Intervention analysis with autoregressive integrated moving average models for time series of urban air pollutants

L A Manco-Perdomo, L A Pérez-Padilla, and C A Zafra-Mejía

1 Grupo de Investigación en Ingeniería Ambiental, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia

E-mail: lamancop@correo.udistrital.edu.co, czafra@udistrital.edu.co

Abstract. The objective of this paper is to show an intervention analysis with autoregressive integrated moving average models for time series of air pollutants in a Latin American megacity. The interventions considered in this study correspond to public regulations for the control of urban air quality. The study period comprised 10 years. Information from 10 monitoring stations distributed throughout the megacity was used. Modelling showed that setting maximum emission limits for different pollution sources and improving fuel were the most appropriate regulatory interventions to reduce air pollutant concentrations. Modelling results also suggested that these interventions began to be effective between the first 4 days-15 days after their publication. The models developed on a monthly timescale had a short autoregressive memory. The air pollutant concentrations at a given time were influenced by the concentrations of up to three months immediately preceding. Moving average term of the models showed fluctuations in time of the air pollutant concentrations (3 months - 14 months). Within the framework of the applications of physics for the air pollution control, this study is relevant for the following findings: the usefulness of autoregressive integrated moving average models to temporal simulate air pollutants, and for its suitable performance to detect and quantify regulatory interventions.

1. Introduction

In urban environments, the effect of air pollution has been related to high health costs due to respiratory complications, a decrease in production due to medical disabilities, and a reduction in life expectancy [1]. Thus, air quality has been considered as a key indicator for urban planning [2]. For example, in metropolises such as Mexico City have improved air quality by promoting regulatory interventions such as the elimination of lead from gasoline and the strengthening of vehicle inspection and maintenance programs [3].

Each territory has a unique atmospheric dynamic due to its particular meteorological and orography conditions, and that is why it is necessary to study in a differentiated way the regulatory interventions in favor of urban air quality [4]. In this particular case of atmospheric physics, regulatory intervention is interpreted as the occurrence of an exogenous and induced event to the historical behavior of air quality variables [5]. Thus, there is a need for statistical tools to address the modeling of time series of air pollutants influenced by external events such as the establishment of public regulations to improve urban air quality.

Autoregressive integrated moving average (ARIMA) models are a non-causal statistical method, which allow describing a value as a linear function of previous data and errors due to random [6]. These models aim to identify a pattern in the data, in such a way that it can be predicted based on historical records [7]. The objective of this study is to show an intervention analysis with ARIMA models for time
series of air pollutants in a Latin American megacity (Bogotá, Colombia). Within the framework of the applications of physics for the control of urban air pollution, this study is relevant for the following aspects: The use of ARIMA models to study time series of air pollutants, and to detect and evaluate the effectiveness of regulatory interventions in air quality.

2. Materials and methods
The methodology has four phases described below.

2.1. Detection of regulatory interventions
Through the computerized compilation of regulations, doctrine, and jurisprudence of the “Secretaría Jurídica Distrital de Bogotá”, Colombia, the regulatory interventions related to air quality were detected for the megacity during the study period (2007-2017). Regulatory interventions were organized according to their dates of issue, regulation, repeal, or adjustment, and were classified according to their approach and type of air quality goal [8].

2.2. Data collection and processing
Based on information from “Red de Monitoreo de Calidad del Aire de Bogotá (RMCAB)”, time series of air pollutants and meteorological variables were downloaded (http://201.245.192.252:81/home/map) from 10 stations located throughout the megacity (Figure 1). These time series were pre-processed following reference studies [9], to then be evaluated in terms of quantity and reliability and ensuring at least 75% availability of information during the study period [10]. The air pollutants analyzed were as follows: particulate matter < 10 µm (PM$_{10}$), particulate matter < 2.5 µm (PM$_{2.5}$), ozone (O$_3$), carbon monoxide (CO), sulfur dioxide (SO$_2$), and nitrogen dioxide (NO$_2$).

![Figure 1. Location in Bogotá, Colombia, of the 10 monitoring stations used in this study.](image)

2.3. Data completion
Initially, missing data in the previously processed time series was completed using the normal ratio method. This method used time series from nearby stations to complete missing information [11]. For those time series, that still showed missing data after applying the above method, regression models were then developed with time series of other variables from the same station with a Pearson’s
correlation coefficient greater than |0.75|. The regression models developed made it possible to complete the missing information. Lastly, the information that was not yet completed at this point was estimated using the maximum expectation algorithm (MEA) in the IBM SPSS V.25.0 software [12]. This method imputed the lost data by estimated values, which were successively re-estimated proceeding iteratively until convergence [13].

2.4. Intervention analysis with autoregressive integrated moving average models

To study the temporal trend of air pollutants, ARIMA models were developed from the completed time series. These time series were smoothed with a monthly moving average, which suppressed seasonal and accidental variations that could generate noise to detect an appropriate ARIMA model [6]. Subsequently, Expert Modeler Tool of the IBM SPSS V.25.0 software [12] was used to identify and estimate the best-fit models for the time series of the variables under study. ARIMA models were validated using the Ljung-Box’s statistic (Q’ p-value > 0.05) and the determination coefficient (R²). In case we had more than one validated model for the same variable, their selection was made using the normalized BIC criterion [14]. The following statistics were also considered to evaluate the fit of the ARIMA models developed: root mean squared error (RMSE) and mean absolute percentage error (MAPE). The temporal trend of air pollutant concentrations with respect to climate variables was also analyzed.

The discontinuous regression design methodology [15] was used in determining the percentage effectiveness of regulatory interventions on air pollutant concentrations. This was a quasi-experimental pretest-posttest design that studied the causal effects of the application of air quality regulations (interventions) by assigning a cut-off value or threshold above or below which an intervention was assigned. Thus, the ARIMA models developed were used to analyze the time series of air pollutants with respect to significant and prolonged gradual discontinuities close to the regulatory dates of the identified interventions. This by identifying innovative or level shift outliers within a maximum period of six months (observation time window) [16]. All the above analyses were performed with IBM SPSS V.25.0 software [12]. Finally, the terms autoregressive (p), difference (d), and moving average (q) of the developed models were analyzed.

3. Results and discussion

From the mode calculation for the terms (p, d, q) of the ARIMA models developed under a monthly timescale, the results showed that the air pollutant concentrations had a short memory. This from the variation of the autoregressive term (p) of the models developed (’p’ between 1 - 2). Namely, concentrations of the air pollutants under study were influenced by the concentrations observed up to two months immediately preceding. This behavior suggested that the persistence of air pollutants in the megacity was between 1 month - 2 months (Table 1). Overall, this autoregressive behavior was similar in both the north and south of the megacity. However, when studying each air pollutant separately, differences were observed in the autoregressive term of the ARIMA models developed monthly. These findings suggested from the viewpoint of the persistence of air pollutants, that ARIMA modelling should be independent for each type of air pollutant. Thus, the PM_{10} and PM_{2.5} concentrations could be simulated together, and the concentrations of the other air pollutants could be simulated independently. Lastly, the results suggested that the autoregressive term of the models did not allow visualizing significant spatial differences for the same air pollutant. This term was possibly not suitable for studying the spatial behavior of air pollutants in the megacity.

In relation to the moving average term (q) of the ARIMA models developed monthly, the results showed greater differences in their magnitude compared to the autoregressive term (p). The findings suggested that this moving average term was possibly best suited to study the spatial variation of air pollutant concentrations in the megacity (Table 1). In the case of PM_{10}, greater fluctuations (‘q’ between 1 month - 14 months) were observed in its concentration in the northern zone compared to the southern zone (‘q’ between 1 month - 3 months). Conversely, the PM_{2.5} concentrations fluctuated more in the southern zone (Figure 2). This differential behavior between the south and north in relation to PM concentrations
concentrations may be associated with the local behavior of climate variables such as wind speed and direction, and relative humidity. For example, it was reported that the northern zone of the megacity historically showed the highest wind speeds [4]. In the case of O₃ and SO₂ concentrations, differences in the moving average term between nearby monitoring stations were observed (‘q’ between 1 month - 18 months). This trend suggested the existence of a differential behavior between nearby monitoring stations, for which it was not possible to visualize a general spatial behavior. Apparently, there were local conditions that caused these air pollutants to vary significantly between nearby monitoring stations.

The results suggested that ARIMA modelling of these air pollutants should be interpreted independently by each monitoring station. In contrast, the concentrations of CO and NO₂ showed lower fluctuations between monitoring stations (‘q’ between 1 month - 3 months), which suggested a more uniform spatial behavior of these air pollutants throughout the megacity.

**Table 1.** Frequency analysis for the terms and number of outliers in ARIMA models developed under a monthly timescale.

| T | MF | F (%) | North zone | South zone | R² | Outliers |
|---|----|-------|-------------|------------|----|----------|
| PM₁₀ | p 2 | 90.0  | 2           | 2          |   |   |
|     | d 1 | 100.0 | 1           | 1          | 1  | 0.96-0.99  |
| q 1 | 50.0 |       | 10          | 10         | 1  | 1        |
| PM₂₅ | p 2 | 50.0  | 3           | 3          |   |   |
|     | d 1 | 100.0 | 1           | 1          | 1  | 0.98-0.99  |
| q 1 | 62.5 |       | 1           | 3          |   | 7       |
| O₃  | d 1 | 100.0 | 1           | 1          |   |   |
| q 1 | 57.1 |       | 2           | 10         | - | 1       |
| CO  | d 1 | 100.0 | 1           | 1          |   |   |
| q 1 | 42.9 |       | 1           | 1          | - | 1       |
| SO₂ | d 1 | 100.0 | 1           | 1          |   |   |
| q 14| 40.0 |       | -           | 14         | - | -       |
| NO₂ | d 1 | 100.0 | -           | -          |   |   |
| q 2 | 50.0 |       | 1           | 2          |   | 1       |

*MF = Most frequent ARIMA term (mode), F = Frequency (%). Monitoring stations: Gu = Guaymaral, Su = Suba, Us = Usaquén, Lf = Las Ferias, Car = Centro de alto rendimiento, Pa = Puente Aranda, Ke = Kennedy, Ca = Carvajal, Tu = Tunal, Sc = San Cristóbal.

The results showed the existence of 34 air quality regulations (interventions) in Bogotá, Colombia, during the study period (2007 - 2017). Analyzing the dynamics of issuance, regulation, repeal, or adjustment of this compendium of interventions, a greater number of changes was observed for the mobile pollution sources (vehicles) and with an average change periodicity of two years. The results suggested that this short periodicity of change in regulations possibly did not allow obtaining adequate results during the management of air quality in the megacity. Base on the ARIMA models developed and prioritizing in the monitoring stations located in high pollution areas (south zone), it was observed that the regulatory interventions on air quality oriented to the restriction of the precursor substances generation of air pollutants were possibly the most effective for the control of these.

The intervention analysis (see outliers in Table 1) with ARIMA models allowed to detect the “Resolución 1304 de 2012” [17] as the most effective in reducing O₃ (-38.2%), by establishing stricter reduction regulations for diesel engines in public transport systems (Figure 3). This resolution was stricter in the emission of NOₓ and thus decreased the formation of tropospheric O₃ [18]. Although this was not effective in controlling this air pollutant. Additionally, the “Resolución 909 de 2008” [19] decreased O₃ (-12.8%), regulating differential emission standards for precursor pollutants (HCl) to...
various industries. “Resolución 180782 de 2007” [20] was also the most effective in reducing PM$_{2.5}$ (-46.9%), by modifying the quality criteria of biofuels for use in diesel engines (Figure 4). This possibly improved oxygenation in the combustion process, which decreased the emission and formation of fine particles [21]. According to the intervention analysis, the results suggested that these regulations began to be effective between the first 4 days-15 days after their publication.

The PM$_{10}$ did not show a significant reduction from the regulatory interventions detected during the study period, with the “Decreto Distrital 623 de 2011” [22] being the most effective for its reduction. This regulatory intervention decreased the PM$_{10}$ concentration by 2.83%, from the prohibition of waste oils in boilers and furnaces (Figure 5). This suggested that the normative instruments implemented in the megacity did not have the same effectiveness in reducing the air pollutants under study. Thus, during the development of regulatory interventions for the monitoring and control of air quality in the megacity, it was essential to identify previously the main emission sources, transport mechanisms, and dispersion dynamics of their air pollutants.

![Figure 2. Monthly spatial variation of the moving average term (q) in ARIMA models of Bogotá, Colombia.](image1)

![Figure 3. Observed and simulated temporal behavior of O$_3$ concentrations after the intervention of “Resolución 1304 de 2012” [17]. “Puente Aranda” monitoring station.](image2)
4. Conclusions

Analysis of the autoregressive term of ARIMA models suggests that air pollutants have a short memory (‘p’ between 1 - 2). Namely, these findings imply that the temporal persistence of pollutants in the atmosphere of the megalcity is up to two months. It appears that this temporal persistence of air pollutants is similar in all sectors of the megalcity. Nevertheless, the findings suggest that the analysis of the autoregressive term of the models should be performed independently for each air pollutant under study. This is because differences between 1 month - 2 months are observed in its persistence.

The findings suggest that the moving average term of ARIMA models is best suited to study the spatial variation of air pollutants in the megalcity. This is compared to the autoregressive term of the models developed. The analysis of the moving average term allows visualizing spatial differences between the urban concentrations of the air pollutants under study (PM$_{10}$, PM$_{2.5}$, O$_3$, CO, SO$_2$, and NO$_2$).

During the study period, 34 air quality regulatory interventions are detected in the megalcity, which focus mainly on the control of mobile emission sources. It is also observed that the most effective interventions are those focused on the restriction of precursor substances of air pollutants.

The intervention analysis allows identifying the most effective regulations to reduce the concentrations of air pollutants under study: O$_3$ = -38.2% (“Resolución 1304 de 2012”), PM$_{2.5}$ = -46.9% (“Resolución 180782 de 2007”), and PM$_{10}$ = -2.83% (“Decreto Distrital 623 de 2011”). Lastly, the ARIMA results suggest that these regulatory interventions begin to be effective between the first 4 days - 15 days after their publication.

Within the framework of the applications of physics for the air pollution control, this study is relevant for the following findings: The usefulness of autoregressive integrated moving average models to temporal simulates air pollutants, and for its suitable performance to detect and quantify regulatory interventions.
References

[1] Hao Y, Deng Y, Lu Z-N, Chen H 2018 Is environmental regulation effective in China? Evidence from city-level panel data Journal of Cleaner Production 188 966

[2] Borrego C, Martins H, Teipel O, Salmim L, Monteiro A, Miranda A I 2006 How urban structure can affect city sustainability from an air quality perspective Environmental Modelling & Software 21(4) 461

[3] Baklanov A, Molina L T, Gauss M 2016 Megacities, air quality and climate Atmospheric Environ. 126 235

[4] Zafra C, Suárez J, Pachón J E 2021 Public health considerations for PM10 in a high-pollution megacity: influences of atmospheric condition and land coverage Atmosphere 12(1) 118

[5] Han Y, Lam J C, Li V O, Reiner D 2021 A Bayesian LSTM model to evaluate the effects of air pollution control regulations in Beijing, China Environmental Science & Policy 115 26

[6] Abhilash M S K, Thakur A, Gupta D, Sreevidya B 2018 Time series analysis of air pollution in Bengaluru using ARIMA model Ambient Communications and Computer Systems ed Perez G, Tiwari S, Trivedi M, Mishra K (Singapore: Springer Nature) p 413

[7] Wijesekara W M L K N, Liyanage L 2020 Comparison of imputation methods for missing values in air pollution data: case study on Sydney air quality index Advances in Information and Communication ed Arai K, Kapoor S, Bhatia R (Cham: Springer Nature Switzerland AG) p 257

[8] Zafra C, Temprano J, Tejero I 2017 The physical factors affecting heavy metals accumulated in the sediment deposited on road surfaces in dry weather: a review Urban Water Journal 14(6) 639

[9] Tsai Y-T, Zeng Y-R, Chang Y-S 2018 Air pollution forecasting using RNN with LSTM International Symposium on Dependable, Autonomic and Secure Computing (Athens: Institute of Electrical and Electronics Engineers)

[10] Kadiyala A, Kumar A 2017 Applications of Python to evaluate environmental data science problems Environmental Progress & Sustainable Energy 36(6) 1580

[11] Montealegre Bocanegra J E 1990 Técnicas Estadísticas Aplicadas en el Manejo de Datos Hidrológicos y Meteorológicos (Bogotá: Instituto Colombiano de Hidrología, Meteorología y Adecuación de Tierras HIMAT)

[12] International Business Machines (IBM) 2021 IBM SPSS Statistics 25.0 (United States of America: International Business Machines)

[13] Fan Y, Ma H, Chen Z, Shen K 2021 Research and application of algorithm based on maximum expectation and collaborative filtering in recommended system Journal Physics: Conference Series 1754(1) 012205:1

[14] Khan F M, Gupta R 2020 ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India Journal of Safety Science and Resilience 126 121336

[15] Yang G, Zhang Y, Li X 2020 Impact of gasoline upgrade policy on particulate matter pollution in China Journal of Cleaner Production 262 121336

[16] Auffhammer M, Kellogg R 2011 Clearing the air? The effects of gasoline content regulation on air quality American Economic Review 101(6) 2687

[17] Secretaría Distrital de Ambiente de Bogotá (SDA) 2012 Resolución 1304 de 2012 (Bogotá: Secretaría Distrital de Ambiente de Bogotá)

[18] Li K, Jacob D, Shen L, Lu X, Smedt I, Liao H 2020 Increases in surface ozone pollution in China from 2013 to 2019: anthropogenic and meteorological influences Atmospheric Chemistry and Physics 20(19) 11423

[19] Ministerio de Ambiente, Vivienda y Desarrollo Territorial de Colombia (MAVDT) 2008 Resolución 909 de 2008 (Bogotá: Ministerio de Ambiente, Vivienda y Desarrollo Territorial de Colombia)

[20] Ministerio de Ambiente, Vivienda y Desarrollo Territorial de Colombia (MAVDT) 2007 Resolución 180782 de 2007 (Bogotá: Ministerio de Ambiente, Vivienda y Desarrollo Territorial de Colombia)

[21] Han J, Wang S, Vittori R, Somers L M T 2020 Experimental study of the combustion and emission characteristics of oxygenated fuels on a heavy-duty diesel engine Fuel 268 117219

[22] Alcaldía Mayor de Bogotá 2011 Decreto 623 de 2011 (Bogotá: Alcaldía Mayor de Bogotá)