



CIMAT

Centro de Investigación en Matemáticas, A.C.

Estimaciones y pronósticos de tendencias de los principales contaminantes atmosféricos monitoreados por regiones de la Zona Metropolitana del Valle de México a través de una perspectiva no paramétrica, 2008-2018

T E S I S

Que para obtener el grado de
Maestro en Análisis Estadístico

PRESENTA

ELBA RAMOS IBARRA

DIRECTOR DE TESIS:

Dr. José Eliud Silva Urrutia

Guanajuato, Gto, 9 de octubre de 2019

Agradecimientos

A Dios por haberme permitido terminar la maestría.

Al Instituto Nacional de Estadística y Geografía (INEGI) y al Centro de Investigación en Matemáticas (CIMAT) por el apoyo en la formación de la Maestría en Análisis Estadístico.

A mi director de tesis Dr. J. Eliud Silva por todo su apoyo para la realización de este trabajo y por cada una de sus observaciones.

A mis sinodales Dr. Rogelio Ramos y Dra. Leticia Ramirez por su tiempo en la lectura de esta investigación y comentarios.

A mis profesores por ampliarme el panorama de conocimiento sobre la estadística.

A mis padres, Gilberto y Elia, y a mis hermanas, Laura y Lourdes, por todo el cariño que me han dado y porque siempre han estado a mi lado.

Resumen

Se estiman tendencias y pronósticos de los principales contaminantes atmosféricos (O_3 , SO_2 , NO_2 , CO , PM_{10} , $PM_{2.5}$, NO y NO_x) por regiones de la Zona Metropolitana del Valle de México (ZMVM); con datos de los máximos diarios desde 2008 hasta 2018, obtenidos de la Red Automática de Monitoreo Atmosférico (RAMA). Se utiliza una técnica estadística no paramétrica de suavizamiento controlado basada en el filtro de Hodrick y Prescott y estimada a través del filtro de Kalman, propuesta por Guerrero (2007 y 2008), ya que no tiene supuestos distribucionales y su implementación y mantenimiento se consideran simples. Se generan tanto estimaciones puntuales e intervalos de las tendencias y sus respectivos pronósticos. Se comparan las estimaciones contra la Norma Ambiental para la Ciudad de México (NADF-009-AIRE-2017) y se evidencia que, en general, se sigue estando distante de una buena calidad del aire en la ZMVM, pues a diferencia del CO y NO_2 el resto de los contaminantes tienen tendencias y pronósticos que distan de los límites permisibles.

Palabras clave: tendencias, suavizamiento controlado, índice de suavidad, parámetro de suavizamiento, filtro de Hodrick-Prescott, filtro de Kalman, pronósticos, contaminantes atmosféricos.

CONTENIDO

Introducción.....	1
1 Antecedentes.....	3
1.1 Breve panorama de la contaminación del aire en la ZMVM	3
1.2 Análisis estadísticos sobre los contaminantes del aire en la ZMVM y otras ciudades de México.....	4
1.3 Otras ciudades del mundo	6
2 Algunas técnicas de suavizamiento	9
2.1 Media móvil	9
2.3 Suavizado exponencial.....	9
2.4 Método de Holt-Winters	10
2.5 Splines	12
3 Metodología.....	18
3.1 Suavizamiento controlado y pronósticos.....	18
3.2 Índice de suavidad.....	20
3.3 Medidas de precisión de los pronósticos.....	21
3.4 Delimitación de la ZMVM y ubicación de los centros de monitoreo	22
3.5 Datos.....	24
4. Resultados.....	26
4.1 Tendencias.....	26
4.2 Pronósticos	31
Conclusiones.....	35
Anexo A.....	37
Anexo B.....	38
Anexo C.....	41
Referencias	47

LISTA DE FIGURAS Y TABLAS

<i>Figura 1</i> Ubicación de las unidades de monitoreo de la ZMVM según región geográfica	23
<i>Tabla 1</i> Catálogo de estaciones de monitoreo por región geográfica.....	24
<i>Tabla 2</i> Datos faltantes de los contaminantes por región.....	25
<i>Tabla 3</i> Criterios de la calidad del aire	26
<i>Figura 2</i> Series observadas, estimación de tendencias e intervalos de estimación a 2 desviaciones estándar por contaminante y región geográfica.....	29
<i>Figura 2</i> Series observadas, estimación de tendencias e intervalos de estimación a 2 desviaciones estándar por contaminante y región geográfica (Continuación)	30
<i>Tabla 4</i> Medidas de precisión de los pronósticos por contaminante y región.....	32
<i>Figura 3</i> Estimación de tendencia y pronósticos de la tendencia en un horizonte de 7 días (del 26 al 31 de octubre de 2018) para cada contaminante según la región geográfica	33
<i>Figura 3</i> Estimación de tendencias y pronósticos de la tendencia en un horizonte de 7 días (del 26 al 31 de octubre de 2018) para cada contaminante según la región geográfica (Continuación).....	34

Introducción

De acuerdo con la Organización para la Cooperación y el Desarrollo Económico (OCDE, 2015), la Zona Metropolitana del Valle de México (ZMVM) es la tercera metrópoli de esta organización que concentra el mayor número de población; según lo reportado por el Consejo Nacional de Población (CONAPO et al., 2018) cuenta con 20.9 millones de personas en una superficie de 7866 km², ubicándose al suroeste de la cuenca de México a una altitud de 2240 metros sobre el nivel del mar y con un total de 76 municipios. Diversos estudios como Bravo et al. (2002), Secretaría del Medio Ambiente (SEDEMA, 2017) y Rodríguez et al. (2016) coinciden que, debido a su geografía y características orográficas, se impide la circulación del viento, lo cual aunado a la dinámica de movilidad de la población e industrialización, se propicia la acumulación de contaminantes atmosféricos.

Por otra parte, la OCDE (2015) refiere que la administración unidimensional y sectorizada de la ZMVM no permiten la generación de políticas públicas integrales a largo plazo, que alienten a la mejora de los problemas sociales y ambientales. La Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT, 2016) señala que la calidad del aire, además de estar afectada por factores climáticos y geográficos, se encuentra relacionada con la concentración y características de los contaminantes en la región, por lo que para entender su comportamiento es indispensable conocer los niveles de concentración de los principales contaminantes.

El análisis aquí expuesto busca estimar la tendencia de los máximos diarios y sus respectivos pronósticos. Se enfoca a ocho de los principales contaminantes atmosféricos: O₃ -ozono-, SO₂ -dióxido de azufre-, NO₂ -dióxido de nitrógeno-, CO -monóxido de carbono-, PM₁₀ -partículas menores a 10 micrómetros-, PM_{2.5} -partículas menores a 2.5 micrómetros-, NO -monóxido de nitrógeno- y NO_x -óxido de nitrógeno-. El estudio se lleva a cabo aplicando la técnica estadística no paramétrica de suavizamiento controlado propuesta por Guerrero (2007 y 2008), misma que está basada en el filtro de Hodrick y Prescott. Se consideran datos diarios para el periodo de enero de 2008 a octubre de 2018, dentro de cinco regiones que constituyen a la ZMVM; los datos fueron extraídos de la Red Automática de Monitoreo Atmosférico (RAMA).

Cabe citar que, en este trabajo, no se pretende evidenciar la eficiencia o ineficiencia de programas formulados por las respectivas autoridades para el control de la calidad del aire en la ZMVM (para ello véase, por ejemplo, Gallego et al. (2013a y 2013b) y Davis (2008), entre otros). Asimismo, se deja de lado el estudio sobre la morbilidad y mortalidad de la población que se encuentra expuesta a la contaminación del aire. Debe notarse que no se trata de todos los contaminantes atmosféricos existentes en la ZMVM, sino solo aquellos que son monitoreados y registrados por la red señalada.

El trabajo está organizado en cuatro capítulos. En el primero se expone brevemente un panorama de los contaminantes en la ZMVM, además de citar diversas investigaciones que tienen de manifiesto el estudio de tendencias y pronóstico de contaminantes, tanto en ciudades de México como en algunas otras del mundo. Para el segundo se abordan algunas técnicas de suavizamiento. En el tercero se trata con detalle la metodología que se sigue para estimar y predecir el comportamiento de los contaminantes atmosféricos seleccionados, posteriormente se expone la procedencia de los datos y su procesamiento para la apropiada aplicación de la metodología. En el cuarto, se ilustran los resultados de las estimaciones de las tendencias y sus pronósticos, se contrastan con la Norma NADF-009-AIRE-2017. Posteriormente, se abordan las conclusiones. En el Anexo A se muestra la formación de los contaminantes que se tratan en este trabajo, el Anexo B hace referencia a un modelo espacio estado y el filtro de Kalman y en el Anexo C contiene el código en R utilizado para el estudio. Finalmente, se muestran las referencias consultadas.

1 Antecedentes

La contaminación del aire en la ZMVM se ha hecho visible desde décadas pasadas, por lo que ha sido un tema de interés de estudio para comprender el fenómeno y tomar las medidas pertinentes, siendo así que este primer capítulo aborda un breve escenario del comportamiento de los contaminantes atmosféricos, asimismo se presentan algunos estudios estadísticos para la ZMVM y otras ciudades mexicanas, y más adelante se hace énfasis en los métodos utilizados para la estimación y pronósticos de contaminantes realizados en otras ciudades del mundo.

1.1 Breve panorama de la contaminación del aire en la ZMVM

Diversos artículos como Bravo et al. (1992 y 2002) mencionan que la contaminación del aire en la ZMVM se ha hecho visible desde los años sesenta y ha ido en aumento; por lo que la expansión demográfica que comienza a darse durante la década de los ochenta, acelera el deterioro de la calidad del aire en esta metrópoli, siendo en este punto que comienza la medición de ciertos contaminantes, para tomar medidas pertinentes y combatir los problemas que en sus momentos se estaban presentando.

En 1986 en la ZMVM los niveles de concentración del O₃ comenzaron a aumentar, por lo que se generaron medidas para el control de la calidad del aire durante ese periodo, una propuesta fue el cambio de fórmula y componentes de la gasolina, sin embargo, ocasionó valores altos del O₃, que aún prevalecen, por ello se pugnaba por proponer estrategias realistas acorde a las características geográficas y urbano-industriales de la ZMVM.

En este contexto, la emisión del O₃ y otros contaminantes atmosféricos, llevaron a que la Ciudad de México realizara medidas de restricción de manejo desde 1989, convirtiéndose en la primera ciudad del mundo en implementar este tipo de políticas, sin embargo, investigaciones realizadas como en Lucas et al. (2017) sobre los días de prohibición de manejo arrojaron resultados desalentadores, ya que no había evidencia de que el programa hubiese mejorado la calidad del aire, por el contrario, se identificó que la contaminación del aire aumento 24 meses después de haberse realizado dicha imposición; pues ante la falta de transporte particular, los conductores utilizaron medios de transporte adicionales y pesar de expandirse el hoy no circula los sábados, solo se observó una disminución moderada de monóxido de carbono (CO) y óxido de nitrógeno (NO_x) los sábados por la tarde, pero no por la mañana.

De acuerdo con las normas de calidad del aire de la ZMVM los contaminantes que rebasan con mayor frecuencia los estándares son O₃ y PM₁₀, el resto de los contaminantes por lo general registran niveles dentro de la norma. Al respecto, Velazco et al. (2017) menciona que en marzo de 2016 las concentraciones de O₃ superaron la norma establecida en la Ciudad de México durante un periodo de estancamiento de tres días, por lo que 40% de los 5.4 millones

de automóviles particulares, incluyendo los nuevos modelos, fueron retirados de circulación en esos tres días, además las industrias experimentaron recortes en sus actividades y algunas actividades al aire libre fueron prohibidas.

Dicha medida fue puesta en marcha durante tres meses más, donde una vez a la semana cada auto quedaría fuera de circulación, sin embargo, fue poco efectiva, pues en fechas anteriores el Tribunal Supremo del país había descartado las restricciones vehiculares en base a la edad del coche, es decir, los vehículos tenían la posibilidad de circular todos los días de la semana, según sus condiciones mecánicas y de emisión de gases y no bajo la condicionante de su edad, lo que provocó la congestión del tráfico en un 20%.

Respecto a las propuestas para implementar el monitoreo que permita mejorar la condiciones de la calidad del aire Jaimes (2017) y Gong et al. (2017) mencionan que el O_3 es uno de los contaminantes fotoquímicos que son partes del denominado Smog¹, y que debido a su nivel y la forma en cómo se genera, proponen la implementación de un programa de monitoreo de los contaminantes precursores del O_3 , aprovechando la infraestructura de los centros de monitoreo, para poder caracterizar el O_3 generado y de esta forma implementar políticas públicas encaminadas al control de dicho contaminante atmosférico.

1.2 Análisis estadísticos sobre los contaminantes del aire en la ZMVM y otras ciudades de México

Ante el escenario anterior, la preocupación por el comportamiento de los contaminantes atmosféricos debido a los efectos que éstos pueden tener en el ambiente y el bienestar de la población, así como el panorama que presentan algunos de ellos, rebasando en ocasiones los niveles máximos permisibles por las normas en materia de salud ambiental, han dado lugar a análisis estadísticos sobre la estimación de su tendencia y pronóstico.

Al respecto, Huerta y Sansó (2007) ilustran el comportamiento de los valores extremos diarios del O_3 para el centro de la ZMVM, en el periodo de 1990 a 2002 a través de un método tanto para el tiempo como para el espacio, bajo un enfoque bayesiano, definiendo el componente temporal como un modelo lineal dinámico (dlm) y el elemento espacial impuesto mediante la matriz de evolución del dlm. Como resultado se tiene que los niveles máximos del O_3 decrecieron en ese periodo y en la región.

Por su parte, Gong y Ordires (2015) predicen las concentraciones máximas diarias del O_3 en la ZMVM, a través de cuatro modelos de inteligencia artificial: regresión lineal múltiple, redes neuronales, máquinas de vectores de soporte, *random forest* y dos técnicas de ensamble. Encuentran que los errores de predicción en relación con el día actual son de alrededor del 50%. Asimismo, con valores máximos del O_3 en la Ciudad de México,

¹ El smog es la mezcla de diversos compuestos gaseosos y aerosoles, de entre los que destaca el Ozono (O_3), Dióxido de Nitrógeno (NO_2) y Nitratos (NO_3).

Rodríguez et al. (2016) presentan un análisis de tendencias para este contaminante en las cinco regiones, en el periodo de 2001 a 2014, mediante un modelo de distribución generalizada de valores extremos, utilizando Open-BUGS/Winbugs. Los resultados muestran que los máximos mensuales del O₃ disminuyeron durante este periodo.

Más adelante, Aguilar y Reyes (2018) realizan un análisis a través de la transformación wavelet de Haar para valores extremos de O₃ y sus precursores (NO₂ y CO), para la Ciudad de México, en el periodo de 2015 a 2016, años en los que se implementaron cambios notorios en las restricciones de manejo en la ciudad; los resultados mostraron que los eventos multidiá del O₃ exhiben periodos T mayores a 4 días, mientras que para el NO₂ y CO son de T mayores a 2. Además, se demuestra que estos contaminantes atmosféricos están correlacionados grupal y multitemporalmente.

Por su parte, Yeongkwon et al. (2018) por medio de una regresión espacio-temporal estudian el comportamiento de las concentraciones para los contaminantes PM_{2.5}, PM₁₀, O₃, NO₂, CO y SO₂. A través de modelos que toman en consideración promedios por hora, diarios, mensuales, semestrales y anuales, deducen que se estiman resultados satisfactorios en los pronósticos, excepto para las series de PM₁₀, PM_{2.5} y SO₂.

En otras ciudades de México, el problema de la contaminación del aire y su monitoreo también se ha hecho visible, al respecto Hernández et al. (2004) presentan un análisis geoespacial del CO con promedios mensuales y anuales de 1995 a 2001, en Toluca (capital del Estado de México), utilizando el software Surfer 7.2 con el que se trazaron isóneas (curvas de nivel) de concentración del contaminante, determinando de esta manera cómo la dispersión del CO se favorece con los vientos y precipitaciones donde solamente se sobrepasa la norma establecida durante el invierno, situación que no ocurre en la ZMVM.

En el centro Occidente de México, para la Zona Metropolitana de Guadalajara, Hernández (2009) realiza un estudio para predecir los máximos diarios del O₃ monitoreados por siete estaciones, para el periodo de 1997 a 2006; por medio de la teoría de valores extremos comprueba que, con altas temperaturas y bajas velocidades de viento, se obtiene un aumento del O₃; además que al usar información no estacionaria y atmosférica, es posible predecir los máximos diarios del O₃ para de esta manera tener un monitoreo que permita proteger a la población de concentraciones peligrosas.

Al Noroeste de México (en Baja California), Corona y Rojas (2009) para la ciudad de Mexicali, realizaron un diagnóstico de la calidad del aire a través de datos de las concentraciones de los contaminantes, los cuales se interpolan con el programa Surfer 8. Se encontró que la calidad del aire (medida a través del número de días en que se rebasa la norma) durante el periodo de 1997 a 2005 no fue satisfactoria.

También en Cruz et al. (2013), para la ciudad de Hermosillo, Sonora, con muestreo de concentraciones de partículas suspendidas, se encontró que aquellas de metales pesados

estaban por debajo de los máximos permisibles (Pb - plomo, Cd - cadmio, Ni - níquel, Cu - cobre, Cr - cromo), sin embargo, la calidad del aire no era satisfactoria. Hernández et al. (2017) realizan estimaciones de la tendencia a largo plazo del O₃ en la Zona Metropolitana de Monterrey (ZMM), la comparan con la ZMVM y la de Guadalajara (ZMG), a través del análisis los promedios por hora, diarios y mensuales. Las estimaciones se elaboran con las herramientas openair WindRose, timeVariation y TheilSen del software R, para el periodo de 1993 a 2014; observan que la tendencia del O₃ para la ZMM es creciente en primavera, verano y otoño, en contraste con la ZMVM con tendencia a la baja, sin embargo, para la ZMG no se distingue una tendencia clara, en el periodo de referencia.

1.3 Otras ciudades del mundo

Hoy en día, la calidad del aire no es satisfactoria en diversas áreas urbanas del mundo, por lo que ha surgido la necesidad de realizar estudios que permitan establecer y anticipar el comportamiento de los contaminantes del aire. En este sentido, Gramsh et al. (2006) analizan las tendencias de contaminantes en Chile, mediante clústers para identificar patrones de comportamiento; en particular se examinan PM₁₀ y O₃, evidenciando que los niveles de concentraciones varían con la estacionalidad, para el caso de las PM₁₀ es más alto en invierno y el O₃ en verano. En los patrones reconocidos juega un papel importante las características topográficas y meteorológicas de las áreas evaluadas.

Más adelante, Jaramillo et al. (2007) hacen referencia a la importancia del estudio de pronóstico del O₃ en la ciudad de Cali, Colombia, y utilizan la metodología de Box-Jenkins, para el periodo de abril a julio de 2003. Se analizan 2496 datos semanales, en donde los primeros 2232 fueron utilizados para la estimación y el resto para corroborar los resultados del modelo. Su mejor modelo tuvo la capacidad predictiva de 8 horas por lo que se podrían generar medidas preventivas con éste.

Al norte de América, Camalier et al. (2007) describen y ejemplifican un modelo lineal generalizado (glm), con el que relacionan el O₃ con variables meteorológicas para inferir las tendencias del O₃ en 39 ciudades de Estados Unidos. Establecen un modelo por separado para cada una de las ciudades y proponen otro con Splines para las relaciones no lineales que existen entre los componentes meteorológicos, el O₃ y los cambios estacionales. De este análisis se evidencia que el O₃ aumenta cuando lo hace la temperatura y la humedad disminuye.

En Atlanta (Georgia, Estados Unidos), mediante la recolección de muestras de varios contaminantes, dentro de los que destacan: CO, CO₂, NO, NO₂, NO_y (óxidos de nitrógeno), PN (número de partículas) y BC (carbono blanco), mediante un monitoreo móvil y para detectar patrones espaciales, Brantley et al. (2019) realizaron un análisis de sus tendencias, aplicando una regresión por cuantiles con Splines cúbicos naturales. Se observaron

tendencias similares en el NO y NO_y, así como PN y BC, además se encontraron concentraciones de contaminantes diferenciadas por el tipo de ubicación.

Para cuatro zonas urbanas de dos ciudades europeas (Bruselas y Londres), Kumar y De Ridder (2010) realizan pronósticos de las concentraciones máximas diarias del O₃ mediante el modelo GARCH en asociación con el modelo FFT-ARIMA. Encontraron que determinados factores como el tráfico, representan un papel importante en la concentración del contaminante, pues la estructura del modelo GARCH depende del sitio, a diferencia del modelo FFY-ARIMA, por lo que el uso del modelo GARCH no solo mejora los intervalos de predicción, sino también hace más precisos los pronósticos de los máximos del O₃.

En Bulgaria, Georgieva et al. (2014) analizan la calidad del aire de 2011 a 2012 a través de Análisis factorial (AF) y de la metodología de Box-Jenkins, para las concentraciones de los contaminantes: NO, NO₂, NO_x, PM₁₀, SO₂ y O₃. Se encontró multicolinealidad y se establecieron tres factores F1={NO₂, NO, NO_x, PM10}, F2={O₃} y F3={SO}, explicando el 90.74 % de la varianza total. Por otra parte, se generaron modelos SARIMA para cada uno de los seis contaminantes, con un horizonte de pronóstico de 72 horas y se observó que los resultados mostraron ser precisos. Se detectó que, en el caso particular de las PM₁₀, se presentan concentraciones más altas en invierno, además de superar las normas nacionales y europeas.

Palomares et al. (2019) utilizan cinco métodos para predecir las concentraciones promedio de las PM₁₀, en tres áreas urbanas de Andalucía (España). Los métodos que emplean los dividen en paramétricos: modelo de persistencia y modelo de regresión lineal múltiple, y, no paramétricos: adaptive linear neuron, multilayer, and radial basis function. Los datos corresponden al periodo de 2005 a 2010 y se utilizan mediciones meteorológicas como variables exógenas para la conformación de los modelos. Se obtuvo que los pronósticos con los modelos no paramétricos fueron los mejores, por lo que incluir información sobre las condiciones meteorológicas mejora su capacidad predictiva. No obstante, tales modelos presentan una desventaja al requerir una gran cantidad de datos.

En Asia, para Hong Kong, So et al. (2007) consideran un estudio con relación al PM_{2.5}, donde se evalúan tendencias a largo plazo y variaciones espaciales, considerando carreteras, entornos urbanos y entornos rurales, donde se analizaron muestras recabadas cada seis días durante 12 meses, en tres sitios estratégicos. Entre varios hallazgos, se encontró que las variaciones estacionales fueron similares: altas en otoño e invierno y bajas en el verano; así como que el comportamiento del PM_{2.5} no fue similar en las áreas consideradas.

Chang y Yao (2008), reconocen una asociación entre el desarrollo económico y las emisiones de contaminantes atmosféricos en ciudades densamente pobladas que concentran el 20% del PIB, donde se identifican los desafíos inmediatos para comprender y controlar tal problemática. Por lo que hacen una revisión de la tendencia a largo plazo de los

contaminantes atmosféricos para el caso de tres ciudades de China, mediante análisis químicos y pruebas ANOVA, encontrando que el aumento de los contaminantes en las ciudades es semejante, es decir, alto en otoño e invierno y menor en verano.

Zamri et al. (2009) obtienen pronósticos del CO y NO₂ de concentraciones máximas mensuales, a través de la metodología de Box-Jenkins, los datos contemplan el periodo de 1997 a 2006 en cuatro zonas de Malasia. Destacan que en general, los pronósticos de 2016 no exceden los límites permisibles. La predicción de las PM_{2.5} en Pekín es vista en Aditya et al. (2018), donde se comparan varios modelos de machine learning y se opta por el de regresión logística. Se emplean datos diarios y se consideran las siguientes variables: temperatura, velocidad del viento, punto de rocío, presión. Asimismo, para fines de predicción se utiliza un modelo autorregresivo (AR) y se afirma que estos modelos son eficientes para predecir los niveles de PM_{2.5}.

Jaiswal et al. (2018), para Varanasi (India), estudian la tendencia anual del CO, NO₂, SO₂, PM_{2.5} y PM₁₀ de 2013 a 2016. Utilizan el test de Mann–Kendall y para realizar su pronóstico emplean un modelo ARIMA. Los resultados mostraron que las PM_{2.5}, CO, NO₂ y SO₂, tienen una tendencia decreciente a diferencia de las PM₁₀ que tienen una tendencia al alza.

Vita et al. (2018) desarrollan un algoritmo a través del método Fractional Kalman Filter (FKF) para pronosticar las concentraciones de los contaminantes: O₃, NO, NO₂, SO₂, PM_{2.5} y PM₁₀. Se parte del modelo “TAPM- CTM” (The Air Pollution Model and Chemical Transport Model), para medir las concentraciones de los contaminantes del aire. Comparan los resultados con el filtro de Kalman y se muestra que el modelo con FKF tiene mejor precisión.

Las metodologías que se han utilizado para la estimación de tendencias y pronósticos de los diferentes contaminantes atmosféricos han sido diversas de acuerdo con el objetivo previsto, siendo entre otras: glm, dlm, valores extremos, Box-Jenkins (ARIMA), clúster, Spline, enfoques bayesianos y machine learning. El método que aquí se propone aplicar, no tiene supuestos distribucionales y su implementación y mantenimiento se considera simple, donde la dinámica de los datos es la protagonista al hacer estimaciones tanto puntuales como por intervalos. Se utiliza exclusivamente la evidencia empírica de los contaminantes objeto de estudio, dejando así de lado variables adicionales como, por ejemplo, de tipo meteorológicas.

2 Algunas técnicas de suavizamiento

Una forma de visualizar la tendencia es mediante suavizamiento de la serie, al respecto el método de medias móviles, suavizamiento exponencial, Holt-Winters y Splines son técnicas que permiten obtener, a partir de la serie observada, una nueva serie que suaviza los efectos ajenos a la tendencia, de manera que se pueda observar ésta.

2.1 Media móvil

Este método de suavizamiento tradicional es uno de los más usados para describir la tendencia. Consiste en fijar un número k , preferentemente impar, y calcular los promedios de todos los grupos de k términos consecutivos de la serie. Se obtiene una nueva serie suavizada por media móviles de orden k . De este modo se tienden a anular las variaciones aleatorias. La fórmula está dada por

$$z_t = \frac{y_{t-k} + y_{t-k+1} + \dots + y_t + y_{t+1} + \dots + y_{t+k}}{2 * k + 1}$$

El suavizamiento de media móvil es muy fácil de aplicar, permite visualizar la tendencia de la serie. Pero tiene dos inconvenientes: (1) No es posible obtener estimaciones de la tendencia en extremos y (2) no entrega un medio para hacer predicciones. Si la serie presenta un efecto estacional de período k , es conveniente aplicar un suavizamiento de media móvil de orden k . En tal caso se elimina el efecto estacional, junto con la variación aleatoria, observándose solamente la tendencia (Ríos y Hurtado, 2008).

2.3 Suavizado exponencial

El objetivo del suavizado exponencial es predecir algún valor futuro x_{n+k} dado la historia pasada $\{x_1, x_2, \dots, x_n\}$ de las observaciones hasta el tiempo n . Se asume que no hay efectos estacionales o que en el proceso se han identificado y eliminado (Covpewartait y Metcalfe, 2009). La media del proceso puede cambiar de un paso del tiempo al siguiente, pero no se tiene información sobre la dirección probable de estos cambios. El modelo es:

$$x_t = u_t + w_t$$

donde u_t es la media no estacionaria del proceso en el tiempo t y w_t son las desviaciones aleatorias independientes con una media 0 y una desviación estándar σ . Dado que no hay una tendencia sistemática, una estimación intuitiva y razonable de la media, la media en el tiempo t es dada por el promedio ponderado de las observaciones en el tiempo y la estimación de la media en el tiempo $t-1$:

$$a_t = \alpha x_t + (1 - \alpha)x_{t-1} \quad 0 < \alpha < 1$$

a_t es la media móvil ponderada exponencialmente en el momento t , el valor de α determina la cantidad de suavizado y se conoce como el parámetro de suavizamiento, si α está cercano a 1 hay poco suavizado y es aproximado a x_t , esto solo sería apropiado si se espera que los cambios en el nivel medio fueran grandes en comparación con σ . En el otro extremo, un valor de α cercano 0 proporciona estimaciones altamente suavizadas y tiene poco en cuenta las observaciones más recientes, esto sería apropiado si los cambios en el nivel medio fueran pequeños en comparación con σ . Una cifra típica para $\alpha = 0.2$, ya que, en la práctica, generalmente, se espera que el cambio en la media entre el tiempo $t-1$ y el tiempo t sea menor que σ . Cómo se ha supuesto que no hay tendencia sistemática y que no hay efectos estacionales en el tiempo para cualquier tiempo, solo la media estimada en el tiempo n . la ecuación de estimación es:

$$\hat{x}_{t+k|n} = a_n \quad k = 1, 2, \dots$$

La ecuación $a_t = \alpha x_t + (1 - \alpha)x_{t-1}$ se puede reescribir en otras dos fórmulas útiles. En primer lugar, se puede escribir la suma de a_{t+1} y una proporción del error de la estimación un paso adelante, $x_t - a_{t-1}$

$$a_t = \alpha(x_t - a_{t-1}) + a_{t-1}$$

Segundo, por sustitución hacia atrás se tiene que:

$$a_t = \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \dots$$

Cuando se reescribe esta forma, se ve que a_t es una combinación lineal de las observaciones actuales y pasadas, con un mayor peso a las observaciones más recientes. La restricción de $0 < \alpha < 1$ garantiza que los pesos $\alpha(1 - \alpha)^i$ sean más pequeños al incrementar i . Cabe notar que los ponderadores tienen la forma de una serie geométrica y la suma de la serie infinita es la unidad. Se puede evitar la regresión infinita especificando $a_1 = x_1$ en la ecuación $a_t = \alpha x_t + (1 - \alpha)x_{t-1}$

Para cualquier α dada, el modelo en la ecuación $a_t = \alpha(x_t - a_{t-1}) + a_{t-1}$ junto con el inicio se puede usar para calcular a_t para $t=2, 3, \dots, n$. los errores de predicción de un paso adelante están dados por $e_t = x_t - \hat{x}_{t|t-1} = x_t - a_{t-1}$

2.4 Método de Holt-Winters

El método de Holt-Winters fue sugerido por Holt y Winters (Cowpertwait y Metcalfe, 2009) y usa promedios móviles ponderados exponencialmente para actualizar las estimaciones de la media ajustadas por estacionalidad (nivel), pendiente y estacionalidad.

Holt-Winters modelo aditivo

El método generaliza la ecuación $a_t = \alpha x_t + (1 - \alpha)x_{t-1}$ y la forma estacional aditiva de sus ecuaciones de actualización para una serie $\{x_t\}$ con periodo p es

$$a_t = \alpha(x_t - s_{t-p}) + (1 - \alpha)(a_{t-1} - b_{t-1})$$

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$$

$$s_t = \gamma(x_t - a_t) + (1 - \gamma)s_{t-p}$$

donde a_t , b_t y s_t son las estimaciones del nivel, pendiente y el efecto estacional en el tiempo, respectivamente y α , β y γ son parámetros de suavizamiento. La primera ecuación de actualización toma un promedio ponderado de la última observación, con el estimado existente del efecto estacional apropiado extraído y el pronóstico del nivel hecho un paso atrás.

El pronóstico de un paso adelante del nivel es la suma de las estimaciones del nivel y pendiente en el momento del pronóstico. Una elección típica del ponderador α es 0.2. La segunda ecuación el promedio ponderado de las estimaciones previas y de la última estimación de la pendiente, la cual es la diferencia en el nivel estimado en el tiempo t y el nivel estimado en el tiempo $t-1$. Nótese que la segunda ecuación solo se puede utilizar después de que se haya aplicado la primera ecuación para conseguir a_t . Finalmente se tiene la estimación del efecto estacional, desde las diferencias entre la observación y la estimación del nivel y se toma el promedio ponderado de las últimas estimaciones del efecto estacional para la estacionalidad, que se hizo en el tiempo $t-p$. Generalmente los ponderadores β y γ son 0.2. Las ecuaciones de actualización pueden iniciar con $a_1 = x_1$ y la pendiente b_1 y los efectos estacionales s_1, \dots, s_p , obtenidos desde la experiencia, estimados desde los datos o establecido en 0.

La ecuación de pronóstico para x_{n+k} hecha después de la observación en el tiempo n es:

$$\hat{x}_{t+k|n} = a_n + kb_n + s_{n+k-p} \quad k \leq p$$

donde a_n es el nivel estimado y b_n es la pendiente estimada, por lo que $a_n + kb_n$ es el nivel esperado en el tiempo $n + k$ y s_{n+k-p} es la estimación ponderada exponencialmente del efecto estacional realizado en el momento $n=k-p$

Holt-Winters modelo multiplicativo

El modelo Holt-Winters multiplicativo es similar al aditivo. Multiplica la previsión con la tendencia por la estacionalidad. Es más adecuado para datos con tendencia y estacionalidad que aumenta a lo largo del tiempo. Su fórmula recursiva es como sigue

$$a_n = \alpha \left(\frac{x_n}{s_{n-p}} \right) + (1 - \alpha)(a_{n-1} - b_{n-1})$$

$$b_n = \beta(a_n - a_{n-1}) + (1 - \beta)b_{n-1}$$

$$s_n = \gamma \left(\frac{x_n}{a_n} \right) + (1 - \gamma)s_{n-p}$$

La ecuación de pronóstico para x_{n+k} hecha después de la observación en el tiempo n es

$$\hat{x}_{t+k|n} = (a_n + kb_n)s_{n+k-p} \quad k \leq p$$

2.5 Splines

De acuerdo con Infante y Rey (2015) la palabra inglesa Spline denota un instrumento flexible usado en dibujo técnico que sirve para trazar curvas suaves. Desde el punto de vista matemático, una función Spline en un intervalo $[a, b]$ está formada por polinomios definidos en subintervalos de $[a, b]$ obedeciendo a ciertas condiciones de regularidad. Es decir

Sea $\Delta = \{a = x_0 < x_1 < \dots < x_n = b\}$ una partición del intervalo $[a, b]$. $S_\Delta: [a, b] \rightarrow \mathbb{R}$ es una función Spline de orden $k \in \mathbb{N}$ asociada a Δ si $S_\Delta \in C^{k-1}([a, b])$ y S_Δ coincide en cada intervalo $[x_i, x_{i+1}]$, $i = 0, 1, \dots, n - 1$ con un polinomio de grado $\leq k$.

Cabe destacar que cuando el valor $k=1$ se obtiene una función lineal a trozos, cuando $k=3$ es una función Spline cúbica. Siendo así que los Splines dependen de tres elementos: 1) Grado del polinomio, 2) Número de nodos y 3) Localización de los nodos.

Un Spline de grado 1 se puede definir por (León y Constanzo, 2006):

$$s(x) = \begin{cases} S_0(x) = a_0x + b_0, & x \in [x_0, x_1) \\ S_1(x) = a_1x + b_2, & x \in [x_1, x_2) \\ \vdots \\ S_{n-1}(x) = a_{n-1}x + b_{n-2}, & x \in [x_{n-1}, x_n) \end{cases}$$

Splines cúbicos

El Spline cubico ($k=3$) es un conjunto de polinomios de tercer grado que se genera a partir de un conjunto de puntos. Es decir, dada una función f en $[a, b]$ y $n+1$ puntos distintos $a = x_0 < x_1 < \dots < x_n = b$,

Para cada $i = 0, 1, \dots, n - 1$, $S_i(x)$ es un polinomio cúbico, en el subintervalo $[x_i, x_{i+1}]$, por tanto

$$s(x) = \begin{cases} S_0(x) = S_0, & x \in [x_0, x_1) \\ S_1(x) = S_1, & x \in [x_1, x_2) \\ \vdots \\ S_{n-1}(x) = S_{n-1}, & x \in [x_{n-1}, x_n) \end{cases}$$

Los polinomios S_{i-1} y S_i interpolan el mismo valor en el punto x_i , es decir se cumple:

a) Condición de continuidad (en nodos internos)

$$S_i(x_{i+1}) = S_{i+1}(x_{i+1}) \text{ para } i = 0, 1, \dots, n-2$$

b) Condición de suavidad (en nodos interiores)

$$S'_{i+1}(x_{i+1}) = S'_i(x_{i+1}) \text{ para } i = 0, 1, \dots, n-2$$

$$S''_{i+1}(x_{i+1}) = S''_i(x_{i+1}) \text{ para } i = 0, 1, \dots, n-2$$

c) Obsérvese que se obtienen en total $4n-2$ ecuaciones, lo que significa que para determinar el Spline $s(x)$ de forma única se requiere imponer 2 condiciones adicionales. Dichas condiciones suelen imponerse sobre los extremos del intervalo siendo las más habituales

$$\text{Spline cubico natural } S''_0(x_0) = S''_{n-1}(x_n) = 0$$

$$\text{Spline cubico sujeto } S'_0(x_0) = f'(x_0) \text{ y } S'_{n-1}(x_n) = f'(x_n)$$

Aplicando las condiciones de continuidad del Splines y de las primeras y segundas derivadas, es posible encontrar la expresión analítica del Spline:

$$S_i(x) = \frac{z_i}{6h_i}(x_{i+1} - x)^3 + \frac{z_{i+1}}{6h_i}(x - x_i)^3 + \left(\frac{y_{i+1}}{h_i} + \frac{z_{i+1}h_i}{6}\right)(x - x_i) + \left(\frac{y_i}{h_i} + \frac{z_i h_i}{6}\right)(x_{i+1} - x)$$

En la expresión anterior, $h_i = x_{i+1} - x_i$ y z_0, z_1, \dots, z_n son incógnitas. Para determinar sus valores, utilizamos las condiciones de continuidad que deben cumplir estas funciones.

$$h_{i-1}z_{i-1} + 2(h_i + h_{i-1})z_i + h_i z_{i+1} = \frac{6}{h_{i-1}}(y_{i+1} - y_i) - \frac{6}{h_{i-1}}(y_i - y_{i-1})$$

La ecuación anterior, con $i = 1, 2, \dots, n-1$ genera un sistema de $n-1$ ecuaciones lineales con $n+1$ incógnitas z_0, z_1, \dots, z_n . Podemos elegir z_0 y z_1 de forma arbitraria y resolver el sistema de ecuaciones resultante para obtener los valores de z_1, z_2, \dots, z_{n-1} . Una elección especialmente adecuada es hacer $z_0 = z_1 = 0$ (Spline cubico natural) y el sistema de ecuaciones lineal expresado en forma matricial es:

$$\begin{bmatrix} u_1 & h_1 & & & & \\ h_2 & u_2 & h_2 & & & \\ & h_2 & u_3 & h_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & h_{n-3} & u_{n-2} & h_{n-2} \\ & & & & h_{n-2} & u_{n-1} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{n-2} \\ z_{n-1} \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_{n-2} \\ v_{n-1} \end{bmatrix}$$

donde

$$\begin{aligned} h_i &= x_{i+1} - x_i \\ 2(h_i + h_{i-1}) - \frac{h_{i-1}^2}{u_{i-1}} \\ b_i &= \frac{6}{h_i} (y_{i+1} - y_i) \\ v_i &= b_i - b_{i-1} - \frac{h_{i-1} v_{i-1}}{u_{i-1}} \end{aligned}$$

Este sistema tridiagonal se puede resolver mediante eliminación gaussiana sin pivoteo. El valor del Spline S en un punto x cualquiera interpolado se puede calcular de forma eficiente empleando la siguiente expresión:

$$S_i(x) = y_i + (x - x_i)[c_i + (x - x_i)[B_i + (x - x_i)A_i]$$

$$\begin{aligned} A_i &= \frac{1}{6h_i} (z_{i+1} - z_i) \\ B_i &= \frac{z_i}{2} \\ c_i &= -\frac{h_i}{6} z_{i+1} - \frac{h_i}{3} z_i + \frac{1}{h_i} (y_{i+1} + y_i) \end{aligned}$$

Spline con Penalizaciones (P-Splines)

El suavizado por Splines penalizados fueron presentados originalmente por O'Sullivan en 1986 y fueron Eilers y Marx en 1996 quienes le dieron el nombre de suavizado P-Spline. La idea es estimar una función de regresión suave, asumiendo una forma paramétrica funcional construida a través de una base de funciones de alta dimensión. La dimensión de la base se elige para lograr la flexibilidad deseada, mientras que los coeficientes de la base se penalizan para garantizar la suavidad de las estimaciones resultantes. La dimensión de la base tiene poca influencia en el ajuste, por lo que a lo más 40 nodos son posibles de definir para todos los tamaños de muestra y todas las funciones de suavidad sin muchas oscilaciones (Krivobokova et al., 2008).

De tal manera que la combinación de B-Spline con penalización basada en diferencias (entre los coeficientes estimados) es lo que se conoce como P-Splines. (véase Eilers y Marx, 1996 y Eilers et al., 2015).

- B-Splines

Los B-Spline son construidos desde piezas polinomiales, unidas a ciertos valores de x (nodos). En los puntos de unión no solo las ordenadas de los polinomios coinciden, sino también sus primeras derivadas son iguales (aunque no sus segundas derivadas)

Sus propiedades son:

- Un B-Spline de grado p consiste en $p+1$ trozos de polinomios de orden p
- Se unen en p nodos internos
- En los puntos de unión las derivadas hasta el orden $p-1$ son continuas
- El B-Splines es positivo en el dominio expandido por $p+2$ nodos y 0 en el resto
- Excepto en los extremos, se solapa con $2p$ trozos de polinomios de sus vecinos y
- Para cada valor de x , $p+1$ B-Spline son no nulos

Se sugiere una forma simple para derivar los B-Splines de acuerdo con Boor:

$$\begin{aligned} h \sum_j a_j B_j'(x; q) &= \sum_j a_j B_j(x; q-1) - \sum_j a_{j+1} B_{j+1}(X; q-1) \\ &= - \sum_j \Delta a_{j+1} B_j(x; q-1) \end{aligned}$$

Donde h es la distancia entre los nodos y $\Delta a_j = a_j - a_{j-1}$

Por inducción se encuentra su segunda derivada:

$$h \sum_j a_j B_j''(x; q) = \sum_j \Delta^2 a_j B_j(x; q-2)$$

Donde $\Delta^2 a_j = \Delta \Delta a_j = a_j - 2a_{j-1} + a_{j-2}$

- Penalizaciones

Considera la regresión de m puntos de datos $(x_i; y_i)$ de un conjunto de n B-Splines $B_j(\cdot)$. La función objetivo de mínimos cuadrados a minimizar es

$$S = \sum_{i=1}^m \{y_i - \sum_{j=1}^n a_j B_j(x_i)\}^2$$

Para la penalización (alto orden) se proponen las diferencias de los coeficientes de los B-Splines adyacentes, es decir:

$$S = \sum_{i=1}^m \left\{ y_i - \sum_{j=1}^n a_j B_j(x_i) \right\}^2 + \lambda \sum_{j=k+1}^n (\Delta^k a_j)^2$$

Esta forma reduce la dimensionalidad del problema a n , el número de B-Splines, en lugar de m , el número de observaciones, teniendo aún λ el parámetro para el control continuo de la suavidad.

Por lo que para minimizar la ecuación anterior se propone reescribirla como:

$$B^T y = (B^T B + \lambda D_k^T D_k) a$$

donde D_k es la representación matricial del operador diferencia de Δ^k y los elementos de B son $b_{ij} = B_j(x_i)$, cuando $\lambda = 0$ se tienen una ecuación normal estándar de regresión lineal con base B-Splines, cuando $k=0$ es un caso especial de regresión ridge. Cuando $\lambda > 0$ la penalización solo influye en la diagonal y subdiagonal k (ambos lados de la diagonal principal) del sistema de ecuaciones. Introduciendo Modelo Lineal Generalizado (GLM), la forma de verisimilitud penalizada es:

$$B^T \tilde{W} (y - \tilde{u}) + B^T \tilde{W} B \tilde{a} = (B^T \tilde{W} B + \lambda D_k^T D_k) a$$

donde \tilde{a} y \tilde{u} son aproximaciones actualizadas de la solución y \tilde{W} es una matriz diagonal de ponderadores $w_{ii} = \frac{1}{v_i} \left(\frac{\partial u_i}{\partial n_i} \right)^2$, donde v_i es la varianza de y_i , dado u_i

Selección del parámetro de suavizamiento

El parámetro de suavizamiento λ en los P-Splines tiene por objeto controlar la suavidad de la curva, penalizando los coeficientes que están muy separados entre sí y cuanto mayor sea λ , más se aproximarán los coeficientes a cero, de modo que si $\lambda \rightarrow \infty$ se aproxima a un ajuste polinómico. Por el contrario, cuando $\lambda \rightarrow 0$ se estará utilizando mínimos cuadrados ordinarios.

Para elegir el valor óptimo de λ , existen diversos criterios automáticos a utilizar, los más populares son: criterio de información Akaike (AIC), validación cruzada (CV), validación cruzada generalizada (GCV) y criterio de información bayesiano (BIC). Cabe mencionar que no necesariamente estos criterios proporcionan valores apropiados de los parámetros de suavizamiento para realizar comparativos en las series de tiempo.

AIC

La idea básica del AIC es corregir la verosimilitud de un modelo ajustado por el número efectivos de parámetros y se define como

$$AIC(\lambda) = dev(y; a, \lambda) + 2 * dim(a, \lambda)$$

donde $dim(a, \lambda)$ es la dimensión (efectiva) del vector de parámetros, a , y $dev(y; a, \lambda)$ es la devianza, para poder obtener la dimensión efectiva se propone $dim(a) = tr(H)$, cabe notar que $tr(H) = n$ cuando $\lambda = 0$ como en una regresión lineal estándar (no singular).

Como $tr(AB) = tr(BA)$ entonces si

$$tr(H) = tr \left\{ B(B^T W B + \lambda D_k^T D_k)^{-1} B^T W \right\} = tr \left\{ (B^T W B + \lambda D_k^T D_k)^{-1} B^T W B \right\}$$

Por lo que

$$AIC = 2 \log \left(\sum_{i=0}^n (y_i - \hat{y}_i)^2 \right) - 2 \log(n) + 2 \log(tr(H))$$

CV

Otra alternativa para hallar la λ óptima es a través de validación cruzada, la cual se define como

$$CV = \sum_{i=0}^m \left(\frac{y_i - \hat{y}_i}{1 - h_{ii}} \right)^2$$

donde h_{ii} son los elementos de la diagonal de la matriz H .

GCV

Bajo el criterio GCV la mejor λ será aquella que minimice el valor:

$$GCV = \sum_{i=0}^n \frac{(y_i - \hat{y}_i)^2}{n - tr(H)}$$

BIC

El criterio de información bayesiano (BIC), se define como:

$$BIC = -2LL + k\log(n)$$

donde LL es la log-verosimilitud del modelo, k es el número de parámetros independientes y n es el tamaño de muestra. La λ que obtenga el valor BIC más bajo será la mejor.

3 Metodología

Maravall (1993) documenta varios métodos para la estimación de tendencias. Dentro de ellos se tiene el filtro de Hodrick y Prescott (1997), mismo que no requiere de la aplicación de un modelo formal estadístico para la estimación de tendencias y pronósticos a diferencia de por ejemplo el modelo ARIMA o los modelos estructurales (Guerrero, 2008).

El filtro de Hodrick y Prescott (1997) tiene su origen en el llamado método Whittaker-Herderson, fue usado primero por actuarios para suavizar tablas de mortalidad y también fue de utilidad en astronomía. Los autores lo definen como las fluctuaciones recurrentes en la actividad real respecto a una tendencia. Las fluctuaciones son desviaciones respecto a un camino suave pero variable (tendencia), el cual es posible estimar mediante un proceso de ajuste a una curva suave. En este trabajo, para el análisis de las tendencias y sus pronósticos de los contaminantes seleccionados de la ZMVM no se desestacionalizan tales series de tiempo y se trabaja exclusivamente con la tendencia de los datos.

3.1 Suavizamiento controlado y pronósticos

El enfoque utilizado para el filtro de Hodrick y Prescott (1997) a través del filtro de Kalman, se sigue como lo expresa Guerrero (2007 y 2008). Sea un conjunto de datos observados en el tiempo, técnicamente una serie de tiempo, compuesta por la suma de una tendencia no observable más un componente aleatorio,

$$y_t = \tau_t + \eta_t, \text{ para } t = 1, \dots, N$$

donde τ_t es la tendencia, η_t es el componente aleatorio del cual no se asume distribución específica y N el número de datos. A través del método de suavizamiento controlado se reduce gradualmente el componente aleatorio, induciendo a una suavidad, y con él se estiman valores suavizados utilizando un índice mismo que se expondrá más adelante. El filtro es sencillo de aplicar para cualquier serie de datos. La estructura del filtro proviene de la siguiente expresión

$$\min_{\tau} M(\lambda) = \sum_{t=1}^N (y_t - \tau_t)^2 + \lambda \sum_{t=3}^N [(\tau_t - \tau_{t-1}) - (\tau_{t-1} - \tau_{t-2})]^2$$

donde λ es el parámetro de suavizamiento que penaliza por la suavidad; en el primer término se busca la bondad de ajuste de la tendencia con los datos observados, en tanto que el segundo a la inducción de la suavidad. Cuando el parámetro, se acerca a cero, $\lambda \rightarrow 0$, la tendencia estimada se aproxima a los datos observados, y, cuando se hace tender al parámetro a infinito, $\lambda \rightarrow \infty$, la tendencia tiende a una recta. De manera matricial se tiene

$$\min_{\tau} M(\lambda) = (\mathbf{y} - \boldsymbol{\tau})' (\mathbf{y} - \boldsymbol{\tau}) + \lambda (K_2 \boldsymbol{\tau})' (K_2 \boldsymbol{\tau})$$

donde $\mathbf{y} = (y_1, \dots, y_N)'$, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N)'$ y K_2 es la matriz de $(N + 2) \times N$ de segundas diferencias dada por

$$K_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 & 1 \end{bmatrix}$$

La solución al problema está dada por

$$\hat{\boldsymbol{\tau}} = (I_N + \lambda K_2' K_2)^{-1} \mathbf{y}$$

donde $\hat{\boldsymbol{\tau}}$ representa la tendencia estimada dado un valor de λ , en tanto que la varianza de $\hat{\boldsymbol{\tau}}$, en términos de error cuadrático medio, está dada por $Var(\hat{\boldsymbol{\tau}}) = (I_N + \lambda K_2' K_2)^{-1} \sigma_{\eta}^{-2}$. Con información de la matriz $Var(\hat{\boldsymbol{\tau}})$ se obtienen los intervalos de las respectivas estimaciones.

Dado que la segunda derivada de $M(\lambda)$ evaluada en $\tau = \hat{\boldsymbol{\tau}}$ es una matriz simétrica y positiva, se tiene que $\hat{\boldsymbol{\tau}} = (I_N + \lambda K_2' K_2)^{-1} \mathbf{y}$ produce un mínimo. Se debe notar que para obtener $\hat{\boldsymbol{\tau}}$, la matriz de $N \times N$ tiene que ser invertida, por lo que este cálculo puede causar inestabilidad y falta de precisión de la solución numérica cuando N es grande, además para obtener $(I_N + \lambda K_2' K_2)^{-1}$ se requiere que no haya datos faltantes en la serie. En este contexto, el filtro de Hodrick y Prescott estimado a través del filtro de Kalman es una solución, en la que se requiere formular un modelo espacio-estado (véase anexo B) como sigue

$$x_t = A_t x_{t-1} + w_t, y_t = c_t' x_t + \eta_t, \text{ con}$$

$$x_t = \begin{pmatrix} \tau_t \\ \tau_{t-1} \end{pmatrix}, A_t = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}, c_t' = (1 \quad 0) \text{ y } w_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}$$

donde ε_t y η_t son dos errores aleatorios independientes de media cero, no correlacionados e idénticamente distribuidos con $var(\varepsilon_t) = \sigma_\varepsilon^2$ y $var(\eta_t) = \sigma_\eta^2$. Por lo que la ecuación de estado tiene la siguiente forma

$$\tau_t = 2\tau_{t-1} - \tau_{t-2} + \varepsilon_t$$

Por otra parte, en Guerrero (2007) el pronóstico puntual de la tendencia, $\hat{\tau}_N(h)$, para el periodo h de τ_{N+h} con origen en N para $h \geq 1$ y $d = 2$, es decir con el filtro de Hodrick y Prescott, se encuentra expresado por medio de

$$\hat{\tau}_N(h) = [h(h+1)/2]\mu + (h+1)\tau_N - h\tau_{N-1}$$

donde se asume $\mu = 0$. Asimismo, para la obtención de los pronósticos por intervalos, se consideran todos los datos disponibles y se recalcula tanto la tendencia como su respectivo intervalo de estimación.

3.2 Índice de suavidad

El índice, propuesto por Guerrero (2008), que mide la suavidad de la tendencia está dado por la siguiente expresión

$$S(\lambda; N) = 1 - \text{tr}[(I_N + \lambda K_2' K_2)^{-1}]/N$$

donde $\text{tr}(\cdot)$ representa la traza de la matriz. Nótese que cuando $\lambda \rightarrow 0$ la suavidad $S(\lambda; N) \rightarrow 0$, en tanto que cuando $\lambda \rightarrow \infty$, se tiene que $S(\lambda; N) \rightarrow 1$. Es decir, el índice está comprendido entre cero y uno. El índice por practicidad se puede expresar en términos porcentuales. Se puede demostrar que el máximo valor del índice de suavidad alcanzable para un conjunto de N datos, está dado por $1 - 2/N$. En particular, dada la cantidad de datos por contaminante disponibles para este estudio $N = 3950$, la máxima suavidad a la que se puede aspirar es de 99.95%.

Nótese que diferentes índices de suavidad, $S(\lambda; N)$, generan diferentes tendencias estimadas y por lo tanto no existe una única tendencia, ni bajo dicha óptica una tendencia *óptima*; de hecho, se podrían estimar infinidad de tendencias a partir de un único conjunto de datos específicos. Con este método, se puede estimar desde una tendencia que sea exactamente el conjunto de datos observados hasta otra que sea una línea recta en el límite. El problema se centra en decidir qué porcentaje de suavidad es el apropiado. Esta técnica se ha utilizado para series de tiempo en economía y en temas de demografía, sin embargo, no se tienen ubicados antecedentes de su uso en temas de contaminantes.

Para decidir el porcentaje de suavidad a través de dicho índice, en Guerrero et al. (2017), a partir de algunas particularidades, se sugiere un determinado índice de suavidad con base en la estructura de los datos. Sin embargo, los tamaños muestrales de las series de tiempo son

muy distintas. En este documento, se cuenta con 3950 datos por serie y en dicha referencia son solo de 48, 100 y a lo más 400 observaciones. Adicionalmente, las aplicaciones que ahí se ilustran poseen una variabilidad sustantivamente menor a la que se observa en las series aquí analizadas. De hecho, debe notarse que el índice, $S(\lambda; N)$, depende de dos parámetros: λ y N . Por lo que, a mayor tamaño de muestra mayor índice de suavidad, es decir cuando $N \rightarrow \infty$ se tiene que $S(\lambda; N) \rightarrow 1$. Con esto queda de manifiesto que con la cantidad de datos aquí empleada se requieren suavidades significativamente grandes.

La propuesta para imponer un índice de suavidad pertinente en este trabajo consiste en elegir aquel que logre que las tendencias estimadas prescindan de comportamientos afectados por observaciones extremas, es decir se impone un índice de suavidad elevado tal que en las tendencias estimadas se mitigen dichos efectos. En el caso de que se quisiera detectar un cambio en la variabilidad del respectivo contaminante, se podría hacer un análisis segmentado como se propone en Guerrero y Silva (2015). En concreto, se impondrá un porcentaje de suavidad cercano a 1 por la cantidad de datos, con lo cual se superará la posible generación de tendencias espurias. Se utiliza entonces, después de múltiples ensayos, para la estimación de las tendencias de los contaminantes, un índice de suavidad de 99.91%, que corresponde a un parámetro de suavizamiento de $\lambda = 1 \times 10^{11}$.

Por otra parte, para la estimación de $Var(\hat{\tau})$ y $S(\lambda; N)$ es preciso mencionar que los efectos de las eventuales discrepancias en cuanto a calidad diferenciada de los datos que se emplean, se mitigan desde el punto de vista estadístico gracias a que se estandarizan las tendencias estimadas con el empleo de un índice de suavidad idéntico para todas las series de contaminantes, elemento que no necesariamente ocurre al emplear otras técnicas estadísticas, donde se tengan criterios automáticos para la selección del parámetro de suavizamiento. También se asume que la calidad de los datos de cada uno de los contaminantes es la misma en toda la RAMA.

3.3 Medidas de precisión de los pronósticos

La precisión de los pronósticos de los contaminantes se valora a partir de las siguientes medidas de precisión como lo señala Petris et al. (2007),

Desviación Absoluta Media (MAD).

$$MAD = \frac{1}{N} \sum_{t=1}^N |e_t|$$

Error Cuadrático Medio (MSE)

$$MSE = \frac{1}{N} \sum_{t=1}^N e_t^2$$

Porcentaje de Error Medio Absoluto (MAPE)

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^N \frac{|e_t|}{y_t}$$

donde $e_t = y_t - \hat{y}_t$, con $\hat{y}_t = \hat{\tau}$, que se refiere al error de pronóstico en el periodo t , y_t y \hat{y}_t , representan el valor real y el valor del pronosticado de la tendencia en el periodo t .

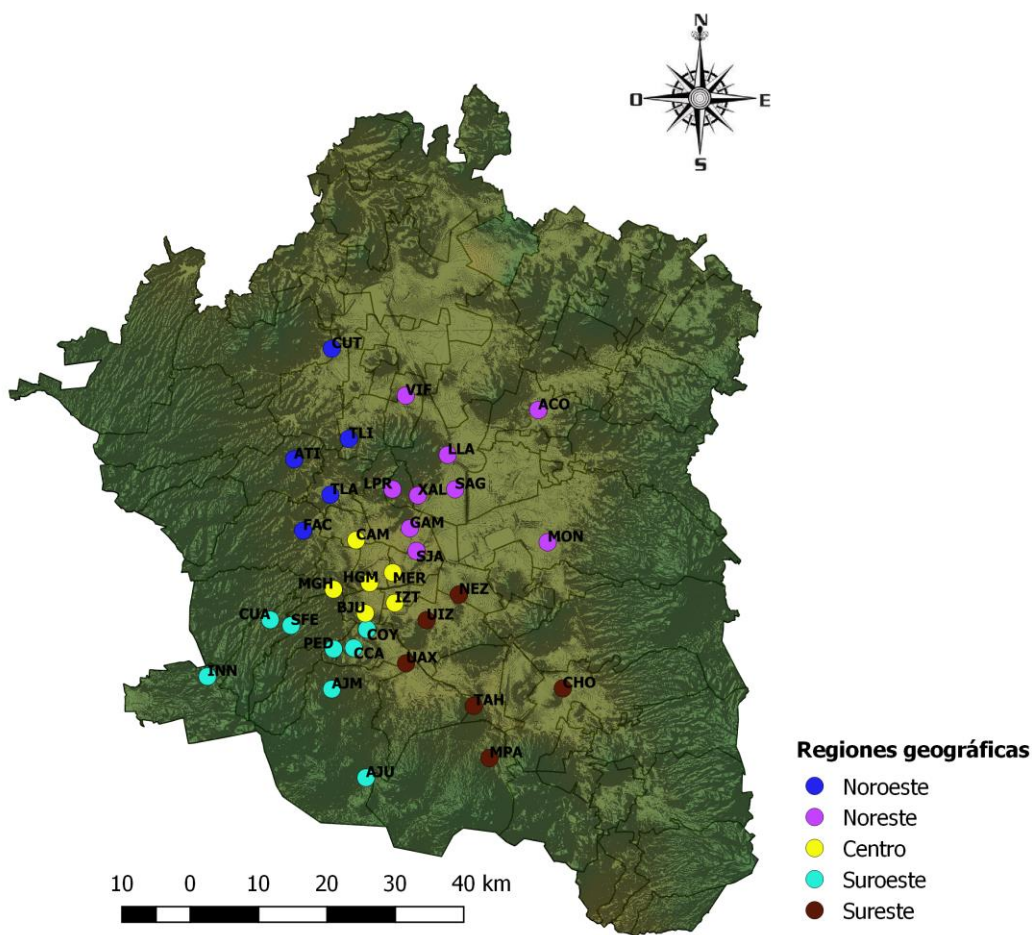
3.4 Delimitación de la ZMVM y ubicación de los centros de monitoreo

El Sistema de Monitoreo Atmosférico (SIMAT) de la Ciudad de México cuenta actualmente con seis diferentes sistemas de monitoreo:

- Red Automática de Monitoreo Atmosférico (RAMA): capta las concentraciones de los contaminantes del aire cada hora.
- Red de Meteorología (REDMET): contienen información de los parámetros meteorológicos que se registran cada hora.
- Red Manual de Monitoreo Atmosférico (REDMA): obtiene las concentraciones de los contaminantes muestreados durante 24 horas, cada 6 días.
- Red de Depósito Atmosférico (REDDA): presenta información de las muestras que se colectan semanalmente en época de lluvias.
- Red de Radiación Solar (RADIACIÓN): tiene información de los parámetros meteorológicos registrados cada hora.
- Presión Atmosférica: presenta información sobre la presión atmosférica que se registrada cada hora.

En este estudio, debido a que el comportamiento de los contaminantes atmosféricos no es homogéneo para toda la ZMVM, se dividió la metrópoli en cinco regiones: noroeste, noreste, centro, suroeste y sureste. Dentro de las cuales cada una de ellas se tienen unidades de monitoreo, en total 34, que gracias a la información del SIMAT y a través de la RAMA se pudieron ubicar, como se muestra en la siguiente Figura

Figura 1 Ubicación de las unidades de monitoreo de la ZMVM según región geográfica



Fuente: Elaboración propia.

Se identificó el tipo de contaminante que es monitoreado en cada uno de los centros, pero debido a que estos no son medidos de manera constante, ya sea porque inicia su medición en alguna estación o se abre un nuevo centro, se realizó una evaluación de cada unidad de monitoreo y contaminante por región. Así se tomaron en cuenta solamente aquellos que registran la información durante los diez años de estudio, es decir, desde 2008 hasta 2018. La tabla 1 presenta por regiones el catálogo de las 20 estaciones que monitorean los contaminantes seleccionados.

Tabla 1 Catálogo de estaciones de monitoreo por región geográfica

Región	Estación	Abreviatura	Ubicación	Contaminante
Noroeste	FES Acatlán	FAC	Naucalpan de Juárez, Estado de México	O ₃ , NO ₂ , NO _x , NO, SO ₂ , CO, PM ₁₀
	Tlalnepantla	TLA	Tlalnepantla de Baz, Estado de México	O ₃ , NO ₂ , NO _x , NO, SO ₂ , CO, PM ₁₀ , PM _{2.5}
	Atizapán	ATI	Atizapán de Zaragoza, Estado de México	NO ₂ , NO _x , NO
	Tultitlán	TLI	Tultitlán, Estado de México	NO ₂ , NO _x , NO, SO ₂ , CO, PM ₁₀
Noreste	Acolman	ACO	Acolman, Estado de México	O ₃ , SO ₂ ,
	Montecillo	MON	Texcoco, Estado de México	O ₃
	San Agustín	SAG	Ecatepec de Morelos, Estado de México	O ₃ , NO ₂ , NO _x , NO, SO ₂ , CO, PM ₁₀ , PM _{2.5}
	Xalostoc	XAL	Ecatepec de Morelos, Estado de México	O ₃ , NO ₂ , NO _x , NO, SO ₂ , CO, PM ₁₀ , PM _{2.5}
	Villa de las Flores	VIF	Coacalco de Berriozábal, Estado de México	NO ₂ , NO _x , NO, SO ₂ , CO, PM ₁₀
	Los Laureles	LLA	Ecatepec de Morelos, Estado de México	SO ₂
	La Presa	LPR	Tlalnepantla de Baz, Estado de México	SO ₃
	San Juan de Aragón	SJA	Gustavo A. Madero, CDMX	PM _{2.5}
Centro	Merced	MER	Venustiano Carranza, CDMX	O ₃ , NO ₂ , NO _x , NO, SO ₂ , CO, PM ₁₀ , PM _{2.5}
	Iztacalco	IZT	Iztacalco, CDMX	O ₃ , NO ₂ , NO _x , NO, SO ₂ , CO, PM ₁₀
	Camaronés	CAM	Azcapotzalco, CDMX	PM _{2.5}
Suroeste	Cuajimalpa	CUA	Cuajimalpa de Morelos, CDMX	O ₃
	Pedregal	PED	Álvaro Obregón, CDMX	O ₃ , NO ₂ , NO _x , NO, SO ₂ , CO, PM ₁₀
	Coyoacán	COY	Coyoacán, CDMX	PM _{2.5}
Sureste	Tláhuac	TAH	Xochimilco, CDMX	O ₃ , SO ₂ , PM ₁₀ , PM _{2.5}
	UAM Iztapalapa	UIZ	Iztapalapa, CDMX	O ₃ , NO ₂ , NO _x , NO, SO ₂ , CO, PM _{2.5}

3.5 Datos

La RAMA monitorea la concentración de los ocho contaminantes cada hora, utilizando tres unidades de medición a saber, para el O₃, SO₂, NO₂, NO y NO_x se miden con partes por billón (ppb), para PM₁₀ y PM_{2.5} con microgramos por metro cúbico (µg/m³) y para el CO con partes por millón (ppm). Debido a que cada contaminante puede ser captado por más de un centro de monitoreo en las cinco regiones geográficas, se identificó en cada uno de ellos primero la concentración máxima diaria alcanzada, para después extraer únicamente el máximo diario del contaminante en cada una de las regiones, esto se realizó para el periodo del 1 enero de 2008 al 24 octubre de 2018, obteniendo un total de 3950 datos diarios.

Cabe mencionar que a excepción del O₃ en la región noreste y del SO₂ en la región noroeste y noreste, todos los demás contaminantes en cada una de las regiones mostraron cierto porcentaje de datos faltantes (véase Tabla 2), ya sea por mantenimiento o validación de la información.

Tabla 2 Datos faltantes de los contaminantes por región

Contaminante	%				
	Noroeste	Noreste	Centro	Suroeste	Sureste
O ₃	0.13	0.00	2.99	0.25	0.25
NO ₂	0.05	0.03	7.16	3.22	5.44
NO _x	0.05	0.03	7.16	3.22	5.44
NO	0.05	0.03	7.16	3.22	5.44
SO ₂	0.00	0.00	3.52	4.68	1.01
CO	0.03	0.03	6.56	4.28	7.54
PM ₁₀	0.33	0.05	7.75	29.57	9.42
PM _{2.5}	11.32	6.63	1.59	13.72	19.54

Ante la presencia de datos faltantes, se justifica el aplicar el filtro de Hodrick y Prescott a través del filtro de Kalman, tal y como lo propone Guerrero (2008), tanto para la estimación de sus tendencias como para sus respectivos pronósticos. Cabe mencionar que para el procesamiento de los datos y la elaboración de las estimaciones se utilizó SAS versión 13.2 y R versión 3.5.2

El Índice de Calidad del Aire (IMECA) para la ZMVM utiliza las normas oficiales mexicanas en materia de salud ambiental vigentes, mismas que determinan los límites permisibles para la preservación de la salud de la población. Este índice puede sufrir cambios por la actualización de alguna norma ambiental (NOM-022-SSA1-2010, NOM-020-SSA1-2014 y NOM-025-SSA1-2014O3) y contempla seis de los ocho contaminantes considerados en este trabajo, denominados criterios, pues pueden afectar el bienestar y salud de la población, siendo ellos: O₃, SO₂, NO₂, CO, PM₁₀ y PM_{2.5}. Además de estos contaminantes se monitorean otros tres contaminantes, los cuales son el NO_x, NO y PM_{10-2.5} aunque este último no se estudia en este trabajo ya que su medición comienza en 2011. La Tabla 3 muestra los criterios para establecer la calidad del aire según la concentración correspondiente a cada uno de los seis contaminantes criterio

Tabla 3 Criterios de la calidad del aire

IMECA	Categoría	O ₃ (ppm)	SO ₂ (ppm)	NO ₂ (ppm)	PM ₁₀ (ug/m ³)	PM _{2.5} (ug/m ³)	CO (ppm)
0-50	Buena	0.000-0.070	0.000-0.025	0.000-0.105	0-40	0.0-12.0	0.0-5.5
51-100	Regular	0.071-0.095	0.026-0.11	0.106-0.210	41-75	12.1-45.0	5.6-11.0
101-150	Mala	0.096-0.154	0.111-0.207	0.211-0.430	76-214	45.1-97.4	11.1-13.0
151-200	Muy mala	0.155-0.204	0.208-0.304	0.431-0.649	215-354	97.5-150.4	13.1-15.4
201-300	Extremadamente mala	0.205-0.404	0.305-0.604	0.650-1.249	355-424	150.5-250.4	15.5-30.4
301-400		0.405-0.504	0.605-0.804	1.250-1.649	425-504	250.5-350.4	30.5-40.4
401-500	Peligrosa	0.505-0.604	0.805-1.004	1.650-2.049	505-604	350.5-500.4	40.5-50.4

Fuente: Gaceta oficial de la ciudad de México, No. 453. 15 de noviembre 2018. Consultado en <https://fondeso.cdmx.gob.mx/storage/app/media/gaceta-oficial-de-la-ciudad-de-mexico-a-15-de-noviembre-de-2018.pdf> el 9 de enero de 2109.

- Notas: a) Para el O₃ y el NO₂, el Índice de Calidad del Aire se obtendrá a partir de concentraciones promedio de una hora;
 b) El SO₂ derivará de una concentración obtenida como promedio móvil de 24 horas;
 c) El CO derivará de una concentración obtenida como promedio móvil de 8 horas; y
 d) Las PM₁₀ y las PM_{2.5} derivarán de concentraciones obtenidas como promedio móvil de 24 horas.

4. Resultados

Se presentan las estimaciones de las tendencias y los pronósticos, divididos tanto por contaminante como por regiones. Para tal efecto se hace una descripción de los hallazgos de manera regional en función de la Figura 3 para el caso de tendencias y Figura 4, junto con Tabla 4, para el caso de pronósticos.

4.1 Tendencias

Región noroeste

La tendencia estimada de los máximos diarios, en general, para los contaminantes PM₁₀, NO₂, NO_x, NO y CO, en la región noroeste son decrecientes, en tanto que O₃ y SO₂ a pesar de mostrar una marcada tendencia decreciente al inicio del periodo, éste ritmo tiende a desacelerarse al final del periodo; las PM_{2.5} muestran oscilaciones que no permiten determinar una tendencia única.

Tomando en cuenta la norma ambiental NADF-009-AIRE-2017, la estimación de la tendencia de los máximos diarios para el O₃ en la región noroeste está en todo momento fuera del rango de buena calidad del aire, es decir en la categoría de calidad regular, lo mismo ocurre con las PM₁₀ y las PM_{2.5}. Respecto a la del SO₂ se ubica como regular, las tendencias del CO y NO₂ son las únicas que en ningún momento rebasan las concentraciones permisibles por la norma. Aunque para el NO y NO_x no existe una norma mexicana que regule su comportamiento, ambas presentan un comportamiento similar en sus tendencias iniciando el periodo con un ligero crecimiento, sin embargo, para 2012 comienzan un descenso notable.

Región noreste

La tendencia, en general, es decreciente para todos los contaminantes en la región noreste a excepción del O₃ y PM_{2.5}; cabe notar que el O₃ inicia con una tendencia a la baja, sin embargo, en 2016 comienza un ligero ascenso, en tanto que para las PM_{2.5} se comienza un patrón similar al observado cuatro años antes.

De acuerdo con la NADF-009-AIRE-2017, la tendencia del O₃ durante todo el periodo de estudio se encuentra en el intervalo de calidad regular, pues las concentraciones máximas se encuentran por debajo de 95 ppm. El SO₂ también muestra una tendencia con criterios regulares. En tanto que las PM₁₀ y PM_{2.5} presentan tendencias con connotaciones de mala calidad del aire en todo el periodo de análisis, mientras que las tendencias del NO₂ y CO muestran criterios de buena calidad. Por su parte el NO_x y NO reflejan una tendencia claramente decreciente.

Región centro

Las series observadas que corresponden a los contaminantes NO₂, NO, NO_x, CO y PM₁₀ muestran segmentos de datos faltantes. La tendencia estimada es decreciente para NO₂, NO, CO, PM₁₀ y NO_x, en tanto que el O₃, y PM_{2.5} presentan aún un comportamiento de tendencia distinguidos por segmentos, pues el O₃ a mediados de 2016 comienza una tendencia creciente; por su parte las PM_{2.5} inician una tendencia decreciente hasta finales de 2010, se convierte en creciente hasta inicios de 2016 y vuelve a decrecer en adelante. El SO₂ muestra apenas una ligera tendencia decreciente.

Por otra parte, las tendencias de las PM₁₀ y PM_{2.5} indican que en el periodo de enero de 2008 a octubre 2018 se encuentran en el rango de una mala calidad, aunque con una tendencia decreciente para las PM₁₀. La tendencia del O₃ por el contrario a las PM₁₀, hasta antes de 2015 venía siendo decreciente para invertirse en ese año, no obstante, toda la tendencia se encuentra dentro el criterio de una calidad regular. Por su parte, la tendencia del NO₂, SO₂ y CO se encuentran en todo el periodo de estudio, bajo el criterio de una buena calidad del aire. El comportamiento de las tendencias del NO_x y NO se asemeja y son decrecientes.

Región suroeste

A pesar de no contar con datos para las PM₁₀ y PM_{2.5} a finales del periodo de estudio y exclusivamente para la región suroeste, el modelo fue capaz de realizar las respectivas estimaciones de sus tendencias e intervalos hasta octubre de 2018. El NO₂ y CO muestran una clara estimación de tendencias decrecientes. Las PM₁₀ y NO_x muestran un comportamiento similar en sus tendencias, aunque con niveles de concentración máxima diferenciadas, el SO₂ a pesar de que inicia con una tendencia creciente marcada, en el

transcurso del tiempo se va haciendo más lenta. Por su parte el O₃ a diferencia del PM_{2.5} inicia con una tendencia decreciente para después invertirse en 2014.

Al observar el comportamiento de las tendencias con respecto a la NADF-009-AIRE-2017, se tiene que el O₃ inicia el periodo con una mala calidad del aire, sin embargo, la mayoría del periodo continua en una situación regular, no obstante, al igual que las otras regiones a partir de 2015 comienza un ascenso en sus tendencias. Lo mismo pasa con las PM₁₀, pues a la mitad del periodo la tendencia se ubica en una mala calidad del aire, pero después comienza a ser regular; contrario a esto las PM_{2.5} inician con una tendencia regular hasta 2010, para después pasar a concentraciones máximas de mala calidad. El SO₂ muestra una tendencia con concentraciones máximas de buena calidad, misma situación se presenta para el NO₂ y CO, respecto al NO_x y el NO al igual que las regiones anteriormente descritas muestran comportamientos semejantes en cuanto a tendencia.

Región sureste

La serie observada de las PM_{2.5} se tiene una proporción considerable de datos faltantes, además de mostrar una tendencia marcada por segmentos. Para el CO es contundente el hecho que su tendencia es decreciente; es importante señalar que, a diferencia de las otras regiones, el O₃ presenta una tendencia claramente decreciente. Respecto al resto de los contaminantes, sus tendencias son decrecientes, aunque inician con un ligero estancamiento.

Las tendencias de los contaminantes PM₁₀ y PM_{2.5} se encuentran en los intervalos de mala calidad del aire, destacando que mientras el primero de ellos muestra una tendencia decreciente y es menos acelerada al final del periodo, en el segundo contaminante la tendencia va al alza. La tendencia del O₃ se encuentra en el intervalo de calidad regular. Los contaminantes que presentan una tendencia a la baja, así como que se encuentran en los intervalos de una buena calidad del aire son el NO₂ y el CO y SO₂. Por su parte, los contaminantes NO_x y NO siguen mostrando los mismos comportamientos en su tendencia.

Figura 2 Series observadas, estimación de tendencias e intervalos de estimación a 2 desviaciones estándar por contaminante y región geográfica

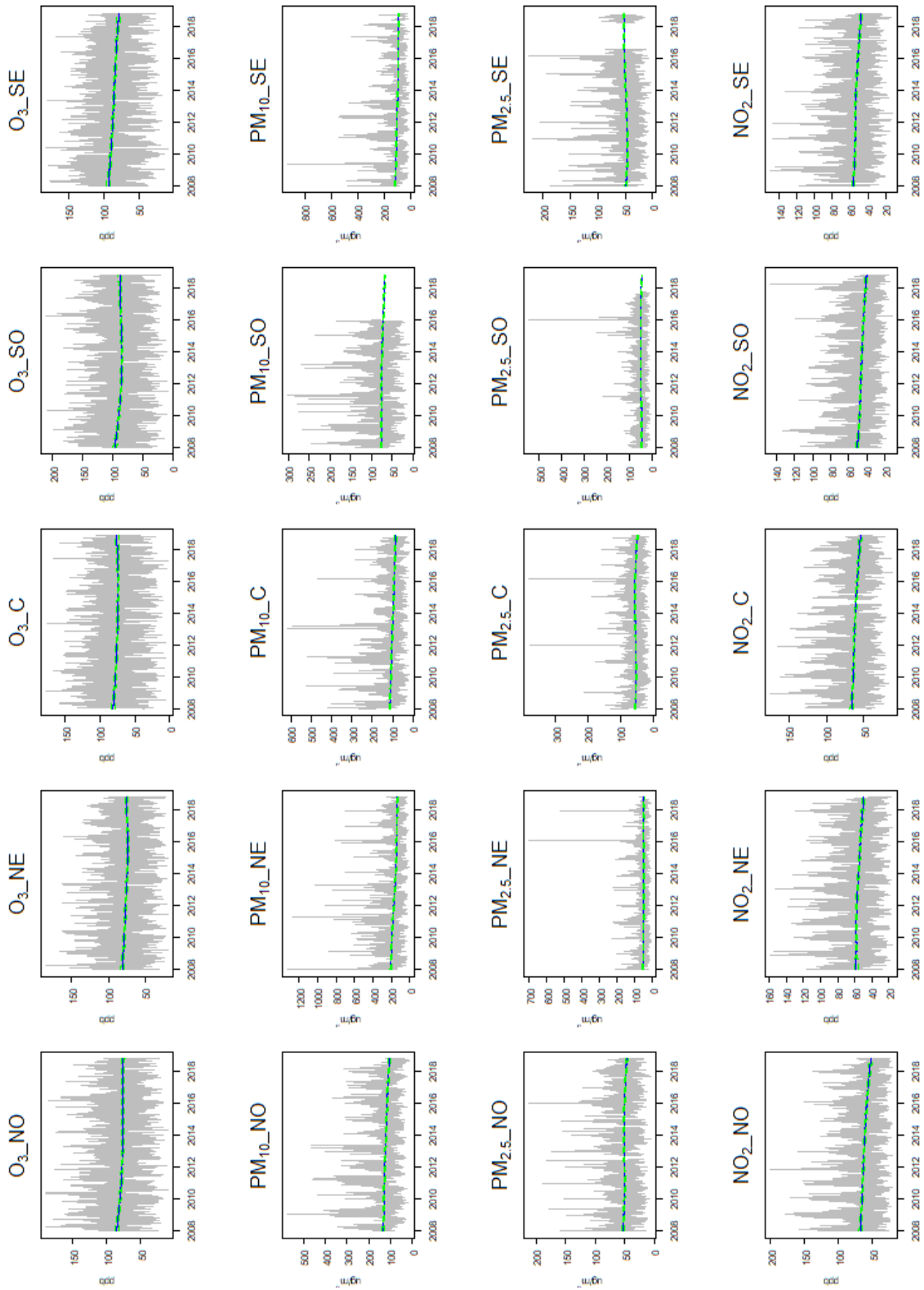
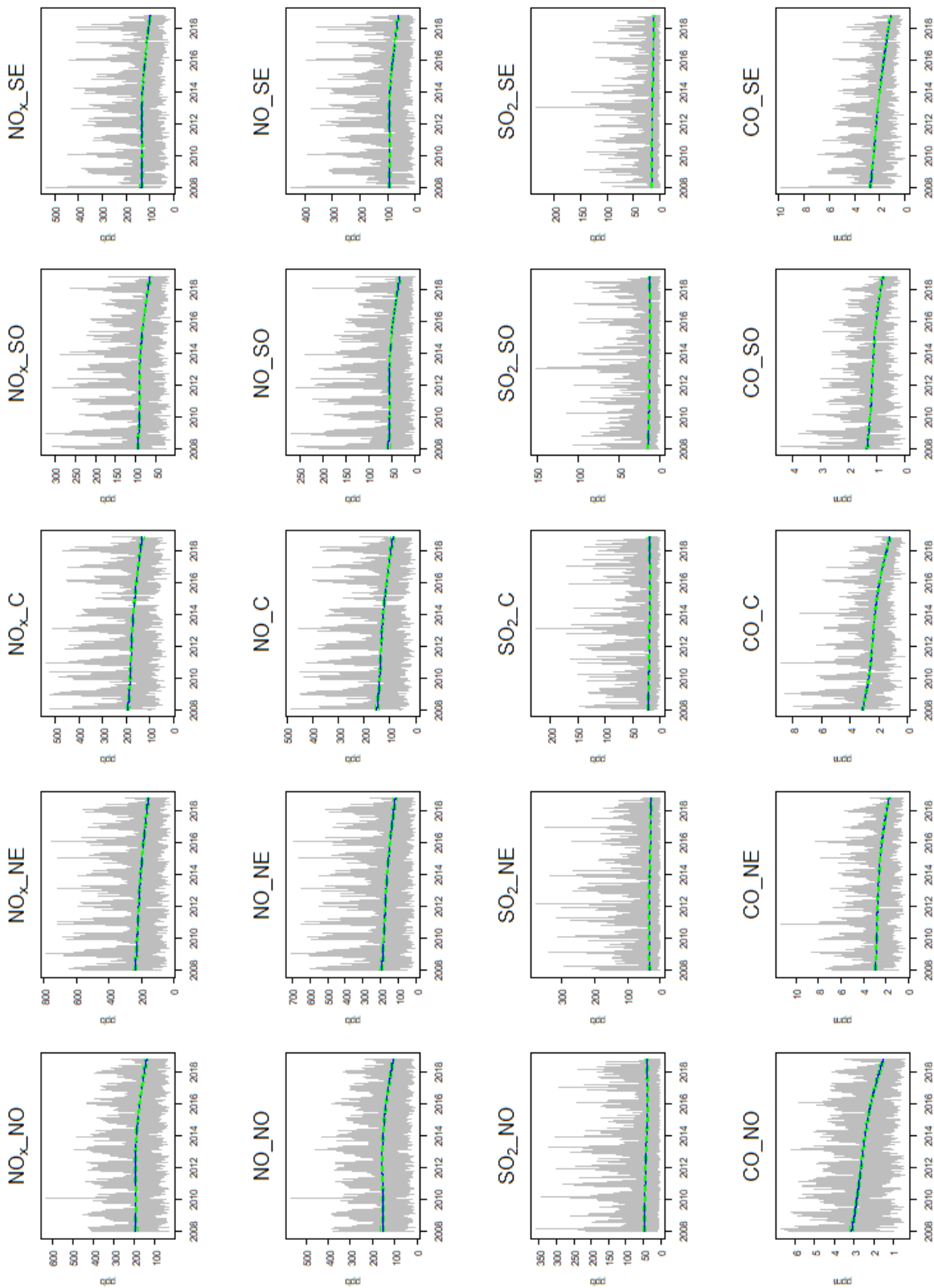


Figura 2 Series observadas, estimación de tendencias e intervalos de estimación a 2 desviaciones estándar por contaminante y región geográfica (Continuación)



4.2 Pronósticos

Es importante señalar que no se realizan pronósticos de las concentraciones máximas observadas de los contaminantes, sino de las tendencias estimadas. Así, dichos pronósticos se realizan en un horizonte de 7 días. Nótese que los datos observados con los que se contó para el análisis fueron del 1 de enero de 2008 hasta el 31 de octubre de 2018, sin embargo, para la estimación de la tendencia solo se contempló hasta el 25 de octubre, para después pronosticar del 26 al 31 de octubre de 2018. Con la intención de verificar la validez de estos, se estimó además la tendencia con todos los datos hasta el 31 de octubre de 2018, además se obtuvieron las medidas de precisión MAD, MSE y MAPE para cada uno de los contaminantes en las regiones.

En general, los pronósticos de las tendencias de los ocho contaminantes para cada una de las cinco regiones mostraron buenas precisiones, ya que a través del MAPE es posible observar que el promedio de los errores entre la estimación de la tendencia y el pronóstico fueron menores a 1.5%, respecto al MAD estas fueron menor a 0.5% y para MSE no rebasan el 0.2%. Además, todos los pronósticos de la tendencia se encuentran dentro de los intervalos de estimación de la tendencia (véase la Figura 2 y la Tabla 4). En los siguientes párrafos se describe a más detalle lo que sucede con los pronósticos por región.

Región noroeste

De acuerdo con los pronósticos de la tendencia, para la semana de referencia, se tiene que el CO, NO₂ y SO₂, en la región noroeste, se mantuvieron por debajo de las concentraciones permisibles por la norma NADF-009-AIRE-2017, en tanto que para las PM₁₀ y PM_{2.5} la calidad del aire se vio afectada por estos contaminantes, calificándola como mala. Respecto al O₃ en dicha semana pronosticada la calidad del aire fue regular.

Región noreste

Para PM₁₀, en la región noreste se pronosticaron concentraciones máximas superiores al resto de las regiones, en la semana de análisis, además la calidad del aire se considera como mala, las PM_{2.5} contaminaron el aire colocándola como mala, por otra parte, el O₃ presenta un escenario regular del aire. En contraste el CO, NO₂ y SO₂ cumplen con las normas.

Región centro

En la región centro los pronósticos de la estimación de la tendencia para el O₃ para los siete días, se ubicaron en una calidad regular, en tanto que las PM₁₀ y PM_{2.5} estuvieron en una mala calidad. Aunque los pronósticos del NO₂ no rebasan las normas establecidas, éste tuvo

concentraciones máximas pronosticadas superiores a la de las otras regiones, el CO y SO₂ tampoco rebasan las normas.

Región suroeste

Los pronósticos de la tendencia advierten que, para las PM₁₀, es la única de las zonas en las que se tiene niveles de calidad regulares, además, aunque para las PM_{2.5}, la calidad del aire es mala, se registran los menores niveles a diferencia de las otras regiones. El O₃ tiene un comportamiento semejante a la región noroeste. Los contaminantes CO, NO₂ y SO₂ se encuentran dentro de las normas establecidas, aunque cabe destacar que el NO₂ en esta región tiene las concentraciones máximas más altas en comparación con las otras regiones.

Región sureste

En la región sureste los únicos pronósticos de la estimación de tendencia que presentan concentración fuera de la norma son las PM₁₀ y PM_{2.5}, mientras tanto para la semana de pronósticos el O₃ parece estar un poco más controlado, aunque esta situación se refleja con mayor éxito en el CO, NO₂ y SO₂.

Tabla 4 Medidas de precisión de los pronósticos por contaminante y región

Región	Medida	O ₃	PM ₁₀	PM _{2.5}	NO ₂	NO _x	NO	SO ₂	CO
Noroeste	MAD	0.05805	0.43961	0.05700	0.01889	0.21703	0.44569	0.28543	0.00678
	MSE	0.00337	0.19325	0.00325	0.00036	0.04710	0.19864	0.08147	0.00005
	MAPE	0.00076	0.00431	0.00121	0.00036	0.00153	0.00416	0.00742	0.00446
Noreste	MAD	0.07371	0.38584	0.19862	0.01665	0.18827	0.00608	0.14881	0.00024
	MSE	0.00543	0.14887	0.03945	0.00028	0.03545	0.00004	0.02215	0.00000
	MAPE	0.00098	0.00285	0.00409	0.00034	0.00121	0.00005	0.00533	0.00014
Centro	MAD	0.01019	0.13284	0.09271	0.05826	0.02055	0.08566	0.25027	0.00252
	MSE	0.00010	0.01765	0.00859	0.00339	0.00042	0.00734	0.06264	0.00001
	MAPE	0.00013	0.00157	0.00191	0.00107	0.00016	0.00097	0.01351	0.00202
Suroeste	MAD	0.15206	0.00000	0.00000	0.02701	0.10974	0.25950	0.17437	0.00102
	MSE	0.02312	0.00000	0.00000	0.00073	0.01204	0.06734	0.03040	0.00000
	MAPE	0.00172	0.00000	0.00000	0.00067	0.00167	0.00787	0.01400	0.00131
Sureste	MAD	0.32526	0.69153	0.00000	0.00000	0.00000	0.00000	0.14300	0.00000
	MSE	0.10579	0.47821	0.00000	0.00000	0.00000	0.00000	0.02045	0.00000
	MAPE	0.00420	0.00759	0.00000	0.00000	0.00000	0.00000	0.01276	0.00000

Figura 3 Estimación de tendencia y pronósticos de la tendencia en un horizonte de 7 días (del 26 al 31 de octubre de 2018) para cada contaminante según la región geográfica

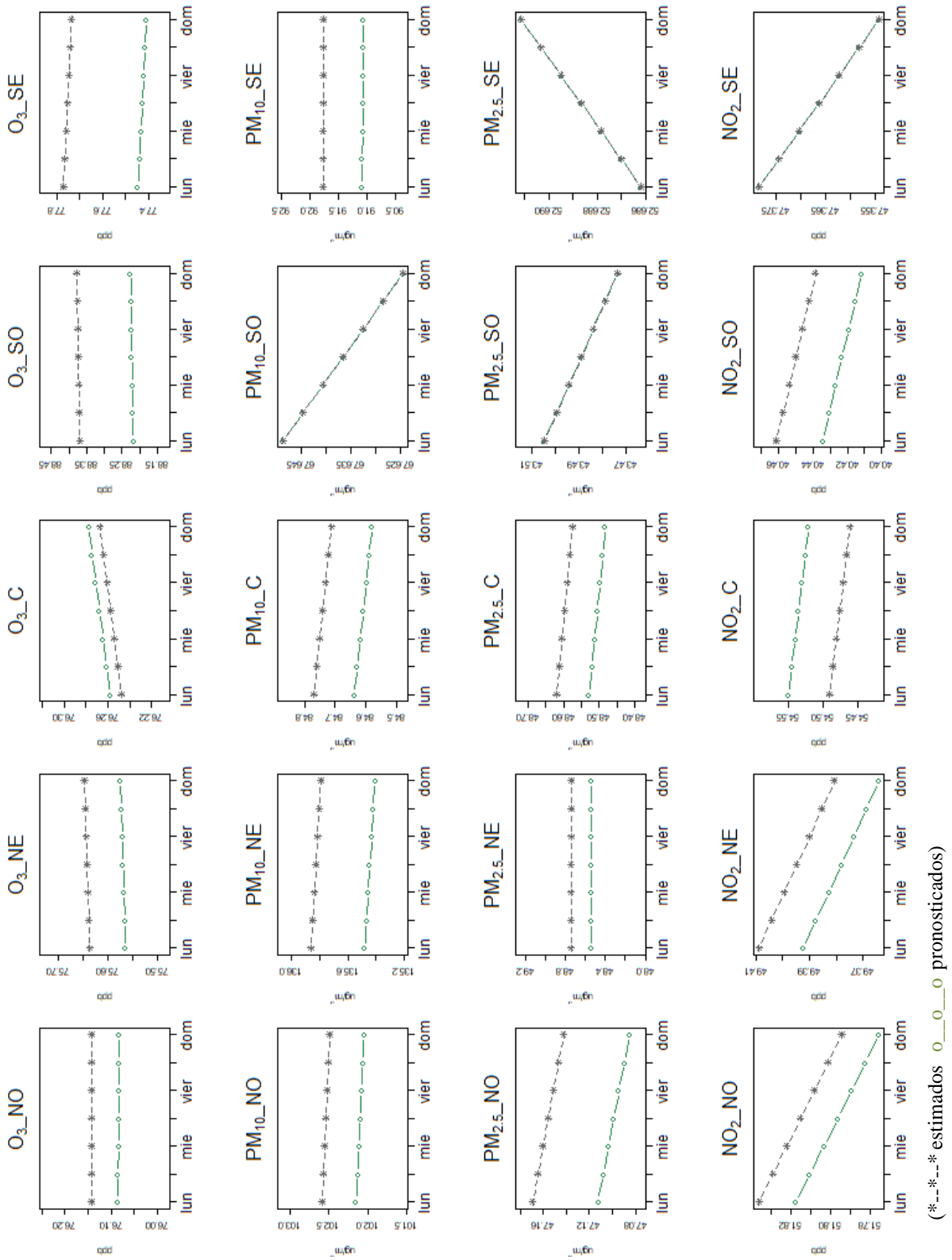
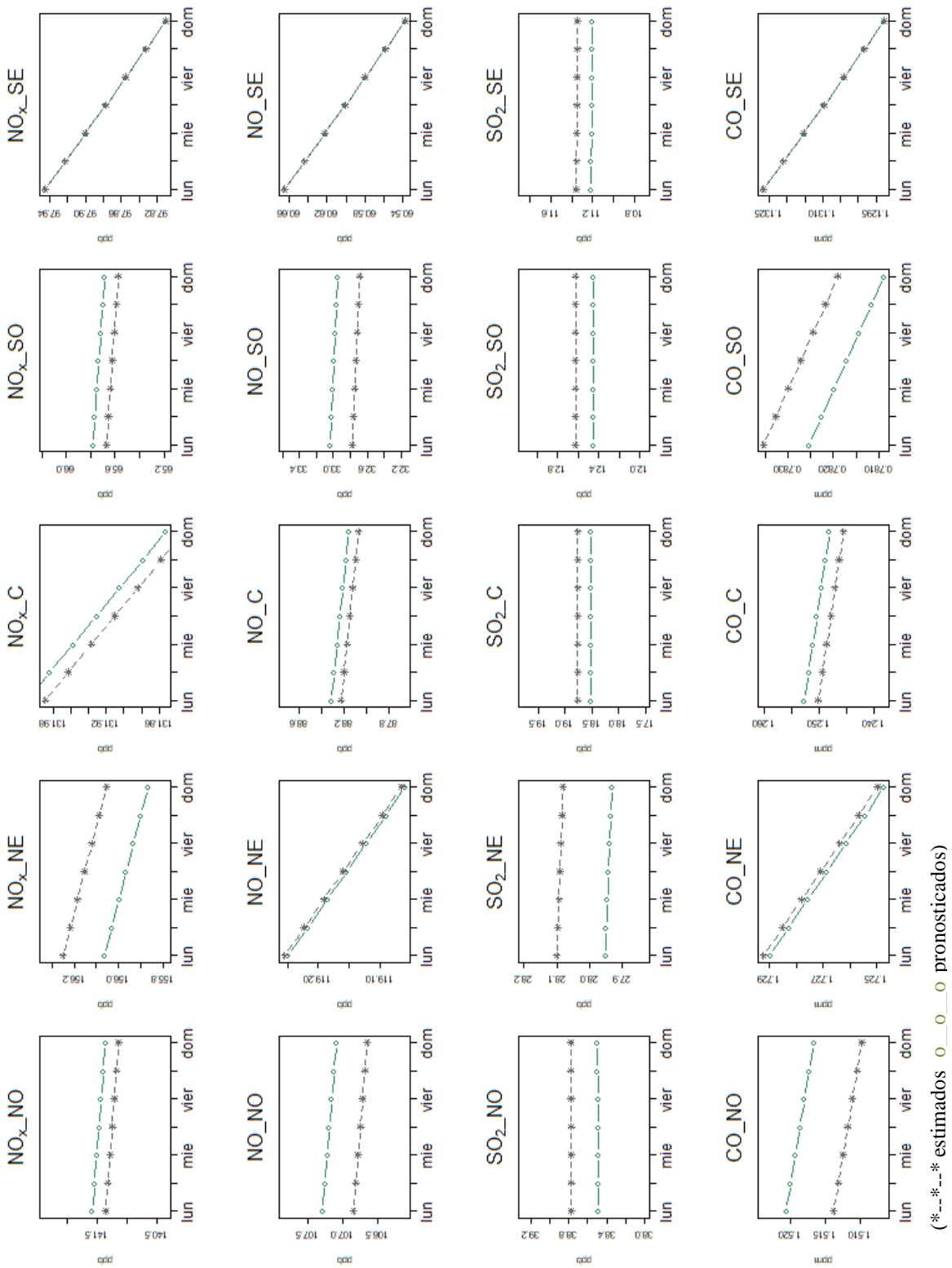


Figura 3 Estimación de tendencias y pronósticos de la tendencia en un horizonte de 7 días (del 26 al 31 de octubre de 2018) para cada contaminante según la región geográfica (Continuación)



En resumen, se obtuvo que las tendencias de las PM_{10} y $PM_{2.5}$ en el periodo de estudio y para cada una de las regiones de la ZMVM, mostraron ubicarse en el umbral de una mala calidad. Por su parte, las tendencias del O_3 para las cinco regiones, se caracterizan por estar dentro del intervalo como una calidad regular del aire, y solo en la región sureste es donde se muestra una tendencia decreciente.

El SO_2 muestra que sus tendencias estimadas en la región noroeste y noreste están dentro del umbral de una calidad regular del aire, mientras que en el centro, suroeste y sureste se encuentra dentro de una buena calidad. EL CO y el NO_2 se pueden considerar como dos de los contaminantes controlados en toda la ZMVM, pues en todas las regiones sus tendencias y pronósticos se encuentran a la baja y nunca rebasando los límites permisibles.

Si bien el NO_x es monitoreado para la ZMVM, no se establece en la norma mexicana de salud las concentraciones permisibles, por lo que se consideraría importante que se integrará este contaminante como parte de ella, pues el NO_x es uno de los precursores del O_3 , contaminante que repercute en la salud de la población en concentraciones elevadas. En este mismo sentido el NO presenta el mismo comportamiento en sus tendencias con el NO_x y tampoco es regulado por la norma.

Conclusiones

En la literatura especializada se han desarrollado y/o aplicado diversas metodologías en torno al comportamiento de las tendencias y pronósticos de las concentraciones de contaminantes atmosféricos, con la intención de comprender el fenómeno o prever escenarios futuros. Se considera que la metodología aplicada y el análisis aquí expuesto por región, es un recurso técnico pertinente, de naturaleza exploratoria, útil y simple, que abona en la materia.

Por medio del suavizamiento controlado, estimado mediante el filtro de Kalman, fue posible estimar tendencias de los contaminantes O_3 , SO_2 , NO_2 , CO, PM_{10} , $PM_{2.5}$, NO y NO_x para las cinco regiones de la ZMVM, aun cuando se presentaban datos faltantes tanto en la parte central de las series como al final de estas. Asimismo, dicha circunstancia no impidió el poder realizar sus respectivos pronósticos, por lo que se considera una herramienta robusta y eficiente para su aplicación.

Aunque no hay un índice óptimo de suavidad, la imposición de este en el modelo para todas las series permitió comparar el comportamiento de las tendencias entre los diferentes contaminantes y para cada una de las regiones de la ZMVM, reflejando que aún se está lejos de poder hablar de una problemática resuelta de manera integral sobre la contaminación atmosférica en esta metrópoli.

Los pronósticos aquí presentados se realizan a partir de las estimaciones de las tendencias por lo que estos no reflejan el comportamiento puntual, sino el pronóstico de la tendencia en la semana del 26 al 31 de octubre comprobando que los resultados de los pronósticos tuvieron una buena precisión para cada contaminante y región. Por último, se considera que este tipo de herramienta estadística sería oportuna para los tomadores de decisiones en los temas de contaminación atmosférica. Sería deseable como futura línea de investigación realizar la estimación de las tendencias y sus pronósticos de los ocho contaminantes considerando su estacionalidad.

Anexo A

Formación de los contaminantes atmosféricos

Contaminante	Formación
Ozono (O ₃)	El ozono es un contaminante secundario que se forma por una reacción fotoquímica entre emisiones primarias de óxidos de nitrógeno (NO _x), compuestos orgánicos volátiles (COVs) o hidrocarburos (HCs) en presencia de la radiación solar, aunado a las condiciones geográficas, climatológicas y meteorológicas del medio ambiente.
Monóxido de Carbono (CO)	Es un gas inodoro e incoloro que se produce por la combustión incompleta de compuestos de carbono, consecuentemente pueden verterlo al aire los vehículos automotores y la industria, aunque en menor escala; algunos procesos naturales son capaces de emitirlo, tales como los incendios forestales o su emisión de los procesos naturales que se llevan a cabo en los océanos. Mención especial debe hacerse de la acumulación intramuros por procesos domésticos y el hábito de fumar
Dióxido de azufre (SO ₂)	Está presente en el petróleo y el carbón en su estado natural, y se señala desde hace décadas que este contaminante es el responsable en buena medida de las lluvias ácidas, recientemente se ha reconocido al SO ₂ por su contribución en aerosoles inorgánicos secundarios. Se genera como emisión tanto de fuentes naturales, como de la combustión de compuestos ricos en azufre.
Dióxido de nitrógeno (NO ₂)	Se deriva de los procesos de combustión es un contaminante primario y tiene un doble papel, ya que se le reconoce efectos potencialmente dañinos de manera directa, pero también es uno de los precursores del ozono
Partículas menores a 10 micrómetros (PM ₁₀)	El material particulado es una mezcla compleja de sustancias en estado líquido o sólido, que permanece suspendida en la atmósfera por periodos variables de tiempo. Por su origen pueden definirse como primarias (aquellas producidas directamente por alguna fuente contaminante) o secundarias (las que se forman en la atmósfera, como resultado de la interacción química entre gases y partículas), de acuerdo con su diámetro aerodinámico, estas puedan clasificarse en menores o iguales a 10 micras o menores o iguales a 2.5 micras
Partículas menores a 2.5 micrómetros (PM _{2.5})	
Monóxido de nitrógeno (NO)	Es un gas sin color, y difícilmente soluble en agua, que constituye uno de los contaminantes de la atmósfera que forma parte de la lluvia ácida. Es expulsado a la atmósfera mediante los tubos de escape de los automóviles.
Óxido de nitrógeno (NO _x)	Los óxidos de nitrógeno son una mezcla de gases compuestos de nitrógeno y oxígeno. El monóxido de nitrógeno y el dióxido de nitrógeno constituyen dos de los óxidos de nitrógeno más importantes toxicológicamente.

Fuente: Elaboración propia con base a las Normas de Calidad del Aire. COFEPRIS.

Anexo B

Representación del modelo espacio-estado y Filtro de Kalman

De acuerdo con Rodríguez (2002), el artículo de Kalman (1960), mostró que una amplia clase de problemas podía representarse mediante un modelo lineal simple a través de espacio-estados, en la que su naturaleza markoviana permite que los cálculos necesarios para su aplicación práctica se efectúen de manera recursiva. Una vez que el modelo espacio-estado ha sido definido, el Filtro de Kalman proporciona el medio para estimar el estado inobservable a partir de alguna magnitud observable relacionada con éste, de forma que la estimación se actualice cada vez que se dispone de nueva información.

Modelo espacio-estado

En la forma general de espacio-estado, para una serie multivariada, Y_t , con N elementos, las variables observables están relacionadas con el vector $m \times 1$, α_t , conocido como el vector de estado, a través de una ecuación de medida u observación (Harvey, 1989).

$$Y_t = Z_t \alpha_t + d_t + \varepsilon_t, \quad t = 1, \dots, T$$

Donde Z_t es una matriz de $N \times m$, d_t es un vector de $N \times 1$ y errores serialmente no correlacionados con media cero y matriz de covarianzas $H_t: E(\varepsilon_t) = 0$ y $Var(\varepsilon_t) = H_t$.

Por lo que para el caso univariado ($N=1$) la ecuación de observación puede ser escrita como:

$$y_t = z_t' \alpha_t + d_t + \varepsilon_t, \quad Var(\varepsilon_t) = h_t \quad t = 1, \dots, T$$

En la que los elementos de α_t son no observables, sin embargo, se sabe que son generados por el proceso de Markov de primer orden, cuya ecuación se conoce como de transición:

$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t \quad t = 1, \dots, T$$

donde T_t es una matriz de $m \times m$, c_t es un vector $m \times 1$, R_t es una matriz de $m \times g$ y η_t es un vector $g \times 1$ de errores serialmente incorrelacionados con media cero y matriz de covarianza Q_t , es decir, $E(\eta_t) = 0$ y $Var(\eta_t) = Q_t$.

La especificación del sistema de espacio-estado presenta además dos supuestos:

- a. El vector inicial, α_0 , tiene una media a_0 y una matriz de covarianzas P_0 :

$$E(\alpha_0) = a_0 \text{ y } Var(\alpha_0) = P_0$$

- b. Los errores ε_t, η_t son no correlacionados entre sí en todos los periodos de tiempo y no correlacionados con el estado inicial:

$$E(\varepsilon_t \eta_s') = 0 \text{ para todo } s, t = 1, \dots, t \text{ y}$$
$$E(\varepsilon_t \alpha_0') = 0, E(\eta_t \alpha_0') \text{ para toda } t = 1, \dots, t$$

Filtro de Kalman

El filtro de Kalman a través de su forma recursiva obtiene el estimador óptimo del vector de estados en el tiempo t , basada en la información disponible hasta tiempo t , su forma general es (Harvey, 1989):

Considerando el modelo de espacio-estado anterior, a_{t-1} denota el estimador óptimo de α_{t-1} basadas en las observaciones incluidas hasta el momento y_{t-1} . Sea P_{t-1} la matriz de covarianza $m \times m$ del error de estimación, es decir.

$$P_{t-1} = E[(\alpha_{t-1} - a_{t-1})(\alpha_{t-1} - a_{t-1})']$$

Considerando a_{t-1} y P_{t-1} , el estimador óptimo de α_t es dado por

$$a_{t|t-1} = T_t a_{t-1} + c_t$$

Mientras que la matriz de covarianzas del error estimado es

$$P_{t|t-1} = T_t P_{t-1} T_t' + R_t Q_t R_t', \quad t = 1, \dots, T$$

Estas dos ecuaciones son conocidas como ecuaciones de predicción.

Una vez que la nueva observación, y_t , es disponible, el estimador α_t , $a_{t|t-1}$, puede ser actualizado; por lo que las ecuaciones de actualización son:

$$a_t = a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} (y_t - Z_t a_{t|t-1} - d_t)$$

y

$$P_t = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1}$$

Donde

$$F_t = Z_t P_{t|t-1} Z_t' + H_t, \quad t = 1, \dots, T$$

Una vez establecidas las ecuaciones de predicción y actualización, se puede constituir el filtro de Kalman, por lo que si se desea se puede reescribir como un único conjunto de recursiones directamente de a_{t-1} a a_t , o alternativamente, desde $a_{t|t-1}$, para este último caso se tiene

$$a_{t|t-1} = (T_{t+1} - K_t Z_t) a_{t|t-1} + K_t y_t + (c_{t+1} - K_t d_t)$$

Donde la matriz de ganancia es dada por

$$K_t = T_{t+1}P_{t|t-1}Z'_tF_t^{-1}, \quad t = 1, \dots, T$$

La recursión para el error de la matriz de covarianza es

$$P_{t|t-1} = T_{t+1}(P_{t|t-1} - P_{t|t-1}Z'_tF_t^{-1}Z_tP_{t|t-1})T'_{t+1} + R_{t+1}Q_{t+1}, \quad t = 1, \dots, T$$

Los valores iniciales para el Filtro de Kalman, pueden ser especificados en términos de a_0 y P_0 o $P_{1|0}$. Dadas dichas condiciones iniciales, el Filtro de Kalman obtiene el estimador óptimo del vector de estado para cada una de las observaciones disponibles. Cuando todas las observaciones T han sido procesadas, el filtro produce el estimador óptimo del vector de estado actual y/o el vector de estado en el siguiente periodo, basado en toda la información del conjunto.

Cabe notar que cuando la serie es univariada, hay un cambio de notación para la ecuación del filtro, por lo que la matriz F_t de $N \times N$, es reemplazada por el escalar

$$f_t = z'_tP_{t|t-1}z_t + h_t$$

Además, es común que al configurar el modelo de espacio estado, para el filtrado de una serie univariada, la variación de las perturbaciones sea proporcional a un escalar positivo, es decir.

$$Var(\varepsilon_t) = \sigma_*^2 h_t \text{ y } Var(\eta_t) = \sigma_*^2 Q_t$$

Si la matriz inicial de covarianzas es también especificada hasta el factor de proporcionalidad, σ_*^2 , el filtro de Kalman se puede ejecutar de manera independiente de σ_*^2 con las cantidades marcadas definidas por

$$P_{t|t-1} = \sigma_*^2 P_{t|t-1}^* \text{ y } f_t = \sigma_*^2 f_t^*$$

Anexo C

```
#-----  
#           Programa del filtro de Hodrick y Prescott  
#           a través del filtro de Kalman  
#-----  
  
#*****  
# DESCRIPCIÓN:  
# La función suaviza las tendencias de las series de tiempo de los  
# contaminantes de la ZMVM utilizando el filtro de Hodrick y Prescott a  
# través del filtro de Kalman, además los pronósticos de sus respectivas  
# estimaciones de tendencia, propuesto en Guerrero (2008 y 2007).  
# Cuyos argumentos son: datos, n1, n2, inicia, lambda, horizonte.  
  
# SELECCIONAR:  
# 1. Ruta de trabajo  
# 2. Base de datos (ruta especificada de Dropbox)  
# 3. Número de datos para estimar tendencia (n1)  
# 4. Total de datos de la serie (n2)  
# 5. Año de inicio de la serie (serie)  
# 6. Valor del parámetro de suavizamiento (lambda)  
# 7. Días de horizonte del pronóstico (horizonte)  
# *****  
  
rm(list = ls(all = TRUE))  
  
#librerías  
library(foreign)  
library(dlm)  
  
#Directorio  
getwd()  
setwd(dir)  
  
#Base de los máximos diarios de los contaminantes de la ZMVM por región  
base2 <-  
read.csv("https://www.dropbox.com/s/a68rnrkuado9bir/MAX_18_octubre1.csv?d  
l=1")  
base3 <- base2[,-1]  
  
#Función del filtro de Hodrick-Prescott a través del filtro de Kalman  
  
hp <- function(datos, n1=3950, n2=3957, inicia=2008, lambda=10000000000,  
horizonte=7){  
  
  obs <- datos[1:n1]  
  n <- length(obs)  
  tsdatos <- ts(obs, start=c(inicia, 1, 1), frequency=365.25)  
  
  # Porcentaje de suavidad elegido Guerrero(2008, p.193)  
  K2 = diag(n)  
  K21 = diff(diff(K2))  
  M = diag(1, n)+lambda*t(K21)%*%K21  
  inv <- chol2inv(chol(M))  
  
}
```

```

d <- sum(diag(inv))

#porcentaje de suavidad elegido
porcsuav <- (1-(d/n))*100

# Conjunto de estados iniciales
level <- tsdatos[1]
slope <- mean(diff(tsdatos), na.rm=TRUE)

# Modelo estado-espacio propuesto en Guerrero (2008,p.189 )
EstParams <- dlm(FF = matrix(c(1, 0), nr = 1),
                 V = 1,
                 GG = matrix(c(2, 1, -1, 0), 2, 2),
                 W = diag(c(0, 1/lambda)),
                 m0 = c(level, slope),
                 C0 = 2*diag(2))

Smooth_Estimates <- dlmSmooth(tsdatos, EstParams)

# Tendencia estimada suavizada con HP-Kalman
trend <- data.frame(Smooth_Estimates$s[2:(n+1), 1])

h <- data.frame(h=1:horizonte)

# Pronósticos propuestos en Guerrero (2007, p.1228)
u <- 0
trend_pron <- (((h*(h+1))/2)*u)+((h+1)*trend[n, ])-(h*trend[n-1, ])

# Tabla para la Tendencia y pronostico
names(trend)[names(trend)=="Smooth_Estimates.s.2..n...1...1."] <-
"tend"
names(trend_pron)[names(trend_pron)=="h"] <- "tend"
tendencia<-rbind(trend, trend_pron)
tendencia_vec<-as.vector(tendencia[, 1])

obs_ten<-data.frame(obs, trend)
obs_ten$obs_com<-ifelse(is.na(obs_ten[, 1]), obs_ten[, 2], obs_ten[, 1])

# Intervalos de estimación
dd = (t(obs_ten[, 3])*(1/lambda))%*%diag(n)%*%obs_ten[, 3]-
t(matrix(trend[, 1]))%*%(((1/lambda)*diag(n))+t(K21)%*%K21)%*%matrix(trend
[, 1])
vare = dd/(n-2)
sqvaredd = sqrt(vare)
M1dd = inv*lambda*(as.vector(vare))
var = diag(M1dd)
LS = trend+2*sqrt(var)
LI = trend-2*sqrt(var)
LILS <- cbind(LI, trend, LS)
colnames(LILS) <- c("LI", "C", "LS")

r <- data.frame(LI=c(rep(NA, horizonte)), C=c(rep(NA, horizonte)),
LS=c(rep(NA, horizonte)))

```

```

LILS1 <- rbind(LILS,r)

# Renombramos el nombre de las variables y ponemos los observados,
# tendencia y pronostico en un data.frame
datos_d <- data.frame(c(obs, rep(NA,horizonte)))
date <- data.frame(seq(as.Date("2008/1/1"), length.out =
length(tendencia[,1]), by="day"))

names(date)[names(date)=="seq.as.Date..2008.1.1....length.out...length.te
ndencia...1...."] <- "Fecha"
names(datos_d)[names(datos_d)=="c.obs..rep.NA..horizonte.."] <- "obs"
tenobs <- cbind(date, datos_d, tendencia, LILS1)

# Se aplica el procedimiento anterior para todos los datos de las
# series de los contaminantes de la ZMVM, para comparar con los
# pronósticos
datos18 <- datos[1:n2]
n18 <- length(datos18)
tsdatos18 <- ts(datos18,start=c(2008,1,1),frequency=365.25)

# Porcentaje de suavidad elegido Guerrero(2008, p.193)
K218 = diag(n18)
K2118 = diff(diff(K218))
M18 = diag(1,n18)+lambda*t(K2118)%*%K2118
inv18 <- chol2inv(chol(M18))
d18 <- sum(diag(inv18))

#porcentaje de suavidad elegido
porcsuav18 <- (1-(d18/n18))*100

# Conjunto de estados iniciales
level18 <- tsdatos18[1]
slope18 <- mean(diff(tsdatos18),na.rm=TRUE)

# Modelo estado-espacio propuesto en Guerrero (2008,p.189 )
EstParams18 <- dlm(FF = matrix(c(1, 0), nr = 1),
V = 1,
GG = matrix(c(2,1,-1,0),2,2),
W = diag(c(0,1/lambda)),
m0 = c(level,slope),
C0 = 2*diag(2))

Smooth_Estimates18 <- dlmSmooth(tsdatos18,EstParams18)

# Tendencia estimada suavizada
trend18 <- data.frame(Smooth_Estimates18$s[2:(n18+1),1])

obs_ten18 <- data.frame(datos18, trend18)
obs_ten18$obs_com <-
ifelse(is.na(obs_ten18[,1]),obs_ten18[,2],obs_ten18[,1])

# Intervalos de estimación

```



```

dd18 = (t(obs_ten18[,3])*(1/lambda))%*%diag(n18)%*%obs_ten18[,3]-
t(matrix(trend18[,1]))%*%(((1/lambda)*diag(n18))+t(K2118)%*%K2118)%*%matr
ix(trend18[,1])
vare18 = dd18/(n18-2)
sqvaredd18 = sqrt(vare18)
M1dd18 = inv18*lambda*(as.vector(vare18))
var18 = diag(M1dd18)
LS18 = trend18+2*sqrt(var18)
LI18 = trend18-2*sqrt(var18)
LILS18 <- cbind(LI18, trend18, LS18)
colnames(LILS18) <- c("LI18", "C18", "LS18")

# Errores de pronósticos
error <- LILS18$C18[3951:3957]-tenobs$tend[3951:3957]
estimados <- data.frame ( LILS18$C18[3951:3957],
LILS18$LI18[3951:3957], LILS18$LS18[3951:3957] )
pronos <- tenobs$tend[3951:3957]
Me <- mean(error)
MAD <- mean(abs(error))
MSE <- mean((error)^2)
MAPE <- mean(abs(error/LILS18$C18[3951:3957]))
med_error <- data.frame(Me, MAD, MSE, MAPE)

return(data.frame(a=tenobs[,c(2:3,4,6)], b=LILS18))
}

# Calculamos para todos los contaminantes la estimación de la
# tendencia, los pronósticos y los intervalos de estimación

x<- list()
for (i in 1:dim(base3)[2]) {
  x[[i]] <- data.frame(hp(base3[,i]))
}

# Graficamos los valores observados, la tendencia estimada y los
# intervalos de estimación

unid <- c(rep("ppb",5),rep(expression('ug/'*m'^3),10),rep("ppb",20)
,rep("ppm",5))

tit<- c(expression('O'[3]*'_NO'), expression('O'[3]*'_NE'),
expression('O'[3]*'_C'), expression('O'[3]*'_SO'),
expression('O'[3]*'_SE'), expression('PM'[10]*'_NO'),
expression('PM'[10]*'_NE'), expression('PM'[10]*'_C'),
expression('PM'[10]*'_SO'), expression('PM'[10]*'_SE'),
expression('PM'[2.5]*'_NO'), expression('PM'[2.5]*'_NE'),
expression('PM'[2.5]*'_C'), expression('PM'[2.5]*'_SO'),
expression('PM'[2.5]*'_SE'), expression('NO'[2]*'_NO'),
expression('NO'[2]*'_NE'), expression('NO'[2]*'_C'),

```

```

expression('NO'[2]*'_SO'), expression('NO'[2]*'_SE'),
expression('NO'[x]*'_NO'), expression('NO'[x]*'_NE'),
expression('NO'[x]*'_C'), expression('NO'[x]*'_SO'),
expression('NO'[x]*'_SE'), expression('NO_NO'), expression('NO_NE'),
expression('NO_C'), expression('NO_SO'), expression('NO_SE'),
expression('SO'[2]*'_NO'), expression('SO'[2]*'_NE'),
expression('SO'[2]*'_C'), expression('SO'[2]*'_SO'),
expression('SO'[2]*'_SE'), expression('CO_NO'), expression('CO_NE'),
expression('CO_C'), expression('CO_SO'), expression('CO_SE'))

```

```

graf_tend <- function(name, min, max){

  png(paste(name, ".png", sep = ""), width = 1024, height = 768 )
  par(mfrow = c(4,5))

  for (i in min:max){

    fecha <- (seq(as.Date("2008/1/1"), length.out =
length(data.frame(x[i])[1:3950,1]), by="day"))

    plot(fecha,data.frame(x[i])[1:3950,1],las=1,col="gray", type = "l",
xlab = "", ylab=unid[i], main=tit[i],cex.main=2,
ylim = c(min(data.frame(x[i])[1:3950,1], na.rm =
T),max(data.frame(x[i])[1:3950,1], na.rm = T)))
lines(fecha,data.frame(x[i])[1:3950,2] ,col="blue", lwd=.000001)
lines(fecha,data.frame(x[i])[1:3950,3] ,col="green", lty=2,
lwd=.000001)
lines(fecha,data.frame(x[i])[1:3950,4] ,col="green", lty=2,
lwd=.000001)

  }
  dev.off()
}

graf_tend("grafica2", 1, 20)
graf_tend("grafica2_cont", 21, 40)

```

#Grafica de la tendencia de todo el periodo y el pronóstico a 7 días

```

graf_pron <- function(name, min, max){

  png(paste(name, ".png", sep = ""), width = 1024, height = 768 )
  par(mfrow = c(4,5))
  rango<-c(.1,.08,.05,.09,.08,.5,.2,.1,0,.7,
           .01,.5,.1,.01,0,0,0,.1,.01,0,1,.08,0,.5,
           0,1,0,.5,.8,0,.5,.1,1,.5,.5,.01,0,.01,0,0)

  for (i in min:max){

    fecha<-seq(as.Date("2018/10/25"), length.out =7 , by="day")

    plot(1:7,data.frame(x[i])[3951:3957,6], col="seagreen", xlab = "",

```

```

        ylab=unid[i], axes=T, ylim = c(
min(min(data.frame(x[i])[3951:3957,6]-rango[i], na.rm =
T)),max(data.frame(x[i])[3951:3957,2]+rango[i], na.rm = T)),
        main=tit[i], xaxt="n", lwd=1, type="b", xlim=c(1, 7), lty=1,
cex.main=2)
        axis(1, at=1:7, labels = c("Mon", "Tue", "wed", "thu", "Fri", "Sat",
"Sun"), cex.axis =1.5)
        lines(1:7,data.frame(x[i])[3951:3957,2] ,col="dimgray", lwd=1, lty=2,
type="b", pch=8)

    }

    dev.off()

}

graf_pron("grafica3", 1, 20)
graf_pron("grafica3_cont", 21, 40)

#----- FIN-----

```

Referencias

- Aditya, C., Chandana, R., Nayana, D. y Praveen, G. (2018) "Detection and Prediction of Air Pollution using Machine Learning Models", *International Journal of Engineering Trends and Technology (IJETT)*, Vol 59, No. 4, pp. 204-207.
- Aguilar, D. y Reyes, I. (2018) "A wavelet analysis of multi day extreme ozone and its precursors in Mexico City during 2015–2016", *Atmospheric Environment*, Vol. 188, pp. 112-119.
- Brantley, H., Hagler, G., Herndon, S., Massoli, P., Bergin, M. and Russell, A. (2019) "Characterization of Spatial Air Pollution Patterns Near a Large Railyard Area in Atlanta, Georgia", *International Journal of Environment Research and Public Health*, Vol 16, No. 535, pp. 1-14.
- Bravo, H., Roy-Ocotla, G., Sánchez, P. y Torres, R. (1992) "Contaminación Atmosférica por Ozono en la Zona Metropolitana de la Ciudad de México: Evolución Histórica y Perspectivas". *Sección de Contaminación Ambiental, Centro de Ciencias de la Atmósfera*, UNAM.
- Bravo, R., Sánchez, P. y Jaimes, M. (2002) "La calidad del aire en la Zona Metropolitana de la Ciudad de México y recomendaciones para mejorarla considerando conceptos básicos de ingeniería ambiental". *Revista de Ingeniería Investigación y Tecnología*, Vol. 3, No.4, pp. 185-193.
- Brockwell, P. J. y Davis, R. A. (2002) *Introduction to Time Series and Forecasting*, 2da Edición, Springer, New York.
- Camalier, L., Cox, W. y Dolwick, P. (2007) "The effects of meteorology on ozone in urban areas and their use in assessing ozone trends", *Atmospheric environment*, Vol. 41, pp. 7127-7137.
- Chang, C. y Yao, X. (2008) "Review Air pollution in megacities in China", *Atmospheric Environment* Vol. 42, pp. 1-42.
- Consejo Nacional de Población (CONAPO), Instituto Nacional de Estadística y Geografía (INEGI) y Secretaria de Desarrollo Social (SEDATU) (2018) "Delimitación de las zonas metropolitanas de México 2015", México, p. 57.
- Cowperrwait, P. y Metcalfe, A (2009) *Introductory Time Series with R*. Springer Dordrecht Heidelberg London New York, pp 55-56.
- Corona, A. y Rojas R. (2009) "Calidad del Aire y su Incorporación en la Planeación Urbana: Mexicali, Baja California, México", *Estudios Fronterizos*, Vol. 10, No.20, pp. 79 – 102.
- Cruz, M., Gómez, A., Quintero, M. y Varela, J. (2013) "Evaluación de la calidad del aire respecto de partículas suspendidas totales (PST) y metales pesados (Pb, Cd, Ni, Cu, Cr) en la ciudad de Hermosillo, Sonora, México, durante un periodo anual", *Revista internacional de contaminación ambiental*, Vol. 29, No. 4, pp. 269-283.
- Davis, L. (2008) "The effect of driving restrictions on air quality in Mexico City", *Journal of Political Economy*, Vol. 116, No.1, pp. 38-81.
- Eilers, P. H. C. y Marx, B. D. (1996), "Flexible Smoothing with B-Splines and Penalties", *Statistical Science*, Vol. 11, No. 2, pp. 89–102.
- Eilers, P. H. C., Marx, B. D. y Durbán, M. (2015), "Twenty Years of P-Splines", *Statistics and Operations Research Transactions*. Vol. 39, No. 2, pp. 149–186.

- Infante, J. y Rey (2015), *Métodos numéricos, teoría, problemas y prácticas con MATLAB*. 4° edición, Madrid. pp. 264-274. Consultado en: <https://ebookcentral.proquest.com/lib/bibliodgbsp/reader.action?docID=4626678#>
- Gallego, F., Montero, J. y Salas, C. (2013a) “The effect of transport policies on car use: a bundling model with applications”, *Energy Economics*, Vol. 40, No. 1, pp. 85–97.
- Gallego F., Montero, J. y Salas, C. (2013b) “The effect of transport policies on car use: evidence from Latin American cities”, *Journal of Public Economics*, Vol. 107, pp. 47–62.
- Georgieva, S., Valev, A., Stoyanova, D. y Todorov, D. (2014) “Time series analysis and forecasting for air pollution in small urban area: an SARIMA and factor analysis approach”, *Stoch Environ Res Risk Assess*, Vol. 28, pp. 1045–1060.
- Gong, B. y Ordires, J. (2015) “Prediction models for ozone in metropolitan area of Mexico City based on artificial intelligence techniques”. *Journal Information and Decision Sciences*, Vol. 7, No 2, pp. 115-138.
- “Reconfiguring existing pollutant monitoring station increasing the value of the gathered information”. *Journal Environmental Modelling & Software*, Vol. 96 pp. 106-122.
- Gramsh, E., Cereceda-Balic, F., Oyola, P. y Von Baer, D. (2006) “Examination of pollution trends in Santiago de Chile with cluster analysis of PM₁₀ and Ozone data”, *Atmospheric Environment*, Vol. 40, pp. 5464-5475.
- Guerrero, V. M. (2007) “Time series smoothing by penalized least squares”. *Statistics & Probability Letters*, Vol. 77, pp. 1225–1234.
- (2008) “Estimating Trends with Percentage of Smoothness Chosen by the User”, *International Statistical Review*, Vol.76, No. 2, pp. 187–202.
- Guerrero, V. M., Islas, A. y Ramirez, L. (2017) “Trend estimation of multivariate time series with controlled smoothness”. *Communications in Statistics -Theory and Methods*, Vol. 46, pp. 6704-6726.
- Guerrero, V. M. y Silva, E. (2015) “Smoothing a Time Series by Segments of the Data Range, Communications in Statistics”, *Theory and Methods*, Vol.44, No.21, pp. 4568-4585.
- Harvey, A. (1990) *Forecasting, Structural Time Series Models and the Kalman Filter* Cambridge: Cambridge University Press.
- Hernández, J., Morales, C. y Madrigal, D. (2004) “Comportamiento del monóxido de carbono y el clima en la ciudad de Toluca, de 1995 a 2001”, *Ciencia Ergo Sum*, Vol. 11, No 3, pp. 263-274.
- Hernández, L. (2009) “Modelado atmosférico para determinar niveles máximos diarios de ozono en la Ciudad de Guadalajara”, *Universidad Autónoma Metropolitana (UAM). División de Ciencias Básicas e Ingeniería. Departamento de matemáticas.*
- Hernández, I., Clemitshaw, K. y Mendoza, A. (2017) “Observed trends in ground-level O₃ in Monterrey, Mexico, during 1993–2014: comparison with Mexico City and Guadalajara”, *Atmos. Chem. Phys*, Vol. 17, 9163–9185.
- Hodrick, R. y Prescott, E. (1997) “Postwar U.S. Business Cycles: An Empirical Investigation”, *Journal of Money, Credit and Banking*, Vol. 29.
- Huerta, G. y Sansó, B. (2007) “Time-varying Models for extreme values”, *Environmental and Ecological Statistics*, Vol. 14, pp. 285-299.
- Jaimes, M. (2017) “Diseño del monitoreo de compuestos precursores de ozono en la atmosfera de la Ciudad de México y su Área Metropolitana”. Tesis doctoral en Ingeniería ambiental-aire. Universidad Nacional Autónoma de México.

- Jaiswal, A., Samuel, C. y Kadabgaon, V. (2018) “Statistical trend analysis and forecast modeling of air pollutants”, *Global Journal of Environmental Science and Management*, Vol. 4, No. 4, pp. 427-438.
- Jaramillo. M., González. D., Nuñez, M., Portilla., y G. Lucio. H. (2007) “Análisis de series de tiempo univariante aplicando metodología de Box-Jenkins para la predicción de ozono en la ciudad de Cali, Colombia”, *Revista Facultad de Ingeniería*, N. 39, pp. 79-88.
- Kalman, R.E. (1960) “A New Approach to Linear Filtering and Prediction Problems”, *Transactions of the ASME—Journal of Basic Engineering*, Vol. 82, pp. 35-45.
- Krivobokova, T, Craineceanu, C.M., y Kauermann (2008) “Fast Adaptative Penalized Splines”. *Journal of computational and Graphical Statistics*, Vol, 17, No. 1, pp. 1-20.
- Kumar, U. and De Ridder, K. (2010) “GARCH modelling in association with FFT e ARIMA to forecast ozone episodes”, *Atmospheric Environment*, Vol. 44, pp. 4252-4265.
- León, R.J y Constanzo, J. (2006) Splines cúbicos. Universidad Técnica Federico Santa María. España Valparaiso.
- Lucas W. Davis (2017) “Saturday Driving Restriction Fail to Improve Air Quality in Mexico City”. *Scientific Reports*. Vol. 7. Article number 41652.
- Maravall, A. (1993) Stochastic linear trends. Models and estimators, *Journal of Econometrics*, Vol. 56, pp. 5–37.
- Nakata, T y Tonetti, C. (2010) “Kalman Filter and Kalman Smoother”. New York University
- OCDE (2015) Territorial Reviews: Valle de Mexico, Mexico.
- Palomares, J., González, J., Agüera, A., Sierra, J. y Florencias, O. (2019) “Forecasting PM10 in the Bay of Algeciras Based on Regression Models”, *Sustainability*, Vol. 11, 968.
- Petris, G., Petrone, S. y Campagnoli, P. (2007) *Dynamic Linear Models with R*, Berlin Heidelberg New York, Hong Kong London y Milan Paris Tokyo, Springer.
- Rodríguez, Gloria. (2002) “Representación en el espacio de los estados y filtro de Kalman en el contexto de las series temporales económicas”. *Documentos de trabajo conjuntos: Facultades de Ciencias Económicas y Empresariales*, No. 5, p.2.
- Rodríguez, S., Huerta, G., Reyes, H. (2016) “A study of trends for Mexico City ozone extremes: 2001-2014”, *Atmósfera* Vol. 29 No. 2, pp. 107-120.
- Secretaría del Medio Ambiente (SEDEMA). (2017) *Calidad de en la Ciudad de México, Informe 2016*. Dirección General de Gestión de la Calidad del Aire, Dirección de Monitoreo Atmosférico. Ciudad de México. 9-10, 52-81.
- Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT) (2016) *Informe de la Situación del Medio Ambiente en México. Compendio de Estadísticas Ambientales. Indicadores Clave de Desempeño Ambiental y de Crecimiento Verde*. Edición 2015 SEMARNAT. México. 2016. P. 261.
- So, K., Guo, H. y Li, Y. (2007) “Long-term variation of PM2.5 levels and composition at rural, urban, and road side sites in Hong Kong: Increasing impact of regional air pollution”, *Atmospheric Environment*, Vol. 4, pp. 9427-9434.
- Velazco, E. y Ratama, A. (2017) “Ozone’s Threat hits back Mexico City”. *Secretaría del Medio Ambiente de la Ciudad de México, Dirección de Monitoreo Atmosférico*.

- Vita, Y., Apriliani, E. y Khusnul, D. (2018) "Fractional kalman filter to estimate the concentration of air Pollution", *Journal of Physics: Conference Series*. Ser. 1008 012008.
- Yeongkwon, S., Osornio-Vargas, A., O'Neill, M., Hystad, P., Texcalac-Sangrador, J., Ohman-Strickland, P., Qingyu, M. y Schwander, S. (2018) "Land use regression models to assess air pollution exposure in Mexico City using finer spatial and temporal input parameters". *Science of The Total Environment*, Vol. 639, pp. 40-48.
- Zamri, M., Zailan, R., Ismail, M. y Safiih, M. (2009) "Forecasting and Time Series Analysis of Air Pollutants in Several Area of Malaysia". *American Journal of Environmental Sciences*, Vol 5 No.5, pp. 625-632.