UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

Modelos dinámicos lineales: una aplicación para el pronóstico de las concentraciones de ozono

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuaria

PRESENTA:

Nayeli Montiel Rodríguez

DIRECTOR DE TESIS:

Dra. Lizbeth Naranjo Albarrán

Ciudad Universitaria, Ciudad de México, 2019

1. Datos del alumno
Montiel
Rodríguez
Nayeli
(55) 58 79 61 48
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
311199415

2. Datos del tutor
Dra.
Lizbeth
Naranjo
Albarrán

3. Datos del sinodal 1
Dra.
Eliane
Regina
Rodrigues

4. Datos del sinodal 2
Dr.
Carlos
Díaz
Ávalos

5. Datos del sinodal 3
Dr.
Eduardo Arturo
Gutiérrez
Peña

6. Datos del sinodal 4
M. en C.
María Fernanda
Gil Leyva
Villa

7. Datos del trabajo escrito
Modelos dinámicos lineales: una aplicación para el pronóstico de las concentraciones de ozono.
104 p.
2019

Agradecimientos

Agradezco a dios por permitirme llegar hasta este punto de mi vida, pues sin él nada sería posible.

A mis padres, por depositar su confianza en mí, por animarme a seguir adelante aún en los momentos más díficiles, por sus consejos, por su paciencia, por su amor y por todas las cosas bellas que me han convertido en la persona que soy ahora. ¡Sí se pudo!

A mi hermana, gracias por cuidarme y escucharme. Te amo.

A ti Julio, gracias por enseñarme que puedo ser más fuerte de lo que pienso, por hacerme ganar independencia, por cuidarme todo este tiempo, por hacerme pensar que siempre habrá cosas más grandes, por tú sinceridad y cariño.

A mis tíos: Ali, Lupe, Ana, Isidro, Lalo, gracias por formar parte de mi vida, por sus consejos y cariño. Por estar ahí cuando más los necesitamos, muchas gracias.

A mis abues Lulú y Pancho, por preocuparse por mí y por motivarme a terminar este proyecto. Soy afortunada de tenerlos conmigo. A mi abuelita Andrea y mi abuelito Pancho, que no pudieron llegar hasta este punto y ver a su nieta titulada, pero que estoy segura se hubiesen sentido muy orgullosos.

A mis primos: Ali, Esme, Isra, Karen, Paty, Fer, gracias por ser parte de mi hermosa familia. ¡Nunca se rindan!

A mi tutora, Lizbeth, gracias por estar detrás de mí y por el tiempo que me brindaste para hacer de este trabajo una realidad. A mis sinodales, gracias por todas sus observaciones y correcciones. A mis maestras: Ruth y Begoña, por su dedicación y paciencia para transmitir en cada alumno sus conocimientos. Las admiro mucho.

A ti Adri, por ser una persona increíble, por brindarme tu amistad y confianza. Todos los días me enseñas algo nuevo.

Finalmente, gracias a la UNAM por ser una institución de calidad y formarnos como profesionistas.

Índice general

Prefacio

En las últimas décadas la contaminación atmosférica se ha convertido en uno de los riesgos latentes que más preocupa a la sociedad actual. Las acciones desmedidas, el mal uso de los recursos naturales, la generación de residuos y emisiones en la atmósfera, etc., han agravado los problemas de contaminación y consigo los problemas de salud a los que estamos expuestos. De acuerdo con la SEMARNAT, la Ciudad de México (CDMX) destaca por ser uno de los casos más conocidos y documentados que se enfrentan a esta problemática, siendo el ozono troposférico (O_3) una de las emisiones con mayor impacto en la calidad del aire de esta ciudad. El ozono es caracterizado por ser un potente oxidante, que se encuentra a nivel de la superficie en áreas urbanas, el cual se produce cuando los óxidos de nitrógeno (NO_x) y los compuestos orgánicos volátiles (COV) reaccionan en la atmósfera en presencia de la luz solar.

Con base en esta temática, se analizan los llamados *modelos de espacio de estados*, como una herramienta para el análisis de series de tiempo que surge alrededor de los años 70's, pues gracias a los avances computacionales y a la gran cantidad de aplicaciones en diversas áreas de estudio, entre las que se encuentran las ciencias ambientales, genética, biología, economía, etc., han ganado popularidad en años recientes. Tomando los modelos dinámicos Lineales Gaussianos, mejor conocidos como *Modelos Dinámicos Lineales* (MDL), como un caso particular.

El objetivo de este trabajo será el estudio de las concentraciones diarias [1] de ozono troposférico en partes por billón (ppb) [2] de la estación el Pedregal (PED), y otras variables meteorológicas: la temperatura, la velocidad del viento y la humedad relativa, a través de un modelo de regresión múltiple dinámico lineal, por su versatilidad para el ajuste de modelos variantes en el tiempo (West and Harrison, 1997). Los datos empleados se obtuvieron de la Red Automática de Monitoreo Ambiental (RAMA) y la Red de Meteorología y Radiación Solar (REDMET). Asimismo se contrastan los resultados del pronóstico temporal del nivel de este contaminante contra aquellos que se obtienen a través de un modelo de regresión lineal múltiple estático. Para dichos fines, este escrito comprende la teoría general de los MDL's. El capítulo 1, muestra los conceptos básicos del enfoque Bayesiano, pues es éste bajo el que se construyen dichos modelos. El capítulo 2, concentra el marco de modelos general, los algoritmos de filtración, suavizamiento y pronóstico, empleados en la inferencia sobre los estados no observables u observaciones futuras. Además,

[1] Concentración diaria: al valor máximo de las concentraciones horarias o de las concentraciones de los promedios móviles de 8 horas de cada día (Secretaria de Salud, 2014).

[2] Partes por billón (ppb): Unidad de medida en la que se mide la concentración del O_3 en un volumen, donde éste es dividido en 1 billón de partes.

en este apartado se incluye el desarrollo de dichos algoritmos para el caso de los MDL's y los métodos de estimación por máxima verosimilitud y métodos de Monte Carlo vía cadenas de Markov para la estimación de parámetros. En el capítulo 3, se introducen las estructuras de modelos básicos bajo el enfoque adoptado: con tendencia, con factores estacionales y covariables. Dedicamos un último capítulo para el desarrollo de una aplicación basada en el objetivo principal de este trabajo. En ésta se incluyen: la descripción de la base de datos de las concentraciones de ozono en la estación el Pedregal, una de las 34 estaciones de monitoreo atmosférico en la CDMX; la especificación de los modelos de regresión lineal estático y dinámico; y su evaluación, para mostrar así las ventajas y desventajas de estos últimos frente a la metodología usual.

Finalmente, se incluye un apéndice con los códigos elaborados en el software estadístico R, sobre la manipulación de la base de datos y los análisis realizados: ajuste de un modelo polinomial, ajuste de un modelo de regresión estático y ajuste de un modelo dinámico lineal.

Capítulo 1

Fundamentos de Inferencia Bayesiana

Una característica esencial de los métodos Bayesianos es el uso explícito de la probabilidad para cuantificar la incertidumbre en inferencias basadas en el análisis estadístico de datos, es decir, hacer inferencia estadística sobre alguna cantidad, significa aprender de esta cantidad desconocida a partir de los datos y explorar cuales podrían ser sus valores posibles.

De manera general el proceso Bayesiano para el análisis de datos puede dividirse en los siguientes pasos:

I. Establecer un modelo de probabilidad completo, es decir, establecer una distribución de probabilidad conjunta para todas las cantidades observables y no observables en el problema en cuestión.

II. Hallar la distribución de probabilidad condicional de las cantidades no observables dada la información observada.

III. Evaluar el ajuste del modelo y las implicaciones resultantes del paso II.

Si bien, gracias a los avances tecnológicos, hoy día resulta una tarea más sencilla el cálculo de probabilidades condicionales, a continuación mencionaremos uno de los pilares bajo el que se construye el enfoque Bayesiano: la *Regla de Bayes*.

1.1. La Regla de Bayes

Definición 1.1.1. ***(Espacio muestral)****. El conjunto, S, de todos los posibles resultados de un experimento en particular será llamado el espacio muestral del experimento.*

Definición 1.1.2. ***(Evento)****. Diremos que un evento es cualquier colección de los posibles resultados de un experimento, es decir, cualquier subconjunto del espacio muestral, S (incluyendo a S).*

Definición 1.1.3. ***(Probabilidad condicional)****. Sean A y B dos eventos en S, y* $P(B) > 0$*, entonces la probabilidad de A dado B, escrita como* $P(A|B)$*, será*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

donde A es el evento de interés, B es el resultado experimental que puede proveer información sobre el evento A, $P(A|B)$ es la probabilidad de A dado que ocurrió el evento B y $P(B)$ es la probabilidad marginal de B.

Ahora bien, reexpresando la igualdad de la Definición 1.1.3 podemos obtener una forma útil de calcular la intersección de la probabilidad de los eventos A y B,

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A), \tag{1.1}$$

donde por simetría obtenemos la segunda igualdad.

Así, utilizando la Definición de probabilidad condicional, 1.1.3, y la igualdad anterior (1.1), obtenemos que

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1.2}$$

Conoceremos a la ecuación 1.2 como la *Regla de Bayes*, en honor al personaje al que se le atribuye su descubrimiento, Sir Thomas Bayes. Una regla simple que relaciona probabilidades condicionales y es utilizada en la inferencia estadística.

Una de las aplicaciones claves de la *Regla de Bayes*, surge al emplearla en un modelo estadístico, definido como un conjunto $(S, \wp)$, tal que, S es el espacio muestral definido en 1.1.1, y $\wp$ es el conjunto de probabilidad de alguna de las distribuciones en S, comúnmente parametrizado como $\wp = \{\wp_\theta : \theta \in \Theta\}$. El conjunto Θ define los parámetros del modelo. Así, con el fin de hacer afirmaciones probabilísticas sobre los parámetros de un modelo, θ, dada la información o datos observados, y, debemos iniciar proporcionando su función de masa de probabilidad conjunta o función de densidad conjunta [1] dada como el producto de dos densidades comunmente conocidas en la literatura como la *distribución a priori*, $p(\theta)$, y la *distribución muestral* (la distribución de los datos), $p(y|\theta)$, respectivamente:

$$p(\theta, y) = p(y|\theta)p(\theta). \tag{1.3}$$

Reescribiendo la ecuación (1.3) y utilizando la *Regla de Bayes* (1.2), obtenemos la probabilidad condicional de θ dada la información observada, y, a la que llamaremos *distribución posterior*:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)}. \tag{1.4}$$

[1] Con las expresiones $p(\bullet, \bullet)$ y $p(\bullet)$ denotaremos a las funciones de probabilidad conjunta y marginal respectivamente, y con $p(\bullet|\bullet)$ a la función de probabilidad condicional. Cabe aclarar que, utilizaremos la misma notación para denotar a las funciones de masa de probabilidad y las funciones de densidad de probabilidad, en el caso de variables aleatorias discretas y variables aleatorias continuas, respectivamente.

donde $p(y)$ será la distribución marginal de y y estará dada por $p(y) = \sum_\theta p(y|\theta)p(\theta)$ con θ una variable que toma todos los valores posibles en la suma en el caso discreto, y

$$p(y) = \int_\theta p(y|\theta)p(\theta)d\theta \tag{1.5}$$

con θ una variable que toma todos los valores posibles sobre la integral en el caso continuo[2].

Nota: Una forma equivalente de la igualdad (1.4) omite la probabilidad $p(y)$, al ser un factor que no depende de θ, por lo que se puede considerar constante, dando como resultado una densidad posterior no normalizada [3],

$$p(\theta|y) \propto p(y|\theta)p(\theta). \tag{1.6}$$

1.1.1. La Distribución a Priori

Dada su importancia en la inferencia Bayesiana, la elección de una *distribución a priori* adecuada según el contexto y aplicación de los datos es fundamental, pues representa la creencia sobre el valor desconocido del parámetro, θ, antes de introducir los datos, y, en el modelo. Además, da lugar a la *distribución posterior*, es decir, a la probabilidad de la distribución de θ dados los datos, y.

En la práctica cuantificar una creencia subjetiva se torna complicado, por lo que, en la literatura se suelen mostrar algunas de las más convencionales, entre las que se encuentran:

a) *A priori informativa*. Es aquella distribución que no es dominada por la función de verosimilitud y cuyo impacto sobre la distribución posterior suele ser mayor. Ésta suele especificarse delicadamente a través de información de la población bajo estudio, de estudios previos, o incluso combinando la opinión de expertos con la información actual.

b) *A priori no informativa*. Es aquella distribución que no posee una base poblacional y cuya densidad a priori es descrita como: vaga, plana, difusa o "no informativa".

   Aún cuando en muchos casos resulta complicado construir este tipo de distribución, se suele justificar el uso de la misma con el siguiente argumento: "dejar que los datos hablen por sí solos", de tal modo que las inferencias no se vean afectadas por información externa a los datos actuales.

c) *A priori propias e impropias*. De manera general llamaremos a una densidad a priori $p(\theta)$ propia, si integra uno. Por el contrario será impropia si $\int p(\theta)d\theta = \infty$. Este último tipo de distribuciones puede llevar, en algunos casos, a distribuciones posteriores propias, es decir, $\int p(\theta|y)d\theta$ será finita para todo valor de y. Finalmente, si no se cae en ninguno de los dos casos anteriores, es decir, $\int p(\theta)d\theta \neq 1$ y $\int p(\theta)d\theta \neq \infty$ [4], entonces lo que se puede hacer es normalizar para caer en el caso de una distribución propia.

[2] En secciones posteriores nos referiremos a la *distribución predictiva a priori*, definida para el caso de continuo.
[3] El símbolo $\propto$ significa "proporcional a".
[4] El simbolo $\neq$ significa "disntinto de".

d) *A priori conjugadas.* Si $\Im$ es una clase de distribuciones muestrales $p(y|\theta)$ y $\mathscr{C}$ es una clase de distribuciones a priori para θ, entonces la clase $\mathscr{C}$ será conjugada para $\Im$ si:

$$p(\theta|y) \in \mathscr{C}, \text{ para toda } p(\bullet|\theta) \in \Im, \text{ y } p(\theta) \in \mathscr{C}.$$

Nos va a interesar tomar a la clase $\mathscr{C}$ como el conjunto de todas las densidades que tienen el mismo kernel que la función de verosimilitud.

En las siguientes secciones trabajaremos primordialmente con las distribuciones a priori conjugadas pues simplifican los cálculos computacionales y en la mayoría de los casos es posible obtener expresiones analíticas cerradas.

1.2. Estructuras de Dependencia Simples

1.2.1. Notación y Convenciones

Antes de indagar sobre las estructuras de dependencia, introduciremos la notación y convenciones que serán utilizadas a lo largo de este escrito. Comenzaremos denotanto por Y, en la mayoría de los casos, a cualquier variable o vector aleatorio, mientras que, y, denotará a cualquier valor posible de Y. Cabe mencionar que se utilizará de forma indistinta dicha notación para una variable aleatoria o vector aleatorio, misma que deberá interpretarse con base en el contexto.

Una serie de tiempo es una sucesión de variables o vectores aleatorios, y la denotaremos por $(Y_t : t = 1,2,...)$, o bien, $(Y_t)_{t\geq 1}$, donde el índice t hará refencia al tiempo. Por simplicidad, pensaremos en una sucesión de observaciones equiespaciadas (datos con periodicidad diaria, mensual, anual, etc.). Además, si la serie de tiempo es finita, la denotaremos como $(Y_t : t = 1,2,...,n)$ o $(Y_t)_{t=1}^{n}$, con $n \in \mathbb{N}$. Ahora bien, supongamos que queremos tomar una muestra de variables o vectores aleatorios, entonces denotaremos como $y_{1:n} = \{y_1, y_2, ..., y_n\}$, tal que $n \in \mathbb{N}$, a la sucesión de observaciones consecutivas tomadas de Y. A partir de este punto, nos referiremos a $y_{1:n}$, como los datos u observaciones.

Ahora bien, bajo la notación anterior podemos reescribir, los siguientes conceptos, nombrados en la sección 1.1:

- La *distribución muestral*, también conocida como la función de verosimilitud, $p(y_{1:n}|\theta)$, es la distribución de los datos, $y_{1:n}$, dado que se conoce el valor del vector de parámetros θ.

- La *distribución a priori*, $p(\theta)$, es la distribución inicial del vector de parámetros θ, o bien, la incertidumbre de θ sin los datos $y_{1:n}$.

- La *distribución posterior*, $p(\theta|y_{1:n})$, refleja nuestra creencia sobre el vector de parámetros θ, dadas nuestras observaciones $y_{1:n}$.

1.2.2. Independencia Condicional

En la práctica suponer independencia[5] de eventos o variables aleatorias es poco común; sin embargo, podemos suponer con frecuencia que dos eventos son independientes dados un tercer evento, de este hecho la importancia que cobra el concepto de independencia condicional en la teoría de probabilidad.

Definición 1.2.1. ***(Independencia Condicional)****. Sean A, B y C eventos aleatorios; se dice que A y B son condicionalmente independientes dado C, con $P(C) > 0$, si dado que el evento C ocurrió la probabilidad condicional de que A ocurra no se ve afectada por la información de que el evento B haya ocurrido. Formalmente la definimos como:*

$$P(A \cap B|C) = P(A|C)P(B|C). \tag{1.7}$$

de forma equivalente se tiene que,

$$P(A|B \cap C) = P(A|C). \tag{1.8}$$

Este concepto puede generalizarse para más de dos eventos, y dado que a lo largo de este documento haremos uso de dicha propiedad, extenderemos tal Definición para una sucesión de variables aleatorias.

Definición 1.2.2. *Se dice que $(Y_t)_{t=1}^n$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (usaremos la notación i.i.d, en lo subsecuente) dado θ si,*

$$p\left(y_{1:n}|\theta\right) = \prod_{i=1}^{n} p(y_i|\theta). \tag{1.9}$$

Además si reescribimos la *Regla de Bayes* (1.2) bajo la nueva notación y aplicamos la Definición 1.2.2, entonces tendremos que:

$$p(\theta|y_{1:n}) = \frac{p(y_{1:n}|\theta)p(\theta)}{p(y_{1:n})} = \frac{\left(\prod_{i=1}^{n} p(y_i|\theta)\right)p(\theta)}{p(y_{1:n})}. \tag{1.10}$$

Finalmente por la ecuación 1.6, obtendremos que:

$$p(\theta|y_{1:n}) \propto \left(\prod_{t=1}^{n} p(y_t|\theta)\right) p(\theta) \tag{1.11}$$

Pues notemos que la distribución de probabilidad conjunta $p(y_{1:n})$ no depende de θ, por lo que toma el rol de constante normalizadora; así la *distribución posterior* es proporcional al producto de la *distribución muestral* y la *distribución a priori*.

Cabe notar que, bajo el supuesto de independencia condicional, podemos obtener una expresión recursiva para la *distribución posterior*, es decir, podemos ir actualizando esta distribución

[5] Sean A y B, dos eventos aleatorios; entonces diremos que son independientes si $P(A \cap B) = P(A)P(B)$. De forma equivalente, si consideramos que $P(A) > 0$ o $P(B) > 0$, entonces el concepto de independencia puede definirse como $P(A|B) = P(A)$.

conforme vayamos recopilando nuevas observaciones. A manera de ejemplo, supongamos que contamos con información disponible al tiempo $(n-1)$, entonces podemos calcular la densidad condicional de θ dado $y_{1:n-1}$ como sigue:

$$p(\theta|y_{1:n-1}) \propto \left(\prod_{t=1}^{n-1} p(y_t|\theta)\right) p(\theta)$$

Dicha expresión jugará el rol de *distribución a priori* al tiempo n. Ahora bien, sea y_n la nueva observación disponible, entonces como consecuencia de la independencia condicional de $Y_{1:n}$ dado θ, podemos calcular la *distribución muestral* como,

$$p(y_n|\theta, y_{1:n-1}) = p(y_n|\theta).$$

De nuevo por el supuesto de independencia condicional y la *Regla de Bayes*, obtenemos la siguiente expresión recursiva para el cálculo de la densidad posterior,

$$p(\theta|y_{1:n-1}, y_n) \propto p(\theta|y_{1:n-1})p(y_n|\theta) \propto \left(\prod_{t=1}^{n-1} p(y_t|\theta)\right) p(\theta)p(y_n|\theta). \tag{1.12}$$

Observe que las expresiones en (1.11) y (1.12) son equivalentes.

Además, utilizando el concepto de independencia condicional, podemos predecir el valor de nuestra nueva observación, y_n, dada la información que teníamos previamente $y_1, y_2, ..., y_{n-1}$, es decir, la *distribución predictiva*, $p(y_n|y_{1:n-1})$. Que podemos expresar de la siguiente forma:

$$\begin{aligned} p(y_n|y_{1:n-1}) &= \int p(y_n, \theta|y_{1:n-1})d\theta \\ &= \int p(y_n|\theta, y_{1:n-1})p(\theta|y_{1:n-1})d\theta \\ &= \int p(y_n|\theta)p(\theta|y_{1:n-1})d\theta. \end{aligned} \tag{1.13}$$

1.2.3. Intercambiabilidad

La estructura de dependencia básica en el análisis Bayesiano es la intercambiabilidad. Ésta nos permite mostrar los aspectos esenciales en la inferencia Bayesiana y resulta apropiada cuando se cree que los datos a analizar son homogéneos. Comenzaremos introduciendo la definición más usual en términos de cursos de probabilidad básica.

Definición 1.2.3. ***(Intercambiabilidad finita)****. Diremos que las variables aleatorias $Y_1, Y_2, ..., Y_n$ son intercambiables bajo una medida de probabilidad P si su distribución conjunta satisface,*

$$P(Y_1, Y_2, ..., Y_n) = P(Y_{i_1}, Y_{i_2}, ..., Y_{i_n}).$$

Para todas las permutaciones i definidas en el conjunto $\{1, ..., n\}$ tal que $n \in \mathbb{N}$. En términos de la correspondiente función de densidad o función de masa, la condición anterior se reduce a la siguiente expresión,

$$p(Y_1, Y_2, ..., Y_n) = p(Y_{i_1}, Y_{i_2}, ..., Y_{i_n}).$$

Para efectos del tema a desarrollar, daremos una definición de intercambiabilidad para una sucesión infinita de variables aleatorias que será a la que nos referiremos en capítulos subsecuentes.

Definición 1.2.4. ***(Intercambiabilidad infinita)***. *Sea $(Y_t)_{t\geq 1}$ una sucesión de variables aleatorias; diremos que es intercambiable si para cualquier $(n \geq 1)$ tal que $n \in \mathbb{N}$, el vector $(Y_1, Y_2, ..., Y_n)$ y cualquier permutación de sus componentes $(Y_{i_1}, Y_{i_2}, ..., Y_{i_n})$, tienen la misma distribución.*

Para concluir esta sección enunciaremos el teorema de representación de Finetti, el cuál nos muestra que el supuesto de intercambiabilidad es equivalente al supuesto de independencia condicional y la distribución idéntica.

Teorema 1.2.1. ***(Teorema de Representación de Finetti)***. *Sea $(Y_t)_{t\geq 1}$ una sucesión infinita de variables aleatorias intercambiables, entonces se cumplen las siguientes igualdades:*

I. Con probabilidad uno, la sucesión de funciones de la distribución empírica,

$$F_n(y) = F_n(y; Y_1, Y_2, ..., Y_n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(-\infty, y]}(Y_i),$$

donde $\mathbb{I}_{(-\infty,y]}(Y_i)$ es la función indicadora, la cual toma el valor 1 si Y cae en el intervalo $(-\infty, y]$ y 0 en otro caso.

Converge débilmente a una función de distribución aleatoria F, cuando $n \to \infty$.

II. Para cualquier $n \geq 1$, la función de distribución de $(Y_1, Y_2, ..., Y_n)$ puede ser representada como:

$$P(Y_1 \leq y_1, Y_2 \leq y_2, ..., Y_n \leq y_n) = \int \prod_{i=1}^{n} F(y_i)\, d_p F.$$

donde p es la ley de probabilidad del límite débil F de la sucesión de funciones de la distribución empírica.

Este teorema nos dice que, si suponemos que la sucesión de variables aleatorias $(Y_t)_{t\geq 1}$, es intercambiable, entonces podemos pensarla como una sucesión de variables aleatorias condicionalmente independientes e idénticamente distribuidas dada la función de distribución F, donde F es el límite débil de las funciones de distribución empíricas. La distribución inicial p se refiere a la ley de probabilidad sobre el espacio $\mathscr{F}$, de todas las funciones de distribución sobre el espacio muestral $\mathscr{Y}$ y expresa nuestras creencias sobre el límite de las funciones de distribución empíricas.

1.2.4. Heterogeneidad

En muchos problemas, debido a la complejidad de la estructura de dependencia, resulta apropiado permitir un grado de heterogeneidad entre los datos. Una manera de hacerlo es asumiendo lo siguiente:

$$(Y_1, Y_2, ..., Y_n|\theta_1, \theta_2, ..., \theta_n) \sim \prod_{t=1}^{n} p_t\left(y_t|\theta_t\right),$$

donde $p_t\left(y_t|\theta_t\right)$ denota la función de densidad de probabilidad condicional de y_t dada θ_t.

Es decir, tenemos que $Y_1, Y_2, ..., Y_n$ son condicionalmente independientes dado un vector $\theta = (\theta_1, \theta_2, ..., \theta_n)$, donde Y_t depende únicamente del parámetro θ_t correspondiente.

1.3. La Distribución Normal

La distribución normal (algunas veces llamada la distribución Gaussiana) juega un papel fundamental en la modelación estadística, es ampliamente utilizada debido a sus propiedades matemáticas y su relativa facilidad para manipularla. Además, esta distribución posee una familiar forma de campana, cuya simetría la hace una distribución atractiva para modelos de población, pues aun cuando existen otras distribuciones con esta forma, resulta más costoso tratarlas analíticamente. Finalmente, otra de las razones a las que debe su importancia recae en el Teorema del Límite Central, que muestra que bajo ciertas condiciones, la distribución normal puede usarse para aproximar una gran variedad de distribuciones en muestras grandes (Casella and Berger, 2002).

En este sentido y con el fin de facilitar el entendimiento bajo el que se construyen los modelos dinámicos lineales que trataremos en el siguiente capítulo, veremos dos resultados de la distribución normal.

Proposición 1.3.1. ***(La distribución normal con varianza conocida)****. Sean* $\theta \sim N(m_0, C_0)$*,* $(Y_t|\theta) \sim N(\theta, \sigma^2)$*, para toda* $t = 1, 2, ...$ *tal que* σ^2 *es una constante conocida y las* Y_i*'s son independientes dado* θ*. Entonces tendremos que:*

$$\theta|y_{1:n} \sim N\left(m_n, C_n\right).$$

donde

$$m_n = E\left(\theta|y_{1:n}\right) = \frac{C_0}{C_0 + \sigma^2/n}\bar{y} + \frac{\sigma^2/n}{C_0 + \sigma^2/n}m_0, \tag{1.14}$$

$$C_n = \mathrm{Var}\left(\theta|y_{1:n}\right) = \left(\frac{n}{\sigma^2} + \frac{1}{C_0}\right)^{-1} = \frac{\sigma^2 C_0}{\sigma^2 + nC_0}. \tag{1.15}$$

Tal que $\bar{y}$ *es la media muestral, definida como* $\frac{1}{n}\sum_{i=1}^{n} y_i$*.*

Demostración. Notemos que, dada información hasta el tiempo n, podemos actualizar nuestro conocimiento sobre θ a través de la densidad posterior (1.11):

$$p(\theta|y_{1:n}) \propto \prod_{t=1}^{n} p(y_t|\theta)p(\theta),$$

donde la función de verosimilitud viene dada por:

$$p(y_{1:n}|\theta) = \prod_{t=1}^{n} p(y_t|\theta) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{1}{2\sigma^2}(y_t-\theta)^2\right\},$$

y la densidad a priori es:

$$p(\theta) = \frac{1}{\sqrt{2\pi C_0}} \exp\left\{-\frac{1}{2C_0}(\theta-m_0)^2\right\}.$$

Una vez identificados los componentes para el cálculo de la densidad posterior, desarrollamos para hallar la expresión analítica correspondiente,

$$\begin{aligned}
p(\theta|y_{1:n}) &\propto \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{1}{2\sigma^2}(y_t-\theta)^2\right\} \frac{1}{\sqrt{2\pi C_0}} \exp\left\{-\frac{1}{2C_0}(\theta-m_0)^2\right\} \\
&\propto \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{1}{2\sigma^2}(y_t^2-2\theta y_t+\theta^2)\right\} \frac{1}{\sqrt{2\pi C_0}} \exp\left\{-\frac{1}{2C_0}(\theta-m_0)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} y_t^2 - 2\theta\sum_{i=1}^{n} y_t + n\theta^2\right) - \frac{1}{2C_0}(\theta^2-2m_0\theta+m_0^2)\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_t^2 + \frac{\theta n}{\sigma^2}\sum_{i=1}^{n}\frac{y_t}{n} - \frac{n\theta^2}{2\sigma^2} - \frac{\theta^2}{2C_0} + \frac{m_0\theta}{C_0} - \frac{m_0^2}{2C_0}\right\} \\
&\propto \exp\left\{\frac{2\theta n C_0}{2\sigma^2 C_0}\overline{y} - \frac{n\theta^2 C_0}{2\sigma^2 C_0} - \frac{\theta^2\sigma^2}{2C_0\sigma^2} + \frac{2m_0\theta\sigma^2}{2C_0\sigma^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2 C_0}\left(\left(nC_0+\sigma^2\right)\theta^2 - 2\left(nC_0\overline{y}+m_0\sigma^2\right)\theta\right)\right\} \\
&\propto \exp\left\{-\frac{\left(nC_0+\sigma^2\right)\theta^2}{2\sigma^2 C_0} + \frac{2\left(nC_0\overline{y}+m_0\sigma^2\right)\theta}{2\sigma^2 C_0}\quad\frac{nC_0+\sigma^2}{nC_0+\sigma^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2 C_0/(nC_0+\sigma^2)}\left(\theta^2 - 2\frac{nC_0\overline{y}+m_0\sigma^2}{nC_0+\sigma^2}\theta + \left(\frac{nC_0\overline{y}+m_0\sigma^2}{nC_0+\sigma^2}\right)^2\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2 C_0/(nC_0+\sigma^2)}\left(\theta^2 - 2\frac{nC_0\overline{y}+m_0\sigma^2}{nC_0+\sigma^2}\theta + \left(\frac{nC_0\overline{y}+m_0\sigma^2}{nC_0+\sigma^2}\right)^2\right)\right\} \\
&\propto exp\left\{-\frac{1}{2\sigma^2 C_0/(nC_0+\sigma^2)}\left(\theta - \frac{nC_0\overline{y}+m_0\sigma^2}{nC_0+\sigma^2}\right)^2\right\}.
\end{aligned}$$

En la última expresión podemos reconocer el kernel de una distribución normal con media m_n y varianza C_n, es decir, $\theta|y_{1:n} \sim N(m_n, C_n)$, con parámetros:

$$m_n = E(\theta|y_{1:n}) = \frac{C_0}{C_0 + \sigma^2/n}\bar{y} + \frac{\sigma^2/n}{C_0 + \sigma^2/n}m_0,$$

$$C_n = \text{Var}(\theta|y_{1:n}) = \left(\frac{n}{\sigma^2} + \frac{1}{C_0}\right)^{-1} = \frac{\sigma^2 C_0}{\sigma^2 + nC_0}.$$

□

De la ecuación (1.14) de la proposición anterior, 1.3.1, observemos que la media posterior es un promedio ponderado de la media muestral $\bar{y}$ y la media inicial m_0, cuyos pesos dependen de los valores de las varianzas C_0 y σ^2. Esto da lugar a los siguientes casos:

1. Si la incertidumbre inicial, C_0, es pequeña comparada con σ^2, la media inicial, m_0, recibirá más peso.
2. Si por el contrario C_0 es muy grande, entonces $m_n \simeq \bar{y}$ y $C_n \simeq \sigma^2/n$.[6]

Como vimos en la sección 1.2.2, al tener una nueva observación, y_n, la *distribución posterior* puede obtenerse de forma recursiva utilizando la ecuación (1.12), tomando la densidad condicional de θ dada la información hasta el tiempo $n-1$ como la *distribución a priori*, es decir, $\theta|y_{1:n-1} \sim N(m_{n-1}, C_{n-1})$ y como la función de verosimilitud a la densidad condicional $Y_n|\theta \sim N\left(\theta, \sigma^2\right)$. Además, por el resultado de la Proposición 1.3.1 sabemos que si la *distribución a priori* y la función de verosimilitud siguen una distribución normal, entonces la *distribución posterior* (de la media) $p(\theta|y_{1:n})$ seguirá una distribución normal con los siguientes parámetros:

$$\begin{aligned} m_n &= \frac{C_{n-1}}{C_{n-1} + \sigma^2} y_n + \frac{\sigma^2}{C_{n-1} + \sigma^2} m_{n-1} \\ &= \frac{C_{n-1}}{C_{n-1} + \sigma^2} y_n + \left(1 - \frac{C_{n-1}}{C_{n-1} + \sigma^2}\right) m_{n-1} \\ &= \frac{C_{n-1}}{C_{n-1} + \sigma^2} (y_n - m_{n-1}) + m_{n-1}, \end{aligned} \tag{1.16}$$

$$C_n = \text{Var}(\theta|y_{1:n}) = \left(\frac{1}{\sigma^2} + \frac{1}{C_{n-1}}\right)^{-1} = \frac{\sigma^2 C_{n-1}}{\sigma^2 + C_{n-1}}. \tag{1.17}$$

Como veremos a continuación, uno de los problemas estadísticos más comunes al trabajar con series de tiempo es hacer predicciones a futuro, por lo que suele ser de interés calcular la *distribución predictiva* de una observación futura, digamos y_n. Utilizando la distribución condicional $y_n|\theta \sim N\left(\theta, \sigma^2\right)$, la distribución posterior $\theta|y_{1:n-1} \sim N(m_{n-1}, C_{n-1})$, y el resultado 1.13 se sigue lo siguiente:

[6] El símbolo $\simeq$ significa "aproximadamente ".

$$p(y_n|y_{1:n-1}) = \int p(y_n|\theta)p(\theta|y_{1:n-1})d\theta$$
$$\propto \int \exp\left(-\frac{1}{2\sigma^2}(y_n-\theta)^2\right)\exp\left(-\frac{1}{2C_{n-1}}(\theta-m_{n-1})^2\right)d\theta.$$

Aunque pareciera un cálculo complicado, se pueden utilizar las propiedades de la distribución normal bivariada para facilitarlo. Observemos que el producto en el integrando es la exponencial de una función cuadrática que depende de (y_n,θ), así por las propiedades de la distribución normal, y_n y θ siguen una distribución posterior conjunta normal y se cumple que la distribución posterior marginal de y_n también seguirá una distribución normal.

Para determinar los parámetros de la *distribución predictiva* haremos uso de la esperanza y varianza condicional, así como del hecho que, $y_n|\theta \sim N\left(\theta,\sigma^2\right)$, es decir, $E\left(y_n|\theta\right)=\theta$ y la $\text{Var}\left(y_n|\theta\right)=\sigma^2$ y $\theta|y_{1:n-1} \sim N\left(m_{n-1},C_{n-1}\right)$, es decir, $E\left(y_n|\theta\right)=m_{n-1}$ y la $\text{Var}\left(y_n|\theta\right)=C_{n-1}$; por lo tanto se tienen los siguientes resultados,

$$E\left(y_n|y_{1:n-1}\right) = E\left(E\left(y_n|\theta,y_{1:n-1}\right)|y_{1:n-1}\right) = E\left(\theta|y_{1:n-1}\right) = m_{n-1}, \tag{1.18}$$

$$\begin{aligned}\text{Var}\left(y_n|y_{1:n-1}\right) &= E\left(\text{Var}\left(y_n|\theta,y_{1:n-1}\right)|y_{1:n-1}\right) + \text{Var}\left(E\left(y_n|\theta,y_{1:n-1}\right)|y_{1:n-1}\right)\\ &= E\left(\sigma^2|y_{1:n-1}\right) + \text{Var}\left(\theta|y_{1:n-1}\right)\\ &= \sigma^2 + C_{n-1}.\end{aligned} \tag{1.19}$$

Note que (1.18) es la media posterior m_{n-1}. Mientras que la ecuación (1.19), de la varianza de la *distribución predictiva*, es la suma de la varianza inicial σ^2 y la varianza posterior C_{n-1}.

Capítulo 2

Una Mirada hacia los Modelos Dinámicos Lineales

Los modelos de espacio de estados se originaron en el campo de la ingeniería en los años 60's; sin embargo el problema de pronóstico de observaciones ha sido un tópico fundamental y fascinante en la teoría de procesos estocásticos y series de tiempo. Es así que, una década después aparecen en la literatura de series de tiempo y comienzan a establecerse durante los años 80's (Campagnoli et al., 2009).

A grandes rasgos esta clase de modelos considera como una serie de tiempo a los resultados de un sistema dinámico perturbado por alteraciones alteatorias, lo que permite la interpretación natural de una serie de tiempo como una combinación lineal de diversos componentes, tales como tendencia, estacionariedad, cambios estructurales y patrones irregulares.

A lo largo de este capítulo describirmos de manera general la estructura de los modelos de espacio de estados, los procedimientos de estimación utilizados: filtración, suavizamiento y predección, y nos enfocaremos en un caso particular de éstos, los llamados Modelos Dinámicos Lineales.

2.1. Modelos de Espacio de Estados

Antes de abordar los modelos de espacio de estados, cabe recordar la definición de una cadena de Markov:

Definición 2.1.1. ***(Cadena de Markov)***. *Sea* $(Y_t)_{t\geq 0}$ *una sucesión de variables aleatorias, diremos que es una cadena de Markov si satisface que,*

$$p\left(y_{t+1}|y_{0:t}\right) = p\left(y_{t+1}|y_t\right).$$

Observe que la Definición 2.1.1 nos dice que la información acerca de una nueva observación (una observación futura), $Y_{t+1} = y_{t+1}$, solo depende de la información proporcionada por una observación anterior $Y_t = y_t$ y que, dada Y_t, la información pasada $Y_{0:t-1} = y_{0:t-1}$ es irrelevante.

Por otro lado, vale la pena recordar que la distribución conjunta de esta serie de variables aleatorias puede escribirse como:

$$\begin{aligned} p(y_{0:t+1}) &= p(y_0)\, p(y_1|y_0) \ldots p(y_t|y_{t-1}) \\ &= p(y_0) \prod_{i=1}^{t+1} p(y_i|y_{i-1}). \end{aligned}$$

Sin más preámbulos veamos la definición de los modelos de espacio de estado.

Definición 2.1.2. ***(Modelo de Espacio de Estados)***. *Sean $(\theta_t)_{t\geq 1}$ y $(Y_t)_{t\geq 1}$ dos series de tiempo en $\mathbb{R}^p$ y $\mathbb{R}^m$ respectivamente. Diremos que forman un Modelo de Espacio de Estados si satisfacen lo siguiente:*

I. $(\theta_t)_{t\geq 1}$ *es una cadena de Markov.*

II. Dada $(\theta_t)_{t\geq 1}$, *las* Y_t'*s son independientes y* Y_t *solo depende de* θ_t.

En este contexto, uno puede pensar en $(\theta_t)_{t\geq 1}$ como una serie de tiempo auxiliar para especificar con mayor facilidad la distribución de probabilidad de la serie de tiempo observable $(Y_t)_{t\geq 1}$. De hecho como consecuencia de I y II en la Definición 2.1.2 un modelo de espacio de estados queda totalmente identificado por la distribución inicial $p(\theta_0)$ y las probabilidades condicionales $p(\theta_t|\theta_{t-1})$ y $p(y_t|\theta_t)$ para toda $t \geq 1$ y se cumple que,

$$p(\theta_{0:t}, y_{1:t}) = p(\theta_0) \left(\prod_{i=1}^{t} p(\theta_i|\theta_{i-1})\, p(y_i|\theta_i) \right), \quad t \geq 1. \tag{2.1}$$

Un dato importante es que, a los modelos de espacio de estados en los que los estados toman valores dentro de un conjunto de variables aleatorias discretas, suele conocerceles como "Modelos Ocultos de Markov".

Por otro lado, la estructura de dependencia de un modelo de espacio de estados se puede representar gráficamente por la Figura 2.1. Esta se basa en el hecho de que existe una cadena de Markov no observable $(\theta_t)_{t\geq 1}$, mejor conocido como proceso de estados, tal que Y_t es una medida imprecisa de θ_t, siendo fundamental la independencia condicional entre Y_t y $(\theta_0, \ldots, \theta_{t-1}, Y_0, \ldots, Y_{t-1})$ dado θ_t.

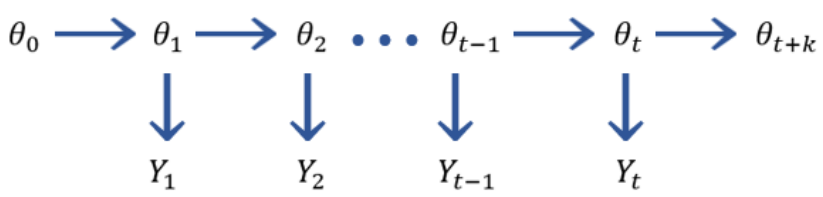


Figura 2.1: Estructura de dependencia de un modelo de espacio de estados.

Dada la gran flexibilidad de los modelos de espacios de estados, éstos han ido extendiendo su popularidad y cada vez son más utilizados en diversas áreas para resolver problemas aplicados; sin embargo, como en cualquier problema estadístico, la especificación del modelo suele ser complicada. De hecho los problemas más comunes a la hora de construir el modelo son:

i) Interpretación difusa sobre los estados físicos.

ii) Que la representación del espacio de estados no sea única.

iii) Un espacio de estados muy grande y pobremente identificable.

iv) Un modelo sumamente complicado.

Por ahora nos enfocaremos en las dos tareas principales de los modelos de espacio de estados, que son:

❶ **Inferencia sobre los estados no observables**. Se resuelve estimando el vector de estados a través de la densidad condicional $p(\theta_s|y_{1:t})$, tal que $t,s \in \mathbb{N}$, utilizando las técnicas de filtración ($s=t$) y suavizamiento ($s<t$).

❷ **Predicción de observaciones futuras**. Se resuelve estimando la evolución del sistema, $p(\theta_{t+k}|y_{1:t})$, en primera instancia, para después calcular la densidad predictiva $p(y_{t+k}|y_{1:t})$ k pasos hacia adelante, con $k \geq 1$.

A continuación detallaremos las técnicas mencionadas y el pronóstico de un modelo de espacio de estados.

2.1.1. Filtración

Antes de definir el problema de filtración formalmente, pensemos en una aplicación financiera para ejemplificarlo. Supongamos que un banco busca estimar la estructura de las tasas de interés colocadas diariamente, actualizando las estimaciones actuales mientras nueva información de mercado esté disponible. En otras palabras busca un procedimiento tal que estime el valor actual del vector de estados, con base en las observaciones hasta el tiempo t (el presente) y que actualice sus estimaciones y pronósticos al incorporarse información nueva al tiempo $t+1$ (futuro). Un grupo de analistas a los que se les presentó tal problema concluyen que para resolverlo basta estimar el vector de estados θ_t dada la información recolectada hasta el momento, y poder actualizar dicha estimación conforme nueva información se vuelva disponible, es decir, calcular la densidad condicional: $p(\theta_s|y_{1:t})$ tal que $s=t$ y actualizarla calculando $p(\theta_{t+1}|y_{1:t+1})$.

Formalmente para resolver este problema en el que los datos llegan secuencialmente, y gracias a la estructura Markoviana de la dinámica de los estados y el supuesto de independencia condicional en el proceso observable, las densidades de filtración para un modelo de espacio de estados pueden calcularse de manera recursiva como se enuncia a continuación.

Proposición 2.1.1. ***(Densidades de Filtración)***.

i) La densidad predictiva un paso hacia adelante para los estados puede calcularse como,

$$p(\theta_t|y_{1:t-1}) = \int p(\theta_t|\theta_{t-1})p(\theta_{t-1}|y_{1:t-1})\,d\theta_{t-1}, \tag{2.2}$$

donde $p(\theta_{t-1}|y_{1:t-1})$ es la densidad de filtración al tiempo t-1.

ii) La densidad predictiva un paso hacia adelante para las observaciones puede calcularse a partir de i) como,

$$p(y_t|y_{1:t-1}) = \int p(y_t|\theta_t)p(\theta_t|y_{1:t-1})\,d\theta_t. \tag{2.3}$$

iii) La densidad de filtración puede calcularse utilizando i) y ii) como,

$$p(\theta_t|y_{1:t}) = \frac{p(y_t|\theta_t)\,p(\theta_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}. \tag{2.4}$$

Demostración. i) Densidad predictiva de los estados.

$$\begin{aligned} p(\theta_t|y_{1:t-1}) &= \int p(\theta_t,\theta_{t-1}|y_{1:t-1})d\theta_{t-1} \\ &= \int p(\theta_t|\theta_{t-1},y_{1:t-1})p(\theta_{t-1}|y_{1:t-1})\,d\theta_{t-1} \\ &= \int p(\theta_t|\theta_{t-1})p(\theta_{t-1}|y_{1:t-1})\,d\theta_{t-1}. \end{aligned}$$

La última igualdad se sigue del hecho de que θ_t y $Y_{1:t-1}$ son condicionalmente independientes dado θ_{t-1}.

ii) Densidad predictiva de las observaciones.

$$\begin{aligned} p(y_t|y_{1:t-1}) &= \int p(y_t,\theta_t|y_{1:t-1})d\theta_t \\ &= \int p(y_t|\theta_t,y_{1:t-1})p(\theta_t|y_{1:t-1})\,d\theta_t \\ &= \int p(y_t|\theta_t)p(\theta_t|y_{1:t-1})\,d\theta_t. \end{aligned}$$

La última igualdad se sigue del hecho de que Y_t y $Y_{1:t-1}$ son condicionalmente independientes dado θ_t.

iii) Densidad de filtración.

$$\begin{aligned} p(\theta_t|y_{1:t}) &= \frac{p(\theta_t, y_{1:t})}{p(y_{1:t})} \\ &= \frac{p(y_t|\theta_t, y_{1:t-1})\, p(\theta_t, y_{1:t-1})}{p(y_{1:t})} \\ &= \frac{p(y_t|\theta_t)\, p(\theta_t|y_{1:t-1})}{p(y_t|y_{1:t-1})}. \end{aligned}$$

La segunda y tercera igualdad se siguen de la *Regla de Bayes* y del hecho de que Y_t y $Y_{1:t-1}$ son condicionalmente independientes dado θ_t, respectivamente.

□

2.1.2. Suavizamiento

El problema de suavizamiento o análisis retrospectivo consiste en obtener estimadores para θ_t con base en una muestra de datos completa, digamos $(Y_t : t = 1, 2, ..., T)$. Estos estimadores son llamados suavizadores, pues una gráfica de tiempo de la serie θ_t es típicamente "más suave" que la serie predictiva o de filtraciones.

En otras palabras, en el análisis de series de tiempo suele pasar que se tienen observaciones Y_t para un cierto período de tiempo y se quiere reconstruir retrospectivamente el comportamiento del sistema, para estudiar la construcción socio-económica o el fenómeno de las observaciones subyacentes. En este caso se busca un algortimo recursivo hacia atrás para calcular la distribución condicional $p(\theta_t|y_{1:T})$ tal que $t < T$. Tomado como punto de partida el cálculo de la densidad de filtración $p(\theta_T|y_{1:T})$ y estimar a partir de esta las densidades hacia atrás de los estados dada la información disponible, así la primera densidad recursiva sería $p(\theta_{T-1}|y_{1:T})$, luego $p(\theta_{T-2}|y_{1:T})$, entre otras. A continuación damos la definición formal del algoritmo recursivo hacia atrás.

Proposición 2.1.2. ***(Algoritmo recursivo de suavizamiento)****. Considere un modelo de espacio de estados como el de la Definición 2.1.2, entonces se cumple lo siguiente:*

i) Las distribuciones condicionales de transición hacia atrás de la secuencia de estados $(\theta_1, ..., \theta_T)$, dada la información disponible hasta el tiempo T, $y_{1:T}$, pueden obtenerse como,

$$p(\theta_t|\theta_{t+1}, y_{1:T}) = \frac{p(\theta_{t+1}|\theta_t)\, p(\theta_t|y_{1:t})}{p(\theta_{t+1}|y_{1:t})} \text{ para toda } t = 1, 2, ..., T-1. \tag{2.5}$$

ii) Las distribuciones de suavizamiento de θ_t dada la información hasta el tiempo T, $y_{1:T}$, pueden calcularse de acuerdo a la siguiente ecuación recursiva hacia atrás en t, comenzando con $p(\theta_T|y_{1:T})$:

$$p(\theta_t|y_{1:T}) = p(\theta_t|y_{1:t}) \int \frac{p(\theta_{t+1}|\theta_t)}{p(\theta_{t+1}|y_{1:t})} p(\theta_{t+1}|y_{1:T})\, d\theta_{t+1}. \tag{2.6}$$

Demostración. i)

$$\begin{aligned} p(\theta_t|\theta_{t+1}, y_{1:T}) &= p(\theta_t|\theta_{t+1}, y_{1:t}) \\ &= \frac{p(\theta_t, \theta_{t+1}, y_{1:t})}{p(\theta_{t+1}, y_{1:t})} \\ &= \frac{p(\theta_{t+1}|\theta_t, y_{1:t})\, p(\theta_t|y_{1:t})}{p(\theta_{t+1}|y_{1:t})} \\ &= \frac{p(\theta_{t+1}|\theta_t)\, p(\theta_t|y_{1:t})}{p(\theta_{t+1}|y_{1:t})}. \end{aligned}$$

Aplicando la *Regla de Bayes* en la tercera igualdad y considerando el hecho de que θ_t y $Y_{t+1:T}$ son condicionalmente independientes dado θ_{t+1} y θ_{t+1} y $Y_{1:T}$ son condicionalmente independientes dado θ_t llegamos a la igualdad deseada.

ii)

$$\begin{aligned} p(\theta_t|y_{1:T}) &= \int p(\theta_t, \theta_{t+1}|y_{1:T})\, d\theta_{t+1} \\ &= \int p(\theta_t|\theta_{t+1}, y_{1:T})\, p(\theta_{t+1}|y_{1:T})\, d\theta_{t+1} \\ &= \int \frac{p(\theta_{t+1}|\theta_t)\, p(\theta_t|y_{1:t})}{p(\theta_{t+1}|y_{1:t})} p(\theta_{t+1}|y_{1:T})\, d\theta_{t+1} \\ &= p(\theta_t|y_{1:t}) \int \frac{p(\theta_{t+1}|y_{1:T})}{p(\theta_{t+1}|y_{1:t})} p(\theta_{t+1}|\theta_t)\, d\theta_{t+1}. \end{aligned}$$

Marginalizando con respecto a θ_{t+1} y utilizando el resultado de i) en la tercera igualdad obtenemos el resultado.

□

2.1.3. Predicción

En el análisis de series de tiempo la predicción de observaciones futuras es de gran interés. Pensemos por ejemplo, en el problema financiero de conocer el precio de alguna acción día a día, con el fin de que los inversionistas obtengan mayores ganancias, sabiendo cuándo vender y/o comprar. Entonces, dada la información disponible sobre el precio de las acciones hasta el tiempo t (el presente) les gustaría conocer el precio futuro de éstas al siguiente día, Y_{t+1}, o mejor aún k días hacia adelante, Y_{t+k}. Además bajo el contexto de los modelos de espacio de estados también es de interés pronosticar los estados k pasos hacia adelante, $k \geq 0$, para estimar la evolución del sistema.

Ambos problemas pueden resolverse partiendo de la densidad condicional un paso hacia adelante, al ir actualizándola con la nueva información secuencialmente, y la densidad de filtración, como veremos a continuación.

Proposición 2.1.3. ***(Algoritmo recursivo de predicción)***. *Considere un modelo de espacio de estados como el de la Definición 2.1.2. Entonces para k>0 se cumplirá lo siguiente:*

i) La distribución predictiva k pasos hacia adelante para los estados es:

$$p(\theta_{t+k}|y_{1:t}) = \int p(\theta_{t+k}|\theta_{t+k-1})\, p(\theta_{t+k-1}|y_{1:t})\, d\theta_{t+k-1}. \tag{2.7}$$

ii) La distribución predictiva k pasos hacia adelante para las observaciones es:

$$p(y_{t+k}|y_{1:t}) = \int p(y_{t+k}|\theta_{t+k})\, p(\theta_{t+k}|y_{1:t})\, d\theta_{t+k}. \tag{2.8}$$

Demostración. i)

$$\begin{aligned} p(\theta_{t+k}|y_{1:t}) &= \int p(\theta_{t+k}, \theta_{t+k-1}|y_{1:t})\, d\theta_{t+k-1} \\ &= \int p(\theta_{t+k}|\theta_{t+k-1}, y_{1:t})\, p(\theta_{t+k-1}|y_{1:t})\, d\theta_{t+k-1} \\ &= \int p(\theta_{t+k}|\theta_{t+k-1})\, p(\theta_{t+k-1}|y_{1:t})\, d\theta_{t+k-1}. \end{aligned}$$

La última igualdad se sigue del hecho de que θ_{t+k} y $Y_{1:t}$ son independientes condicionalmente dada θ_{t+k-1}.

ii)

$$\begin{aligned} p(y_{t+k}|y_{1:t}) &= \int p(y_{t+k}, \theta_{t+k}|y_{1:t})\, d\theta_{t+k} \\ &= \int p(y_{t+k}|\theta_{t+k}, y_{1:t})\, p(\theta_{t+k}|y_{1:t})\, d\theta_{t+k} \\ &= \int p(y_{t+k}|\theta_{t+k})\, p(\theta_{t+k}|y_{1:t})\, d\theta_{t+k}. \end{aligned}$$

La última igualdad se sigue del hecho de que Y_{t+k} y $Y_{1:t}$ son independientes condicionalmente dada θ_{t+k}.

□

2.2. Modelos Dinámicos Lineales

Los modelos dinámicos lineales, también conocidos como "Modelos de Espacio de Estados Lineales Gaussianos" por ser un caso particular de los modelos de espacio de estados cuando se cumplen las propiedades de linealidad y normalidad, han ganado popularidad en años recientes en diversas áreas de estudio, siendo la ecología una de ellas, pues ofrecen un campo versátil y robusto para el ajuste de modelos que varían a través del tiempo (West and Harrison 1997).

En este trabajo, adoptaremos dicho enfoque y nos adentraremos en el estudio de éstos para modelar la dependencia temporal de las concentraciones de ozono en la estación el Pedregal en la Ciudad de México.

Definición 2.2.1. ***(Modelo Dinámico Lineal)***. *Sean Y_t y θ_t vectores aleatorios de dimensiones m y p, respectivamente. Diremos que un modelo dinámico lineal está dado por las siguientes ecuaciones,*

$$Y_t = F_t\theta_t + \nu_t, \quad \nu_t \sim N_m(0, V_t). \tag{2.9}$$
$$\theta_t = G_t\theta_{t-1} + w_t, \quad w_t \sim N_p(0, W_t). \tag{2.10}$$

donde F_t y G_t son matrices conocidas de dimensiones $m \times p$ y $p \times p$ respectivamente, $(v_t)_{t\geq 1}$ y $(w_t)_{t\geq 1}$ son series independientes de vectores aleatorios con una distribución normal con media cero y matrices de varianza $(V_t)_{t\geq 1}$ y $(W_t)_{t\geq 1}$ respectivamente. Además sea,

$$\theta_0 \sim N_p(m_0, C_0).$$

con θ_0 independiente de los ruidos aleatorios, $(v_t)_{t\geq 1}$ y $(w_t)_{t\geq 1}$.

Diremos que un modelo dinámico lineal está totalmente especificado si se conocen los siguientes componentes: m_0, C_0, F_0, G_0, V_t y W_t.

Una representación alternativa de las ecuaciones (2.9) y (2.10) es la siguiente,

$$Y_t|\theta_t \sim N(F_t\theta_t, V_t). \tag{2.11}$$
$$\theta_t|\theta_{t-1} \sim N(G_t\theta_{t-1}, W_t). \tag{2.12}$$

La ecuación (2.9) se llama la *ecuación de observación* del modelo, ya que corresponde a la distribución muestral de los datos, mientras que la ecuación (2.10) se conoce como *ecuación del sistema* o *ecuación de estados*, pues define la evolución en el tiempo del vector de estados.

2.2.1. El Filtro de Kalman

En la sección anterior el problema de filtración en el que los datos llegan secuencialmente podía resolverse a través de ecuaciones recursivas un tanto complejas. En este caso gracias a la normalidad de los modelos dinámicos lineales dichos cálculos se simplifican y pueden ser resueltos por el filtro de Kalman.

Para facilitar su cálculo demostraremos la siguiente proposición como una generalización del caso multivariado de la distribución condicional de $\theta|y_{1:n} \sim N(m_n, C_n)$, vista en la Proposición 1.3.1. (página 8).

Proposición 2.2.1. *Sea Y un vector aleatorio tal que $Y|\theta \sim N_m(F\theta, \Sigma)$, donde F es una matriz conocida de dimensiones $m \times p$, θ un vector aleatorio de dimensión p y Σ es la matriz de covarianzas, si la distribución a priori $\theta \sim N_p(a, R)$, entonces*

$$\theta|Y \sim N_p(\mu, C).$$

donde

$$\mu = C\left(F^T\Sigma^{-1}y + R^{-1}a\right). \tag{2.13}$$

y

$$C = \left(F^T\Sigma^{-1}F + R^{-1}\right)^{-1}. \tag{2.14}$$

Demostración. Por la ecuación (1.11) sabemos que la densidad posterior de θ es proporcional al producto de la función de verosimilitud y la *distribución a priori*, es decir,

$$p(\theta|y) \propto p(y|\theta)\, p(\theta),$$

donde la función de verosimilitud está dada por:

$$\begin{aligned}
p(y|\theta) &= \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(y - F\theta)^T \Sigma^{-1}(y - F\theta)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(y^T - \theta^T F^T\right)\Sigma^{-1}(y - F\theta)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(y^T\Sigma^{-1} - \theta^T F^T\Sigma^{-1}\right)(y - F\theta)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(y^T\Sigma^{-1}y - y^T\Sigma^{-1}F\theta - \theta^T F^T\Sigma^{-1}y + \theta^T F^T\Sigma^{-1}F\theta\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(y^T\Sigma^{-1}y - 2\theta^T F^T\Sigma^{-1}y + \theta^T F^T\Sigma^{-1}F\theta\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\theta^T F^T\Sigma^{-1}F\theta - 2\theta^T F^T\Sigma^{-1}y\right)\right\}.
\end{aligned}$$

Note que $|\Sigma|$ denota el determinante de Σ, mientras que la quinta igualdad se sigue del hecho de que Σ^{-1} es una matriz simétrica[1], por lo que $y^T\Sigma^{-1}F\theta = (\theta^T F^T\Sigma^{-1}y)^T$.

Por otro lado, la densidad a priori es:

$$\begin{aligned}
p(\theta) &= \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{|R|^{1/2}} \exp\left\{-\frac{1}{2}(\theta - a)^T R^{-1}(\theta - a)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\theta^T R^{-1}\theta - 2\theta^T R^{-1}a\right)\right\}.
\end{aligned}$$

Analogamente note que $|R|$ denota el determinante de R, y dado que R^{-1} también es simétrica, se cumple que $a^T R^{-1}\theta = (\theta^T R^{-1}a)^T$.

[1] Sea $A_{nxm} = (a_{ij})$, diremos que es simétrica si cumple que $A^T = A$

De los resultados anteriores se sigue que,

$$\begin{aligned} p(\theta|Y) &\propto \exp\left\{-\frac{1}{2}\left(\theta^T F^T \Sigma^{-1} F\theta - 2\theta^T F^T \Sigma^{-1} y\right)\right\} \exp\left\{-\frac{1}{2}\left(\theta^T R^{-1}\theta - 2\theta^T R^{-1} a\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\theta^T F^T \Sigma^{-1} F\theta - \theta^T F^T \Sigma^{-1} y - \frac{1}{2}\theta^T R^{-1}\theta - \theta^T R^{-1} a\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left(\theta^T\left(F^T \Sigma^{-1} F + R^{-1}\right)\theta\right) - 2\theta^T\left(F^T \Sigma^{-1} y + R^{-1} a\right)\right\}. \end{aligned}$$

Si definimos a:

$$C = \left(F^T \Sigma^{-1} F + R^{-1}\right)^{-1},$$

$$\mu = C\left(F^T \Sigma^{-1} y + R^{-1} a\right),$$

entonces,

$$\begin{aligned} p(\theta|Y) &\propto \exp\left\{-\frac{1}{2}\left(\theta^T C^{-1}\theta - 2\theta^T C^{-1}\mu\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\theta-\mu)^T C^{-1}(\theta-\mu)\right\}, \end{aligned}$$

por lo tanto,

$$\theta|Y \sim N_p(\mu, C).$$

□

Proposición 2.2.2. ***(Filtro de Kalman)****. Considere un modelo dinámico lineal como el de la Definición 2.2.1, existe t tal que $\theta_{t-1}|y_{1:t-1}$ sigue una distribución normal com media m_{t-1} y varianza C_{t-1}, es decir que:*

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

Entonces se cumple lo siguiente:

i) La distribución predictiva un paso hacia adelante de $\theta_t|y_{1:t-1}$ sigue una distribución normal con media a_t y varianza R_t, dados por las siguientes expresiones :

$$a_t = E(\theta_t|y_{1:t-1}) = G_t m_{t-1}. \tag{2.15}$$

$$R_t = \mathrm{Var}(\theta_t|y_{1:t-1}) = G_t C_{t-1} G_t^T + W_t. \tag{2.16}$$

ii) La distribución predictiva un paso hacia adelante de $Y_t|y_{1:t-1}$ sigue una distribución normal con media f_t y varianza Q_t, dados por las siguientes expresiones:

$$f_t = E(Y_t|Y_{1:t-1}) = F_t a_t. \tag{2.17}$$

$$Q_t = \mathrm{Var}(Y_t|Y_{1:t-1}) = F_t R_t F_t^T + V_t. \tag{2.18}$$

iii) La distribución de filtración de $\theta_t|y_{1:t}$ sigue una distribución normal con media m_t y varianza C_t, dados por las siguientes expresiones:

$$m_t = E(\theta_t|y_{1:t}) = a_t + R_t F_t^T Q_t^{-1} e_t. \tag{2.19}$$
$$C_t = \text{Var}(\theta_t|y_{1:t}) = R_t - R_t F_t^T Q_t^{-1} F_t R_t. \tag{2.20}$$

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

Demostración. Antes de dar la demostración a los incisos i), ii) y iii), cabe recordar que como mencionamos al inicio de este capítulo, los modelos dinámicos lineales son un caso especial de los modelos de espacio de estados, por lo que al igual que en los primeros nos va a interesar resolver los problemas de filtración, suavizamiento y pronóstico. Sin embargo, gracias a los supuestos de linealidad y normalidad bajo los que se construyen es más sencillo calcular las densidades recursivas generales. De hecho por las propiedades de la distribución normal multivariada, se puede probar que el vector aleatorio $(\theta_0, ..., \theta_t, Y_1, ..., Y_t)$ sigue una distribución normal para toda $t \geq 1$. Y se sigue que las distribuciones condicionales y marginales también resultan seguir tal distribución (Anderson, 2003).

i) Sea $\theta_t|y_{1:t-1} \sim N(a_t, R_t)$, además utilizando la Definición 2.2.1, tenemos que $\theta_t = G_t\theta_{t-1} + w_t$ por la ecuación (2.10), G_t es una matriz conocida y $(w_t)_{t\geq 1}$ (el ruido) es un vector aleatorio independiente con media cero y varianza W_t; entonces, aplicando la esperanza y varianza iteradas, se tiene que:

$$\begin{aligned}
a_t &= E[\theta_t|y_{1:t-1}] \\
&= E[E[\theta_t|\theta_{t-1}, y_{1:t-1}]|y_{1:t-1}]. \\
&= E[E[G_t\theta_{t-1} + w_t|\theta_{t-1}, y_{1:t-1}]|y_{1:t-1}] \\
&= G_t E[\theta_{t-1}|y_{1:t-1}] + E[w_t] \\
&= G_t E[\theta_{t-1}|y_{1:t-1}] \\
&= G_t m_{t-1}.
\end{aligned}$$

Para obtener la varianza consideremos lo siguiente,

Usando la ecuación alternativa para la ecuación de estados (2.12), tenemos que:

$$\text{Var}[\theta_t|\theta_{t-1}, y_{1:t-1}] = W_t.$$

Entonces,

$$\begin{aligned}
R_t &= \text{Var}\left[\theta_t | y_{1:t-1}\right] \\
&= E\left[\text{Var}\left[\theta_t | \theta_{t-1}, y_{1:t-1}\right] | y_{1:t-1}\right] + \text{Var}\left[E\left[\theta_t | \theta_{t-1}, y_{1:t-1}\right] | y_{1:t-1}\right] \\
&= E\left[W_t | y_{1:t-1}\right] + \text{Var}\left[E\left[G_t\theta_{t-1} + w_t | \theta_{t-1}, y_{1:t-1}\right] | y_{1:t-1}\right] \\
&= W_t + \text{Var}\left[G_t\theta_{t-1} | y_{1:t-1}\right] \\
&= W_t + G_t \,\text{Var}\left[\theta_{t-1} | y_{1:t-1}\right] G_t^T \\
&= W_t + G_t C_{t-1} G_t^T .
\end{aligned}$$

Note que la última igualdad es válida, pues por hipótesis $\theta_{t-1}|y_{1:t-1}$ sigue una distribución normal con media m_{t-1} y varianza C_{t-1}.

ii) Sea $Y_t|y_{1:t-1} \sim N(f_t, Q_t)$, además utilizando la Definición 2.2.1, tenemos que $Y_t = F_t\theta_t + v_t$ por la ecuación (2.9), F_t es una matriz conocida y $(v_t)_{t\geq 1}$ (el ruido) es un vector aleatorio independiente con media cero y varianza V_t; entonces, aplicando la esperanza y varianza iteradas, se tiene que:

$$\begin{aligned}
f_t &= E\left[Y_t | y_{1:t-1}\right] \\
&= E\left[E\left[Y_t | \theta_t, y_{1:t-1}\right] | y_{1:t-1}\right] \\
&= E\left[E\left[F_t\theta_t + v_t | \theta_t, y_{1:t-1}\right] | y_{1:t-1}\right] \\
&= F_t E\left[\theta_t | y_{1:t-1}\right] \\
&= F_t a_t .
\end{aligned}$$

La última igualdad se cumple por el inciso i) de la demostración anterior, que nos dice que $\theta_t|y_{1:t-1}$ sigue una distribución normal con media a_t.

Por otro lado, para obtener la varianza considere la ecuación alternativa para la ecuación de observaciones (2.11), entonces tenemos que:

$$\text{Var}\left[Y_t | \theta_t, y_{1:t-1}\right] = V_t .$$

$$\begin{aligned}
Q_t &= \text{Var}\left[Y_t | y_{1:t-1}\right] \\
&= E\left[\text{Var}\left[Y_t | \theta_t, y_{1:t-1}\right] | y_{1:t-1}\right] + \text{Var}\left[E\left[F_t\theta_t + v_t | \theta_t, y_{1:t-1}\right] | y_{1:t-1}\right] \\
&= E\left[V_t | y_{1:t-1}\right] + \text{Var}\left[F_t\theta_t | y_{t-1}\right] \\
&= V_t + F_t \,\text{Var}\left[\theta_t | y_{t-1}\right] F_t^T \\
&= V_t + F_t R_t F_t^T .
\end{aligned}$$

La última igualdad se cumple por el inciso i) de la demostración anterior, que nos dice que $\theta_t|y_{1:t-1}$ sigue una distribución normal con varianza R_t.

iii) Consideremos la Proposición 2.2.1 tal que $(\theta|Y) \sim N(m_t, C_t)$ cuyos parámetros están dados por las ecuaciones (2.13) y (2.14), en particular tomemos a $p(\theta|y_{1:t-1})$ como la *distribución a priori* que sigue una distribución normal con media a_t y varianza R_t, demostrado en el inciso i). Además sea $p(y_t|\theta_t)$ la función de verosimilitud con distribución normal con media $F_t\theta_t$ y varianza V_t, (2.12).

Entonces tendremos que,

$$\theta_t|y_{1:t} \sim N_p(m_t, C_t).$$

donde

$$\begin{aligned} m_t &= C_t(F_t^T V_t^{-1} y_t + R_t^{-1} a_t). \\ C_t &= (F_t^T V_t^{-1} F_t + R_t^{-1})^{-1}. \end{aligned}$$

Ahora bien, para llegar a la expresión de C_t deseada, consideremos $Q_t = F_t R_t F_t^T + V_t$ (Proposición 2.2.2 ecuación (2.16)), entonces debemos comprobar que:

$$\begin{aligned} (F_t^T V_t^{-1} F_t + R_t^{-1})^{-1} &= R_t - R_t F_t^T (F_t R_t F_t^T + V_t)^{-1} F_t R_t \\ &= R_t - R_t F_t^T Q_t^{-1} F_t R_t. \end{aligned}$$

Es decir, verificar que $R_t - R_t F_t^T Q_t^{-1} F_t R_t$ es la matriz inversa de $(F_t^T V_t^{-1} F_t + R_t^{-1})^{-1}$.

$$\begin{aligned} (F_t^T V_t^{-1} F_t + R_t^{-1})(R_t - R_t F_t^T (F_t R_t F_t^T + V_t)^{-1} F_t R_t) &= (F_t^T V_t^{-1} F_t R_t) + I \\ &\quad - \left(F_t^T V_t^{-1} F_t^T V_t^{-1} F_t R_t F_t^T + F_t^T\right) \left(F_t R_t F_t^T + V_t\right)^{-1} F_t R_t \\ &= (F_t^T V_t^{-1} F_t R_t) + I \\ &\quad - \left(F_t^T V_t^{-1})(F_t R_t F_t^T + V_t\right) \left(F_t R_t F_t^T + V_t\right)^{-1} F_t R_t \\ &= (F_t^T V_t^{-1} F_t R_t) + I \\ &\quad - \left(F_t^T V_t^{-1} F_t R_t\right) \\ &= I. \end{aligned}$$

Ahora considerando la expresión para C_t, vamos a reescribir el vector de medias m_t como sigue:

$$\begin{aligned}
m_t &= \left(R_t - R_t F_t^T Q_t^{-1} F_t R_t\right)\left(F_t^T V_t^{-1} y_t + R_t^{-1} a_t\right) \\
&= R_t F_t^T V_t^{-1} y_t + a_t - R_t F_t^T Q_t^{-1} F_t R_t V_t^{-1} y_t - F_t^T Q_t^{-1} F_t R_t a_t \\
&= a_t - F_t^T Q_t^{-1} F_t R_t a_t + R_t F_t^T \left(I + Q_t^{-1} F_t R_t F_t^T\right) V_t^{-1} y_t \\
&= a_t - F_t^T Q_t^{-1} F_t R_t a_t + R_t F_t^T Q_t^{-1}\left(Q_t + F_t R_t F_t^T\right) V_t^{-1} y_t \\
&= a_t - F_t^T Q_t^{-1} F_t R_t a_t + R_t F_t^T Q_t^{-1} V_t V_t^{-1} y_t \\
&= a_t - F_t^T Q_t^{-1} F_t R_t a_t + R_t F_t^T Q_t^{-1} y_t \\
&= a_t -_t R_t F_t^T Q_t^{-1}\left(y_t - F_t a_t\right)
\end{aligned}$$

Por lo tanto concluimos que,

$$\begin{aligned}
m_t &= E\left(\theta_t | y_{1:t}\right) = a_t + R_t F_t^T Q_t^{-1} e_t. \\
C_t &= \mathrm{Var}\left(\theta_t | y_{1:t}\right) = R_t - R_t F_t^T Q_t^{-1} F_t R_t.
\end{aligned}$$

□

Cabe agregar que las distribuciones predictiva y de filtración se pueden calcular de forma recursiva iniciando con $\theta_0 \sim N(m_0, C_0)$ y calculando $p(\theta_t | y_t)$ tal que $t \geq 1$, mientras se tenga nueva información disponible.

2.2.2. Filtración con Observaciones Faltantes

En el análisis de series de tiempo es común enfrentarse con el problema de datos faltantes. Es decir, la ausencia de datos para algún tiempo t del vector de observaciones. A manera de ejemplo, consideremos la base de datos de las concentraciones de ozono de la estación el Pedregal en la Ciudad de México, con la que estaremos trabajando en secciones posteriores y que analizaremos con mayor detalle en el último capítulo de este trabajo. Es un ejemplo claro de este problema, que se deriva del proceso de medición de los datos, pues existen períodos de interrupción para la calibración de los intrumentos de medición. Considerando las observaciones faltantes de la base, bajo el contexto de los modelos de espacio de estados, este problema puede ser resuelto calculando las filtraciones de recursión de manera adecuada,

$$p\left(\theta_t | y_{1:t}\right) = p\left(\theta_t | y_{1:t-1}\right).$$

Es decir que, la distribución de filtración al tiempo t, será la distribución predictiva un paso hacia adelante al tiempo $t-1$. En particular para un modelo dinámico lineal, tal que $\theta_t | y_{1:t-1} \sim N(a_t, R_t)$, lo que se necesita hacer es $m_t = a_t$ y $C_t = R_t$, para lo cual, basta con establecer $F_t = 0$ o $V_t = \infty$. En el primer caso y_t no se relaciona con θ_t de ninguna forma, en el segundo la observación es tan ruidosa para proveer información significativa sobre θ_t. Resultando así en una matriz de ganancia $K_t = 0$ y por consecuencia $m_t = a_t$ y $C_t = R_t$.

2.2.3. Suavizador de Kalman

Anteriormente se vio el algoritmo de suavizamiento para un modelo de espacio de estados. En este apartado revisaremos el suavizador de Kalman, otro algoritmo de estimación robusto y bastante útil en el caso de los MDLs.

Proposición 2.2.3. ***(Suavizador de Kalman).*** *Considere un modelo dinámico lineal como el de la Definición 2.2.1, y sea* $\theta_{t+1}|y_{1:T} \sim N_p(s_{t+1}, S_{t+1})$. *Entonces se tendrá que*

$$\theta_t|y_{1:T} \sim N_p(s_t, S_t),$$

donde

$$s_t = m_t + C_t G_{t+1}^T R_{t+1}^{-1}(s_{t+1} - a_{t+1}) \tag{2.21}$$
$$S_t = C_t - C_t G_{t+1}^T R_{t+1}^{-1}(R_{t+1} - S_{t+1}) R_{t+1}^{-1} G_{t+1} C_t. \tag{2.22}$$

Demostración. Por hipótesis tenemos que $\theta_{t+1}|y_{1:T} \sim N(s_{t+1}, S_{t+1})$. Además las propiedades de la distribución normal multivariada se puede probar que el vector aleatorio $(\theta_0, ..., \theta_t, Y_1, ..., Y_t)$ sigue una distribución normal para toda $t \geq 1$. Y que las distribuciones condicionales y marginales también resultan seguir tal distribución. En particular se cumplirá que $\theta_t|y_{1:T}$ también sigue una distribución normal, por lo tanto basta con verificar su media y su varianza.

$$\begin{aligned} s_t &= E(\theta_t|y_{1:T}) \\ &= E(E(\theta_t|\theta_{t+1}, y_{1:T})|y_{1:T}). \\ S_t &= \text{Var}(\theta_t|y_{1:T}) \\ &= \text{Var}(E(\theta_t|\theta_{t+1}, y_{1:T})|y_{1:T}) + E(\text{Var}(\theta_t|\theta_{t+1}, y_{1:T})|y_{1:T}). \end{aligned}$$

Como vimos en la prueba de la Proposición 2.1.2 inciso i), θ es independiente de $Y_{t+1:T}$ dado θ_{t+1}; entonces

$$p(\theta_t|\theta_{t+1}, y_{1:T}) = p(\theta_t|\theta_{t+1}, y_{1:t}).$$

Esta distribución puede calcularse utilizando la *Regla de Bayes* y observando que la función de verosimilitud $p(\theta_{t+1}|\theta_t, y_{1:t}) = p(\theta_{t+1}|\theta_t)$, expresada por la ecuación de estados (2.12), sigue la siguiente distribución:

$$\theta_{t+1}|\theta_t \sim N(G_{t+1}\theta_t, W_{t+1}).$$

Y si consideremos a $p(\theta_t|y_{1:t})$ como la *distribución a priori* y la Proposición 2.2.2 inciso iii) (distribución de filtración) que nos dice que dicha densidad sigue una distribución normal con media m_t y varianza C_t.

Entonces por la Proposición 2.2.1 (página 20) tendremos que:

$$\begin{aligned} E(\theta_t|\theta_{t+1}, y_{1:t}) &= (G_{t+1}^T W_{t+1}^{-1} G_t + C_t^{-1})(G_{t+1}^T W_{t+1}^{-1}\theta_{t+1} + C_t^{-1} m_t) \\ &= m_t + C_t G_{t+1}^T \left(G_{t+1} C_t G_{t+1}^T + W_{t+1}\right)^{-1} (\theta_{t+1} - G_{t+1} m_t) \\ &= m_t + C_t G_{t+1}^T R_{t+1}^{-1} (\theta_{t+1} - a_{t+1}). \end{aligned}$$

y,

$$\begin{aligned} \text{Var}(\theta_t|\theta_{t+1}, y_{1:t}) &= (G_{t+1}^T W_{t+1}^{-1} G_{t+1} + C_t^{-1})^{-1} \\ &= C_t - C_t G_{t+1}^T (G_{t+1} C_t G_{t+1}^T + W_{t+1})^{-1} G_{t+1} C_t \\ &= C_t - C_t G_{t+1}^T R_{t+1}^{-1} G_{t+1} C_t. \end{aligned}$$

Utilizando ambas expresiones, obtenemos lo siguiente:

$$\begin{aligned} s_t &= E(\theta_t|y_{1:T}) \\ &= E(E(\theta_t|\theta_{t+1}, y_{1:T})|y_{1:T}) \\ &= E\left(m_t + C_t G_{t+1}^T R_{t+1}^{-1} (\theta_{t+1} - a_{t+1})|y_{1:T}\right) \\ &= m_t + C_t G_{T+1}^T R_{t+1}^{-1} (s_{t+1} - a_{t+1}). \\ S_t &= \text{Var}(\theta_t|y_{1:T}) \\ &= \text{Var}(E(\theta_t|\theta_{t+1}, y_{1:T})|y_{1:T}) + E(\text{Var}(\theta_t|\theta_{t+1}, y_{1:T})|y_{1:T}) \\ &= \text{Var}\left(m_t + C_t G_{t+1}^T R_{t+1}^{-1} (\theta_{t+1} - a_{t+1})|y_{1:T}\right) + E\left(C_t - C_t G_{t+1}^T R_{t+1}^{-1} G_{t+1} C_t|y_{1:T}\right) \\ &= C_t - C_t G_{t+1}^T R_{t+1}^{-1} G_{t+1} C_t + C_t G_{t+1}^T R_{t+1}^{-1} S_{t+1} R_{t+1}^{-1} G_{t+1} C_t \\ &= C_t - C_t G_{t+1}^T R_{t+1}^{-1} (R_{t+1} - S_{t+1}) R_{t+1}^{-1} G_{t+1} C_t. \end{aligned}$$

Por lo tanto,

$$\theta_t|y_{1:T} \sim N_p(s_t, S_t).$$

□

2.2.4. Predicción de un Modelo Dinámico Lineal

En la proposición 2.1.3. vimos diversas expresiones generales sobre el algoritmo recursivo de predicción, en el caso de los modelos dinámicos lineales todas éstas pueden calcularse de forma explícita. Sin embargo, como fue para los procesos de filtración y suavizamiento, dado que todas las distribuciones de predicción siguen una distribución normal, es suficiente con calcular sus medias y varianzas.

Antes de enunciar nuestra siguiente proposición, consideremos la siguiente notación. Sea $k \geq 1$, entonces definimos las siguientes igualdades:

$$a_t(k) = E[\theta_{t+k}|y_{1:t}] \tag{2.23}$$
$$R_t(k) = E[\theta_{t+k}|y_{1:t}] \tag{2.24}$$
$$f_t(k) = E[Y_{t+k}|y_{1:t}] \tag{2.25}$$
$$Q_t(k) = E[Y_{t+k}|y_{1:t}] \tag{2.26}$$

Proposición 2.2.4. *Consideremos un modelo dinámico lineal definido como en 2.2.1, y sean* $a_t(0) = m_t$ *y* $R_t(0) = C_t$*, entonces para* $k \geq 1$*, se cumplirá lo siguiente:*

i) La distribución de θ_{t+k} *dado* $y_{1:t}$ *es normal, con*

$$a_t(k) = G_{t+k}a_{t,k-1}, \tag{2.27}$$
$$R_t(k) = G_{t+k}R_{t,k-1}G_{t+k}^T + W_{t+k}; \tag{2.28}$$

ii) La distribución de Y_{t+k} *dado* $y_{1:t}$ *es normal, con*

$$f_t(k) = F_{t+k}a_t(k), \tag{2.29}$$
$$Q_t(k) = F_{t+k}R_t(k)F_{t+k}^T + V_t. \tag{2.30}$$

Demostración. Las igualdades en i) y ii) son ciertas para $k = 1$, por la demostración de la Proposición 2.2.2 (filtro de Kalman). Por otro lado, para $k > 1$ tendremos que

i)

$$\begin{aligned} a_t(k) &= E[\theta_{t+k}|y_{1:t}] \\ &= E\left[E\left[\theta_{t+k}|y_{1:t}, \theta_{t+(k-1)}\right]|y_{1:t}\right] \\ &= E\left[G_{t+k}\theta_{t+(k-1)}|y_{1:t}\right] \\ &= G_{t+k}E\left[\theta_{t+(k-1)}|y_{1:t}\right] \\ &= G_{t+k}a_t(k-1). \end{aligned}$$

$$\begin{aligned} R_t(k) &= \mathrm{Var}[\theta_{t+k}|y_{1:t}] \\ &= \mathrm{Var}\left[E\left[\theta_{t+k}|y_{1:t}, \theta_{t+(k-1)}\right]|y_{1:t}\right] + E\left[\mathrm{Var}\left[\theta_{t+k}|y_{1:t}, \theta_{t+(k-1)}\right]|y_{1:t}\right] \\ &= \mathrm{Var}\left[G_{t+k}\theta_{t+(k-1)}|y_{1:t}\right] + E\left[W_{t+k}|y_{1:t}\right] \\ &= G_{t+k}R_{t,k-1}G_{t+k}^T + W_{t+k}. \end{aligned}$$

ii)

$$\begin{aligned} f_t(k) &= E[Y_{t+k}|y_{1:t}] \\ &= E[E[Y_{t+k}|y_{1:t},\theta_{t+k}]|y_{1:t}] \\ &= E[F_{t+k}\theta_{t+k}|y_{1:t}] \\ &= F_{t+k}E[\theta_{t+k}|y_{1:t}] \\ &= F_{t+k}a_t(k). \end{aligned}$$

$$\begin{aligned} Q_t(k) &= \mathrm{Var}[Y_{t+k}|y_{1:t}] \\ &= \mathrm{Var}[E[Y_{t+k}|y_{1:t},\theta_{t+k}]|y_{1:t}] + E[\mathrm{Var}[Y_{t+k}|y_{1:t},\theta_{t+k}]|y_{1:t}] \\ &= \mathrm{Var}[F_{t+k}\theta_{t+k}|y_{1:t}] + E[V_{t+k}|y_{1:t}] \\ &= F_{t+k}R_t(k)F_{t+k}^T + V_t. \end{aligned}$$

□

2.2.5. Proceso de Innovaciones

Como es usual en la construcción de un modelo estadístico, la verificación de supuestos es un factor clave para determinar si éste es adecuado o no para explicar el problema en cuestión. En el caso de los modelos dinámicos lineales serán de gran relevancia los errores de predicción conocidos comúnmente como innovaciones, pues es a través de éstos con los que se valida el ajuste del modelo.

En secciones anteriores de este trabajo vimos cómo hacer predicciones del vector de observaciones un paso hacia adelante, $f_t = E[Y_t|y_{1:t-1}] = F_t E(\theta_t|y_{1:t-1})$, y definimos el error de predicción como:

$$e_t = Y_t - E[Y_t|y_{1:t-1}] = Y_t - f_t. \tag{2.31}$$

Esta definición tiene sentido si pensamos en Y_t como la suma de un componente que es predecible, f_t, dado un conjunto de observaciones pasadas, más un componente, e_t, independiente del pasado que contiene nueva información sobre la serie de datos Y_t.

Algunas de las propiedades que satisface el proceso de innovaciones están dadas en la siguiente proposición:

Proposición 2.2.5. ***(Propiedades del proceso de innovación)****. Sea $(e_t)_{t\geq 1}$, la serie de errores de predicción de un modelo dinámico lineal. Entonces se cumplen las siguientes propiedades:*

i) El valor esperado del vector aleatorio e_t es cero.

ii) El vector aleatorio e_t no está correlacionado con cualquier función de $Y_1,\ldots,Y_{t-1}$.

iii) e_t *y* Y_s *no están correlacionados para cualquier s<t.*

iv) e_t *y* e_s *no están correlacionados para cualquier s<t.*

v) e_t *es una función lineal de* $Y_1, ..., Y_t$.

vi) $(e_t)_{t\geq 1}$ *es un proceso Gaussiano.*

Demostración. i) Tomando esperanza iterada tenemos que:

$$\begin{aligned} E(e_t) &= E[E[e_t|y_{1:t-1}]] \\ &= E[E[Y_t - f_t|y_{1:t-1}]] \\ &= E[E[Y_t - E[Y_t|y_{1:t-1}]|y_{1:t-1}]] \\ &= E[E[Y_t|y_{1:t-1}] - E[Y_t|y_{1:t-1}]] \\ &= 0. \end{aligned}$$

ii) Sea $Z = g(Y_1, ..., Y_{t-1})$,

$$\begin{aligned} \text{Cov}(e_t, Z) &= E[e_t Z] - E[e_t]E[Z] \\ &= E[e_t Z] \\ &= E[E[e_t Z|y_{1:t-1}]] \\ &= E[ZE[Y_t - E[Y_t|y_{1:t-1}]|y_{1:t-1}]] \\ &= E[Z(E[Y_t|y_{1:t-1}] - E[Y_t|y_{1:t-1}])] \\ &= 0. \end{aligned}$$

iii) Notemos que en el caso de observaciones univariadas podemos tomar $Z = Y_s$ y por el inciso ii) de la Proposición 2.2.4 demostrada previamente se sigue que $\text{Cov}(e_t, Y_s) = 0$ con $s < t$. Por otro lado en el caso observaciones multivariadas, se debe aplicar dicha proposición a cada uno de los componentes.

iv) De manera similar a la propiedad iii) notemos que en el caso en el que las observaciones son univariadas podemos tomar $Z = e_s$ y por la proposición ii) demostrada anteriormente se sigue que $\text{Cov}(e_t, e_s) = 0$ con $s < t$. Por otro lado, en el caso en el que las observaciones son multivariadas se debe aplicar la proposición ii) a cada uno de los componentes.

v) A lo largo de este trabajo hemos visto que el vector $(Y_1, ..., Y_t)$ sigue una distribución conjunta normal; además $f_t = E[Y_t|y_{1:t-1}]$ es una función lineal de $(Y_1, ..., Y_{t-1})$ y como $e_t = Y_t - f_t$ se sigue que es una función lineal de $(Y_1, ..., Y_t)$, es decir que:

$$\begin{bmatrix} e_1 \\ \vdots \\ e_t \end{bmatrix} = A \begin{bmatrix} Y_1 \\ \vdots \\ Y_t \end{bmatrix} + c.$$

con A una matriz constante y c un vector constante.

vi) De la proposición iv) tenemos que el vector $(e_1,...,e_t)$ es una transformación lineal de $(Y_1,...,Y_t)$, y sabemos que $(Y_1,...,Y_t)$ sigue una distribución conjunta normal. Por lo tanto, dado que todas las distribuciones finitas dimensionales son normales, el proceso $(e_t)_{t\geq 1}$ es Gaussiano.

□

2.2.6. Validación de un Modelo Dinámico Lineal

En el caso de un modelo dinámico lineal con observaciones univariadas, la sucesión de innovaciones estandarizada viene dada por la expresión: $\tilde{e}_t = \frac{e_t}{\sqrt{Q_t}}$, que es un ruido blanco Gaussiano [2], y suele ser utilizada para verificar el ajuste del modelo. Si éste es adecuado, la sucesión de innovaciones, $(\tilde{e}_t)_{t\geq 1}$, calculada a partir de los datos debe asemejarse a una muestra de variables aleatorias normales estándar de tamaño t. Para lo cual debe verificarse que:

i) La sucesión de variables aleatorias $(\widetilde{e}_t)_{t\geq 1}$ sigue una distribución normal estándar.

ii) Las $\widetilde{e}_t's$ no están correlacionadas para toda $s < t$.

En este trabajo emplearemos la paquetería MARSS() que se encuentra dentro del software estadístico R, la cual nos provee de la función MARSSkfss(), útil para la obtención de las innovaciones de los errores de predicción de un modelo dinámico lineal.

Asimismo realizaremos pruebas empíricas para la validación del proceso de innovaciones, siendo algunas de ellas: la gráfica Q-Q plot e histograma para verificar el supuesto de normalidad. Usaremos las funciones: qqnorm( ), para graficar los cuantiles de las innovaciones en el eje y contra los cuantiles teóricos de una distribución normal en el eje x, y la función hist( ), como una representación gráfica del proceso de innovaciones, para obtener así una vista general de la distribución de nuestros datos. Adicional a éstas emplearemos la prueba no paramétrica Anderson Darling implementada en el paquete nortest( ) con la función ad.test( ), y la prueba t de Student, implementada en la función t.test( ), para determinar si nuestra muestra de datos posee media cero o no, es decir, sí $E[\tilde{e}_t] = 0$. Para verificar nuestro segundo supuesto, usaremos la función de autocorrelación de los residuales, acf( ), para visualizar si las observaciones del proceso covarían con ellas mismas en el tiempo, detectar outliers, puntos de cambio, entre otros. Diremos que se cumple el supuesto en ii) si ninguna de las observaciones excede un intervalo de 95 % de confianza [3]. Finalmente utilizamos la prueba de Ljung-Box, Box.test( ), como complemento a ésta última prueba estadística para determinar si nuestro proceso se distribuye de forma independiente; es decir, si las correlaciones de una muestra de nuestro proceso de innovaciones es cero.

Cabe agregar que para observaciones multivariadas usualmente se aplican los mismos diagnósticos gráficos univariados por componentes a las secuencias de innovaciones. Sin embargo,

[2] Ruido blanco Gaussiano: sucesión de variables aleatorias independientes e idénticamente distribuidas normales con media cero y varianza constante.

[3] Bajo el enfoque clásico, el intervalo de predicción al $100(1-alpha)\,\%$, donde el valor $1-\alpha$ es conocido como el nivel de confianza, para una variable aleatoria Y, es un intervalo aleatorio de la forma $[L(y)-U(y)]$, que cumple la siguiente propiedad: $P_{\psi}(L(y) \leq Y \leq U(y)) \geq 1-\alpha$ para todos los valores posibles del parámetro ψ.

otro enfoque no tan popular consiste en definir la estandarización de vectores $\widetilde{e}_t = B_t e_t$, donde B_t es una matriz de dimensiones $p \times p$, tal que $B_t Q_t B_t^T$ es igual a la matriz identidad. Así se tendrá que los componentes de $\widetilde{e}_t$ serán independientes e idénticamente distribuidos de acuerdo a una distribución normal estándar. No obstante se empleará el primer enfoque para la validación del modelo.

2.3. Estimación de Parámetros

Antes de concluir este capítulo daremos una introducción sobre la estimación de parámetros. Si bien en algunos de los ejemplos que se presentarán a lo largo de este escrito se supone que las matrices F_t, G_t, V_t y W_t son conocidas, en la práctica no se suele trabajar con modelos completamente especificados. Es por esto que se muestran dos enfoques usuales para la estimación de la varianza observacional y matriz de varianzas del sistema de un modelo dinámico lineal.

2.3.1. Estimación por Máxima Verosimilitud

El método de estimación por máxima verosimilitud, EMV, es por mucho la técnica más popular, utilizada para obtener estimadores puntuales de los parámetros de un modelo. Suponga que tiene n vectores aleatorios, $Y_1, ..., Y_n$, cuya distribución depende de los parámetros desconocidos, digamos ψ. Sea $p(y_1, ..., y_n; \psi)$ la densidad conjunta de las observaciones para un valor dado del parámetro. Entonces, podemos definir a la función de densidad de probabilidad de los datos observados como función de ψ, conocida como la función de verosimilitud, L, como:

$$L(\psi) = k \cdot p(y_1, ..., y_n; \psi) = k \cdot \prod_{t=1}^{n} p(y_t | y_{1:t-1}; \theta) \tag{2.32}$$

donde k es una constante y $p(y_t | y_{1:t-1}; \psi)$, para toda $t = 1, 2, ..., n$, sigue una distribución normal con media f_t y varianza Q_t (por el inciso ii) de la Proposición 2.2.2).

La función de log-verosimilitud puede ser escrita como:

$$\begin{aligned}
\ell(\psi) &= \log(L(\psi)) \\
&= \log\left(\prod_{t=1}^{n} p(y_t | y_{1:t-1}; \psi)\right) \\
&= \log\left(\prod_{i=1}^{n} \frac{1}{(2\pi)^{p/2} |Q_t|^{1/2}} \exp\left\{-\frac{1}{2}(y_t - f_t)^T Q_t^{-1} (y_t - f_t)\right\}\right)
\end{aligned}$$

por lo tanto,

$$\ell(\psi) = -\frac{1}{2}\sum_{t=1}^{n} \log|Q_t| - \frac{1}{2}\sum_{t=1}^{n} (y_t - f_t)^T Q_t^{-1} (y_t - f_t). \tag{2.33}$$

Maximizando numéricamente la expresión (2.33), podemos obtener el estimador máximo verosímil de ψ,

$$\widehat{\psi} = \underset{\psi}{\operatorname{argmax}} \left(\ell \left(\psi \right) \right).$$

Usualmente cuando se busca el estimador máximo verosímil, solemos pensar en minimizar la función de verosimilitud negativa[4]. En el contexto de los modelos dinámicos lineales podemos pensar esto en dos pasos, en primer lugar construir el MDL y posteriormente evaluar su función de verosimilitud negativa como función de las matrices definidas en éste.

$$\psi \overset{\text{construcción}}{\Longrightarrow} MDL \overset{\text{log-verosimilitud}}{\Longrightarrow} -\ell \left(\psi \right).$$

Sea H la matriz Hessiana de $-\ell(\psi)$ evaluada en $\psi = \hat{\psi}$. Entonces podemos obtener una aproximación de la varianza del estimador máximo verosímil, $\operatorname{Var}(\widehat{\psi})$, a partir de H^{-1}.

En los modelos dinámicos lienales más utilizados, se suelen tener las propiedades de consistencia y normalidad asintótica. Sin embargo hay dos problemas a considerar relacionados con la optimización numérica. El primero de ellos es que la función de verosimilitud de un MDL puede presentar más de un máximo local al inicializarse con diferentes valores a priori, por lo que es recomendable probar con diversos puntos iniciales y comparar los máximos en cada caso. El segundo suele ser que el estimador de la varianza del estimador máximo verosímil sea muy grande, lo cual puede ser señal de un modelo pobremente identificable. Una posible solución es eliminar algunos de lo parámetros, especialmente cuando uno está interesando en hacer inferencia o bien busca interpretar los parámetros del modelo.

2.3.2. Métodos de Monte Carlo Vía Cadenas de Markov

Como vimos en el apartado anterior, el enfoque de máxima verosimilitud tiene por objeto encontrar el valor del parámetro que maximiza la función de verosimilitud para los datos observados, y producir estimadores puntuales de los errores estándar con base en el teorema del Límite Central. En contraste el enfoque Bayesiano ofrece una formulación del problema más consistente, en el que los parámetros desconocidos, ψ, son considerados como un vector aleatorio. Además, en este enfoque la distribución posterior sobre un parámetro busca resumir la información completa.

Así, trataremos el caso más general en el que los cálculos se tornan analíticamente intratables. Se introducirá el método de Monte Carlo vía cadenas de Markov (MCMC, por sus siglas en inglés) por su eficiencia en la aproximación de la distribución posterior de interés. En particular utilizaremos el muestreo de Gibbs para analizar la distribución posterior de un vector de parámetros desconocidos y el vector de estados no observables de un MDL, dadas nuestras observaciones:

$$p \left(\psi, \theta_{0:T} | y_{1:T} \right).$$

[4]Recuerde que $min(f(x)) = max(-f(x))$.

Introducción a los Métodos de Muestreo MCMC

Los métodos de Monte Carlo vía cadenas de Markov, son métodos de simulación para generar muestras de distribuciones posteriores y estimar cantidades de interés. Comúnmente son utilizados para realizar inferencia Bayesiana, pues en la mayoría de los casos $p(\psi|y)$ es desconocida y no es computacionalmente eficiente obtener una muestra del vector de parámetros directamente de ésta distribución. El enfoque básico de estos métodos de muestreo se basa en simular secuencialmente valores de ψ de distribuciones aproximadas hasta obtener cadenas que se asemejen más a la distribución posterior.

De hecho el nombre de "Monte Carlo vía Cadenas de Markov" implica este proceso. Por un lado "Monte Carlo" hace referencia al proceso de simulación aletorio. Por otro lado, "Cadenas de Markov" se refiere al proceso de muestreo de un valor nuevo de la distribución posterior, dado el valor previo. Así este proceso iterativo produce una cadena de valores que constituyen una muestra de la distribución posterior.

En el caso más general de un modelo dinámico lineal con un vector de parámetros desconocidos, ψ, y distribución a priori, $p(\psi)$, uno de los métodos de muestreo MCMC para el análisis de la distribución posterior, es el muestreo de Gibbs, el cual es un caso especial de un algoritmo más general conocido como "Metropolis Hastings" (Gelman et al., 2014). Dicho algoritmo consiste en obtener muestras de la distribución de estados, dados el vector de parámetros desconocidos y las observaciones, y de la distribución del vector de parámetros desconocidos, dados los estados y las observaciones.

A grandes rasgos, el proceso que sigue el muestreo de Gibss puede describirse de la siguiente manera. Supongamos que el vector de parámetros desconocidos ψ se divide en d componentes o subvectores, $\psi = (\psi_1, \psi_2, ..., \psi_d)$. Entonces, en cada iteración, t, del muestreo de Gibbs se recorren de manera ciclíca los subvectores y se elige un ordenamiento de ψ, sucesivamente, cada subconjunto ψ_j^t se muestrea de la distribución condicional dado el valor de todos los demás componentes de ψ, es decir,

$$p\left(\psi_j|\psi_{-j}^{t-1}, y\right)$$

donde $\psi_{-j}^{t-1} = \left(\psi_1^t, ..., \psi_{j-1}^t, \psi_{j+1}^{t-1}, ..., \psi_d^{t-1}\right)$ representa todos los componentes de ψ, excepto el j-ésimo, ψ_j.

Por lo tanto, cada subvector del vector de parámetros desconocidos, se actualiza dados los últimos valores del resto de los componentes de ψ.

Antes de enunciar de manera formal este proceso aplicable a los modelos dinámicos lineales bajo estudio, introduciremos el algoritmo *Forward Filtering Backward Sampling* (FFBS), el cual se originó para dar respuesta a lo siguiente; sea $p(\theta_{0:T}|\psi, y_{1:T})$ y suponga que tiene el rol de condicional completa en el muestreo de Gibbs, entonces a partir de $p(\theta_{0:T}, \psi|y_{1:T})$, ¿cómo

podemos generar una muestra de la distribución de $\theta_{0:T}$ dado $(y_{1:T}, \psi)$?.

Considere la siguiente expresión para la distribución conjunta de $\theta_{0:T}$ dada toda la información hasta el tiempo T, $y_{1:T}$:

$$p(\theta_{0:T}|y_{1:T}) = \prod_{t=0}^{T} p(\theta_t|\theta_{t+1:T}, y_{1:T}), \tag{2.34}$$

donde el último factor en el proceso es simplemente $p(\theta_T|y_{1:T})$; es decir, la distribución de filtración de θ_T que sigue una distribución $N(m_T, C_T)$. Así, a partir de la ecuación (2.34) podemos empezar a obtener la muestra de la distribución del lado izquierdo de manera recursiva, iniciando con θ_T de una $N(m_T, C_T)$ y después para $t = T-1, T-2, ..., 0$ hasta tener una muestra de θ_t de $p(\theta_t|\theta_{t+1:T}, y_{1:T})$. Pues por la Proposición 2.1.2 inciso i) sobre el algoritmo de suavizamiento, se tiene que θ_t y $Y_{t+1:T}$ son condicionalmente independientes dado θ_{t+1}, entonces $p(\theta_t|\theta_{t+1:T}, y_{1:T}) = p(\theta_t|\theta_{t+1}, y_{1:t})$ sigue una distribución $N(h_t, H_T)$ con los siguientes parámetros:

$$\begin{aligned} h_t &= m_t + C_t G_{t+1}^T R_{t+1}^{-1} (\theta_{t+1} - a_{t+1}) \\ H_t &= C_t - C_t G_{t+1}^T R_{t+1}^{-1} G_{t+1} C_t \end{aligned}$$

Por lo tanto, una vez que conseguimos obtener la distribución conjunta de $(\theta_{t+1}, ..., \theta_T)$, solo nos queda obtener la distribución de θ_t de una $N(h_t, H_t)$. Note que la media de la distribución normal h_t depende explícitamente del valor de θ_{t+1} generado previamente.

Así el algortimo FFBS puede resumirse en los siguientes pasos:

1. Correr el filtro de Kalman.
2. Simular $\theta_T \sim N(m_T, C_T)$.
3. Para $t = T-1, T-2, ..., 0$, simular $\theta_t \sim N(h_t, H_t)$.

Como podremos notar a continuación, este algorimo es un pilar en la construcción del muestreo de Gibbs, el cual puede resumirse de la siguiente manera:

0. Asignar un vector de valores iniciales: $\psi = \psi^{(0)}$.
1. Para $i = 1, ..., N$:
   a) Obtener una muestra de $\theta_{0:T}^{(i)}$, a partir de $p\left(\theta_{0:T}|y_{1:T}, \psi = \psi^{(i-1)}\right)$ utilizando el algoritmo FFBS.
   b) Obtener una muestra de $\psi^{(i)}$, a partir de $p\left(\psi|y_{1:T}, \theta_{0:T} = \theta_{0:T}^{(i)}\right)$.

Cabe agregar que cuando ψ es un vector r-dimensional, suele ser más simple realizar muestreo de Gibbs para cada componente, en vez de obtener una muestra del vector mismo.

Una vez que se obtienen los resultados de los métodos de muestreo MCMC, deben verificarse para asegurar que los valores simulados han convergido aproximadamente a la distribución estacionaria $p(\psi|y)$. Para esto se suelen utilizar pruebas empíricas para determinar si los valores simulados provienen de una muestra de valores independientes e idénticamente distribuidos. En lo que a ello se refiere existen extensos procedimientos en la literatura para estudiar la convergencia de una cadena. Dentro del software estadístico R, nos referimos a la paquetería *coda*, la cual provee un conjunto de funciones útiles para implementar muchos de estos diagnósticos.

Capítulo 3

Modelos Dinámicos Lineales en el Análisis de Series de Tiempo

En este capítulo introduciremos uno de los enfoques más populares para especificar completamente un modelo dinámico lineal. Partiremos de pensar en una serie de tiempo como una combinación de componentes elementales simples, pues por la estructura aditiva de estos modelos se puede asumir que cada componente captura alguna de las siguientes características: tendencia, estacionalidad y dependencia sobre covariables, sujetos a un error observacional. Esto nos dará la pauta para introducir las estructuras de modelos básicos para el análisis de series de tiempo bajo el enfoque adoptado en este trabajo. Comenzaremos por los modelos con tendencia, posteriormente, daremos unda breve introducción sobre los modelos con factores estacionales, y por último pasaremos a los modelos de regresión dinámicos.

3.1. Técnica de Descomposición Aditiva para Modelos con Observaciones Univariadas

El enfoque general para el análisis de series de tiempo con observaciones univariadas consiste en representar cada uno de sus componentes (tendencia, ciclos estacionales y dependencia sobre covariables) como un modelo dinámico lineal individual, y posteriormente combinarlos en uno solo que explique la serie de tiempo dada. De manera formal, sea Y_t una serie de tiempo univariada; entonces podemos asumir que ésta puede ser escrita como la suma de componentes mutuamente independientes:

$$Y_t = Y_{1,t} + \ldots + Y_{h,t},$$

donde el i-ésimo elemento, $Y_{i,t}$ con $i = 1, \ldots, h$, describirá cada uno de los componentes de la serie y estará dado por la siguiente representación:

$$\begin{aligned} Y_{i,t} &= F_{i,t}\theta_{i,t} + v_{i,t} \qquad v_{i,t} \sim N(0, V_{i,t}) \\ \theta_{i,t} &= G_{i,t}\theta_{i,t-1} + w_{i,t} \qquad w_{i,t} \sim N(0, W_{i,t}) \end{aligned}$$

donde los vectores de estados, $\theta_{i,t}$, p_i-dimensionales, son distintos y las series $(Y_{i,t}, \theta_{i,t})$ y $\left(Y_{j,t}, \theta_{j,t}\right)$ son independientes para toda $i \neq j$. Así, cada uno de los modelos dinámicos lineales es combinado hasta obtener aquel que define la serie $Y_t = \sum_{i=1}^{h} Y_{i,t}$. De hecho, por el supuesto de independencia de los componentes, es sencillo mostrar que la serie Y_t es descrita por el siguiente modelo:

$$\begin{aligned} Y_t &= F_t\theta_t + \nu_t, \quad \nu_t \sim N_m(0, V_t), \\ \theta_t &= G_t\theta_{t-1} + w_t, \quad w_t \sim N_p(0, W_t), \end{aligned}$$

donde,

$$\theta_t = \begin{pmatrix} \theta_{1,t} \\ \vdots \\ \theta_{h,t} \end{pmatrix}, \quad F_t = \left(F_{1,t}, F_{2,t}, \ldots, F_{h,t}\right),$$

G_t, W_t son un bloque de matrices diagonales,

$$G_t = \begin{bmatrix} G_{1,t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & G_{h,t} \end{bmatrix}, \quad W_t = \begin{bmatrix} W_{1,t} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & W_{h,t} \end{bmatrix},$$

y $V_t = \sum_{i=1}^{j} V_{i,t}$.

En las siguientes secciones trataremos los modelos más usuales para describir cada uno de los componentes de una serie de tiempo.

3.2. Modelos con Tendencia

Los modelos dinámicos lineales polinomiales son los más usados para describir la tendencia de una serie de tiempo, la cual al tiempo tiempo t puede pensarse como el comportamiento esperado de Y_{t+k} con $k \geq 1$, dada la información hasta el tiempo t. Así, definimos un modelo polinomial de orden n, como:

Definición 3.2.1. ***(Modelo polinomial de orden n).*** *Un modelo polinomial de orden n es un modelo dinámico lineal, descrito por las ecuaciones usuales* (2.9) *y* (2.10)*, con matrices $F_t = F$, $G_t = G$ constantes dadas por:*

$$F = (1, 0, ..., 0),$$

$$G = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & 1 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix},$$

y función predictiva:

$$f_t(k) = E(Y_{t+k}|y_{1:t}) = a_{t,0} + a_{t,1}k + ... + a_{t,n-1}k^{n-1},\ k \geq 0,$$

donde $a_{t,0} + a_{t,1} + ... + a_{t,n-1}$ *son funciones lineales de* $m_t = E(\theta_{t+k}|y_{1:t})$*, independientes de k.*

La función predictiva es un polinomio de orden $n-1$ en k, con n la dimensión del vector de estados θ y no el grado del polinomio.

Dentro de estos modelos, los más conocidos y empleados en problemas prácticos son: el modelo polinomial de primer orden, también conocido como modelo de nivel local o modelo de caminata aleatoria más ruido; y el modelo de crecimento lineal. Por ser los modelos polinomiales más simples, son de gran utilidad para describir características relevantes asociadas al marco de modelos bajo estudio.

3.2.1. Modelo Polinomial de Primer Orden

El modelo polinomial de primer orden, es comúnmente utilizado para pronósticos a corto plazo de series de tiempo que no muestren una tendencia clara ni variaciones estacionales. Se construye bajo el supuesto de que las observaciones Y_t se modelan como ruidos aleatorios de un nivel μ_t sujeto a cambios aleatorios. De ahí que se le conozca también como "modelo de caminata aleatoria más ruido". Dentro de sus aplicaciones se encuentran, por ejemplo, la modelación de la demanda de mercado para un producto, al caracterizarse por ser un proceso local en el tiempo; es decir que, la demanda de un producto "x" tiende a ser constante. A continuación describimos formalmente este modelo para un mejor entendimiento del mismo.

Definición 3.2.2. ***(Modelo Polinomial de primer orden)****. Sea* $(Y_t)_{t\geq 1}$ *una serie de tiempo con observaciones univariadas, definimos el modelo polinomial de primer orden por las siguientes ecuaciones:*

$$\begin{aligned} Y_t &= \mu_t + \nu_t, & \nu_t &\sim N(0,V). \\ \mu_t &= \mu_{t-1} + w_t, & w_t &\sim N(0,W). \end{aligned}$$

$$\mu_0 \sim N(m_0, C_0)$$

donde μ_t *es el nivel de la serie al tiempo t,* $(\nu_t)_{t\geq 1}$ *la serie de errores observables al tiempo t y* $(w_t)_{t\geq 1}$ *la serie de errores de evolución de la serie, siendo éstas últimas mutuamente independientes. Es decir, para toda s y t con* $t \neq s$*,* $\nu_s \perp \nu_t$*,* $w_s \perp \nu_t$ *y* $\nu_t \perp w_s$[1]*. Finalmente V y W son las varianzas de los errores observacional* v_t *y del sistema* w_t*.*

Cabe observar que como caso particular de un modelo polinomial de orden n (Definición 3.2.1), un modelo polinomial de primer orden se caracteriza porque $\theta_t = \mu_t$ es un son vector aleatorio de dimensión 1 (de ahí su nombre) y $F_t = G_t$ para toda t, son constantes iguales a 1. De ahí que la función predictiva sea un polinomio constante en el tiempo (un polinomio de grado 0) cuya media μ_t depende de la información pasada hasta el tiempo t:

[1] El símbolo $\perp$ significa "independiente de".

$$f_t(k) = E(Y_{t+k}|y_{1:t}) = E(\theta_{t+k}|y_{1:t}) = \mu_t.$$

Para este modelo en particular es interesante ver que el comportamiento del proceso se puede ver influenciado por el cociente de las varianzas de los errores, $r = W/V$, usualmente conocido como *signal to noise ratio*. El cual es un concepto comunmente utilizado en ingeniería y el cual hace referencia a la varianza del sistema con respecto a la varianza de las observaciones (West and Harrison, 1997).

Ejemplo: Un modelo con tendencia constante

Para ejemplificar un modelo con tendencia constante o modelo polinomial de primer orden, utilizaremos la serie de tiempo de las concentraciones de ozono en la estación el Pedregal, que será descrita en el capítulo 4. Propondremos dos modelos similares en los que el papel central de las estimaciones recae en el cociente de los valores de las varianzas de los errores, *signal to noise ratio*. Para ello utilizaremos la Definición 3.2.2 de un modelo polinomial de primer orden.

Sea $(Y_t)_{t\geq 1}$ la serie de las concentraciones diarias de ozono en partes por billón (ppb), con $t = 1,2,...,365$ y μ_t el nivel de la serie al tiempo t. Entonces, si especificamos dos listas en R: la primera con el valor del parámetro inicial del vector de estados $\mu_0 = 0$, la segunda con el resto de los parámetros que definen el modelo $G_t = F_t = 1$, $V_t = V$ y $W_t = W$, tal que, V y W son escalares. Utilizamos la función MARSS para generar las estimaciones de las varianzas de los errores observacional y del sistema, siendo $\widehat{V} = 444.1$ y $\widehat{W} = 46.2$ los estimadores máximo verosímiles. Para el segundo modelo se hacen los mismos supuestos con la diferencia de que la varianza de los errores del vector de estados es 10 veces mayor, es decir que, modificamos este valor accediendo al parámetro *par* de nuestro primer modelo por el siguiente $\widehat{W} = 462$.

Una vez especificados los modelos procedemos a calcular los valores filtrados del vector de estados para cada uno y por consiguiente las estimaciones un paso hacia adelante de las concentraciones de ozono de la estación el Pedregal, $f_t(1) = E(Y_t|y_{1:t-1})$. En la Figura 3.1 y la Figura 3.2 se muestran ambas series para $r = 0.10403$ y $r = 1.0403$, respectivamente. Es claro que para el modelo 2, con $r = W/V$ diez veces mayor, los valores predichos tienden a parecerse más al comportamiento de la serie de las concentraciones de ozono; es decir, se observa que, entre menor sea el ruido, el valor del *signal-to-noise-ratio* será mayor.

Como se mencionó anteriormente, una forma de verificar empirícamente qué tan bueno es un modelo, es a través de pruebas gráficas como son el qq-plot, el ACF de los residuales y la prueba de Ljung-Box. En la Figura 3.3 y la Figura 3.4 se muestran los resultados obtenidos para los modelos 1 y 2, respectivamente. En ambas figuras se puede apreciar que los p valores resultantes de la prueba de Ljung-Box son muy pequeños, relativamente cero, por lo que no hay evidencia suficiente para aceptar la hipótesis nula; es decir, los residuales estandarizados están correlacionados. Lo cual puede deberse a que las concentraciones de ozono no son las mismas mes con mes, es decir, se tiene una tendencia creciente en algunos, por ejemplo en el mes de Mayo, mientras que en otros es decreciente. Asimismo podemos notar que la distribución de los

residuales en ambos casos es leptocúrtica; tales resultados se acentúan en la gráfica qq-Plot, por lo que se concluye que los residuales no se distribuyen normal.

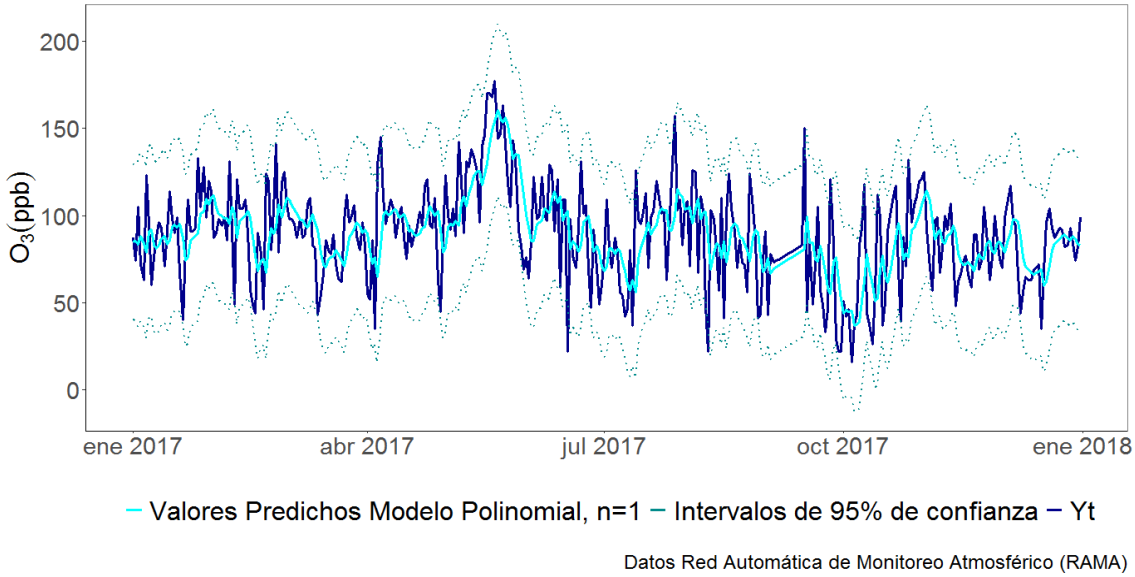


Figura 3.1: Valores filtrados de la serie de tiempo de las concentraciones diarias de ozono, tal que, $\widehat{V} = 444.1$, $\widehat{W} = 46.2$ y $r = \hat{W}/\hat{V} = 0.10403$.

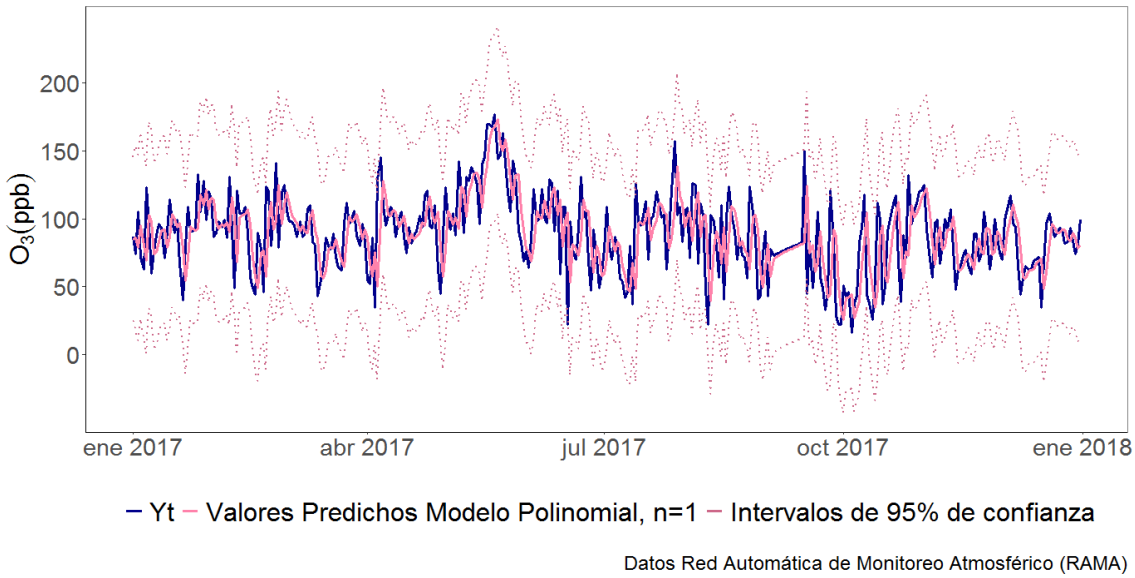


Figura 3.2: Valores filtrados de la serie de tiempo de las concentraciones diarias de ozono, tal que, $\widehat{V} = 444.1$, $\widehat{W} = 462$ y $r = \hat{W}/\hat{V} = 1.0403$.

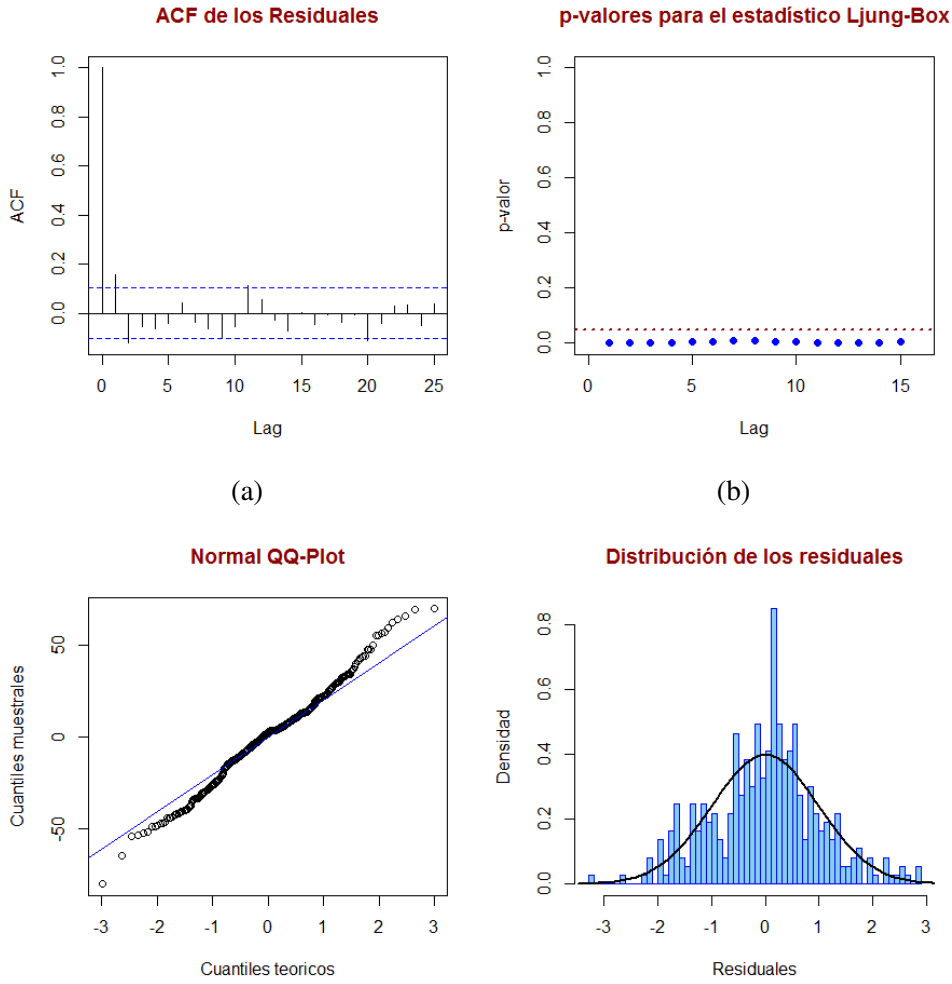


Figura 3.3: Validaciones Empíricas para el Modelo 1.

En el capítulo 4 de este trabajo, se mostrará el ajuste de un modelo dinámico cuya tendencia no es constante en el tiempo, como una alternativa para solucionar el problema de los modelos anteriores.

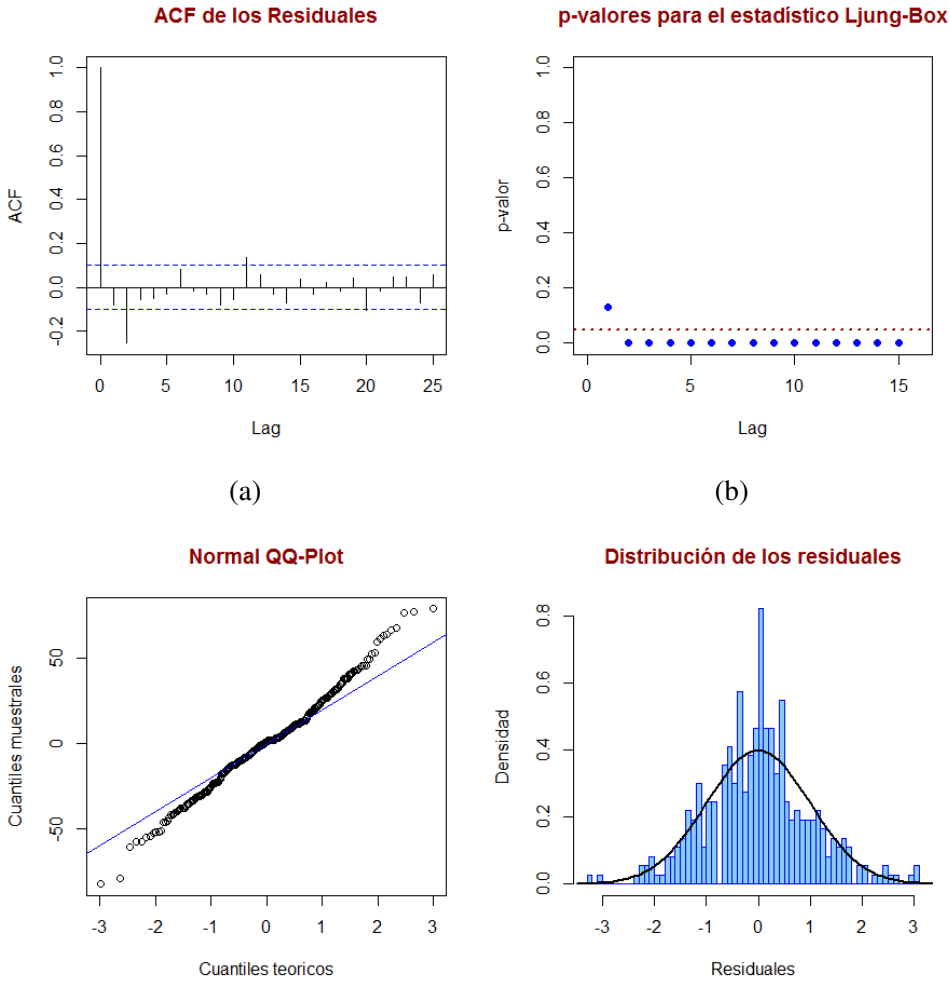


Figura 3.4: Validaciones Empíricas para el Modelo 2.

3.3. Modelo Polinomial de Segundo Orden

Este modelo, también conocido como modelo de tendencia lineal local o modelo de crecimiento lineal, ha sido utilizado históricamente para modelar series de tiempo con tendencia lineal con componentes estocásticos y parámetros que pueden ser interpretados con facilidad. De la Definición 3.2.1 de un modelo polinomial de orden n, se desprenden aquellas que definen a éste con $n = 2$,

$$Y_t = F_t\theta_t + \nu_t, \quad \nu_t \sim N(0, V_t).$$
$$\theta_t = G_t\theta_{t-1} + w_t, \quad w_t \sim N_2\left(0, \text{diag}(\sigma_\beta^2, \mu_\beta^2)\right).$$

donde las matrices $F_t = F$ y $G_t = G$ son constantes y están dadas por:

$$F = (1,0)^T, \quad G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Además, el vector de estados es igual a $\theta_t = (\mu_t, \beta_t)^T$ y μ_t suele interpretarse como el nivel local de la serie y β_t como la tasa de crecimiento local. Así, el modelo supone que el nivel actual de la serie cambia linealmente a través del tiempo y la tasa de crecimiento evoluciona.

Finalmente, la función predictiva es $f_t(k) = E(Y_{t+k}|y_{1:t}) = \hat{\mu}_t + k\hat{\beta}_t$.

3.4. Modelos Estacionales

En esta sección describiremos el modelo factor estacional, por ser la forma más simple de modelación del componente de estacionalidad [2] para un MDL. Como una alternativa para la representación de patrones cíclicos, se sugiere también el conocido modelo estacional forma de Fourier, el cual aparece en (West and Harrison, 1997). Dicha forma de modelación debe su nombre al uso de las funciones trigonométricas que llevan a representaciones de estacionalidad en formas de Fourier. Sin embargo, por practicidad y para los fines de este trabajo, dicho enfoque no se presenta.

3.4.1. Modelo Factor Estacional

Sea Y_t una serie de tiempo con período s; supongamos que ésta tiene media cero y no tiene componentes adicionales, es decir, no muestra tendencia ni posee covariables explicativas. Entonces puede ser modelada a través de un vector de estados de desviaciones estacionales, θ_t, de dimensión s y un MDL con las siguientes especificaciones: $F_t = F = (1,0,...,0)$, $G_t = G$ una matriz de permutación[3] de dimensiones $s \times s$ y un vector de factores estacionales α_j de dimensión s tal que $j = 1,...,s$.

Existen algunas restricciones de identificabilidad que deben ser impuestas al vector de factores estacionales; el más común es $\sum_{j=1}^{s} a_j = 0$. Este último supuesto implica que solo existen $s-1$ factores estacionales libres, pues el último queda determinado. Además, sugiere el uso de un modelo más parsimonioso, con un vector de estados de dimensión menor, $s-1$, y las matrices $F_t = F$ y $G_t = G$ constantes dadas por:

$$F = (1,0,...,0), \quad G = \begin{pmatrix} -1 & -1 & \cdots & -1 & -1 \\ 1 & 0 & & 0 & 0 \\ 0 & 1 & & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

[2] Se refiere a comportamientos cíclicos o periódicos.

[3] Sea φ una permutación, entonces la matriz de permutación φ, denotada por P_φ, estará dada por: $P_\varphi = \left(\delta_{\varphi(i),j}\right)_{1 \le i,j \le n}$. Es decir, en la i-ésima fila de la matriz P_φ la $\varphi(i)$-ésima entrada de ϕ es uno y todas las demás son cero.

donde F tene dimensión $s-1$ y G dimensiones $s-1 \times s-1$.

Bajo estas representaciones también es posible incluir variaciones dinámicas en los componentes estacionales, a través de la varianza del error de evolución del sistema, el cual puede definirse como: $W_t = \text{diag}\left(\sigma^2, 0, ..., 0\right)$.

Ahora bien, con el fin de ejemplificar el modelo descrito, supongamos que se tiene una serie de tiempo con media cero y sin tendencia aparente, Y_t tal que $t = 1, 2, ...$, con datos trimestrales. Además suponga que las desviaciones respecto a la media estarán expresadas por los coeficientes α_1, α_2, α_3 y α_4, para cada uno de los trimestres del año. Así por ejemplo, si Y_{t-1} se refiere al primer trimestre del año, Y_t al segundo trimestre, etc., supondremos que,

$$Y_{t-1} = \alpha_1 + v_{t-1}. \tag{3.1}$$

$$Y_t = \alpha_2 + v_t. \tag{3.2}$$

Entonces el MDL quedará descrito por el vector de estados $\theta_t = (\alpha_1, \alpha_4, \alpha_3, \alpha_2)$ y las ecuaciones usuales:

$$Y_t = F_t \theta_t + \nu_t, \quad \nu_t \sim N_1(0, V_t).$$
$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim N_4(0, W_t).$$

con $F_t = F$ y $G_t = G$ matrices constantes:

$$F = (1,0,0,0), \quad G = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Note que la multiplicación de matrices $G\theta_{t-1}$ da como resultado la permutación del vector de estados al tiempo $t-1$, es decir que, G va rotando a los elementos del vector de estados θ_t, así $\theta_t = (\alpha_2, \alpha_1, \alpha_4, \alpha_3)^T + w_t$.

Para que se cumplan las ecuaciones (3.1) y (3.2) la matriz de varianzas de la de los errores de la ecuación del sistema W_t, debe ser igual a una matriz de ceros de dimensiones 4×4, es decir estamos bajo el caso de un modelo factor estacional estático. Sin embargo, de manera general los efectos estacionales α_j tal que $j = 1, 2, ..., s$, pueden variar en el tiempo, entonces la matriz W_t es distinta de cero y debe ser especificada con mucho cuídado.

3.5. Modelos de Regresión Dinámicos Lineales

En esta sección incorporaremos componentes de regresión a un modelo dinámico lineal. Cabe mencionar que, aunque la estructura de los modelos dinámicos lineales permite una gran cantidad de formas posibles de relaciones a través de la elección adecuada de variables de regresión o una combinación de éstas, nos centraremos en los modelos de regresión múltiple con variables cuantitativas.

Definición 3.5.1. ***(Modelos de regresión múltiple dinámicos)***. *Sea* $(Y_t)_{t\geq 1}$ *una serie de tiempo afectada por un conjunto de p variables independientes* $(X_t)_{t\geq 1}$. *Considere el valor de la i-ésima variable explicativa conocido, al cual denotaremos por* $X_{i,t}$, $(i = 1,...,p; t = 1,...)$. *Observe que suele incluirse un término constante en el modelo,* $X_{1,t} = 1$, *para toda t, el cual está asociado al parámetro de intercepción. Así el modelo de regresión múltiple dinámico lineal queda definido como:*

$$Y_t = F_t\theta_t + v_t, \quad v_t \sim N_1(0, V_t).$$
$$\theta_t = G_t\theta_{t-1} + w_t, \quad w_t \sim N_p(0, W_t).$$
$$\theta_0 \sim N_p(m_0, C_0).$$

donde $F_t = (X_{1,t},...,X_{p,t})$ *es un vector columna de p variables regresoras,* $\theta_t = (\theta_{1,t},...,\theta_{p,t})^T$ *es el vector de parámetros de regresión asociado de dimensiones* $p \times 1$, V_t *la varianza observacional y* W_t *la matriz de varianzas del sistema* θ_t.

Cabe observar que una elección comunmente utilizada para la ecuación de estados es tomar la matriz de evolución del sistema, G_t, como la matriz identidad, y a la matriz de varianzas, W_t, como una matriz diagonal. Es decir, modelar los coeficientes de regresión como una caminata aleatoria. Además note que el modelo de regresión lineal estático es un caso particular del modelo anterior, donde $W_t = 0$ para toda t, lo cual implica que $\theta_t = \theta$ es constante en el tiempo.

Dada la Definición 3.5.1 es importante notar que, en una regresión dinámica la evolución en el tiempo es posible a través del término de error de la ecuación del sistema, w_t, pues describe los cambios en los elementos del vector de parámetros de regresión entre $t-1$ y t. Asimismo, que el vector de medias de su distribución sea cero, refleja las creencias esperadas de que θ_t sea constante sobre el intervalo, mientras que la matriz de varianzas W_t gobierna los movimientos del vector de parámetros de regresión θ_t.

3.5.1. Características en el Análisis de un Modelo de Regresión Dinámico

Como en cualquier modelo estadístico existen varios puntos que suelen ser relevantes en el ajuste, por lo que se mencionarán brevemente algunas características asociadas a los modelos bajo estudio.

(I) Estabilidad. En algunos contextos donde el objetivo principal del modelo es realizar pronósticos, la estabilidad en el mismo es primordial. En el caso de los modelos dinámicos lineales es deseable que los términos de error de la ecuación del sistema sean pequeños, pues de lo contrario; es decir que, si los elementos de la matriz de varianzas de evolución resultan ser demasiado grades pueden presentarse los siguientes problemas: las distribuciones predictivas se vuelven difusas; en la actualización, el peso de observaciones nuevas suele ser alto, por lo que la distribución posterior se adapta de forma marcada de una observación a otra, lo que ocasiona que aún cuando predicciones a corto plazo sean precisas en términos de localización; las prediciones a mediano y largo plazo puedan ser pobres en términos de localización y muy difusas.

(II) Problema de multicolinealidad. En la práctica suele ocurrir que la matriz de precisión esté cerca de ser singular[4], al tener un determinante positivo muy pequeño. Lo cual indica fuertes relaciones de dependencia entre las variables regresoras, o bien, el conocido problema de multicolinealidad[5] en regresión. Tal situación lleva a modelos en los que los parámetros estimados tienen grandes errores, es decir, donde los parámetros pueden variar ampliamentre de una muestra a otra disminuyendo la precisión de la información que tenemos acerca del valor verdadero de los parámetros. Esto puede resultar erróneamente en que muchas de las covariables sean no significativas para el modelo.

Una de las acciones que suele tomarse ante dicha problematica consiste en reducir el número de regresores del modelo, eliminando aquellas que son redundantes para éste; sin embargo, al no realizarse de forma adecuada, puede provocar que los parámetros estimados puedan estar sesgados cuando una o más covariables relevantes sean excluidas del mismo.

(III) Ortogonalidad. En modelos de regresión estáticos, estandarizar los regresores suele ser una práctica común. Dada una muestra de observaciones y un conjunto de regresores se les resta la media aritmética y se divide el resultado entre la desviación estándar. La razón por la que se resta la media es que los efectos de regresión son separados claramente del parámetro de intercepción en el modelo, siendo esencialmente ortogonales a este, lo cual permite una interpretación más sencilla.

Para ejemplificar el término de ortogonalidad considere un modelo de regresión simple dado por $Y_t = \alpha + \beta X_t$, que puede reescribirse como $Y_t = \left(\alpha + \beta\overline{X}\right) + \beta\left(X_t - \overline{X}\right) = \alpha^* + \beta X_t^*$, donde $\overline{X}$ es la media aritmética y $\alpha^* = \alpha + \beta\overline{X}$ el nuevo intercepto del modelo. El nuevo vector de regresión $F_t = (1, X_t)^T$, es tal que $\sum_t F_t F_t^T$ es diagonal y ahora fácilmente invertible.

En un modelo de regresión dinámico, y más aún cuando se consideran series de tiempo con observaciones tomadas en un período relativamente corto, la forma de estandarización descrita aplica bajo los mismos principios.

Si bien en esta última sección describimos los modelos dinámicos lineales más simples para series de tiempo con observaciones univariadas, nos referimos a Campagnoli et al. (2009), para el estudio de otros problemas en el análisis de series de tiempo multivariantes. Entre ellos: modelos dinámicos lineales para datos longitudinales; es decir, problemas donde se tiene una cantidad y observada para m unidades en el tiempo, tal que $(Y_t)_{t\geq 1}$ con $Y_t = (Y_{t,1}, ..., Y_{m,t})^T$; modelos dinámicos jerárquicos, como una extensión de los sistemas dinámicos; las ecuaciones de series de tiempo aparentemente no relacionadas (SUTSE, por sus siglas en inglés), como una clase de modelos caracterizados por la especificación de la estructura de dependencia entre los vectores de estados $\theta_t^{(1)}, ..., \theta_t^{(m)}$, entre otros.

[4] Sea A una matriz cuadrada de orden N, diremos que es singular si su determinante es cero, es decir, $\det(A) = 0$. Asimismo una matriz singular es no invertible.

[5] Se refiere al caso en el que dos o más covariables en el modelo de regresión están correlacionadas una con la otra. Diremos que tendrán colinealidad exacta, si el coeficiente de correlación $\rho = 1$ y colinealidad aproximada, cuando $|\rho| \simeq 1$.

En el siguiente capítulo pondremos en práctica los conocimientos adquiridos y ejemplificaremos el uso de los modelos dinámicos lineales. A grandes rasgos construiremos un modelo de regresión dinámico para datos sobre las concentraciones de ozono de la estación el Pedregal en la Ciudad de México.

Capítulo 4

Análisis del Pronóstico de las Concentraciones de Ozono

En los últimos años el crecimiento demográfico y junto con éste el desarrollo industrial que ha venido experimentando la Ciudad de México, han ocasionado que problemas como la contaminación atmosférica y el deterioro de la calidad del aire se vean agravados. Es por esto que en la actualidad el estudio de medidas para contrarrestar sus efectos ha tomado mayor relevancia. El ozono ha formado parte de diversos estudios e investigaciones debido a que es uno de los contaminantes más dañinos y los efectos adversos con los que se relaciona. Su tiempo de vida depende de las condiciones geográficas, climatológicas, meteorológicas, entre otras, registrándose las concentraciones más elevadas cuando la temperatura aumenta.

4.1. Estaciones de Monitoreo RAMA

En sus esfuerzos por coadyuvar con el monitoreo y la vigilancia de la calidad del aire de la Ciudad de México y su área conurbada, la Secretaría de Desarrollo Urbano y Ecología (SEDUE), en 1986 puso en marcha la Red Automática de Monitoreo Atmosférico (RAMA) para la medición de los principales contaminantes del aire en la Ciudad de México, entre los que se encuentra el ozono. Cabe mencionar que tales mediciones son tomadas automáticamente segundo por segundo; sin embargo, los valores que se registran para realizar estadísticas corresponden al promedio de las concentraciones del contaminante registradas en una hora (concentración horaria[1]) en partes por billón (ppb). La RAMA inició operaciones con veinticinco estaciones automáticas de monitoreo, pero continuó ampliando su cobertura y sufriendo modificaciones a lo largo del tiempo, hasta conformarse por treinta y cuatro estaciones de monitoreo y formar parte del Sistema de Monitoreo Atmosférico de la Ciudad de México (SIMAT).

Para los fines de este escrito, nos restringimos al análisis de las concentraciones horarias de

[1]Concentración horaria, dato horario o promedio horario: al promedio o media aritmética de las concentraciones registradas en el intervalo de tiempo de 60 minutos delimitado por los minutos 0 y 59 de la hora. Para efectos del manejo de datos se considerará válido, cuando se calcule con al menos el 75 % de las concentraciones registradas en la hora.

ozono del 1° de enero al 31 de diciembre de 2017. Como un primer acercamiento en la Figura 4.1 mostramos un histograma de nuestros datos, en el que podemos observar que su distribución es sesgada hacia la derecha, es decir, de colas pesadas. Por esta razón, la media siempre queda por arriba del valor de la mediana, lo cual podría considerarse como una sobreestimación de las concentraciones horarias de O_3. Por ello tomamos la mediana de nuestras observaciones al ser una medida de tendencia central más robusta. Adicional a esto, en la Tabla 4.1 se muestran las principales medidas de tendencia central para nuestras observaciones.

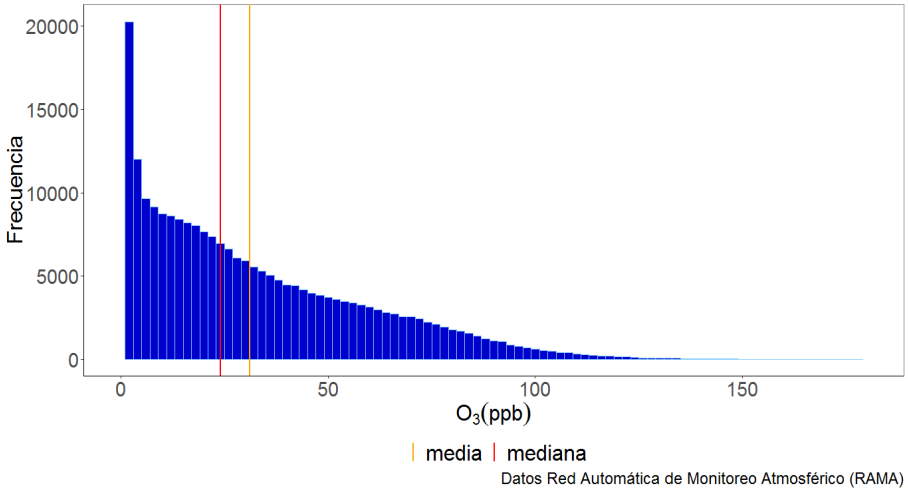


Figura 4.1: Histograma de la distribución de las concentraciones de ozono como promedio horario en la Ciudad de México y zona conurbada, del 1° de Enero al 31 de Diciembre 2017.

N	Min	Q1	Mediana	Media	Q3	Máx	Obs. Faltantes
297840.00	0.00	8.00	24.00	31.14	48.00	190.00	44833.00

Tabla 4.1: Medidas de tendencia central de las concentraciones de ozono como promedio horario en la Ciudad de México y zona conurbada, del 1° de Enero al 31 de Diciembre 2017.

Uno de los objetivos de la RAMA es evaluar el cumplimiento de normas estándares. De acuerdo a la Norma Oficial Mexicana vigente (NOM-020-SSA1-2014) el valor límite permisible para la concentración de ozono en el aire ambiente es de 0.095 partes por millón (ppm) como promedio horario (Secretaria de Salud, 2014).

Con base en este valor obtendremos el subconjunto de las observaciones cuya concentración de ozono es mayor que 95 ppb. En estadística a este tipo de datos se les conoce como excesos sobre un umbral (*Exceedances over a Threshold*) y existe una vasta Teoría de Valores Extremos (TVE) para cuantificarlos, así como medir sus consecuencias o efectos futuros. Si bien en este escrito trabajaremos desde otro enfoque, resulta interesante observar el comportamiento de este

tipo de datos, para determinar las estaciones de monitoreo cuyas concentraciones de ozono en el tiempo tuvieron mayor variabilidad, hasta alcanzar valores por arriba del permisible.

La Figura 4.2 muestra la distribución de las observaciones por estación de monitoreo, mientras que el número de horas en las que se superó el valor permisible establecido por la NOM pueden verse en la Figura 4.3.

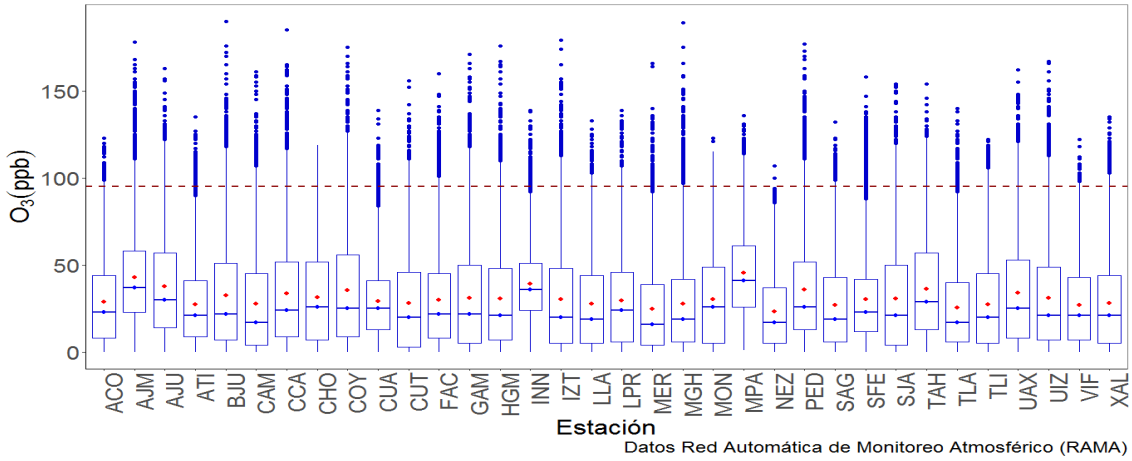


Figura 4.2: Boxplot del promedio de las concentraciones de ozono como promedio horario que sobrepasaron las 95 ppb durante el 1° de Enero al 31 de Diciembre 2017. En rojo se marca la media de la distribución, mientras que la línea punteada señala el valor límite permisible por la NOM.

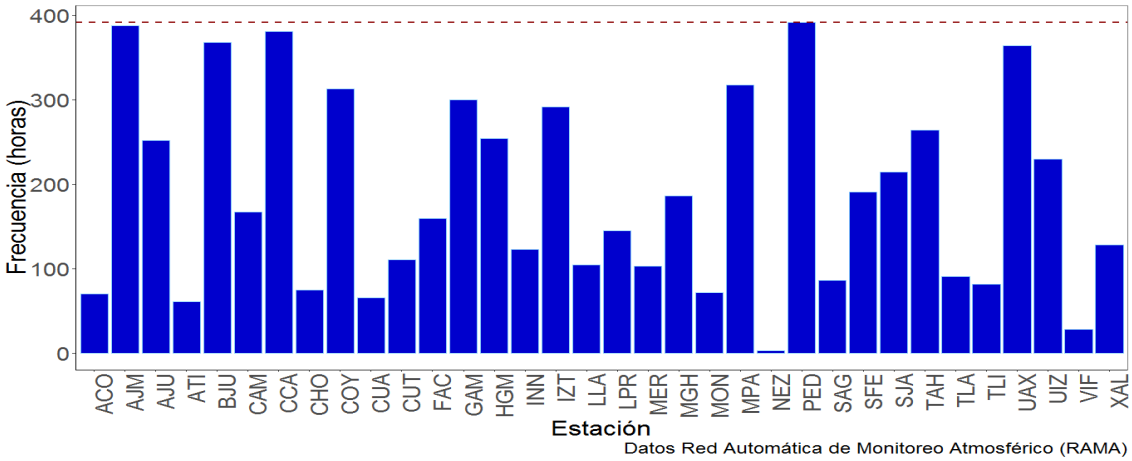


Figura 4.3: Frecuencia de las concentraciones de ozono como promedio horario que sobrepasaron las 95 ppb del 1° de Enero al 31 de Diciembre 2017.

En la Tabla 4.2 se indica el número de horas mayores a los valores límite de la NOM por estación de monitoreo, aquellas con el mayor número por arriba del valor permisible en partes

Estación de monitoreo	Clave	Obs. >95 (ppb)	Obs. día	Máx. día (ppm)
Acolman	ACO	70	26	0.123
Ajusco Medio	AJM	388	146	0.178
Ajusco	AJU	252	99	0.163
Atizapán	ATI	61	28	0.135
Benito Juarez	BJU	368	144	0.190
Camarones	CAM	167	67	0.161
Centro de Ciencias de la Atmósfera	CCA	381	136	0.185
Chalco	CHO	75	38	0.119
Coyoacán	COY	313	106	0.175
Cuajimalpa	CUA	66	33	0.139
Cuautitlán	CUT	111	38	0.156
FES Acatlán	FAC	160	60	0.160
Gustavo A. Madero	GAM	300	102	0.171
Hospital General de México	HGM	254	95	0.176
Investigaciones Nucleares	INN	123	58	0.139
Iztacalco	IZT	292	110	0.179
Los Laureles	LLA	105	42	0.133
La Presa	LPR	145	60	0.139
Merced	MER	103	39	0.166
Miguel Hidalgo	MGH	186	74	0.189
Montecillo	MON	72	33	0.123
Milpa Alta	MPA	318	105	0.136
Nezahualcóyotl	NEZ	3	2	0.107
Pedregal	PED	392	148	0.177
San Agustín	SAG	86	36	0.132
Santa Fe	SFE	191	68	0.158
San Juan de Aragón	SJA	215	77	0.154
Tláhuac	TAH	264	101	0.154
Tlalnepantla	TLA	91	39	0.140
Tultitlán	TLI	82	34	0.122
UAM Xochimilco	UAX	364	138	0.162
UAM Iztapalapa	UIZ	230	85	0.167
Villa de las Flores	VIF	28	12	0.122
Xalostoc	XAL	128	57	0.135

Tabla 4.2: Obs. >95 (ppb): Número de horas en las que se súpero el valor límite permisible como promedio de las mediciones de O_3 en una hora (concentración horaria) en ppb. Obs. día: Número de días en los que se sobrepaso el límite permisible como promedio de las mediciones de O_3 en una hora (concentración horaria) en ppb. Máx. día (ppm): Concentraciones diarias de O_3.

por billón durante el 2017 fueron: el Pedregal (PED), Ajusco Medio (AJM), Centro de Ciencias de la Atmósfera (CCA) y Miguel Hidalgo (MGH). La estación de monitoreo el Pedregal (PED) ubicada en la delegación Álvaro Obregón, reportó los niveles del contaminante más altos durante el 2017 con 392 horas distribuidas en 148 días por arriba del nivel permisible. Mientras que la estación Nezahualcóyotl (NEZ) en el Estado de México, tuvo el mínimo número de horas por arriba de las 95 ppb con solo 3 excesos en dos días. Es interesante notar que, según estadísticas de la Dirección de Monitoreo Ambiental, históricamente la estación el Pedregal ha sido aquella con mayores concentraciones de ozono como promedio horario. Finalmente con respecto a las concentraciones diarias, mostradas en la última columna de la Tabla 4.2, la estación Benito Juárez (BJU) presentó el valor máximo con 0.190 en ppm.

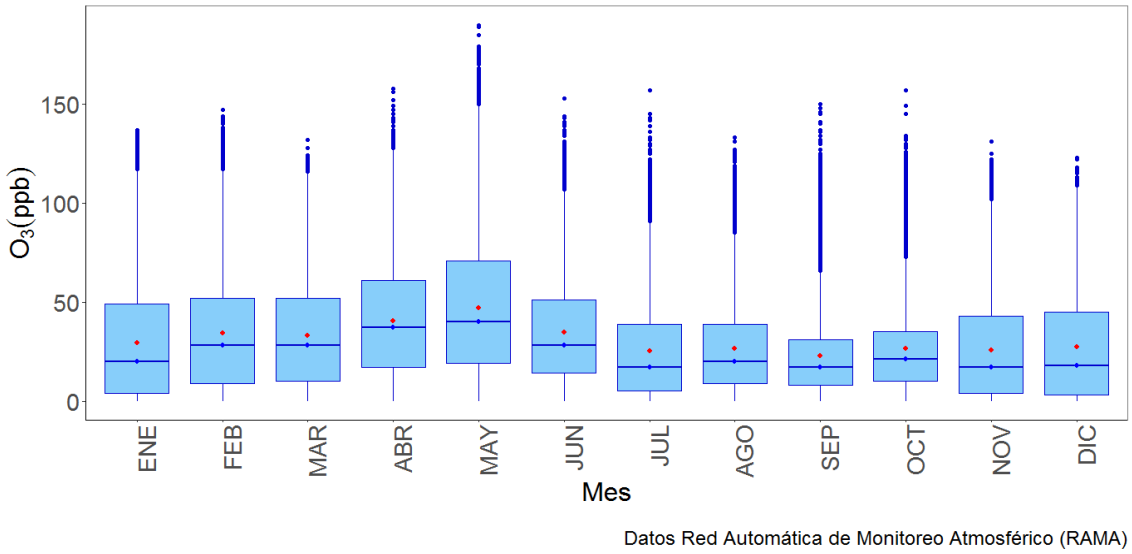


Figura 4.4: Boxplot de las concentraciones de ozono como promedio horario en la Ciudad de México y zona conurbada, del 1° de Enero al 31 de Diciembre 2017.

Cabe agregar que durante los meses de abril y mayo los niveles de dicho contaminante parecen aumentar, ver Figura 4.4, mostrando una tendencia decreciente durante el mes de junio, la cual se mantiene hasta el mes de diciembre. Si bien es cierto que la Ciudad de México y zona conurbada presentan características diversas, podemos pensar que se caracterizan por tener un clima templado con temperaturas moderadas a lo largo del año; sin embargo es la primavera, calenderizada del 1° de marzo al 30 de mayo, la estación que suele ser la más cálida y seca con temperaturas hasta de 30°C por las tardes. Por lo que podemos suponer que mientras más alta sea la temperatura, mayores serán las concentraciones de ozono.

4.2. Estación de Monitoreo el Pedregal (PED)

En este estudio, utilizaremos la serie de las concentraciones diarias de ozono en la estación de monitoreo automática el Pedregal, Figura 4.5, la cual hemos identificado como la estación con

mayores concentraciones en la Ciudad de México. Aunque, en la literatura podemos encontrar distintas medidas para el análisis de series de tiempo de este contaminante, como son: los promedios de las concentraciones de ozono (Huerta et al., 2004), las concenttraciones máximas diarias de 8 horas (Sahu et al., 2007), los máximos de las concentraciones del promedio móvil de 8 horas (Sahu and Bakar, 2012), entre otras.

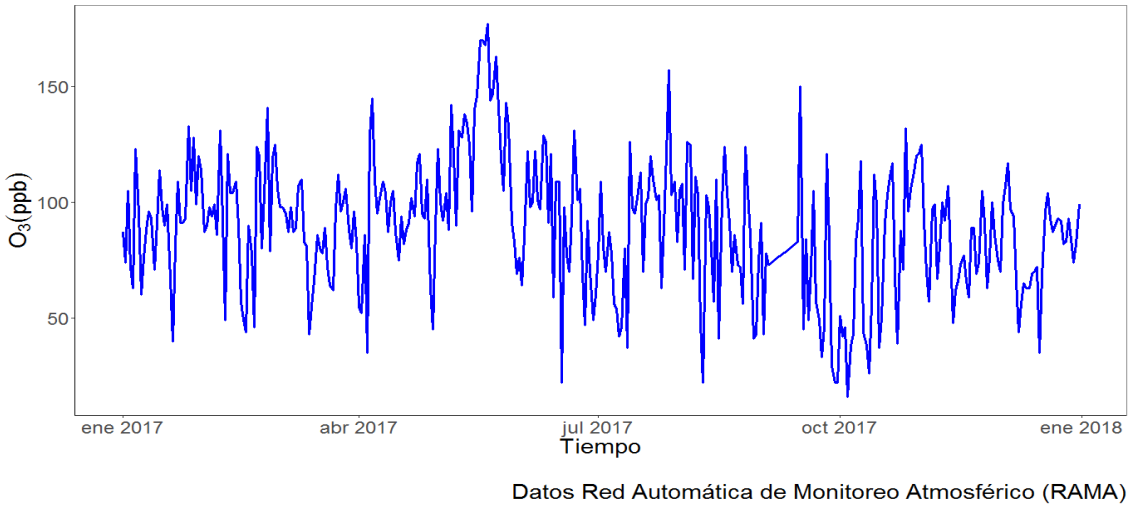


Figura 4.5: Serie de tiempo de las concentraciones diarias de ozono en ppb de la estación el Pedregal, del 1° de Enero 2017 al 31 de Diciembre 2017.

Variable	Min	Q1	Mediana	Media	Q3	Máx	Obs. Faltantes
O_3	16.00	70.00	91.00	89.21	105.00	177.00	16
TMP	14.30	21.10	23.00	22.90	24.60	29.50	5
WSP	29.00	66.00	74.50	73.24	84.00	94.00	5
RH	1.90	3.00	3.80	3.81	4.50	8.40	6

Tabla 4.3: Medidas de tendencia central de las concentraciones diarias de ozono en la estación el Pedregal y de las covariables: temperatura (TMP), velocidad del viento (WSP) y humedad relativa (RH), del 1° de Enero al 31 de Diciembre 2017.

Nuestro objetivo principal es establecer un modelo de regresión múltiple dinámico lineal para el pronóstico temporal del nivel de ozono. Para dicho modelo emplearemos como covariables: la temperatura (TMP) medida en grados celcius; la velocidad del viento (WSP) en m/s; y la humedad relativa (RH) medida en %. En la Tabla 4.3 se muestran las principales medidas de tendencia central del ozono y las tres variables meteorológicas. En el apéndice A se muestra el código empleado para la obtención de la base de datos transformada (concentraciones diarias).

Un diagrama de dispersión de las concentraciones diarias de ozono contra las variables meteorológicas bajo estudio, se muestra en la Figura 4.6. El diagrama de dispersión entre la temperatura y las concentraciones de ozono presenta una forma triangular, es decir que, pareciera que a mayor temperatura, mayor es el nivel de las concentraciones de ozono registradas. Aunque el resto de los regresores no presenta una relación aparente contra la variable dependiente (O_3), se considerarán en la siguiente sección, para el ajuste de un modelo de regresión múltiple y se verificará si son o no significativas.

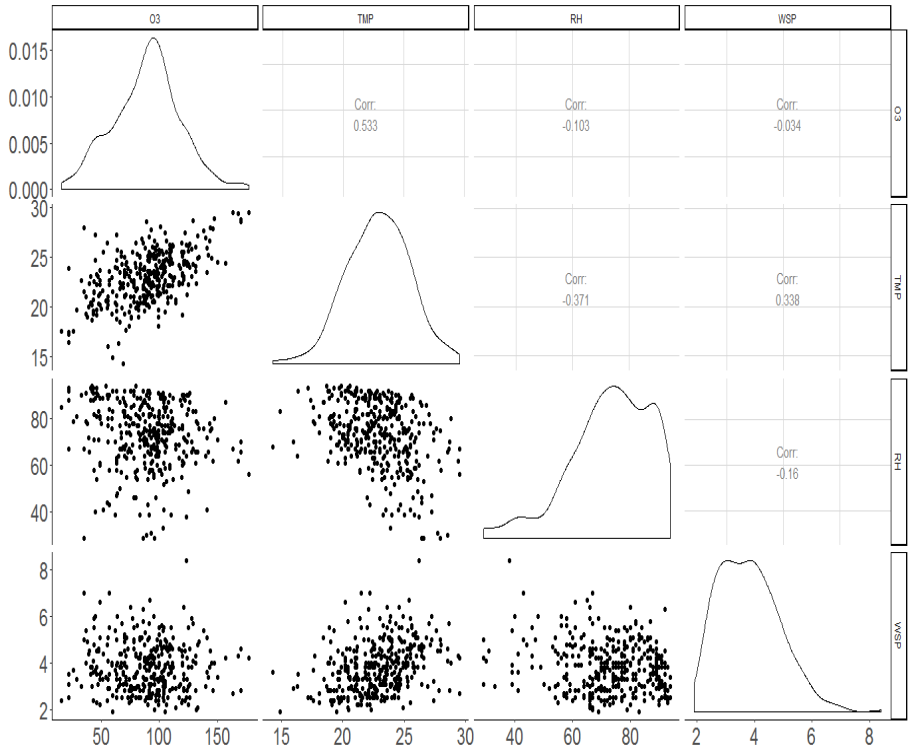


Figura 4.6: Diagrama de dispersión de las concentraciones de ozono contra las mediciones de las variables meteorológicas: temperatura, humedad relativa y velocidad del viento, de la estación PED.

Otro punto relevante en el tratamiento de la base de datos es la presencia de valores faltantes. Actualmente las estaciones de monitoreo de la RAMA operan los 365 días del año; sin embargo el proceso suele interrumpirse debido a períodos de calibración y medición de los instrumentos, entre otras causas inmersas, que provocan la ausencia de datos.

Como se puede observar en la Tabla 4.3, el número de observaciones faltantes respecto al total pareciera no ser significativo para el análisis; sin embargo para nuestros fines, y con el objeto de no repercutir en la inferencia o el desempeño del modelo predictivo, consideramos un método de imputación para estimarlos. El paquete empleado *imputTS*, desarrollado para el software estadístico R, ofrece un vasto conjunto de algoritmos especializados en el proceso de reemplazo de valores desconocidos en series de tiempo univariadas, entre ellos: interpolación, promedios móviles, sustitución por alguna medida de tendencia central (media, mediana y moda),

suavizamiento de Kalman, entre otros.

El método utilizado fue el algoritmo de suavizamiento de Kalman, descrito a grandes rasgos en el capítulo 2 de este trabajo, por ser uno de los más robustos para series de tiempo univariadas con una tendencia marcada y ciclos estacionales.

4.3. Modelo de Regresión Lineal Múltiple

Con el fin de explicar y contrastar los resultados de un modelo de regresión dinámico lineal, primero mostraremos el ajuste de un modelo de regresión múltiple para nuestros datos. Como vimos en el capítulo previo, una práctica común es la normalización de las covariables (TMP, RH y WSP), y dado que las nuestras están medidas en distintas escalas según su naturaleza, estandarizaremos cada una de ellas, restando la media para centrar los datos y dividiendo entre la desviación estándar para escalarlos. Una de las ventajas de dicho procedimiento es que facilita la interpretación de los parámetros de regresión estimados.

El modelo de regresión lineal múltiple con las tres covariables consideradas queda especificado de la siguiente manera:

Sea $(Y_t)_{t\geq 0}$ la serie de concentraciones diarias de ozono O_3 de la estación el Pedregal, con $t = 1,...,365$ y $(\beta_1,\beta_2,\beta_3)^T$ el vector de parámetros de regresión asociado, dónde α es el parámetro de intercepción, entonces se tiene la siguiente ecuación:

$$Y_t = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \text{ tal que } i = 1,...,n$$

donde

$x_1 =$ Observaciones diarias de la temperatura en la estación el Pedregal.

$x_2 =$ Observaciones diarias de la humedad relativa en la estación el Pedregal.

$x_3 =$ Observaciones diarias de la velocidad del viento en la estación el Pedregal.

```
Call:
lm(formula = Yt.s ~ x1 + x2 + x3)

Residuals:
     Min       1Q   Median       3Q      Max
-77.243  -15.835    2.078   15.711   57.469

Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)    88.682       1.207   73.488  < 2e-16 ***
x1             18.284       1.366   13.389  < 2e-16 ***
```

```
x2            2.481      1.299    1.909     0.057 .
x3           -6.555      1.286   -5.096 5.61e-07 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

Residual standard error: 23.06 on 361 degrees of freedom
Multiple R-squared:  0.3416,    Adjusted R-squared:  0.3361
F-statistic: 62.42 on 3 and 361 DF,  p-value: < 2.2e-16
```

Utilizamos la función *lm()* de la librería *stats* dentro del software R para el ajuste de un primer modelo con todas las covariables. A primera vista, la covariable humedad relativa (x_{2i}) pareciera no ser significativa para el modelo; sin embargo, emplearemos métodos de selección de variables para determinar aquel con el que trabajaremos. Existen dos clases de métodos comúnmente utilizados para la selección de variables, estos son: *best subset selection* y *stepwise selection*.

El método conocido como *Best Subset Selection*, se caracteriza por ser un algoritmo para encontrar la mejor combinación de p predictores entre todas las combinaciones posibles y consiste en lo siguiente:

i) Definir M_0 como el modelo con ningún predictor, es decir, aquel que predice la media muestral para cada observación.

ii) Ajustar para cada una de las combinaciones posibles de p predictores un modelo de regresión por mínimos cuadrados; es decir, para $k = 1,2,...,p$, obtener las $\binom{p}{k}$ combinaciones con exactamente k predictores. Posteriormente obtener el mejor modelo entre éstas, M_k, definido como aquel que minimiza la suma de cuadrados residual [2] (RSS, por sus siglas en inglés) o equivalentemente aquel que maximiza el coeficiente de determinación [3].

iii) Seleccionar un solo modelo entre $M_0,...,M_p$ utilizando alguna medida estadística como el criterio de información de Akaike (AIC, por sus siglas en inglés) o el criterio de información bayesiano (BIC, por sus siglas en inglés).

Por otra parte, el método *Stepwise Selection* o selección de variables paso a paso, caracterizado por ser un algoritmo computacionalmente eficiente. Suele utilizarse cuando el número de covariables p es grande. Son parte de esta clase: el agoritmo de selección de variables hacia adelante, comocido como *Forward Stepwise* y el algoritmo de selección de variables hacia atrás

[2] La suma de de los cuadrados residual, también conocida como suma del cuadrado de los errores de predicción (SSE), para un modelo con una sola variable explicativa se define como: $\sum_{i=1}^{n}(\varepsilon_i)^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, es decir la suma del cuadrado de las diferencias entre la i-ésima observación y los valores predichos para cada observación y_i.

[3] En estadística el coeficiente de determinación R^2 es una medida para cuantificar si los valores ajustados describen bien los datos, y_i, pues determina la proporción de la variación total explicada por el modelo de regresión. Se define como $R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$. Es decir, compara la suma de los cuadrados de regresión contra la suma total de cuadrados (es la suma de los errores de regresión más la suma de los cuadrados de los residuales).

conocido como *Backward Stepwise*. El primero de ellos debe su nombre al proceso de selección de variables, ya que inicia con un modelo con cero predictores y en cada paso añade uno hasta incluir los p predictores totales en el modelo, en particular en cada paso se asignará una covariable al mejor modelo si ésta brinda una mejoría en el ajuste. Por su parte, el segundo realiza el proceso inverso, es decir, incia con un modelo con el total de predictores p y en cada paso elimina uno hasta determinar el mejor modelo ajustado. Si bien esta clase de modelos son útiles en muchos casos, para los fines de este trabajo y dado el número reducido de covariables, emplearemos la primera clase de modelos de selección de variables.

Para implementarlo utilizaremos la función *regsubsets* que forma parte de la paquetería *leaps*, cuyo objeto es identificar el mejor modelo para un número de predictores dado. Los parámetros que recibe son: el modelo de regresión, la base de datos de la que proviene cada variable y el número k de predictores para efectuar las combinaciones posibles. Por default la función genera los subconjuntos posibles de modelos hasta con ocho predictores; sin embargo, el parámetro *nvmax* permite especificar modelos con $k > 8$.

El siguiente código en R, nos muestra los siete modelos posibles por número de predictores ordenados del 1 al 3 en cada caso, siendo aquel con el número 1 el mejor modelo ajustado. Es decir que, para el grupo con solo una covariable, el mejor modelo es aquel que incluye la covariable temperatura; para el segundo grupo aquel que incluye las covariables temperatura y velocidad del viento; y finalmente el tercer grupo aquel que contiene estas dos últimas más la humedad relativa.

```
>   combinaciones <- regsubsets(Yt.s ~ x1 + x2 + x3,
+                               data=base.s, nbest = 7)
> summary(combinaciones)
Subset selection object
Call: regsubsets.formula(Yt.s ~ x1 + x2 + x3, data = base.s, nbest =
    7)
3 Variables  (and intercept)
   Forced in Forced out
x1     FALSE      FALSE
x2     FALSE      FALSE
x3     FALSE      FALSE
7 subsets of each size up to 3
Selection Algorithm: exhaustive
         x1  x2  x3
1  ( 1 ) "*" " " " "
1  ( 2 ) " " "*" " "
1  ( 3 ) " " " " "*"
2  ( 1 ) "*" " " "*"
2  ( 2 ) "*" "*" " "
2  ( 3 ) " " "*" "*"
3  ( 1 ) "*" "*" "*"
```

La Figura 4.7 muestra el valor del coeficiente de determinación para cada una de las combinaciones posibles obtenidas con la función *regsubsets*. Como se puede apreciar, este coeficiente se maximiza para el modelo completo, $Y_t \sim x_1 + x_2 + x_3$, seguido del modelo que considera únicamente a las covariables temperatura y velocidad del viento, $Y_t \sim x_1 + x_2 + x_3$.

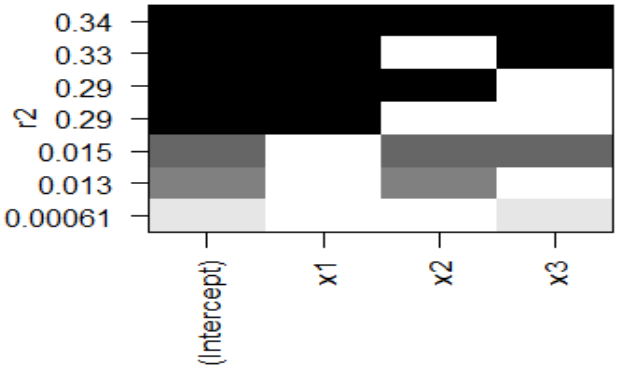


Figura 4.7: Resultados del algoritmo de selección de variables basado en el coeficiente de determinación, R^2.

Una vez que hemos obtenido el mejor modelo con $p = 1,2,3$ predictores, vamos a determinar cuál de ellos es el mejor sin importar el número de predictores en cada caso. Para lo que utilizaremos un método de comparación de modelos llamado validación cruzada de K iteraciones, el cual consiste en generar K subconjuntos de valores de la muestra, regularmente 10, siendo uno de ellos el de prueba y los $K-1$ restantes los modelos de entrenamiento. El proceso de entrenamiento se ejecuta $K = 10$ veces para cada combinación obtenida con los datos de entrenamiento hasta obtener los errores de predicción (MSE)[4] para cada uno de los modelos con p covariables; es decir, obtenemos una matriz de errores de predicción de dimensiones 10×3. Finalmente obtenemos la media aritmética de los errores resultantes de las 10 iteraciones para los modelos con uno, dos y tres predictores.

Como se puede observar el modelo completo es aquel que minimiza el promedio de los errores de predicción, con un valor de 549.9251, seguido del modelo con las covariables temperatura y velocidad del viento, 550.8936. Sin embargo, en la Tabla 4.4 se puede ver que, aunque el modelo completo fue aquel con el menor MSE, no se puede aceptar la hipótesis nula de que cada uno de los parámetros de regresión es distinto de cero, pues el intervalo de confianza para el coeficiente

[4]Los valores de predicción se obtienen como el promedio de la diferencia entre las concentraciones de ozono del conjunto de prueba y los valores predichos al cuadrado, ver Figura 4.8.

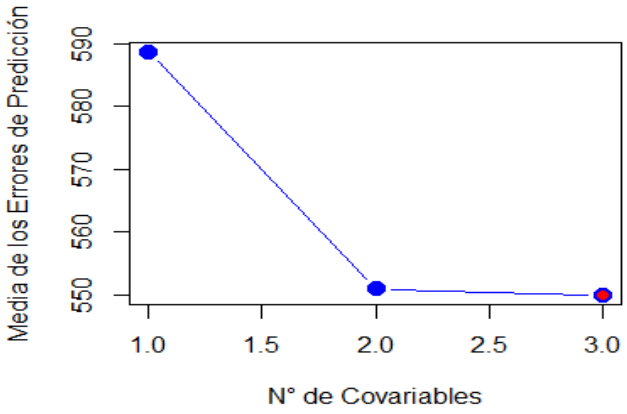


Figura 4.8: Errores de predicción del método de validación cruzada con $K = 10$ iteraciones. En rojo se señala el modelo con el menor error cuadrático medio (MSE).

lo contiene, con lo que se puede concluir que el parámetro de regresión $\hat{\beta}_2$ no es significativo para el modelo.

Modelo	Parámetro	Valor Estimado	I.C.
$Y_t = \alpha + \beta_1 x_{1i}$	$\hat{\alpha}$	88.682	(86.2182, 91.1462)
	$\hat{\beta}_1$	15.138	(12.6703, 17.6051)
$Y_t = \alpha + \beta_1 x_{1i} + \beta_3 x_{3i}$	$\hat{\alpha}$	88.682	(86.3004, 91.0640)
	$\hat{\beta}_1$	17.405	(14.8671, 19.9420)
	$\hat{\beta}_3$	−06.642	(−9.1792,−4.1044)
$Y_t = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}$	$\hat{\alpha}$	88.682	(86.3090, 91.0554)
	$\hat{\beta}_1$	18.284	(15.5983, 20.9694)
	$\hat{\beta}_2$	2.481	(−0.0743, 5.0366)
	$\hat{\beta}_3$	−06.555	(−9.0852,−4.0256)

Tabla 4.4: Estimadores de los parámetros de regresión e intervalos de 95 % de confianza.

La Figura 4.9 contiene los gráficos de las pruebas empíricas empleadas para los residuales del modelo seleccionado, $Y_t \sim x_1 + x_3$, pues al concluir se utilizarán para compararlo contra el modelo ajustado bajo la teoría de los modelos dinámicos lineales, tema central de este trabajo.

La Tabla 4.5 muestra un resumen de los resultados de las pruebas empíricas que fueron

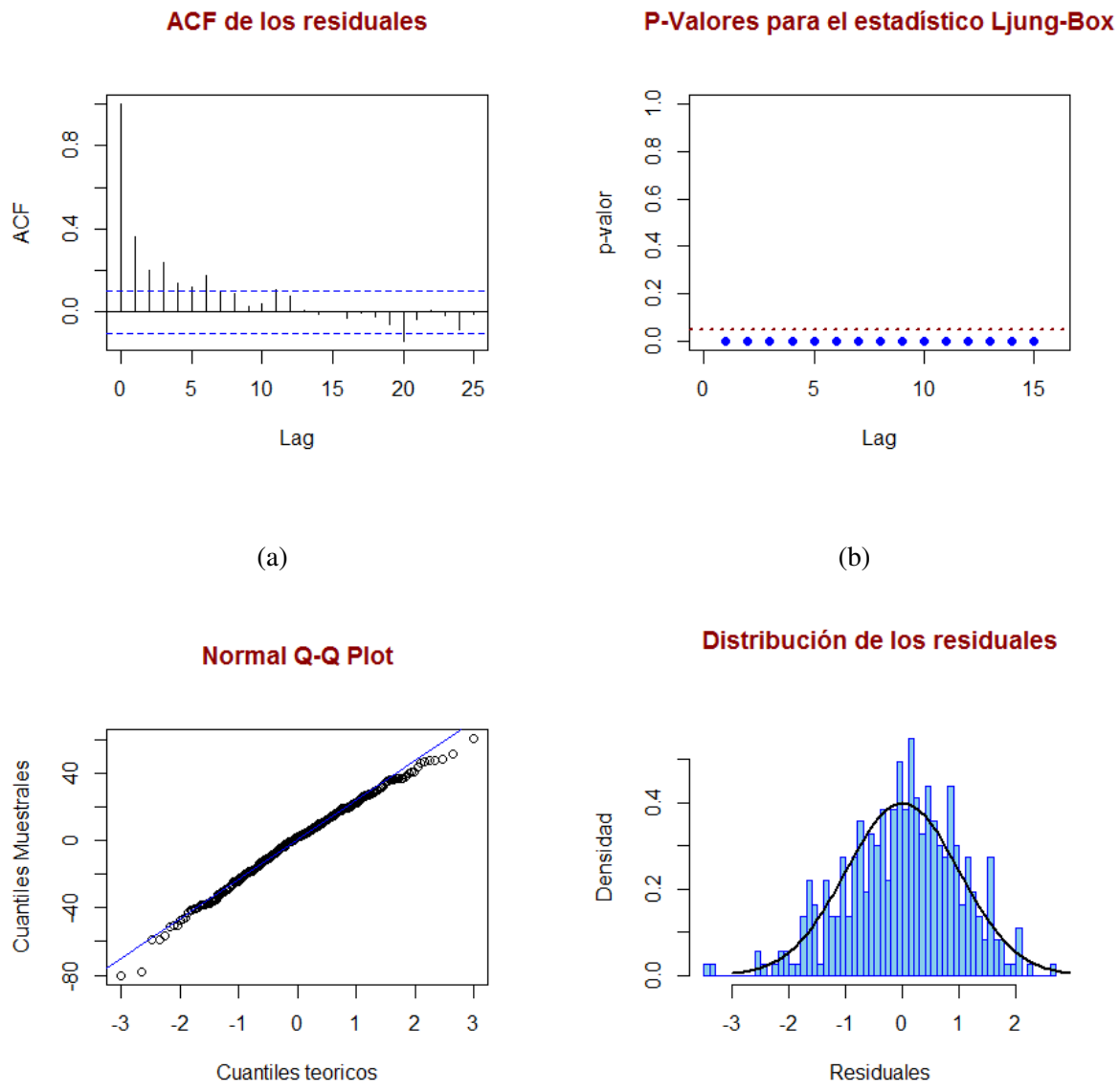


Figura 4.9: Diagnósticos para los residuales del modelo de regresión estático con dos covariables (temperatura y velocidad del viento).

empleadas, para verificar los supuestos de independencia y normalidad de residuales para cada uno de los tres mejores modelos con p predictores. Note que, bajo la prueba de Ljung Box, donde la hipótesis nula es H_0 : los datos se distribuyen de forma independiente, no hay evidencia suficiente para aceptar el supuesto de independencia de los residuales para ninguno de los tres modelos propuestos. Adicional a esto al aplicar las pruebas: Anderson Darling, donde la hipótesis

nula es H_0 : los datos provienen de una distribución normal; y la prueba Breush-Pagan, donde la hipótesis nula es H_0 : los datos son homocedasticos (varianza constante), podemos probar que el modelo que incluye solo la covariable temperatura no cumple los supuestos de normalidad ni varianza constante de los residuales por lo que puede ser descartado. Ya que el modelo completo parece ser tan competitivo como el modelo con dos covariables, elegimos este último pues se determinó que para el modelo con tres covariables la humedad relativa pareciera no ser significativa.

N° predictores en el modelo	Independencia	Normalidad	Homocedasticidad
1	Rechazar H_0	Rechazar H_0	Rechazar H_0
2	Rechazar H_0	Aceptar H_0	Aceptar H_0
3	Rechazar H_0	Aceptar H_0	Aceptar H_0

Tabla 4.5: Resumen de las pruebas empleadas para verificar los supuestos de independencia, normalidad y varianza constante (homocedasticidad) de los residuales.

4.4. Modelo de Regresión Dinámico Lineal

Si bien los modelos de regresión estáticos permiten modelar la relación entre las concentraciones de ozono en la estación el Pedregal y otras covariables (TMP, WSP y RH), se ven limitados por las dinámicas relacionadas con éstos. Como una alternativa, proponemos un modelo de regresión múltiple dinámico lineal (MDLM), pues en contraste con los primeros, que suponen una relación funcional constante entre las covariables y la variable respuesta, estos últimos permiten que los parámetros varíen en el tiempo.

Con base en los resultados obtenidos en la sección anterior, definimos nuestro modelo como sigue:

Sea $(Y_t)_{t\geq 0}$ la serie de concentraciones diarias de O_3, con $t = 1,...,365$ y $\theta_t = (\beta_{0,t}, \beta_{1,t}, \beta_{2,t})^T$ el vector de parámetros de regresión asociado, dónde $\beta_{0,t}$ es el parámetro de intercepción, $\beta_{1,t}$ el parámetro asociado a la covariable temperatura y $\beta_{2,t}$ el parámetro asociado a la covariable velocidad del viento. Entonces las ecuaciones observacional y del sistema están dadas por:

$$Y_t = F_t\theta_t + v_t, \qquad v_t \sim N_1(0,V), \tag{4.1}$$

$$\theta_t = G_t\theta_{t-1} + w_t, \qquad w_t \sim N_3(0,W), \tag{4.2}$$

$$\theta_0 \sim N_3(m_0, C_0).$$

Aquí, $F_t = (1, x_{1,t}, x_{2,t})$ un vector columna de variables regresoras de dimensiones 365×3, tal que $x_{0,t} = 1$ pues está asociado al parámetro de intercepción y $x_{i,t}$ con $i = 1,2$, son las covariables

temperatura y velocidad del viento al tiempo t, v_t y $w_t = \left(w_t^{(1)}, w_t^{(2)}, w_t^{(3)}\right)^T$, son las series de errores aleatorios de las ecuaciones observacional (4.1) y del sistema (4.2) y considere a V como la varianza observacional. Además como en la literatura Campagnoli et al. (2009), suponga que la matriz de covarianzas del sistema está dada por $W = diag\left(\sigma_0^2, \sigma_1^2, \sigma_2^2\right)$ y G_t como una matriz identidad de dimensiones 3×3.

Bajo esta Definición, los parámetros de regresión se actualizan continuamente con base en la historia de las variables de entrada (las concentraciones de ozono de la estación PED) y las de salida (la temperatura y la velocidad del viento). Además, la ecuación del sistema modela la forma en la que varían, es decir, como una caminata aleatoria. En otras palabras, los parámetros de regresión al tiempo t son iguales a ellos mismos al tiempo $t-1$, más el error de regresión.

Para que el modelo quede completamente específicado es necesario estimar la varianza observacional, V, y la matriz de covarianzas del sistema, W. En este sentido y para análisis posteriores del modelo, introduciremos la paquetería de modelos de espacio de estados autorregresivos multivariados (MARSS, por sus siglas en inglés) implementada en el software estadístico R. Dentro de sus múltiples aplicaciones es útil para la estimación de parámetros de modelos lineales MARSS con errores gaussianos. Algunos de sus campos de aplicación son: economía, ingeniería, genética, física y ecología. Según el campo de estudio suelen conocerse como modelos dinámicos lineales (MDL) o modelos de espacio de estados de vectores autorregresivos (VAR). De manera general este paquete permite ajustar modelos MARSS con variaciones temporales con o sin covariables valiéndose del algoritmo esperanza-maximización (EM, por sus siglas en inglés) (Casella and Berger, 2002).

Utilizando la paquetería MARSS podemos escribir nuestro modelo como:

$$\mathbf{x}_t = \mathbf{B}_t \mathbf{x}_{t-1} + \mathbf{u}_t + \mathbf{C}_t \mathbf{c}_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim N_1(0, \mathbf{Q}_t). \tag{4.3}$$

$$\mathbf{y}_t = \mathbf{Z}_t \mathbf{x}_{t-1} + \mathbf{a}_t + \mathbf{D}_t d_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim N_3(0, \mathbf{R}_t). \tag{4.4}$$

$$\mathbf{x}_0 \sim N_3(\mathbf{m}, \mathbf{C}),$$

donde $\mathbf{x}_t = \theta_t, \mathbf{B}_t = \mathbf{G}_t, \mathbf{u}_t = \mathbf{C}_t = \mathbf{c}_t = 0$, $\mathbf{Q}_t = V$, $\mathbf{y}_t = Y_t$, $\mathbf{Z}_t = \mathbf{F}_t, \mathbf{a}_t = \mathbf{D}_t = \mathbf{d}_t = 0$ y $\mathbf{R}_t = W$, tal que $\mathbf{R}_t$ es de dimensiones 1×1.

Para obtener los estimadores máximo verosímiles de la varianza observacional y la matriz de covarianzas de la ecuación del sistema, bajo el modelo descrito por las ecuaciones (4.1) y (4.2), utilizamos la función "MARSS", que recibe como parámetros: el vector de concentraciones de ozono de la estación PED, y_t, los parámetros iniciales del vector de estados, $\theta_0 = (0,0,0)$, y el modelo especificado en forma matricial. Así, $\hat{W}_t = \text{diag}(10.342, 6.970, 0.435)$ y $\hat{V}_t = 322.45$.

4.4.1. Estimación de Estados y Evaluación del Modelo

Uno de los objetivos centrales de este estudio es determinar si los modelos dinámicos lineales son adecuados para la predicción de observaciones futuras y contrastar los resultados obtenidos

con los modelos de regresión estáticos. En este sentido, compararemos los modelos descritos en las secciones previas bajo ambas metodologías, para predecir las concentraciones diarias de ozono en la estación el Pedregal, de este contaminante al día siguiente.

Como vimos en capítulos previos, la inferencia sobre el vector de estados no observable, θ_t, se obtiene al calcular la densidad condicional $p(\theta_s|y_{1:t})$, a través de dos procedimientos comúnmente utilizados: el filtro de Kalman, $s = t$, y el suavizamiento de Kalman, $s < t$. La Figura 4.10 muestra los estimadores máximo verosímiles del recorrido de los parámetros de regresión $\hat{\beta}_{0,t}, \hat{\beta}_{1,t}, \hat{\beta}_{2,t}$.

Por otro lado, para resolver el problema de predicción bajo el marco teórico de los MDLs, basta con estimar la distribución predictiva del vector de estados un paso hacia adelante, $p(\theta_t|y_{t-1})$. Así, una vez actualizados los parámetros de regresión podemos obtener la distribución predictiva para nuestras observaciones, $p(y_t|y_{t-1})$. Se puede observar que empleando la técnica de filtración obtenemos los valores predichos. La Figura 4.11, muestra la serie de las concentraciones diarias de ozono, la serie de los valores predichos e intervalos de 95 % de confianza para el modelo dinámico.

Para evaluar la precisión de ambos modelos en el pronóstico a un día de las concentraciones de ozono, utilizaremos tres medidas de error comúnmente empleadas en series de tiempo: el error absoluto promedio (MAE, por sus siglas en inglés), que se define por:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| = \frac{1}{n}\sum_{i=1}^{n}|e_i|,$$

donde n es el número de días, Y_i son las concentraciones diarias de ozono en la estación el Pedregal y $\hat{Y}_i$ son los valores predichos a un día de las Y_i. Mide el promedio de los errores de regresión absolutos; es decir, la distancia entre nuestras observaciones y los valores predichos.

El error cuadrático medio (MSE, por sus siglas en inglés), definido como:

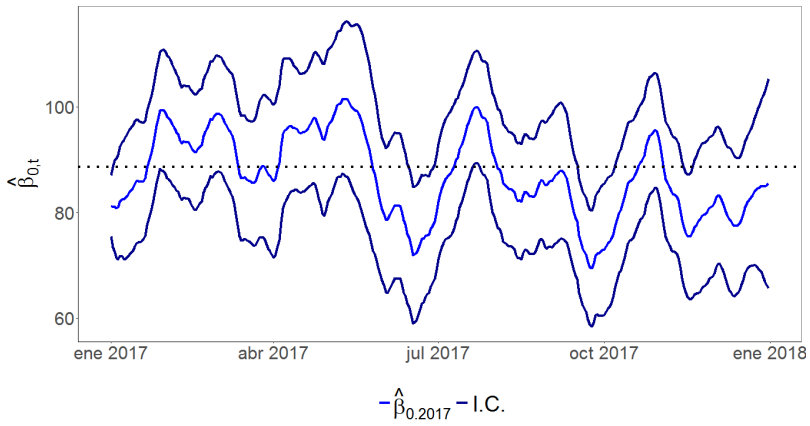
$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}e_i^2.$$

y el error absoluto porcentual promedio (MAPE, por sus siglas en inglés):

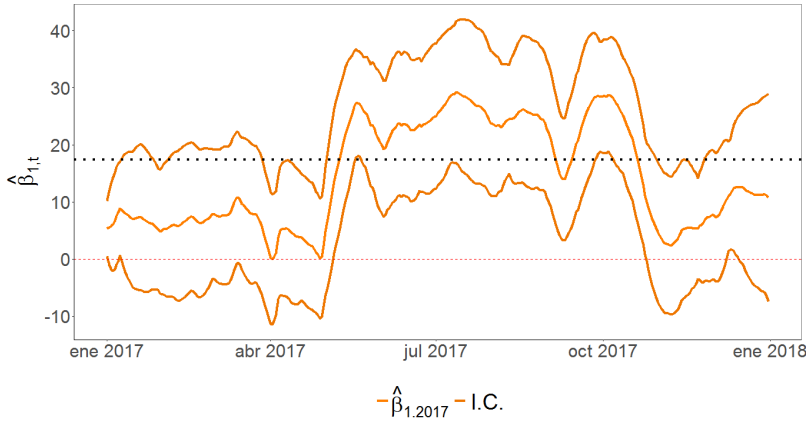
$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Y_i - \hat{Y}_i\right|}{|Y_i|} \times 100 = \frac{1}{n}\sum_{i=1}^{n}\frac{|e_i|}{|Y_i|} \times 100.$$

Esta última, a diferencia de las dos anteriores, mide el tamaño de los errores en términos de porcentaje, lo cual hace que sea más fácil de interpretar. Además, para las tres diremos que un modelo es "más adecuado", si las medidas de error son más cercanas a cero.

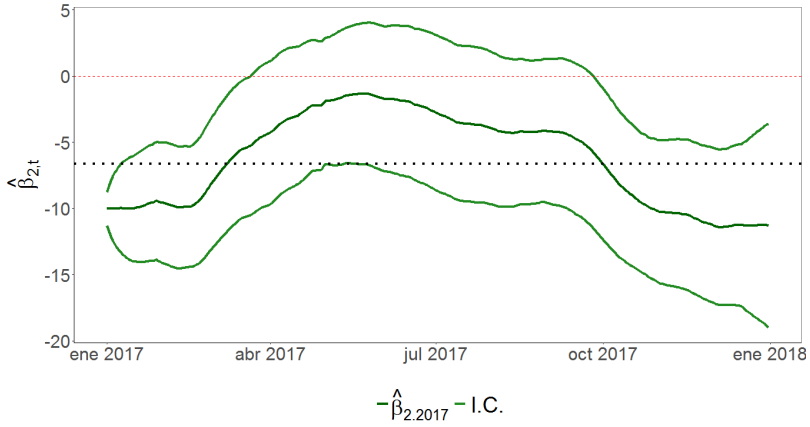
La Tabla 4.6 muestra los valores obtenidos para cada una de las medidas descritas, como se puede observar las tres medidas de error para el modelo ajustado bajo el marco teórico de los modelos dinámicos lineales son menores que aquellos que se obtuvieron para el modelo de regresión estático.



(a)



(b)



(c)

Figura 4.10: Valores suavizados e intervalos de 95 % de confianza para los estados del MDL. La línea negra punteada muestra el valor del coeficiente predicho bajo el modelo de regresión estático.

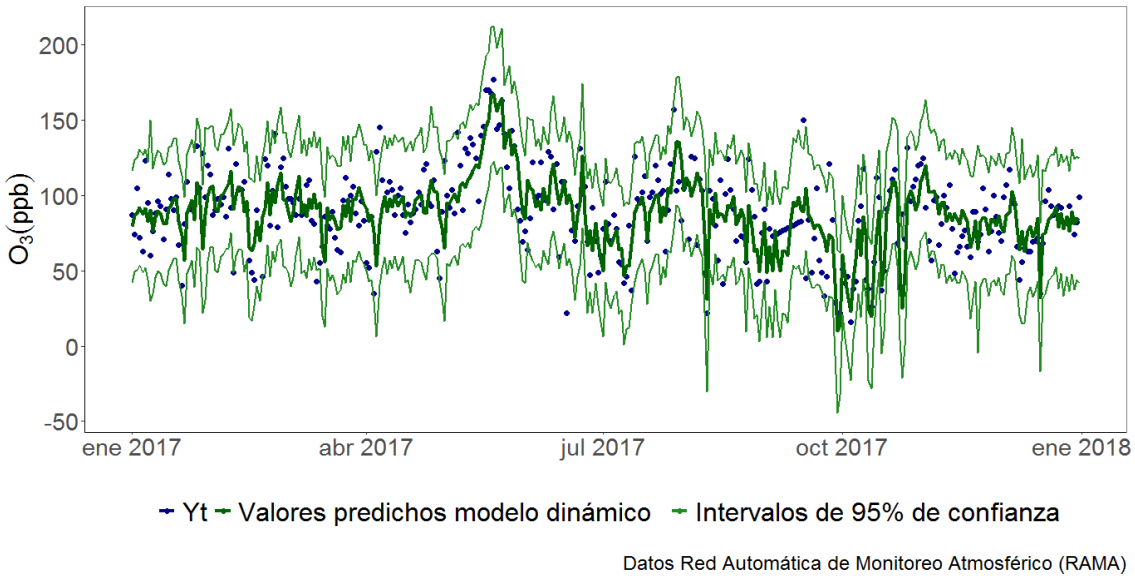


Figura 4.11: Serie de tiempo de las concentraciones diarias de ozono de la estación el Pedregal, valores predichos un paso hacia adelante, $f_t(1)$, e intervalos de 95 % de confianza.

Modelo	MSE	MAE	MAPE (%)
Regresión Dinámica Lineal	441.8437	16.1748	21.6017
Regresión Lineal Múltiple	531.0243	18.3004	26.8923

Tabla 4.6: Medidas para los errores de predicción para el modelo dinámico lineal y el modelo de regresión estático.

Adicional a éstas calcularemos la *cobertura* bajo ambos modelos, es decir, la proporción de observaciones que caen dentro de los intervalos de 95 % de confianza. Definimos dicha medida de confiabilidad como: $\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(L_i \leq Y_i \leq U_i)$, donde $\mathbb{I}()$ es la función indicatriz; es igual a uno si el valor predicho a un día cae dentro de los intervalos de confianza y cero en cualquier otro caso.

Note que aún cuando la proporción del número de observaciones que caen dentro de los intervalos de confianza es mayor para el modelo de regresión estático, ver Tabla 4.7, los valores resultantes por las medidas de los errores, nos dicen que el modelo de regresión dinámico lineal múltiple es más preciso. Lo anterior es consistente con los resultados de la Figura 4.12, pues para el MDLM las observaciones se encapsulan en intervalos de confianza más estrechos, mientras que los del estático son mucho más amplios.

Para concluir esta sección, en la Figura 4.13 se muestran los resultados de las pruebas empíricas más utilizadas en el análisis de series de tiempo para verificar el supuesto de normalidad de los residuales estandarizados o errores de predicción. En la Figura 4.13(a) se puede apreciar la función de autocorrelación (ACF); note que los valores de dicha función decaen a cero rápidamen-

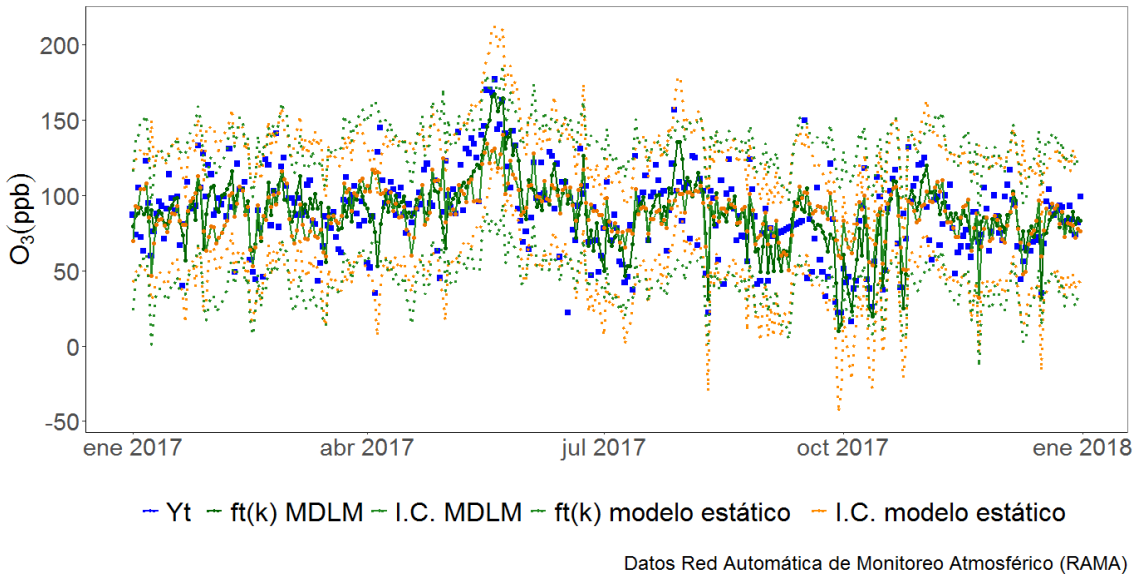


Figura 4.12: Serie de tiempo de las concentraciones diarias de ozono de la estación el Pedregal, valores estimados bajo el MDLM un paso hacia adelante, $f_t(1)$, y valores predichos bajo el modelo de regresión múltiple.

Modelo	N° obs	Proporción del n° obs
Regresión Dinámica Lineal	347	0.95068
Regresión Lineal Múltiple	349	0.95616

Tabla 4.7: Proporción del número de observaciones, Y_t, que caen dentro de los intervalos de 95 % de confianza para el modelo dinámico lineal y el modelo de regresión estático.

te, por lo que podríamos decir que el proceso de innovaciones se comporta como un ruido blanco, es decir que los valores parecen independientes e idénticamente distribuidos. Adicionalmente con la función *t.test( )* en R, podemos probar si la media de la distribución de las innovaciones es cero, obteniendo que el p- valor es 0.6553, por lo que no hay evidencia suficiente para rechazar que $E(e_t) = 0$. Con la prueba de Ljung Box 4.13(b), podemos corroborar el suspuesto de independencia, mientras que el qqplot e histograma de los residuales, Figura 4.13(c) y Figura 4.13(d) respectivamente, nos dan indicios de problemas en las colas de la distribución, pues se observan valores extremos. Utilizamos la prueba de Anderson-Darling para probar el supuesto de normalidad y obtenemos que el p- valor es 0.05898 por lo que no hay evidencia suficiente para rechazar este supuesto con un nivel de significancia del 5 %.

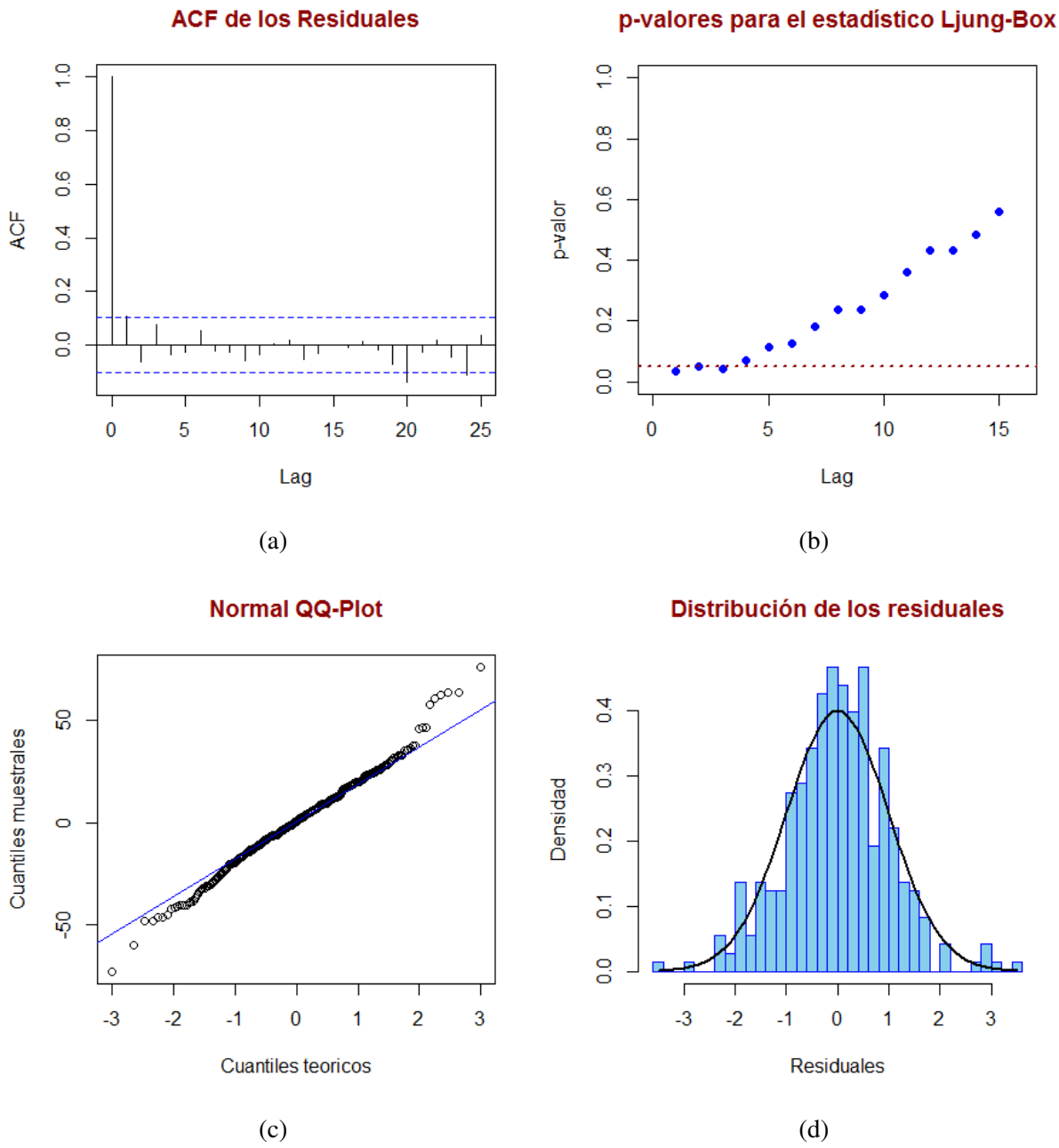


Figura 4.13: Diagnósticos para los residuales de la ecuación de observaciones del modelo dinámico lineal múltiple.

4.4.2. Pronóstico de Observaciones Futuras

En esta sección realizaremos la predicción de observaciones para un mes hacia adelante, bajo el MDL. Recordemos que a partir de la densidad condicional de un paso hacia adelante, al ir actualizando la información secuencialmente utilizando un algoritmo recursivo, podemos obtener la distribución predictiva $k = 31$ días hacia adelante de los estados, $p(\theta_{t+k}|y_{1:t})$, y posteriormente

la distribución predictiva de las concentraciones de ozono un mes hacia adelante, $p(y_{t+k}|y_{1:t})$.

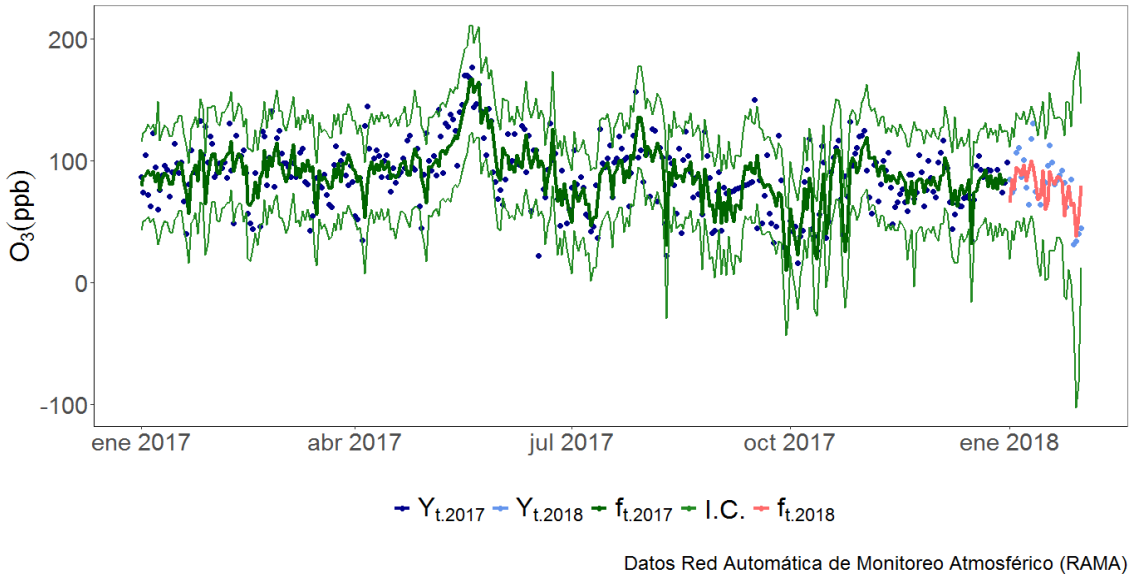
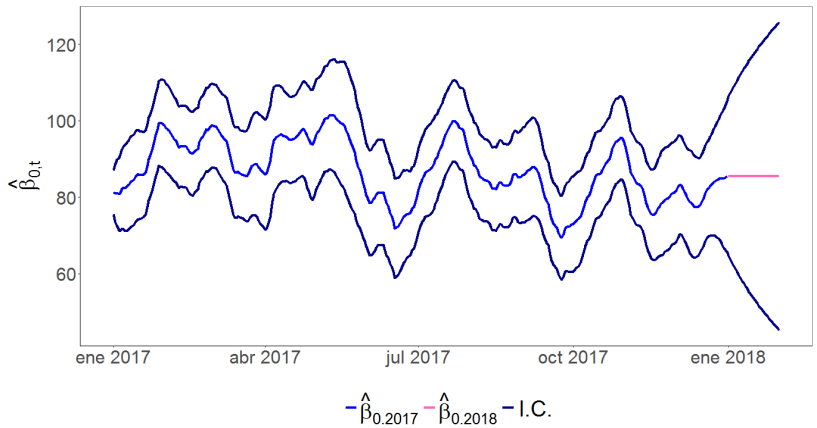


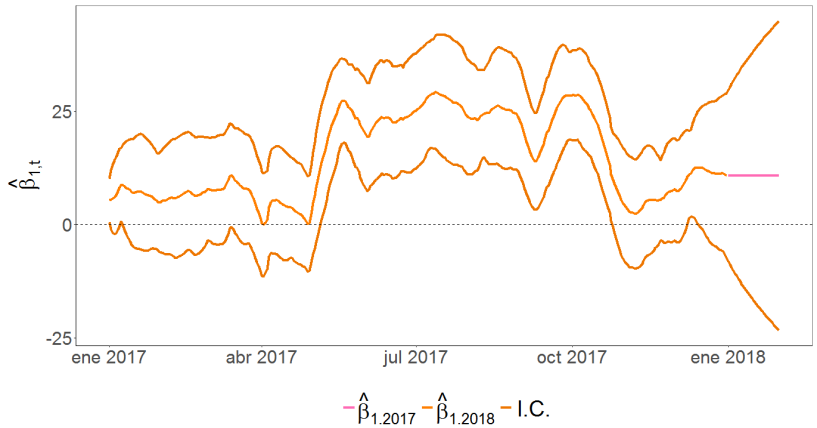
Figura 4.14: Serie de tiempo de las concentraciones diarias de ozono en la estación el Pedregal durante 2017 (puntos azul marino), serie de tiempo de las concentraciones diarias de ozono en la estación el Pedregal durante 2018 (puntos azul cielo), valores estimados bajo el MDLM un mes hacia adelante (línea roja) e intervalos al 95 % de confianza (líneas color verde).

Así, para obtener las estimaciones futuras del vector de observaciones, θ_t, en un horizonte de tiempo mayor, $k > 1$, en R, debemos conocer en primera instancia los valores de las covariables temperatura y velocidad del viento del 1° al 31 de Enero de 2018, es decir, $x_{1,t}, x_{2,t}$ tal que $t = 1, 2, ..., 31$, ver Figura 4.15. Dichos valores, al igual que las series de tiempo de las observaciones meteorológicas de 2017 empleadas a lo largo de este trabajo, están disponibles en la REDMET. Una vez que se conocen los valores de las covariables para el período de tiempo deseado basta con extender el vector de las concentraciones diarias de ozono, $y_1, y_2, ..., y_{365}$, añadiendo 31 valores faltantes y ajustando el modelo dinámico lineal bajo dichas consideraciones. Entonces, la secuencia de valores faltantes al final de la serie de observaciones dará lugar a los valores predichos un mes hacia adelante. La Figura 4.14 muestra los valores resultantes un mes hacia adelante e intervalos de 95 % de confianza para nuestras observaciones.

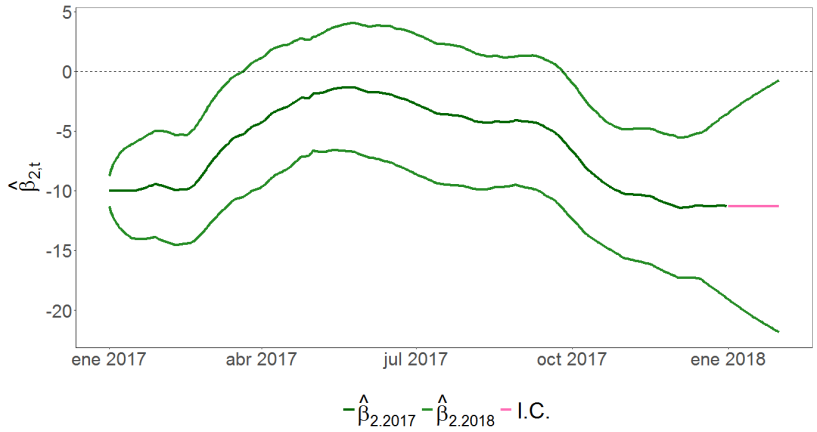
Al igual que en el pronóstico de observaciones a un día, la Tabla 4.8 muestra los valores resultantes, para las medidas de los errores de las innovaciones del modelo propuesto para el pronóstico de observaciones de un mes. Note que aún para horizontes de tiempo más largos el MDLM parece brindarnos mayor precisión en contraste con el modelo de regresión múltiple.



(a)



(b)



(c)

Figura 4.15: Valores predichos un mes hacia adelante para los estados del MDLM e intervalos de 95 % de confianza.

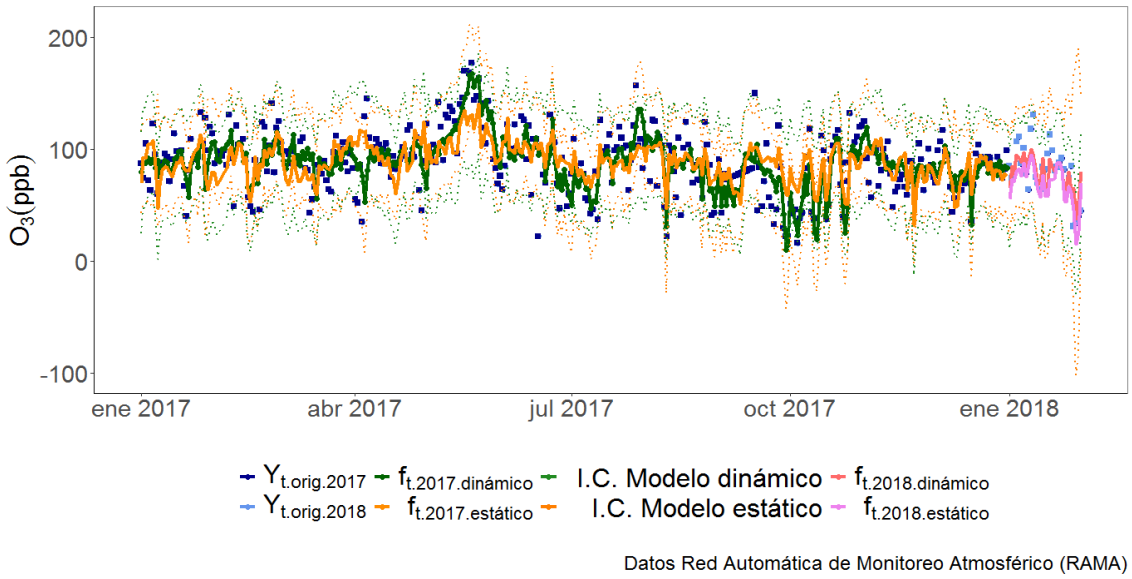


Figura 4.16: Serie de tiempo de las concentraciones diarias de ozono de la estación el Pedregal durante 2017 (puntos azul marino), concentraciones de diarias ozono de la estación el Pedregal durante 2018 (puntos azul cielo), valores estimados un mes hacia adelante bajo el MDLM (línea roja), intervalos de 95 % de confianza bajo el MDLM (líneas color verde), valores estimados un mes hacia adelante bajo el modelo de regresión múltiple (línea rosa) e intervalos de 95 % de confianza bajo el modelo de regresión estático (líneas color naranja).

Modelo	MSE	MAE	MAPE (%)
Regresión Dinámica Lineal	431.4538	15.9858	21.4595
Regresión Lineal Múltiple	522.1366	18.1755	26.4868

Tabla 4.8: Medidas para los errores de predicción del modelo dinámico lineal para observaciones futuras un mes hacia adelante.

Modelo	N° obs	Proporción del n° obs
Regresión Dinámica Lineal	377	0.9520
Regresión Lineal Múltiple	378	0.9545

Tabla 4.9: Proporción del número de observaciones, Y_t, que caen dentro de los intervalos de 95 % de confianza para el pronóstico a un mes, bajo los modelos dinámico lineal y modelo de regresión estático.

Lo anterior puede deberse a que, en contraste con el modelo de regresión estático, aún si permitimos que el período o intervalo de tiempo se actualice, las predicciones futuras estarán basadas en el mejor de los casos en datos de años previos, mientras que para un modelo de

regresión dinámico se basarán en toda la información disponible hasta el tiempo $t-1$.

Finalmente, la Tabla 4.9 muestra que la proporción de observaciones que caen dentro de los intervalos de confianza bajo los dos modelos propuestos en ambos casos es aproximadamente de 95 %, estando apenas por arriba aquella que se obtuvo bajo el modelo de regresión múltiple; sin embargo, como se pude apreciar en la Figura 4.16, los intervalos de confianza para este último son más amplios. Por lo tanto, con base en nuestros resultados, los modelos dinámicos lineales pueden funcionar como una herramienta alternativa para el análisis de series de tiempo donde la naturaleza de los datos tiende a verse influenciada por efectos en el tiempo que no pueden ser capturados bajo la estructura de los modelos de regresión estáticos.

Conclusiones

En este trabajo se exploraron los Modelos Dinámicos Lineales como una herramienta alternativa para el análisis de series de tiempo bajo el enfoque Bayesiano. Si bien éste no es tan conocido como el enfoque tradicional, los avances tecnológicos han impulsado su desarrollo, al permitir la creación de algoritmos cada vez más potentes y precisos. Tal es el caso del filtro de Kalman que nos provee de resultados comparables a los que obtendríamos con el enfoque clásico.

Aunque la teoría mostrada no fue más allá de los aspectos básicos que podemos encontrar en textos afines al tema, se espera que el lector encuentre de su interés dichos modelos, que por la forma en que se construyen (datos, observaciones y parámetros) tienen la característica de poder explicarse con relativa facilidad. Asimismo, considere que los algoritmos mostrados (filtración, suavizamiento y predicción) en conjunto con los modelos con tendencia, estacionales y con covariables, funcionan como la base para el entendimiento y desarrollo de otros con una complejidad mayor, véase Campagnoli et al. (2009) y West and Harrison (1997).

Con el conocimiento adquirido sobre los modelos bajo estudio, fue posible el análisis de observaciones relativas a las concentraciones de ozono, una de las emisiones con mayor impacto en la contaminación atmosférica en el valle de la Ciudad de México. El estudio de dicha base de datos fue una de las motivaciones que dio origen al estudio del tópico abordado y que se espera tratar como un problema espacio-temporal en trabajos futuros. En este sentido considere la propuesta que se muestra en Huerta et al. (2004), pues parece funcionar con este tipo de datos.

En particular, el problema del pronóstico de las concentraciones de ozono, visto a través de un modelo de regresión dinámico lineal nos permitió comprender las características básicas del marco de referencia bajo estudio, vimos que estos modelos son capaces de capturar la variación de los parámetros sin que estos pierdan interpretabilidad, además de que el pronóstico de datos se hará siempre considerando toda la información disponible hasta el tiempo anterior. Éstas son solo algunas de las ventajas que nos proveen en contraste con un modelo de regresión estático.

Desde sus orígenes la estadística ha dado respuesta a múltiples preguntas y formado parte de aplicaciones diversas. La que aquí se ha mostrado es solo una pieza del rompecabezas que forma parte de la problemática ambiental a la que nos enfrentamos diariamente en todo el mundo. De ahí que estadísticos y científicos de áreas relacionadas, deben colaborar para el desarrollo de planes de acción para combatirla.

Apéndices

A. Manipulación de la Base de Datos

```
# Transformación y manipulación de la base de datos Ozono
# Troposférico (O3)

# Librerías
suppressMessages(suppressWarnings(library(plyr)))
suppressMessages(suppressWarnings(library(dplyr)))
suppressMessages(suppressWarnings(library(readxl)))
suppressMessages(suppressWarnings(library(lubridate)))
suppressMessages(suppressWarnings(library(ggplot2)))
suppressMessages(suppressWarnings(library(reshape2)))

setwd("C:\\Users\\DELL\\Desktop\\TESIS_NMR\\Insumos Tesis NMR\\Datos
    Crudos v1")

# Cargamos las bases de datos del contaminante O3
# Observaciones del año 2017
O3 <- read_excel("201703.xls")
# Observaciones del año 2016
O3_2016 <- read_excel("201603.xls")

# Se tienen 8760 observaciones de los niveles de ozono
# estratosférico de 37 estaciones y las variables: hora,
# fecha y mes, en que fueron tomadas las mediciones.

etiquetas <- function(base) {

  # Formato días y meses
  dia <- format(base$FECHA, format="%a");
  mes <- month(as.POSIXlt(base$FECHA, format="%d/%m/%Y"))
  base <- data.frame(dia, mes, base)

  # Etiquetas meses
  for(i in 1:length(base$mes)){
    base$mes[i][which(base$mes[i]==1)] <- "ENE"
    base$mes[i][which(base$mes[i]==2)] <- "FEB"
    base$mes[i][which(base$mes[i]==3)] <- "MAR"
    base$mes[i][which(base$mes[i]==4)] <- "ABR"
```

```
37      base$mes[i][which(base$mes[i]==5)] <- "MAY"
38      base$mes[i][which(base$mes[i]==6)] <- "JUN"
39      base$mes[i][which(base$mes[i]==7)] <- "JUL"
40      base$mes[i][which(base$mes[i]==8)] <- "AGO"
41      base$mes[i][which(base$mes[i]==9)] <- "SEP"
42      base$mes[i][which(base$mes[i]==10)] <- "OCT"
43      base$mes[i][which(base$mes[i]==11)] <- "NOV"
44      base$mes[i][which(base$mes[i]==12)] <- "DIC"
45    }
46
47    # Transformando variables
48    base$FECHA <- as.Date(base$FECHA)
49    base$HORA <- as.integer(base$HORA)
50    base$mes <- as.factor(base$mes)
51    base$mes <- factor(base$mes, levels=c("ENE", "FEB", "MAR",
52                                          "ABR", "MAY", "JUN",
53                                          "JUL", "AGO", "SEP",
54                                          "OCT", "NOV", "DIC"))
55    base$dia <- as.factor(base$dia)
56    base$dia <- factor(base$dia, levels=c("lun", "mar", "mié",
57                                          "jue", "vie", "sáb",
58                                          "dom"))
59
60    # Se modifica el valor de los valores faltantes de la
61    # base (-99) por NA's.
62
63    for(i in 4:dim(base)[2]){
64      base[,i][which(base[,i]==-99)]<-NA
65    }
66     return(base)
67  }
68
69  O3 <- etiquetas(O3)
70  # Estructura de la base de datos
71  str(O3)
72  summary(O3[,-c(1:4)])
73
74  # Se aplica la función "melt( )" de la librería "reshape2"
75  # del software estadístico R a nuestra base de datos para
76  # tener una observación por cada variable y poder manipularla
77  # con mayor facilidad.
78
79  melt_O3 = melt(O3, id = c("dia","mes","FECHA", "HORA"))
80  colnames(melt_O3) <- c("dia","mes", "fecha", "hora",
81                         "estacion", "concentracionO3" )
82  str(melt_O3)
83
84  # Frecuencia de observaciones que sobrepasan el valor permisible por hora
85
86  melt_O3 <- data.frame(cont=rep(1, dim(melt_O3)[2]), melt_O3)
87  mayor_95ppb <- melt_O3%>%
88    group_by(cont, estacion)%>%
89    filter(concentracionO3>95)%>%
```

```
90    summarize(mayores=sum(cont))
91
92  # Concentraciones Diarias: Máximos de las concentraciones horarias
93
94  melt_O3_max <- melt_O3%>%
95    group_by(cont, fecha, dia, mes, estacion)%>%
96    summarize(maximos=max(concentracionO3, na.rm = T))
97
98  melt_O3_max <- arrange(melt_O3_max, melt_O3_max$estacion)
99
100 #write.csv(melt_O3_max, file = "C:\\Users\\DELL\\Desktop\\TESIS_NMR\\
      Insumos Tesis NMR\\Datos Transformados v2\\O3_max_diarios_ma.csv")
101
102 # ------------------------------------------------
103
104 # Gráficos Estaciones de Monitoreo
105 # Unidad de medición de las observaciones: Promedio de la concentraciones
      del contaminante
106 # O3 registradas en una hora (concentración horaria), cuya unidad de
      medición se denomina como
107 # partes por billón.
108
109 theme_set(theme_classic())
110 theme_update(plot.title = element_text(hjust = 0.5, size = 24, face = "
      bold"),
111              plot.subtitle=element_text(size=20, hjust=0.5),
112              plot.caption = element_text(size = 18),
113              axis.title=element_text(size=24),
114              axis.text = element_text(size=22),
115              legend.text=element_text(size=24),
116              legend.position="bottom",
117              legend.box = "horizontal")#centrar el titulo en las gráficas
118
119
120 # Histograma de la distribución del promedio de las concentraciones de
      ozono registradas en una hr.
121
122 ggplot(melt_O3, aes(x=concentracionO3)) + geom_histogram(col="lightskyblue
      ",fill="mediumblue", binwidth=2) +
123   labs(title=" ", caption="Datos Red Automática de Monitoreo Atmosférico (
        RAMA)",
124        x=expression(O[3](ppb)), y=expression(Frecuencia)) + theme(panel.
             background = element_rect(fill = "white", colour = "grey50")) +
125   scale_x_continuous(limits=c(0, 180)) +
126   geom_vline(aes(xintercept = mean(concentracionO3, na.rm = T), color="
        media"),size=1) +
127   geom_vline(aes(xintercept = median(concentracionO3, na.rm = T), color="
        mediana"),size=1) +
128   scale_color_manual(name = "", values = c(media = "orange", mediana = "
        red"))
129
130 summary(melt_O3$concentracionO3)
131
```

```
132 # Boxplot concentración horaria de ozono por estación de monitoreo
133 ggplot(melt_O3, aes(x=estacion, y=concentracionO3)) + geom_boxplot(col="
        mediumblue") +
134   labs(title=" ", y=expression(O[3](ppb)), x="Estación", caption="Datos
        Red Automática de Monitoreo Atmosférico (RAMA)") +
135   stat_summary(fun.y=mean, geom="point", shape=20, size=3, color="red",
        fill="red")+
136   stat_summary(fun.y=median, geom="point", shape=20, size=3, color="blue",
         fill="blue") + theme(panel.background = element_rect(fill = "white",
         colour = "grey50")) +
137   theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
138   geom_hline(yintercept = 95, linetype="dashed",
139              color = "darkred", size=0.8)
140
141 # Frecuencia del número de horas en las que se superó el valor 95ppb
142 mayor_95ppb <- melt_O3%>%
143   filter(concentracionO3>95)
144
145 by_hora_mayor_95ppb <- as.data.frame(table(mayor_95ppb$hora))
146 by_dia_mayor_95ppb <- as.data.frame(table(mayor_95ppb$dia))
147 by_mes_mayor_95ppb <- as.data.frame(table(mayor_95ppb$mes))
148 by_estacion_mayor_95ppb <- as.data.frame(table(mayor_95ppb$estacion))
149
150 ggplot(by_estacion_mayor_95ppb, aes(x=Var1, y=Freq)) + geom_col(col="
        lightskyblue",fill="mediumblue") +
151   labs(title=" ", x="Estación", y="Frecuencia (horas)", caption="Datos Red
         Automática de Monitoreo Atmosférico (RAMA)") +
152   guides(fill=guide_legend(title="")) + theme(panel.background = element_
        rect(fill = "white", colour = "grey50")) +
153   theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
154   geom_hline(yintercept = max(by_estacion_mayor_95ppb$Freq), linetype="
        dashed",
155              color = "darkred", size=0.8)
156
157 # Boxplot de las concentraciones de ozono como promedio horario por mes de
        monitoreo
158
159 melt_O3$mes <- as.factor(melt_O3$mes)
160 melt_O3$mes <- factor(melt_O3$mes, levels=c("ENE", "FEB", "MAR",
161                                             "ABR", "MAY", "JUN",
162                                             "JUL", "AGO", "SEP",
163                                             "OCT", "NOV", "DIC"))
164
165 ggplot(melt_O3, aes(x=mes, y=concentracionO3)) + geom_boxplot(col="
        mediumblue", fill="lightskyblue") +
166   labs(title=" ", subtitle=" ",
167        caption="\nDatos Red Automática de Monitoreo Atmosférico (RAMA)", x
             ="Mes", y=expression(O[3](ppb))) +
168   theme(panel.background = element_rect(fill = "white", colour = "grey50")
        )+
169   stat_summary(fun.y=mean, geom="point", shape=20, size=3, color="red",
        fill="red")+
```

```
170     stat_summary(fun.y=median, geom="point", shape=20, size=3, color="blue",
            fill="blue") + theme(panel.background = element_rect(fill = "white",
            colour = "grey50")) +
171     theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

B. Modelo Polinomial de Primer Orden

```
1  # Modelo Polinomial
2
3  # Librerías
4  suppressMessages(suppressWarnings(library(MARSS)))
5  suppressMessages(suppressWarnings(library(imputeTS)))
6  suppressMessages(suppressWarnings(library(ggplot2)))
7  suppressMessages(suppressWarnings(library(forecast)))
8  suppressMessages(suppressWarnings(library(GGally)))
9  suppressMessages(suppressWarnings(library(changepoint)))
10 suppressMessages(suppressWarnings(library(strucchange)))
11 suppressMessages(suppressWarnings(library(ggpmisc)))
12 suppressMessages(suppressWarnings(library(ggfortify)))
13 suppressMessages(suppressWarnings(library(lubridate)))
14 suppressMessages(suppressWarnings(library(readxl)))
15 suppressMessages(suppressWarnings(library(nortest)))
16
17 setwd("C:\\Users\\DELL\\Desktop\\TESIS_NMR\\Insumos Tesis NMR\\Datos
       Finales v3")
18
19 base <- read.csv("Datos_v3.csv")
20 attach(base)
21 fecha <- as.Date(as.character(fecha), "%d/%m/%Y")
22
23 # Diseño Gráficas
24 theme_set(theme_classic())
25 theme_update(plot.title = element_text(hjust = 0.5, size = 24, face = "
       bold"),
26              plot.subtitle=element_text(size=20, hjust=0.5),
27              plot.caption = element_text(size = 18),
28              axis.title=element_text(size=24),
29              axis.text = element_text(size=22),
30              legend.text=element_text(size=24),
31              legend.position="bottom",
32              legend.box = "horizontal")#centrar el titulo en las gráficas
33
34 # Número de días de los datos
35 TT = length(fecha)
36
37 # Variable respuesta Yt
38 Yt = matrix(Yt.s,nrow=1)
39
40 #--- Modelo de Polionomial de Orden 1 ---
```

```
41
42 # Definición de los parámetros de la ecuación del sistema
43
44 B = matrix(1)               ## Gt es igual a la constante 1
45 U = matrix(0)               ## escalar 0; 1x1
46 Q = matrix(list("Q"),1,1)   ## Matriz de varianzas del sistema;
47
48 # Definición de los parámetros de la ecuación de observaciones
49 Z = matrix(1)               ## Z=Ft'=1; escalar 1
50 A = matrix(0)               ## escalar 0; 1x1
51 R = matrix("r")             ## escalar r; 1x1;
52
53 # Valores iniciales
54 inits.list = list(x0=matrix(c(0), nrow=1))
55 # Lista de los parámetros que definen el modelo
56 mod.list = list(B=B, U=U, Q=Q, Z=Z, A=A, R=R)
57 # Ajuste del modelo polinomial de primer grado
58 poly = MARSS(Yt, inits=inits.list, model=mod.list)
59
60 cat("Estimador máximo verosimil de W:", poly$par$Q,
61     "\nEstimador máximo versosimil de V:", poly$par$R,
62     "\nSignal to noise ratio:", poly$par$Q/poly$par$R)
63
64 #--- Predicción del modelo polinimial de grado 1 ---
65
66 # Salida del filtro de Kalman
67 kf.out = MARSSkfss(poly)
68 # Prónostico a un paso
69 eta = kf.out$xtt1
70 # Media de la distribución predictiva un paso adelante para los estados
71 # at=E(thetha_1:t-1|Yt-1)=Gt*mt-1=mt-1
72 fore.mean = vector()
73 for(t in 1:TT) {
74   fore.mean[t] = eta[,t,drop=F]
75 }
76
77
78 # Varianza un paso adelante de los parámetros de regresión al tiempo t; 1
     x2xT
79 Phi = kf.out$Vtt1
80 # Varianza de la ec. de observación; matriz 1x1
81 R.est = coef(poly, type="matrix")$R
82 # Varianza de la distribución predictiva
83 # Rt=Var(thetha_1:t-1|Yt-1)=Gt*Ct-1*Gt'+Wt=Ct-1+Wt
84 fore.var = vector()
85 for(t in 1:TT) {
86   fore.var[t] = Phi[,,t]+ R.est
87 }
88
89 fk_poly = fore.mean
90 fk_lower = fore.mean-2*sqrt(fore.var)
91 fk_upper = fore.mean+2*sqrt(fore.var)
```

```
92  forecast_poly <- data.frame(fecha=fecha, Obs=Yt.s, fk1=fore.mean, low=fk_
        lower, upp=fk_upper)
93  names(forecast_poly) <- c("fecha", "Yt.s", "fkpoly", "lower", "upper")
94
95  ggplot(forecast_poly, aes(x=fecha)) +
96    geom_line(aes(y=Yt.s, colour="darkblue"), size=1.3) +
97    #geom_point(aes(y=Yt.s, colour="red"), size=1.3) +
98    geom_line(aes(y=fk_poly, colour="cyan"), size=1.3) +
99    geom_line(aes(y=lower, colour="cyan4"), size=1, linetype=3) +
100   geom_line(aes(y=upper, colour="cyan4"), size=1, linetype=3) +
101   scale_colour_manual("", values = c("cyan", "cyan4", "darkblue", "cyan4")
          ,
102                       labels = c("Valores Predichos Modelo Polinomial, n=1
                              ",
103                                  "Intervalos al 95% de confianza", "Yt"))
                                    +
104   labs(title=" ", x=" ", y=expression(O[3](ppb)),
105        caption="\nDatos Red Automática de Monitoreo Atmosférico (RAMA)") +
106   theme(panel.background = element_rect(fill = "white", colour = "grey50")
          ) +
107   scale_y_continuous(breaks=seq(-50, 250, 50)) #Marcas del -50 al 250,
        cada 50.
108
109
110 #--- Modelo Polinomial 2 ---
111
112 poly2 <- poly
113 poly2$par$Q <- 462
114
115 #--- Predicción del modelo polinomial de grado 1, Wt=462 ---
116
117 # Salida del filtro de Kalman
118 kf.out2 = MARSSkfss(poly2)
119 # Prónostico a un paso
120 eta2 = kf.out2$xtt1
121 # Media de la distribución predictiva un paso adelante para los estados
122 # at=E(thetha_1:t-1|Yt-1)=Gt*mt-1=mt-1
123 fore.mean2 = vector()
124 for(t in 1:TT) {
125   fore.mean2[t] = eta2[,t,drop=F]
126 }
127
128
129 # Varianza un paso adelante de los parámetros de regresión al tiempo t; 1
      x2xT
130 Phi2 = kf.out2$Vtt1
131 # Varianza de la ec. de observación; matriz 1x1
132 R.est2 = coef(poly2, type="matrix")$R
133 # Varianza de la distribución predictiva
134 # Rt=Var(thetha_1:t-1|Yt-1)=Gt*Ct-1*Gt'+Wt=Ct-1+Wt
135 fore.var2 = vector()
136 for(t in 1:TT) {
137   fore.var2[t] = Phi2[,,t]+ R.est2
```

```
}

fk_poly2 = fore.mean2
fk_lower2 = fore.mean2-2*sqrt(fore.var2)
fk_upper2 = fore.mean2+2*sqrt(fore.var2)
forecast_poly2 <- data.frame(fecha=fecha, Obs=Yt.s, fk1=fore.mean2, low=fk_lower2, upp=fk_upper2)
names(forecast_poly2) <- c("fecha", "Yt.s", "fkpoly2", "lower2", "upper2")

ggplot(forecast_poly2, aes(x=fecha)) +
  geom_line(aes(y=Yt.s, colour="darkblue"), size=1.3) +
  #geom_point(aes(y=Yt.s, colour="red"), size=1.3) +
  geom_line(aes(y=fk_poly2, colour="palevioletred1"), size=1.3) +
  geom_line(aes(y=lower2, colour="palevioletred3"), size=1, linetype=3) +
  geom_line(aes(y=upper2, colour="palevioletred3"), size=1, linetype=3) +
  scale_colour_manual("", values = c( "darkblue","palevioletred1", "palevioletred3", "palevioletred3"),
                      labels = c("Yt",
                                 "Valores Predichos Modelo Polinomial, n=1",
                                 "Intervalos de 95% de confianza" )) +
  labs(title=" ", x=" ", y=expression(O[3](ppb)),
       caption="\nDatos Red Automática de Monitoreo Atmosférico (RAMA)") +
  theme(panel.background = element_rect(fill = "white", colour = "grey50")) +
 scale_y_continuous(breaks=seq(0, 200, 50)) #Marcas del 0 al 70, cada 10.

#--- Diagnósticos para los residuales del modelo 1 ---

# Errores de prediccion o innovaciones del modelo
innov1 = kf.out$Innov

# ACF de los residuales
acf(t(innov1), main="")
title(main="ACF de los Residuales", col.main="darkred")

# Ho: La media de la distribución de las innovaciones es igual a cero
# H1: La media de la distribución de las innovaciones es distinta de cero

t.test(t(innov1))

# Ljung Box
Box.test(t(innov1), lag=20, type = "Ljung")
p_values_LB1 <- sapply(1 : 15, function(i)
  Box.test(t(innov1), lag=i, type="Ljung-Box")$p.value)

plot(p_values_LB1, type="p",  pch=19, ylim=c(0,1), xlim=c(0,16), lwd=1, col="blue", main=" ", xlab="Lag", ylab="p-valor")
abline(h=0.05, col="darkred", lty=3, lwd=2)
title(main="p-valores para el estadístico Ljung-Box", col.main="darkred")

# Histograma de los Residuales
x=(-35:35)/10
```

```
hist((t(innov1)-mean(t(innov1)))/sd(t(innov1)),nclass=50,freq=F,
     border="blue", col="skyblue", xlab="Residuales",
     ylab="Densidad", main=" ")
title(main="Distribución de los residuales", col.main="darkred")
lines(x,dnorm(x), lwd=2)

# Q-Q plot de los residuales
qqnorm(t(innov1), main=" ", ylab="Cuantiles muestrales", xlab="Cuantiles
    teoricos")
qqline(t(innov1), col="blue", lwd=1.8, lty=1.3)
title(main="Normal QQ-Plot", col.main="darkred")

# Prueba de Normalidad Anderson Darling
ad.test(t(innov1))
t.test(t(innov1))

#--- Diagnósticos para los residuales del modelo 2 ---

# Errores de prediccion o innovaciones del modelo
innov2 = kf.out2$Innov

# ACF de los residuales
acf(t(innov2), main="")
title(main="ACF de los Residuales", col.main="darkred")

# Ho: La media de la distribución de las innovaciones es igual a cero
# H1: La media de la distribución de las innovaciones es distinta de cero

t.test(t(innov2))

# Ljung Box
Box.test(t(innov2), lag=20, type = "Ljung")
p_values_LB2 <- sapply(1 : 15, function(i)
  Box.test(t(innov2), lag=i, type="Ljung-Box")$p.value)

plot(p_values_LB2, type="p",  pch=19, ylim=c(0,1), xlim=c(0,16), lwd=1,
    col="blue", main=" ", xlab="Lag", ylab="p-valor")
abline(h=0.05, col="darkred", lty=3, lwd=2)
title(main="p-valores para el estadístico Ljung-Box", col.main="darkred")

# Histograma de los Residuales
x=(-35:35)/10
hist((t(innov2)-mean(t(innov2)))/sd(t(innov2)),nclass=50,freq=F,
     border="blue", col="skyblue", xlab="Residuales",
     ylab="Densidad", main=" ")
title(main="Distribución de los residuales", col.main="darkred")
lines(x,dnorm(x), lwd=2)

# Q-Q plot de los residuales
qqnorm(t(innov2), main=" ", ylab="Cuantiles muestrales", xlab="Cuantiles
    teoricos")
qqline(t(innov2), col="blue", lwd=1.8, lty=1.3)
title(main="Normal QQ-Plot", col.main="darkred")
```

```
236
237 # Prueba de Normalidad Anderson Darling
238 ad.test(t(innov2))
239 t.test(t(innov2))
```

C. Modelo de Regresión Lineal Estático

```
1
2  # Regresión estática y selección de variables
3
4  # Librerías
5     suppressMessages(suppressWarnings(library(imputeTS)))
6     suppressMessages(suppressWarnings(library(ggplot2)))
7     suppressMessages(suppressWarnings(library(forecast)))
8     suppressMessages(suppressWarnings(library(GGally)))
9     suppressMessages(suppressWarnings(library(changepoint)))
10    suppressMessages(suppressWarnings(library(strucchange)))
11    suppressMessages(suppressWarnings(library(ggpmisc)))
12    suppressMessages(suppressWarnings(library(ggfortify)))
13    suppressMessages(suppressWarnings(library(lubridate)))
14    suppressMessages(suppressWarnings(library(nortest)))
15    suppressMessages(suppressWarnings(library(lmtest)))
16 # Selección de variables
17    suppressMessages(suppressWarnings(library(leaps)))
18    suppressMessages(suppressWarnings(library(MASS)))
19    suppressMessages(suppressWarnings(library(car)))
20    suppressMessages(suppressWarnings(library(ggvis)))
21
22    #setwd("C:/Users/SIEC/Documents/TESIS/R/Insumos Tesis NMR/Datos
          Transformados v2")
23    setwd("C:\\Users\\DELL\\Desktop\\TESIS_NMR\\Insumos Tesis NMR\\Datos
          Transformados v2")
24    getwd()
25
26 # Base de las concentraciones de ozono máximas del promedio
27 # móvil de 8 horas y las covariables: TMP, RH y WSP.
28
29    base <- read.csv("O3_TMP_PED.csv")
30    fecha <- as.Date(as.character(base$fecha), "%d/%m/%Y")
31    attach(base)
32
33    str(base)
34    summary(base)
35
36    o3 = O3_max
37    tmp = TMP_max
38    rh = RH_max
39    wsp = WSP_max
40
```

```
41 # Varianzas
42   sapply(base[2:5], function(x){var(x,na.rm = T)})
43
44 # Proporción valores faltantes
45   sapply(base[2:5], function(x){length(which(is.na(x)))})
46   sapply(base[2:5], function(x){length(which(is.na(x)))/length(x)})
47
48   theme_set(theme_classic())
49   theme_update(plot.title=element_text(hjust = 0.5,
50               size = 24, face="bold"),
51               plot.subtitle=element_text(size=19, hjust=0.5),
52               plot.caption=element_text(size = 22),
53               axis.title=element_text(size=22),
54               axis.text=element_text(size=18),
55               legend.text=element_text(size=18),
56               legend.position="bottom",
57               legend.box="horizontal")
58
59   # Scatterplot
60   ggpairs(data.frame(o3, tmp, rh, wsp),
61         columnLabels = c("O3", "TMP", "RH", "WSP"),
62         axisLabels = "show")
63
64 # Imputar valores faltantes
65   Yt = na.kalman(base$O3_max)
66   xt = na.kalman(base$TMP_max)
67   x2t = na.kalman(base$RH_max)
68   x3t = na.kalman(base$WSP_max)
69
70   base.sin.na <- cbind(Yt, xt, x2t, x3t)
71
72 # Conversión de cada una de las variables en objetos de series de tiempo
73   Yt.s <- ts(Yt, start = c(2017,1), frequency = 365)#concentraciones de O3
74   xt.s <- ts(xt, start = c(2017,1), frequency = 365)#covariable
       temperatura
75   x2t.s <- ts(x2t, start = c(2017,1), frequency = 365)#covariable humedad
       relativa
76   x3t.s <- ts(x3t, start = c(2017,1), frequency = 365)#covariable WSP
77
78 # Estandarizar covariables
79   xt.s <- ts((xt - mean(xt))/sqrt(var(xt)),
80               start = c(2017,1), frequency = 365)
81   x2t.s <- ts((x2t - mean(x2t))/sqrt(var(x2t)),
82                start = c(2017,1), frequency = 365)
83   x3t.s <- ts((x3t - mean(x3t))/sqrt(var(x3t)),
84                start = c(2017,1), frequency = 365)
85
86   fecha <- as.character(base$fecha)
87   base.s <- as.data.frame(cbind(Yt.s, xt.s, x2t.s, x3t.s))
88   base.s <- cbind(fecha, base.s)
89   summary(base.s)
90
91   # Base de datos a emplear para el modelo predictivo
```

```
92    write.csv(base.s, file="C:\\Users\\DELL\\Desktop\\TESIS_NMR\\Insumos
          Tesis NMR\\Datos Finales v3\\Datos_v3.csv")
93
94  # Serie de tiempo O3
95    ggplot(Yt.s, as.numeric = FALSE) +
96    geom_line(col="blue", size=1.3) +
97    #  geom_point(col="red", size=1.8) +
98    labs(title="Concentraciones de Ozono Máximas Diarias del Promedio Móvil
          de 8 hrs",
99           subtitle="Estación Pedregal (PED), Delegación Alvaro Obregón, CDMX
                 (2017)\n",
100          x="Tiempo", y="Yt",
101          caption="\nDatos Red Automática de Monitoreo Atmosférico (RAMA)")
                 +
102   theme(panel.background = element_rect(fill = "white", colour = "grey50")
          )
103
104
105
106 # Modelo estático
107   mod1 <- lm(Yt.s ~ xt.s+x2t.s+x3t.s)
108   summary(mod1)
109
110 # Algoritmos de selección de variables
111
112 # Encuentra las mejores combinaciones con "p" predictores, i.e.,
113 # Para cada uno de los modelos generados
114 # (combinaciones con el mismo número de covariables t=1 o 2 o 3)
115 # nos devuelve el "mejor" con base en el valor de rss
116 # (suma de cuadrados de los residuales).
117   combinaciones <- regsubsets(Yt.s ~ xt.s+x2t.s+x3t.s,
118                               data=base.s, nbest = 7)
119
120
121   summary(combinaciones)
122   plot(combinaciones, scale="r2", xlab="Coeficientes de regresión")
123
124
125
126   # Comparación de los 3 mejores modelos
127
128   #--- Validación cruzada
129
130   #Elegiremos el mejor modelo sin importar el número de predictores p
131
132   predict.regsubsets <- function(object, newdata, id ,...) {
133     form <- as.formula(object$call[[2]])
134     mat <- model.matrix(form, newdata)
135     coefi <- coef(object, id = id)
136     xvars <- names(coefi)
137     mat[, xvars] %*% coefi
138   }
139
```

```
140
141   k <- 10
142   set.seed(1)
143   folds <- sample(1:k, nrow(base.s), replace = TRUE)
144   table(folds)
145
146   # n° de obs en cada subconjunto de la muestra
147   # 1  2  3  4  5  6  7  8  9 10
148   # 25 43 41 44 44 32 32 39 31 34
149
150   #Creamos nuestra matriz de errores de predicción
151   cv_errors <- matrix(NA, k, 3, dimnames = list(NULL, paste(1:3)))
152
153   #Realizamos el método de validación cruzada de 10 iteraciones
154   for(j in 1:k) {
155
156     # Obtenemos nuestros mejores modelos del conjunto de entrenamiento
157     best_subset <- regsubsets(Yt.s ~ xt.s+x2t.s+x3t.s, base.s[folds != j,
          ])
158
159     # Método de validación cruzada
160     for( i in 1:3) {
161       #Valores predichos para los mejores modelos de p predictores con
            base en el conjunto de prueba
162       pred_x <- predict.regsubsets(best_subset, base.s[folds == j, ], id =
            i)
163       cv_errors[j, i] <- mean((base.s$Yt.s[folds == j] - pred_x)^2) #
            errores de predicción
164     }
165   }
166
167   # Obtenemos la media de los errores de predicción, resultantes de las 10
        iteraciones,
168   # para cada uno de los 7 modelos
169
170   mean.cv.errors = colMeans(cv_errors)
171
172   #1        2        3
173   #588.7107 550.8936 549.9251
174
175   which.min(mean.cv.errors)
176
177   plot(mean.cv.errors, pch = 19, cex=1.5, type="b", col="blue",
178        xlab = "N° de Covariables", ylab = "Media de los Errores de
           Predicción")
179   points(3, mean.cv.errors[3], pch = 20, cex=1.5, col = "red")
180
181   # El modelo que minimiza el error es el modelo 7 -  Yt.s ~ TMP + WSP +
        RH,
182   # seguido del modelo 4 - Yt.s ~ TMP + WSP
183
184   # Basados en el AIC
185     step <- stepAIC(mod1, direction="both")
```

```
186     step$anova # Muestra los resultados
187
188
189 #--- Validaciones de los residuales de los 3 mejores modelos ---
190
191
192 # Estimación de parámetros para los 3 mejores modelos
193   coef(object = combinaciones, id = 1)
194   coef(object = combinaciones, id = 4)
195   coef(object = combinaciones, id = 7)
196
197 # Modelo Completo: Yt.s ~ xt.s + x2t.s + x3t.s
198
199 # Resumen del modelo
200   summary(mod1)
201 # Intervalos de confianza
202   confint(mod1)
203 # ACF de los residuales
204   acf(mod1$residuals, main="")
205   title(main="ACF de los residuales", col.main="darkred",
206       cex.main=1.5)
207 # Pruebas de normalidad de los residuales
208   qqnorm(mod1$residuals, main="")
209   title(main="Normal Q-Q Plot", col.main="darkred",
210       cex.main=1.5)
211   qqline(mod1$residuals, col="blue")
212   shapiro.test(mod1$residuals)
213   ad.test(mod1$residuals)
214   x=(-30:30)/10
215   hist((mod1$res-mean(mod1$res))/sd(mod1$res), nclass=50,
216      freq=F, border="blue", col="skyblue", xlab="Residuales",
217      cex.axis=1.3, cex.lab=1.3, ylab="Densidad", main=" ")
218   title(main="Distribución de los residuales", col.main="darkred",
219       cex.main=1.5)
220   lines(x,dnorm(x), lwd=2)
221 # Prueba de homocedasticidad de los residuales
222   bptest(mod1)
223   gvlma(mod1)
224
225 # Modelo con 2 covariables: Yt.s ~ xt.s + x3t.s
226
227   mod2 <- lm(Yt.s ~ xt.s + x3t.s)
228
229 # Resumen del modelo
230   summary(mod2)
231 # Intervalos de confianza
232   confint(mod2)
233 # ACF de los residuales
234   acf(mod2$residuals, main="")
235   title(main="ACF de los residuales", col.main="darkred")
236 # Ljung Box
237   Box.test(mod2$residuals, lag=20, type = "Ljung")
238   p_values_LB <- sapply(1 : 15, function(i)
```

```
239       Box.test(mod2$residuals, lag=i, type="Ljung-Box")$p.value)
240
241     plot(p_values_LB, type="p", pch=19, ylim=c(0,1), xlim=c(0,16), lwd=1,
          col="blue", main=" ", xlab="Lag", ylab="p-valor")
242     abline(h=0.05, col="darkred", lty=3, lwd=2)
243     title(main="P-Valores para el estadístico Ljung-Box", col.main="darkred"
          )
244   # Pruebas de normalidad de los residuales
245     qqnorm(mod2$residuals, main="", xlab = "", ylab = "")
246     title(main="Normal Q-Q Plot", col.main="darkred",
247           xlab = "Cuantiles teoricos", ylab = "Cuantiles Muestrales")
248     qqline(mod2$residuals, col="blue")
249     shapiro.test(mod2$residuals)
250     ad.test(mod2$residuals)
251     jarque.bera.test(mod2$residuals)
252     x=(-30:30)/10
253     hist((mod2$res-mean(mod2$res))/sd(mod2$res),nclass=50,
254           freq=F, border="blue", col="skyblue", xlab="Residuales",
255           ylab="Densidad", main=" ")
256     title(main="Distribución de los residuales\n", col.main="darkred")
257     lines(x,dnorm(x), lwd=2)
258   # Prueba de homocedasticidad de los residuales
259     bptest(mod2)
260     gvlma(mod2)
261
262   # Modelo con 1 covariable: Yt.s ~ xt.s
263
264     mod3 <- lm(Yt.s ~ xt.s)
265
266   # Resumen del modelo
267     summary(mod3)
268   # Intervalos de confianza
269     confint(mod3)
270   # ACF de los residuales
271     acf(mod3$residuals, main="")
272     title(main="ACF de los residuales\n Modelo de regresión
273           estático una covariable", col.main="darkred",
274           cex.main=1.5)
275   # Pruebas de normalidad de los residuales
276     qqnorm(mod3$residuals, main="")
277     title(main="Normal Q-Q Plot\n Modelo de regresión lineal
278           estático una covariable", col.main="darkred",
279           cex.main=1.5)
280     qqline(mod3$residuals, col="blue")
281     shapiro.test(mod3$residuals)
282     jarque.bera.test(mod3$residuals)
283     ad.test(mod3$residuals)
284     x=(-30:30)/10
285     hist((mod3$res-mean(mod3$res))/sd(mod3$res), nclass=50,
286           freq=F, border="blue", col="skyblue", xlab="Residuales",
287           cex.axis=1.3, cex.lab=1.3, ylab="Densidad", main=" ")
288     title(main="Distribución de los residuales", col.main="darkred",
289           cex.main=1.5)
```

```
290    lines(x,dnorm(x), lwd=2)
291  # Prueba de homocedasticidad de los residuales
292    bptest(mod3)
293    gvlma(mod3)
294
295  # Transformación de Box - Cox
296
297    boxCox(Yt.s ~ xt.s + x2t.s + x3t.s,
298           lambda=seq(-2,2,length=21), family="yjPower")
299    title(main="Modelo completo", col.main="darkred",
300          cex.main=1.5)
301    boxCox(Yt.s ~ xt.s + x3t.s , lambda=seq(-2,2,length=21),
302           family="yjPower")
303    title(main="Modelo dos covariables", col.main="darkred",
304          cex.main=1.5)
305    boxCox(Yt.s ~ xt.s, lambda=seq(-2,2,length=21),family="yjPower")
306    title(main="Modelo una covariable", col.main="darkred",
307          cex.main=1.5)
```

D. Modelo Dinámico Lineal Multivariado

```
1  # Regresión dinámica lineal Multivariada
2
3  # Librerías
4    suppressMessages(suppressWarnings(library(MARSS)))
5    suppressMessages(suppressWarnings(library(imputeTS)))
6    suppressMessages(suppressWarnings(library(ggplot2)))
7    suppressMessages(suppressWarnings(library(forecast)))
8    suppressMessages(suppressWarnings(library(GGally)))
9    suppressMessages(suppressWarnings(library(changepoint)))
10   suppressMessages(suppressWarnings(library(strucchange)))
11   suppressMessages(suppressWarnings(library(ggpmisc)))
12   suppressMessages(suppressWarnings(library(ggfortify)))
13   suppressMessages(suppressWarnings(library(lubridate)))
14   suppressMessages(suppressWarnings(library(readxl)))
15   suppressMessages(suppressWarnings(library(nortest)))
16
17   #setwd("C:/Users/SIEC/Documents/TESIS/R/Insumos Tesis NMR/Datos Finales
         v3")
18   setwd("C:\\Users\\DELL\\Desktop\\TESIS_NMR\\Insumos Tesis NMR\\Datos
         Finales v3")
19
20   base <- read.csv("Datos_v3.csv")
21   attach(base)
22   fecha <- as.Date(as.character(fecha), "%d/%m/%Y")
23
24 # Diseño Gráficas
25   theme_set(theme_classic())
```

```
26    theme_update(plot.title = element_text(hjust = 0.5, size = 24, face = "
          bold"),
27                 plot.subtitle=element_text(size=20, hjust=0.5),
28                 plot.caption = element_text(size = 18),
29                 axis.title=element_text(size=24),
30                 axis.text = element_text(size=22),
31                 legend.text=element_text(size=24),
32                 legend.position="bottom",
33                 legend.box = "horizontal")#centrar el titulo en las
                       gráficas
34
35  # Número de días de los datos
36    TT = length(fecha)
37
38  # Número de parámetros de regresión (B0, B1, B2)
39    m = 3
40
41  # Variable respuesta Yt
42    Yt = matrix(Yt.s,nrow=1)
43
44  # Covariables
45    xt <- matrix(xt.s, nrow = 1)
46    x2t <- matrix(x2t.s, nrow = 1)
47    x3t <- matrix(x3t.s, nrow = 1)
48
49  #--- Modelo de regresión Dinámico ---
50
51  # Definición de los parámetros de la ecuación del sistema
52
53  # Matriz identidad; 3x3
54    B = diag(m)
55  # Matriz de 0's; 3x1
56    U = matrix(0,nrow=m, ncol=1)
57  # Matriz de varianzas del sistema = 0's por ahora; 3x3
58    Q = matrix(list(0),m,m)
59    # Regresión múltiple: diag(Wt)=[0,0,0] -> thetha_t=thetha
60      Q2 = matrix(list(0),m,m)
61  # diag(Matriz de varianzas del sistema) = (q1,q2,q3); 3x3
62    diag(Q) = c("q.beta0","q.beta1", "q.beta2" )
63    # Regresión múltiple: diag(Wt)=[0,0,0] -> thetha_t=thetha
64      diag(Q2) = c(0,0,0)
65
66  # Definición de los parámetros de la ecuación de observaciones
67
68  Z = array(NA, c(1,m,TT))    ## Z=Ft' vacía por ahora; NxMxT
69  Z[1,1,] = rep(1,TT)         ## 1's para el intercepto; Nx1
70  Z[1,2,] = xt                ## covariable xt; Nx1
71  Z[1,3,] = x3t               ## covariable x3t; Nx1
72  A = matrix(0)               ## escalar 0; 1x1
73  R = matrix("r")             ## escalar r; 1x1;
74
75  # Valores iniciales
76    inits.list = list(x0=matrix(c(0, 0, 0), nrow=m))
```

```
77  # Lista de los parámetros que definen el modelo
78    mod.list = list(B=B, U=U, Q=Q, Z=Z, A=A, R=R)
79    # Regresión múltiple: diag(Wt)=[0,0,0] -> thetha_t=thetha
80      mod.list2 = list(B=B, U=U, Q=Q2, Z=Z, A=A, R=R)
81  # Ajuste del modelo multivariado dinámico lineal: Yt ~ xt + x3t
82    dlm1 = MARSS(Yt, inits=inits.list, model=mod.list)
83    # Regresión múltiple: diag(Wt)=[0,0,0] -> thetha_t=thetha
84      dlm2 = MARSS(Yt, inits=inits.list, model=mod.list2)
85
86  #--- Gráficos modelo de regresión dinámico lineal multivariado ---
87
88  # Intervalos de confianza para alpha, beta1 y beta2
89    b0.u = dlm1$states[1,] + 1.96*dlm1$states.se[1,]
90    b0.l = dlm1$states[1,] - 1.96*dlm1$states.se[1,]
91    b1.u = dlm1$states[2,] + 1.96*dlm1$states.se[2,]
92    b1.l = dlm1$states[2,] - 1.96*dlm1$states.se[2,]
93    b2.u = dlm1$states[3,] + 1.96*dlm1$states.se[3,]
94    b2.l = dlm1$states[3,] - 1.96*dlm1$states.se[3,]
95
96  # Gráficas de los valores suavizados para alpha, beta1 y beta2 con
97  # intervalos al 95% de confianza
98    b0 <- data.frame(dlm1$states[1,], b0.u, b0.l)
99    names(b0) <- c("b0.s", "b0.u", "b0,l")
100   ggplot(b0, aes(x=fecha)) +
101   geom_line(aes(y=b0.s, colour="blue"), size=1.5) +
102   geom_line(aes(y=b0.u, colour="darkblue"), size=1.5) +
103   geom_line(aes(y=b0.l, colour="darkblue"), size=1.5) +
104   geom_hline(yintercept = 88.62, linetype=3, size=1.5) +
105   scale_colour_manual("", values = c("blue", "darkblue", "darkblue"),
106                       labels = c(expression(hat(beta)[0.2017]), "I.C.")) +
107                       labs(title=" ", x=" ", y=expression(hat(beta)["0,t"
                            ]))+
108                       theme(panel.background = element_rect(fill = "white"
                            , colour = "grey50"))
109
110
111   b1 <- data.frame(dlm1$states[2,], b1.u, b1.l)
112   names(b1) <- c("b1.s", "b1.u", "b1,l")
113   ggplot(b1, aes(x=fecha)) +
114   geom_line(aes(y=b1.s, colour="darkorange1"), size=1.5) +
115   geom_line(aes(y=b1.u, colour="darkorange2"), size=1.5) +
116   geom_line(aes(y=b1.l, colour="darkorange2"), size=1.5) +
117   geom_hline(yintercept = 0, linetype=2, colour="red") +
118   geom_hline(yintercept = 17.405, linetype=3, size=1.5) +
119   scale_colour_manual("", values = c("darkorange1", "darkorange2", "
        darkorange2"),
120             labels = c(expression(hat(beta)[1.2017]), "I.C.")) +
121             labs(title=" ", x=" ", y=expression(hat(beta)["1,t"])) +
122             theme(panel.background = element_rect(fill = "white", colour
                    = "grey50"))
123
124
125   b2 <- data.frame(dlm1$states[3,], b2.u, b2.l)
```

```
126    names(b2) <- c("b2.s", "b2.u", "b2,l")
127    ggplot(b2, aes(x=fecha)) +
128    geom_line(aes(y=b2.s, colour="darkgreen"), size=1.5) +
129    geom_line(aes(y=b2.u, colour="forestgreen"), size=1.5) +
130    geom_line(aes(y=b2.l, colour="forestgreen"), size=1.5) +
131    geom_hline(yintercept = 0, linetype=2, colour="red") +
132    geom_hline(yintercept = -6.642, linetype=3, size=1.5) +
133              scale_colour_manual("", values = c("darkgreen", "forestgreen",
                   "forestgreen"),
134              labels = c(expression(hat(beta)[2.2017]), "I.C.")) +
135              labs(title=" ", x=" ", y=expression(hat(beta)["2,t"]))+
136              theme(panel.background = element_rect(fill = "white", colour =
                   "grey50"))
137
138
139  #--- Predicción del modelo dinámico lineal multivariado k=1 ---
140
141
142  # Salida del filtro de Kalman
143    kf.out = MARSSkfss(dlm1)
144
145  # Predicción de los parámetros de regresión; matriz de 3xT
146    eta = kf.out$xtt1
147  # Media de la distribución predictiva un paso adelante para los estados
148  # at=E(thetha_1:t-1|Yt-1)=Gt*mt-1
149    fore.mean = vector()
150    for(t in 1:TT) {
151        fore.mean[t] = Z[,,t] %*% eta[,t,drop=F]
152    }
153    #--- Modelo de regresión múltiple-----------
154      kf.out2 = MARSSkfss(dlm2)
155      eta2 = kf.out2$xtt1
156      fore.mean2 = vector()
157      for(t in 1:TT) {
158        fore.mean2[t] = Z[,,t] %*% eta2[,t,drop=F]
159      }
160    #--------------------------------------------
161
162  # Varianza un paso adelante de los parámetros de regresión al tiempo t; 1
       x2xT
163    Phi = kf.out$Vtt1
164  # Varianza de la ec. de observación; matriz 1x1
165    R.est = coef(dlm1, type="matrix")$R
166  # Varianza de la distribución predictiva
167  # Rt=Var(thetha_1:t-1|Yt-1)=Gt*Ct-1*Gt'+Wt
168    fore.var = vector()
169    for(t in 1:TT) {
170        tZ = matrix(Z[,,t],m,1) ## transpose of Z
171        fore.var[t] = Z[,,t] %*% Phi[,,t] %*% tZ + R.est
172    }
173
174      #--------------------------------------------
```

```
      # Varianza un paso adelante de los parámetros de regresión al tiempo
          t; 1x2xT
        Phi2 = kf.out2$Vtt1
      # Varianza de la ec. de observación; matriz 1x1
        R.est2 = coef(dlm2, type="matrix")$R
      # Varianza de la distribución predictiva
      # Rt=Var(thetha_1:t-1|Yt-1)=Gt*Ct-1*Gt'+Wt
        fore.var2 = vector()
        for(t in 1:TT) {
          tZ2 = matrix(Z[,,t],m,1) ## transpose of Z
          fore.var2[t] = Z[,,t] %*% Phi2[,,t] %*% tZ2 + R.est2
        }
    #-------------------------------------------------


# Predicción el vector de observaciones un paso hacia adelante
  fk_dlm1 = fore.mean
  fk_lower = fore.mean-2*sqrt(fore.var)
  fk_upper = fore.mean+2*sqrt(fore.var)
  forecast_dlm1 <- data.frame(fecha=fecha, Obs=Yt.s, fk1=fore.mean, low=fk
      _lower, upp=fk_upper)
  names(forecast_dlm1) <- c("fecha", "Yt.s", "fkdlm1", "lower", "upper")

  ggplot(forecast_dlm1, aes(x=fecha)) +
  geom_point(aes(y=Yt.s, colour="darkblue"), size=2.1) +
  geom_line(aes(y=fkdlm1, colour="darkgreen"), size=1.5) +
  geom_line(aes(y=lower, colour="forestgreen"), size=1, linetype=1) +
  geom_line(aes(y=upper, colour="forestgreen"), size=1, linetype=1) +
  scale_colour_manual("", values = c("darkblue", "darkgreen", "forestgreen
      ",
                                     "forestgreen"),
                           labels = c("Yt","Valores predichos modelo
                              dinámico",
                                     "Intervalos de 95% de confianza")) +
                           labs(title=" ", x=" ", y=expression(O[3](ppb)),
                           caption="\nDatos Red Automática de Monitoreo
                              Atmosférico (RAMA)") +
                           theme(panel.background = element_rect(fill = "
                              white", colour = "grey50"))

#---------------- Predicciones un paso adelante ------------------------
# Modelo de regresión múltiple vs Modelo Dinámico Lineal Multivariado

  fk_lower_est = fore.mean2-2*sqrt(fore.var2)
  fk_upper_est = fore.mean2+2*sqrt(fore.var2)

  forecast1 <- data.frame(fecha=fecha, Obs=Yt.s, fk1=fore.mean, fk2=fore.
      mean2, low=fk_lower, upp=fk_upper,
                           low_est=fk_lower_est, upp_est=fk_upper_est)

  names(forecast1) <- c("fecha", "Yt.s", "fkdlm", "fkest", "lower", "upper
      ", "low_est", "upp_est")

```

```
220     ggplot(forecast1, aes(x=fecha)) +
221     geom_point(aes(y=Yt.s, colour="blue"), size=1.2) + geom_point(aes(y=Yt.s
            ), colour="blue", size=2.2, shape=15) +
222     geom_line(aes(y=fkdlm, colour="darkgreen"), size=1) + geom_point(aes(y=
            fkdlm), colour="darkgreen", size=1.5) +
223     geom_line(aes(y=lower, colour="forestgreen"), size=1.4, linetype=3) +
224     geom_line(aes(y=upper, colour="forestgreen"), size=1.4, linetype=3) +
225     geom_line(aes(y=fkest, colour="darkorange"), size=1, linetype=1) + geom_
            point(aes(y=fkest), colour="darkorange2", size=1.5) +
226     geom_line(aes(y=low_est, colour="darkorange1"), size=1.4, linetype=3) +
227     geom_line(aes(y=upp_est, colour="darkorange1"), size=1.4, linetype=3) +
228     scale_colour_manual("", values = c("blue", "darkgreen", "forestgreen","
            forestgreen","darkorange", "darkorange1", "darkorange1"),
229                         labels = c("Yt ","fk MDLM", "I.C. MDLM", "fk
                                modelo estático", "I.C. modelo estático")) +
230                         labs(title=" ",
231                         caption="\nDatos Red Automática de Monitoreo
                                Atmosférico (RAMA)", x=" ", y=expression(O
                                [3](ppb)))+
232     theme(panel.background = element_rect(fill = "white", colour = "grey50")
            )
233
234     # MDLM: Conteo del número de observaciones que caen dentro del intervalo
            al 95% de confianza
235     cuenta = ifelse(forecast1$Yt.s>=forecast1$lower & forecast1$Yt.s<=
            forecast1$upper,1,0)
236     sum(cuenta)
237     sum(cuenta)/sum(!is.na(cuenta))
238
239
240     # Modelo estático: Conteo del número de observaciones que caen dentro
            del intervalo al 95% de confianza
241     cuentaMEST = ifelse(forecast1$Yt.s>=forecast1$low_est & forecast1$Yt.s<=
            forecast1$upp_est,1,0)
242     sum(cuentaMEST)
243     sum(cuentaMEST)/sum(!is.na(cuentaMEST))
244
245   #------------------ Validaciones del Modelo ------------------------------
246
247   # Modelo Dinámico Lineal Multivariado
248     MAE <- mean(abs(fore.mean-t(Yt)))
249     MSE1 <- mean((fore.mean-t(Yt))^2)
250     MAPE1 <- mean(abs(fore.mean-t(Yt))/abs(t(Yt)))*100
251     sqrt(sum((fore.mean - t(Yt))[-(1:5)]^2)/sum(diff(t(Yt)[-(1:4)])^2))
252     #sqrt(sum((fore.mean - t(Yt))^2)/sum(diff(t(Yt))^2))
253
254   # Modelo de Regresión Lineal Multivariado
255     MAE2 <- mean(abs(fore.mean2-t(Yt)))
256     MSE2 <- mean((fore.mean2-t(Yt))^2)
257     MAPE2 <- mean(abs(fore.mean2-t(Yt))/abs(t(Yt)))*100
258     sqrt(sum((fore.mean2 - t(Yt))[-(1:5)]^2)/sum(diff(t(Yt)[-(1:4)])^2))
259     #sqrt(sum((fore.mean2 - t(Yt))^2)/sum(diff(t(Yt))^2))
260
```

```
261
262 # Diagnósticos para los residuales del modelo
263
264
265 # Errores de prediccion o innovaciones del modelo
266   innov = kf.out$Innov
267
268 # ACF de los residuales
269   acf(t(innov), main="")
270   title(main="ACF de los Residuales", col.main="darkred")
271
272 # Ho: La media de la distribución de las innovaciones es igual a cero
273 # H1: La media de la distribución de las innovaciones es distinta de cero
274
275   t.test(t(innov))
276
277 # Ljung Box
278   Box.test(t(innov), lag=20, type = "Ljung")
279   p_values_LB <- sapply(1 : 15, function(i)
280   Box.test(t(innov), lag=i, type="Ljung-Box")$p.value)
281
282   plot(p_values_LB, type="p", pch=19, ylim=c(0,1), xlim=c(0,16), lwd=1,
        col="blue", main=" ", xlab="Lag", ylab="p-valor")
283   abline(h=0.05, col="darkred", lty=3, lwd=2)
284   title(main="p-valores para el estadístico Ljung-Box", col.main="darkred"
        )
285
286 # Histograma de los Residuales
287   x=(-35:35)/10
288   hist((t(innov)-mean(t(innov)))/sd(t(innov)),nclass=50,freq=F,
289        border="blue", col="skyblue", xlab="Residuales",
290        ylab="Densidad", main=" ")
291   title(main="Distribución de los residuales", col.main="darkred")
292   lines(x,dnorm(x), lwd=2)
293
294 # Q-Q plot de los residuales
295   qqnorm(t(innov), main=" ", ylab="Cuantiles muestrales", xlab="Cuantiles
        teoricos")
296   qqline(t(innov), col="blue", lwd=1.8, lty=1.3)
297   title(main="Normal QQ-Plot", col.main="darkred")
298
299 # Prueba de Normalidad Anderson Darling
300   ad.test(t(innov))
301   t.test(t(innov))
302
303
304 #--- Predicción de observaciones nuevas k=31 días hacia adelante ---
305
306
307 # Si bien la evalución de ambos modelos y el prónostico de un paso
308 # hacia adelante de las observaciones sugieren que el modelo dinámico
309 # brinda un mejor ajuste para nuestros datos, mostraremos las
310 # predicciones del modelo k pasos hacia adelante.
```

```
311
312 # Pronóstico de los estados (theta_t) k=31 pasos hacia adelante
313
314   # Valores verdaderos para las series O3, TMP y WSP
315     base_2018 <- read.csv("base_2018.csv")
316     fecha_2018 <- as.Date(as.character(base_2018$fecha), "%d/%m/%Y")
317     Yt_2018 <- base_2018[1:31,]
318
319   # Estandarizar covariables
320     xt_2018 <- (base_2018$TMP_max - mean(base_2018$TMP_max, na.rm=T))/sqrt
            (var(base_2018$TMP_max, na.rm=T))
321     x3t_2018 <- (base_2018$WSP_max - mean(base_2018$WSP_max, na.rm=T))/
            sqrt(var(base_2018$WSP_max, na.rm=T))
322
323   # Covariable temperatura
324     xtnew = c(xt, xt_2018[1:31])
325     xtnew = ts(xtnew, start = c(2017,1), frequency = 365)
326   # Covariable WSP
327     x3tnew = c(x3t, x3t_2018[1:31])
328     x3tnew <- ts(x3tnew, start = c(2017,1), frequency = 365)
329
330   # Número de obs. nuevas a predecir
331     TTnew = 31
332   # Vector de NA's de longitud 31
333     Ytnew = rep(NA,TTnew)
334   # Matriz de observaciones concatenada con los 31 NA's
335     Yt3 = matrix(c(Yt,Ytnew),nrow=1)
336   # Matriz de dimensiones NxMxT; vacía por ahora
337     Z3 = array(NA, c(1,m,(TT+TTnew)))
338   # Matriz de 1's para el intercepto;  Nx1
339     Z3[1,1,] = rep(1,(TT+TTnew))
340   # Covariables temperatura; Nx1
341     Z3[1,2,] = xtnew
342   # Covariables velocidad del viento; Nx1
343     Z3[1,3,] = x3tnew
344
345   # Definición y ajuste del modelo dinámico lineal
346     mod.list3 = list(B=B, U=U, Q=Q, Z=Z3, A=A, R=R)
347     dlm3 = MARSS(Yt3, inits=inits.list, model=mod.list3)
348     #--- Definición y ajuste del modelo de regresión múltiple ---
349     mod.list3_est = list(B=B, U=U, Q=Q2, Z=Z3, A=A, R=R)
350     dlm3_est = MARSS(Yt3, inits=inits.list, model=mod.list3_est)
351
352 # Gráficos Modelo Predictivo k=31 pasos hacia adelante
353
354   # Predicción de Estados MDLM
355
356 # Intervalos de confianza para alpha, beta1 y beta2
357   b0.u.new = dlm3$states[1,] + 1.96*dlm3$states.se[1,]
358   b0.l.new = dlm3$states[1,] - 1.96*dlm3$states.se[1,]
359   b1.u.new = dlm3$states[2,] + 1.96*dlm3$states.se[2,]
360   b1.l.new = dlm3$states[2,] - 1.96*dlm3$states.se[2,]
361   b2.u.new = dlm3$states[3,] + 1.96*dlm3$states.se[3,]
```

```
362   b2.1.new = dlm3$states[3,] - 1.96*dlm3$states.se[3,]
363
364 # Valores suavizados para alpha, beta1 y beta2 e intervalos de confianza
365   fecha.new = c(fecha, fecha_2018[1:31])
366
367   b0.new <- data.frame(c(dlm3$states[1,1:365], rep(NA,31)), c(rep(NA,365),
          dlm3$states[1,366:dim(dlm3$states)[2]]),
368                        b0.u.new, b0.l.new)
369   names(b0.new) <- c("b0.s1", "b0.s2", "b0.u", "b0.l")
370   ggplot(b0.new, aes(x=fecha.new)) +
371   geom_line(aes(y=b0.s1, colour="blue"), size=1.5) +
372   geom_line(aes(y=b0.s2, colour="cornflowerblue"), size=1.5) +
373   geom_line(aes(y=b0.u, colour="darkblue"), size=1.5) +
374   geom_line(aes(y=b0.l, colour="darkblue"), size=1.5) +
375   #geom_vline(aes(xintercept=fecha.new[365]), size=1.5, linetype=2, color
          ="darkred")+
376   scale_colour_manual("", values = c("blue", "hotpink", "darkblue", "
          darkblue"),
377   labels = c(expression(hat(beta)[0.2017]), expression(hat(beta)[0.2018]),
           "I.C.")) +
378   labs(title=" ", x=" ", y=expression(hat(beta)["0,t"]))+
379   theme(panel.background = element_rect(fill = "white", colour = "grey50")
          )
380
381   b1.new <- data.frame(c(dlm3$states[2,1:365], rep(NA,31)), c(rep(NA,365),
          dlm3$states[2,366:dim(dlm3$states)[2]]),
382                        b1.u.new, b1.l.new)
383   names(b1.new) <- c("b1.s1", "b1.s2", "b1.u", "b1.l")
384   ggplot(b1.new, aes(x=fecha.new)) +
385   geom_line(aes(y=b1.s1, colour="darkorange1"), size=1.5) +
386   geom_line(aes(y=b1.s2, colour="cornflowerblue"), size=1.5) +
387   geom_line(aes(y=b1.u, colour="darkorange2"), size=1.5) +
388   geom_line(aes(y=b1.l, colour="darkorange2"), size=1.5) +
389   #geom_vline(aes(xintercept=fecha.new[365]), size=1.5, linetype=2, color
          ="darkred")+
390   geom_hline(yintercept = 0, linetype=2) +
391   scale_colour_manual("", values = c("hotpink", "darkorange1", "
          darkorange2", "darkorange2"),
392   labels = c(expression(hat(beta)[1.2017]), expression(hat(beta)[1.2018]),
           "I.C.")) +
393   labs(title=" ", x=" ", y=expression(hat(beta)["1,t"]))+
394   theme(panel.background = element_rect(fill = "white", colour = "grey50")
          )
395
396   b2.new <- data.frame(c(dlm3$states[3,1:365], rep(NA,31)), c(rep(NA,365),
          dlm3$states[3,366:dim(dlm3$states)[2]]),
397                        b2.u.new, b2.l.new)
398   names(b2.new) <- c("b2.s1", "b2.s2", "b2.u", "b2.l")
399   ggplot(b2.new, aes(x=fecha.new)) +
400   geom_line(aes(y=b2.s1, colour="darkgreen"), size=1.5) +
401   geom_line(aes(y=b2.s2, colour="hotpink"), size=1.5) +
402   geom_line(aes(y=b2.u, colour="forestgreen"), size=1.5) +
403   geom_line(aes(y=b2.l, colour="forestgreen"), size=1.5) +
```

```
404     #geom_vline(aes(xintercept=fecha.new[365]), size=1.5, linetype=2, color
            ="darkred")+
405     geom_hline(yintercept = 0, linetype=2) +
406     scale_colour_manual("", values = c( "darkgreen", "forestgreen", "hotpink
            ", "forestgreen"),
407                         labels = c(expression(hat(beta)[2.2017]), expression
                                (hat(beta)[2.2018]), "I.C.")) +
408     labs(title=" ", x=" ", y=expression(hat(beta)["2,t"]))+
409     theme(panel.background = element_rect(fill = "white", colour = "grey50")
            )
410
411     #Prónostico MDLM k=31
412
413     # Salida del filtro de Kalman
414       kf.out3 = MARSSkfss(dlm3)
415     # Predicción de los parámetros de regresión; matriz de 3xT
416       at3 = kf.out3$xtt1
417     # Media de la distribución predictiva un paso adelante para los estados
418       ft3 = vector()
419       for(t in 1:(TT+TTnew)) {
420           ft3[t] = Z3[,,t] %*% at3[,t,drop=F]
421       }
422
423       #--- Prónostico modelo de regresión múltiple ---
424        kf.out3_est = MARSSkfss(dlm3_est)
425        at3_est = kf.out3_est$xtt1
426        ft3_est = vector()
427        for(t in 1:(TT+TTnew)) {
428          ft3_est[t] = Z3[,,t] %*% at3_est[,t,drop=F]
429        }
430
431     # Varianza un paso adelante de los parámetros de regresión al tiempo t;
            1x2xT
432       Rt3 = kf.out3$Vtt1
433     # Varianza de la ec. de observación; matriz 1x1
434       Vt3 = coef(dlm3, type="matrix")$R
435     # Varianza de la distribución predictiva
436       Qt3 = vector()
437       for(t in 1:(TT+TTnew)) {
438           tZ3 = matrix(Z3[,,t],m,1) #Matriz transpuesta de Z
439           Qt3[t] = Z3[,,t] %*% Rt3[,,t] %*% tZ3 + Vt3
440       }
441
442       #--- Prónostico modelo de regresión múltiple ---
443       Rt3_est = kf.out3_est$Vtt1
444       Vt3_est = coef(dlm3_est, type="matrix")$R
445       Qt3_est=vector()
446       for(t in 1:(TT+TTnew)) {
447         tZ3_est = matrix(Z3[,,t],m,1) #Matriz transpuesta de Z
448         Qt3_est[t] = Z3[,,t] %*% Rt3_est[,,t] %*% tZ3_est + Vt3_est
449       }
450
451
```

```
452   # Valores predichos MDLM
453     fk_dlm3 = ft3
454     #--- Valores predichos modelo de regresión múltiple
455       fk_dlm3_est = ft3_est
456   # Intervalos de confianza
457     fk_lower3 = ft3 -1.96*sqrt(Qt3)
458     fk_upper3 = ft3 +1.96*sqrt(Qt3)
459     # --- Intervalos de confianza modelo de regresión múltiple
460       fk_lower3_est = ft3_est -1.96*sqrt(Qt3_est)
461       fk_upper3_est = ft3_est +1.96*sqrt(Qt3_est)
462
463   # Serie de datos observado (reales)
464     Yt.s3 <- c(Yt.s, rep(NA,31))
465     Yt_nuevo <- c(rep(NA,365), base_2018$PED_max[1:31])
466     Yt_comp <- c(Yt.s, base_2018$PED_max[1:31])
467     fk_dlm3_1 <- c(fk_dlm3[1:365], rep(NA, 31))
468     fk_dlm3_2 <- c(rep(NA, 365), fk_dlm3[366:length(fk_dlm3)])
469
470     forecast_dlm3 <- data.frame(fecha=fecha.new, Obs=Yt.s3, new_real=Yt_
          nuevo , fk1=fk_dlm3_1,
471                                 fk1=fk_dlm3_2, low=fk_lower3, upp=fk_
                                      upper3)
472     names(forecast_dlm3) <- c("fecha.new", "Yt3", "new_real", "fkdlm3_1",
          "fkdlm3_2", "lower", "upper")
473
474     ggplot(forecast_dlm3, aes(x=fecha.new)) +
475     geom_point(aes(y=Yt3, colour="cornflowerblue"), size=2.2) +
476     geom_point(aes(y=new_real, colour="darkblue"), size=2.2) +
477     geom_line(aes(y=fkdlm3_1, colour="darkgreen"), size=1.5) +
478     geom_line(aes(y=fkdlm3_2, colour="indianred1"), size=1.5) +
479     geom_line(aes(y=lower, colour="forestgreen"), size=1, linetype=1) +
480     geom_line(aes(y=upper, colour="forestgreen"), size=1, linetype=1) +
481     #geom_vline(aes(xintercept=fecha.new[365]), size=1.5, linetype=2,
          color="darkred") +
482     scale_colour_manual("", values = c("darkblue", "cornflowerblue", "
          darkgreen","forestgreen", "indianred1", "forestgreen"),
483                         labels = c(expression(Y[t.2017]), expression(Y[t
                                .2018]), expression(f[t.2017]),"I.C.",
                                expression(f[t.2018]))) +
484     labs(title="Predicción de las concentraciones de ozono k=31 días",
485          subtitle="Serie de las concentraciónes de Ozono Máximas Diarias
                 del Promedio Móvil de 8 hrs\nEstación Pedregal (PED),
                 Delegación Alvaro Obregón, CDMX (2017)\n",
486          caption="\nDatos Red Automática de Monitoreo Atmosférico (RAMA)",
487          x=" ", y=expression(O[3](ppb))) +
488     theme(panel.background = element_rect(fill = "white", colour = "grey50
          "))
489
490     #--- Prónostico MDLM vs modelo de regresión múltiple
491
492       fk_dlm3_1_est <- c(fk_dlm3_est[1:365], rep(NA, 31))
493       fk_dlm3_2_est <- c(rep(NA, 365), fk_dlm3_est[366:length(fk_dlm3_est)
            ])
```

```
494
495         forecast_dlm3_est <- data.frame(fecha=fecha.new, Obs=Yt.s3, new_real
                =Yt_nuevo ,
496                                         fk_1=fk_dlm3_1,fk_2=fk_dlm3_2, low=
                                              fk_lower3, upp=fk_upper3,
497                                         fk_1_est=fk_dlm3_1_est, fk_2_est=fk_
                                              dlm3_2_est,
498                                         low_est=fk_lower3_est, upp_est=fk_
                                              upper3_est)
499
500         names(forecast_dlm3_est) <- c("fecha.new", "Yt3", "new_real",
501                                       "fkdlm3_1", "fkdlm3_2", "lower", "upper",
502                                       "fkdlm3_1_est", "fkdlm3_2_est", "lower_est
                                            ", "upper_est")
503
504
505       ggplot(forecast_dlm3_est, aes(x=fecha.new)) +
506         geom_point(aes(y=Yt3, colour="darkblue"), size=2.2) + geom_point(aes
              (y=Yt3), colour="darkblue", size=2.2, shape=15) +
507         geom_point(aes(y=new_real, colour="cornflowerblue"), size=2.2) +
              geom_point(aes(y=new_real), colour="cornflowerblue", size=2.2,
              shape=15) +
508         geom_line(aes(y=fkdlm3_1, colour="darkgreen"), size=1.5) + geom_
              point(aes(y=fkdlm3_1), colour="darkgreen", size=2.2) +
509         geom_line(aes(y=fkdlm3_2, colour="indianred1"), size=1.5) +
510         geom_line(aes(y=lower, colour="forestgreen"), size=1, linetype=3) +
511         geom_line(aes(y=upper, colour="forestgreen"), size=1, linetype=3) +
512         geom_line(aes(y=fkdlm3_1_est, colour="darkorange"), size=1.5) +
513         geom_line(aes(y=fkdlm3_2_est, colour="violet"), size=1.5) +
514         geom_line(aes(y=lower_est, colour="darkorange1"), size=1, linetype
              =3) +
515         geom_line(aes(y=upper_est, colour="darkorange1"), size=1, linetype
              =3) +
516         #geom_vline(aes(xintercept=fecha.new[365]), size=1.5, linetype=2,
              color="darkred") +
517         scale_colour_manual("", #values = c("darkblue", "cornflowerblue", "
              darkgreen",
518                                 #           "forestgreen", "indianred1", "
                                        forestgreen"),
519                             values = c("darkblue", "cornflowerblue", "
                                    darkgreen", "darkorange",
520                                        "forestgreen", "darkorange1",
521                                        "indianred1", "violet","darkorange1",
                                            "darkorange1"),
522                             labels = c(expression(Y[t.orig.2017]),
                                    expression(Y[t.orig.2018]), expression(f[t
                                    .2017.dinámico]),
523                                        expression(f[t.2017.estático]),"I.C.
                                            Modelo dinámico", "I.C. Modelo
                                            estático",
524                                        expression(f[t.2018.dinámico]),
                                            expression(f[t.2018.estático]))) +
525         labs(title="Predicción de las concentraciones de ozono k=31 días",
```

```
526             subtitle="Serie de las concentraciónes de Ozono Máximas Diarias
                    del Promedio Móvil de 8 hrs\nEstación Pedregal (PED),
                   Delegación Alvaro Obregón, CDMX (2017)\n",
527             caption="\nDatos Red Automática de Monitoreo Atmosférico (RAMA)
                   ",
528             x=" ", y=expression(O[3](ppb))) +
529       theme(panel.background = element_rect(fill = "white", colour = "
              grey50"))
530
531     #--- Evaluación del Modelo k=31 ---
532
533     # Modelo Dinámico Lineal Multivariado
534     MAE3 <- mean(abs(fk_dlm3-Yt_comp))
535     MSE3 <- mean((fk_dlm3-Yt_comp)^2)
536     MAPE3 <- mean(abs(fk_dlm3-Yt_comp)/abs(Yt_comp))*100
537     sqrt(sum((fk_dlm3 - Yt_comp)[-(1:5)]^2)/sum(diff(Yt_comp[-(1:4)])^2))
538     #sqrt(sum((fore.mean - t(Yt))^2)/sum(diff(t(Yt))^2))
539
540     # Modelo de Regresión Lineal Multivariado
541     MAE3_est <- mean(abs(fk_dlm3_est-Yt_comp))
542     MSE3_est <- mean((fk_dlm3_est-Yt_comp)^2)
543     MAPE3_est <- mean(abs(fk_dlm3_est-Yt_comp)/abs(Yt_comp))*100
544     sqrt(sum((fore.mean2 - t(Yt))[-(1:5)]^2)/sum(diff(t(Yt)[-(1:4)])^2))
545     #sqrt(sum((fore.mean2 - t(Yt))^2)/sum(diff(t(Yt))^2))
546
547
548
549     # MDLM: Conteo del número de observaciones que caen dentro del
            intervalo al 95% de confianza
550     cuenta3 = ifelse(Yt_comp<fk_upper3 & Yt_comp>fk_lower3,1,0)
551     sum(cuenta3)
552     sum(cuenta3)/sum(!is.na(cuenta3))
553
554       # Modelo estático: Conteo del número de observaciones que caen
              dentro del intervalo al 95% de confianza
555       cuenta3_est = ifelse(Yt_comp<fk_upper3_est & Yt_comp>fk_lower3_est
              ,1,0)
556       sum(cuenta3_est)
557       sum(cuenta3_est)/sum(!is.na(cuenta3_est))
```

Bibliografía

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, 3rd edition.

Bernardo, J. M. and Smith, A. F. (1994). *Bayesian theory*. Wiley Series in Probability and Statistics.

Campagnoli, P., Petris, G., and Petrone, S. (2009). *Dynamic Linear Models with R*. Springer.

Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*. Chapman and Hall/CRC, 3rd edition.

Huerta, G., Sansó, B., and Stroud, J. R. (2004). A spatio–temporal model for Mexico City ozone levels. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(2):231–248.

Moritz, S. and Bartz-Beielstein, T. (2017). imputeTS: time series missing value imputation in R. *The R Journal*, 9(1):207–218.

Osthus, D., Caragea, P., Higdon, D., Morley, S., Reeves, G., and Weaver, B. (2014). Dynamic linear models for forecasting of radiation belt electrons and limitations on physical interpretation of predictive models. *Space Weather*, 12(6):426–446.

Sahu, S. K. and Bakar, K. (2012). A comparison of Bayesian models for daily ozone concentration levels. *Statistical Methodology*, 9(1-2):144–157.

Sahu, S. K., Gelfand, A. E., and Holland, D. M. (2007). High-resolution space–time ozone modeling for assessing trends. *Journal of the American Statistical Association*, 102(480):1221–1234.

Secretaria de Salud (2014). NORMA Oficial Mexicana NOM-020-SSA1-2014, Salud ambiental. Valor límite permisible para la concentración de ozono (O3) en el aire ambiente y criterios para su evaluación. Al margen un sello con el Escudo Nacional, que dice: Estados Unidos Mexicanos.- Secretaría de Salud. *Diario Oficial de la Federación*.

West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*. Springer-Verlag, 2nd edition.