



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Análisis Bayesiano para las Máximas
Concentraciones de Ozono de la Ciudad
de México

T E S I S

QUE PARA OPTAR POR EL GRADO DE:

Actuaria

PRESENTA:

Diana Alvarado Lima

DIRECTORA DE TESIS:

Dra. Lizbeth Naranjo Albarrán



Ciudad Universitaria, CDMX, 2018



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

-
1. Datos del alumno
Alvarado
Lima
Diana
55 14 29 20 31
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
311129658
 2. Datos del tutor
Dra.
Lizbeth
Naranjo
Albarrán
 3. Datos del sinodal 1
Dra.
Eliane R.
Rodrigues
 4. Datos del sinodal 2
Dr.
Carlos
Díaz
Ávalos
 5. Datos del sinodal 3
Dra.
Ruth
Selene
Fuentes
García
 6. Datos del sinodal 4
Act.
Jaime
Vázquez
Alamilla
 7. Datos del trabajo escrito
Análisis Bayesiano para las Máximas Concentraciones de Ozono de la Ciudad de México
78p.
2018

*El éxito no se logra con la suerte
es el resultado de un esfuerzo constante.*

Agradecimientos

Un trabajo de investigación es siempre producto de ideas y esfuerzos previos pero también es fruto del apoyo incondicional que nos brindan las personas que nos quieren, sin el cual no tendríamos la fuerza que nos alienta a seguir adelante.

A mi mamá Rosy. Gracias mamá por siempre guiarme a lo largo de mi vida, por darme fortaleza incluso en los momentos más difíciles, por todos los sacrificios que hiciste para que yo llegaré hasta aquí, eres un ejemplo de vida y eres la mujer que más admiro en el mundo, te amo con todo mi ser.

A mi papá Rogelio. Gracias papá porque a pesar de todo siempre estuviste con nosotros y nunca nos dejaste, por tu apoyo y por entender el concepto de responsabilidad por seguir luchando y no rendirte ante esto llamado vida, gracias por ser un buen padre, te amo papá.

A mis hermanos, Itzel y Carlos. Gracias Itzel por demostrarme que se puede luchar ante cualquier adversidad, por siempre tener una sonrisa y brindarme una en mis peores momentos, por ser fuerte, responsable y muy trabajadora, gracias por ser mi hermana gemela. Gracias Carlitos por escucharme, aconsejarme y darme ánimos cuando sentía que ya no podía, ser responsable y dedicado te llevará a lograr todas tus metas. Gracias hermanos me siento muy orgullosa de ustedes y no puedo describir el amor que siento por ustedes.

A mis abuelitos Blandina y Victor por su apoyo incondicional y ser un ejemplo de vida y perseverancia, por mantener a la familia unida, los quiero mucho. En general gracias a toda la familia Lima por apoyarme y ser mi familia en toda la extensión de la palabra, los amo a todos.

A Kaly por ser más que mi mejor amiga, por apoyarme incondicionalmente y estar ahí siempre, por motivarme a estudiar la mejor carrera que pude haber elegido, gracias por formar parte de mi vida, acuérdate que vamos por los 100 años te quiero mucho.

A mis amigos de la preparatoria con los que compartí experiencias académicas y de vida, porque de cada uno aprendí algo y con los años perdura la amistad, siempre los llevo conmigo. También agradezco a mis amigos de la Facultad con los que recorrí esta travesía llamada Actuaría son grandes personas, grandes amigos y grandes actuarios. En ambos casos ustedes saben quiénes son.

A mis maestros Sandra, Ismael, Rocio, Lilia, Graciela y Olivia que siempre me motivaron y me transmitieron el gusto por aprender, que confiaron en mí desde muy pequeña. También quiero agradecer al profesor Javier Fernández García por enseñarme a nunca tirar la toalla antes de tiempo y a Ruth S. Fuentes García por transmitirme el gusto por la hermosa estadística.

Asimismo agradezco a mis sinodales, la Dra. Ruth S. Fuentes, la Dra. Eliane Rodrigues, al Dr. Carlos Díaz y al Act. Jaime Vázquez por sus correcciones, comentarios y recomendaciones que fueron fundamentales para enriquecer y culminar este trabajo, sepan que les guardo mucho respeto y admiración.

Por último y no menos importante, al contrario, estoy infinitamente agradecida con la directora de esta tesis, la Dra. Lizbeth Naranjo Albarrán. Gracias Liz por la dirección, dedicación y apoyo que me brindaste a lo largo de este trabajo por contagiarme el amor por la estadística y por compartir tus conocimientos conmigo. Gracias por confiar en mí y motivarme, por tu paciencia y tu tiempo pero sobre todo gracias por tu amistad que así me hiciste sentir siempre, te admiro y te respeto mucho, eres una excelente persona y una gran profesora, te quiero mucho.

Gracias a la UNAM y a la Facultad de Ciencias por darme todo, conocimientos, sueños, amigos y experiencias de vida y en general gracias a todos los que hicieron posible este trabajo.

Con cariño, Diana.

Índice general

1. Introducción	1
2. Paradigma Bayesiano	3
2.1. Conceptos Básicos	3
2.1.1. Regla de Bayes	3
2.1.2. Distribución a Priori	4
2.1.3. Estimación Puntual	6
2.1.4. Funciones de Pérdida	6
2.1.5. Intervalos de Credibilidad	7
2.1.6. Predicción	8
2.1.7. Comparación de Modelos	8
2.2. Métodos Monte Carlo Vía Cadenas de Markov	10
2.2.1. Algoritmo Metropolis Hastings	10
2.2.2. Muestreo de Gibbs	11
2.2.3. Diagnóstico de Convergencia	12
2.3. Modelos Jerárquicos Bayesianos	13
2.3.1. Estructura General	13
2.3.2. Intercambiabilidad	14
2.3.3. Inferencia	15
2.3.4. Predicción	15
3. Valores Extremos	17
3.1. Modelos Asintóticos	17
3.2. Teorema de Valores Extremos	18
3.3. Distribución Generalizada de Valores Extremos	19
3.3.1. Inferencia para la Distribución <i>GEV</i>	20
3.3.2. Niveles de Retorno	22
3.3.3. Características de la Distribución <i>GEV</i>	23
3.4. Inferencia Bayesiana en Valores Extremos	24
4. Datos de Ozono en la Ciudad de México	27
4.1. Descripción de los Datos	27
4.2. Estación Coyoacán	34

ÍNDICE GENERAL

4.3. Valores Faltantes	36
4.4. Modelo Bayesiano para los Datos de Ozono en la Estación Coyoacán . . .	39
5. Valores Extremos en Datos Dependientes	45
5.1. Valores Extremos en Secuencias Estacionarias	45
5.2. Valores Extremos en Secuencias No Estacionarias	46
5.3. Modelo Bayesiano Jerárquico Temporal para la Estación Coyoacán . . .	48
6. Conclusiones	57
A. Apéndice	61
A.1. Introducción a BUGS	61
A.2. Introducción a JAGS	64
A.3. Códigos	65
A.3.1. Modelo 1	65
A.3.2. Modelo 2	67
Bibliografía	69

Capítulo 1

Introducción

Seguramente hemos oído o leído la palabra *contaminación ambiental*, pero el hecho de saber qué es supone algo muy importante, ya que estar informado es la mejor manera de evitarla o al menos tratar de disminuirla. Se entiende por *contaminación ambiental* a la alteración del estado natural de un medio por cualquier agente de origen biológico, físico o químico, riesgoso para la salud de los seres humanos, los animales y las plantas.

Hay distintos tipos de contaminación ambiental y se clasifican de acuerdo al tipo de variables que se consideren, las variables más importantes son el tipo de contaminante que lo produce y el medio afectado. Cuando se habla del tipo de contaminante se refiere a que pueden ser contaminantes físicos, químicos o incluso biológicos; y respecto al medio afectado se tienen tres: el agua, suelo y aire, las tres son importantes, sin embargo, para fines de este trabajo solo describiremos la contaminación del aire.

La contaminación del aire sucede cuando se incorporan a la atmósfera gases tóxicos como monóxido de carbono (CO), dióxido de azufre (SO₂), ozono (O₃), óxido de nitrógeno (NO_x) o partículas suspendidas alterando la composición natural del aire que respiramos. Las consecuencias de la contaminación ambiental suelen ser irreversibles y son un problema a nivel mundial, México no es la excepción siendo uno de los principales países con problemas de contaminación.

La presente tesis se enfoca en el contaminante ozono (O₃) y tiene como objetivo ajustar la distribución Generalizada de Valores Extremos a las concentraciones máximas de ozono que reportan las estaciones de monitoreo de la Ciudad de México, haciendo uso de herramientas estadísticas descriptivas y de inferencia, con un enfoque Bayesiano. La idea principal es modelar la dependencia temporal de este contaminante durante los últimos 5 años y obtener resultados consistentes. Mi principal motivación es aplicar mis conocimientos adquiridos durante mis cursos de estadística para analizar una problemática real del país.

El contenido de la tesis se estructura en 5 capítulos que se describirán a continuación.

1. INTRODUCCIÓN

En el segundo capítulo se presentan los conceptos básicos del paradigma Bayesiano que servirán como herramienta para el desarrollo de este trabajo. Se describe el uso del Teorema de Bayes en la estadística Bayesiana y los conceptos para hacer inferencia desde un enfoque Bayesiano. Se describen los métodos Monte Carlo vía Cadenas de Markov que serán útiles para obtener las cantidades de interés a posteriori cuando es difícil muestrear directamente de la distribución a posteriori. Se explican de manera muy general los algoritmos más conocidos como el algoritmo de Metropolis Hastings y el de Gibbs, así como algunas medidas para revisar la convergencia de las cadenas simuladas.

En el tercer capítulo se presenta un panorama general de la teoría de valores extremos enfocándonos en la distribución del máximo, como son los modelos asintóticos y la convergencia a las distribuciones de valores extremos por medio del teorema de valores extremos. Se presenta la distribución Generalizada de Valores Extremos (*GEV* por sus siglas en inglés), cómo hacer inferencia sobre los parámetros y las características estadísticas más importantes de esta distribución.

En el cuarto capítulo se describen los datos de ozono de la Ciudad de México, se da un panorama general del monitoreo de este contaminante en la Ciudad de México, se menciona la Norma de contaminación ambiental y los límites permisibles. Se muestra un análisis descriptivo de los datos en todas las estaciones de monitoreo y se selecciona la estación Coyoacán como ejemplo, se trata el problema de los valores faltantes y se presenta un método de imputación para resolver este problema. Por último se incluye un primer modelo en el que se ajusta la distribución Generalizada de Valores Extremos y la estimación de los parámetros se realiza desde una perspectiva Bayesiana. Es importante mencionar que en este primer modelo los parámetros son constantes en el tiempo.

Finalmente en el capítulo cinco se muestran los modelos de valores extremos para secuencias de datos dependientes en procesos estacionarios y no estacionarios, algunas generalidades de éstos y cómo se hace la estimación de los parámetros de la distribución *GEV* en estos casos. En la última sección se encuentra un segundo modelo que trata de explicar la dependencia temporal de los datos de ozono a través de una jerarquía en el parámetro de localización.

Por último se presentan las conclusiones de esta tesis y también se incluye un Apéndice que contiene los códigos de los programas con los que se obtuvieron los resultados en este trabajo.

Capítulo 2

Paradigma Bayesiano

La inferencia estadística es el conjunto de métodos y técnicas matemáticas que permiten deducir conclusiones a partir de los datos observados de una muestra y extrapolar a toda la población. Distinguimos dos enfoques para hacer inferencia: el frecuentista y el Bayesiano. Usualmente se trabaja bajo un enfoque frecuentista, sin embargo, actualmente se ha divulgado mucho el uso de la inferencia Bayesiana.

La inferencia Bayesiana es el proceso de ajustar un modelo de probabilidad a un conjunto de datos y resumir el resultado mediante una distribución de probabilidad en los parámetros del modelo y en cantidades no observadas, así como obtener predicciones para nuevas observaciones. En tanto que los métodos frecuentistas se cuestionan qué nos dicen los datos acerca del valor del parámetro, los métodos Bayesianos indagan directamente en cómo se modifica el estado de la información acerca del valor del parámetro con los datos observados. Además el enfoque Bayesiano permite realizar declaraciones de probabilidad directas sobre los parámetros.

2.1. Conceptos Básicos

A nivel conceptual, una de las principales ventajas del enfoque Bayesiano es la facilidad con la que se ponen en práctica las ideas básicas de este enfoque. ¹

2.1.1. Regla de Bayes

Denotamos a θ como un vector de parámetros desconocidos y sea $x = (x_1, x_2, \dots, x_n)$ una muestra de datos. Si queremos hacer afirmaciones de probabilidad sobre θ dado x debemos comenzar con un modelo que proporcione una distribución de probabilidad

¹Este capítulo es una introducción a los conceptos básicos de la Estadística Bayesiana, se han escrito libros introductorios a estos temas como [11] y [21] y posteriormente se puede profundizar en [15].

2. PARADIGMA BAYESIANO

conjunta para θ y para x , ésta puede escribirse como el producto de dos densidades que a menudo se conocen como la distribución a priori $p(\theta)$ y la distribución de los datos o verosimilitud $p(x|\theta)$, respectivamente

$$p(\theta, x) = p(\theta)p(x|\theta),$$

ahora simplemente condicionando el valor conocido de los datos y usando la propiedad de probabilidad condicional conocida como regla de Bayes se obtiene la densidad a posteriori

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{p(x)},$$

donde $p(x) = \sum_{\Theta} p(\theta)p(x|\theta)$, la suma sobre todos los posibles valores de θ o $p(x) = \int_{\Theta} p(\theta)p(x|\theta)d\theta$ en el caso de que θ sea continuo.

Una forma equivalente de la expresión anterior es omitiendo el factor $p(x)$, que no depende de θ y con x fijo puede considerarse una constante, así tenemos de lado derecho la densidad posteriori no normalizada

$$p(\theta|x) \propto p(\theta)p(x|\theta).$$

El término $p(x|\theta)$ se toma aquí como una función de θ y no de x . En el enfoque Bayesiano toda la inferencia se hace a partir de la distribución a posteriori.

2.1.2. Distribución a Priori

El Teorema de Bayes nos dice cómo actualizar nuestras creencias sobre θ con los datos observados, podemos considerarlo como un método general de inducción:

Distribución a priori + Datos \rightarrow Distribución a posteriori

Se basa en la interpretación subjetiva de la probabilidad, para ello utiliza la distribución a priori que es la percepción existente por parte del investigador como una variable que modifica los datos dando lugar a la distribución a posteriori. La distribución a priori es aquella que proporciona los parámetros para modelar los datos observados a partir de información previa existente, por ejemplo estudios realizados sobre el evento de interés, es decir, pretende aportar información adicional.

Una pregunta que debería surgir es ¿cómo construir la distribución a priori? Existen distribuciones a priori informativas que se basan en estudios empíricos previos o por conocimientos del investigador y las no informativas como las impropias, las de Jeffrey's o distribuciones poco informativas. A continuación se describirán de manera muy general.

Distribuciones a Priori Informativas

Una distribución a priori informativa es aquella que no es dominada por la verosimilitud y tiene un impacto sobre la distribución a posteriori y es útil cuando se tiene información previa de estudios anteriores.

Distribuciones a Priori No Informativas

Una distribución a priori es no informativa cuando tiene un mínimo impacto sobre la distribución a posteriori de θ .

Distribuciones a Priori Impropias

Una distribución a priori propia es aquella que no depende de los datos y su integral es igual a uno, es decir,

$$\int_{\Theta} p(\theta) d\theta = 1.$$

Por lo tanto una distribución a priori es impropia si

$$\int_{\Theta} p(\theta) d\theta = \infty.$$

Distribución a Priori de Jeffreys

La distribución a priori de Jeffreys es

$$p(\theta) \propto [I(\theta)]^{1/2},$$

donde $I(\theta)$ es la información de Fisher para θ

$$I(\theta) = E \left[\left(\frac{d}{d\theta} \log(p(x|\theta)) \right)^2 \right].$$

En la mayoría de los casos, la distribución a priori de Jeffreys es una distribución a priori impropia, sin embargo la distribución a posteriori si es propia¹.

El uso de ciertas distribuciones a priori suele ser ventajoso al obtener la distribución a posteriori y a este tipo de distribuciones se les conoce como distribuciones conjugadas. Suponga que una distribución a priori pertenece a cierta familia de distribuciones paramétricas D entonces esa distribución es conjugada respecto a la verosimilitud si la distribución a posteriori pertenece a la familia D de distribuciones, es decir, se mantiene dentro de la misma familia de distribuciones paramétricas y los valores muestrales sólo

¹Para estudiar con detalle las distribuciones a priori no informativas y revisar ejemplos consulte [11].

2. PARADIGMA BAYESIANO

afectan los valores de los parámetros de la distribución. En la Tabla 2.1 se muestran algunas distribuciones conjugadas.

Familia	Priori Conjugada
$Binomial(n, \theta)$	$\theta \sim Beta(a, b)$
$Poisson(\theta)$	$\theta \sim Gamma(\alpha_0, \lambda_0)$
$N(\mu, \sigma^2)$, σ^2 conocido	$\mu \sim N(\mu_0, \sigma_0^2)$
$N(\mu, \sigma^2)$, μ conocido	$1/\sigma^2 \sim Gamma(\alpha_0, \lambda_0)$
$Gamma(\alpha, \lambda)$, α conocido	$\lambda \sim Gamma(\alpha_0, \lambda_0)$
$Beta(a, b)$, b conocido	$\lambda \sim Gamma(\alpha_0, \lambda_0)$

Tabla 2.1: Distribuciones a priori conjugadas para una función de verosimilitud de las familias estándar.

El uso de distribuciones conjugadas facilita los cálculos pero no necesariamente son las mejores elecciones a priori pues depende del contexto de los datos y se pueden usar otras alternativas como las distribuciones a priori antes mencionadas.

2.1.3. Estimación Puntual

La estimación se vuelve un problema de decisión en el que elegimos $\hat{\theta}$ como estimador de θ . Asociada con cada estimador hay una pérdida que refleja la diferencia entre θ y $\hat{\theta}$. Se determina la función de pérdida como $L(\theta, \hat{\theta})$ que cuantifica la posible penalización en estimar θ por $\hat{\theta}$, sin embargo, θ es desconocido, solo tenemos su distribución a posteriori $p(\theta|x)$ por lo tanto minimizaremos la pérdida esperada a posteriori

$$\min_{\hat{\theta}} E[L(\theta, \hat{\theta})|x] = \int_{\Theta} L(\theta, \hat{\theta})p(\theta|x)d\theta,$$

el estimador Bayesiano será

$$\hat{\theta} = \arg \min_{\hat{\theta}} E[L(\theta, \hat{\theta})|x].$$

2.1.4. Funciones de Pérdida

Existen distintas funciones de pérdida y la selección particular dependerá del contexto del problema pero las dos más utilizadas se presentan a continuación:

- *Pérdida cuadrática:*

$$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2,$$

el estimador posteriori Bayesiano es la media a posteriori

$$\hat{\theta} = E(\theta|x) = \int_{-\infty}^{\infty} \theta \cdot p(\theta|x) d\theta.$$

- *Pérdida de error absoluto:*

$$L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|,$$

el estimador Bayesiano posteriori es la mediana a posteriori

$$\hat{\theta} = \int_{-\infty}^{\hat{\theta}} p(\theta|x) d\theta = 0.5.$$

2.1.5. Intervalos de Credibilidad

Los estimadores puntuales no proporcionan una medida de la precisión de la estimación y a menudo es deseable identificar regiones del espacio parametral que probablemente contengan el verdadero valor del parámetro. El paradigma Bayesiano propone una noción de intervalos de confianza que es más natural que su contraparte frecuentista y estos se llaman intervalos de credibilidad.

A partir de la distribución a posteriori se puede determinar el intervalo de los valores del parámetro θ que son a posteriori más plausibles con probabilidad $(1 - \alpha)$ donde usualmente se toma $\alpha = 0.05$. Formalmente $[a, b]$ es un intervalo de credibilidad al $100(1 - \alpha)\%$ para θ si

$$p(a \leq \theta \leq b|x) = \int_a^b p(\theta|x) d\theta = 1 - \alpha. \quad (2.1)$$

La interpretación de este intervalo es que describe su información sobre la ubicación del verdadero valor de θ después de haber observado $X = x$. La expresión (2.1) no define de manera única un intervalo de credibilidad, cualquier intervalo con probabilidad $(1 - \alpha)$ cumple lo anterior entonces vamos a tener una infinidad de intervalos. Esto sugiere preguntarse, ¿con cuál de estos intervalos trabajar? Hay varias maneras de elegirlo, pero el interés se centra básicamente en el llamado intervalo de alta densidad o por su nombre en inglés (*Highest Posterior Density, HPD*).

Un intervalo de credibilidad al $100(1 - \alpha)\%$ es de alta densidad a posteriori (*HPD*) si para todo $\theta_1 \in [a, b]$ y para todo $\theta_2 \notin [a, b]$ se tiene que

$$p(\theta_1|x) \geq p(\theta_2|x).$$

Por lo tanto, el intervalo *HPD* contiene los valores de θ que son a posteriori más plausibles, es decir, $p(\theta|x)$ es más alto para todos los θ dentro del intervalo que para los

2. PARADIGMA BAYESIANO

valores fuera del intervalo. Una propiedad importante del *HPD* es que su longitud es la más pequeña de tal forma que $p(a \leq \theta \leq b|x) = 1 - \alpha$.¹

Gráficamente, el intervalo *HPD* se construye determinando la intersección de una línea horizontal con la densidad a posteriori hasta que se acumule el $(1 - \alpha)$ de probabilidad. Las coordenadas x de los dos puntos de intersección definen el intervalo y el tamaño correcto del *HPD* se obtiene ajustando la altura de la línea horizontal hasta que el área bajo la curva sea igual a $(1 - \alpha)$. Esta es la base para determinar el intervalo pero generalmente el *HPD* es encontrado por métodos numéricos.

2.1.6. Predicción

La predicción es un tema importante en todas las áreas de aplicación de la estadística, además la comparación de las respuestas predichas con las observadas es un método predeterminado para la verificación del modelo. Para hacer inferencias predictivas consideremos la distribución marginal de x

$$p(x) = \int p(x, \theta) d\theta = \int p(\theta)p(x|\theta) d\theta.$$

A esto a menudo se le llama distribución predictiva a priori, porque no está condicionada a una observación precedente del proceso y porque es la distribución de una cantidad que es observable. Después de que se hayan observado los datos x , podemos predecir un dato desconocido, \tilde{x} . La distribución de \tilde{x} se le conoce como la distribución predictiva a posteriori, porque ya está condicionado al x observado:

$$\begin{aligned} p(\tilde{x}|x) &= \int p(\tilde{x}, \theta|x) d\theta \\ &= \int p(\tilde{x}|\theta, x)p(\theta|x) d\theta \\ &= \int p(\tilde{x}|\theta)p(\theta|x) d\theta. \end{aligned}$$

La segunda y tercer líneas muestran la distribución predictiva posteriori como un promedio de predicciones condicionales sobre la distribución posterior de θ , el último paso se sigue de la independencia condicional asumida de x y \tilde{x} dada θ .

2.1.7. Comparación de Modelos

Se pueden considerar diferentes modelos para un conjunto de datos y hay una serie de metodologías para la comparación o selección de modelos. Entre los principales criterios están el AIC, BIC y DIC que a continuación se describen de manera general.

¹Para revisar la prueba de este resultado consulte [28].

Criterio de Información de Akaike (AIC)

Es una medida que toma en cuenta la bondad del ajuste del modelo y su complejidad. La idea clave es la de penalizar un exceso de parámetros ajustados. Es un estimador muestral de la esperanza de la log-verosimilitud $E[\log(f(X|\theta))]$ que viene dado por la expresión

$$\text{AIC} = -2 \log L(\hat{\theta}|x) + 2k,$$

donde $L(\hat{\theta}|x)$ es la verosimilitud valuada en la estimación máximo verosímil del vector de parámetros θ , y k es el número de parámetros estimados dentro del modelo.

Criterio de Información Bayesiano (BIC)

El BIC es muy similar al AIC, se basa en la máxima verosimilitud como forma de medida de la bondad de ajuste. Con este criterio se penaliza el número de parámetros con $\log(n)$, en lugar de 2 como en el AIC. Su fórmula es

$$\text{BIC} = -2 \log L(\hat{\theta}|x) + k \log(n),$$

donde $L(\hat{\theta}|x)$ es la verosimilitud evaluada en la estimación máximo verosímil del vector de parámetros θ , k es el número de parámetros estimados dentro del modelo, y n es el tamaño muestral. Por lo tanto el BIC acabará eligiendo el modelo más simple o sencillo y cuyas predicciones son hechas a menor detalle.

Criterio de Información de la Devianza (DIC)

El DIC es una generalización del AIC y del BIC. El DIC utiliza cadenas de Monte Carlo para buscar la distribución de los parámetros. El DIC se define como

$$\text{DIC} = \bar{D}(\theta) + p_D = D(\bar{\theta}) + 2p_D.$$

Se trabaja con la devianza que es una estadística de bondad de ajuste la cual toma en cuenta la función completa de verosimilitud y se define como

$$D(\theta) = -2 \log(p(x|\theta)).$$

Así la medida de ajuste será el promedio de la devianza $\bar{D}(\theta) = E(D(\theta))$ donde

$$\bar{D}(\theta) = -2 \int \log(p(x|\theta)) d\theta.$$

El DIC calcula el número efectivo de parámetros, p_D , que se define como el promedio de la devianza menos la devianza evaluada en el promedio de los parámetros

$$p_D = \bar{D}(\theta) - D(\bar{\theta}).$$

Para medir la complejidad del modelo, el DIC utiliza de nuevo la devianza de las funciones de los parámetros para penalizar la complejidad. Para la implementación del

DIC se incluye tanto el tamaño muestral como el número de parámetros y la relación de covariación entre estos.¹

2.2. Métodos Monte Carlo Vía Cadenas de Markov

Las siglas MCMC combina dos propiedades importantes: integración Monte Carlo y cadenas de Markov. Monte Carlo² es la práctica de estimar por medio de integración numérica las características de una distribución mediante muestras aleatorias de esta distribución, y la propiedad de la cadena de Markov da la idea a que las muestras aleatorias se generan mediante un proceso secuencial y aunque cada muestra aleatoria que se genera depende de la anterior las nuevas muestras no dependen de ninguna muestra anterior a la anterior, ésta es la denominada propiedad de Markov.³

En Estadística Bayesiana, en la mayoría de los casos, la forma analítica de la distribución a posteriori es desconocida, por tanto se usan métodos de simulación, en particular los métodos de Monte Carlo vía Cadenas de Markov (MCMC) ya que son métodos de simulación para generar muestras de las distribuciones a posteriori y estimar cantidades de interés a posteriori.

Dada una distribución $p(\theta)$ completamente conocida salvo por su constante de proporcionalidad la aplicación de los métodos MCMC consiste básicamente en generar una o varias realizaciones de una cadena de Markov cuya distribución estacionaria sea la buscada $p(\theta)$. Cada valor generado depende sólo del valor anterior simulado de ahí viene la noción de cadena de Markov, y por medio de la integración Monte Carlo hacer inferencia sobre esta muestra generada. Cabe aclarar que los métodos MCMC no son exclusivos de la inferencia Bayesiana pero son usados por la complejidad analítica de las distribuciones a posteriori.

2.2.1. Algoritmo Metropolis Hastings

El algoritmo de Metropolis Hastings (MH) consiste en simular de una cadena de Markov cuya distribución estacionaria es $p(\theta|x)$. Se comienza con un valor inicial θ_0 :

1. Dado el valor actual θ_t , simular un valor candidato θ^* , de una densidad propuesta $q(\theta^*|\theta_t)$.

¹Estos son algunos de los criterios más importantes en la comparación de modelos, sin embargo puede adoptar otros de su interés, en [39] se discuten otros criterios.

²Puede estudiar los métodos Monte-Carlo en [20].

³Para profundizar en las propiedades de las cadenas de Markov consulte [14] y [34].

³Si es de su interés profundizar en los algoritmos MCMC se sugiere consultar [30] y [10].

2. Calcular la probabilidad de aceptar el valor generado:

$$\alpha = \min \left\{ 1, \frac{p(\theta^*|x) q(\theta_t|\theta^*)}{p(\theta_t|x) q(\theta^*|\theta_t)} \right\}.$$

3. Simular u de una distribución uniforme $U(0, 1)$.

4. Si $u \geq \alpha$, tomar $\theta_{t+1} = \theta^*$, en caso contrario tomar $\theta_{t+1} = \theta_t$.

5. Volver a 1.

La probabilidad de aceptar, α , no depende de la constante de integración de la distribución a posteriori, de modo que

$$\alpha = \min \left\{ 1, \frac{p(\theta^*) p(x|\theta^*) q(\theta_t|\theta^*)}{p(\theta_t) p(x|\theta_t) q(\theta^*|\theta_t)} \right\}.$$

Por ello el algoritmo de MH puede ser una herramienta muy útil en la aplicación de las técnicas Bayesianas¹.

2.2.2. Muestreo de Gibbs

El muestreo de Gibbs es un caso particular del algoritmo MH por bloques en el cual las densidades propuestas coinciden con las distribuciones a posteriori condicionadas de modo que la probabilidad de aceptar es siempre uno. Supongamos que el conjunto de parámetros es $\theta = (\theta_1, \dots, \theta_k)$ y que para todo $i = 1, \dots, k$, es fácil simular de la distribución a posteriori condicional $p(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$.

Fijar un valor inicial $(\theta_{1,0}, \dots, \theta_{k,0})$, repetir:

1. Simular $\theta_{1,t+1} \sim p(\theta_1 | \theta_{2,t}, \dots, \theta_{k,t})$.
2. Simular $\theta_{2,t+1} \sim p(\theta_2 | \theta_{1,t+1}, \theta_{3,t}, \dots, \theta_{k,t})$.
3. Simular $\theta_{3,t+1} \sim p(\theta_3 | \theta_{1,t+1}, \theta_{2,t+1}, \theta_{4,t}, \dots, \theta_{k,t})$
- \vdots
4. Simular $\theta_{k,t+1} \sim p(\theta_k | \theta_{1,t+1}, \dots, \theta_{k-1,t+1})$.

Cuando la cadena converge, los valores resultantes de θ_t son una muestra de la distribución $p(\theta|x)$ ².

¹Estos algoritmos se pueden consultar con más detalle en [31].

²Este algoritmo se puede consultar con más detalle en [31].

2.2.3. Diagnóstico de Convergencia

Cuando se usa un método MCMC, es importante analizar si los valores simulados, θ_t , han convergido a la distribución estacionaria $p(\theta|x)$. Y para verificar la convergencia es necesario revisar algunos criterios de convergencia y aquí se presentan algunos.

1. Examinar qué tan bien está explorando el algoritmo MCMC el espacio de estados.
2. Probar la convergencia de las medias de los valores simulados en el MCMC, por ejemplo $\frac{1}{N} \sum_{t=1}^N \theta_t \rightarrow E[\theta|x]$.
3. Analizar si los valores simulados son aproximadamente una muestra de valores independientes e idénticamente distribuidos.
4. Gelman y Rubin proponen un enfoque general para monitorear la convergencia cuando se tiene más de una cadena, formulan un factor de reducción de escala basado en la comparación de varianzas dentro de la cadena y entre cadenas. Primero se simulan de forma independiente $m \geq 2$ cadenas y luego para cada parámetro de interés se calculan

$$B/n = \sum_{i=1}^m \frac{(\bar{x}_i - \bar{x}_{..})^2}{m-1},$$
$$W = \sum_{i=1}^m \frac{s_i^2}{m},$$

donde n es el número de iteraciones, B/n es la varianza empírica entre las cadenas y W estima la varianza de la distribución estacionaria como la media de la varianza empírica dentro de cada cadena. Por lo que la varianza empírica de todas las cadenas combinadas puede expresarse como

$$\hat{\sigma}^2 = \frac{(n-1)W}{n} + \frac{B}{n}.$$

Estas estimaciones son imparciales cuando las cadenas han convergido, los valores sustancialmente superiores a 1 indican falta de convergencia¹.

Aunque existen muchos procedimientos para evaluar la convergencia de las cadenas aquí sólo se mencionan algunos de manera muy general.²

¹Si quiere conocer más detalles sobre la construcción del factor de reducción de escala de Gelman y Rubin puede consultar [12].

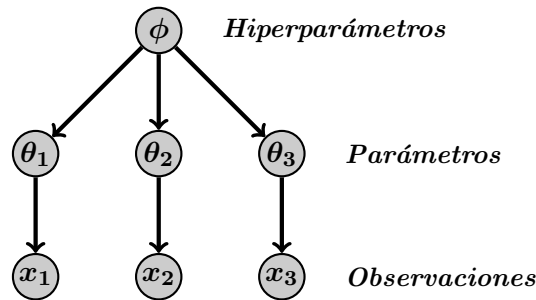
²Otros criterios de convergencia más específicos se pueden encontrar en [31].

2.3. Modelos Jerárquicos Bayesianos

En general los modelos jerárquicos son modelos estadísticos que proporcionan un marco formal para el análisis de datos con una estructura compleja en la que se presenta una jerarquía. Dependiendo del contexto puede haber distintos niveles de agregación, por ejemplo cuando los datos están distribuidos en grupos demográficos, distribuidos temporalmente o espacialmente, estos modelan la dependencia entre las observaciones.¹ Es preferible usar la terminología de *modelo jerárquico* porque alude la forma en que se usan los modelos dentro de los modelos para reflejar los niveles dentro de los datos.

2.3.1. Estructura General

El método consiste en suponer que los datos dependen de una serie de parámetros que, a su vez, dependen de otros parámetros y lo ideal sería elegir un modelo de probabilidad conjunta que describa esta dependencia entre los parámetros con diversas componentes denominadas niveles o etapas. El modelo clásico jerárquico se ve así:



La modelación desde este enfoque se basa en el hecho de tratar a la distribución conjunta como una colección de variables aleatorias que se pueden descomponer en una serie de modelos condicionales por niveles. Es útil escribir el modelo jerárquico en tres niveles:

Nivel 1: Modelo para los datos

$$p(\mathbf{x}|\boldsymbol{\theta}).$$

Nivel 2: Modelo para los parámetros

$$p(\boldsymbol{\theta}|\phi).$$

¹Si es de su interés adentrarse en los modelos jerárquicos Bayesianos revise [11].

2. PARADIGMA BAYESIANO

Nivel 3: Modelo para los hiperparámetros

$$p(\phi).$$

La idea es resolver un problema dividiéndolo en sub-problemas más simples. Los métodos Bayesianos permiten una estimación de forma natural en la modelación jerárquica. Deseamos hacer inferencias sobre modelos con muchos parámetros $\theta_1, \dots, \theta_J$ medidos en J unidades (individuos, áreas, puntos en el tiempo, ensayos, etc.) que están relacionados o conectados por la estructura del problema. Podemos identificar tres supuestos de modelado diferentes:

1. *Parámetros idénticos*: Todos los θ 's son idénticos, en cuyo caso los datos pueden agruparse y las unidades individuales pueden ignorarse.
2. *Parámetros independientes*: Todos los θ 's no guardan relación alguna en cuyo caso los resultados de cada unidad pueden analizarse independientemente (por ejemplo usando una distribución a priori totalmente especificada dentro de cada unidad).
3. *Parámetros intercambiables*: Se supone que los θ 's son similares en el sentido de que los índices no aportan información. En general la suposición de que las unidades son intercambiables equivale matemáticamente a que los parámetros $\theta_1, \dots, \theta_J$ se extraen de una distribución a priori común con parámetros desconocidos.

2.3.2. Intercambiabilidad

Para crear un modelo de probabilidad conjunta para todos los parámetros θ , utilizamos la idea de intercambiabilidad. Cuando no se tiene información para distinguir cualquiera de los parámetros $(\theta_1, \dots, \theta_J)$ y además no se pueden ordenar ni agruparse debe asumirse cierta simetría entre los parámetros en su distribución a priori. Esta simetría se denota por el concepto de intercambiabilidad, es decir, los parámetros $(\theta_1, \dots, \theta_J)$ son intercambiables en su distribución conjunta si $p(\theta_1, \dots, \theta_J)$ es invariante a las permutaciones de los índices $(1, \dots, J)$.

La forma más simple de una distribución intercambiable tiene cada uno de los parámetros θ_j como una muestra independiente de una distribución a priori guiada por algún vector de parámetros desconocido ϕ , así

$$p(\theta | \phi) = \prod_{j=1}^J p(\theta_j | \phi).$$

En general ϕ es desconocido, por lo que la distribución de θ es

$$p(\theta) = \int \left(\prod_{j=1}^J p(\theta_j | \phi) \right) p(\phi) d\phi. \quad (2.2)$$

Obsérvese que la intercambiabilidad no implica independencia, ya que el producto y la integral en la expresión (2.2) no pueden intercambiarse. Pero la independencia de las variables aleatorias que tienen la misma distribución marginal sí implica intercambiabilidad. Por lo tanto la mezcla de distribuciones independientes e idénticamente distribuidas es lo que usualmente se requiere en la práctica para inducir la intercambiabilidad.

2.3.3. Inferencia

Lo importante de estos modelos es que ϕ no se conoce y bajo el enfoque Bayesiano este tiene su propia distribución a priori $p(\phi)$. Así la distribución a priori conjunta es

$$p(\phi, \theta) = p(\phi)p(\theta|\phi),$$

y la distribución a posteriori conjunta es

$$\begin{aligned} p(\phi, \theta|x) &\propto p(\phi, \theta)p(x|\phi, \theta) \\ &= p(\phi, \theta)p(x|\theta), \end{aligned}$$

esta última igualdad se da porque la distribución de los datos $p(x|\phi, \theta)$ depende solo de θ , los hiperparámetros ϕ afectan solo a través de θ .

Para crear una distribución a priori conjunta para (ϕ, θ) se debe asignar una distribución a priori a ϕ . Como en los modelos no jerárquicos usualmente es práctico comenzar dando distribuciones a priori no informativas o poco informativas y añadir información a priori si se cuenta con ella.

2.3.4. Predicción

Para los modelos jerárquicos hay dos distribuciones predictivas:

1. La distribución de las observaciones futuras \tilde{x} correspondiente a un existente θ_j .
2. La distribución de las observaciones \tilde{x} correspondiente a los θ_j futuros tomados de la misma superpoblación y lo denotamos como $\tilde{\theta}$.

Capítulo 3

Valores Extremos

En las últimas décadas se ha desarrollado la teoría de valores extremos tanto desde el punto de vista metodológico como el de aplicaciones, siendo así una de las ramas más importantes de la estadística. Para [Gumbel](#) el objetivo de la teoría de valores extremos es analizar valores extremos observados y predecir valores extremos en el futuro.

Formalmente la teoría de valores extremos centra su estudio en los eventos asociados a las colas de la distribución, es decir, en los valores más altos o más bajos de la variable bajo estudio. El análisis de valores extremos tiene muchas aplicaciones en la práctica, por ejemplo en hidrología es necesario estudiar las precipitaciones máximas, en meteorología los cambios extremos de la temperatura. En cuestiones de contaminación es de interés analizar los niveles altos de los contaminantes del aire, en ingeniería la resistencia máxima de ciertos materiales, o incluso en finanzas dentro del mercado financiero los valores mínimos o máximos de un portafolio de inversión, entre otros.¹

3.1. Modelos Asintóticos

El problema consiste en dada una muestra aleatoria independiente de una distribución desconocida F queremos estimar la cola de la distribución F . Sean X_1, X_2, \dots, X_n una muestra aleatoria de observaciones independientes con distribución común F . El modelo se centra en el comportamiento estadístico de $M_n = \max\{X_1, \dots, X_n\}$. En la práctica las observaciones X_i representan los valores de un proceso medido por cierto periodo de tiempo, por ejemplo temperaturas promedio diarias. Por lo tanto M_n representa el máximo del proceso en n unidades de tiempo. La función de distribución de M_n es:

¹Para profundizar en la teoría de valores extremos se sugieren los libros: [2],[7], [13] y [19].

3. VALORES EXTREMOS

$$\begin{aligned} P[M_n \leq z] &= P[X_1 \leq z, X_2 \leq z, \dots, X_n \leq z] \\ &= P[X_1 \leq z] \cdot P[X_2 \leq z] \cdots P[X_n \leq z] \\ &= F(z) \cdot F(z) \cdots F(z) \\ &= [F(z)]^n, \end{aligned}$$

dicha función converge en probabilidad a 0 si $z < z^*$ y a 1 si $z \geq z^*$ donde $z^* = \sup\{x : F(x) < 1\}$. Así que para que la distribución límite no sea degenerada hay que estandarizar la variable M_n mediante una transformación lineal, hay que encontrar sucesiones de constantes $\{a_n > 0\}$ y $\{b_n\}$ de tal forma que la expresión

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

tenga una distribución no degenerada. Se tiene una completa variedad de distribuciones límite para M_n^* que se describirán de manera muy general en la siguiente sección.

3.2. Teorema de Valores Extremos

Ahora el objetivo es encontrar distribuciones límite para M_n^* . Para ello se tiene el teorema conocido como *Teorema de Valores Extremos* o *Teorema de Fisher – Tippett – Gnedenko*¹. Si existen sucesiones de constantes $\{a_n > 0\}$ y $\{b_n\}$ tal que

$$P \left[\frac{(M_n - b_n)}{a_n} \leq z \right] \rightarrow F(z) \quad \text{cuando } n \rightarrow \infty,$$

donde F es una función de distribución no degenerada, entonces F pertenece a una de las siguientes familias:

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

¹El boceto de la demostración del Teorema lo puede consultar en [2].

para parámetros $a > 0$, $b \in \mathbb{R}$ y en el caso de las familias II y III $\alpha > 0$.

Estas tres clases de distribuciones son conocidas como las distribuciones de valores extremos, donde las de tipo I son la familia *Gumbel*, las de tipo II de *Fréchet*, y las de tipo III de *Weibull*, cada una con su parámetro de localización b y de escala a ; y además, las familias de *Fréchet* y de *Weibull* tienen un parámetro de forma α .

Cuando se pueden encontrar dichas sucesiones que verifiquen lo anterior, entonces la distribución asintótica de la variable transformada M_n^* es de alguno de los tres tipos anteriores. A su vez, se deduce que la distribución de M_n^* sólo puede ser una de esas tres. Sin duda, este teorema es posiblemente considerado como el más importante dentro de la teoría de valores extremos.

3.3. Distribución Generalizada de Valores Extremos

Todas las distribuciones anteriores se pueden resumir en una sola que es la *Distribución Generalizada de Valores Extremos (GEV)*, cuya función de distribución es

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

donde $\mu \in \mathbb{R}$ es el parámetro de localización, $\sigma \in \mathbb{R}^+$ es el parámetro de escala y $\xi \in \mathbb{R}$ es el parámetro de forma y se denota como $GEV(\mu, \sigma, \xi)$. El signo $+$ indica la parte positiva de la expresión y la función de densidad de probabilidad es

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

definida en $\{z : 1 + \xi \left(\frac{z - \mu}{\sigma} \right) > 0\}$. Esta distribución abarca la *Fréchet* ($\xi > 0$), *Weibull* ($\xi < 0$) y *Gumbel* ($\xi \rightarrow 0$). La Figura 3.1 muestra las funciones de densidad de acuerdo al comportamiento del parámetro de forma ξ .

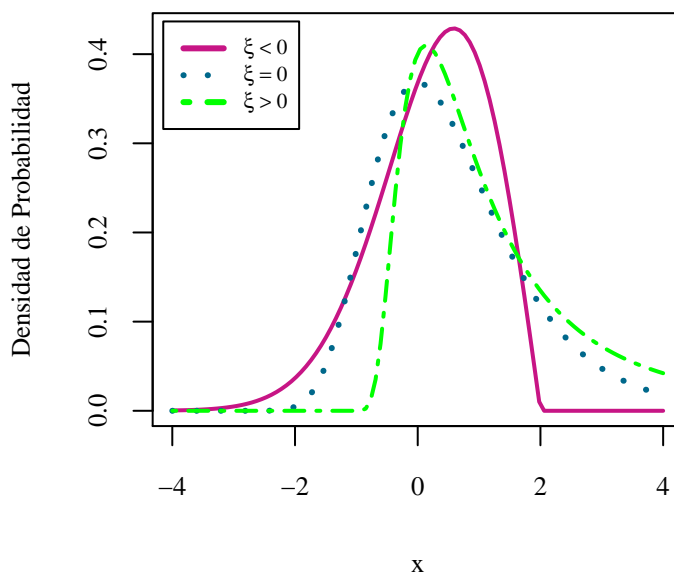


Figura 3.1: Distribución Generalizada de Valores Extremos, $GEV(\mu, \sigma, \xi)$ con $\mu = 0$ y $\sigma = 1$.

La unificación de las tres familias originales de distribución de valores extremos en una sola familia simplifica enormemente la implementación estadística. A través de la inferencia sobre ξ , los datos mismos determinan el tipo más apropiado del comportamiento de la cola, y no hay necesidad de hacer juicios subjetivos a priori sobre qué familia de valores extremos individuales adoptar.

3.3.1. Inferencia para la Distribución GEV

Ahora sean Z_1, Z_2, \dots, Z_n los máximos por bloques de la misma longitud de una muestra de variables aleatorias independientes e idénticamente distribuidas. Si las X'_i s variables aleatorias son independientes entonces los máximos Z'_i s también lo serán y se asume que siguen una distribución GEV cuyos parámetros hay que estimarse.

Se supone que Z_1, Z_2, \dots, Z_n son independientes y siguen una distribución GEV con función de densidad de probabilidad dada por la expresión

$$f(z_i|\mu, \sigma, \xi) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]^{-1/\xi - 1} \exp \left\{ - \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}.$$

La función de verosimilitud es

$$\begin{aligned} L(\mu, \sigma, \xi; \mathbf{z}) &= \prod_{i=1}^n f(z_i | \mu, \sigma, \xi) \\ &= \prod_{i=1}^n \frac{1}{\sigma} \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]^{-1/\xi - 1} \exp \left\{ - \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\}. \end{aligned}$$

Un inconveniente con el uso de métodos de máxima verosimilitud para la *GEV* se refiere a las condiciones de regularidad¹ que son necesarias para que las propiedades asintóticas asociadas al estimador máximo verosímil sean válidas. Tales condiciones no son satisfechas por el modelo porque los puntos finales de la distribución *GEV* están en función de los valores de los parámetros: $\mu - \frac{\sigma}{\xi}$. Cuando $\xi > 0$ es un punto final superior de la distribución y cuando $\xi < 0$ es un punto final inferior, a fin de resolver estos problemas se obtuvieron los siguientes resultados²:

- Cuando $\xi > -0.5$, los estimadores máximo verosímiles son regulares, en el sentido de tener las propiedades asintóticas habituales.
- Cuando $-1 < \xi < -0.5$, los estimadores máximo verosímiles se pueden obtener pero no tienen las propiedades asintóticas habituales.
- Cuando $\xi < -1$, es poco probable que se puedan obtener los estimadores máximo verosímiles.

Pero en la práctica es mejor utilizar la función de log-verosimilitud, la cual se define como

$$\log L(\mu, \sigma, \xi; \mathbf{z}) = l(\mu, \sigma, \xi; \mathbf{z}) = \sum_{i=1}^n \log f(z_i | \mu, \sigma, \xi).$$

Por lo tanto la función de log-verosimilitud de la *GEV* cuando $\xi \neq 0$ es

$$\begin{aligned} l(\mu, \sigma, \xi; \mathbf{z}) &= -n \log(\sigma) - (1 + 1/\xi) \sum_{i=1}^n \log \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right] - \\ &\quad \sum_{i=1}^n \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]^{-1/\xi}. \end{aligned}$$

Y en el caso de que $\xi = 0$ se tiene

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

¹Las condiciones de regularidad para los estimadores máximo verosímil se pueden consultar en [4].

²Los detalles de cómo se obtuvieron estos resultados y profundizar en la estimación de máxima verosimilitud en casos no regulares se pueden ver en [37].

3. VALORES EXTREMOS

siempre que

$$1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) > 0, \quad \text{para } i = 1, \dots, n.$$

No existe una solución analítica para la log-verosimilitud, pero para cualquier conjunto de datos dado, la maximización es sencilla usando algoritmos de optimización numérica estándar. Sujeto a las limitaciones anteriores sobre ξ la distribución aproximada de $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$ es asintóticamente normal multivariada (debido a las propiedades asintóticas del estimador máximo verosímil) con vector de medias (μ, σ, ξ) y matriz de varianzas-covarianzas igual a la matriz inversa de la información de Fisher evaluada en el estimador máximo verosímil. Aunque esta matriz puede calcularse analíticamente, es más fácil utilizar técnicas de diferenciación numérica para evaluar las segundas derivadas, y las rutinas numéricas estándar para llevar a cabo la inversión. Los intervalos de confianza y otras formas de inferencia se siguen de la normalidad aproximada del estimador máximo verosímil.

3.3.2. Niveles de Retorno

Se llama *niveles de retorno* a los cuantiles de la distribución Generalizada de Valores Extremos y se denotan por z_p con $0 < p < 1$ estos se obtienen como la inversa de la función de distribución de la distribución *GEV*

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - \{-\log(1-p)\}^{-\xi}] & \text{si } \xi \neq 0, \\ \mu - \sigma \log\{-\log(1-p)\} & \text{si } \xi = 0, \end{cases}$$

de manera que z_p es el nivel de retorno asociado al periodo de retorno $1/p$, es decir, se espera que el nivel z_p sea excedido en promedio una vez cada $1/p$ unidades de tiempo. Dicho de otra forma, z_p será excedido en una unidad de tiempo con probabilidad p .

La estimación de \hat{z}_p se obtiene mediante la sustitución directa de los estimadores máximo verosímil de la distribución *GEV* por lo que se tiene

$$\hat{z}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} [1 - \{-\log(1-p)\}^{-\hat{\xi}}] & \text{si } \hat{\xi} \neq 0, \\ \hat{\mu} - \hat{\sigma} \log\{-\log(1-p)\} & \text{si } \hat{\xi} = 0. \end{cases}$$

Cabe mencionar que los periodos de retorno largos son los que son de mayor interés y por ello juegan un papel importante en la modelación de valores extremos. Por ejemplo, si nos interesa estimar la probabilidad de que un evento excediera una vez cada 100 años entonces se puede escribir como

$$P(\text{Evento} > z_{100}) = \frac{1}{100},$$

esto implica

$$1 - P(\text{Evento} \leq z_{100}) = \frac{1}{100},$$

y en términos de la función de distribución GEV tenemos que

$$\begin{aligned} 1 - P(\text{Evento} \leq z_{100}) &= 1 - F(z_{100}; \hat{\mu}, \hat{\sigma}, \hat{\xi}), \\ \Rightarrow 1 - \exp \left\{ - \left[1 + \hat{\xi} \left(\frac{z_{100} - \hat{\mu}}{\hat{\sigma}} \right) \right]_+^{-1/\hat{\xi}} \right\} &= 0.01, \\ \Rightarrow \exp \left\{ - \left[1 + \hat{\xi} \left(\frac{z_{100} - \hat{\mu}}{\hat{\sigma}} \right) \right]_+^{-1/\hat{\xi}} \right\} &= 0.99. \end{aligned}$$

Resolviendo para z_{100} , una estimación de retorno para 100 años es

$$\hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left[1 - \{-\log(0.99)\}^{-\hat{\xi}} \right],$$

y en el caso de que $\hat{\xi} = 0$

$$\hat{\mu} - \hat{\sigma} \log\{-\log(0.99)\}.$$

Sabemos que la estimación puntual conlleva a tener un error estándar en la estimación, en este caso el error es mayor ya que la estimación está en función de los parámetros estimados de la distribución *GEV*, sin embargo, se puede usar el método delta para obtener el error estándar de la estimación de \hat{z}_p .¹

3.3.3. Características de la Distribución *GEV*

Otras características importantes de la distribución *GEV* se presentan a continuación ²:

$$\text{Media} \quad \left\{ \begin{array}{ll} \mu + \sigma \left(\frac{\Gamma(1-\xi)-1}{\xi} \right) & \text{si } \xi \neq 0, \quad \xi < 0 \\ \mu + \sigma \gamma & \text{si } \xi = 0 \\ \text{No existe} & \text{si } \xi \geq 1 \end{array} \right.$$

¹El procedimiento de cómo usar el método delta para estimar el error estándar de \hat{z}_p lo puede consultar en [7].

²Éstas se obtuvieron de [19] y se pueden consultar otras características de interés.

3. VALORES EXTREMOS

donde $\Gamma(\cdot)$ es la función *Gamma* y γ es la constante de *Euler*.

$$\text{Varianza} \quad \begin{cases} \sigma^2 \left(\frac{g_2 - g_1^2}{\xi^2} \right) & \text{si } \xi \neq 0, \quad \xi < \frac{1}{2} \\ \sigma^2 \left(\frac{\pi^2}{6} \right) & \text{si } \xi = 0 \\ \text{No existe} & \text{si } \xi \geq \frac{1}{2} \end{cases}$$

donde $g_k = \Gamma(1 - k\xi)$.

$$\text{Cuantil de orden } p \quad \begin{cases} \mu - \frac{\sigma}{\xi} [1 - \{-\log(1-p)\}^{-\xi}] & \text{si } \xi \neq 0 \\ \mu - \sigma \log\{-\log(1-p)\} & \text{si } \xi = 0 \end{cases}$$

$$\text{Mediana} \quad \begin{cases} \mu + \sigma \left(\frac{\log(2)^{-\xi} - 1}{\xi} \right) & \text{si } \xi \neq 0 \\ \mu - \sigma \log(-\log(2)) & \text{si } \xi = 0 \end{cases}$$

$$\text{Moda} \quad \begin{cases} \mu + \sigma \left(\frac{(1+\xi)^{-\xi} - 1}{\xi} \right) & \text{si } \xi \neq 0 \\ \mu & \text{si } \xi = 0 \end{cases}$$

3.4. Inferencia Bayesiana en Valores Extremos

En la sección anterior para la estimación de parámetros en modelos de valores extremos argumentamos que los métodos basados en la máxima verosimilitud son preferibles pues de todos los modelos considerados adopta el que asigna mayor probabilidad a los datos observados. Pero ésta no es la única forma, dado que el objetivo de un análisis de valores extremos generalmente es una estimación de la probabilidad de que eventos futuros alcancen niveles extremos esto se puede hacer a través de distribuciones predictivas.

Las técnicas Bayesianas ofrecen una forma alternativa de obtener conclusiones de la función de verosimilitud. La diferencia como se vió en el Capítulo 2 es que los parámetros de la distribución se tratan como variables aleatorias, para lo cual establecemos distribuciones a priori que nos permite complementar la información provista por los datos, que en análisis de valores extremos a menudo es muy limitada con otras fuentes

de información.

Por ejemplo un modelo adecuado para observaciones máximas diarias de algún proceso Z es suponer que $Z_i \sim GEV(\mu, \sigma, \xi)$. La estimación de $\theta = (\mu, \sigma, \xi)$ podría hacerse desde una perspectiva Bayesiana dando a los parámetros distribuciones a priori. Sin embargo la distribución GEV no admite ninguna distribución a priori conjugada, por lo que se debe elegir de acuerdo al contexto del problema, si tiene información a priori incorporarla, en caso contrario se pueden usar distribuciones a priori no informativas como se mencionó en el Capítulo 2. A manera de ejemplo podríamos elegir las siguientes distribuciones a priori para los parámetros

$$\begin{aligned} p(\theta) &= p(\mu)p(\sigma)p(\xi), \\ \mu &\sim Normal(\delta_0, v_0), \\ \sigma &\sim InverseGamma(\alpha_0, \beta_0), \\ \xi &\sim U(a_0, b_0), \end{aligned}$$

y la verosimilitud es

$$L(\mu, \sigma, \xi; \mathbf{z}) = \prod_{i=1}^n \frac{1}{\sigma} \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]^{-1/\xi - 1} \exp \left\{ - \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]_+^{-1/\xi} \right\},$$

así la distribución posteriori conjunta está dada por

$$p(\mu, \sigma, \xi | \mathbf{z}) \propto L(\mu, \sigma, \xi; \mathbf{z}) p(\mu) p(\sigma) p(\xi). \quad (3.1)$$

Un inconveniente frecuente en la inferencia Bayesiana es la enorme complejidad computacional que presenta en muchos casos el cálculo de cantidades a posteriori de interés, como en este ejemplo la expresión (3.1) no tiene una solución analítica por tanto recurrimos al uso de métodos MCMC los cuales se describieron en el Capítulo 2.

Capítulo 4

Datos de Ozono en la Ciudad de México

La contaminación atmosférica es un problema a nivel mundial que afecta a diferentes países, en particular a México. Durante las últimas décadas la Ciudad de México ha sido una de las ciudades con más contaminación. La quema de combustibles fósiles para el transporte y la generación de energía, tanto a nivel industrial como doméstico, produce miles de toneladas de contaminantes que diariamente son emitidos a la atmósfera. Los vehículos son la principal fuente de emisión, le sigue la industria y los hogares. El deterioro de la calidad del aire por presencia de contaminantes tiene un impacto negativo en la salud humana y el medio ambiente- Durante los últimos años se han realizado estudios que demuestran la relación entre el incremento en la concentración de los contaminantes y el aumento de enfermedades respiratorias y cardiovasculares.

4.1. Descripción de los Datos

El ozono se considera el mayor representante de los contaminantes debido a su mayor porcentaje en la atmósfera. En altas concentraciones reduce la función pulmonar e irrita las vías respiratorias, es por ello que el análisis se centra en las concentraciones de ozono. El ozono es uno de los muchos compuestos que constituyen al smog fotoquímico y se conforma por compuestos orgánicos (COV) y los óxidos de nitrógeno (NO_x) y estas reacciones son activadas por la energía solar. A manera de proteger la salud de la población se crea en la Ciudad de México el Sistema de Monitoreo Atmosférico (SIMAT) que se encarga del monitoreo y la difusión continua del estado de la calidad del aire.

Actualmente el SIMAT cuenta con más de 40 sitios de monitoreo distribuidos en el área metropolitana, comprendiendo la Ciudad de México y la zona conurbada del Es-

4. DATOS DE OZONO EN LA CIUDAD DE MÉXICO

tado de México. Estos sitios se conocen como estaciones de monitoreo de la calidad del aire, y en la mayoría se utilizan equipos continuos para realizar la medición de los contaminantes. Está conformado por la Red Manual de Monitoreo Atmosférico (REDMA), la Red de Meteorología y Radiación Solar (REDMET), la Red de Depósito Atmosférico (REDDA), el Laboratorio de Análisis Ambiental (LAA), el Centro de Información de la Calidad del Aire (CICA) y la Red Automática de Monitoreo Atmosférico (RAMA).¹

La RAMA es la responsable de la medición de dióxido de azufre, monóxido de carbono, dióxido de nitrógeno, ozono y PM₁₀, PM_{2.5}. Está integrada por 29 estaciones de las cuales 19 son las correspondientes a la Ciudad de México (CDMX). La Tabla 4.1 presenta información sobre las 19 estaciones, el acrónimo, el nombre de la estación y los años de operación.

Abreviación	Estación	Años de operación
AJU	Ajusco	2015-actual
AJM	Ajusco Medio	2015-actual
BJU	Benito Juárez	2015-actual
CAM	Camarones	2003-actual
CCA	Centro de Ciencias de la Atmósfera	2014-actual
COY	Coyoacán	2012-actual
CUA	Cuajimalpa	1994-actual
HGM	Hospital General de México	2012-actual
IZT	Iztacalco	2007-actual
MCM	Museo de la Ciudad de México	1989-2002
MER	Merced	1986-actual
MGH	Miguel Hidalgo	2015-actual
MPA	Milpa Alta	2016-actual
PED	Pedregal	1986-actual
SJA	San Juan Aragón	2012-actual
SFE	Santa Fe	2012-actual
TAH	Tlahúac	1994-actual
UAX	UAM Xochimilco	1987-actual
UIZ	UAM Iztapalapa	2012-actual

Tabla 4.1: Información de las 19 estaciones de monitoreo de la CDMX.

En la Figura 4.1 se muestra un mapa de la localización de las estaciones de monitoreo en la CDMX.²

¹Para más información sobre la estructura del SIMAT consultar: <http://www.aire.cdmx.gob.mx>.

²Las coordenadas de localización de las estaciones se obtuvieron de <http://www.aire.cdmx.gob.mx>.

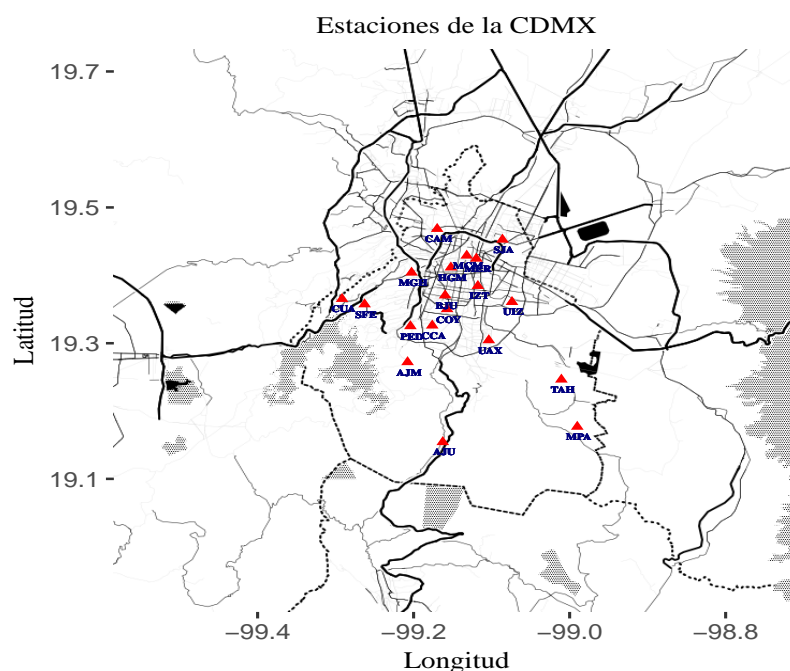


Figura 4.1: Mapa de localización de las estaciones de monitoreo de la CDMX.

Las normas de calidad del aire establecen las concentraciones máximas de contaminantes en el ambiente que no deben ser excedidas con determinada frecuencia. Son un medio regulatorio que pretende que se cumplan los objetivos de mejoramiento de la calidad del aire en el país. En México la Norma Oficial Mexicana (NOM-020-SSA1-2014)¹ establece los criterios para evaluar la calidad del aire con respecto al ozono (O_3). Ésta recomienda concentraciones de ozono menores a 0.095 partes por millón (ppm) para el promedio de 1 hora, sin embargo las concentraciones registradas están en partes por billón (ppb), la conversión es $x_{ppb} = x_{ppm} * 1000$ así 0.095 ppm equivalen a 95 ppb para el promedio de 1 hora. Por lo tanto los valores que sobrepasen este límite se consideran dañinos para la salud.

Las bases de datos anuales² contienen información de las concentraciones de ozono que se registran a cada hora, sin embargo, como se puede observar en la Tabla 4.1 no todas las estaciones iniciaron sus operaciones en el mismo año por lo tanto se seleccionaron las estaciones que cumplen con el periodo de tiempo de estudio. En las Figuras 4.2, 4.3, 4.4 se muestran las gráficas y los histogramas de las concentraciones por hora de ozono de las estaciones de monitoreo de la CDMX durante el periodo del 1 de Enero del 2012 al 31 de Mayo del 2017.

¹La Norma Oficial Mexicana la puede consultar en [8].

²Los datos de las concentraciones de ozono fueron obtenidos de: <http://www.aire.cdmx.gob.mx>

4. DATOS DE OZONO EN LA CIUDAD DE MÉXICO

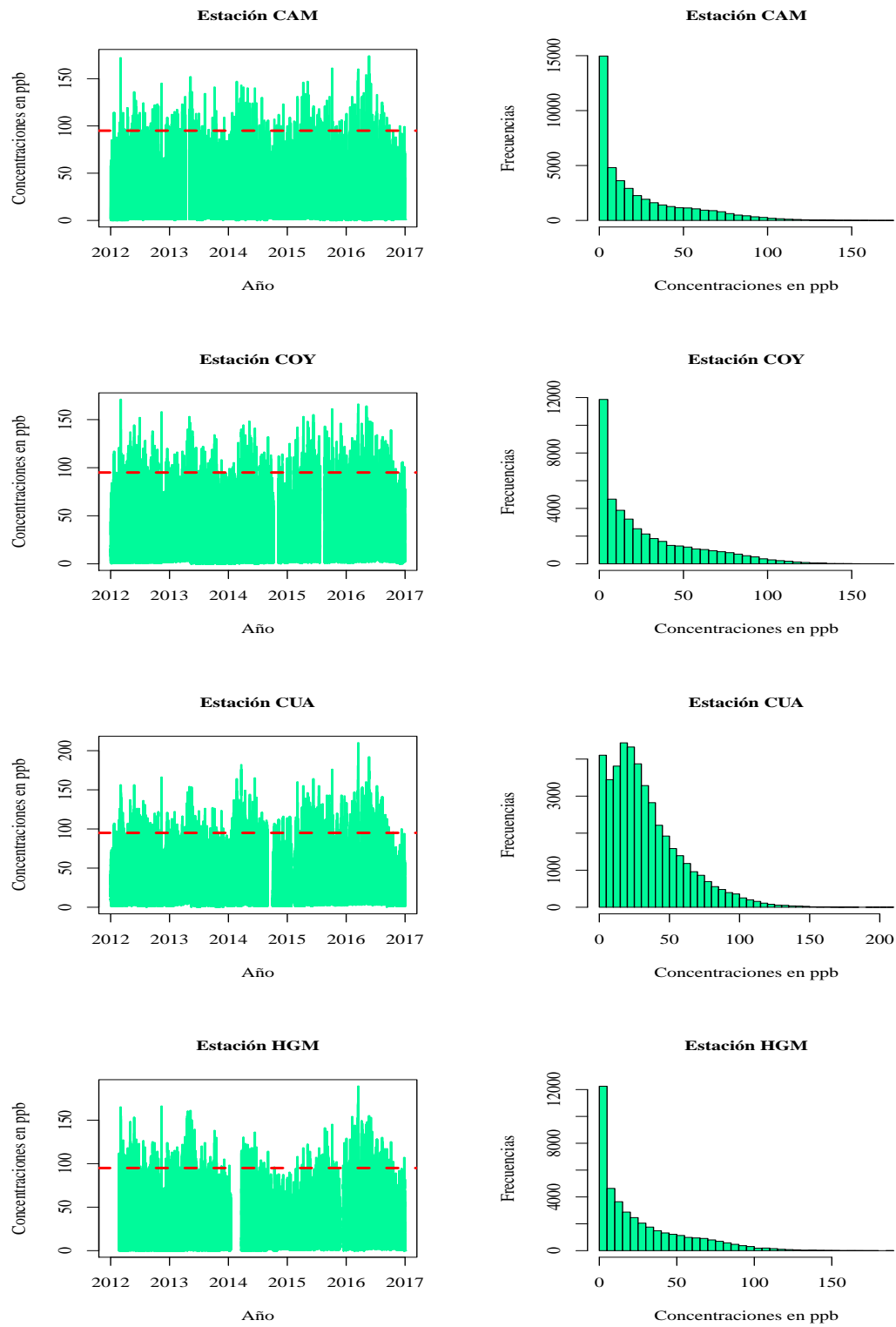


Figura 4.2: Distribución de O_3 de las estaciones CAM, COY, CUA, HGM.

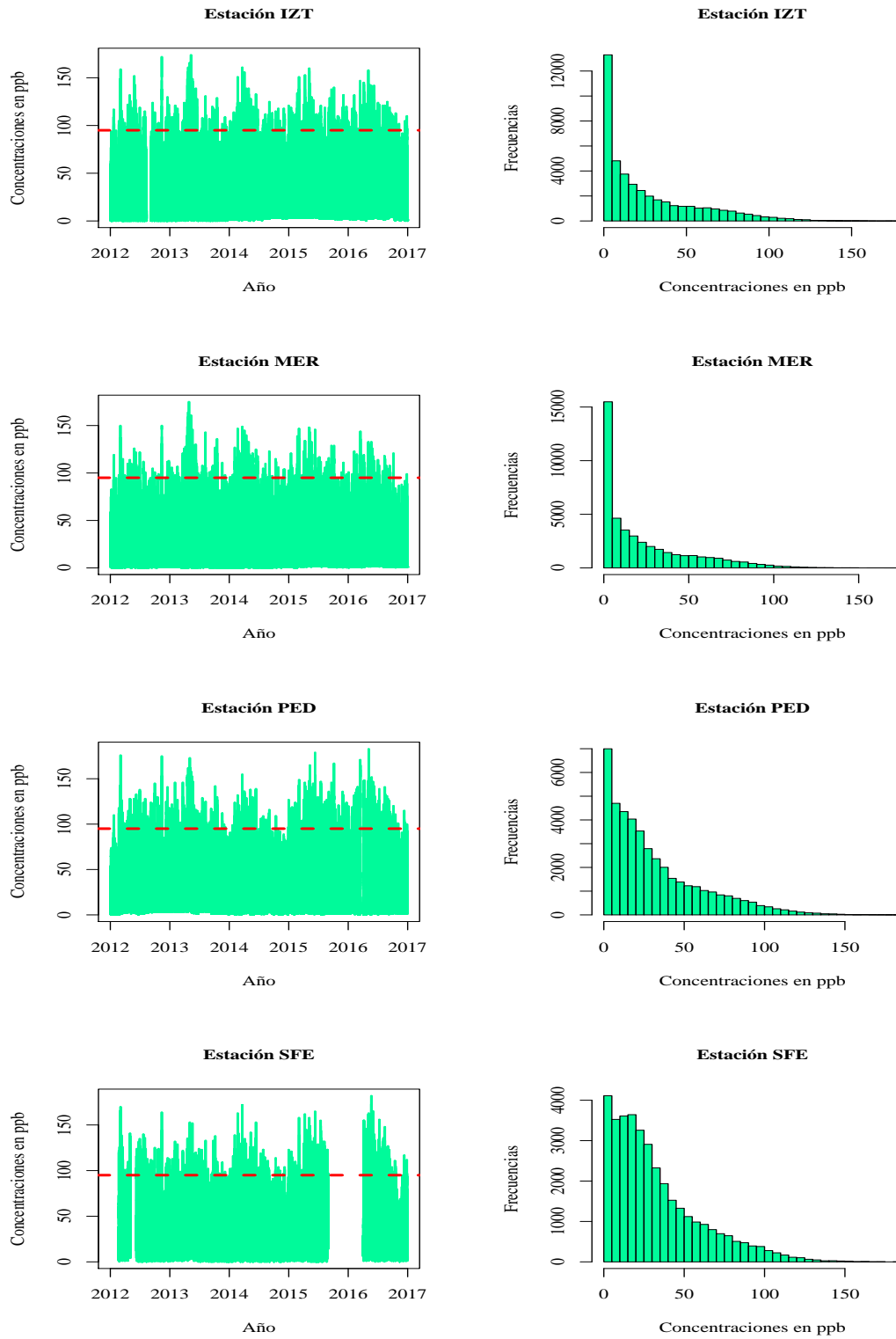


Figura 4.3: Distribución de O₃ de las estaciones IZT, MER, PED, SFE.

4. DATOS DE OZONO EN LA CIUDAD DE MÉXICO

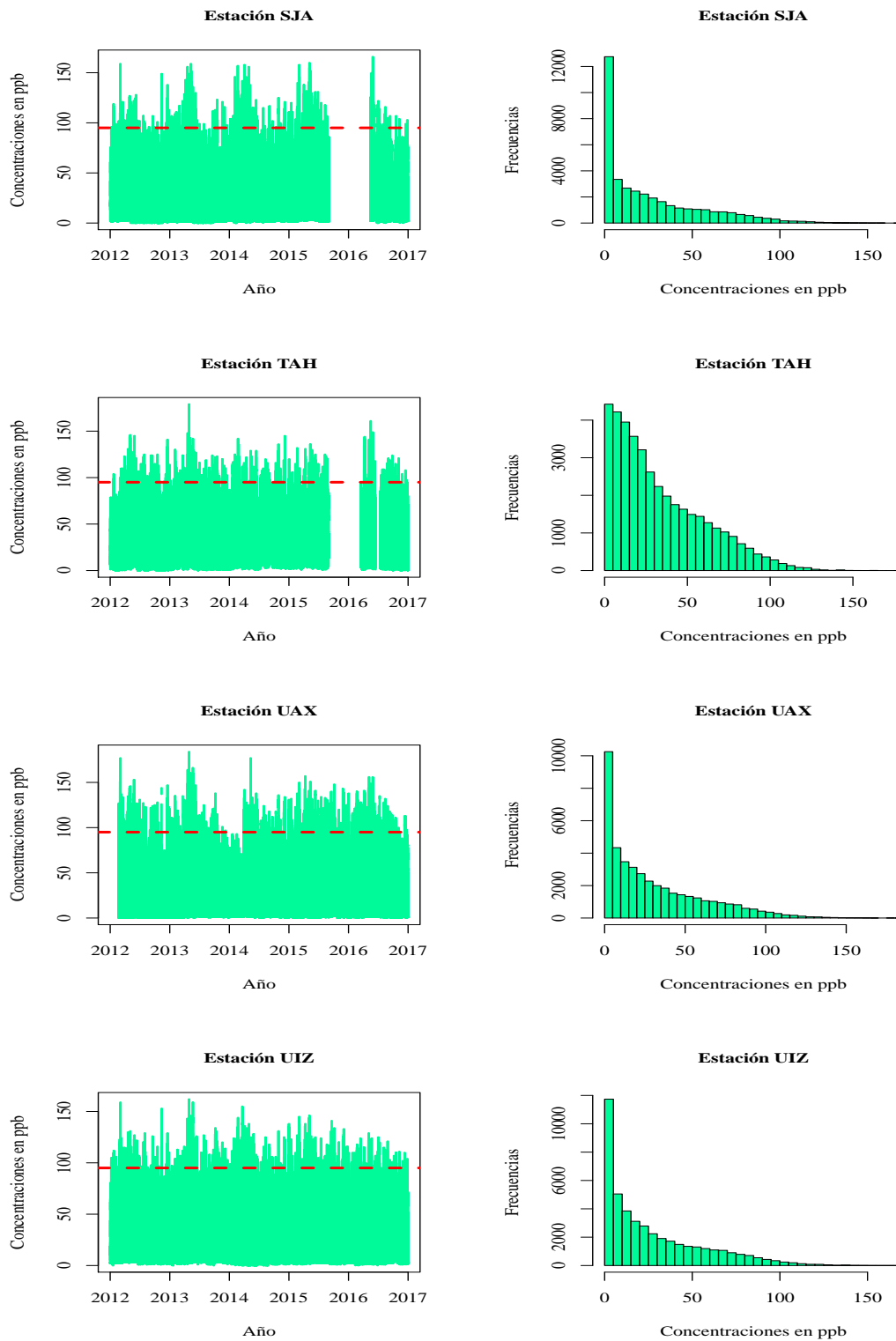


Figura 4.4: Distribución de O_3 de las estaciones SJA, TAH, UAX, UIZ.

Como puede observar en las Figuras 4.2, 4.3, 4.4 en la mayoría de los histogramas de las concentraciones de ozono diarias por hora son asimétricos con un sesgo positivo¹, además de que se observa una gran cantidad de observaciones en la cola derecha de la distribución por lo que esta cola es más larga. Cabe mencionar que en varios casos se tiene un número limitado de observaciones debido a cierres o aperturas de estaciones. Se tienen 44,472 observaciones durante el periodo del 1 de Enero del 2012 al 31 de Mayo del 2017, la Tabla 4.2 muestra el número de datos faltantes en las 12 estaciones de monitoreo.

Estación	Valores Faltantes
CAM	3854
COY	3834
CUA	3728
HGM	5673
IZT	3760
MER	3295
PED	4076
SJA	11261
SFE	9222
TAH	7657
UAX	4225
UIZ	2982

Tabla 4.2: Información de los datos faltantes de las 12 estaciones de monitoreo.

Como puede notar el número de datos faltantes puede parecer indefenso, sin embargo para el presente trabajo se les dará un trato especial pues son importantes al momento de hacer inferencia. En cuanto al análisis estadístico no se pretende analizar todas las estaciones solo se analizarán los datos reportados de la estación Coyoacán.

¹ Decimos que hay *sesgo positivo* o *hacia la derecha* si la distribución esta concentrada a la izquierda y la cola derecha es más larga. Por el contrario cuando hay *sesgo negativo* o *hacia la izquierda* la distribución esta concentrada a la derecha y la cola izquierda es más larga.

4. DATOS DE OZONO EN LA CIUDAD DE MÉXICO

4.2. Estación Coyoacán

En cuanto a la distribución espacial de los contaminantes con el mayor riesgo para la salud, durante los últimos años la delegación Coyoacán ha sido una de las más afectadas por el ozono. En la Tabla 4.3 se muestran las características más importantes de los datos de esta estación y puede observar que el 32.55% de los datos reportados por esta estación rebasan el límite de ozono permitido (95 ppb), por lo que es fácil deducir que la concentración de ozono si representa un problema serio para esta estación.

En la Figura 4.5 se muestran las gráficas de las concentraciones diarias por hora de ozono (O_3) de Enero de 2012 a Mayo de 2017.

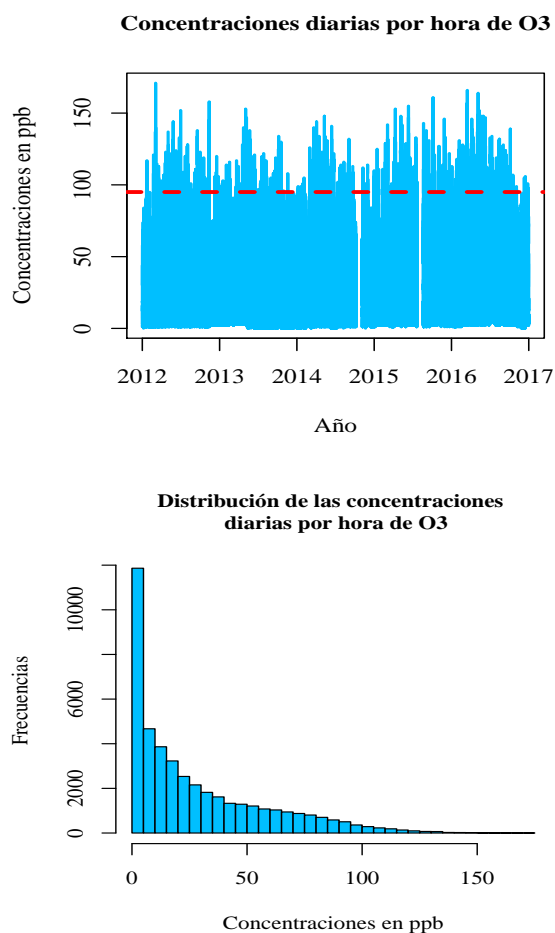


Figura 4.5: Concentraciones diarias por hora de O_3 de la estación COY.

En la gráfica de las concentraciones diarias por hora, la línea punteada en rojo indica

el nivel de concentración de ozono por hora permitido (95 ppb), observe que hay valores muy extremos que sobrepasan el nivel de ozono indicado y en el histograma note que la distribución de las concentraciones de ozono está sesgado hacia la derecha ya que se tiene una gran cantidad de observaciones que se encuentran en la cola derecha de la distribución, por lo que se extrajeron las concentraciones máximas por día y vamos a analizar la distribución de estos valores extremos durante el mismo periodo de tiempo. La Figura 4.6 muestra sus respectivas gráficas.

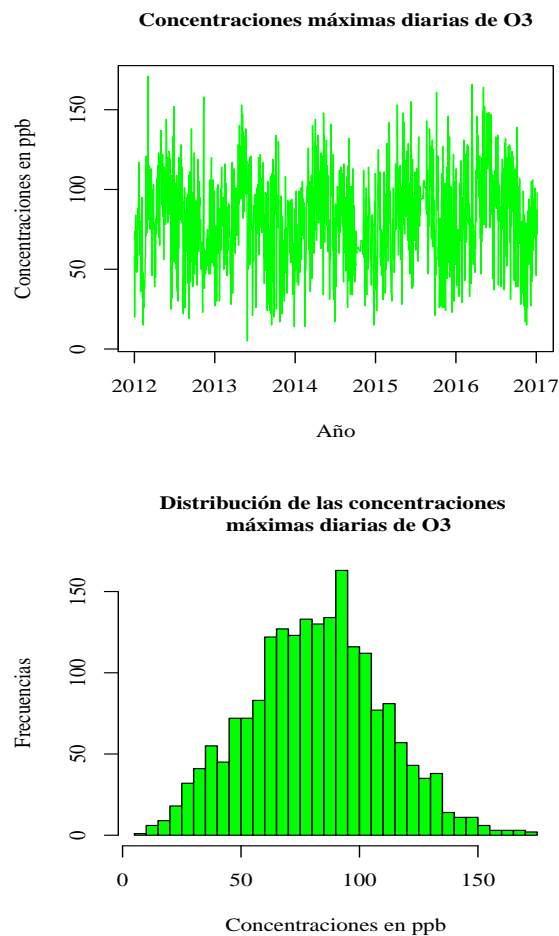


Figura 4.6: Concentraciones máximas diarias de O₃ de la estación COY.

Observe cómo cambia la distribución de las concentraciones máximas por día en comparación de las concentraciones por hora, en este caso la distribución está centrada y es aproximadamente simétrica. Algo más que puede notar es que la gráfica no es continua pues hay días en los que no se tiene registro de ozono, todo el análisis estadístico se hará sobre esta distribución que representa las máximas concentraciones de ozono

4. DATOS DE OZONO EN LA CIUDAD DE MÉXICO

por día de la estación Coyoacán. En la Tabla 4.3 se reportan las características más importantes de las concentraciones máximas diarias de ozono de Enero de 2012 a Mayo de 2017.

Núm. obs.	Mín.	Mediana	Media	Máx.	Obs. faltantes	Obs.> 95ppb
1978	5 ppb	83 ppb	82.1 ppb	175 ppb	312	644

Tabla 4.3: Características importantes de las concentraciones máximas de ozono.

Note que la mediana está cerca de la media por ello la distribución es aproximadamente simétrica y además se encuentran dentro del nivel de ozono permitido, sin embargo, cerca del 32% de los datos rebasan el límite de ozono permitido. Por otro lado como ya se había mencionado hay valores faltantes lo cual sigue siendo un problema y corresponde al 15.77% del total de los datos. En la siguiente sección se imputarán estos valores faltantes.

4.3. Valores Faltantes

El concepto de valores faltantes es importante de entender para poder hacer un análisis más completo de los datos. Si se omiten dichos valores puede terminar haciendo una inferencia incorrecta sobre los datos. Sin embargo existen métodos de imputación que consisten en estimar los valores faltantes a partir de la información del conjunto de datos o bien del uso de algunas variables para estimar estos valores.

Se mencionó en la sección anterior que hay un número de valores faltantes debido principalmente a la calibración y mantenimiento de los instrumentos con los que se lleva a cabo la medición de las concentraciones de ozono. Para imputar los datos se empleó el paquete `imputeTS`¹ en R, especializado en la imputación de valores faltantes en series de tiempo univariadas. Los algoritmos de imputación disponibles son promedios móviles ponderados, interpolación, suavizamiento de Kalman, entre otros.

Para la imputación se usó el algoritmo de promedios móviles ponderados pero primero se explicará de manera muy general qué son los promedios móviles y posteriormente la función con la que se hizo la imputación. Un promedio móvil se usa comúnmente con datos de series de tiempo para suavizar las fluctuaciones a corto plazo y resaltar las tendencias o ciclos estacionales a largo plazo. Se construye sustituyendo cada valor de una serie por la media obtenida con esa observación. Consiste en fijar un número k y calcular los promedios de todos los grupos de k términos consecutivos de la serie, si la

¹Se recomienda leer la documentación del paquete en [23] y estudiar los demás algoritmos que se pueden implementar.

serie presenta un efecto estacional de periodo d es conveniente usar este valor o algún múltiplo de él, así el pronóstico de promedio móvil de orden k es

$$\hat{X}_{t+1} = \frac{X_t + X_{t-1} + \dots + X_{t-k+1}}{k},$$

donde \hat{X}_{t+1} es el pronóstico para el periodo $t + 1$ y X_t es el valor actual para el periodo t . En este método de promedios móviles cada observación en el cálculo del promedio móvil recibe el mismo peso. Una variación conocida como promedios móviles ponderados implica elegir un peso diferente para cada dato y luego calcular un promedio ponderado de los k valores más recientes como el pronóstico. En la mayoría de los casos la observación más reciente recibe el mayor peso y disminuye para valores más lejanos.

La ponderación más usada es el suavizamiento exponencial el cual aplica factores de ponderación que disminuyen exponencialmente. La ponderación para cada observación antigua se vuelve más pequeña pero nunca llegando a cero. El promedio móvil exponencial puede calcularse recursivamente como

$$\hat{X}_{t+1} = \alpha X_t + (1 - \alpha)\hat{X}_t,$$

donde \hat{X}_{t+1} es el pronóstico para el periodo $t + 1$, X_t es el valor de la serie en el periodo t , \hat{X}_t es el pronóstico al tiempo t y α es la constante de suavizamiento la cual $0 \leq \alpha \leq 1$. El pronóstico para el periodo $t + 1$ es un promedio ponderado del valor en el periodo t y el pronóstico $t + 1$. Resulta que el suavizamiento exponencial para cualquier periodo es en realidad un promedio ponderado de todos los valores anteriores.

Mediante el uso de la función `na.ma`¹ en la cual los valores faltantes se reemplazan por promedios móviles, la media en este algoritmo se toma de un número igual de observaciones en cualquier lado de un valor central, es decir, que para un valor NA en la posición i de una serie de tiempo se usará el promedio

$$\hat{Z}_i = \frac{Z_{i-k} + Z_{i-k+1} + \dots + Z_{i-1} + Z_{i+1} + \dots + Z_{i+k-1} + Z_{i+k}}{2k}.$$

Además la función permite opciones para usar el promedio móvil simple (SMA), el promedio móvil ponderado lineal (LWMA) y el promedio móvil ponderado exponencial (EWMA).

- *Promedio móvil simple (SMA)*: todas las observaciones son igualmente ponderadas para calcular la media.

- *Promedio móvil ponderado lineal (LWMA)*: las observaciones están ponderadas y estos pesos disminuyen en progresión aritmética, es decir, las observaciones al lado de un valor central i tienen un peso de $1/2$, las observaciones más alejadas ($i - 2, i + 2$)

¹Esta función se encuentra en el paquete [23].

4. DATOS DE OZONO EN LA CIUDAD DE MÉXICO

tienen un peso de $1/3$, el siguiente $(i - 3, i + 3)$ tienen un peso i y así sucesivamente.

- *Promedio móvil ponderado exponencial (EWMA)*: las observaciones están ponderadas por pesos que disminuyen exponencialmente, las observaciones a lado de un valor central i tienen un peso de $(1/2)^1$, las observaciones más alejadas $(i - 2, i + 2)$ tienen un peso de $(1/2)^2$, el siguiente $(i - 3, i + 3)$ tienen un peso de $(1/2)^3$ y así sucesivamente.

Es importante mencionar cómo están distribuidos los valores faltantes pues en estos datos existen brechas largas de NA's en las que los valores al lado del valor central NA también son valores NA. Sin embargo el algoritmo es semiadaptivo pues si ocurre lo anterior el número de términos en la ponderación aumentará de forma incremental hasta que haya valores que no son NA. Ahora para determinar los k términos para el promedio móvil es necesario analizar si la serie de las concentraciones máximas de ozono presenta ciclos estacionales y cada cuándo se llevan a cabo. Para ello la Figura 4.7 muestra el comportamiento de los datos durante los últimos 5 años.

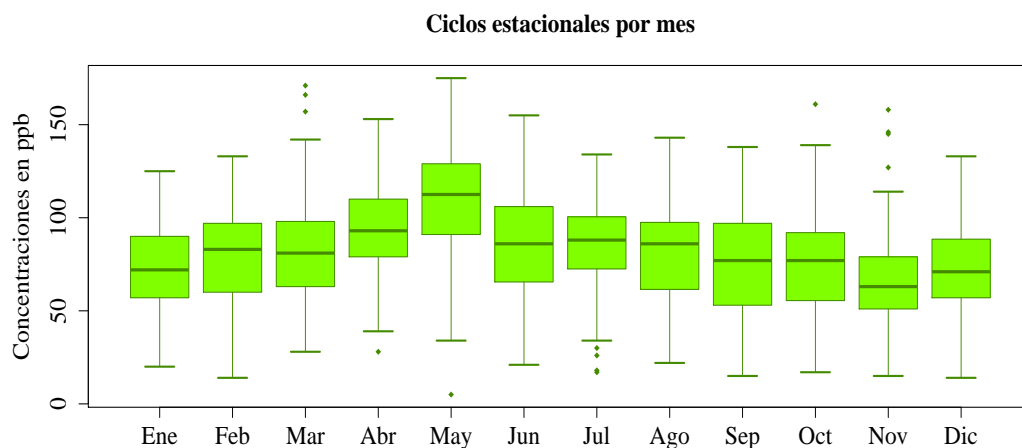


Figura 4.7: Ciclos estacionales por mes de las concentraciones máximas de ozono, estación Coyoacán.

Observe en la figura 4.7 que la serie presenta ciclos estacionales. En el mes de Mayo se registran mayores concentraciones de ozono y en todos los años se presenta el mismo comportamiento de hecho a este patrón se le conoce como la *temporada de ozono* y se refiere a los meses en los que aumenta la concentración del ozono en el aire debido a la combinación de diferentes factores meteorológicos. Generalmente inicia en la segunda quincena de Febrero y concluye con la temporada de lluvia en Junio. De tal forma que el periodo del ciclo estacional es anual, es decir, $d = 365$ ya que las observaciones son diarias. Tomando en cuenta las sugerencias de cómo determinar k y además de que tenemos brechas largas de NA se toma a $k = 365$ y adicionalmente la ponderación

será exponencial. Esta técnica es más eficiente que la media móvil simple pues como los factores decaen exponencialmente le da mayor ponderación a los datos más recientes. En la siguiente tabla se muestran las características estadísticas más importantes después de la imputación.

Núm. obs.	Mín.	Mediana	Media	Máx.	Obs. faltantes
1978	5 ppb	83 ppb	82.01 ppb	175 ppb	0

Tabla 4.4: Características importantes de las concentraciones máximas de ozono con los datos imputados.

Es claro que los valores no varían en gran medida con los de la tabla 4.3. Sin embargo como se tenía un número representativo de valores faltantes no se podían ignorar y así se tendrá mayor certeza a la hora de hacer inferencia.

4.4. Modelo Bayesiano para los Datos de Ozono en la Estación Coyoacán

De acuerdo al tipo de preguntas que se busque responder se pueden usar diferentes metodologías para estudiar el comportamiento del ozono. Como se han analizado los registros de la estación Coyoacán, en esta sección se propone un primer modelo que describa el comportamiento de las concentraciones máximas por día de ozono del 1 de Enero de 2012 al 31 de Mayo de 2017 de esta estación de monitoreo. En este primer modelo vamos a ajustar la distribución GEV a nuestros datos y la estimación de los parámetros será desde un enfoque Bayesiano. Es importante mencionar que vamos a suponer en este primer modelo los parámetros de la GEV son constantes a través del tiempo, por lo que el modelo es el siguiente.

Sean $Z_1, Z_2, \dots, Z_{1978}$ la serie de observaciones que corresponden a las concentraciones máximas de O_3 diarias en ppb (partes por billón) de la estación Coyoacán de manera que siguen una distribución GEV tal que

$$Z_t \sim GEV(\mu, \sigma, \xi) \quad \text{con } t = 1, \dots, 1978.$$

Como ya se mencionó para estimar los parámetros de la distribución GEV se usará un enfoque Bayesiano en el cual asignaremos distribuciones a priori a los parámetros. Para la distribución GEV no hay distribuciones a priori conjugadas, por tanto la elección de las distribuciones a priori será de manera no informativa ya que no se tiene información inicial. Aunque no lo parezca usar estas distribuciones aporta información vaga o general sobre los parámetros de una forma más objetiva, pues sólo respeta el espacio paramétral

4. DATOS DE OZONO EN LA CIUDAD DE MÉXICO

en el que están los parámetros $\mu \in \mathbb{R}$, $\sigma > 0$ y $\xi \in \mathbb{R}$ por lo que las distribuciones a priori elegidas son

$$\begin{aligned} Z_t | \mu, \sigma, \xi &\sim GEV(\mu, \sigma, \xi) \quad t = 1, \dots, 1978, \\ \mu &\sim Normal(0, 0.01), \\ \sigma &\sim Gamma(0.01, 0.01), \\ \xi &\sim U(-0.5, 0.5). \end{aligned}$$

Por lo visto en el Capítulo 2 la distribución a posteriori es

$$p(\mu, \sigma, \xi | \mathbf{Z}) \propto \overbrace{L(\mu, \sigma, \xi; \mathbf{Z})}^{\text{Verosimilitud}} \underbrace{p(\mu, \sigma, \xi)}_{\text{Distribución a priori}}.$$

Como estamos suponiendo que los parámetros son a priori independientes

$$p(\mu, \sigma, \xi | \mathbf{z}) \propto L(\mu, \sigma, \xi; \mathbf{z}) p(\mu) p(\sigma) p(\xi),$$

y como asumimos que los datos son condicionalmente independientes dados los parámetros la verosimilitud es

$$\begin{aligned} L(\mu, \sigma, \xi; \mathbf{z}) &= \prod_{t=1}^{1978} \frac{1}{\sigma} \left[1 + \xi \left(\frac{z_t - \mu}{\sigma} \right) \right]^{-1/\xi - 1} \exp \left\{ - \left[1 + \xi \left(\frac{z_t - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \\ &= \frac{1}{\sigma^{1978}} \prod_{i=1}^{1978} \left[1 + \xi \left(\frac{z_t - \mu}{\sigma} \right)^{-1/\xi - 1} \right] \\ &\quad \times \exp \left\{ - \sum_{t=1}^{1978} \left[1 + \xi \left(\frac{z_t - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \\ &= \frac{1}{\sigma^{1978}} \exp \left\{ \left(-\frac{1}{\xi} - 1 \right) \sum_{t=1}^{1978} \log \left[1 + \xi \left(\frac{z_t - \mu}{\sigma} \right) \right] \right\} \\ &\quad \times \exp \left\{ - \sum_{t=1}^{1978} \left[1 + \xi \left(\frac{z_t - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \end{aligned}$$

y las distribuciones a priori son

$$\begin{aligned} p(\mu) &= \frac{1}{\sqrt{2\pi(0.01)^2}} e^{-\frac{\mu^2}{2(0.01)^2}}, \\ p(\sigma) &= \frac{(0.01\sigma)^{0.99}}{\Gamma(0.01)} 0.01 e^{-0.01\sigma}, \\ p(\xi) &= \mathbb{1}_{[\xi \in (-0.5, 0.5)]}. \end{aligned}$$

La distribución a posteriori es

$$\begin{aligned}
 p(\mu, \sigma, \xi | \mathbf{Z}) &\propto \frac{1}{\sigma^{1978}} \exp \left(\left(-\frac{1}{\xi} - 1 \right) \sum_{t=1}^{1978} \log \left(1 + \xi \left(\frac{z_t - \mu}{\sigma} \right) \right) \right) \\
 &\times \exp \left(- \sum_{t=1}^{1978} \left(1 + \xi \left(\frac{z_t - \mu}{\sigma} \right) \right)^{-1/\xi} \right) \times \frac{1}{\sqrt{2\pi(0.01)^2}} e^{-\frac{\mu^2}{2(0.01)^2}} \\
 &\times \frac{(0.01\sigma)^{0.99}}{\Gamma(0.01)} 0.01 e^{-0.01\sigma} \times \mathbf{1}_{[\xi \in (-0.5, 0.5)]}.
 \end{aligned}$$

Habíamos explicado que en inferencia Bayesiana el objeto de estudio se centra en la distribución a posteriori $p(\theta | Z)$, que es de donde queremos obtener muestras para realizar inferencia. Como no hay un método general para muestrear de distribuciones de altas dimensiones, en este caso haremos uso de los métodos MCMC, donde se construye una cadena de Markov cuya distribución estacionaria sea nuestra distribución a posteriori y por medio de la integración Monte Carlo obtenemos las características muestrales de interés.

El MCMC produce muestras a posteriori para todas las cantidades desconocidas de nuestro modelo a través de la interface de R con JAGS¹, el algoritmo y el código se anexa en el Apéndice, se realizó un periodo de calentamiento de 10,000 iteraciones con 50,000 iteraciones adicionales que se utilizaron para producir las inferencias a posteriori, se simularon tres cadenas y la convergencia se monitoreó y se verificó a través de las trazas de cada parámetro estimado, en la Figura 4.8 se muestran las trazas correspondientes a cada parámetro.

¹En [26] y [40] puede encontrar los respectivos manuales de estos software.

4. DATOS DE OZONO EN LA CIUDAD DE MÉXICO

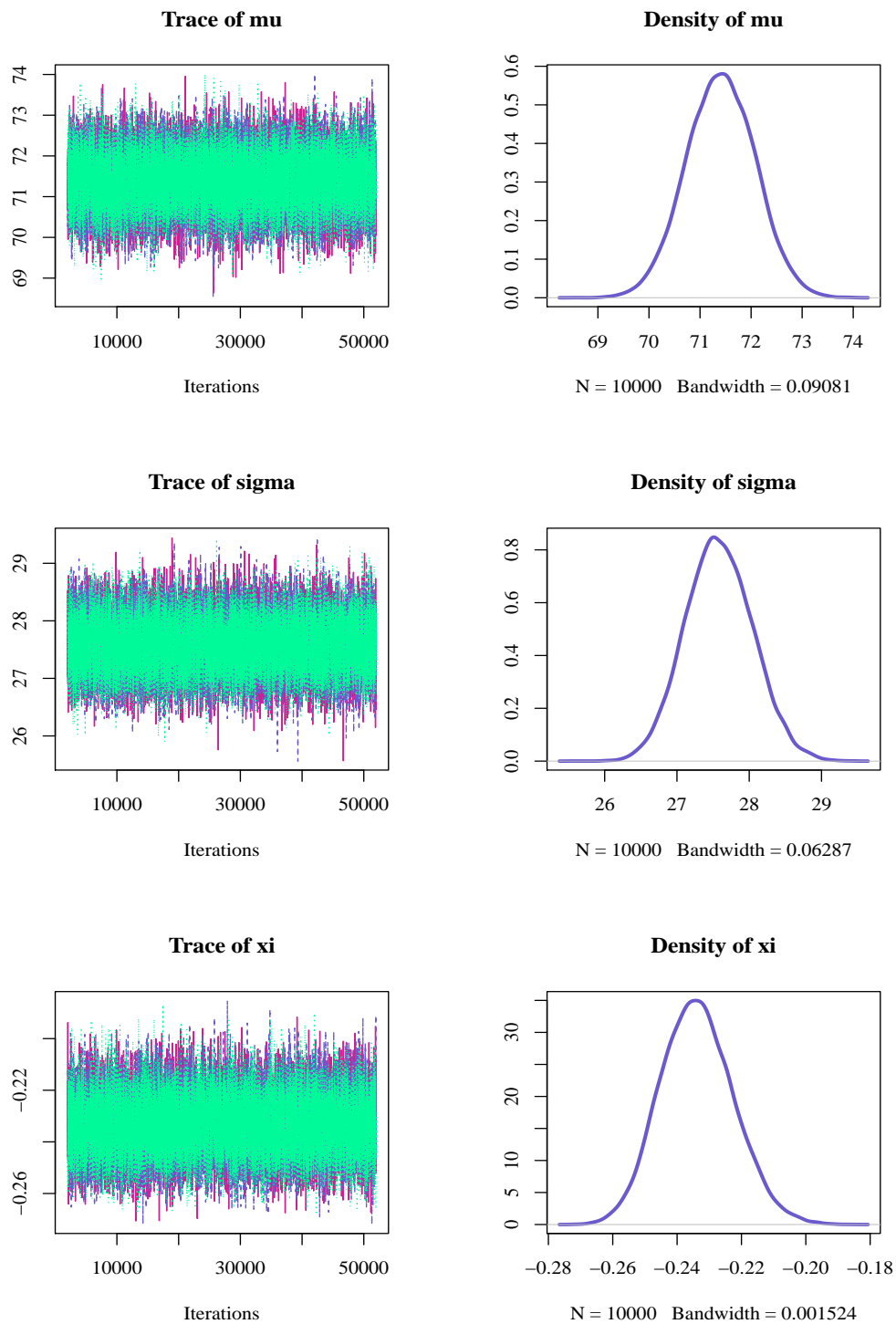


Figura 4.8: Traza y densidad de las cadenas simuladas para cada parámetro de la estación Coyoacán.

La herramienta más sencilla para visualizar la convergencia de una cadena de Markov es el gráfico de la izquierda que se denomina la traza del parámetro y muestra los valores que el parámetro tomó durante el tiempo de ejecución de la cadena quitando el periodo de calentamiento. En las trazas mostradas para μ , σ y ξ se puede observar que las tres cadenas simuladas presentan un comportamiento estable y el gráfico de la derecha es la densidad marginal que básicamente es el histograma suavizado de los valores de las trazas, es decir, la distribución de los valores del parámetro en la cadena. Aunque las cadenas parecen converger debemos hacer la prueba formal, usaremos la prueba de Gelman y Rubin por medio de la función `gelman.diag` implementada en la paquetería `coda`¹ de R, con un 95 % de confianza la Tabla 4.5 muestra los resultados obtenidos.

Parámetro	Estimación puntual	IC superior
μ	1	1
σ	1	1
ξ	1	1

Tabla 4.5: Factor de reducción de escala para cada parámetro estimado.

Por lo visto en la Subsección 2.2.3 como todos los valores son 1 podemos asegurar la convergencia de las cadenas. Ahora en la Tabla 4.6 se muestran las características a posteriori más importantes para cada parámetro.

Parámetro	Media	SD	Intervalo de credibilidad
μ	71.4059	0.6733	(70.0884, 72.7130)
σ	27.5827	0.4662	(26.7030, 28.5207)
ξ	-0.2338	0.0113	(-0.2551, -0.2111)

Tabla 4.6: Resumen de las características a posteriori para cada parámetro estimado de la estación Coyoacán.

La Tabla 4.6 muestra las estimaciones medias a posteriori, desviaciones estándar e intervalos de credibilidad al 95 %. Observe que el estimador puntual para μ es 71.4059, para el de escala σ es 27.5827 y para el parámetro de forma ξ es negativo lo cual corresponde a un comportamiento de cola tipo *Weibull* y para terminar con el análisis del primer modelo como se mencionó en la Subsección 2.1.7, una medida que combina

¹La documentación de esta paquetería la puede consultar en [27].

4. DATOS DE OZONO EN LA CIUDAD DE MÉXICO

la bondad del ajuste con la complejidad del modelo en un resumen numérico es el DIC el cual fue $DIC = 18,811.96$, más adelante servirá para la comparación de modelos.

Capítulo 5

Modelos de Valores Extremos en Secuencias de Datos Dependientes

Hasta ahora en la modelación de valores extremos se supone que hay un proceso implícito de una secuencia de variables aleatorias independientes. Sin embargo en muchas aplicaciones uno encuentra una dependencia serial o espacial que no se debe ignorar y para el cual suponer la independencia es poco realista. En este contexto tienen sentido preguntas como: ¿Son válidos los resultados del caso de independencia para este caso? ¿Bajo qué condiciones? o, si no lo son, ¿cuáles son entonces las distribuciones límite? En estos casos lo usual será considerar que el proceso observado es estacionario.

5.1. Valores Extremos en Secuencias Estacionarias

Intuitivamente un proceso de variables aleatorias se considera estacionario si sus características estadísticas como la media, varianza, entre otras se mantienen constantes a través del tiempo. De manera más formal:

Sea X_1, X_2, \dots una secuencia de variables aleatorias. Se dice que dicha secuencia es estacionaria si para cada conjunto de índices $1 \leq t_1 < t_2 < \dots < t_m$ la distribución conjunta $(X_{t_1}, X_{t_2}, \dots, X_{t_m})$ coincide con la distribución conjunta de $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_m+h})$ para todo entero $m > 0$ y $h > 0$.

Algunas secuencias estacionarias son ejemplos de observaciones dependientes. Ahora se enunciará una condición que establece que si los eventos extremos están lo suficiente-

mente distantes en el tiempo son casi independientes. Esta condición es conocida como la *condición de dependencia* $D(u_n)$ ¹.

Condición de dependencia $D(u_n)$: Para una serie estacionaria X_1, X_2, \dots decimos que satisface la condición $D(u_n)$ si para todo $i_1 < \dots < i_p < j_1 < \dots < j_q$ con $j_1 - i_p > l$ se tiene que

$$|P\{X_{i_1} \leq u_n, \dots, X_{i_p} \leq u_n, X_{j_1} \leq u_n, \dots, X_{j_q} \leq u_n\} - P\{X_{i_1} \leq u_n, \dots, X_{i_p} \leq u_n\} P\{X_{j_1} \leq u_n, \dots, X_{j_q} \leq u_n\}| \leq \alpha(n, l_n), \quad (5.1)$$

donde $\alpha(n, l_n) \rightarrow 0$ para alguna secuencia l_n tal que $l_n/n \rightarrow 0$ como $n \rightarrow \infty$. Para secuencias de variables aleatorias independientes la diferencia en probabilidad de la expresión (5.1) es igual a cero. Si no son independientes las variables, esta condición asegura que si las variables están lo suficientemente alejadas entre sí, la diferencia en probabilidad es cercana a cero. Todo lo anterior se resume en el siguiente resultado.

Sean X_1, X_2, \dots un proceso estacionario y sean $M_n = \max\{X_1, \dots, X_n\}$. Entonces si $a_n > 0$ y b_n son secuencias de constantes tales que

$$P \left[\frac{(M_n - b_n)}{a_n} \leq z \right] \rightarrow G(z),$$

donde G es una función de distribución no degenerada y la condición $D(u_n)$ se satisface con $u_n = a_n z + b_n$ para cualquier número real z . G pertenece a la familia de distribuciones de valores extremos generalizada.

Este resultado es de gran importancia ya que dada una sucesión de variables que tenga una dependencia de largo alcance limitada en niveles extremos, los valores máximos de las secuencias estacionarias siguen las mismas reglas del límite de distribución que las de las secuencias independientes, es decir, se distribuyen igual que la distribución límite de las secuencias de variables independientes. No obstante, los parámetros de la distribución se verán afectados por la dependencia².

5.2. Valores Extremos en Secuencias No Estacionarias

Los procesos no estacionarios tienen características que cambian a través del tiempo. Contextualizando un poco, el cambio climático en los datos ambientales provoca patrones climáticos en diferentes meses del año generando efectos estacionales en los

¹Para más detalles la condición $D(u_n)$ y el resultado se obtuvieron de [5], [7].

²Una forma de analizar cómo se ven los nuevos parámetros es comparando los máximos del proceso estacionario con respecto al proceso de variables independientes y el desarrollo lo puede consultar en [7].

datos o incluso tendencias que se generan a través del tiempo. En la sección anterior se vio que en cierto sentido y sujeto a limitaciones específicas los modelos de distribución límite para valores extremos son aplicables en los procesos estacionarios, pero para los procesos no estacionarios no se pueden establecer los mismos modelos.

En este caso se suponen procesos no estacionarios con cambios a lo largo del tiempo. La no estacionariedad se puede expresar por medio de cambios en los parámetros del modelo. Por ejemplo se podría expresar al parámetro de localización como una función lineal

$$\mu(t) = \theta_1 t,$$

o como un polinomio de segundo grado

$$\mu(t) = \theta_0 + \theta_1 t + \theta_2 t^2,$$

así el modelo que se tiene para el proceso no estacionario es

$$GEV(\mu(t), \sigma, \xi).$$

En otros casos la no estacionariedad también se puede expresar por medio del parámetro de escala, por ejemplo

$$\sigma(t) = \exp^{\theta_0 + \theta_1 t}.$$

Sin embargo el parámetro de forma ξ no es recomendable expresarlo en función del tiempo ya que como se vió en la Sección 3.3 este parámetro tiene ciertas restricciones para su estimación. No obstante un modelo con todos sus parámetros expresados en función del tiempo es

$$GEV(\mu(t), \sigma(t), \xi(t)).$$

De manera general, sea Z_1, \dots, Z_t un proceso que sigue una distribución GEV tal que

$$Z_t \sim GEV(\mu(t), \sigma(t), \xi(t)),$$

donde cada uno de los parámetros tiene una estructura de dependencia en el tiempo, la función de verosimilitud es

$$L(\boldsymbol{\mu}(t), \boldsymbol{\sigma}(t), \boldsymbol{\xi}(t); z) = \prod_{t=1}^n f(Z_t; \mu(t), \sigma(t), \xi(t)),$$

donde $f(Z_t; \mu(t), \sigma(t), \xi(t))$ denota la función de densidad de probabilidad de Z_t la cual tiene una distribución GEV con parámetros $\mu(t), \sigma(t), \xi(t)$ tal que tienen una estructura de dependencia temporal. Al igual que con el modelo GEV sin dependencia del tiempo, también se pueden hacer estimaciones por máxima verosimilitud o usando otros métodos de estimación¹.

¹Para profundizar en los modelos de valores extremos para procesos no estacionarios consulte [7].

5.3. Modelo Bayesiano Jerárquico Temporal

Históricamente el mes con el promedio máximo de ozono es Mayo y en la Figura 5.1 puede observar que en efecto, las concentraciones de ozono aumentan considerablemente en Mayo a diferencia de los otros meses. Esto se debe a que el ozono se forma por la reacción fotoquímica con la luz solar y otros contaminantes procedentes de emisiones de vehículos o de la industria. Por lo tanto en temporadas de calor las concentraciones tienden a elevarse. Además ya se había mencionado la famosa *temporada de ozono* que inicia en la segunda quincena de Febrero y concluye con la temporada de lluvia en Junio.

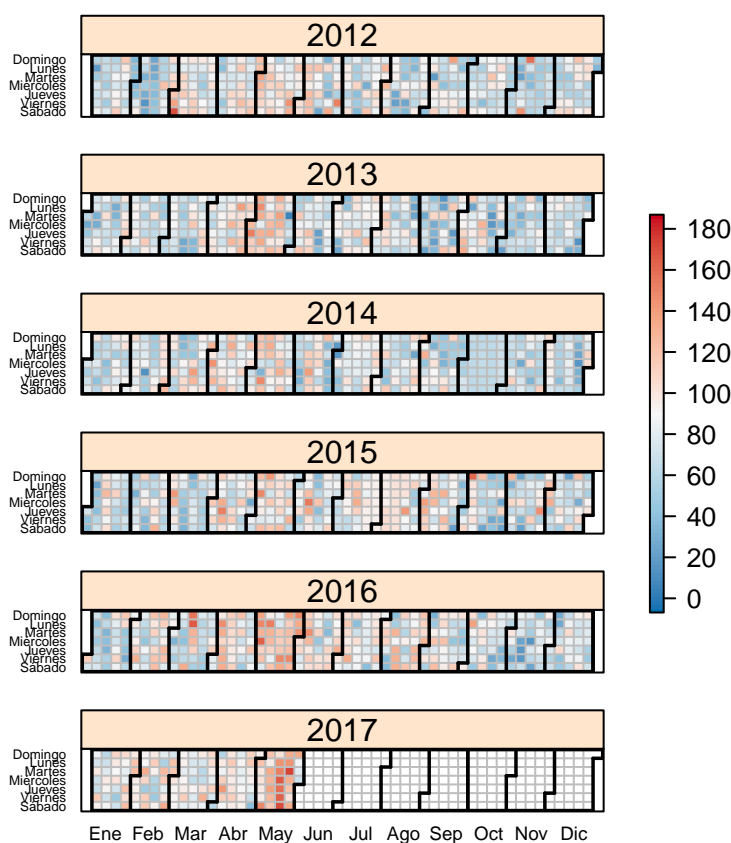


Figura 5.1: Calendario de las concentraciones máximas de ozono por día de la estación Coyoacán.

Debido a estos patrones que se forman en el calendario y a los ciclos estacionales que se observan en la Figura 4.7 nos sugiere que la serie de observaciones puede tener una

estructura de dependencia temporal. En esta sección se propone un modelo Bayesiano jerárquico temporal para las máximas concentraciones de ozono en la CDMX, el cual se adaptará a los datos reportados por la estación Coyoacán, que es con la que se ha hecho el análisis estadístico.

Sean $Z_1, Z_2, \dots, Z_{1978}$ la serie de observaciones que corresponden a las concentraciones máximas de O_3 diarias en ppb (partes por billón) de la estación Coyoacán, de manera que siguen una distribución $GEV(\mu, \sigma, \xi)$ y para modelar la dependencia temporal se modifica el parámetro de localización en el que se asume una estructura de dependencia en el tiempo, así

$$Z_t \sim GEV(\mu_t, \sigma, \xi) \quad t = 1, \dots, 1978,$$

cuya función de densidad es

$$f(z_t | \mu_t, \sigma, \xi) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{z_t - \mu_t}{\sigma} \right) \right]^{-1/\xi - 1} \exp \left\{ - \left[1 + \xi \left(\frac{z_t - \mu_t}{\sigma} \right) \right]_+^{-1/\xi} \right\},$$

para $t = 1, \dots, 1978$ donde $\sigma \in \mathbb{R}^+$ es el parámetro de escala, $\xi \in \mathbb{R}$ es el parámetro de forma y μ_t es el parámetro de localización que guarda una dependencia autorregresiva de orden 1, $AR(1)$ ¹. La idea de los modelos autorregresivos $AR(p)$ es que los valores actuales de la serie dependen de los p valores anteriores, en este caso el valor actual sólo dependerá de $p = 1$ la observación anterior.

Para incluir más información al estimar los parámetros del modelo GEV nuevamente asignamos a los parámetros distribuciones a priori poco informativas de acuerdo a las creencias que suponemos sobre los parámetros de la distribución GEV , respetando los espacios parametrales. Determinar una distribución a priori es un elemento necesario incluso si no hay información previa. Asumimos que las observaciones son condicionalmente independientes dados los parámetros y además se tiene la siguiente jerarquía:

$$\begin{aligned} Z_t | \mu_t, \sigma, \xi &\sim GEV(\mu_t, \sigma, \xi), \\ \mu_t &\sim Normal(\alpha + \beta \mu_{t-1}, 1/\phi), \\ \log(\sigma) &\sim Normal(0, 0.01), \\ \xi &\sim U(-0.5, 0.5). \end{aligned}$$

Ya que se asume una estructura de dependencia temporal $AR(1)$ para el parámetro de localización, éste queda expresado como

$$\mu_t = \alpha + \beta \mu_{t-1} + \varepsilon_t \quad \text{donde} \quad \varepsilon_t \sim Normal(0, 1/\phi),$$

¹Para estudiar este y otros modelos de series de tiempo consultar [3], [6] y [18].

5. VALORES EXTREMOS EN DATOS DEPENDIENTES

con distribuciones a priori

$$\alpha \sim Normal(0, 0.001), \quad \beta \sim Normal(0, 0.001), \quad 1/\phi \sim Gamma(0.01, 0.01).$$

Determinar una distribución a priori es un elemento necesario incluso si no hay información previa. La elección de densidades normales es arbitraria y como los datos no tienen información previa, la varianza¹ es muy alta para reflejar la ausencia de información inicial. Por lo tanto trabajar con la parametrización $\log(\sigma)$ resulta ser más sencillo y se sigue respetando la positividad de σ .

Para simplificar notación definiremos a $\theta = (\mu_t, \sigma, \xi)$ y a $\eta = (\alpha, \beta, \phi)$, así la distribución conjunta a priori es

$$\begin{aligned} p(\eta, \theta) &= p(\eta)p(\theta|\eta) \\ &= p(\alpha)p(\beta)p(1/\phi)p(\sigma)p(\xi), \end{aligned}$$

la función verosimilitud es

$$L(\theta, \eta; \mathbf{Z}) = \prod_{t=1}^{1978} p(z_t|\mu_t, \sigma, \xi)p(\mu_t|\mu_{t-1}, \alpha, \beta, \phi).$$

Por tanto como se vió en el Capítulo 2 la distribución a posteriori es proporcional a la verosimilitud por la distribución conjunta a priori

$$p(\theta, \eta|\mathbf{z}) \propto L(\theta, \eta; \mathbf{Z})p(\theta|\eta)p(\eta)$$

Si la desglozamos tenemos que la distribución posterior es

$$p(\theta, \eta|\mathbf{z}) \propto \prod_{t=1}^{1978} p(z_t|\mu_t, \sigma, \xi)p(\mu_t|\mu_{t-1}, \alpha, \beta, \phi)p(\alpha)p(\beta)p(1/\phi)p(\sigma)p(\xi).$$

La inferencia a posteriori será realizada usando los métodos estándar Monte Carlo vía cadenas de Markov (MCMC) que se describieron anteriormente. El MCMC produce muestras a posteriori para todas las cantidades desconocidas de nuestro modelo a través de la interface de R con JAGS, se realizó un periodo de calentamiento de 10,000 iteraciones con 50,000 iteraciones adicionales que se utilizaron para producir las inferencias a posteriori, se simularon tres cadenas y la convergencia se monitoreó y se verificó a través de las trazas de cada parámetro estimado, a continuación se muestran las trazas correspondientes a cada parámetro.

¹En muchos casos cuando se usa la distribución *Normal* se utiliza la precisión en lugar de la varianza que se define como $\tau = 1/\sigma^2$.

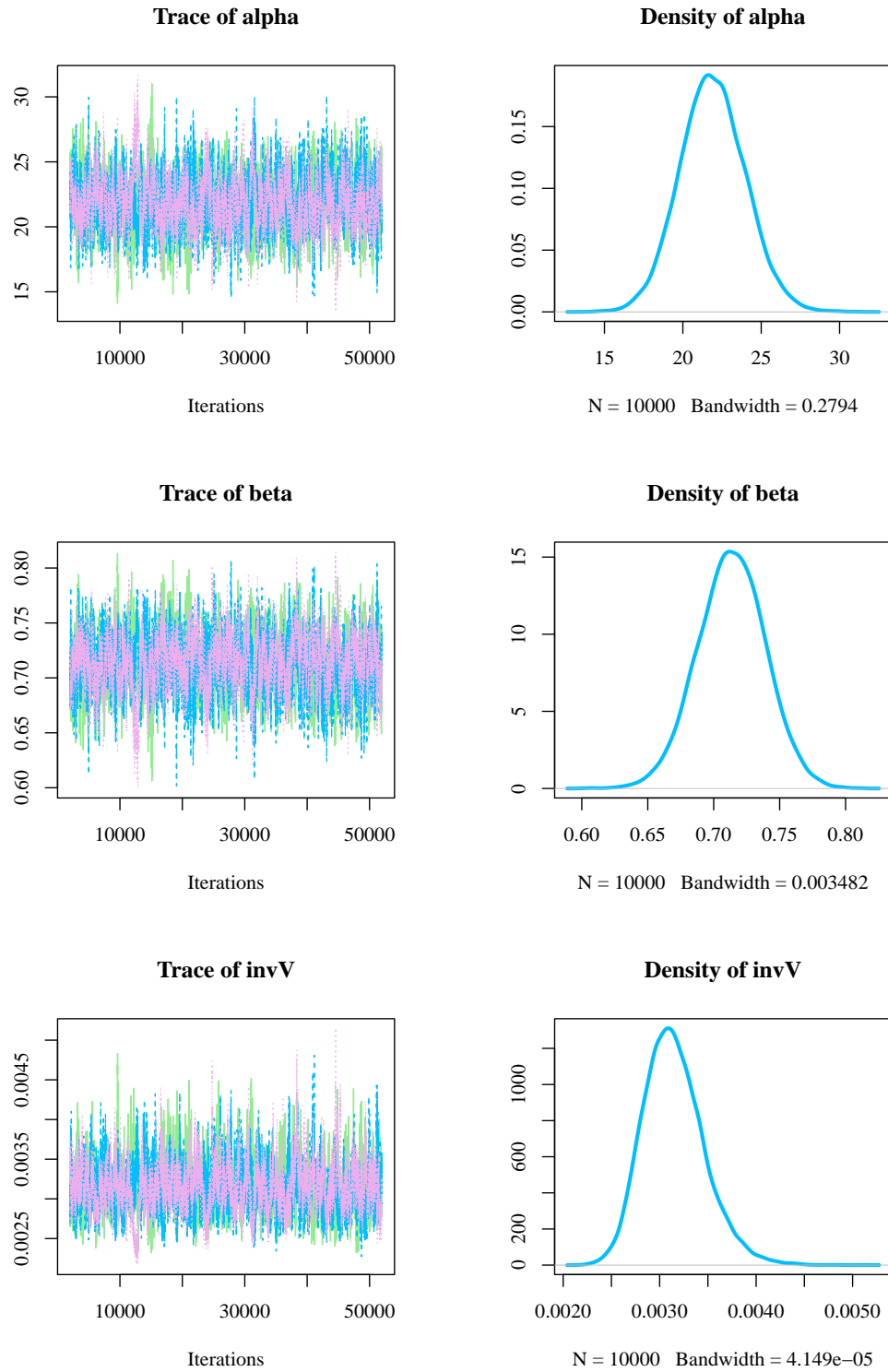


Figura 5.2: Traza y densidad de las cadenas simuladas para los parámetros α , β y $1/\phi$ de la estación Coyoacán.

5. VALORES EXTREMOS EN DATOS DEPENDIENTES

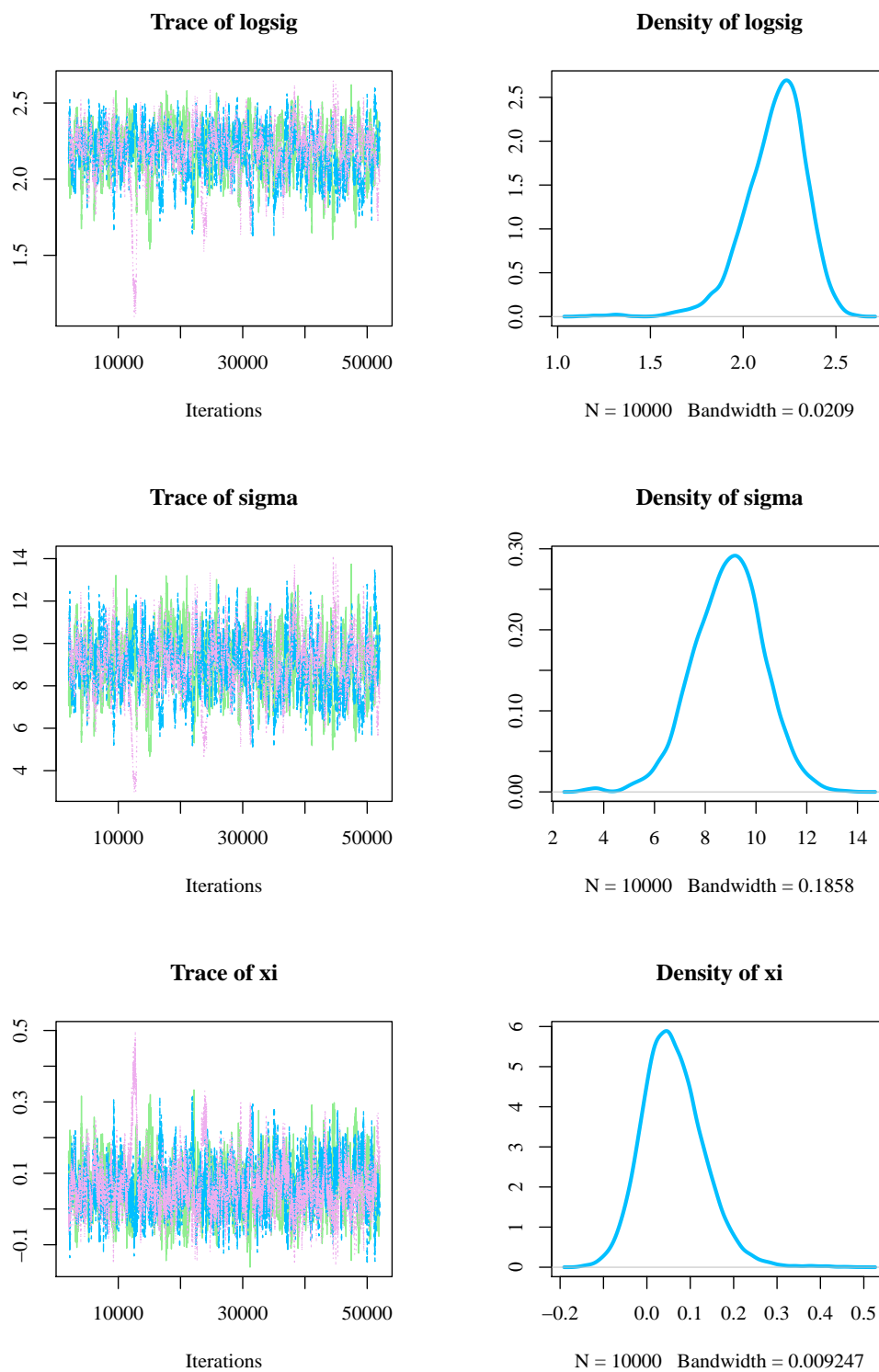


Figura 5.3: Traza y densidad de las cadenas simuladas para los parámetros $\log(\sigma)$, σ y ξ de la estación Coyoacán.

Como se puede observar en las Figuras 5.2 y 5.3 las cadenas se logran estabilizar y se superponen lo cual es un buen indicio de convergencia a la distribución estacionaria. Producto de la convergencia de las trazas en la Figura 5.2 se pueden observar las densidades estimadas para cada parámetro. Éstas se estabilizan casi simétricamente alrededor de la media a posteriori. Sin embargo, visualmente no se puede asegurar dicha convergencia. Recurrimos a la prueba de Gelman y Rubin, los valores sustancialmente superiores a 1 indican falta de convergencia, haciendo uso de la paquetería coda en R con un 95 % de confianza se obtuvieron los resultados de la Tabla 5.1.

Parámetro	Estimación puntual	IC superior
α	1.01	1.03
β	1	1.01
$1/\phi$	1.01	1.03
$\log(\sigma)$	1.01	1.03
σ	1.01	1.03
ξ	1.01	1.02

Tabla 5.1: Factor de reducción de escala para cada parámetro estimado.

Observe en la Tabla 5.1 que el factor en todos los casos es aproximadamente 1 entonces las cadenas sí convergen a la distribución estacionaria. Finalmente en la Tabla 5.2 se muestra un resumen de las características estadísticas obtenidas para cada parámetro.

Parámetro	Media	SD	Intervalo de credibilidad
$1/\phi$	0.0031	0.0003	(0.0025, 0.0038)
α	21.9100	2.0850	(17.9743, 26.1304)
β	0.7130	0.0258	(0.6630, 0.7636)
$\log(\sigma)$	2.1740	0.1671	(1.8638, 2.4753)
σ	8.9130	1.3960	(6.1207, 11.4910)
ξ	0.06108	0.0718	(-0.0708, 0.1930)

Tabla 5.2: Resumen de las características a posteriori para cada parámetro estimado de la estación Coyoacán.

5. VALORES EXTREMOS EN DATOS DEPENDIENTES

La Tabla 5.2 muestra las estimaciones medias a posteriori, desviaciones estándar e intervalos de credibilidad al 95 %, observe que el valor estimado de σ es 8.9130 y el parámetro de forma ξ estrictamente hablando es positivo lo cual corresponde a un comportamiento de cola tipo *Fréchet* pero también es un valor muy cercano a 0 lo cual también podría ser del tipo *Gumbel*. Para terminar con el capítulo en la subsección 2.1.7 se mencionó la importancia de comparar modelos por medio del *DIC* como una medida que combina la bondad del ajuste, para este modelo se obtuvo un valor $DIC = 16,900$.

Una consideración importante en cualquier análisis complejo es la necesidad de elegir entre modelos para los mismos datos. Aquí nos enfocaremos en un método llamado el Criterio de Información de Devianza, o el *DIC* para abreviar. Una forma de comparar el ajuste de dos modelos diferentes es comparar los valores de la probabilidad evaluados en las estimaciones de los parámetros. El modelo con mayor probabilidad se ajusta mejor a los datos. Nos gustaría seleccionar el modelo que mejor se ajuste a los datos con el menor número de parámetros. Este es el principio de parsimonia. En la Tabla 5.3 presentamos los resultados obtenidos para cada modelo donde DIC_{p_D} denota el criterio de la información de la devianza usando la penalización del número de parámetros estimados y el $DIC_{p_{opt}}$ usa como penalización el “optimismo“ (optimism) que más adelante se explicará.

Modelo 1 \Rightarrow Todos los parámetros son constantes en el tiempo.

Modelo 2 \Rightarrow Asumimos una estructura de dependencia *AR*(1) en el parámetro de localización.

Modelo	$\bar{D}(\theta)$	p_D	p_{opt}	DIC_{p_D}	$DIC_{p_{opt}}$
1	18,811	0.445	0.96	18,811.4	18,811.96
2	14,834	229.400	2,066.00	15,063.4	16,900.00

Tabla 5.3: Comparación de modelos de acuerdo al *DIC*.

Como recordará en la Subsección 2.1.7 el *DIC* se obtiene con la devianza esperada más una penalización que generalmente es el número de parámetros estimados. La función `dic.samples` en el paquete `rjags` en R extrae la devianza penalizada de un modelo de *JAGS*, esta función permite elegir dos alternativas para el cálculo de la penalización de la devianza, uno de ellos es el número de parámetros estimados p_D y el p_{opt} o conocido como *optimism*¹ que en general es una definición alternativa para el número efectivo de parámetros y en este caso el *DIC* se calcula sumando a la devianza esperada el optimismo, la penalización del p_{opt} es al menos el doble del tamaño de la

¹Los detalles del cálculo del p_{opt} o “optimismo“ es recomendable consultar [25].

penalización del p_D y penaliza más severamente a los modelos complejos.

Pasando a la comparación del DIC el modelo 2 es superior al modelo 1 en términos del rendimiento del DIC y a fin de tener un modelo parsimonioso nos quedamos con el modelo 2, pues en los dos tipos de penalización el DIC es menor además pensaríamos que es correcto establecer una estructura de dependencia temporal de acuerdo al contexto.

En este capítulo se han presentado dos modelos en el que se incluye la aplicación de valores extremos y los métodos de inferencia Bayesiana para las máximas concentraciones diarias de ozono (O_3) en ppb de la estación Coyoacán, con el fin de modelar la dependencia temporal que se ha observado en los últimos 5 años. Cabe mencionar que no es la única forma en la que se puede modelar esta dependencia temporal incluso se puede determinar la dependencia en los otros parámetros y no sólo en el de localización. Sin embargo, con los resultados obtenidos se puede concluir que resulta ser un modelo parsimonioso que describe las concentraciones de ozono y se puede adaptar a los datos de otra estación de monitoreo.

Capítulo 6

Conclusiones

En este trabajo de investigación se presentaron dos modelos. El modelo 1 consistía en ajustar la distribución *GEV* con los parámetros μ , σ y ξ constantes a través del tiempo y en el modelo 2 incluimos en el parámetro de localización μ una estructura autoregresiva de orden $p = 1$. Con base en la comparación de modelos, el modelo 2 resultó ser mejor que el modelo 1, ya que este representaba adecuadamente las variaciones estacionales de los datos, demostrando así, que existe una estructura de dependencia temporal en las concentraciones máximas de ozono de la estación Coyoacán.

Es primordial mencionar que el análisis de estos datos representó un reto interesante ya que se contaba con un gran número de observaciones. Sin embargo, por el cierre o apertura de las estaciones en ciertos meses se tenían cantidades representativas de datos faltantes que se pudieron imputar. Los datos faltantes son omnipresentes en muchas áreas de la investigación científica y pueden conducir a inferencias sesgadas o ineficientes si se ignoran o se manejan de manera inapropiada. Se han propuesto una variedad de enfoques para analizar tales datos, y utilizar un método de imputación adecuado para el tipo de datos que se presenten.

Si bien el presente trabajo abordó la problemática de la contaminación del aire, en particular los efectos que tiene el ozono en el aire, se buscó describir su comportamiento a través del tiempo. Estos modelos son un primer acercamiento a los lectores para que comprendan la importancia del monitoreo de ozono en la Ciudad de México, y no sólo de este contaminante sino de los demás contaminantes que se adhieren al aire para que se tomen las medidas necesarias para evitar que éstos rebasen los límites permisibles por las normas de salud y con ello reducir la contaminación atmosférica.

Estos datos se pueden estudiar desde una perspectiva clásica, sin embargo era fundamental recalcar que la estadística Bayesiana está en el auge del análisis de datos siendo una herramienta bastante importante que permite que los investigadores incluyan conocimientos previos sobre el tema y hoy en día se pueden hacer análisis interesantes de datos usando estos métodos.

6. CONCLUSIONES

Por último debe tener en cuenta que se pueden ajustar otros modelos a estos datos. En este trabajo sólo se presentaron dos propuestas para tratar la dependencia temporal pero no quiere decir que sean los únicos modelos. Como recordará en el Capítulo 4 se mostró un mapa de la localización de las estaciones de monitoreo, entonces cada observación está ubicada en el espacio por lo que se puede hacer un análisis espacial de este contaminante. Inclusive existen modelos más complejos que nos permiten incorporar tanto la estructura temporal como la estructura espacial y son conocidos como modelos espacio-temporales, para los cuales se recomienda revisar otros temas que pueden ser alternativas para modelar estos datos.

A continuación se presenta una propuesta de análisis espacial de las concentraciones máximas de ozono y lo considere como tema subsecuente a esta tesis.

Los variogramas permiten analizar el comportamiento espacial de una variable sobre un área definida. La función se define como:

$$\bar{\gamma}(h) = \frac{1}{2N(h)} \left[\sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \right],$$

donde $Z(x_i)$ es el valor de la variable en un sitio x_i , $Z(x_i + h)$ es otro valor muestral separado del anterior por una distancia h y $N(h)$ es el número de parejas que se encuentran separadas por dicha distancia. En la Figura 6.1 se muestra la distribución espacial y el variograma estimado de las estaciones de monitoreo de la Ciudad de México.

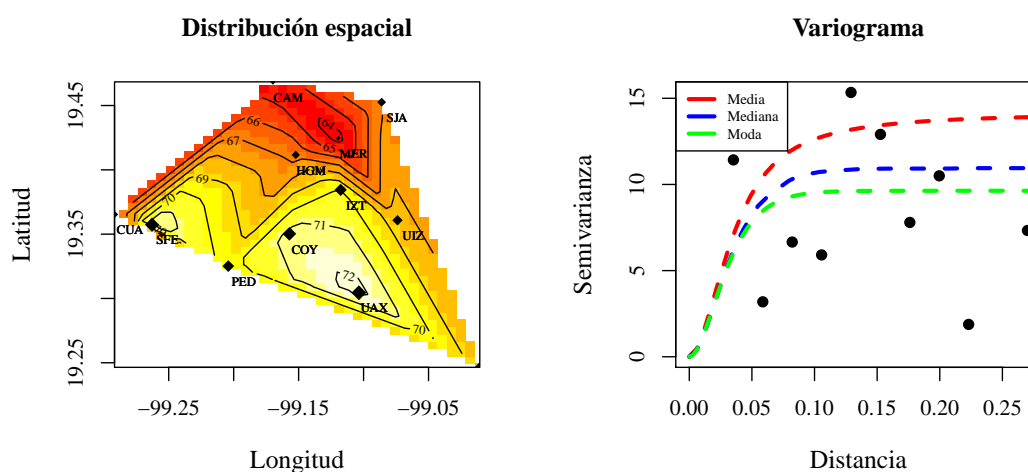


Figura 6.1: Distribución espacial y variograma de las estaciones de monitoreo de la CDMX.

Se requiere predecir el valor $Z(x_0)$ entonces se asume que el proceso espacial se puede expresar como $Z(x) = \mu + \epsilon(x)$, $x \in D$ donde D es un subconjunto fijo y continuo de \mathbb{R}^m , es decir, $D \subset \mathbb{R}^m$, $\mu \in \mathbb{R}$ y $\epsilon(x)$ es el error aleatorio en el punto x_0 . El predictor lineal que el *kriging* utiliza es:

$$\hat{Z}(x_0) = \sum_{i=1}^n \varphi Z(x_i) = \varphi' \mathbf{Z}_x,$$

donde $\varphi' = (\varphi_1, \dots, \varphi_n)'$ es el peso asignado a cada uno de los datos, si denotamos a $\sigma^2(x_0)$ el error cuadrático y se utilizan distribuciones a priori no informativas para los parámetros de interés $\theta = (\mu, \sigma^2, \varphi)$

$$p(\mu) \propto 1, \quad p(\sigma^2) \propto \frac{1}{\sigma^2}, \quad p(\varphi) \propto U(0, 1).$$

La distribución predictiva es

$$p(Z(x_0) | Z) = \iiint p(Z(x_0), \mu, \sigma^2, \varphi | Z) d\mu d\sigma^2 d\varphi.$$

En la Figura 6.2 se muestran los mapas de los valores esperados y su varianza de las concentraciones máximas de ozono para las estaciones de la Ciudad de México, cabe mencionar que la escala de colores va de rojo a blanco, donde el rojo representa valores bajos y blanco valores altos.

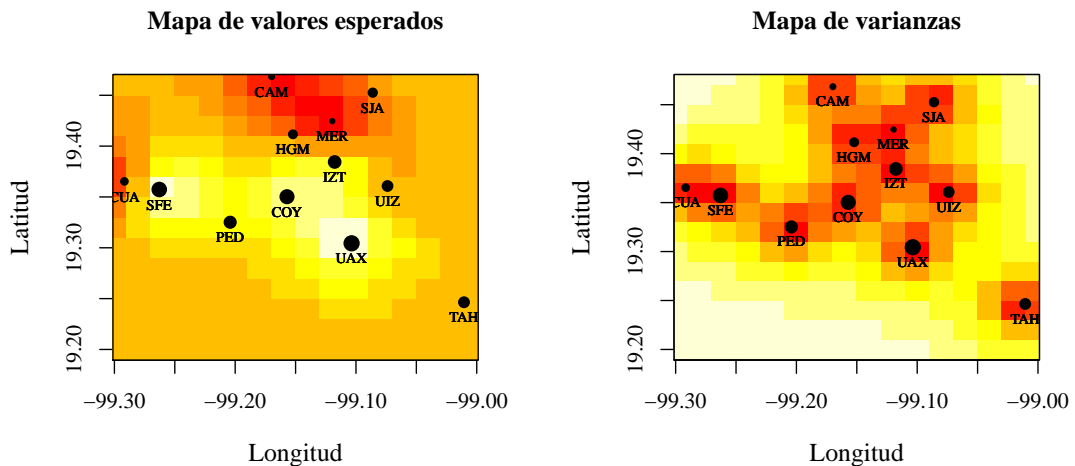


Figura 6.2: Mapas de valores esperados y varianzas de la distribución predictiva para las estaciones de monitoreo de la CDMX.

6. CONCLUSIONES

Se puede extender el modelo a los datos en el espacio y el tiempo añadiendo un proceso Gaussiano de tal forma que el modelo Bayesiano jerárquico espacio temporal queda como

$$\begin{aligned}Z_{it} &\sim GEV(\mu_{it}, \sigma, \xi), \\ \mu_{it} &= \alpha + \beta * altura_i + \gamma * t + W_{it}, \\ W_{it} &\sim Normal_n(\mathbf{0}, \mathbf{\Omega}), \\ \Omega_{ii} &= \text{var}(W_{it}) = \lambda_w^2, \\ \Omega_{ij} &= \text{cov}(W_{it}, W_{jt}) = \tau_w^{-1} \exp\{-(\phi d)^{\kappa}\} \quad i \neq j.\end{aligned}$$

Para profundizar en estos modelos y en estadística espacial puede consultar [9], [17], [16], [32], [33] y [41].

Apéndice A

Apéndice

A.1. Introducción a BUGS

BUGS (Bayesian Inference Using Gibbs Sampling) es un lenguaje de programación de análisis Bayesiano que utiliza métodos MCMC para adaptarse a modelos estadísticos. Se puede usar en problemas simples como estimar medias y varianzas o tan complejos como adaptar modelos jerárquicos y modelos de datos faltantes.

Versiones actuales:

- a) WinBUGS 1.4.3 (se ejecuta en Windows)
 - Se puede ejecutar en el mismo programa o desde otro software.
 - Interfaces desarrolladas para R, Excel, Splus, SAS, Matlab.
 - Disponible en <https://www.mrc-bsu.cam.ac.uk/software/bugs/> y otros recursos como el manual, numerosos ejemplos y la subscripción a la lista de correo de usuarios.
- b) OpenBUGS (se ejecuta en Windows, Unix/Linux y Mac a través de Wine)
 - Versión de código abierto.
 - También es libre y se puede descargar en el link anterior.

Descargue el archivo .exe y siga las instrucciones para la instalación, una vez instalado a continuación se enuncia un ejemplo para ejecutarlo en WinBUGS.

Por ejemplo supongamos que queremos hacer inferencia para las medias de dos distribuciones normales independientes con varianzas desconocidas de tamaño $n = 10$. Tenemos $X_1, \dots, X_{10}, Y_1, \dots, Y_{10}$.

Modelo estadístico:

$$x_i \sim \text{Normal}(\mu_1, \sigma_1^2) \quad i = 1, \dots, 10$$

$$y_i \sim \text{Normal}(\mu_2, \sigma_2^2) \quad i = 1, \dots, 10$$

Distribuciones a priori:

$$\mu_1 \sim \text{Normal}(0, 0.01) \quad 1/\sigma_1^2 \sim \text{Gamma}(0.01, 0.01)$$

$$\mu_2 \sim \text{Normal}(0, 0.01) \quad 1/\sigma_2^2 \sim \text{Gamma}(0.01, 0.01)$$

Un programa de WinBUGS incluye tres secciones: modelo, datos y valores iniciales. Traduciendo el modelo estadístico a lenguaje de WinBUGS.

Modelo:

```

model{
for (i in 1:10){
x[i] ~ dnorm (mu[1], prec[1])
y[i] ~ dnorm (mu[2], prec[2])
}
mu[1] ~ dnorm (0, 0.0001)
mu[2] ~ dnorm (0, 0.0001)
prec[1] ~ dgamma (0.01, 0.01)
prec[2] ~ dgamma (0.01, 0.01)
s2[1] <- 1/prec[1]
s2[2] <- 1/prec[2]
}

```

- "*model{...}*" especifica el modelo estadístico que estamos ajustando.
- "*for (i in 1:10){...}*" denota que las dos afirmaciones en {...} se realicen 10 veces.
- Los corchetes permiten indexar un vector de valores, son equivalentes a los subíndices en el modelo anterior.
- WinBUGS usa la precisión como parámetro en lugar de una varianza de una distribución *Normal*, es decir, $prec. = 1/var.$

Los **datos** se cargan como:

```
list(x = c(6.49,2.65,9.14,5.31,6.58), y=c(9.06,8.95,5.66,9.78,8.09))
```

Por último necesitamos dar **valores iniciales** para los parámetros

```
list(mu=c(0,0),prec=c(1,1))
```

Y la forma de ejecutar en WinBUGS es

A. Cargar el modelo, los datos y los valores iniciales.

1. Abra un nuevo documento en WinBUGS y escriba las tres partes (modelo, datos, valores iniciales) en él.
2. Guarde el archivo como .odc.
3. Desde el panel superior: abrir *Model* → *Specification*. Se abrirá un cuadro de diálogo.
4. Haga doble clic (resalte) la palabra *model* en su archivo y haga clic en *check model* en el cuadro de diálogo.
5. Mire la esquina inferior izquierda y debe decir *model is syntactically correct*.
6. Resalte la palabra *list* para la sección de datos y haga clic en *load data* en el cuadro de diálogo. Mire la esquina inferior izquierda para *data loaded*.
7. Haga clic en *compile* y debe aparecer *model compiled*.
8. Resalte la palabra *list* para la sección de valor inicial y haga clic en *load inits* en el diálogo. Busque *model is initialized*.

B. Ejecutar el muestreador.

1. Abra la ventana *Sample monitor tool: Menu* → *Inference* → *Sample*.
2. Escriba los parámetros que le interesan en el cuadro *node* y haga clic en *set*. En este ejemplo nos interesa tanto *mu* como *s2*.
3. Abra la ventana *Update tool: Menu* → *Model* → *Update*.
4. En el cuadro *update* también, escriba la cantidad de muestras posteriores que desea en el cuadro de *updates*. Por ejemplo: 10000.
5. Haga clic en *Update* y mire cómo se ejecuta en el cuadro de iteración.

C. Inferencia posteriori.

1. En *Sample Monitor Tool*, seleccione *mu* en el cuadro desplegable en *node*.
2. Seleccione la cantidad de muestras iniciales que queremos colocar en el cuadro *beg*. Esto también se conoce como *burn-in*. Elija 1000 aquí.
3. Seleccione el número *j* de iteración que desea mantener en el cuadro *thin*. Por ejemplo, si eliges 10, entonces solo se usará la décima parte de la muestra de las iteraciones, es decir, 2000 al 10000 para la inferencia posterior. Vamos a elegir 10 aquí.

Y a partir de todos los resultados generados por las muestras a posteriori puede hacer el análisis correspondiente.

Dentro del lenguaje viene incluida una lista de distribuciones estándar como la distribución Normal, Poisson, Gamma, Binomial, etc. sin embargo, supongamos que deseamos utilizar una distribución de muestreo que no esté incluida en la lista de distribuciones estándar, en la que una observación x_i contribuye con un término de verosimilitud L_i . Podemos usar el ”**truco de ceros**”: una observación $Poisson(\lambda)$ de cero tiene verosimilitud $\exp(-\lambda)$, por lo que si nuestros datos observados son un conjunto de O 's, y λ_i se establece en $-\log(L_i)$, obtendremos la contribución de verosimilitud correcta. (Tenga en cuenta que λ_i siempre debe ser > 0 ya que es la media de una Poisson, por lo que puede que necesitemos agregar una constante adecuada para asegurar que sea positiva.) Este truco se ilustra a continuación en código BUGS.

```
C <- 10000 # debe ser lo suficientemente grande para que lambda[i]>0
for (i in 1:N) {
zeros[i] <- 0
lambda[i] <- -log(L[i]) + C
zeros[i] ~ dpois(lambda[i])
}
```

Este truco permite el uso de distribuciones de muestreo arbitrarias que no están en la lista de distribuciones estándar, y es particularmente adecuado cuando se trata de distribuciones truncadas.

A.2. Introducción a JAGS

JAGS (Just Another Gibbs Sampler) es un programa para el análisis de modelos jerárquicos Bayesianos utilizando métodos MCMC, no es muy diferente de BUGS, fue escrito con tres objetivos:

- Tener un motor multiplataforma para el lenguaje BUGS.
- Ser extensible, permitiendo a los usuarios escribir sus propias funciones, distribuciones y samplers.
- Ser una plataforma de experimentación con ideas en modelado Bayesiano.

Ejecutar un modelo se refiere a generar muestras a partir de la distribución posterior de los parámetros del modelo. Esto se lleva a cabo en cinco pasos:

1. Definición del modelo.
2. Compilación.
3. Inicialización.
4. Adaptación y burn-in.

5. Monitoreo.

El modelo se define en un archivo de texto usando un dialecto del lenguaje **BUGS**, es decir, es la misma estructura del modelo que en **WinBUGS**, los datos y los valores iniciales. Cuando se compila un modelo, se crea un gráfico que representa el modelo en la memoria de la computadora, por defecto, **JAGS** solo ejecuta una sola cadena. Existen algunas diferencias de sintaxis entre **BUGS** y **JAGS** que puede consultar en los respectivos manuales que se encuentran en la Bibliografía.

JAGS está diseñado para trabajar estrechamente con el lenguaje **R** y el entorno para el cálculo estadístico y los gráficos. Encontrará útil instalar el paquete **coda** para **R** para analizar la salida. También puede usar el paquete **rjags** para trabajar directamente con **JAGS** desde dentro **R**.

Los resultados que se presentan en este trabajo de investigación se obtuvieron con **JAGS** mediante su interfaz con **R** y cabe mencionar que la distribución *GEV* no se encuentra dentro de la lista de distribuciones estándar por lo que se hizo uso del **"truco de ceros"**.

A.3. Códigos

A.3.1. Modelo 1

```
# Estimación GEV Bayesiano sin jerarquía

# Cargamos las librerías
library(R2jags)
library(rjags) # interfaz con JAGS
library(coda) # sirve para analizar las muestras a posteriori

# Cargamos los datos
datos<- read.csv('COYSINNA.csv', header=T)
summary(datos)
attach(datos)

# Lista con los datos
data <- list(
Z = COY,
zeros = rep(0,length(COY)),
N = length(COY)
)
# Valores iniciales para los parámetros
```

A. APÉNDICE

```
inits <- function() {
list(mu=0, sigma=1, xi=0.1)
}
# Los parámetros que vamos a incluir en el muestreador
parameters <- c("mu","sigma","xi")

# Aquí definimos el modelo
model= ‘‘model {
# Para muestrear los valores de la dist. GEV usamos el truco de ceros
for(i in 1:N){
zeros[i] ~ dpois(lambda[i])
cota[i] <- 1 + xi*(y[i] - mu)/sigma
lambda[i] <- step(cota[i]) * ( log(sigma) + (1+1/xi)*log(abs(cota[i])) +
pow(abs(cota[i]),-1/xi) ) + step(-cota[i])*100000
}
# Distribuciones a priori
sigma ~ dgamma(0.01,0.01)
xi ~ dunif(-0.5,0.5)
mu ~ dnorm(0,0.001)
}‘‘
# Hacemos el ajuste con la función jags.model
fit1<- jags.model(textConnection(model), data, inits=inits, n.chains=2)
update(fit1,1000) # actualiza la cadena asociada con el modelo
sample1<- coda.samples(fit1, parameters, n.iter=50000, thin=1)

# En esta parte hacemos los gráficos y obtenemos las características
más importantes de las muestras a posteriori
dev.new(height=4,width=4)
par(family="serif")
traceplot(sample1, col=c(".orange1","turquoise3"),family="serif", cex=0.7,
cex.main=0.9)
densplot(sample1, col="turquoise4", lwd=2, family="serif", cex=0.7,
show.obs = F)
summary(sample1)
HPDinterval(sample1) # intervalos de credibilidad

# Con el paquete coda revisamos la convergencia de las cadenas
gelman.diag(sample1,confidence = 0.95)
gelman.plot(sample1)
geweke.diag(sample1, frac1=0.1, frac2=0.5)
geweke.plot(sample1)
autocorr.plot(sample1)
autocorr.diag(sample1)
```

```

effectiveSize(sample1)
# Por último obtenemos el DIC para este modelo
dic1<-dic.samples(fit1, n.iter = 1000, type = "pD")
dic2<-dic.samples(fit1, n.iter = 1000, type = "popt")

```

A.3.2. Modelo 2

```

# Estimación GEV Bayesiano Jerárquico

# Cargamos las librerías
library(R2jags)
library(rjags) # interfaz con JAGS
library(coda) # sirve para analizar las muestras a posteriori

# Cargamos los datos
datos<- read.csv('COYSINNA.csv', header=T)
summary(datos)
attach(datos)

# Lista con los datos
data <- list(
  Z = COY,
  zeros = rep(0,length(COY)),
  N = length(COY)
)
# Valores iniciales para los parámetros
inits <- function() {
  list( xi=0.1, alpha=20, beta=0.65, logsig=2, invV=0.1)
}
# Los parámetros que vamos a incluir en el muestreador
parameters <- c("xi", ".alpha", "beta", "logsig", "sigma",
  invV", "V")

# Aqui definimos el modelo
model='model{
for(t in 1:N){
zeros[t] ~ dpois(lambda[t])
cota[t] <- 1 + xi*(Z[t] - mu[t])/sigma
lambda[t] <- step(cota[t]) * ( logsig + (1+1/xi)*log(abs(cota[t])) +
pow(abs(cota[t]), -1/xi) ) + step(-cota[t])*1000000
}
for(t in 2:N){
mu[t] ~ dnorm(eta[t],invV)

```

A. APÉNDICE

```
eta[t]<- alpha + beta*mu[t-1]
}
mu[1] ~ dnorm(eta[1],invV)
eta[1] ~ dnorm(0,0.001)
# Distribuciones a priori
logsig ~ dnorm(0,0.001)
sigma <- exp(logsig)
xi ~ dunif(-0.5,0.5)
alpha ~ dnorm(0,0.001)
beta ~ dnorm(0,0.001)
invV ~ dgamma(0.01,0.01)
V <- 1/invV
}''

# Hacemos el ajuste con la función jags.model
fit2<- jags.model(textConnection(model), data, inits=inits, n.chains=2)
update(fit2,1000) # actualiza la cadena asociada con el modelo
sample2<- coda.samples(fit2, parameters, n.iter=50000, thin=1)

# En esta parte hacemos los gráficos y obtenemos las características más
importantes de las muestras a posteriori
dev.new(height=4,width=4)
par(family="serif")
traceplot(sample2, col=c("firebrick1","dodgerblue2"),family="serif", cex=0.7,
cex.main=0.9)
densplot(sample2, col="firebrick3", family="serif", cex=0.7, show.obs = F)
summary(sample2)
HPDinterval(sample2) # intervalos de credibilidad

# Con el paquete coda revisamos la convergencia de las cadenas
gelman.diag(sample2,confidence = 0.95)
gelman.plot(sample2)
geweke.diag(sample2, frac1=0.1, frac2=0.5)
geweke.plot(sample2)
autocorr.plot(sample2)
autocorr.diag(sample2)
effectiveSize(sample2)

# Por último obtenemos el DIC para este modelo
dic3<-dic.samples(fit2, n.iter = 1000, type = "pD")
dic4<-dic.samples(fit2, n.iter = 1000, type = "popt")
```

Bibliografía

- [1] Bauer, I. d., Hemández Tejeda, T., and Manning, W. J. (1985). Ozone causes needle injury and tree decline in *pinus hartwegii* at high altitudes in the mountains around Mexico city. *Journal of the Air Pollution Control Association*, 35(8):838–840.
- [2] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.
- [3] Brockwell, P. J. and Davis, R. A. (2013). *Time series: theory and methods*. Springer Science & Business Media.
- [4] Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.
- [5] Castillo, E., Hadi, A. S., Balakrishnan, N., and Sarabia, J.-M. (2005). *Extreme value and related models with applications in engineering and science*. Wiley Hoboken, NJ.
- [6] Chatfield, C. (2016). *The analysis of time series: an introduction*. CRC press.
- [7] Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer.
- [8] de Salud, S. (2014). Norma Oficial Mexicana NOM-020-SSA1-2014. *Diario Oficial de la Federación*.
- [9] Diggle, P. J., Tawn, J., and Moyeed, R. (1998). *Model based geostatistics*, volume 47. Wiley.
- [10] Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press.
- [11] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. CRC press Boca Raton, FL.
- [12] Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472.

- [13] Gumbel, E. J. (2012). *Statistics of extremes*. Courier Corporation.
- [14] Hoel, P. G., Port, S. C., and Stone, C. J. (1986). *Introduction to stochastic processes*. Waveland Press.
- [15] Hoff, P. D. (2009). *A first course in Bayesian statistical methods*. Springer Science & Business Media.
- [16] Huerta, G. and Sansó, B. (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics*, 14(3):285–299.
- [17] Huerta, G., Sansó, B., and Stroud, J. R. (2004). A spatio-temporal model for Mexico city ozone levels. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(2):231–248.
- [18] Kendall, M. G. and Ord, J. K. (1990). *Time-series*, volume 296. Edward Arnold London.
- [19] Kotz, S. and Nadarajah, S. (2000). *Extreme value distributions: theory and applications*. World Scientific.
- [20] Kroese, D. P., Taimre, T., and Botev, Z. I. (2013). *Handbook of Monte Carlo Methods*, volume 706. Gobierno Federal.
- [21] Lee, P. M. (2012). *Bayesian statistics: an introduction*. John Wiley & Sons.
- [22] Lesaffre, E. and Lawson, A. B. (2012). *Bayesian biostatistics*. John Wiley & Sons.
- [23] Moritz, S. and Bartz-Beielstein, T. (2015). ImputeTS: Time series missing value imputation. *R package version 0.4*.
- [24] Mudelsee, M. (2013). *Climate time series analysis*. Springer.
- [25] Plummer, M. (2008). Penalized loss functions for bayesian model comparison. *Biostatistics*, 9(3):523–539.
- [26] Plummer, M. (2009). JAGS version 1.0. 3 manual. URL: http://www-ice.iarc.fr/~martyn/software/jags/jags_user_manual.pdf.
- [27] Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R news*, 6(1):7–11.
- [28] Press, S. J. (2009). *Subjective and objective Bayesian statistics: principles, models, and applications*, volume 590. John Wiley & Sons.
- [29] Reiss, R.-D., Thomas, M., and Reiss, R. (2007). *Statistical analysis of extreme values*, volume 2. Springer.
- [30] Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.

- [31] Robert, C. and Casella, G. (2010). *Introducing Monte Carlo Methods with R*. Springer Science & Business Media.
- [32] Rodrigues, E. R., Gamerman, D., Tarumoto, M. H., and Tzintzun, G. (2015). A non-homogeneous Poisson model with spatial anisotropy applied to ozone data from Mexico city. *Environmental and Ecological Statistics*, 22(2):393–422.
- [33] Rodríguez, S., Huerta, G., and Reyes, H. (2016). A study of trends for Mexico city ozone extremes: 2001-2014. *Atmósfera*, 29(2):107–120.
- [34] Ross, S. M. (1996). *Stochastic processes. 1996*. Wiley, New York.
- [35] Sahu, S. K., Yip, S., and Holland, D. M. (2011). A fast Bayesian method for updating and forecasting hourly ozone levels. *Environmental and Ecological Statistics*, 18(1):185–207.
- [36] Seinfeld, J. H. (1988). Ozone air quality models. *JAPCA*, 38(5):616–645.
- [37] Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1):67–90.
- [38] Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). WinBUGS user manual.
- [39] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.
- [40] Verzani, J. (2014). *Using R for introductory statistics*. CRC Press.
- [41] Wikle, C. K., Berliner, L. M., and Cressie, N. (1998). Hierarchical bayesian space-time models. *Environmental and Ecological Statistics*, 5(2):117–154.