Multitemporal analysis with statistical models: influence of the atmospheric condition on urban concentrations of particulate matter

L Rodríguez-Garavito¹, K J Romero-Corredor¹, and C A Zafra-Mejía¹
¹ Facultad del Medio Ambiente y Recursos Naturales, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia

E-mail: czafra@udistrital.edu.co

Abstract. This paper shows a multitemporal analysis with autoregressive integrated moving average models of the influence of atmospheric condition on concentrations of particulate matter \( \leq 10 \, \mu m \) in Bogotá city, Colombia. Information was collected from six monitoring stations distributed throughout the city. The study period was nine years. Autoregressive component of the models suggests that urban areas with greater atmospheric instability show a lower hourly persistence of particulate matter (one hour) compared to urban areas with lower atmospheric instability (two hours). Moving average component of the models hints those urban areas with greater atmospheric instability show greater hourly variability in particulate matter concentrations (5-10 hours). The models also suggest that a high degree of air pollution decreases the temporal influence of the atmospheric condition on particulate matter concentrations; in this case, the temporal behavior of particulate matter possibly depends on the urban emission sources of this pollutant rather than on the existing atmospheric condition. This study is relevant to deepen the knowledge in relation to the following aspects of atmospheric physics: The use of statistical models for the time series analysis of atmospheric condition, and the analysis by statistical models of the influence of atmospheric condition on air pollutant concentrations.

1. Introduction
The progressive deterioration of air quality in large cities has been a cause for concern in the context of public health due to its relationship with respiratory diseases [1]. This problem has its genesis mainly in factors such as accelerated economic growth and the consequent increase in urban atmospheric emissions [2]. Studies also report the significant influence of climate variables (e.g., wind speed and solar radiation) on urban air quality [3]. Thus, the stability and instability conditions of the atmosphere play a significant role in the dispersion of urban air pollutants [4]. The study of the influence of atmospheric condition (AC) on the urban concentrations of particulate matter is complex and requires specialized tools for its representation and understanding within the framework of atmospheric physics [5]. For this reason, it is appropriate to consider the time series characteristics of climate variables associated with AC and urban atmospheric pollutants of interest.

The autoregressive, integrated, and moving average (ARIMA) models have had an exceptional development, especially to analyze the temporal trends of urban air pollutants [6]. ARIMA models consist of three components (p,d,q) and each models a different behavior in the time series [7]. The main function of the difference component (d) in ARIMA models is to eliminate a possible trend (not seasonality) in the time series under study. The autoregressive component represents the influence of
previous observations on 'p' periods. Lastly, the moving average component represents the deviation in 'q' periods of the observations from their mean level [8].

The main objective of this paper is to show a multitemporal analysis with ARIMA models of the influence of AC on concentrations of particulate matter ≤ 10 µm (PM$_{10}$) in Bogotá city, Colombia. This study is relevant to deepen the knowledge in relation to the following aspects of atmospheric physics: (i) the use of ARIMA models for the time series analysis of AC, and (ii) the analysis by ARIMA models of the influence of AC on air pollutant concentrations.

2. Materials and methods
The study area was delimited by six monitoring stations distributed throughout the city of Bogotá, Colombia (4°35’56” N; 74°04’51 W). The stations were named as follows: Kennedy (KE), Carvajal (CAR), Guaymaral (GU), Corporas (CO), Barrios Unidos (BU), and Puente Aranda (PA) (see Figure 1).

![Figure 1. Location of monitoring stations in Bogotá city, Colombia. Adapted from Google Earth Pro, 2021.](image)

The methodology used in this study considered the following five phases. Phase 1: information collection. In this phase, hourly information on rainfall, temperature, wind speed, solar radiation, and PM$_{10}$ concentration was collected at each of the monitoring stations considered. The sampling period was nine years (01/01/2007-31/12/2015); it was also ensured that the time series of the variables under study had at least 75% of their data during the period considered. A total of 3286-day data was collected, corresponding to 78888 hourly data for each study variable and for each monitoring station. Hourly information of the variables under study was transformed by a moving average at the daily, weekly, and monthly timescales.

Phase 2: application of the Pasquill–Guifford model [9]. This model was used to determine the hourly AC at each monitoring station. The model considered wind speed and solar radiation as main variables [10]. In this study, the following quantitative scale was used to identify each AC; 1 = stable, 2 = slightly stable, 3 = neutral, 3.5 = neutral to slightly unstable, 4 = slightly unstable, 4.5 = slightly unstable to unstable, 5 = unstable, 5.5 = unstable to very unstable, and 6 = very unstable.

Phase 3: comparative analysis between climate information and AC. In this phase descriptive statistics were initially obtained for each time series. Subsequently, Spearman's correlation coefficient was used to study the possible relationship between climate variables and AC. Non-normality of the time series was verified by a Kolmogorov-Smirnov test.
Phase 4: development of ARIMA models. This phase followed the methodology proposed by Box and Jenkins [11] for the identification, parameter estimation, and assumption verification of ARIMA models. The software used during this phase was IBM SPSS V.25.0.

The statistics used to evaluate the fit of the models were the following, determination coefficient (R²), root-mean square error (RMSE), and mean absolute percentage error (MAPE). The Ljung-Box (Q’) statistic was also used to evaluate the ARIMA models developed. A p-value > 0.05 in this statistic indicated that the models were properly developed [11]. In addition, the normalized Bayesian information criterion was used to select the best ARIMA model. This is for when two or more ARIMA models of the same time series complied with the stage of assumption verification. Thus, the model with the lowest normalized BIC was selected [12].

Phase 5: analysis of ARIMA models. During this phase the autoregressive (p) and moving average (q) components of the ARIMA models selected for each time series were analyzed. Thus, a comparative analysis of the ARIMA temporal structure was carried out between the time series of AC and PM_{10} concentrations. This comparative analysis was performed with respect to the magnitude of the autoregressive and moving average components of ARIMA models.

3. Results and discussion

The results of the application of Pasquill-Gifford model to study the AC showed that at the CO, GU, and BU monitoring stations the dominant daytime AC was unstable (5 = unstable; hourly frequency: 23.4%). At the CAR, KE, and PA monitoring stations the dominant daytime AC was slightly unstable (4 = slightly unstable; hourly frequency: 17.5%). The findings suggested that during the daytime the dominant AC in the study city was between slightly unstable and unstable. This trend also hinted at favorable daytime AC for the dispersion of urban air pollutants. Overall, the condition of atmospheric instability in monitoring stations located in the north of the city was higher compared to monitoring stations located in the south (Table 1). Namely, in the northern zone there were possibly better ACs to disperse air pollutants. In contrast, stable ACs prevailed at night for all monitoring stations (1 = stable; hourly frequency: 39.0%). This nighttime trend implied unfavorable ACs for the dispersion of urban air pollutants.

| Zone | Station | AC 6.0 | AC 5.5 | AC 5.0 | AC 4.5 | AC 4.0 | AC 3.5 | AC 3.0 | AC 2.0 | AC 1.0 |
|------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| North| GU      | 6.18   | 13.8   | 25.20  | 1.01   | 3.81   | 0.02   | 0.13   | 0.97   | 48.90  |
|      | CO      | 5.60   | 11.0   | 20.50  | 3.70   | 8.51   | 0.34   | 2.30   | 4.66   | 43.30  |
|      | BU      | 3.02   | 11.8   | 24.40  | 2.43   | 8.25   | 0.04   | 0.39   | 3.05   | 46.60  |
|      | PA      | 1.54   | 5.75   | 15.10  | 6.05   | 15.70  | 2.49   | 9.30   | 11.80  | 32.30  |
| South| KE      | 0.81   | 3.80   | 16.50  | 7.33   | 18.70  | 1.65   | 7.76   | 11.80  | 31.60  |
|      | CAR     | 1.17   | 4.15   | 17.60  | 8.23   | 18.10  | 0.25   | 5.32   | 13.70  | 31.40  |

Frequency of occurrence in percentage (%). In bold the dominant frequencies.

Spearman’s correlation analysis between climate variables and AC was performed in daytime and nighttime, and for the four timescales considered in this study (hourly, daily, weekly, and annual). The results showed that the hourly timescale was the one that showed the best correlations; thus, the analyses focused on this timescale. The results revealed that temperature in daytime was the climate variable with the best correlations compared to rainfall.

On average, a significant direct correlation was observed between temperature and AC (rs > 0.495; p-value < 0.001). Indeed, as the temperature increased in daytime, urban atmospheric instability also increased. The monitoring station that best evidenced this behavior was GU, in the north of the city. As expected, temperature showed significant medium to strong correlations with solar radiation (rs > 0.712; p-value < 0.001) and wind speed (rs > 0.642; p-value < 0.001); therefore, these two climate variables were the basis for applying the AC model proposed by Pasquill-Gifford [10].
Figure 2 shows the hourly results of the AC simulation and the observed information of solar radiation and wind speed at the GU (north zone) and KE (south zone) stations. Lastly, rainfall in daytime and nighttime did not show significant correlations with the simulated AC.

**Figure 2.** Observed solar radiation and wind speed, and simulated AC at the GU (a) and (b), KE (c) and (d) stations. GU station = north zone, and KE station = south zone. RD = solar radiation, WS = wind speed, SD = standard deviation, AC = atmospheric condition, UL = upper limit, and LL = lower limit.

ARIMA modelling results showed that hourly PM$_{10}$ concentrations at monitoring stations located in the north zone of the city had a shorter memory (AR = 1) compared to monitoring stations located in the south zone (AR = 2), Table 2. In the north zone, PM$_{10}$ concentrations at a given time were influenced by concentrations observed during the immediately preceding hour. In contrast, in the southern zone, PM$_{10}$ concentrations were influenced by what happened until two hours earlier. This indicated a greater hourly persistence of this pollutant in the atmosphere of the southern zone.

In other words, PM$_{10}$ concentrations in the south zone persisted for two hours and in the north zone for one hour. These findings were consistent with the results of AC simulation with the Pasquill-Gifford model. In the north zone, the AC simulated in daytime was unstable (AC = 5) and in the south zone it was slightly unstable (AC = 4), see Table 1. Therefore, this study hinted at increased atmospheric instability in the northern zone of the city, which probably decreased the PM$_{10}$ persistence. Coccia [4] also reported a similar urban trend.

Analysis of the difference component in ARIMA models suggested that the PM$_{10}$ concentration time series showed a possible trend (d = 1) in all the monitoring stations considered (Table 2). Namely, the time series were not stationary. In this study, PM$_{10}$ concentrations tended to decrease during the sampling period. The low magnitude of this component for all monitoring stations implied a weak decline trend during the study period; on average, PM$_{10}$ concentrations in the city decreased 3.74%/year during the sampling period (9 years). Other researchers [7] have reported similar trends in urban areas.
### Table 2. ARIMA models for hourly PM$_{10}$ concentrations.

| Zone | Station | AR | I | MA | R$^2$ | RMSE | MAPE | P-value | BIC |
|------|---------|----|---|----|------|------|------|---------|-----|
| North | GU | 1 | 1 | 5 | 0.989 | 1.402 | 1.711 | 0.328 | 0.683 |
|       | CO | 1 | 1 | 10 | 0.997 | 0.891 | 1.863 | 0.285 | -0.221 |
|       | BU | 1 | 1 | 10 | 0.997 | 0.891 | 1.863 | 0.285 | -0.221 |
|       | PA | 2 | 1 | 5 | 0.997 | 1.777 | 1.141 | 0.159 | 1.157 |
| South | KE | 2 | 1 | 6 | 0.994 | 2.015 | 1.177 | 0.071 | 1.402 |
|       | CAR | 2 | 1 | 6 | 0.993 | 2.324 | 1.325 | 0.036 | 1.681 |

In relation to the moving average component of ARIMA models, the findings showed differences between the monitoring stations located in the north and south of the city. In the north zone, greater hourly variations in PM$_{10}$ concentrations were observed. These variations lasted up to 10 hours (MA between 5 hours – 10 hours). In the southern monitoring stations, hourly variations in the PM$_{10}$ concentration were observed that lasted up to 6 hours (MA between 5 hours – 6 hours), Figure 3. These results were dependable with the AC simulation performed using the Pasquill-Gifford model.

In the northern zone, the simulated AC in daytime was unstable (AC = 5) and in the southern zone of the city it was slightly unstable (AC = 4) (Table 1). Therefore, the findings suggested greater atmospheric instability in the northern zone of the city, which possibly increased hourly variations in PM$_{10}$ concentrations.

![GU Station](image1.png)  
(a)  

![KE Station](image2.png)  
(b)

**Figure 3.** Hourly average concentrations of PM$_{10}$ observed and simulated during the study period at the GU (a) and KE (b) stations. GU station = north zone, and KE station = south zone. UL = upper limit, and LL = lower limit.

The results of the hourly comparative analysis between the ARIMA models for PM$_{10}$ concentrations and AC showed differences in their temporal structure. This in relation to the monitoring stations located in the north and south of the city. In the north zone, a similar hourly structure was observed for each monitoring station between PM$_{10}$ concentrations and AC. Namely, the magnitude of the autoregressive
components of the ARIMA models developed was similar for both PM$_{10}$ concentrations and AC, Table 3. This trend suggested that AC influenced the temporal behavior of PM$_{10}$ concentrations in the north of the city. In contrast, in the southern zone, the temporal structure was not similar for each monitoring station.

The results showed that the autoregressive components of the ARIMA models for PM$_{10}$ concentrations were higher (AR = 2), compared to the same components obtained for the AC time series (AR = 1). These findings hinted that in the southern zone of the city the AC did not have the same influence on PM$_{10}$ concentrations, this compared to the northern zone. Apparently, PM$_{10}$ persisted longer compared to the AC. This behavior may have been associated with the high PM$_{10}$ concentrations observed in the south of the city during the study period (average: 76.4 µg/m$^3$). On average, PM$_{10}$ concentrations in the southern city were 1.81 times higher compared to concentrations observed in the northern zone. Other researchers [13] have also reported this trend in cities with high levels of air pollution.

In relation to the moving average components of the ARIMA models developed, the results also showed differences between the north and south of the city. In the northern zone, the magnitude of these components was similar at each monitoring station (Table 3). Namely, during the study period, similar variations in PM$_{10}$ concentrations and AC were observed (periods between 5 hours - 13 hours). These two phenomena were possibly related.

In contrast, in the southern zone, the variations in these two-time series were not similar at each monitoring station. The PM$_{10}$ concentration time series showed smaller variations (between 5 hours - 6 hours) compared to the AC time series (between 8 hours - 14 hours). The above trend could be explained by the high air pollution levels by PM$_{10}$ in the south of the city. As there were high PM$_{10}$ concentrations, its variation was possibly smaller over time compared to the AC.

| Zone | Station | PM$_{10}$ (p,d,q) | AC (p,d,q) | $R^2$ | RMSE | MAPE | P-value | BIC |
|------|---------|------------------|-----------|-------|-------|-------|---------|-----|
| North | GU      | (1,1,5)          | (1,1,5)   | 0.961 | 0.017 | 0.285 | 0.059   | -8.106 |
|       | CO      | (1,1,5)          | (1,1,5)   | 0.965 | 0.026 | 0.539 | 0.127   | -7.325 |
|       | BU      | (1,1,10)         | (1,1,13)  | 0.996 | 0.003 | 0.061 | 0.075   | -7.547 |
| South | PA      | (2,1,5)          | (1,1,14)  | 0.967 | 0.032 | 0.783 | 0.163   | -6.890 |
|       | KE      | (2,1,6)          | (1,1,8)   | 0.969 | 0.033 | 0.747 | 0.076   | -6.823 |
|       | CAR     | (2,1,6)          | (1,1,8)   | 0.982 | 0.042 | 0.792 | 0.083   | -6.317 |

ARIMA models developed in day and nighttime.

4. Conclusions

The findings of this multitemporal study with ARIMA models to analyze the influence of the AC on urban PM$_{10}$ concentrations allow us to visualize that the AC simulation using the Pasquill-Gifford model in the northern zone of the city there is a greater degree of atmospheric instability (AC = 5, unstable) compared to the southern zone (AC = 4, slightly unstable). This scenario of greater atmospheric instability in the northern zone, possibly offers conditions more favorable for the dispersion of urban air pollutants; in this case for PM$_{10}$.

Findings from ARIMA models suggest that AC influences the temporal persistence of PM$_{10}$; urban areas with greater atmospheric instability show a lower temporal persistence of PM$_{10}$ (one hour); in urban areas with less atmospheric instability, the temporal persistence of this air pollutant increases to two hours. ARIMA multitemporal analysis shows that AC also influences the variability of urban PM$_{10}$ concentrations. Indeed, urban areas with greater atmospheric instability show greater variability in PM$_{10}$ concentrations; a high degree of air pollution appears to decrease the temporal influence of AC on PM$_{10}$ concentrations. Namely, the temporal behavior of PM$_{10}$ concentrations in this case possibly depends on the emission sources of this pollutant rather than on the existing AC.
References

[1] Patella V, Florio G, Magliacane D, Giuliano A, Crivellaro M A, Di Bartolomeo D, Genovese A, Palmieri M, Postiglione A, Ridolo E, Scalleto C, Ventura M T, Zollo A 2018 Urban air pollution and climate change: “The Decalogue: Allergy Safe Tree” for allergic and respiratory diseases care Clinical and Molecular Allergy 16 20

[2] Mudway I S, Dundas I, Wood H E, Marlin N, Jamaludin J B, Brenner S A, Cross L, Grieve A, Nanzer A, Barratt B M, Beevers S, Dajnak D, Fuller G W, Font A, Colligan G, Sheikh A, Walton R, Grigg J, Kelly F J, Lee T H, Griffiths C J 2019 Impact of London’s low emission zone on air quality and children’s respiratory health: a sequential annual cross-sectional study The Lancet Public Health 4 28

[3] Mika J, Fogo P, Lakatos L, Olah A B, Rapi S, Utsa Z 2018 Impact of 1.5K global warming on urban air pollution and heat island with outlook on human health effects Current Opinion in Environmental Sustainability 30 151

[4] Coccia M 2021 The effects of atmospheric stability with low wind speed and of air pollution on the accelerated transmission dynamics of COVID-19 International Journal of Environmental Studies 78 1

[5] Alimissis A, Philippopoulos K, Tzanis C G, Deligiorgi D 2018 Spatial estimation of urban air pollution with the use of artificial neural network models Atmospheric Environment 191 205

[6] Hosamane S N, Prashanth K S, Virupakshi A S 2020 Assessment and prediction of PM$_{10}$ concentration using ARIMA Journal of Physics: Conference Series. 1706(1) 012132:1

[7] Kulkarni G E, Muley A A, Deshmukh N K, Bhalchandra P U 2018 Autoregressive integrated moving average time series model for forecasting air pollution in Nanded city, Maharashtra, India Model. Earth Syst. Environ. 4 1435

[8] Ruchiraset A, Tantrakarnapa K 2018 Time series modeling of pneumonia admissions and its association with air pollution and climate variables in Chiang Mai Province, Thailand Environ Sci Pollut Res 25 33277

[9] Kahl J D W, Chapman H L 2018 Atmospheric stability characterization using the Pasquill method: a critical evaluation Atmospheric Environment 187 196

[10] Chambers S D, Williams A G, Crawford J, Griffiths A D 2015 On the use of radon for quantifying the effects of atmospheric stability on urban emissions Atmospheric Chemistry and Physics 15 1175

[11] Box G E P, Jenkins G M, Reinsel G C, Ljung G M 2016 Time Series Analysis: Forecasting and Control (New Jersey: John Wiley & Sons)

[12] Ljung G M, Box G E P 1978 On a measure of lack of fit in time series models Biometrika 65 297

[13] Chen Z, Barros C P, Gil-Alana L A 2016 The persistence of air pollution in four mega-cities of China Habitat International 56 103