Developing a methodology to predict PM₁₀ concentrations in urban areas using Generalized Linear Models

J.M. Garcia^{1, *}, F. Teodoro^{1,2}, R. Cerdeira¹, L. M. R. Coelho¹, Prashant Kumar^{3, 4}, M.G. Carvalho⁵

¹ *Escola Superior de Tecnologia de Setúbal, Instituto Politécnico, Setúbal, Portugal*

² *CEMAT, Instituto Superior Técnico, Portugal*

³ *Department of Civil and Environmental Engineering, Faculty of Engineering and Physical Sciences (FEPS), University of Surrey Guildford GU2 7XH, Surrey, United Kingdom*

⁴ *Environmental Flow Research Centre, FEPS, University of Surrey Guildford GU2 7XH, Surrey, United Kingdom*

⁵ *Instituto Superior Técnico, Portugal*

**Corresponding author: joao.garcia@estsetubal.ips.pt*

Abstract

A methodology to predict PM₁₀ concentrations in urban outdoor environments is developed based on the Generalized Linear Models (GLM). The methodology is based on the relationship developed between atmospheric concentrations of air pollutants (i.e. CO, NO₂, NO_x, VOCs, SO₂) and meteorological variables (i.e. ambient temperature, relative humidity and wind speed) for a city (Barreiro) of Portugal. The model uses air pollution and meteorological data from the Portuguese monitoring air quality station networks. The developed GLM model considers PM₁₀ concentrations as a dependent variable, and both the gaseous pollutants and meteorological variables as explanatory independent variables. A logarithmic link function was considered with a Poisson probability distribution. Particular attention was given to cases with air temperatures both below and above 25 °C. The best

performance for modelled results against the measured data was achieved for model with values of air temperature above 25 °C compared with model considering all range of air temperatures and with model considering only temperature below 25 °C. The model was also tested with similar data from another Portuguese city, Oporto, and results found to behave similarly. It is concluded that this model and the methodology could be adopted for other cities to predict PM₁₀ concentrations when this data is not available by measurements from air quality monitoring stations or other acquisition means.

Keywords: Outdoor air quality; PM10; Generalized linear methods; SPSS; Methodology

1. Introduction

Concern about air quality has grown due to increase in respiratory problems, especially in children, elderly and people with respiratory diseases, related with air pollution [1]. Also, the economic and social development has led to the increase of urban traffic and industry that emit a wide variety of pollutants, namely carbon monoxide (CO), nitrogen oxides (NO_x), volatile organic carbons (VOCs), particulate matter (PM) and sulphur dioxide (SO₂) [2]. In the last decades, air pollution related to road traffic and associated health problems have increased [3]. It is now accepted that air pollutants can trigger allergies and respiratory problems, particularly in children [4-6]. In this context, PM concentration in urban environments, especially in street canyons is a major issue [7]. PM contains microscopic components, and some of its fraction such as nanoparticles (<100 nm in diameter) are so small that they can penetrate deep into the lungs and cause serious health problems [8].

In fact, particular attention was dedicated to both PM₁₀ and PM_{2.5} [3, 9] and more recently to nanoparticles [10, 11] among the range of airborne pollutants. Unfortunately, the spatial distribution of PM₁₀ concentrations is not always easy to understand since there are no urban

air quality stations in some of the urban areas. Therefore a need remains to measure them at a greater number of locations or model them using detailed modelling systems such as CFD [12] or numerical models [13]. The monitoring methods include acquisition of PM₁₀ concentrations using the scientific instruments while the numerical simulation of pollutants dispersion using computational tools, physical modelling using wind tunnel experiments [14], or through the statistical methods [15, 16]. Statistical models based on multiple regression analysis and classification and regression trees analysis have been developed and applied in the forecasting of average daily concentrations for PM and ozone levels [17-18]. In studies based on the estimation of PM concentrations using satellite remote sensing techniques, some statistical tools have also been widely used. In this field, the Aerosol Optical Thickness (AOT) is the satellite derived parameter most commonly used as the basis for PM estimation using statistics techniques [19]. Several methods have been used to correlate this satellite remote sensing (AOT) with the PM concentrations based on ground measurements from air quality stations. These include linear relations [20], statistical and chemical transport models [21], multiple regression analysis [22] and neural networks [23].

Also, statistical methods were developed and used in the past to determine relationships between air pollution concentrations and meteorological parameters. Among these, methods such as multiple linear regression analysis [24], nonlinear multiple regressions [25], artificial neural networks [26, 27], and generalized additive models and fuzzy-logic-based models [28] were used. These models were tested in a perspective of daily or long-term forecasting and focused in the perspective of the exploring relationship between O₃ and PM. However, in some situations it would be useful to know (or at least to estimate) unknown concentrations of PM based on the values of other air pollutants and on meteorological variables. This could be carried out based on known air concentrations from other air pollutants and meteorological parameters using data from monitoring sites or from specific acquisition data equipment. This

is particularly useful in urban environments, where there is no data from monitoring sites and when it is important to know outdoor PM concentrations, particularly in high traffic urban areas.

A well-known documented and tested tool like General Linear Models (GLM) [29] is used to develop a methodology to estimate outdoor PM₁₀ concentrations based on known values of other air pollutant concentrations from the same site. We have therefore used this method on the hourly data of air pollutants (CO, NO_x, NO₂, O₃, SO₂ and PM₁₀) that are hourly monitored by several stations, to build a model that is subsequently used to predict PM₁₀ concentrations at the same site. To build this model, it was taken into consideration that atmospheric PM are very different in their constitution, origin and governing mechanisms. Generically are grouped under the designation of particle matter (PM), a group of air pollutants considerably extended and different, and that may have their origin in sources as diverse as automobiles, steel mills, power stations, heating systems, factories cement, volcanoes, deserts and oceans. In general terms, this is common to consider particulate matter as the definition from NIST [30] as “any condensed-phase tri-dimensional discontinuity in a dispersed system may generally be considered a particle”. In terms of classification, PM are usually classified based on two distinct criteria. They can be classified by their mechanism of formation, and in this case they are called primary particles or secondary particles, or can be classified by their physical size. According to the criterion of the formation mechanism, the primary particles are those that are directly emitted as particles, whereas secondary particles are those which are formed from gaseous precursors in the atmosphere through a mechanism of formation gas-to-particle conversion. PM are also often classified by their physical size. Their characteristic dimensions vary from a range of few nanometres (nm) up to dozens of micrometres (µm) in diameter. The particles larger than 2.5 µm (coarse particles) are produced by mechanically breaking of the larger solid particles. This PM can include dust

originating from agricultural processes transported by wind, dust originating from the bare soil, dust originating from unpaved roads or dust from other processes such as mining or stone quarrying. Smaller particles (fine particles) are mainly formed from gases. The smaller ones (less than 0.1 microns) are formed by nucleation, i.e. the condensation of substances formed by high temperature steaming or by chemical reactions in the atmosphere [31]. The particles below 1 μm may be formed by condensation of metal or condensation of organic compounds that are evaporated in combustion processes, or they can also be produced by condensation reactions resulting from atmospheric gases. The particles produced by these reactions of gases in the atmosphere are called secondary particles. Sulphate and nitrate particles are usually the predominant component of these fine particles. Other important aspect in the definition of the characteristics of PM concentrations in the atmosphere is the meteorological variables such as wind speed and direction, atmospheric temperature, precipitation and atmospheric boundary layer height. Higher concentrations of particle concentrations are often registered under weather conditions with atmospheric stability, especially in situations of inversion with low wind speeds. Also chemical and physical processes of particle formation are regulated largely by meteorological variables [32]. Chaloulakou [33] found that $\text{PM}_{2.5}$ and PM_{10} concentrations were highly correlated with carbon monoxide, black carbon and nitrogen oxides and inversely correlated with local wind speed. Also, solar radiation and temperature have major importance in the mechanisms of formation of secondary particles. Results from Anderson et al. [34] indicate that 25 °C is a key air temperature value from which the occurrence of summertime air pollution episodes are promoted.

The purpose of this paper is to study the relationship between atmospheric pollutants and develop a methodology that can be used to estimate PM_{10} concentrations in the city of Barreiro in Portugal, by using an Generalised Linear Model (GLM) on the data of CO, NOx,

VOCs, and SO₂ available from air quality stations. The predict values are compared with real measured values of PM₁₀ outside air concentrations in the city. Despite the fact that the study uses a localised case study, the methodology proposed and the model developed allow a broad understanding of the interrelationships between the gaseous pollutants and PM₁₀ in urban environments. Thus the work contributes to the basis of development of more complex model in future.

2 Methodology

2.1 Location

Barreiro is a medium-size city located 40 km south of Lisbon, Portugal, with 34 km² area and about 80000 inhabitants, with industry near the centre and typical suburbs important car traffic fluxes. The city is almost flat, with highest point at approximately 10 meters above sea level. The weather is temperate, with no severe seasons. The main industrial activity in Barreiro city is developed in the industrial area. A natural gas power plant and some chemical industries are the main industrial sources. The most important pollutants released from these industrial sources are NO_x, SO₂ and PM.

2.2 Meteorological and Air Quality data

Meteorological data was obtained from the Instituto Português do Mar e Atmosfera (IPMA). The prevailing wind direction is NW (frequency 35.1%). The highest wind speed registered corresponds to the prevailing direction NW (14.1 km/h). The NW wind is particularly frequent in the summer months (June, July and August), with a maximum occurring in August (58.5%) and a minimum frequency recorded in December (15.6%). The average wind speed is relatively constant throughout the year. Air Quality data from pollutants concentrations (CO, NO_x, NO, NO₂, O₃, SO₂ and PM₁₀) are hourly monitored by

seven air quality stations that are managed by the Portuguese government. Data from September 2003 to December 2005 was statistically treated, according to the pollutant in question. A twenty-four hour mean was calculated for NO_x, NO, NO₂, SO₂ and PM₁₀ and 8 hours mean to CO and O₃. Daily averages of each pollutant were related with each other and with meteorological data.

2.3 The GLM methodology

A GLM was used to building a methodology to estimate PM outside concentrations based on known values of other outside air pollutant concentrations [29]. GLM are based on the assumption that there are K independent values $Y_1, \dots, Y_K$, from a variable of interest or response variable (effect) that follows an exponential family distribution with expected value $E(Y_i) = \mu_i$ [35]. Considering K vectors $x_i = (1 \ x_{i1} \ x_{i2} \ \dots \ x_{ip})^t$, $i=1, \dots, K$, containing the values of p explanatory variables, independent or covariates (variables candidate to "causes"). Considered also a link differentiable function g, such that:

$$g(\mu_i) = x_i^t \beta, \ i = 1, \dots, K \quad (1)$$

Where $(\beta = \beta_1 \ \beta_2 \ \dots \ \beta_p)$ are the values of parameters to be estimated. Thus, if we consider for the function g the identity function we have:

$$g(\mu_i) = \mu_i \quad (2)$$

then

$$\mu_i = E(Y_i) = x_i^t \beta \quad (3)$$

The resulting model is the Gaussian linear regression model. If alternatively, consider the function g as a logarithmic function and Y_i has a Poisson distribution, then the model will

result in a Poisson regression model and each term β_i is the effect of variable X_i in $g(\mu_i)$. Each β_i represents the “effect” of variable X_i in the function $g(\mu_i)$.

In this case, the objective was to estimate PM_{10} concentration values based on other variables, such as air pollutant concentration in $\mu g/m^3$ (i.e. CO, NO₂, NO_x, O₃ and SO₂) and meteorological variables such as air temperature (Temp, °C), relative humidity (RH, %) and wind velocity (WV, m/s). The statistical analysis performed earlier on individual variables showed that the wind direction was not related to PM_{10} concentrations (Pearson correlation coefficient $r = 0.01$), since for the number of data and according with a performed t-student distribution, there is a significant correlation (99.9% probability, 700 cases) for $r > 0.114$, so it was decided not to include wind direction as a variable in the interest of not overloading the modelling calculations. The general model parameter used in GLM models are resumed in Table 1. Statistical Package software for Social Sciences (SPSS 10.0) for windows was used to build and analyse the model.

3 Results and discussion

3.1 Analysis of Models representativeness

GLM models were used to investigate the complex relationships between the concentration of 5 air pollutant concentrations, meteorological and PM_{10} concentration levels in the Barreiro city.

$$\ln[PM_{10}] = \alpha + \beta_1 var_1 + \beta_2 var_2 + \beta_3 var_3 + \dots + \beta_i var_i \quad (4)$$

Based on these results, estimations of PM_{10} can be expressed as the product of the exponential terms:

$$[PM_{10}] = e^{(\alpha + \beta_1 var_1 + \beta_2 var_2 + \beta_3 var_3 + \dots + \beta_i var_i)} \quad (5)$$

The first term contains the regression intercept and the other terms contain variables, originated from GLM model as explained above. This methodology as applied to three tested models A, B, and C. The three models presented in table. 3 differ only in data considered. In model A, we considered the total number of observations recorded. In model B we considered the observations recorded in days with maximum air temperature of day above 25 °C (maximum). In model C we consider only observations with maximum air temperature of day less or equal to 25 °C. These considerations are shortly resumed in Table 2. The β coefficients obtained with methodology implemented for the three models are:

Model A:

$$\ln (\text{PM}_{10}) = 2.425652 - 0.000357 [\text{CO}] + 0.001821 [\text{O}_3] - 0.000364 [\text{SO}_2] + 0.028348 [\text{NO}_2] + 0.000093 [\text{NO}_x] + 0.016820 [\text{Temp}] - 0.000490 [\text{RH}] + 0.002821 [\text{WV}] \quad (6)$$

Model B:

$$\ln (\text{PM}_{10}) = 1.957605 - 0.000204 \times [\text{CO}] + 0.001931 [\text{O}_3] - 0.003097 [\text{SO}_2] + 0.024388 [\text{NO}_2] + 0.000309 [\text{NO}_x] + 0.043356 [\text{Temp}] - 0.000960 [\text{RH}] + 0.003548 [\text{WV}] \quad (7)$$

Model C:

$$\ln (\text{PM}_{10}) = 2.419685 - 0.000219 [\text{CO}] + 0.000863 [\text{O}_3] + 0.002149 [\text{SO}_2] + 0.019767 [\text{NO}_2] + 0.001449 [\text{NO}_x] + 0.021912 [\text{Temp}] + 0.000153 [\text{RH}] + 0.003008 [\text{WV}] \quad (8)$$

Figs.1, 2 and 3 shows the scattered plot with measured versus the predicted PM_{10} concentrations by the three models (A, B and C). The values of measured PM_{10} concentrations ($\mu\text{g}/\text{m}^3$) are identified from measured data and the PM_{10} concentrations values predicted ($\mu\text{g}/\text{m}^3$) by the three models.

Knowing that the correlation coefficient $R^2 (\times 100)$ gives the percentage of variability explained by the model [i.e. $R^2 = \text{sum of squared explained (SSE)} / \text{total sum of squares (SST)}$], the calculation of R^2 results that the model B gives the best R^2 values from the three

selected models (Fig. 1) ($R^2=0.65$). Therefore model A (Fig. 2) ($R^2=0.39$) and model C (Fig. 3) ($R^2=0.15$) have a weak explanatory capacity compared with the model B.

Table 3 shows a resume of the statistical model results performance for the three models (A, B and C). First column of Table 3 presents the statistics tests, most often used in generalized linear models, representing measures of dispersion (generalized and / or corrected), which permit to test the quality of models. Values from Table 3 confirm that model B is the one with the best performance shown by statistical tests. These statistical tests are obtained using all the deviations obtained between the estimated and recorded residuals for each observation. Considering the Akaike Information Criterion, the objective is to minimize AIC. From the three models, model B is the one with lowest AIC, which means that evidence for the model B is the best. The same can be concluded when analysing AICC (Akaike Information Criterion corrected by minimizing the number of model parameters).

When comparing with the quantile of a chi-square distribution with $n-p$ degrees of freedom (n -number of observations, p -number of estimated parameters), it is possible to measure the suitability of models. Results of deviance show that the three are suitable. Another measure of goodness of fit is the Pearson chi-square test, which leads to the same conclusions when compared with the quantile of the chi-square distribution with $n-p$ degrees of freedom. Table 4 shows the likelihood ratio chi-square test, which compares each model with the null model. Regardless of model B is considered the best, each model individually, has a greater explanation of the dependent variable using some of the explanatory than any other model without explanatory variables, where NV is the number of variables and sig is the p-value associated.

We observe from Fig. 5 that the residues associated with the model B are those with a more adequate to the expected aspect: dispersed values without standard and with homogeneous

variability (white noise). Either model A (Fig. 4) or model C (Fig. 6), the residues appear to have a functional relationship and not look like white noise. The variability is also not constant as would be expected. Some diagnostic tests have been made (independence, heteroscedescidade, normality) and models A and C are rejected. Only after validation of residuals has behavior of white noise with normal distribution is that it can and should consider the inference using models.

One last step for evaluating the quality of the model is to perform simple tests using the Wald Chi-Square statistic (Table 5). This test serves to verify that some independent variable (explanatory) in particular, contributes significantly to the explanation of the response variable, testing in the form $H_0: \beta_i = 0$ versus $H_1: \beta_i \neq 0$. If we reject the null hypothesis, we have evidence that the variable is a good explanatory variable. From Table 5, the p-values (sig in Table 5) associated with the nullity test of each parameter, the sig values are zero in majority, indicating rejection of the null hypothesis, showing that the associated variables should be considered. Note that the model HR variable B is statistically significant (p-value = 0.021), not being in model A (p-value = 0.080) and model C (p-value = 0.720). Remember that we rejects the null hypothesis if p-value < significance level (the level of significance is usually 5%). It is concluded that the relative humidity is important when considering the higher temperatures.

3.2 Physical representativeness of beta (β) coefficients

The results in Section 3.1 suggest that the model B show the best performance in estimating PM_{10} concentrations. It is however necessary to discuss and understand the physical representativeness of β coefficients integrating the three models, especially for model B. It is known that the β coefficients show the different weights of the variables under study (CO, NO₂, NO_x, O₃, SO₂, Temp, RH and WV) in PM_{10} concentration. Thus, if the value of β is greater than 0, the concentration of PM_{10} increases with the increase in value of

this variable and vice-versa. It is visible from the analysis of the results obtained for the coefficients β in model B (Table 5) that PM_{10} usually increase with O_3 , NO_2 , NO_x , wind speed (WV) and ambient temperature (Temp). One of the important variables for all these models is the Temp, particularly for Model B. Another important variable responsible for the increase of PM_{10} concentration is the NO_2 ; it is also visible that PM_{10} concentrations also increases (but with minor importance) with WV and O_3 . This is explained by the importance of air temperature and solar radiation combined with gaseous pollutants (O_3 , NO_2) in the formation of secondary particles. The increase in PM_{10} concentration with the increasing WV could be explained by the fact that a major wind speed would promote the physical-chemical reaction for the formation of secondary particles by increasing the mixing ratio of precursor gases. It is also found that for model B, PM_{10} concentration decline with increasing SO_2 and with increasing relative humidity (RH), but this cannot be considered significant or representative considering the low values for β . It can be expected that the PM_{10} concentration increased with SO_2 , however it is known that for forming secondary particles originated by SO_2 , the presence of other components such as ammonia is often required. In fact, it is noted that in model C (Temp $<25^\circ C$) PM_{10} concentrations increase slightly with SO_2 , which could be explained by the existence of ammonia in the atmosphere during the measurement days, due to thermal inversion conditions that promote the permanence of ammonia from industrial emissions. Since there is a relationship between RH and Temp, there is an inverse indirect effect on PM_{10} concentrations; this is because the increase of Temp generally results in lowering of the RH.

It is also important to analyze the correlation between the different variables tested. For this purpose before the development of three models, the Pearson correlation coefficients were calculated for the eight studied variables (covariates). Table 6 shows these values for Pearson correlation coefficient between the considered variables. Analyzing Table 6 and the results of

the correlations, the importance of the correlation between some of these variables is visible, including an obvious high correlation between NO_x and NO₂ ($r = 0.79$), between NO_x and CO ($r = 0.69$), and between Temp and O₃ ($r = 0.52$). There was an obvious negative correlation between Temp and RH ($r = -0.43$). The previous analysis of correlations between these variables was important for subsequent development of the model.

In the case of temporal time series, an important aspect to consider is the level of autocorrelation between the variables. Therefore, a previous study of Autocorrelation Function (ACF) was performed before the model development. This test was performed to make an initial estimation and to decide the type of model to be used. The ACF values for the variables are shown in Fig.7. As expected, significant ACF were evident for all variables and weekly effects are visible for some pollutants (NO, NO_x). As expected, Temp and RH showed a stable time behavior with a strong ACF for earlier days (Fig.7a e 7b). The WV values were found to be auto correlated with a 3 days lag period (Fig.7c). There is clearly a weekly behavior (7 days lag) on the levels of NO₂ and NO_x, which may be due to the type of road traffic profiles over the week.

4. Model implementation to Oporto data

Knowing that model B (T_{max} of air $>25^{\circ}\text{C}$) is the model that best predicts PM₁₀ concentrations based in measured concentrations of CO, NO₂, NO_x, O₃, SO₂, T, RH, WV, model B was tested with the data from a different Portuguese city (Oporto). Data from the Portuguese air quality network (Campanhã air quality station in Oporto), managed by CCDR, was used, considering values from January 2011 to December 2011. Also meteorological data from FEUP meteorological acquisition station, at the same period was used. Results showing PM₁₀ concentrations predicted by the model and PM₁₀ concentrations measured are showed

in Fig.8. This figure suggests that the model predicts the PM_{10} concentrations with reasonable accuracy ($R^2 = 0.47$) when used on an independent data set for the Oporto city.

5. Summary and conclusions

With the objective of improving the model accuracy, two sub models with the criteria of maximum ambient temperature above 25 °C (model B) and below 25 °C (model C) were developed. Results show that the PM_{10} predictions by model A (all values) are poor ($R^2=0.39$).

Using the information on secondary PM formation (Section 1), it is expected that PM concentration could be correlated with gaseous pollutants mainly NO_x , SO_2 , VOC and the ambient temperature. For this specific case we have no VOC data and therefore VOC concentrations were not used in the analysis. For O_3 , knowing that this pollutant is also a result of photochemical oxidation and it is expected that O_3 could also be correlate with secondary PM, even if it is not a precursor for secondary particles. Comparisons of the three models show that best performance results are achieved for model B that considers only data with values of ambient T_{max} above 25 °C ($R^2=0.65$). These findings are in accordance with the results from Anderson et al. (2001), concluding that “simultaneous occurrence of daily maximum temperatures above 25°C and low wind speed conditions favour the occurrence of summertime air pollution episodes”. When comparing model A (all data) and model B (T_{max} air >25 °C) and model C (T_{max} air <25 °C) the best fit prediction is achieved from model B, showing the importance of higher air temperature in the formation of the secondary particles in air. This can also be concluded by observation of the highest coefficient values (β) in temperature variable (Temp) observed.

Results show a good accuracy for situations where solar radiation is an important factor, which is reflected in the outside ambient temperature parameter ($T_{max} > 25$ °C). These models

are an important tool in situations where there are no measurements of PM concentrations, but it is possible to achieve data from other gaseous air pollutants (e.g. CO, NO₂, NO_x, O₃, SO₂) and also meteorological parameters (e.g. T, RH and WV).

6. Acknowledgments

The authors wish to acknowledge Comissão de Coordenação e Desenvolvimento Regional de Lisboa e Vale do Tejo (CCDR-LVT) and Instituto Português do Mar e Atmosfera (IPMA) by the information provided.

7. References

- [1] João Garcia, Rita Cerdeira, Luís Coelho, Prashant Kumar, and Maria da Graça Carvalho, (2014), Influence of Pedestrian Trajectories on School Children Exposure to PM₁₀, Journal of Nanomaterials, Volume 2014, Article ID 505649, 9 pages
- [2] Kumar, P., Jain, S., Gurjar, B.R., Sharma, P., Khare, M., Morawska, L., Britter, R., (2013). Can a “Blue Sky” return to Indian megacities? Atmospheric Environment 71, 198-201.
- [3] WHO, (2004). World Health Organization, “Outdoor air pollution: assessing the environment burden of disease at national and local levels”.
- [4] E.E.A., (2005), “Environment and Health”. EEA report, Denmark – Copenhagen , No 10/2005
- [5] EPA (2002). “Child-specific exposure factors handbook”. U.S. Environmental Protection Agency
- [6] Pekkanen J, Remes ST, Husman T, Lindberg M, Kajosaari M, Koivikko A, Soininen L, (1997). “Prevalence of asthma symptoms in video and written questionnaires among children in four regions of Finland”, Eur Respir J; 10 : 1787-1794.

- [7] Kumar, P., Fennell, P., Langley, D., Britter, R., (2008). Pseudo-simultaneous measurements for vertical variation of coarse, fine and ultra fine particles in an urban street canyon. *Atmospheric Environment* 42, 4304–4319.
- [8] Heal, M.R., Kumar, P., Harrison, R.M., (2012). Particles, Air Quality, Policy and Health. *Chemical Society Reviews* 41, 6606-6630.
- [9] EEA, (2011), European Environment Agency “Air quality in Europe — 2011 report”, European Environment Agency.
- [10] Kumar, P., Ketzel M., Vardoulakis S., Pirjola L., Britter R., (2011), “Dynamics and dispersion modelling of nanoparticles from road traffic in the urban atmospheric environment – a review”, *Journal of Aerosol Science*, 42, pp.580-603.
- [11] Kumar, P., Morawska, L., Birmili, W., Paasonen, P., Hu, M., Kulmala, M., Harrison, R.M., Norford, L., Britter, R., (2014). Ultrafine particles in cities. *Environment International* 66, 1-10
- [12] Garcia, J.; Cerdeira, R.; Tavares, N.; Coelho, L.M.R.; Kumar, P.; Carvalho, M.G., (2013). Influence of virtual changes in building configurations of a real street canyon on the dispersion of PM₁₀
- [13] Holmes, L. Morawska, (2006). A review of dispersion modelling and its application to the dispersion of particles: An overview of different dispersion models available. *Atmospheric Environment*, 40: 5902-5928, 2006.
- [14] Carpentieri, M., Kumar, P., Robins, A., (2011). An overview of experimental results and dispersion modelling of nanoparticles in the wake of moving vehicles. *Environmental Pollution* 159, 685-693.
- [15] Gilbert, R. O. (1987). “Statistical Methods for Environmental Pollution Monitoring”. John Wiley & Sons. ISBN 0-471-28878-0

- [16] Goel, A., Kumar, P., (2014). A review of fundamental drivers governing the emissions, dispersion and exposure to vehicle-emitted nanoparticles at signalised traffic intersections. *Atmospheric Environment* 97, 316-331.
- [17] Neto, J., Torres P., Ferreira F. Boavida F., (2009). "Lisbon air quality forecast using statistical methods". *Int. J of Environmental Pollution*. Vol.39, N3/4, pp.333-339
- [18] Demuzere M., van Lipzig N. P. M. (2010). "A new method to estimate air-quality levels using a synoptic-regression approach. Part I: Present-day O₃ and PM₁₀ analysis". *Atmospheric Environment* 44, 1341–1355
- [19] Themistocleous, K., Hadjimitsis, D.G., Retalis, A., Chrysoulakis, N., (2012). "The development of air quality indices through image-retrieved AOT and PM₁₀ measurements in Limassol Cyprus". *Proceedings of SPIE Remote Sensing 2012: Remote Sensing of Clouds and the Atmosphere XVII and Lidar Technologies, Techniques, and Measurements for Atmospheric Remote Sensing VIII*. Proc.. of SPIE, vol. 8534, p. 85340B. <http://dx.doi.org/10.1117/12.974701>.
- [20] Yap, X.Q., Hashim, M., (2012). "A robust calibration approach for PM₁₀ prediction from MODIS aerosol optical depth". *Atmospheric Chemistry and Physics Discussions* 12, 31483e31505. <http://dx.doi.org/10.5194/acpd-12-31483-2012>.
- [21] Kloog, I., Nordio, F., Coull, B.A., Schwartz, J., (2012). "Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM_{2.5} exposures in the Mid-Atlantic States". *Environmental Science and Technology* 46 (21), 11913e11921. <http://dx.doi.org/10.1021/es302673e>.
- [22] Gupta, P., Christopher, S.A., (2009a). "Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: multiple regression

approach”. *Journal of Geophysical Research* 114 (D14), 1e13.
<http://dx.doi.org/10.1029/2008JD011496>.

- [23] Gupta, P., Christopher, S.A., (2009b). “Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach.” *Journal of Geophysical Research* 114 (D20), 1e14.
<http://dx.doi.org/10.1029/2008JD011497>.
- [24] Barrero, M.A., Grimalt, J.O., Canton, L., (2006). “Prediction of daily ozone concentration maxima in the urban atmosphere”. *Chemometrics and Intelligent Laboratory Systems* 80, 67–76.
- [25] Cobourn, W.G., (2007). “Accuracy and reliability of an automated air quality forecast system for ozone in seven Kentucky metropolitan areas”. *Atmospheric Environment*, 41, 5863–5875.
- [26] Reggente, M., Peters, J., Theunis, J., Poppel, M.V., Kumar, P., De Baets, B. (2014). Prediction of ultrafine particle number concentration in urban environments by means of Gaussian process regression based on measurements of oxides of nitrogen. *Environment Modelling & Software* 61, 135-150.
- [27] Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., Brasseur, O., (2005). “A neural network forecast for daily average PM10 concentrations in Belgium”. *Atmospheric Environment* 39, 3279–3289.
- [28] Cobourn, W.G., Dolcine, L., French, M., Hubbard, M.C., (2000). “A comparison of nonlinear regression and neural network models for ground-level ozone forecasting”. *Journal of the Air & Waste Management Association* 50, 1999–2009.
- [29] Nelder J.A., Wedderburn R.W.M. (1972). “Generalized linear models”. *J R Stat Soc A*; 35: pg 370-84.

- [30] Vincent A., Hackley, Chiara F., Ferraris, (2001). “The Use of Nomenclature in Dispersion Science and Technology”, NIST Recommended Practice Guide, NIST, Special Publication 960-3.
- [31] Ketzler, M., Berkowicz, R., (2004). “Modelling the fate of ultrafine particles from exhaust pipe to rural background: an analysis of time scales for dilution, coagulation and deposition”. *Atmospheric Environment* 38, 2639-2652
- [32] Pohjola M, Kousa A, Aarnio P, Koskentalo T, Kukkonen J, Härkönen J, Karppinen A. (2000). “Meteorological Interpretation of Measured Urban PM_{2.5} and PM₁₀ Concentrations in the Helsinki Metropolitan Area”. In: Longhurst JWS, Brebbia CA, Power H (eds.), *Air Pollution VIII*. Wessex Institute of Technology Press.
- [33] Chaloulakou A., Kassomenos P., Spyrellis N., Demokritou P., Koutrakis P. (2003). “Measurements of PM₁₀ and PM_{2.5} particle concentrations in Athens, Greece”. *Atmospheric Environment* 37, pp649–660.
- [34] Anderson H.R., Derwent R.G., Stedman J., (2001). “Air Pollution and Climate Change”. St George's Hospital Medical School. 2 Met Office.
- [35] Conceição G.M.S, Saldiva P. H. N., Singer J. M., (2001). “Modelos MLG e MAG para análise da associação entre poluição atmosférica e marcadores de morbi-mortalidade: uma introdução baseada em dados da cidade de São Paulo”. *Rev. Bras. Epidemiol.* 206 Vol. 4, pp 206-219.

List of Tables

Table 1. General model parameter information resume.

Dependent Variable	PM ₁₀ concentration ($\mu\text{g}/\text{m}^3$)
Covariates	CO concentration ($\mu\text{g}/\text{m}^3$)
	NO ₂ concentration ($\mu\text{g}/\text{m}^3$)
	NO _x concentration ($\mu\text{g}/\text{m}^3$)
	O ₃ concentration ($\mu\text{g}/\text{m}^3$)
	SO ₂ concentration ($\mu\text{g}/\text{m}^3$)
	Temp (°C)
	RH (%)
	WV (m/s)
Probability Distribution	Poisson
Link Function	Logarithmic

Table 2. Specific models short description.

Model	Restriction	Dependent variable	Covariates
A	No [all values]	PM ₁₀ (µg/m ³)	CO, NO ₂ , NO _x , O ₃ SO ₂ Temp, RH, WV
B	T _{max} >25 °C	PM ₁₀ (µg/m ³)	CO, NO ₂ , NO _x , O ₃ SO ₂ Temp, RH, WV
C	T _{max} ≤ 25 °C	PM ₁₀ (µg/m ³)	CO, NO ₂ , NO _x , O ₃ SO ₂ Temp, RH, WV

Table 3. Resume of models results performance.

	Model A			Model B			Model C		
	Value	n	Value /n	Value	n	Value /n	Value	n	Value /n
Deviance	26868,641	6718	4,000	6450,747	3263	1,977	17576,459	3220	5,459
Pearson Chi-Square	27483,708	6718	4,091	6471,573	3263	1,983	18044,189	3220	5,604
Log Likelihood	31497,004			12102,567			17428,694		
Akaike's									
Information	63012,008			24223,134			34875,391		
Criterion (AIC)									
Finite Sample									
Corrected AIC	63012,034			24223,189			34875,447		
(AICC)									
Bayesian									
Information	63073,333			24277,973			34930,11		
Criterion (BIC)									
Consistent AIC									
(CAIC)	63082,333			24286,973			34939,11		

Table 4. Models likelihood ratio chi-square test performance.

			Model A			Model B			Model C		
			Value	NV	Sig	Value	NV	Sig	Value	NV	Sig
Likelihood	Ratio	Chi-Square	17979,21	8	0,00	14532,59	8	0,00	3537,43	8	0,00

Table 5. Results of models hypotheses tests.

Parameter	Model A			Model B			Model C		
	β	Wald Chi-Square	Sig	β	Wald Chi-Square	Sig	β	Wald Chi-Square	Sig
(Intercept)	2.425652	7677.888	0.000	1.957605	1146.149	0.000	2.419685	2184.525	0.000
CO	-0.000357	643.828	0.000	-0.000204	109.603	0.000	-0.000219	84.910	0.000
O3	0.001821	195.808	0.000	0.001931	146.327	0.000	0.000863	7.100	0.008
SO2	0.000364	22.990	0.000	-0.003097	359.387	0.000	0.002149	530.754	0.000
NO2	0.028348	8976.730	0.000	0.024388	3932.244	0.000	0.019767	785.657	0.000
NOx	0.000093	10.957	0.001	0.000309	83.629	0.000	0.001449	200.609	0.000
Temp	0.016820	613.923	0.000	0.043356	686.462	0.000	0.021912	250.801	0.000
RH	-0.000490	3.509	0.080	-0.000960	5.355	0.021	0.000153	0.129	0.720
WV	0.002821	1136.725	0.000	0.003548	167.289	0.000	0.003008	979.439	0.000

Table 6. Pearson correlation coefficients between variables.

	CO	O3	SO2	NO2	NOx	Temp	RH	WV
CO	1.00							
O3	-0.48	1.00						
SO2	-0.21	0.36	1.00					
NO2	0.52	-0.08	0.00	1.00				
NOx	0.69	-0.44	-0.06	0.79	1.00			
Temp	-0.41	0.52	0.35	-0.15	-0.32	1.00		
RH	0.34	-0.50	-0.22	0.02	0.22	-0.43	1.00	
WV	0.07	-0.17	0.02	-0.05	0.04	-0.17	0.19	1.00

List of Figure Captions

Fig.1 - Comparison between PM concentrations predicted versus measured for models A.

Fig.2 - Comparison between PM concentrations predicted versus measured for model B.

Fig.3 - Comparison between PM concentrations predicted versus measured for model C.

Fig.4 – Scattered plot of residuals for model A

Fig.5 – Scattered plot of residuals for model B

Fig.6 – Scattered plot of residuals for model

Fig.7 – Autocorrelation Function (ACF) values for variables

Fig.8 – Comparison between PM concentrations predicted versus measured for Oporto data

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

Figure 8