



# COLEGIO DE POSTGRADUADOS

---

---

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN  
EN CIENCIAS AGRÍCOLAS

**CAMPUS MONTECILLO**

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA  
ESTADÍSTICA

**Estimación Bayesiana de Tendencia en los Niveles  
muy Altos de un Contaminante de Aire (Ozono)**

Gabriel Castillo Flores

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA  
OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO  
2013

La presente tesis titulada: **Estimación Bayesiana de Tendencia en los Niveles muy Altos de un Contaminante de Aire (Ozono)**, realizada por el alumno: **Gabriel Castillo Flores**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

## MAESTRO EN CIENCIAS

### SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA ESTADÍSTICA

#### CONSEJO PARTICULAR

CONSEJERO

  
Dr. Gilberto Rendón Sánchez

ASESOR

  
Dra. Elizabeth González Estrada

ASESOR

  
Dr. Barry C. Arnold

# Estimación Bayesiana de Tendencia en los Niveles muy Altos de un Contaminante de Aire (Ozono)

Gabriel Castillo Flores

Colegio de Postgraduados, 2013

En este estudio, se presenta un análisis del comportamiento espacio-temporal de los niveles altos de ozono urbano en la Zona Metropolitana del Valle México, usando los registros de la base de datos del Sistema de Monitoreo Atmosférico (SIMAT) de la Ciudad de México; en particular, la información es de 17 estaciones de Monitoreo durante el periodo comprendido del año 1986 al año 2012. El análisis utiliza los cuantiles de la Distribución Generalizada de Valores Extremos, Estimadores de Bayes, y el método de interpolación "Kriging.ºrdinario. En el presente estudio se muestra una propuesta de modelo estadístico basado en la función cuantil para investigar la tendencia de los niveles muy altos de ozono a lo largo del tiempo, y el uso de métodos geoestadísticos para predecir en espacio tiempo la condición de ozono en el Valle de México. Los resultados muestran una tendencia decreciente de los niveles muy altos de ozono en el Valle de México, siendo las zonas centro y suroeste las que presentan una mayor disminución de dichos niveles.

**Palabras clave:** Estadística Bayesiana; Eventos Extremos; Función Cuantil; Niveles muy Altos de Ozono.

# Bayesian Estimation of trend in high levels of air pollution (ozone)

Gabriel Castillo Flores

Colegio de Postgraduados, 2013

In this study, we present an analysis of the spatio-temporal behavior of high urban ozone levels in the Mexico City Metropolitan Area by using the Atmospheric Monitoring System (SIMAT) database of Mexico City; in particular, the information comes from 17 monitoring stations during the period of 1986 to 2012. The analysis uses the quantiles of the Generalized Extreme Value distribution, Bayes estimators, and the interpolation method known as ordinary "Kriging". In the present study we show a proposed statistical model based on quantile functions to investigate the trend of very high levels of ozone over time, and use geostatistical methods to predict the condition of ozone in the Valley of Mexico with regard to space and time. The results show a decreasing trend of very high ozone levels in the Valley of Mexico, with the greatest decrease being found in the central and southwestern areas.

**Key words:** Bayesian Statistics; Generalized Extreme Value Distribution (DGEV) Quantile Function; very high ozone levels.

## AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado para la realización de mis estudios de Maestría.

Al Colegio de Postgraduados, en particular al Programa de Estadística por haberme brindado la oportunidad de continuar con mi formación académica en sus aulas.

Al Dr. Humberto Vaquera Huerta mi más sincero agradecimiento por haberme brindado esta oportunidad, su apoyo incondicional, confianza, amabilidad y paciencia durante mi estancia en el Programa de Estadística.

A cada uno los integrantes de mi Consejo Particular:

Al Dr. Gilberto Rendón Sánchez, por el apoyo brindado durante mi estancia académica y por las oportunas observaciones realizadas para llevar a buen puerto el presente trabajo.

A la Dra. Elizabeth González Estrada, por las acertadas sugerencias, observaciones y comentarios realizados en este trabajo, y por su valioso apoyo para resolver cada una de mis dudas.

Al Dr. Barry C. Arnold, por el apoyo y facilidades brindadas para la culminación de este trabajo.

A cada uno de los profesores de quienes en las distintas áreas de estudio he recibido parte de su conocimiento para mi preparación.

A todos los integrantes del personal administrativo del ISEI por su amabilidad y apoyo que siempre me brindaron, en especial a Emmita, Mary y Toñita.

A todos mis compañeros y amigos que me acompañaron para alcanzar el mismo objetivo, por cada ayuda proporcionada y los ánimos brindados cuando alguna dificultad se presentaba, Annel, Esperanza, Amanda, Laura, Lucy, Enrique y Julio, a todos uds. mi más grande agradecimiento y admiración.

## **DEDICATORIA**

A mis padres

**Ma. Elena y Jesús**

Gracias por su esfuerzo, por su gran apoyo brindado para culminar otro ciclo más en mi preparación. Por su amor incondicional y por brindarme la mejor familia que jamás hubiera imaginado.

A mis hermanos

**Daniel y Perla**

Quiero expresarles que todos mis esfuerzos y logros han sido suyos también, porque me han acompañado en cada paso de mi vida.

Los amo mucho.

A mis tíos y primos

**Ana, Javier, Eduardo y Ricky**

Gracias por las infinitas muestras de cariño, siempre los tengo presentes.

A ti **Adrián**

Mi confidente y mejor amigo por tu apoyo en esta nueva etapa de mi vida, por los momentos vividos y los que están por venir.

# Índice

<b>1. Introducción</b>	<b>1</b>
<b>2. Objetivos</b>	<b>2</b>
<b>3. Marco teórico</b>	<b>3</b>
3.1. Teoría de Valores Extremos . . . . .	3
3.1.1. Aplicaciones . . . . .	5
3.2. Distribución de Valores Extremos Generalizada . . . . .	6
3.2.1. Función Cuantil . . . . .	9
<b>4. Estimación Bayesiana</b>	<b>11</b>
4.1. Aplicaciones de la Teoría de Valores Extremos basada en Estimación Bayesiana . . . . .	12
4.2. Enfoque Bayesiano . . . . .	12
4.3. Función a priori . . . . .	13
4.4. Función a posterior . . . . .	14
4.5. Función de Verosimilitud . . . . .	16
4.6. Estimador de Bayes . . . . .	17
4.7. Cadenas de Markov Monte Carlo (MCMC) . . . . .	18

# Índice

---

<b>5. Aplicación a Niveles Máximos de Ozono</b>	<b>20</b>
5.1. El ozono y su importancia . . . . .	20
5.2. El problema de contaminación por ozono . . . . .	20
5.3. Metodología . . . . .	22
5.4. Aplicación al análisis de Datos del Valle de México . . . . .	22
5.4.1. Obtención de los máximos diarios y máximos anuales . . . . .	23
5.4.2. Modelo Propuesto . . . . .	24
5.4.3. Obtención de los Estimadores de Bayes . . . . .	25
5.4.4. Obtención de los Cuantiles . . . . .	28
<b>6. Resultados y Conclusiones</b>	<b>38</b>
6.1. Resultados . . . . .	38
6.2. Conclusiones y Recomendaciones . . . . .	39
<b>7. Apéndice</b>	<b>40</b>
Apéndice A . . . . .	40
Apéndice B . . . . .	45
Apéndice C . . . . .	46
<b>Referencias</b>	<b>47</b>



# Capítulo 1

## Introducción

Durante las últimas décadas la Teoría de Valores Extremos ha emergido como una de las principales áreas de la estadística que apoya la generación de conocimiento científico en temas relacionados con el cambio climático global (Katz et al. (2002), Scotto et al. (2011)), inundaciones (Coles (2003)), estudio de contaminantes, economía, finanzas (Bali (2007)), entre otros. Además de que la inferencia sobre el comportamiento estocástico de eventos extremos (máximos y mínimos) es de creciente interés.

En el presente trabajo se hace uso de la función cuantil para la Distribución Generalizada de Valores Extremos (DGVE) con el propósito de estimar la tendencia en los niveles máximos de ozono en la Ciudad de México y su Zona Metropolitana; para lo cual se emplea Estimación Bayesiana aplicada en la Teoría de Valores Extremos. Esto debido a que con base en estudios recientemente realizados como los de Papadakis and Tsonas (2012) y Rodríguez R. et al. (2012), se ha demostrado que el uso del marco inferencial Bayesiano reporta resultados más satisfactorios para el análisis de datos, sobre todo en aquellos casos en los que se hace uso de datos extremos; aunado al hecho de que los estimadores de Bayes, correspondientes a una función de densidad a priori arbitraria, satisfacen varias de las propiedades deseables de cualquier estimador, tales como: la de ser función de la estadística suficiente mínima, así como ser Consistentes y Mejor Asintóticamente Normales, (Mood et al. (1974)).

La metodología desarrollada sugiere la estimación de los cuantiles de la DGVE por medio de los Estimadores de Bayes para estimar la tendencia de los niveles de ozono muy altos, puesto que estos afectan de manera importante la salud de plantas y habitantes de grandes zonas urbanas, como la Ciudad de México y su Zona Metropolitana. De donde deriva la importancia de monitorear la tendencia que presenta este contaminante a lo largo del tiempo.

# Capítulo 2

## Objetivos

### General

Investigar la tendencia en los niveles altos de ozono en el Valle de México utilizando Estimación Bayesiana y Teoría de Valores Extremos.

### Particulares

1. Generar estimadores bayesianos en la Distribución de Valores Extremos Generalizada (DGEV) usando distribuciones no informativas.
2. Aplicar los resultados al caso de niveles máximos de ozono en el Valle de México.
3. Generar mapas de predicción es espacio tiempo.

# Capítulo 3

## Marco teórico

### 3.1. Teoría de Valores Extremos

Cuando se habla de Valores Extremos se está haciendo referencia a aquellos eventos u observaciones que se presentan muy rara vez, por lo que su probabilidad de ocurrencia es muy baja (por ejemplo: un registro de temperatura alta en plena época invernal, un día completo de lluvia, presión baja en un paciente sano, una severa tormenta de nieve, etc.). La importancia de realizar investigaciones que involucren dichos valores radica en el hecho de que cuando se presentan, generalmente vienen acompañados de situaciones de alto impacto que suelen alterar de manera importante el entorno donde ocurren.

En términos estadísticos, los Valores Extremos se pueden encontrar al realizar cualquier análisis de información y se conocen como observaciones aberrantes, las cuales generalmente suelen ser eliminadas u omitidas. Y es debido a que son poco comunes el principal problema al que se han enfrentado los expertos en estadística para diseñar métodos, técnicas y herramientas que permitan cuantificar la probabilidad, la frecuencia y la intensidad con la que estos fenómenos suelen presentarse.

Es complicado definir con precisión el punto en que la estadística acerca de Valores Extremos comenzó a desarrollarse; sin embargo, se tiene

### 3.1. Teoría de Valores Extremos

---

conocimiento de que en 1709 Nicolas Bernoulli planteó el problema de la distancia media máxima desde el origen de "n" puntos distribuidos aleatoriamente en una línea recta de distancia fija t. Posteriormente, [Chaplin \(1880\)](#) planteó el problema del efecto del tamaño en la resistencia de materiales, es decir, un problema de mínimos. No fue si no hasta 1920 cuando distintos investigadores comenzaron a desarrollar la Teoría de Valores Extremos al mismo tiempo, de los cuales ? desarrollaron la forma límite de la distribución del valor máximo o mínimo de una muestra aleatoria.

Desde entonces la Teoría del Valores Extremos (TVE) se ha desarrollado rápidamente pero es en las últimas dos décadas que ha tenido un crecimiento importante tanto desde el punto de vista metodológico como del de las aplicaciones. La mayoría de los estudios estadísticos tratan de la modelación del promedio de la distribución de la variable de interés, dicho promedio se estima a partir de la media muestral. En la TVE el interés principal no está en el promedio, sino en los valores más bajos o más altos de la variable bajo estudio, es decir, el interés está en los eventos asociados a la cola de la distribución ([Velasco Luna \(2007\)](#)). Ejemplo de ello puede encontrarse en investigaciones sobre Climatología ([Etienne et al. \(2010\)](#), [Jagger \(2006\)](#), [Coles \(2003\)](#)), Ecología ([Katz et al. \(2005\)](#)), Oceanografía ([Mestre \(2009\)](#), [Macafee \(2007\)](#)), Finanzas ([Bali \(2007\)](#)), Estadística Ambiental ([Rodríguez R. et al. \(2012\)](#)), entre otras. En donde se estudia el comportamiento a lo largo del tiempo de los valores mínimos o máximos del fenómeno en cuestión, los cuales pueden llegar a provocar o han provocado situaciones catastróficas.

Como bien se puede observar, la TVE trata de evaluar a partir de una muestra ordenada de una determinada variable aleatoria, la probabilidad de ocurrencia de eventos que son más extremos que cualquier otro evento observado anteriormente.

En su versión más sencilla el problema es el siguiente: dada una muestra independiente e idénticamente distribuida  $x_1, x_2, \dots, x_n$  de una distribución desconocida F, se requiere hacer inferencias precisas en la cola de F ([Coles \(2001\)](#)). Lo anterior puede sonar fácil de llevar a cabo; sin

### 3.1. Teoría de Valores Extremos

---

embargo, los problemas a los que se enfrenta el investigador son:

- Hay muy pocas observaciones en la cola de la distribución.
- Es común que se desee estimar valores que van más allá del máximo valor de la muestra.
- Las técnicas estándar de estimación de densidades pueden tener sesgos importantes al estimar las colas, aunque se ajusten bien donde los datos tienen mayor densidad.

Por lo tanto, la TVE proporciona técnicas simples para estimar las probabilidades de futuros niveles extremos de determinado fenómeno dados sus datos históricos. Existen muchas variantes sobre este tema, pero el más simple es cuando los datos históricos consisten de una secuencia de observaciones máximas anuales (Coles (2003)).

De esta manera, los valores máximos y mínimos son denominados los valores extremos de las  $n$  variables aleatorias independientes e idénticamente distribuidas (i.i.d)  $X_1, X_2, \dots, X_n$  (Landaverde Vaca (2005)). Los cuales pueden modelarse por medio de dos enfoques :

- Distribución de Valores Extremos Generalizada.
- Análisis de excedentes sobre umbrales.

#### 3.1.1. Aplicaciones

Como ya se ha mencionado anteriormente, el análisis de Valores Extremos es ampliamente utilizado en muchas disciplinas, que van desde la ingeniería estructural, finanzas, ciencias de la tierra, la predicción de tráfico, ingeniería geológica, entre otras. Sin embargo, es en las ciencias ambientales en donde recientemente ha tenido mayor auge debido a que muchos fenómenos meteorológicos involucran valores extremos (Gumbel and Lieblein (1954)), por ejemplo:

### 3.2. Distribución de Valores Extremos Generalizada

---

- Presión atmosférica
- Temperaturas
- Precipitaciones
- Nevadas
- Rachas de viento
- Inundaciones
- Sequías y demás.

Recientemente, un área de aplicación que está siendo de creciente interés principalmente para las grandes urbes del mundo, es en la detección de niveles máximos de ozono urbano ([Smith \(1989\)](#), [Coles \(1996b\)](#)). Expertos en medio ambiente y autoridades tienen especial interés en estudiar el comportamiento del ozono troposférico debido al impacto que tiene en la salud humana, en el deterioro de materiales y en el daño a los ecosistemas terrestres([Huerta et al. \(2004\)](#)).

### 3.2. Distribución de Valores Extremos Generalizada

Las distribuciones de valores extremos surgen naturalmente como las distribuciones de máximos o mínimos de un gran número de variables aleatorias independientes e idénticamente distribuidas ([Gumbel \(1935, 1958\)](#)).

Esto es, sea  $X_1, X_2, \dots, X_n$  una muestra de variables aleatorias independientes con función de distribución acumulada común  $F(x)$ , es decir,  $F(x) = P(X_i \leq x)$ . Y sean  $X_{1n} = \min(X_1, X_2, \dots, X_n)$  y  $X_{nn} = \max(X_1, X_2, \dots, X_n)$ , los extremos definidos como el mínimo y el máximo de las  $n$  variables aleatorias, respectivamente, con función de distribución:

### 3.2. Distribución de Valores Extremos Generalizada

---

$$F_{X_{1n}}(x) = P(X_{1n} \leq x) = 1 - P(X_{1n} > x) = 1 - P(X_1 > x_1, \dots, X_n > x_n)$$

$$F_{X_{1n}}(x) = 1 - [1 - F(x)]^n$$

con  $i \leq n$ ,

$$F_{X_{nn}}(x) = P(X_{nn} \leq x) = P(X_1 \leq x_1, \dots, X_n \leq x_n) = [F(x)]^n = F^n(x)$$

donde  $F$  y  $F_{X_{nn}}$  suelen ser desconocidas en la práctica.

Ahora, siguiendo el Teorema de Fisher & Tippett el cual sugiere a la DGVE como la única distribución límite para máximos ([Landaverde Vaca \(2005\)](#)):

#### **Teorema de Fisher & Tippett**([McNeil \(1997\)](#))

Suponga que se tiene una secuencia de variables aleatorias independientes e idénticamente distribuidas de una distribución desconocida  $F$ . Si el máximo de las primeras  $n$  observaciones se denota por  $M_n = \max(X_1, X_2, \dots, X_n)$ , y suponga además que es posible encontrar secuencias de números reales  $a_n > 0$  y  $b_n$  tales que  $(M_n - b_n)/a_n$ , la secuencia de máximos normalizados converge a una distribución límite  $G$  cuando  $n$  se incrementa. Esto es,

$$P\{(M_n - b_n)/a_n \leq x\} = F^n(a_n x + b_n) \longrightarrow G(x),$$

cuando  $n \longrightarrow \infty$ , para una distribución límite no degenerada  $G(x)$ .

Y si la condición se mantiene, entonces  $F$  está en el dominio máximo de atracción de  $G$ , y entonces  $G$  puede ser una de las tres familias de funciones de distribución de Valores Extremos:

### 3.2. Distribución de Valores Extremos Generalizada

---

- Frechét  $(G_1) : \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0, \alpha > 0 \end{cases}$
- Weibull  $(G_2) : \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & x \leq 0, \alpha > 0 \\ 1, & x > 0 \end{cases}$
- Gumbel  $(G_3) : \Lambda(x) = \exp\{-\exp(-x)\}, x \in \mathbb{R}$

De esta manera, siguiendo el siguiente Teorema:

**Teorema** (Barry et al. (1992))

Se pueden escoger  $a_n, b_n > 0$  de modo que  $F^n(a_n x + b_n) \longrightarrow G(x)$ , tal que:

1.  $b_n = 0, a_n = F^{-1}(1 - n^{-1})$ , si  $G = G_1$
2.  $b_n = F^{-1}, a_n = F^{-1}(1) - F^{-1}(1 - n^{-1})$ , si  $G = G_2$
3.  $b_n = F^{-1}(1 - n^{-1}), a_n = E(X - b_n | X > b_n)$  ó  $F^{-1}(1 - (ne)^{-1}) - F^{-1}(1 - n^{-1})$  si  $G = G_3$

Así el modelado estadístico de eventos extremos está basado en tres tipos de distribuciones límite (Davison et al. (2012)), las cuales, juegan un importante papel en el análisis de datos ambientales extremos.

De acuerdo con Velasco Luna (2007), un enfoque para modelar dichos eventos es a partir de la Distribución de Valores Extremos Generalizada (DGVE), la cual es una combinación de las tres familias anteriores en una sola distribución con parametrización común (Von Mises (1936), Jenkinson (1955)), realizando la reparametrización  $\xi = 1/\alpha$ ; así, la representación es conocida como la Distribución Generalizada de Valores Extremos (DGEV) y se ajusta a los valores máximos o mínimos de los datos:

$$G(x; \mu, \sigma, \xi) = \begin{cases} \exp\{-[1 + \xi(\frac{x-\mu}{\sigma})]^{-\frac{1}{\xi}}\}, & \text{si } 1 + \xi(\frac{x-\mu}{\sigma}) > 0 \\ \exp(-\exp^{-\frac{x-\mu}{\sigma}}), & \text{si } \xi = 0 \end{cases} \quad (3.1)$$



## 3.2. Distribución de Valores Extremos Generalizada

---

donde  $\sigma > 0$ ,  $-\infty < \mu < \infty$ ,  $1 + \xi(x - \mu)/\sigma > 0$ ,  $x_+ = \max\{x, 0\}$ .

Con función de densidad:

$$g(x; \mu, \sigma, \xi) = \begin{cases} \frac{1}{\sigma} [1 + \xi(\frac{x-\mu}{\sigma})]^{-\frac{1}{\xi}-1} \exp\{-[1 + \xi(\frac{x-\mu}{\sigma})]^{-\frac{1}{\xi}}\}, & 1 + \xi(\frac{x-\mu}{\sigma}) > 0 \\ \exp(-\exp^{-\frac{x-\mu}{\sigma}}) \exp^{-\frac{x-\mu}{\sigma}} \frac{1}{\sigma}, & \xi = 0 \end{cases} \quad (3.2)$$

El parámetro  $\mu$  es el parámetro de localización,  $\sigma$  es el parámetro de escala y  $\xi$  es el parámetro de forma. Los casos en los cuales  $\xi = 0$ ,  $\xi > 0$  y  $\xi < 0$ , corresponden a las Distribuciones Gumbel, Fréchet y Weibull, respectivamente. Así, la forma de la DGVE cambia con base en el valor de  $\xi$  ([Sexto Monroy et al. \(2010\)](#)):

- Gumbel: distribución de cola moderadamente pesada cuando  $\xi = 0$
- Fréchet: distribución de cola pesada cuando  $\xi > 0$
- Weibull: distribución de cola ligera cuando  $\xi < 0$

En las distribuciones de tipo Gumbel,  $\xi = 0$  se interpreta como  $\xi \rightarrow 0$ . Este tipo de distribuciones tienen una cola superior no acotada que disminuye de manera relativamente rápida (ejemplo de ello son la normal, exponencial, gamma y log-normal). Las distribuciones asociadas a las tipo Fréchet no tienen límite superior y disminuyen de manera lenta; tal que sus momentos son infinitos para todas las ordenes mas grandes que  $1/\xi$ . Ejemplo de este tipo de distribuciones son la Pareto, Cauchy, t-student y las mixtas. Por último las distribuciones tipo Weibull tienen límite superior en  $x = \mu - \frac{\sigma}{\xi}$ .

### 3.2.1. Función Cuantil

Es muy común que al trabajar con Valores Extremos nos enfoquemos en los cuantiles extremos de la DGVE; específicamente, en el nivel de

### 3.2. Distribución de Valores Extremos Generalizada

---

retorno  $(1-p)$  ésimo cuantil de la DGVE, asociado al periodo de retorno  $\frac{1}{p}$  (Monroy (2010)).

Así, dado un valor  $0 < \alpha = 1 - p < 1$ , se define el  $\alpha$ -cuantil de la función de distribución  $F_X(x; \mu)$  como el valor  $q_\alpha$  tal que  $\alpha = P(Z \leq q_\alpha) = G_Z(q_\alpha) \Leftrightarrow G_Z^{-1}(\alpha; \theta) = Z_\alpha(p) = q_\alpha$ .

A  $q_\alpha$  se le llama Función Cuantil de  $G_z(x; \mu)$ . Para valores muy grandes de  $q_\alpha$ , se considera que hay un  $(1 - \alpha)$  % de observaciones más grandes.

En el caso de la distribución GVE, es posible obtener una expresión para los cuantiles invirtiendo la función; así, con  $\alpha$  fijo (Coles (2001)) se tiene:

$$Z_\alpha(p) = \begin{cases} \mu - \frac{\sigma}{\xi} \{1 - (y_p)^{-\xi}\}, & \xi \neq 0 \\ \mu - \sigma \log\{y_p\}, & \xi = 0 \end{cases} \quad (3.3)$$

donde  $y_p = -\log(1 - p), 0 \leq p \leq 1$

De esta manera, a  $z_p$  se le conoce como el nivel de retorno asociado al periodo de retorno  $\frac{1}{p}$ . Note que para el cálculo de los cuantiles se emplean los estimadores de los parámetros. Y en general, es común esperar que el nivel  $z_p$  sea excedido en promedio por algún máximo anual con probabilidad  $p$  (Reyes Cervantes (2008)).

# Capítulo 4

## Estimación Bayesiana

Un enfoque alternativo cuando se trata de valores extremos es considerar un enfoque bayesiano. En términos generales, el teorema de Bayes convierte una distribución a priori, sea  $\pi(\theta)$ , para un vector de parámetros  $\theta := (\theta_1, \dots, \theta_p)$ , de acuerdo con la disponibilidad de datos históricos  $x := (x_1, \dots, x_n)$ , en una distribución a posteriori  $\pi(\theta | x) \propto L(x | \theta)\pi(\theta)$ , donde  $L(x | \theta)$  denota la verosimilitud de los datos observados. La distribución a priori expresa el grado de certeza acerca de la situación antes de que los datos han sido tomados. Por lo tanto, el teorema de Bayes proporciona los medios para actualizar nuestro conocimiento sobre el fenómeno en estudio, expresado en términos de una función de densidad de probabilidad. Es importante hacer hincapié en el hecho de que el resultado de un análisis Bayesiano es una distribución completa en  $\theta$ , lo que representa una ventaja considerable respecto a los métodos clásicos, en lugar de sólo una estimación puntual, se obtiene una distribución probabilística completa de los parámetros. Las estimaciones puntuales pueden obtenerse fácilmente tomando la media o la mediana de la distribución posterior (Scotto et al. (2011)).

Otra razón por la cual es conveniente llevar a cabo un análisis bayesiano de valores extremos es debido a la escasez de datos y la facilidad con la que se pueden incluir otras fuentes de información por medio de la distribución a priori (Coles (2001)).

## 4.1. Aplicaciones de la Teoría de Valores Extremos basada en Estimación Bayesiana

Muchas han sido las aplicaciones que se han planteado en diversos estudios relacionados con la Teoría de Valores Extremos; sin embargo, es el Enfoque Bayesiano quien durante las últimas décadas ha mostrado creciente interés entre diversos investigadores debido tal vez a las ventajas que este enfoque proporciona. Ejemplo de ello es el trabajo realizado por [Papadakis and Tsionas \(2012\)](#), en el cual se emplea la función Gumbel y una función a priori informativa para obtener la función posterior y así proponer una nueva metodología que permita llevar a cabo un modelo de regresión utilizando los datos de los valores de las acciones al cierre de DELL.

Por otro lado [Rodríguez R. et al. \(2012\)](#) emplea la DGVE para ingresar en el parámetro de escala una función de regresión lineal con liga exponencial, con el propósito de verificar la tendencia decreciente del ozono en el tiempo; empleando Estimación Bayesiana para obtener los estimadores de cada uno de los parámetros.

De esta manera, se puede observar como la Estimación Bayesiana combinada con la Teoría de Valores Extremos tienen diversas aplicaciones en distintos campos que van desde lo financiero y económico hasta diversos fenómenos ambientales.

## 4.2. Enfoque Bayesiano

En el enfoque clásico el parámetro  $\theta$ , se piensa que es desconocido, pero fijo. Una muestra aleatoria  $X_1, \dots, X_n$  se extrae de una población indexada por  $\theta$ , y basados en los valores observados en la muestra, el valor de  $\theta$  es estimado. En el enfoque Bayesiano  $\theta$  se considera que es una cantidad cuya variación puede ser descrita por una distribución de probabilidad (llamada la distribución a priori). Esta es una distribución subjetiva, basada en la experiencia del investigador, y se formula antes

### 4.3. Función a priori

---

de que los datos sean vistos (de ahí el nombre de distribución a priori). Una muestra es entonces tomada de una población indexada por  $\theta$  y la distribución a priori es actualizada con esta información de la muestra. La a priori actualizada es llamada la distribución posterior. Esta actualización es hecha con el uso de la Regla o Teorema de Bayes, de ahí el nombre de Estadística Bayesiana (Casella and Berger (2002)).

#### Teorema de Bayes

Sea  $A_1, A_2, \dots$  una partición del espacio muestral y sea B cualquier conjunto. Entonces, para cada  $i = 1, 2, \dots$ ,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)} \quad (4.1)$$

### 4.3. Función a priori

Considere un problema de Estadística Inferencial en el cual las observaciones son tomadas de una distribución para la cual la función de densidad es  $f(x|\theta)$ , donde  $\theta$  es un parámetro que tiene un valor desconocido. Se supone que el valor desconocido de  $\theta$  debe estar en un espacio paramétrico específico  $\Omega$ . El problema entonces es determinar en que parte del espacio de parámetros  $\Omega$  se encuentra el valor de  $\theta$  basados en las observaciones de la función de densidad. Así en muchos problemas, antes de que cualquier observación de  $f(x|\theta)$  esté disponible, el investigador o estadístico debe ser capaz de resumir la información previa disponible y su conocimiento para saber en que regiones se pudiera encontrar el valor del parámetro  $\theta$  dentro del espacio paramétrico  $\Omega$  construyendo una distribución de probabilidad para  $\theta$  en el conjunto  $\Omega$ . Esta distribución es nombrada la distribución a priori de  $\theta$  porque representa la probabilidad relativa de que el verdadero valor de  $\theta$  se encuentre en varias regiones de  $\Omega$  previo a la observación de cualquier valor de  $f(x|\theta)$ .

#### 4.4. Función a posterior

---

De esta manera, la función a priori es un punto clave en el uso del análisis bayesiano ya que en ella se encuentra toda la información previa y actualizada que se pudiera conocer del fenómeno en estudio; sin embargo, uno de los problemas con los que el investigador se enfrenta es el de elegir una adecuada función a priori para los parámetros que se suponen desconocidos debido a que no siempre se cuenta con el apoyo de un experto en el tema ([Landaverde Vaca \(2005\)](#)).

Así, al emplear una distribución a priori no informativa, la información resultante en la función posterior  $\pi(\theta|x)$  surge de los datos, y en consecuencia la inferencia resultante será completamente objetiva ([Carlin and Louis \(2000\)](#)).

#### 4.4. Función a posterior

Suponga ahora que las  $n$  variables aleatorias  $X_1, \dots, X_n$  forman una muestra aleatoria de una distribución para la cual su función de densidad es  $f(x|\theta)$ . Suponga además que el valor del parámetro  $\theta$  es desconocido y que la función de probabilidad a priori es  $\pi(\theta)$ . Para simplificar, se asume que el espacio paramétrico  $\Omega$  es, o bien, un intervalo de la recta real o toda la recta real; que  $\pi(\theta)$  es una función de densidad de probabilidad a priori en  $\Omega$ ; y que  $f(x|\theta)$  es una función de densidad de probabilidad.

Entonces, las variables aleatorias  $X_1, \dots, X_n$  forman una muestra aleatoria de la distribución para la cual  $f(x|\theta)$  es su función de densidad de probabilidad; su función de probabilidad conjunta  $f_n(x_1, \dots, x_n|\theta)$  es dada por la ecuación:

$$f_n(x_1, \dots, x_n|\theta) = f(x_1|\theta) \dots f(x_n|\theta) \quad (4.2)$$

Si se usa notación vectorial  $x = (x_1, \dots, x_n)$ ; entonces, la función de probabilidad conjunta en (4.2) puede ser escrita simplemente como  $f_n(x|\theta)$ .

#### 4.4. Función a posterior

---

Dado que el propio parámetro  $\theta$  es ahora considerado que tiene una distribución para la cual su función de probabilidad es  $\pi(\theta)$ ; la función de probabilidad conjunta  $f_n(x|\theta)$  debe ser considerada propiamente como la función condicional conjunta de  $X_1, \dots, X_n$  para un valor dado de  $\theta$ . Si se multiplica esta función condicional conjunta por la función de probabilidad  $\pi(\theta)$ , se obtiene la  $(n + 1)$ -dimensional función de probabilidad conjunta de  $X_1, \dots, X_n$  y  $\theta$  en la forma  $f_n(x|\theta)\pi(\theta)$ . La función marginal conjunta de  $X_1, \dots, X_n$  puede ser obtenida integrando esta función de probabilidad conjunta sobre todos los valores de  $\theta$ . Por lo tanto, la  $n$ -dimensional función marginal de probabilidad conjunta  $g_n(x)$  de  $X_1, \dots, X_n$  puede ser escrita de la siguiente forma (DeGroot (1975)):

$$g_n(x) = \int_{\Omega} f_n(x|\theta)\pi(\theta)d\theta \quad (4.3)$$

De acuerdo con Casella and Berger (2002) si se denota la distribución a priori por  $\pi(\theta)$  y la distribución de la muestra por  $f(x|\theta)$ , entonces la distribución a posterior es la distribución condicional de  $\theta$  dada la muestra  $x$ :

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)}, \quad f(x|\theta)\pi(\theta) = f(x, \theta) \quad (4.4)$$

donde  $m(x)$  es la distribución marginal de  $X$ , esto es,

$$m(x) = \int f(x|\theta)\pi(\theta)d\theta$$

Note que la distribución a posterior es una distribución condicional, condicionada a la observación de la muestra. La distribución posterior es ahora usada para hacer declaraciones acerca de  $\theta$ , la cual aún es considerada una cantidad aleatoria. Por ejemplo, la media de la distribución posterior puede ser usada como una estimación puntual de  $\theta$ .

Mood et al. (1974) menciona que la distribución posterior reemplaza

## 4.5. Función de Verosimilitud

---

la función de verosimilitud como una expresión que incorpora toda la información. Sin embargo, a diferencia de la función de verosimilitud (como una función de  $\theta$ ), la distribución posterior es una función de distribución, de modo que se podrá igualmente estimar  $\theta$  con la mediana o la media de la distribución posterior. Así entonces, se usará la media de la distribución posterior como nuestro estimador de  $\theta$ .

## 4.5. Función de Verosimilitud

Note que el denominador en la parte derecha de la ecuación (4.4) es simplemente la integral del numerador sobre todos los valores posibles de  $\theta$ . Aunque el valor de esta integral depende de los valores de  $x_1, \dots, x_n$ , no depende de  $\theta$  y puede ser tratada como una constante. Por lo tanto, la ecuación (4.4) puede reemplazarse por la siguiente relación:

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \quad (4.5)$$

El símbolo de proporcionalidad  $\propto$  es usado para indicar que el lado izquierdo de la ecuación (4.5) es igual al lado derecho excepto posiblemente por un factor constante, cuyo valor puede depender de los valores observados  $x_1, \dots, x_n$  pero no de  $\theta$ . El factor constante apropiado para establecer la igualdad de los dos lados en la relación (4.5) puede ser determinado en cualquier momento basados en el hecho de que  $\int_{\Omega} \pi(\theta|x)d\theta = 1$ , porque  $\pi(\theta|x)$  es una función de densidad de probabilidad de  $\theta$ .

Cuando la función de probabilidad conjunta de  $f(x|\theta)$  de las observaciones en una muestra aleatoria es considerada como una función de  $\theta$  para los valores dados de  $x_1, \dots, x_n$ , se le denomina función de verosimilitud. En esta terminología, la ecuación (4.5) establece que la función posterior de  $\theta$  es proporcional al producto de la función de verosimilitud por la función a priori de  $\theta$  (DeGroot (1975)).

Así, utilizando la proporcionalidad expresada en (4.5), es a menudo



posible determinar la función posterior de  $\theta$ .

## 4.6. Estimador de Bayes

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la densidad  $f(x|\theta)$ , donde  $\theta$  es un valor de la variable aleatoria  $\Theta$  con función conocida  $\pi(\theta)$ . El estimador posterior de Bayes de  $\tau(\theta)$  con respecto a la a priori  $\pi(\theta)$  según [Mood et al. \(1974\)](#) es definido por (4.6):

$$E[\tau(\Theta)|X_1, \dots, X_n] \tag{4.6}$$

Esto es:

$$\begin{aligned} E[\tau(\Theta)|X_1 = x_1, \dots, X_n = x_n] &= \int \tau(\theta) f_{\Theta|X_1=x_1, \dots, X_n=x_n}(\theta|x_1, \dots, x_n) d\theta \\ &= \frac{\int \tau(\theta) [\prod_{i=1}^n f(x_i|\theta)] g_{\Theta}(\theta) d(\theta)}{[\prod_{i=1}^n f(x_i|\theta)] g_{\Theta}(\theta) d(\theta)} \end{aligned}$$

Ahora bien, como ya se ha visto el Análisis Bayesiano basa sus inferencias en la densidad posterior; sin embargo, tal como lo menciona [Landa-verde Vaca \(2005\)](#) existen ciertas complicaciones al realizar el cálculo del estimador de Bayes. Estas son: que la densidad posterior por lo general no esta disponible en forma cerrada debido a que la densidad predictiva a priori tampoco tiene una forma analítica cerrada; otra es que en muchos casos el cálculo del  $E[\tau(\Theta)|X_1 = x_1, \dots, X_n = x_n]$  no se puede hacer analíticamente.

Derivado de lo anterior, actualmente se han desarrollado diversas rutinas computacionales como las Cadenas de Markov Monte Carlo (MCMC por sus siglas en inglés) y aproximaciones numéricas que facilitan y sir-

ven de gran apoyo para el Análisis Bayesiano.

## 4.7. Cadenas de Markov Monte Carlo (MCMC)

La complicación técnica en un análisis Bayesiano es la evaluación del factor de proporcionalidad en la ecuación (4.5), ya que esto requiere integrar en el espacio del parámetro  $\theta$ . Y aunque este problema puede ser abordado desde distintas perspectivas, son los avances computacionales; en especial la metodología de las MCMC, los que permiten llevar a cabo la inferencia estadística de una forma más sencilla (Coles and Tawn (2005)).

Con base en investigaciones anteriores se tiene conocimiento de que el primer análisis Bayesiano llevado a cabo en los Valores Extremos fue hecho por Smith (1987). Tiempo después Coles (1996a) y Coles (1996c); mostraron como el algoritmo de las MCMC puede ser aplicado para llevar a cabo Inferencia Bayesiana en el modelado de Valores Extremos.

De acuerdo con DeGroot (1975), una cadena de Markov es un tipo especial de proceso estocástico, el cual puede describirse como sigue: En cualquier momento dado  $n$ , cuando el estado actual  $X_n$  y todos los estados anteriores  $X_1, \dots, X_{n-1}$  del proceso son conocidos, las probabilidades de todos los futuros estados  $X_j (j > n)$  dependen solamente del estado actual  $X_n$  y no dependen de los estados anteriores  $X_1, \dots, X_{n-1}$ . Formalmente, una cadena de Markov es un proceso estocástico tal que para  $n = 1, 2, \dots$  y para cualquier secuencia posible de estados  $x_1, x_2, \dots, x_{n-1}$ ,

$$Pr(X_{n+1} = x_{n+1} | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = Pr(X_{n+1} = x_{n+1} | X_n = x_n)$$

Note que, las probabilidades en una cadena de Markov deben satisfacer la siguiente relación:

## 4.7. Cadenas de Markov Monte Carlo (MCMC)

---

$$Pr(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) =$$

$$Pr(X_1 = x_1)Pr(X_2 = x_2|X_1 = x_1)Pr(X_3 = x_3|X_2 = x_2)\dots$$

$$\dots Pr(X_n = x_n|X_{n-1} = x_{n-1})$$

La salida del algoritmo de una MCMC, es una serie simulada  $\theta_1, \dots, \theta_m$ ; la cual, aunque no es independiente, es aproximadamente estacionaria con distribución marginal igual a la distribución posterior objetivo  $\pi(\theta|x)$ .

# Capítulo 5

## Aplicación a Niveles Máximos de Ozono

### 5.1. El ozono y su importancia

El ozono troposférico es un contaminante de aire secundario originado por procesos químicos atmosféricos y altas temperaturas; los cuales actúan sobre contaminantes primarios (gases emitidos) y por otros gases de la atmósfera (oxígeno). Niveles altos de ozono afectan de manera importante la salud de plantas y habitantes de grandes zonas urbanas, como la Ciudad de México y su zona metropolitana, Monterrey y Guadalajara. De ahí la importancia de estudiar dichos niveles y de esta manera entender y mejorar la calidad del aire en la principales zonas urbanas ([Huerta et al. \(2004\)](#)).

### 5.2. El problema de contaminación por ozono

Se tiene conocimiento que antes de la mitad del siglo pasado y con base en estudios de monitoreo, la contaminación por ozono en la Ciudad de México no representaba un problema grave ([Bravo et al. \(1992\)](#)). Sin embargo, esta situación cambió y recientemente la Zona Metropolitana de la Ciudad de México es considerada como la metrópoli más conta-

## 5.2. El problema de contaminación por ozono

---

minada del país, y quizás en el mundo ([Anders \(1988\)](#), [Torres \(1992\)](#), [Bravo et al. \(1992\)](#)). Se estima que más de 9.5 millones de habitantes están expuestos a diferentes grados de contaminación por ozono, que frecuentemente se encuentra por arriba de la NOM-020-SSA1-1993 (0.11 ppm promedio máximo en una hora)([García \(2004\)](#)).

En 1986, las autoridades municipales reconocieron la magnitud del problema e instalaron una red de estaciones de monitoreo para medir el ozono, monóxido de carbono e hidrocarburos. La red se denomina Red Automática de Monitoreo Ambiental (RAMA) de la Ciudad de México y es parte del Sistema de Monitoreo Atmosférico de la Secretaría del Medio Ambiente del Gobierno del Distrito Federal; este sistema es empleado para monitorear la calidad del aire de la Zona Metropolitana del Valle de México. En la actualidad la RAMA cuenta con 34 estaciones de monitoreo ubicadas en puntos estratégicos de la Ciudad de México, 21 están localizadas en el Distrito Federal y 13 en el Estado de México. Los criterios técnicos para la ubicación de las estaciones son: densidad de población, distribución de fuentes de emisión, meteorología y topografía. Estas estaciones operaran los 365 días del año, con periodos cortos de interrupciones para la calibración de los instrumentos de medición. Cada estación toma medidas de contaminantes de forma automática, segundo a segundo, y los correspondientes promedios por hora son reportadas al público. En general, las unidades de las mediciones son partes por millón (ppm), es decir, la concentración de la sustancia en un volumen, en donde el volumen es dividido entre 1 millón de partes ([Huerta et al. \(2004\)](#)); esta unidad de medición es utilizada para conocer concentraciones diminutas de elementos presentes por unidad de volumen.

La red RAMA es un sistema dinámico, lo que significa que se va adaptando a las circunstancias de calidad del aire, que se presentan con el tiempo en la Zona Metropolitana del Valle de México, es por ello que algunas estaciones se dan de baja o se incorporan nuevas.

## 5.3. Metodología

1. Seleccionar las Estaciones de Monitoreo que cuenten con mayor cantidad de registros disponibles.
2. Obtener los máximos por día.
3. Posteriormente obtener los máximos anuales.
4. Aplicar la Teoría de Valores Extremos a los datos; para lo cual, se supone que los máximos anuales siguen una Distribución Generalizada de Valores Extremos (DGVE).
5. Emplear Estimación Bayesiana para obtener los Estimadores de Bayes.
6. Estimar los cuantiles de la DGVE para los 10 años siguientes utilizando los Estimadores de Bayes.
7. Generar los mapas de predicción para observar la tendencia de los niveles muy altos de ozono en cada una de las Estaciones de Monitoreo.

## 5.4. Aplicación al análisis de Datos del Valle de México

Con base en los registros provenientes de las bases de datos de la Red Automática de Monitoreo Atmosférico (RAMA) del Sistema de Monitoreo Atmosférico (SIMAT); cuya información se encuentra disponible en este enlace <http://www.calidadaire.df.gob.mx/calidadaire/productos/basesdedatos/bases.php>, se seleccionaron 17 estaciones de monitoreo, las cuales contaban con mayor cantidad de datos disponibles (Ver Cuadro 5.1). En general, la mayoría de las estaciones contaron con registros en todos los años a excepción de las estaciones San Agustín y Tlalpan, las cuales durante un lapso de un año no reportaron ningún valor. De manera que se estudiaron periodos de hasta 27 años que van desde 1986 hasta el primer semestre del 2012:

## 5.4. Aplicación al análisis de Datos del Valle de México

---

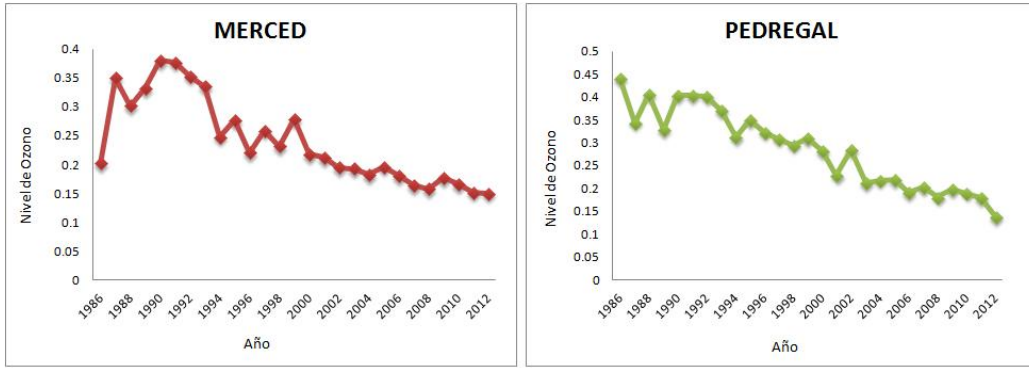
**Cuadro 5.1:** Periodo de Estudio por Estación Meteorológica

<b>ESTACIÓN</b>	<b>AÑO</b>	<b>PERIODO</b>
MERCED	27	1986-2012
PEDREGAL	27	1986-2012
XALOSTOC	27	1986-2012
TLALNEPANTLA	27	1986-2012
SAN AGUSTIN	25	1986-2012
AZCAPOTZALCO	26	1986-2011
CERRO DE LA ESTRELLA	26	1986-2011
PLATEROS	26	1986-2011
FES ACATLÁN	23	1990-2012
UAM IZTAPALAPA	23	1990-2012
LAGUNILLA	22	1990-2011
TAXQUEÑA	19	1993-2011
TACUBA	19	1993-2011
TLÁHUAC	19	1994-2012
MONTECILLO	19	1994-2012
CUAJIMALPA	19	1994-2012
TLAPAN	18	1994-2012

### 5.4.1. Obtención de los máximos diarios y máximos anuales

Para cada una de las estaciones meteorológicas, se trabajó con la información disponible en las bases de datos para obtener los máximos diarios de los niveles de ozono; una vez logrado lo anterior, se tomaron los máximos anuales por año (Ver Apéndice B).

## 5.4. Aplicación al análisis de Datos del Valle de México



**Figura 5.1:** Máximos Anuales de la Estación Merced y Pedregal

En la figura 5.1 se muestra, a manera de ejemplo, la distribución de los máximos anuales para la estación meteorológica Merced y Pedregal, respectivamente. En ambas gráficas puede observarse una tendencia a la baja en los niveles máximos de ozono conforme el transcurso de los años; lo cual se corroborará al estimar la tendencia en los cuantiles de la distribución.

De igual manera, se observa un comportamiento similar en los máximos anuales de las demás estaciones, cuyas gráficas pueden ser consultadas en la sección de Apéndices (Apéndice A).

### 5.4.2. Modelo Propuesto

Después de haber obtenido los máximos anuales  $Y_i$ ; lo cuales, para este trabajo se supone tienen el comportamiento de una Distribución de Valores Extremos Generalizada (DGEV), donde el parámetro de localización  $\mu$  está en función de la variable tiempo, denotado por  $\mu_t$ , se propone el siguiente modelo:

$$f(Y_i/\mu_t, \sigma, \xi) \sim dgev(\mu_t, \sigma, \xi), \quad i = 1, 2, \dots, n$$



## 5.4. Aplicación al análisis de Datos del Valle de México

---

Donde:

$$\mu_t = \beta_0 + \beta_1 t, \quad \beta_0, \beta_1 \in \mathbb{R}$$

Y cuya función de Distribución tiene la forma de la ecuación (3.1),

$$G(x; \mu_t, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu_t}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \quad \text{si } 1 + \xi\left(\frac{x - \mu}{\sigma}\right) > 0$$

donde  $\sigma > 0$ ,  $-\infty < \mu < \infty$ ,  $1 + \xi(x - \mu)/\sigma > 0$ ,  $x_+ = \max\{x, 0\}$ .

### 5.4.3. Obtención de los Estimadores de Bayes

Una vez planteado nuestro modelo se procedió entonces a aplicar el enfoque Bayesiano a los datos de máximos anuales para posteriormente estimar la tendencia en los cuantiles de la distribución. Así, basándonos en el Teorema de Bayes expresado en la ecuación (4.1) se obtuvo la distribución posterior de cada uno de los parámetros; para lo cual, se requería calcular la función de Verosimilitud de los parámetros dada la información y contar con las distribuciones a priori. En particular para este trabajo se emplearon distribuciones a priori no informativas para la estimación de los parámetros, las cuales se presentan a continuación:

$$\beta_0, \beta_1, \xi \sim dflat(\text{ d. plana - no informativa})$$

$$\sigma \sim GammaInversa(\alpha, \beta)$$

$$\text{Es decir, } f(\sigma) = \frac{\beta^\alpha}{\Gamma(\alpha)} \sigma^{-(\alpha+1)} e^{-\frac{\beta}{\sigma}}, \sigma > 0$$

Después de haber establecido lo anterior y mediante el uso de herramientas numéricas (Cadenas de Markov Monte Carlo (MCMC)) nos

#### 5.4. Aplicación al análisis de Datos del Valle de México

apoyamos en el software libre OpenBugs para obtener la Distribución posterior de cada uno de los parámetros en las 17 estaciones de monitoreo y estimar los valores del parámetro en función del tiempo  $\beta_1$ , obteniéndose los siguientes resultados:

**Cuadro 5.2:** Valores del Estimador Bayesiano  $\beta_1$  asociado a la variable tiempo, incluido en el parámetro de localización  $\mu_t$  de la DGVE e Intervalo de Confianza.

ESTACIÓN	$\beta_1$	Intervalo de Credibilidad del 95 %	
MERCED	-0.009618	-0.01117	-0.006752
PEDREGAL	-0.01094	-0.01191	-0.008478
XALOSTOC	-0.001584	-0.004927	6.36E-04
TLALNEPANTLA	-0.005845	-0.008746	-0.003205
SAN AGUSTIN	-0.001822	-0.003838	-3.54E-04
AZCAPOTZALCO	-0.00778	-0.0102	-0.004046
CERRO DE LA ESTRELLA	-0.005494	-0.007488	-0.003864
PLATEROS	-0.007223	-0.01053	-0.004696
FES ACATLÁN	-0.007964	-0.01046	-0.006016
UAM IZTAPALAPA	-0.007048	-0.008122	-0.005835
LAGUNILLA	-0.007066	-0.01094	-0.004423
TAXQUEÑA	-0.008247	0.01008	-0.006181
TACUBA	-0.0102	-0.01202	-0.008233
TLÁHUAC	-0.007022	0.009665	-0.004475
MONTECILLO	-0.003318	-0.004667	-0.001976
CUAJIMALPA	-0.006219	-0.007849	-0.004587
TLALPAN	-0.008872	-0.01063	-0.006671

#### 5.4. Aplicación al análisis de Datos del Valle de México

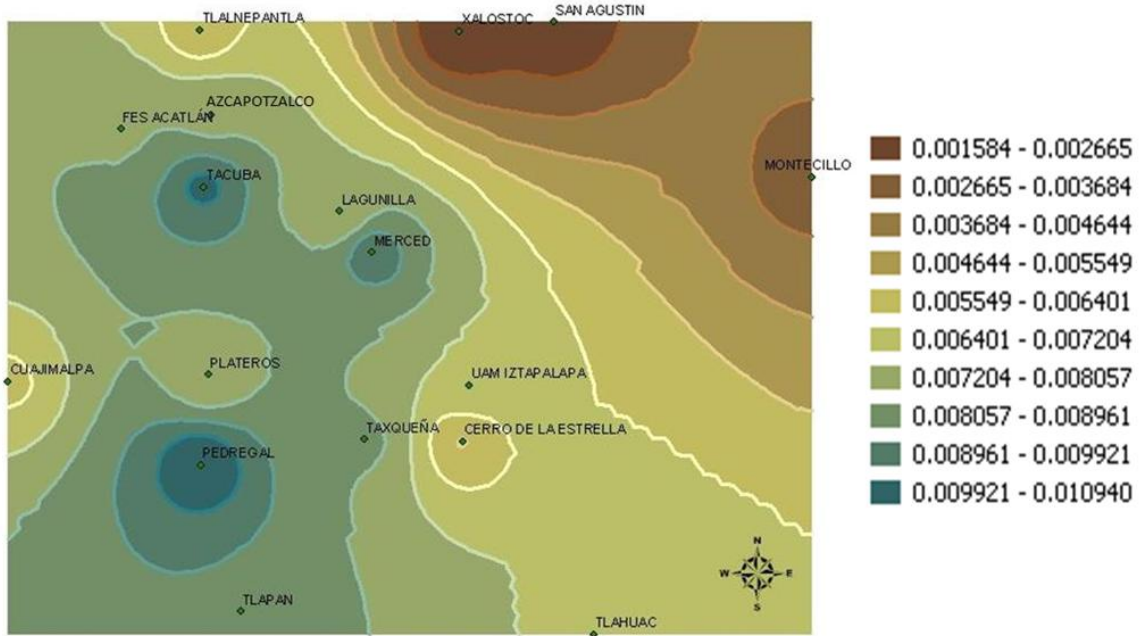
Con los resultados reportados en el Cuadro 5.2 se probó el siguiente juego de hipótesis:  $H_0 : \beta_1 = 0$  cuando no hay dependencia en el tiempo y  $H_1 : \beta_1 \neq 0$  cuando existe dependencia en el tiempo; de esta manera de los resultados obtenidos se observa que todos los valores para  $\beta_1$  son diferentes de cero, con lo que se corroboró que en efecto existe dependencia de los datos con la variable tiempo. Además se puede ver que el parámetro ligado a la variable tiempo para la estación Merced está comprendido entre -0.01117 y -0.006752 con un 95 % de probabilidad; de manera similar se observa en los demás casos.

Por otra parte, para ubicar las estaciones de monitoreo se emplearon los valores reportados del parámetro  $\beta_1$  y las coordenadas geográficas de cada estación (ver Cuadro 5.3) para, por medio del método de interpolación “kriging” elaborar un mapa con el paquete ArcMAP que nos permitiera identificar aquellas zonas y en particular las estaciones donde los niveles de ozono son más altos.

**Cuadro 5.3:** Coordenadas Geográficas de las Estaciones de Monitoreo

ESTACIÓN	LONGITUD(UTM)	LATITUD(UTM)
AZCAPOTZALCO	479141.9433	2154935.837
TLÁHUAC	498926.0809	2128017.235
TLALPAN	480671.6294	2129195.723
CUAJIMALPA	468635.4877	2141111.938
FES ACATLÁN	474470.3812	2154170.617
LAGUNILLA	485809.9968	2149915.478
UAM IZTAPALAPA	492525.5242	2140927.218
CERRO DE LA ESTRELLA	492156.396	2137994.694
MERCED	487478.276	2147790.317
PLATEROS	478985.5724	2141451.244
PEDREGAL	478600.8304	2136771.854
SAN AGUSTIN	496859.0599	2159721.413
XALOSTOC	491979.9041	2159225.075
TLALNEPANTLA	478573.394	2159309.729
TAXQUEÑA	487059.1698	2138103.435
TACUBA	478750.3965	2151193.447
MONTECILLO	510235.6854	2151708.034

## 5.4. Aplicación al análisis de Datos del Valle de México



**Figura 5.2:** Localización de las 17 estaciones de monitoreo y detección de las zonas más afectadas con niveles de ozono muy altos (ppm)

En el mapa de la figura 5.2 observa con tonalidades en azul las estaciones que se encuentran dentro de las zonas donde los niveles de ozono son muy altos, siendo las más afectadas: Pedregal, Tacuba, Merced, Tlalpan, y Taxqueña. Con tonalidades en café las zonas con menores niveles de ozono: Xalostoc y San Agustín en la Ciudad de México, y Montecillo en la Zona Metropolitana.

### 5.4.4. Obtención de los Cuantiles

Para obtener los cuantiles de la Distribución se utilizó la Función Cuantil para la DGEV denotada en la ecuación (3.3),

$$Z_{\alpha}(p) = \begin{cases} \mu - \frac{\sigma}{\xi} \{1 - (-\log(1 - p))^{-\xi}\}, & \xi \neq 0 \\ \mu - \sigma \log\{-\log(1 - p)\}, & \xi = 0 \end{cases}$$

#### 5.4. Aplicación al análisis de Datos del Valle de México

---

donde  $y_p = -\log(1 - p)$ ,  $0 \leq p \leq 1$

Como se puede observar los parámetros de la Función Cuantil son los mismos que integran la DGVE por lo que con ayuda del software Open-Bugs se obtuvieron los valores de los Estimadores Bayesianos  $\xi$ ,  $\sigma$  y  $\mu_t$  para la DGVE. Los resultados para  $\xi$  y  $\sigma$  se presentan en el Cuadro 5.4.

**Cuadro 5.4:** Valores Estimados para  $\xi$  y  $\sigma$  por estación meteorológica

ESTACIÓN	$\xi$	$\sigma$
MERCED	-0.5534	0.04678
PEDREGAL	-0.5416	0.03138
XALOSTOC	-0.1033	0.03829
TLALNEPANTLA	-0.2853	0.03697
SAN AGUSTIN	-0.02714	0.03187
AZCAPOTZALCO	-0.5958	0.05502
CERRO DE LA ESTRELLA	-0.1642	0.03423
PLATEROS	-0.122	0.05103
FES ACATLÁN	0.1412	0.02647
UAM IZTAPALAPA	-0.279	0.01862
LAGUNILLA	-0.147	0.03928
TAXQUEÑA	-0.2823	0.0237
TACUBA	-0.1471	0.02339
TLÁHUAC	-0.2126	0.02663
MONTECILLO	-0.06394	0.01595
CUAJIMALPA	-0.573	0.02959
TLALPAN	-0.4017	0.02198

Para la media  $\mu_t$  se estimaron los valores de los 10 años siguientes a partir del 2013, teniendo en consideración que dicho parámetro se encuentra en función de la variable tiempo. Los resultados se muestran en el Cuadro 5.5 y 5.6.

#### 5.4. Aplicación al análisis de Datos del Valle de México

**Cuadro 5.5:** Valores Estimados de la media ( $\mu_t$ ) en función del tiempo

ESTACIÓN	2013	2014	2015	2016	2017
MERCED	0.094096	0.084478	0.07486	0.065242	0.055624
PEDREGAL	0.12468	0.11374	0.1028	0.09186	0.08092
XALOSTOC	0.160748	0.159164	0.15758	0.155996	0.154412
TLALNEPANTLA	0.13844	0.132595	0.12675	0.120905	0.11506
SAN AGUSTIN	0.151228	0.149406	0.147584	0.145762	0.14394
AZCAPOTZALCO	0.12624	0.11846	0.11068	0.1029	0.09512
CERRO DE LA ESTRELLA	0.119962	0.114468	0.108974	0.10348	0.097986
PLATEROS	0.148179	0.140956	0.133733	0.12651	0.119287
FES ACATLÁN	0.137064	0.1291	0.121136	0.113172	0.105208
UAM IZTAPALAPA	0.133648	0.1266	0.119552	0.112504	0.105456
LAGUNILLA	0.125982	0.118916	0.11185	0.104784	0.097718
TAXQUEÑA	0.10726	0.099013	0.090766	0.082519	0.074272
TACUBA	0.1218	0.1116	0.1014	0.0912	0.081
TLÁHUAC	0.12176	0.114738	0.107716	0.100694	0.093672
MONTECILLO	0.12384	0.120522	0.117204	0.113886	0.110568
CUAJIMALPA	0.15292	0.146701	0.140482	0.134263	0.128044
TLALPAN	0.131232	0.12236	0.113488	0.104616	0.095744

**Cuadro 5.6:** Valores Estimados de la media ( $\mu_t$ ) en función del tiempo (Continuación)

ESTACIÓN	2018	2019	2020	2021	2022
MERCED	0.046006	0.036388	0.02677	0.017152	0.007534
PEDREGAL	0.06998	0.05904	0.0481	0.03716	0.02622
XALOSTOC	0.152828	0.151244	0.14966	0.148076	0.146492
TLALNEPANTLA	0.109215	0.10337	0.097525	0.09168	0.085835
SAN AGUSTIN	0.142118	0.140296	0.138474	0.136652	0.13483
AZCAPOTZALCO	0.08734	0.07956	0.07178	0.064	0.05622
CERRO DE LA ESTRELLA	0.092492	0.086998	0.081504	0.07601	0.070516
PLATEROS	0.112064	0.104841	0.097618	0.090395	0.083172
FES ACATLÁN	0.097244	0.08928	0.081316	0.073352	0.065388
UAM IZTAPALAPA	0.098408	0.09136	0.084312	0.077264	0.070216
LAGUNILLA	0.090652	0.083586	0.07652	0.069454	0.062388
TAXQUEÑA	0.066025	0.057778	0.049531	0.041284	0.033037
TACUBA	0.0708	0.0606	0.0504	0.0402	0.03
TLÁHUAC	0.08665	0.079628	0.072606	0.065584	0.058562
MONTECILLO	0.10725	0.103932	0.100614	0.097296	0.093978
CUAJIMALPA	0.121825	0.115606	0.109387	0.103168	0.096949
TLALPAN	0.086872	0.078	0.069128	0.060256	0.051384

#### 5.4. Aplicación al análisis de Datos del Valle de México

Una vez obtenidos los valores estimados de los parámetros mostrados en los cuadros 5.4, 5.5 y 5.6, se fijó un  $\alpha = 0.95$  para estimar los cuantiles de la DGVE, por medio de la Función Cuantil, para los diez años siguientes y cuyos resultados se muestran en los Cuadros 5.7 y 5.8.

**Cuadro 5.7:** Cuantiles estimados para la DGVE del periodo comprendido de 2013 - 2022

ESTACIÓN	2013	2014	2015	2016	2017
MERCED	0.168330801	0.158712801	0.149094801	0.139476801	0.129858801
PEDREGAL	0.175237557	0.164297557	0.153357557	0.142417557	0.131477557
XALOSTOC	0.281198606	0.279614606	0.278030606	0.276446606	0.274862606
TLALNEPANTLA	0.224251758	0.218406758	0.212561758	0.206716758	0.200871758
SAN AGUSTIN	0.266419836	0.264597836	0.262775836	0.260953836	0.259131836
AZCAPOTZALCO	0.209013028	0.201233028	0.193453028	0.185673028	0.177893028
CERRO DE LA ESTRELLA	0.216805266	0.211311266	0.205817266	0.200323266	0.194829266
PLATEROS	0.303489741	0.296266741	0.289043741	0.281820741	0.274597741
FES ACATLÁN	0.270376649	0.262412649	0.254448649	0.246484649	0.238520649
UAM IZTAPALAPA	0.177296337	0.170248337	0.163200337	0.156152337	0.149104337
LAGUNILLA	0.24044073	0.23337473	0.22630873	0.21924273	0.21217673
TAXQUEÑA	0.162529613	0.154282613	0.146035613	0.137788613	0.129541613
TACUBA	0.1899448	0.1797448	0.1695448	0.1593448	0.1491448
TLÁHUAC	0.191228269	0.184206269	0.177184269	0.170162269	0.163140269
MONTECILLO	0.177701435	0.174383435	0.171065435	0.167747435	0.164429435
CUAJIMALPA	0.198721932	0.192502932	0.186283932	0.180064932	0.173845932
TLALPAN	0.17407933	0.16520733	0.15633533	0.14746333	0.13859133

## 5.4. Aplicación al análisis de Datos del Valle de México

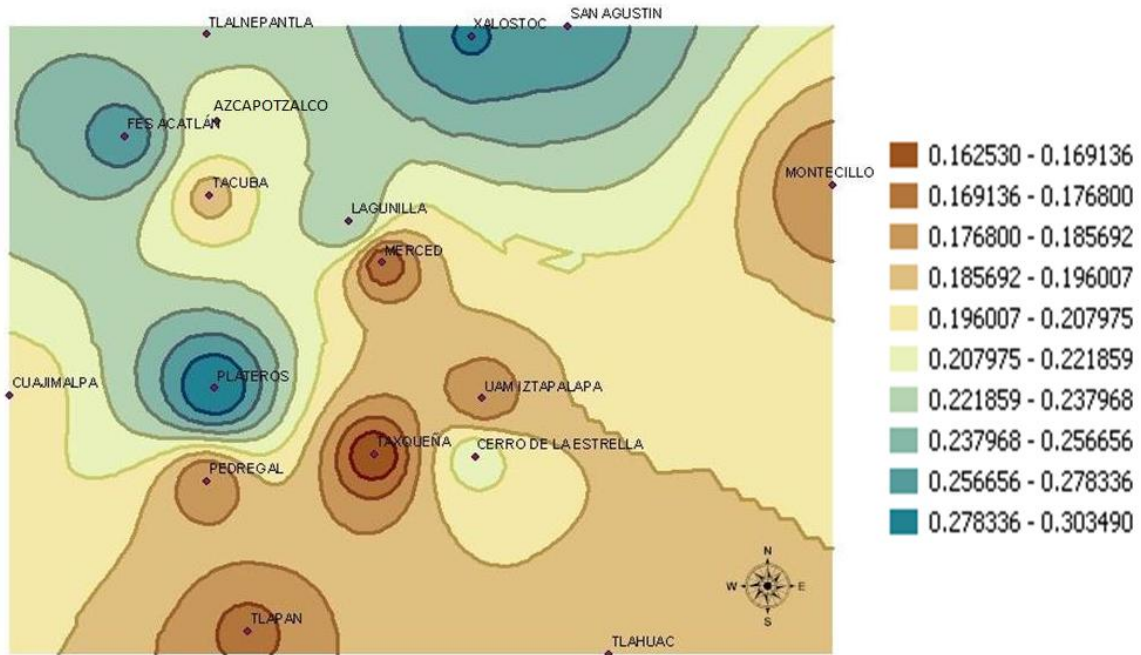
**Cuadro 5.8:** Cuantiles estimados para la DGVE del periodo comprendido de 2013 - 2022 (Continuación)

ESTACIÓN	2018	2019	2020	2021	2022
MERCED	0.120240801	0.110622801	0.101004801	0.091386801	0.081768801
PEDREGAL	0.120537557	0.109597557	0.098657557	0.087717557	0.076777557
XALOSTOC	0.273278606	0.271694606	0.270110606	0.268526606	0.266942606
TLALNEPANTLA	0.195026758	0.189181758	0.183336758	0.177491758	0.171646758
SAN AGUSTIN	0.257309836	0.255487836	0.253665836	0.251843836	0.250021836
AZCAPOTZALCO	0.170113028	0.162333028	0.154553028	0.146773028	0.138993028
CERRO DE LA ESTRELLA	0.189335266	0.183841266	0.178347266	0.172853266	0.167359266
PLATEROS	0.267374741	0.260151741	0.252928741	0.245705741	0.238482741
FES ACATLÁN	0.230556649	0.222592649	0.214628649	0.206664649	0.198700649
UAM IZTAPALAPA	0.142056337	0.135008337	0.127960337	0.120912337	0.113864337
LAGUNILLA	0.20511073	0.19804473	0.19097873	0.18391273	0.17684673
TAXQUEÑA	0.121294613	0.113047613	0.104800613	0.096553613	0.088306613
TACUBA	0.1389448	0.1287448	0.1185448	0.1083448	0.0981448
TLÁHUAC	0.156118269	0.149096269	0.142074269	0.135052269	0.128030269
MONTECILLO	0.161111435	0.157793435	0.154475435	0.151157435	0.147839435
CUAJIMALPA	0.167626932	0.161407932	0.155188932	0.148969932	0.142750932
TLALPAN	0.12971933	0.12084733	0.11197533	0.10310333	0.09423133

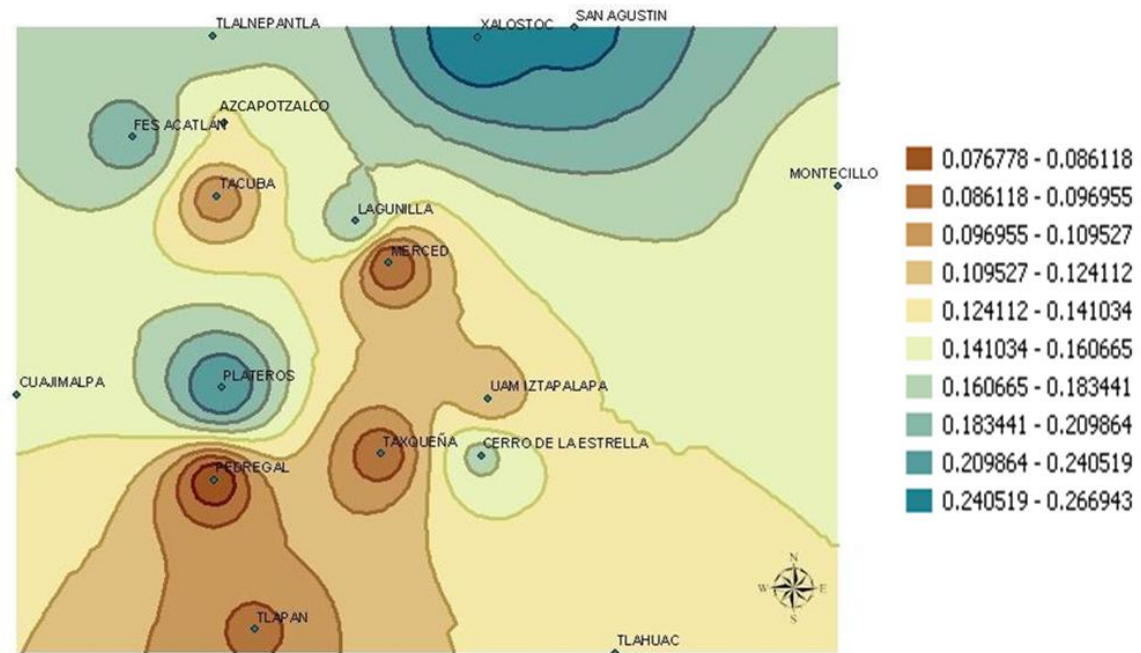
De esta manera, con los cuantiles estimados en los cuadros anteriores se aplicó nuevamente el método de interpolación “kriging” para realizar los mapas de predicción en el programa ArcMap correspondientes a los años 2013 y 2022 con la finalidad de observar el comportamiento de los niveles muy altos de ozono en la Ciudad de México y su Zona Metropolitana.



## 5.4. Aplicación al análisis de Datos del Valle de México



**Figura 5.3:** Cuantiles de la DGVE estimados para el año 2013, en las 17 estaciones de monitoreo.

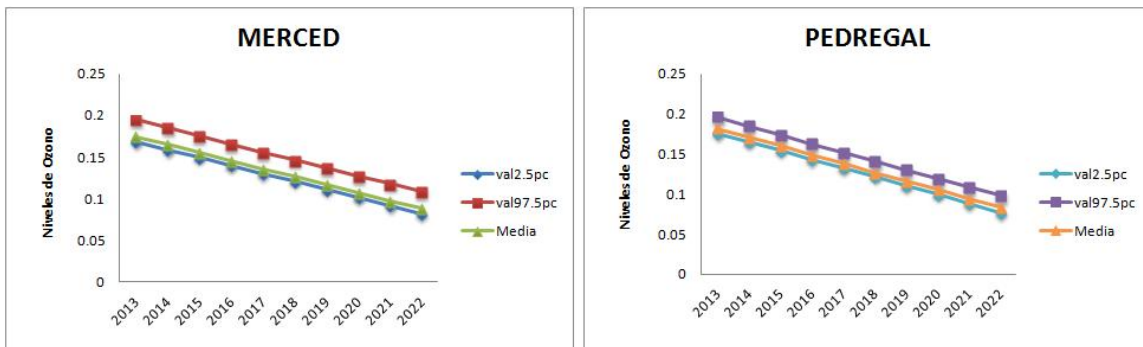


**Figura 5.4:** Cuantiles de la DGVE estimados para el año 2022, en las 17 estaciones de monitoreo.

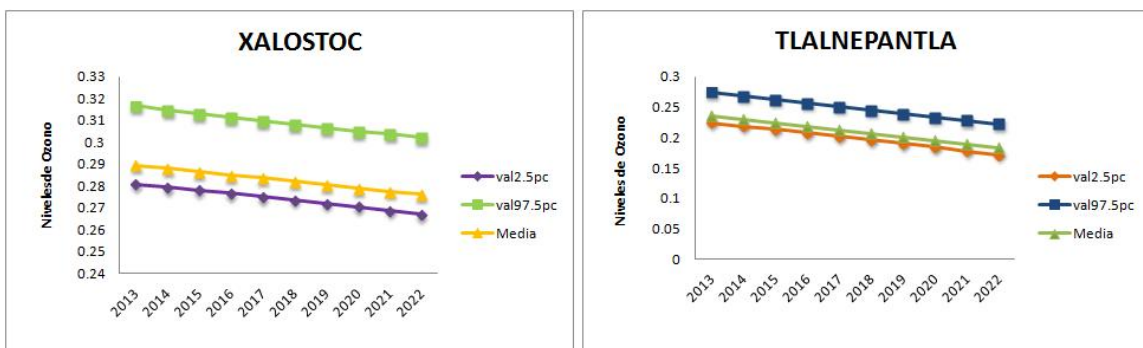
#### 5.4. Aplicación al análisis de Datos del Valle de México

Así, se puede observar con distintas tonalidades en las figuras 5.3 y 5.4 como los niveles máximos de ozono muestran una tendencia a la baja con el transcurso de los años. Por lo que se estima que para el año 2022 las estaciones Pedregal, Merced, Taxqueña, Tlalpan y Tacuba alcanzaran niveles de ozono menores a las 0.110 ppm.

Posteriormente se obtuvieron los Intervalos de Credibilidad del 95 % para las estimaciones de los cuantiles de cada estación meteorológica del año 2013 al año 2022, cuyos resultados se presentan en las figuras siguientes:



**Figura 5.5:** Intervalo de Credibilidad del 95 % para los cuantiles estimados de las estaciones Merced y Pedregal



**Figura 5.6:** Intervalo de Credibilidad del 95 % para los cuantiles estimados de las estaciones Xalostoc y Tlalnepantla

## 5.4. Aplicación al análisis de Datos del Valle de México

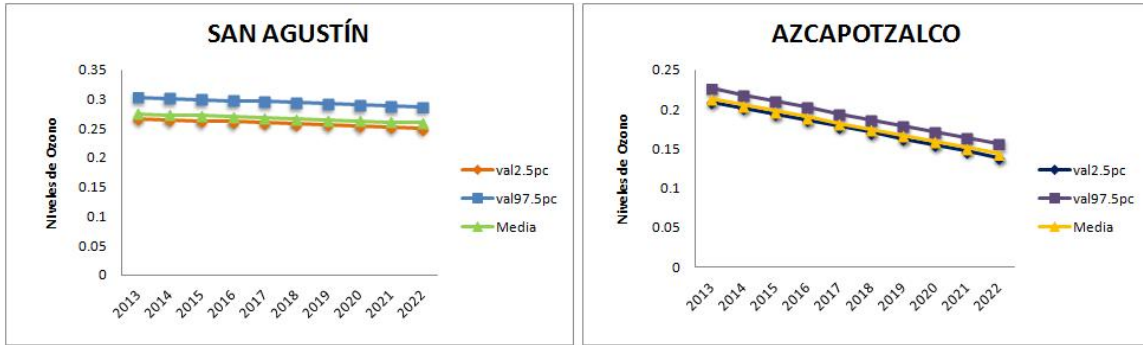


Figura 5.7: Intervalo de Credibilidad del 95 % para los cuantiles estimados de las estaciones San Agustín y Azcapotzalco

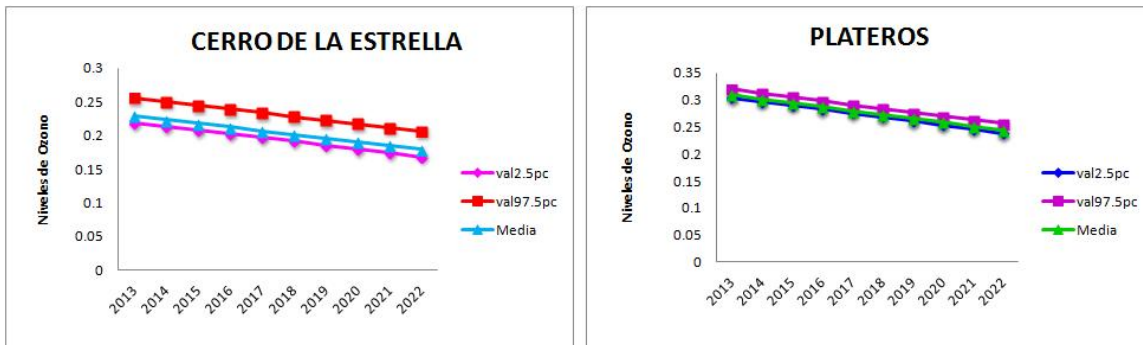


Figura 5.8: Intervalo de Credibilidad del 95 % para los cuantiles estimados de las estaciones Cerro de la Estrella y Plateros

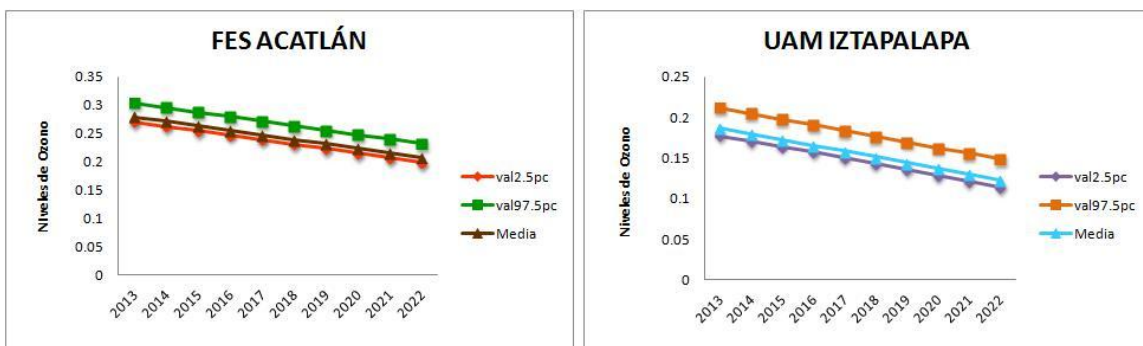


Figura 5.9: Intervalo de Credibilidad del 95 % para los cuantiles estimados de las estaciones FES Acatlán y UAM Iztapalapa

## 5.4. Aplicación al análisis de Datos del Valle de México

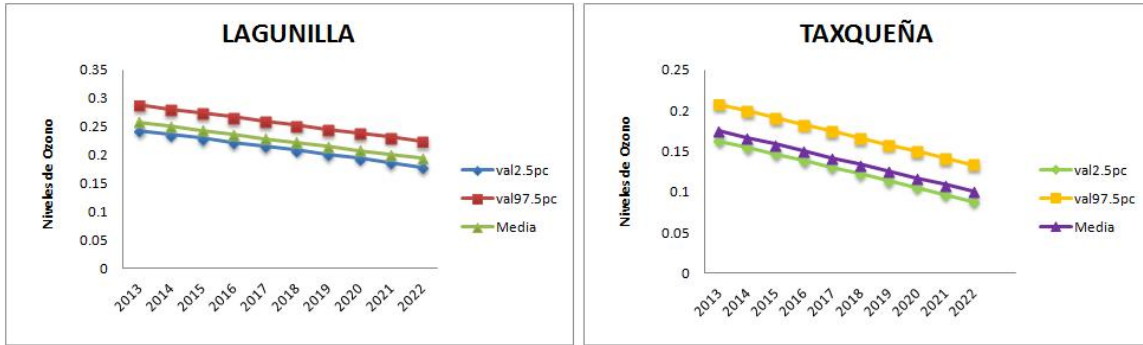


Figura 5.10: Intervalo de Credibilidad del 95 % para los cuantiles estimados de las estaciones Lagunilla y Taxqueña

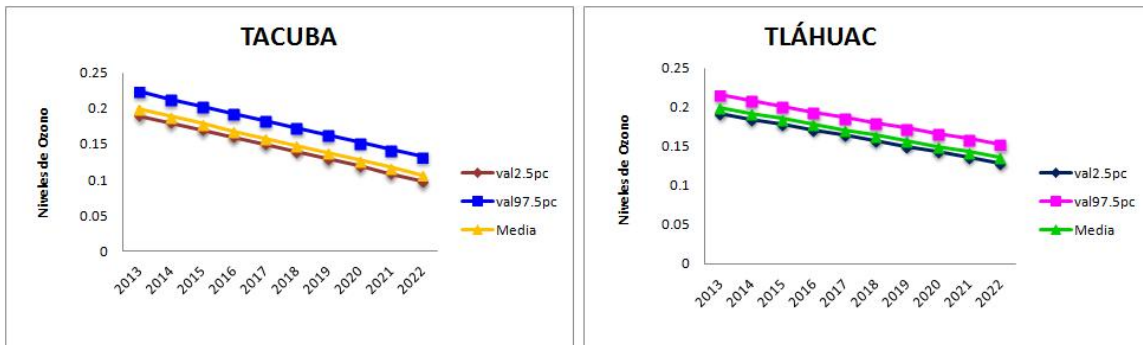


Figura 5.11: Intervalo de Credibilidad del 95 % para los cuantiles estimados de las estaciones Tacuba y Tláhuac

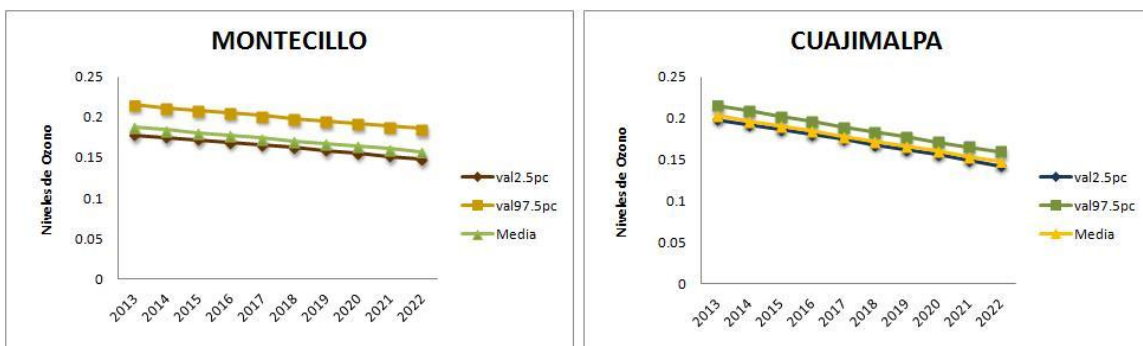
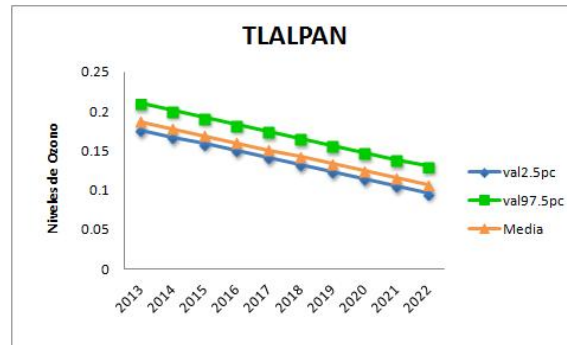


Figura 5.12: Intervalo de Credibilidad del 95 % para los cuantiles estimados de las estaciones Montecillo y Cuajimalpa

## 5.4. Aplicación al análisis de Datos del Valle de México

---



**Figura 5.13:** Intervalo de Credibilidad del 95 % para los cuantiles estimados de la estación Tlalpan

Finalmente, en las figuras anteriores (de la 5.5 a la 5.13) se presenta el gráfico del Intervalo de Credibilidad del 95 % de los cuantiles estimados para las 17 estaciones de monitoreo, los cuales muestran una tendencia a la baja.

Así, a manera de ejemplo se puede observar que para el caso de la estación Pedregal el cuantil del 95 % para el año 2022 está comprendido entre 0.07676 ppm y 0.09821 ppm, con un 95 % de probabilidad. De manera similar, se realiza la interpretación para los demás años en cada una de las estaciones.

# Capítulo 6

## Resultados y Conclusiones

### 6.1. Resultados

- Con base en la información proporcionada por los datos, las horas en que se registran mayores niveles de concentración de ozono se encuentran entre las 15 y 17 hrs; debido tal vez, a que la temperatura ambiente alcanza su máximo en ese horario.
- Con los datos disponibles fue posible obtener los estimadores de Bayes para cada uno de los parámetros de la DGVE.
- En la mayoría de las estaciones de monitoreo analizadas se observa que la tendencia en los niveles muy altos de ozono es decreciente con el transcurso de los años; principalmente en las estaciones Pedregal, Merced y Tacuba que en el año 2012 presentaban mayores niveles de contaminación por ozono, lo cual se puede constatar con los cuantiles estimados para los 10 años posteriores al 2012.
- La importancia de continuar con el monitoreo de los niveles máximos de ozono radica en el hecho de que aún se registran niveles superiores a la NOM-020-SSA1-1993 (0.110 ppm), incluso en los resultados reportados al estimar los cuantiles para el año 2022 se observa que 12 de las 17 estaciones analizadas aún presentan niveles de ozono muy altos.
- Las zonas más afectadas son las cercanas a la zona nor-poniente

de la Ciudad de México.

## 6.2. Conclusiones y Recomendaciones

El modelo propuesto permitió llevar a cabo el análisis de los niveles máximos de ozono en la Ciudad de México y su Zona Metropolitana; sin embargo, como trabajo futuro es recomendable incluir otras variables como temperatura, altitud de la zona, velocidad del viento, transporte, emisiones de ozono por fábrica, entre otras para contar con más herramientas que nos permitan tener un mejor panorama del comportamiento de los niveles de ozono.

El programa desarrollado en el software OpenBugs permite ver la utilidad que tienen las Cadenas de Markov Monte Carlo (MCMC) cuando se desea trabajar con Estadística Bayesiana, principalmente porque todas las inferencias se basan en la función posterior y en la mayoría de los casos el cálculo de los estimadores no se puede realizar de forma analítica.

Con el apoyo de las políticas públicas implementadas por el Estado se ha contribuido a la disminución de los niveles muy altos de ozono; sin embargo, es importante que se propongan nuevas políticas o que las ya existentes sean reestructuradas para lograr una mayor aceleración en la disminución de dichos niveles, principalmente para aquellas zonas más afectadas. Para así evitar situaciones de contingencia ambiental como las presentadas recientemente en el Valle de México.

# Capítulo 7

## Apéndice

### Apéndice A. Gráficas de los Máximos Anuales por Estación Meteorológica

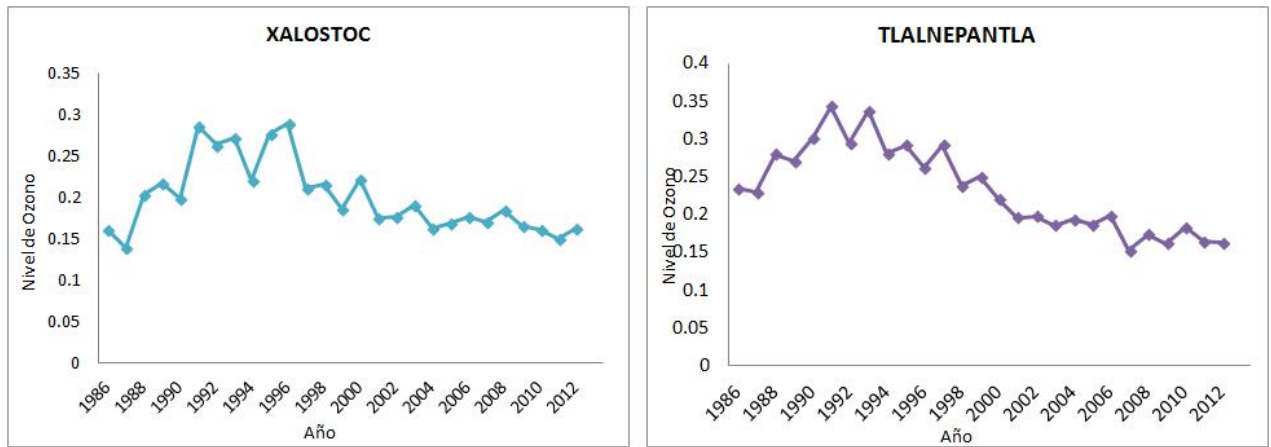


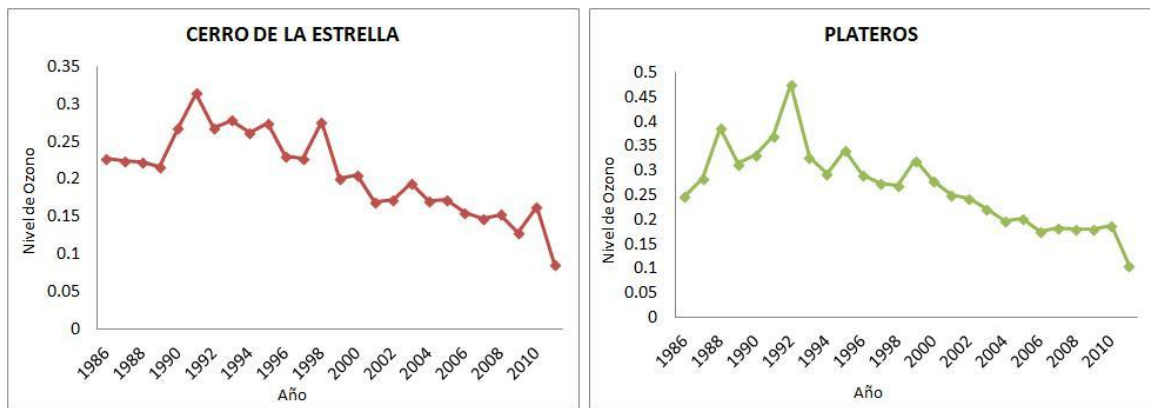
Figura 7.1: Máximos Anuales de las Estaciones Xalostoc y Tlalnepantla



## 7. Apéndice



**Figura 7.2:** Máximos Anuales de las Estaciones San Agustín y Azcapotzalco



**Figura 7.3:** Máximos Anuales de las Estaciones Cerro de la Estrella y Plateros

## 7. Apéndice

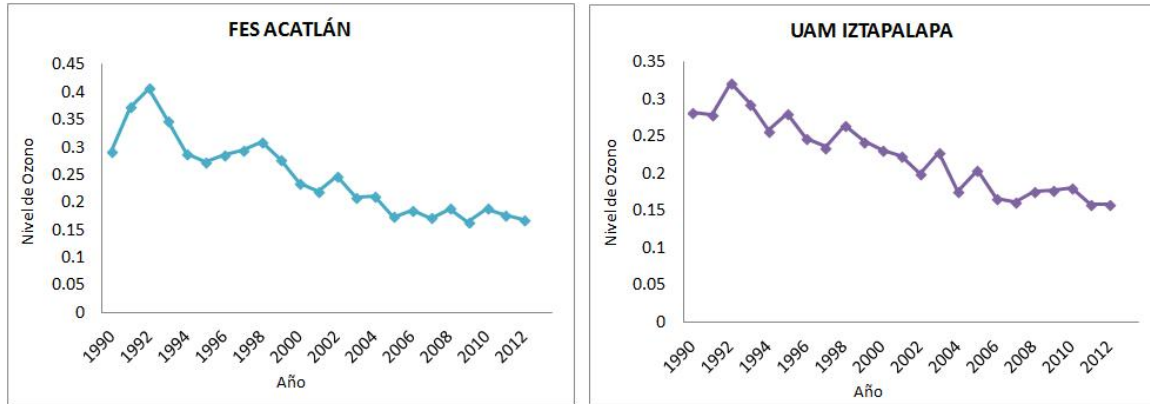


Figura 7.4: Máximos Anuales de las Estaciones FES Acatlán y UAM Iztapalapa

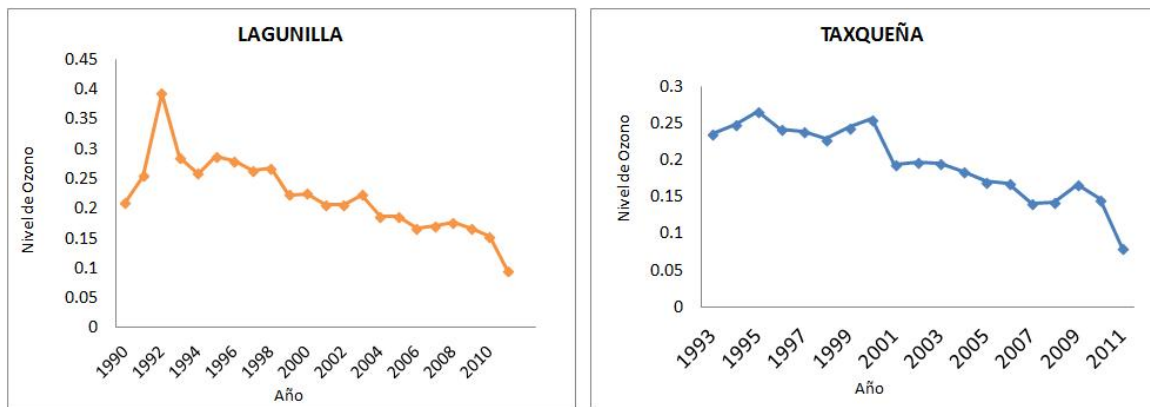
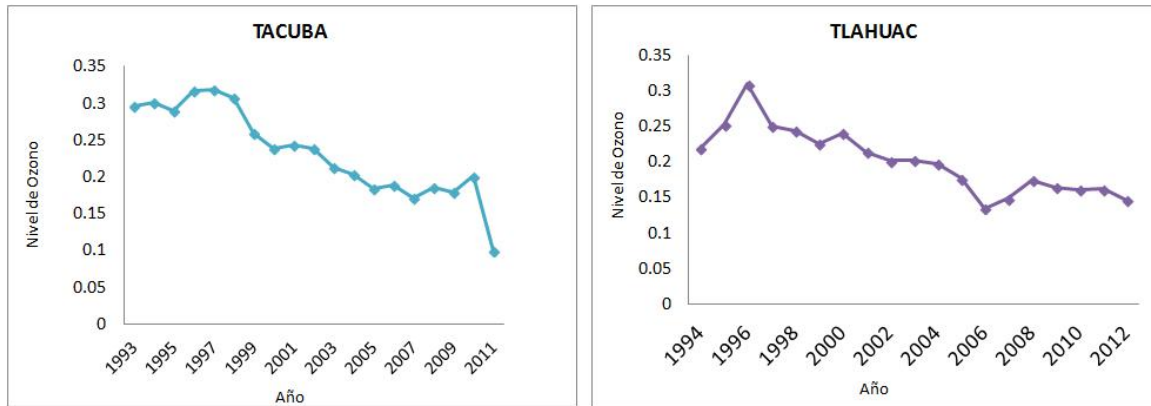
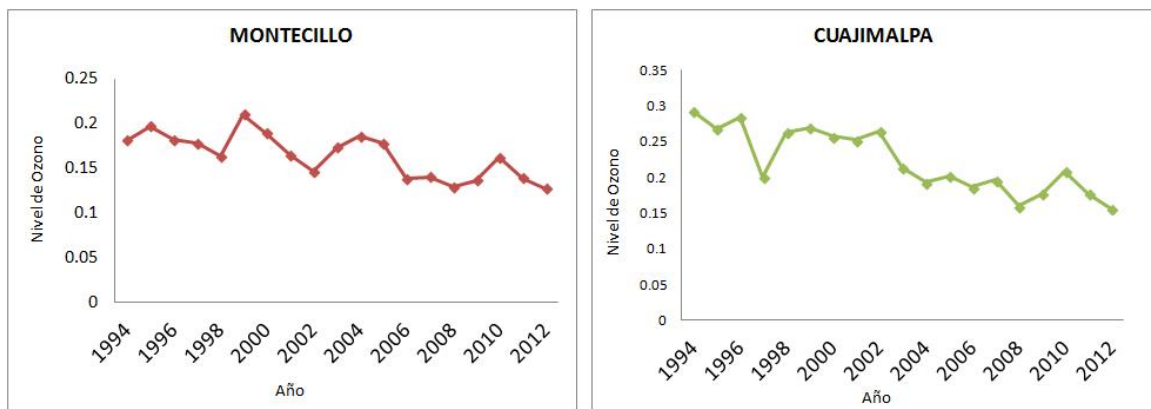


Figura 7.5: Máximos Anuales de las Estaciones Lagunilla y Taxqueña

## 7. Apéndice



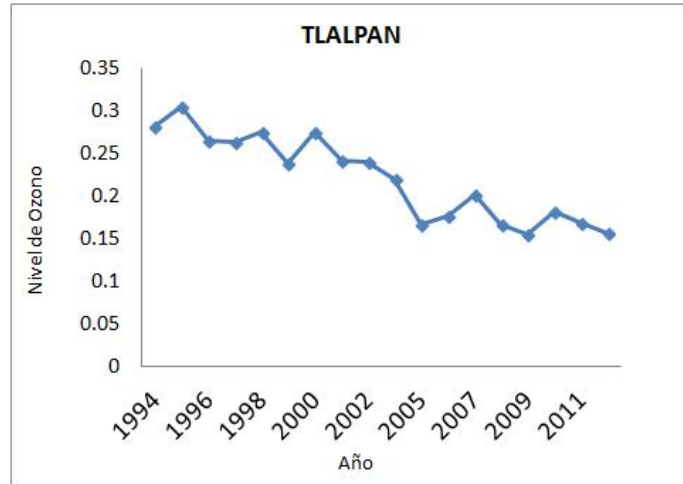
**Figura 7.6:** Máximos Anuales de las Estaciones Tacuba y Tláhuac



**Figura 7.7:** Máximos Anuales de las Estaciones Montecillo y Cuajimalpa

## 7. Apéndice

---



**Figura 7.8:** Máximos Anuales de las Estacion Tlalpan

## Apéndice B. Datos de los Niveles Máximos de Ozono Anuales por Estación Meteorológica

Cuadro 7.1: Niveles Máximos de Ozono Anuales

ESTACIÓN	1986	1987	1988	1989	1990	1991	1992	1993	1994
MERCED	0.202	0.35	0.302	0.332	0.38	0.377	0.353	0.335	0.248
PEDREGAL	0.441	0.344	0.405	0.329	0.403	0.404	0.402	0.37	0.312
XALOSTOC	0.161	0.139	0.203	0.217	0.198	0.285	0.263	0.272	0.22
TLALNEPANTLA	0.235	0.229	0.28	0.27	0.301	0.343	0.294	0.337	0.281
SAN AGUSTIN	0.17	0.154	NR	NR	0.171	0.212	0.244	0.208	0.198
AZCAPOTZALCO	0.156	0.254	0.282	0.263	0.275	0.354	0.364	0.313	0.287
CERRO DE LA ESTRELLA	0.227	0.224	0.223	0.216	0.268	0.314	0.268	0.278	0.262
PLATEROS	0.247	0.284	0.387	0.314	0.332	0.37	0.475	0.328	0.293
FES ACATLÁN	NR	NR	NR	NR	0.293	0.373	0.407	0.349	0.289
UAM IZTAPALAPA	NR	NR	NR	NR	0.283	0.28	0.322	0.295	0.258
LAGUNILLA	NR	NR	NR	NR	0.211	0.255	0.393	0.286	0.259
TAXQUEÑA	NR	NR	NR	NR	NR	NR	NR	0.236	0.249
TACUBA	NR	NR	NR	NR	NR	NR	NR	0.297	0.301
TLÁHUAC	NR	NR	NR	NR	NR	NR	NR	NR	0.219
MONTECILLO	NR	NR	NR	NR	NR	NR	NR	NR	0.181
CUAJIMALPA	NR	NR	NR	NR	NR	NR	NR	NR	0.292
TLAPAN	NR	NR	NR	NR	NR	NR	NR	NR	0.283

NR: No Registrado

## 7. Apéndice

**Cuadro 7.2:** Niveles Máximos de Ozono Anuales (Continuación)

ESTACIÓN	1995	1996	1997	1998	1999	2000	2001	2002	2003
MERCED	0.277	0.221	0.259	0.232	0.279	0.218	0.212	0.195	0.193
PEDREGAL	0.349	0.323	0.309	0.295	0.311	0.282	0.229	0.284	0.213
XALOSTOC	0.276	0.288	0.21	0.215	0.185	0.222	0.175	0.177	0.191
TLALNEPANTLA	0.292	0.262	0.292	0.238	0.25	0.22	0.196	0.198	0.186
SAN AGUSTIN	0.273	0.208	0.207	0.285	0.215	0.212	0.194	0.203	0.21
AZCAPOTZALCO	0.294	0.28	0.278	0.27	0.284	0.23	0.271	0.235	0.243
CERRO DE LA ESTRELLA	0.274	0.231	0.227	0.276	0.201	0.205	0.169	0.173	0.195
PLATEROS	0.341	0.291	0.274	0.27	0.321	0.279	0.25	0.242	0.222
FES ACATLÁN	0.274	0.287	0.294	0.309	0.277	0.235	0.221	0.247	0.209
UAM IZTAPALAPA	0.281	0.248	0.236	0.265	0.244	0.232	0.224	0.2	0.229
LAGUNILLA	0.287	0.279	0.264	0.268	0.223	0.225	0.207	0.207	0.223
TAXQUEÑA	0.266	0.242	0.239	0.229	0.245	0.256	0.194	0.198	0.196
TACUBA	0.29	0.317	0.318	0.307	0.259	0.239	0.244	0.238	0.213
TLÁHUAC	0.252	0.31	0.251	0.244	0.226	0.241	0.214	0.202	0.203
MONTECILLO	0.197	0.182	0.178	0.163	0.21	0.189	0.165	0.146	0.173
CUAJIMALPA	0.268	0.284	0.2	0.263	0.27	0.257	0.252	0.264	0.213
TLAPAN	0.305	0.266	0.264	0.276	0.239	0.275	0.243	0.24	0.22

**Cuadro 7.3:** Niveles Máximos de Ozono Anuales (Continuación)

ESTACIÓN	2004	2005	2006	2007	2008	2009	2010	2011	2012
MERCED	0.183	0.196	0.181	0.164	0.158	0.177	0.166	0.152	0.15
PEDREGAL	0.218	0.22	0.191	0.203	0.181	0.198	0.19	0.18	0.138
XALOSTOC	0.162	0.169	0.177	0.17	0.184	0.166	0.16	0.15	0.162
TLALNEPANTLA	0.193	0.187	0.199	0.152	0.174	0.162	0.183	0.164	0.163
SAN AGUSTIN	0.163	0.159	0.157	0.177	0.18	0.146	0.146	0.135	0.135
AZCAPOTZALCO	0.212	0.179	0.211	0.188	0.173	0.164	0.188	0.101	NR
CERRO DE LA ESTRELLA	0.171	0.172	0.155	0.148	0.153	0.129	0.163	0.086	NR
PLATEROS	0.197	0.202	0.176	0.183	0.18	0.18	0.187	0.106	NR
FES ACATLÁN	0.211	0.174	0.185	0.172	0.189	0.165	0.189	0.177	0.168
UAM IZTAPALAPA	0.177	0.205	0.167	0.162	0.177	0.178	0.181	0.159	0.159
LAGUNILLA	0.186	0.187	0.167	0.171	0.177	0.167	0.153	0.096	NR
TAXQUEÑA	0.185	0.171	0.169	0.141	0.143	0.167	0.147	0.08	NR
TACUBA	0.203	0.184	0.189	0.171	0.186	0.179	0.201	0.1	NR
TLÁHUAC	0.198	0.177	0.135	0.149	0.174	0.164	0.161	0.162	0.146
MONTECILLO	0.186	0.178	0.138	0.14	0.129	0.136	0.162	0.139	0.127
CUAJIMALPA	0.193	0.202	0.186	0.196	0.16	0.177	0.208	0.177	0.156
TLAPAN	NR	0.168	0.178	0.202	0.167	0.155	0.182	0.169	0.158

## Apéndice C. Código en OPENBUGS del Modelo Propuesto

```
model;
{
  for (i in 1 : N){
    Y[i] ~ dgev(media[i], sigma, xi)
  }
  for (i in 1 :N) {
    media[i] <- Const + beta.T*T[i]
  }
  sigma <- 1/sqrt(tau)
  Const ~ dflat()
  beta.T ~ dflat()
  tau ~ dgamma(0.0001,0.0001)
  xi ~ dflat()
}
```

# Referencias

- Anders, C. (1988). Interview: the effects are present. *México Journal*, 2:3–4.
- Bali, T. G. (2007). A generalized extreme value approach to financial risk measurement. *Journal of Money, Credit and Banking*, 39:1613 – 1649.
- Barry, C. A., Balakrishnan, N., and Nagaraja, H. (1992). *A First Course in Order Statistics*. Jhon Wiley & Sons, Inc.
- Bravo, H., Sosa, E., and Torres, R. (1992). *La Contaminación Atmosférica por Ozono en la Zona Metropolitana de la Ciudad de México, La contaminación atmosférica en México. Sus causas y sus efectos*. Comisión Nacional de Derechos Humanos México.
- Carlin, B. and Louis, T. (2000). *Bayes and Empirical Bayes Methods for Data Analysis. Second Edition*. New York: Chapman & Hall/CRC.
- Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Advanced Series.
- Chaplin, W. S. (1880). The relation between the tensile strengths of long and short bar. *Van Nostrand's Engineering Magazine*, 23:441–444.
- Coles, S. G. & Powell, E. A. (1996a). Bayesian methods in extreme value modelling: a review and new developments. *Int. Stat. Rev.*, 64:119–136.
- Coles, Stuart G. & Pan, F. (1996b). The analysis of extreme pollution levels: A case study. *Journal of Applied Statistics*, 23(2-3):333–348.



## Referencias

---

- Coles, S. G. & Tawn, J. A. (1996c). A bayesian analysis of extreme rainfall data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 45, No.4:463–478.
- Coles, Stuart & Pericchi, L. (2003). Anticipating catastrophes through extreme value modelling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 52:405–416.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag.
- Coles, S. and Tawn, J. (2005). Bayesian modelling of extreme surges on the uk east coast. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 363, No. 1831:1387–1406.
- Davison, A. C., Padoan, S. A. P., and Ribatet, M. (2012). Statistical modelling of spatial extremes. *Statistical Science*, 27:161–186.
- DeGroot, M. H. (1975). *Probability and Statistics*. ADDISON - WESLEY.
- Etienne, C., Lehmann, A., and Goyette, S. (2010). Spatial predictions of extreme wind speeds over switzerland using generalized additive models. *Journal of Applied Meteorology and Climatology*, 49:1956–1970.
- García, A. C. (2004). *Contaminantes Atmosféricos en la Zona Metropolitana de la Ciudad de México*. Secretaría de Educación Pública.
- Gumbel, E. J. (1935). Les valeurs extremes des distributions statistiques. *Annales d Institute H. Poincare*, 5, No.2:115–158.
- Gumbel, E. J. (1958). *Statistics of Extremes*. Columbia University Press.
- Gumbel, E. J. and Lieblein, J. (1954). Some applications of extreme value methods. *The American Statistician*, 8, No. 5:14–17.
- Huerta, G., Sansó, B., and Stroud, J. R. (2004). A spatiotemporal model for mexico city ozone levels. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 53, No. 2:231–248.

## Referencias

---

- Jagger, Thomas H. & Elsner, J. B. (2006). Climatology models for extreme hurricane winds near the united states. *Journal of Climate*, 19:3220–3236.
- Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81:158–171.
- Katz, R. W., Brush, G. R., and Parlange, M. B. (2005). Statistics of extremes: Modeling ecological disturbances. *Ecological Society of America*, 86:1124–1134.
- Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. *Elsevier Science Ltd.*, 25:12871304.
- Landaverde Vaca, J. G. (2005). Estimadores bayesianos en la distribución de valores extremos tipo gumbel. Master's thesis, Colegio de Postgraduados.
- Macafee, Allan W. & K. Wong, S. W. (2007). Extreme value analysis of tropical cyclone trapped-fetch wave. *Journal of Applied Meteorology and Climatology*, 46:1501–1522.
- McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27, No. 1:117–137.
- Mestre, Oliver & Hallegatte, S. (2009). Predictors of tropical cyclone numbers and extreme hurricane intensities over the north atlantic using generalized additive and linear models. *Journal of Climate*, 22:633–648.
- Monroy, B. S. (2010). Modelación de eventos extremos usando la distribución dagum. Master's thesis, Colegio de Postgraduados.
- Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics*. McGraw-Hill, Inc.
- Papadakis, E. N. and Tsionas, E. G. (2012). Bayesian analysis of extreme value regression. *Applied Economics Letters*, 19(17):1707–1710.

- Reyes Cervantes, H. J. (2008). *Estimación de Tendencias en Niveles Máximos de Contaminación usando Regresión por Cuantiles ajustando el efecto por variables meteorológicas*. PhD thesis, Colegio de Postgraduados.
- Rodríguez R., S., Reyes Cervantes, H. J., Vaquera Huerta, H., and Tajonar Sanabria, F. (2012). Estimación bayesiana para valores extremos de ozono en dos estaciones del simat. *Academia Journals*, 4, No.3:2599–2604.
- Scotto, M. G., Barbosa, S. M., and Alonso, A. M. (2011). Extreme value and cluster analysis of european daily temperature series. *Journal of Applied Statistics*, Vol. 38:2793–2804.
- Sexto Monroy, B., Vaquera Huerta, H., and Arnold, B. C. (2010). Use of the dagum distribution for modeling tropospheric ozone levels. *Journal of Environmental Statistics*, VV:140.
- Smith, R. L. & Naylor, J. C. (1987). A comparison of maximum likelihood and bayesian estimators for the three-parameter weibull distribution. *Applied Statistics*, 36:358–369.
- Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, 4, No. 4:367–377.
- Torres, J. (1992). Construcción y aplicación de trayectorias de retroceso de parcelas de aire al estudio de la contaminación atmosférica por ozono en la zona metropolitana de la ciudad de México. Master's thesis, División de Estudios de Postgrado, Facultad de Ingeniería, Universidad Nacional Autónoma de México.
- Velasco Luna, F. & Hernández González, S. (2007). Teoría de valores extremos: Una introducción. *Revista de Ciencias Básicas UJAT*, 6, Num. 1:10 – 16.
- Von Mises, R. (1936). La distribution de la plus grande de n valeurs. *Revue Mathématique de l'Union Interbalkanique*, 1:141–160.