

# Use of multivariate time series techniques to estimate the impact of particulate matter on the perceived annoyance

Milena Machado, Valdério Anselmo Reisen, Jane Meri Santos, Neyval Costa Reis Júnior, Séverine Frère, Pascal Bondon, Marton Ispany, Higor Henrique Aranda Cotta

► **To cite this version:**

Milena Machado, Valdério Anselmo Reisen, Jane Meri Santos, Neyval Costa Reis Júnior, Séverine Frère, et al.. Use of multivariate time series techniques to estimate the impact of particulate matter on the perceived annoyance. *Atmospheric Environment*, Elsevier, 2020, 222, pp.117080. 10.1016/j.atmosenv.2019.117080 . hal-02501972

**HAL Id: hal-02501972**

**<https://hal-centralesupelec.archives-ouvertes.fr/hal-02501972>**

Submitted on 20 Aug 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Use of multivariate time series techniques to estimate the impact of particulate**  
2 **matter on the perceived annoyance**

3

4 Milena Machado<sup>a\*</sup>, Valdério Anselmo Reisen<sup>bce</sup>, Jane Meri Santos<sup>c</sup>, Neyval Costa Reis  
5 Junior<sup>c</sup>, Severine Frère<sup>d</sup>, Pascal Bondon<sup>e</sup>, Márton Ispány<sup>f</sup>, Higor Henrique Aranda Cotta<sup>be</sup>

6

7 <sup>a</sup> Instituto Federal de Ciência e Tecnologia do Espírito Santo, Guarapari -E.S.- Brazil

8 <sup>b</sup> Department of Statistics, Universidade Federal do Espírito Santo, Vitoria, Brazil

9 <sup>c</sup> Department of Environmental Engineering, Universidade Federal do Espírito Santo,  
10 Vitoria, Brazil

11 <sup>d</sup> Université du Littoral Côte d'Opale, Maison de la Recherche en Science de l'homme,  
12 Dunkerque, France

13 <sup>e</sup> Laboratoire des Signaux et Systems (L2S), CNRS-CentraleSupélec-Université Paris-  
14 Sud, Gif-sur-Yvette, France

15 <sup>f</sup> University of Debrecen, Debrecen, Hungary

16 \* Tel: +55 (27) 988527717, e-mail: milenammm@ifes.edu.br

17

18 **Abstract**

19 *As well known, Particulate matter (PM) is an air pollutant that causes damage to the*  
20 *health of humans, other animals, plants, affects the climate and is a potential cause of*  
21 *annoyance through deposition on various surfaces. The perceived annoyance caused by*  
22 *particulate matter is related mainly to the increase of settled dust in urban and residential*  
23 *environments. PM can originate from many sources, i.e., paved and unpaved roads,*  
24 *buildings, agricultural operations and wind erosion represent the largest contributions*  
25 *beyond the relatively minor vehicular and industrial sources emissions. The aim of this*  
26 *paper is to quantify the relationship between perceived annoyance and particulate matter*  
27 *concentration and to estimate the relative risk (RR). The data was collected in the*  
28 *Metropolitan Region of Vitoria (MRV), Brazil. For this purpose, the variables of interest*  
29 *were modelled using vector time series model (VAR), principal component analysis*  
30 *(PCA), and logistic regression (LOG). The combination of these techniques resulted in a*  
31 *hybrid model denoted as LOG-PCA-VAR which allows to estimate RR by handling*  
32 *multipollutant effects. This study shows that there is a strong association between the*

33 *perceived annoyance and different sizes of PM. The estimates of RR indicate that an*  
34 *increase in air pollutant concentrations significantly contributes in increasing the*  
35 *probability of being annoyed.*

36

37 **Key words:** Annoyance, principal component analysis, logistic regression, relative risk.

38

## 39 **1- Introduction**

40 Particulate matter, such as dust, dirt, soot, and smoke, are environmental stressors that  
41 can cause annoyance, disturbance, stress and impairs well-being (Colls, 2002; Cox, 2000;  
42 Dockery and Pope, 1994; Farfel et al., 2005). According to Nordin and Lidén (2006),  
43 perceived annoyance can be considered as a community problem even if only a small  
44 proportion of the population is annoyed on sparse occasions. The World Health  
45 Organization (WHO, 1946) defines health as a state of complete physical, mental and  
46 social well-being and not merely the absence of disease.

47

48 PM is formed by particles with different composition, form and sizes: ultrafine particles  
49 (PM<sub>0.1</sub>) whose effects on human health are still poorly studied, fine particles (PM<sub>2.5</sub>) that  
50 are housed in the terminal bronchiole, inhalable particles (PM<sub>10</sub>) that penetrate the  
51 respiratory system, total suspended particles (TSP) which are represented by all particles  
52 suspended in the atmosphere (size range from 0.005µm to 100µm), and the sediment  
53 particles matter (SPM) that result from the sedimentation or deposition of particles  
54 previously suspended in the atmosphere, with different sizes and origin, that accumulate  
55 on the surfaces and cause annoyance (Holgate *et al.*, 1999).

56

57 The association between air pollutants and perceived annoyance is the subject of interest  
58 in several studies. Most of them, have considered regression models to quantify this  
59 relationship, for example, in the cases of odours (Blanes-Vidal, 2012), gases (Klaeboe  
60 *et al.*, 2000; Oglesby *et al.* (2000a), and particles (Klaeboe *et al.*, 2003; Rotko *et al.* 2002;  
61 Jacquemin *et al.*, 2007; Llop *et al.*, 2008; Klaeboe, 2008; Amundsen *et al.* 2008;  
62 Nikolopoulou *et al.*, 2011).

63

64 Klaeboe *et al.* (2000) have considered logistic regression to correlate NO<sub>2</sub> concentration  
65 and degrees of annoyance due to traffic, and they have found that people are more likely  
66 to be annoyed when they are exposed to high air pollution levels. Oglesby *et al.* (2000)  
67 have applied a linear regression model to correlate annoyance and concentration levels of  
68 NO<sub>2</sub> and PM<sub>10</sub>, and they have found significant correlations between these variables.  
69 Rotko *et al.* (2002) have compared exposures to PM<sub>2.5</sub> and NO<sub>2</sub> concentrations and  
70 perceived annoyance using a linear regression model, and they have observed a high  
71 correlation between personal 48h-PM<sub>2.5</sub> and 48h-NO<sub>2</sub> concentrations exposure and  
72 perceived annoyance at home. Jacquemin *et al.* (2007) have applied a linear regression,  
73 and they have found a strong positive correlation between the PM<sub>2.5</sub> concentration and  
74 perceived annoyance reported by people. Amundsen *et al.* (2008) have quantified  
75 exposure–response relationships between perceived annoyance and PM<sub>10</sub>, PM<sub>2.5</sub> and NO<sub>2</sub>  
76 concentrations, and they have observed a significant correlation between these variables.  
77 Nikolopoulou *et al.* (2011) have used a logistic regression model to correlate air quality  
78 perception of pedestrians and PM<sub>1-10</sub> concentration measured on sidewalks close to  
79 streets, and they have found a positive correlation in this study.

80  
81 Note that, the above-mentioned studies have applied simple linear regression and logistic  
82 regression but have not considered a synergistic effect among pollutants and perceived  
83 annoyance. As pointed out by Vanhatalo *et al.* (2016), Souza *et al.* (2018) among others,  
84 this analysis becomes very restrictive and may lead to biased regression estimates because  
85 air pollutants covariates are physically and statistically correlated phenomena. In  
86 addition, to estimate any multiple regression model without considering the multi-  
87 collinearity, the parameter estimates may lead to a spurious model. One way to mitigate  
88 the multi-collinearity problem is to apply principal component analysis (PCA). However,  
89 as pointed out by Zamprogno *et al.* (2019), to use PCA technique the variables have to  
90 be uncorrelated in time.

91  
92 As well known, the air pollutants concentrations are time series and they can't be assumed  
93 to be temporally uncorrelated. Thus, it is necessary to use the autocorrelation (ACF) and  
94 partial autocorrelation (PACF) functions of the pollutants to identify the existence of  
95 serial correlation, and to apply a Vector Autoregressive Model (VAR) as a filter to  
96 mitigate the temporal correlation in the covariates.

97  
98 In this context, this paper proposes a combination of multivariate statistical techniques to  
99 investigate the joint effect of different sizes of particulate matter to the perceived  
100 annoyance. Thus, the combination of the statistic tools LOG model, PCA and time series  
101 analysis can lead to an estimate of the relative risk of perceived annoyance by handling  
102 multipollutant effects. The relative risk is usually the parameter of interest to measure the  
103 impact of the covariates, especially the air pollutants on the population health (Zou,  
104 2004). The proposed methodology results in a model called LOG-PCA-VAR. To our  
105 knowledge, this is the first work which uses logistic regression with PCA and multivariate  
106 time series models to quantify the relationships between particulate matter (PM<sub>10</sub>, TSP  
107 and SPM) and perceived annoyance to estimate the relative risk (RR), which is the ratio  
108 of the probability of an outcome in an exposed group to the probability of an outcome in  
109 an unexposed group. In the air pollution problems, it is usually to measure the impact of  
110 atmospheric pollutants on the health of the exposed population see, for example, (Martin  
111 *et al.*, 1987).

112  
113  
114

## 2- Material and methods

### 115 2.1. Metropolitan Region of Vitoria

116  
117 The Metropolitan Region of Vitoria (MRV) is located on the east coast of Brazil, in the  
118 state of Espirito Santo (Figure 1). MRV is a densely populated region, with 1,500,000  
119 inhabitants and it is a highly industrialized and expanding urban region with various air  
120 pollutants emission sources such as steel, pelletizing, mining, cement industries, vehicles,  
121 road re-suspension, port and airport operations, and construction (Santos *et al.*, 2017).

122 In the MRV area, there is an interest to investigate the impact caused by PM due to  
123 population reports of being constantly annoyed (approximately 25% of the complaints to  
124 environmental agency in 2008 are about air pollution), specially by the amount of dust in  
125 surfaces (Souza, 2014; Melo *et al.*, 2015). Recently, Machado *et al.* (2018) have  
126 developed a survey where showed that, in the MRV, more than 90% of the respondents

127 have complained about perceived annoyance caused by the air pollution and, the most of  
128 these complaints were related to the amount of dust in their houses.

129

## 130 **2.2. The particulate matter data**

131 In the MRV area the weather conditions and the air quality are monitored via two  
132 complementary sets of monitoring network stations: automatic air quality monitoring and  
133 the manual SPM monitoring. Figure 1 shows the map of the urbanized area divided by  
134 municipality (Cariacica, Serra, Viana, Vila Velha e Vitoria), the main roads, the main  
135 industrial sources of PM (point red) and the air quality monitoring stations networks (blue  
136 points). They are: (M1) Laranjeiras, (M2) Carapina, (M3) Jardim Camburi, (M4)  
137 Enseada, (M5) Vitória, (M6) Vila Velha, (M7) Ibes, (M8) Cariacica. The coverage areas  
138 are 1.5 km around of each air quality monitoring station.

139 The monitoring station networks are managed by the local environmental agency (IEMA)  
140 that measure automatically hourly concentrations of different pollutants, specifically the  
141 PM<sub>10</sub> (particulate matter less than 10µg/m<sup>3</sup>) and TSP (total suspend particles). The SPM  
142 (sediment particulate matter) are measured monthly only. Therefore, for a coherence  
143 analysis, the maximum mean of PM<sub>10</sub> and TSP concentrations were also monthly  
144 computed and used in the regression model.

145 The datasets used are the flow of monthly average sediment particulate matter (SPM) as  
146 well as monthly maximum and average values of particulate matter (PM<sub>10</sub>) and total  
147 suspended particle (TSP) from the eight air quality monitoring stations measured during  
148 3 years (from July 11 to July 2014).

149

## 150 **2.3. The Survey**

151

152 Measurements of PM and perceived annoyance were performed monthly from July 11 to  
153 July 2014. Perceived annoyance was collected in two steps: face-to-face interview to the  
154 first contact with respondent and monthly telephone updates (panel survey). The face-to-  
155 face interviews randomly selected surrounding 1.5 km of each air-quality monitoring  
156 station (Figure 1). On the face-to-face interview the respondent confirmed in continuing  
157 the interviews in the following months (panel survey) about perceived annoyance (details  
158 in Machado, 2018).

159 The monthly panel survey questionnaire only included two questions were applied to 220  
160 respondents (over 16 years old) from July 11 to July 2014. Telephone questions aimed at  
161 monitoring the evolution of perceived annoyance over time-related to PM in the  
162 environment.

163 To quantify the perceived annoyance, categorical and numerical scales were considered  
164 and applied according to the context of the question (for example, “*Do you feel annoyed*  
165 *by dust during this last month?*” With the categorical answers option: *not annoyed,*  
166 *slightly annoyed, moderate annoyed, very annoyed, extremely annoyed* and “*do not*  
167 *know*”. And a second question with a numerical scale: “*What is the score that represents*  
168 *your perceived annoyance last month? from 1 to 10 points scales, where 1 is not annoyed*  
169 *and 10 is extremely annoyed.*”). These questions were formulated based on the following  
170 studies Rotko *et al.* (2002), Klæboe, (2008) and Amundsen *et al.* (2008).

171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218

From these questions, the average levels of perceived annoyance reported by all respondents was calculated. The results were dichotomized to be used as the dependent variable in the logistic regression model discussed in Section 2.4. The cut-off sample score of the perceived annoyance was the median 7, i.e., the scores levels of perceived annoyance attributed high scores ( $\geq 7$ ) was codified by 1 while the average levels of annoyed reported low scores ( $< 7$ ) was codified as 0. Similar approach was used by Rotko *et al.* (2002), Egondi *et al.* (2013) and Whittle *et al.* (2014).

## 2.4. Statistical Techniques

As previously mentioned, the main objective of this paper is to quantify the association between perceived annoyance (response) and pollutants (covariates) variables using data observed in the Metropolitan Region of Vitoria (MRV). The response variable is binary. Therefore, the logistic regression becomes the appropriate regression method to describe the association among variables. However, for this statistic model, some assumptions are required, and, among them, the covariates should be independent from each other and independent of time. And, the air pollutants do not follow these assumptions. From this matter raised one of the main contribution of this papers which is to proposed a hybrid logistic regression model (LOG\_VAR\_PCA) to quantify the association between the perceived annoyance and pollutant variables using the data set referred in the previous section.

Since the covariates (air pollutants) are time series, the use of time series models can help to understand the dynamic of the data and, additionally, to give a more precise statistical support in quantifying and discussing the association between particulate matter concentrations and perceived effects (Schwartz *et al.*, 2000, Gouveia *et al.*, 2004).

Multivariate techniques are also required for the purpose of this paper as justified as follows. To analyse the perceived annoyance caused by particulate matter a joint analysis of sediment particulate matter (SPM), particulate matter (PM<sub>10</sub>) and total suspended particles (TSP) is required. In this context, an analysis of the multivariate data set will be performed without simply isolating the effects of a single pollutant.

Since the covariates are time series and cross-correlated, the data requires a prior treatment using principal component analysis, see Zamprogno *et al.* (2019), Souza *et al.* (2018) Vanhatalo *et al.* (2016) and reference therein. Although the components obtained from PCA are not correlated, they can also present autocorrelation, which is transferred to the residuals of the fitted model. Thus, in this work, data are filtered through a multivariate time series model (the VAR model see, for example, Wei (2006)) before applying the PCA technique, as suggested by Souza *et al.* (2018) and Zamprogno *et al.* (2019). The models and techniques are summarized in the next subsections.

### 2.4.1 The Logistic Regression model

In many practical situations, the response variable in a regression model is categorical, for example, when the variable is binary, indicating the presence or absence of a

219 characteristic. Therefore, the logistic regression model becomes an important statistical  
 220 tool to measure and quantify the relationship between perceived annoyance and a set of  
 221 explanatory variables (particulate matter).

222 The logistic regression model and its parameter estimates are summarized. For more  
 223 details see, for example, Abraham and Ledolter (2006).

224 Let  $\mathbf{X} = (X_1, X_2, \dots, X_p)^t$  be a vector containing  $p$  explanatory variables. Suppose that the  
 225 response variable  $Y$  is dichotomic (binary), that is,  $Y = 1$  or  $Y = 0$  for the outcome to  
 226 be success or failure, respectively. Let the probability of  $Y$  to have success or failures,  
 227 with respect to  $\mathbf{X}$ , be defined as  $P(Y = 1|\mathbf{X}) = \pi(\mathbf{X})$  and  $P(Y = 0|\mathbf{X}) = 1 - \pi(\mathbf{X})$ ,  
 228 respectively.

229  
 230 For the explanatory vector  $\mathbf{X}$ , with the parameter vector  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^t$ , and the  
 231 response  $Y$ , the probability of success is parameterized as  
 232

$$P(Y = 1) = \pi(\mathbf{X}) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}. \quad (1)$$

233  
 234 Since this probability is a logistic function of the vector  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^t$ , it can be  
 235 shown that the logit of the multiple logistic regression model is given by

$$\ln\left(\frac{\pi(\mathbf{X})}{1 - \pi(\mathbf{X})}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p. \quad (2)$$

236 The parameter  $\beta_i, i = 0, \dots, p$ , are unknown and have to be estimated based on sample  
 237 data by the iteratively reweighted least squares approach. Let now  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be a sample  
 238 of observations of the vector of covariates  $\mathbf{X}$  and  $Y_1, \dots, Y_n$  are the corresponding  
 239 response variables. It can be shown that the vector parameter  $\boldsymbol{\beta}$  can be estimated by

$$\hat{\boldsymbol{\beta}} = (\mathbf{P}'\widehat{\mathbf{W}}\mathbf{P})^{-1}\mathbf{P}'\widehat{\mathbf{W}}\mathbf{Z}, \quad (3)$$

240 where the matrix  $\mathbf{P}$  is the matrix of regressors which has one in the first column for the  
 241 intercept parameter and  $\widehat{\mathbf{W}}$  is a diagonal matrix of dimension  $n \times n$  with elements given  
 242 by  $\hat{\pi}_i(1 - \hat{\pi}_i), i = 1, \dots, n$ , where  $\hat{\pi}_i$  have to be estimated using the maximum  
 243 likelihood method based on sample data,  $\mathbf{Z}$  is a  $n \times 1$  matrix which elements are

$$Z_i = \ln\left\{\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right\} + \left\{\frac{Y_i - \hat{\pi}_i}{\hat{\pi}_i(1 - \hat{\pi}_i)}\right\}. \quad (4)$$

244 It can be demonstrated that

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{P}'\widehat{\mathbf{W}}\mathbf{P})^{-1} \quad (5)$$

245 Regarding to Equations (4) and (5) it is possible to identify a problem that may occur: the  
 246 multicollinearity. The exact multicollinearity occurs when the matrix of covariates is not  
 247 a full rank matrix, i.e., when the maximal number of linearly independent columns of  $\mathbf{P}$   
 248 is less than the number of columns. Hence, the determinant of the matrix  $(\mathbf{P}'\widehat{\mathbf{W}}\mathbf{P})^{-1}$  is 0  
 249 and the matrix is not invertible.

250 This problem can be seen by writing  $\widehat{W} = \widehat{W}^2 \widehat{W}^1$  and  $L = \widehat{W}^2 P$  then

$$\text{Var}(\widehat{\beta}) = (L'L)^{-1}. \quad (6)$$

251 It can be shown that  $\text{rank}(L) = \text{rank}(P)$ , where  $\text{rank}(\cdot)$  denotes the operator which  
252 counts the quantity of linear independent lines. Therefore, if  $P$  has not full rank or its  
253 columns are very close to being linearly dependent (highly correlated), this will have an  
254 effect on  $(L'L)^{-1}$  matrix, thus, affecting the estimated parameters (Lutkepohl, 1991).

255

#### 256 2.4.2 Principal Component Analysis

257

258 As well known, Principal Component Analysis (PCA) is a multivariate statistical  
259 technique that aims, in general, to reduce the dimensionality of a data matrix space  
260 through linear transformations of the original variables.

261 In this study, the PCA technique is used to circumvent the problem of pollutants that are  
262 correlated with each other, i.e., the multicollinearity phenomenon. In general, the whole  
263 variability of a system determined by  $p$  variables can only be explained using all the  $p$   
264 principal components. However, a large part of this variability can be explained using a  
265 lower number  $r$  of components ( $r < p$ ) see for example, Johnson and Wichern (2007).

266 As mentioned before, the use of PCA requires attention regarding the covariates that are  
267 correlated in time (serial correlation) as it is the case of air pollutants. The time correlation  
268 of the vector  $X$  will lead to PCs auto-correlated and cross-correlated in time. As pointed  
269 by Souza *et al.* (2018) and Zamprogno *et al.* (2019), the effect of time correlation in  
270 atmospheric pollutants strongly influences the estimates of the principal components,  
271 increasing the total variability of the data and increasing the retained variability of the  
272 first component. This can be mitigate using a multivariate time series to filter the data, as  
273 suggested in Souza *et al.* (2018) and Zamprogno *et al.* (2019).

274 In Equation (2), the vector  $X$  will be the PCA variables generated from the sample  
275 covariance matrix of the filtered pollutants using a multivariate autoregressive time series  
276 model of order 1 (VAR(1)) (see, for example, Wei (2006)).

277

278 This is addressed in the Result and discussion Section. More details of the use of PCA  
279 in regression models can be recently found in Souza *et al.* (2018) Zamprogno *et al.*  
280 (2019), Hu and Tsay (2014) and Roberts and Martin (2006).

281

#### 282 2.4.3 Relative Risk

283

284 The relative risk (RR) is frequently used in epidemiological studies to measure the impact  
285 of atmospheric pollutant concentrations on the health of the exposed population. The RR  
286 can be defined as the association that an effect (annoyance) can occur following a certain  
287 exposure to a risk factor, which corresponds to the exposure to particulate matter  
288 concentration levels in this study. The relative risk is used in data analysis with binary  
289 outcomes (0 or 1) as in the case of annoyance. According to Bishop (2007) the relative



290 risk is the result of dividing the probability of the event (being annoyed when exposed –  
 291  $A|B$ ) by the probability of the event (being annoyed when not exposed –  $A|B^C$ ), i.e.:

$$RR(A, B) = \frac{P(A|B)}{P(A|B^C)} \quad (7)$$

292

293 According to Baxter (1997), by analogy, the relative risk function at level  $x$  of the desired  
 294 pollutant, denoted  $RR(x)$ , is defined as:

$$RR(x) = \frac{E(Y|X = x)}{E(Y|X = 0)} \quad (8)$$

295

296 It is the ratio of the expected value of the response variable at level  $x$  of the independent  
 297 variable to the expected of the response if the independent variable was 0.

298 In this context, for the logistic regression, it can be shown that the RR can be estimated  
 299 by

$$\widehat{RR}(x_i) \approx e^{x_i \widehat{\beta}_i} \quad (9)$$

300

301 where  $x_i$  is the interquartile variation (3st quantile - 1st quantile from Table 1) in the  $i$ th  
 302 pollutant concentration and  $\widehat{\beta}_i$  is represented by:

$$\widehat{\beta}_i = \sum_{j=1}^r \widehat{\alpha}_{ij} \widehat{\gamma}_j \quad i = 1, 2, \dots, p, \quad (10)$$

303 where  $\widehat{\alpha}_j = (\widehat{\alpha}_{ji})$  is the  $j$ -th estimated eigenvector of the covariates matrix (from Table  
 304 3);  $\widehat{\gamma}_j$  is the estimated coefficient of the  $j$ -th PC calculated in the logistic regression (from  
 305 Table 4). Through the coefficient  $\widehat{\beta}_i$  it is computed the individual contribution of each  
 306 pollutant to the perceived annoyance see, for example, Souza *et al.* (2018).

307

## 308 1- Results and discussion

309 Table 1 presents the descriptive statistics (minimum, maximum, average and standard  
 310 deviation) of the pollutants monthly measured in the Vitoria region from 2011 to 2014.  
 311 Note that, the maximum particulate matter concentrations observed for  $PM_{10}$  and TSP  
 312 pollutants can be very dangerous for the health system since its values are above the limits  
 313 set by the World Health Organization (WHO, 2006). The maximum value for SPM is also  
 314 higher than the annoyance standard values considered in many countries see, for example,  
 315 (Vallack and Shilitto, 1998; Melo *et al.*, 2018).

316

317 In the standard regression model, the basic assumption is that the covariates are not  
 318 correlated and not time-dependent. However, in the case studied here, the predicable  
 319 variables do not satisfy these properties, since the pollutant variables are serially and time

320 dependent. As shown in Table 2, the pollutants are contemporaneously correlated, for  
321 example, the sample correlation between SPM x PM<sub>10</sub> is  $\hat{\rho}_{SPM,PM_{10}} = 0.424$ . The  
322 pollutants are time series, and their behaviours over time are displayed in Figures 2 to 6.  
323 These figures show the monthly data time series of each air pollutant (particles deposition  
324 rate, monthly averages of PM<sub>10</sub> and TSP, monthly maximum averages of PM<sub>10</sub> and TSP)  
325 from July 2011 to October 2014. These also display the sample autocorrelation (ACF)  
326 and partial autocorrelation (PACF) functions which clearly show that the pollutants are  
327 time-dependent. In the ACF and Partial ACF plots (Figures 2-11), the vertical axis  
328 measures the strength of the correlation and the horizontal axis is the time lag at which  
329 the correlation was calculated. The dashed lines represent the 95% confidence intervals  
330 for uncorrelated data.

331 The sample ACF measures the dependence between the observations of the same time  
332 series at different delays, usually denoted as lags in time series methods. Figures 7 to  
333 11 show that the VAR (1) removed the time correlations. From these, it appears that the  
334 series have a very weak yearly seasonality. However, it should be noted that the seasonal  
335 yearly effect (if any) may be reduced by the smoothing of the monthly mean average of  
336 the pollutants PM<sub>10</sub> and TSP.

337 Since the covariates do not meet the regression basic assumption, one way to mitigate the  
338 problem is to remove the time correlation (serial-correlation) of the series. In this context,  
339 it is suggested here to use a linear time series filter as a procedure to transform the data  
340 into a “white noise” process. This problem and how to mitigate it are well-addressed in  
341 the recent publications Souza *et al.* (2018), Vahatalo and Kulahci (2016), and Zamprogno  
342 *et al.* (2019).

343 Based on the sample ACF plots, the residual analysis and the Akaike information  
344 criterion (AIC), which is an estimator of the relative quality of statistical models for a  
345 given set of data, a Vector Autoregressive Model of order 1, denoted by VAR (1), was  
346 chosen to model the vector of all pollutants time series (particles deposition rate, monthly  
347 averages of PM<sub>10</sub> and TSP, monthly maximum averages of PM<sub>10</sub> and TSP). The sample  
348 ACF plots of the filtered data are displayed in Figures 7 to 11. From these plots, it can be  
349 seen that the time-correlation of the series was removed, and the filtered data displays a  
350 similar behaviour of a white noise process, that is, the correlations of the residuals are  
351 nulls. In addition, the residuals do not show any anomaly (results are available upon  
352 request). Therefore, this indicates that the VAR (1) model well-fitted the data. For a more  
353 details of multivariate linear time series models see, for example, Wei (2006).

354 Table 3 displays the results of the PCA technique applied to the filtered series. The total  
355 cumulative variance was used as a criterion for choosing the number of components  
356 resulted by the PCA. Thus, the first three components were chosen, which explain 86%  
357 of the total variability. In the PC1, the higher contributions come from TSP, PM<sub>10</sub> TSP.  
358 In the case of PC2, SP gives most of the variability and, for the PC3, PM<sub>10</sub> gives the  
359 highest contribution. The pollutants indicated by (\*) are the ones that give more  
360 contributions to the variability of the PC. For more details on PCA and its application  
361 see, for example, Cadima and Jolliffe, (1995).

362 In the multiple logistic regression model, the response variable (perceived annoyance)  
363 was associated with the covariates PC1, PC2 and PC3 resulting in the hybrid LOG-PCA-  
364 VAR fitted model and its parameter estimates are in Table 4.

365 The relative risk (RR) of annoyance results were expressed by the interquartile variation  
366 range. The RR analysis was performed for different levels of pollutants concentrations to  
367 test the null hypotheses  $H_0: RR = 1$  against  $H_1: RR > 1$ , using significance level of 5%.  
368 For each pollutant, Table 5 displays the results of the estimates of RR and the respectively  
369 confidence interval (CI), for the standard and the proposed methodology, that is,  $\widehat{RR}^*$   
370 refers to the estimated RR using the standard logistic regression, and  $\widehat{RR}$  corresponds to  
371 RR estimate based on the LOG-PCA-VAR model. Note that, the  $\widehat{RR}^*$  was considered  
372 in the study for comparison purpose, that is, to quantify (if any) the impact on the RR  
373 when the multivariate time series properties (multicollinearity and time and cross-  
374 correlation structures) of the covariates are ignored.

375 According to Table 5, the estimate of the RR for SPM increases approximately by a factor  
376 of 1.5 considering the interquartile variation equal to  $2\text{g/m}^2$  30 days whereas, for  $\text{PM}_{10}$   
377 (monthly mean),  $\widehat{RR}$  increases by a factor of 1.6 considering the interquartile variation  
378 equal to  $5\mu\text{g/m}^3$ . In the case of TSP (monthly mean),  $\widehat{RR}$  can be interpreted as a factor  
379 that increases 2.2 when exposed to the interquartile variation equal to  $13\mu\text{g/m}^3$ . For  
380  $\text{PM}_{10}$  (monthly maximum) variable,  $\widehat{RR}$  grows by a factor of 2.4 considering the  
381 interquartile variation equal to  $8\mu\text{g/m}^3$  whereas, for the variable TSP (monthly  
382 maximum),  $\widehat{RR}$  is equal to 1.8 considering the interquartile variation equal to  $20\mu\text{g/m}^3$ .  
383 The estimated confidence intervals were calculated based on the central limit theorem as  
384 showed by Souza *et al.* (2018). The  $\widehat{RR}$  values indicate that, all pollutants contributes  
385 significantly for the increase of the probability of being annoyed with 95% of confidence.  
386 It is interesting to note that the values of  $\widehat{RR}^*$  was not significant in any case. This is not  
387 a surprising result since the temporal correlation in data was not considered in the  
388 regression model which lead to underestimating the regression parameter and inflating  
389 the intercept. Consequently, this gives a spurious result in the sense that the pollutants  
390 don't make any impact on the perceived annoyance.

391 The proposed hybrid LOG-PCA-VAR model, in addition to the estimation of the impact  
392 of particulate matter on the perceived annoyance, which indicated significantly  
393 contribution of the pollutant to this response variable, it contributed to show the spurious  
394 result when the temporal correlation structure in the data is not considered to obtain the  
395 estimates of a logistic regression model. This corroborates the use of the proposed  
396 methodology when dealing with regression models in which the covariates are  
397 multivariate time series and all results are in accordance with Souza *et al.* (2018).

398  
399

## 2- Conclusion

400 This study proposes the application of multivariate statistical techniques (time series  
401 models, principal component analysis and logistic regression) to estimate the effect  
402 between exposure to particulate matter concentrations (SPM,  $\text{PM}_{10}$  and TSP) and  
403 response of the population measured by the perceived annoyance levels.

404 The descriptive and graphical analysis motivated the use of the PCA technique for the air  
405 pollutant data by the initial indication of cross-correlation between the covariates  
406 (pollutants). The VAR(1) model was used to transform the original time series of air  
407 pollutants, resulting in time uncorrelated data (white noise) before applying the PCA  
408 technique. Based on these modelling steps, the PCA variables becomes uncorrelated and  
409 not cross-correlated.

410 The logistic regression model was applied with the level of annoyance as the dependent  
411 variable and the air pollutants as covariates. Moreover, by the new methodology  
412 developed in this study (*LOG-PCA-VAR*), the combined effect of particulate matter was  
413 analysed and the relative risk of annoyance for each original air pollutants was calculated.  
414 The estimates of relative risk, i.e.,  $\widehat{RR}$ , showed that, in general, an increase in air pollutant  
415 concentrations (i.e., the particulate matter metrics examined here: TSP, PM<sub>10</sub> and SPM)  
416 significantly contributes in increasing the probability of being annoyed.

417 In summary, the results obtained in this study provide evidence of a significant correlation  
418 between particulate matter and perceived annoyance levels, also indicating that, at least  
419 for particulate matter, perceived annoyance is not only related to one pollutant but to a  
420 group of pollutant. In future work, this methodology should be used to analysis with other  
421 pollutants. Other methodologies, such as bootstrap techniques, could also be used to  
422 estimate the confidence intervals more precisely, and GLARMA modelling could be used  
423 to solve the data autocorrelation problem.

424  
425  
426

### 3- Acknowledgements

427  
428 The authors would like to thank the anonymous reviewers for their helpful and  
429 constructive comments that greatly contributed to improving the final version of the  
430 manuscript. The results in this paper were part of the PhD thesis of the first author under  
431 supervision of Valderio A. Reisen and Jane M. Santos, at PPGEA-UFES, Brazil, 2015  
432 (Machado 2015). The authors would like to thank CNPq, CAPES and FAPES for their  
433 financial support. Part of this paper was revised when Valdério Reisen, Márton Ispány  
434 and Milena Machado were visiting CentraleSupélec in July 2018, January and July 2019.  
435 These authors are indebted to CentraleSupélec and Université Paris-Sud for their financial  
436 supports. This research was also partially supported by the iCODE Institute, research  
437 project of the IDEX Paris-Saclay, and by the Hadamard Mathematics LabEx (LMH)  
438 through the grant number ANR-11-LABX-0056-LMH in the Programme des  
439 Investissements d'Avenir. The work of Márton Ispány is supported by the EFOP-3.6.1-  
440 16-2016-00022 project. The project is co-financed by the European Union and the  
441 European Social Fund.

442

### 6- References

443

- 444
- 445 1) Abraham B. & Ledolter J. Introduction to Regression Modeling. Thomson  
446 Brooks/Cole, 2006.
  - 447 2) Amundsen A.H., Klaeboe R. & Fyhri A. (2008). Annoyance from vehicular air  
448 pollution: Exposure–response relationships for Norway. Atmospheric  
449 Environment, 42, 679-688.

- 450 3) Bishop, Y., Fienberg, S., Holland, P. (2007). *Discrete Multivariate Analysis:*  
451 *Theory and Practice.* Cambridge: MIT, 575 p.
- 452 4) Baxter L., Finch S., Lipfert F., and Yu Q. (1997). Comparing estimates of the  
453 effects of air pollution on human mortality obtained using different regression  
454 methodologies. *Risk Analysis*, 17, 273–278.
- 455 5) Blanes-Vidal, V., Suh H., Nadimi E. S., Løfstrøm P., Ellermann T., Andersen H.  
456 V., Schwartz J., (2011) Residential exposure to outdoor air pollution from  
457 livestock operations and perceived annoyance among citizens. *Environment*  
458 *International*, 40, 44-50.
- 459 6) Cadima J., Jolliffe I.T. (1995). Loadings and correlations in the interpretation of  
460 principal components. *Journal of Applied Statistics*, 22(2), 203-214.
- 461 7) Colls, J. *Air Pollution.* 2ed. USA: SPON Press Taylor & Francis Group, 2002.
- 462 8) Cox, L. (2000). Statistical issues in the study of air pollution involving airborne  
463 particulate matter. *Environmetrics* 11, 611-626.
- 464 9) Dockery, D.W. and Pope, C.A. (1994). Acute respiratory effects of particulate air  
465 pollution. *Annual Review of Public Health*, 15, 107–132.
- 466 10) Egondi, T., Kyobutungi, C., Ng, N., Muindi, K., Oti, S., van de Vijver, S., Ettarh,  
467 R., Rocklöv, J., 2013. Community perceptions of air pollution and related health  
468 risks in Nairobi slums. *Int. J. Environ. Res. Publ. Health* 10, 4851–4868.
- 469 11) Farfel, M.R., Orlova, A.O., Lees, P.S.J., Rohde, C., Ashley, P.J., Julian Chisolm,  
470 J., 2005. A study of urban housing demolition as a source of lead in ambient dust  
471 on sidewalks, streets, and alleys. *Environ. Res.* 99, 204–213.  
472 doi:10.1016/j.envres.2004.10.005
- 473 12) Gouveia, N., Bremner, S.A., Novaes, H.M. (2004). Association between ambient  
474 air pollution and birth weight in São Paulo, Brazil. *Journal of Epidemiology and*  
475 *Community Health*, 58, 11-17.
- 476 13) Holgate, S.T., Samet, J.M., Koren, H.S., Maynard, R.L., 1999. *Air pollution and*  
477 *health.* Academic Press.
- 478 14) Jacquemin B., Sunyer J., Forsberg B., Gotschi T., Oglesby L., Ackermann-  
479 Liebrich U., De Marco R., Heinrich J., Jarvis D., Toren K., Kunzli N. 2007.  
480 Annoyance due to air pollution in Europe. *International Journal of Epidemiology*,  
481 36, 809–820.
- 482 15) Johnson, R.A., Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis.*  
483 6th edition. Prentice Hall, New Jersey, 800 p.
- 484 16) Klæboe, R., Kolbenstvedt, M., Clench-Aas, J., Bartonova, A. (2000). Oslo traffic  
485 study - part 1: an integrated approach to assess the combined effects of noise and  
486 air pollution on annoyance. *Atmospheric Environment*, 34, 4727-4736.
- 487 17) Klæboe R., Öhrström E., Turunen-Rise I., Bendsten H., Nykänen H. (2003).  
488 Vibration in dwellings from road and rail traffic – Part III: towards a common  
489 methodology for socio-vibrational surveys. *Applied Acoustics*, 64, 111–120.
- 490 18) Klæboe, R., Amundsen A.H., Fyhri A. (2008). Annoyance from vehicular air  
491 pollution: A comparison of European exposure–response relationships.  
492 *Atmospheric Environment*, 42, 7689-7694.
- 493 19) Hu Y-P., Tsay R.S. (2014) Principal volatility component analysis. *Journal of*  
494 *Business and Economic Statistics*, 32(2), 153-164.
- 495 20) Llop S., Ballester F., Estarlich M., Esplugues A., Fernández-Patier R., Ramón R.,  
496 Marco A., Aguirre A., Sunyer J., Iñiguez C., on behalf of INMA-Valencia cohort

- 497 (2008). Ambient air pollution and annoyance responses from pregnant women.  
498 Atmospheric Environment, 42, 2982-2992.
- 499 21) Lutkepohl, H. (1991). Introduction to Multiple Time Series Analysis. Springer-  
500 Verlag, Berlin.
- 501 22) Machado M., Santos J.M., Reisen V. A., Reis N.C., Mavroidis I. Lima A. T. A  
502 new methodology to derive settleable particulate matter guidelines to assist  
503 policy-makers on reducing public nuisance. Atmospheric Environment 182  
504 (2018) 242–251.
- 505 23) Martin, S.W., Meek, A.H., Willeberg, P., 1987. Veterinary epidemiology.  
506 Principles and methods. Iowa State University Press, Ames, IA, p. 343.
- 507 24) Melo, M.M., Santos, J.M., Frere, S., Reisen, V.A., Jr., N.C.R., Leite, M.F.S. de  
508 F.S., 2015. Annoyance Caused by Air Pollution: A Comparative Study of Two  
509 Industrialized Regions. World Acad. Sci. Eng. Technol. Int. J. Environ. Ecol.  
510 Eng. 2, 182–187.
- 511 25) Nikolopoulou M., Kleissl J., Linden P.F., Lykoudis S. (2011). Pedestrians'  
512 perception of environmental stimuli through field surveys: Focus on particulate  
513 pollution. Science of the Total Environment, 409(13), 2493-202.
- 514 26) Nordin S., Lidén E. (2006). Environmental odor annoyance from air pollution  
515 from steel industry and bio-fuel processing. Journal of Environmental  
516 Psychology, 26, 141–145.
- 517 27) Oglesby, L., Kunzli, N., Monn, C., Schindler, C., Ackermann-Liebrich, U.,  
518 Leuenberger, P. (2000). Validity of annoyance scores for estimation of long term  
519 air pollution exposure in epidemiologic studies: The Swiss study on air pollution  
520 and lung diseases in adults (SAPALDIA). American Journal of Epidemiology,  
521 152, 75–83.
- 522 28) Roberts S, Martin M. Using supervised principal components analysis to assess  
523 multiple pollutant effects. Environmental Health Perspectives, Vol. 116, No. 12.  
524 2006.
- 525 29) Rotko T., Oglesby L., Kunzli N., Carrer P., Nieuwenhuijsen M.J., Jantunen M.  
526 (2002). Determinants of perceived air pollution annoyance and association  
527 between annoyance scores and air pollution (PM2.5, NO2) concentrations in the  
528 European EXPOLIS study. Atmospheric Environment, 36, 4593–4602.
- 529 30) Santos, J.M., Reis Jr, N.C., Galvão, E.S., Silveira, A., Goulart, E.V., Lima, A.T.,  
530 2017. Source apportionment of settleable particles in a mining-impacted urban  
531 and industrialized region in Brazil. Environ. Sci. Pollut. Res. doi:10.1007/s11356-  
532 017-9677-y.
- 533 31) Schwartz, J. (2000). Harvesting and long-term exposure effects in the relationship  
534 between air pollution and mortality. American Journal of Epidemiology, 151(5),  
535 440- 448.
- 536 32) Stenlund, T., Lidén, E., Anderson, K., Garvill, J., Nordin, S. (2009). Annoyance  
537 and health symptoms and their influencing factors: A population-based air  
538 pollution intervention study. Public Health. Vol. 123, p. 339-345.
- 539 33) Souza, J. B., Reisen, V. A., Santos, J.M., Franco, G. C. (2014). Principal  
540 components and generalized linear modeling in the correlation between hospital  
541 admissions and air pollution. Rev Saúde Pública 48(3):451-458.

- 542 34) Souza, J. B., Reisen, V. A., Franco, G. C., Ispány, M., Bondon, P., Santos, J. M.  
543 (2018). Generalized additive models with principal component analysis: an  
544 application to time series of respiratory disease and air pollution data. *Journal of*  
545 *the Royal Statistical Society: Series C (Applied Statistics)* (67), 453-480, 2018.
- 546 35) Vallack, H., Shillito, D., 1998. Suggested guidelines for deposited ambient dust.  
547 *Atmos. Environ.* 32, 2737–2744. doi:10.1016/S1352-2310(98)00037-5
- 548 36) Vanhatalo E., Kulahci M., Impact of autocorrelation on principal components and  
549 their use in statistical process control, *Quality and Reliability Engineering*  
550 *International* 32 (2016) 1483–1500.
- 551 37) Wei, W.W.S. (2006). *Time Series Analysis: Univariate and Multivariate Methods.*  
552 Pearson Addison Wesley.
- 553 38) Whittle, N., Peris, E., Condie, J., Woodcock, J., Brown, P., Moorhouse, A.T.,  
554 Waddington, D.C., Steele, A., (2015). Development of a social survey for the  
555 study of vibration annoyance in residential environments: good practice guidance.  
556 *Appl. Acoust.* 87, 83–93.
- 557 39) WHO, 1946. Constitution of the World Health Organization as adopted by the  
558 International Health Conference, New York, 19-22 June 1946; signed on 22 July  
559 1946 by the representatives of 61 States (Official Records of the World Health  
560 Organization, no. 2, p. 100) and entered into force on 7 April 1948.
- 561 40) WHO, 2005. WHO Air Quality Guidelines for Particulate Matter, Ozone,  
562 Nitrogen Dioxide and Sulphur Dioxide. Summary of Risk Assessment. Geneva,  
563 2006.
- 564 41) Zamprogno, B., Reisen, V. A., Reis Junior, Neyval Costa, C. H. H. A., and  
565 Bondon,P. (2019). Principal component analysis with autocorrelated data (pre-  
566 print available with the authors).
- 567 42) ZOU, G. A modified Poisson regression approach to prospective studies with  
568 binary data. *American Journal of Epidemiology* 159(7), 702–706. 2004.
- 569 43)

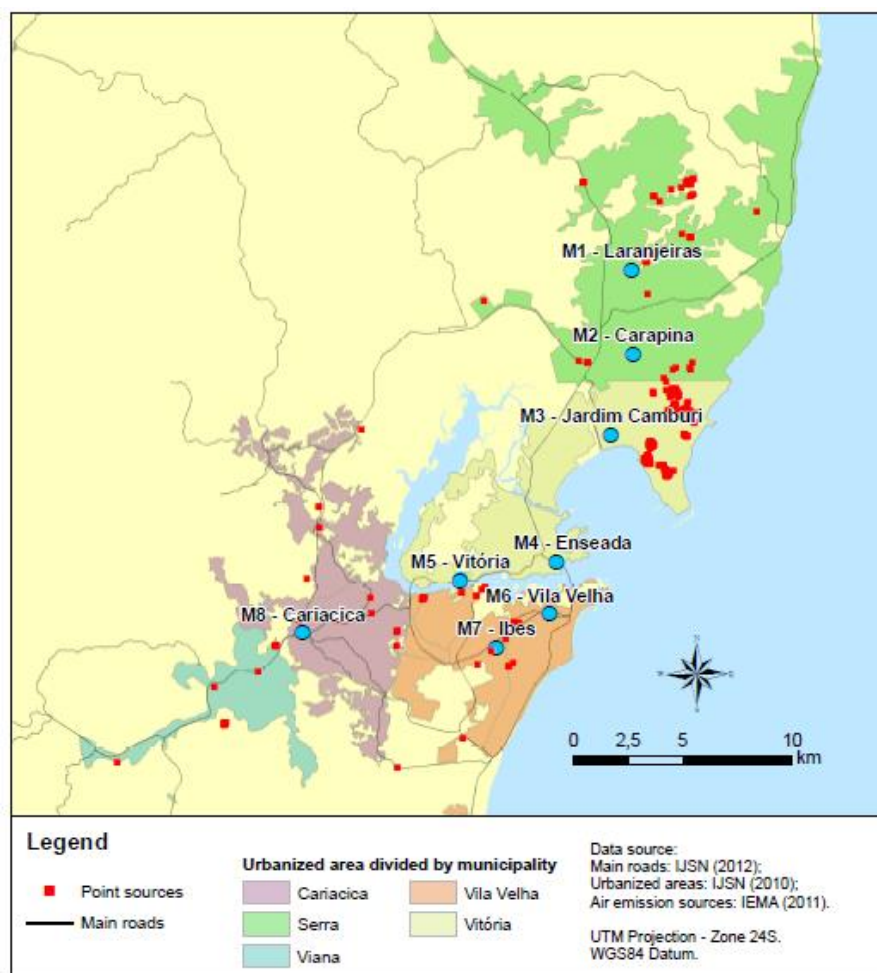


Figure 1- Metropolitan Region of Vitória, the main sources, the main roads and the air quality monitoring stations network: (M1) Laranjeiras, (M2) Carapina, (M3) Jardim Camburi, (M4) Enseada, (M5) Vitória, (M6) Vila Velha, (M7) Ibés, (M8) Cariacica.

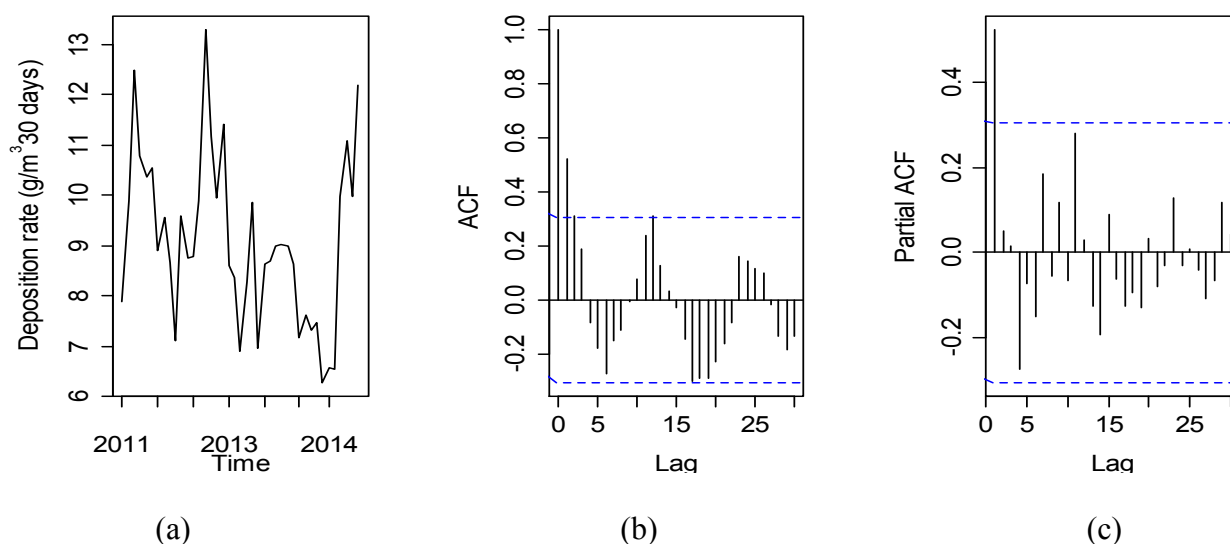
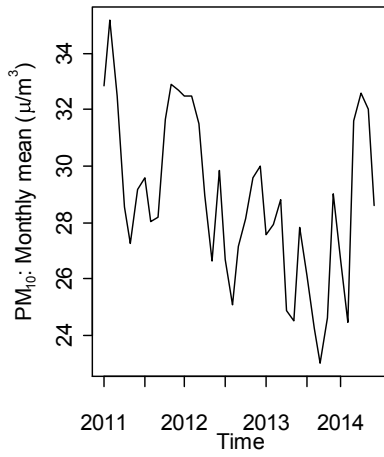
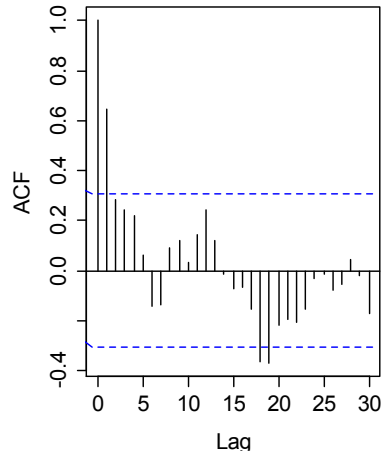


Figure 2 – Time series (a), autocorrelation function (b) and partial autocorrelation function (c) for SPM from 2011 to 2014.

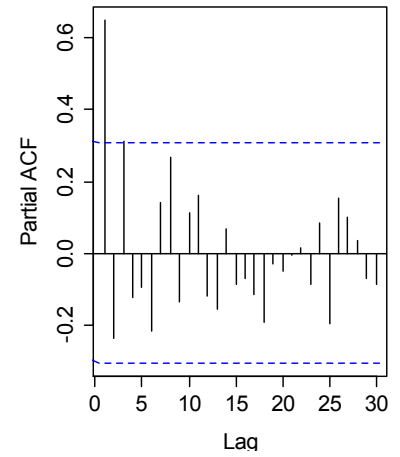




(a)

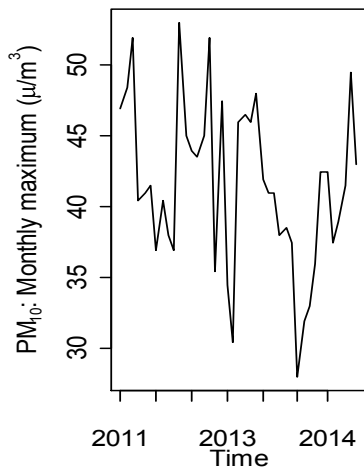


(b)

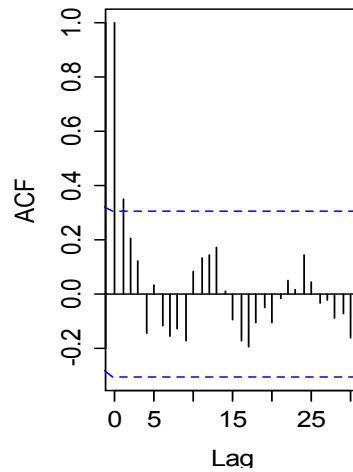


(c)

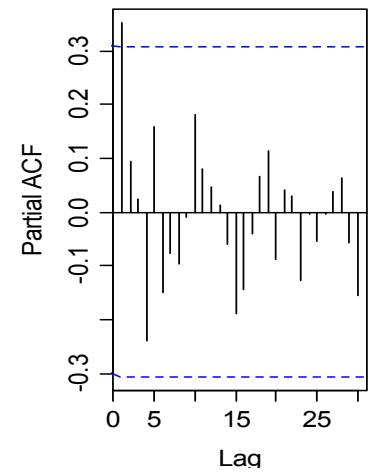
Figure 3- Time series (a), autocorrelation function (b) and partial autocorrelation function (c) for monthly mean concentration of  $PM_{10}$  from 2011 to 2014.



(a)

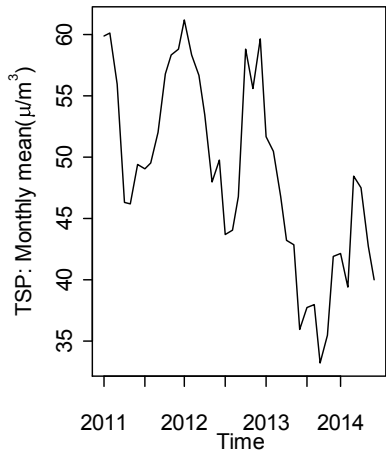


(b)

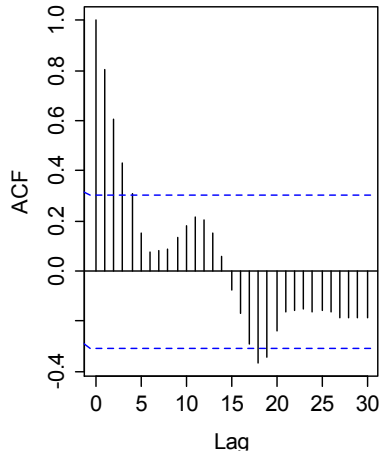


(c)

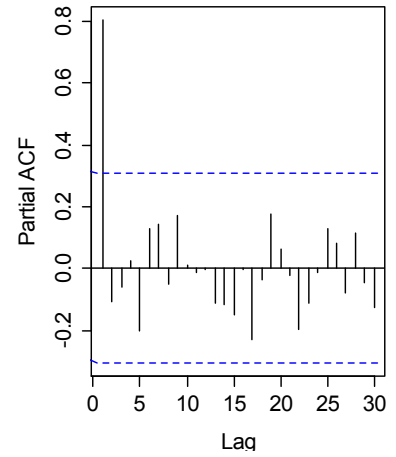
Figure 4- Time series, autocorrelation function and partial autocorrelation function for monthly maximum  $PM_{10}$  concentration from 2011 to 2014.



(a)

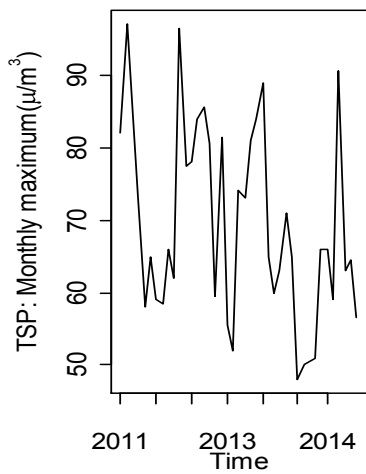


(b)

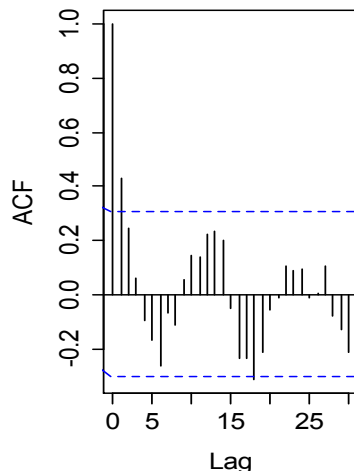


(c)

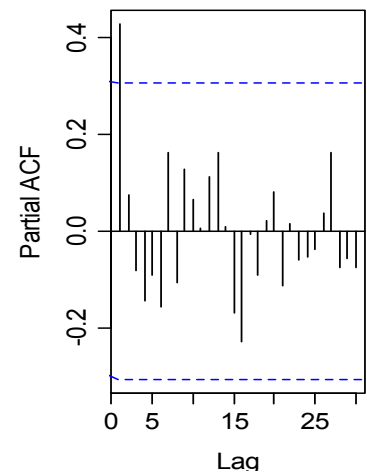
Figure 5- Time series, autocorrelation function and partial autocorrelation function for monthly mean TSP concentration from 2011 to 2014.



(a)

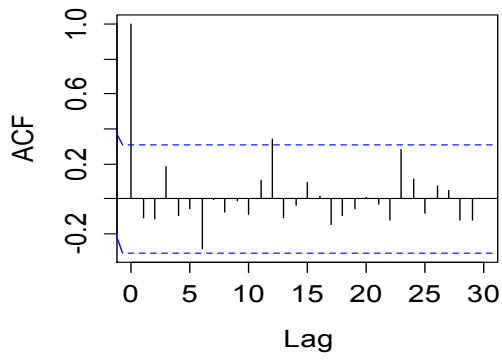


(b)

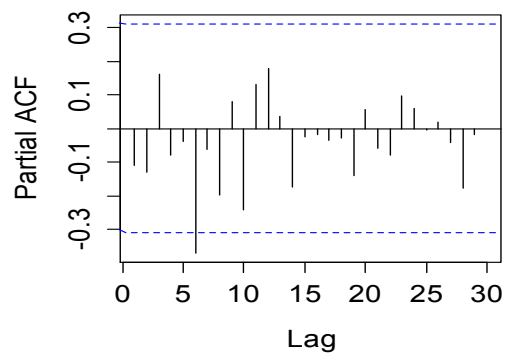


(c)

Figure 6- Time series, autocorrelation function and partial autocorrelation function for monthly maximum TSP concentration from 2011 to 2014.

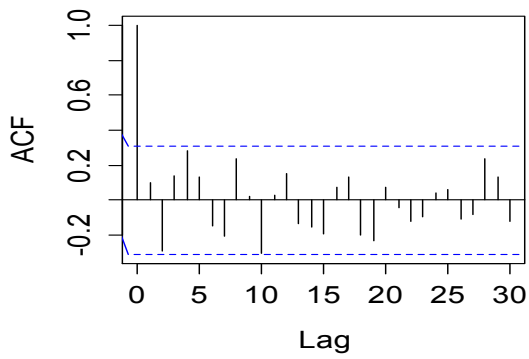


(a)

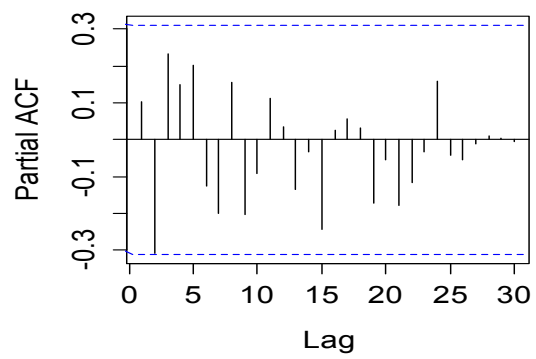


(b)

Figure 7 - Autocorrelation function (a) and partial autocorrelation function (b) for particles deposition rate from 2011 to 2014 after filtering.

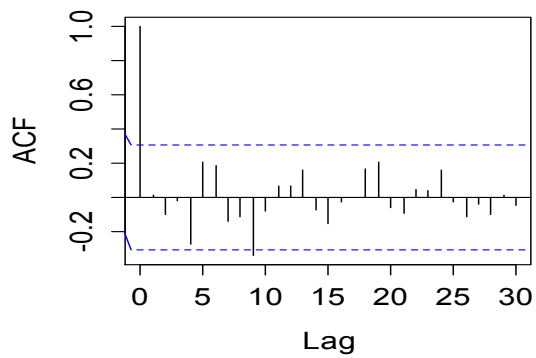


(a)

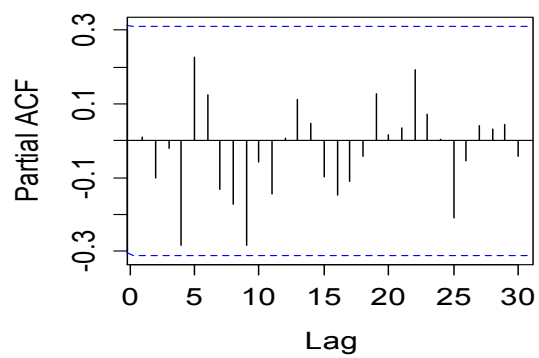


(b)

Figure 8- Autocorrelation function (a) and partial autocorrelation function (b) for monthly mean concentration of PM<sub>10</sub> from 2011 to 2014 after filtering.

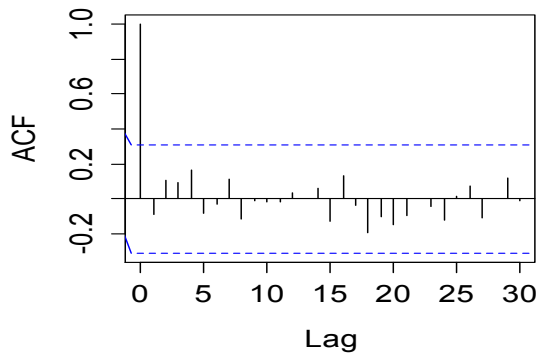


(a)

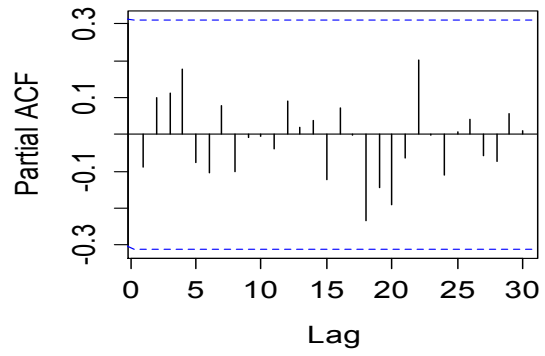


(b)

Figure 9- Autocorrelation function (a) and partial autocorrelation function (b) for monthly maximum PM<sub>10</sub> concentration from 2011 to 2014 after filtering.

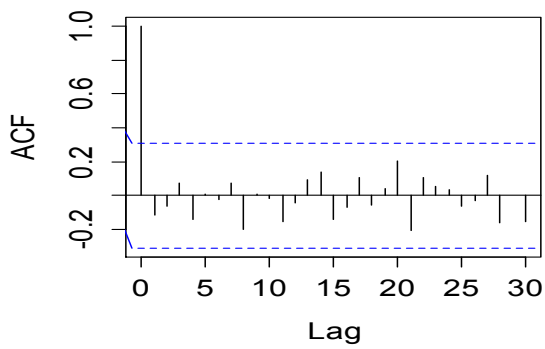


(a)

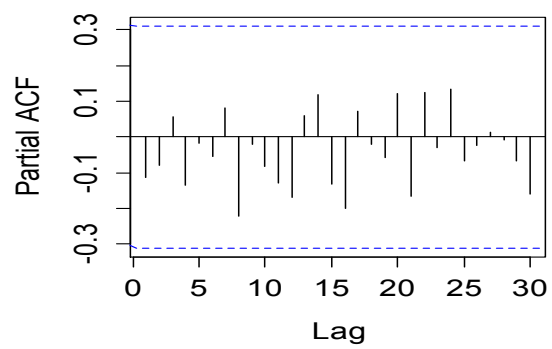


(b)

Figure 10- Autocorrelation function (a) and partial autocorrelation function (b) for monthly mean TSP concentration from 2011 to 2014 after filtering.



(a)



(b)

Figure 11- Autocorrelation function (a) and partial autocorrelation function (b) for monthly maximum TSP concentration from 2011 to 2014 after filtering.

Table 1 – Descriptive statistics of air pollutants (from July 2011 to November 2014)

Variable	Minimum	Maximum	Mean	Std. Dev.	1st quantile	3st quantile	90th percentile
SPM (g/m <sup>2</sup> 30 days)	6.267	13.283	9.097	1.680	7.683	9.969	11.173
PM <sub>10</sub> (µg/m <sup>3</sup> )	23.002	35.167	28.818	2.962	26.670	31.575	32.590
TSP (µg/m <sup>3</sup> )	33.166	61.167	48.665	7.808	42.705	55.899	58.830

Table 2 – Correlation matrix for the original variables (before time series analysis)

Variables	SPM	PM <sub>10</sub> (mean)	TSP (mean)	PM <sub>10</sub> (maxim)	TSP (maxim)
SPM	1.				
PM <sub>10</sub> (mean)	0.424**	1			
TSP (mean)	0.278	0.764**	1		
PM <sub>10</sub> (maxim)	0.409**	0.681**	0.654**	1	
TSP (maxim)	0.342*	0.701**	0.754**	0.772**	1

\*\*p-value=0,01

\*p-value=0,05

Table 3- Results of factor loadings statistics and application of PCA

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	2.576	1.071	0.681	0.396	0.276
Variability (%)	51.528	21.426	13.622	7.913	5.510
Cumulative %	<b>51.528</b>	<b>72.955</b>	<b>86.577</b>	94.490	100.000
SP (monthly rate)	0.267	0.733*	-0.554	-0.269	-0.112
PM <sub>10</sub> (monthly mean)	0.495*	-0.257	-0.365	0.674	-0.319
TSP (monthly mean)	0.400*	-0.583	-0.318	-0.607	0.172
PM <sub>10</sub> (monthly maxim)	0.492*	0.104	0.611*	-0.254	-0.557
TSP (monthly maxim)	0.531*	0.214	0.293	0.200	0.739

\*High contributions

Table 4- Parameters estimated by the multiple logistic model estimated for the first three components

	$\hat{\beta}$	Standard error	Exp( $\hat{\beta}$ )
PC1	0.053	0.202	1.054
PC2	0.058	0.309	1.060
PC3	-0.245	0.390	0.783
Intercept	0.204	0.320	-

Table 5- The estimate RR of annoyance for each pollutant and the respective interval confidence

Pollutants	$\widehat{RR}^*$	CI (95%)	$\widehat{RR}$	CI (95%)
	(standard methodology)		(LOG-PCA-VAR)	
SPM	0.865	(0.582;1.283)	1.462	(1.070; 1.854)
PM <sub>10</sub> (monthly mean)	0.819	(0.650; 1.031)	1.649	(1.061; 2.237)
TSP (monthly mean)	0.953	(0.875; 1.037)	2.181	(1.471; 2.891)
PM <sub>10</sub> (monthly maxim)	0.977	(0.877; 1.088)	2.411	(1.401; 3.421)
TSP (monthly maxim)	0.965	(0.918; 1.014)	1.822	(1.52; 3.052)