A METHODOLOGY TO ESTIMATE PM_{10} OUTDOOR URBAN CONCENTRATIONS USING GLM

J.M. Garcia (1), F. Teodoro (1), R. Cerdeira (1), L.M.R. Coelho (1), M.G. Carvalho (2)(3)

(1) Escola Superior de Tecnologia de Setubal, Instituto Politécnico de Setúbal, Campus do IPS, 2910-761 Setúbal, Portugal
(2) Instituto Superior Técnico, Portugal, (3) European Parliament Brussels, Belgium
Presenting author email: joao.garcia@estsetubal.ips.pt

Summary
This study presents a methodology that uses Generalized Linear Model (GLM) to predict PM_{10} concentrations based in the previous study of the relations between PM_{10} air concentrations and CO, NO_2, NOx, VOCs, SO_2 atmospheric concentrations, but also meteorological variables as air temperature, relative humidity and wind speed. The study is applied to a particular city (Barreiro, Portugal) and the model uses data from the monitoring air quality stations Portuguese network, and meteorological data. The developed GLM consider as dependent variable (or response variable) PM_{10} outside air concentrations, and considers as explanatory independent variables or covariates the air concentrations of pollutants NO_2, NO_x, CO, O_3 but also meteorological variables, air temperature, relative humidity and wind speed. A logarithmic link function was considered with a Poisson probability distribution. The studied model includes detailed inspection for cases with maximum air temperature below 25ºC and maximum air temperature above 25ºC. Results indicate that best performance is achieved for model considering only data with maximum air temperature values above 25ºC (R^2=0,649) when compared with model considering all data (R^2=0,386) and also when compared with model that considerers only data with maximum air temperature below 25ºC (R^2=0,149). The best performance model was tested with data from other Portuguese city (Oporto) showing reasonable fit results.

Introduction
In past recent, the concern about air quality has growth mainly due to the increasing knowledge of relations between respiratory problems, especially in children's and air pollution. In the case of PM, is known that airborne particulate matter has its origin in a wide variety of natural and anthropogenic sources. Some particles are directly emitted into the atmosphere and others may be formed in the air as a result of the chemical reaction of gases. Articles dedicated to study the contribution to air pollution from different sources, conclude that there is a strong relation between gaseous air pollutants and atmospheric PM.

Methodology and Results
General Linear Model (GLM) (Nelder et Wedderburn, 1972) was used to building a methodology to estimate PM_{10} outside concentrations, based on known values of other outside gaseous air pollutant concentrations. GLM are based in the assumption that there are K independent values Y_1, ..., Y_K, from a variable of interest or response variable (effect) that follows an exponential family distribution with expected value E (Yi) = μi [Conceição et al, 2001]. Considering K vectors x_i = (1 xi_1 $x1_2$ … xi_p)t, i=1, ..., K, containing the values of p explanatory variables, independent or covariates. A link logarithmic function was considered and Yi has a Poisson distribution, the model results in a Poisson regression model and each term β_i is the effect of variable X_i in g (μ_i). In fact, β_i represents the "effect" of variable X_i in the function g(μ_i). In this case the objective is to estimate PM_{10} concentration values based on other variables, which are gaseous air pollutant concentrations namely CO, NO_2, NO_x, O_3 and SO_2 but also meteorological variables namely temperature (T,ºC), relative humidity (RH,%) and wind velocity (WV, m/s). Statistical Package software for Social Sciences SPSS 10.0 for windows was used to build and analyse the model.

Conclusions
Results show that model considering all data predicts, poorly results for PM_{10} concentrations (R^2=0,386) but when two sub models were developed, considering only data with the criteria of maximum temperature of air i) above 25ºC and ii) below 25ºC results improved substantially. Comparisons of the three models show that best performance results are achieved for model with values of T_{max} air>25ºC (R^2=0,649) showing the importance of air temperature in the formation of the secondary particles in air.

Acknowledgement
The authors wish to acknowledge Comissão de Coordenação e Desenvolvimento Regional de Lisboa e Vale do Tejo (CCDR-LVT) and Instituto de Metereologia (IM) by the information provided.

References
Nelder J.A., Wedderburn R.W.M. (1972). "Generalized linear models". J R Stat Soc A; 35: pg 370-84.

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, Nº 3, pp 206-219.