13

Methodology to Assess Air Pollution Impact on Human Health Using the Generalized Linear Model with Poisson Regression

Yara de Souza Tadano[1], Cássia Maria Lie Ugaya[2]
and Admilson Teixeira Franco[2]
[1]*State University of Campinas – Sao Paulo,*
Department of Mechanical Engineering
[2]*Federal University of Technology – Paraná,*
Department of Mechanical Engineering
Brazil

1. Introduction

The growth of urban areas increased the access to many facilities such as transportation, energy, education, water supply, etc. As a consequence, there was a vehicular and industrial growth that combined with unfavorable meteorological conditions caused several worldwide episodes of excessive air pollution with life losses and health damage. Some examples were the well known Donora disaster in October, 1948 and fog episodes occurred after December, 1952 in London (Lipfert, 1993).

Since then, the researchers started to worry about air pollutants impact. Many epidemiological studies of air pollution have been conducted showing that air pollution affects human health, especially in respiratory and cardiovascular diseases, even where concentration levels of pollutants are below the air quality standard levels (Braga et al., 1999; Braga et al., 2001; Burnett et al., 1998; Ibald-Mulli et al., 2004; Peng et al., 2006; Peters et al., 2001; Pope III et al., 2002; Samet et al., 2000a, 2000b).

The evaluation of the impact of air pollution on human health is complex due to the fact that several personal characteristics (age, genetics, social conditions, etc.) influence on the response to a given air pollutant concentration. For instance, several studies have shown that a higher air pollution concentration increases the number of respiratory diseases in elderly and children (Braga et al., 1999; Braga et al., 2001). These studies show that children are more susceptible because they need twice the amount of air inhaled by adults and the elderly are more affected due to their weak immune and respiratory systems in addition to the fact they have been exposed to a great amount of air pollution throughout their lives. Another characteristic is genetics. The studies showed that people with chronic diseases or allergies, such as bronchitis and asthma are more sensitive to air pollution.

In this chapter, it will be presented a summary of four kinds of studies usually used to assess air pollution impact on human health and emphasizing the most used one: the time

series studies. In time series studies, a model very useful is the Generalized Linear Model (GLM). Then, the steps to apply the GLM to air pollution impact on human health studies will be presented in details, including a case study as an example. The results have shown that the GLM with Poisson regression fitted well to the database of the case study considered.

It is relevant to emphasize that the concepts included in this chapter are available in the literature, but the methodology presented to assess air pollution impact on human health employing the GLM with Poisson regression has no precedents.

2. Statistical methods

To assess air pollution impact on human health, epidemiological studies often use statistical methods that are extremely useful tools to summarize and interpret data.

The health effects (acute or chronic), type of exposure (short or long term), the nature of the response (binary or continuous) and data structure lead to model selection and the effects to be estimated. Regression models are generally the method of choice.

The exposure to ambient air pollution varies according to temporal and/or spatial distribution of pollutants. Most air pollution studies have used measures of ambient air pollution instead of personal exposure because estimating relevant exposures for each person can be daunting. According to this approximation; "misclassification of exposure is a well-recognized limitation of these studies" (Dominici et al., 2003).

According to Dominici et al. (2003), epidemiological studies of air pollution fall into four: time series; case-crossover; panel and cohort. The time series, case-crossover and panel studies are more appropriate for acute effects estimation while the cohort studies are used for acute and chronic effects combined.

2.1 Case-crossover studies

Case-crossover studies are conducted to estimate the risk of a rare event associated with a short-term exposure. It was first proposed by Maclure (1991) cited in Dominici et al. (2003) to "study acute transient effects of intermittent exposures". In practice, this design is a modification of the matched case-control design. The difference between a case-crossover and a case-control design is that in case-control designs, each case acts as his/her own control and the exposure distribution is then compared between cases and controls and in case-crossover design "exposures are sampled from an individual's time-varying distribution of exposure". In particular, "the exposure at the time just prior to the event (the *case* or *index time*) is compared to a set of *control* or *referent times* that represent the expected distribution of exposure for non-event follow-up times". In such a way, the unique characteristics of each individual such as gender, age and smoking status; are matched, reducing possible confounding factors (Dominici et al., 2003).

According to Maclure & Mittleman (2000) cited in Dominici et al. (2003), "in the last decade of application, it has been shown that the case-crossover design is best suited to study intermittent exposures inducing immediate and transient risk, and abrupt rare outcomes".

2.2 Panel studies

Panel studies collect individual time and space varying exposures, outcomes counts and confounding factors. Consequently they include all other epidemiological designs which are based on temporally and/or spatially aggregated data. Actually, panel studies also rely on group-level data.

In panel designs, the goal is to follow a cohort or panel of individuals to investigate possible changes in repeated outcome measures. This design shows to be more effective in short-term health effects of air pollution studies, mainly for a susceptible subgroup of the population. Usually, panel studies involve the collection of repeated health outcomes measures for all considered subjects of a susceptible subpopulation over the entire time of study. The measure of pollution exposure could be from a fixed-site ambient monitor or from personal monitors (Dominici et al., 2003).

Some care should be taken when designing a panel study, because the main goal of estimating the health effect of air pollution exposure sometimes can be less clear. It happens whenever the panel members do not share the same observation period, so parameterization and estimation of exposure effects need to be considered with much care (Dominici et al., 2003).

2.3 Cohort studies

The cohort studies are frequently used to associate long-term exposure to air pollution with health outcomes. Prospective or retrospective designs are possible. The first one consists of participants' interview at the beginning of the research containing particular information such as age, sex, education, smoking history, weigh, and so on. After that, the participants are followed-up over time for mortality or morbidity events. The retrospective design consists of using already available database information. Cohort designs are frequently used to multicity studies, as it ensures "sufficient variation in cumulative exposure, particularly when ambient air pollution measurements are used" (Dominici et al., 2003).

2.4 Time series studies

The time series impact studies are often used as they demand simply data such as the amount of hospital admission or mortality in a given day, being easy to obtain on health government departments. So it is unnecessary to follow-up the group of people involved in the study, which demands much time (Schwartz et al., 1996).

Another key advantage of the time series approach is the use of daily data and while the underlying risk on epidemiological studies of air pollution varies with some factors such as age distribution and smoking history, these factors will not have influence on the expected number of deaths or morbidity on any day, since they do not vary from day to day (Schwartz et al., 1996).

Regression models are usually chosen in time series studies, as they are useful tools to assess the relationship between one or more explanatory variables (independent, predictor variables or covariates) (x_1, x_2,…,x_n) and a single response variable (dependent or predicted variable) (y) (Dominici et al., 2003). The simplest regression analysis consisting of more than one explanatory variable is the multiple linear regression and is given by:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon\,, \tag{1}$$

where y is the response variable and x_i (i = 1, 2, …, n) are the explanatory variables. β_0 represents the value of y when all the explanatory variables are null, β_i terms are called regression coefficients and the residual (ε) is the prediction error (the difference between measured and adjusted values of the response variable).

The regression models goal is to find an expression that better predicts the response variable as a combination of the explanatory variables. It means to find the $\beta's$ that better fits to the database.

Due to the non-linearity of the response variable in time series studies of air pollution impacts on human health, the Generalized Linear Models (GLM) with parametric splines (e.g. natural cubic splines) (McCullagh & Nelder, 1989) and the Generalized Additive Models (GAM) with non-parametric splines (such as smoothing splines or lowess smoothers) (Hastie & Tibishirani, 1990) are usually applied.

According to the studies conducted in the last decade, GAM was the most widely applied method as it allows for non-parametric adjustment of the non-linear confounding factors such as seasonality, short-term trends and weather variables. It is also a more flexible approach than fully-parametric models like GLM with parametric splines. Nevertheless, recently the GAM implementation in statistical softwares, like S-Plus has been called into question (Dominici et al., 2003).

To evaluate the impact of default implementation of the GAM software on published analyses, Dominici et al. (2002) reanalyzed the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) data (Samet et al., 2000a, 2000b) using three different methods: The GLM (Poisson regression) with natural cubic splines to achieve nonlinear adjustments for confounding factors; the GAM with smoothing splines and default convergence parameters; and the GAM with smoothing splines and more stringent convergence parameters than the default settings. The authors found that "estimates obtained under GLMs with natural cubic splines better detect true relative rates than GAMs with smoothing splines and default convergence parameter". The authors also added that: "although GAM with nonparametric smoothers provides a more flexible approach for adjusting for nonlinear confounders compared with fully parametric alternatives in time series studies of air pollution and health, the use and implementation of GAMs requires extreme caution".

In such a way, in this chapter it will be presented all the steps that should be followed to conduct a time series study using GLM with Poisson regression, from data collection to measure the goodness of fit. More details about design comparisons between time series; case-crossover; panel and cohort studies are in Dominici et al. (2003).

3. Generalized Linear Models (GLM)

The GLMs are a union of linear and non-linear models with a distribution of the exponential family, which is formed by the normal, Poisson, binomial, gamma, inverse normal distributions including the traditional linear models, as well as logistic models (Nelder & Wedderburn, 1972).

Since 1972, many researches on GLMs where conducted and as a consequence several computational skills were created such as, GLIM (Generalized Linear Interactive Models), S-Plus, R, SAS, STATA and SUDAAN (Dobson & Barnett, 2008; Paula, 2004).

The GLMs are defined by a probability distribution of the exponential distribution family, and are formed by the following components (McCullagh & Nelder, 1989):

- Random component: n explanatory variables $(y_1,\ldots,y_n)$ of a response variable which follows a distribution of the exponential family with expected value $E(y_i)=\mu$;
- Systematic component: concerns a linear structure for the regression model $(\boldsymbol{\eta}=\boldsymbol{\beta}\mathrm{x}^T)$, called linear predictor, where $\mathrm{x}^T=(x_{i1},x_{i2},\cdots,x_{ip})^T$, $i=1,\cdots,n$ are the so-called explanatory variables and;
- Link function: a monotone and differentiable function g, called link function, capable of connecting the random and systematic components, relating the response variable mean (μ) to the linear structure, defined in GLMs as $g(\mu)=\eta$, where:

$$\eta=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n, \tag{2}$$

or in matrix form:

$$\boldsymbol{\eta}=\boldsymbol{\beta}\mathrm{x}^T, \tag{3}$$

where the regression coefficients $\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_n)$ represents the vector of parameters to be estimated (McCullagh & Nelder, 1989).

Each distribution has a special link function, called canonical link function which occurs when $\eta=\theta_i$, where θ is called the local or canonical parameter. Table 1 shows the canonical function for some distributions of the exponential family (McCullagh & Nelder, 1989).

Distribution	**Canonical link function** (η)
Normal	μ
Poisson	$\ln(\mu)$
Binomial	$\ln\{\mu/(1-\mu)\}$
Gamma	μ^{-1}
Inverse Gaussian	μ^{-2}

Table 1. Canonical link functions of some distributions of the exponential family (McCullagh & Nelder, 1989).

According to Myer & Montgomery (2002), using the canonical link function implies some interesting properties, although it does not mean it should be always used. This choice is convenient because, besides the simplification of the estimative of the model parameter, it also becomes easier to obtain the confidence interval of the response variable mean. However, the convenience do not necessarily implies in goodness of fit.

In studies of air pollution impact on human health with non-negative count data as response variable, the GLM with Poisson regression is broadly applied (Dockery & Pope III, 1994; Dominici et al., 2002; Lipfert, 1993; Metzger et al., 2004).

The GLM with Poisson regression consists in relating the response variable (y) (mortality or morbidity), which can take on only non-negative integers, with the explanatory variables (x_1,x_2,…,x_n) (pollutants concentration, weather variables, etc.) according to:

$$\ln(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n , \tag{4}$$

Usually, the regression coefficients ($\beta's$) are estimated using the Fisher score method of maximizing the likelihood function (maximum likelihood method), which is the same as the Newton-Raphson method when the canonical link function is considered. For the Poisson regression, the likelihood density function is given by (Dobson & Barnett, 2008; McCullagh & Nelder, 1989):

$$f_y(y;\theta;\phi) = \exp\{y\ln(\mu) - \mu - \ln(y!)\} , \tag{5}$$

where y is the response variable, θ is the canonical parameter and ϕ is the dispersion parameter. When the link function is the canonical link ($\ln \mu = \eta$).

One feature of the GLM with Poisson regression is that even if all the explanatory variables were known and measured without error, there would still be considerable unexplained variability in the response variable. This is a result of the fact that even if the response variable is more precise, the Poisson process ensures stochastic variability around that expected count. In a classic stationary Poisson regression, the variance is equal to the mean (ϕ = 1), but in many actual count processes there is overdispersion, when the variance is greater than the mean or, underdispersion the other way round. In these cases, it is still possible to apply the GLM with Poisson regression (Everitt & Hothorn, 2010; Schwartz et al., 1996). One way to adjust the over or underdispersion is to assume that the variance is a multiple of the mean and estimate the dispersion parameter using the quasi-likelihood method. Details of this method are in McCullagh & Nelder (1989).

4. Steps to fit GLM with poisson regression

To apply the GLM with Poisson regression, four main steps should be followed: development of the database; adjustment of the temporal trends; goodness of fit analysis and results analysis. The details of each step are at the following topics.

4.1 Database

Usually, in time series studies of air pollution impact on human health using the GLM with Poisson regression, the data used are air pollutants concentration, weather measures, outcome counts and some confounding terms. The data must be collected daily and for at least two years, to capture the seasonal trends.

Pollutants concentrations are usually obtained by fixed-site ambient monitors. The weather measures frequently used are temperature (or dewpoint temperature) and air relative humidity.

The outcome depends on the purpose of the study. For example, in some studies mortality is used, in others, morbidity. The outcome can be stratified by type of disease (such as respiratory or cardiovascular diseases), age (children, young and elderly people) and any other factor of interest. The confounders can be of long-term (such as seasonality) or short-term (day of the week, holiday indicator, etc.) and will be shown in Section 4.2.

4.2 Temporal trends adjustment

A common feature of epidemiological studies is biases due to confounding factors and correlations among covariates that can never be completely ruled out in observational data. Confounding factors are present when a covariate is associated with both the outcome and the exposure of interest but is not a result of the exposure. So, in all epidemiological studies, a basic issue in modeling is to control properly for all the potential confounders. Time series studies have some unique features in this regard (Dominici et al., 2003; Peng et al., 2006).

The intercorrelation of different pollutants in the atmosphere is one source of biases. One way to address this intercorrelation "has been to conduct studies in locations where one or more pollutants are absent or nearly so" (Dominici et al., 2003).

The sources of potential confounding factors in time series studies of air pollution impact on human health can be broadly classified as measured or unmeasured. Measured confounding factors such as weather variables (temperature; dewpoint temperature; humidity and others) are of unique importance in this kind of studies. Some studies have demonstrated a relationship between temperature and mortality being positive for warm summer days and negative for cold winter days, like in Curriero et al. (2002) cited in Peng et al. (2006). One approach to adjust confounding factors by temperature or humidity is to include non-linear functions or a mean of current and previous days temperature (or dewpoint temperature) in the model (Peng et al., 2006). Unmeasured confounding factors are those factors that have influence in outcome counts and have a similar variation in time as air pollutants concentration. These confounding factors produce seasonal and long-term trends in outcome counts that can confound its relationship with air pollution. Some important examples are influenza and respiratory infections (Peng et al., 2006).

4.2.1 Seasonality

In time series studies, the primarily concern is about potential confounding by factors that vary on timescales in a similar manner as pollution and health outcomes. This attribute is usually called seasonality. "A common approach to adjust this trend is to use semi-parametric models which incorporate a smooth function of time". The smooth function serves as a linear filter for the mortality (morbidity) and pollution series and "removes any seasonal or long-term trends in the data" (Peng et al., 2006). Several methods to deal with this trend are being used such as smoothing splines, penalized splines, parametric (natural cubic) splines and less common LOESS smoothers or harmonic functions (Dominici et al., 2002; Peng et al., 2006; Samet et al., 2000a, 2000b; Schwartz et al., 1996).

The spline function provides an approximation for the behavior of functions which has local and abrupt changes. The most used spline to smoothing curves in GLMs is the natural cubic

spline (Chapra & Canale, 1987; Samoli et al., 2011; Schwartz et al., 1996), the other ones are usually applied in GAM.

Using splines, polynomial functions will be provided for each defined interval instead of a single polynomial for the whole database. The natural cubic spline is based on third order polynomials derived for each interval between two knots at fixed locations throughout the range of the data (Chapra & Canale, 1987; Peng et al., 2006). The choice of knots locations can result in substantial effect on the resulting smooth. So, in Peng et al. (2006) study the authors "provided a comprehensive characterization of model choice and model uncertainty in time series studies of air pollution and mortality, focusing on confounding factors adjustment for seasonal and long-term trends". According to their results, for natural splines, the bias drops suddenly between one and four degrees of freedom (df) per year and is stable afterwards, suggesting that at least 4 degrees of freedom per year of data should be used. In such way, in time series studies of air pollution and mortality (or morbidity) usually is used four to six knots per year, as the seasonality trend is due to the different behavior of variables during the seasons of the year (Tadano, 2007). Their results show that "both fully parametric and nonparametric methods perform well, with neither preferred. A sensitivity analysis from the simulation study indicates that neither the natural spline nor the penalized spline approach produces any systematic bias in the estimates of the log-relative-rate β" (Peng et al., 2006).

The smooth functions of time accounts only for potential confounding factors which vary smoothly with time, such as seasonality. Some potential confounders which vary on shorter timescales are also important, as they confound the relationship between air pollution and health outcomes, such as day of the week and holiday indicator (Peng et al., 2006).

4.2.2 Day of the week and holiday indicator

Important potential confounding factors that may bias time series studies of air pollution and mortality (or morbidity) are factors which vary on shorter timescales like calendar specific days, such as day of the week and holiday indicator (Lipfert, 1993). These trends are not necessarily present, but they occur often enough that they should be checked (Samoli et al., 2011; Schwartz et al., 1996). For example, on weekends the number of hospital admissions can be lower than on weekdays and can also be lower during holidays.

One way to adjust according the week day trend is to add qualitative explanatory variable for each day of the week (varying from one to seven) starting at Sundays. To adjust the holiday indicator, it can be considered an additional binomial explanatory variable in which one means holidays and zero means workdays (Tadano et al., 2009).

Adding all the time trends mentioned and explanatory variables in the GLM with Poisson regression, the expression used in some studies of air pollution impact on population's health is as follows (Tadano, 2007):

$$\ln(y) = \beta_0 + \beta_1 T + \beta_2 RH + \beta_3 PC + \beta_4 H + \beta_5 dow + \beta_6 ns\,, \tag{6}$$

where y = health outcome of interest; T = air temperature or dewpoint temperature (°C); RH = air relative humidity (%); PC = pollutant concentration ($\mu g/m^3$); H = time trend

variable for holidays; *dow* = time trend variable for days of the week; *ns* = natural cubic spline to adjust for seasonality.

Some of these short-term trends can lead to autocorrelation between data from one day to previous days, even after its adjustment. In this regard, partial autocorrelation functions are used.

4.2.3 Partial autocorrelation functions

The short-term trends such as days of the week and holiday indicator can lead to an autocorrelation between data from one day and previous days, even using the adjustment. One way to analyze this time trend is plotting the partial autocorrelation function (Partial ACF) against lag days.

The autocorrelation function of the model's residuals is as follows:

$$ACF = \frac{c_k}{c_0}, \tag{7}$$

where $c_k = \frac{1}{n}\sum_{i=1}^{n-k}(y_i - \mu)(y_{i+k} - \mu)$, with n = number of observations and k = lag days (Box et al., 1994). In the partial autocorrelation function plot, the residuals should be as smaller as possible, ranging from $-2n^{-1/2}$ to $2n^{-1/2}$ (dashed lines) as shown in Fig. 1.

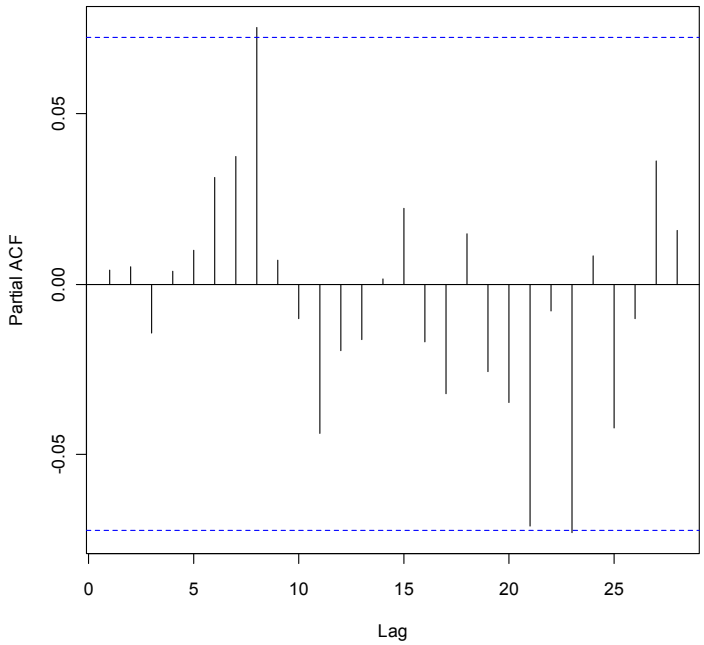


Fig. 1. Example of the partial autocorrelation function (Partial ACF) plot against lag days where there are no autocorrelations between data for less than five lag days.

In epidemiological studies of air pollution, the important autocorrelations are those occurring in the first five days, which are usually caused by the decrease of health outcomes in weekends and holidays (Tadano, 2007). If the database has autocorrelations, then the model should consider them by including the residuals in the model.

In R or S-Plus language, the residuals to be included are the working residuals. These residuals are returned when extracting the residuals component directly from the *glm* comand. They are defined as:

$$r_i^W = (y_i - \hat{\mu}_i)\frac{\partial \eta}{\partial \hat{\mu}_i}, \tag{8}$$

where η = the link function which in canonical form of the Poisson regression $\eta = \ln \mu$; y_i = measured values of the response variable; $\hat{\mu}_i$ = adjusted value by modeling and $i = 1, 2, ..., n$ with n = number of observations.

After adjusting the GLM with Poisson regression including all time trends and explanatory variables, the fitting model need to be tested to assure that this is the best model to be applied to the database.

4.3 Goodness of fit

The GLM with Poisson regression has been widely applied in epidemiological studies of air pollution (Dockery & Pope III, 1994; Dominici et al., 2002; Lipfert, 1993; Metzger et al., 2004; Tadano et al., 2009) but it needs caution, as sometimes this model may not fit well to the database. There are two statistical methods that can be used to evaluate goodness of fit in GLMs, as follows.

4.3.1 Pseudo R^2

One interesting and easy to apply goodness of fit test for GLM with Poisson regression is the statistic called pseudo R^2 which is similar to the determination coefficient of classic linear models. It is defined as:

$$Pseudo\, R^2 = \frac{l(\boldsymbol{b}_{\min}) - l(\boldsymbol{b})}{l(\boldsymbol{b}_{\min})}, \tag{9}$$

where l =log-likelihood function; $l(\boldsymbol{b}_{\min})$ = maximum value of the log-likelihood function for a minimal model with the same rate parameter for all y's and no explanatory variables (null model) and $l(\boldsymbol{b})$ = maximum value of the log-likelihood function for the model with p parameters (complete model) (Dobson & Barnett, 2008).

This statistic measures the deviance reduction due to the inclusion of explanatory variables and can be applied in R (R Development Core Team, 2010) throughout the Anova Table with chi-squared test where the residual deviance values indicates the maximum value of the log-likelihood function for the complete model and the null one.

According to Faraway (1999), a good value of R^2 depends on the area of application. The author suggests that in biological and social sciences it is expected lower values for R^2.

Values of 0.6 might be considered good, because in these studies the variables tend to be more weakly correlated and there is a lot of noise. The author also advises that it is a generalization and "some experience with the particular area is necessary for you to judge your R^2's well".

4.3.2 Chi-squared statistic

Another statistical test used as goodness of fit in GLM with Poisson regression is the chi-squared (χ^2) or Pearson statistic, which is used to evaluate the model fit comparing the measured distribution to that obtained by modeling. The expression that represents the chi-squared statistic is:

$$\chi^2 = \sum_{i=1}^{n} \left(y_i - \hat{\mu}_i\right)^2 / \hat{\mu}_i \,, \tag{10}$$

where y_i = measured values of the response variable; $\hat{\mu}_i$ = adjusted value by modeling and $i = 1, 2, ..., n$ with n = number of observations.

The chi-squared statistic is the sum of the Pearson residual of each observation. According to this statistic a model that fits well to the data has a chi-squared statistic close to the degrees of freedom (df) (χ^2/df~1), where df = n – p (n = number of observations and p = number of parameters) (Wang et al., 1996). There is no evidence of which goodness of fit (Pseudo R^2 or χ^2) is preferred.

After the confirmation of GLM with Poisson regression fitting, the results must be analyzed to find the nature of the correlation between air pollutants concentration and health outcomes.

4.4 Results analysis

In epidemiological studies of air pollution it is common to find a relation between the air pollutants concentration of one day to the health outcomes of the next day, two days later or even after one week. Then, researchers usually fit the model to different arrangements of the same database with lags. In time series studies, lags of one day to seven days are frequently applied and then the one that best fits is chosen. One criteria to select the best option is the Akaike Information Criterion.

4.4.1 Akaike information criterion

The Akaike Information Criterion (AIC) is very useful when choosing between models from the same database. The smallest is the AIC, the better is the model. The AIC is automatically calculated in R software when applying the GLM algorithm and is calculated by:

$$AIC = -2l(\boldsymbol{b}) + 2(df)\hat{\phi} \,, \tag{11}$$

where $l(\boldsymbol{b})$ = maximum log-likelihood value for the complete model; df = degrees of freedom of the model and $\hat{\phi}$ = estimated dispersion parameter (Peng et al., 2006).

After choosing for the model that better fits the database and which has the best relationship between air pollution and health outcomes, a method to verify the strength of this

relationship is applied. One method frequently applied (used in S-Plus and R software) is the Student *t* Test.

4.4.2 Student *t* test for statistical significance

In time series studies, the confirmation of any relation between air pollution and health outcomes is obtained throughout a hypothesis test that can show if the regression coefficients are statistically significant or not.

The statistical hypothesis to be tested is the null one (H_0) expressed by an equality. The alternative hypothesis is given by an inequality (Mood et al., 1974).

The hypothesis test has several goals; one of them is to verify if the estimated regression coefficient can be discredited. In this case, the following hypotheses are considered H_0: $\beta = 0$ and H_1: $\beta \neq 0$. The statistical test used to verify these hypotheses is given by:

$$t_0 = \beta / \varepsilon \,, \tag{12}$$

where ε is the standard error of the estimated regression coefficient (β). The rejection of the null hypothesis occurs when $|t_0| > t_{\alpha/2,n-k-1}$ (n = number of observations, k = number of explanatory variables, α is the considered significance level), indicating that the estimated regression coefficient is statistically significant. In other words, the explanatory variable influences in the response variable (Bhattacharyya & Johnson, 1977; Mood et al., 1974).

The values $t_{\alpha/2,n-k-1}$ are presented in Student t distribution table, where n-k-1 is the degrees of freedom (df) and α is the considered significance level (Bickel & Doksum, 2000).

If the study results in a statistically significant relation between air pollutant concentration and the health outcome of interest, then some analysis and projections are made using the relative risk (RR).

4.4.3 Relative risk

The relative risk (called rate ratio by statisticians) (Dobson & Barnett, 2008) is used to estimate the impact of air pollution on human health, making some projection according to pollutants concentration.

The relative risk is a measure of the association between an explanatory variable (e.g. air pollutant concentration) and the risk of a given result (e.g. the number of people with respiratory injury) (Everitt, 2003).

In a specific way, the relative risk function at level x of a pollutant concentration, denoted as RR(x), is defined as (Baxter et al., 1997):

$$RR(x) = \frac{E(y|x)}{E(y|x=0)} \,. \tag{13}$$

It is the ratio of the expected number of end points at level x of the explanatory variable to the expected number of end points if the explanatory variable was 0 (Baxter et al., 1997). For the Poisson regression, the relative risk is given by:

$$RR(x) = e^{\beta x}, \tag{14}$$

indicating, for example, that the risk of a person exposed to some pollutant concentration (x) having a specific injury is RR(x) times greater than someone who has not been exposed to this concentration. A RR(x) = 2 for a pollutant concentration of 100 $\mu g/m^3$, indicates that a person exposed to this concentration has two times more chance to get a health problem than someone who has not been exposed to any concentration.

5. Case-study

To exemplify the appliance of GLM with Poisson regression, a case study will be presented. It will be evaluated the impact of air pollution on population's health of Sao Paulo city, Brazil, from 2007 to 2008. Sao Paulo is the largest and most populated city of Brazil, and one of the most populated in the world.

In this study, it was evaluated the impact of PM_{10} (particles with an aerodynamic diameter less or equal to 10 μm) on the number of hospital admissions for respiratory diseases, according to the International Classification of Diseases (ICD-10).

PM_{10} was chosen because according to WHO (2005) as cited in Schwarze *et al*. (2010), particulate air pollution is regarded as a serious health problem and some studies reported that reductions in particulate matter levels decrease health impact of air pollution. According to Braga *et al*. (2001) the health outcomes had high correlation with PM_{10} concentration in Sao Paulo (Brazil) population.

5.1 Case-study database

The data collected in this study consisted of daily values from January 1st, 2007 to December 31st, 2008 to Sao Paulo city, Brazil.

The hospital admissions for respiratory diseases, according to the ICD-10, were considered as response variable. The data was obtained from the Health System (SUS) website (2011). The explanatory variables consisted of PM_{10} concentration and weather variables (air temperature and air relative humidity), also including parametric splines for long-term trend (seasonality), qualitative variable for days of the week and binomial variable for holiday indicator.

The PM_{10} concentration, air temperature and humidity where obtained from QUALAR system in Cetesb (Environmental Company of Sao Paulo State) website (2011).

The fixed-site monitoring network of Sao Paulo city, held by Cetesb, has twelve automatic stations, and PM_{10} concentration is collected in all but one of the stations and temperature and humidity data are acquired in eight of them. This network also contains nine manual stations, but none of them monitors PM_{10} concentration, just TSP (Total Suspended Particles

- particles with an aerodynamic diameter less or equal to 50 μm) and $PM_{2.5}$ (particles with an aerodynamic diameter less or equal to 2.5 μm).

The PM_{10} concentration was monitored in eight fixed-site automatic monitoring stations during the study period (from 2007 to 2008); the air temperature and the air relative humidity were monitored only at two stations. The daily data of these variables used in this study comprise the mean of the available data.

The descriptive analysis of the variables considered in this study is presented in Table 2. The values in Table 2 show that the maximum daily PM_{10} concentration (103 $\mu g/m^3$) did not overcome the national air quality standard (150 $\mu g/m^3$).

Variable	Mean	Standard Deviance	Minimum	Maximum
RD	165.53	42.43	57.00	267.00
PM_{10}	39.25	18.27	11.00	103.00
Temperature	20.49	2.94	10.10	27.20
Humidity	73.36	9.13	40.40	99.50

Table 2. Descriptive statistics for hospital admissions for respiratory diseases (RD), concentration of PM_{10} and weather variables.

To have an initial idea of the relation between the response variable and the explanatory ones, the Pearson correlation matrix was constructed (Table 3). The Pearson correlation between hospital admissions for respiratory diseases (RD) and PM_{10} was positive and statistically significant. It means the number of RD increases as PM_{10} concentration increases. This table also shows that the number of RD increases as temperature and humidity decreases, but the Pearson correlation was not statistically significant in this case. Consequently, as shows Table 3, the PM_{10} concentration increases in days with low temperature and humidity indexes.

	RD	PM_{10}	Temperature	Humidity
RD	1.00			
PM_{10}	0.41*	1.00		
Temperature	-0.14	0.01	1.00	
Humidity	-0.15	-0.57*	-0.27*	1.00

* statistically significant ($p<0.05$)

Table 3. Pearson correlation matrix between hospital admissions for respiratory diseases (RD), concentration of PM_{10} and weather variables.

5.2 Long and short-term trend adjustment

The long-term trend usually included in time series studies of air pollution impact on human health is seasonality and in this case study it was considered a natural cubic spline, the most used parametric smooth in GLMs.

To apply the natural cubic spline in GLM with Poisson regression, an explanatory variable for the days is added to the model, consisting of values from 1 to 731, comprising all two years of data.

The short-term trends usually considered in epidemiological studies of air pollution are the day of the week and holiday indicator. The day of the week variable was considered as a qualitative variable which varies from one to seven, starting at Sundays. The holiday indicator was adjusted adding a binomial variable in which one means holidays and zero means workdays.

According to the considerations above, Table 4 brings an example of the first lines of the database used.

Data	RD	PM10	T	RH	day	dow	H
01/01/2007	104	15	21.7	85.7	1	2	1
02/01/2007	171	14	22.1	82.7	2	3	0
03/01/2007	140	17	21.9	87.2	3	4	0
04/01/2007	155	20	22	92	4	5	0
05/01/2007	130	13	21.5	89.9	5	6	0
06/01/2007	93	11	21.9	84.2	6	7	0
07/01/2007	101	18	23.2	80.4	7	1	0

Table 4. First values of the database considered in this study (RD = daily number of hospital admissions for respiratory diseases, PM10 = concentration of PM_{10} in $\mu g/m^3$; T = air temperature in ^oC; RH = air relative humidity; day = variable to consider seasonality; dow = days of the week; H = holidays indicator).

With the database considered, the expression used to apply the GLM with Poisson regression in R software (R Development Core Team, 2010) is:

$$\text{m}.name <- glm\left(\begin{array}{l} RD \sim ns(day, df) + as.factor(dow) + as.factor(H) + T + RH + PM10, \\ data = database.name, family = poisson, na.action = na.omit \end{array}\right), \quad (15)$$

where *m.name* = is the name given to the analysis; *ns* = natural cubic spline; *df* = degrees of freedom; *database.name* = name given to the database file.

To apply this model, one important decision is about the number of degrees of freedom (*df*) to be considered in the natural cubic spline of days of study. In epidemiological studies of air pollution, the common values are four, five or six degrees of freedom per year of data. To decide which one to use, three analyses where made considering four, five and six degrees of freedom (*df*) in Equation (15) and the results were compared using the AIC, as shown in Table 5.

Number of df per year	AIC
4	6,911.4
5	6,909.2
6	6,798.7

Table 5. Comparison of models with different numbers of degrees of freedom for seasonality adjustment.

According to the results indicated in Table 5, the model with 6 degrees of freedom per year of data is the one that better fits the data. Then, in the following analyses, it was considered $df = 6$ in Equation (15).

The short-term trends considered in this study (days of the week and holiday indicator) can lead to autocorrelation between data from one day and the previous days, so the Partial ACF plot against lag days must be analyzed. The lines of each lag day until five lags must be between $-2n^{-1/2}$ and $2n^{-1/2}$. In this case study the number of observations (n) is equal to 731, so the lines in Partial ACF plot out of the range (-0.07;0.07) indicates a strong autocorrelation between data from one day and previous days.

The Partial ACF plots against lag days for the model with six degrees of freedom and considering the effects of PM_{10} concentration on the same day for the model with no residual inclusion is shown Fig. 2 and Fig. 3 shows for the model after including residuals.

For the model with 6 degrees of freedom, the Partial ACF plot (Fig. 2) shows autocorrelations between one day and the previous 1, 2, 3 and 4 days, as the lines for these lag days are out of the range. To adjust for this time trend, it is necessary to include the residuals for these lag days in the model.

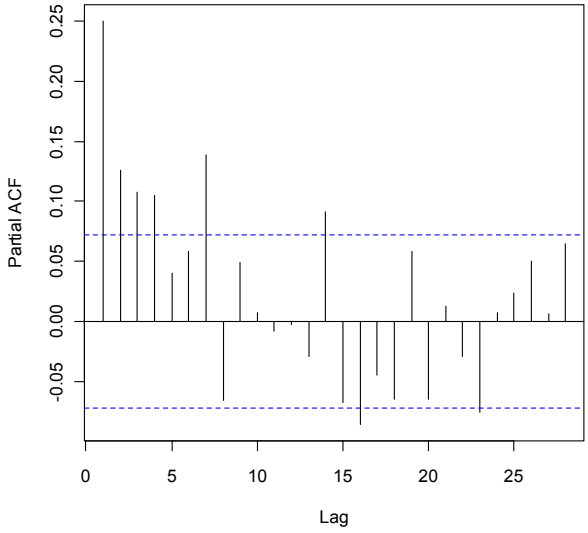


Fig. 2. Partial ACF plot against lag days with no residuals included.

After do so, the Partial ACF plot (Fig. 3) shows no more autocorrelations between data for the first five days, indicating that it is the best fitted model.

After adjusting the GLM with Poisson regression including all the time trends and explanatory variables and choosing the degrees of freedom that better fits the data; the fitted model was tested using the pseudo R^2 and the chi-squared statistic to assure that it is the right one to be applied to the case-study.

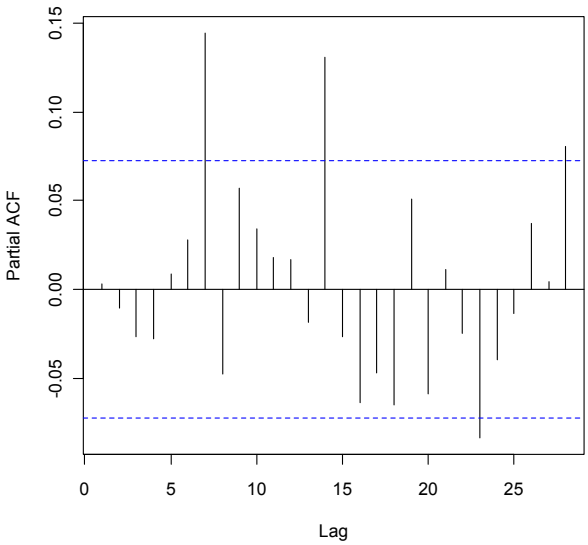


Fig. 3. Partial ACF plot against lag days including residuals.

5.3 Model adjustment results

In epidemiological studies of air pollution it is common to find a relation between the air pollutants concentration of one day to the health outcomes of some lag days. In this case-study, analyses of the relation between PM_{10} concentration of one day and the number of hospital admission for respiratory diseases for the same day until one week later was performed.

All models were fitted with no residual inclusion and also with inclusion of residuals due to autocorrelation.

The goodness of fit to the analyses from no lag to seven lag days is shown in Table 6 (A) without residual and, (B) with residuals. The ACF plots for all models adjusted will not be shown, but they are similar to that of Fig. 2. All of them (from no lag to seven lag days) indicated the need of residuals inclusion for 1, 2, 3 and 4 lag days. The models with residuals inclusion did not show anymore autocorrelations.

A

Lag days	R^2	χ^2	χ^2/df
0	0.82	1,509	2.16
1	0.82	1,530	2.19
2	0.81	1,528	2.19
3	0.81	1,526	2.19
4	0.81	1,529	2.20
5	0.81	1,531	2.20
6	0.81	1,531	2.21
7	0.81	1,525	2.20

B

Lag days	R^2	χ^2	χ^2/df
0	0.80	1,701	2.40
1	0.79	1,718	2.43
2	0.79	1,709	2.42
3	0.79	1,708	2.42
4	0.79	1,715	2.44
5	0.79	1,721	2.45
6	0.79	1,721	2.45
7	0.79	1,724	2.46

Table 6. Goodness of fit results for the analyses from no lag to seven lag days for models with no residual inclusion (A) and including residuals in the model (B).

According to the analysis of the goodness of fit shown in Table 6, all the models presented a pseudo R^2 greater than 0.6, showing that the models fitted well to the data, but the chi-squared statistic analysis showed values much greater than the degrees of freedom. As there is no evidence of which statistic is suitable for each situation, we can conclude the model fitted well to the data, according to Pseudo R^2 statistic results.

After verifying the models fitted well to the data, the analysis of the regression coefficients was held. The results are shown in Table 7 without residuals and Table 8 with residuals. Analyzing the AIC, the model that include the residuals seems to fit better than the one which was not included and the model considering the effect of seven days lag shows better results, but the regression coefficient did not show statistical significance.

Furthermore, the AIC value with three days lag is lower than for two, one or no lags; showed no autocorrelation and with a regression coefficient statistically significant.

In conclusion, the chosen model was the one with the effect of three days lag in which residuals was included. The relative risk results are therefore only presented for this model (with # symbol in Table 8).

Lag days	AIC	β	ε	t-value
0	6,798.7	$1.60x10^{-03}$	$2.54x10^{-04}$	6.30***
1	6,809.6	$9.88x10^{-04}$	$2.17x10^{-04}$	4.56***
2	6,793.6	$9.07x10^{-04}$	$1.98x10^{-04}$	4.59***
3	6,786.0	$8.91x10^{-04}$	$1.96x10^{-04}$	4.55***
4	6,785.8	$6.20x10^{-04}$	$1.94x10^{-04}$	3.19**
5	6,785.0	$3.78x10^{-04}$	$1.91x10^{-04}$	1.97*
6	6,778.2	$3.46x10^{-04}$	$1.91x10^{-04}$	1.82
7	6,774.9	$7.36x10^{-05}$	$1.92x10^{-04}$	0.38

***= 0; **= 0.001; *= 0.01 (Statistical significance level - α).

Table 7. Results analysis for no lag to seven lag days for models with no residual inclusion.

Lag days	AIC	β	ε	t-value
0	6,588.0	$1.45x10^{-03}$	$2.54x10^{-04}$	5.71***
1	6,602.4	$6.98x10^{-04}$	$2.18x10^{-04}$	3.20**
2	6,593.7	$6.22x10^{-04}$	$2.00x10^{-04}$	3.12**
3#	6,584.8	$5.82x10^{-04}$	$1.98x10^{-04}$	2.94**
4	6,581.2	$2.71x10^{-04}$	$1.96x10^{-04}$	1.38
5	6,576.1	$5.50x10^{-05}$	$1.94x10^{-04}$	0.29
6	6,569.6	$6.54x10^{-05}$	$1.93x10^{-04}$	0.34
7	6,556.6	$-1.94x10^{-04}$	$1.94x10^{-04}$	-1.00

***= 0; **= 0.001; *= 0.01 (Statistical significance level - α); # = the better model

Table 8. Results analysis for no lag to seven lag days for models including residuals.

5.4 Relative risk analysis

To analyze and estimate the PM_{10} impact on Sao Paulo's population health, the relative risk for the model considering the effects of three lag days including residuals was calculated. The expression that represents it is given by:

$$RR(x) = e^{0.000582x}. \tag{16}$$

The relative risks where calculated according to Equation (16). The plot of it against PM_{10} concentration is shown in Fig. 4.

In Fig. 4 it can be seen that the RR has a linear relation with PM_{10} concentration, then the greater the PM_{10} concentration, the higher the RR. Thus, when the PM_{10} concentration increases from 10 to 100 $\mu g/m^3$, the RR increases 5%. It may mean someone exposed to a concentration ten times greater has 5% more chance of getting a respiratory disease.

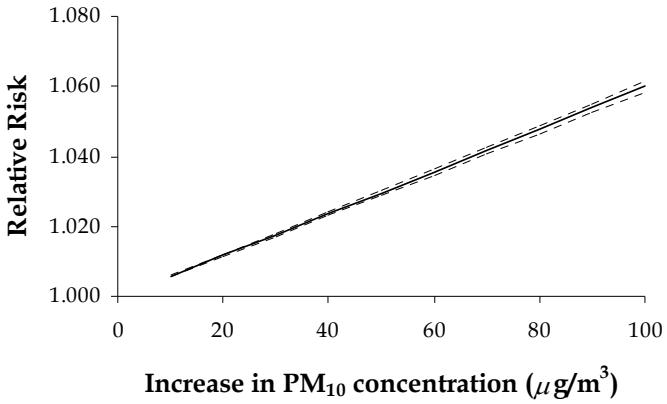


Fig. 4. Estimates of relative risk for the model considering the effect of three days lag and including residuals according to the increase in PM_{10} concentration (the dashed lines are the confidence interval).

6. Conclusion

Concluding, previous studies have not found yet a single model that can explain the impact of all kinds of air pollution on human health. Lipfert (1993) made a comparison among approximately 100 studies involving air pollution and demands for hospital services and concluded that this comparison is hampered due to the diversity encountered. The studies vary in design, diagnoses studied, air pollutant investigated; lag periods considered and the ways in which potentially confounding variables are controlled.

Lipfert (1993) also concluded that study designs have evolved considerably over the 40 years of published findings on this topic. The early studies tended to emphasize the need to limit the populations studied to those living near air pollution monitors, but more recent studies employed the concept of regional pooling, in which both hospitalization and air monitoring data are pooled over a large geographic area.

In this chapter it was emphasized times series studies appliance using Generalized Linear Models (GLM) with Poisson regression. This model is often used when response variables are countable, which demands less time then studies of follow-up kind.

The four steps to be followed to fit GLM with Poisson regression (database construction, temporal trends adjustment, goodness of fit analysis and results analysis) were applied to a case study comprising the PM_{10} concentration impact on the number of hospital

admissions for respiratory diseases in Sao Paulo city from 2007 to 2008. The results showed that GLM with Poisson regression is useful as a tool for epidemiological studies of air pollution.

According to the case study, the model fitted well to the data as the pseudo R^2 statistic has shown good results (around 0.8 > 0.6) for all adjustments. The models without residual inclusion for effects of PM_{10} concentration of the same day (no lag) to five days later (five lag days) showed regression coefficients statistically significant, but autocorrelations for one, two, three and four lag days was identified, suggesting a correlation between data from one day until four days later even after adding variables for day of the week and holiday confounders. So, the models that fitted well to the data were those with residuals inclusions for one, two, three and four lag days for effects of PM_{10} concentration for the same day (no lag) to three days later (three lag days), but the better fit was for the effect after three days of exposure according to Akaike Information Criterion (AIC) analysis that has shown the lowest AIC (6584.8).

In this way, an analysis of the relative risk (RR) for the model with residual inclusion and considering the effects of exposures after three days (three lag days) showed that the risk of someone get sick with a respiratory disease increases 5% as the concentration goes from 10 to 100 $\mu g/m^3$.

The results of the study showed that the risk of getting sick due to PM_{10} concentration can occur up to three days after the exposure and the more concentration, the higher the risk.

Finally, the steps to apply the GLM with Poisson regression to studies of air pollution impact on human health presented in this chapter had not been found in the literature and can be extended to all air pollutants and health outcomes.

7. Acknowledgment

This chapter was developed with financial support of CNPQ (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and ANP (Agência Nacional do Petróleo).

8. References

Baxter, L.A.; Finch, S.J.; Lipfert, F.W. & Yu, Q. (1997). Comparing estimates of the effects of air pollution on human mortality obtained using different regression methodologies. *Risk Analysis*, Vol. 17, No. 3, pp. 273-278.

Bhattacharyya, G.K. & Johnson, R.A. (1977). *Statistical Concepts and Methods*, John Wiley & Sons, Inc., ISBN: 0471072044, USA.

Bickel, P.J. & Doksum, K.A. (2000). *Mathematical Statistics: Basic Ideas and Selected Topics. (Volume I)* (second edition), Pearson Prentice Hall, ISBN: 013850363X, USA.

Box, G.E.P.; Jenkins, G.M. & Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control* (third edition), Prentice-Hall, ISBN: 0470272848, USA.

Braga, A. L. F.; Conceição, G. M. S.; Pereira, L. A. A.; Kishi, H. S.; Pereira, J. C. R.; Andrade, M. F.; Gonçalves, F. L. T.; Saldiva, P. H. N. & Latorre, M. R. D. O. (1999). Air pollution and pediatric respiratory hospital admissions in Sao Paulo, Brazil. *Journal of Occupational and Environmental Medicine*, Vol. 1, pp. 95-102.

Braga, A.L.F.; Saldiva, P.H.N.; Pereira, L.A.A.; Menezes, J.J.C.; Conceição, G.M.S.; Lin, C.A.; Zanobetti, A.; Schwartz, J. & Dockery, D.W. (2001). Health effects of air pollution exposure on children and adolescents in Sao Paulo, Brazil. *Pediatric Pulmonology*, Vol. 31, pp. 106-113.

Burnett, R.T.; Cakmak, S. & Brook, J.R. (1998). The effect of the urban ambient air pollution mix on daily mortality rates in 11 Canadian cities. *Canadian Journal of Public Health*, Vol. 89, pp. 152-156.

Cetesb – Environmental Company of Sao Paulo. June 5th, 2011. Available from: <http://www.cetesb.sp.gov.br/ar/qualidade-do-ar/32-qualar>: In Portuguese.

Chapra, S. C. & Canale, R. P. (1987). *Numerical Methods for Engineers with Personal Computer Applications* (second edition), McGraw-Hill International Editions, USA.

Dobson, A.J & Barnett, A.G. (2008). *An Introduction to Generalized Linear Models* (third edition), Chapman & Hall. ISBN: 1584889500, USA.

Dockery, D.W. & Pope III, C.A. (1994). Acute respiratory effects of particulate air pollution. *Annual Review of Public Health*, Vol. 15, pp. 107-132.

Dominici, F.; McDermott, A.; Zeger, S.L. & Samet, J.M. (2002). On the use of Generalized Additive Models in time-series studies of air pollution and health. *American Journal of Epidemiology*, Vol. 156, No. 3, pp. 193-203.

Dominici, F.; Sheppard, L. & Clyde, M. (2003). Health effects of air pollution: A statistical review. *International Statistical Review*, Vol. 71, No. 2, pp. 243-276.

Everitt B. S. (2003). *Modern Medical Statistics,* Oxford University Press Inc., 0340808691, USA.

Everitt, B.S. & Hothorn, T. (2010). *A Handbook of Statistical Analyses Using R* (second edition), Chapman & Hall, ISBN: 9781420079333, USA.

Faraway, J.J. *Practical Regression and Anova using R*. (1999), June 15th, 2006. Available from: <http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>.

Hastie, T.J. & Tibishirani, R.J. (1990). *Generalized Additive Models* (first edition), Chapman & Hall, ISBN: 0412343908, USA.

Ibald-Mulli, A.; Timonen, K. L.; Peters, A.; Heinrich, J.; Wölke, G.; Lanki, T.; Buzorius, G.; Kreyling, W. G.; Hartog, J.; Hoek, G.; Brink, H. M & Pekkanen, J. (2004). Effects of particulate air pollution on blood pressure and heart rate in subjects with cardiovascular disease: A multicenter approach. *Environmental Health Perspectives*, Vol. 112, No. 3, pp. 369-377.

Lipfert, F.W. (1993). A critical review of studies of the association between demands for hospital services and air pollution. *Environmental Health Perspectives Supplements*, Vol. 101, Suppl. 2, pp. 229-268.

McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models* (second edition). Chapman & Hall. ISBN: 0412317605, USA.

Metzer, K.B.; Tolbert, P.E.; Klein, M.; Peel, J.L.; Flanders, W.D.; Todd, K.; Mulholland, J.A.; Ryan, P.B. & Frumkin, H. (2004). Ambient air pollution and cardiovascular emergency department visits. *Epidemiology*, Vol. 15, Iss. 1, pp. 46-56.

Mood, A.M.; Graybill, F.A. & Boes, D.C. (1974). *Introduction to the Theory of Statistics* (third edition), McGraw-Hill, ISBN: 0070428646, USA.

Myers, R.H. & Montgomery, D.C. (2002). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments,* John Wiley & Sons, USA.

Nelder, J. A. & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A,* Vol. 135, No. 2, pp. 370-384.

Paula, G. A. *Regression Models with Computational Support*. Sao Paulo: Math and Statistic Institute, Universidade de Sao Paulo (2004). January 25th, 2006. Available from: <http://www.ime.usp.br/~giapaula/livro.pdf>: In Portuguese.

Peng, R.D.; Dominici, F. & Louis, T.A. (2006). Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society Series A*, Vol. 169, pp. 179-203.

Peters, A.; Dockery, D. W.; Muller, J. E. & Mittleman, M.A. (2001). Increased particulate air pollution and the triggering of myocardial infarction. *Circulation*, Vol. 103, pp. 2810-2815.

Pope III, C. A.; Burnett, R. T.; Thun, M. J.; Calle, E. E.; Krewski, D.; Ito, K. & Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Journal of the American Medical Association*, Vol. 287, No. i9, pp. 1132-1142.

R Development Core Team. (2010). *R: a language and environment for statistical computing*/ R version 2.12.0. The R Foundation for Statistical Computing.

Samet, J.M.; Zeger, S.L.; Domini, F.; Curriero, F.; Coursac, I.; Dockery, D.; Schwartz, J. & Zanobetti, A. (2000a). *The National Morbidity, Mortality, and Air Pollution Study, Part I, Methods and Methodological Issues (Report No. 94, I),* Cambridge: Health Effects Institute.

Samet, J.M.; Zeger, S.L.; Dominici, F.; Curriero, F.; Coursac, I.; Dockery, D.; Schwartz, J. & Zanobetti, A. (2000b). *The National Morbidity, Mortality, and Air Pollution Study, Part II, Morbidity and Mortality from Air Pollution in the United States (Report No. 94, II),* Cambridge: Health effects Institute.

Samoli, E.; Nastos, P.T.; Paliatsos, A.G.; Katsouyanni, K. & Priftis, K.N. (2011). Acute effects of air pollution on pediatric asthma exacerbation: evidence of association and effect modification. *Environmental Research*, Vol. 111, pp. 418-424.

Schwartz, J.; Spix, C.; Touloumi, G.; Bachárová, L.; Barumamdzadeh, T.; Tetre, A. Le; Piekarksi, T.; Ponce de Leon, A.; Pönkä, A.; Rossi, G.; Saez, M. & Schouten, J.P. (1996). Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *Journal of Epidemiology and Community Health*, Vol. 50, Suppl. 1, pp. S3-S11.

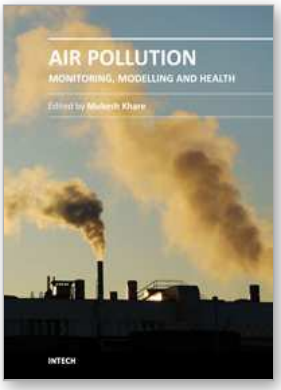
Schwarze P.E.; Totlandsdal A.I.; Herseth J.I.; Holme J.A.; Lag M.; Refsnes M.; Øvrevik J.; Sandberg W.J. & Bølling A.K. (2010). Importance of sources and components of particulate air pollution for cardio-pulmonary inflammatory responses, In: *Villanyi, V. Intech,* p. 47-74.

SUS – Health System. June 6th, 2011, Available from: <http://www2.datasus.gov.br/DATASUS/index.php?area=0701&item=1&acao=11>: In Portuguese.

Tadano, Y.S. (2007) *Analysis of PM_{10} Impact on Population's Health: Case Study in Araucaria, PR,* Parana, Brazil, 120p, Thesis (Master in Mechanical and Material Engineering), Federal University of Technology - Parana: In Portuguese.

Tadano, Y.S.; Ugaya, C.M.L. & Franco, A.T. (2009). Methodology to assess air pollution impact on the population's health using the Poisson regression method. *Ambiente & Sociedade*, Vol. XII, No. 2, pp. 241-255: In Portuguese.

Wang, P.; Puterman, M.L.; Cockburn, I. & Le, N. (1996). Mixed Poisson regression models with covariate dependent rates. *Biometrics*, Vol. 52, pp. 381-400.



Air Pollution - Monitoring, Modelling and Health

Edited by Dr. Mukesh Khare

ISBN 978-953-51-0424-7
Hard cover, 386 pages
Publisher InTech
Published online 23, March, 2012
Published in print edition March, 2012

Air pollution has always been a trans-boundary environmental problem and a matter of global concern for past many years. High concentrations of air pollutants due to numerous anthropogenic activities influence the air quality. There are many books on this subject, but the one in front of you will probably help in filling the gaps existing in the area of air quality monitoring, modelling, exposure, health and control, and can be of great help to graduate students professionals and researchers. The book is divided in two volumes dealing with various monitoring techniques of air pollutants, their predictions and control. It also contains case studies describing the exposure and health implications of air pollutants on living biota in different countries across the globe.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Yara de Souza Tadano, Cássia Maria Lie Ugaya and Admilson Teixeira Franco (2012). Methodology to Assess Air Pollution Impact on Human Health Using the Generalized Linear Model with Poisson Regression, Air Pollution - Monitoring, Modelling and Health, Dr. Mukesh Khare (Ed.), ISBN: 978-953-51-0424-7, InTech, Available from: http://www.intechopen.com/books/air-pollution-monitoring-modelling-and-health/methodology-to-assess-air-pollution-impact-on-human-health-using-the-generalized-linear-model-with-p

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821