Quid N° 23, pp. 13-24, jul - dic, 2014, ISSN: 1692-343X, Medellín-Colombia

ANÁLISIS DE DEPENDENCIA USANDO REGRESIÓN LINEAL MÚLTIPLE Y REGRESIÓN DE COMPONENTES PRINCIPALES EN CONTAMINANTES DEL AIRE Y PARÁMETROS METEOROLÓGICOS: CASO DE ESTUDIO

DEPENDENCE ANALYSIS USING MULTIPLE LINEAR REGRESION AND PRINCIPAL COMPONENT REGRESSION IN AIR POLLUTANTS AND METEOROLOGICAL PARAMETERS: CASE STUDY

(Recibido el 22 octubre de 2014 - Aprobado el 07 de noviembre de 2014)

Julián Novoa Betancur
Instituto Tecnológico Metropolitano (ITM),
Grupo de Investigación e Innovación Biomédica (GI2B), Medellín-Colombia
novoa.julian@gmail.com

Daniel Parra
Instituto Tecnológico Metropolitano (ITM),
Grupo de Investigación e Innovación Biomédica (GI2B), Medellín-Colombia
Danielparra91457@correo.itm.edu.co

PhD. Silvia Quijano
Instituto Tecnológico Metropolitano (ITM),
Docente investigadora, Ciencias Exactas y Aplicadas, Medellín-Colombia
silviaquijano@itm.edu.co

Mg. Diana Orrego
Instituto Tecnológico Metropolitano (ITM),
Docente investigadora, Grupo de Investigación e Innovación Biomédica (GI2B), Medellín-Colombia
Dianaorrego@itm.edu.co

Mg. Miguel Becerra
Institución Universitaria Salazar y Herrera (IUSH)
Docente investigador, Escuela de Ingeniería, Grupo de Electrónica y Automática (GEA). Medellín-Colombia.
m.becerra@iush.edu.co

Resumen. La contaminación del aire ha sido un problema ambiental y de salud en el mundo, se han propuesto varios métodos de regresión lineal multivariable para determinar niveles de asociación y dependencia entre contaminantes del aire y parámetros meteorológicos. Estos parámetros se caracterizan por tener información redundante e irrelevante, que es perjudicial para un modelo de regresión. En este contexto, un análisis de la dependencia mediante Regresión Lineal Múltiple y Regresión de Componentes Principales de los contaminantes del aire y parámetros meteorológicos fue desarrollado, encontrando una fuerte correlación entre el ozono y la temperatura (0.8) para la estación Tlalnepantla, así como un coeficiente de determinación de 82%. El modelo de PCR demostró que parámetros meteorológicos como temperatura y humedad relativa tienen un impacto significativo en la contaminación del aire en la Ciudad de México.

Palabras clave: análisis de dependencia; contaminación del aire; parámetros meteorológicos; regresión de componentes principales; regresión lineal múltiple.

Abstract. Air pollution has been an environmental and health problem in the world, several methods of multivariable linear regression has been proposed to determine levels of association and dependence between air pollutants and meteorological parameters. These parameters are characterized by redundant and irrelevant information, which is detrimental to a regression model. In this context a dependence analysis using Multiple Linear Regression and Principal Component Regression in air pollutants and meteorological parameters was developed, found a strong correlation between ozone and temperature (0.8) for Tlalnepantla station, as well as a determination coefficient of 82%. The PCR model proved that meteorological parameter such as temperature and relative humidity have a significant impact on the air pollution in Mexico City.

Keywords: air pollution; dependence analysis; meteorological parameters; multiple linear regression; principal component regression.

Citar, estilo APA: Novoa, J., Parra, D., Quijano, S., Orrego, D., Becerra, M. (2014) Análisis de dependencia usando regresión lineal múltiple y regresión de componentes principales en contaminantes del aire y parámetros meteorológicos: caso de estudio. *Revista QUID,* (23), 13-20.

1. INTRODUCTION

The high levels of air pollutants cause adverse effects on human health and environment (Bhaskar, 2010), producing 3.7 million premature death in the world per year, mainly heart disease and pulmonary, i.e. as well as lung cancer. In Latin America and the Caribbean, at least 100 million people are exposed to air pollution above recommended levels on World Health Organization (WHO), where the majority of pollutants are estimated to come from motor vehicles, poor fuel quality, stationary and natural sources(Green & Sanchez, 2012). A report of the United Nations Environment Programme (UNEP) and the WHO about air pollution in Megacities states that the air in Mexico City is the worst of the 15 metropolitan areas with more than 10 million habitants. The city has serious problems for particulate matter, sulfur dioxide, carbon monoxide and ozone pollution, plus moderate problems for lead and nitrogen oxides (WHO & UNEP, 1992). The increase of the air pollutant in Mexico City is associated with the geographical and meteorological characteristics of the valley (Nebot & Mugica, 2014). Certain parts of the metropolitan area, such as Merced, Pedregal, Tlalnepantla, and Xalostoc are exposed to local industrial sources such as factories and power plants (O'Neill, Loomis, Torres Meza, Retama, & Gold, 2002).

In (L. Li et al., 2014) an spatial and temporal relation of Air Pollution Index (API) with meteorological parameters is presented during 2001–2011 in Guangzhou China, where it is determined that temperature, relative humidity, precipitation and wind speed were negatively correlated with API, while diurnal temperature range and atmospheric pressure were positively correlated with API in the annual cycle. Other studies address only the behavior of a single air pollutant compared to selected meteorological parameters, observing a negative correlation between daily total precipitation and the contribution of PM10 to API, aAs well as a positive correlation between mean surface temperature and the contribution of O3 (Buchholz, Junk, Krein, Heinemann, & Hoffmann, 2010). Although more and more studies show temporal correlations between air quality and meteorological parameters, comparing their dependency analysis are contradictory (Maraziotis, Sarotis, Marazioti, & Marazioti, 2008; Pearce, Beringer, Nicholls, Hyndman, & Tapper, 2011). For example, (Maraziotis et al., 2008) stated that temperature was negatively correlated with PM10 using statistical Methods, but these factors were found to be positively correlated in another study using generalized additive models (Pearce et al., 2011). The differences might be caused by trends and seasonal variations due to study place, mainly the distribution of local pollution sources, the presence of distant pollution sources that impact the city, and the meteorological and topographical conditions of the area (Morawska, Vishvakarman, Mengersen, & Thomas, 2002). Such as relationship have been examined in several studies which have used a combination of statistical regression, graphical analysis, fuzzy logic method and cluster analysis (Abdul-wahab, Bakheit, & Al-alawi, 2005). Artificial Neural Networks (ANN) has been one of the most known and used methods to predict air pollutant concentrations (Arhami, Kamali, & Rajabi, 2013), however, it is also known for its lack of precision due the presence of factors such as noise, the high computational cost and a high selection of inputs and model architecture difficulty (Voukantsis et al., 2011). Linear regression is an important analysis tool in statistics and artificial intelligent (Su, Gao, Li, & Tao, 2012), in recent decades has been widely used for forecasting in several applications of pattern recognition and machine learning (Bishop, 2007). Multiple Linear Regression (MLR) is a statistical approach that allows the formulation of simple equations to predict the percentage or degree of association between the independent variable and the dependent variable, without detailing the causes of these relationships (Paschalidou, Kassomenos, & Bartzokas, 2009) and Principal Component Regression (PCR) is a machine learning technique that allows reduce the data dimensions, the high computational cost and eliminate the multicollinearity, when is not clear what is the predictor (Jianfeng, Myodo & Sakazawa, 2013). In this context, and due to the fact that not too many dependent analysis regarding pollution and meteorological parameters in Mexico city are known, a dependent analysis with PCR and MLR is proposed to determine the dependence between air pollutants and meteorological parameters in 4 stations of Mexico city's metropolitan area.

2. MATERIALS AND METHODS

2.1 Theoretical Background

2.1.1 Handling Missing Date

Real world data sets are almost always accompanied by missing data due to different factors such equipment malfunction, communication noise, lack of equipment availability and/or unknown reasons; this missing data on several analysis fields cause inconclusive study results. Different methods have been proposed to handle missing data; two of the most commonly used are interpolation methods and machine learning techniques (D. Li, Deogun, Spaulding, & Shuart, 2005). The popular approaches for spatial interpolation include inverse distance weighting (IDW), kriging, spline interpolation, and interpolation polynomials (Bhattacharjee, Mitra, & Ghosh, 2014).

2.1.1.1 Inverse Distance Weighting (IDW)

The IDW is one of the mostly applied interpolation techniques in the field of environmental science.

$$Z(s_0) = \sum_i^n W_i Z(s_i) \qquad (1)$$

Where $Z(s_0)$ is the intended value to be predicted in terms of a particular S_0 place, n will be taken to account as the number of sample points around the site to be predicted, W_i is the weight assigned to each sample point and $Z(s_i)$ is the value observed on place S_i.

Weights are determined by the equation:

$$W_i = \frac{d_{io}^{-p}}{\sum_{i=1}^{n} d_{io}^{-p}} \qquad (2)$$

Where d_{i0} is the distance between the place of prediction and each sample place; as the distance becomes larger, the weight is reduced by p factor (Cañada, Moreno, & Gonzalez, 2014).

2.1.2 Pearson and Spearman Correlations Coefficients

Common measure of linear association between two variables. It is defined as ratio of the two variables to the product of their respective standard deviation. Tthe Pearson correlation coefficient ranges from -1 to 1, where if the coefficient is greater than 1 indicate positive association, if coefficient is less than 1, indicate negative association and if coefficient is 0 indicate no correlation (Shong, 2010).

Spearman correlation coefficient is a rank based version of the Pearson correlation coefficient. Unlike the Pearson correlation coefficient, Spearman coefficient is used also for the variables that are nonlinearly related, and is robust to the presence of outliers. The Spearman coefficient interpretation is the same as the Pearson correlation coefficient (Restrepo & González, 2007).

2.1.3 Multiple Linear Regression (MLR)

Multiple linear regression (MLR) is a statistical technique that has been used in air pollution analysis to predict the variability and model the mathematics relationship between the dependent variable and independent variable (Wang, 2007). The regression model is given as:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_k \qquad (3)$$

Where $\beta i(i=0,\ldots,n)$ are the parameters generally estimated by least squares, $xi(i=1,\ldots,n)$ are the explanatory variables (predictors), and ε is the error associated with regression. MLR is a simple implementation model based on linear and additive associations of explanatory variables which has been extensively used due to its satisfactory results, nevertheless, in regression equations, the collinearity between independent variables can lead to incorrect identification of the most important predictors (Heo & Kim, 2004; Thompson, Reynolds, Cox, Guttorp, & Sampson, 2001), which is a serious problem which can dramatically influence the effectiveness of a regression model.

2.1.4 Principal Component Regression

Principal Component Analysis (PCA) is a multivariate statistic tools, widely used in reduction from dimensions for air pollution analysis. Its goal is to reduce the number of predictive variables and transform them into new variables, called principal components (PC). These new variables are independent linear combinations of the original data and retain the maximum possible variance of the original set (Sousa, Martins, & Pereira, 2007). The Principal components (PCs) can be expressed as in Eq. (4) below:

$$Z_{ij} = a_{1i}x_{1j} + a_{2i}x_{2j} + \cdots + a_{ni}x_{nj} \qquad (4)$$

Where Z is component score, a is component loading (correlation coefficient), x is the measured value of the variable, i is the component number, j is the sample number and n is the total number of variables (Dominick, Talib, Zain, & Zaharin, 2012). Principal Component Regression (PCR) is methods that combine PCA with linear regression for reduce the number of regressor variables, while retaining as much as possible of the variation present in the data. This reduction is achieved by transforming the data into uncorrelated principal components (PCs) and by discarding the PCs which contribute least to the variance in the original data (Jolliffe, 2002).

Hence, in PCR, all the PCs of the independent variables $x_1 \dots xp$ are transformed to new independent variables. Thus the corresponding linear regression model will be

$$y_i = \beta_0 + \beta_1 PC_{1i} + ... + \beta_r PC_{ri} + \varepsilon_i,\ i = 1, ..., n. \quad (5)$$

Where β are the coefficients of regression, PC the principal components and ε the error. The literature reports Principal Component Regression (PCR) as a known estimator designed to deal with multicollinearity or near-linear dependence of regressors, because it extracts factors from the variables with serious multicollinearity and establishes a regression model on the orthogonal variables.

2.1.5 R-Squared

Statistical measure that shows how close the data are to the fitted regression line, the r squared also called determination coefficient is between 0 and 100%, where 0% indicates that the model not explains the variability of the response data around its mean, and 100% indicates that the model explains all the variability of the response data around its mean (IDRE, n.d.).

2.2 Database

The Automatic Air Quality Monitoring Network (RAMA) with 24 stations for continuously measure of air pollutions, and Meteorology and Sun Radiation Network (RETMET) that providing information of the meteorological parameters in 16 stations, belonging to the Monitoring System of the Cite of Mexico (SIMAT) are considered to form the database for this study. Four (4) stations were selected (Tlalnepantla, Xalostoc, Mercedes y Pedregal) under the method of data completeness and parallel measurements of RAMA and RETMET, with 12 characteristic distributed in meteorological parameters: temperature (TMP), relative humidity (RH), wind direction (WDR) and wind speed (WSP), reported in degrees Celsius (°C), percentage (%), degrees azimuth (°azimuth), meters per second (m/s) respectively, and air pollutants: ozone (O3), nitrogen dioxide (NO2), nitrogen oxides (NOx), nitrogen monoxide (NO), sulfur dioxide (SO2), carbon monoxide (CO), particles smaller than 10 micrometers (PM10), particles smaller than 2.5 micrometers (PM2,5), reported in parts per million (ppm) for the first 5 variables and micrograms per cubic meter (μg/m3) for the next 2. The measurements were collected per hour during the years 2005-2013. A total amount of twelve (12) characteristics with 78888 observations were provided, leading to a 78888x12 dimension matrix.

2.3 Proposed procedure

A comparison analysis with the maximum and minimum values of the meteorological variables and pollutants is performed for remove outlier early and avoid its impact over analysis dependence methods. Further the interpolation approach is applied for IDW to attain dates completeness, since the power parameter in IDW interpolation controls the significance of surrounding points. Next the database normalization is applied between [0, 1], which is considered as the preprocessing according with the Figure 1. Consequently the relationship between meteorological and air pollution features is determined from correlation matrix using the Pearson and Spearman coefficient. Then, two regression multivariate approaches are applied for each day. MLR and PCR were performed for making the dependence analysis and determine which environmental factors may influence the high levels of pollution. The dependents variable is the daily values (O3, NO2, NOx, NO, SO2, CO, PM10, PM2.5,) based on a large number of predictor variables (TMP, RH, WDR, WSP) and those predictors are highly correlated or even collinear. Both methods construct new predictor variables, as linear combinations of the original predictor variables, but they construct those components in different ways. PCR creates components through eigenvalues, to explain the observed variability in the predictor variables, without considering the response variable at all. In this context, the regression models for the feature matrix are obtained, and the relationship of dependency among specific for each pollutants in function of meteorological variables is determined. The least-square method is used for minimizing the error.

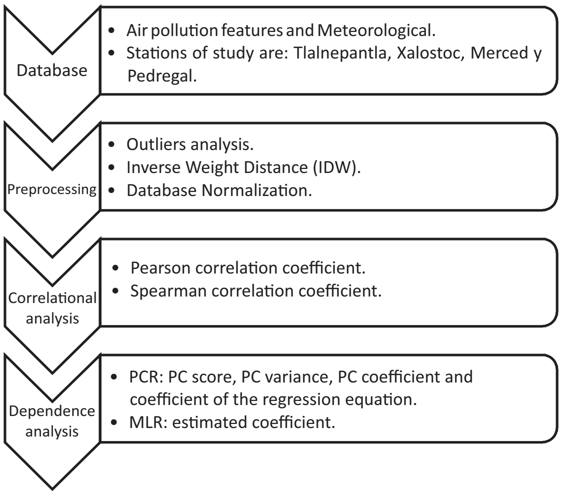

Figure 1. Block diagram of the proposed method

3. RESULTS AND DISCUSSION

During the study period, we found that all air pollution concentrations were within permissible values for metropolitan area Mexico, according to quality norms for atmospheric contaminants from SEMARNAT (Secretaría de Medio Ambiente y Recursos Naturales) (15). Two correlation matrices are obtained for 4 monitoring station. The correlations (R) between O3 and RH, O3 and TMP, CO and WSP, NO and TMP are shown in Table 2, where evidenced positive association among O3 and temperature with high correlation coefficients, also was found that ozone was negatively related with Relative Humidity (RH) showing moderated correlation coefficients of Spearman. In (Rasmussen et al., 2012) a strong correlation between surface temperature and O3 concentrations on multiple time scales is found in US where is attributed to the high population density and engines of vehicles and poor fuel quality as well as in (Sánchez et al., 2008), however for analysis in Mexico City did not find a full registry of the temperature profile in the stations that represent the Metropolitan Area. The wind speed may transport air pollutants (O3, CO, NOx) from distant sources, also high CO and NOx levels occurring when temperatures and winds speed are low (Hosseinibalam & Hejazi, 2012; Pollutants, Ocak, & Turalioglu, 2008), this explains a moderate negative association coefficients between CO and WSP, NO and TMP shown in Table 1.

Table 1. Correlation (R) between O3 and RH, O3 and TMP, CO and WSP, NO and TMP

	$R(O_3$-TMP)	$R(O_3$-RH)	R(CO-WSP)	R(NO-TMP)
Xalostoc	0,75668	-0,53567		-0,40870
Pedregal	0,70761	-0,50553	-040059	
Merced	0,79747	-0,56693	-0,40136	-0,43179
Tlalnepantla	0,79833	-0,54176	-0,41326	

Historical trends of the most significant correlations in Table 1 sampled on the 4 seasons are shown in Figure 2 for month average values of O3 and temperature and in Figure 3 for O3 and RH.

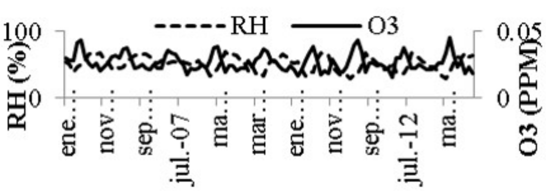

Figure 2. Month average values of O3 and TMP

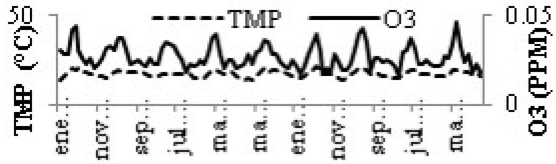

Figure 3. Month average values of O3 and RH

The regression equation with less error and high coefficient of determination were obtained using PCR and are expressed for the pollutants with high and moderated coefficient correlation as follow:

- Pedregal station:

$$CO = 1{,}1775 - 0{,}0014881RH + 0{,}0026825TMP - 0{,}00061557WDR - 0{,}23117WSP$$

R-squared: 0.3624
Root Mean Squared Error: 0.1135

- Merced station:

$$NO = 0{,}15726 - 0{,}00038581RH - 0{,}0046527TMP - 6{,}2445e^{-06}WDR - 0{,}010362WSP$$

Root Mean Squared Error: 0.0398
R-squared: 0.4248

- Tlalnepantla station:

$$O3 = -0{,}009199 - 0{,}000243RH + 0{,}003155TMP - 1{,}3465e^{-05}WDR - 0{,}001072WSP$$

Root Mean Squared Error: 0.0178
R-squared: 0.8237

High dependence of O3 respect to RH, TMP, WDR and WSP is determined in 82%, also there is a moderated dependence of NO and CO regarding meteorological parameters as a 42% and 36% respectively. These equations reveal that O3 and CO increases with increasing TMP and decreasing of RH, WDR and WSP, while NO increase with decreasing TMP and others parameters.

The literature report in (Hosseinibalam & Hejazi, 2012) a dependence of O3 respect to meteorological parameters of 59%, negative correlation with wind direction a strong correlation whit SO2 and O3 in range from 0.51 to 0.89. MLR models of CO, NOx and O3 with meteorological parameters and pollutants concentrations previous shows determination coefficients of 48%, 28% and 75% respectively (Pollutants et al., 2008). There are limited studies of dependence analysis of pollutants and meteorological parameters described in this paper in metropolitan area of Mexico City with which we can directly compare, because the environmental conditions are different according to location of the city.

4. CONCLUSIONS

A dependence analysis is used to determine the importance of meteorological parameters respect to air pollutants, where is found that ozone is highly correlated with temperature and relative humidity in Tlalnepantla station, evidencing increasing in ozone concentration when temperature is high and relative humidity is low.

Also this study shows the usefulness of PCR and MLR as tool for analyzing a large multivariate data set. The dependency analysis revealed that the PCR model presented better performance in comparison to the MLR model, since their results showed lower mean square error and more coefficient of determination, due to use of PCs that allow eliminating of collinearity problems in MLR and reduction of the number of predictors.

The results from this study would be used in providing the necessary information to the general public, to protect their health and take necessary precautionary measures.

ACKNOWLEDGMENTS

To The Monitoring System of the Cite of Mexico (SIMAT) for providing information of the meteorological parameters and pollutants variables to form the database for this study.

REFERENCES

Abdul-wahab, S. A., Bakheit, C. S., & Al-alawi, S. M. (2005). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software*, (20), 1263–1271. doi:10.1016/j.envsoft.2004.09.001

Arhami, M., Kamali, N., & Rajabi, M. M. (2013). Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. *Environmental Science and Pollution Research International*, 20(7), 4777–4789. doi: 10.1007/s11356-012-1451-6

Bhaskar, B. V. (2010). Atmospheric Particulate Pollutants and their Relationship with Meteorology in Ahmedabad. Aerosol and Air Quality Research. *Aerosol and Air Quality Research*, (10), 301–315. doi:10.4209/aaqr.2009.10.0069

Bhattacharjee, S., Mitra, P., & Ghosh, S. K. (2014). Spatial interpolation to predict missing attributes in gis using semantic kriging. *IEEE Transactions on Geoscience and Remote Sensing*, 52(8), 4771–4780.

Bishop, C. M. (2007). Pattern Recognition and Machine Learning. Retrieved from http://www.springer.com/computer/image+processing/book/978-0-387-31073-2

Buchholz, S., Junk, J., Krein, A., Heinemann, G., & Hoffmann, L. (2010). Air pollution characteristics associated with mesoscale atmospheric patterns in northwest continental Europe. *Atmospheric Environment*. 44, 5183–5190. doi:10.1016/j.atmosenv.2010.08.053

Cañada, M., Moreno, A., & González, H. (2014). Modelado de la calidad del aire urbano: un experimento metodológico con técnicas de interpolación espacial. *Boletín de La Asociación de Geógrafos Españoles*, 65, 317–342.

Dominick, D., Talib, M., Zain, S. M., & Zaharin, A. (2012). Spatial assessment of air quality patterns in Malaysia using multivariate analysis. *Atmospheric environment*, (60), 172–181. doi:10.1016/j.atmosenv.2012.06.021

Clean Air Institute. (2012) Air Quality in Latin America: An Overview. (1) Retrieved from http://www.cleanairinstitute.org/calidaddelaireamericalatina/cai-report-english.pdf

Heo, J., & Kim, D. (2004). A new method of ozone forecasting using fuzzy expert and neural network systems. *Science of the Total Environment*, 325, 221–237. doi:10.1016/j.scitotenv.2003.11.009

Hosseinibalam, F., & Hejazi, A. (2012). Influence of Meteorological Parameters on Air Pollution in Isfahan. *3rd international conference on biology, Environment and chemistry,* (46). doi:10.7763/IPCBEE.

Institute for Digital Research and Education -IDRE. (2014) *FAQ: What are pseudo R-squareds?* California: UCLA. Retrieved from http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm

Jianfeng, X., Myodo, E., & Sakazawa, S. (2013). Motion synthesis for affective agents using piecewise principal component regression. *IEEE International Conference on Multimedia and Expo (ICME)*, California: USA.

Jolliffe, I. T. (2002). *Principal Component Analysis.* New York, USA: Springer-Verla. doi: 10.1007/b98835

Li, D., Deogun, J., Spaulding, W., & Shuart, B. (2005). Dealing with Missing Data : Algorithms Based on Fuzzy Set and Rough Set Theories. *Springer-Verlag Berlin Heidelberg* 2005, 4, 37–57.

Li, L., Qian, J., Ou, C.-Q., Zhou, Y.-X., Guo, C., & Guo, Y. (2014). Spatial and temporal analysis of Air Pollution Index and its timescale-dependent relationship with meteorological factors in Guangzhou, China, 2001-2011. *Environmental Pollution,* 190, 75–81. doi:10.1016/j.envpol.2014.03.020

Maraziotis, E., Sarotis, L., Marazioti, C., & Marazioti, P. (2008). Statistical Analysis of Inhalable (Pm 10) and Fine Particles (Pm 2. 5) Concentrations In Urban Region Of Patras, Greece. *Global nest journal*, 10(2), 123–131.

Morawska, L., Vishvakarman, D., Mengersen, K., & Thomas, S. (2002). Spatial variation of airborne pollutant concentrations in Brisbane, Australia and its potential impact on population exposure assessment Centre for Medical, Health and Environmental Physics, Queensland University of Technology, PO Box 2434 Brisbane,. *Atmospheric Environment*, (36), 3545–3555.

Nebot, À., & Mugica, F. (2014). Simulation and Modeling Methodologies, Technologies and Applications. *Springer International Publishing Switzerland*, (256), 239–252. doi:10.1007/978-3-319-03581-9

O'Neill, M. S., Loomis, D., Torres Meza, V., Retama, A., & Gold, D. (2002). Estimating particle exposure in the Mexico City metropolitan area. *Journal of Exposure Analysis and Environmental Epidemiology*, 12(2), 145–56. doi:10.1038/sj.jea.7500212

Paschalidou, A. K., Kassomenos, P. A., & Bartzokas, A. (2009). A comparative study on various statistical techniques predicting ozone concentrations: implications to environmental management. *Environmental Monitoring and Assessment,* 148(1-4), 277–89. doi:10.1007/s10661-008-0158-0

Pearce, J. L., Beringer, J., Nicholls, N., Hyndman, R. J., & Tapper, N. J. (2011). Quantifying the in fl uence of local meteorology on air quality using generalized additive models. Atmospheric *Environment. Atmospheric Environment,* 45(6), 1328–1336. oi:10.1016/j.atmosenv.2010.11.051

Pollutants, R., Ocak, S., & Turalioglu, F. S. (2008). Effect of Meteorology on the Atmospheric Concentrations of Traffic. *Enviromental application & science*, 3(5), 325–335.

Rasmussen, D. J., Fiore, A. M., Naik, V., Horowitz, L. W., Mcginnis, S. J., & Schultz, M. G. (2012). Surface ozone-temperature relationships in the eastern US : A monthly climatology for evaluating chemistry-climate models. *Atmospheric Environment*, (47), 142–153. doi:10.1016/j.atmosenv.2011.11.021

Restrepo, L., & González, J. (2007). De Pearson a Spearman. *Revista Colombiana de Ciencias Pecuarias*, 20(2), 183–192.

Sánchez, J. A., Javier, F., Ayala, G., Química, F. De, Fisicoquímica, D. De, Nacional, U., … México, D. F. (2008). Recent Trend in Ozone Levels in the Metropolitan zone of Mexico City. *Sociedad Química de México*, 52(4), 256–262.

Shong, N. (2010). Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data. *University of Pittsburgh*.

Sousa, S. I. V, Martins, F. G., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, (22). doi:10.1016/j.envsoft.2005.12.002

Su, Y., Gao, X., Li, X., & Tao, D. (2012). Multivariate multilinear regression. IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics : *A Publication of the IEEE Systems, Man, and Cybernetics Society,* 42(6), 1560–73. doi:10.1109/TSMCB.2012.2195171

Thompson, M. Lou, Reynolds, J., Cox, L. H., Guttorp, P., & Sampson, P. D. (2001). A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmospheric environment*, (35). 617-630. doi:10.1016/S1352-2310(00)00261-2

Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., & Kolehmainen, M. (2011). Intercomparison of air quality data using principal component analysis, and forecasting of PM10 and PM2.5 concentrations using artificial neural networks, in Thessaloniki and Helsinki. *The Science of the Total Environment,* 409(7), 1266–76. doi:10.1016/j.scitotenv.2010.12.039