



**UNIVERSIDAD
SERGIO ARBOLEDA**

ESTUDIO DE LA CORRELACIÓN DE CONTAMINANTES DEL AIRE DE BOGOTÁ.

Autor:

Michael Stiven Pinzón Rodríguez.

Universidad Sergio Arboleda

Facultad de ciencias exactas e ingeniería

Maestría en Matemáticas Aplicadas

Bogotá D.C - 2022



UNIVERSIDAD
SERGIO ARBOLEDA

ESTUDIO DE LA CORRELACIÓN DE
CONTAMINANTES DEL AIRE DE
BOGOTÁ.

Autor:

Michael Stiven Pinzón Rodríguez.

Tutora:

Biviana Marcela Suárez Sierra.

*Tesis de grado presentada como requisito parcial para optar
al título de:*

Magister en Matemáticas Aplicadas

Universidad Sergio Arboleda

Facultad de ciencias exactas e ingeniería

Maestría en Matemáticas Aplicadas

Bogotá D.C - 2022

Agradecimientos

En primer lugar, quiero agradecer a la Doctora Biviana Marcela Suárez Sierra, por su compromiso, paciencia y dedicación conmigo en la realización de este trabajo de grado, por haber confiado en mi y por compartirme sus ideas para poder llevar a cabo este trabajo. Por aceptar ser mi tutora aún sabiendo que no tenía conocimientos en programación y por haberme motivado a aprender al respecto para poderlo sacar adelante. También agradezco a la universidad Sergio Arboleda por permitirme la oportunidad de pertenecer a esta maestría y a los profesores, quienes aportaron innumerables cosas a mi profesión.

Contenido

Introducción	III
Agradecimientos	III
Bibliografía	VII
1. Preliminares	3
1.1. Inferencia bayesiana	3
1.2. Teorema de Bayes	4
1.3. Método Monte Carlo vía Cadenas de Markov (MCMC)[Jiménez (2015)]	6
1.3.1. Algoritmo de Metropolis Hastings	6
1.3.2. Muestrador de Gibbs	7
1.4. Algunos diagnósticos de convergencia para los parámetros estimados . .	8
1.4.1. Método Geweke [Geweke (1992)]	8
1.4.2. Método Heidelberger-Welch	8
1.4.3. Método Gelman and Rubin	9
1.4.4. Método Raftery and Lewis	10
1.5. Funciones cópula	11
1.6. Definición de función Cópula[Nelsen (2006) [2]]	11
1.7. Cópula Farlie-Gumbel-Morgenstern (FGM)	12
2. Modelos univariados para los contaminantes PM_{10}, $PM_{2.5}$ y O_3	14
2.1. Descripción de los datos	14
2.1.1. Material particulado de $10\mu m$ y $2.5\mu m$ (PM_{10}) y ($PM_{2.5}$) respec- tivamente.	14
2.1.2. Ozono (O_3)	16
2.2. Reglamentación	16
2.3. Datos de los tres contaminantes:	16
2.4. Sin puntos de cambio	19

2.5. Un punto de cambio	22
2.6. Dos puntos de cambio:	26
2.7. Política pública:	29
3. Diagnóstico de convergencias	31
3.1. Sin puntos de cambio	31
3.2. Un punto de cambio	31
3.3. Dos puntos de cambio	33
4. Modelos univariados para los pares de los contaminantes PM_{10}, $PM_{2.5}$ y O_3	40
4.1. Correlación de dos procesos estocásticos	40
4.1.1. Correlación, θ , entre los contaminantes PM_{10} y O_3	42
4.1.2. Correlación, θ , entre los contaminantes $PM_{2.5}$ y O_3	43
4.1.3. Correlación, θ , entre los contaminantes $PM_{2.5}$ y PM_{10}	44
5. Conclusiones y consideraciones	47
5.1. Conclusiones	47
5.2. Consideraciones	48

Índice de figuras

2.1. Comparación del tamaño del material particulado.	15
2.2. Niveles de concentración diaria de contaminantes O_3 , PM_{10} y $PM_{2.5}$, la recta horizontal representa el umbral máximo permitido por las autoridades ambientales en Colombia.	18
2.3. Media acumulada observada del número de excedencias para cada uno de los tres contaminantes.	19
2.4. Función de riesgo $\lambda(t)$, estimada para el modelo en ausencia de puntos de cambio.	21
2.5. Función de media acumulada (azul), junto con los intervalos de confianza estimados con un 95 % de confianza (verde) y los datos observados (negro).	22
2.6. Función de media acumulada (azul), junto con los intervalos de confianza estimados con un 95 % de confianza (verde) y los datos observados (negro), para un punto de cambio.	23
2.7. Función de riesgo $\lambda(t)$, estimada para el modelo construido bajo un punto de cambio.	25
2.8. Función de media acumulada (azul), junto con los intervalos de confianza estimados con un 95 % de confianza (verde) y los datos observados (negro), para dos puntos de cambio.	27
2.9. Función de riesgo $\lambda(t)$, estimada para el modelo construido bajo dos puntos de cambio.	28
4.1. Curvas de nivel para los contaminantes PM_{10} y O_3 . [12]	43
4.2. Curvas de nivel para los contaminantes $PM_{2.5}$ y O_3 . [12]	45
4.3. Curvas de nivel para los contaminantes PM_{10} y $PM_{2.5}$	46

Índice de tablas

2.1. Normativa establecida en la resolución 2254 para cada uno de los tres contaminantes.	16
2.2. Número de rebases de cada contaminante.	17
2.3. Información estadística de cada modelo sin puntos de cambio.	20
2.4. Información estadística de cada modelo con un punto de cambio.	24
2.5. Información estadística de cada modelo con dos puntos de cambio.	26
3.1. Información de convergencia para el modelo PM_{10} , sin puntos de cambio. H-W=Heidelberger-Welch	32
3.2. Información de convergencia para el modelo $PM_{2.5}$, sin puntos de cambio. H-W=Heidelberger-Welch	32
3.3. Información de convergencia para el modelo O_3 , sin puntos de cambio. H-W=Heidelberger-Welch	33
3.4. Información de convergencia para el modelo PM_{10} , un punto de cambio. H-W=Heidelberger-Welch	34
3.5. Información de convergencia para el modelo $PM_{2.5}$, un punto de cambio. H-W=Heidelberger-Welch	35
3.6. Información de convergencia para el modelo O_3 , un punto de cambio. H-W=Heidelberger-Welch	36
3.7. Información de convergencia para el modelo PM_{10} , dos puntos de cambio. H-W=Heidelberger-Welch	37
3.8. Información de convergencia para el modelo $PM_{2.5}$, dos puntos de cambio. H-W=Heidelberger-Welch	38
3.9. Información de convergencia para el modelo O_3 , dos puntos de cambio. H-W=Heidelberger-Welch	39
4.1. Prueba del modelo bivariado	41
4.2. Correlación entre los contaminantes PM_{10} y O_3	42
4.3. Correlación entre los contaminantes PM_{10} y O_3	44

4.4. Correlación entre los contaminantes PM_{10} y $PM_{2.5}$	44
---	----

Introducción

La asociación o correspondencia entre dos variables, ha sido estudiada desde distintos puntos de vista y probada por diferentes autores. De acuerdo con Farlie (1960) en [16] algunos de estos coeficientes de correlación son:

1. Coeficiente de correlación de Pearson.
2. Coeficiente de correlación de Spearman.
3. Coeficiente de correlación de Kendall τ
4. Probabilidad en concordancia.

Farlie (1960), afirma en [16] que la eficiencia de cada uno de estos coeficientes no es muy conocida, sin embargo, es que ninguno es uniformemente mejor que el otro. De manera particular, el coeficiente de correlación es una herramienta útil. Sólo en algunos casos, se utiliza la Cópula de Farlie-Gumbel-Morgenstern (FGM) para determinar o estimar el parámetro θ que representará la correlación entre dos parejas de contaminantes, estos son PM_{10} vs O_3 , $PM_{2.5}$ vs O_3 y $PM_{2.5}$ vs PM_{10} , para poder hacer esto, se hará de la siguiente manera. En el capítulo 1 se describirán algunos elementos de la inferencia bayesiana como el teorema de bayes, la creación de cadenas de Markov mediante el método de Monte Carlo (MCMC) y los algoritmos utilizados para lograr generar dichas cadenas.

También se presentan algunos criterios de convergencia para los parámetros obtenidos mediante las simulaciones por el método MCMC, esto para determinar su validez y generar confianza en el modelo, estos criterios de convergencia son el de Geweke, que analiza dos partes de las cadenas construidas, el 10 % inicial y el 50 % final, verificando que la diferencia de sus medias sea mínima, cuando esto sucede, los parámetros son aceptados. De la misma manera se estudian otros criterios de convergencia como lo son el de Heidelberger-Welch, el de Raftery and Lewis y el de Gelman & Rubin. También se hace una descripción detallada del concepto de cópula junto con sus propiedades y conceptos previos necesarios para entenderlas y al finalizar este capítulo se describe la

cópula utilizada para determinar la correlación de los tres pares de contaminantes.

En el segundo capítulo se abordan las características de los datos analizados. Para esto se inicia con una descripción de cada material contaminante junto con la reglamentación establecida por el ministerio de ambiente y desarrollo sostenible de Colombia. Esta descripción, está acompañada de los porcentajes en los cuales, durante el periodo comprendido entre el 2018 y el 2020, estos tres contaminantes analizados rebasan la norma establecida y se representa todo esto en tres series de tiempo diferentes construidas con el software R. Adicionalmente, se realiza el análisis de la gráfica de medias acumuladas que representa los datos de cada contaminante y se empieza a elaborar un modelo con el fin de encontrar los parámetros que se ajustan de la mejor forma a los datos conocidos. Se realizan tres modelos diferentes buscando un ajuste óptimo de los datos. El primero, **sin puntos de cambio**, no tiene en cuenta alguna afectación ocurrida durante los tres años, que muestre un cambio brusco en el comportamiento de los datos. En el segundo análisis, con **un punto de cambio**, se tiene en cuenta uno de los cambios mas representativos que hay en la serie de tiempo, para tomarlo como punto de cambio y hacer que el modelo cambie su comportamiento en ese punto dado. De igual manera en el tercer análisis **dos puntos de cambio**, se observa mejor el ajuste, dado que se configura el código programado para que asuma dos puntos específicos en los que hubo algún comportamiento drástico en los datos.

El capítulo tres, corresponde al diagnóstico de convergencia para verificar si los resultados obtenidos en el capítulo dos, eran óptimos o debían ser modificados, para cada uno de los tres modelos se realizaron los tests hasta obtener resultados favorables que permitieran parámetros adecuados. Estos resultados dan paso al cuarto capítulo, donde se toman como referencia los parámetros obtenidos para el primer ajuste, **sin puntos de cambio** y se comienza el análisis bivariado. Buscando, mediante la cópula utilizada, el parámetro que determinará la correlación entre cada par de contaminantes, aquí se presentan los resultados numéricos obtenidos sobre esta correlación, encontrando cosas particulares sobre cada par de contaminantes.

Finalmente, en el capítulo número cinco, se presentan algunas conclusiones de todo lo realizado en el desarrollo del presente trabajo de grado y se sugiere además la realización del modelo bivariado teniendo en cuenta los dos ajustes faltantes.

Capítulo 1

Preliminares

1.1. Inferencia bayesiana

En el presente trabajo de grado, son necesarios algunos resultados importantes del curso Inferencia Bayesiana, es por eso que a continuación se presentan algunas definiciones y teoremas que son de gran utilidad para poder encontrar los diferentes parámetros que modelan los contaminantes y su respectiva correlación.

~~El análisis de datos Bayesiano es el nombre que se le da a cada uno de los métodos prácticos que se construyen para hacer inferencias a partir de datos conocidos, utilizando modelos de probabilidad [9].~~ Gracias a la experiencia o al conocimiento previo sobre ~~algún tema particular,~~ se pueden construir modelos que permiten realizar inferencias y tomar decisiones futuras, siempre enfocados en ~~la mejora continua de los procesos.~~ A este conocimiento previo se le denomina distribución a priori $P(\theta)$, cuya definición es una representación por medio de distribuciones que dependen de los datos ya conocidos.

En el presente trabajo de grado, el interés está en encontrar conclusiones sobre la relación que tienen, dos a dos, de tres contaminantes elegidos y de los cuales se tiene una información previa que en este caso será la experiencia necesaria para hacer uso de la teoría Bayesiana y enfocar el objetivo principal que será identificar la correlación entre dos procesos estocásticos. Esto se podrá hacer, siempre y cuando se establezca un modelo para cada uno de los tres contaminantes y también para los pares, para hacerlo, se utiliza la siguiente notación, θ denota cantidades de vectores no observables o parámetros poblacionales de interés y X los datos observados o conocidos con lo cuales se construirá el modelo, para este caso, el número de excedencias de alguno de los tres contaminantes durante los tres años analizados.

En este sentido, la inferencia bayesiana se encargará de realizar conclusiones estadísticas sobre el parámetro θ a través de la probabilidad condicional, lo cual se escribe simplemente como $P(\theta|X)$ [9]. En adelante, la expresión $p(\cdot|\cdot)$ denota una densidad de probabilidad condicional cuyos argumentos se podrán inferir del contexto

1.2. Teorema de Bayes

En [9], Gelman, A. et al, hacen una descripción un tanto sencilla para comprender lo que aporta a esta investigación, el teorema de Bayes, pues mencionan que para formular enunciados de probabilidad sobre θ dado X se debe iniciar con un modelo que proporcione una distribución de probabilidad conjunta para θ y X . La función de densidad de probabilidad conjunta, como se menciona en [9] se puede escribir como el producto de dos densidades, la distribución a priori $P(\theta)$ y la distribución de muestreo $p(X|\theta)$ esto es:

$$p(\theta, X) = p(\theta)p(X|\theta) \quad (1.1)$$

Aquí, aplicando la definición de probabilidad condicional a la ecuación 1.1, se obtiene lo siguiente:

$$p(\theta|X) = \frac{p(\theta, X)}{p(X)} = \frac{p(\theta)p(X|\theta)}{p(X)} \quad (1.2)$$

En la cual $p(X) = \sum_{\theta} p(\theta)p(X|\theta)$ para el caso discreto y $p(X) = \int p(\theta)p(X|\theta)d\theta$ en el caso continuo de θ .

Una forma equivalente a la ecuación 1.2 no tiene en cuenta el factor $p(X)$, dado que este no depende de θ y con X fijo, se puede considerar una constante, esto produce la *densidad posterior no normalizada* [9]:

$$p(\theta|y) \propto p(\theta)p(y|\theta) \quad (1.3)$$

Antes de que los datos X sean considerados, la distribución de lo desconocido pero

observado X , es:

$$p(X) = \int p(X|\theta)d\theta = \int p(\theta)p(X|\theta)d\theta \quad (1.4)$$

La ecuación 1.4 es conocida como la *distribución marginal de X* sin embargo un nombre que la describe mejor es *distribución predictiva a priori*, priori, porque no esta condicionada a una observación previa del proceso y predictiva porque es la distribución para una cantidad que es observable.

Ya una vez se tienen los datos X observados, se pueden predecir datos observables desconocidos, \tilde{X} , la distribución de estos datos es llamada *distribución predictiva posterior*, posterior porque esta condicionada a los X observados y predictiva porque es una predicción para los \tilde{X} observables.

$$\begin{aligned} p(\tilde{X}|X) &= \int p(\tilde{X}, \theta|X)d\theta \\ &= \int p(\tilde{X}|\theta, X)p(\theta|X)d\theta \\ &= \int p(\tilde{X}|\theta)p(\theta|X)d\theta \end{aligned} \quad (1.5)$$

Cuando se observa X_1, X_2, \dots, X_n , estos datos ya no representan una variable aleatoria, de esta manera la distribución $p(x|\theta)$ se convierte en una función solo de θ . A esta función se le conoce como *verosimilitud* y se denota al igual que en [1] como $L(D|\theta)$.

Dado que el presente trabajo de grado es una consecuencia de la tesis de doctorado de Suarez, B. (2020) [1], la notación referente a los datos observados es la misma, por la tanto aquí también, si N es el número total de excedencias de alguno de los tres contaminantes estudiados, localizados en los días $D = d1, d2, \dots, d_N$ y θ el parámetro desconocido del modelo, entonces de la formula de Bayes se tiene qué:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \propto L(D|\theta)p(\theta) = \prod_{i=1}^N p(d_i|\theta)p(\theta) \quad (1.6)$$

1.3. Método Monte Carlo vía Cadenas de Markov (MCMC)[Jiménez (2015)]

Los métodos MCMC, cuyas siglas provienen de su nombre en inglés: *Markov Chain Monte Carlo*, son algoritmos que simulan el comportamiento de variables aleatorias para poder estimar parámetros desconocidos de las funciones de probabilidad de dichas variables. En el presente trabajo de grado se realizará la inferencia sobre la función aposteriori, $p(\theta|X)$

En [10] se presentan los siguientes dos pasos sobre la aplicación de las técnicas MCMC:

1. Generar una muestra X_1, \dots, X_n mediante una cadena de Markov cuya distribución estacionaria sea la buscada.
2. Tomar medias muestrales (integración Monte Carlo) y realizar inferencias sobre la muestra. Para construir estas cadenas de Markov, existen dos familias de algoritmos que son: El algoritmo de Metropolis-Hastings y el muestreo de Gibbs, los cuales serán descritos a continuación.

1.3.1. Algoritmo de Metropolis Hastings

Suárez, B. (2020), en [1] cita a Metropolis (1949), Hastings (1970) de la siguiente manera:

Sea $f(\cdot)$ una distribución de la cual se quiere generar valores, pero que no es posible hacerlo directamente. El algoritmo Metropolis-Hastings dice que apartir de una distribución condicional $q(y|x)$, de la que se puede generar valores facilmente, se crea una secuencia de observaciones X_1, X_2, \dots , cuya distribución estacionaria es $f(\cdot)$. El algoritmo es:

Sea $X_0 = x_0$ el valor inicial, generado a partir de una distribución inicial adecuada. Para $n = 0, 1, 2, \dots$, si el estado actual es $X_n = x$, se obtiene X_{n+1} así:

1. Se genera un valor candidato y usando $q(\cdot|X_n = x)$
2. Se evalúa $\alpha(x, y) = \min \left[\frac{f(y)q(x/y)}{f(x)q(y/x)}, 1 \right]$
3. Asigne el valor a X_{n+1} de la siguiente manera:

$$X_{n+1} = \begin{cases} y & \text{con probabilidad } \alpha(x, y) \\ x & \text{con probabilidad } 1 - \alpha(x, y) \end{cases}$$

Aquí, para ejecutar el paso 3, se genera U usando una distribución *uniforme*(0, 1) si $U < \alpha$, $X_{n+1} = y$; en otro caso $X_{n+1} = x$

1.3.2. Muestrador de Gibbs

Sea $X = X_1, X_2, \dots, X_{k-1}$ un vector aleatorio. El muestreador de Gibbs, arroja un resumen inferencia de la distribución conjunta $p(X) = p(X_1, X_2, \dots, X_k)$. Suponga que sea relativamente fácil generar valores de las funciones de densidad condicional completas denotadas por $p(X_1|X_2, \dots, X_k), p(X_i|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k)$, para $i = 2, \dots, k-1$ y $p(X_k|X_1, \dots, X_{k-1})$.

Dado un vector inicial $X^{(0)} = (X_1^{(0)}, X_2^{(0)}, \dots, X_k^{(0)})$ obtenido de acuerdo a alguna distribución adecuada, se implementa el siguiente proceso iterativo.

Para $n = 1, 2, \dots$, sea $X^{(n-1)} = (X_1^{(n-1)}, \dots, X_k^{(n-1)})$, entonces

1. Genere $X_1^{(n)}$ usando $P(\cdot|X_2^{(n-1)}, \dots, X_k^{(n-1)})$
2. Para $i = 2, 3, \dots, k-1$, genere $X_i^{(n)}$ usando $P(\cdot|X_1^{(n)}, \dots, X_{i-1}^{(n)}, X_{i+1}^{(n)}, \dots, X_k^{(n-1)})$
3. genere $X_k^{(n)}$ usando $P(\cdot|X_1^{(n)}, \dots, X_{k-1}^{(n)})$,
4. Haga $X^{(n)} = (X_1^{(n)}, \dots, X_k^{(n)})$ y regrese al ítem 1, con n en el lugar de $n-1$ y $n+1$ en lugar de n .

La sucesión $X^{(1)}, X^{(2)}, \dots$, con $X^{(i)} = (X_1^{(i)}, X_2^{(i)}, \dots, X_k^{(i)})$, $i = 1, 2, \dots$ es una realización de una cadena de Markov cuya distribución de transición está dada por:

$$\begin{aligned}
 P(X^{(t+1)}|X^{(t)}) &= p(X_1^{(t+1)}|X_2^{(t)}, \dots, X_k^{(t)}) \\
 &\times \prod_{i=2}^{k-1} p(X_i^{(t+1)}|X_1^{(t+1)}, \dots, X_{i-1}^{(t+1)}, X_{i+1}^{(t)}, \dots, X_k^{(t)}) \\
 &\times p(X_k^{(t+1)}|X_1^{(t+1)}, \dots, X_{k-1}^{(t+1)}),
 \end{aligned} \tag{1.7}$$

Lo cual bajo condiciones adecuadas converge a $f(\cdot)$ cuando n tiende a infinito.

1.4. Algunos diagnósticos de convergencia para los parámetros estimados

Como ya se mencionó, para la elaboración del presente trabajo de grado, se generan cadenas de Markov para estimar los parámetros deseados. En esta sección se presentan algunos criterios que permiten generar confianza en los modelos, pues se evalúa mediante su convergencia si son fiables o no. A continuación se presentan algunas de estas pruebas de convergencia:

1.4.1. Método Geweke [Geweke (1992)]

Geweke, J. (1992), [11] citado por Suárez, B. (2020) [1] brinda un diagnóstico de convergencia en el cual establece que si el comportamiento del 10 % inicial de la cadena de Markov inicial es igual al 50 % final, entonces el estadístico de Geweke tiene una distribución normal estándar asintótica, es decir, Z_n , es tal que cuando $n \rightarrow \infty$ su distribución es una normal con parámetros $(0, 1)$, aquí:

$$Z_n = \frac{\sqrt{T}(\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{r_A} + \frac{\sigma_B^2}{r_B}}} \quad (1.8)$$

Con T el número de iteraciones, μ_A , μ_B medias y σ_A^2 , σ_B^2 varianzas del primer y último segmento respectivamente, además $T_A = r_A T$ y $T_B = r_B T$ con r_A y r_B proporciones de la longitud de la cadena del primer y último segmento.

1.4.2. Método Heidelberger-Welch

Este modelo calcula un test estadístico para aceptar o rechazar la hipótesis nula de que la cadena de Markov corresponde a una distribución estacionaria. El método consiste de dos partes, en la primera se establece una cadena de N iteraciones y se analiza, mediante el estadístico de prueba, si la cadena es estacionaria o no, cuando la hipótesis nula es rechazada, se descarta el 10 % inicial de la cadena y se repite el proceso, esto se hace hasta que la hipótesis nula es aceptada o hasta llegar al 50 % de la cadena. Si no se puede aceptar la hipótesis nula se dice que la cadena falla y que se debe elegir un mayor número de iteraciones.

En la segunda parte del test, cuando la cadena o una parte de ella pasó la primera parte, se calcula el error estándar de la media y se calcula un intervalo de confianza

para la media.

Si la mitad del ancho de este intervalo de confianza construido es menor que un ϵ se puede decir que el tamaño de la nueva cadena es suficiente para estimar la media con precisión. [20]

1.4.3. Método Gelman and Rubin

En 1992, Gelman and Rubin plantearon una prueba de convergencia para determinar la veracidad de los parámetros estimados mediante cadenas de Markov. Este criterio de convergencia, consiste analizar n cadenas independientes, y esta basado en estimar las varianzas σ^2 de θ por medio de un estimador V , el cual es construido promediando los estimadores obtenidos en cada cadena de Markov. Matemáticamente hablando, El Adlouni et al. (2006), citando a Gelman and Rubin (2006), presentan en [17] una descripción del algoritmo:

Denotan, en primer lugar, $\theta_i^t = \theta(x_i^t)$ como la funcional evaluada en la t -ésima observación de la cadena i ,

$$\bar{\theta}_i = \frac{1}{n} \sum_{t=n+1}^{2n} \theta_i^t, \bar{\theta} = \frac{1}{m} \sum_{i=1}^m \bar{\theta}_i, s_i^2 = \frac{1}{n-1} \sum_{t=n+1}^{2n} (\theta_i^t - \bar{\theta}_i)^2 \quad (1.9)$$

A partir de lo anterior, en [17] definen el estimador de la siguiente manera:

$$V = \frac{n-1}{W} + \left(1 + \frac{1}{m}\right) \frac{B}{n} \quad (1.10)$$

Donde $\frac{B}{n}$ representa las medias de la secuencia m y es de la siguiente forma:

$$\frac{B}{n} = \frac{1}{m-1} \sum_{i=1}^m (\bar{\theta}_i - \bar{\theta})^2, \quad (1.11)$$

W , es la media de las m varianzas de la secuencia, s_i^2 y esta dada por:

$$W = \frac{1}{m} \sum_{i=1}^m s_i^2 \quad (1.12)$$

Con la información anterior, es posible construir un estimador R_c , el cual según Gelman and Rubin (1998), citados en [17], si tiene valores muy grandes, hay dos posibilidades.

1. La varianza V se puede disminuir para mas simulaciones.
2. W puede disminuir para mas simulaciones.

En cambio, si el estimador de Gelman y Rubin, se encuentra cerca de 1, se puede concluir que cada una de los m conjuntos de n simulaciones observadas esta cerca de la simulación objetivo.

1.4.4. Método Raftery and Lewis

Raftery and Lewis (1992), proponen un método apropiado para determinar el número de iteraciones necesarias para calcular un cuantil de la cadena de markov construida. Además analizan la longitud de ejecución para que la probabilidad resultante se encuentre en un intervalo $[q - \epsilon, q + \epsilon]$, esta precisión requerida ϵ se encuentra con una probabilidad $(1 - \alpha)$ [18]

En el software R, por medio de la función *Raftery.Diagnostic* se obtienen los resultados de este criterio de convergencia, uno de estos resultados es la matriz *resmatrix* en la cual se observan los resultados: burn-in sugerido, un número N que es el número de iteraciones realizadas, un N_{min} que es el número minimo de iteraciones sugeridas y un factor de dependencia I el cual es determinado de la siguiente manera: $I = \frac{N}{N_{min}}$, el cual tiene la siguiente interpretación: $I > 5$ indica que hay relación de dependencia entre cada uno de los estados de la cadena, sugiriendo entonces que se deben modificar algunos de los valores propuestos en el modelo. [1], [19]

1.5. Funciones cópula

El interés en el presente trabajo está, en determinar la correlación de dos procesos estocásticos relacionados con la contaminación del aire de la ciudad de Bogotá, para eso se hará uso de las funciones cópula, las cuales, dos a dos, darán un valor θ que determinará si la correlación es negativa, positiva o cero, arrojando resultados en el intervalo $[-1, 1]$

1.6. Definición de función Cópula[Nelsen (2006) [2]]

Las cópulas son funciones que determinan distribuciones multivariadas a partir de las distribuciones marginales de una dimensión.

Preliminares: Durante el proyecto se trabajará con la siguiente notación:

Se define \mathbb{R} como la recta real usual, y $\overline{\mathbb{R}}$ como la recta real $[-\infty, \infty]$, en este caso $\overline{\mathbb{R}}^2$ denota el plano real extendido $\overline{\mathbb{R}} \times \overline{\mathbb{R}}$. Un rectángulo en $\overline{\mathbb{R}}^2$ es el producto cartesiano B de dos intervalos cerrados:

$$B = [x_1, y_1] \times [x_2, y_2]$$

Se define también el cuadrado unitario \mathbf{I}^2 como el producto $\mathbf{I} \times \mathbf{I}$, donde $\mathbf{I} = [0, 1]$.

Definición 1: Una función real 2 – place H se define como una función cuyo dominio $DomH$ es un subconjunto de $\overline{\mathbb{R}}^2$ y cuyo rango, $RanH$ es un subconjunto de \mathbb{R} .

Definición 2: Sea S_1 y S_2 subconjuntos no vacíos de $\overline{\mathbb{R}}$, y sea H una función real "2 – place" tal que su dominio es $S_1 \times S_2$. Sea $B = [x_1, y_1] \times [x_2, y_2]$ un rectángulo cuyos vértices están en el $DomH$. Entonces el H – volume de B está dado por:

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1)$$

Definición 3: Una función real 2 – place es 2 – creciente si $V_H(B) \geq 0$ para todos los rectángulos B cuyos vértices están en el dominio de H .

Dada la información anterior, es posible ahora definir lo que son las Cópulas, pues son un concepto fundamental en el desarrollo del presente trabajo. Sin embargo para definir las, antes es necesario definir lo que es una subcópula:

Definición 4: Una subcópula 2 – dimensional es una función C con las siguientes propiedades:

1. $DomC' = S_1 \times S_2$, donde S_1 y S_2 son subconjuntos de \mathbf{I} que contienen a 0 y a 1;

2. C' es cerrado y 2 – *creciente*
3. Para cada u en S_1 y cada v en S_2 ,

$$C'(u, 1) = u \text{ y } C'(1, v) = v$$

Definición 5: Una cópula 2 – *dimensional* es una 2-*subcopula* C cuyo dominio es I^2 . De forma equivalente, una cópula es una función C de \mathbf{I}^2 en \mathbf{I} que satisface las siguientes propiedades.

1. Para cada u, v en \mathbf{I} ,

$$C(u, 0) = 0 = C(0, v)$$

y

$$C(u, 1) = u \text{ y } C(1, v) = v;$$

2. Para cada u_1, u_2, v_1, v_2 en \mathbf{I} tal que $u_1 \leq u_2$ y $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

Teorema 2.1 (Teorema de Sklar [15]). *Sea $F(\cdot, \cdot)$ una distribución conjunta con distribuciones marginales $F_1(\cdot)$ y $F_2(\cdot)$ respectivamente. Entonces, existe una cópula $C(\cdot, \cdot)$ tal que:*

$$F(x, y) = C(F_1(x), F_2(y)), \text{ para todo } x, y \in (-\infty, \infty) \quad (1.13)$$

La demostración de este teorema se puede encontrar en [15].

En [1], Suarez (2020), hace uso de las funciones cópulas para determinar de manera detallada la función de riesgo bivariada $\lambda(t_1, t_2)$, la función de media acumulada (aproximada) $M(t_1, t_2)$ y la función de riesgo acumulada bivariada $m(t_1, t_2)$.

1.7. Cópula Farlie-Gumbel-Morgenstern (FGM)

Existen diferentes cópulas propuestas, dependiendo de la situación en cuestión, para este estudio, al igual que en [1] se utilizará la copula propuesta por Morgenstern (1956), citado por Fraile(1960) en [16] cuyas marginales son Weibull.

Distribución bivariada Weibull:

$$F(x)G(x)[1 + \theta\{1 - F(x)\}\{1 - G(x)\}] \quad (1.14)$$

En la cual, $F(x), G(x) \in [0, 1]$ y $\theta \in [-1, 1]$

Suárez. B, (2020) en [1] en el capítulo 3.2 realiza una descripción detallada de la manera como se obtienen las funciones mencionadas anteriormente. Esto lo hace inicialmente, partiendo de que las funciones $F(x) = F_1(t_1), G(x) = F_2(t_2)$ sean las funciones marginales definidas de la siguiente manera:

$$F_i(t_i) = 1 - \exp \left\{ - \left(\frac{t_i}{\sigma_i} \right)^{\alpha_i} \right\}, \text{ con } \alpha_i, \sigma_i > 0, i = 1, 2. \quad (1.15)$$

De esta manera, al sustituir las marginales 1.15 en la ecuación 1.14, se obtiene:

$$\begin{aligned} F(t_1, t_2) = & \left(1 - \exp \left\{ - \left(\frac{t_1}{\sigma_1} \right)^{\alpha_1} \right\} \right) \left(1 - \exp \left\{ - \left(\frac{t_2}{\sigma_2} \right)^{\alpha_2} \right\} \right) \\ & \times \left(1 + \theta \exp \left\{ - \left(\frac{t_1}{\sigma_1} \right)^{\alpha_1} - \left(\frac{t_2}{\sigma_2} \right)^{\alpha_2} \right\} \right) \end{aligned} \quad (1.16)$$

Función de densidad conjunta Weibull:

$$f(t_1, t_2) = f_1(t_1)f_2(t_2)[1 + \theta[1 - 2F_1(t_1)][1 - 2F_2(t_2)]] \quad (1.17)$$

Donde $f_i(\cdot)$ es la función de densidad asociada a $F_i(\cdot), i = 1, 2$.

Función de riesgo acumulada:

La ecuación 1.18 es una de las más importantes en el presente documento, pues es la cópula que determina el modelo de programación utilizado para determinar los parámetros de correlación entre los dos contaminantes.

Para llegar a ella, Suárez. B, (2020), muestra detalladamente en [1] la manera de obtenerla y realiza además su demostración.

$$\lambda(t_1, t_2) = \left(\frac{\alpha_1}{\sigma_1} \right) \left(\frac{\alpha_2}{\sigma_2} \right) \left(\frac{t_1}{\sigma_1} \right)^{\alpha_1-1} \left(\frac{t_2}{\sigma_2} \right)^{\alpha_2-1} \frac{\left[1 + \theta \left(1 - 2 \exp \left\{ - \left(\frac{t_1}{\sigma_1} \right)^{\alpha_1} \right\} \right) \left(1 - 2 \exp \left\{ - \left(\frac{t_2}{\sigma_2} \right)^{\alpha_2} \right\} \right) \right]}{\left[1 + \theta \left(1 - \exp \left\{ - \left(\frac{t_1}{\sigma_1} \right)^{\alpha_1} \right\} \right) \left(1 - \exp \left\{ - \left(\frac{t_2}{\sigma_2} \right)^{\alpha_2} \right\} \right) \right]} \quad (1.18)$$

Capítulo 2

Modelos univariados para los contaminantes PM_{10} , $PM_{2.5}$ y O_3

2.1. Descripción de los datos

Para el estudio de la correlación de los contaminantes del aire de Bogotá, se trabajará con las concentraciones de 1096 días, correspondientes a los años 2018 a 2020, de material particulado de 10 micrómetros de tamaño, PM_{10} , material particulado de 2.5 micrómetros de tamaño, $PM_{2.5}$ y Ozono, O_3 .

2.1.1. Material particulado de $10\mu m$ y $2.5\mu m$ (PM_{10}) y ($PM_{2.5}$) respectivamente.

El material particulado (PM) es una mezcla de partículas sólidas y líquidas que se encuentran en el aire y que son imposibles de observar a simple vista.

Esta contaminación por partículas contiene, entre otros, los dos analizados en el presente estudio:

1. PM_{10} que son partículas que el ser humano puede inhalar, de 10 o menos μm de diámetro.
2. $PM_{2.5}$ al igual que el anterior, es un contaminante de acceso respiratorio de 2.5 o menos μm de diámetro.

En [12], la Agencia de protección Ambiental de Estados Unidos, presenta una comparación del tamaño de estas partículas, por ejemplo, el diámetro del grosor de un cabello humano promedio, es de $50\mu m$ a $70\mu m$, un grano de arena fina de playa, $90\mu m$ esta

comparación la realizan mediante la ilustración de la figura 2.1, aquí, de color azul se observa el tamaño aproximado del PM_{10} y de color fucsia el tamaño del $PM_{2.5}$, claramente se ve la diferencia y es obvio el porqué son imperceptibles a simple vista.



Comparación de tamaño de las partículas de PM

Figura 2.1: Comparación del tamaño del material particulado.

De acuerdo con la organización mundial de la salud (2018), la contaminación del aire aumenta cada día más, en países de bajo y medio ingreso, afectando la salud de los niños, principalmente. En el año 2016 las muertes de niños a causa de enfermedades respiratorias como consecuencia de la contaminación atmosférica, ascendieron a las 600.000 en todo el mundo.

En [13], se menciona que una de las principales causas de que los niños sean los que mas sufren por la contaminación atmosférica, es porque ellos respiran mas rápido que los adultos, razon por la cual, absorben mas micropartículas.

A causa de lo anterior y de los descubrimientos sobre cancer de pulmón causados por el material particulado, en [14] la organización mundia de la salud, establece algunas medidas que se deben tomar en todos los países para contrarestar esta situación y evitar que la contaminación continúe cobrando vidas.

Estas medidas establecidas estan relacionadas con la toma de acciones y la creación de políticas públicas que eviten la contaminación y reglamenten el uso adecuado de los recursos en las industrias públicas y privadas para disminuir su aporte de contaminación.

2.1.2. Ozono (O_3)

El Ozono, es un gas, que al rededor del ser humano puede generar daños a la salud respiratoria. Exposiciones por horas continuas a altas concentraciones de este contaminante pueden reducir la capacidad pulmonar, inflamar las vías respiratorias, producir infecciones respiratorias, etc.

2.2. Reglamentación

Para el presente análisis, es necesario estudiar la reglamentación establecida por el ministerio de ambiente y desarrollo sostenible de Colombia, el cual, en la resolución N 2254 del 1 de noviembre de 2017: *Por la cual se adopta la norma del aire y se dictan otras disposiciones*, establece que a partir del 1 de julio del 2018 los niveles máximos permisibles de contaminantes criterio en el aire son los que se presentan en la tabla 2.1:

Contaminante	Nivel máximo permisible
PM_{10}	$75\mu g/m^3$
$PM_{2.5}$	$37\mu g/m^3$
O_3	$100ppb$

Tabla 2.1: Normativa establecida en la resolución 2254 para cada uno de los tres contaminantes.

2.3. Datos de los tres contaminantes:

Estos 1096 datos diarios de los tres contaminantes, fueron obtenidos a partir de los datos horarios registrados en las 13 estaciones de monitoreo de calidad del aire de Bogotá. Para el caso del PM_{10} y $PM_{2.5}$ se determinaron promedios por cada 24 horas y de estos promedios se elegía el valor máximo; para el caso del O_3 , algo similar, solo que, en lugar de promedio se tomo el máximo de cada periodo de un día.

A continuación, se verifica mediante el software R, la cantidad de veces que estos umbrales se rebasan en los 1096 días de los años 2018 a 2020, en la figura 2.2 se muestran las series de tiempo que describen el comportamiento diario de los contaminantes durante los tres años analizados.

Para el caso del PM 10, sucede algo particular y es que sus valores máximo y mínimo suceden en el mismo mes, aquí se observa que el día 24 de junio del 2020, se presentó

una medición de 172.37, sobrepasando considerablemente el máximo permitido y para el día 15 de junio del mismo año, se obtuvo una medición mínima en los tres años de 14.06. Para este contaminante, se observaron un total de 318 rebases.

En cuanto al material particulado $PM_{2.5}$ se observó que este rebasaba los umbrales permitidos un 30,3 % de veces, con un total de 332 días en los que la norma no se estaba cumpliendo, se puede observar en la figura 2.2 que el día 28 de marzo del 2019, este contaminante sobrepasó el umbral permitido en más del 200 %.

Para la serie de tiempo que muestra los máximos diarios de concentración de Ozono en la ciudad de Bogotá, se debe tener en cuenta que, en la resolución número 2254 del ministerio de ambiente y desarrollo sostenible, se presenta el umbral de este contaminante medido en $\mu g/m^3$ y en las estaciones de monitoreo se recibe la información en partes por billón (*ppb*), de manera que se debe realizar la revisión de rebases, cuando los datos recogidos superen el valor aproximado de 50,9375 *ppb* que corresponde a los $100\mu g/m^3$ que establece la norma, esto nos da un total de 219 rebases en los tres años analizados.

En la tabla 2.2 se muestra un resumen de los rebases obtenidos para cada uno de los contaminantes de interés, teniendo en cuenta que para el O_3 los rebases se determinaron sobre un valor de 50.93 luego de realizar la respectiva conversión de partes por billón a $\mu g/m^3$.

En la figura 2.3 se presentan los graficos de las medias acumuladas de los rebases de los umbrales máximos permitidos. En ellos se puede evidenciar que hubo meses en los cuales no se presentaron rebases, esto se observa cuando las gráficas se muestran como una constante, un ejemplo es el caso del Ozono, en los meses comprendidos entre marzo y julio de 2018, no hubo ni un solo rebase, así como también entre los meses de abril y agosto del año 2019. De esta misma manera ocurre con los otros dos contaminantes, sin embargo es el Ozono, en el cual los periodos en los que no se rebasan los umbrales son mas extensos.

Contaminante	Umbral	Rebases	Porcentaje
PM_{10}	$70\mu g/m^3$	318	29 %
O_3	100 <i>ppb</i>	219	20 %
$PM_{2.5}$	$37\mu g/m^3$	332	30.3 %

Tabla 2.2: Número de rebases de cada contaminante.

Para estudiar el número de rebases de los umbrales ambientales, se considera un proceso de Poisson con función de intensidad $\lambda(t)$, con la forma Weibull, que depende de los parámetros α y β de acuerdo con la siguiente expresión:

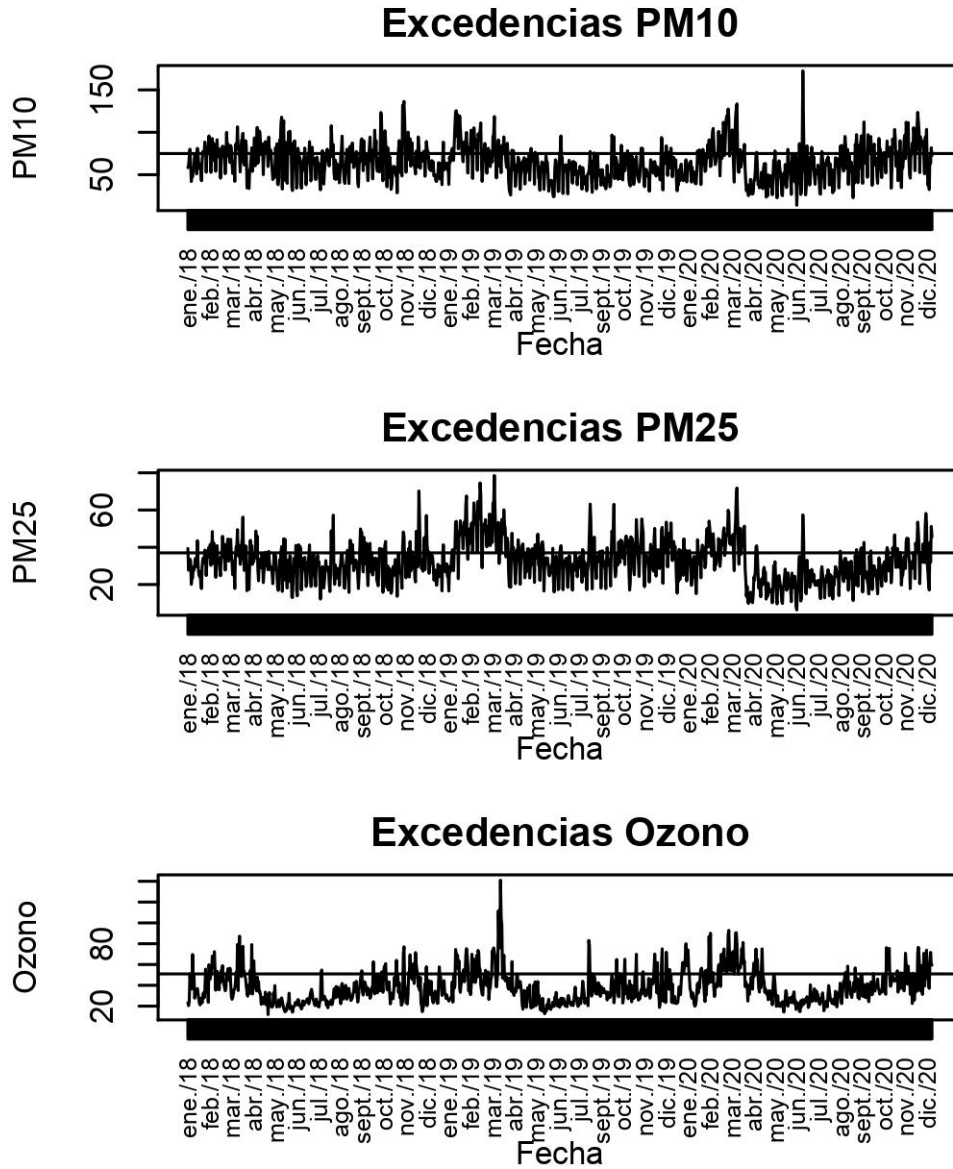


Figura 2.2: Niveles de concentración diaria de contaminantes O_3 , PM_{10} y $PM_{2.5}$, la recta horizontal representa el umbral máximo permitido por las autoridades ambientales en Colombia.

$$\lambda(t) = \frac{\alpha}{\beta} \left(\frac{t}{\beta} \right)^{\alpha-1}, \alpha, \beta > 0 \text{ y } t \geq 0$$

A partir de esta expresión se construyen las funciones de riesgo para cada uno de los modelos, teniendo en cuenta que cuando el parámetro $\alpha < 1$ la función es decreciente, si $\alpha > 1$ será creciente y constante cuando $\alpha = 1$

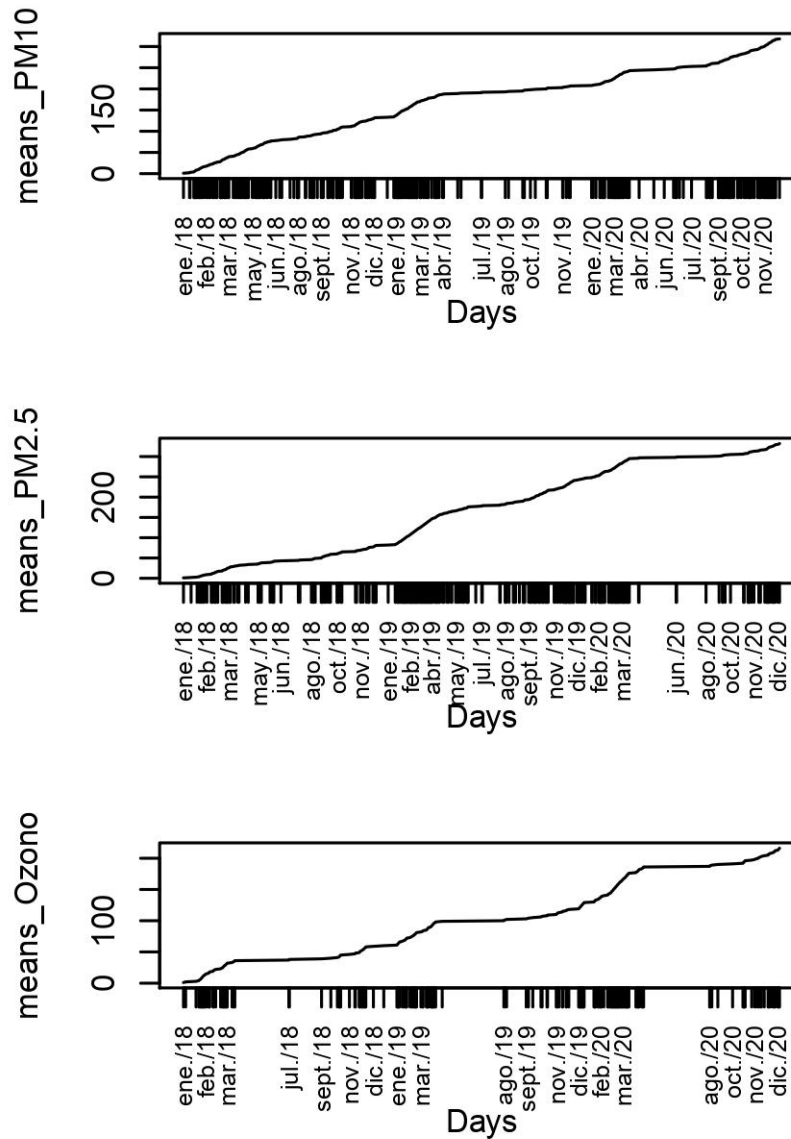


Figura 2.3: Media acumulada observada del número de excedencias para cada uno de los tres contaminantes.

2.4. Sin puntos de cambio

Para estimar los parámetros de este modelo, se realizaron 20.000 iteraciones y se eliminaron 2.000 de ellas. En esta implementación se encontró que para el PM_{10} , los parámetros son $\alpha = 0.89$ y $\beta = 1.76$, lo cual indica que la función de riesgo será decreciente, esto lo que quiere decir, es que el número de rebases tiende a disminuir con el paso del tiempo.

Para el caso del $PM_{2.5}$, se obtuvo: $\alpha = 1.16$ y $\beta = 1.75$ y para el ozono, $\alpha = 1.21$

y $\beta = 13.37$, estos últimos valores de α demuestran que las funciones son crecientes, de manera que con el paso del tiempo los niveles de contaminación de $PM_{2.5}$ y O_3 aumentarán.

En ausencia de puntos de cambio, se determinaron los siguientes parámetros α y β con los cuales se podrán construir las funciones de riesgo acumuladas:

	Información estadística				
	Parámetros	dist. inicial	Media	sd	intervalo 95 %
PM_{10}	α_1	$unif(1, 3)$	0.89	0.05	(0.79; 0.99)
	β_1	$unif(1, 10)$	1.76	0.65	(0.77; 3.28)
$PM_{2.5}$	α_1	$dunif(1, 2.8)$	1.25	0.02	(1.22; 1.31)
	β_1	$dunif(10; 100)$	10.85	0.83	(10.02; 13.06)
O_3	α_1	$dunif(1, 5.99)$	1.21	0.03	(1.16; 1.29)
	β_1	$dunif(11.9, 141)$	13.37	1.41	(11.94; 17.12)

Tabla 2.3: Información estadística de cada modelo sin puntos de cambio.

Teniendo estos parámetros, en la figura 2.4 se presentan las diferentes funciones de riesgo $\lambda(t)$ para cada uno de los tres contaminantes, esto sin tener en cuenta algún punto de cambio.

El comportamiento de los rebases de cada contaminante se puede observar a través de la función de media acumulada $m(t)$, observada y estimada. En nuestro caso se observa en la figura 2.5, la función de media acumulada estimada en color azul, frente a lo estimado con un intervalo de confianza del 95 % de credibilidad, lo cual se representa mediante las líneas verdes, buscando hacer un ajuste y contraste con lo observado, que se presenta de color negro.

Se puede observar, que para el caso del PM_{10} , el modelo inicia con un buen ajuste, sin embargo para los meses posteriores a enero del 2019, se ve un desajuste considerable. Caso contrario el del $PM_{2.5}$ cuya estimación no es favorable en la mayoría de los meses estudiados, al igual que en el caso del ozono, en el cual se ve un buen ajuste durante los tres primeros meses y durante el mes de abril del año 2020.

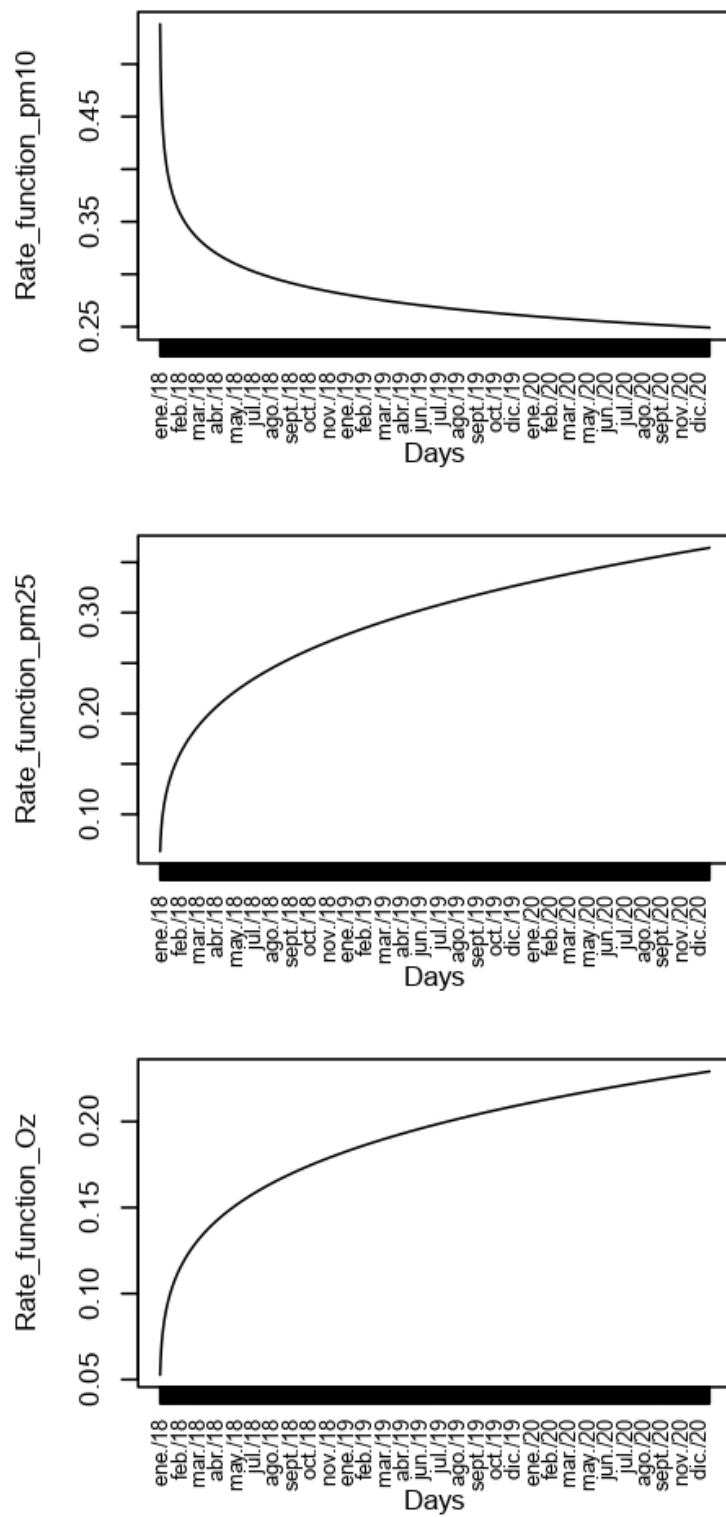


Figura 2.4: Función de riesgo $\lambda(t)$, estimada para el modelo en ausencia de puntos de cambio.

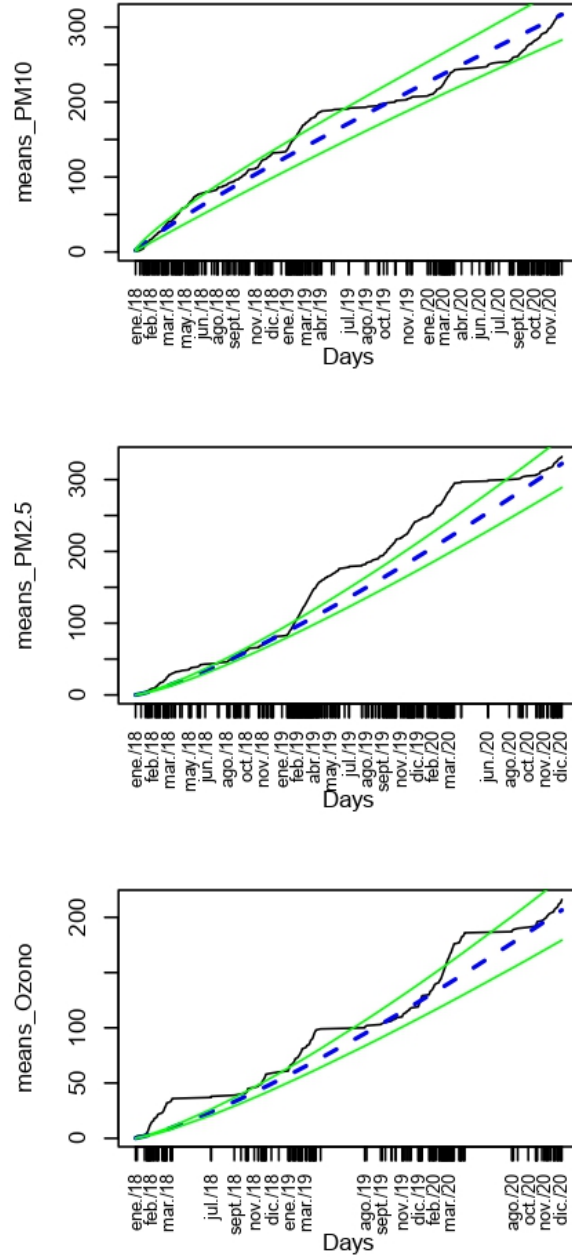


Figura 2.5: Función de media acumulada (azul), junto con los intervalos de confianza estimados con un 95 % de confianza (verde) y los datos observados (negro).

2.5. Un punto de cambio

Este primer punto se encontró gráficamente, observando cambios en la figura 2.3, e identificando acontecimientos sucedidos en las fechas donde se observan dichos cambios. Estos puntos pueden ser aproximadamente los siguientes: $\tau_1 = 380$, $\tau_2 = 415$ y $\tau_3 = 115$, para PM_{10} , $PM_{2.5}$ y O_3 respectivamente.

Para estos modelos se tendrán en cuenta dos tramos de las gráficas de medias acumuladas, que estarán determinados por el punto de cambio elegido. Es decir que antes de cada punto de cambio se determinarán un α y un β y para después del punto de cambio, otros α y β diferentes. Estos modelos se presentan en la figura 2.6

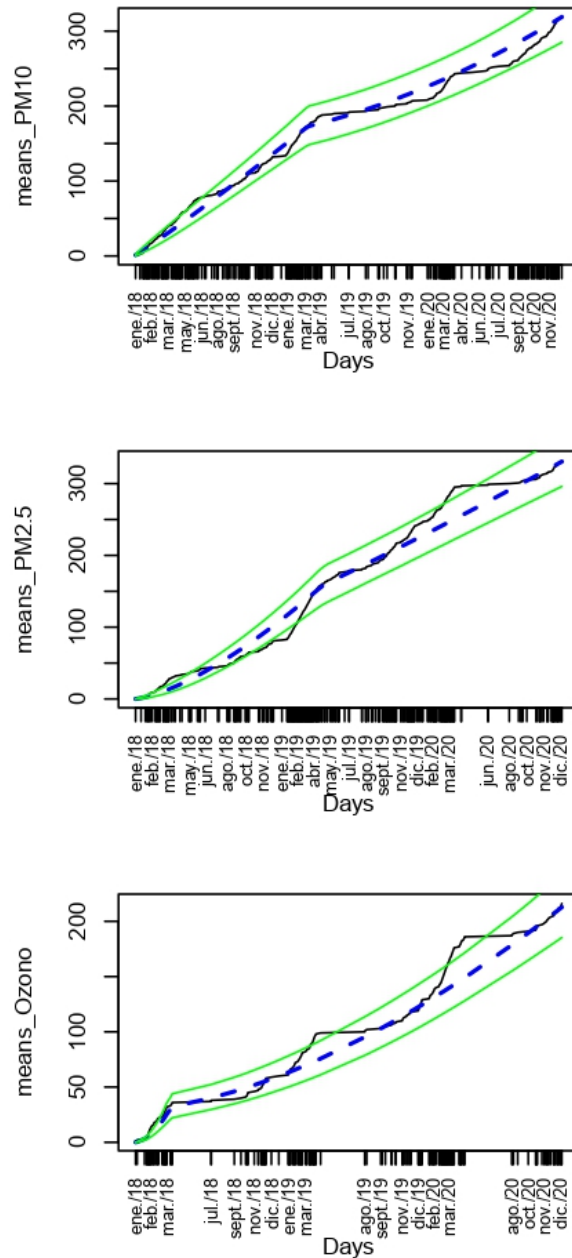


Figura 2.6: Función de media acumulada (azul), junto con los intervalos de confianza estimados con un 95 % de confianza (verde) y los datos observados (negro), para un punto de cambio.

Los parámetros estimados para este modelo, fueron obtenidos realizando 20.000 itera-

ciones, al igual que en sin un punto de cambio. Se observa aquí que los parámetros α para el contaminante PM_{10} fueron mayores que 1, lo que indica que el comportamiento del contaminante, tiende a aumentar su concentración con el paso del tiempo. De la misma forma, para los contaminantes $PM_{2.5}$ y O_3 ambos parámetros son mayores que 1, de manera que, así como se observó en el modelo sin puntos de cambio, antes y después del punto establecido, los contaminantes mantienen la predicción de aumento con el paso del tiempo. Estos resultados se encuentran resumidos en la tabla 2.4 y la gráfica correspondiente a las funciones de riesgo se muestran en la figura

	Información estadística				
	Parámetros	dist. inicial	Media	sd	intervalo 95 %
PM_{10}	τ_1	$dunif(350, 450)$	445.13	4, 34	(432.79; 449.89)
	α_1	$dunif(1, 3)$	1.06	0.08	(0.91; 1.21)
	β_1	$dunif(1, 100)$	3, 63	1.25	(1.56; 6.48)
	α_2	$dunif(1, 3)$	2.01	0.13	(1.68; 2.17)
	β_2	$dunif(1, 100)$	84.23	13.54	(49.34; 99.55)
$PM_{2.5}$	τ_1	$dunif(350, 500)$	483.53	9.44	(465.86; 498.37)
	α_1	$dunif(1, 2.5)$	1.45	0.11	(1.25; 1.68)
	β_1	$dunif(6.4, 78.5)$	14.90	4.04	(8.19; 23.72)
	α_2	$dunif(1, 2.5)$	1.28	0.15	(1.10; 1.65)
	β_2	$dunif(6.4, 78.5)$	15.21	9.15	(6.58; 41.13)
O_3	τ_1	$dunif(50, 200)$	96.40	5.89	(83, 56; 104, 13)
	α_1	$dunif(1, 6)$	1.83	0.17	(1.57; 2.24)
	β_1	$dunif(11.9, 141)$	14.69	2.35	(11.99; 20.71)
	α_2	$dunif(1, 5.9)$	1.66	0.15	(1.38; 1.97)
	β_2	$dunif(11.9, 141)$	48.45	13.98	(24.36; 79.27)

Tabla 2.4: Información estadística de cada modelo con un punto de cambio.

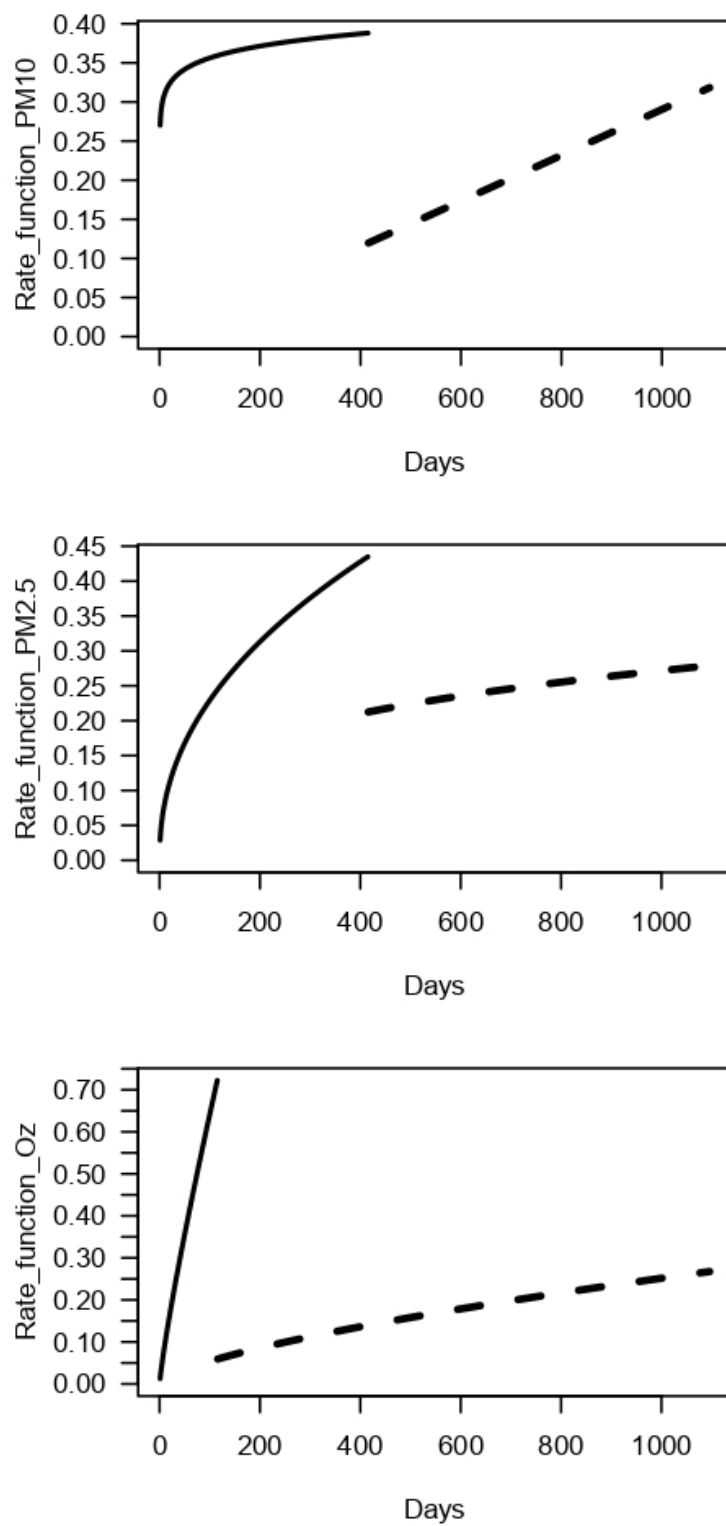


Figura 2.7: Función de riesgo $\lambda(t)$, estimada para el modelo construido bajo un punto de cambio.

2.6. Dos puntos de cambio:

Al igual que en la sección anterior, de manera que se eligieron dos puntos de cambio que visualmente parecen ser representativos y en los cuales, según las fechas, hubo implementaciones de normas para combatir la contaminación ambiental, estos son por ejemplo, para el caso del contaminante PM_{10} los días $\tau_1 = 446$ y $\tau_2 = 754$ aproximadamente, para el $PM_{2.5}$ los días $\tau_1 = 460$ y $\tau_2 = 821$ y para el O_3 los días $\tau_1 = 99$ y $\tau_2 = 468$ Abril del 2018 [4], Marzo del 2019[3], Enero del 2020 y marzo del 2020 [5] y [6].

	Información estadística				
	Parámetros	dist. inicial	Media	sd	intervalo 95 %
PM_{10}	τ_1	$dunif(250, 450)$	446.22	3.63	(437.42; 449.93)
	τ_2	$dunif(700, 900)$	754.19	18.09	(726.01; 777.40)
	α_1	$dunif(0, 3)$	1.07	0.08	(0.92; 1.23)
	β_1	$dunif(0, 100)$	3.71	1.32	(1.61; 6.74)
	α_2	$dunif(0, 3)$	1.33	0.32	(0.74; 1.23)
	β_2	$dunif(0, 100)$	37.39	27.07	(1.43; 94.18)
	α_3	$dunif(0, 3)$	1.89	0.26	(1.24; 2.21)
	β_3	$dunif(0, 100)$	66.40	24.67	(11.47; 98.78)
$PM_{2.5}$	τ_1	$dunif(300, 500)$	460.23	43.77	(368.97; 498.00)
	τ_2	$dunif(698, 898)$	821.60	3.36	(817.50; 829.87)
	α_1	$dunif(1, 3.5)$	1.37	0.18	(1.02; 1.67)
	β_1	$dunif(1, 100)$	13.19	4.94	(5.02; 23.27)
	α_2	$dunif(1, 3.5)$	2.05	0.47	(1.06; 2.50)
	β_2	$dunif(1, 100)$	66	33.44	(2.94; 99.07)
	α_3	$dunif(1, 3.5)$	1.66	0.19	(1.18; 1.92)
	β_3	$dunif(1, 100)$	71.063	22.56	(18.20; 99.07)
O_3	τ_1	$dunif(50, 200)$	99.18	1.81	(97.51; 104.09)
	τ_2	$dunif(350, 500)$	468.10	4.02	(462.65; 479.67)
	α_1	$dunif(1, 6)$	1.83	0.16	(1.57; 2.20)
	β_1	$dunif(12, 141)$	14.64	2.25	(12.08; 20.35)
	α_2	$dunif(1, 6)$	3.30	0.19	(2.86; 3.59)
	β_2	$dunif(12, 141)$	131.81	8.48	(109.41; 140.74)
	α_3	$dunif(1, 6)$	2.12	0.21	(1.62; 2.41)
	β_3	$dunif(12, 141)$	107.05	25.23	(47.89; 139.68)

Tabla 2.5: Información estadística de cada modelo con dos puntos de cambio.

Los resultados de utilizar estos puntos de cambio correspondientes a las siguientes fechas aproximadas, Abril del 2018, Marzo del 2019, Enero del 2020 y marzo del 2020, se presentan en la figura 2.8, aquí se observa la manera en que el ajuste es cada vez mas perfecto.

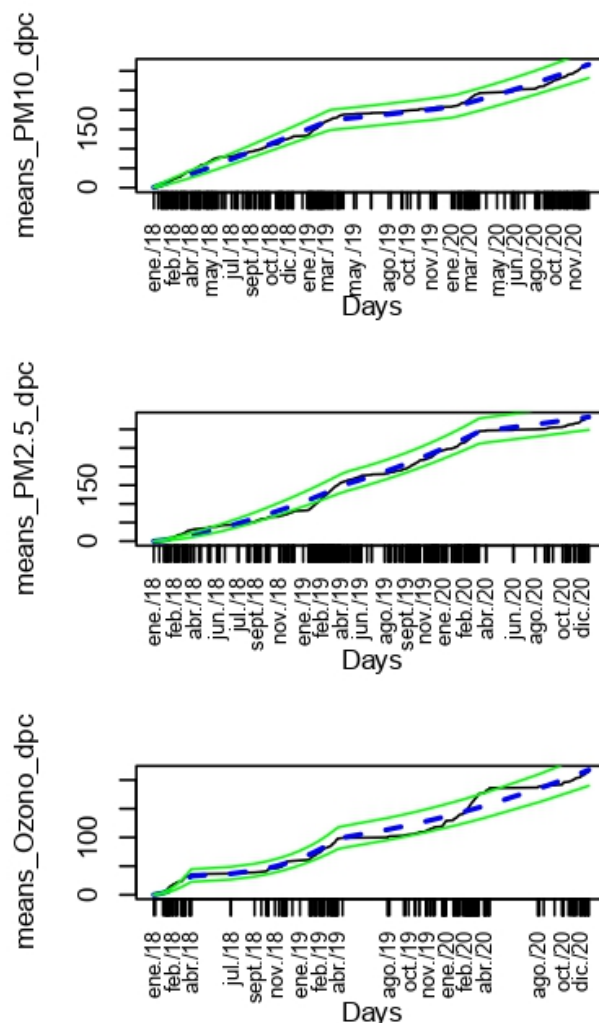


Figura 2.8: Función de media acumulada (azul), junto con los intervalos de confianza estimados con un 95 % de confianza (verde) y los datos observados (negro), para dos puntos de cambio.

Con los parámetros obtenidos para cada uno de los intervalos de tiempo determinados por los dos puntos de cambio, tabla 2.5, se determinan las funciones de riesgo que describen el comportamiento de los tres contaminantes en cada periodo de tiempo. En la figura 2.9 se observa qué las medidas implementadas por la alcaldía de Bogotá funcionan, pero de manera momentánea, por ejemplo para el caso del PM_{10} se observa como los niveles de contaminación en el primer tramo son altos, y luego para el segundo momento disminuyen y tienden a permanecer 'constantes'; sin embargo para el momento que sigue, la implementación de restricciones de movilidad y control de contaminantes en grandes industrias de febrero y marzo del 2020, cuando se declaró emergencia ambiental, ocasionaron, según la grafica, que el nivel de concentración de PM_{10} empezara a

aumentar y a mostrar una tendencia creciente.

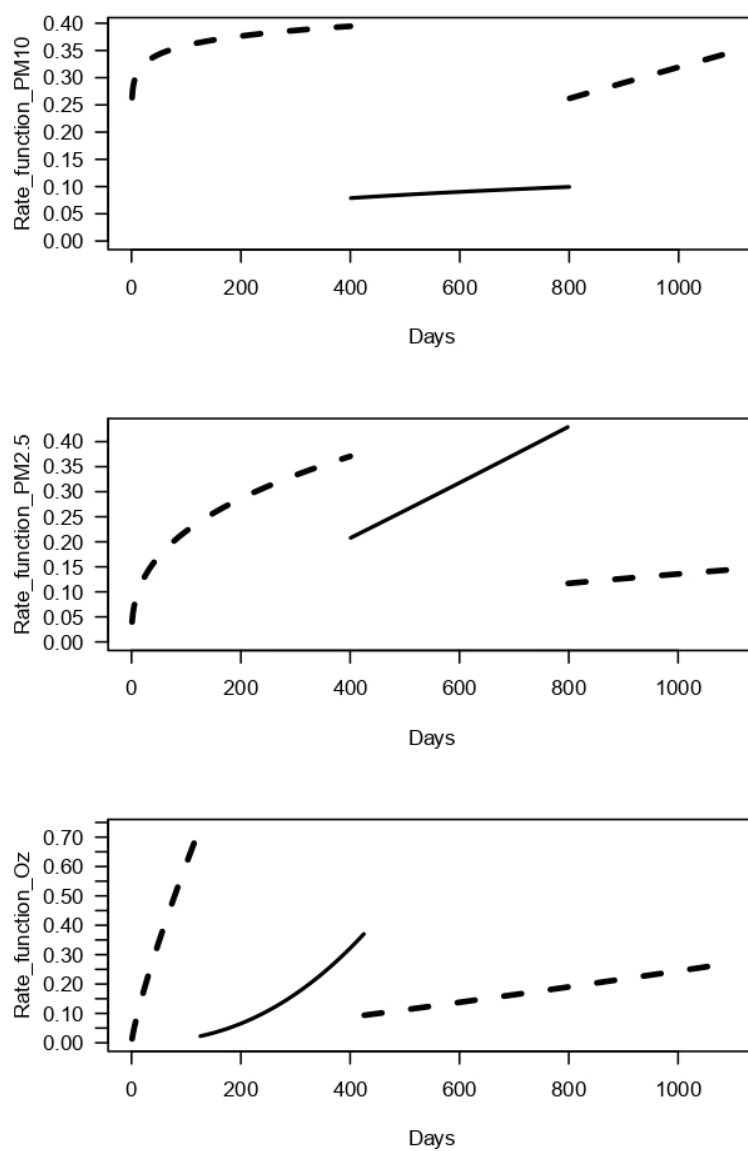


Figura 2.9: Función de riesgo $\lambda(t)$, estimada para el modelo construido bajo dos puntos de cambio.

2.7. Política pública:

Como ya se mencionó, los puntos de cambio de los modelos construidos se determinaron principalmente con ayuda de las gráficas de las medias acumuladas de cada uno de los contaminantes, para validar estos puntos, una posibilidad es identificar algunas de las estrategias del gobierno nacional para mitigar la contaminación ambiental durante estas fechas. Uno de los insumos mas importantes que permite identificar esto, es el *plan decenal de desconaminación del aire de Bogotá* que se elabora cada 10 años y al cual se le realizan revisiones periódicas para verificar su alcance y cumplimiento. El documento que compete a esta investigación, es el elaborado para la década del 2010 al 2020 [8], en el cual, al igual que en la anterior década, se centran en la implementación de las siguientes acciones para mejorar la calidad del aire:

- Pico y placa ambiental.
- Pico y placa de movilidad.
- Mejoramiento del ACPM.
- Operativos en Vía.
- Control a fuentes industriales.
- Inventarios de emisiones.

Estas acciones, se organizaron para ser implementadas durante los 10 años en diferentes etapas, para el caso del periodo de tiempo que se esta teniendo en cuenta en el presente documento, las acciones planteadas son, para el 2019 y 2020 únicamente, en el sector industrial, se estableció el proyecto *Uso de sistemas de control de emisiones*, para los años 2012 al 2020, en cuestiones de transporte, se debían implementar los proyectos *Uso de sistemas de control de emisiones en vehículos de transporte de carga*, y *Uso de sistemas de control de emisiones en motocicletas* y para el periodo comprendido entre 2011 y 2020 el proyecto *Implementación del sistema integrado de transporte público*.

En la ciudad de Bogotá, se han reducido los niveles de concentración de PM_{10} y $PM_{2.5}$ entre los años de 2012 y 2017 gracias a las propuestas implementadas en el plan decenal de decontaminación del aire para Bogotá, como por ejemplo la integración del transporte público SITP y la discontinuación de algunas busetas que no cumplían con las características establecidas para contrarrestar la contaminación; el seguimiento y control en la industria, aumento en los días sin carro, entre otras. Puntualmente, para

el caso del PM_{10} y el $PM_{2.5}$ su primer punto de cambio ocurre al rededor del dato 400, el cual corresponde a una fecha aproximada al 4 de febrero del 2019, este cambio se puede explicar en parte, por la resolución 383 del año 2019 [7], en la cual se declara alerta amarilla porque el material particulado excede los umbrales máximos permitidos. Así mismo, para el caso del Ozono, sucede algo similar solo que su primer punto de cambio ocurre en marzo del 2018, cuando en la resolución 831 de 2018 [4] también es decretada la alerta amarilla por contaminación ambiental y se establecen algunas restricciones en la ciudad de Bogotá, como por ejemplo establecer restricciones en el sector de transporte y movilidad y restringir la operaciones de fuentes fijas industriales que operan con combustibles sólidos y líquidos.

En el año 2020, para los meses de febrero y marzo se decretó alerta amarilla en la ciudad de Bogotá, por medio del decreto 077 del 4 de marzo de 2020 [6] y [5], el cual esta relacionado con la restricción de vehiculos de carga, dado que los niveles de concentración de material particulado se encontraban elevados en algunas estaciones de monitoreo y por esta fecha se estaban presentando incendios forestales, aun cuando la movilidad se empezaba a restringir a causa de la pandemia ocasionada por el Covid 19.

Capítulo 3

Diagnóstico de convergencias

3.1. Sin puntos de cambio

De acuerdo a las simulaciones generadas para cada modelo, se realizaron los análisis de convergencia correspondientes obteniendo resultados favorables que indican que el modelo muestra buenos resultados pues los parámetros cumplen la mayoría de los criterios, esto se muestra en las tablas 3.1, 3.2 y 3.3.

En todos los casos, para el criterio de convergencia propuesto en Geweke (1992), se observa que cada una de las pruebas aplicadas a las cinco cadenas realizadas, satisface la condición de que el valor Z calculado, se encuentre entre el intervalo $(-2, 2)$ esto da un buen indicio sobre la convergencia de los parámetros.

Por otra parte, el criterio Heidelberg-Welch de estacionalidad fue satisfactoria para todas las cadenas, por lo cual se identifica si el ancho medio para cada una de ellas es menor que $\varepsilon = 0.1$, y esto, en efecto es cierto en todas las cadenas realizadas para los tres contaminantes.

Y finalmente el criterio de Gelman Rubin, menciona que valores cercanos a 1, indicarían una posible convergencia y en cada uno de los tres contaminantes su valor fue cercano a 1. De esta manera, cumpliendo la mayoría de condiciones de cada uno de los criterios de convergencia estudiados, se puede afirmar con seguridad que los parámetros estimados funcionan adecuadamente.

3.2. Un punto de cambio

El diagnóstico de convergencia para un punto de cambio, arrojó algunos resultados importantes que se deben mencionar. Iniciando con el criterio de Geweke (1992), este es satisfactorio en las cinco cadenas realizadas para el material particulado de $10\mu m$

	Diagnóstico de convergencia					
	Parámetro	cadena 1	cadena 2	cadena 3	cadena 4	cadena 5
Geweke	α	0.03089	1.585	-1.765	-0.5473	-0.1294
	β	-0.24968	1.296	-1.641	-0.3181	-0.23
Raftery - Lewis	α	16.9	16.3	18.7	22.5	16.7
	β	16.8	26.7	17.3	22.0	18.2
H-W Estacionalidad	α	0.877	0.642	0.903	0.916	0.680
	β	0.890	0.598	0.826	0.984	0.747
H-W 1/2 Ancho	α	0.00107	0.00106	0.00112	0.000986	0.000891
	β	0.02705	0.02698	0.02746	0.024899	0.022256
Gelman - Rubin	α	1.00				
	β	1.00				

Tabla 3.1: Información de convergencia para el modelo PM_{10} , sin puntos de cambio. H-W=Heidelberger-Welch

	Diagnóstico de convergencia					
	Parámetro	cadena 1	cadena 2	cadena 3	cadena 4	cadena 5
Geweke	α	0.0959	0.5630	-1.785	-1.455	-0.2081
	β	0.1362	0.4231	-1.836	-1.423	-0.1948
Raftery - Lewis	α	2.51	1.85	2.50	2.56	2.56
	β	2.13	2.41	2.43	2.29	2.34
H-W Estacionalidad	α	0.229	0.764	0.795	0.484	0.0534
	β	0.308	0.507	0.776	0.474	0.8327
H-W 1/2 Ancho	α	0.00291	0.00253	0.00253	0.00232	0.00246
	β	0.09838	0.08906	0.08830	0.07683	0.07932
Gelman - Rubin	α	1				
	β	1				

Tabla 3.2: Información de convergencia para el modelo $PM_{2.5}$, sin puntos de cambio. H-W=Heidelberger-Welch

de tamaño pues la comparación de los valores de las medias iniciales y finales de las cadenas, están entre -2 y 2. Para el caso del $PM_{2.5}$ se observa en la tabla 3.5 que en algunas de las cadenas los valores no se encuentran entre -2 y 2, es el caso de los parámetros α_2 y σ_2 , en los cuales en dos y tres de las cadenas, respectivamente, los resultados se encuentran fuera del intervalo que se considera adecuado, sin embargo, estos números no son significativamente grandes. Para el caso del O_3 sucede lo mismo que en el PM_{10} pues todos sus valores se encuentran dentro del intervalo que genera confianza, para poder hablar de estacionalidad y posiblemente de convergencia.

Otro de los criterios que proporciona información sobre la convergencia de los parámetros es el test de estacionalidad de Heidelberger-Welch, pues los tres contaminantes lo pasaron, de manera que se procede a hacer la revisión del criterio del Ancho medio,

	Diagnóstico de convergencia					
	Parámetro	cadena 1	cadena 2	cadena 3	cadena 4	cadena 5
Geweke	α	0.7882	0.2856	0.7351	0.9028	-0.07696
	β	0.7052	0.2041	0.9396	0.8609	-0.14975
Raftery - Lewis	α	1.72	1.67	2.40	1.94	1.80
	β	1.31	1.26	1.34	1.22	1.25
H-W Estacionalidad	α	0.408	0.907	0.234	0.351	0.273
	β	0.4185	0.924	0.113	0.374	0.335
H-W 1/2 Ancho	α	0.00139	0.00172	0.00163	0.0018	0.00151
	β	0.06699	0.08421	0.07863	0.0883	0.07574
Gelman - Rubin	α	1				
	β	1.01				

Tabla 3.3: Información de convergencia para el modelo O_3 , sin puntos de cambio. H-W=Heidelberger-Welch

el cual es satisfactorio para todos los parámetros α_1 y α_2 de los tres contaminantes, a excepción de una cadena del contaminante Ozono, sin embargo al ser solo una de cinco cadenas, no se tiene en cuenta, de forma que todos los parámetros a_1 y a_2 tienen una confirmación más de que tienen convergencia. Para el caso de los parámetros β_1 y β_2 este criterio, muestra que en algunas cadenas, sus valores no son menores a 0.1, por tanto este criterio no es suficiente o de apoyo para argumentar convergencia de ellos, de forma similar con el parámetro τ en el caso de los tres contaminantes.

Finalmente, para el caso de los modelos con un solo punto de cambio, se observó que el criterio Gelman- Rubin, es satisfactorio en los tres modelos, pues sus valores son cercanos a 1, toda la información anterior, se encuentra resumida en las tablas 3.4, 3.5 y 3.6.

3.3. Dos puntos de cambio

Al igual que para cero y un punto de cambio, se realiza el diagnostico de convergencia para cada uno de los parametros obtenidos. Dado que cuando hay dos puntos de cambio, se determinan tres momentos para los cuales hay parámetros α_i y β_i para $i = 1, 2, 3$, se observa en algunas de las pruebas de convergencia que los resultados para los parámetros de la tercera sección no son satisfactorios. Sin embargo, son aceptados tres de los cinco tests, por tanto se aceptan como válidos.

Para el caso del criterio Geweke, las cinco cadenas realizadas para los 3 contaminantes arrojaron resultados positivos pues la diferencia entre la parte inicial (10%) y la parte final (50%) de las cadenas es mínima, encontrándose la mayoría de veces en un rango

	Diagnóstico de convergencia					
	Parámetro	cadena 1	cadena 2	cadena 3	cadena 4	cadena 5
Geweke	α_1	-0.2161	-0.3797	0.8125	-0.6749	1.4716
	β_1	-0.2583	-0.3447	0.8189	-0.6263	1.5856
	α_2	-0.3066	-0.1314	-1.1402	-0.4315	1.0628
	β_2	-0.3165	0.1059	-1.1862	-0.4369	1.0672
	τ	-1.8701	-0.6091	-1.4073	0.1858	-0.7393
Raftery - Lewis	α_1	4.49	5.25	2.88	2.67	2.83
	β_1	3.64	3.41	3.51	3.01	36.30
	α_2	18.60	30.10	13.60	27.40	4.69
	β_2	19.00	28.50	13.80	29.10	44.70
	τ	5.81	6.90	6.85	6.32	5.95
H-W Estacionalidad	α_1	0.5609	0.986	0.193	0.186	0.996
	β_1	0.5359	0.983	0.259	0.212	0.995
	α_2	0.0573	0.509	0.651	0.122	0.653
	β_2	0.0973	0.593	0.639	0.125	0.646
	τ	0.4983	0.637	0.645	0.327	0.384
H-W 1/2 Ancho	α_1	0.00642	0.00547	0.00679	0.00705	0.00758
	β_1	0.12170	0.10158	0.12922	0.13551	0.14583
	α_2	0.01301	0.01932	0.01404	0.01324	0.01780
	β_2	1.57577	1.86337	1.58302	1.43715	1.75334
	τ	0.16809	0.17736	0.19509	0.18970	0.21971
Gelman - Rubin	α_1	1.01				
	β_1	1.01				
	α_2	1.00				
	β_2	1.00				
	τ	1.00				

Tabla 3.4: Información de convergencia para el modelo PM_{10} , un punto de cambio. H-W=Heidelberger-Welch

entre $(-2, 2)$, esto se puede observar claramente en las tablas 3.7, 3.8 y 3.9.

	Diagnóstico de convergencia					
	Parámetro	cadena 1	cadena 2	cadena 3	cadena 4	cadena 5
Geweke	α_1	-0.2860	-1.0051	-17.774	0.7692	-0.5075
	β_1	-0.3071	-0.9200	0 - 12.524	0.7046	-0.5437
	α_2	-2.6915	-3.8182	-1.997	-1.5921	-1.3143
	β_2	-2.5533	-3.4243	-4.756	-1.4134	-1.2864
	τ	0.2518	-0.2185	-13.928	0.6264	1.0813
Raftery - Lewis	α_1	13.60	11.20	8.37	14.30	12.10
	β_1	12.00	13.00	6.89	11.40	12.70
	α_2	4.54	3.14	4.19	2.85	6.01
	β_2	3.87	3.99	2.60	29.10	5.17
	τ	7.59	4.33	14.40	8.44	5.05
H-W Estacionalidad	α_1	0.713	0.2035	0.970	0.381	0.746
	β_1	0.714	0.2018	0.982	0.361	0.743
	α_2	0.161	0.0714	0.104	0.445	0.217
	β_2	0.203	0.1176	0.166	0.647	0.322
	τ	0.951	0.1457	0.989	0.563	0.328
H-W 1/2 Ancho	α_1	0.0161	0.0150	0.0184	0.0148	0.0167
	β_1	0.5514	0.5158	0.5726	0.4953	0.5719
	α_2	0.0239	0.0206	0.0149	0.0154	0.0120
	β_2	1.4992	1.1535	0.8434	0.8537	0.6364
	τ	0.2058	0.1376	5.0098	0.4146	0.1456
Gelman - Rubin	α_1	1.03				
	β_1	1.01				
	α_2	1.03				
	β_2	1.05				
	τ	2.65				

Tabla 3.5: Información de convergencia para el modelo $PM_{2.5}$, un punto de cambio. H-W=Heidelberger-Welch

	Diagnóstico de convergencia					
	Parámetro	cadena 1	cadena 2	cadena 3	cadena 4	cadena 5
Geweke	α_1	0.1738	-1.8966	0.06743	-3.1571	-1.49105
	β_1	-0.7537	-0.8614	0.23790	-2.5854	-1.0130
	α_2	-1.7965	-1.6817	1.22712	0.8546	1.02269
	β_2	-1.9557	-1.6606	1.19464	0.8934	0.9521
	τ	-0.7835	1.1726	-0.22160	0.9531	1.7600
Raftery - Lewis	α_1	2.54	2.67	1.76	1.80	1.91
	β_1	1.34	1.24	1.37	11.40	34.00
	α_2	16.60	29.60	29.40	32.50	2.51
	β_2	20.70	18.50	27.60	35.70	42.00
	τ	2.90	2.52	32.30	2.11	9.89
H-W Estacionalidad	α_1	0.360	0.000394	0.421	0.0629	0.222
	β_1	0.193	0.070700	0.2334	0.2334	0.808
	α_2	0.634	0.731613	0.529	0.6852	0.340
	β_2	0.582	0.777112	0.504	0.6699	0.286
	τ	0.729	0.206458	0.574	0.3733	0.183
H-W 1/2 Ancho	α_1	0.0134	NA	0.00964	0.0143	0.0107
	β_1	0.1459	0.1456	0.13497	0.1554	0.1403
	α_2	0.0283	0.0256	0.02509	0.0264	0.0265
	β_2	2.5965	2.3807	2.32836	2.3940	2.4613
	τ	1.2920	2.0953	0.56484	1.9224	0.9394
Gelman - Rubin	α_1	1.01				
	β_1	1.00				
	α_2	1.03				
	β_2	1.03				
	τ	1.16				

Tabla 3.6: Información de convergencia para el modelo O_3 , un punto de cambio. H-W=Heidelberger-Welch

	Diagnóstico de convergencia					
	Parámetro	cadena 1	cadena 2	cadena 3	cadena 4	cadena 5
Geweke	α_1	-1.52267	0.2015	1.2013	1.488	1.115
	β_1	-1.63797	0.2498	1.0321	1.404	1.041
	α_2	-0.02648	0.4891	0.8232	1.870	1.740
	β_2	0.12896	0.5491	0.8213	1.920	1.477
	α_3	-0.99249	-0.1763	3.3275	3.353	2.521
	β_3	-1.06002	-0.2584	3.5730	3.338	2.766
	τ_1	-1.62130	1.5736	0.5697	1.020	1.515
	τ_2	1.95434	-0.7360	-0.8417	-1.122	-0.652
Raftery - Lewis	α_1	17.30	15.70	23.50	27.30	17.10
	β_1	17.30	21.00	28.10	22.70	18.40
	α_2	63.40	26.50	54.50	36.80	41.50
	β_2	38.10	28.70	48.80	32.10	49.60
	α_3	80.10	113.00	72.10	68.10	121.00
	β_3	91.90	132.00	61.00	68.20	129.00
	τ_1	7.54	4.65	7.35	10.00	6.28
	τ_2	3.87	2.85	3.58	3.96	4.64
H-W Estacionalidad	α_1	0.6142	0.850	0.878	0.442	0.5959
	β_1	0.5893	0.872	0.880	0.505	0.5159
	α_2	0.9387	0.950	0.226	0.751	0.1613
	β_2	0.9613	0.922	0.143	0.701	0.0831
	α_3	0.4788	0.111	0.180	0.184	0.3721
	β_3	0.3953	0.087	0.119	0.169	0.2929
	τ_1	0.0741	0.230	0.423	0.954	0.7865
	τ_2	0.4755	NA	0.962	0.773	0.2687
H-W 1/2 Ancho	α_1	0.00859	0.00805	0.00881	0.00902	0.00802
	β_1	0.14425	0.13628	0.12922	0.15081	0.12333
	α_2	0.05006	0.05475	0.14883	0.05362	0.07562
	β_2	4.06275	4.51499	3.88306	4.18159	6.14082
	α_3	0.14425	0.05542	0.05968	0.04159	0.08012
	β_3	4.64227	4.74765	5.20534	4.02355	6.63150
	τ_1	0.13175	0.10341	0.13202	0.15353	0.12206
	τ_2	0.24266	0.17736	0.33214	1.75243	0.27591
Gelman - Rubin	α_1	1.00				
	β_1	1.00				
	α_2	1.00				
	β_2	1.04				
	α_3	1.04				
	β_3	1.00				
	τ_1	1.00				
	τ_2	1.20				

Tabla 3.7: Información de convergencia para el modelo PM_{10} , dos puntos de cambio. H-W=Heidelberger-Welch

	Diagnóstico de convergencia					
	Parámetro	cadena 1	cadena 2	cadena 3	cadena 4	cadena 5
Geweke	α_1	-0.3746	0.9093	2.0729	-1.1998	-1.05503
	β_1	-0.3087	0.7571	1.4881	-1.3089	-0.94787
	α_2	0.3705	-0.2982	0.4372	-1.6210	0.78496
	β_2	0.4360	-0.3576	0.3181	-1.7367	0.76589
	α_3	3.0666	-1.8681	-1.0074	-0.7896	0.38751
	β_3	3.0499	-1.9577	-0.9945	-0.8740	0.28443
	τ_1	0.6974	0.4718	1.1884	-0.5721	-0.06915
	τ_2	1.3431	1.1733	0.8947	0.4712	-0.50966
Raftery - Lewis	α_1	23.30	63.50	65.20	3.62	14.30
	β_1	17.90	59.40	37.20	2.60	14.00
	α_2	26.70	138.00	61.10	12.50	45.10
	β_2	31.80	643.00	99.70	11.20	51.30
	α_3	28.20	35.30	36.10	19.70	31.40
	β_3	32.50	36.70	24.30	20.90	34.70
	τ_1	2.00	118.00	33.90	9.30	1.95
	τ_2	2.38	2.26	2.81	2.32	2.37
H-W Estacionalidad	α_1	0.622	0.464	0.0977	0.3278	0.475
	β_1	0.662	0.688	0.1093	0.3297	0.507
	α_2	0.497	0.268	0.6616	0.5856	0.311
	β_2	0.457	0.250	0.5127	0.2951	0.285
	α_3	0.346	0.179	0.9512	0.8914	0.478
	β_3	0.325	0.195	0.9309	0.9208	0.354
	τ_1	0.150	0.405	0.8667	0.0740	0.886
	τ_2	0.290	0.566	0.9381	0.0575	0.610
H-W 1/2 Ancho	α_1	0.0112	0.0137	0.0228	0.0042	0.0134
	β_1	0.4060	0.4696	0.5871	0.1242	0.4753
	α_2	0.0210	0.0342	0.1070	0.0379	0.0218
	β_2	1.8244	2.7514	7.1463	1.4405	1.8541
	α_3	0.0202	0.0214	0.0208	0.0178	0.0169
	β_3	2.2830	2.4528	2.3331	2.0932	1.9352
	τ_1	0.1903	2.3093	14.1131	1.4328	0.1835
	τ_2	0.1726	0.1001	0.1139	0.1517	0.1645
Gelman - Rubin	α_1	2.96				
	β_1	2.00				
	α_2	1.00				
	β_2	1.04				
	α_3	2.04				
	β_3	1.00				
	τ_1	6.31				
	τ_2	1.20				

Tabla 3.8: Información de convergencia para el modelo $PM_{2.5}$, dos puntos de cambio. H-W=Heidelberger-Welch

	Diagnóstico de convergencia					
	Parámetro	cadena 1	cadena 2	cadena 3	cadena 4	cadena 5
Geweke	α_1	-0.2309	0.7046	0.37025	-0.7834	0.1310
	β_1	-0.5452	1.1900	0.47824	-0.7366	-0.1229
	α_2	-0.7818	-1.01922	0.31625	1.8306	0.1882
	β_2	-0.5366	-0.8262	0.37046	1.6133	-0.1370
	α_3	1.4362	1.7939	0.03348	1.3482	1.7756
	β_3	1.4085	1.6871	0.05774	1.2667	1.7408
	τ_1	0.1269	0.6347	0.56679	-0.9945	1.2870
	τ_2	1.4221	0.4750	0.65729	-1.8853	-1.0234
Raftery - Lewis	α_1	1.74	2.66	2.60	1.84	1.84
	β_1	2.29	2.14	2.29	2.20	1.22
	α_2	10.60	8.95	15.00	10.30	13.30
	β_2	15.90	9.87	13.50	11.80	12.80
	α_3	60.80	81.60	28.30	75.00	27.30
	β_3	58.00	91.50	28.00	86.30	28.30
	τ_1	1.69	3.69	2.01	2.68	2.19
	τ_2	1.53	2.41	1.56	1.57	2.82
H-W Estacionalidad	α_1	0.846	0.0777	0.202	0.1048	0.6148
	β_1	0.852	0.1015	0.159	0.0774	0.6247
	α_2	0.357	0.6049	0.683	0.2314	0.9614
	β_2	0.474	0.7155	0.855	0.2932	0.8402
	α_3	0.400	0.3084	0.127	0.3574	0.0919
	β_3	0.380	0.3380	0.126	0.3619	0.0919
	τ_1	0.377	0.5603	0.442	0.7607	0.0704
	τ_2	0.371	0.7673	0.417	0.3295	0.4528
H-W 1/2 Ancho	α_1	0.00692	0.00831	0.00647	0.00655	0.00638
	β_1	0.10168	0.12152	0.09463	0.11495	0.09249
	α_2	0.00896	0.00849	0.00947	0.00861	0.00880
	β_2	0.44361	0.41473	0.47466	0.41301	0.42244
	α_3	0.02759	0.03222	0.02518	0.03738	0.02571
	β_3	3.38161	3.77077	3.037821	4.07588	3.15376
	τ_1	0.05536	0.05117	0.05628	0.05118	0.05425
	τ_2	0.11938	0.13950	0.11812	0.10954	0.10968
Gelman - Rubin	α_1	1.00				
	β_1	1.00				
	α_2	1.00				
	β_2	1.0				
	α_3	1.03				
	β_3	1.03				
	τ_1	1.00				
	τ_2	1.00				

Tabla 3.9: Información de convergencia para el modelo O_3 , dos puntos de cambio. H-W=Heidelberger-Welch

Capítulo 4

Modelos univariados para los pares de los contaminantes PM_{10} , $PM_{2.5}$ y O_3

4.1. Correlación de dos procesos estocásticos

En el presente capítulo se realizará la aplicación del modelo bivariado, haciendo uso de la función cópula construida para determinar el parámetro θ que describirá la correlación entre los dos procesos estocásticos en cuestión. Para este caso particular se revisará de la siguiente manera:

1. PM_{10} vs O_3
2. $PM_{2.5}$ vs O_3
3. PM_{10} vs $PM_{2.5}$

Como primer paso se realizará una prueba que valide el funcionamiento correcto del modelo, para esto se realizará la simulación para el Ozono versus él mismo, para esto, se determinan los parámetros de las distribuciones a priori de la siguiente manera: Se retoma la información de la tabla 2.3 particularmente la media y la desviación estandar, con estos dos datos y sabiendo que para la distribución uniforme con parámetros (a, b) se tiene:

$$E(x) = \frac{(a+b)}{2} \qquad Var(x) = \frac{(b-a)^2}{12}$$

De esta manera es inmediato identificar los valores de a y b

$$1.21 = \frac{(a+b)}{2} \quad (0.03)^2 = \frac{(b-a)^2}{12}$$

Luego entonces, para el parámetro α los valores son $a = 1.16$ y $b = 1.26$, de manera análoga se realizan los cálculos para el parámetros β y también para los otros dos contaminantes. Al ejecutar el programa de computo se obtienen los siguientes resultados:

Modelo	Información estadística O_3 vs O_3				
	Parámetros	dist. inicial	Media	sd	intervalo 95 %
Sin puntos de cambio	θ	$unif(-1, 1)$	0.9487761	0.004088	(0.8906871; 0.9570268)
	α_1	$unif(1.2, 1.3)$	1.2000047	0.0000497	(1.2000001; 1.2000183)
	β_1	$unif(10.9, 15.8)$	13.3206233	0.10380	(13.1292662; 13.5385384)
	α_2	$unif(1.2, 1.3)$	1.2000047	0.0000465	(1.2000001; 1.2000163)
	β_2	$unif(10.9, 15.8)$	13.3157448	0.10250	(13.1135846; 13.5139078)

Tabla 4.1: Prueba del modelo bivariado

Como se puede observar en la tabla 4.1 el parámetro θ arroja un valor muy cercano a 1 lo cual consistente y tiene sentido pues los datos estudiados son exactamente iguales, esta prueba genera confianza para poder determinar la correlación, ahora sí, entre cada par de contaminantes.

4.1.1. Correlación, θ , entre los contaminantes PM_{10} y O_3

En esta sección se presentan los resultados obtenidos al ejecutar el modelo que relaciona el material particulado de 10 micras de tamaño, junto con el ozono.

Como distribución a priori se elige, uniforme con parámetros a, b y estos valores fueron encontrados tal como se mencionó anteriormente. La tabla 4.2 resume los resultados obtenidos para estos dos procesos estocásticos.

Modelo	Información estadística PM_{10} y O_3				
	Parámetros	dist. inicial	Media	sd	intervalo 95 %
Sin puntos de cambio	θ	$unif(-1, 1)$	-0.9999994	0.0000069	$(-1.0000000; -0.99999758)$
	α_1	$unif(0.803, 0.977)$	0.8030031	0.0002969	$(0.8030001; 0.8030110)$
	β_1	$unif(0.6, 2.9)$	0.6000459	0.000472	$(0.6000013; 0.6001711)$
	α_2	$unif(1.2, 1.3)$	1.2000027	0.0000277	$(1.2000001; 1.2000100)$
	β_2	$unif(10.9, 15.8)$	15.4519679	0.000493	$(15.3486982; 15.5417151)$

Tabla 4.2: Correlación entre los contaminantes PM_{10} y O_3

Para este primer análisis, el resultado del parámetro θ arroja un resultado completamente correlacionado negativamente, pues su valor medio fue -0.9999994 , esto lo que indica es que, con una alta probabilidad, el aumento de la contaminación de PM_{10} afecta de manera inversa a la concentración de partículas de O_3 en el aire. Por otra parte en la tabla 4.2 se puede observar que el intervalo de confianza es reducido, esto se puede explicar, porque los parámetros de las distribuciones iniciales también lo son, principalmente los α_i de cada uno de los contaminantes. Para representar esta información de forma gráfica, se realizaron las curvas de nivel que se observan en la figura 4.1. En esta imagen, de color gris se representa la suma de los datos observados de PM_{10} y O_3 , las graficas de color verde y azul, representan el intervalo de confianza estimado para la función $m(t_1, t_2)$ y la grafica de color verde, representa la estimación propuesta. Este análisis se realiza sin tener en cuenta algún punto de cambio, al igual que en el modelo univariado, si se empiezan a tener en cuenta, los resultados iran siendo óptimos en su proceso.

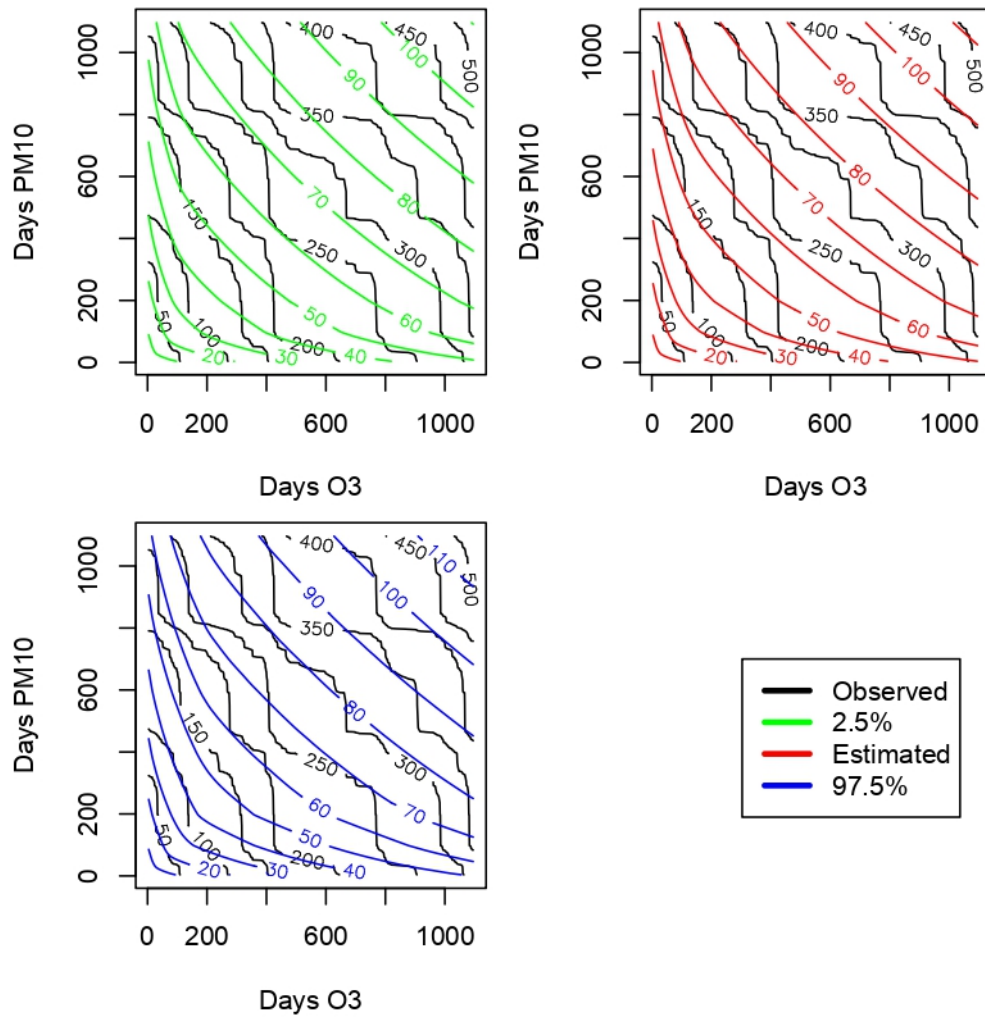


Figura 4.1: Curvas de nivel para los contaminantes PM_{10} y O_3 . [12]

4.1.2. Correlación, θ , entre los contaminantes $PM_{2.5}$ y O_3

En la tabla 4.3, se muestran los resultados obtenidos al ejecutar el código en R, se observa aquí algo interesante y es que la correlación es negativa y casi perfecta. Lo que quiere decir esto, es que a medida que los niveles de concentración de material particulado de $2.5\mu m$ de tamaño, los niveles de concentración de ozono disminuyen.

Modelo	Información estadística $PM_{2.5}$ y O_3				
	Parámetros	dist. inicial	Media	sd	intervalo 95 %
Sin puntos de cambio	θ	$unif(-1, 1)$	-0.9999321	0.000678	(-0.9999980; -0.9997630)
	α_1	$unif(1.2, 1.3)$	0.7645434	0.0000206	(0.7596495; 0.7696293)
	β_1	$unif(10.9, 15.8)$	5.0000769	0.00808	(5.0000019; 5.0002863)
	α_2	$unif(1.2, 1.3)$	0.6867448	0.0000248	(0.6820837; 0.6916936)
	β_2	$unif(10.9, 15.8)$	5.0000812	0.00808	(5.0000025; 5.0003026)

Tabla 4.3: Correlación entre los contaminantes PM_{10} y O_3

Al igual que en el caso anterior, a la información contenida en la tabla 4.3 se le realizó la respectiva representación por medio de curvas de nivel, esta información se puede observar en la figura 4.2, en ella se puede observar el ajuste del modelo, similar a como ocurría en el modelo univariado, donde los datos se ajustaban de buena manera, en el inicio de los días analizados, pero que luego, por los puntos de cambio el modelo se aleja un poco.

4.1.3. Correlación, θ , entre los contaminantes $PM_{2.5}$ y PM_{10}

El material particulado dentro del análisis presenta justo lo que se espera, pues por tratarse del mismo contaminante pero con distinto tamaño, se espera que los resultados de uno, de alguna manera, están contenidos dentro del otro, esto significa que su correlación debe ser positiva y cercana a 1, y es justo lo que se puede observar en la tabla 4.4 donde el valor θ dio un resultado de 0.7546785, esto continúa resultados interesantes dentro del estudio porque concuerda con lo esperado.

Modelo	Información estadística $PM_{2.5}$ y PM_{10}				
	Parámetros	dist. inicial	Media	sd	intervalo 95 %
Sin puntos de cambio	θ	$unif(-1, 1)$	0.7546785	0.008007	(0.7390866; 0.7693154)
	α_1	$unif(1, 3)$	1.0000078	0.000789	(1.0000002; 1.0000301)
	β_1	$unif(7, 20)$	8.0001299	0.000129	(8.0000029; 8.0004811)
	α_2	$unif(1, 3)$	1.0000105	0.0000981	(1.0000003; 1.0000364)
	β_2	$unif(7, 20)$	8.0001377	0.0001347	(8.0000046; 8.0005117)

Tabla 4.4: Correlación entre los contaminantes PM_{10} y $PM_{2.5}$

Para este par de contaminantes también se construyeron las curvas de nivel, figura 4.3, que asocian estos resultados. En este caso el modelo se ajusta de manera adecuada y de forma similar a los pares estudiados anteriormente, se pueden construir los modelos teniendo en cuenta algunos puntos de cambio en común para obtener un ajuste óptimo.

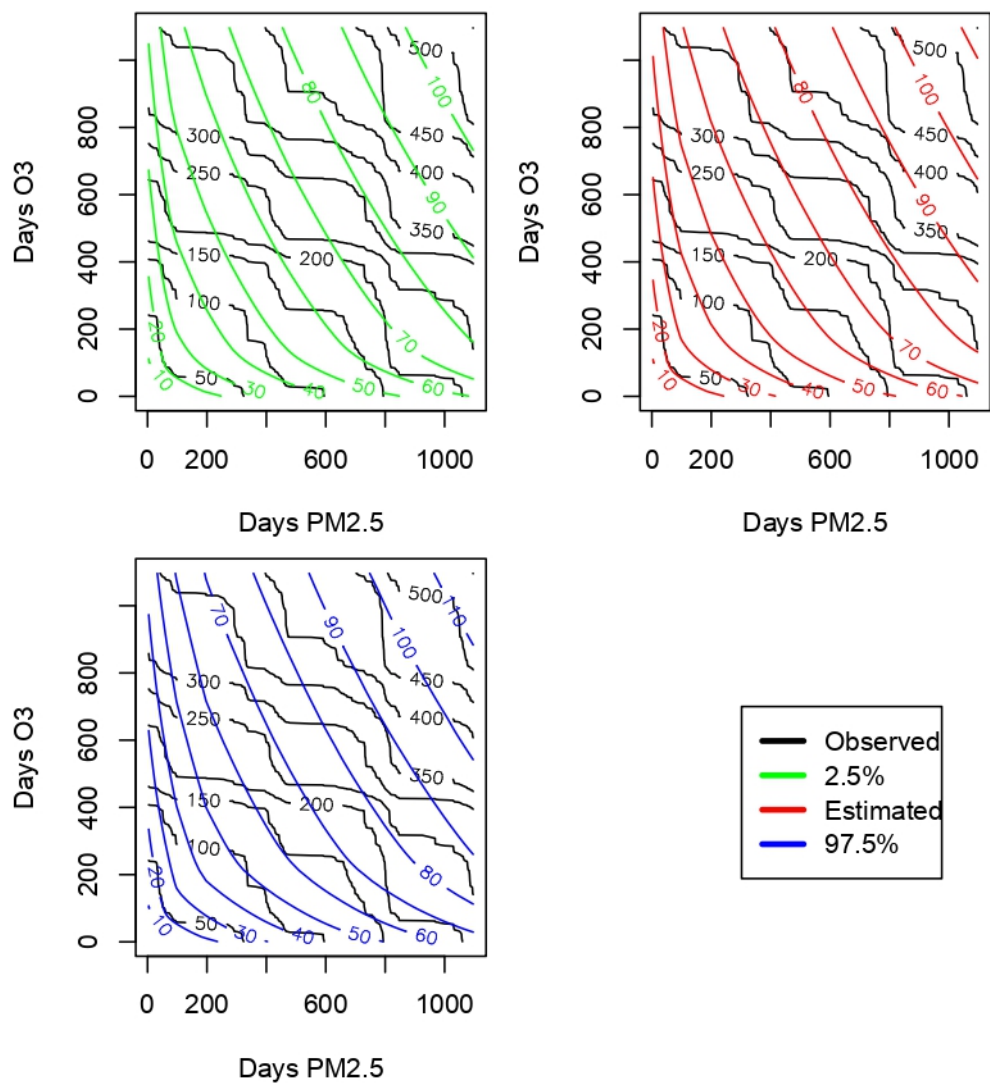


Figura 4.2: Curvas de nivel para los contaminantes $PM_{2.5}$ y O_3 . [12]

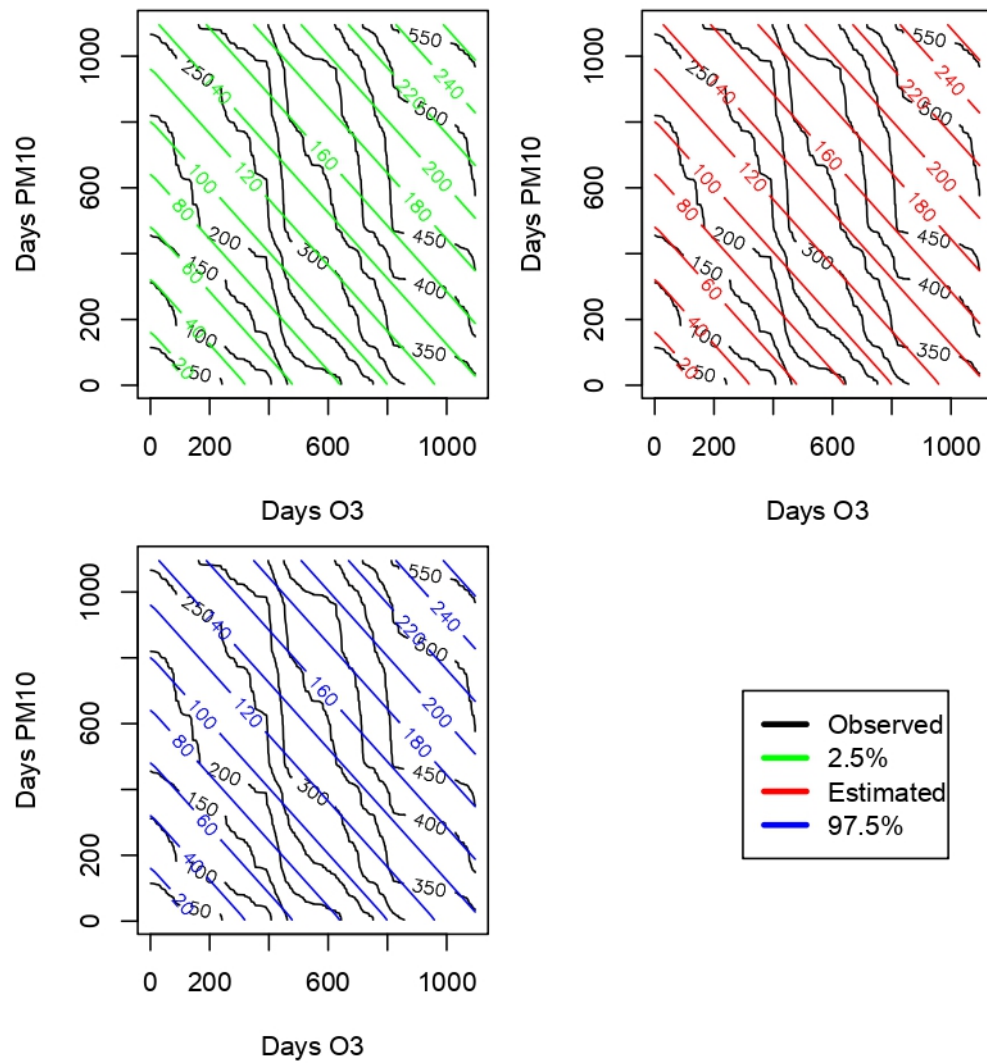


Figura 4.3: Curvas de nivel para los contaminantes PM_{10} y $PM_{2.5}$.

Capítulo 5

Conclusiones y consideraciones

5.1. Conclusiones

En el desarrollo del presente trabajo de grado se encontraron algunos aspectos que describen la contaminación del aire de Bogotá, en los años 2018 a 2020, específicamente de los contaminantes PM_{10} , $PM_{2.5}$ y O_3 . En primer lugar se evidenció, que durante estos tres años, se sobrepasó la norma establecida por la secretaria de ambiente de Bogotá en mas del 20 % de los días analizados, lo cual es una alerta que se debe considerar en la toma de desiciones para los planes de descontaminación del aire de Bogotá, además, en las series de tiempo presentadas en la gráfica 2.2, se evidencia el crecimiento de los niveles de contaminación hacia el final del año 2020, año en el cual se mostró una mejoría ocasionada por las restricciones establecidas por la pandemia causada por el COVID 19.

Se evidencia en el capítulo 2, que la modelación del comportamiento de los datos de contaminación en el aire, mejora considerablemente al tener en cuenta algún punto de cambio y es algo interesante observar la manera en que estos puntos de cambio coinciden justo con alertas decretadas por la secretaria de ambiente de Bogotá, o con días en los cuales hubo día sin carro en la ciudad.

El objetivo principal del presente trabajo de grado era determinar la correlación entre dos contaminantes relacionados con el aire de la ciudad de Bogotá, luego de construir los modelos bivariados para las tres posibles parejas de contaminantes y de encontrar su correlación a partir del parámetro θ se obtuvieron algunos resultados que vale la pena resaltar, como por ejemplo que los contaminantes PM_{10} y O_3 tienen una correlación negativa y muy cercana a -1 , lo cual indica que en los lugares de Bogotá, en los que hay altos niveles de concetración de material particulado de 10 micrometros de tamaño, la cantidad de particulas de Ozono, se reducen o al contrario, si se observa una cantidad

elevada de Ozono, el material particulado aumenta. Esto lo que puede sugerir es que las estrategias para contrarestar los altos niveles de concentración, pueden estar fallando, en el sentido que, si bien es cierto que se disminuye un contaminante, se está fortaleciendo el otro, lo cual no es lo esperado. De la misma manera, para el par $PM_{2.5}$ vs O_3 sucede igual, la concentración de una reduce la del otro.

Para el par de contaminantes material particulado de $2.5\mu m$ y $10\mu m$ de tamaño, sucede matemáticamente, justo lo esperado y es que por ser el mismo material, pero con distinto tamaño, la lectura de su concentración en las estaciones de monitoreo deberían estar correlacionadas pues uno esta contenido dentro del otro. En efecto la correlación presentada por el modelo realizado fue de $\theta = 0.7546785$ tal como se observa en la tabla 4.4.

5.2. Consideraciones

El modelo bivariado realizado, arrojo valores de correlación para el ozono vs cada uno de los dos materiales particulados muy exactos, esta correlación puede llegar a ser mas exacta si se tienen en cuenta los otros dos modelos realizados de manera univariada, es decir, ajustar el modelo y buscar la correlación de los dos procesos teniendo en cuenta un o dos puntos de cambio y porqué no, más momentos en los cuales hubo cambios extremos.

Bibliografía

- [1] Suárez, B. M. (2020). Modelos de Poisson no homogéneos en el estudio de contaminantes en la ciudad de Bogotá, Estadística, Universidad Nacional de Colombia, Tesis de doctorado.
- [2] Nelsen, R. B. (2006) An Introduction to Copulas (Springer Series in Statistics). Springer-Verlag, Berlin, Heidelberg.
- [3] Barreto, L. (2019) Se declara alerta ambiental en Bogotá Mi-ciudad
- [4] RES(2018). Resolución 831 Por el cual se Declara la Alerta Amarilla por contaminación atmosférica en la ciudad de Bogotá D.C. Secretaria Distrital de Ambiente.
- [5] DEC(2019). Decreto 840, Por medio del cual se establecen las condiciones y restricciones para el tránsito de los vehículos de transporte de carga en el Distrito Capital, y se dictan otras disposiciones. Alcaldía Mayor de Bogotá.
- [6] DEC(2020). Decreto 077, Por medio del cual se modifica el Decreto Distrital 840 de 2019 y se dictan otras disposiciones. Alcaldía Mayor de Bogotá.
- [7] RES(2019). Por la cual se Declara Alerta Amarilla en la Ciudad de Bogotá D.C, y Alerta Naranja en el Sur Occidente de la Ciudad por contaminación atmosférica, y se toman otras determinaciones. Secretaria Distrital de Ambiente.
- [8] SDA (2010). Plan de desconaminación del aire.
- [9] Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2003) Bayesian Data Analysis. A CRC Press Company.
- [10] Jiménez, J. (2015) Métodos Monte Carlo basados en cadenas de Markov.
- [11] Geweke, J.(1992). Evaluating the accuract of sampling-based approaches to the calculation of posterior moments. Bayesian Statistics, Oxford Univ. Press., 4:169-193.

-
- [12] EPA (SF) Conceptos básicos sobre el material particulado (PM, por sus siglas en inglés), Agencia de Protección Ambiental de Estados Unidos.
- [13] Organización Mundial de la Salud, (2018). Más del 90 % de los niños del mundo respiran aire tóxico a diario. Comunicado de prensa.
- [14] Outdoor air pollution a leading environmental cause of cancer deaths. Press release 221, World Health Organization, Geneva, Switzerland.
- [15] Sklar, A. (1996). Random variables, distribution functions, and copulas—a personal look backward and forward, colume 28 of Lecture Notes-Monograph Series, pages 1-14. Institute of Mathematical Satatistics, Hayward, CA.
- [16] Farlie, D. J. G. (1960). the performance of some correlation coefficients for a general bivariate distribution. Biometrika, 47(3/4):307-323
- [17] El Adlouni, S., Favre, A.-C, and Bobée, B. (2006). Comparison of methodologies to assess the convergence of markov chain monte carlo methods. Computational Statistics Data Analysis, 50:2685-2701.
- [18] Roy, V. (2019). Convergence diagnostics for Markov chain Monte Carlo. Departamento de estadística, Iowa state University.
- [19] R core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- [20] Romo, V, (2013). Implementación del muestreo de Gibbs en R” para la estimación de parámetros del modelo de volatilidad estocástica con aplicación a datos de contaminación. Universidad Nacional Autónoma de México, Tesis de maestría.