayton, que se ajustaron en los meses de junio, julio, agosto y octubre, se caracterizan por tener una fuerte densidad y concentración de los datos en la cola izquierda. Es decir, existe una relación con los eventos negativos, en este caso los estados de sequía, por lo que se podría esperar que cuando haya eventos de sequía en un mes, al siguiente mes las probabilidades de que ocurra uno se incrementan (Mendoza, 2014).

3.2.2 Probabilidades del modelo de Redes Bayesianas de primer orden

Una vez seleccionadas las cópulas que mejor se ajustan por mes, se procedió a aplicar la Ecuación 9, en la cual se realiza la unión de los valores de las cópulas con las distribuciones observadas mensualmente. Este procedimiento se realizó con los datos del periodo de validación y el resultado de las probabilidades por mes en la época lluviosa, se muestra en la Figura 9, mientras que en el Anexo 4 se añaden los meses de la época seca.

Como se observa en la Figura 9, en todos los meses la categoría de No sequía es la que presenta la mayor probabilidad de ocurrencia, mientras que los estados de sequía, tienen una probabilidad mucho menor, usualmente por debajo del 20%. Esto sucede porque en las distribuciones observadas en el periodo de calibración tienden a estar por arriba del umbral de No sequía (Figura 8). Sin embargo, la categoría No sequía muestra una disminución en sus probabilidades de ocurrencia cuando se observa algún evento de sequía. Mientras que las probabilidades de que se presente algún estado de sequía aumentan levemente.

Por ejemplo, en el mes de mayo el promedio de las probabilidades de la categoría No sequía es de 77%, pero en el 2008, esa probabilidad bajó a 65%, y en el 2015 la probabilidad fue del 60%; en esos dos años ocurrieron eventos de sequía meteorológica, y como se puede observar, las probabilidades de que se presente alguna sequía aumentaron levemente con respecto a los otros años.

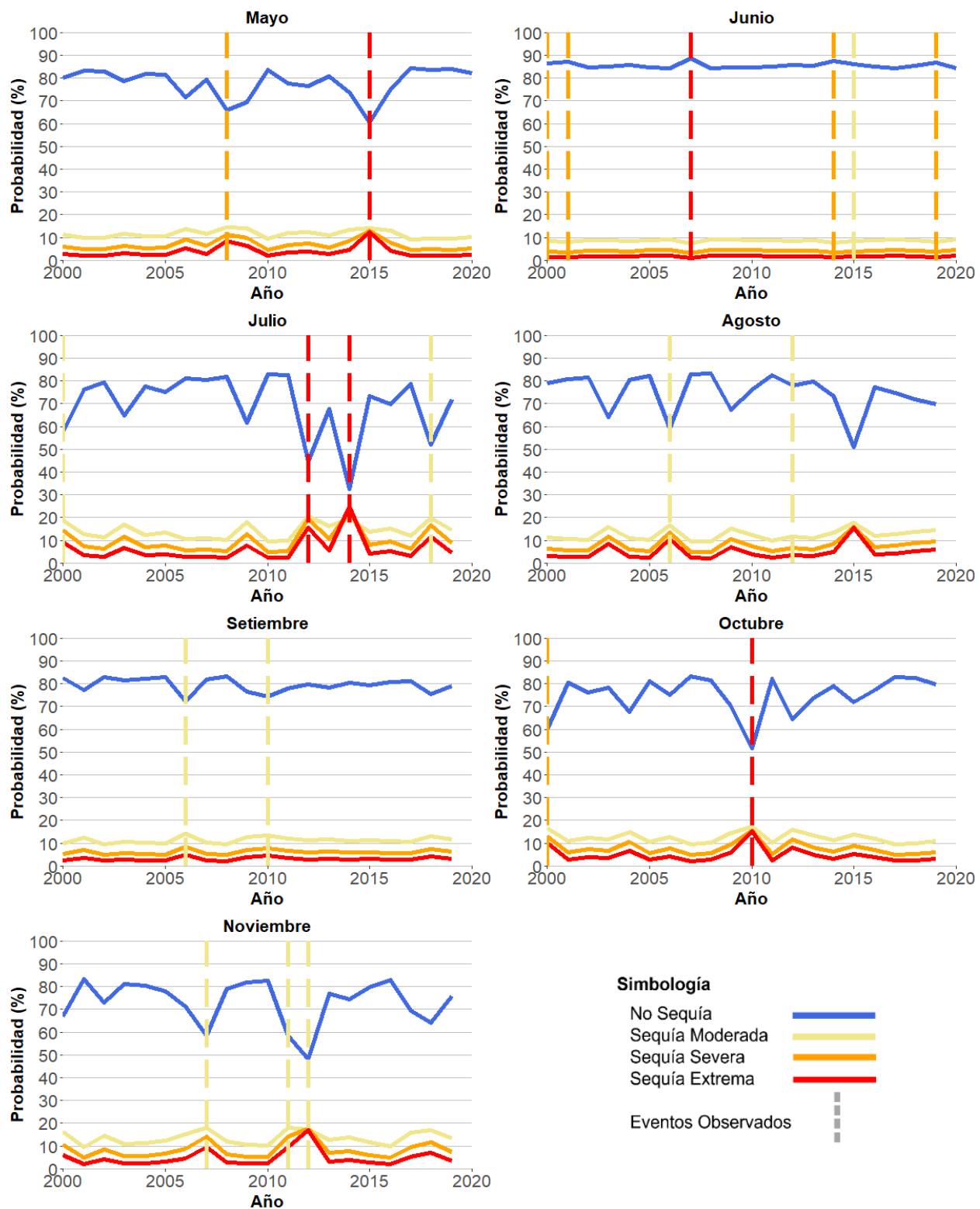


Figura 9. Probabilidades por mes del modelo de Redes Bayesianas de primer orden, en la época lluviosa del 2000 al 2020.

Este comportamiento también se presenta en otros meses, siendo muy notorio en julio, ya que en ese mes se presentaron dos eventos de sequía (2012 y 2014), y se observa como las probabilidades de sequía aumentaron a casi un 25% de probabilidad, mientras que las de categoría de No sequía bajaron hasta el 35% de probabilidad. El único mes que no presenta este comportamiento es junio, esto se debe a que la correlación de las pseudo-observaciones entre mayo y junio es débil (Tabla 8), por lo que se genera una mayor incertidumbre en la transición de esos meses.

Tabla 8. Correlaciones de las pseudo-observaciones para cada mes en la Estación Aeropuerto.

Método	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov	Dic
Pearson	-0.09	-0.78	0.75	0.80	0.16	0.02	0.20	0.66	0.79	0.49	0.22	0.13
Kendall	-0.03	-0.59	0.62	0.65	0.11	0.03	0.12	0.47	0.61	0.35	0.14	0.09
Spearman	-0.09	-0.78	0.75	0.80	0.16	0.02	0.20	0.66	0.79	0.49	0.22	0.13

3.3 Redes Bayesianas de segundo orden

3.3.1 Ajuste de cópulas de tres dimensiones

Al igual que el modelo de primer orden, las Redes Bayesianas de segundo orden necesitan que los parámetros de las cópulas se ajusten, en este caso en tres dimensiones, ya que se toma en cuenta la distribución de dos meses anteriores. Es por esto que la selección de cópulas para este modelo es diferente (Tabla 9) e incluye tres parámetros.

Tabla 9. Resultado del ajuste de cópulas de tres dimensiones.

Mes	Copula	Parámetro 1	Parámetro 2	Parámetro 3
Enero	Normal	0.762	0.640	0.609
Febrero	T	0.609	0.297	0.058
Marzo	Normal	0.058	0.036	0.016
Abril	T	0.016	0.176	0.226
Mayo	T	0.226	-0.222	-0.040
Junio	T	-0.040	0.141	0.386
Julio	Clayton	0.370	0.370	0.370
Agosto	Normal	0.249	0.474	0.201
Setiembre	T	0.201	0.085	0.244
Octubre	T	0.244	0.262	0.399
Noviembre	T	0.399	0.115	0.565
Diciembre	Normal	0.565	0.484	0.779

En este modelo las cópulas elípticas (Normal y T) son las de mejor ajuste, excepto para junio. Esto significa que cuando se agrupan dos meses, las distribuciones tienen forma de elipse. Como se observa en el Anexo 5, la diferencia que había en el valor estadístico de las cópulas Normales y las T era muy pequeño, y por esto la selección se realizó considerando más de cinco decimales.

3.3.2 Probabilidades del modelo de Redes Bayesianas de segundo orden

Al igual que el modelo de primer orden, se aplicó la Ecuación 9, para conocer las probabilidades de las categorías, en este caso las de segundo orden. Las probabilidades de los meses lluviosos se muestran en la Figura 10, mientras que en el Anexo 6 adjuntan los meses de la época seca.

Se observa que la categoría de No sequía es la que mayor probabilidad tiene en todos los meses, con valores promedio del 80% de probabilidad. Mientras que los estados de sequía presentan probabilidades menores al 10%. En el caso de este modelo, las probabilidades se mantienen estables a través del periodo de validación, esto se debe a que las cópulas que se ajustan a este modelo son las elípticas, las cuales se caracterizan por ser una generalización de las distribuciones normales.

Julio, es el único mes que no cumple con lo anterior, ya que las probabilidades son más inconsistentes durante el periodo de validación. Esto sucede porque la copula Clayton que se ajustó a este mes se caracteriza por tener una concentración de datos en la cola izquierda, lo que ocasiona que haya una relación con los eventos negativos, es decir con la sequía. Por eso este mes demuestra que las probabilidades de los estados de sequía aumentan cuando se observó un evento de sequía, mientras que la probabilidad de No Sequía disminuye con respecto a su comportamiento histórico.

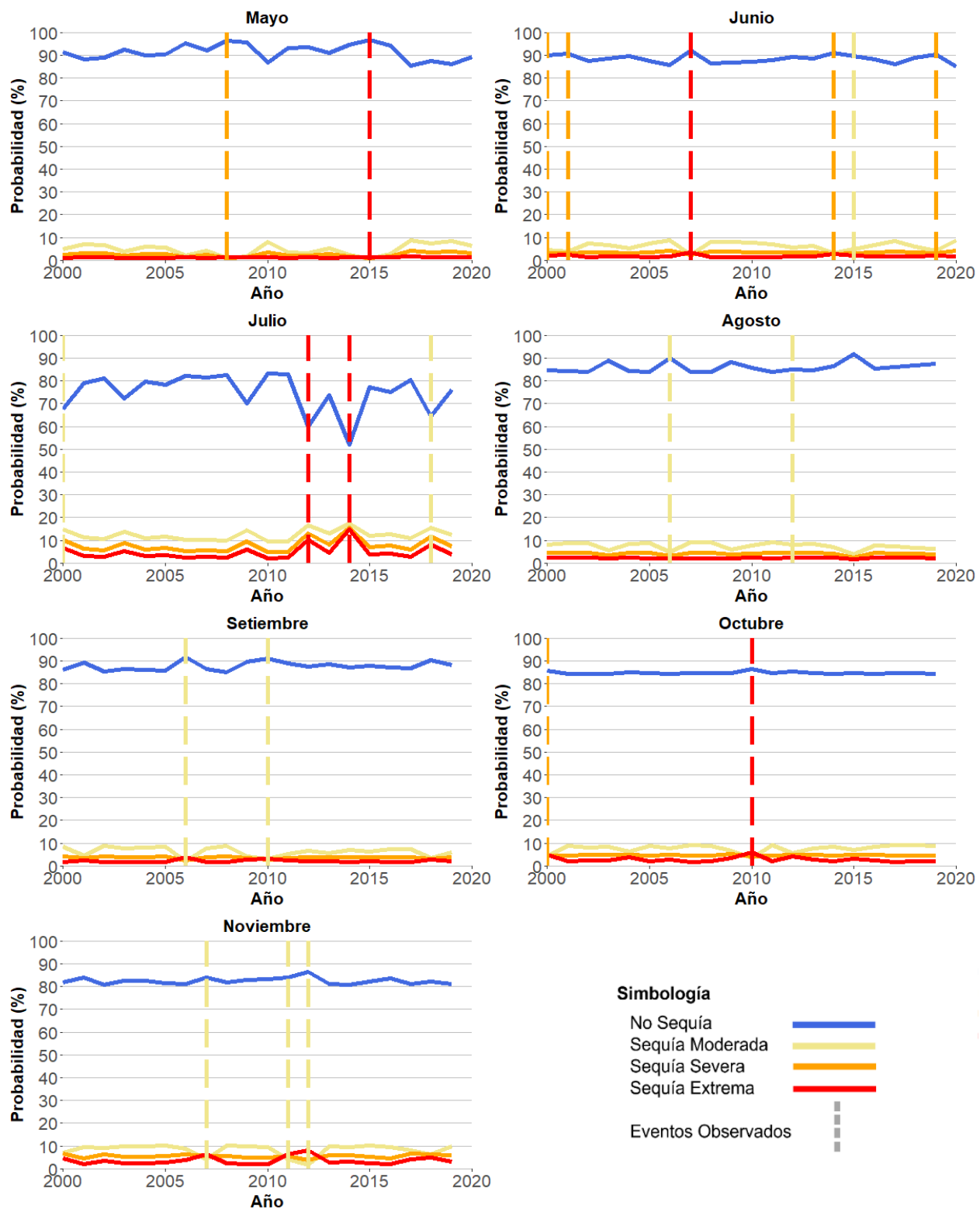


Figura 10. Probabilidades por mes del modelo de Redes Bayesianas de segundo orden, en la época lluviosa del 2000 al 2020.

Capítulo 4. Evaluación de los modelos probabilísticos

Con el fin de conocer el rendimiento de los modelos se aplicó dos tipos de evaluaciones, la primera es el *RPS*, el cual se aplica con la Formula 13; mientras que la segunda se llama *RPSS* y utiliza la Formula 14. Para ambos casos se realizaron dos evaluaciones, una que toma en cuenta todos los estados y otra que solo considera los estados de sequía. Esto porque el objetivo de esta tesis es el pronóstico de sequía, por lo que es más importante conocer el rendimiento de los modelos solo con los estados de sequía.

4.1 *Ranked Probability Score* para todas las categorías

El primer método de validación utilizado fue el *Ranked Probability Score (RPS)*. Este método evalúa la diferencia que hay entre los datos pronosticados y los datos observados. En este caso un valor $RPS = 0$ significa que no hay diferencia entre el pronóstico y lo observado, es decir, el modelo lograría un pronóstico perfecto. Si el resultado del *RPS* es diferente a 0, implica que existe una diferencia entre los resultados pronosticados y los datos observados. En ese sentido, cuanto más cerca se encuentre del valor 0, mejor rendimiento tiene el modelo.

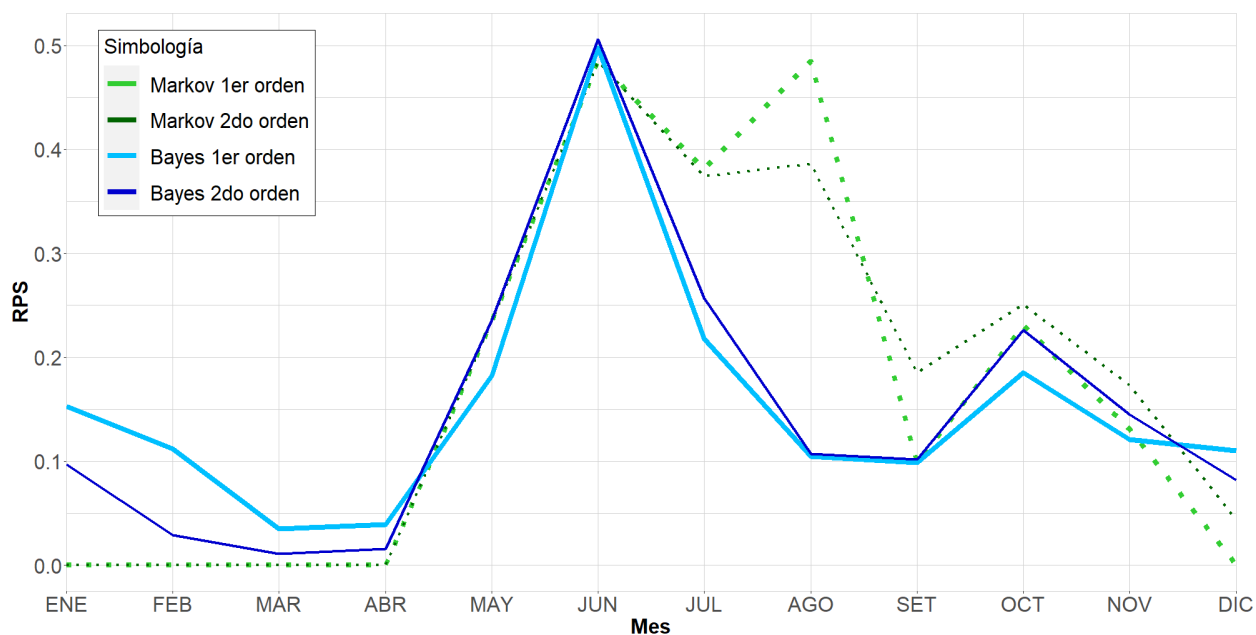


Figura 11. Ranked Probability Score (RPS) de los modelos de Cadenas de Markov y Redes Bayesianas de primer y segundo orden, para las cuatro categorías.

En la Figura 11 se muestra el resultado de la validación RPS para los modelos de Cadenas de Markov y Redes Bayesianas, tanto de primer y segundo orden. En esta

validación se toman en cuenta los cuatro estados de sequía, por eso se presentan los resultados para todos los meses, pues se incluye la categoría de No Sequía. Se observa que de diciembre a abril los modelos de Cadenas de Markov presentan un $RPS = 0$, esto quiere decir que se logra pronosticar perfectamente las observaciones *in situ*.

Sin embargo, esto se debe a que los modelos de Markov pronostican que en los meses secos hay una probabilidad del 100% de que se presente la categoría No Sequía, lo cual es esperable ya que, como se presentó en la Figura 7, históricamente los meses de estación seca no presentan sequía meteorológica. En el caso de los modelos de Redes Bayesianas se muestra que hay una diferencia del pronóstico con lo observado para esos mismos meses, esto se debe a que los modelos bayesianos asignan probabilidades de ocurrencia a las cuatro categorías de sequía.

Para los meses de estación lluviosa se observan diferencias entre los pronósticos y las observaciones, lo cual es esperable ya que las condiciones son más cambiantes con respecto a la estación seca. En los dos primeros meses de la época lluviosa los modelos presentan un comportamiento similar, sin embargo, de julio a noviembre las Redes Bayesianas obtienen un rendimiento más cercano a cero, lo que significa que tiene menos diferencias con las categorías observadas.

4.2 Ranked Probability Score únicamente para los estados de sequía

Como el objetivo de esta tesis se enfoca en la capacidad de pronosticar la sequía meteorológica, se decidió realizar una validación que incluyera únicamente los eventos de sequía. Es decir, en este análisis se dejó por fuera los meses que presentaron la categoría de No Sequía, así el RPS calculado para la Figura 12, muestra las diferencias que entre el pronóstico y las sequías observadas.

Como se observa los modelos de Cadenas de Markov tienen un comportamiento muy similar, excepto en julio, donde las de primer orden muestran una mayor diferencia. En octubre los pronósticos de ambas Cadenas tienen una mayor incertidumbre, pues existe una mayor diferencia con los eventos observados, pero logran un mejor rendimiento para el mes de noviembre.

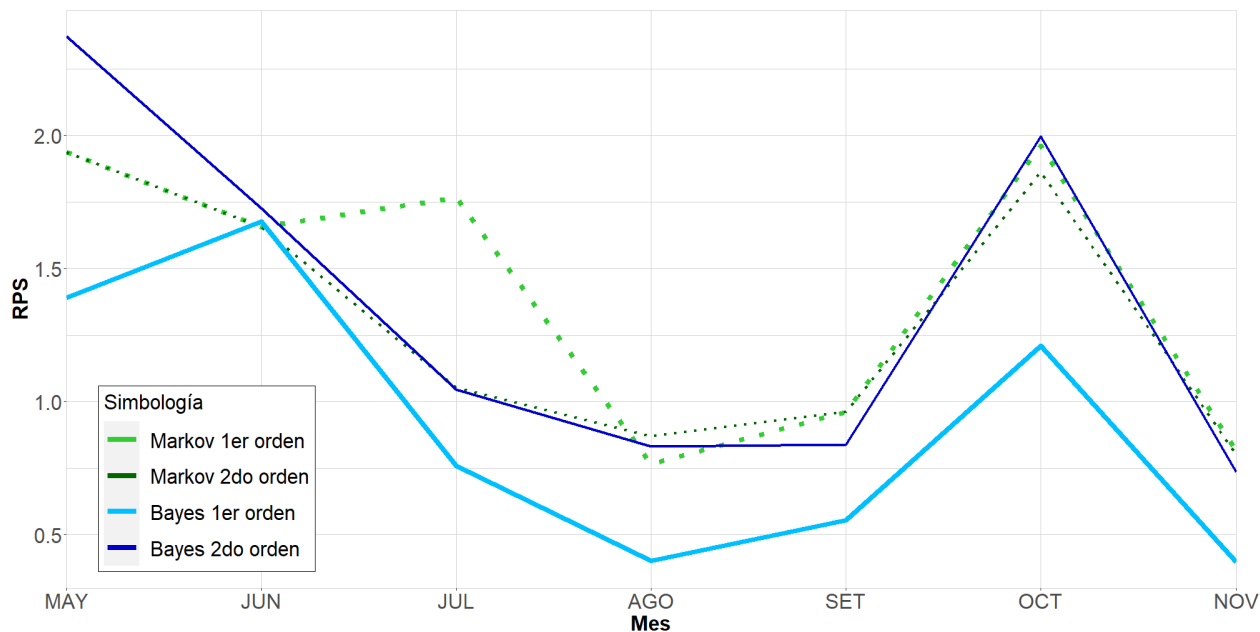


Figura 12. Ranked Probability Score (RPS) de los modelos de Cadenas de Markov y Redes Bayesianas de primer y segundo orden, únicamente para los estados de sequía.

En cuanto a las Redes Bayesianas, el modelo de segundo orden presenta un rendimiento muy similar a los de las Cadenas de Markov, donde en el mes de mayo hay una mayor incertidumbre, pero en los siguientes meses mejora su pronóstico y, al igual que las Cadenas de Markov tiene problemas en el mes de octubre. Por otro lado, las Redes Bayesianas de primer orden son el modelo de mejor rendimiento en estas condiciones y lo mantiene durante toda la época lluviosa.

Si bien es cierto sigue el patrón de los demás modelos, las Redes Bayesianas de primer orden presentan el mejor rendimiento, ya que el pronóstico tiene menor diferencia con las observaciones. Los mejores resultados se obtuvieron en agosto y noviembre, esto es importante, pues si se logra pronosticar sequías meteorológicas en el último mes de la época lluviosa, se podría utilizar como una alerta temprana de otros tipos de sequía, ya que la época seca se extendería.

4.3 Ranked Probability Skill Score para todas las categorías

La segunda forma de validación utilizada fue la *Ranked Probability Skill Score (RPSS)*, este método indica la precisión relativa de los pronósticos modelados con respecto a un pronóstico de referencia, así se compara los resultados con una referencia esperable del comportamiento de los estados de sequía. Si el *RPSS* es igual a 1 quiere decir que los

modelos mejoran en un 100% los pronósticos de referencia. Si el *RPSS* es igual o menor a 0, significa que no hay una mejora. Al igual que en el primero método de evaluación se calcularon dos *RPSS*, el primero incluye todas las categorías (Figura 13), mientras que la segunda muestra la validación para los meses que presentaron un evento de sequía (Figura 14).

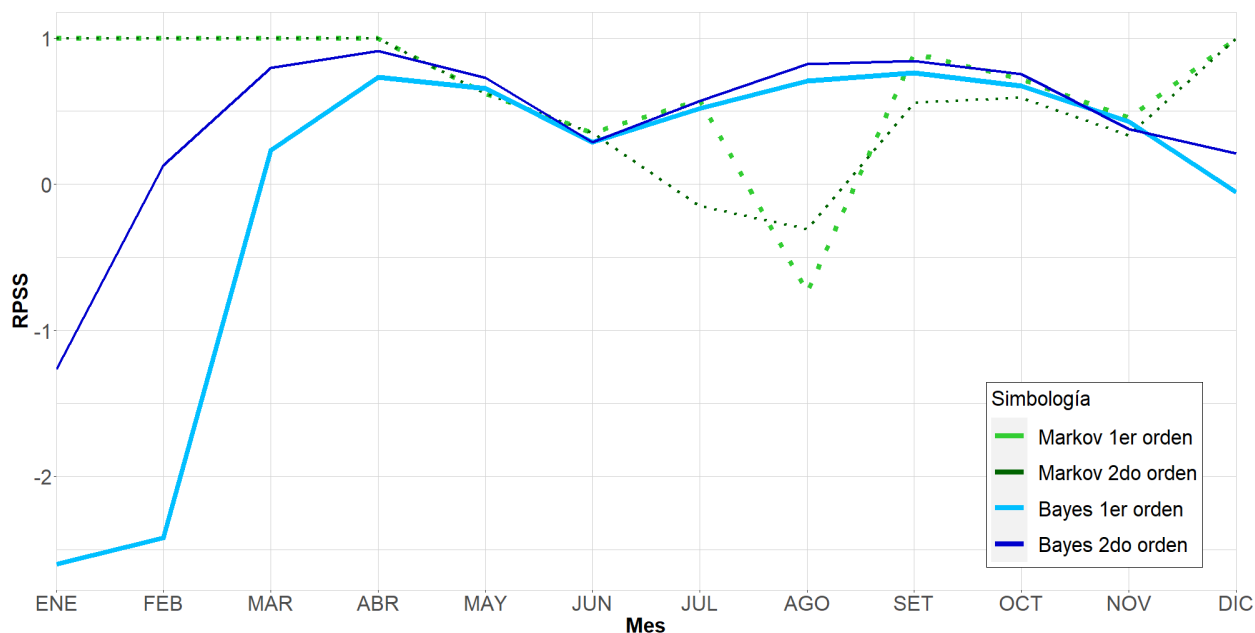


Figura 13. Ranked Probability Skill Score (RPSS) de los modelos de Cadenas de Markov y Redes Bayesianas de primer y segundo orden, para las cuatro categorías.

Como se puede observar, los modelos de Cadenas de Markov obtienen resultados perfectos en los meses de época seca, esto porque pronostican eventos de No Sequía, sin embargo, esto es lo esperable para esos meses. Por otro lado, las Redes Bayesianas muestran un rendimiento pobre en estos meses, esto se debe a que los cuatro estados tienen probabilidad de ocurrencia, pero para el pronóstico de referencia solo existe una opción, que es el estado de No Sequía, por lo que el pronóstico de los modelos no mejora el pronóstico de referencia.

En los meses de abril, mayo y junio, los cuatro modelos tienen un rendimiento similar, en julio Markov de segundo orden es el modelo que baja su rendimiento comparado con los demás modelos. En agosto son los modelos de Redes Bayesianas que obtienen el mejor rendimiento, mientras que las Cadenas de Markov no logran mejorar el pronóstico de referencia. Para finales de la época lluviosa, los dos modelos de Bayes y Markov de

primer orden tienen valores *RPSS* aproximados a 0.7, lo que quiere decir que mejoran el pronóstico de referencia en un 70%.

4.4 *Ranked Probability Skill Score* únicamente para los estados de sequía

Por último, se realizó la validación *RPSS* de los cuatro modelos únicamente teniendo en consideración los meses que presentaron estados de sequía y el resultado se muestra en la Figura 14. Se puede observar que de manera general los modelos no obtienen una mejora con respecto al pronóstico de referencia, ya que el valor más alto fue de un *RPSS* = 0.54. Además, los modelos de Cadenas de Markov de primer y segundo orden y Redes Bayesianas de segundo orden obtuvieron valores que no sobrepasaron un *RPSS* = 0.06.

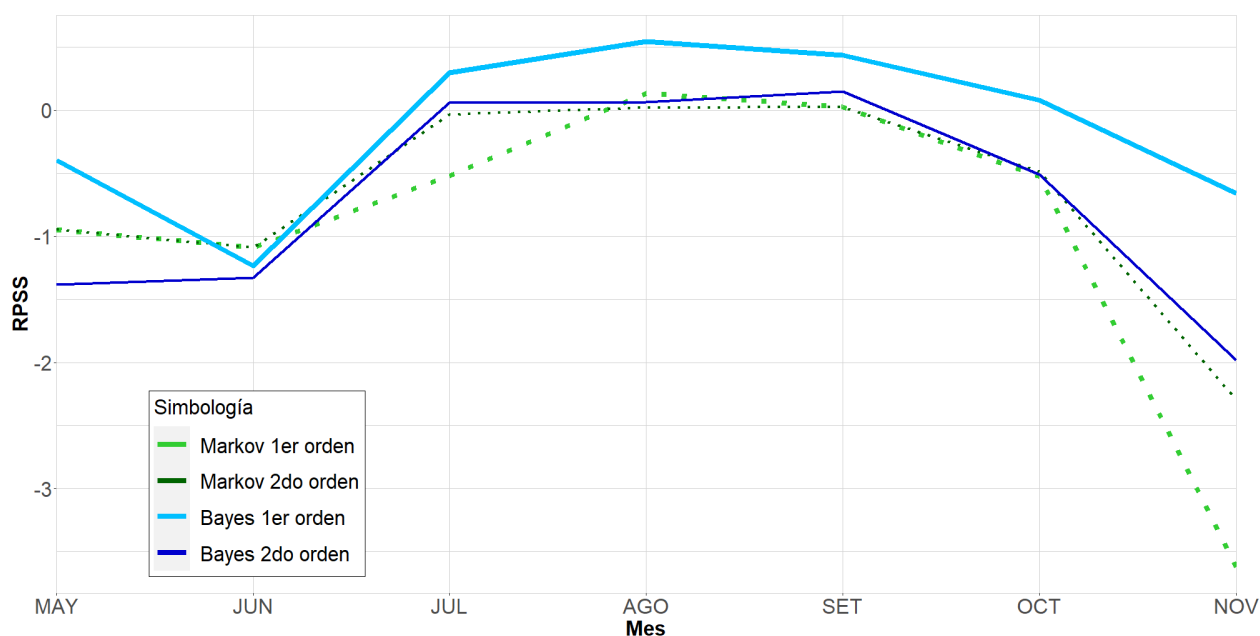


Figura 14. Ranked Probability Skill Score (*RPSS*) de los modelos de Cadenas de Markov y Redes Bayesianas de primer y segundo orden, únicamente para los estados de sequía.

En ese sentido, para esta validación los modelos no presentan una mejora con respecto al pronóstico de referencia. Sin embargo, el modelo de Redes Bayesianas de primer orden destaca, pues en la época lluviosa obtuvo el mejor rendimiento de todos los modelos, principalmente en los meses de julio a octubre. Y aunque en noviembre obtiene un rendimiento negativo, la separación con respecto a los demás modelos es importante, pues de los cuatro modelos es el que mejor logra aproximar un pronóstico más preciso.

Capítulo 5. Aplicación del modelo de Redes Bayesianas de primer orden en las estaciones Meteorológicas Cañas Dulces y Talolinga

A partir del análisis de las evaluaciones de los modelos, se puede responder que el modelo de Redes Bayesianas de primer orden es el que obtiene el mejor rendimiento con los datos de la Estación Aeropuerto. Es por esto que se seleccionó para aplicarlo en las estaciones Meteorológicas de Cañas Dulces y Talolinga. Como se muestra en la Figura 1, la Estación de Cañas Dulces se encuentra al norte de la Estación Aeropuerto, mientras que la Estación Talolinga se ubica al sur. De esta manera se tienen tres referencias espaciales dentro de la cuenca del Río Tempisque, esto con el objetivo de obtener un panorama más diverso del comportamiento del modelo datos diferentes.

5.1 Datos de precipitación de las estaciones Cañas Dulces y Talolinga

La selección de estas dos estaciones se debió a que se buscaba representar de alguna forma la variabilidad de precipitación que cae en la cuenca. Por eso se escogieron primero estaciones que tuvieran una serie temporal completa y segundo que estuvieran en la parte alta y baja de la cuenca. Así las estaciones de Cañas Dulces y Talolinga fueron las que cumplieron con esos criterios.

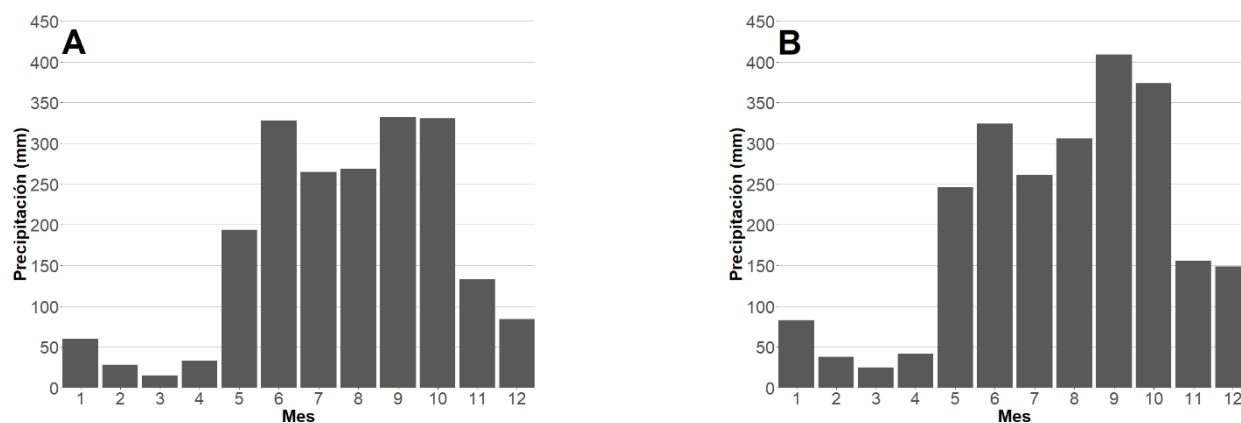


Figura 15. Precipitación promedio mensual para las estaciones Cañas Dulces (A) y Talolinga (B). Ambas con un periodo de 1950 a junio, 2020.

Como se observa en la Figura 15, ambas estaciones presentan el patrón climático típico de la Región Pacífico Norte, con una estación seca de diciembre a abril y una época lluviosa de mayo a noviembre. También se observa que ninguna de las dos estaciones presenta un mes de transición, pues los cambios de épocas se dan de un mes a otro. No obstante, estas estaciones presentan un mes de diciembre con mucha más precipitación

en comparación con la estación Aeropuerto, pues en esta llueve alrededor de 25 mm en promedio, mientras que en Cañas Dulces precipita en promedio 80 mm y en Talolinga 150 mm. Con esto en cuenta, es esperable que, para estas estaciones, se presenten valores de *SPI* menores a -1 para el mes de diciembre, lo que indicaría condiciones más secas con respecto al promedio.

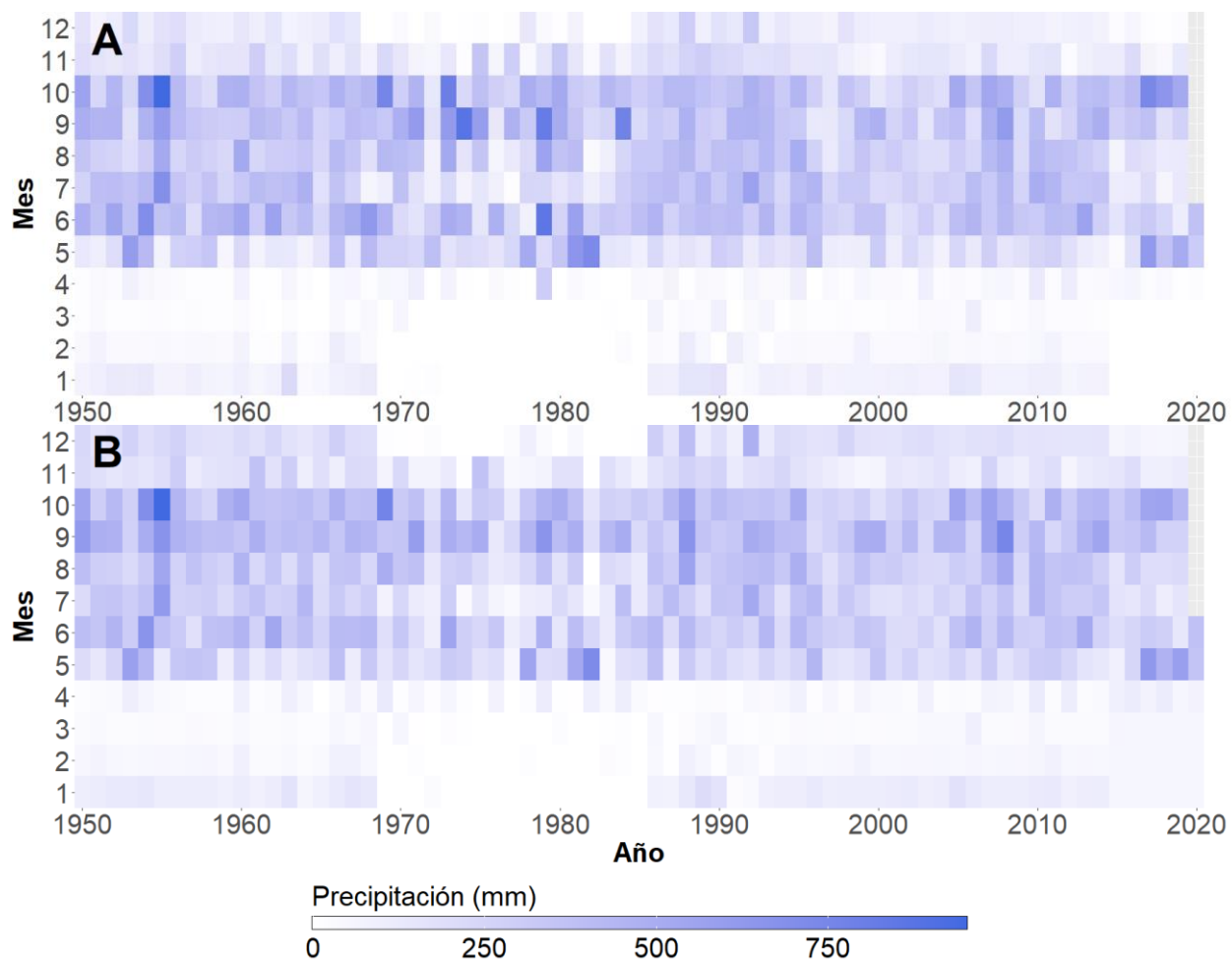


Figura 16. Mapa de calor que representa la precipitación promedio mensual (1950- junio, 2020) para las estaciones Cañas Dulces (A) y Talolinga (B). El tono blanco representa menor precipitación, mientras que los azules una mayor cantidad de lluvia.

Con la Figura 16 se logra visualizar la estacionalidad típica del clima tropical seco, así mismo se confirma que durante el tiempo de registro, no se aprecian meses de transición entre las épocas lluviosas y secas. También se puede visualizar que los meses de diciembre siguen la tendencia de la época lluviosa, esto en comparación de la Estación Aeropuerto (Figura 5). Para ambas estaciones se aprecia que desde 1970 hasta 1985 y

del 2015 al 2020, las épocas secas están mucho más marcadas a diferencia de los otros años, esto se puede deber a una subestimación de los datos con los que se rellenaron los datos faltantes. Ya que como en esta región la precipitación disminuye considerablemente de diciembre a abril, las imágenes satelitales lo muestran de esta manera para la región.

5.2 Estados de sequía para las estaciones Cañas Dulces y Talolinga

Al igual que con la Estación Aeropuerto, se decidió calcular el *SPI* a escala mensual (Anexo 7) para obtener el panorama general del comportamiento mensual de la precipitación y así poder identificar los meses con déficit de precipitación. Como se muestra en la Figura 17, se observan pocos eventos de sequía entre 1950 y 1965, la gran mayoría de los eventos fueron sequías moderadas y solo un evento de sequía severa para cada estación en ese periodo.

Entre 1965 y finales de los años 80, se nota que los eventos de sequía fueron constantes y de diferentes severidades. Incluso se presentaron sequías moderadas en los meses de diciembre marzo y abril, sin embargo, la razón de que estos últimos dos meses presenten este comportamiento se puede deber a la subestimación de los datos. En la década de los 90, los eventos de sequía fueron escasos y se presentaron hasta el final, es decir en 1998, esto debido a impacto de El Niño en ese año. Aunque en las dos estaciones la severidad fue distinta, pues en Cañas Dulces se presentaron sequías severas, mientras que en Talolinga el impacto fue de sequías moderadas.

Las series temporales de estas estaciones se dividieron en dos periodos, el primero es el de calibración que incluye los años entre 1950 y 1999, mientras que el segundo va del 2000 al 2020 y es el periodo de validación. En este último periodo se muestra que la frecuencia de las sequías es similar, pero la severidad es diferente. Ya que Cañas Dulces presenta mayor cantidad de eventos de sequía moderada e incluso eventos de sequía extrema. Por otro lado, en Talolinga no se observa la sequía extrema en ese periodo y en su gran mayoría los eventos fueron moderados.

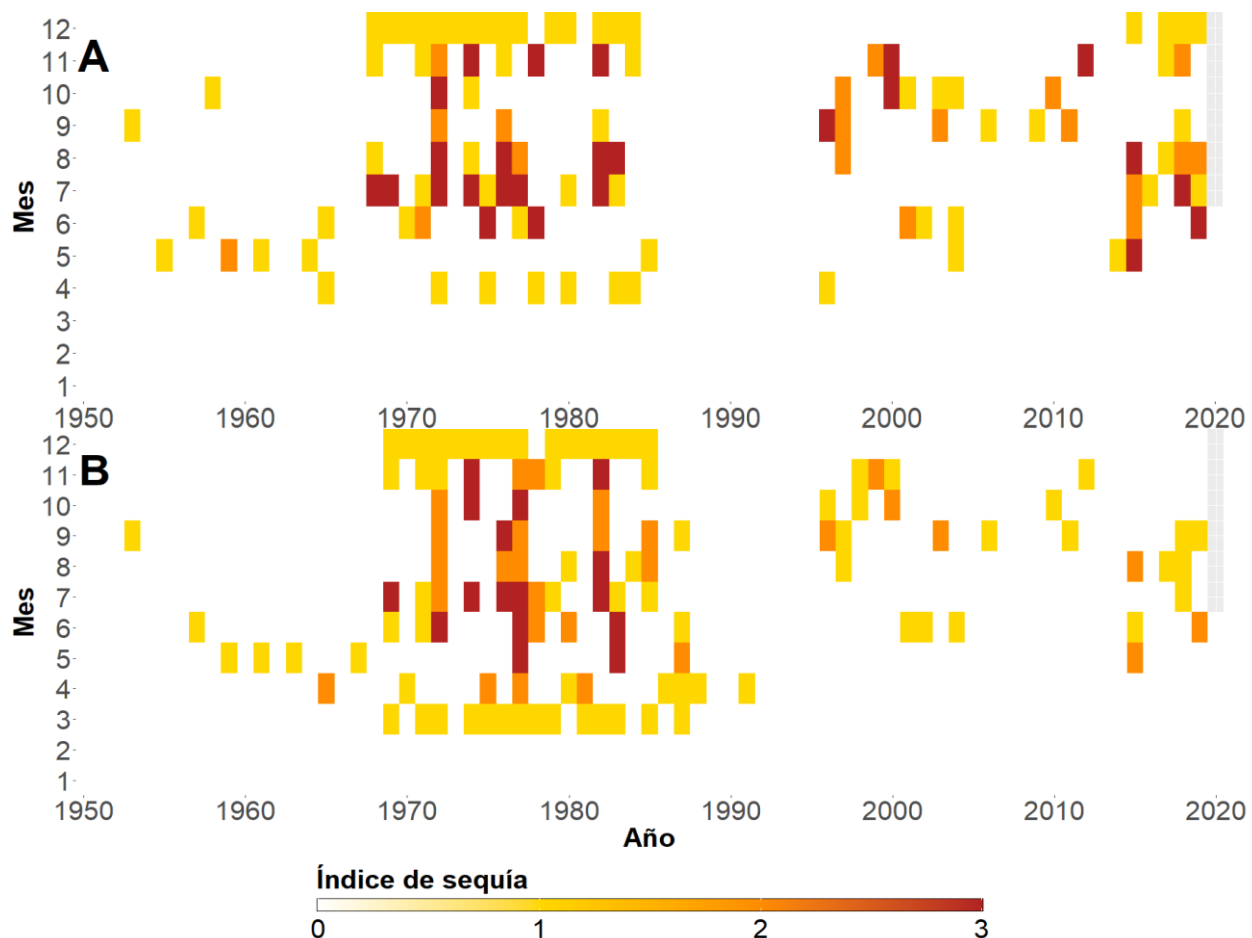


Figura 17. Mapa de calor que representa los estados de sequía por mes de toda la serie temporal para las estaciones Cañas Dulces (A) y Talolinga (B).

De manera general Talolinga es la estación que presenta la mayor frecuencia de sequías meteorológicas, pues en todo el periodo de estudio se observaron 120 eventos, mientras que en Cañas Dulces conto con 107. Sin embargo, esta última mostro 25 sequías extremas, y Talolinga 16. Es importante destacar que, en el periodo de validación, Talolinga no presento sequías extremas, mientras que Cañas Dulces obtuvo 7. Las tablas que presentan la cantidad y porcentaje de la frecuencia de los eventos de sequía se muestran en el Anexo 8 para la Estación Cañas Dulces y en el Anexo 9 para Talolinga.

5.3 Ajuste de cópulas para las estaciones Cañas Dulces y Talolinga.

Luego de tener claro el comportamiento de la sequía meteorológica se procedió a aplicar la función de distribución acumulada (*CDF*) para las dos estaciones. Como los datos son diferentes se debe primero ajustar el umbral de *CDF* que corresponde a cada categoría de sequía, esto se presenta en la Tabla 10. Este paso es importante, pues, aunque los

valores de los umbrales parezcan similares, en realidad la diferencia puede significar que los eventos se clasifiquen en una u otra categoría cuando su distribución se encuentra cerca de un umbral.

Tabla 10. Umbrales de la distribución acumulada de las estaciones Cañas Dulces y Talolinga según cada categoría de sequía.

Categoría	Umbrales <i>CDF</i>	
	Cañas Dulces	Talolinga
No sequía	Mayor a 0.1583	Mayor a 0.158
Sequía moderada	0.1582 a 0.068	0.157 a 0.068
Sequía severa	0.067 a 0.025	0.067 a 0.024
Sequía Extrema	Menor a 0.024	Menor a 0.023

Una vez establecidos los umbrales se aplicó el ajuste de cópulas de dos dimensiones para conocer qué tipo de cópulas se ajustan a cada mes en los periodos de calibración (1950-1999). Los resultados de los valores estadísticos S , los p -values y los parámetros de las cópulas se muestran en el Anexo 10 para Cañas Dulces y en el Anexo 11 para Talolinga. Mientras que en las Tablas 11 y 12 se presenta el resumen de las cópulas seleccionadas y sus respectivos parámetros por mes y para las dos estaciones meteorológicas. Ambas estaciones presentan diferentes tipos de ajuste de cópulas y aun cuando presentan la misma Copula el parámetro es diferente. Esto asegura que el modelo se ajusta a los datos de cada estación, así el resultado de los pronósticos corresponde al comportamiento histórico de cada lugar.

En el caso de Cañas Dulces la copula Clayton es la que mejor se ajusta en la época seca, mientras que en la época lluviosa existe una variabilidad entre Clayton y T, lo que quiere decir que los valores se distribuyen de manera amplia y no poseen una cola en sus extremos. Para Talolinga sucede un comportamiento similar, donde la copula Clayton se ajusta en la época seca, mientras que en la época lluviosa hay cambios entre Clayton, Frank y T.

Tabla 11. Resultado del ajuste mensual de cópulas de dos dimensiones para la Estación de Cañas Dulces.

Mes	Copula	Parámetro
Enero	Clayton	2.057
Febrero	Clayton	2.212
Marzo	Clayton	0.562
Abril	Clayton	-0.103
Mayo	Frank	0.140
Junio	Clayton	0.188
Julio	Clayton	0.562
Agosto	T	0.465
Setiembre	T	0.383
Octubre	Clayton	0.295
Noviembre	T	0.276
Diciembre	Clayton	2.068

Tabla 12. Resultado del ajuste mensual de cópulas de dos dimensiones para la Estación de Talolinga.

Mes	Copula	Parámetro
Enero	Clayton	2.770
Febrero	Clayton	2.386
Marzo	T	0.029
Abril	T	-0.066
Mayo	Frank	2.021
Junio	Clayton	0.664
Julio	Clayton	0.633
Agosto	T	0.392
Setiembre	T	0.442
Octubre	Clayton	0.527
Noviembre	Frank	2.155
Diciembre	Clayton	2.333

5.4 Probabilidades de Redes Bayesianas de primer orden para las estaciones Cañas Dulces y Talolinga

A partir de la selección de las cópulas de mejor ajuste, sus parámetros y los umbrales definidos para cada estación, se obtuvieron las probabilidades de cada categoría en el periodo de validación. Así en la Figura 15 se muestra las probabilidades para Cañas

Dulces y en la Figura 16 las de Talolinga, cada una solamente para la época lluviosa. En los Anexos 12 y 13, se presentan las probabilidades para los doce meses.

Como se observa en la Figura 18, el mes de mayo presenta un comportamiento muy regular, pues no presenta cambios en las probabilidades de ocurrencia de cada categoría de sequía. Esto se debe a la baja correlación que tiene la copula Frank con los datos del mes anterior (abril) el cual tiene ajustado una copula T, es decir que abril presenta una distribución de tipo normal, por lo que su distribución no afecta lo suficiente en mayo para que el modelo pronostique cambios fuertes.

En cuanto a los demás meses, el modelo, si bien mantiene la tendencia de otorgar una mayor probabilidad a la categoría de No Sequía, logra generar indicios de la presencia de algún grado de sequía. Ya que cuando se observaron sequías, las probabilidades de la categoría No Sequía disminuyeron, mientras que las categorías de Sequía Moderada, Severa y Extrema aumentan ligeramente.

Un ejemplo de esto se observa en agosto del 2015, en el cual se observó una sequía extrema, el modelo identificó esto y otorgó una probabilidad del 43% a la categoría No Sequía, sin embargo, cuando se compara ese pronóstico con los agostos anteriores se observa que en promedio esa categoría mantiene una probabilidad superior al 60%. Mientras que la categoría de Sequía Extrema históricamente tiene probabilidades del 4% en promedio, en agosto del 2015 el modelo le dio una probabilidad del 20%.

Para el caso de Talolinga (Figura 19), el mes de mayo es el que presenta mayor estabilidad en las probabilidades, sin embargo, fluctúan más que el mismo mes en Cañas Dulces. Si bien Talolinga presentó menos sequías observadas que Cañas Dulces en el periodo de validación, el modelo también logra interpretar cambios en las distribuciones que pueden indicar categorías de sequía. Esto se observa muy claro en los meses de junio, julio, agosto y setiembre. Particularmente en junio del 2019 las probabilidades de que se presentara la categoría de No Sequía cayeron al 33%, es decir, había un 66% de probabilidad de que se presentara alguna categoría de sequía. En este caso se observó una sequía moderada, es decir, el modelo logró pronosticar el evento de sequía.

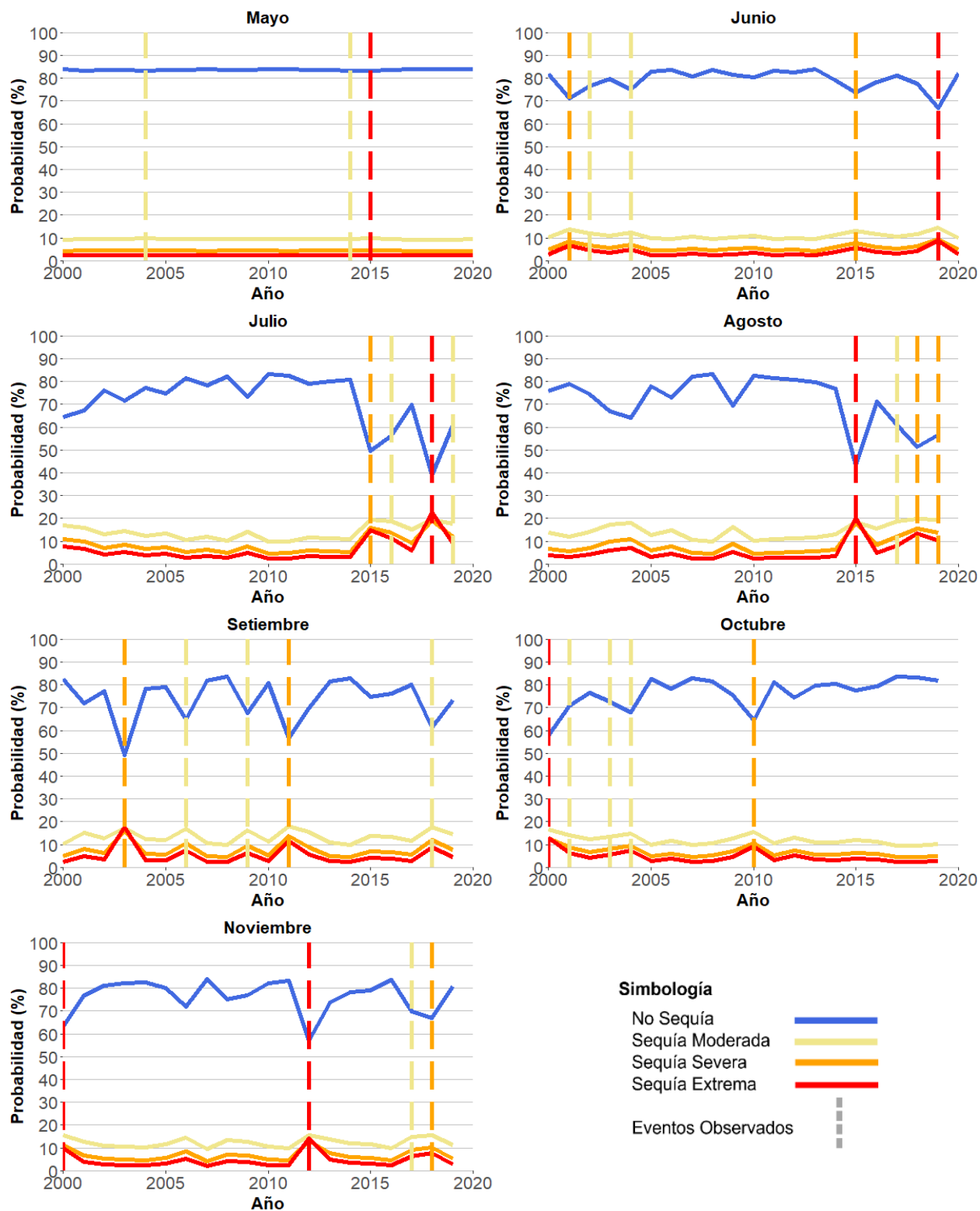


Figura 18. Probabilidades por mes del modelo de Redes Bayesianas de primer orden, en la época lluviosa del 2000 al 2020, en la Estación Cañas Dulces.

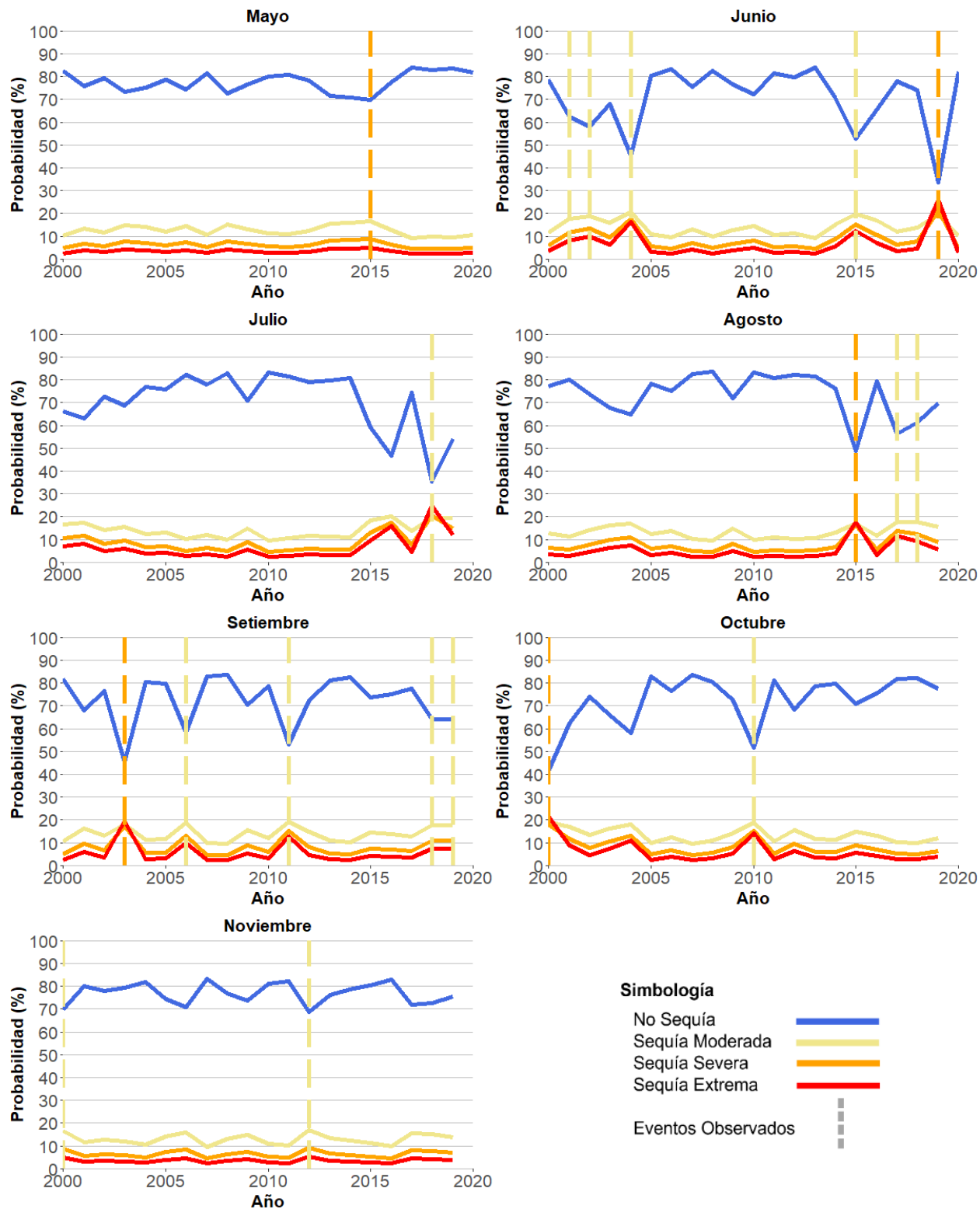


Figura 19. Probabilidades por mes del modelo de Redes Bayesianas de primer orden, en la época lluviosa del 2000 al 2020, en la Estación Talolinga.

5.5 Validación *RPS* para las estaciones Cañas Dulces y Talolinga

Para determinar el comportamiento del modelo de Redes Bayesianas de primer orden en las dos estaciones meteorológicas, se les aplicó una validación *RPS* y otra *RPSS*. En el primer caso, la evaluación *RPS* (Figura 20), muestra que el rendimiento en Cañas Dulces (Figura 20A) es mejor en la mitad de la época lluviosa, mientras que, para el inicio y final de la misma, muestras mayores diferencias con respecto a las observaciones.

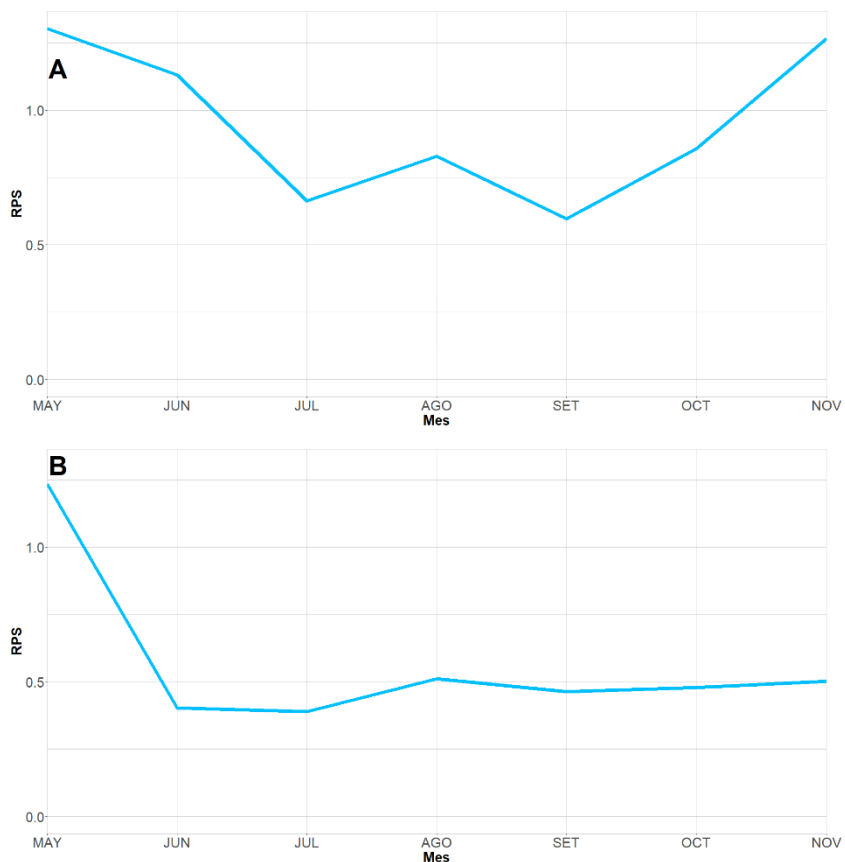


Figura 20. Ranked Probability Score (RPS) del modelo Redes Bayesianas de primer, únicamente para los estados de sequía en las estaciones de Cañas Dulces (A) y Talolinga (B).

En el caso de la estación Talolinga (Figura 20B), el mes de mayo es el que tiene una mayor diferencia con las observaciones, mientras que en los siguientes meses de la estación lluviosa mejora su rendimiento e incluso se mantiene estable, es decir que el modelo en esta estación logra pronosticar los eventos de manera constante y sin mucha diferencia con las observaciones. En el Anexo 14 se muestran los resultados para ambas estaciones a lo largo del año.

5.5 Validación *RPSS* para las estaciones Cañas Dulces y Talolinga

La evaluación *RPSS* (Figura 21), muestra que el modelo en la Estación de Cañas Dulces (Figura 21A), no logra mejorar el pronóstico de referencia en los meses de mayo y octubre, sin embargo, logra una leve mejoría para el resto de la temporada lluviosa, obteniendo en julio su mejor rendimiento.

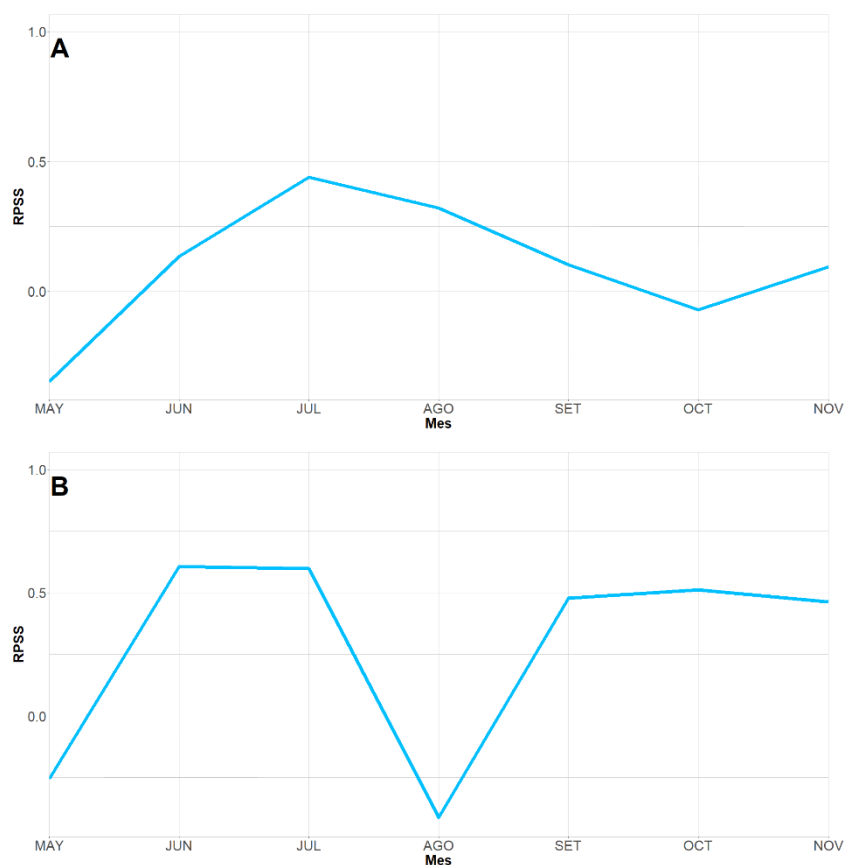


Figura 21. Ranked Probability Skill Score (*RPSS*) del modelo Redes Bayesianas de primer, únicamente para los estados de sequía en las estaciones de Cañas Dulces (A) y Talolinga (B).

En Talolinga (Figura 21B), el rendimiento es bastante bueno, ya que los únicos meses en donde no logró una mejoría con respecto al pronóstico de referencia fue en mayo y agosto. Pero en los otros meses logra una mejorar ese pronóstico en un 50% en promedio, lo que demuestra que las Redes Bayesianas no solo logra adaptarse a los patrones característicos de cada serie de datos, sino que también da buenos resultados al momento de generar las probabilidades de las categorías de sequía.

Conclusiones y recomendaciones

En esta tesis se exploró la utilización de los modelos estadísticos de Cadenas de Markov y Redes Bayesianas para generar pronósticos de sequía meteorológica en tres estaciones meteorológicas situadas en la cuenca del Río Tempisque. La utilización de los modelos probabilísticos permite conocer la incertidumbre de los resultados, así los tomadores de decisiones tienen un margen de seguridad con el cual pueden decidir y mejorar la utilización de los recursos, tanto económicos como ambientales.

Antes de la aplicación de los modelos se realizó un análisis del comportamiento mensual de la precipitación. Esto con el objetivo de distinguir los eventos de sequía, para lo cual se aplicó un *SPI* a escala mensual. De este análisis se concluye que las sequías meteorológicas se presentan durante la época lluviosa. Esto tiene sentido porque la precipitación es muy escasa durante los meses secos, por lo que no ocurre un déficit en el promedio de precipitación.

Los resultados de las Cadenas de Markov, muestran que durante la época seca logran un rendimiento perfecto. Ya que el modelo pronostica que la categoría de No Sequía tiene una probabilidad del 100% durante estos meses. Sin embargo, esto es esperado ya que con la serie temporal se demostró que las sequías meteorológicas no ocurren en la época seca. Por otro lado, en la época lluviosa tienen problemas para generar pronósticos, pues no logran predecir la variabilidad de precipitación en esos meses. Esto se debe a que las Cadenas de Markov se basan en los cambios de estados de un mes a otro, entonces, si no se observa un cambio, el modelo seguirá pronosticando la categoría de No Sequía, pues esta es la de mayor frecuencia y probabilidad.

En el caso de las Redes Bayesianas la utilización de funciones de cópulas resultó importante pues estas se adaptan a las distribuciones de los datos. Así cada mes tiene un ajuste diferente, y aun cuando se cuenta con el mismo tipo de copula ajustada para dos meses, el parámetro de ajuste es diferente. De esta manera las Redes Bayesianas dan una probabilidad a las cuatro categorías, pero su distribución se va ajustando según el comportamiento de los datos.

De la comparación y validación de estos modelos en la Estación Aeropuerto, se concluye que las Redes Bayesianas de primer orden son las que tienen el mejor rendimiento. Esto

porque sus pronósticos lograron diferenciar los meses secos de los lluviosos. Y durante la época húmeda se ajustaron las probabilidades de acuerdo al comportamiento mensual, aun cuando la categoría No Sequía era la de mayor probabilidad, se notó que cuando había una sequía observada esas probabilidades disminuyeron y las probabilidades de las categorías de sequía aumentaron. Esto es importante ya que permitiría utilizar este pronóstico para conocer el estado de sequía y continuar su monitoreo si las probabilidades de sequía empiezan a aumentar.

La aplicación del modelo de Redes Bayesianas de primer orden en las estaciones Cañas Dulce y Talolinga, resultó ser eficiente, pues se adaptó a las condiciones particulares de cada estación. Esto quiere decir que las Redes Bayesianas es un buen modelo para aplicar en diferentes series temporales, pues sus cálculos se basan únicamente en los datos de entrada. Por eso es importante conocer el comportamiento del área de estudio y de los datos, para así lograr una interpretación correcta de los pronósticos.

Una de las limitantes de esta metodología es que se necesita conocer los valores de *SPI* mensual para conocer las probabilidades. Es decir, para pronosticar los siguientes meses se necesita saber la precipitación con antelación, lo cual le agregaría una mayor incertidumbre al modelo. En este sentido este pronóstico solo se puede aplicar al inicio de cada mes, para conocer las probabilidades de ese mismo mes, pues solo así se tendría certeza de los datos de precipitación del mes anterior.

En conclusión, los modelos probabilísticos analizados en esta tesis lograron diferenciar la marcada estacionalidad de la región. Los modelos de Redes Bayesianas de primer orden logran el mejor rendimiento principalmente en la época lluviosa. Lo cual es el criterio más importante, pues es en esta época cuando se presentan las sequías meteorológicas, por lo que se recomienda el uso de estos modelos para la cuenca del Tempisque. También se recomienda explorar la aplicabilidad de estos modelos con *SPI* a diferentes escalas, esto para tratar de detectar y pronosticar otros tipos de sequías que tengan un impacto más directo en las comunidades, como lo son las sequías hidrológicas y agrícolas.

Referencias bibliográficas

- Alam, N. *et ál.* (2015). Probabilistic Drought Analysis of Weekly Rainfall Data Using Markov Chain Model. *Journal of Reliability and Statistical Studies*. 8 (1): 105-114.
- Andreu, J. (Ed). (2015). *Drought: Research and Science-Policy Interfacing*. Leiden: CRC Press/Balkema.
- Avilés, A., Célleri, R., Paredes, J., Solera, A. (2015). Evaluation of Markov Chain Based Drought Forecasts in an Andean Regulated River Basin Using the Skill Scores RPS and GMSS. *Water Resour Manag*. 29:1949–1963. doi: 10.1007/s11269-015-0921-2
- Avilés, A., Célleri, R., Solera, A., Paredes, J. (2016). Probabilistic Forecasting of Drought Events Using Markov Chain- and Bayesian Network- Based Models: A Case Study of an Andean Regulated River Basin. *Water*. 8, 37; doi:10.3390/w8020037
- Avilés, A. (2017). *Pronóstico probabilístico de eventos de sequías y evaluación del riesgo en la gestión de sistemas de recursos hídricos. Caso de estudio en una cuenca andina regulada*. (Tesis doctoral en Ingeniería del Agua y Medioambiental). Universidad Politécnica de Valencia.
- Barrantes, R. (2002). *Investigación: un camino al conocimiento, un enfoque cualitativo y cuantitativo*. San José. EUNED.
- Birkel, C. (2005). *Temporal and Spatial Variability of Drought Indices in Costa Rica*. (Tesis de maestría). Institut für Hydrologie, Universität Freiburg.
- Birkel, C. (2006). Sequía Hidrológica en Costa Rica ¿Se han vuelto más severas y frecuentes en los últimos años? *Revista Reflexiones*. 85 (1-2): 107-116.
- Birkel, C., Brenes, A., Sánchez-Murillo, R. (2017). The Tempisque-Bebedero catchment system: energy-water-food consensus in the seasonally dry tropics of northwestern Costa Rica, in Al-Saidi, Mohammad, Ribbe, Lars (Eds.). *Nexus*

Outlook: assessing resource use challenges in the water, energy and food nexus. Nexus Research Focus, TH-Koeln, University of Applied Sciences.

Decreto Ejecutivo N°15321-MAG. Crea Distritos de Riego Avenamiento y Control de Inundaciones. San José, Costa Rica, 16 de marzo de 1984.

Díaz, R., Alfaro, E. (2016). El Mega-Niño de 1877-1878 y su impacto social en Costa Rica. *Tópicos Meteorológicos y Oceanográficos*. Obtenido de https://www.researchgate.net/publication/314864820_El_Mega-Nino_de_1877-1878_y_su_impacto_social_en_Costa_Rica

Durham, R., Rivera-Araya, M., Davis, M., Quesada, L., Venegas, N., Cui, D. (2016). *Monitoring Drought and Water Balance in the Guanacaste Province to Enhance Decision Making and Response Planning in Costa Rica*. NASA DEVELOP National Program. Recuperado de https://www.researchgate.net/publication/310141860_Monitoring_Drought_and_Water_Balance_in_the_Guanacaste_Province_to_Enhance_Decision_Making_and_Response_Planning_in_Costa_Rica/ [Consulta 22 mayo 2017].

FAO. (Food and Agriculture Organization). (2016). *Corredor Seco América Central. Informe de Situación*. Recuperado de <http://www.fao.org/3/a-br092s.pdf> [Consulta 4 mayo 2017].

Funk, C.C., Peterson, P.J., Landsfeld, M.F., Pedreros, D.H., Verdin, J.P., Rowland, J.D., Romero, B.E., Husak, G.J., Michaelsen, J.C., and Verdin, A.P. (2014). A quasi-global precipitation time series for drought monitoring: U.S. Geological Survey Data Series 832, 4 p. <http://pubs.usgs.gov/ds/832/>

Guzmán-Arias, I. (2013). Recurso Hídrico Superficial en la Cuenca del río Tempisque, Costa Rica: Escenarios de Disponibilidad y Demanda. *Tecnología en Marcha*. Número Especial: 118-132. doi: 10.18845/tm.v0i0.1661

Heckerman, D. (1995). A tutorial on learning with Bayesian networks. In *Learning in Graphical Models*; Springer: Dordrecht, The Netherlands, 1998; pp. 301–354.

- IMN. (Instituto Meteorológico Nacional). (2008). *Clima, variabilidad y cambio climático en Costa Rica*. [Archivo PDF]. Recuperado de http://www.cambioclimaticocr.com/multimedia/recursos/mod1/Documentos/el_clima_variabilidad_y_cambio_climatico_en_cr_version_final.pdf [Consulta 22 de mayo 2017].
- IMN. (Instituto Meteorológico Nacional). (2015). *Boletín Meteorológico Mensual*. [Archivo PDF]. Recuperado de <https://www.imn.ac.cr/documents/10179/14641/JUNIO/ef2e9527-e02e-446e-becc-9f19ad241d78?version=1/> [Consulta 4 mayo 2017].
- Ki-moon, B. (2013). *Día Mundial de Lucha contra la Desertificación y la Sequía*. Recuperado de <http://www.cepal.org/es/articulos/dia-mundial-de-lucha-contra-la-desertificacion-y-la-sequia/> [Consulta 4 mayo 2017].
- Madadgar, S., Moradkhani, H. (2013). A Bayesian Framework for Probabilistic Seasonal Drought Forecasting. *Journal of Hydrometeorology*. 14: 1685-1705. doi: 10.1175/JHM-D-13-010.1
- Mena-Benavides, M. (2019). *Evaluación de la sustentabilidad hídrica de las cuencas Tempisque y Bebedero mediante el modelo Water Evaluation and Planning System (WEAP). Guanacaste, Costa Rica, 1993-2016*. (Tesis de licenciatura en Geografía). Universidad de Costa Rica.
- Mendoza Velázquez, Alfonso, & Galvanovskis, Evalds. (2014). La cópula GED bivariada. Una aplicación en entornos de crisis. *El trimestre económico*, 81(323), 721-746. Recuperado en 15 de marzo de 2021, de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2448-718X2014000300721&lng=es&tlng=es
- Mishra, A., Desai, V. (2006). Drought forecasting using stochastic models. *Stoch Environ Res Risk Assess*. 19: 326–339. doi: 10.1007/s00477-005-0238-4
- Mishra, A.; Singh, V. (2011). Drought modeling – A review. *Journal of Hydrology*. doi: 10.1016/j.jhydrol.2011.03.049
- Mora, E. (2010). *Curso intermedio de probabilidades*. San José, Costa Rica: Edit. UCR.

- Mora, J. (2016). Más allá de la sequía. El Tempisque, un gigante que muere en silencio. *Seminario Universidad*. Recuperado de <https://semanariouniversidad.com/pais/tempisque-gigante-muere-silencio/> [Consulta 28 junio 2017].
- Patterson, O. (1992). Riesgo por sequías en Costa Rica. *Revista Geográfica de América Central*. 25-26, pp. 385-411
- PIAAG. (Programa Integral de Abastecimiento de Agua para Guanacaste). (s.f.). *Proyecto Cuenca Media Río Tempisque y Comunidades Costeras*. [Archivo PDF]. Recuperado de <http://www.da.go.cr/informacion-general-piaag/#/> [Consulta 22 de mayo].
- Schmidt, T. (2006). Coping with copulas. *Copulas - From Theory to Applications in Finance*. Londres. Risk Books: 3–34
- Steyn, D., Moisseeva, N., Harari, O., & Welch, W. J. (2016). *Temporal and Spatial Variability of Annual Rainfall Patterns in Guanacaste, Costa Rica*. Vancouver: University of British Columbia Library. doi: 10.14288/1.0340318
- Van Loon, A. (2015). *Hydrological drought explained*. *WIREs Water*, 2: 359–392. doi: 10.1002/wat2.108
- Vargas, G. (2001). Fitogeografía de ecosistemas secos en la meseta de ignimbritas de Guanacaste, Costa Rica. *Revista Biología Tropical*. 49(1): 227-238.
- Wilhite, D.A.; y Glantz, M.H. (1985). Understanding the Drought Phenomenon: The Role of Definitions. *Water International* 10:3 (1985), pp. 111–120.
- Wilks, D.S. (2006). *Statistical Methods in the Atmospheric Sciences, 3rd ed.*; Academic Press: Cambridge, MA, USA, 2011.
- Willmott, C. J. y K. Matsuura. (2001). *Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950 - 1999)*. Recuperado de http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts2.html/

WMO. (World Meteorological Organization). (2012). *Standardized Precipitation Index User Guide*. WMO-No. 1090.

Wolter, K., and M. S. Timlin. (1998). Measuring the strength of ENSO events - how does 1997/98 rank? *Weather*, 53, 315-324. DOI: 10.1002/j.1477- 8696.1998.tb06408.x.

Yan, J. (2006). Enjoy the joy of copulas: With a Package copula. *Journal of Statistical Software*. 21, 1–21.

Anexos

Anexo 1. Probabilidades de transición de los estados del modelo *MC1*, para todo el año en la Estación Aeropuerto.

Estado mes anterior <i>i</i>	Estado mes actual <i>j</i>	Probabilidades del mes actual <i>j</i>											
		ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC
0	0	1	1	1	1	0.78	0.88	0.86	0.87	0.96	0.84	0.90	1
	1	0	0	0	0	0.16	0.08	0.02	0.07	0.02	0.14	0.06	0
	2	0	0	0	0	0.02	0	0.02	0.02	0	0.02	0.02	0
	3	0	0	0	0	0.05	0.04	0.11	0.04	0.02	0	0.02	0
1	0	1	1	1	1	1	1	1	0	1	0.50	0.56	1
	1	0	0	0	0	0	0	0	0	0	0.50	0.22	0
	2	0	0	0	0	0	0	0	0	0	0	0.22	0
	3	0	0	0	0	0	0	0	1	0	0	0	0
2	0	1	1	1	1	1	1	1	0	0.50	0.50	1	1
	1	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	1	0	0.50	0	0
	3	0	0	0	0	0	0	0	0	0.50	0	0	0
3	0	1	1	1	1	1	0.67	1	0.33	0.33	0.67	1	1
	1	0	0	0	0	0	0	0	0.17	0.17	0	0	0
	2	0	0	0	0	0	0.33	0	0	0.33	0	0	0
	3	0	0	0	0	0	0	0	0.50	0.17	0.33	0	0

Anexo 2. Probabilidades de transición de los estados del modelo *MC2*, para todo el año en la Estación Aeropuerto.

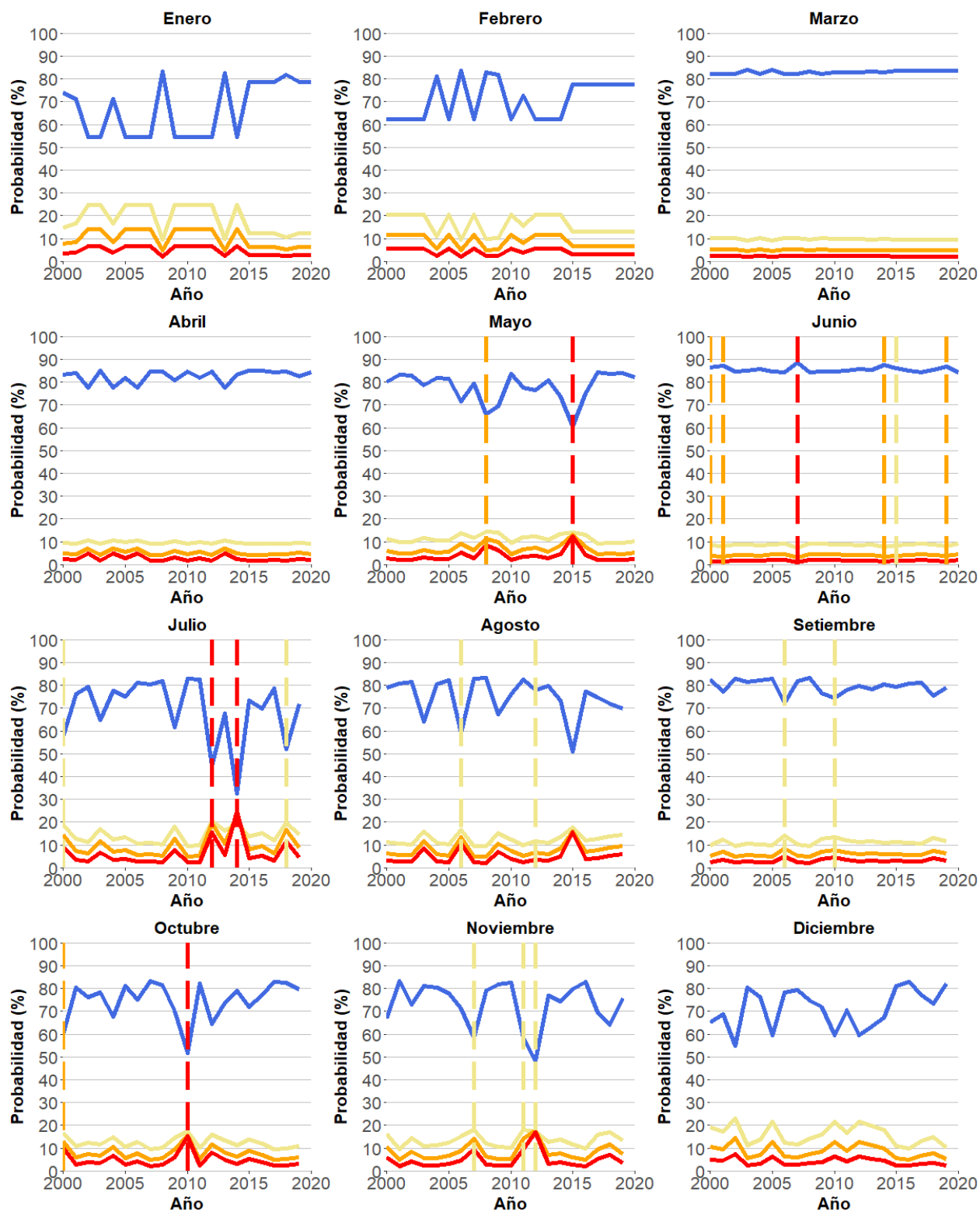
Estados meses anteriores <i>i</i>	Estado mes actual <i>j</i>	Probabilidades del mes actual <i>j</i>											
		ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC
0-0	0	0.25	0.25	0.25	0.25	0.78	0.88	0.84	0.85	0.96	0.88	0.89	0.25
	1	0.25	0.25	0.25	0.25	0.16	0.08	0.02	0.08	0.02	0.10	0.06	0.25
	2	0.25	0.25	0.25	0.25	0.02	0	0.02	0.02	0	0.02	0.02	0.25
	3	0.25	0.25	0.25	0.25	0.05	0.04	0.12	0.04	0.02	0	0.02	0.25
0-1	0	0.25	0.25	0.25	0.25	0.25	1	1	0	1	0	0.63	0.25
	1	0.25	0.25	0.25	0.25	0.25	0	0	0	0	1	0.25	0.25
	2	0.25	0.25	0.25	0.25	0.25	0	0	0	0	0	0.13	0.25
	3	0.25	0.25	0.25	0.25	0.25	0	0	1	0	0	0	0.25
0-2	0	0.25	0.25	0.25	0.25	0.25	1	0.25	0	0	0.25	1	0.25
	1	0.25	0.25	0.25	0.25	0.25	0	0.25	0	0	0.25	0	0.25
	2	0.25	0.25	0.25	0.25	0.25	0	0.25	1	0	0.25	0	0.25
	3	0.25	0.25	0.25	0.25	0.25	0	0.25	0	1	0.25	0	0.25
0-3	0	0.25	0.25	0.25	0.25	0.25	0.67	1	0.33	0.5	1	0.25	0.25
	1	0.25	0.25	0.25	0.25	0.25	0	0	0.17	0	0	0.25	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.33	0	0	0	0	0.25	0.25
	3	0.25	0.25	0.25	0.25	0.25	0	0	0.5	0.5	0	0.25	0.25
1-0	0	0.25	0.25	0.25	0.25	0.25	0.25	0.9	1	0.25	0.6	1	0.25
	1	0.25	0.25	0.25	0.25	0.25	0.25	0	0	0.25	0.4	0	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.25	0	0	0.25	0	0	0.25
	3	0.25	0.25	0.25	0.25	0.25	0.25	0.1	0	0.25	0	0	0.25
1-1	0	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25
	1	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	1	0.25
	3	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25
1-2	0	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
	1	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
	3	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
1-3	0	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	0.25	0.25
	1	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	0.25	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	1	0.25	0.25	0.25
	3	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	0.25	0.25
2-0	0	0.25	0.25	0.25	0.25	0.25	0.25	1	1	0.25	1	1	0.25
	1	0.25	0.25	0.25	0.25	0.25	0.25	0	0	0.25	0	0	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.25	0	0	0.25	0	0	0.25
	3	0.25	0.25	0.25	0.25	0.25	0.25	0	0	0.25	0	0	0.25

2-1	0	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
	1	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
	3	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
2-2	0	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	1	0.25	1	0.25
	1	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	0	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	0	0.25
	3	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	0	0.25
2-3	0	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	1	0.25	0.25
	1	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	0.25
	3	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	0.25
3-0	0	0.25	0.25	0.25	0.25	0.25	0.25	1	1	1	0.5	1	0.25
	1	0.25	0.25	0.25	0.25	0.25	0.25	0	0	0	0.5	0	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.25	0	0	0	0	0	0.25
	3	0.25	0.25	0.25	0.25	0.25	0.25	0	0	0	0	0	0.25
3-1	0	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	1	1	0.25	0.25
	1	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0	0.25	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0	0.25	0.25
	3	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	0	0.25	0.25
3-2	0	0.25	0.25	0.25	0.25	0.25	0.25	1	0.25	0.25	0.5	0.25	0.25
	1	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	0.25	0	0.25	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	0.25	0.5	0.25	0.25
	3	0.25	0.25	0.25	0.25	0.25	0.25	0	0.25	0.25	0	0.25	0.25
3-3	0	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.33	0	1	0.25
	1	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.33	0	0	0.25
	2	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.33	0	0	0.25
	3	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0	1	0	0.25

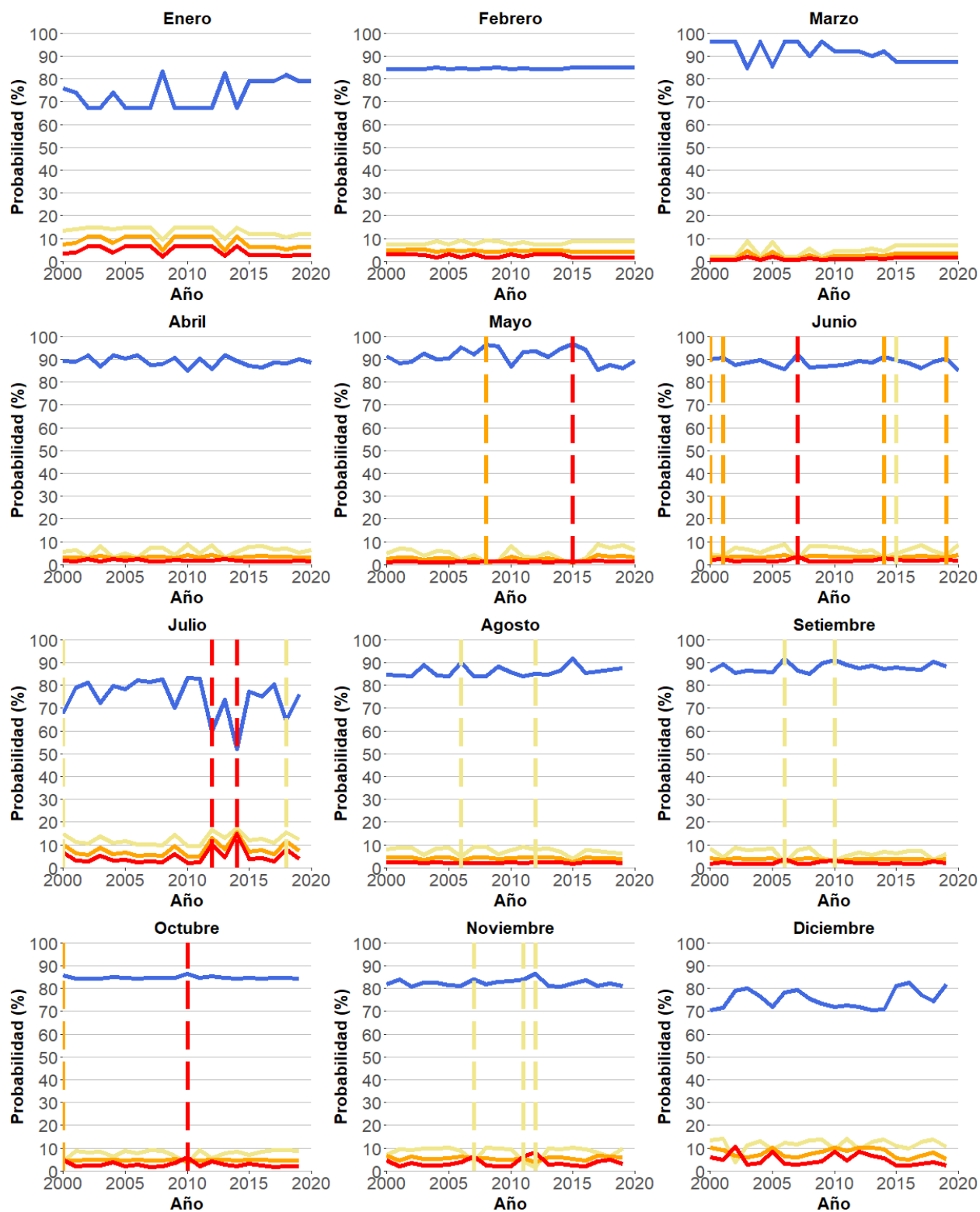
Anexo 3. Resultado del ajuste mensual de cópulas de dos dimensiones, para la Estación Aeropuerto. En negrita y subrayado se resaltan los valores que cumplen con el criterio de selección de mejor ajuste para cada mes.

<i>p-Value</i>												
Copula	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC
Normal	0.583	0.417	0.083	0.417	0.250	0.083	0.750	0.417	0.083	0.583	0.083	0.083
T	0.417	0.417	0.083	0.417	0.750	0.250	0.583	0.750	0.250	0.583	0.417	0.250
Clayton	0.750	0.250	0.417	0.917	0.417	0.083	0.750	0.583	0.250	0.583	0.083	0.083
Frank	0.917	0.583	0.583	0.917	0.583	0.250	0.083	0.417	0.250	0.583	0.083	0.417
Valor estadístico S												
Copula	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC
Normal	0.022	0.026	<u>0.036</u>	0.023	0.027	0.030	0.023	0.025	<u>0.035</u>	0.024	0.031	0.030
T	0.023	0.025	0.037	<u>0.022</u>	<u>0.026</u>	0.030	0.022	0.025	0.036	0.024	<u>0.030</u>	0.032
Clayton	0.056	0.063	0.038	0.023	0.035	<u>0.029</u>	<u>0.019</u>	<u>0.022</u>	0.044	<u>0.021</u>	0.046	0.037
Frank	<u>0.012</u>	<u>0.021</u>	0.036	0.023	0.027	0.030	0.025	0.028	0.037	0.024	0.034	<u>0.030</u>
Parámetros de las cópulas												
Copula	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC
Normal	0.762	0.609	<u>0.058</u>	0.016	0.226	-0.040	0.386	0.249	<u>0.201</u>	0.244	0.399	0.565
T	0.762	0.609	0.058	<u>0.016</u>	<u>0.226</u>	-0.040	0.386	0.249	0.201	0.244	<u>0.399</u>	0.565
Clayton	2.462	1.431	0.077	0.021	0.340	<u>-0.049</u>	<u>0.675</u>	<u>0.381</u>	0.296	<u>0.373</u>	0.707	1.236
Frank	<u>6.766</u>	<u>4.401</u>	0.336	0.092	1.331	-0.227	2.396	1.471	1.178	1.442	2.492	<u>3.917</u>

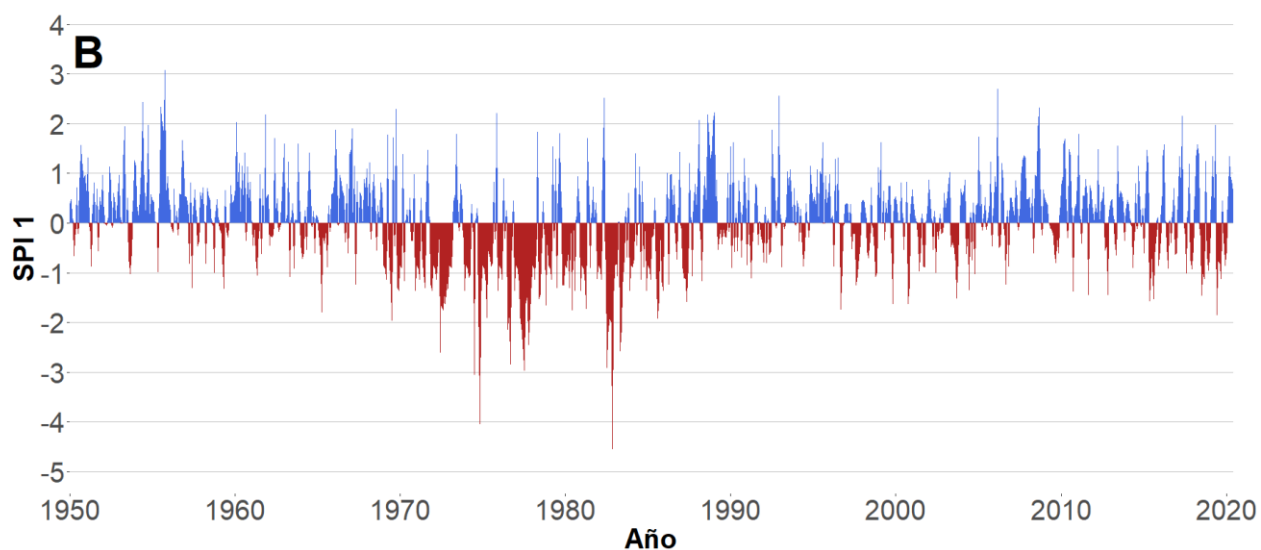
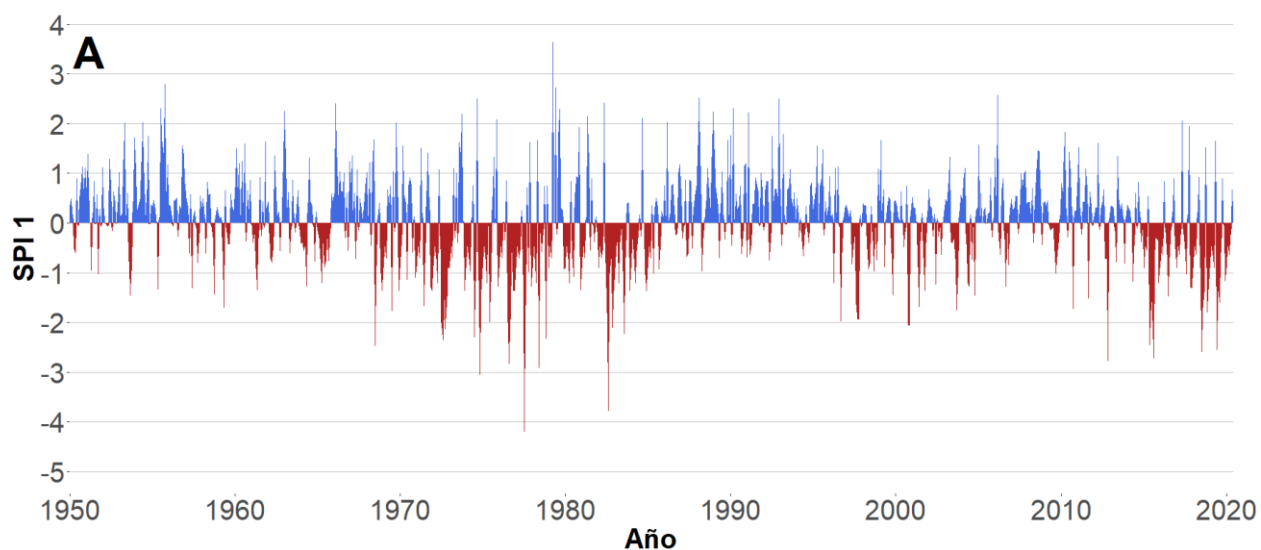
Anexo 4. Probabilidades del modelo de Redes Bayesianas de primer orden, para todos los meses del 2000 al 2020, en la Estación Aeropuerto.



Anexo 6. Probabilidades del modelo de Redes Bayesianas de segundo orden, para todos los meses del 2000 al 2020, en la Estación Aeropuerto.



Anexo 7. Índice Estandarizado de Precipitación para las estaciones Cañas Dulces (A) y Talolinga (B). Ambas con una serie temporal desde 1950 hasta junio, 2020.



Anexo 8.

Frecuencia de los eventos de sequía en la Estación Cañas Dulces (1950-2020).

Categoría	Cantidad de meses	Porcentaje de los datos
No sequía	739	0.87
Sequía moderada	63	0.07
Sequía severa	19	0.02
Sequía extrema	25	0.03
Total	846	1

Frecuencia de los eventos de sequía para el periodo de calibración (1950-1999), en la Estación Cañas Dulces.

Categoría	Cantidad de meses	Porcentaje de los datos
No sequía	527	0.88
Sequía moderada	45	0.08
Sequía severa	10	0.02
Sequía extrema	18	0.03
Total	600	1

Frecuencia de los eventos de sequía para el periodo de validación (2000-2020), en la Estación Cañas Dulces.

Categoría	Cantidad de meses	Porcentaje de los datos
No sequía	212	0.86
Sequía moderada	18	0.07
Sequía severa	9	0.04
Sequía extrema	7	0.03
Total	246	1

Anexo 9.

Frecuencia de los eventos de sequía en la Estación Talolinga (1950-2020).

Categoría	Cantidad de meses	Porcentaje de los datos
No sequía	726	0.86
Sequía moderada	73	0.09
Sequía severa	31	0.04
Sequía extrema	16	0.02
Total	846	1

Frecuencia de los eventos de sequía para el periodo de calibración (1950-1999), en la Estación Talolinga.

Categoría	Cantidad de meses	Porcentaje de los datos
No sequía	499	0.83
Sequía moderada	59	0.10
Sequía severa	26	0.04
Sequía extrema	16	0.03
Total	600	1

Frecuencia de los eventos de sequía para el periodo de validación (2000-2020), en la Estación Talolinga.

Categoría	Cantidad de meses	Porcentaje de los datos
No sequía	227	0.92
Sequía moderada	14	0.06
Sequía severa	5	0.02
Sequía extrema	0	0.00
Total	246	1

Anexo 10. Resultado del ajuste mensual de cópulas de dos dimensiones, para la Estación Cañas Dulces. En negrita y subrayado se resaltan los valores que cumplen con el criterio de selección de mejor ajuste para cada mes.

<i>p-Value</i>												
Copula	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC
Normal	0.083	0.083	0.417	0.250	0.417	0.417	0.083	0.583	0.417	0.250	0.250	0.083
T	0.083	0.083	0.417	0.250	0.250	0.083	0.083	0.750	0.417	0.417	0.750	0.083
Clayton	0.083	0.083	0.417	0.583	0.583	0.083	0.417	0.583	0.250	0.250	0.083	0.083
Frank	0.083	0.250	0.083	0.417	0.417	0.083	0.250	0.583	0.083	0.250	0.417	0.083

Valor estadístico S												
Copula	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC
Normal	0.097	0.079	0.038	0.029	0.030	0.039	0.050	0.025	0.033	0.032	0.029	0.056
T	0.095	0.078	0.038	0.031	0.031	0.037	0.050	0.022	0.031	0.032	0.025	0.053
Clayton	0.050	0.046	0.029	0.028	0.030	0.036	0.038	0.027	0.035	0.029	0.032	0.034
Frank	0.086	0.058	0.035	0.028	0.030	0.039	0.052	0.028	0.036	0.032	0.030	0.058

Parámetros de las cópulas												
Copula	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC
Normal	0.715	0.734	0.338	-0.085	0.024	0.135	0.338	0.465	0.383	0.201	0.276	0.716
T	0.715	0.734	0.338	-0.085	0.024	0.135	0.338	0.465	0.383	0.201	0.276	0.716
Clayton	2.057	2.212	0.562	-0.103	0.032	0.188	0.562	0.888	0.668	0.295	0.433	2.068
Frank	5.865	6.213	2.054	-0.488	0.140	0.779	2.056	3.004	2.377	1.173	1.645	5.889

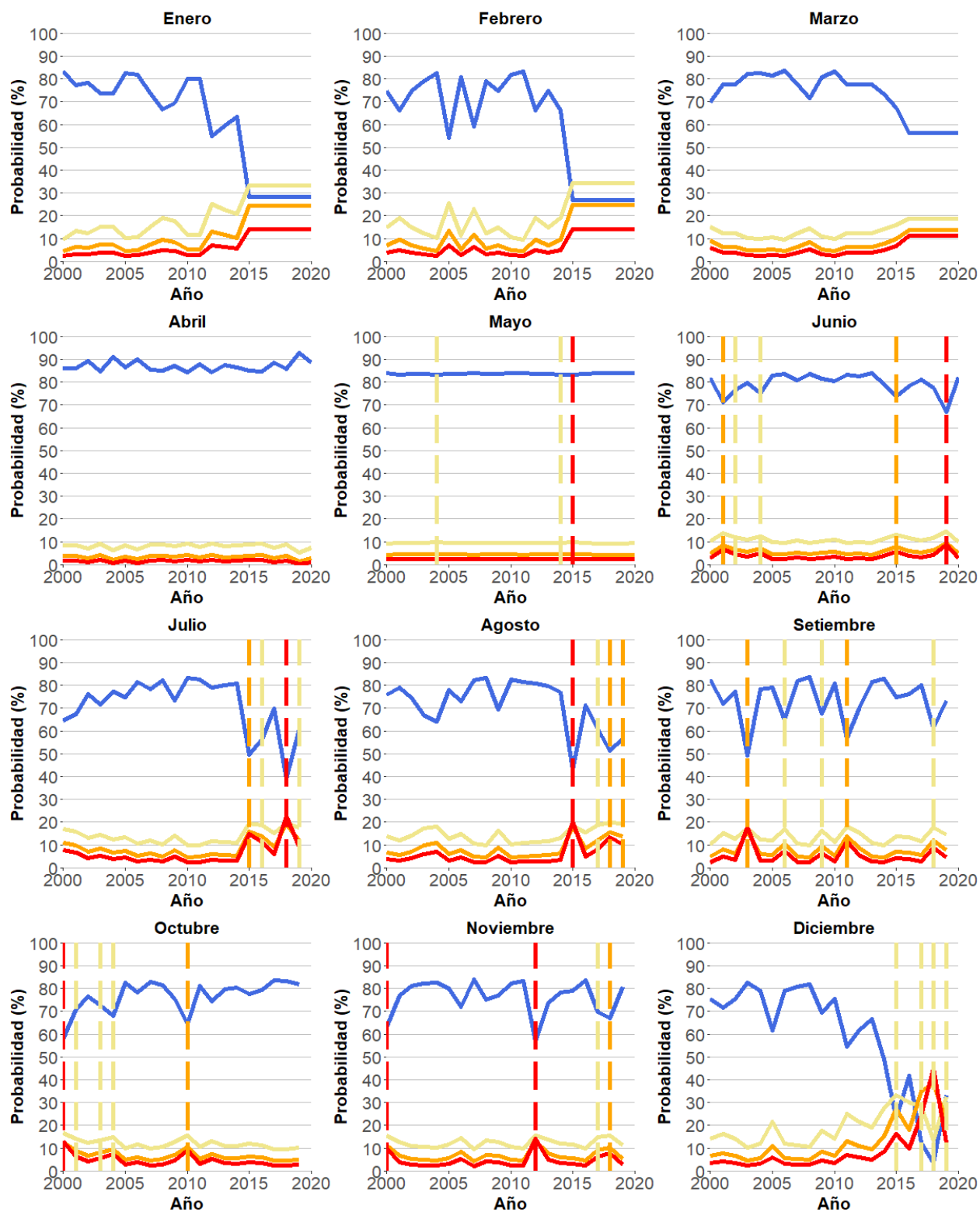
Anexo 11. Resultado del ajuste mensual de cópulas de dos dimensiones, para la Estación Talolinga. En negrita y subrayado se resaltan los valores que cumplen con el criterio de selección de mejor ajuste para cada mes.

<i>p-Value</i>												
Copula	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC
Normal	0.083	0.250	0.083	0.583	0.417	0.250	0.750	0.083	0.083	0.083	0.750	0.250
T	0.083	0.083	0.250	0.750	0.417	0.083	0.583	0.083	0.417	0.083	0.417	0.083
Clayton	0.417	0.750	0.250	0.750	0.250	0.917	0.583	0.083	0.417	0.417	0.250	0.250
Frank	0.083	0.250	0.250	0.750	0.750	0.250	0.583	0.083	0.250	0.250	0.583	0.083

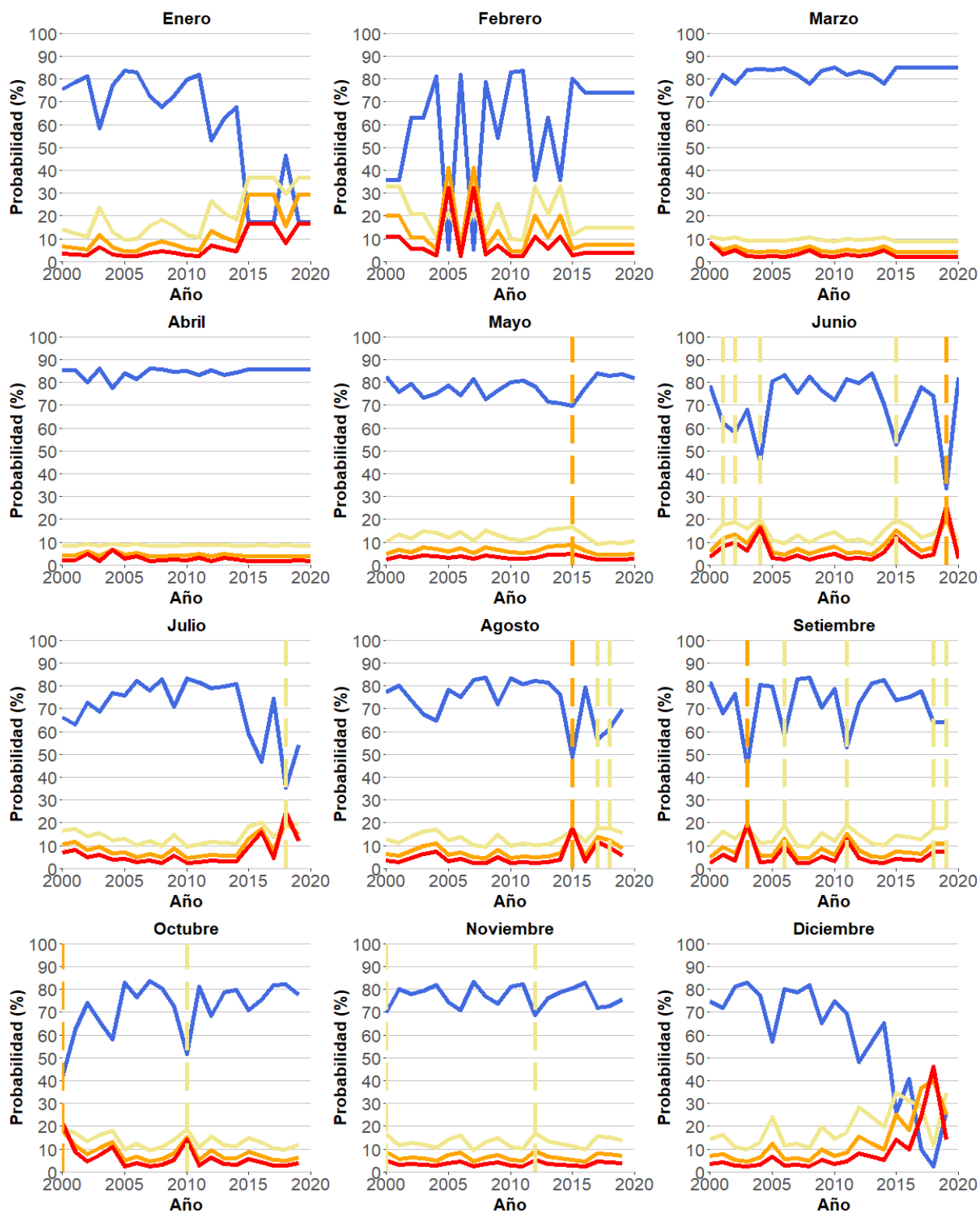
Valor estadístico S												
Copula	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC
Normal	0.069	0.042	0.038	0.025	0.028	0.029	0.025	0.053	0.028	0.040	0.026	0.043
T	0.067	0.041	<u>0.034</u>	<u>0.025</u>	0.027	0.028	0.024	<u>0.049</u>	<u>0.025</u>	0.039	0.025	0.042
Clayton	<u>0.033</u>	<u>0.022</u>	0.037	0.025	0.030	<u>0.023</u>	<u>0.021</u>	0.056	0.029	<u>0.031</u>	0.028	<u>0.028</u>
Frank	0.056	0.035	0.038	0.025	<u>0.025</u>	0.030	0.027	0.062	0.032	0.042	<u>0.025</u>	0.041

Parámetros de las cópulas												
Copula	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SET	OCT	NOV	DIC
Normal	0.791	0.754	0.029	-0.066	0.333	0.382	0.369	0.392	0.442	0.322	0.352	0.748
T	0.791	0.754	<u>0.029</u>	<u>-0.066</u>	0.333	0.382	0.369	<u>0.392</u>	<u>0.442</u>	0.322	0.352	0.748
Clayton	<u>2.770</u>	<u>2.386</u>	0.038	-0.080	0.551	<u>0.664</u>	<u>0.633</u>	0.690	0.824	<u>0.527</u>	0.594	<u>2.333</u>
Frank	7.437	6.598	0.166	-0.377	<u>2.021</u>	2.365	2.270	2.442	2.824	1.946	<u>2.155</u>	6.482

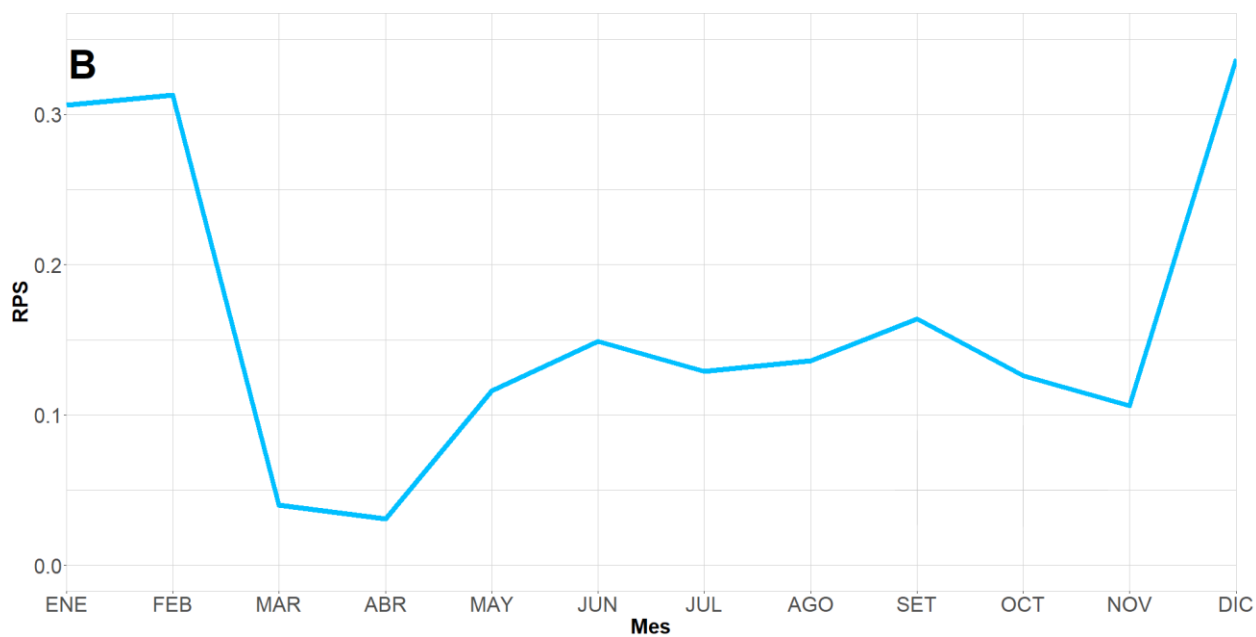
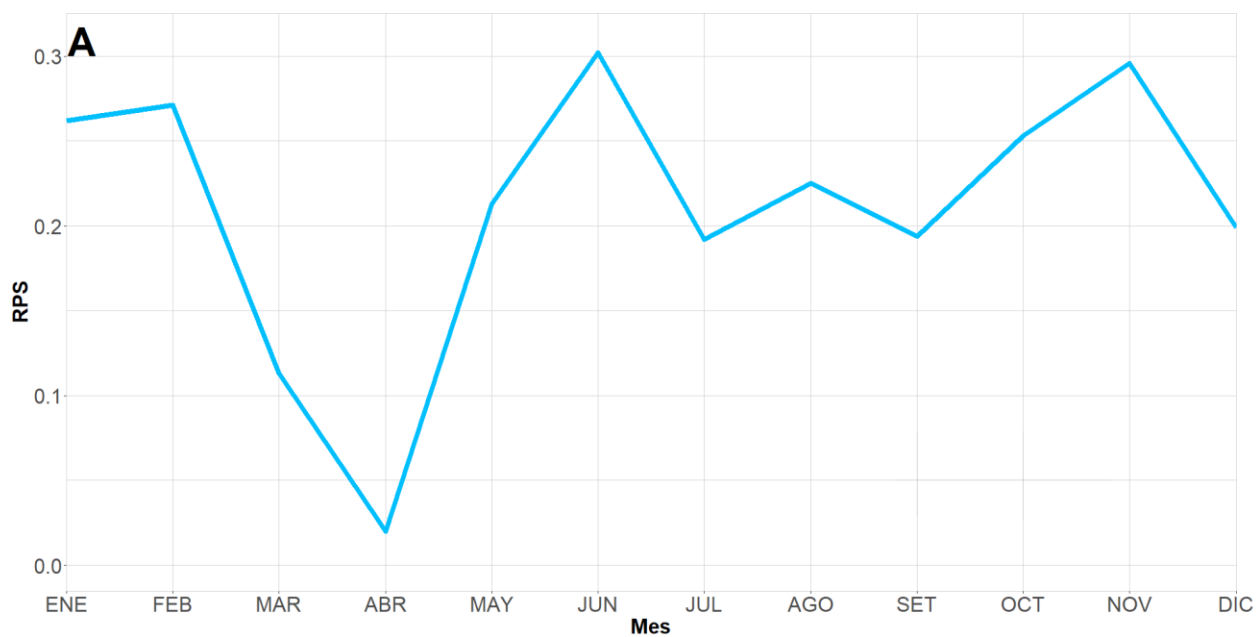
Anexo 12. Probabilidades del modelo de Redes Bayesianas de segundo orden, para todos los meses del 2000 al 2020, en la Estación Cañas Dulces.



Anexo 13. Probabilidades del modelo de Redes Bayesianas de segundo orden, para todos los meses del 2000 al 2020, en la Estación Talolinga.



Anexo 14. *Ranked Probability Score (RPS)* del modelo Redes Bayesianas de primer orden, para las cuatro categorías en las estaciones Cañas Dulces (A) y Talolinga (B).



Anexo 15. *Ranked Probability Skill Score (RPSS)* del modelo Redes Bayesianas de primer orden, para las cuatro categorías en las estaciones Cañas Dulces (A) y Talolinga (B).

