



Universidad Nacional Autónoma de Honduras



Facultad de Ciencias

**“Modelos Lineales Dinámicos para el análisis  
de valores extremos”**

Tesis para obtener el grado de:

**Maestría en Matemática con orientación en  
Estadística Matemática**

Presenta:

**Elvis Rafael Arrazola Acosta**

Asesor de Tesis:

**Ph.D. Cristian Andrés Cruz Torres**

Co-Asesor de Tesis:

**Ms.C. Asael Alonzo Matamoros**

Tegucigalpa, M.D.C

Diciembre 2023

# Agradecimientos

Quiero agradecer primeramente a Dios por darme la vida, por guiarme e iluminar para poder concluir mi tesis, por darme fortaleza y la calma en los momentos complejos del camino.

A mis padres Miguel Arrazola y Regina Acosta por apoyarme siempre en mi vida y seguirme impulsando a seguir desarrollándome como profesional. Gracias por su compañía y por estar luchando para que logremos juntos este objetivo.

A mis hermanos Reny y Antonio por su apoyo, por luchar desde el otro lado del mundo con su trabajo y esfuerzo para que pueda ser mejor profesional.

A mis familiares, en especial mi abuela Mariana Moncada gracias por su cariño, consejos y oraciones que son motor y fortaleza para no rendirse, a usted ejemplo de lucha un abrazo.

A mi esposa Rubenia Zúniga, por estar siempre vigilante y luchando con cada una de las peleas a mi lado. Por no rendirse y seguir creyendo que se puede aún a pesar de las dificultades, gracias por estar siempre a mi lado un beso.

A mis amigos, hermanos de estudio, de profesión que siempre estuvieron allí exigiendo y motivando para poder culminar este proceso, a ustedes todos sin excepción gracias por siempre escuchar y ayudar cuando más se necesitaba.

Quiero agradecer a mi Asesor PhD. Cristian Cruz y a mi Co-Asesor Asael Alonzo por su dedicación, compromiso y paciencia a lo largo de este trabajo. Gracias a su esfuerzo he podido comprender muchos elementos en la dinámica de la estadística matemática que sin su apoyo serían aún más complejos de desarrollar. Gracias por confiar y por su tiempo orientando cada dificultad y re-orientando cada aspecto de este trabajo. Gracias por ser parte de este logro.

# Índice general

<b>Agradecimiento</b>	<b>II</b>
<b>1. Introducción</b>	<b>3</b>
<b>2. Marco Teórico</b>	<b>6</b>
2.1. Teoría de valores extremos univariados . . . . .	6
2.1.1. Modelo asintótico y teorema para valores extremos . . . . .	6
2.1.2. Distribución generalizada de valores extremos . . . . .	7
2.2. Inferencia para valores extremos . . . . .	9
2.2.1. Verosimilitud para la distribución GEV . . . . .	10
2.2.2. Estimación de parámetros para valores extremos . . . . .	10
2.3. Modelos Lineales Dinámicos . . . . .	15
2.3.1. Estimación y Predicción de Modelos en Espacios Estado . . . . .	17
2.4. Forward Filtering Backward Sampling (FFBS) . . . . .	19
2.5. Estadística espacial . . . . .	21
2.6. Modelado de espacio-tiempo . . . . .	23
2.6.1. Modelado espacial . . . . .	25
2.6.2. Modelos dinámicos para datos espacio-temporales . . . . .	26
<b>3. Modelos espacio temporales de valores extremos</b>	<b>28</b>
3.1. Modelos GEV-DLM . . . . .	29
3.1.1. Modelo de Huerta and Sansó (2007) . . . . .	30
3.1.2. Modelo propuesto . . . . .	32
3.2. Selección del Método MCMC . . . . .	33
3.3. Caso Simulación . . . . .	36
3.3.1. Recuperación de los parámetros . . . . .	36
3.4. Análisis de las precipitaciones en Washington . . . . .	40
<b>4. Conclusiones y Trabajos Futuros</b>	<b>47</b>
4.1. Conclusiones . . . . .	47
4.2. Trabajos Futuros . . . . .	48
<b>Bibliografía</b>	<b>49</b>
<b>Apéndices</b>	<b>54</b>

# Índice de cuadros

3.1. Tamaños de muestra efectivos (ESS) para cada parámetro utilizando los algoritmos M-H y HMC. . . . .	34
3.2. Estimación de los parámetros $\sigma$ y $k$ . En la segunda columna se presenta la media de los parámetros, en la segunda se determina la mediana. La tercera columna muestra la desviación estándar, la cuarta columna la desviación media absoluta. En las sexta y séptima columnas representamos el intervalo de credibilidad mediante el cuartil 5 y el cuartil 95, respectivamente. La séptima columna indica la convergencia en las cadenas. Las columnas novena y décima muestran los tamaños de muestras efectivos para la distribución a posteriori del modelo vía Metrópolis-Hasting. . . . .	38
3.3. Estaciones Meteorológicas del estado de Washington U.S. Con sus respectivas latitudes y longitudes. . . . .	42
3.4. Estimación de los parámetros $\sigma$ para las 18 estaciones meteorológicas. En la segunda columna se presenta la media de los parámetros, en la segunda se determina la mediana. La tercera columna muestra la desviación estándar, la cuarta columna la desviación media absoluta. En las sexta y séptima columnas representamos el intervalo de credibilidad mediante el cuartil 5 y el cuartil 95, respectivamente. La séptima columna indica la convergencia en las cadenas. Las columnas novena y décima muestran los tamaños de muestras efectivos para la distribución a posteriori del modelo. . . . .	43
3.5. Estimación de los parámetros $k$ para las 18 estaciones meteorológicas. En la segunda columna se presenta la media de los parámetros, en la segunda se determina la mediana. La tercera columna muestra la desviación estándar, la cuarta columna la desviación media absoluta. En las sexta y séptima columnas representamos el intervalo de credibilidad mediante el cuartil 5 y el cuartil 95, respectivamente. La octava columna indica la convergencia en las cadenas. Las columnas novena y décima muestran los tamaños de muestras efectivos para la distribución a posteriori del modelo. . . . .	44
3.6. Niveles de retorno en un periodo de 5 a 25 años. . . . .	46

# Índice de figuras

2.1. Estructura de dependencia para un modelo de espacio de estado, (Lindsten, 2013). . . . .	15
3.1. Gráfico de auto-correlación ACF para los parámetros de escala sigma (primera columna) y forma k (segunda columna). La primera fila representa los ACF de las cadenas obtenidas por el algoritmo Metropolis Hasting, y la segunda fila representa los ACF de las cadenas obtenidas por el Hamiltonian Monte Carlo.	35
3.2. La primera columna muestra las simulaciones de las locaciones obtenidas a partir del Filtro de Kalman. La segunda columna muestra los datos simulados ( $y$ ) a partir del modelo y los parámetros de escala ( $\sigma$ ) y forma ( $k$ ). . . . .	37
3.3. La primera columna presenta la densidad de los parámetros de escala $\sigma$ y forma $k$ , respectivamente. En la segunda columna se grafican las cadenas correspondientes a los parámetros de escala ( $\sigma$ ) y forma $k$ . . . . .	38
3.4. Validación del modelo utilizando datos reales ( $y$ ), representados en color negro, y datos de entrenamiento para el modelo ( $y_{rep}$ ), que se muestran mediante intervalos en color amarillo con líneas verdes. . . . .	39
3.5. Validación del modelo mediante el uso de datos reales ( $y$ ), representados en color negro, y datos de entrenamiento para el modelo ( $y_{rep}$ ), los cuales se presentan mediante intervalos en color amarillo con líneas naranjas, calculados a través del algoritmo de Huerta. . . . .	39
3.6. Gráfico estado de Washington USA con sus respectivas estaciones meteorológicas.	41
3.7. Validación del modelo espacio-temporal mediante el uso de datos reales ( $y$ ), representados en color azul oscuro, y datos de entrenamiento para el modelo ( $y_{rep}$ ), los cuales se presentan mediante intervalos en color azul claro. . . . .	45
A.1. Funciones de densidad para las 18 estaciones para el parámetro de escala. . .	54
A.2. Funciones de densidad para las 18 estaciones para el parámetro de forma. . .	55
A.3. Cadenas para las 18 estaciones para el parámetro de escala. . . . .	56
A.4. Cadenas para las 18 estaciones para el parámetro de forma. . . . .	57

# Capítulo 1

## Introducción

Cuando se hace referencia a fenómenos extremos, estamos aludiendo a eventos poco comunes cuya ocurrencia, debido a su naturaleza inusual, puede acarrear consecuencias ambientales, económicas y humanas devastadoras. Este trabajo presenta el desarrollo de un modelo híbrido que fusiona los modelos lineales dinámicos con la teoría de valores extremos, con un enfoque particular en su implementación computacional. El objetivo principal radica en la estimación de parámetros y, sobre todo, en la recuperación de la estructura temporal de los datos. Para validar la eficacia del modelo propuesto, se realiza un proceso de simulación de datos. Es importante destacar que este modelo es altamente versátil y puede ser aplicado en el análisis computacional de una amplia gama de fenómenos extremos, como el análisis de riesgos financieros, la detección de anomalías climáticas y la evaluación de la seguridad de infraestructuras, entre otros.

En el contexto de la investigación de eventos extremos se vincula en este documento los niveles de precipitación, como medida de la aplicación del modelo computacional, es fundamental resaltar que la recopilación de datos se realiza a partir de múltiples estaciones meteorológicas ubicadas en la región de análisis. Estas estaciones, influenciadas por factores espaciales como la longitud y la latitud, así como elementos adicionales como la altitud, constituyen puntos notables en la obtención de información relevante sobre fenómenos de precipitaciones.

La metodología propuesta adopta un enfoque integral para modelar valores extremos que varían tanto en el tiempo como en el espacio. Inicialmente, se parte del supuesto de que las observaciones obtenidas mediante los datos siguen una distribución de Valor Extremo Generalizado (GEV), donde los parámetros de ubicación de esta distribución desempeñan un papel crucial en la construcción de la estructura espacio-temporal de los fenómenos. El componente temporal se conceptualiza mediante un Modelo Lineal Dinámico (DLM) o una representación de espacio de estados, basándose en el trabajo previo de Huerta et al. (2004). Para el componente espacial, se introduce a través de la matriz de evolución del DLM, implementando mediante un núcleo que es lo que define una forma de convolución de proceso Higdon (2004).

Se destaca la utilización de una simulación de Markov Chain Monte Carlo (MCMC) para generar estimaciones temporales y espaciales de nuestro modelo. Este enfoque proporciona una herramienta robusta para analizar la dinámica espacio-temporal de los eventos extremos,

permitiendo una comprensión más profunda de su comportamiento y facilitando la toma de decisiones informadas en contextos como la gestión del riesgo y la planificación del uso de la tierra. La propuesta, en su conjunto, constituye un aporte valioso para la comprensión y modelado de fenómenos climáticos extremos en el marco de la investigación meteorológica y climática.

Este trabajo se compone de 4 capítulos y 3 apéndices. En primer lugar, el Capítulo 2 aborda el marco teórico iniciando con la teoría probabilística relacionada con eventos inusuales, conocida como la Teoría de los Valores Extremos. Su objetivo principal es desarrollar métodos y modelos matemáticos capaces de describir y prever la ocurrencia de fenómenos raros. Inicialmente, la teoría de valores extremos fue introducido en la hidrología y meteorología para estudiar los niveles de inundación y desastres naturales. Sin embargo, en los últimos años, esta metodología estadística ha encontrado aplicaciones en diversas disciplinas y ciencias aplicadas, tales como finanzas, predicción del tráfico, seguros e ingeniería estructural. En este capítulo, se presentan los algoritmos de Metropolis-Hastings 1 y Hamiltoniano-Monte Carlo 2, los cuales serán cruciales para la estimación de parámetros de la distribución propuesta por la teoría de valores extremos.

En la Sección 2.3, se ofrece una introducción preliminar a los modelos de espacio estado, que sirven como base para los modelos lineales dinámicos (DLM). Los modelos lineales dinámicos (DLM) ofrecen una mayor flexibilidad en el tratamiento de series temporales no estacionarias, una interpretación más accesible en comparación con los modelos auto-regresivos de medias móviles (ARIMA) y una mayor versatilidad en su implementación. Además, se presenta el algoritmo FFBS 3, el cual es un filtro utilizado para resolver el modelo espacio-temporal.

En la sección 2.5, se presenta una descripción de la Estadística Espacial, con un enfoque especial en la rama de la Geoestadística, dada la naturaleza de nuestros datos. Esta disciplina nos proporciona herramientas fundamentales para el análisis del proceso espacial latente. Sin embargo, optaremos por emplear un enfoque basado en procesos de convolución Higdon (2004) para abordar su estudio de manera más eficaz y versátil.

En la Capítulo 3, se presenta inicialmente la estructura temporal del modelo combinándolo con la distribución generalizada de valores extremos (GEV-DLM). Posteriormente, se amplía para definir la estructura espacio-temporal del modelo, abordando así el desafío de modelar los DLMs para valores extremos que operan en espacio y tiempo. Se proporciona una descripción detallada del algoritmo que resuelve el problema espacio-temporal, mediante la implantación de un modelo propio que toma como referencia los algoritmos propuestos por Huerta and Sansó (2007) y Nakajima et al. (2012).

En la sección 3.2, se presentan los resultados obtenidos mediante el uso de algoritmos MCMC basados en 1 y 2, con el objetivo de determinar cuál es más apropiado para abordar el problema bajo estudio. Este análisis se inicia con una exploración de los algoritmos, evaluando su idoneidad y eficiencia en el contexto específico de la investigación.

En la sección 3.3 se realiza una simulación de datos con el objetivo principal de validar el método descrito en este capítulo. Esta fase de simulación no solo busca confirmar la robustez del modelo propuesto, sino también explorar su capacidad predictiva y comprender su comportamiento bajo distintas condiciones simuladas.

El énfasis recae en la recuperación de la estructura temporal del modelo, desglosando

minuciosamente cada componente para evaluar su eficacia en la captura de las complejidades inherentes a los fenómenos analizados. Se utiliza un enfoque detallado para analizar cómo el modelo se adapta a las variaciones temporales y espaciales, proporcionando una comprensión profunda de su desempeño en condiciones dinámicas.

En la sección 3.4 se muestra una aplicación del modelo propuesto, utilizando un conjunto de datos que consiste en mediciones de precipitaciones recopiladas en el estado de Washington durante las últimas dos décadas. El objetivo es no solo demostrar la aplicabilidad del modelo propuesto, sino también obtener percepciones valiosas sobre el comportamiento de los fenómenos extremos en esta región específica. En esta sección se incluyen además, información detallada en la tabla 3.3 la cual contiene los nombres, latitudes y longitudes de cada estación meteorológica utilizada en el estudio y se finaliza con la presentación de la tabla 3.6 que ofrece los niveles de retorno para las 18 estaciones meteorológicas estudiadas, consolidando así la información esencial derivada de la aplicación del modelo propuesto en diferentes ubicaciones geográficas. Este análisis detallado enriquece el conocimiento sobre el clima y la meteorología de la zona, proporcionando herramientas prácticas para la toma de decisiones informadas en gestión de riesgos y planificación territorial.

En el Capítulo 4, se presentan las conclusiones derivadas del modelo, enfatizando la recuperación de la componente temporal. Este capítulo no solo resume los resultados numéricos, sino que también proporciona un análisis de las implicaciones prácticas y teóricas de las conclusiones alcanzadas, contribuyendo así al entendimiento global del modelo propuesto.

Posteriormente, en el Apéndice A, se ofrecen detalles adicionales que enriquecen la comprensión del trabajo. Se presentan los gráficos de las funciones de densidad y trazas de cadenas para los parámetros de forma y escala, brindando una visión más completa de la variabilidad y comportamiento de estos parámetros.

Este apéndice no solo ofrecen detalles técnicos adicionales, sino que también sirven como recursos valiosos para investigadores y profesionales interesados en profundizar en la metodología y en los resultados específicos obtenidos en este estudio. El enfoque detallado de este apéndice complementa y respalda las conclusiones presentadas en el capítulo principal, proporcionando una visión más completa y detallada del trabajo realizado.

# Capítulo 2

## Marco Teórico

### 2.1. Teoría de valores extremos univariados

El objetivo principal de este capítulo es explorar la teoría clásica de valores extremos EVT<sup>1</sup> como un enfoque para abordar la problemática planteada. Aunque el problema puede extenderse hacia un análisis multivariado, esta sección inicial se concentrará exclusivamente en la secuencia univariada de variables aleatorias independientes e idénticamente distribuidas (i.i.d). Para aquellos interesados en una perspectiva más abarcadora, el enfoque multivariado se discute detalladamente en Coles (2013).

Es fundamental destacar que el abordaje de este problema se realiza desde una perspectiva Bayesiana, la cual será explorada con mayor detalle en el transcurso de este capítulo.

#### 2.1.1. Modelo asintótico y teorema para valores extremos

El estudio de valores extremos se aborda de forma asintótica utilizando el análogo del teorema del límite central. Considere la secuencia  $X_1, X_2, \dots, X_n$  de variables aleatorias tal que son independientes e idénticamente distribuidas (i.i.d.), donde  $n \in \mathbb{Z}_+$  y  $X_i \in \mathbb{R}$  para todo  $i = 1, 2, 3, \dots, n$ . En este caso,  $\mathbb{R}$  y  $\mathbb{Z}_+$  representan el conjunto de números reales y enteros positivos, respectivamente. El propósito fundamental es explorar los máximos de la muestra, centrándonos especialmente en su comportamiento.

$$M_n = \max \{X_1, X_2, \dots, X_n\}.$$

Se supone que existe una función de distribución  $F(X)$ . La distribución de los máximos  $M_n$  se puede encontrar de la siguiente manera:

$$\begin{aligned} \mathbb{P}(M_n \leq z) &= \mathbb{P}(X_1 \leq z, X_2 \leq z, \dots, X_n \leq z), \\ &= \mathbb{P}(X_1 \leq z)\mathbb{P}(X_2 \leq z) \dots \mathbb{P}(X_n \leq z), \\ &= F^n(z). \end{aligned} \tag{2.1}$$

Esta forma de definir  $M_n$  en la práctica no es posible de utilizar, ya que se necesita conocer la distribución de  $F(X)$  para poder hacerlo. Sustituir por las estimaciones empíricas de  $F(X)$

---

<sup>1</sup>[1] (por sus siglas en inglés, *extreme value theory*)

en (2.1) es inviable, puesto que una leve variación conlleva a malas aproximaciones y generar errores significativos al estimar  $F^n(X)$ . Coles (2013) propuso estudiar el comportamiento asintótico de  $F^n(X)$ , encontrando familias adecuadas de modelos que puedan estimarse utilizando observaciones extremas. Definiendo  $z_+$  como el valor más pequeño de  $z$  para el cual  $F(z) = 1$ , entonces:

$$\lim_{n \rightarrow \infty} F^n(z) = 0, \quad \forall z < z_+.$$

Podemos resolver el problema normalizando  $M_n = \max \{X_1, X_2, \dots, X_n\}$  mediante una transformación lineal:

$$M_n^* = \frac{M_n - b_n}{a_n},$$

donde  $\{a_n > 0\}$  y  $\{b_n\}$  son secuencias en donde la ubicación y la escala de  $M_n^*$  se estabilizan a medida que  $n$  sigue creciendo. El teorema fundamental para la teoría de valores extremos busca caracterizar la distribución de los máximos mediante la familia de valores extremos generalizada (GEV, por sus siglas en inglés) Fisher and Tippett (1928); Gnedenko (1943).

**Teorema 1.** *Si existe una sucesión constante  $\{a_n > 0\}$  y  $\{b_n\}$  tal que:*

$$\mathbb{P} \left\{ \frac{M_n - b_n}{a_n} \leq z \right\} \rightarrow G(z), \quad n \rightarrow \infty.$$

Donde  $G(z)$  es una función de distribución no degenerada, entonces  $G$  pertenece a una de las siguientes familias:

$$\begin{aligned} I. \quad & G(z) = \exp \left[ - \exp \left( \frac{z - b}{a} \right) \right], \quad -\infty < z < \infty. \\ II. \quad & G(z) = \begin{cases} 0, & \text{si } z \leq b, \\ \exp \left[ - \left( \frac{z-b}{a} \right)^{-\xi} \right], & \text{si } z > b. \end{cases} \\ III. \quad & G(z) = \begin{cases} 1, & \text{si } z \geq b, \\ \exp \left\{ - \left( \frac{z-b}{a} \right)^{-\xi} \right\}, & \text{si } z < b. \end{cases} \end{aligned}$$

Fisher and Tippett (1928) menciona que existen tres familias, denominadas familias I, II y III, que se conocen como *distribuciones de valores extremos*. Estas son las distribuciones *Gumbel*, *Fréchet* y *Weibull*, respectivamente. Las tres familias están denotadas por tres parámetros, dos de ellos,  $a$  y  $b$ , corresponden a la *escala* y *ubicación*, respectivamente. El tercero, denominado  $\xi$ , corresponde a la *forma*, y con este parámetro se caracterizan las distribuciones de Fréchet y Weibull.

### 2.1.2. Distribución generalizada de valores extremos

En el Teorema 1, se describen las tres familias para tratar los valores extremos, las cuales difieren según el comportamiento de la cola de la distribución  $F(x)$ . Notemos que podemos describirlas simplemente señalando el estudio del comportamiento de  $G(z)$  cerca de su punto extremo superior  $z_+$ .

En la práctica, no es tan útil mantener esta formulación, ya que no es conveniente comprender qué familia de las tres utilizar antes de pasar a la estimación de parámetros. Sin

embargo, en Coles (2013), se presenta una reformulación de cada una de las distribuciones de  $G(z)$ , como se muestra mediante la siguiente distribución:

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right)^{\frac{-1}{\xi}} \right] \right\}, \quad (2.2)$$

definido en el conjunto  $\left\{ z : 1 + \xi \left( \frac{z - \mu}{\sigma} \right)^{\frac{-1}{\xi}} > 0 \right\}$  donde los parámetros  $\mu, \xi \in \mathbb{R}$ , y  $\sigma > 0$ . La ecuación (2.2) se conoce como distribución GEV <sup>2</sup> donde:

- $\mu$  parámetro de ubicación,
- $\sigma > 0$  parámetro de escala,
- $\xi$  parámetro de forma.

En los casos donde  $\xi > 0$  y  $\xi < 0$ , corresponderá a las familias Fréchet y Weibull, respectivamente. Para la distribución Gumbel, se debe observar el parámetro de forma, que en este caso es  $\xi = 0$ . Sin embargo, la distribución Gumbel se obtiene tomando el límite de (2.2) cuando  $\xi \rightarrow 0$  y tiene la siguiente forma:

$$G(z) = \exp \left[ - \exp \left\{ - \left( \frac{z - \mu}{\sigma} \right) \right\} \right].$$

Ahora se puede formular el Teorema 1 de la siguiente manera:

**Teorema 2** (Coles (2013)). *Si existen las sucesiones constantes  $\{a_n > 0\}$  y  $\{b_n\}$  tales que:*

$$\mathbb{P} \left\{ \frac{M_n - b_n}{a_n} \leq z \right\} \rightarrow G(z), \quad n \rightarrow \infty. \quad (2.3)$$

donde  $G$  es una distribución no degenerada, entonces  $G$  pertenece a una de las familias GEV:

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right)^{\frac{-1}{\xi}} \right] \right\},$$

definido en el conjunto  $\left\{ z : 1 + \xi \left( \frac{z - \mu}{\sigma} \right)^{\frac{-1}{\xi}} > 0 \right\}$  donde los parámetros satisfacen las condiciones:  $\mu, \xi \in \mathbb{R}$ , y  $\sigma > 0$ .

Con esto, se obtiene que no es necesario escoger una familia antes de estimar los parámetros; por lo tanto, la selección de un modelo estadístico se limita una estimación correcta de los parámetros. Además, con esto obtenemos que en el límite en (2.3), permite modelar la secuencia de máximos mediante la distribución GEV de la forma:

$$\mathbb{P} \left\{ \frac{M_n - b_n}{a_n} \leq z \right\} \approx G(z),$$

para  $n$  grande, entonces,

$$\mathbb{P} \{M_n \leq z\} \approx G \left( \frac{z - b}{a} \right) \approx G^*(z). \quad (2.4)$$

---

<sup>2</sup>[2] (por sus siglas en inglés, *generalized extreme values*)

Coles (2013) define que  $G^*$  es igual a una de las distribuciones  $F$  y  $F^*$ , donde son del mismo tipo si existen constantes  $a > 0$  y  $b$ , tales que  $F^*(ax + b) = F(x)$  para todo  $x \in \mathbb{R}$ .

Ahora, para modelar las descripciones mencionadas anteriormente, es necesario definir una buena estrategia que pueda analizar la secuencia de los extremos  $X_1, X_2, \dots$ , conocido como enfoque de bloques máximos. Los bloques de máximos consisten en dividir los valores extremos en bloques  $M_{n,1}, \dots, M_{n,m}$  con longitud de  $n$  relativamente grande. Normalmente, la elección de la longitud se hace considerando bloques anuales Coles (2013).

### El modelo de superación de umbrales

Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias i.i.d., con distribución  $F(x)$  y observamos los valores que excedan algún umbral  $u$ , estas excedencias pueden leerse como observaciones extremas para alguna serie de datos dada. Sea  $X$  un término arbitrario de la sucesión, la siguiente probabilidad condicional puede usarse para describir el comportamiento de eventos extremos;

$$\mathbb{P}\{X > u + y | X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0. \quad (2.5)$$

En muchas ocasiones la ecuación (2.5) es imposible de aplicar directamente, dado que  $F(x)$  es desconocida en la mayoría de los casos. Por lo tanto, sería deseable encontrar aproximaciones de la distribución principal de los excesos, que sean válidas para valores de umbral bastante grandes.

Un umbral demasiado alto dará lugar a un modelo caracterizado por una gran varianza, ya que solo se utilizarán unas pocas observaciones para ajustarlo, mientras que un valor demasiado pequeño para  $u$  producirá un sesgo, ya que es posible que ya no se cumpla la suposición asintótica detrás de la aproximación.

## 2.2. Inferencia para valores extremos

En las secciones anteriores, se ha proporcionado un resumen de los principales resultados teóricos del EVT. Dado que nuestra perspectiva es Bayesiana, en esta sección presentamos la inferencia de valores extremos y la evaluación de modelos.

Para llevar a cabo la inferencia, es crucial comprender la verosimilitud de los valores extremos, por lo tanto, debemos ajustar la distribución GEV. Supongamos que los datos  $x = (x_1, \dots, x_n)$  son independientes de una colección i.i.d. de una variable aleatoria  $X$ , donde la función de densidad pertenece a una familia paramétrica  $\{f(x|\theta) : \theta \in \Theta\}$ . La función de verosimilitud se define como:

$$L(\theta; x) = \prod_{i=1}^n f(x_i|\theta),$$

en este caso,  $f(x_i|\theta)$  es la función de densidad de probabilidad de  $X$ . El valor de  $\theta$  para el cual se maximiza  $L(x; \theta)$  se denomina estimador de máxima verosimilitud.

Sea  $l(x; \theta)$  la transformación logarítmica de  $L(x; \theta)$ , también llamada log-verosimilitud:

$$l(\theta; x) = \log L(x; \theta) = \sum_{i=1}^n \log f(x_i|\theta).$$

Se utiliza la función de log-verosimilitud, ya que a menudo resulta más fácil de manejar:  $l(\theta; x) = \log \{L(\theta; x)\}$ . La estimación de máxima verosimilitud  $\hat{\theta}$  es el valor en el que  $l(\theta; x)$  alcanza su máximo.

En un enfoque Bayesiano, se asume que, sin referencia a los datos, es posible formular creencias acerca de  $\theta$ , donde  $\theta \in \Theta$ . Mientras que, en un enfoque clásico, estas creencias son tratadas como frecuencias. Dado que el enfoque es enfoque Bayesiano, los parámetros que definen  $\theta$  se considerarán como variables aleatorias que tendrán una distribución de probabilidad con alguna información conocida.

Para la información inicial que se debe cuantificar, se prevé suponer una priori para  $\theta$ , donde  $f(\theta)$  representa la representación inicial subjetiva respecto al parámetro de interés.

### 2.2.1. Verosimilitud para la distribución GEV

Para profundizar en el análisis para la distribución GEV, se debe tener presente que la función de distribución GEV está dada por la ecuación (2.2), donde  $\theta = (\mu, \sigma, \xi)$  son los parámetros de ubicación, escala y forma, respectivamente. Si asumimos que los datos  $\mathbf{z} = (z_1, \dots, z_n)$  son una colección de variables aleatorias independientes e idénticamente distribuidas (i.i.d.) como GEV( $\theta$ ), la probabilidad logarítmica es:

$$l(\theta; \mathbf{z}) = -n \log(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log \left(1 + \frac{\xi(z_i - \mu)}{\sigma}\right) - \sum_{i=1}^n \left\{1 + \frac{\xi(z_i - \mu)}{\sigma}\right\}^{-\frac{1}{\xi}},$$

siempre que  $\left\{z : 1 + \frac{\xi(z_i - \mu)}{\sigma} > 0\right\}$  sea positivo para cada  $i = 1, \dots, n$ . Si alguno de estos términos no es positivo, la probabilidad es cero. El caso  $\xi = 0$  (tomando  $\xi \rightarrow 0$ ) define la distribución Gumbel. Debido a un argumento asintótico, este modelo se usa a menudo cuando los datos  $x$  consisten en máximos (o mínimos negativos que son valores extremadamente bajos) de algún proceso subyacente Coles (2013).

### 2.2.2. Estimación de parámetros para valores extremos

En esta sección, la relevancia radica en determinar los métodos de simulación que se usarán para la estimación de parámetros de la distribución GEV. Cabe destacar que esta es una distribución no lineal y no Gaussiana, amplificando el problema ya que métodos clásicos de estimación son poco efectivos y cuantifican pobremente la incertidumbre. En este caso, es importante denotar  $z$  como las observaciones del problema, con función de densidad  $L(\theta; z)$ , donde  $\theta$  son los parámetros de la ecuación (2.2), y  $f(\theta)$  es la función a priori definida por el investigador. En este enfoque estamos interesados en la función a posterior de los parámetros, que por el teorema de Bayes se expresa mediante la siguiente ecuación:

$$f(\theta|z) \propto L(\theta; z)f(\theta).$$

La posterior  $f(\theta|x)$  actualiza la incertidumbre de los parámetros, una vez la muestra ha sido analizada y tomada en consideración. El mayor problema es que el computo de la posterior no siempre es factible de forma analítica y métodos de simulación son necesarios para obtener una aproximación a la incertidumbre final del parámetro.

## Métodos de Markov Chain Monte Carlo

Los métodos de Markov Chain Monte Carlo (MCMC) son técnicas estadísticas utilizadas para aproximar distribuciones de probabilidad y realizar inferencias en problemas complejos. MCMC utiliza cadenas de Markov para muestrear de manera efectiva distribuciones de probabilidad. La idea principal es generar secuencias de muestras que convergen a la distribución de interés.

Para abordar el problema de inferencia para la distribución GEV, es crucial caracterizar los algoritmos utilizados en la simulación del muestreo. Entre los diversos algoritmos de MCMC disponibles, destacamos dos en particular que se emplearán en nuestro análisis:

- **Metropolis-Hastings (MH):** Introduce una propuesta de transición y acepta o rechaza según una regla de aceptación.
- **Hamiltonian Monte Carlo (HMC):** Utiliza principios de la mecánica de Hamilton para mejorar la eficiencia del muestreo.

Aunque existen otros algoritmos, nos centraremos en el M-H y el HMC para los propósitos de este análisis. Para obtener información detallada sobre otros métodos, se recomienda consultar Gamerman and Lopes (2006).

### Metropolis-Hastings

El algoritmo de Metropolis-Hastings es una técnica de MCMC que se emplea para generar muestras de una distribución de probabilidad, es útil cuando no es posible extraer muestras directas de dicha distribución. La propuesta de este algoritmo fue desarrollada por Metropolis et al. (1953) y fue extendido posteriormente por Hastings (1970).

Para este algoritmo, se requieren dos elementos fundamentales: una distribución propuesta con alguna densidad, denotada como  $q(\theta^*|\theta)$ , y una probabilidad de aceptación  $u(\theta^*|\theta)$ . El algoritmo sigue los siguientes pasos:

---

#### Algorithm 1 Metropolis-Hastings

---

1. Inicializar la cadena con  $\theta^0$ . establecer  $j = 1$ .
2. Proponer  $\theta^* \sim q(\theta^*|\theta^{(j-1)})$ .
3. Calcule la probabilidad de aceptación  $\alpha(\theta^*|\theta^{(j-1)})$  del movimiento propuesto, donde:

$$\begin{aligned} \alpha(\theta^*|\theta^{(j-1)}) &= \min \{ u(\theta^*|\theta^{(j-1)}), 1 \}, \\ &= \min \left\{ \frac{f(\theta^*|y_{1:n})q(\theta^{(j-1)}|\theta^*)}{f(\theta^{(j-1)}|y_{1:n})q(\theta^*|\theta^{(j-1)})}, 1 \right\}, \end{aligned}$$

4. Con probabilidad  $\alpha(\theta^*|\theta^{(j-1)})$ , establecer  $\theta^j = \theta^*$ , de otra forma establecer  $\theta^j = \theta^{(j-1)}$ .
  5. Establecer  $j = j + 1$ , regresar al paso 2.
- 

El M-H simula una cadena de Markov cuya distribución estacionaria coincide con la distribución de interés, para ello se asume que los movimientos están definidos por:

$$f_i(\theta|\tilde{\theta}, y_{1:n}) = q(\theta|\tilde{\theta})\alpha(\theta|\tilde{\theta}).$$

Entonces, se puede probar que:

$$f(\theta|\tilde{\theta}, y_{1:n})f(\tilde{\theta}|y_{1:n}) = f_i(\tilde{\phi}|\phi, y_{1:n})f(\theta|y_{1:n}),$$

Donde la ecuación anterior se conoce como la ecuación de equilibrio. En el algoritmo, es crucial garantizar su convergencia, para lo cual Roberts and Smith (1994) y Tierney (1994) proponen las mejores estrategias de ajuste. La probabilidad de aceptación ser por:

$$u(\theta^*|\theta^{(j-1)}) = f(\theta^*)f(y_{1:n}|\theta^*)q(\theta^{(j-1)}|\theta^*)f(\theta^{(j-1)})f(y_{1:n}|\theta^{(j-1)})q(\theta^*|\theta^{(j-1)}),$$

Si suponemos que las componentes de  $\theta$  son independientes se obtiene:

$$u(\theta^*|\theta^{(j-1)}) = f(y_{1:n}|\theta^*)q(\theta^{(j-1)}|\theta^*) \prod_{i=1}^{n_{par}} f(\theta_i^*)f(y_{1:n}|\theta^{(j-1)})q(\theta^*|\theta^{(j-1)}) \prod_{i=1}^{n_{par}} f(\theta_i^{(j-1)}).$$

Factorizando las ecuaciones  $f(y_{1:n}|\theta)$ , se obtiene que la probabilidad de aceptación será:

$$u(\theta^*|\theta^{(j-1)}) = \frac{q(\theta^{(j-1)}|\theta^*)}{q(\theta^*|\theta^{(j-1)})} \frac{f(y_1|\theta^*)}{f(y_1|\theta^{(j-1)})} \prod_{t=2}^n \frac{f(y_t|y_{1:t-1}, \theta^*)}{f(y_t|y_{1:t-1}, \theta^{(j-1)})} \prod_{i=1}^{n_{par}} \frac{f(\theta^*)}{f(\theta^{(j-1)})}.$$

A continuación, se presentan algunas opciones para definir la distribución propuesta.

## Cadenas simétricas

Si la distribución propuesta es simétrica, entonces  $q(\theta^{(j-1)}|\theta^*) = q(\theta^*|\theta^{(j-1)})$ . En este caso, la probabilidad de aceptación se simplifica como,

$$u(\theta^*|\theta^{(j-1)}) = \frac{f(\theta^*|y_{1:n})}{f(\theta^{(j-1)}|y_{1:n})}, \quad (2.6)$$

que solo requiere la distribución posterior (no normalizada) Lai (2019).

## Cadenas de paseo aleatorio

Frecuentemente se elige la distribución de propuesta de paseo aleatorio, mediante:

$$\theta^* = \theta^{(j-1)} + \epsilon, \epsilon \sim N(0, \Sigma).$$

Aquí es relevante destacar el término del efecto no medido  $\epsilon$ , el cual dependerá de la varianza  $\Sigma$ . Si  $\Sigma$  es pequeña, la cadena dará pasos pequeños, mientras que si  $\Sigma$  es grande, los pasos serán extensos, lo que podría resultar en la no aceptación de saltos muy lejanos. Por lo tanto, para elegir  $\Sigma$  de manera más apropiada, Roberts et al. (1997) y Roberts and Rosenthal (2001) proponen la siguiente fórmula:

$$\Sigma = \frac{2,382}{n_{par}} \text{Var}(\phi|y_{1:n}),$$

lo que conduce a una tasa de aceptación adecuada para algunas distribuciones. Los valores típicos de la tasa de aceptación suelen estar entre 0.1 y 0.4, lo que se considera razonable. No siempre se obtienen los valores deseados, ya que, en algunos casos, los parámetros deben ser positivos. En tales situaciones, se realiza una transformación logarítmica normal de la forma:

$$\log \theta^* = \log \theta^{(j-1)} + \tilde{\epsilon}, \quad \tilde{\epsilon} \sim N(0, \tilde{\Sigma}),$$

donde para  $\log\theta$  se tiene  $\tilde{\Sigma}$  es su estimación de la varianza posterior. La probabilidad de aceptación es:

$$u(\theta^*|\theta^{(j-1)}) = \frac{f(y_1|\theta^*)}{f(y_1|\theta^{(j-1)})} \prod_{t=2}^n \frac{f(y_t|y_{t-1}, \theta^*)}{f(y_t|y_{t-1}, \theta^{(j-1)})} \prod_{i=1}^{n_{par}} \frac{f(\theta_i^*)}{f(\theta_i^{(j-1)})} \prod_{i=1}^{n_{par}} \frac{\theta_i^*}{\theta_i^{(j-1)}}.$$

Con esta parte se concluye la relevancia del método de MH mediante el algoritmo 1, destacando que este aporte está bien definido en Lai (2019).

## Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) corresponde a una instancia del algoritmo Metropolis-Hastings, donde se tiene un comportamiento dinámico mediante un integrador de salto para que genere un nuevo punto en el espacio de estado.

Kennedy (1987) propuso el HMC con el nombre de Hybrid Monte Carlo. Se detalla el método en Brooks et al. (2011), sin embargo, es importante destacar su utilidad en el contexto de los modelos GEV Coles (2013).

Defina un vector de parámetros con dimensión  $d$ , que será de la forma  $\theta \in \mathbb{R}^d$ . La densidad posterior para  $\theta$  está definida por  $f(\theta)$ , y el vector auxiliar paramétrico  $p \in \mathbb{R}^d$  donde, este vector tiene una distribución de la forma  $p \sim N(0, M)$ . En este caso,  $\theta$  se interpreta como la posición de una partícula en un momento dado,  $-\log p(\theta)$  describe su energía potencial, mientras que  $p$  es el momento definido como energía cinética  $p^T M^{-1} p/2$ .

La función de Hamilton se define como un sistema cerrado está se define a continuación:

$$H(\theta, p) = -L(\theta) + p^T M^{-1} p/2,$$

donde  $L(\theta) = \log f(\theta)$ . La densidad conjunta de  $(\theta, p)$  está dada por:

$$f(\theta, p) \propto f(\theta) \exp(p^T M^{-1} p/2) \times \exp[-H(\theta, p)].$$

Para un tiempo continuo  $t$ , mediante las ecuaciones dinámicas de Hamilton se busca mantener constante energía total de la partícula:

$$\begin{aligned} \frac{\partial \theta}{\partial p} &= \frac{\partial(\theta, p)}{\partial p} = M^{-1} p \\ \frac{\partial p}{\partial t} &= -\frac{\partial(\theta, p)}{\partial \theta} = \nabla_{\theta} L(\theta), \end{aligned}$$

con  $\nabla_{\theta} L(\theta)$  definido como el gradiente de  $L(\theta)$  con respecto a  $\theta$ . Se busca lograr una exploración más eficiente con la introducción el gradiente y la variable auxiliar  $p$ . El problema se complica, ya que necesitamos resolver las ecuaciones de manera numérica, es decir, a través de métodos numéricos para ello se siguen los siguientes pasos:

$$\begin{aligned} p^{\tau+\frac{\epsilon}{2}} &= p^{\tau} + \frac{\epsilon}{2} \nabla_{\theta} L(\theta^{\tau}) \\ \theta^{\tau+\epsilon} &= \theta^{\tau} + \epsilon M^{-1} p^{\tau+\frac{\epsilon}{2}} \\ p^{\tau+\epsilon} &= p^{\tau+\frac{\epsilon}{2}} + \frac{\epsilon}{2} \nabla_{\theta} L(\theta^{\tau+\epsilon}), \end{aligned}$$

Defina un tamaño de paso pequeño para  $\epsilon > 0$ . Luego de un número de pasos de tiempo, esto da como resultado  $(\theta^*, p^*)$ . Esta expresión se la nueva propuesta que al igual que en el algoritmo de M-H, se deberá considerarse una probabilidad de aceptación y verificar la convergencia entonces se sigue:

$$\begin{aligned}\alpha[(\theta, p), (\theta^*, p^*)] &= \min \left[ \frac{f(\theta^*, p^*)}{f(\theta, p)}, 1 \right] \\ &= \min \left[ \exp[H(\theta, p) - H(\theta^*, p^*)], 1 \right].\end{aligned}$$

Donde  $M$  es una matriz de masa simétrica definida positiva, generalmente es diagonal con valores constantes, que se vuelve simple cuando  $M = 1$ .

---

**Algorithm 2** Hamiltonian Montecarlo
 

---

1. Dar una posición inicial  $\theta^{(0)}$  y establecer  $j = 1$ .
  2. Calcular  $p^* \sim N(0, I^d)$  y  $u \sim U(0, 1)$ .
  3. Establecer  $(\theta^{(j)}, p^{(j)}) = (\theta^{(j-1)}, p^*)$  y  $H^{(0)} = H(\theta^{(j)}, p^{(j)})$ .
  4. Repetir  $L$  veces la solución:
    - $p^* = p^* + \frac{\epsilon}{2} \nabla_{\theta} L(\theta^{(j-1)})$ .
    - $\theta^{(j-1)} = \theta^{(j-1)} + \epsilon p^*$ .
    - $p^* = p^* + \frac{\epsilon}{2} \nabla_{\theta} L(\theta^{(j-1)})$ .
  5. Establecer  $(\theta^{(L)}, p^{(L)}) = (\theta^{(j-1)}, p^*)$  y  $H^{(1)} = H(\theta^{(L)}, p^{(L)})$ .
  6. Calcular  $\alpha[(\theta^{(j)}, p^{(j)}), (\theta^{(L)}, p^{(L)})] = \min[\exp(H^{(0)} - H^{(1)}), 1]$ .
  7. Establecer  $\theta^{(j)} = \theta^{(L)}$  si  $\alpha[(\theta^{(j)}, p^{(j)}), (\theta^{(L)}, p^{(L)})] > u$  y caso contrario  $\theta^{(j)} = \theta^{(j)}$ .
  8. Establecer  $j = j + 1$  y regrese al paso 2 hasta la convergencia.
- 

El algoritmo necesita las densidades logarítmicas posteriores, y se busca que con este método las cadenas se vuelvan estacionarias rápidamente. Tener en cuenta que para el HMC se requiere encontrar los valores de  $\epsilon$  y  $L$ ; para ello, Hoffman and Gelman (2011) proponen el algoritmo No-U-Turn Sampler (NUTS), que adapta automáticamente la longitud de los pasos de muestreo evitando así la necesidad de especificar el número de pasos o iteraciones. Esto se logra al seguir una trayectoria que no tiene un punto de retorno ("No-U-Turn"), lo que significa que la cadena de Markov continúa explorando hasta que se detecta que retroceder no sería beneficioso.

## 2.3. Modelos Lineales Dinámicos

En el ámbito de las series temporales  $Y_{s:t}$  denota la secuencia de la variable aleatoria  $Y$  entre los periodos de tiempo  $s$  a  $t$ , incluyendo tanto el momento  $s$  como el momento  $t$ . De manera análoga,  $y_{s:t}$  representa sus posibles valores. Denote las funciones de densidad de probabilidad mediante  $f(\cdot)$ , donde el argumento indica la distribución especificada; por ejemplo,  $f(y)$  denota la densidad de  $y$ .

Para una comprensión más profunda de los elementos siguientes, se proporcionará una breve introducción a los modelos en espacio estado.

### Modelos de espacio y estados

Los procesos estocásticos son una familia de variables aleatorias indexada. Formalmente, se representa como  $\{Y_t \mid t \in T\}$ , donde  $T$  es el conjunto de índices y  $Y_t$  es la variable aleatoria asociada al índice  $t$  de  $T$ . En este contexto, una serie de tiempo se conforma por una secuencia de resultados observados  $(y_1, \dots, y_t)$  generados a partir de procesos estocásticos.

Un proceso estocástico, representado por  $\{Y_t\}$ , presenta su comportamiento generado por las colecciones finitas de la distribución conjunta  $F(Y_{i_1}, Y_{i_2}, Y_{i_3}, \dots, Y_{i_t})$ . La obtención de dicha distribución resulta una tarea desafiante sin la asunción de independencia o intercambiabilidad entre las variables. Sin embargo, estas suposiciones son difíciles de mantener, ya que implicarían que el factor tiempo carece de relevancia.

Naturalmente, se puede extender el vector  $Y_t$  a mayores dimensiones para incluir lugares o estaciones en una región geográfica.  $Y_t = (Y_{1,t}, \dots, Y_{s,t})$  denota  $s$  estaciones en alguna región en el momento  $t$ . Una serie de tiempo  $\{y_t, t \in 1, 2, \dots\}$ , es un vector observable de un proceso  $\{Y_t\}$ . Con el fin de realizar inferencias sobre la serie de tiempo, especialmente para predecir el próximo valor  $Y_{t+1}$  a partir de las observaciones previas  $(Y_1, \dots, Y_t)$ , es necesario especificar la ley de probabilidad del proceso  $Y_t$ . Esto implica definir la estructura de dependencia entre las variables de  $Y_t$ .

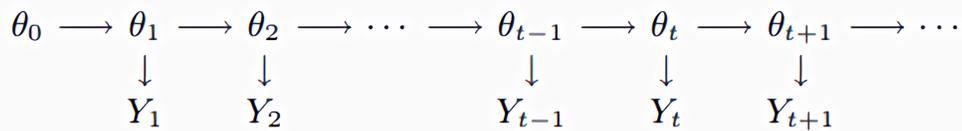


Figura 2.1: Estructura de dependencia para un modelo de espacio de estado, (Lindsten, 2013).

Los modelos de espacio y estados (SSM)<sup>3</sup> descansan en la premisa de que la serie temporal  $\{Y_t\}$  se configura como una representación incompleta y ruidosa de un proceso subyacente no observado  $\{\theta_t\}$ , identificado como el proceso de estados. En aplicaciones de ingeniería,  $\theta_t$  comúnmente describe el estado de un sistema físico que genera la salida  $Y_t$  junto con perturbaciones aleatorias. En términos generales, es posible concebir a  $\{\theta_t\}$  como un proceso aleatorio adicional que simplifica la tarea de definir la distribución de probabilidad de la serie

<sup>3</sup>[3](por sus siglas en inglés, *state space models*)

temporal. El proceso observable ( $Y_t$ ) está vinculado al proceso de estados latentes ( $\theta_t$ ), que sigue una dinámica Markoviana más simple. Se puede suponer de manera plausible que la observación  $Y_t$  solo está condicionada al estado del sistema en el momento de la medición, representado por  $\theta_t$ . La Figura 2.1 ilustra de manera gráfica la estructura de datos propuesta.

Las ecuaciones fundamentales de un SSM se expresan de la siguiente manera:

$$\begin{aligned} Y_t &= h_t(\theta_t, v_t), \\ \theta_t &= g_t(\theta_{t-1}, w_t); \end{aligned}$$

donde las funciones  $g_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$  y  $h_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$  funciones lineales continuas, para cualquier entero positivo  $d$ ,  $v_t$  y  $w_t$  son los efectos no medidos para  $h_t$  y  $g_t$  respectivamente. Los supuestos de un SSM son:

1.  $\{\theta_t\}$  es una cadena de Markov; es decir,  $\theta_t$  depende únicamente del estado anterior
2. Condicional a  $\{\theta_t\}$ , los  $Y_t$  son independientes entre sí y depende solamente de  $\theta_t$ . Se sigue que, para cualquier  $n \geq 1$ ,  $(Y_1, \dots, Y_n | \theta_1, \dots, \theta_n)$  tienen densidad condicional conjunta  $\prod_{t=1}^n f(y_t | \theta_t)$ .

con estos elementos el modelo queda completamente especificado mediante la distribución inicial o distribución a priori,  $p(\theta_0)$ , y las densidades condicionales  $p(\theta_t | \theta_{t-1})$  y  $p(y_t | \theta_t)$ . Esto es, que para todo  $t > 0$ , se cumple que:

$$\begin{aligned} (\theta_0, \theta_1, \dots, \theta_n, Y_0, Y_1, \dots, Y_n) &\sim p_0(\theta_0) \prod_{t=1}^n p(\theta_t, Y_t | (\theta_0, \theta_1, \dots, \theta_{t-1}, Y_0, Y_1, \dots, Y_{t-1})), \\ &= p_0(\theta_0) \prod_{t=1}^n f(Y_t | (\theta_0, \theta_1, \dots, \theta_{t-1}, Y_0, Y_1, \dots, Y_{t-1})), \\ &\times p(\theta_t | (\theta_0, \theta_1, \dots, \theta_{t-1}, Y_0, Y_1, \dots, Y_{t-1})), \\ &= p_0(\theta_0) \prod_{t=1}^n f(Y_t | \theta_t) p(\theta_t | \theta_{t-1}). \end{aligned} \tag{2.7}$$

Es notable que el proceso  $\{(\theta_t, Y_t)\}$  exhibe una propiedad fundamental de Markov. La densidad de  $(Y_1, \dots, Y_n)$  se obtiene mediante al integrar las variables  $\theta$  de la densidad conjunta, como se indica en la ecuación (2.7). Aunque las secciones subsiguientes demuestran que los cálculos son simples en el contexto del SSM lineal Gaussiano, la densidad de  $(Y_1, \dots, Y_n)$  generalmente no se encuentra en una forma cerrada, y el proceso observable ( $Y_t$ ) no posee la propiedad Markoviana de manera general. Sin embargo, destaca una propiedad significativa:  $Y_t$  exhibe independencia condicional respecto a las observaciones previas  $(Y_1, \dots, Y_{t-1})$  dada la condición de  $\theta_t$ . Este aspecto proporciona una interpretación del estado  $\theta_t$ , ya que representa información cuantitativa que resume la historia pasada del proceso observable y resulta suficiente para anticipar su comportamiento futuro.

## Modelos lineales dinámicos

Los modelos lineales dinámicos (DLMs)<sup>4</sup> emergen como una variante especial de los modelos de SSM, presentando propiedades particulares al ser tanto lineales como Gaussianos.

---

<sup>4</sup>[4] (por sus siglas en inglés, *Dynamic Linear Models*)

La aplicación del filtro de Kalman (Triantafyllopoulos, 2021) facilita la obtención recursiva de estimaciones y pronósticos en el contexto de los DLM. Las ecuaciones fundamentales de un DLM son:

$$\theta_0 \sim N_p(m_0, c_0), \quad (2.8)$$

junto con las ecuaciones para cada tiempo  $t \geq 1$ ,

$$Y_t = F_t \theta_t + v_t, \quad v_t \sim N_m(0, V_t), \quad (2.9)$$

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim N_p(0, W_t); \quad (2.10)$$

Suponiendo un escenario en el que  $Y_t$  adopta la estructura de un vector de dimensión  $m$ , mientras que  $\theta_t$  se presenta como un vector de dimensión  $p$ . Aquí, las matrices  $G_t$  y  $F_t$  juegan un papel crucial, siendo conocidas y con dimensiones  $p \times p$  y  $m \times p$  respectivamente. Las secuencias  $\{v_t\}$  y  $\{w_t\}$  exhiben independencia y siguen una distribución Gaussiana, con medias nulas y varianzas  $\{V_t\}$  y  $\{W_t\}$ , según Pole et al. (1994). Además, se asume la independencia entre  $\theta_0$  y  $\{v_t\}$ ,  $\{w_t\}$ . Se conocen las ecuaciones (2.9) y (2.10), como ecuaciones de observación y estados respectivamente. Se adoptará la notación propuesta por (West and Harrison, 1997). La cuádrupla DLM  $\{F_t, G_t, V_t, W_t\}$  representa el DLM propuesto por las ecuaciones (2.9) y (2.10), y es importante destacar que cuando las matrices en esta cuádrupla son constantes, se clasifica al DLM como invariable en el tiempo.

Una perspectiva interesante de los DLM es considerarlos como una generalización de los modelos de regresión lineal, ya que posibilitan la variación temporal de los coeficientes de regresión, según Pole et al. (1994).

### 2.3.1. Estimación y Predicción de Modelos en Espacios Estado

Suponga que se dispone de información hasta el tiempo  $t > 0$  (es decir,  $y_{1:t}$ ) y que el principal objetivo es realizar inferencias sobre el vector de estados. En este contexto, resulta crucial obtener la densidad condicional  $p(\theta_s | y_{1:t})$ . Para lograr este propósito, es fundamental distinguir entre tres situaciones: la *filtración* (cuando  $s = t$ ), el *pronóstico* (si  $s > t$ ) y el *suavizado* (si  $s < t$ ).

En el ámbito de la filtración, se entiende que la información se actualiza de manera continua con el objetivo de derivar la densidad condicional de  $\theta_t$  dado  $y_{1:t}$ . En este mismo contexto, el *Filtro de Kalman* (Kalman and Bertram, 1960), al simplificar la inferencia del vector de estado, facilita la actualización con la llegada de nuevos datos, transitando de  $p(\theta_t | y_{1:t})$  a  $p(\theta_{t+1} | y_{1:t+1})$  (Durbin and Koopman, 2012). La capacidad del filtro para aplicarse de forma recursiva destaca como una de sus características fundamentales, convirtiéndolo en una elección idónea para implementaciones computacionales Durbin and Koopman (2012).

El suavizado representa un progreso significativo en la teoría y aplicación de los modelos de espacio de estados, discutido en trabajos como Anderson and Moore (2005); Catlin (2011); de Jong (1989); Harvey (1991); Koopman (1993); Durbin and Koopman (2012). El método más utilizado es el suavizado de intervalos fijos, derivado por Rauch et al. (1965).

En el análisis de series temporales, el *pronóstico* asume usualmente el papel principal. La estimación del estado se convierte en un paso previo para predecir las observaciones futuras. En el contexto específico del pronóstico de un paso adelante, es decir, la predicción de la siguiente observación  $Y_{t+1}$  basándose en los datos  $y_1, \dots, y_t$ , se busca estimar el próximo valor

$\theta_{t+1}$  del vector de estado. Luego, utilizando estas estimaciones, calculamos el pronóstico para  $Y_{t+1}$ .

Es fundamental resaltar que la incertidumbre en los pronósticos se intensifica a medida que se avanza de manera temporal en  $t + k$  (similar a los pronósticos meteorológicos). No obstante, es posible cuantificar la incertidumbre mediante una densidad de probabilidad, específicamente la densidad predictiva de  $Y_{t+1}$  dada  $(Y_1, \dots, Y_t)$ .

## Filtro de Kalman

Aunque la forma de espacio de estados de un sistema variable en el tiempo era conocida, no estaba directamente enlazada al problema de filtrado, según Zadeh and Desoer (2008). Fue Kalman and Bertram (1960) quien sugirió la idea de utilizar variables de estado y el SSM lineal para abordar el problema de filtrado.

Inicialmente utilizado en procesos de tiempo discreto, el filtro de Kalman fue concebido por Kalman como el equivalente al filtro de Wiener. Este enfoque se adapta a la estructura de espacio de estado en tiempo discreto, como se detalla en Durbin and Koopman (2012). El filtro de Kalman se distingue por dos aspectos esenciales:

1. Se basa en la forma de espacio de estado, una elección que hoy en día se considera altamente beneficiosa para describir una amplia gama de fenómenos físicos.
2. Abandonó de la necesidad de suponer estacionariedad, un detalle significativo dado que la mayoría de los sistemas en la realidad no cumplen con esta condición.

El filtro de Kalman proporciona una solución óptima, siempre y cuando se conozcan y especifiquen el vector de diseño  $\theta_t$  y la matriz de transición  $F_t$ , junto con las varianzas de  $V_t$  y la matriz de covarianza de  $W_t$ . Aunque Kalman sugiere que estos componentes pueden ser conocidos por el sistema o derivados de experimentos anteriores, no aborda cómo especificarlos o estimarlos a partir de los datos. A pesar de la utilidad del filtro de Kalman, el cálculo de las densidades condicionales relevantes no es sencillo. En el caso de los DLMS, esta tarea se simplifica considerablemente gracias a las suposiciones de distribuciones normales Durbin and Koopman (2012).

Se establece la demostración de que el vector aleatorio  $(\theta_0, \theta_1, \dots, \theta_t, Y_1, \dots, Y_t)$  sigue una distribución normal para  $t \geq 1$ , gracias a la naturaleza Markoviana intrínseca del modelo. Este hallazgo implica que tanto las distribuciones marginales como las condicionales adoptan la forma de la distribución gaussiana. Dado que todas las distribuciones relevantes se modelan como normales, resulta suficiente calcular sus valores medios y covarianzas. La resolución del problema de filtrado se fundamenta en el teorema que se presenta a continuación.

**Teorema 3.** *Considere el DLM  $\{F_t, G_t, V_t, W_t\}$ , y suponga que*

$$\theta_{t-1} | y_{1:t} \sim N(m_{t-1}, C_{t-1}).$$

*Entonces se cumplen los siguientes enunciados.*

- i: La densidad predictiva del estado de un paso adelante de  $\theta_t$ , dado que  $y_{1:t-1}$  es normal, con parámetros.*

$$a_t = \mathbb{E}(\theta_t | y_{1:t-1}) = G_t m_{t-1},$$

$$R_t = \text{Var}(\theta_t | y_{1:t-1}) = G_t C_{t-1} G_t' + W_t.$$

ii: La densidad predictiva de un paso adelante de  $Y_t$  dado  $y_{1:t-1}$  es normal, con parámetros:

$$f_t = \mathbb{E}(Y_t|y_{1:t-1}) = F_t a_t,$$

$$Q_t = \text{Var}(Y_t|y_{1:t-1}) = F_t R_t F_t' + V_t.$$

iii: la distribución filtrada de  $\theta_t$  dado  $y_{1:t}$  es normal, con parámetros:

$$m_t = \mathbb{E}(\theta_t|y_{1:t}) = a_t + R_t F_t' Q_t^{-1} e_t,$$

$$C_t = \text{Var}(\theta_t|y_{1:t}) = R_t - R_t F_t' Q_t^{-1} F_t R_t.$$

donde  $e_t = Y_t - f_t$ , es el error de predicción.

El filtro de Kalman permite el cálculo recursivo de las densidades predictivas y de filtrado, comenzando con  $\theta_0 \sim N(m_0, C_0)$ , y luego calculando  $p(\theta_1|y_1)$ . Este proceso se repite de manera recursiva a medida que se incorpora nueva información de datos. La densidad condicional  $\theta_t|y_{1:t}$  aborda eficientemente el problema de filtrado.

Los resultados y discusiones proporcionados por West and Harrison (1997) son notoriamente similares, aunque se limitan a DLMs, facilitando así la comprensión de los modelos y la importancia de los resultados originalmente propuestos por Kalman.

## 2.4. Forward Filtering Backward Sampling (FFBS)

En el contexto del DLM univariado definido por  $DLM \{F_t, G_t, V_t, W_t\}$ , con la elección de varianzas constantes ( $V_t = V$  y  $W_t = W$ ), la generación de muestras  $\theta_{0:n} \sim p(\theta_t|y_{1:t-1})$  plantea ciertos desafíos. La solución a este problema se encuentra en el FFBS, un método de inferencia utilizado para llevar a cabo el filtrado y el muestreo hacia atrás en SSM, como en DLM. Este procedimiento ha sido detalladamente explicado en investigaciones previas, resaltando las contribuciones de Carter and Kohn (1994) y Frühwirth-Schnatter (1994).

De manera recursiva, se calcula el filtrado de las distribuciones para  $p(\theta_t|y_{1:n})$  en los instantes  $t = 1, \dots, n$ . Este proceso se realiza mediante un filtro directo, específicamente el filtro de Kalman and Bertram (1960). En el tiempo  $t = n$ , se lleva a cabo un muestreo de  $\theta_n$  desde la posterior  $p(\theta_t|y_{1:t-1})$ , la cual representa una marginal. Finalmente, se realiza un muestreo hacia atrás en los estados dinámicos.

El algoritmo FFBS está definido por las ecuaciones (2.9), (2.10), y el Teorema 3. Este teorema es fundamental para la eficiente generación del algoritmo, el cual expondremos en 3, para más detalles ver Lai (2019).

Iniciando con la distribución a priori del estado latente  $\theta$ , procedemos al cálculo del proceso latente mediante  $\theta_t|y_{1:t-1}$ . Es importante señalar que esta distribución sigue una forma normal con media  $G_t m_{t-1}$  y varianza  $R_t$ , donde  $R_t = G_t C_{t-1} C_{t-1}' + W$ . Seguidamente, calculamos la probabilidad marginal  $p(y_t|y_{1:t-1})$ , que también adopta una distribución normal con media  $F_t G_t m_{t-1}$  y varianza  $Q_t$ , siendo  $Q_t = F_t R_t F_t' + V$ .

Continuando con la posterior en  $t$ , donde  $\theta_t|y_{1:t}$  sigue una distribución normal con media  $m_t$  y varianza  $C_t$ , siendo:

$$m_t = G_t m_{t-1} + R_t F_t' (Q_t)^{-1} (y_t - F_t G_t m_{t-1}),$$

$$C_t = R_t - R_t F_t' (Q_t)^{-1} F_t R_t.$$

---

**Algorithm 3** Forward Filtering Backward Sampling (FFBS)
 

---

**Forward Filtering**

1. Distribución priori  $\theta_0 \sim N(m_0, C_0)$ . Almacene los valores de  $m_0$  y  $C_0$ .
2. Para  $t = 1, \dots, n$ 
  - a: Mediante la ecuación del sistema, se obtiene:

$$\theta_t | y_{1:t-1} \sim N(G_t m_{t-1}, R_t).$$

Almacenar  $R_t = G_t C_{t-1} G_t' + W$ .

- b: Predicción de un paso. Mediante la ecuación de observación, tenemos que:

$$y_t | y_{1:t-1} \sim N(F_t G_t m_{t-1}, Q_t).$$

Obtenga la probabilidad marginal:

$$p(y_t | y_{1:t-1}) = N(F_t G_t m_{t-1}, Q_t).$$

- c: La posterior en  $t$ :  $\theta_t | y_{1:t} \sim N(m_t, C_t)$ , donde,

$$m_t = G_t m_{t-1} + R_t F_t' (Q_t)^{-1} (y_t - F_t G_t m_{t-1}),$$

$$C_t = R_t - R_t F_t' (Q_t)^{-1} F_t R_t.$$

Almacene los valores  $m_t$  y  $C_t$ .

**Backward Sampling**

3. Muestrear  $\theta_n | y_{1:n} \sim N(m_n, C_n)$ .
4. Para  $t = n - 1, \dots, 0$ 
  - a: Distribución hacia atrás:  $\theta_t | \theta_{t+1}, y_{1:t} \sim N(h_t, H_t)$ , donde

$$h_t = m_t + C_t G_{t+1}' (G_{t+1} C_t G_{t+1}' + W)^{-1} (\theta_{t+1} - G_{t+1} m_t),$$

$$H_t = C_t - C_t G_{t+1}' (G_{t+1} C_t G_{t+1}' + w)^{-1} G_{t+1} C_t.$$

- b: Muestrear  $\theta_t | \theta_{t+1}, y_{1:t} \sim N(h_t, H_t)$ .

5. Devolver  $\theta_t$
-

En el contexto del FFBS, el proceso de filtrado tiene como objetivo estimar la distribución de los estados latentes dada la información observada hasta el instante  $t$ . Una vez completada esta fase, se procede al suavizado. En este paso, se realiza el muestreo de  $\theta_n|y_{1:n}$ , cuya distribución normal tiene como parámetros la media  $m_t$  y la varianza  $C_t$ , obtenidas en el paso c del algoritmo.

Luego, se implementa un remuestreo hacia atrás mediante la distribución  $\theta_t|\theta_{t+1}, y_{1:t}$ , que sigue una distribución normal con media  $h_t$  y varianza  $H_t$ , donde:

$$\begin{aligned} h_t &= m_t + C_t G'_{t+1} (G_{t+1} C_t G'_{t+1} + W)^{-1} (\theta_{t+1} - G_{t+1} m_t), \\ H_t &= C_t - C_t G'_{t+1} (G_{t+1} C_t G'_{t+1} + W)^{-1} G_{t+1} C_t. \end{aligned}$$

Finalmente, retorne el estado latente  $\theta_t$  en el tiempo  $t$ . Este proceso de muestreo de estados latentes retrospectivamente mejora las estimaciones de los estados anteriores dadas las observaciones.

## 2.5. Estadística espacial

En el análisis climático, es frecuente examinar datos provenientes de diversas ubicaciones o modelos dentro de una región de estudio específica. Este capítulo se enfoca en modelar extremos en puntos particulares, aunque en algunas instancias se exploran otras formas de ubicación, como las celdas de una cuadrícula mediante pesos espaciales. Para desarrollar los elementos necesarios, resulta fundamental apoyarse en los principios de la *geoestadística*.

### Geoestadística

En el contexto de la *geoestadística*, los datos pueden derivarse de ubicaciones puntuales dispuestas en una cuadrícula regular. Sin embargo, es común asumir que las estaciones de datos se distribuyen de manera no uniforme en el dominio espacial, como señala Koopman (1993).

La *geoestadística*, que se deriva de la estadística, se centra en fenómenos espaciales (Howarth, 1979). Su objetivo principal radica en la estimación, predicción y simulación de tales fenómenos (Warrick and Myers, 1987). Esta disciplina adapta las técnicas clásicas de regresión para describir la continuidad espacial, un atributo esencial en muchos fenómenos naturales (Isaaks and Srivastava, 2022).

El modelado espacial, una adición reciente en la literatura estadística, resulta fundamental para disciplinas como geología, ciencias del suelo, agronomía y otras que gestionan datos de diversas ubicaciones espaciales (Cressie, 1989). Este enfoque se dedica a la creación de modelos que identifiquen la dependencia entre medidas en distintos lugares, abordando aspectos como la predicción espacial, la simulación y el diseño muestral.

En el ámbito de la predicción espacial, la geoestadística desempeña un papel crucial en un proceso de dos fases. En la primera etapa, conocida como análisis estructural, se detalla la correlación entre puntos en el espacio. La segunda etapa consiste en realizar predicciones

en sitios no muestreados mediante la técnica de *kriging*. Esta técnica calcula un promedio ponderado de observaciones muestrales, ajustando los pesos según la estructura espacial y la configuración de muestreo (Petitgas, 1996).

## Variable Regionalizada

Una variable regionalizada es aquella medida en el espacio con una estructura de correlación. Formalmente, se define como un proceso estocástico en un espacio euclidiano  $\mathbb{R}^d$ ,  $\{Z(x) : x \in D \subset \mathbb{R}^d\}$ . Cuando  $d = 2$ ,  $Z(x)$  se relaciona con una variable medida en un punto  $x$  del plano. En la práctica,  $Z(x)$  puede representar la medición de una variable aleatoria (como la concentración de un contaminante) en un punto  $x$  de una región de estudio.

Sea  $\{Z(x) : x \in D \subset \mathbb{R}^d\}$  el proceso estocástico que define una variable regionalizada. Para  $n$  puntos  $x_1, x_2, \dots, x_n$ , el vector aleatorio  $\mathbf{Z}(\mathbf{x})$  tiene una función de distribución conjunta  $F[z_1, z_2, \dots, z_n]$ . Con densidades marginales conocidas, se pueden establecer la función media  $m(x_i)$ , la varianza  $\sigma^2(x_i)$ , la autocovarianza  $c(x_i, x_j)$  y la semivarianza  $\gamma(x_i, x_j)$ .

Una variable regionalizada es *estacionaria* si su función de distribución conjunta permanece invariante ante cualquier translación del vector  $h \in \mathbb{R}^d$ . En otras palabras, la función de distribución del vector aleatorio  $\mathbf{Z}(\mathbf{x})$  es idéntica a la del vector  $\mathbf{Z}(\mathbf{x} + \mathbf{h})$  para cualquier  $h > 0$ . En estadística espacial, un proceso estocástico  $\{Z(\mathbf{x}), \mathbf{x} \in D\}$  se considera estacionario de primer orden si cumple con  $E[Z(\mathbf{x})] = \mu$ , donde  $E$  denota el operador de esperanza matemática,  $Z(\mathbf{x})$  es el valor del proceso en la ubicación espacial  $\mathbf{x}$ , y  $\mu$  es una constante independiente de  $\mathbf{x}$ .

## Variograma y Semi-variograma

En el desarrollo de un análisis geoestadístico, es importante determinar la dependencia espacial de una variable. Llevar este proceso a cabo necesita, emplear tres funciones fundamentales basadas en la información muestral: el semivariograma, el covariograma y el correlograma. Veamos que propone Giraldo (2002) para estos elementos.

Para establecer la estacionariedad débil se asume que la varianza de los incrementos de la variable regionalizada es finita. Esta función, denotada por  $2\gamma(h)$ , recibe el nombre de variograma Segura (2018). Mediante la definición teórica de la varianza determinada por el valor esperado de una variable aleatoria, se expresa como:

$$2\gamma(h) = E[Z(x+h) - Z(x)]^2.$$

La semi-varianza,  $\gamma(h)$ , es la mitad del variograma y sirve para caracterizar las propiedades de dependencia espacial en un proceso. En la estimación de esta función a partir de una realización del fenómeno, se emplea el semi-variograma experimental. Su cálculo se realiza a través del método de momentos, conforme a la propuesta de Wackernagle (1998):

$$\hat{\gamma}(h) = \frac{E[Z(x+h) - Z(x)]^2}{2n},$$

donde  $Z(x)$  representa el valor de la variable en un punto  $x$ ,  $Z(x+h)$  indica otro valor muestral ubicado a una distancia  $h$ , y  $n$  es el número elementos pares que están separados la distancia  $h$ . La interpretación del semi-variograma es sencilla a distancias más cortas entre los puntos, hay una mayor similitud en los datos y a distancias mayores, la similitud de los datos es menor.

## Covariograma y correlograma

La función de covarianza muestral entre pares de observaciones separadas por una distancia  $h$  se calcula utilizando la fórmula clásica de covarianza muestral:

$$C(h) = \frac{\sum_{i=1}^n (Z(x+h))(Z(x))}{n} - m^2,$$

donde  $m$  es el promedio, y  $n$  nuevamente es el número elementos pares que están separados la distancia  $h$ . Este cálculo se conoce como covariograma. Si se estima la varianza y se asume estacionariedad, obtenemos el correlograma, definido como:

$$r(h) = \frac{C(h)}{C(0)}.$$

Con la condición de estacionariedad, las tres funciones de dependencia espacial: semivariograma, covariograma y correlograma, pueden ser utilizadas para determinar la relación espacial entre los datos. Comúnmente se prefiere el semivariograma, ya que no requiere la estimación de parámetros, a diferencia de las otras dos funciones que sí lo necesitan, lo que las hace más complejas. Para obtener más información sobre esta temática, consulte Giraldo (2002).

## 2.6. Modelado de espacio-tiempo

El propósito de esta sección es presentar la teoría necesaria para desarrollar un modelo capaz de manejar datos espacio-temporales en conjunto con los Modelos Lineales Dinámicos (DLMs). Para lograr esto, se emplean procesos de convolución y núcleos en un contexto espacio-temporal. La construcción de un modelo espacial continuo se lleva a cabo mediante la convolución de un proceso elemental con una función kernel o de dispersión de puntos. Esta estrategia ofrece diversas ventajas en comparación con la especificación basada en un covariograma espacial. En especial, la formulación de procesos de convolución no solo facilita los cálculos computacionales, sino que también se adapta con facilidad a modelos estacionarios más complejos Ver Hoef et al. (2004).

El modelado de datos espaciales mediante procesos Gaussianos representa un pilar fundamental en los análisis geostatísticos. Entre las fuentes de referencia más influyentes en este campo se encuentran trabajos como el de Matheron (1963), Howarth (1979), Cormack and Cressie (1991), Wackernagle (1998) y Stein (1999). En estos estudios, el enfoque más común se centra en la modelación de la dependencia espacial a través del covariograma  $C(h)$ . Bajo este enfoque, la covarianza entre dos puntos se determina exclusivamente por la distancia euclidiana entre ellos, aunque en ocasiones se emplean otras métricas.

Una estrategia alternativa y constructiva para generar un proceso Gaussiano sobre  $\mathbb{R}^d$  consiste en utilizar un conjunto (i.i.d.) de variables aleatorias Gaussianas dispuestas en una malla en  $\mathbb{R}^d$  y realizar una convolución con un kernel arbitrario.

### Modelos de convolución de procesos

La evaluación de la dinámica espacio-temporal ha dado lugar a diversas técnicas en la literatura. Entre ellas, destacan los modelos de covarianza no paramétricos Barry and Ver Hoef (1996), procesos latentes no gaussianos Ickstadt et al. (2008), la restricción del dominio de

$x(s)$  Higdon (2004), la covarianza espacial no estacionaria Higdon (2002), los modelos de espacio-tiempo Higdon et al. (1998), los procesos espaciales según su construcción Ver Hoef and Barry (1998); Ver Hoef et al. (2004), la reducción de dimensiones Higdon (2002), entre otros.

En su trabajo, Higdon (2002) propone un enfoque para desarrollar procesos de convolución basado en el modelo de media móvil. En este contexto, el modelo para el proceso espacial  $Z(s)$  se define mediante la especificación del modelo para el proceso latente  $x(s)$  y el kernel de suavizado  $k(s)$ .

La técnica de convolución del proceso, conocida también como media móvil o suavizado de kernel, se destaca por su aplicación extensiva en la especificación de procesos espaciales de manera general mediante un enfoque constructivo. En términos generales, un proceso estocástico espacial  $Y(\cdot)$  en  $R_1 \subset \mathbb{R}^d$  puede definirse mediante la operación de convolución kernel:

$$Y(s) = \int_{R_1} K(s - u)dW(u). \quad (2.11)$$

donde  $W(\cdot)$  es un proceso estocástico de dimensión  $d$  y  $K(\cdot)$  es una función kernel, si quiere saber más, consulte Thiebaux (1976) y Thiebaux and Pedder (1987). El trabajo de Higdon (2002) resume la clase extremadamente flexible de modelos estadísticos espaciales definidos mediante (2.11). Se pueden encontrar ejemplos en Ver Hoef and Barry (1998), Barry and Ver Hoef (1996), Ickstadt et al. (2008), Higdon et al. (1998), y Ver Hoef et al. (2004). Este enfoque tiene su relevancia en que es mucho más sencillo especificar funciones kernel (posiblemente paramétricas) en lugar de una función de covarianza directamente, ya que las funciones kernel solo requieren:

$$\int_{\mathbb{R}^d} k(u)du < \infty,$$

y

$$\int_{\mathbb{R}^d} k(u^2)du < \infty.$$

Mientras que una función de covarianza debe ser una función definida positiva, el proceso  $Y(\cdot)$  en (2.11) se convierte en un proceso Gaussiano. Entonces, la integral de la ecuación (2.11) puede expresarse como:

$$Y(s) = \int_G K(s - u)V(u)du. \quad (2.12)$$

Se escoge  $V(\cdot)$  como ruido blanco Gaussiano, entonces la función de covarianza es de la forma:

$$C(s, s') = \mathbb{E}[Y(s)Y(s')] = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K(s - u)K(s' - t)\mathbb{E}[V(u)V(t)]dudt.$$

Como  $\mathbb{E}[V(u)V(t)] = 0$  para todo  $u \neq t$  y, sin pérdida de generalidad,  $\mathbb{E}[V^2(u)] = 1$ , lo anterior se convierte en:

$$C(s, s') = \int_{\mathbb{R}^d} K(s - u)K(s' - u)\mathbb{E}[V^2(u)]du = \int_{\mathbb{R}^d} K(s - u)K(s' - u)du. \quad (2.13)$$

Cuando la función kernel  $K(\cdot)$  es constante, (2.13) se puede re-escribir usando la transformación lineal  $t = u - s$  (con Jacobiano unitario) como:

$$C(s, s') = \int_{\mathbb{R}^d} K(t)K([s' - s] - t)dt, \quad (2.14)$$

que es una función de covarianza estacionaria.

La elección del proceso latente y el kernel de suavizado puede conducir a una serie de enfoques de modelado interesantes, algunos de los cuales se detallan en (Higdon, 2004).

## Pesos Espaciales

Los pesos espaciales son una herramienta fundamental para modelar la dependencia espacial entre observaciones en un conjunto de datos georreferenciados. Uno de los enfoques es usar valores que representan la influencia relativa que tiene cada ubicación sobre las ubicaciones circundantes, y se utilizan para realizar interpolación espacial, estimar valores en ubicaciones no muestreadas y analizar patrones espaciales. Los pesos espaciales se pueden definir de diferentes maneras, dependiendo del enfoque y del tipo de dependencia espacial que se desea capturar. A continuación, se presentan algunas formas comunes.

Los pesos basados en la distancia asignan importancias inversamente proporcionales a la distancia entre las ubicaciones y son utilizados en la interpolación por Kriging y el método de los vecinos más cercanos, Cressie (1993). Los pesos basados en la vecindad se definen según la relación de vecindad entre las ubicaciones; se utilizan en métodos como la interpolación por polígonos de Thiessen y los modelos de autoregresión espacial (SAR)<sup>5</sup> Bivand et al. (2013). Los pesos basados en funciones de influencia se utilizan funciones de influencia espacial para asignar pesos a las ubicaciones en función de su relación espacial. Son comúnmente utilizados en métodos como la regresión espacial y los modelos de tendencia espacial Chou (1992).

Es importante destacar que la elección de los pesos espaciales depende del objetivo y de las características específicas del conjunto de datos y del fenómeno estudiado. Diferentes métodos y enfoques pueden dar lugar a diferentes pesos espaciales y, por lo tanto, a diferentes resultados en el análisis espacial.

### 2.6.1. Modelado espacial

Suponga que el modelo espacial tiene la siguiente forma:

$$Y(\mathbf{x}) = S(\mathbf{x}; \theta) + \epsilon(\mathbf{x}), \quad (2.15)$$

donde  $Y(\mathbf{x})$  representa una observación tomada en la ubicación  $\mathbf{x}$  dentro del dominio  $D$ ,  $S(\mathbf{x}; \theta)$  es una función de media espacial con el parámetro  $\theta$ , y  $\epsilon(\mathbf{x})$  es un proceso de ruido Gaussiano. La función de media  $S(\mathbf{x}; \theta)$  se define como una mezcla ponderada localmente de regresiones lineales:

$$S(\mathbf{x}; \theta) = \prod_{j=1}^J \pi_j(\mathbf{x}) f'_j(\mathbf{x}) \theta_j, \quad (2.16)$$

donde  $f'(\mathbf{x}) = \{f_{j1}(x), \dots, f_{jq}(x)\}^T$  es un conjunto de funciones de base conocidas,  $\theta_j = (\theta_{j1}, \dots, \theta_{jq})^T$  representa un vector de parámetros aleatorios desconocidos,  $\pi_j(\mathbf{x})$  es un núcleo de ponderación no negativo centrado en la ubicación  $\mu_j$ , y  $J$  denota el número total de componentes.

---

<sup>5</sup>[5] (por sus siglas en inglés, *spatial autorregressive*)

Considere una realización del proceso observando  $\mathbf{Y} = (Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))^T$  en las ubicaciones  $\mathbf{x}_1, \dots, \mathbf{x}_n$ . Se define  $\pi_j(\mathbf{x}) = (\pi_j(\mathbf{x}_1), \dots, \pi_j(\mathbf{x}_n))^T$  como el vector de ponderaciones para el kernel  $j$ , y  $\mathbf{X}_j = (f_j(\mathbf{x}_1), \dots, f_j(\mathbf{x}_n))^T$  como la matriz de diseño para el  $j$ -ésimo componente de la mezcla. Utilice  $\text{diag}(r)$  para denotar una matriz con  $r_1, \dots, r_n$  en la diagonal y ceros en el resto de las entradas Christakos (1992).

Es posible escribir el modelo como una regresión lineal:  $\mathbf{Y} = \mathbf{X}\theta + \epsilon(\mathbf{X})$ , donde

$$\mathbf{X} = (\text{diag}(\pi_1)\mathbf{X}_1, \dots, \text{diag}(\pi_j)\mathbf{X}_j),$$

y

$$\theta = (\theta_1, \dots, \theta_j)^T.$$

En términos generales, el proceso  $S(\mathbf{X}; \theta)$  no es estacionario. Al asumir que el parámetro  $\theta$  tiene una distribución priori con media cero y matriz de covarianza  $W$ , el proceso  $S(\mathbf{X}; \theta)$  tiene una covarianza:

$$K(x, y) = \text{COV}[S(x), S(y)] = \sum_{i=1}^J \sum_{j=1}^J \pi_j(x)\pi_j(y)\theta_{ij}, \quad (2.17)$$

donde  $x$  e  $y$  son ubicaciones, y  $\theta_{ij}$  denota el elemento  $ij$  de  $\theta$ . Se elige superficies constantes, es decir,  $q = 1$  y  $f'(x) = 1$ , entonces la ecuación (2.17) se reduce a:

$$K(x, y) = \sum_{i=1}^J \sum_{j=1}^J \pi_j(x)\pi_j(y)W_{ij}.$$

Si además se escoge una priori independiente para los coeficientes con matriz de covarianza  $W = \tau^2 I$ , entonces se obtiene:

$$K(x, y) = \tau^2 \sum_{i=1}^J \sum_{j=1}^J \pi_j(x)\pi_j(y).$$

Esta formulación es similar al enfoque de convolución de procesos de Higdon et al. (1998). Su modelo espacial se puede considerar como un caso especial del modelo (2.15), donde las superficies locales son constantes y el número de núcleos  $J$  es grande.

## 2.6.2. Modelos dinámicos para datos espacio-temporales

La modelación de datos espacio-temporales a menudo implica la fusión de enfoques entre modelos de series temporales y modelos basados en variogramas para estadísticas espaciales.

Una parte significativa de la literatura se centra en los modelos isotrópicos Cressie (1993). Estos modelos se originaron en aplicaciones de geoestadística, donde la predicción, o *kriging*, es el objetivo principal. Por esta razón, a menudo nos referimos a ellos como modelos basados en *kriging*. Sansó and Guenni (1999) adoptaron un enfoque en el que asumieron una función paramétrica de covarianza para los errores de observación. Tonellato (1998) propuso un modelo vectorial  $AR(p)$  que incorpora choques temporales con correlaciones espaciales, enfrentando limitaciones en la dimensión del vector de estado que crece con el número de estaciones de observación. Mardia et al. (1998) fusionaron *kriging* y SSM para modelar la

covarianza espacial y la dependencia temporal. Adoptaron la optimización numérica iterativa para estimar los parámetros del variograma, y el filtro de Kalman se utilizó para estimar los campos de tendencias temporales. Otros procedimientos similares fueron propuestos por Wikle and Cressie (1999) y Huang and Cressie (1996).

Este trabajo se basa en la propuesta Huerta and Sansó (2007) que extienden el DLM para espacio-tiempo usando las ecuaciones:

$$y_t = K^t x_t + \epsilon_t, \quad (2.18)$$

$$x_t = x_{t-1} + v_t. \quad (2.19)$$

Donde  $K^t$  es una matriz  $n_t \times k$  con  $K^t = k(s_i - s_j), t = 1, \dots, m$ , y  $k$  es obtenido por un núcleo dado y los  $s_i$  son las estaciones en la ubicación  $i$ . En este contexto, el DLM univariado se transforma en un modelo multivariado, donde el error de observación  $\epsilon_t \sim N(0, \sigma_\epsilon^2 I_n)$  y el error de evolución  $v_t \sim N(0, \sigma_V^2 I_k)$ , para  $t = 1, \dots, m$ . Para el vector de estado  $x_0$ , se tiene  $x_0 \sim N(0, \sigma_{x_0}^2 I_k)$ .

Los elementos  $k(\cdot - \omega_j)$  están definidos por un núcleo de suavizado, donde  $\omega_1, \dots, \omega_k$  son los sitios o pesos espaciales. Los errores no están correlacionados espacialmente; la dependencia espacial de este modelo está completamente definida por  $k$ .

Algunos núcleos recomendados son:

- Gaussiano:  $k(s) \propto \exp\{-\|s\|^2/2\eta^2\}$ .
- Exponencial:  $k(s) \propto \exp\{-\|s\|/\eta\}$  para alguna  $\eta > 0$ .
- Esférica  $k(s) \propto (1 - \|s\|^3/r^3)^3 I[s \leq r]$ , para una  $r$  dada.

Hacer una convolución en  $\omega_1, \dots, \omega_k$  proporciona una aproximación finita a la convolución continua. Para la convolución integral es posible encontrar una equivalencia entre los núcleos y la función de covarianza del proceso. Ver Higdon (2002) para más detalles.

# Capítulo 3

## Modelos espacio temporales de valores extremos

El objetivo de este trabajo es adaptar la estructura de un DLM a valores extremos cuando las observaciones tienen dependencia espacio-temporal. Inicialmente, identificaremos a las observaciones que superan un umbral y se les asigna la distribución GEV, donde  $\sigma$ ,  $\xi$  son los parámetros de la distribución y las locaciones ( $\mu$ ) definirán la estructura espacio-temporal. El componente temporal será un caso particular de los modelos de espacio y estado y se definirá mediante un modelo lineal dinámico según West and Harrison (1997). Se tomarán valores a lo largo del tiempo y mediante el DLM se podrá estimar la tendencia o estacionalidad de los datos. Para la estructura espacial, seguiremos un proceso de convolución medido mediante una matriz de evolución del DLM y un núcleo para la importancia o dependencia espacial. Huerta and Sansó (2007) desarrolló un trabajo significativo en este enfoque, y este trabajo utiliza su propuesta como base para resolver el problema de valores extremos generalizados con modelos lineales dinámicos (GEV-DLM). La inferencia de los valores desconocidos se realizará con un enfoque Bayesiano, y la aproximación de las funciones a posteriori se obtendrán con simulaciones de Monte Carlo Markov Chain. Ilustramos nuestra metodología con valores extremos de las precipitaciones medidos en el estado de Washington, Estados Unidos NOAA.

Es necesario tomar los dos enfoques de identificación de umbrales bajo el supuesto de independencia. El primero es el método de bloque por máximos, que puede ser muestreado mediante la ecuación (2.2) y asignar una distribución priori a los parámetros desconocidos Coles and Tawn (1996). El segundo método consiste en considerar los elementos que exceden un umbral, como lo define claramente Pickands (1971); este método es conocido como Peaks Over Threshold (POT).

Un enfoque alternativo para modelar datos extremos es considerar la distribución de exceedencias sobre un umbral alto. Para obtener una explicación más detallada, Smith et al. (1997); Smith (1989); Davison and Smith (1990) proporcionan detalles, y recientemente Behrens et al. (2004) lo explica. En el modelado de los datos de ubicaciones, Casson and Coles (1999) propone obtener un umbral basado en procesos puntuales, mientras que Gilleland et al. (2005) se basa en la Distribución de Pareto Generalizada para desarrollar un modelo jerárquico.

En el dominio de series temporales, Gaetan and Grigoletto (2004) presentan un modelo

con parámetros dinámicamente variables para el análisis de extremos de una serie temporal univariada. Su modelo incluye cambios dinámicos para los parámetros de escala/forma y actualización secuencial con filtros de partículas, pero no hay una estructura de espacio o espacio-tiempo.

En el contexto del análisis, se enfoca en explorar la versatilidad de la distribución GEV, permitiendo que sus parámetros evolucionen tanto temporal como espacialmente. En la Sección 3.1, se presenta enfoques innovadores para abordar series temporales de valores extremos, excluyendo consideraciones espaciales en esta etapa. Investigamos la dinámica temporal de los parámetros que gobiernan la distribución de los datos, utilizando DLMS. Demostrando la capacidad de capturar tendencias cambiantes en el tiempo a través de un modelo dinámico que facilita la detección de cambios tanto a corto como a largo plazo.

El problema o desafío está en conectar estas tendencias con covariables que varían en el tiempo, permitiéndonos describir la evolución de la fuerza de la relación lineal a lo largo del tiempo. En la Sección 3.1, expandimos el modelo para incorporar variación espacial en los parámetros. El modelo propuesto para los extremos se fundamenta en convoluciones de procesos (Higdon, 2002), ofreciendo una reducción efectiva del espacio de parámetros y facilitando el manejo eficaz de un gran número de ubicaciones.

### 3.1. Modelos GEV-DLM

Se explora una perspectiva novedosa para modelar la evolución temporal de los máximos utilizando la distribución GEV Coles (2013). Este enfoque consiste en definir los cambios a lo largo del tiempo mediante funciones deterministas, las cuales pueden estar asociadas con la tendencia o estacionalidad de los datos. Considere una secuencia de observaciones condicionalmente independientes  $y_1, y_2, \dots, y_m$ , donde cada observación sigue una distribución GEV con un parámetro de ubicación variable en el tiempo. Es decir:

$$z_t \sim \text{GEV}(\mu_t, \sigma, \xi).$$

Considere la obtención de  $y_t$  al fijar un período de tiempo o bloque de datos y calcular el máximo de las observaciones durante dicho intervalo. Para modelar los cambios temporales, Coles (2013) presenta diversas alternativas tanto para covariables como para modelos no estacionarios.

Considerando un modelo dinámico, se asume que los datos  $y_1, y_2, \dots, y_m$  constituyen una secuencia condicionalmente independiente, donde  $y_t$  sigue una distribución GEV con parámetros  $\mu_t$ ,  $\sigma$ , y  $\xi$ . La función de distribución acumulativa para cada  $y_t$  se expresa como:

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z_t - \mu_t}{\sigma} \right)^{\frac{-1}{\xi}} \right] \right\}; \quad t = 1, \dots, m. \quad (3.1)$$

Ahora, en lugar de considerar una función determinista en  $\mu_t$ , suponga que esta varía en función del tiempo  $t$  de forma desconocida. El DLM que modela la variable latente de forma temporal adopta la siguiente estructura temporal:

$$\mu_t = F_t \theta_t + v_t, \quad v_t \sim N(0, V), \quad (3.2)$$

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim N(0, W). \quad (3.3)$$

Donde:

- $\theta_t$  es un vector de estado de dimensiones  $k \times 1$ .
- $F_t$  es un “regresor” de dimensiones  $k \times 1$ .
- $G_t$  es una matriz de evolución  $k \times k$ .
- $V$  es la varianza observacional.
- $W$  es una matriz de covarianza de evolución  $k \times k$ .

Esto define la cuádrupla DLM  $(F_t, V, G_t, W)$  tal como lo define West and Harrison (1997). Al emplear un DLM, ganamos mayor flexibilidad ya que las tendencias no están limitadas a una forma paramétrica específica. Esto nos permite evaluar tanto los cambios a corto plazo como los de largo plazo. A esta familia de modelos para evaluar extremos de forma dinámica los denominaremos GEV-DLM por sus siglas en inglés, *dynamic linear models with generalized extreme value distributions*.

Conociendo la estructura temporal del GEV-DLM, se extiende el modelo para incorporar la estructura espacial. Cabe destacar que Giraldo (2002) proporciona detalles sobre el análisis espacial, y la geoestadística se encarga de definir rigurosamente la teoría. En este enfoque, trabajamos con procesos de convolución, tal como se describe en Higdon (2002).

Finalmente, la log-verosimilitud para los parámetros del modelo se define mediante la siguiente ecuación:

$$l(\theta) = -m \log(\sigma) - (1 + 1/\xi) \sum_{t=1}^m \sum_{s=1}^S \log(1 + \xi(y_{s,t} - \mu_{s,t})/\sigma) - \sum_{t=1}^m \sum_{s=1}^S \{1 + \xi(y_{s,t} - \mu_{s,t})/\sigma\}^{-\frac{1}{\xi}}.$$

### 3.1.1. Modelo de Huerta and Sansó (2007)

El modelo de Huerta and Sansó (2007) es un caso particular al modelo presentado en la sección anterior, las observaciones en espacio y tiempo se definen por  $z_{s,t} \sim GEV(\mu_{s,t}, \sigma, \xi)$  para  $s = 1, \dots, S$  y  $t = 1, \dots, m$ . La función de distribución para  $z_{s,t}$  es:

$$G_{s,t}(z, t) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z_{s,t} - \mu_{s,t}}{\sigma} \right)^{\frac{-1}{\xi}} \right] \right\}.$$

Para cada  $t$  sea  $\mu_t = (\mu_{1,t}, \mu_{2,t}, \dots, \mu_{S,t})$ . Ahora el DLM para el parámetro de ubicación  $\mu_t$  es:

$$\mu_t = K^t \theta_t + \epsilon_t, \quad (3.4)$$

$$\theta_t = \theta_{t-1} + v_t. \quad (3.5)$$

donde el vector de estado  $\theta_t = (\theta_{t,1}, \dots, \theta_{t,k})$ ,  $\epsilon_t = (\epsilon_{t,1}, \dots, \epsilon_{t,k})$ ,  $v_t = (v_{t,1}, \dots, v_{t,k})$ . Además,  $\epsilon_t \sim N(0, \sigma_\epsilon^2 I_k)$  y  $V_t \sim N(0, \sigma_V^2 I_k)$ . En cuanto a los procesos de convolución, los núcleos se centran alrededor de puntos en una cuadrícula regular, y el parámetro de rango se fija como  $d = c\phi$ , donde  $\phi$  representa la distancia entre los puntos de la cuadrícula y  $c$  es una constante

generalmente entre  $1/2$  y  $2$ . Las convoluciones son estimadas mediante un núcleo Gaussiano, donde la matriz  $K$  de dimensión  $S \times k$  viene dada por:

$$K_{ij} = \exp(-d||s_i - \omega_j||^2/2),$$

$s_i$  es la posición de la estación  $i$ ;  $\omega_j$  es el centro del núcleo  $j$ -ésimo, con  $j = 1, \dots, k$  y  $d$  es un parámetro de rango.

Las convoluciones de procesos ofrecen una representación lineal útil en algunos casos dependiendo de las propiedades específicas de los procesos involucrados y del contexto particular de la aplicación. Para el modelo GEV espacial, necesitamos datos de las observaciones en cada estación, es decir,  $s_1, \dots, s_{n_t}$ . Cada lugar contiene datos temporales; por lo tanto, generamos un vector  $y_t = (y_{1,t}, \dots, y_{n_t,t})$  con observaciones en espacio-tiempo para un tiempo dado.

Para los parámetros en la primera etapa del modelo especificamos las distribuciones a priori:

$$\sigma \sim LN(\mu_\sigma, s_\sigma) \quad y, \quad \xi \sim N(\mu_\xi, s_\xi).$$

Para los parámetros de la segunda etapa:

$$\begin{aligned} \theta_0 &\sim N(0, \sigma_\theta^2 I_k), \\ \sigma_\epsilon^{-2} &\sim \text{Gamma}(\alpha_\epsilon, \beta_\epsilon), \\ \sigma_v^{-2} &\sim \text{Gamma}(\alpha_v, \beta_v) \quad y, \\ \sigma_\theta^{-2} &\sim \text{Gamma}(\alpha_\theta, \beta_\theta). \end{aligned}$$

El algoritmo utilizado para aproximar las posteriors es un MCMC con pasos FFBS en la primer etapa se definen las prioris para el MCMC y en la segunda etapa se define las prioris para el FFBS. Se definen los vectores  $\mathbf{y} = (y_1, y_2, \dots, y_m)$ ,  $\mu = (\mu_1, \mu_2, \dots, \mu_m)$ , y  $\theta = (\theta_1, \theta_2, \dots, \theta_m)$ . En cada paso de el algoritmo es:

---

**Algorithm 4** MCMC-FFBS
 

---

- Muestrear  $V$  mediante una priori gamma.
  - Muestrear  $W$  mediante una priori inversa-gamma.
  - Para  $i = 1, 2, \dots, iter$ , hacer:
    1. Muestrear  $\sigma$  mediante  $p(\sigma|\mu, y, \xi)$ , con MCMC.
    2. Muestrear  $\xi$  mediante  $p(\xi|\mu, y, \sigma)$  con MCMC.
    3. Muestrear  $\mu = (\mu_1, \dots, \mu_n)$  mediante  $p(\mu|\xi, y, \sigma)$  con MCMC.
  - Muestrear  $\theta_t$  mediante un FFBS, para  $t = 1, \dots, m$ .
- 

Huerta and Sansó (2007) propone una solución ingeniosa para integrar valores extremos y modelos lineales dinámicos, pero en este trabajo encontramos varias limitaciones importantes. La primera es que el modelo posee múltiples niveles, haciendo que la cantidad de parámetros sea superior a la cantidad de información obtenida. Esta problemática dificulta el proceso de estimación e identificación. Otra limitante mayor es que los modelos de espacio y estados asumen que la variables observadas (en este caso son  $\mu_t$ ) son independientes, haciendo que la dependencia espacio-temporal se incorpore al modelo de forma inadecuada.

### 3.1.2. Modelo propuesto

El modelo propuesto busca reducir el número de parámetros disminuyendo a un nivel de latencia y por efectos de estabilidad numérica se basa en uno de los resultados de Nakajima et al. (2012). Las observaciones en espacio y tiempo se definen por  $z_{s,t} \sim GEV(\mu_{s,t}, \sigma, \xi)$  mediante la siguiente ecuación

$$z_{s,t} = K_s \cdot \mu_{s,t} + \frac{\sigma(\exp(\xi \cdot \eta_s) - 1)}{\xi} + \epsilon_t, \quad \epsilon_t \sim N(0, v). \quad (3.6)$$

donde el parámetro de ubicación  $\mu_t = (\mu_{1,t}, \mu_{2,t}, \dots, \mu_{S,t})$  se rige mediante la siguiente ecuación:

$$\mu_t = \mu_{t-1} + w_t \quad w_t \sim N(0, W). \quad (3.7)$$

En este caso el proceso latente sigue un modelo auto regresivo espacial, donde  $K_s$  es la  $s$ -ésima columna de la matriz de convoluciones con núcleo Gaussiano, y las distribuciones a prior de los parámetros son:

- $\mu_0 \sim N(0, W)$ ,
- $\sigma \sim \text{half-t}(5)$ ,
- $\xi \sim N(0, 10)$ , y es distinta de cero.
- $W \sim \text{Gamma}(10, 2)$ , y
- $\eta_s \sim \text{Gumbel}$ , para todo  $s = 1, 2, \dots, S$ .

Es importante resaltar que si eliminamos la parte no lineal de la ecuación (3.6), el modelo es un DLM Gaussiano espacial y puede ser muestreado mediante el algoritmo FFBS. Bajo el supuesto de que  $\epsilon_t$  es muy pequeño Nakajima et al. (2012) muestra que el modelo se rige principalmente por la componente no lineal que genera una distribución GEV. Para garantizar que dicho supuesto se cumple elegimos una prior para  $v$  muy cercana a cero; por simplicidad del modelo elegimos a  $v = 0,05$ . Las distribuciones a priori se eligieron de tal forma que sean débilmente informativas al modelo, al no proveer mucha información y solo regularizar el ajuste del MCMC utilizado, ver Gelman et al. (2013).

Pese la fuerte no linealidad del modelo propuesto, presenta un número menor de estimación de parámetros y la posterior puede ser perfectamente muestreada mediante una combinación de FFBS en cada iteración del MCMC. El algoritmo utilizado para aproximar las posteriores es:

---

#### Algorithm 5 Hybrid FFBS-MCMC

---

Para  $i = 1, 2, \dots, \text{iter}$ , hacer:

1. Muestrear  $W \sim \text{Gamma}(10, 2)$  de forma condicional,
2. Muestrear  $\mu = (\mu_1, \dots, \mu_n)$  mediante un FFBS.
3. Muestrear  $\sigma$  mediante  $p(\sigma|y, \eta, \mu, \xi)$ , con paso individual MCMC.
4. Muestrar  $\eta$  mediante  $p(\eta|y, \sigma, \mu, \xi)$ , con paso individual MCMC.
5. Muestrear  $\xi$  mediante  $p(\xi|\mu, y, \sigma)$  con paso individual MCMC.
6. Si  $\xi = 0$ , definir  $z$  como:

$$z_{s,t} = K_s \cdot \mu_{s,t} + \eta_s + \epsilon_t.$$


---

En este documento, se desarrolló una implementación propia del algoritmo mencionado anteriormente. En este caso, se utilizaron diversos elementos, siendo importante destacar que se empleará el paquete *spatialStremes* para definir la estructura del modelo de valores extremos generalizados, es decir, para la función GEV. Esto nos permitirá estimar simultáneamente los parámetros  $\sigma$  y  $\xi$  mediante el algoritmo MCMC implementado por propiamente. Esto marca una diferencia con respecto al enfoque de Huerta and Sansó (2007), que propone pasos individuales y realiza la estimación de los parámetros mediante el uso del paquete *evdBayes* (Stephenson and Ribatet, 2023) que contiene internamente el MCMC. Este paquete produjo resultados poco satisfactorios en nuestros experimentos. Con el algoritmo MCMC desarrollado propiamente, mejoraremos las estimaciones al calcular todos los parámetros de manera conjunta en lugar de dejar algunos estáticos.

Esta variante permite calcular una matriz de valoración, a diferencia de Huerta and Sansó (2007), que se centra en pesos espaciales representados por puntos en una cuadrícula en el espacio observado. Este enfoque genera únicamente un vector, que contiene los valores de las estimaciones de los núcleos, delimitando cada estación con su correspondiente punto vecino en la cuadrícula definida por los pesos espaciales. A través de este método, solo se analizan las estaciones cercanas al mismo punto espacial.

## 3.2. Selección del Método MCMC

Como primer aporte, buscamos seleccionar el método de Markov Chain Monte Carlo (MCMC) a utilizar para definir de manera rigurosa los resultados. Recuerde que se han definido dos MCMC, M-H y HMC. Sin embargo, en este análisis de resultados solo se tomará en cuenta la estructura del modelo de valores extremos generalizados (GEV); es decir, no se considerará la parte temporal. Además, renombraremos al parámetro de forma como  $k$  y se destaca que, debido a que solo definiremos cuál método será utilizado, la comparación no será robusta. Partiendo de las ecuaciones:

$$z_1, z_2, \dots, z_m, z_t \sim GEV(\mu, \sigma, k).$$

Realizamos una comparación detallada entre dos métodos M-H y HMC. El objetivo es determinar la eficiencia relativa de uno en comparación con el otro en un escenario donde no se involucra la componente temporal ni la componente espacial del modelo. En esta evaluación, consideramos una distribución a priori que adopta una forma normal trivariada en  $(\mu, \log(\sigma), k)$ , con un vector media cero para cada parámetro y una matriz diagonal de covarianza. Esto implica que asumimos independencia priori y asignamos varianzas priori idénticas. Aunque normalmente empleamos distribuciones condicionales completas, para efectos de comparación optamos por utilizar el paquete R llamado *evdbayes* (Stephenson and Ribatet, 2006). Dicho paquete está disponible de manera gratuita en el sitio web <http://cran.r-project.org/web/packages/evdbayes> y ofrece funciones diseñadas para llevar a cabo el análisis Bayesiano de modelos de valores extremos utilizando métodos MCMC. Es importante resaltar que este paquete se basa en el algoritmo Metrópolis-Hastings para su funcionamiento.

El algoritmo HMC se implementó en R y, tras algunos ajustes, se encontró que el parámetro  $\epsilon$  era altamente sensible a ellos. En respuesta, se optó por adaptar el algoritmo No-U-Turn

Sampler (NUTS), propuesto por Hoffman and Gelman (2011). Nuestro enfoque se centró exclusivamente en afinar de manera más precisa los valores de  $\epsilon$  y  $L$ , dejando pendiente para futuros trabajos el desarrollo del DLM-GEV con el algoritmo NUTS para su comparación con M-H/HMC. Observamos que el algoritmo HMC alcanza un régimen estacionario mucho más rápido que M-H. Además, prácticamente no hay auto-correlación en la cadenas de salida, ver figura 3.1.

Para comparar la eficiencia relativa de estos métodos, calculamos el tamaño de muestra efectivo ( $ESS$ ) utilizando las muestras posteriores para cada parámetro. Esta medida se define como  $ESS = \frac{N}{(1+2\sum_j \gamma(j))}$  donde  $N$  es el número de muestras posteriores y  $\gamma(j)$  son las auto-correlaciones monótonas de retraso  $j$  muestra Geyer (1992). Por lo tanto, puede interpretarse como el número de muestras efectivamente independientes.

Para una comparación, primero descartamos 500 iteraciones de las muestras generadas por los algoritmos M-H y HMC. El  $ESS$  se puede obtener de manera más simple con el paquete R-coda Plummer et al. (2020) que proporciona herramientas para análisis y diagnóstico para cualquier salida del MCMC.

Iteraciones	MH		HMC	
	$\sigma$	$k$	$\sigma$	$k$
5,000	72.55443	300.5083	43.10983	87.788
8,000	174.7401	442.2094	91.908232	112.06989
10,000	162.1764	666.5499	82.49394	120.8534
12,000	214.7907	843.4945	51.18565	102.1039
15,000	360.3301	967.1631	127.8727	149.2896
18,000	487.3365	1252.599	278.1207	601.61

Cuadro 3.1: Tamaños de muestra efectivos ( $ESS$ ) para cada parámetro utilizando los algoritmos M-H y HMC.

Con el objetivo de evaluar y comparar el desempeño de los algoritmos HMC y M-H, se llevaron a cabo dos estudios de simulación para la estimación de parámetros en un modelo GEV. Ambos estudios se realizaron con simulaciones de  $n = 5000, 8000, 10000, 12000, 15000$  y  $18000$ . La Tabla 3.1 revela que los valores efectivos del algoritmo HMC son inferiores a los del MH, es decir, el tamaño de muestra necesario para garantizar la convergencia es menor en HMC. La rapidez y eficiencia en la obtención de estimaciones de parámetros atribuidas al algoritmo HMC se deben a su naturaleza explosiva, que permite saltos significativos en el MCMC. Sin embargo, es crucial señalar que este método requiere más tiempo para completar las simulaciones en comparación con otros enfoques. Inicialmente, el HMC tomaba 24 horas para calcular las estimaciones de los parámetros. Sin embargo, tras mejoras sustanciales en el algoritmo, este tiempo se redujo drásticamente a un total de 18 minutos para estimar los parámetros sin considerar la estructura espacio temporal. Por otro lado, el algoritmo MH requería solo 2 minutos para estimar los parámetros sin la estructura espacio-temporal. Esta variación en los tiempos se debe a la mayor exigencia computacional del cálculo del gradiente en el HMC en cada etapa de la simulación.

Con el propósito de asegurar lo mencionado anteriormente, resulta muy esclarecedor el análisis de los valores de auto-correlación en función de diversos desfases para la variable en estudio. Para el retraso  $j$ , definimos Auto-correlación ( $ACF$ ) como la correlación muestral

entre dos conjuntos de observaciones  $\theta_1, \dots, \theta_{n-j}$  y  $\theta_{1+j}, \dots, \theta_n$ . En este contexto, una auto-correlación cercana a cero, incluso con un desfase de uno, indicará que los sucesivos pasos de cada cadena están prácticamente no correlacionados.

Las figuras 3.1 presentan los gráficos de auto-correlación para los parámetros  $\sigma$  y  $k$ , utilizando los métodos M-H y HMC, respectivamente. Se observa que en ambos casos, los valores de auto-correlación descienden rápidamente a cero con un desfase de uno. Es importante destacar que los resultados del método HMC muestran un descenso más pronunciado, indicando así una mayor velocidad de convergencia al implementar nuestro modelo.

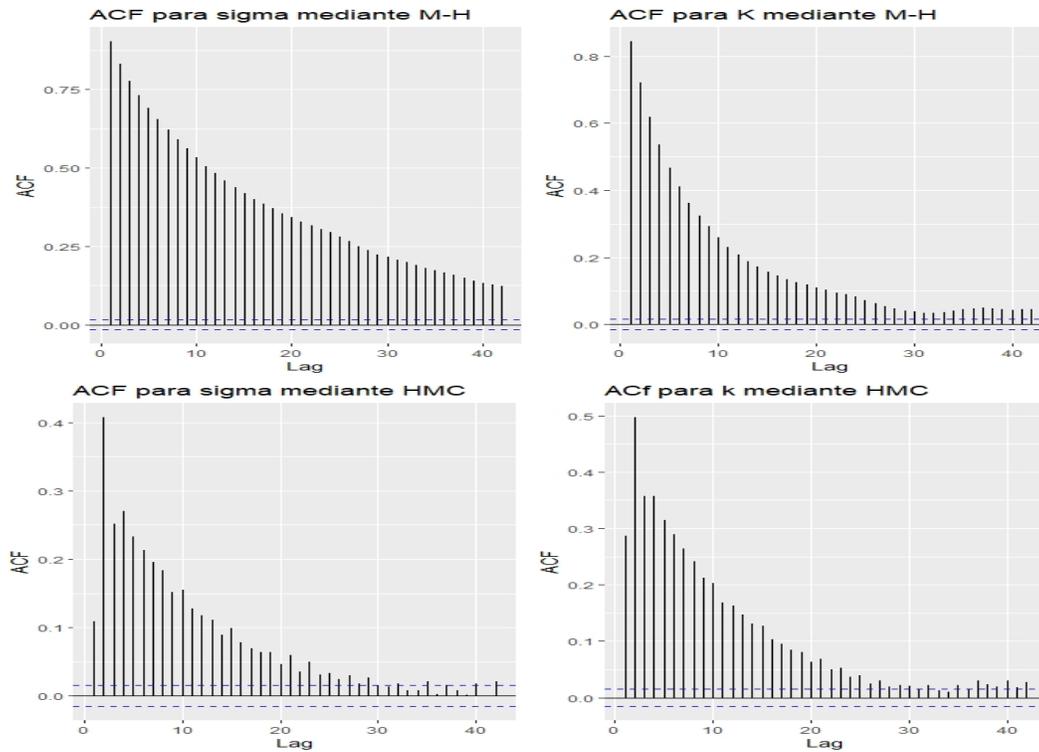


Figura 3.1: Gráfico de auto-correlación ACF para los parámetros de escala sigma (primera columna) y forma k (segunda columna). La primera fila representa los ACF de las cadenas obtenidas por el algoritmo Metropolis Hasting, y la segunda fila representa los ACF de las cadenas obtenidas por el Hamiltonian Monte Carlo.

Finalizamos esta sección con la conclusión de que, en este escenario específico, el HMC muestra ser más eficiente en comparación con el método M-H. Observamos que la auto-correlación en el HMC decae más rápidamente hacia cero, lo que indica una convergencia más eficaz de las cadenas de Markov generadas. Además, los valores del tamaño efectivo de la muestra en el HMC son más pequeños, lo que sugiere que se requiere una cantidad menor de muestras para obtener estimaciones confiables.

Apreciamos que la estacionariedad se establece más rápidamente en el HMC en comparación con el M-H, lo que respalda aún más su eficiencia en la exploración del espacio de parámetros. No obstante, es importante tener en cuenta que el HMC demanda la estimación precisa de los valores de  $\epsilon$  y  $L$  para lograr una convergencia óptima, lo que puede implicar cierta sensibilidad en su configuración.

Adicionalmente, observamos que el proceso de finalización de las simulaciones en el HMC tiende a ser más lento en comparación con el M-H. Esta diferencia en la velocidad de finalización puede depender de varios factores, incluyendo la complejidad del modelo, la elección de parámetros y la implementación del algoritmo.

En resumen, en esta comparación específica, el HMC demuestra su superioridad en términos de eficiencia. Sin embargo, la elección entre estos métodos debe realizarse considerando diversos aspectos, como la precisión requerida, la facilidad de implementación y los recursos computacionales disponibles. Como trabajo futuro, se propone probar los algoritmos utilizando un ordenador más potente.

### 3.3. Caso Simulación

El estudio de simulación tiene como objetivo evaluar si los modelos propuestos e implementados recuperan los parámetros, los cuales, al tratarse de una simulación, son conocidos. Además, en este caso particular, se busca analizar el rendimiento del modelo GEV-DLMs para valores extremos mediante la aproximación de algunas distribuciones paramétricas estándar. En particular, nos centraremos en su desempeño para validar el modelo en su componente temporal mediante revisiones de la distribución predictiva. La metodología a utilizar consiste en: (i) simular valores de los parámetros mediante las distribuciones a priori, (ii) simular una muestra del modelo a partir de los parámetros simulados, (iii) estimar las distribuciones a posteriori y (iv) comparar la predictiva del modelo con los datos simulados. A partir de este estudio, renombraremos el parámetro de forma  $\xi$  como  $k$ . Los datos simulados del modelo GEV-DLMs se obtienen a partir de las siguientes ecuaciones:

$$z_1, z_2, \dots, z_m, z_t \sim GEV(\mu, \sigma, k).$$

Las ecuaciones de la componente temporal son:

$$\begin{aligned} \mu_t &= \theta_t + \epsilon_t, \quad \epsilon_t \sim N(0, V); \\ \theta_t &= \theta_{t-1} + w_t, \quad w_t \sim N(0, W). \end{aligned}$$

Las distribuciones a priori definidas para el modelo son:

$$\begin{aligned} \sigma &\sim \text{half-t}(5) \\ k &\sim N(0, 1), \\ \theta_0 &\sim N(0, 100) \\ w_t &\sim N(0, W) \\ \epsilon_t &\sim N(0, V) \end{aligned}$$

Donde  $W$  y  $V$  son varianzas pequeñas.

#### 3.3.1. Recuperación de los parámetros

En esta etapa fundamental de nuestro análisis, nos enfocamos en la sección 3.3 que marca el punto de partida de nuestra investigación. En esta fase inicial, nuestro enfoque se centra en la

estimación de los estados, y lo hacemos a través de la implementación del filtro de Kalman en un contexto que excluye deliberadamente la componente espacial. Específicamente, optamos por emplear el enfoque del DLM y recurrimos al algoritmo de suavización basado en el FFBS para la estimación del parámetro  $\mu$ .

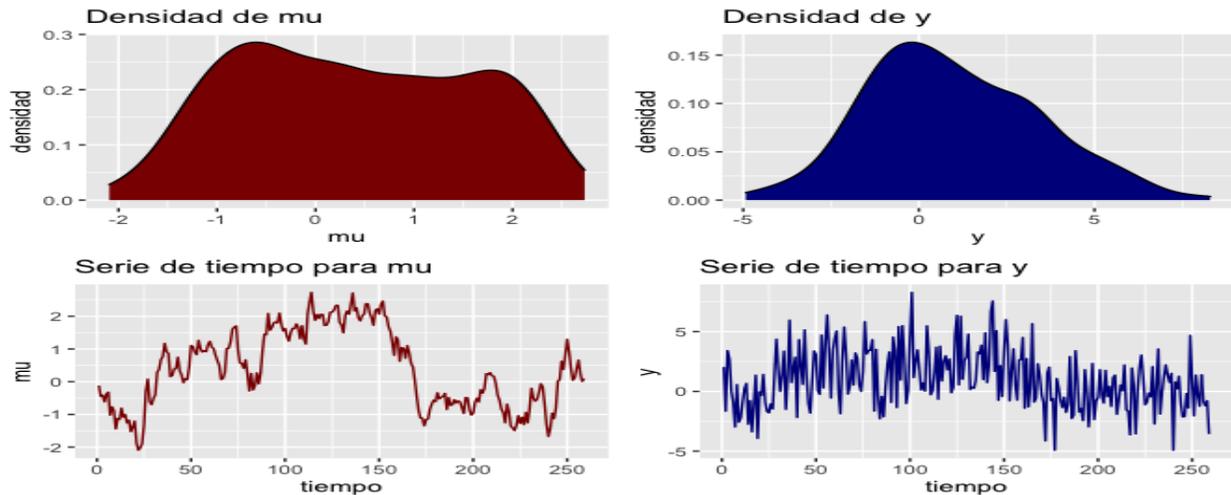


Figura 3.2: La primera columna muestra las simulaciones de las locaciones obtenidas a partir del Filtro de Kalman. La segunda columna muestra los datos simulados ( $y$ ) a partir del modelo y los parámetros de escala ( $\sigma$ ) y forma ( $k$ ).

En este marco analítico, construimos un modelo que se ajusta a una distribución GEV. Esta elección se justifica por la necesidad de abordar las interdependencias temporales inherentes a nuestros datos. Para este propósito, realizamos una modificación sustancial en el modelo original. Introducimos una ecuación adicional específicamente dirigida al parámetro de locación  $\mu_t$ , con el fin de hacer frente de manera adecuada y precisa a la estructura temporal observada en los datos.

Es importante subrayar que este proceso de modelización implica que cada una de las observaciones, denotadas como  $y_t$ , se ajusta a una distribución GEV. En consonancia con esta elección, introducimos un nivel adicional de detalle al modelar el parámetro  $\sigma$ . Optamos por emplear una distribución t-Student truncada con 5 grados de libertad (half-t(5)) para  $\sigma$ . Esta elección garantiza que  $\sigma$  se mantenga en valores positivos, tal como dicta su definición teórica.

Para simular los datos del modelo primero simulamos las locaciones en cada tiempo usando un  $DLM(F = 1, G = 1, V = 1, W = 0,1)$  constante, y extraemos los valores de  $\mu_t$ . Los parámetros de escala y forma se simulaban con variables log-normal y normal estándar respectivamente. Finalmente, simular las observaciones mediante la ecuación 3.3 junto con los valores obtenidos por el DLM y las simulaciones de las distribuciones a priori.

La Figura 3.2 presenta el muestreo para el parámetro  $\mu$  mediante FFBS partiendo del DLM, en esta parte se observa que el modelo no presenta estacionariedad ni tendencia. Además, mostramos los gráficos de los datos correspondientes.

variable	mean	median	sd	mad	q5	q95	rhat	ess_bulk	ess_tail
tasa-aceptación	0.24	0.00	0.43	0.00	0.00	1.00	1.00	-	-
$\sigma$	1.36	1.36	0.06	0.06	1.27	1.46	1.00	1,982.43	2,650.32
$k$	-0.37	-0.37	0.03	0.03	-0.41	-0.32	1.00	2,395.81	3,297.38

Cuadro 3.2: Estimación de los parámetros  $\sigma$  y  $k$ . En la segunda columna se presenta la media de los parámetros, en la segunda se determina la mediana. La tercera columna muestra la desviación estándar, la cuarta columna la desviación media absoluta. En las sexta y séptima columnas representamos el intervalo de credibilidad mediante el cuartil 5 y el cuartil 95, respectivamente. La séptima columna indica la convergencia en las cadenas. Las columnas novena y décima muestran los tamaños de muestras efectivos para la distribución a posteriori del modelo vía Metrópolis-Hasting.

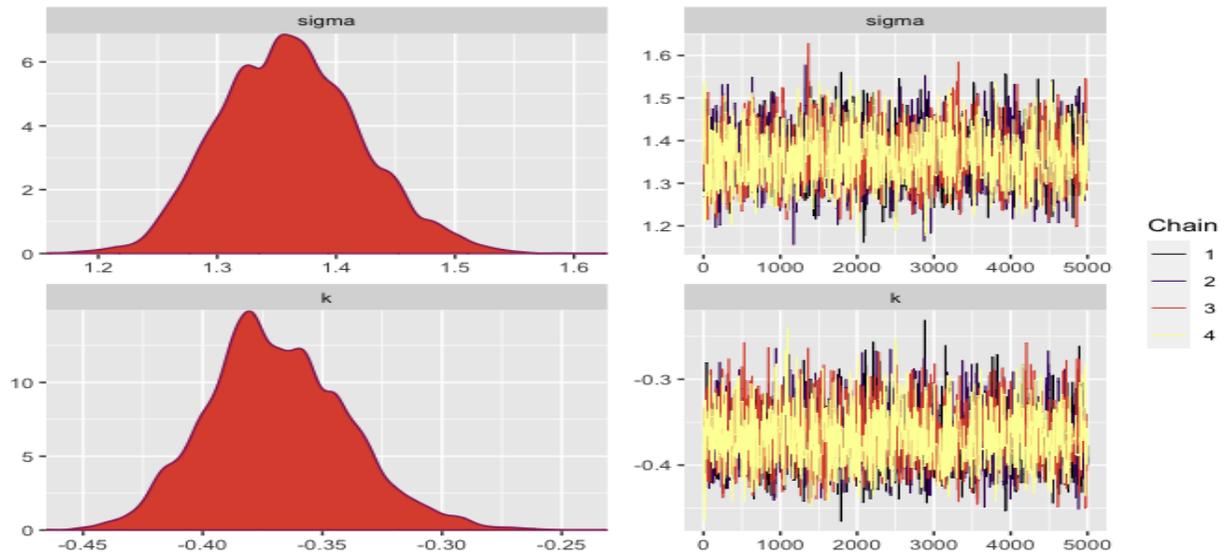


Figura 3.3: La primer columna presenta la densidad de los parámetros de escala  $\sigma$  y forma  $k$ , respectivamente. En la segunda columna se grafican las cadenas correspondientes a los parámetros de escala ( $\sigma$ ) y forma  $k$ .

A partir, de los datos simulados, obtenemos las distribuciones a posterior del modelo aplicando el algoritmo FFBS y el algoritmo MH simultáneamente. Primero se extrae una muestra para la posterior marginal de  $\mu_1, \mu_2, \dots, \mu_n | y$  mediante el algoritmo FFBS asumiendo una aproximación normal. Seguido, ajustar un algoritmo de MH para los otros dos parámetros restantes. En este caso se corren cuatro cadenas de 5,000 iteraciones por cadena, donde las primeras 2,500 iteraciones se descartan por entrenamiento (*warm-up*).

El cuadro 3.2 muestra que las cadenas han convergido con tamaños de muestra efectivos aceptables y factores convergencia (rhat) cercanos a uno, indicando convergencia. La tasa de aceptación es baja pero aceptable para un algoritmo de Metrópolis.

En la figura 3.3 observamos que las densidades y las trazas de las cadenas, en ambos casos no se distinguen comportamientos anómalos por lo tanto aceptamos el modelo y procedemos a analizar el comportamiento temporal del parámetro de ubicación  $\mu$ .

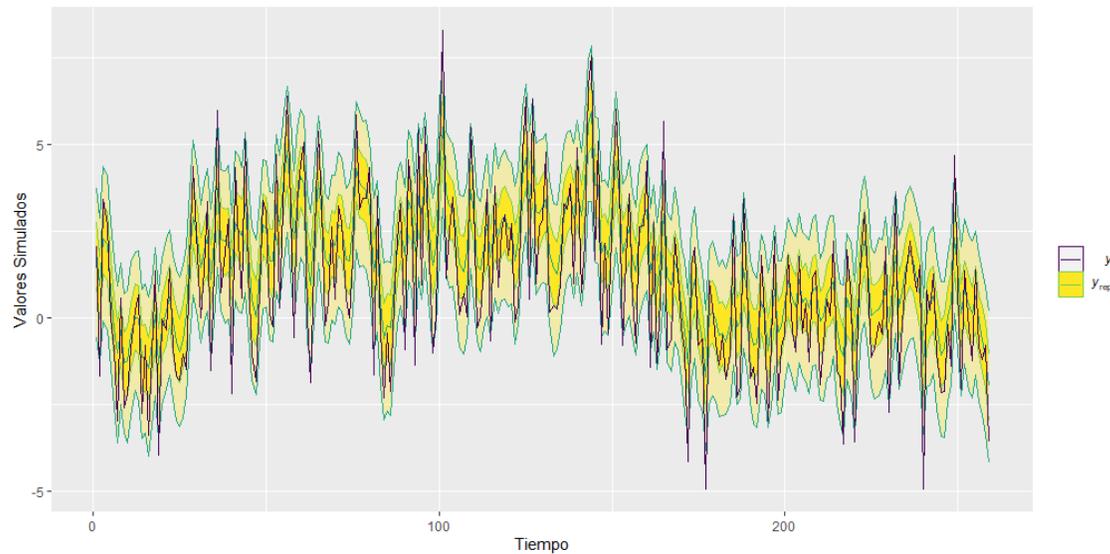


Figura 3.4: Validación del modelo utilizando datos reales ( $y$ ), representados en color negro, y datos de entrenamiento para el modelo ( $y_{rep}$ ), que se muestran mediante intervalos en color amarillo con líneas verdes.

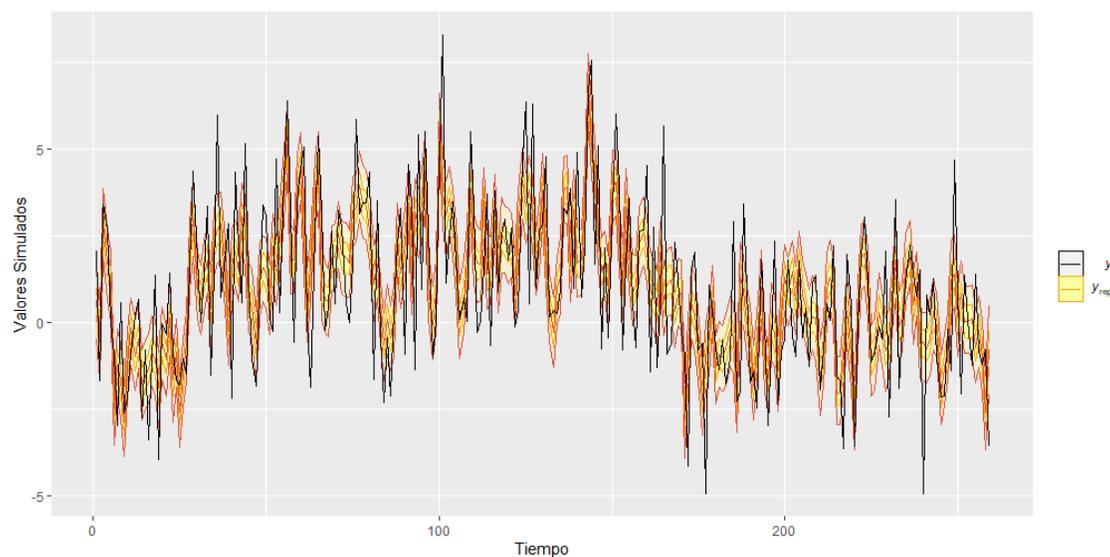


Figura 3.5: Validación del modelo mediante el uso de datos reales ( $y$ ), representados en color negro, y datos de entrenamiento para el modelo ( $y_{rep}$ ), los cuales se presentan mediante intervalos en color amarillo con líneas naranjas, calculados a través del algoritmo de Huerta.

Con el propósito de someter a una evaluación la calidad del ajuste de nuestro modelo, hemos empleado un enfoque de revisión de la distribución predictiva, y sus resultados se muestran en las figuras 3.4 y 3.5. Este enfoque implica la generación de muestras a partir de nuestro modelo utilizando las distribuciones posteriores de los parámetros. Es importante resaltar que, en este caso, la media posterior se emplea de manera clave para informar sobre las locaciones de los datos modelados. Para llevar a cabo este proceso de validación, hemos extraído una muestra específica, compuesta por 500 iteraciones del MCMC. Esta selección se

basa en una ponderación cuidadosa entre el deseo de lograr resultados representativos y la necesidad de manejar eficientemente los costos computacionales involucrados en este análisis exhaustivo Lynch and Western (2004).

La figura 3.4 destaca observaciones cruciales al evaluar la eficacia de nuestro modelo en capturar y reflejar los patrones inherentes a los datos reales. Se observa claramente que los valores se mantienen dentro del intervalo de credibilidad, asegurando una recuperación eficiente de la estructura temporal del modelo. En el contexto de series temporales, este proceso es esencial, ya que potencia la capacidad del modelo para realizar predicciones de manera eficiente. La habilidad para recuperar valores reales no solo favorece el análisis presente, sino que también sienta las bases para proyecciones futuras cercanas. Esta precisión en la representación de la estructura temporal no solo es clave para el análisis actual, sino que también establece un fundamento sólido para predicciones precisas en futuros cercanos.

La figura 3.5, al igual que la figura 3.4, ilustra las observaciones cruciales para evaluar la capacidad de nuestro modelo en capturar y reflejar los patrones inherentes a los datos reales. Esta representación corresponde al algoritmo propuesto por Huerta and Sansó (2007). Se destaca que la recuperación de la estructura temporal en esta instancia es menos efectiva, evidenciando patrones donde la eficacia en la recuperación de los datos reales es débil, y algunos incluso quedan excluidos del intervalo de credibilidad. Por lo tanto, mantenemos la preferencia por nuestro modelo, ya que ofrece resultados más prometedores.

Al profundizar en un análisis más detallado, emerge un aspecto sumamente prometedor: el modelo recupera de manera precisa los parámetros del GEV-DLMs como primer elemento. Además, se refuerza la afirmación anterior sobre la exitosa capacidad del modelo para capturar el comportamiento temporal inherente a los datos reales  $y$  en comparación con los datos generados  $y_{rep}$ . Este éxito se destaca especialmente al considerar el parámetro  $\mu$ . La aplicación de la técnica de suavización basada en el Filtro de Avance y Retroceso (FFBS) emerge como la herramienta clave que posibilita esta recuperación precisa y detallada del aspecto temporal del parámetro  $\mu$ . Este hallazgo no solo valida la robustez del modelo, sino que también resalta la eficacia del enfoque FFBS en el ámbito de la modelización de series temporales, especialmente cuando se trata de componentes dinámicas y temporales de vital importancia.

### 3.4. Análisis de las precipitaciones en Washington

Consideramos el problema de modelar los valores que superen el umbral, para ello, en este trabajo se tomaron los datos de estado de Washington (Estados Unidos) tomando valores de las precipitaciones de 18 estaciones meteorológicas. Las estaciones fueron proporcionadas por Administración Nacional Oceánica y Atmosférica (NOAA, por sus siglas en inglés), agradecemos su valiosa colaboración por proporcionar los datos necesarios para este trabajo. De los datos completos de las precipitaciones diarias por hora seleccionamos las estaciones que tienen datos desde agosto de 2002 a agosto del año 2022 para cada estación seleccionada se tomaron registros diarios en un periodo de 20 años.

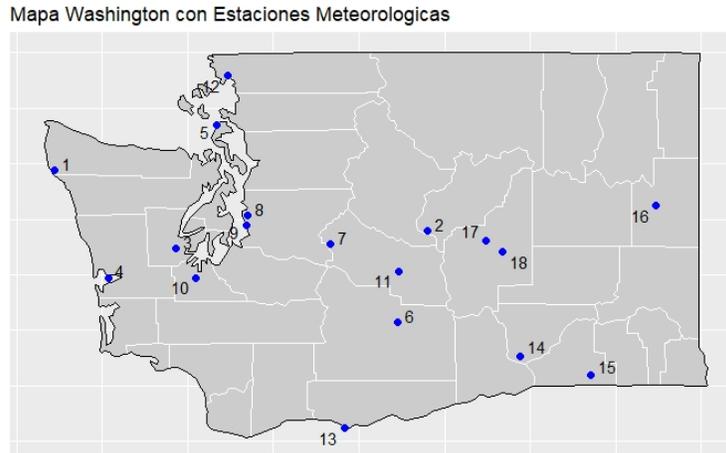


Figura 3.6: Gráfico estado de Washington USA con sus respectivas estaciones meteorológicas.

El estado Washington es nuestro punto de enfoque donde este contiene los datos de 50 estaciones meteorológicas sin embargo, buscamos la misma cantidad de datos en un mismo periodo de tiempo y tenemos que 32 estaciones no proporcionan los datos de manera correspondiente, es decir, muchas de ellas sus periodos de tiempos no eran uniformes entre sí. Por ello se tomó solamente 18 estaciones en las cuales el periodo de tiempo de observación de datos entre estaciones era correspondiente, es decir, mismo periodo de tiempo para todas, la figura 3.6 se muestra en azul cada una de las estaciones para poder desarrollar la estructura del proceso espacio-temporal junto con proceso de convolución. En la tabla 3.3 se muestra el nombre, latitud y longitud de las estaciones estudiadas. Cada una de las estaciones proporciona un total de 414 mediciones por año, es decir, 8208 mediciones para cada una de las 18 estaciones mencionadas.

Es importante recordar en nuestro análisis que no en todos los días del año hay precipitaciones por lo tanto es importante analizar los días en donde las mismas son más elevadas, para ello hacemos uso del cálculo del umbral adecuado, este se tomó haciendo uso del gráfico de vida media residual, ese umbral proporcionará los valores de excedencia para cada estación, ya que los que se toman los que superen el umbral seleccionados. En este punto cada estación tiene un número diferente de registros, ahora nos interesan únicamente aquellos valores que exceden el umbral, recordemos que los datos varían según la estación que estamos estudiando lo que significa que también el umbral cambiará a medida escogemos otra estación.

Naturalmente, se puede extender el modelo presentado en la sección anterior a datos en el espacio y el tiempo. Además, Huerta et al. (2004) presenta un enfoque similar donde podemos hacer uso de un proceso de convolución como se muestra en Higdon (2002). De hecho, las convoluciones de procesos proporcionan una representación lineal conveniente de los procesos Gaussianos. Considere que en el tiempo  $t$  tenemos un vector  $y_t = (y_{1,t}, \dots, y_{n_t,t})$  para el cual se registraron observaciones en los sitios  $s_1, \dots, s_{n_t}$ . Definimos el modelo espacio-tiempo para  $y_t$  mediante las ecuaciones (2.18) y (2.19).

Estaciones	Nombre	Latitud	Longitud
1	QUILLAYUTE STATE AP	47.9375	-124.5550
2	WENATCHEE PANGBORN AP	47.3978	-120.2014
3	SHELTON SANDERSON FLD	47.2381	-123.1408
4	HOQUIAM BOWERMAN AP	46.9728	-123.9303
5	WHIDBEY ISLAND NAS	48.3500	-122.6667
6	YAKIMA AIR TERMINAL	46.5683	-120.5428
7	STAMPEDE PASS	47.2767	-121.3372
8	SEATTLE BOEING FL	47.5303	-122.3008
9	SEATTLE TACOMA INTL AP	47.4444	-122.3139
10	OLYMPIA AP	46.9733	-122.9033
11	ELLENSBURG BOWERS FLD	47.0339	-120.5303
12	BELLINGHAM INTL AP	48.7939	-122.5372
13	THE DALLES MUNI AP	45.6194	-121.1661
14	PASCO TRI CITIES AP	46.2667	-119.1167
15	WALLA WALLA RGNL AP	46.0947	-118.2869
16	SPOKANE INTL AP	47.6217	-117.5281
17	EPHRATA MUNI AP	47.3078	-119.5153
18	MOSES LAKE GRANT CO AP	47.2078	-119.3192

Cuadro 3.3: Estaciones Meteorológicas del estado de Washington U.S. Con sus respectivas latitudes y longitudes.

Partimos de la estructura de los pesos espaciales definidos en 2.6. Inicialmente, Huerta et al. (2004) presenta un enfoque que explora un proceso de convolución con pesos espaciales específicos para cada estación meteorológica. En este sentido, cada peso en el mapa incorpora su propia longitud y latitud. Siguiendo este método, se compara individualmente cada peso con las coordenadas geográficas de las estaciones meteorológicas. La consideración principal es seleccionar el valor mínimo obtenido de estas comparaciones, que se almacena en un vector de pesos mínimos.

En este análisis, nuestra atención se concentra en la ampliación de este proceso de convolución. En específico, llevamos a cabo una exploración profunda de las longitudes y latitudes en relación con cada estación meteorológica. En lugar de simplemente comparar los pesos espaciales, realizamos un proceso de convolución de tipo Gaussiano en las coordenadas longitudinales y latitudinales. Esta aproximación nos permite generar una matriz de auto-correlación que captura las relaciones interdependientes. Cada fila de esta matriz refleja una comparación exhaustiva, donde los núcleos de la estación  $i$  se contrastan con las demás estaciones.

A través de este enfoque, conseguimos analizar con mayor precisión y riqueza la compleja dependencia espacial presente en nuestro modelo. La incorporación de la convolución Gaussiana en el tratamiento de las coordenadas longitudinales y latitudinales nos proporciona un entendimiento más sólido y matizado de las relaciones espaciales, enriqueciendo así nuestra capacidad de modelización y análisis.

La matriz de auto-correlación de pesos es la que se desempeña como vector  $K^t$  de nuestro modelo espacio-temporal, particularmente asociado al DLM. Como las estaciones son medidas de forma independiente, se toma solamente las filas de la matriz, por lo tanto, la matriz  $F$

es de dimensión  $1 \times 18$ , y cada columna contiene los núcleos de la estación 1, respecto a las demás estaciones. Este proceso se hace para cada una de las filas de matriz puesto que cada estación contiene su propia data que debe ser analizada de manera correspondiente.

Las distribuciones a priori y los valores iniciales con las que se iniciaran las cadenas de Markov son una densidad t-student con 5 grados de libertad para la escala del modelo y una densidad normal estándar para la forma. Para generar valores iniciales utilizaremos una densidad log-normal para la escala y una normal estándar para la forma.

Estaciones	Media	Mediana	Sd	MAD	Q5	Q95	Rhat	ess_bulk	ess_tail
1	-0.21	-0.21	0.01	0.01	-0.23	-0.19	1.00	2877.46	3225.93
2	0.04	0.04	0.03	0.03	0.00	0.09	1.00	4810.53	5528.28
3	-0.10	-0.10	0.01	0.01	-0.12	-0.08	1.00	4101.58	4859.98
4	-0.13	-0.13	0.01	0.01	-0.14	-0.11	1.00	4086.31	4685.06
5	-0.10	-0.10	0.01	0.01	-0.12	-0.08	1.00	4237.02	4633.12
6	0.02	0.02	0.03	0.03	-0.02	0.07	1.00	3739.40	3927.91
7	-0.01	-0.01	0.02	0.02	-0.04	0.01	1.00	4895.67	5228.18
8	-0.09	-0.09	0.01	0.01	-0.11	-0.07	1.00	4200.98	4957.61
9	-0.08	-0.08	0.01	0.01	-0.10	-0.06	1.00	4023.40	4589.21
10	-0.11	-0.11	0.01	0.01	-0.13	-0.09	1.00	4401.74	5227.41
11	-0.05	-0.05	0.02	0.02	-0.08	-0.01	1.00	4792.97	5606.37
12	-0.06	-0.06	0.02	0.02	-0.09	-0.03	1.00	4479.99	5434.75
13	-0.15	-0.15	0.01	0.01	-0.17	-0.13	1.00	4237.85	4704.19
14	0.03	0.03	0.02	0.01	0.01	0.05	1.00	5181.11	5489.02
15	-0.14	-0.14	0.01	0.01	-0.16	-0.11	1.00	3529.24	4775.87
16	-0.09	-0.09	0.01	0.01	-0.10	-0.07	1.00	4884.80	5467.90
17	-0.00	-0.00	0.02	0.02	-0.04	0.04	1.00	3869.55	4660.39
18	-0.05	-0.06	0.02	0.02	-0.08	-0.02	1.00	4404.65	5836.95

Cuadro 3.4: Estimación de los parámetros  $\sigma$  para las 18 estaciones meteorológicas. En la segunda columna se presenta la media de los parámetros, en la segunda se determina la mediana. La tercera columna muestra la desviación estándar, la cuarta columna la desviación media absoluta. En las sexta y séptima columnas representamos el intervalo de credibilidad mediante el cuartil 5 y el cuartil 95, respectivamente. La séptima columna indica la convergencia en las cadenas. Las columnas novena y décima muestran los tamaños de muestras efectivos para la distribución a posteriori del modelo.

El cuadro 3.4 muestra las estimaciones de los parámetros de forma  $k$ , la primera columna define cada una de las 18 estaciones meteorológica. La segunda columna contiene la media a posteriori obtenida mediante el algoritmo M-H, representamos mediante el cuantil  $Q5$  y el cuantil  $Q95$  los intervalos de Credibilidad de las estaciones. La columna representada con Rhat define el factor de convergencia el cual aceptamos cuando son cercanas a uno. En el caso del modelo ajustado para los datos de precipitación, todos los Rhat tienen valores aceptables, y concluimos que las estimaciones son aceptables. Una exploración de la distribución a posterior de  $k$  muestra que su mayoría los valores son negativos, esto indica que pertenecen a la familia Weibull. Podemos notar que las estaciones 2, 6 y 14 son positivos esto nos dice que pertenecen a la familia Fréchet, que son importantes para describir eventos de colas pesadas. De igual forma, el cuadro 3.5 muestra las estimaciones de los parámetros de escala  $\sigma$  de cada

una de las 18 estaciones meteorológicas. Los factores de convergencia son cercanos a uno, y los tamaños de muestra efectivos son aceptables, concluyendo que las estimaciones obtenidas para las escalas de las estaciones son aceptables.

Estaciones	Media	Mediana	Sd	MAD	Q5	Q95	Rhat	ess_bulk	ess_tail
1	3.54	3.54	0.11	0.11	3.36	3.71	1.00	3119.61	3332.09
2	0.31	0.31	0.01	0.01	0.29	0.33	1.00	4484.75	5274.34
3	2.39	2.39	0.08	0.08	2.27	2.52	1.00	3656.31	3576.24
4	2.24	2.24	0.07	0.07	2.14	2.36	1.00	4279.69	4095.67
5	0.64	0.64	0.02	0.02	0.61	0.67	1.00	4490.32	4694.12
6	0.32	0.32	0.01	0.01	0.30	0.34	1.00	3790.04	4555.72
7	2.27	2.27	0.08	0.08	2.15	2.41	1.00	4669.16	5496.85
8	1.28	1.28	0.04	0.04	1.21	1.35	1.00	4465.49	4560.22
9	1.41	1.40	0.04	0.04	1.33	1.48	1.00	4197.51	4902.56
10	1.79	1.79	0.06	0.06	1.69	1.88	1.00	4517.99	4958.12
11	0.40	0.40	0.01	0.01	0.38	0.43	1.00	5110.44	5706.56
12	0.57	0.57	0.02	0.02	0.53	0.60	1.00	4651.17	4828.08
13	1.24	1.23	0.04	0.04	1.17	1.30	1.00	4328.36	4717.51
14	0.32	0.32	0.01	0.01	0.31	0.34	1.00	5001.46	6178.90
15	0.78	0.78	0.03	0.03	0.73	0.82	1.00	3484.45	4473.94
16	0.64	0.64	0.02	0.02	0.61	0.68	1.00	4953.89	5171.98
17	0.30	0.30	0.01	0.01	0.28	0.32	1.00	3856.09	4095.03
18	0.33	0.33	0.01	0.01	0.32	0.35	1.00	4325.84	5630.49

Cuadro 3.5: Estimación de los parámetros  $k$  para las 18 estaciones meteorológicas. En la segunda columna se presenta la media de los parámetros, en la segunda se determina la mediana. La tercera columna muestra la desviación estándar, la cuarta columna la desviación media absoluta. En las sexta y séptima columnas representamos el intervalo de credibilidad mediante el cuartil 5 y el cuartil 95, respectivamente. La octava columna indica la convergencia en las cadenas. Las columnas novena y décima muestran los tamaños de muestras efectivos para la distribución a posteriori del modelo.

Para comprender mejor el proceso de estimación, es importante destacar el papel de los parámetros de ubicación, denominado  $\mu$ . Estos parámetros son crucial porque son los que dictan la variación temporal de los datos. En otras palabras, determinan los cambios de las precipitaciones en el tiempo, de cada estación. Las estimaciones de los parámetros de ubicación se pueden evaluar mediante las revisiones a posteriori de las distribuciones predictivas. Esto nos permite comprobar qué tan bien se ajusta nuestro modelo a los datos extremos y si el modelo captura la estructura temporal de los datos. Estas dos características son especialmente relevantes en situaciones en las que las precipitaciones superan ciertos umbrales.

La figura 3.7 representa la recuperación de la estructura temporal del modelo en cada una de las 18 estaciones estudiadas. La estructura temporal se refiere a cómo las precipitaciones varían a lo largo del tiempo en cada ubicación específica. El primer componente de la figura 3.7, representado en un tono azul oscuro y etiquetado como  $y$ , corresponde a los datos observados en cada estación. En otras palabras, son las mediciones reales de precipitación que superaron cierto umbral. Estos datos son la base de nuestro análisis y representan la

realidad que estamos tratando de modelar. El segundo componente, en un tono azul más claro y etiquetado como  $y_{rep}$ , representa cómo la capacidad predictiva del modelo.

Los resultados indican que  $y_{rep}$  es capaz de recuperar de manera efectiva la estructura temporal del modelo. Esto nos proporciona la confianza necesaria para validar el desempeño de nuestro modelo en términos espacio-temporales. Además, ofrecemos intervalos de credibilidad específicos para cada estación, evaluando la incertidumbre de las predicciones obtenidas. Este análisis nos permite comprender mejor cómo nuestro modelo se comporta en diferentes ubicaciones y momentos en el tiempo, lo que es fundamental para la toma de decisiones basadas en datos climáticos.

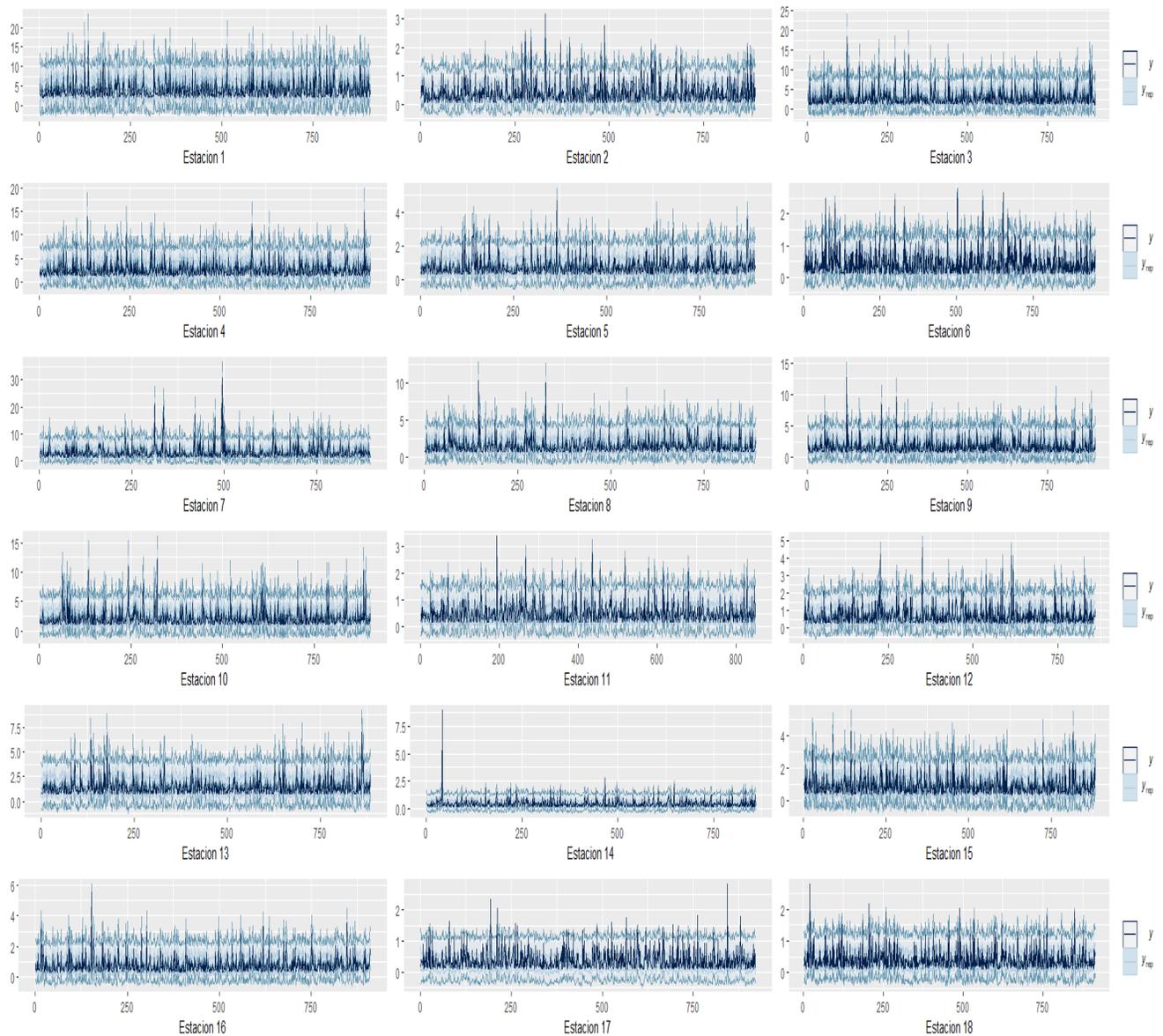


Figura 3.7: Validación del modelo espacio-temporal mediante el uso de datos reales ( $y$ ), representados en color azul oscuro, y datos de entrenamiento para el modelo ( $y_{rep}$ ), los cuales se presentan mediante intervalos en color azul claro.

Para una mayor visualización de las densidades a posterior del modelo espacial. El Apéndice A muestra los gráficos de convergencia del modelo espacial para cada uno de los parámetros de forma y escala, con el objetivo de presentar una mejor visualización de las densidades obtenidas vía MH para todas las 18 estaciones meteorológicas. Se muestran la convergencia de los parámetros de escala  $\sigma$  y forma  $k$  respectivamente. Observamos que las densidad y las trazas de las cadenas, en ambos casos no se distinguen comportamientos anómalos por lo tanto aceptamos el modelo.

Estación	5	10	15	20	25
1	6.461629	8.240257	9.125179	9.698955	10.11686
2	0.745607	0.995233	1.139792	0.995233	1.322794
3	6.178569	7.662997	8.452168	7.662997	9.384644
4	5.657707	6.975541	7.664248	6.975541	8.466927
5	1.597163	1.993304	2.203782	1.993304	2.452362
6	0.846170	1.098622	1.243158	1.098622	1.424444
7	6.321423	7.985747	8.917934	7.985747	10.06660
8	3.314380	4.122885	4.555001	4.122885	5.067765
9	3.668424	4.577904	5.067209	4.577904	5.650921
10	4.587242	5.682501	6.262033	5.682501	6.944214
11	1.016516	1.293440	1.445422	1.293440	1.629654
12	1.458693	1.836952	2.042627	1.836952	2.290085
13	3.049656	3.750662	4.113347	3.750662	4.532666
14	0.875347	1.130854	1.277405	1.130854	1.461489
15	1.850828	2.299749	2.533112	2.299749	2.803939
16	1.632164	2.041439	2.260711	2.041439	2.521412
17	0.752032	0.976372	1.102937	0.976372	1.259806
18	1.367218	1.761978	1.955028	1.761978	2.168426

Cuadro 3.6: Niveles de retorno en un periodo de 5 a 25 años.

Por último, mostramos en la tabla 3.6 muestra las predicciones de los niveles de retorno hacia los siguientes 5, 10, 15, 20 y 25 años, los cuales se espera que ocurra en estos períodos de tiempo específicos, podemos calcularlos estos niveles tal y como lo determina Coles (2013). Esto proporciona información valiosa para diseñar infraestructuras de drenaje, evaluar riesgos de inundación y planificar medidas de adaptación al cambio climático.

# Capítulo 4

## Conclusiones y Trabajos Futuros

### 4.1. Conclusiones

En este trabajo, se presenta una metodología para evaluar datos con valores extremos y estructuras espacio-temporales. Se propone un modelo que es una adaptación del enfoque propuesto por Huerta et al. (2004) con una adaptación del modelo de Nakajima et al. (2012). En este contraste se propone un modelo híbrido que reducen los niveles de latencia del modelo línea dinámica de ? pasando de tres niveles de latencia a un nivel de latencia.

En este contexto para esta metodología se busca recuperar la estructura temporal y estimar los parámetros desconocidos de nuestro modelo, primero filtramos las ubicaciones mediante el algoritmo FFBS y luego estimamos los parámetros adicionales del modelo con un algoritmo de Markov Chain Monte Carlo. Esta metodología nos permitió lograr la convergencia de las estimaciones en tiempos aceptables, obteniendo parámetros identificables y reproducibles, aspecto vital para cuantificar la incertidumbre del modelo propuesto.

Es importante resaltar la implementación del algoritmo suavizado FFBS desarrollado propiamente, este, para la estimación de estados latentes en el tiempo, lo cual asegura que el modelo pueda manejar conjuntos de datos temporales extensos de manera eficiente. Esto permite capturar patrones de dependencia espacial, que son fundamentales en la modelización de eventos climáticos extremos que pueden variar de forma estacional o anual.

Al estimar los parámetros mediante el MCMC, los resultados indicaron que la mayoría de las estaciones pertenecen a la familia Weibull, mientras que algunas son mejor descritas por la familia Fréchet. La validación del modelo mostró que nuestro enfoque es efectivo para recuperar la estructura temporal de las precipitaciones, incluso cuando los datos se distribuyen con distribuciones poco tratables como los son las que pertenecen a la familia de valor extremo generalizado.

El modelo propuesto presenta ventajas significativas en relación a la teoría fundamental, ya que su enfoque combina de manera efectiva dos dimensiones críticas en el análisis de datos climáticos extremos: la variabilidad temporal y la dependencia espacial. Se implementó el proceso de convolución con la una leve diferencia respecto al utilizado por Huerta et al. (2004), en donde se tomaron las coordenadas longitudinales y latitudinales lo cual permite capturar la interdependencia espacial estación por estación. Esta característica genera una matriz

representada por la comparación entre las estaciones, lo cual enriquece nuestra capacidad de modelación. Esta mejora es especialmente relevante para comprender cómo las precipitaciones extremas pueden propagarse o estar correlacionadas con áreas geográficas específicas.

La obtención de resultados confiables en las aproximaciones del modelo es esencial, y el enfoque propuesto muestra un desempeño prometedor en varios aspectos. En primer lugar, se realiza una validación del modelo mediante una revisión de la distribución predictiva. Se generan muestras a partir del modelo utilizando las distribuciones posteriores de los parámetros, destacando la importancia de la media posterior para indicar las ubicaciones de los datos modelados. Nuestro enfoque captura efectivamente el comportamiento temporal de los datos, especialmente en relación a los parámetros de locación ( $\mu$ ). En conclusión, el modelo demuestra una notable capacidad para manejar la dimensión temporal de los datos, proporcionando una base sólida para futuras iteraciones y mejoras, además de una sólida recuperación de los parámetros del GEV-DLM.

Finalmente, destacamos que nuestros resultados demostraron que el enfoque propuesto de modelado fue efectivo el análisis de las precipitaciones extremas en las 18 estaciones meteorológicas estudiadas, para los datos de precipitación de Washington, Estados Unidos, proporcionadas por NOAA. La validación del modelo mostró que nuestro método es capaz de representar adecuadamente cómo las precipitaciones varían a lo largo del tiempo en cada ubicación específica. Esto fortaleció la confianza en la calidad de nuestras estimaciones y resaltó la utilidad práctica de nuestro enfoque.

## 4.2. Trabajos Futuros

- Se plantea la incorporación de covariables climáticas al modelo de precipitación. Se sabe que fenómenos climatológicos como la altitud, la temperatura y la humedad relativa pueden ejercer una influencia significativa en los patrones de precipitación. Por lo tanto, nuestra intención es extender nuestra investigación para comprender mejor cómo estas variables interactúan con la precipitación, mejorando la comprensión de la dinámica climática en nuestra región de estudio.
- La investigación actual se basa evaluar la capacidad de análisis del modelo propuesto con datos de precipitación en el estado de Washington, Estados Unidos, nuestro objetivo a largo plazo es aplicar nuestros modelos a estaciones meteorológicas en Honduras. Reconocemos que la obtención de datos climáticos precisos en este país puede plantear desafíos, pero estamos comprometidos a superar estas dificultades.
- Una parte fundamental de nuestra investigación futura será llevar a cabo una comparación exhaustiva de nuestros modelos con los enfoques previamente propuestos por Nakajima et al. (2012). Estos investigadores han trabajado en el desarrollo de modelos GEV (Valores Extremos Generalizados) con estructuras temporales similares a las nuestras.
- Expandir la teoría de Valores Extremos Generalizados (GEV) con estructuras temporales y volatilidad estocástica con aplicaciones en datos financieros enfocándonos en datos de Honduras. Este nuevo enfoque nos permitirá analizar la variabilidad y la tendencia en los datos financieros en la región.

# Bibliografía

- B. O. Anderson and J. Moore. Optimal Filtering. Dover Publications, illustrated edition, 2005.
- R. P. Barry and J. M. Ver Hoef. Blackbox kriging: Spatial prediction without specifying variogram models. Journal of Agricultural Biological and Environmental Statistics, 1:297, 1996.
- C. N. Behrens, H. F. Lopes, and D. Gamerman. Bayesian analysis of extreme events with threshold estimation. Statistical Modeling, 4:227 – 244, 2004.
- R. Bivand, E. Pebesma, and V. Gómez-Rubio. Applied spatial data analysis with R. 1 2013. doi: 10.1007/978-1-4614-7618-4. URL <https://doi.org/10.1007/978-1-4614-7618-4>.
- S. P. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng. Handbook of markov chain monte carlo. 2011.
- C. Carter and R. Kohn. On gibbs sampling for state space models. Biometrika, 81:541–553, 1994.
- E. Casson and S. Coles. Spatial regression models for extremes. Extremes, 1:449–468, 1999.
- D. Catlin. Estimation, Control, and the Discrete Kalman Filter. Springer, 2011.
- Y. Chou. Spatial autocorrelation analysis and weighting functions in the distribution of wildland fires. International Journal of Wildland Fire, 2:169–176, 1992.
- G. Christakos. Random field models in earth sciences. 1992. URL <https://api.semanticscholar.org/CorpusID:118471193>.
- S. Coles. An Introduction to Statistical Modeling of Extreme Values. Springer Publishing, New York, Estados Unidos, 2013.
- S. Coles and J. A. Tawn. A bayesian analysis of extreme rainfall data. Applied statistics, 45: 463–478, 1996.
- R. M. Cormack and N. Cressie. Statistics for spatial data. Biometrics, 48:1300, 1991.
- N. Cressie. Geostatistics. The American Statistician, 43(4):197, 11 1989. doi: 10.2307/2685361. URL <http://dx.doi.org/10.2307/2685361>.
- N. Cressie. Statistics for spatial data. Biometrics, 50:319, 1993.

- A. C. Davison and R. L. Smith. Models for exceedances over high thresholds. Journal of the royal statistical society series methodological, 52:393–425, 1990.
- P. de Jong. Smoothing and interpolation with the state-space model. 84(408):1085–1088, 1989. doi: 10.1080/01621459.1989.10478876. URL <http://dx.doi.org/10.1080/01621459.1989.10478876>.
- J. Durbin and S. Koopman. Time Series Analysis by State Space Methods: Second Edition (Oxford Statistical Science Series Book 38) (English Edition). OUP Oxford, 2 edition, 2012.
- R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society, 24(2):180–190, 1928. doi: 10.1017/s0305004100015681.
- S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. Journal of Time Series Analysis, 15:183–202, 1994.
- C. Gaetan and M. Grigoletto. Smoothing sample extremes with dynamic models. Extremes, 7:221–236, 2004.
- D. Gamerman and H. F. Lopes. Markov chain monte carlo: Stochastic simulation for bayesian inference, second edition. 2006.
- A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. Bayesian Data Analysis, Third Edition. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013. ISBN 9781439840955. URL <https://books.google.nl/books?id=ZXL6AQAAQBAJ>.
- C. J. Geyer. Practical markov chain monte carlo. Statistical Science, 7:473–483, 1992.
- E. Gilleland, D. W. Nychka, and U. Schneider. Spatial models for the distribution of extremes. 2005.
- R. Giraldo. Introducción a la geoestadística-teoría y aplicación. Technical report, Universidad Nacional de Colombia, 2002.
- B. Gnedenko. Sur la distribution limite du terme maximum d’une serie aleatoire. The Annals of Mathematics, 44(3):423, 1943. doi: 10.2307/1968974.
- A. Harvey. Forecasting, Structural Time Series Models and the Kalman Filter: Volume 0. Cambridge University Press, revised ed. edition, 1991.
- W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. Biometrika, 57:97–109, 1970.
- D. M. Higdon. Space and space-time modeling using process convolutions. 2002.
- D. M. Higdon. A process-convolution approach to modelling temperatures in the north atlantic ocean. Environmental and Ecological Statistics, 5:173–190, 2004.
- D. M. Higdon, J. L. Swall, and J. C. Kern. Non-stationary spatial modeling. 1998.

- M. D. Hoffman and A. Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. J. Mach. Learn. Res., 15:1593–1623, 2011.
- R. J. Howarth. Mining geostatistics. Mineralogical Magazine, 43:563 – 564, 1979.
- H.-C. Huang and N. Cressie. Spatio-temporal prediction of snow water equivalent using the kalman filter. Computational Statistics & Data Analysis, 22:159–175, 1996.
- G. Huerta and B. Sansó. Time-varying models for extreme values. Environmental and Ecological Statistics, 14(3):285–299, 7 2007. doi: 10.1007/s10651-007-0014-3. URL <https://doi.org/10.1007/s10651-007-0014-3>.
- G. Huerta, B. Sansó, and J. R. Stroud. A spatiotemporal model for mexico city ozone levels. Journal of the Royal Statistical Society: Series C (Applied Statistics), 53, 2004.
- K. Ickstadt, , and R. L. Wolpert. Spatial regression for marked point processes. 2008.
- E. H. Isaaks and M. Srivastava. An Introduction to Applied Geostatistics. Oxford University Press, 10 2022.
- R. Kalman and J. Bertram. Control system analysis and design via the “second method” of lyapunov: I—continuous-time systems. 82(2):371–393, 1960. doi: 10.1115/1.3662604. URL <http://dx.doi.org/10.1115/1.3662604>.
- A. D. Kennedy. Hybrid monte carlo. 1987.
- S. Koopman. Disturbance smoother for state space models. 80(1):117–126, 1993. doi: 10.1093/biomet/80.1.117. URL <http://dx.doi.org/10.1093/biomet/80.1.117>.
- Y. Lai. Sequential bayesian inference for dynamic linear models of sensor data. 2019.
- J. Lindsten. Backward Simulation Methods for Monte Carlo Statistical Inference. Now Publishers Inc, illustrated edition, 8 2013.
- S. M. Lynch and B. Western. Bayesian posterior predictive checks for complex models. Sociological Methods & Research, 32:301 – 335, 2004.
- K. V. Mardia, C. R. Goodall, E. Redfern, and F. J. Alonso. The kriged kalman filter. Test, 7:217–282, 1998.
- G. Matheron. Principles of geostatistics. Economic Geology, 58:1246–1266, 1963.
- N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. Journal of Chemical Physics, 21:1087–1092, 1953.
- J. Nakajima, T. Kuniyama, Y. Omori, and S. Frühwirth-Schnatter. Generalized extreme value distribution with time-dependence using the ar and ma models in state space form. Comput. Stat. Data Anal., 56:3241–3259, 2012.
- NOAA. NOAA: National Oceanic and Atmospheric Administration. URL <https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt>. Data.

- P. Petitgas. Geostatistics and their applications to fisheries survey data. 1996.
- J. Pickands. The two-dimensional poisson process and extremal processes. Journal of Applied Probability, 8:745 – 756, 1971.
- M. Plummer, N. Best, K. Cowles, K. Vines, D. Sarkar, D. M. Bates, R. G. Almond, and A. Magnusson. Output analysis and diagnostics for mcmc [r package coda version 0.19-4]. 2020.
- A. Pole, M. West, and J. Harrison. Applied Bayesian Forecasting and Time Series Analysis. Chapman and Hall/CRC, 1 edition, 9 1994.
- H. Rauch, F. Tung, and C. Striebel. Maximum likelihood estimates of linear dynamic systems. 3(8):1445–1450, 1965. doi: 10.2514/3.3166. URL <http://dx.doi.org/10.2514/3.3166>.
- G. O. Roberts and J. S. Rosenthal. Optimal scaling for various metropolis-hastings algorithms. Statistical Science, 16:351–367, 2001.
- G. O. Roberts and A. F. M. Smith. Simple conditions for the convergence of the gibbs sampler and metropolis-hastings algorithms. Stochastic Processes and their Applications, 49:207–216, 1994.
- G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk metropolis algorithms. Annals of Applied Probability, 7:110–120, 1997.
- B. Sansó and L. Guenni. Venezuelan rainfall data analysed by using a bayesian space–time model. Journal of the Royal Statistical Society: Series C (Applied Statistics), 48, 1999.
- S. C. Segura. Evaluación geoestadística de la energía geotérmica en marangani - cusco. 2018. URL <https://api.semanticscholar.org/CorpusID:164782174>.
- R. L. Smith. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. Statistical Science, 4:367–377, 1989.
- R. L. Smith, J. A. Tawn, and S. Coles. Markov chain models for threshold exceedances. Biometrika, 84:249–268, 1997.
- M. L. Stein. Statistical interpolation of spatial data: Some theory for kriging. 1999.
- A. Stephenson and M. Ribatet. A user’s guide to the evdbayes package (version 1.1). 2006.
- A. Stephenson and M. Ribatet. evdbayes: Bayesian Analysis in Extreme Value Theory, 2023. URL <https://CRAN.R-project.org/package=evdbayes>. R package version 1.1-3.
- H. J. Thiebaux. Anisotropic correlation functions for objective analysis. 104(8):994–1002, 1976. doi: 10.1175/1520-0493. URL <http://journals.ametsoc.org/doi/10.1175/1520-0493>.
- H. J. Thiebaux and M. Pedder. Spatial objective analysis: with applications in atmospheric science. Academic Press, AP, 1987.
- L. Tierney. Markov chains for exploring posterior distributions. Annals of Statistics, 22: 1701–1728, 1994.

- S. F. Tonellato. A bayesian approach to the analysis of spatial time series. 1998.
- K. Triantafyllopoulos. Bayesian Inference of State Space Models: Kalman Filtering and Beyond. Springer, 1st ed. 2021 edition, 11 2021.
- J. M. Ver Hoef and R. P. Barry. Constructing and fitting models for cokriging and multivariable spatial prediction. Journal of Statistical Planning and Inference, 69:275–294, 1998.
- J. M. Ver Hoef, N. Cressie, and R. Barry. Flexible spatial models for kriging and cokriging using moving averages and the fast fourier transform (fft). Journal of Computational and Graphical Statistics, 13:265 – 282, 2004.
- H. Wackernagle. Multivariate geostatistics: an introduction with applications. 1998.
- A. Warrick and D. Myers. Optimization of sampling locations for variogram calculations. Water Resources Research, 23(3):496–500, 3 1987. doi: 10.1029/wr023i003p00496. URL <http://dx.doi.org/10.1029/wr023i003p00496>.
- M. West and J. Harrison. Bayesian Forecasting and Dynamic Models (Springer Series in Statistics). Springer-Verlag, February 1997. ISBN 0387947256.
- C. K. Wikle and N. Cressie. A dimension-reduced approach to space-time kalman filtering. Biometrika, 86:815–829, 1999.
- L. Zadeh and C. Desoer. Linear System Theory: The State Space Approach (Dover Civil and Mechanical Engineering). Dover Publications, first thus edition, 2008.

# A Apéndice A

Gráficos para cada estaciones meteorológicas, mostramos la funciones de densidad del parámetro de forma y escala respectivamente. Además, de visualizaciones de las cadenas generadas mediante el algoritmo de Metropolis-Hasting.

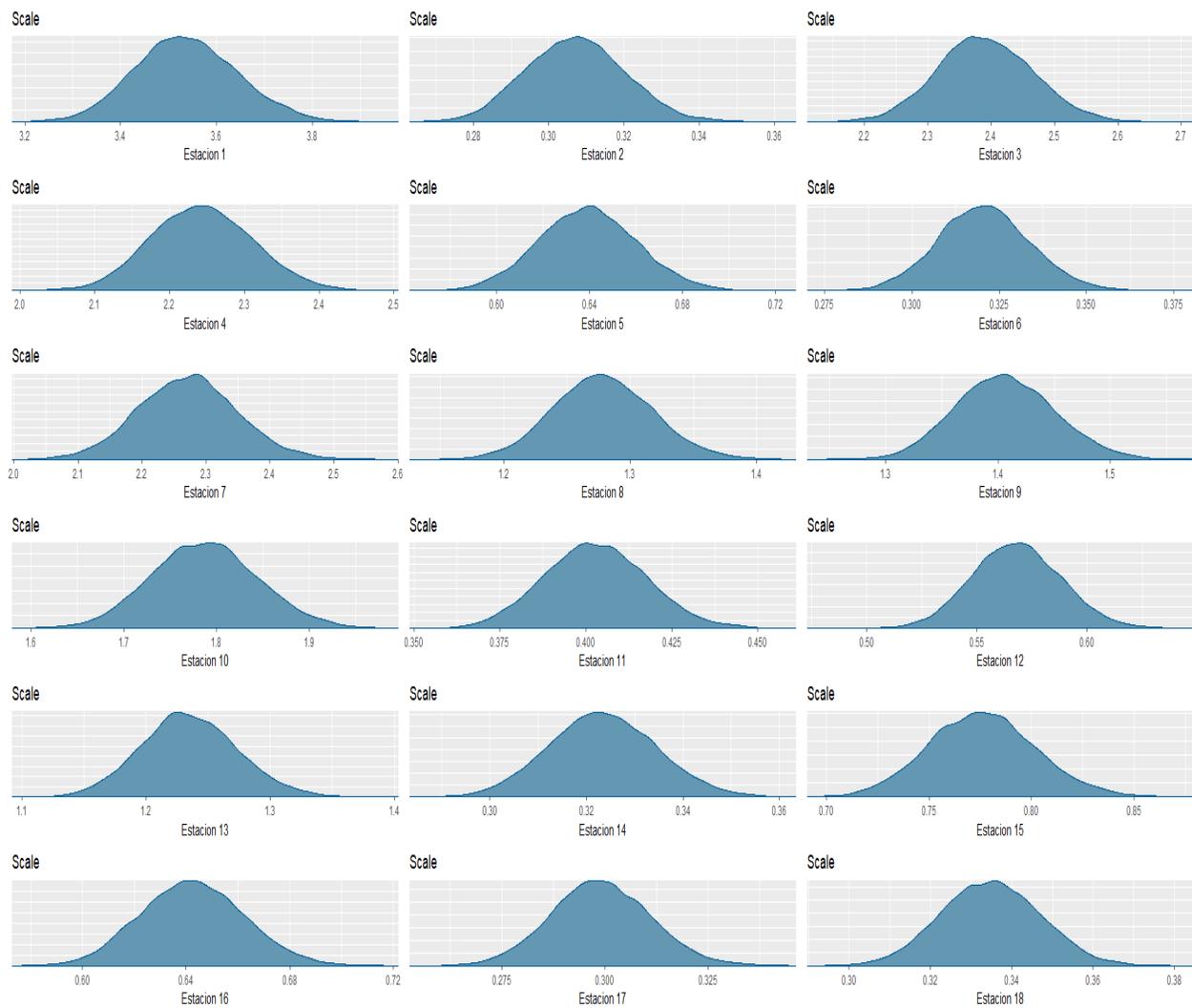


Figura A.1: Funciones de densidad para las 18 estaciones para el parámetro de escala.

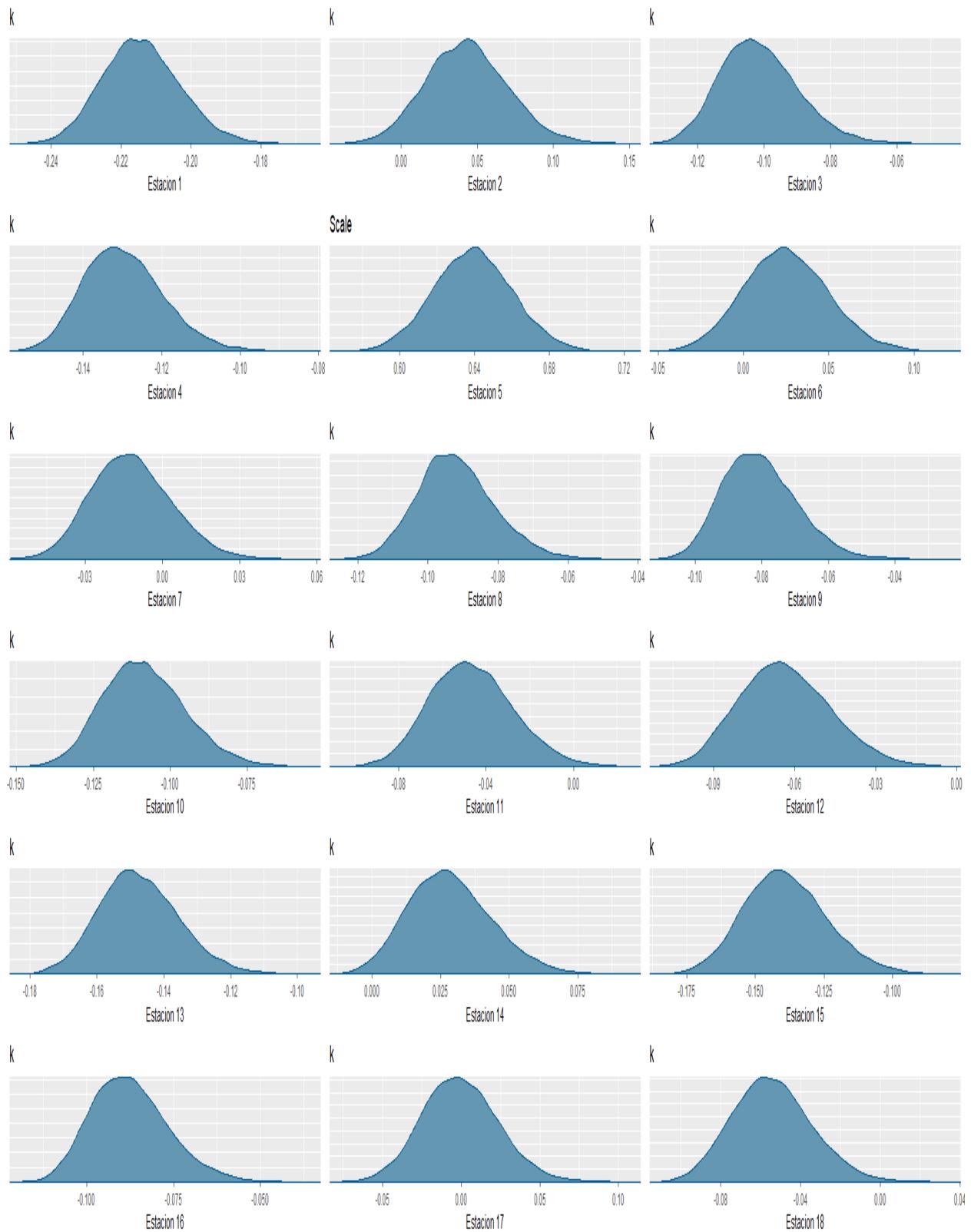


Figura A.2: Funciones de densidad para las 18 estaciones para el parámetro de forma.

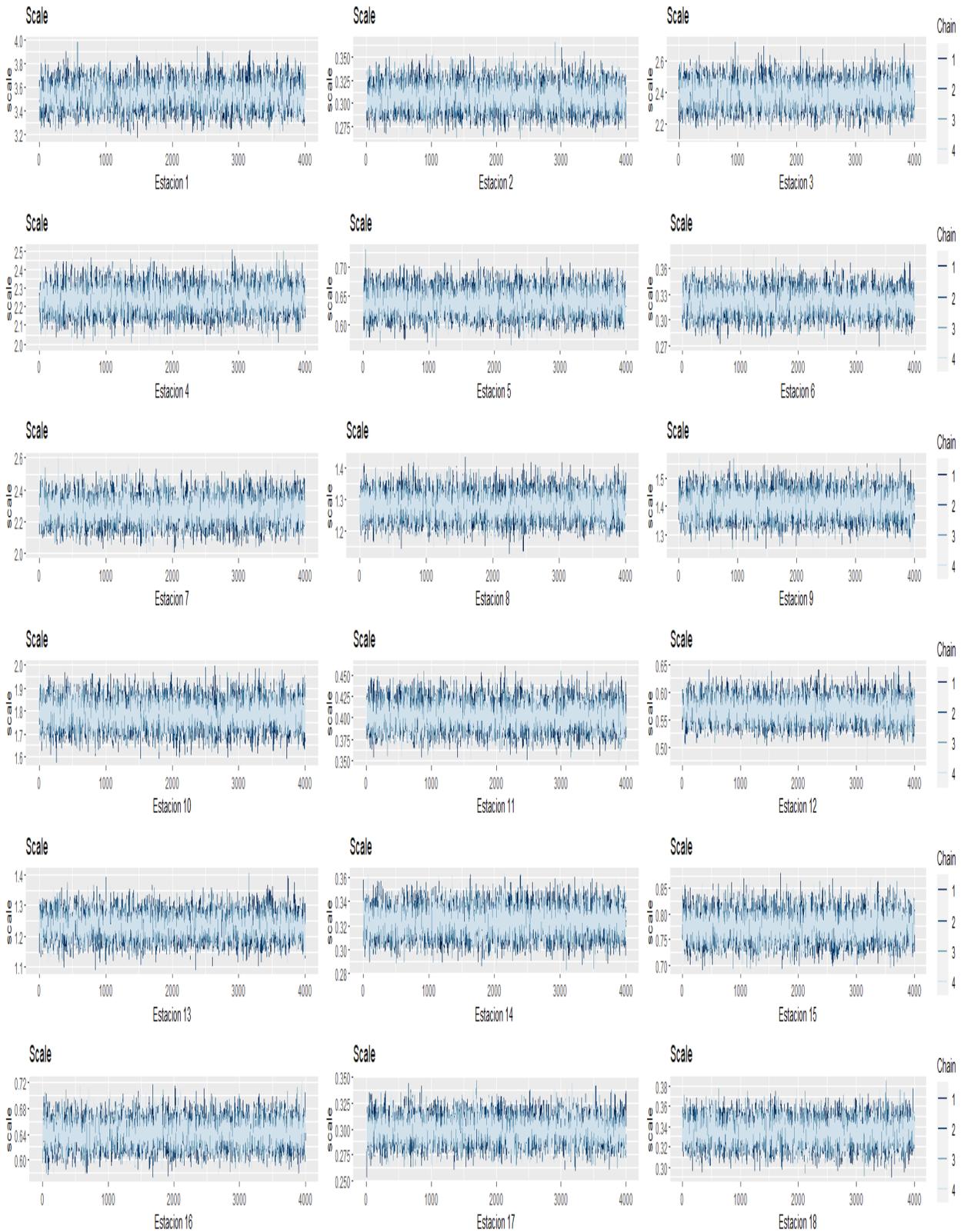


Figura A.3: Cadenas para las 18 estaciones para el parámetro de escala.

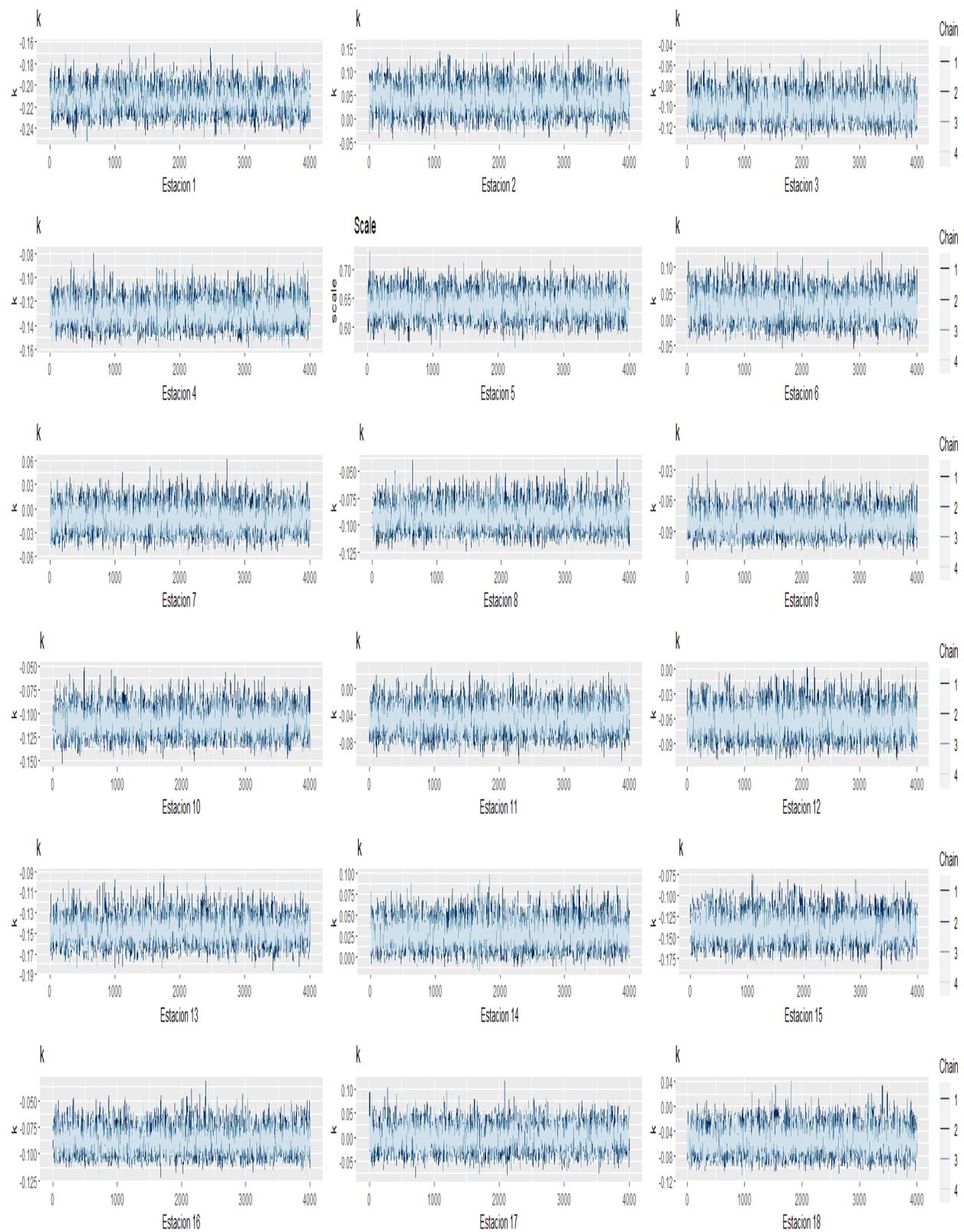


Figura A.4: Cadenas para las 18 estaciones para el parámetro de forma.