

# Trend estimation and forecasting of atmospheric pollutants in the Mexico City Metropolitan Area through a non-parametric perspective

Elba RAMOS-IBARRA<sup>1</sup> and Eliud SILVA<sup>2\*</sup>

<sup>1</sup>Centro de Investigación en Matemáticas A.C (CIMAT), 36023 Guanajuato, Guanajuato, México.

<sup>2</sup>Facultad de Ciencias Actuariales, Universidad Anáhuac México, 52786 Huixquilucan, Estado de México, México.

\*Corresponding author; e-mail: jose.silva@anahuac.mx

Received: August 20, 2019; accepted: February 4, 2020

## RESUMEN

Se estiman los pronósticos y tendencias de los principales contaminantes atmosféricos ( $O_3$ ,  $SO_2$ ,  $NO_2$ ,  $CO$ ,  $PM_{10}$ ,  $PM_{2.5}$ ,  $NO$  y  $NO_x$ ) por regiones en la Zona Metropolitana de la Ciudad de México (ZMCM) con valores máximos diarios de 2008 a 2018. Se utilizó una técnica controlada no paramétrica de suavizado basada en el filtro de Hodrick y Prescott y estimada mediante el filtro de Kalman. Se generan tanto estimaciones puntuales como de intervalo, junto con sus respectivos pronósticos. Las estimaciones se confrontaron con la Norma Ambiental para la Ciudad de México (NADF-009-AIRE-2017) y es evidente que, en general, aún están lejos de cumplir con un estándar de buena calidad del aire en la ZMCM, en comparación con el  $CO$  y el  $NO_2$ . Las tendencias y pronósticos del resto de los contaminantes superan ampliamente los límites permitidos.

## ABSTRACT

Trends and forecasts of the main atmospheric pollutants ( $O_3$ ,  $SO_2$ ,  $NO_2$ ,  $CO$ ,  $PM_{10}$ ,  $PM_{2.5}$ ,  $NO$  and  $NO_x$ ) are estimated by regions in the Mexico City Metropolitan Area (MCMA) with maximum daily data from 2008 to 2018. A non-parametric statistical smoothing controlled technique based on the Hodrick and Prescott filter and estimated through the Kalman filter, is used. Both point and interval estimates, as well as their respective forecasts are generated. Estimates are compared against the environmental standard for Mexico City (NADF-009-AIRE-2017), and it is evident that, in general, they are still distant from good air quality in the MCMA, as opposed to  $CO$  and  $NO_2$ . The remaining pollutants have trends and forecasts that are far from the permissible limits.

**Keywords:** trend, controlled smoothing, smoothness index, smoothing parameter, Hodrick and Prescott filter, Kalman filter, forecasts, atmospheric pollutants.

## 1. Introduction

According to the Organization for Economic Cooperation and Development (OECD, 2015), the Mexico City Metropolitan Area (MCMA) is the third metropolis with greatest population concentration. The Consejo Nacional de Población (National Population Council) (CONAPO-INEGI-SEDATU, 2018) reports that it has 20.9 million people in an area of 7866 km<sup>2</sup>, located southwest of the basin of Mexico, at an

altitude of 2240 masl, and comprises 76 municipalities. Several papers such as Bravo et al. (2002), SEDEMA (2017) and Rodríguez et al. (2016) agree that, due to its geography and orographic characteristics, wind circulation is impeded. With the added dynamics of population mobility and industrialization, this favors the accumulation of atmospheric pollutants.

On the other hand, the OECD (2015) suggests that the one-dimensional and sectorized administration

of the MCMA does not allow the generation of comprehensive public policies in the long term. The Secretaría de Medio Ambiente y Recursos Naturales (Ministry of Environment and Natural Resources) (SEMARNAT, 2016) states that air quality, in addition to being affected by climatic and geographical factors, is related to the concentration and characteristics of pollutants in the region, therefore understanding its behavior is essential to determine the concentration levels of the main pollutants.

The main objective of this study is to estimate trends of atmospheric pollutants in the MCMA and their respective forecasts. The considered pollutants are ozone ( $O_3$ ), sulphur dioxide ( $SO_2$ ), nitrogen dioxide ( $NO_2$ ), carbon monoxide (CO), particles less than 10 and 2.5  $\mu m$  ( $PM_{10}$  and  $PM_{2.5}$ , respectively), nitrogen oxide (NO) and nitrogen oxide ( $NO_x$ ). The study applies the non-parametric statistical technique of controlled smoothing proposed by Guerrero (2008), which is based on the Hodrick and Prescott filter. Daily observations are taken from January 2008 to October 2018. The trends are estimated within five regions that constitute the MCMA, all of which have estimation intervals. Data are extracted from the Red Automática de Monitoreo Ambiental (Automatic Network of Atmospheric Monitoring, RAMA).

It should be mentioned that this work does not intend to show the efficiency or inefficiency of programs formulated by authorities in charge of air quality control in the MCMA (for this see, e.g., Gallego et al. [2013a, b] and Davis [2008], among others). Likewise, the study of the adverse effects of air pollution on the morbidity and mortality of exposed population is left aside. It is important to note that this study only addresses those air pollutants monitored by RAMA.

The work is structured as follows. In the next section, several researches that show the study of trends and forecasting of pollutants, both in Mexico and some other cities around the world are cited. The methodology then estimates and predicts the behavior of selected air pollutants, which is discussed in detail below. Subsequently, the origin of data and its processing for the appropriate application of the methodology is exposed. Estimates of the trends are illustrated in the results, and their forecasts are contrasted with NADF-009-AIRE-2017 (Gaceta Oficial de la Ciudad

de México, 2018). Finally, conclusions are addressed.

## 2. Background

This section presents various statistical analyses on estimates of pollutants carried out for the MCMA and other Mexican cities. Emphasis is consequently placed on the methods used to estimate and forecast pollutants in other cities around the world. This exposition is given chronologically in the following regional order: Mexico, Latin America, North America, Europe and Asia.

### 2.1 MCMA and other Mexican cities

Huerta and Sansó (2007) illustrate the behavior of the daily extreme values of  $O_3$  for the center of the MCMA from 1990 to 2002 through a method for both time and space. Under a Bayesian approach, they define the temporal component as a dynamic linear model (DLM) and the spatial element imposed by the evolution matrix of the DLM. As a result, the maximum levels of  $O_3$  decreased in that period, as well as in the region.

On the other hand, Gong and Ordires (2015) predict the maximum daily concentrations of ozone in the MCMA through several models of artificial intelligence: multiple linear regression, neural networks, support vector machines, random forest and two assembly techniques. They find that prediction errors in relation to the current day are around 50%. Likewise, with maximum values of  $O_3$  in Mexico City, Rodríguez et al. (2016) present a trend analysis for this pollutant in the five regions, between 2001 and 2014, by means of a generalized distribution model of extreme values, using Open-BUGS/Winbugs. The results show that the monthly maximum of  $O_3$  decreased during this period.

Later, Aguilar and Reyes (2018) perform an analysis through Haar wavelet transformation for extreme values of  $O_3$  and its precursors ( $NO_2$  and CO) in Mexico City from 2015 to 2016. During these years, notorious changes were implemented in driving restrictions by the local government. The results showed that for multi-day events of  $O_3$  exhibit, T periods are greater than four days, while for  $NO_2$  and CO they are greater than two. In addition, it is shown that these air pollutants are multi-temporal and comprise a correlated group.

By means of a space-time regression, Yeongkwon et al. (2018) study the behavior of the concentrations of  $PM_{2.5}$ ,  $PM_{10}$ ,  $O_3$ ,  $NO_2$ ,  $CO$  and  $SO_2$  in Mexico City. Through models that consider hourly, daily, monthly, semiannual and annual averages, they deduce satisfactory results in the forecasts, except for the series of  $PM_{10}$ ,  $PM_{2.5}$  and  $SO_2$ .

In other cities of Mexico, the problem of air pollution and its monitoring has also become visible; in this regard, Hernández et al. (2004) present a geospatial analysis of  $CO$  with monthly and annual averages from 1995 to 2001 in Toluca (capital of the State of Mexico). They use the software Surfer 7.2 and determine how the dispersion of the  $CO$  is favored with winds and precipitations, where the norm established is only surpassed during the winter, a situation that does not happen in the MCMA.

Hernández (2009) carries out a study for the Metropolitan Zona of Guadalajara (MZG), in center-west Mexico, to predict the daily maxima of  $O_3$  monitored by seven stations from 1997 to 2006. By means of the theory of extreme values, this study verifies that with high temperatures and low wind speeds, an increase in  $O_3$  is obtained. In addition, when using non-stationary and atmospheric information (wind direction, relative dampness, temperature, wind speed), it is possible to predict the daily maximum of  $O_3$  in order to have a monitoring activity that protects the population from dangerous concentrations.

Corona and Rojas (2009) perform an air quality diagnosis of the city of Mexicali, Baja California (in the northwest of Mexico) through concentration data of pollutants, which were interpolated with the Surfer 8 software. It was found that air quality (measured through the number of days in which the standard is exceeded) was not satisfactory in the period from 1997 to 2005.

For the city of Hermosillo, Sonora, Cruz et al. (2013) found that concentrations of heavy metals (Pb, Cd, Ni, Cu, and Cr) in samplings of suspended particles were below the maximum permissible; however, air quality was not satisfactory. Hernández et al. (2017) made estimates of long-term trends of  $O_3$  in the Metropolitan Zone of Monterrey (MZM) and compared it with the MCMA and MZG through the analysis of hourly, daily and monthly averages. The estimates were elaborated with the

tools openair, WindRose, timeVariation and TheilSen of the R software, for the period from 1993 to 2014. It was observed that the trend of  $O_3$  for the MZM increases in spring, summer and autumn, compared to the downward trend of the MCMA. There was not a clear trend for the MZG in the reference period.

## 2.2 Other cities

Air quality is currently not satisfactory in various urban areas around the world, so there is a need to conduct studies that establish and anticipate the behavior of air pollutants. In this sense, Gramsh et al. (2006) analyze contaminant trends in Chile using clusters to identify patterns; in particular,  $PM_{10}$  and  $O_3$  are examined, evidencing that concentration levels vary with seasonality. For the case of  $PM_{10}$ , concentrations are higher in winter and  $O_3$  in summer. The topographic and meteorological characteristics of the evaluated areas play an important role in the recognized patterns.

Later, Jaramillo et al. (2007) refer to the importance of  $O_3$  forecast studies in the city of Cali, Colombia, using the Box-Jenkins methodology to analyze the period from April to July 2003. From 2496 pieces of weekly data analyzed, the first 2232 were used for the estimate, while the rest were used to corroborate the results of the model. The best model had the predictive capacity of 8 h, so that preventive measures could be taken beforehand.

In North America, Camalier et al. (2007) describe and exemplify a generalized linear model (GLM) in which they relate  $O_3$  with meteorological variables to infer its trends in 39 cities of the USA. They establish a separate model for each city and propose another model with splines for the non-linear relationships that exist among the meteorological components,  $O_3$  and seasonal changes. From this analysis, it is evident that  $O_3$  rises when temperature increases and humidity decreases.

Brantley et al. (2019) applied regression by quantiles with natural cubic splines to estimate the trends of  $CO$ ,  $NO$ ,  $NO_2$ ,  $NO_y$  (oxides of nitrogen), BC (black carbon) and other air pollutants, in a railyard area of Atlanta (Georgia, United States). Among several findings, they observed similar trends in  $NO$  and  $NO_y$  as well as differentiated concentrations of pollutants according to their location.

Kumar and de Ridder (2010) made forecasts of O<sub>3</sub> maximum daily concentrations using the GARCH model in association with the FFT-ARIMA model for four urban areas of Brussels and London. They found that certain factors, such as traffic, play an important role in the concentration of the pollutant, since the structure of the GARCH model depends on the site, unlike the FFT-ARIMA model; therefore, the use of the GARCH model not only improves the intervals of prediction, but also makes more accurate forecasts of O<sub>3</sub> maxima.

In Bulgaria, Georgieva et al. (2014) analyze the concentration of pollutants (NO, NO<sub>2</sub>, NO<sub>x</sub>, PM<sub>10</sub>, SO<sub>2</sub> and O<sub>3</sub>) from 2011 to 2012 to assess air quality, using factor analysis (FA) and the Box-Jenkins methodology. Multicollinearity was found and three factors were established: F1 = {NO<sub>2</sub>, NO, NO<sub>x</sub>, PM<sub>10</sub>}; F2 = {O<sub>3</sub>}; and F3 = {SO}, explaining 90.74% of the total variance. On the other hand, SARIMA models were estimated for each of the six pollutants using a forecast horizon of 72 h, with results observed to be accurate. It was detected that in the case of PM<sub>10</sub>, higher concentrations are present in winter, exceeding national and European standards.

To predict average concentrations of PM<sub>10</sub> in three urban areas of Andalusia (Spain), Palomares et al. (2019) use five methods, divided into parametric (persistence model and multiple linear regression) and nonparametric (adaptive linear neuron, multi-layer and radial basis function). The data correspond to the period from 2005 to 2010 and meteorological measurements are used as exogenous variables for the conformation of the models. It was found that the forecasts with non-parametric models were better, therefore including information on weather conditions improves their predictive capacity. However, such models have a disadvantage, since they require a large amount of data.

In Asia, for Hong Kong, So et al. (2007) carried out an ANOVA analysis of PM<sub>2.5</sub> where long-term trends and spatial variations were evaluated, considering roads, urban environments and rural environments. Samples collected every six days for 12 months were analyzed in strategic sites. Among several results, seasonal variations were found similar: high in autumn and winter and low in summer; moreover, the behavior of PM<sub>2.5</sub> was not similar in the areas considered.

Chang and Yao (2008) recognize an association between economic development and emissions of air pollutants in densely populated cities that account for 20% of the GDP, which identifies the immediate challenges to understand and control such problems. They review long-term trends of air pollutants for the case of three cities in China, through chemical analysis and ANOVA, finding that the increase in pollutants in cities is similar: high in autumn and winter and lower in summer.

Zamri et al. (2009) obtain forecasts of CO and NO<sub>2</sub> of maximum monthly concentrations through the Box-Jenkins methodology. The data covers the period from 1997 to 2006 in four areas of Malaysia. They emphasize that, in general, the 2016 forecasts do not exceed the permissible limits. The prediction of PM<sub>2.5</sub> in Beijing is seen in Aditya et al. (2018), where several machine learning models are compared, and logistic regression is chosen. Daily data of the following variables are considered: temperature, wind speed, dew point and pressure. Likewise, for prediction purposes, an autoregressive model (AR) is used and it is affirmed that these models are efficient to predict PM<sub>2.5</sub> levels.

Jaiswal et al. (2018) study the annual trend of CO, NO<sub>2</sub>, SO<sub>2</sub>, PM<sub>2.5</sub> and PM<sub>10</sub> from 2013 to 2016 for Varanasi, India. They use the Mann-Kendall test and perform their forecast using an ARIMA model. The results show that PM<sub>2.5</sub>, CO, NO<sub>2</sub> and SO<sub>2</sub> have a decreasing tendency, unlike PM<sub>10</sub>, which shows an upward trend.

Vita et al. (2018) develop an algorithm using the fractional Kalman filter (FKF) method to predict the concentrations of the following pollutants: O<sub>3</sub>, NO, NO<sub>2</sub>, SO<sub>2</sub>, PM<sub>2.5</sub> and PM<sub>10</sub>. The Air Pollution Model and Chemical Transport Model Model (TAPM-CTM) is used to measure the concentrations of air pollutants. They compare the results with the Kalman filter and show that the model with FKF has better precision.

The diverse methodologies that have been used to estimate trends and forecasts of different atmospheric pollutants have varied according to the intended objective, being the most important GLM, DLM, extreme values, Box-Jenkins (ARIMA), clusters, splines, Bayesian approaches and machine learning algorithms. The method applied in this study could be considered as simpler, because it has no distributional assumptions to satisfy and its implementation and



maintenance are economic. In fact, the analyst with this proposal has a fast description of the pollutant trends and their dynamics. Only empirical evidence of the pollutants under study is used, thus leaving additional variables aside.

### 3. Methodology

Maravall (1993) documents several methods for estimating trends. Among them is the Hodrick and Prescott (1997) filter, which does not require the application of a formal statistical model for the estimation of trends and forecasts, unlike for example, ARIMA or the structural models (Guerrero, 2008).

The Hodrick and Prescott filter has its origin in the so-called Whittaker-Henderson method (Hodrick and Prescott, 1997), which was first used by actuaries to smooth mortality tables, which was also useful in astronomy. The authors define it as the recurrent fluctuations in real activity with respect to a trend. Fluctuations are deviations from a smooth but variable path (trend), which can be estimated by a process of adjustment to a smooth curve. In this sense, for the analysis of trends and forecasts of the selected pollutants in the MCMA within this work, it should be noted that such time series are not seasonally adjusted and only the data trend is developed.

#### 3.1 Controlled smoothing and forecasts

The approach used for the Hodrick and Prescott filter through the Kalman filter is followed as suggested by Guerrero (2007 and 2008). It encapsulates a set of data observed over time (technically a time series) composed by the sum of a non-observable trend and a random component:

$$y_t = \tau_t + \eta_t, \quad \text{for } t = 1, \dots, N \quad (1)$$

where  $\tau_t$  is the trend,  $\eta_t$  is the random component from which no specific distribution is assumed with  $Var(\eta_t) = \sigma_\eta^2$  and  $N$  is the number of data. Through the controlled smoothing method, the random component is gradually reduced, inducing smoothness. Smoothed values are estimated with this method using an index that will be discussed later. The filter is simple to apply for any data series. The structure of the filter comes from the following expression:

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

where  $\lambda$  is the smoothing parameter that penalizes for smoothness. In the first term, the goodness of fit of the trend with the observed data is sought, while the second is the induction of smoothness. When the parameter approaches zero ( $\lambda \rightarrow 0$ ) the estimated trend approaches the observed data, and when the value of the parameter is  $\infty$  ( $\lambda \rightarrow \infty$ ) the trend tends to a straight line. In a matrix notation, we have

$$\min_{\tau} M(\lambda) = (\mathbf{y} - \boldsymbol{\tau})' + (\mathbf{y} - \boldsymbol{\tau}) + \lambda (\mathbf{K}_2 \boldsymbol{\tau})' (\mathbf{K}_2 \boldsymbol{\tau}) \quad (3)$$

where  $\mathbf{y} = (y_1, \dots, y_N)'$ ,  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N)'$  and  $\mathbf{K}_2$  is the matrix  $(N - 2) \times N$  given by

$$\mathbf{K}_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 & 1 \end{bmatrix}$$

The solution to the problem is

$$\hat{\boldsymbol{\tau}} = (\mathbf{I}_N + \lambda \mathbf{K}_2' \mathbf{K}_2)^{-1} \mathbf{y} \quad (4)$$

where  $\hat{\boldsymbol{\tau}}$  represents the estimated trend given a value of  $\lambda$ , while the variance of  $\hat{\boldsymbol{\tau}}$  in terms of mean square error is given by  $Var(\hat{\boldsymbol{\tau}}) = (\mathbf{I}_N + \lambda \mathbf{K}_2' \mathbf{K}_2)^{-1} \sigma_\eta^{-2}$ , where  $\mathbf{I}_N$  is the  $N$  dimensional identity matrix. Using the positive square root of the diagonal elements from  $Var(\hat{\boldsymbol{\tau}})$ , it is possible to construct intervals for estimated trends.

Given that the second derivative of  $M(\lambda)$  evaluated in  $\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}$  is a symmetric and positive matrix, we see where (4) produces a minimum. It should be noted that to obtain  $\hat{\boldsymbol{\tau}}$  the matrix of dimension  $N \times N$  has to be inverted, so this calculation can cause instability and lack of precision of the numerical solution when  $N$  is large; in addition to obtaining (4), it is required that there is no missing data in the series. In this context, the Hodrick and Prescott filter estimated through the Kalman filter is a solution, in which it is required to formulate a state-space model as follows:

$$x_t = A_t x_{t-1} + w_t, y_t = c_t' x_t + \eta_t, \quad (5)$$

with  $x_t = \begin{pmatrix} \tau_t \\ \tau_{t-1} \end{pmatrix}$ ,  $A_t = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$ ,  $c_t' = (1 \ 0)$  and  $w_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}$ , where  $\varepsilon_t$  and  $\eta_t$  are two independent random errors of zero mean, uncorrelated and identically distributed with  $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$  and  $\text{var}(\eta_t) = \sigma_\eta^2$ . Therefore, the equation of state has the following form:

$$\tau_t = 2\tau_{t-1} - \tau_{t-2} + \varepsilon_t \quad (6)$$

where, according to Guerrero (2008),  $\lambda = \sigma_\eta^2 / \sigma_\varepsilon^2$ . To equate the results of the Kalman filter with smoothing with those obtained by (4), it is assumed that  $\sigma_\varepsilon^2 = 1$  and  $\sigma_\eta^2 = \lambda$ .

On the other hand, in Guerrero (2007) the forecast of the trend,  $\hat{\tau}_N(h)$ , for the period  $h$  of  $\tau_{N+h}$  with origin in  $N$  for  $h \geq 1$  and  $d = 2$ , is

$$\hat{\tau}_N(h) = [h(h+1)/2]\mu + (h+1)\tau_N - h\tau_{N-1} \quad (7)$$

where we assume  $\mu = 0$ . Likewise, in order to obtain forecasts by intervals, all the available data is considered, and both the trend and its respective estimation interval are recalculated.

### 3.2 Smoothness index

The smoothness index (proposed by Guerrero, 2008), which measures the smoothness of the trend, is given by

$$S(\lambda; N) = 1 - \text{tr} \left[ (I_N + \lambda K_2' K_2)^{-1} \right] / N \quad (8)$$

where  $\text{tr}(\cdot)$  represents the matrix trace. Note that when  $\lambda \rightarrow 0$  the smoothness  $S(\lambda; N) \rightarrow 0$ , therefore when  $\lambda \rightarrow \infty$  we have that  $(\lambda; N) \rightarrow 1$ . In other words, the index is between zero and one. The index for practicality can be expressed in percentage terms. It can be shown that the maximum value of the achievable smoothness index for a set of  $N$  data is given by  $1 - 2/N$ . In particular, given the amount of data per pollutant available for this study ( $N = 3950$ ), the maximum smoothness achievable is 99.95%.

Note that different indexes of smoothness,  $S(\lambda; N)$ , generate different trends; therefore, there is no single trend or, under such optic, an optimal trend. In fact, an infinite number of trends could be estimated from a single set of specific data. With this method, estimations can be made from a trend that is exactly

the set of observed data to another that is a straight line at the limit. The problem focuses on deciding what percentage of smoothness is appropriate. This technique has been used for time series in economics and demographic issues; however, there is no history of its use in environmental issues.

To decide the percentage of smoothness through the said index, Guerrero et al. (2017) suggest a certain index of smoothness based on the structure of the data. However, the sample sizes of the time series in that work very different: in this document there are 3950 observations per series compared to 48, 100 and at most 400 observations in that reference. Additionally, the applications illustrated by these authors have substantially less variability than that observed in the series analyzed here. In fact, it should be noted that the index,  $S(\lambda; N)$ , depends on two parameters:  $\lambda$  and  $N$ . Therefore, the larger the sample size, the greater the index of smoothness, i.e., when  $N \rightarrow \infty$  we have that  $S(\lambda; N) \rightarrow 1$ . With this, it is clear that with the amount of data used here, significantly large smocks are required.

The proposal to impose a pertinent smoothness index in this work consists of choosing one that generates trends that are not affected by extreme observations or seasonal effects; this is, a high index of smoothness is imposed such that in the estimated trends, it mitigates the said effects. When a change in the variability of the respective pollutant is detected, a segmented analysis can be done as proposed by Guerrero and Silva (2015). In particular, a percentage of smoothness close to 1 will be imposed by the amount of data, which will overcome the possible generation of spurious trends. After multiple tests, a smoothness index of 99.91% (which corresponds to a smoothing parameter of  $\lambda = 1 \times 10^{11}$ ) is then used for the estimation of contaminant trends.

To choose the appropriate  $\lambda$ , we first selected the observed time series with the highest variance and with seasonality effects. In this case, it was the  $O_3$  time series. Secondly, we tested different values of  $\lambda$ . With the chosen value it was possible to appreciate the underlying trend (see Fig. SM1 in the supplementary material). Third, we estimated the rest of the trends with a similar  $\lambda$ , so that the same smoothness index was imposed for all the pollutants series. This enabled us to make valid comparisons between all the estimated trends.

On the other hand, for the estimation of  $Var(\hat{\tau})$  and  $S(\lambda; N)$  the effects of any discrepancies in terms of differentiated quality of the data used are mitigated from the statistical point of view, since the estimated trends are standardized with the use of an identical smoothness index for all pollutant series. This does not necessarily occur when using other statistical techniques, where there are automatic criteria for the selection of the smoothing parameter. It is also assumed that the quality of the data for each pollutant is the same throughout the RAMA.

### 3.3 Precision measurements of forecasts

The accuracy of pollutant forecasts is assessed using the following precision measures (see Petris et al., 2007):

Mean absolute deviation (MAD):

$$MAD = \frac{1}{N} \sum_{t=1}^N |e_t| \quad (9)$$

Mean square error (MSE):

$$MSE = \frac{1}{N} \sum_{t=1}^N e_t^2 \quad (10)$$

Mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{|e_t|}{y_t} \quad (11)$$

where  $e_t = y_t - \hat{y}_t$ ,  $\hat{y}_t = \hat{\tau}$ , which refers to the forecast error in period  $y_t$  and  $\hat{y}_t$ . This represents the real and predicted values of the trend in period  $t$ . For all these measures, the smaller the value, the better fit the data set to the estimated trend (see Klimberg et al., 2010).

### 3.4 Delimitation of the MCMA and monitoring network location

The spatial behavior of air pollutants is not homogeneous for the entire MCMA (e.g., SEDEMA, 2016, 2018, as well as other references available at [shorturl.at/cuAG1](http://shorturl.at/cuAG1)). For this reason, the MZCM was divided into five regions according to the localization of the monitoring units (Rodríguez et al., 2016) and air pollutant reports from the Atmospheric Monitoring System (Mexico City's atmospheric monitoring system, SIMAT) ([shorturl.at/yzQ03](http://shorturl.at/yzQ03)). This regions are: northwest (NW), northeast (NE), center (C), southwest (SW) and southeast (SE). It is considered that the estimated trends are important to describe the

main characteristics of the five regions studied, each of which contain 34 monitoring units that, according to information provided by SIMAT and RAMA, could be located as shown in Figure 1.

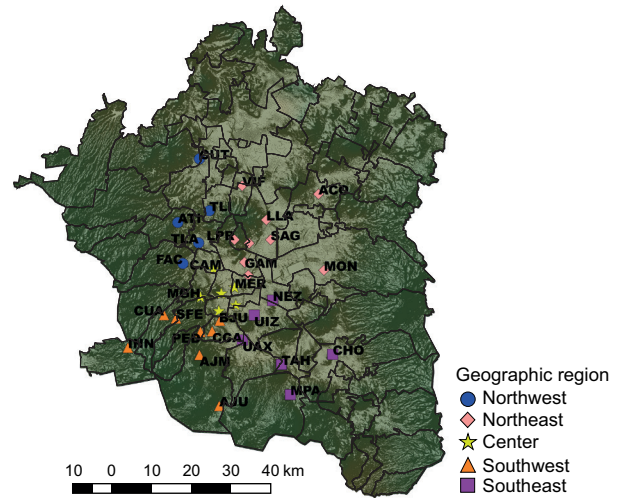


Fig. 1. Location of the monitoring network. Source: own elaboration.

The type of pollutant monitored in each of the centers was identified. However, since they are not measured constantly, an evaluation was made of each monitoring unit and pollutant by region. Thus, only those centers that recorded information during the 10 years of the study (2008-2018) were taken into account. The catalogue of the 20 stations by regions that monitor the selected pollutants is shown in Table I.

### 3.5 Data

RAMA monitors the concentrations of eight pollutants ( $O_3$ ,  $SO_2$ ,  $NO_2$ ,  $NO$ ,  $NO_x$ ,  $PM_{10}$ ,  $PM_{2.5}$  and  $CO$ ) every hour. Since each pollutant can be captured by more than one monitoring station in the five geographical regions, the maximum daily concentration reached was identified in each station in order to extract its daily maximum in a specific region. This was done during the period from January 1, 2008 to October 24, 2018, obtaining a total of 3950 daily data. Except for  $O_3$  in the northeast region and  $SO_2$  in the northwest and northeast regions, the rest of the pollutants had a certain percentage of missing data (Table II).

Table I. Monitoring stations by geographic region.

Region	Station	Abbreviation	Location	Pollutant
Northwest	FES Acatlán	FAC	Naucalpan de Juárez, Estado de México	O <sub>3</sub> , NO <sub>2</sub> , NO <sub>X</sub> , NO, SO <sub>2</sub> , CO, PM <sub>10</sub>
	Tlalnepantla	TLA	Tlalnepantla de Baz, Estado de México	O <sub>3</sub> , NO <sub>2</sub> , NO <sub>X</sub> , NO, SO <sub>2</sub> , CO, PM <sub>10</sub> , PM <sub>2.5</sub>
	Atizapán	ATI	Atizapán de Zaragoza, Estado de México	NO <sub>2</sub> , NO <sub>X</sub> , NO
	Tultitlán	TLI	Tultitlán, Estado de México	NO <sub>2</sub> , NO <sub>X</sub> , NO, SO <sub>2</sub> , CO, PM <sub>10</sub>
Northeast	Acolman	ACO	Acolman, Estado de México	O <sub>3</sub> , SO <sub>2</sub> ,
	Montecillo	MON	Texcoco, Estado de México	O <sub>3</sub>
	San Agustín	SAG	Ecatepec de Morelos, Estado de México	O <sub>3</sub> , NO <sub>2</sub> , NO <sub>X</sub> , NO, SO <sub>2</sub> , CO, PM <sub>10</sub> , PM <sub>2.5</sub>
	Xalostoc	XAL	Ecatepec de Morelos, Estado de México	O <sub>3</sub> , NO <sub>2</sub> , NO <sub>X</sub> , NO, SO <sub>2</sub> , CO, PM <sub>10</sub> , PM <sub>2.5</sub>
	Villa de las Flores	VIF	Coacalco de Berriozábal, Estado de México	NO <sub>2</sub> , NO <sub>X</sub> , NO, SO <sub>2</sub> , CO, PM <sub>10</sub>
	Los Laureles	LLA	Ecatepec de Morelos, Estado de México	SO <sub>2</sub>
	La Presa	LPR	Tlalnepantla de Baz, Estado de México	SO <sub>3</sub>
	San Juan de Aragón	SJA	Gustavo A. Madero, Mexico City	PM <sub>2.5</sub>
Center	Merced	MER	Venustiano Carranza, Mexico City	O <sub>3</sub> , NO <sub>2</sub> , NO <sub>X</sub> , NO, SO <sub>2</sub> , CO, PM <sub>10</sub> , PM <sub>2.5</sub>
	Iztacalco	IZT	Iztacalco, Mexico City	O <sub>3</sub> , NO <sub>2</sub> , NO <sub>X</sub> , NO, SO <sub>2</sub> , CO, PM <sub>10</sub>
	Camarones	CAM	Azcapotzalco, Mexico City	PM <sub>2.5</sub>
Southwest	Cuajimalpa	CUA	Cuajimalpa de Morelos, Mexico City	O <sub>3</sub>
	Pedregal	PED	Álvaro Obregón, Mexico City	O <sub>3</sub> , NO <sub>2</sub> , NO <sub>X</sub> , NO, SO <sub>2</sub> , CO, PM <sub>10</sub>
	Coyoacán	COY	Coyoacán, Mexico City	PM <sub>2.5</sub>
Southeast	Tláhuac	TAH	Xochimilco, Mexico City	O <sub>3</sub> , SO <sub>2</sub> , PM <sub>10</sub> , PM <sub>2.5</sub>
	UAM Iztapalapa	UIZ	Iztapalapa, Mexico City	O <sub>3</sub> , NO <sub>2</sub> , NO <sub>X</sub> , NO, SO <sub>2</sub> , CO, PM <sub>2.5</sub>

With missing values, it is justified to apply the Hodrick and Prescott filter through the Kalman filter as proposed by Guerrero (2008), both for estimating their trends and for their respective forecasts. It is worth mentioning that SAS software version 13.2

and R version 3.5.2 were used to process the data and prepare the estimates.

In the MCMA, the so-called Índice Metropolitano de Calidad del Aire (metropolitan air quality index, IMECA), relies on the Official Mexican



Table II. Missing data on pollutants by region.

Pollutant	%				
	Northwest	Northeast	Center	Southwest	Southeast
O <sub>3</sub>	0.13	0.00	2.99	0.25	0.25
NO <sub>2</sub>	0.05	0.03	7.16	3.22	5.44
NO <sub>x</sub>	0.05	0.03	7.16	3.22	5.44
NO	0.05	0.03	7.16	3.22	5.44
SO <sub>2</sub>	0.00	0.00	3.52	4.68	1.01
CO	0.03	0.03	6.56	4.28	7.54
PM <sub>10</sub>	0.33	0.05	7.75	29.57	9.42
PM <sub>2.5</sub>	11.32	6.63	1.59	13.72	19.54

Standards to determine permissible limits for the preservation of the population's health. This index may suffer changes due to the updating of some environmental standards such as NOM-025-SSA1-2014O3 (Secretaría de Salud, 2010), NOM-022-SSA1-2010 (Secretaría de Salud, 2014a) and NOM-020-SSA1-2014 (Secretaría de Salud, 2014b). The IMECA takes O<sub>3</sub>, SO<sub>2</sub>, NO<sub>2</sub>, PM<sub>10</sub>, PM<sub>2.5</sub> and CO into account. It also considers the pollutants NO<sub>x</sub> and NO. Table III shows the criteria to establish air quality according to the concentration corresponding to each of the six criteria pollutants.

#### 4. Results

Estimates of trends and forecasts are presented, divided both by pollutant and by region. For this purpose, a description of findings is made in a regional manner according to Figure 2, for the case of trends and Figure 3, together with Table IV, for the case of forecasts.

##### 4.1 Trends

###### 4.1.1 Northwest (NW)

In general, the estimated trend of daily maxima for PM<sub>10</sub>, NO<sub>2</sub>, NO<sub>x</sub>, NO and CO in the northwest region is decreasing. In addition, despite showing a marked

Table III. Air quality standards.

IMECA	Category	O <sub>3</sub> (ppm)	SO <sub>2</sub> (ppm)	NO <sub>2</sub> (ppm)	PM <sub>10</sub> (µg m <sup>-3</sup> )	PM <sub>2.5</sub> (µg m <sup>-3</sup> )	CO (ppm)
0-50	Good	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	Bad	0.096-0.154	0.111-0.207	0.211-0.430	76-214	45.1-97.4	11.1-13.0
151-200	Very bad	0.155-0.204	0.208-0.304	0.431-0.649	215-354	97.5-150.4	13.1-15.4
201-300	Extremely bad	0.205-0.404	0.305-0.604	0.650-1.249	355-424	150.5-250.4	15.5-30.4
301-400	Dangerous	0.405-0.504	0.605-0.804	1.250-1.649	425-504	250.5-350.4	30.5-40.4
401-500		0.505-0.604	0.805-1.004	1.650-2.049	505-604	350.5-500.4	40.5-50.4

Source: Gaceta Oficial de la Ciudad de México (2018).

Notes: the air quality index for O<sub>3</sub> and NO<sub>2</sub> was obtained from 1-h average concentrations; SO<sub>2</sub> was derived from a concentration obtained as a moving average of 24 h; CO was derived from a concentration obtained as a moving average of 8 h; PM<sub>10</sub> and PM<sub>2.5</sub> were derived from concentrations obtained as a 24-h moving average.

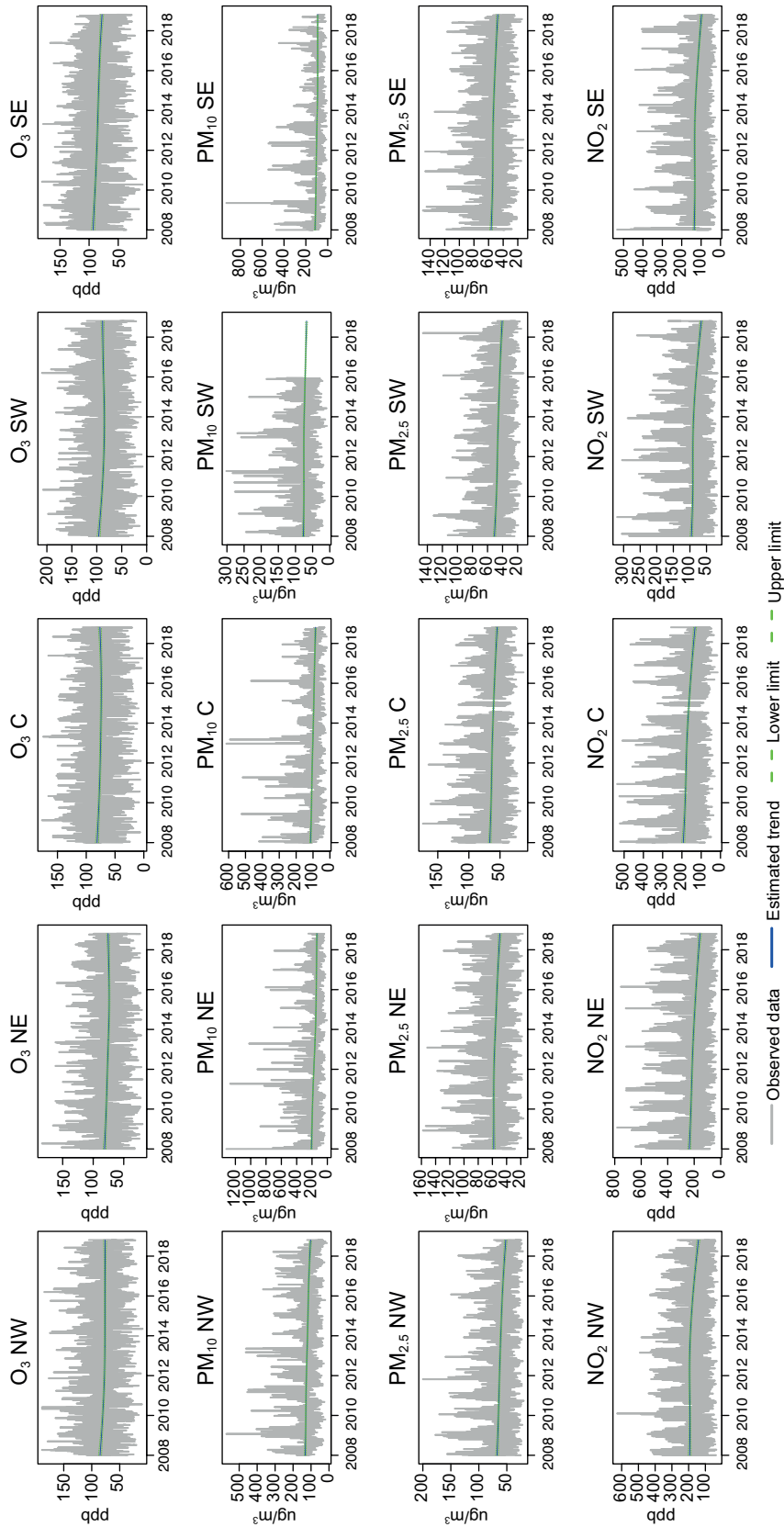


Fig. 2. Observed series, trends and their intervals ( $\pm 2$  standard deviations) per pollutant and geographical region.

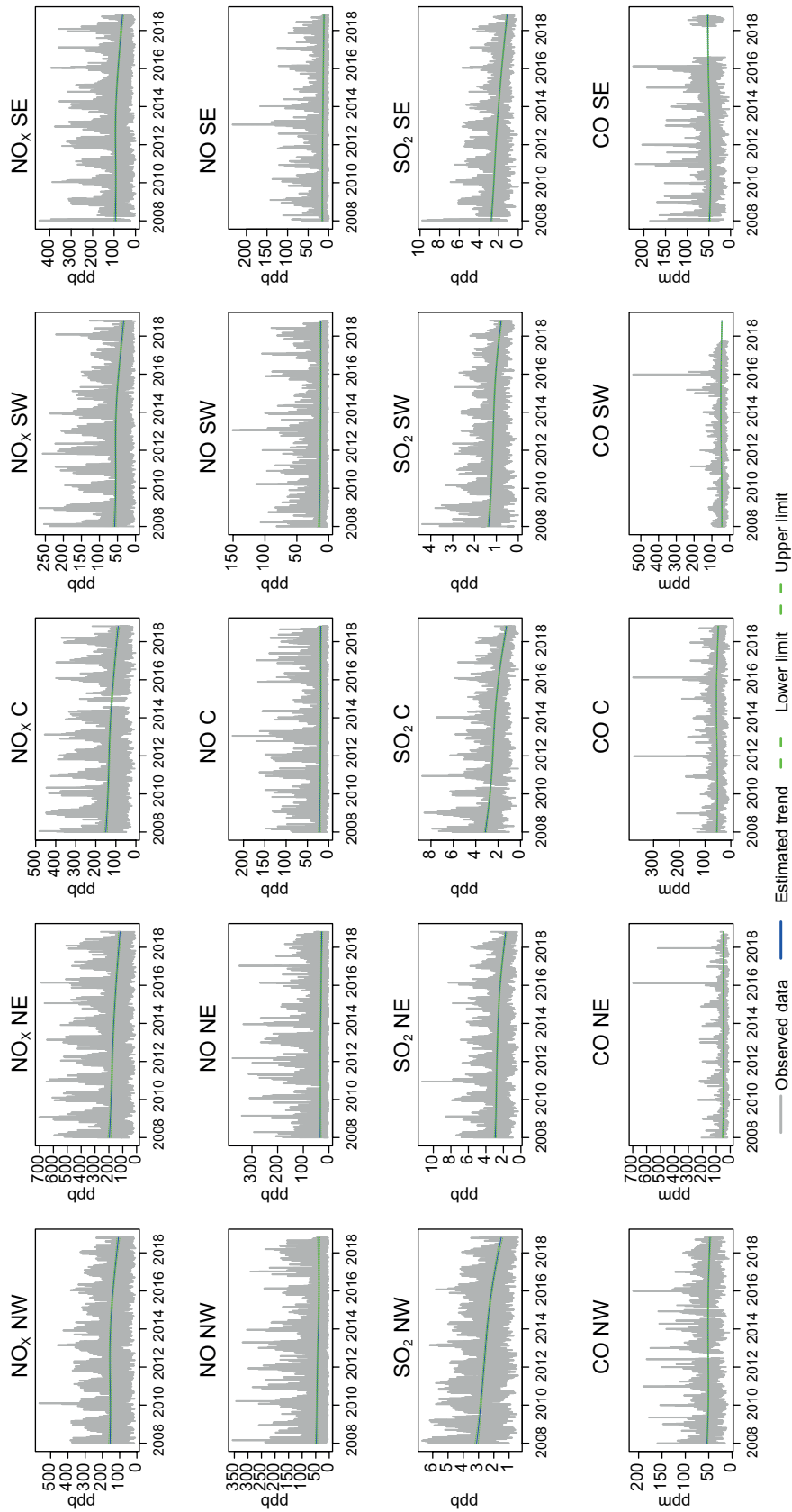


Fig. 2. Observed series, trends and their intervals ( $\pm 2$  standard deviations) per pollutant and geographical region.

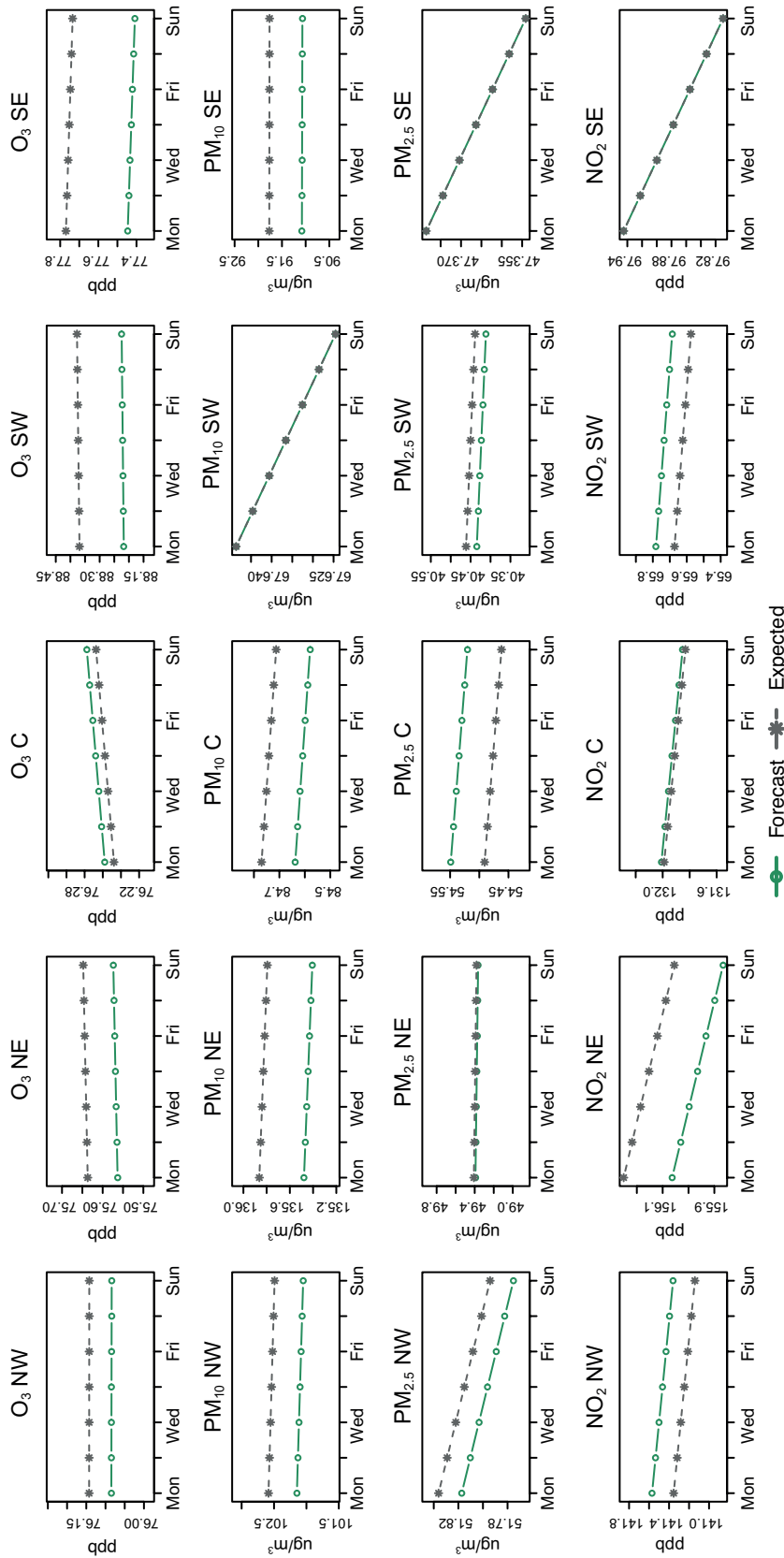


Fig. 3. Trend and forecasting over a 7-day horizon (from October 26 to 31, 2018) for pollutant by region

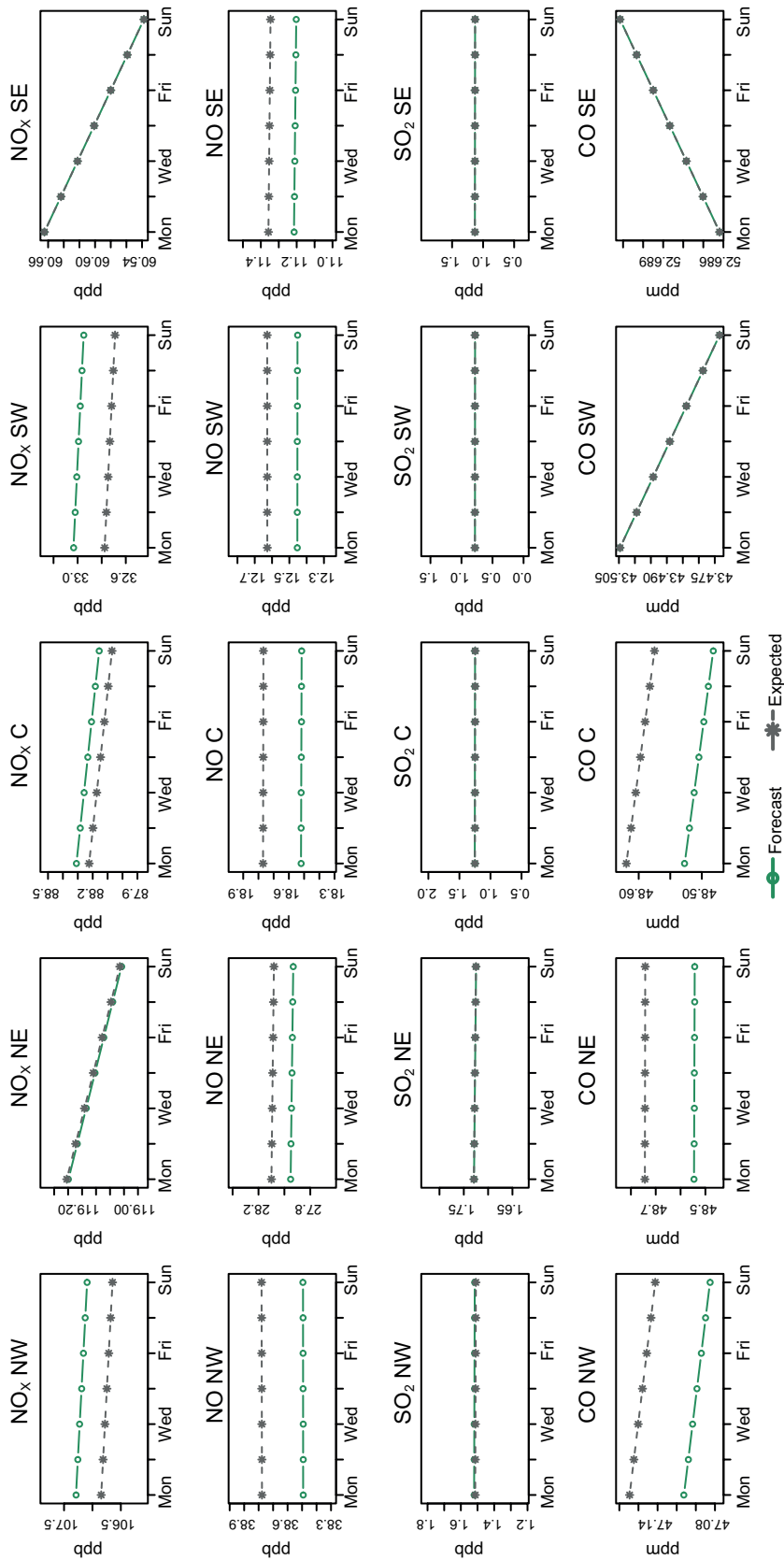


Fig. 3. Trend and forecasting over a 7-day horizon (from October 26 to 31, 2018) for pollutant by region.



decreasing trend at the beginning, the rhythm of O<sub>3</sub> and SO<sub>2</sub> tends to slow down at the end of this period, while; PM<sub>2.5</sub> shows oscillations that do not allow the determination of a single trend.

Taking into account the environmental norm NADF-009-AIRE-2017 (Gaceta Oficial de la Ciudad de México, 2018), the estimate of the trend of daily maxima for O<sub>3</sub> in the northwest region is outside the range of good air quality standards at all times (it remains in the regular quality category). The same happens with PM<sub>10</sub> and PM<sub>2.5</sub>. Regarding SO<sub>2</sub>, it is accredited as regular. The trends of CO and NO<sub>2</sub> are the only ones that, at no time, exceed the permissible concentrations standards. Although no Mexican standard regulates the NO and NO<sub>x</sub> behavior, both present a similar trend pattern in the beginning of the period with a slight growth; however, for 2012, they initiate a remarkable descent.

#### 4.1.2 Northeast (NE)

The trend, in general, is decreasing for all pollutants in the northeast region, with the exception of O<sub>3</sub> and PM<sub>2.5</sub>. It should be noted that the O<sub>3</sub> starts with a downward trend; however, in 2016, a slight rise begins, while for PM<sub>2.5</sub> a similar pattern to that observed 4 yrs earlier begins.

According to the environmental standard NADF-009-AIRE-2017 (Gaceta Oficial de la Ciudad de México, 2018), the tendency of O<sub>3</sub> during the whole study period is in the regular quality interval, since maxima concentrations are below 95 ppm. SO<sub>2</sub> also shows a regular trend, while the PM<sub>10</sub> and PM<sub>2.5</sub> trends place these pollutants in the poor air quality interval throughout the period of analysis. NO<sub>2</sub> and CO trends show good air quality criteria. On their part, NO<sub>x</sub> and NO reflect a clearly decreasing trend.

#### 4.1.3 Center (C)

The observed series corresponding to NO<sub>2</sub>, NO, NO<sub>x</sub>, CO and PM<sub>10</sub> show missing data segments. The estimated trend is decreasing for NO<sub>2</sub>, NO, CO, PM<sub>10</sub> and NO<sub>x</sub>, while O<sub>3</sub> and PM<sub>2.5</sub> still present a trend behavior distinguished by segments, as O<sub>3</sub> in mid 2016 initiates a growing trend. On the other hand, PM<sub>2.5</sub> initiates a downward trend until the end of 2010, when it began increasing until the beginning of 2016 and decreased again thereafter. SO<sub>2</sub> shows only a slight downward trend.

Alternatively, the trends of PM<sub>10</sub> and PM<sub>2.5</sub> indicate that during the period from January to October 2008 these pollutants were in the range of poor quality, although with a decreasing trend for PM<sub>10</sub>. The tendency of O<sub>3</sub> was contrary to that of PM<sub>10</sub>; up to 2015 it had been decreasing to be reversed in that year. However, the whole trend remains within the criterion of regular quality. On their part, the trends of NO<sub>2</sub>, SO<sub>2</sub> and CO are found throughout the study period under the criterion of good air quality. The behavior of NO<sub>x</sub> and NO trends are similar and decreasing.

#### 4.1.4 Southwest (SW)

Despite the absence of data for PM<sub>10</sub> and PM<sub>2.5</sub> at the end of the study period and exclusively for the southwest region, the model was able to perform the respective trend of estimates and intervals up to October 2018. NO<sub>2</sub> and CO show clear decreasing trends, while PM<sub>10</sub> and NO<sub>x</sub> show a similar pattern in their trends, although with differentiated maxima concentration levels. It can also be observed that, despite the fact that SO<sub>2</sub> starts with a marked growing trend, this becomes slower in the course of time. With regard to O<sub>3</sub>, unlike PM<sub>2.5</sub> it starts with a decreasing trend and then reverses in 2014.

When observing the behavior of trends with respect to the environmental standard NADF-009-AIR-2017, O<sub>3</sub> initiates the period with poor air quality; however, it continues with a regular situation during most of the period. Like in the other regions, starting in 2015 there has been a rise in the trends of this pollutant. The same happens with PM<sub>10</sub>, since in the middle of the period its trend corresponds to poor air quality, but then it begins to regularize. Contrary to this, PM<sub>2.5</sub> begins with a regular trend until 2010, and then ascends to maximum concentrations of poor quality. SO<sub>2</sub> shows a trend with maximum concentrations within the good quality interval; the same occurs with NO<sub>2</sub> and CO with respect to NO<sub>x</sub> and NO, as the regions above mentioned show similar behavior in terms of trends.

#### 4.1.5 Southeast (SE)

The observed series of PM<sub>2.5</sub> has a considerable proportion of missing data, in addition to showing a trend marked by segments. For CO, it is evident that its trend is decreasing. Unlike the other regions, O<sub>3</sub>

presents a clearly decreasing trend. Regarding the rest of the pollutants, their trends are decreasing, although they begin with a slight stagnation.

The trends of  $PM_{10}$  and  $PM_{2.5}$  lay within the intervals of poor air quality, highlighting that while the first one shows a decreasing trend and is less accelerated at the end of the period, the trend of the second pollutant is ascending. The  $O_3$  trend is in the regular quality range. The pollutants that show a downward trend, as well as those in the intervals of good air quality, are  $NO_2$ ,  $CO$  and  $SO_2$ . On its part,  $NO_x$  and  $NO$  continue to show the same behavior in their trends.

#### 4.2 Forecasts

It is important to emphasize that no forecasts are made of the maximum observed pollutant concentrations, but of the estimated trends. Thus, these forecasts are made in a seven-day horizon. Note that the observed data for the analysis were registered from January 1, 2008 to October 31, 2018; however, the estimation of the trend was only contemplated until October 25, to then predict the period from 26 to 31 October, 2018. With the intention of verifying the validity of the forecasts, we also estimated the trend with all the data until October 31, 2018. We also obtained the precise MAD, MSE and MAPE for each pollutant by region.

In general, the forecasts of the eight pollutants trends showed a significant proximity between both series (see Table IV), since through the MAPE it is possible to observe that the average error between the estimate of the trend and the forecast was less than 1.5%. Compared to the MAD, these were less than 0.5%, and for MSE they did not exceed 0.2%. The following paragraphs describe in detail what happens with the forecasts by region.

##### 4.2.1 Northwest (NW)

According to the forecasts of daily maxima trends,  $CO$ ,  $NO_2$  and  $SO_2$  remained below the permissible concentrations established by the environmental standard NADF-009-AIRE-2017 (Gaceta Oficial de la Ciudad de México, 2018). Regarding,  $PM_{10}$  and  $PM_{2.5}$ , air quality was affected by these pollutants, qualifying it as bad. As for  $O_3$ , air quality was regular.

##### 4.2.2 Northeast (NE)

For  $PM_{10}$ , higher concentrations were predicted than for the rest of the regions, and air quality was considered non satisfactory.  $PM_{2.5}$  contaminated the air, conferring it a poor quality.  $O_3$  contributed to a regular air scenario. In contrast,  $CO$ ,  $NO_2$  and  $SO_2$  complied with the standards.

Table IV. Precision forecast measurements by pollutant and region.

Region	Statistic	$O_3$	$PM_{10}$	$PM_{2.5}$	$NO_2$	$NO_x$	$NO$	$SO_2$	$CO$
Northwest	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
Northeast	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
Center	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
Southwest	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
Southeast	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

#### 4.2.3 Center (C)

The forecasts of trend estimates for O<sub>3</sub> were located in a regular quality, while PM<sub>10</sub> and PM<sub>2.5</sub> were in the poor quality range. Although the predictions for NO<sub>2</sub> did not exceed the established norms, its predicted maximum concentrations were higher than for other regions; CO and SO<sub>2</sub> did also not exceed the norms.

#### 4.2.4 Southwest (SW)

The forecasts showed that this region was the only one with regular quality levels for PM<sub>10</sub>. In addition, although air quality was poor for PM<sub>2.5</sub>, the lowest levels for this contaminant were registered in this region. O<sub>3</sub> showed a similar behavior to that of the northwest region. CO, NO<sub>2</sub> and SO<sub>2</sub> remained within the established norms, although it should be noticed that NO<sub>2</sub> had the highest concentrations compared to the other regions.

#### 4.1.5 Southeast (SE)

In the trend estimates for this region, the only contaminants that showed concentrations outside the norm were PM<sub>10</sub> and PM<sub>2.5</sub>, while O<sub>3</sub> seemed to be a little more controlled. However, this was more clearly reflected in CO, NO<sub>2</sub> and SO<sub>2</sub>.

In summary, the trends for PM<sub>10</sub> and PM<sub>2.5</sub> during the study period and for each of the regions of the MCMA showed that these contaminants were at the threshold of poor quality. On the other hand, O<sub>3</sub> trends for the five regions were characterized by being in the regular air quality interval, with a decreasing trend found only in the southeast.

SO<sub>2</sub> trends in the northwest and northeast regions were within the threshold of regular air quality, while in the center, southwest and southeast they were in the good air quality range. CO and NO<sub>2</sub> can be considered two pollutants that have been controlled in the entire MCMA, since in all regions their trends and forecasts were low and never exceeded the permissible limits.

Although NO<sub>x</sub> is monitored for the MCMA, permissible concentrations are not established in the Mexican health standard, therefore it is important that this contaminant is included in the environmental norms. NO<sub>x</sub> is one of the precursors of O<sub>3</sub>, a pollutant that in high concentrations affects the health of the population. NO trends present a similar behavior

than those of NO<sub>x</sub>, and this contaminant is also not regulated by the norm.

## 5. Conclusions

In the specialized literature, various methodologies have been developed and/or applied regarding the behavior of trends and forecasts of atmospheric pollutant concentrations, with the intention of understanding the phenomenon or anticipating future scenarios. It is considered that the applied methodology and the analysis here exposed by region, represent a pertinent technical resource, of exploratory, useful and simple nature, that contribute to this matter.

By means of controlled smoothing, estimated by the Kalman filter, it was possible to estimate trends of O<sub>3</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO, PM<sub>10</sub>, PM<sub>2.5</sub>, NO and NO<sub>x</sub> for the five regions of the MCMA, even when the data presented missing values both in the central part and at the end of the series. However, this circumstance did not prevent the elaboration of their respective forecasts, which is why it is considered to be a robust and efficient tool.

Although there is no optimal smoothness index, the imposition of this measure in the model for all the series allowed to compare the behavior of the trends between different pollutants and for each region within the MCMA. This evidences the need to solve the problem of pollution from an integral perspective.

The forecasts presented here are based on trend estimates, so they do not reflect point estimates; however, the trend forecast of the period from October 26 to 31, 2018 verifies that the forecasts are numerically close to the expected values for each pollutant and region. Finally, it is considered that this type of statistical tool would be useful for decision makers regarding air pollution issues. It would be desirable as a future line of research to estimate the trends and forecasts of the eight pollutants, considering their seasonality.

## Acknowledgments

The authors gratefully acknowledge the comments and suggestions from two anonymous reviewers and the editor of this journal. Eliud Silva dedicates this article to his family.

## References

- Aditya C, Chandana R, Nayana D, Praveen G. 2018. Detection and prediction of air pollution using machine learning models. *International Journal of Engineering Trends and Technology* 59, 204-207. <https://doi.org/10.14445/22315381/IJETT-V59P238>
- Aguilar D, Reyes I. 2018. A wavelet analysis of multiday extreme ozone and its precursors in Mexico City during 2015-2016. *Atmospheric Environment* 188, 112-119. <https://doi.org/10.1016/j.atmosenv.2018.06.017>
- Brantley H, Hagler G, Herndon S, Massoli P, Bergin M, 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* 16, 1-14. <https://doi.org/10.3390/ijerph16040535>
- Bravo H, Sosa P, Sánchez P, 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* 3, 185-193. <http://doi.org/10.22201/fi.25940732e.2002.03n4.018>
- Camalier L, Cox W, Dolwick P. 2007. The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. *Atmospheric environment* 41, 7127-7137. <https://doi.org/10.1016/j.atmosenv.2007.04.061>
- Chang Ch, Yao X. 2008. Air pollution in megacities in China. *Atmospheric Environment* 42, 1-42. <https://doi.org/10.1016/j.atmosenv.2007.09.003>
- CONAPO-INEGI-SEDATU. 2018. Delimitación de las zonas metropolitanas de México 2015. Consejo Nacional de Población-Instituto Nacional de Estadística y Geografía-Secretaría de Desarrollo Agrario, Territorial y Urbano, Mexico. Available at: <https://www.gob.mx/conapo/documentos/delimitacion-de-las-zonas-metropolitanas-de-mexico-2015> (last accessed on January 9, 2019).
- Corona E, Rojas R. 2009. Calidad del aire y su incorporación en la planeación urbana: Mexicali, Baja California, México. *Estudios Fronterizos* 10, 79-102.
- Cruz M, Gómez A, Quintero M, 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* 29, 269-283.
- Davis L. 2008. The effect of driving restrictions on air quality in Mexico City. *Journal of Political Economy* 116, 38-81. <https://doi.org/10.1086/529398>
- Gaceta Oficial de la Ciudad de México. 2018. Aviso por el que se da a conocer la norma ambiental para el Distrito Federal NADF-009-AIRE-2017, que establece los requisitos para elaborar el índice de calidad del aire en la Ciudad de México. Available at: <http://www.aire.cdmx.gob.mx/descargas/monitoreo/normatividad/NADF-009-AIRE-2017.pdf> (last accessed on January 9, 2019).
- Gallego F, Montero J, Salas C. 2013a. The effect of transport policies on car use: A bundling model with applications. *Energy Economics* 40, 85-97. <https://doi.org/10.1016/j.eneco.2013.09.018>
- Gallego F, Montero J, Salas C. 2013b. The effect of transport policies on car use: Evidence from Latin American cities. *Journal of Public Economics* 107, 47-62. <https://doi.org/10.1016/j.jpubeco.2013.08.007>
- Georgieva S, Valev A, Stoyanova D, Todorov D. 2014. Time series analysis and forecasting for air pollution in small urban area: An SARIMA and factor analysis approach. *Stochastic Environmental Research and Risk Assessment* 28, 1045-1060. <https://doi.org/10.1007/s00477-013-0800-4>
- Gong B, Ordires J. 2015. Prediction models for ozone in metropolitan area of Mexico City based on artificial intelligence techniques. *Journal Information and Decision Sciences* 7, :115-138. <https://doi.org/10.1504/IJIDS.2015.068756>
- Gramsh E, Cereceda F, Oyola P, von Baer D. 2006. Examination of pollution trends in Santiago de Chile with cluster analysis of PM10 and ozone data. *Atmospheric Environment* 40, 5464-5475. <https://doi.org/10.1016/j.atmosenv.2006.03.062>
- Guerrero V. 2007. Time series smoothing by penalized least squares. *Statistics & Probability Letters* 77, 1225-1234. <https://doi.org/10.1016/j.spl.2007.03.006>
- Guerrero V. 2008. Estimating trends with percentage of smoothness chosen by the user. *International Statistical Review* 76, 187-202. <https://doi.org/10.1111/j.1751-5823.2008.00047.x>
- Guerrero V, Silva E. 2015. Smoothing a time series by segments of the data range. *Communications in Statistics – Theory and Methods* 44, 4568-4585. <https://doi.org/10.1080/03610926.2014.901372>
- Guerrero V, Islas A, Ramírez L. 2017. Trend estimation of multivariate time series with controlled smoothness. *Communications in Statistics – Theory and Methods* 46, 6704-6726. <https://doi.org/10.1080/03610926.2015.1133826>



- Hernández J, Morales C, Madrigal D. 2004. Comportamiento del monóxido de carbono y el clima en la ciudad de Toluca, de 1995 a 2001. *Ciencia Ergo Sum* 11, 263-274.
- Hernández L. 2009. Modelado atmosférico para determinar niveles máximos diarios de ozono en la Ciudad de Guadalajara. M.Sc. thesis. Universidad Autónoma Metropolitana, Mexico.
- Hernández I, Clemitshaw K, Mendoza A. 2017. Observed trends in ground-level O<sub>3</sub> in Monterrey, Mexico, during 1993-2014: Comparison with Mexico City and Guadalajara. *Atmospheric Chemistry and Physics* 17, 9163-9185. <https://doi.org/10.5194/acp-17-9163-2017>
- Hodrick R, Prescott E. 1997. Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking* 29, 1-16.
- Huerta G, Sansó, B. 2007. Time-varying models for extreme values. *Environmental and Ecological Statistics* 14, 285-299. <https://doi.org/10.1007/s10651-007-0014-3>
- Jaiswal A, Samuel C, Kadabgaon V. 2018. Statistical trend analysis and forecast modeling of air pollutants. *Global Journal of Environmental Science and Management* 4, 427-438. <https://doi.org/10.22034/gjesm.2018.04.004>
- Jaramillo M, González D, Nuñez M, Portilla G, Lucio J. 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* 39, 79-88.
- Klimberg R, Sillup G, Boyle K, Tavva V. 2010. Forecasting performance measures – what are their practical meaning? *Advances in Business and Management Forecasting* 7, 137-147. [https://doi.org/10.1108/S1477-4070\(2010\)0000007012](https://doi.org/10.1108/S1477-4070(2010)0000007012)
- Kumar U, De Ridder K. 2010. GARCH modelling in association with FFT e ARIMA to forecast ozone episodes. *Atmospheric Environment* 44, 4252-4265. <https://doi.org/10.1016/j.atmosenv.2010.06.055>
- Maravall A. 1993. Stochastic linear trends. Models and estimators. *Journal of Econometrics* 56, 5-37. [https://doi.org/10.1016/0304-4076\(93\)90099-Q](https://doi.org/10.1016/0304-4076(93)90099-Q)
- OECD. 2015. Valle de México, Mexico. Territorial reviews. Organisation for Economic Co-operation and Development Publishing, Paris. Available at: <https://www.oecd.org/mexico/oecd-territorial-reviews-valle-de-mexico-mexico-9789264245174-en.htm> (last accessed on January 9, 2019).
- Palomares J, de la Rosa J, Agüera A, Sierra J, Florencias O. 2019. Forecasting PM<sub>10</sub> in the Bay of Algeciras based on regression models. *Sustainability* 11, 968. <https://doi.org/10.3390/su11040968>
- Petris G, Petrone S, Campagnoli P. 2007. Dynamic linear models with R. Springer. Springer-Verlag, Berlin, 186 pp.
- Rodríguez S, Huerta G, Reyes H. 2016. A study of trends for Mexico City ozone extremes: 2001-2014. *Atmósfera* 29, 107-120. <https://doi.org/10.20937/ATM.2016.29.02.01>
- Secretaría de Salud. 2010. Norma Oficial Mexicana NOM-022-SSA1-2010. Salud ambiental. Criterio para evaluar la calidad del aire ambiente con respecto al dióxido de azufre (SO<sub>2</sub>). Diario Oficial de la Federación, Mexico, September 8. Available at: [http://dof.gob.mx/nota\\_detalle.php?codigo=5158348&fecha=08/09/2010](http://dof.gob.mx/nota_detalle.php?codigo=5158348&fecha=08/09/2010) (last accessed on January 9, 2019).
- Secretaría de Salud. 2014a. Norma Oficial Mexicana NOM-025-SSA1-2014. Salud ambiental. Valores límites permisibles para la concentración de partículas suspendidas PM<sub>10</sub> y PM<sub>2.5</sub> en el aire ambiente y criterios para su evaluación. Diario Oficial de la Federación, Mexico, August 20. Available at: [http://www.dof.gob.mx/nota\\_detalle.php?codigo=5357042&fecha=20/08/2014](http://www.dof.gob.mx/nota_detalle.php?codigo=5357042&fecha=20/08/2014) (last accessed on January 9, 2019).
- Secretaría de Salud. 2014b. Norma Oficial Mexicana NOM-020-SSA1-2014. Valor límite permisible para la concentración de ozono (O<sub>3</sub>) en el aire ambiente y criterios para su evaluación. Diario Oficial de la Federación, Mexico, August 19. Available at: [http://www.dof.gob.mx/nota\\_detalle.php?codigo=5356801&fecha=19/08/2014](http://www.dof.gob.mx/nota_detalle.php?codigo=5356801&fecha=19/08/2014) (last accessed on January 9, 2019).
- SEDEMA. 2016. Inventario de Emisiones de la Ciudad de México 2014. Dirección General de Gestión de la Calidad del Aire, Dirección de Programas de Calidad del Aire e Inventario de Emisiones. Ciudad de México. Available at: <http://www.aire.cdmx.gob.mx/descargas/publicaciones/flippingbook/inventario-emisiones-cdmx2014-2/IE-CDMX-2014.pdf> (last accessed on January 9, 2019).
- 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. Available at: <http://www.aire.cdmx.gob.mx/informe-2016>



- cdmx.gob.mx/descargas/publicaciones/flippingbook/informe2016calidad-del-aire-en-la-ciudademexico/mobile/informe\_anual\_calidad\_aire\_2016.pdf (last accessed on January 9, 2019).
- SEDEMA. 2018. Inventario de Emisiones de la Ciudad de México 2016. Dirección General de Gestión de la Calidad del Aire, Dirección de Programas de Calidad del Aire e Inventario de Emisiones. Ciudad de México. Available at: <http://www.aire.cdmx.gob.mx/descargas/publicaciones/flippingbook/inventario-emisiones-2016/mobile/inventario-emisiones-2016.pdf> (last accessed on January 9, 2019).
- SEMARNAT. 2016. Informe de la Situación del Medio Ambiente en México. Compendio de Estadísticas Ambientales. Indicadores Clave de Desempeño Ambiental y de Crecimiento Verde. Available at [https://apps1.semarnat.gob.mx:8443/dgeia/informe15/tema/pdf/Informe15\\_completo.pdf](https://apps1.semarnat.gob.mx:8443/dgeia/informe15/tema/pdf/Informe15_completo.pdf) (last accessed on January 9, 2019).
- So K, Guo H, Li Y. 2007. Long-term variation of PM<sub>2.5</sub> levels and composition at rural, urban, and roadside sites in Hong Kong: Increasing impact of regional air pollution. *Atmospheric Environment* 41, 9427-9434. <https://doi.org/10.1016/j.atmosenv.2007.08.053>
- Vita Y, Apriliani E, Khusnul D. 2018. Fractional Kalman filter to estimate the concentration of air pollution. *Journal of Physics* 1008, 012008. <https://doi.org/10.1088/1742-6596/1008/1/012008>
- Yeongkwon S, Osornio A, O'Neill M, Hystad P, Texcalac J, Ohman P, Qingyu M, 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* 639, 40-48. <https://doi.org/10.1016/j.scitotenv.2018.05.144>
- Zamri M, Zailan R, Ismail M, Safiih M. 2009. Forecasting and time series analysis of air pollutants in several areas of Malaysia. *American Journal of Environmental Sciences* 5, 625-632. <https://doi.org/10.3844/ajessp.2009.625.632>

### Supplementary material

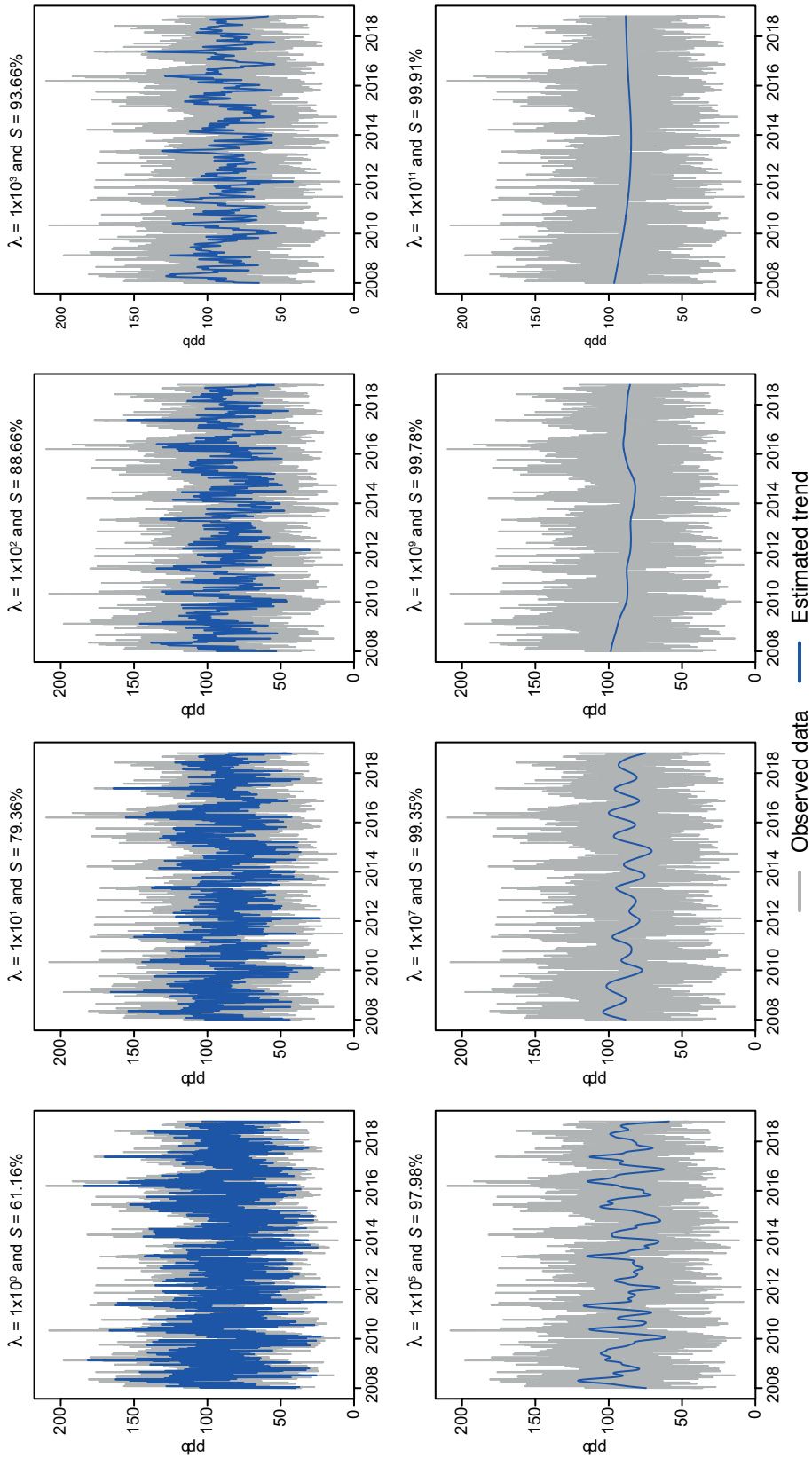


Fig. SM1. Observed O3 time series and its trends with different  $\lambda$  values