Temporal analysis of the heavy metal concentration in road sediment and dust using statistical models

C A Zafra-Mejía¹, H A Rondón-Quintana¹, and L C Echeverry-Prieto¹
¹ Facultad del Medio Ambiente y Recursos Naturales, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia

E-mail: czafra@udistrital.edu.co, harondonq@udistrital.edu.co

Abstract. The objective of this paper is to show a temporal analysis using autoregressive integrated moving average models of the heavy metal concentration in road sediment and dust of Soacha, a Colombian locality. The representative size fractions in the road sediment and dust were <250 μm and ≤10 μm, respectively. The results suggest that lead is the best metallic element to study the relationship between the heavy metal concentration in the road sediment and dust (r-Pearson = 0.90). Univariate models (R² ≥ 0.58) suggest that the time series of lead concentrations in road sediment and dust have the same temporal structure. Namely, because they are first-order autoregressive processes, concentrations at a given moment of time are influenced by the immediately preceding concentrations. The transfer function model (R² = 0.91) suggests that there is no delay in impulse transfer from road dust concentration to lead concentration in the road sediment. The effect is immediate for a sampling interval of 3 days. The results show that modeling has a better fit during the rainy season compared to the dry season. In the context of the simulation of physical phenomena in engineering, this study is relevant to deepen knowledge in relation to the use of autoregressive integrated moving average models.

1. Introduction

Urban surface pollution with heavy metals (HMs) required special attention due to its toxic effects on human health [1]. Road sediments (RSs) accumulated in dry time associated HMs that affected water resources when they were transported by surface runoff [2]. In addition, RSs affected the air quality when were re-suspended by wind and traffic-induced turbulence. This re-suspended material was commonly identified as road dust (RD) [3].

The available literature on methodologies for studying the HMs associated with RS from information on air quality or RD is scarce. The RS has been recognized as an important urban source of atmospheric particulate matter ≤10 μm (PM₁₀) [4]. This contribution of particulate atmospheric material was due to the mechanical disintegration of the RS by crushing, grinding, and explosion processes [5]. Thus, exposure to atmospheric particulate material generated on roads involved increased exposure to ultrafine particles and PM₁₀ emissions from vehicle exhaust, and from brake and tire wear [6]. Therefore, it is then necessary to study the HM content in RS and RD from new approaches to analysis. In this way, the time-series analysis using autoregressive integrated moving average (ARIMA) models can be useful to deepen knowledge in relation to the phenomenon of HM accumulation in RS and RD.

The ARIMA models [7] were effectively used in the study of HM content in different environmental matrices [8] and were widely used for their simplicity and statistical characteristics [9]. These models also offer an effective method for obtaining accurate outcomes without considerable calibration time.
ARIMA models examine the probable temporal association between present and previous information and its errors and shoulder a linear relationship between the data. Hence, each observation can be described as a linear function of its previous data with an extra error term [10].

Explicitly, these models consider the information sequence over time as a random series and analyze a fraction of the information in the sequence to calculate definite parameters that specify the mathematical model, and then model the time series. The ARIMA models consider the residual information in the sequence to validate the mathematical model, and the validated model is utilized to forecast successive data in the time series [11].

The objective of this paper is to show a temporal analysis with ARIMA models of the HM concentration (Pb, Zn, Cu, Ba, and Fe) associated with RS and RD in Soacha, a Colombian locality. This research will be relevant to deepen knowledge in relation to the following aspects in the context of the simulation of physical phenomena in engineering: (i) the temporal behavior of HMs associated with RS, (ii) the possible relationship between HM concentrations in RS and RD, and (iii) the use of ARIMA models to study the temporal relationship of HMs in RS and RD.

2. Materials and methods
This chapter shows a description of the study site and sampling system, and the protocols for laboratory and data analysis.

2.1. Study site
The study road was in the municipality of Soacha, Colombia, in central Colombia, 4°35’05” N, 74°13’12” W. Its tropical mountain climate (cold climate) was characterized by an average annual temperature of 14 °C, wide daily temperature variation (up to 18.0 °C), average annual relative humidity of 66.0%, and average annual rainfall of 900 mm; all climatological data were obtained from a station located at a 340 m distance from the road surface under study. The study road was in the urban center of Soacha, Colombia, adjacent to the central municipal hospital. The road basin had an average elevation of 2560 m.a.s.l., average slope of 1.70%, and waterproof coverage in 90% of its area. The road surface had a one-way street with two lanes for vehicle traffic. The road gives access to residential and institutional areas but was occasionally used by commercial vehicles. In order of importance the land use in the study area was residential, institutional, and commercial.

2.2. Sampling system
The RS collection system consisted of three components: (i) a wooden frame with an area of 0.50 m² (0.707 m x 0.707 m), (ii) a manual brush of synthetic fibers, and (iii) a manual plastic collector [12]. Samples of RS were taken in dry time (every 3 days), at 0.40 m from the curb, between 11:00 - 13:00 hours, and over a continuous period of 127 days (07/01/2018-14/05/2018). The road surface was swept with a brush of synthetic fibers (3 repetitions) so that the sediments most strongly adhered to it were available to be collected. The total number of samples collected on the road surface was 44.

The collected RS was stored in airtight plastic bags to later be taken for lab analysis. The sampling protocol for total suspended particle (TSP) and PM₁₀ (RD) was based on the U.S. Environmental Protection Agency (EPA/625/R-96/010a-IO-2.1) [13]. The TSP sampling system consisted of high-volume manual equipment with mass flow control (Thermo Environmental Instruments Inc., GMW TSP high volume air sampler - mass flow). The sampling system for PM₁₀ consisted of a high-volume manual cyclone-type equipment with volumetric flow control (Wedding & Associates, 600 PM₁₀ critical flow high volume air sampler - volumetric flow).

2.3. Lab analysis
The particle size distribution of RS (< 63 μm - 2800 μm) was determined from ISO 2591-1 standard [14]. The HM extraction from RS and RD was carried out from the ISO 11466 method [15]. The method used to determine the HM content (Pb, Zn, Cu, Ba, and Fe) in RS and RD was as follows: SM 3125 B, inductively coupled plasma-mass spectrometry (ICP-MS) Method [16]. The selection, preparation, and
gravimetric determination method for TSP and PM$_{10}$ filters was based on the U.S. Environmental Protection Agency (EPA/625/R-96/010a-IO-3.1) [13].

2.4. Data analysis
The data analysis considered two stages; stage 1: A descriptive statistical assessment was applied to identify anomalies in HM concentration time series. The normal distribution of HM concentration time series was estimated by a Kolmogorov/Smirnov test (p-value > 0.05). Associations between HMs were examined using Pearson's correlation coefficient ($r$) to identify probable control HMs of their behavior in RS and RD. In this study, we supposed that the HMs with the best associations were those that perhaps described their behavior in RS and RD. Lastly, the influence of rainfall on the HM concentrations in RS and RD was studied. Thus, two rainfall periods were studied: dry season or decrease in rainfall (January/07-February/08) and rainy season (February/09-May/14).

Stage 2: ARIMA models for HM concentration time series were generated applying the methodology of Box, et al. [7]. The three phases stated by Box, et al. [7] were considered in the construction of the models: identification/estimation/verification. During the model identification phase, the orders of the autoregressive and moving average polynomials were calculated, as well as the degree of differentiation to eliminate the non-seasonality of the time series under study. Specifically, the orders ‘p’, ‘d’, and ‘q’ of the models were calculated. In the second phase, the parameters of the autoregressive and moving average polynomials were calculated by the maximum likelihood method for each model found.

The eight hypotheses suggested by Box, et al. [7] were checked to choose the best models for each time series. The appropriateness of modelling was determined using the Ljung and Box (Q’) parameter. A p-value > 0.05 meant that the model was appropriately generated to explain the association information in time series [17]. Another parameter applied to distinguish the best model was the normalized Bayesian information criterion (normalized BIC). The best model was chosen when the BIC was the lower [18]; All data analysis was implemented using the IBM-SPSS V.18.0 software;

3. Results and discussion
The following are the results and discussions related to the HM concentration in RS and RD, and to the ARIMA models developed.

3.1. Heavy metal concentration in road sediment and road dust
The finer fraction of the collected RS (< 63 μm) showed the highest HM concentrations (Table 1). Overall, the HM concentration tended to decrease as the particle size of the RS increased. When comparing the HM concentrations between PM$_{10}$ and TSP, it was observed that these were higher in the potentially respirable fraction (PM$_{10}$). For all study HMs, the concentrations associated with PM$_{10}$ were 3.09 times higher than TSP (Table 2). Therefore, the results suggested that the HM sources mainly emitted particles of size less than or equal to 10 μm.

Chen et al. [4] reported similar findings. Pearson's correlation analysis was performed between HM concentrations in the RS (< 250 μm) and PM$_{10}$. The results showed that Pb was the HM that showed the best correlation coefficient ($r = 0.90$), followed by Cu (0.66), Ba (0.62), Fe (0.40), and Zn (0.12). The HMs with the best correlation coefficients were possibly best suited to study the relationship between HM concentrations in the RS and PM$_{10}$ (Pb > Cu > Ba). In this way, Pb was selected as the representative element to study the relationship between the HM concentration in RS and PM$_{10}$ (RD).

3.2. Autoregressive integrated moving average analysis
Univariate ARIMA models were developed for the time series of Pb concentrations in the RS and PM$_{10}$ (RD). The time series of PM$_{10}$ concentration showed a decreasing trend during the study period. This decrease was possibly associated with rainfall behavior. Namely, in the dry season (January/07 - February/08) the highest PM$_{10}$ concentrations were observed while in the rainy season (February/09 - May/14) the concentrations tended to decrease. In addition, it was observed that the variation in PM$_{10}$ concentration showed a pattern of decrease. In the dry season, the variation in the PM$_{10}$ concentration
was greater than in the rainy season. The time series for the Pb concentration in RS showed the same trends (Figure 1). Thus, the results suggested that the time series under study did not correspond to stationary processes.

Table 1. HM concentration according to the size fraction of RS (95% confidence interval).

| HM  | Size fraction (µm) | < 63 | 63-125 | 125-250 | 250-500 | 500-1000 | 1000-2000 | 2000-2800 |
|-----|-------------------|------|--------|---------|---------|----------|----------|----------|
| Pb  |                   | 99±26| 79±21  | 74±19   | 81±21   | 56±15    | 44±11    | 32±80    |
| Zn  |                   | 137±19| 86±12  | 65±90   | 57±80   | 29±40    | 18±30    | 7±20     |
| Cu  |                   | 46±11| 39±90  | 38±90   | 33±80   | 17±40    | 12±30    | 10±20    |
| Ba  |                   | 192±46| 139±33 | 105±25  | 90±22   | 47±11    | 26±60    | 12±30    |
| Fe  |                   | 11.69±0.9| 8.79±0.7 | 8.25±0.7 | 9.60±0.8 | 7.45±0.6 | 4.59±0.4 | 2.74±0.2 |

*HM concentration in mg/kg of dry matter. *HM; † concentration in g/kg of dry matter.

Table 2. HM concentration in RD: PM10 and TSP (95% confidence interval).

| HM  | PM10 (≤ 10 µm) | TSP (≤ 100 µm) |
|-----|----------------|----------------|
| Pb  | 1818±473       | 86±184         |
| Zn  | 16312±5256     | 3065±1201      |
| Cu  | 1197±273       | 458±700        |
| Fe  | 10515±2688     | 9331±1766      |
| Ba  | 7001±2247      | 1625±5730      |

*Concentration in µg/g of dry matter.

From the viewpoint of the ARIMA process, it was observed that the time series of Pb concentration in the RS had a short memory. Namely, being a first-order autoregressive process (AR-p = 1), it was observed that the Pb concentration in the RS was influenced by the concentration immediately preceding (Table 3). In addition, the time series of PM10 concentration showed a similar ARIMA structure (AR-p = 1). Thus, it was probably a phenomenon of short memory as well as the phenomenon of Pb concentration in the RS. Vicente, et al. [19] obtained similar results by studying daily air pollution by particulate matter, NO2, As, Cd, Ni, and Pb. These researchers identified a short-memory ARIMA model, that is, with a second-order autoregressive term (AR-p = 2).

As observed, there was the possibility of representing the two-time series under study (Pb and PM10) using the same ARIMA time structure (1,0,0). It is important to mention, that in this study the immediately preceding observation corresponded to three days. This is based on the frequency of RS collection and PM10 sampling. In the ARIMA models developed, a moving average term (MA-q = 0) was not evidenced. In other words, no fluctuations were visualized in the time series under study. This with respect to the average value of Pb and PM10 concentrations observed during the study period.

Table 3. Univariate ARIMA models for the time series of Pb concentrations in RS and PM10 (RD).

| Time series | Model (p,d,q) | Transf. | R² | RSME | MAPE | MAE | Q² p-value | BIC | Outliers |
|-------------|---------------|---------|----|------|------|-----|------------|-----|----------|
| Pb          | (1,0,0)       | LN      | 0.577 | 34.17 | 32.01 | 23.69 | 0.577 | 7.313 | 1        |
| PM10        | (1,0,0)       | NT      | 0.831 | 10.45 | 30.78 | 8.05 | 0.944 | 5.028 | 2        |

*Transf. = transformation; LN = natural logarithm; NT = no transformation. When a transformation was applied, the generated model included a constant.

After analyzing the univariate models, a transfer function model was developed to relate the two-time series (Pb and PM10) using an ARIMA model. The hypothesis for the development of the model was the probable existence of a relationship between the Pb concentrations in RS and PM10 (RD). Namely, it could be expected that there was a dynamic relationship between the two variables under study. Thus, a unidirectional causal relationship was suggested from the PM10 concentration to the Pb
The transfer function model considered a single input and output, corresponding to a system related by a linear filter as proposed by Box, et al. [7]. The transfer function modeling allowed the identification and estimation of the transfer function and the noise model, based on the information provided by the time series of PM$_{10}$ (independent variable) and Pb (dependent variable) concentrations in RS. The univariate models obtained for the input variable (PM$_{10}$ concentration) were used to perform the pre-bleaching of the output variable (Pb concentration in RS). Thus, the model transfer function [20].

The estimated cross-correlation function between the two study variables was also calculated to determine the impulse response function. The results showed that the '0' delay was the only one statistically significant, according to the limit line obtained using the Bartlett criterion [21]. Namely, in this study, there was no delay in the impulse transfer from the PM$_{10}$ concentration to the Pb concentration in RS. It is important to mention that the PM$_{10}$ monitoring station was located 5.0 m from the curb of the roadway where the RS was collected. Once the transfer function was identified, the noise series was estimated, and the ARMA process generating the residues was identified through its simple and partial autocorrelation functions.

The results showed that the residues followed an ARMA process (0,0). In addition, an outlier was identified that was included in the model of the complete transfer function (dummy variable). The outlier was in data 12 (February/03). This data corresponded to a minimum value in the concentrations of PM$_{10}$ (RD) and Pb (RS), which preceded the maximum concentrations of the variables under study, respectively (Figure 1). This outlier was of the innovative type, that is, in the time series it affected each observation from a specific starting point in the time series.

The statistics of the transfer function model were studied to evaluate the goodness of fit in time series modeling (Table 4). In this regard, an estimate was made of the proportion of total variance in the series explained by the model ($R^2 = 0.910$). Thus, the PM$_{10}$ concentration explained 91.0% of the variation in the Pb concentration in RS. The results showed that the adjustment of the transfer function model was better compared with the adjustment obtained for each of the univariate models of the time series of concentrations of PM$_{10}$ ($R^2 = 0.831$) and Pb ($R^2 = 0.577$).

Additionally, the average absolute error (MAE) of the ARIMA modeling showed that the predicted values for the time series of Pb concentration in the RS deviated on average 11.36 mg/kg with respect to the observed values (Table 4). The average absolute percentage error (MAPE) of the modeling showed an average deviation of 15.2% between the observed and predicted values for the Pb concentration in RS. In order to determine the worst case in the prediction, we proceeded to determine the maximum mean absolute error (MaxAE: 52.1 mg/kg; MaxAPE: 53.6%).

The results suggested that the maximum errors in the forecast tended to occur in the dry season (e.g., February/10), and the smallest errors tended to occur in the rainy season (e.g., May/10) (Figure 1). This from the periods identified with respect to rainfall (dry season and rainy season). As is known, the highest Pb concentrations in RS occurred in the dry season, and the lowest concentrations occurred in the rainy season. Thus, the ARIMA model of transfer function developed tended to show a better fit during the rainy season.

| Variable | Time series | Model (p,d,q) | Transf. | $R^2$ | RSME | MAPE | MAE | $Q^2$ | p-value | BIC | Outliers |
|----------|-------------|---------------|---------|-------|-------|-------|-----|-------|---------|-----|----------|
| Dependent | Pb | PM$_{10}$ (1,0,0) | LN | 0.910 | 16.17 | 15.21 | 11.35 | 0.795 | 5.983 | 1 |
| Independent | Pb | PM$_{10}$ (0,0,0) | NT | 0.910 | 16.17 | 15.21 | 11.35 | 0.795 | 5.983 | 1 |

*Transf. = transformation; LN = natural logarithm; NT = no transformation. When a transformation was applied, the generated model included a constant.

Finally, the confidence limits (95%) for the predictions of Pb concentration in RS were determined to evaluate possible scenarios in the time series (Figure 1). The results showed that in the dry season the highest Pb concentrations were observed in RS. From the upper limit of the confidence interval, the
concentration could reach values of up to 378 mg/kg. Namely, in dry weather there was a probability of reaching concentrations of up to 4.68 times above the average Pb concentration for the entire study period (80.7 mg/kg). Indeed, in the rainy season the lowest values in the Pb concentration in RS were observed. From the lower limit of the confidence interval, Pb concentration in RS could reach values of up to 20.2 mg/kg. Thus, in rainy weather there was a probability of reaching concentrations of up to 4.0 times below the average Pb concentration observed during the study period.

![Figure 1. Pb concentrations observed and simulated with the ARIMA transfer function model. UL = upper limit and LL = lower limit (95% confidence interval).](image)

4. Conclusions
The results of this study on the HM concentration in RS and RD using ARIMA models allow to visualize the following conclusions.

The results suggest that Pb is the best element to study the relationship between HM concentration in the RS and PM$_{10}$ ($r = 0.90$). Overall, there is a coincidence in the correlations between HMs in the RS and PM$_{10}$ (RD). In this regard, Pb, Cu, and Ba are suggested as possible indicators to study the behavior of HMs in road environments (RS and RD).

The univariate ARIMA models ($R^2 \geq 0.58$) suggest that the time series of Pb concentrations in the RS and PM$_{10}$ have the same temporal structure. These two-time series are short memory. Namely, because they are first-order autoregressive processes (AR = 1), the concentrations at a given instant of time are influenced by the immediately previous concentrations. In this study, the analysis interval was three days. According to the ARIMA approach, no fluctuations are observed in the time series under study (MA = 0).

The ARIMA model of transfer function suggests that there is no delay in the impulse transfer from the PM$_{10}$ concentration to the Pb concentration in RS. Namely, the effect is immediate from the analysis time interval considered in this study (3 days). The transfer function model shows a good fit ($R^2 = 0.910$). The average absolute percentage error of the modeling is 15.2%. The results suggest that ARIMA modeling has a better fit during the rainy season compared to the dry season. Indeed, the innovative atypical data detected occurs during the dry season.

From the confidence limits generated by ARIMA modeling, it is observed that during the dry season the Pb concentrations in RS can be up to 4.68 times higher compared to the average concentration value calculated for the whole study period. In the rainy season, Pb concentrations can be up to 4.0 times lower compared to the calculated average concentration value.

Finally, this study is relevant to deepen knowledge in relation to the use of ARIMA models for the simulation and interpretation of physical phenomena in engineering, in this specific case, associated with environmental and earth sciences.
References

[1] Kaonga C C, Kosamu I B M, Utembé W R 2021 A review of metal levels in urban dust, their methods of determination, and risk assessment Atmosphere 12 891

[2] Jeong H, Choi J Y, Lee J, Lim J, Ra K 2020 Heavy metal pollution by road-deposited sediments and its contribution to total suspended solids in rainfall runoff from intensive industrial areas Environmental Pollution 265 115028

[3] Aguilera A, Bautista F, Gutiérrez-Ruiz M, Ceniceros-Gómez A E, Cejudo R, Goguitchaichvili A 2021 Heavy metal pollution of street dust in the largest city of Mexico, sources and health risk assessment environ monit Assess 193 193

[4] Chen S, Zhang X, Lin J, Huang J, Zhao D, Yuan T, Huang K, Luo Y, Jia Z, Zang Z, Qiu Y, Xie L 2019 Fugitive road dust PM2.5 emissions and their potential health impacts Environ. Sci. Technol. 53 8455

[5] Amato F, Pandolfi M, Viana M, Querol X, Alastuey A, Moreno T 2009 Spatial and chemical patterns of PM10 in road dust deposited in urban environment Atmospheric Environment 43 1650

[6] Schauer J J, Lough G C, Shafer M M, Christensen W F, Arndt M F, DeMinter J T, Park J-S 2006 Characterization of metals emitted from motor vehicles Res. Rep. Health Eff. Inst. 133 77–88

[7] Box G E P, Jenkins G M, Reinsel G C 1978 Estimating the dimension of a model The Annals of Statistics 6 461

[8] Piñilla Barreiro M del P, Pina-Pina R 2021 A review of metal levels in urban dust, their methods of determination, and risk assessment Assess 193 193

[9] Nourani V, Asghari P, Sharghi E 2021 Artificial intelligence based ensemble modeling of wastewater treatment plant using jittered data Journal of Cleaner Production 291 125772

[10] Valipour M 2015 Long-term runoff study using SARIMA and ARIMA models in the United States Meteorological Applications 22 592

[11] Mejía C Z, González J T, Monzón I T 2011 Concentración y distribución de metales pesados (Pb, Zn, Cu, Cd y Cr) en sedimentos viarios urbanos Revista Facultad de Ingeniería Universidad de Antioquia 58 53

[12] United States Environmental Protection Agency (U.S. EPA) 2019 Compendium of Methods for the Determination of Inorganic Compounds in Ambient Air, EPA/625/R-96/010a (United States of America: United States Environmental Protection Agency)

[13] International Organization for Standardization (ISO) 2007 Test Sieving - Part 1: Methods using Test Sieves of Woven Wire Cloth and Perforated Metal Plate, ISO 2591-1 (Switzerland: International Organization for Standardization)

[14] International Organization for Standardization (ISO) 2007 Soil Quality - Extraction of Trace Elements Soluble in Aqua Regia, ISO 11466 (Switzerland: International Organization for Standardization)

[15] American Public Health Association (APHA) 2018 Metals by Inductively Coupled Plasma-Mass Spectrometry, Standard Methods for the Examination of Water and Wastewater, SM 3125 (United States of America: American Public Health Association)

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

[17] Schwarz G 1978 Estimating the dimension of a model The Annals of Statistics 6 461

[18] Vicente A B, Jordán Vidal M M, Sanfeliu Montolio T, Sánchez À, Esteban M D 2012 Air pollution prediction models of particles, As, Cd, Ni and Pb in a highly industrialized area in Castellón (NE, Spain) Environmental Earth Sciences 66 879

[19] Finkenratt A 1991 Forecasting with Dynamic Regression Models (New Jersey: John Wiley & Sons)

[20] Bartlett M S 1946 On the theoretical specification and sampling properties of autocorrelated time-series Journal of the Royal Statistical Society 8 27