Assessment of Air quality in the Portuguese Tâmega e Sousa region

F´ATIMA DE ALMEIDA
Polytechnic of Porto
CIICESI, ESTG
Margaride Felgueiras
PORTUGAL
mff@estg.ipp.pt

ALDINA CORREIA
Polytechnic of Porto
CIICESI, ESTG
Margaride Felgueiras
PORTUGAL
aic@estg.ipp.pt

ELIANA COSTA E SILVA
Polytechnic of Porto
CIICESI, ESTG
Margaride Felgueiras
PORTUGAL
eos@estg.ipp.pt

Abstract: This paper, presents the evaluation of the air quality in several regions of Portugal, with special focus on the region of *Tâmega e Sousa* where ESTG/P.Porto is located. The ANOVA and MANOVA techniques are applied to study the differences between air quality from 2009 to 2012 in Portugal. The data used in this analysis includes altitude, area, expenditure of environmental measures on protection of air quality and climate, expenditure on protection of biodiversity and landscape, burned area, number of forest fires, extractive and manufacturing industries, per municipality and per year. A air quality indicator with five levels is proposed based on information gathered by the project *QualAr* about concentrations of the pollutants: CO, NO_2, O_3, PM_{10} and SO_2. The results point to significant differences in the air quality both for the different regions and the years considered. Furthermore, for identifying the factors that influence the air quality in 2012 a multivariate regression model was used. The results show statistical evidences that air quality in 2011, number of forest fires in 2012 and 2010, number of manufacturing industries per km^2 in 2012 and number of forest fires in 2010 are the variables that present a larger contribution to the quality of the air in 2012.

Key–Words: ANOVA, MANOVA, Multivarite Regression, Air Quality, *Tâmega e Sousa* region.

1 Introduction

Air pollution is a major environmental problem which may present a significant risk for human health [6]. During recent decades, increasing atmospheric pollution has given rise to concern that the natural cycles of the northern hemisphere may be seriously affected. In fact, concentration of 90% of the total anthropogenic emissions in certain regions of this part of the globe have been reported [2].

While in the past the major cause for poor air quality were industrial activity and domestic heating, in nowadays, the major urban air pollutants come from road trafic [5] [9]. This change is the result of the rapid increase in mobility. In fact, urban areas, characterized by high population density and economic development, present an increasing pressure in the air quality due to the resulting pollutant emissions [5].

Contemporary air pollution problems, such as the London smog episodes, have shown to have effects on human health and hence a major economic impact [3]. According to [3] the main atmospheric pollutants are sulphur dioxide, nitrogen dioxide, particulate matter, lead, carbon monoxide, benzene, ozone, polyaromatic hydrocarbons, cadmium, arsenic, nickel and mercury. The National Air Quality Strategy (NAQS) has established the health-based standards for eight specific pollutants [3].

The current Directive 2008/50/EC on ambient air quality and cleaner air for Europe establishes the need to reduce pollution to levels which minimize harmful effects on human health [6]. The European Commission (EC) defined an overall strategy through the setting of long-term air quality objectives. Therefore, to assess the air quality at a regional level, verifying the fulfillment of the limit targets and threshold values imposed by the EC directives, and to understand the causes and origin of air pollution, numerical modeling should be used [7]. Air pollution arises from a variety of diverse sources operating over different temporal scales and spatial areas. In response to this, legislative processes are attempting to tackle air pollution holistically, through the process of Air Quality Management (AQM) [3].

Aenab et. al in [2], purpose to evaluate air quality by studying the presence of heavy metals (Ni, Cu and Pb) using ANOVA two-way test (space and time) within Baghdad city. This authors found high concentrations of total suspended particles in the soils coming from industrial areas. More than 50% of the concentrations exceeded the rates for the public.

Several ozone modeling studies were already performed for the Portuguese territorial, in β-mesoscale domains and during some specific and episode days. Large-scale simulations including Portugal have also been done, but with a coarse grid resolution [6]. Regional air quality simulations for historical and future periods were conducted by Sá *et al.* [8] to investigate the impacts of future climate and anthropogenic emission projections on air quality over Portugal and the Porto metropolitan area in 2050. Modeling results show that climate change will impact NO_2, PM_{10} and O_3 concentrations over Portugal. These authors believe that the air quality degradation is likely to be related with the trends found for the 2046–2065 climate and with the increase of background concentrations of pollutants and particulate matter. The results demonstrate the need for Portuguese authorities and policymakers to design and implement air quality management strategies that take climate change impacts into account.

To the best of our knowledge no study was conducted to the *Tâmega e Sousa* region where ESTG/P.Porto is located. In this paper ANOVA and MANOVA techniques are applied to study the differences between air quality in 2009, 2010, 2011 and 2012 in several regions of Portugal. Furthermore, for identifying the factors that influence the air quality in 2012 a multivariate regression model was used.

Paper is organized as follows. In Section 2 the database is presented. Section 3 and 4 present the data analysis. ANOVA and MANOVA are used in Section 3 for analyzing the differences between air quality in NUTS I, in particular the *Tâmega e Sousa* region is analyzed, and NUTS II are also considered and the difference of the air quality analyzed. In Section 4 a multivariate regression is employed to identify the factors that influence the air quality in 2012. Finally Section 5 presents the discussion.

2 Data Collection and Treatment

Data concerning air quality in Portugal was gathered from INE[1], the Portuguese National Institute of Statistics.

Maximum and minimum altitude, total area, total expenditure of environmental measures, total expenditure on protection of air quality and climate, total expenditure on protection of biodiversity and landscape, burned area, number of forest fires, number of extractive and manufacturing industries, per municipality in a total of 308, were collected from Pordata (`www.pordata.pt`) databases from the most current period available, 2009 to 2012.

The project "QualAr" collected data to calculate air quality indicators, that varies from "Poor" to "Very Good", taking into account the concentrations of the following pollutants: CO, NO_2, O_3, PM_{10} and SO_2 (see [1]). Using this data two independent research groups have developed deterministic numerical methods and statistical forecasting methods for predicting the air quality in Portugal [4, 7].

INE provides the number of days in each level of classification per year and regions, defined accordingly to the project "QualAr". However the regions are different from the municipality borders (see Figure 1), and therefore an formalization of the information was done.

Based on the information gathered, the following indicator of the air quality, per each year, and in each region, was computed:

$$AQI_{ij} = \frac{1n_{P,i} + 2n_{F,i} + 3n_{A,i} + 4n_{G,i} + 5n_{V,i}}{n_i}, \quad (1)$$

where

- $i = 2009, 201, 2011, 2012$ is the year;
- j is the region in Portugal (see Figure 1);
- $n_{P,i}, n_{F,i}, n_{A,i}, n_{G,i}, n_{V,i}$ is the number of days in year i with "Poor", "Fair", "Average", "Good" and "Very Good" air quality, respectively, and
- $n_i = n_{P,i} + n_{F,i} + n_{A,i} + n_{G,i} + n_{V,i}$ is the total number of days in year i for which a classification was provided.

For each municipality, the indicator of the air quality of the corresponding region is considered. Furthermore, for each year and each municipality the $\%$ of burned area, the number of extractive and manufacturing industries per km^2, were computed.

3 Assessment of Differences between Air Quality in Different Regions

All statistical analysis presented in this paper was performed using IBM SPSS. Furthermore, a 5% significance level was considered for all tests.

3.1 Differences between the air quality in NUTS I

Let us start by considering the NUTS I regions:

- Cont - Continente;
- RAM - Região Autónoma da Madeira ;

[1] See www.ine.pt for further information

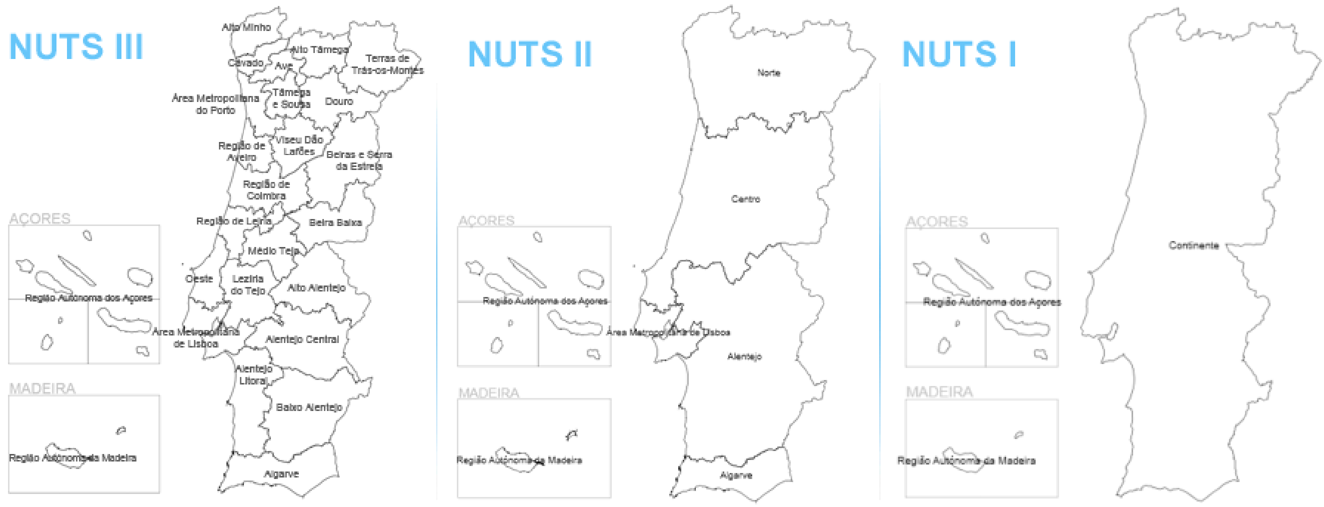

Figure 1: Portuguese NUTS (Nomenclature of Territorial Units for Statistics) classification. Source: http://www.pordata.pt

- RAA - Região Autónoma dos Açores .

In order to have a general overview of the air quality, the average of the AQI_{ij} with $i = 2009, 2010, 2011, 2012$, for each j region was computed.

The one-way Analysis of Variance (ANOVA) test, whose results are depicted in Table 1, suggests the existence of significant differences between averages of the air quality in NUTS I, since p-value < 0.05.

	Sum of Squares	df	Mean Square	F	p-value
Between Groups	1.278	2	0.639	40.22	0.00
Within Groups	4.844	305	0.016		
Total	6.122	307			

Table 1: ANOVA test for comparison of the average air quality between NUTS I.

Performing Post-Hoc Tukey Tests and using the homogeneous subsets definition (Table 2), it is possible to observe the existence of two groups:

- **(1)** RAA (*Região Autónoma dos Açores*), with the best air quality; and
- **(2)** Cont (*Continente*) and RAM (Região Autónoma da Madeira) with very similar air quality.

When considering the region of *Tâmega and Sousa* (TS) separately (see Table 3), the air quality of this region does does not differ significantly from the air quality of the *Continente*, but it differs from the air quality of the autonomous regions of Madeira (RAM) and Açcores (RAA). In fact, the TS region presents the worst air quality, with a average AIQ of 3.770, while RAA presents the best average air quality with 4.070.

		Subset for alpha =0.05	
NUTS	N	1	2
Con	278	3.807	
RAM	11	3.890	
RAA	19		4.070
p-value		0.090	1.000

Table 2: Post-Hoc Tukey Tests for comparison of air quality average between NUTS I.

		Subset for alpha =0.05		
NUTS	N	1	2	3
TS	11	3.770		
Con	267	3.808	3.808	
RAM	11		3.890	
RAA	19			4.070
p-value		0.813	0.241	1.000

Table 3: Post-Hoc Tukey Tests for comparison of air quality average between NUTS I and TS region.

Next, Multivariate Analysis of Variance (MANOVA) was performed in order to analyze the possible existence of differences between the averages of the air quality in NUTS I, during the period between 2009 to 2012.

Performing the Box's Test of Equality of Covariance Matrices, whose null hypothesis is H_0 :the ob-

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent Parameter	Observed Power
NUTS II	Pillai's Trace	1.554	31.866	24.000	1024.000	0.000	0.388	764.778	1.000
	Wilks' Lambda	0.109	38.600	24.000	1040.808	0.000	0.426	772.627	1.000
	Hotelling's Trace	3.451	42.633	24.000	1186.000	0.000	0.463	1023.184	1.000
	Roy's Largest Root	1.870	93.792	6.000	301.000	0.000	0.652	562.753	1.000

Table 4: Multivariate Analysis of Variance (MANOVA) considering the average of air quality.

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent Parameter	Observed Power
NUTS II	2012	1.304	6	0.217	18.753	0.000	0.272	112.521	1.000
	2011	3.853	6	0.642	56.459	0.000	0.530	338.755	1.000
	2010	3.105	6	0.517	43.260	0.000	0.463	259.560	1.000
	2009	2.643	6	0.441	24.017	0.000	0.324	144.100	1.000

Table 5: Tests of Between-Subjects Effects, from 2009 to 2012.

served covariance matrices of the dependent variables are equal across groups, the significance level of the test is approximately $0 < 0.05$, therefore one may assume the lack of the multivariate homoscedasticity.

While from the Bartlett's Test of Sphericity, with a significance level of approximately $0 < 0.05$ the null hypothesis that the residual covariance matrix is proportional to an identity matrix is rejected. Therefore the dependent variables are correlated, which justifies the use of MANOVA.

Table 4 shows the results for testing the effect of the independent variable (NUTS I) on the average air quality between 2009 and 2012. Independently of the statistics used, the factor (or independent variable) has a significant effect on the air quality, i.e., the dependent variables (air quality for 2009, 2010, 2011 and 2012) vary depending on the NUTS I, for all years.

Analyzing the Tests of Between-Subjects Effects, one may conclude the existence of significant differences between means of air quality for all the years (Table 5).

To study which regions differs in each year, an ANOVA for each year was performed. The results depicted in Table 6 suggest that there are significant differences between the average of the air quality in NUTS I.

Performing Post-Hoc Tukey Tests (Table 7) it is possible to observe that the best air quality in Portugal is, in general, in the *Região Autónoma dos Açores* - RAA, while *Região Autónoma da Madeira* - RAM, presents, in general, the worst air quality. In fact, the average AIQ in *Região Autónoma dos Açores* was above 4 for all years, while in *Região Autónoma da Madeira* presented values below 3.9 for all year except for 2009 when presented the best air quality with 4.06.

		Sum of Squares	df	Mean Square	F	Sig.
2012	Between Groups	.551	2	.275	19.802	.000
	Within Groups	4.241	305	.014		
	Total	4.792	307			
2011	Between Groups	1.738	2	.869	47.864	.000
	Within Groups	5.539	305	.018		
	Total	7.277	307			
2010	Between Groups	2.007	2	1.004	65.159	.000
	Within Groups	4.698	305	.015		
	Total	6.705	307			
2009	Between Groups	.936	2	.468	19.736	.000
	Within Groups	7.229	305	.024		
	Total	8.164	307			

Table 6: ANOVA tests for comparison of the air quality between NUTS I, for the 4 years considered.

3.2 Differences between averages of the air quality in NUTS II

Considering now the NUTS II division:

- Nor - *Norte*
- Cen - *Centro*
- AML - *Área Metropolitana de Lisboa*
- Ale - *Alentejo*
- Alg - *Algarve*
- RAM - *Região Autónoma da Madeira*
- RAA - *Região Autónoma dos Açores*

from the one-way Analysis of Variance test, whose results are depicted in Table 8, one may conclude that there are significant differences between averages of the air quality in NUTS II.

2012 AIQ

NUTS I	N	Subset for alpha =0.05	
		1	2
RAM	11	3.8900	
Con	278	3.9164	
RAA	19		4.0900
Sig.		0.754	1.000

2011 AIQ

NUTS I	N	Subset for alpha =0.05	
		1	2
RAM	11	3.7800	
Con	278	3.8323	
RAA	19		4.1400
Sig.		0.431	1.000

2010 AIQ

NUTS I	N	Subset for alpha =0.05	
		1	2
Con	278	3.8374	
RAM	11	3.9200	
RAA	19		4.1700
Sig.		0.087	1.000

2009 AIQ

NUTS I	N	Subset for alpha =0.05	
		1	2
Con	278	3.8386	
RAM	19		4.0000
RAA	11		4.0600
Sig.		1.000	0.428

Means for groups in homogeneous subsets are displayed.

Tukey HSD uses Harmonic Mean Sample Size = 20.389.

The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Table 7: Homogeneous subsets of NUTS I, between 2009 and 2012.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.848	6	0.475	43.639	0.000
Within Groups	3.274	301	0.011		
Total	6.122	307			

Table 8: ANOVA test for comparison of the air quality average between NUTS II.

The Post-Hoc Tukey Tests (Table 9), shows the existence of three groups:

- (1) *Região Autónoma dos Açores* - with the best air quality, i.e., 4.070;
- (2) *Centro, Região Autónoma da Madeira* and *Alentejo* - with a median air quality, between 3.9076 and 3.8467; and
- (3) *Área Metropolitana de Lisboa, Norte* and *Algarve* – with the worst air quality, with 3.6944, 3.7294 and 3.6944, respectively.

On the other hand, when considering the TS region separately (Table 9, right) it is possible to identify four groups. The air quality in the TS region does not differ significantly from the air quality of *Área Metropolitana de Lisboa, Norte* and *Algarve*, nor from *Centro* region.

4 Multivariate regression

In order to determine the factors that influence the air quality in the year 2012 a multivariate regression model was used. The 21 initially considered variables were:

- air quality;
- minimum and maximum altitude,
- percentage of burned area,
- number of extractive industries per km^2,
- number of manufacturing industries per km^2,
- number of forest fires,

in 2009, 2010 and 2011 .

Using the stepwise method, a significant linear model was found, according to the ANOVA linear coefficients test, i.e., there exists at least an independent variable with significant effect in the air quality in 2012.

The linear multivariate model has $R^2 = 0.77$ (Table 11), therefore approximately 77% of the variability of the air quality in 2012 is explained by:

- 2011 Air Quality;
- 2012 number of forest fires;
- 2012 number of manufacturing industries per km^2;
- 2011 percentage of burned area;
- 2009 Air Quality;
- 2010 number of forest fires; and

		Subset for alpha =0.05		
NUTS II	N	1	2	3
AML	18	3.6944		
Nor	86	3.7294		
Alg	16	3.7363		
Cen	100		3.8467	
RAM	11		3.8900	
Ale	58		3.9076	
RAA	19			4.070
Sig.		0.818	0.422	1.000

Means for groups in homogeneous subsets are displayed.

Tukey HSD uses Harmonic Mean Sample Size = 23.297.

The group sizes are unequal. The harmonic mean of the group sizes is used.

Type I error levels are not guaranteed.

		Subset for alpha =0.05			
NUTS II	N	1	2	3	4
AML	18	3.6944			
Nor	75	3.7235			
Alg	16	3.7362			
TS	11	3.7700	3.7700		
Cen	100		3.8467	3.8467	
RAM	11			3.8900	
Ale	58			3.9076	
RAA	19				4.070
Sig.		0.289	0.271	0.576	1.00

Table 9: Post-Hoc Tukey Tests for comparison of the air quality between NUTS II.

- minimum altitude.

Analyzing the standardized coefficients one may conclude that the 2011 Air Quality is the variable with more contribution to the 2012 Air Quality, followed by 2012 Number of forest fires and 2012 number of manufacturing industries per km^2, 2011 % of burned area, 2009 Air Quality, 2010 number of forest fires, and minimum altitude.

On the other hand, there is no statistical evidence that the air quality in 2010, number of forest fires in 2009 and 2011, % of burned area in 2009 and 2011, number of manufacturing in 2009, 2010 and 2011, and maximum altitude, are factor that contribute to the air quality in 2012.

In conclusions, the air quality in 2011, the number of forest fires in 2012, number of manufacturing industries per km^2 in 2012, the % of burned area in 2011, Air Quality in 2009, and minimum altitude, contribute positively to the air quality in 2012. While the number of forest fires in 2010 present a negative impact on the air quality in 2012.

5 Discussion

In this paper, the differences of air quality different the regions of Portugal, according to the NUTS I and NUTS II division were studied. The data was obtained from Pordata and INE, concerning the years 2009, 2010, 2011 and 2012. The particular region of *Tâmega e Sousa* region, in the North of Portugal, was analyzed.

The results show the existence of significant differences in the air quality for the different regions. The region of *Região Autónoma dos Açores* presented the best air quality, while, in general the *Área Metropolitana de Lisboa, Norte* and *Algarve* present the worst air quality in the Portuguese territory. The *Tâmega e Sousa* region presents air quality similar to the air quality of *Área Metropolitana de Lisboa, Norte* and *Algarve*.

In addition to the methods presented in this paper the nonparametric Kruskal-Wallis and Mann-Whitney-Wilcoxon were performed. The results were very similar to the ones presented, for this reason they were not presented.

From the multivariate regression model it was possible to conclude that there is statistical evidence that the variables that present a larger contribution to the air quality in 2012 are the air quality in 2011, the number of forest fires in 2012 and number of manufacturing industries perkm^2 in 2012, the % of burned area in 2011, air quality in 2009, the number of forest fires in 2010, and minimum altitude. The factor that contribute positively to the air quality in 2012 are the air quality in 2011, the number of forest fires in 2012, number of manufacturing industries per km^2 in 2012, the % of burned area in 2011, Air Quality in 2009, and minimum altitude. On the other hand the number of forest fires in 2010 presents a negative impact on the air quality in 2012.

The data used in this study were the most recent one available in INE and Pordata. In the future we expect to obtain more recent data and more spatially detailed data for the particular region of *Tâmega e Sousa*.

Acknowledgments

This work was supported by CIICESI – Center for Research and Innovation in Business Sciences and Information Systems, ESTG/P.Porto – School of Technology and Management, Polytechnic of Porto.

References:

[1] *http://qualar.apambiente.pt/*.

[2] Allaa M Aenab, SK Singh, and Ali Jabir Lafta. Critical assessment of air pollution by anova test

and human health effects. *Atmospheric Environment*, 71:84–91, 2013.

[3] CI Beattie, JWS Longhurst, and NK Woodfield. Air quality management: evolution of policy and practice in the uk as exemplified by the experience of english local government. *Atmospheric Environment*, 35(8):1479–1490, 2001.

[4] C Borrego, A.I Miranda, M Coutinho, J Ferreira, and A.C Carvalho. Air quality management in portugal: example of needs and available tools. *Environmental Pollution*, 120(1):115 – 123, 2002.

[5] C Borrego, O Tchepel, N Barros, and AI Miranda. Impact of road traffic emissions on air quality of the lisbon region. *Atmospheric Environment*, 34(27):4683–4690, 2000.

[6] Carlos Borrego, M Coutinho, AM Costa, J Ginja, C Ribeiro, A Monteiro, I Ribeiro, J Valente, JH Amorim, H Martins, et al. Challenges for a new air quality directive: the role of monitoring and modelling techniques. *Urban Climate*, 14:328–341, 2015.

[7] A Monteiro, AI Miranda, C Borrego, and R Vautard. Air quality assessment for portugal. *Science of the Total Environment*, 373(1):22–31, 2007.

[8] E Sá, H Martins, J Ferreira, M Marta-Almeida, A Rocha, A Carvalho, S Freitas, and C Borrego. Climate change and pollutant emissions impacts on air quality in 2050 over portugal. *Atmospheric Environment*, 131:209–224, 2016.

[9] RLR Salcedo, MCM Alvim Ferraz, CA Alves, and FG Martins. Time-series analysis of air pollution data. *Atmospheric Environment*, 33(15):2361–2372, 1999.

2012 Air Quality

		Subset			
NUTSII	N	1	2	3	4
Alg	16	3.7650			
AML	18	3.8383	3.8383		
RAM	11		3.8900	3.8900	
Nor	86		3.8919	3.8919	
Cen	100			3.9522	
Ale	58			3.9572	
RAA	19				4.0900
Sig.		.235	.619	.337	1.000

2011 Air Quality

		Subset			
NUTSII	N	1	2	3	4
AML	18	3.7078			
Alg	16	3.7375			
Nor	86	3.7463			
RAM	11	3.7800	3.7800		
Cen	100		3.8717	3.8717	
Ale	58			3.9569	
RAA	19				4.1400
Sig.		.242	.055	.095	1.000

2010 Air Quality

		Subset			
NUTSII	N	1	2	3	4
Alg	16	3.7500			
AML	18	3.7783	3.7783		
Cen	100	3.7847	3.7847		
Nor	86		3.8591	3.8591	
RAM	11			3.9200	
Ale	58			3.9386	
RAA	19				4.1700
Sig.		0.933	0.156	0.169	1.000

2009 Air Quality

		Subset		
NUTSII	N	1	2	3
AML	18	3.7261		
Alg	16	3.7363	3.7363	
Nor	86	3.7809	3.7809	
Cen	100		3.8481	
Ale	58			3.9710
RAA	19			4.0000
RAM	11			4.0600
Sig.		.811	.075	.276

Table 10: Homogeneous subsets of NUTS II, from 2009 to 2012.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	.905	.103		8.810	.000		
2011 Air Quality	.673	.032	.830	20.927	.000	.488	2.049
2012 Number of forest fires	.000	.000	.289	4.046	.000	.151	6.643
2012 Number of manufacturing industries/km^2	.004	.001	.134	4.390	.000	.828	1.207
2011 % of Burned area	.008	.003	.096	3.010	.003	.760	1.315
2009 Air Quality	.102	.030	.133	3.377	.001	.498	2.007
2010 Number of forest fires	.000	.000	-.204	-2.868	.004	.152	6.568
Minimum altitude	1.831E-05	.000	.069	2.278	.023	.845	1.183

Table 11: Linear Coefficients of the multivariate regression.