SPATIAL PREDICTION OF AIR QUALITY IN AN URBAN SENSITIVE AREA: A SEASONAL OVERVIEW

Tânia Fontes and Nelson Barros

1. INTRODUCTION

In most urban areas, as Porto, road traffic is the main source of ambient air concentration of pollutants such as nitrogen dioxide (NO_2), carbon monoxide (CO), benzene (C_6H_6), and particles (Alpopi, 2010; Keuken et al., 2004). In order to monitor this problem urban air quality monitoring and forecast has become an important issue for many environmental protection agencies around the world in the last decades (Horálek et al., 2005). To generate maps of concentrations means that interpolating and extrapolating methods have to be used.

Spatial prediction techniques, also known as spatial interpolation techniques, differ from classical modelling approaches in that they incorporate information on the geographic position of the sample data points (Cressie, 1993). The interpolation is the procedure of predicting unknown values using the known values at neighbouring locations which may be regularly or irregularly spaced. The values derived in this way are not necessarily the true value; they are a mathematical "best guess" based on the known values (Aranoff, 1995). Thus this method is necessary when the ground truth data does not cover the whole area. Demers (2005), classifies interpolation methods into linear and non-linear.

Usually, interpolation methods use the weighted average of nearby data to calculate the estimates. The weights could be assigned according to deterministic or statistical criteria. The quality of the interpolation results depends on the accuracy, number, and distribution of the known points used in the calculation and on how well the mathematical function correctly models the phenomenon (Aranoff, 1995). Also, the pollutants are governed by different mechanism, of acting on a different spatial scale: regional and local effects. Fluctuations in the concentration pattern are mainly driven by meteorological phenomena; however, air pollution can have a distinct local character due to local emission sources and their temporal variability. In the dense urbanized region of Asprela in Porto, the study area, the latter effects are significant. This is a sensitive area with several universities and two hospitals, with high traffic density and mobility problems, yet it does not have any air quality station within the national network.

Among statistical methods, geostatistical kriging-based techniques, including simple and ordinary kriging, universal kriging and simple cokriging have been used for spatial analysis. In the deterministic interpolation methods, inverse distance weighting method and its modifications are the most often applied. Kriging and Inverse Distance Weight (IDW) are the most commonly used methods when measuring air quality (Wong et al., 2004; Horálek, 2005; Mesquita, 2010; Sánchez, 2009). Both methods estimate values at unsampled locations based on the measurements at surrounding locations with certain assigned weights for each measurement. However, while kriging requires the preliminary modelling step of a variance-distance relationship, IDW does not require this step. Many studies have compared IDW and kriging, and the performance of kriging was generally better (Wong et al., 2004).

In this paper, two interpolation models have been used – IDW and Ordinary Kriging – which can incorporate both the regional and local aspects of the air pollution phenomenon. The main objective of this study was to describe and analyse the relative performance of these interpolation methods to predict the field concentration of air pollutants, such as benzene (C_6H_6) and nitrogen dioxide (NO_2), in Porto's Asprela area. This work summarised below was performed twice, once in winter 2009 and once in summer 2010.

The outline of this paper is as follows. In Section 2 the methodology is described in detail and in Section 3 and 4 model results are discussed and validated. A conclusion is presented in Section 5.

2. MATERIALS AND METHODS

The present study focuses on the atmospheric concentration of C6H6 and NO2 in Porto's Asprela area. These two pollutants are considered to be good indicators for pollution caused by traffic. Hence, they can be regarded as the most important emission source in the area. Measurement of these pollutants was conducted in the scope of the CIVITAS ELAN project (TREN/FP7TR/218954 - ELAN), co-funded by the European Commission between September 2008 and September 2012.

Statistical analyses were done in two stages. First, a geostatistical analysis was performed and then the distribution of data was described using conventional statistics such as mean, maximum, minimum and standard deviation (SD). In this second step the results were also compared with the annual limit value for the protection of human health defined in the Directive 2008/50/CE (5 µg.m^{-3} for C_6H_6 and 40 µg.m^{-3} for NO_2). Although these limits are linked to annual averages and the results are based on a relatively short measuring period (three weeks), this approach represents a good indicator for the potential impact of the air quality in Asprela (study area) on human health.

Regarding the first step, firstly the number of observation points needed for a correct spatial representation of atmospheric concentrations field in the area was defined. To do this, a statistic approach based on a screen run of a numerical dispersion model was used.

Due to its ability to reproduce mesoscale atmospheric circulations and photochemical production, the possibility to work with long data series, as well as its speed of data processing, the air quality modelling was performed using the TAPM model (Hurley, 2008). The model was applied using a three level nesting technique with 25 x 25 points, centred on the city of Porto (41°9'34, 94''N, 8º37'19, 32''W) and 25 vertical grid levels, between 10 and 8,000 metres. The larger grid uses cells of 30 km x 30 km, the intermediate grid, cells of 10 km x 10 km and the finer grid, cells of 3 km x 3 km (Figure 1). The synoptic forcing was done based on the ECMWF data for 2006, a year that can be considered a good approximation to the usual climatic conditions within the study region (Fontes, 2010). The simulation of air quality was made using CO and NO_2 UNECE emissions (Boavida et al., 2008) for the finer grid. Because the UNECE emission inventory does not include C_6H_6 emissions the TAPM was used to simulate, as a tracer of C_6H_6, the CO concentrations. Thus, two types of air quality simulations were done, one in non-reactive mode for CO concentrations, and another in reactive mode for NO_2. In the reactive mode simulation a value of reactivity pattern of VOC emissions (R_{smog}) of 0.0067 (Hurley et al., 2005) was considered. Additionally, in order to get a refinement of the concentration field over the study area, the output grid was processed to a sub-grid with cells of 300 m x 300 m. The simulation was performed for both measuring periods (winter 2009 and summer 2010).

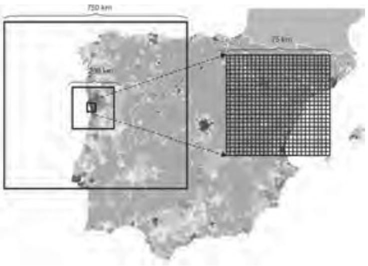
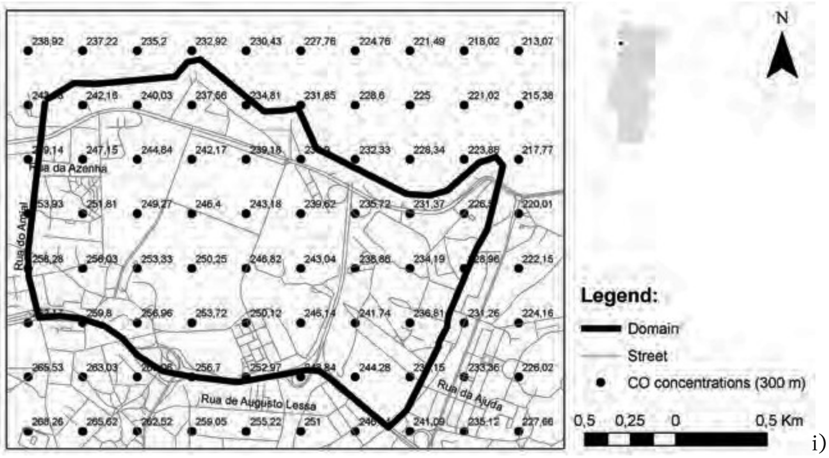


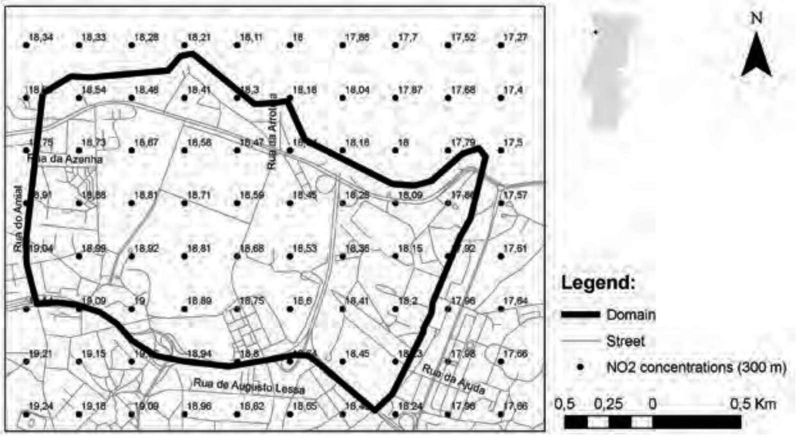
Figure 1: Meteorological areas considered in the air quality simulation using TAPM model in Metropolitan Area of Porto (MAO).

Using the estimate concentrations over the Asprela area (Figure 2), a CO average concentration of 241.86 ± 13.80 µg.m^{-3} and a NO_2 concentration of 18.43 ± 0.51 µg.m^{-3} was predicted. Based on these confidence intervals (S), the minimum number of observation points for each pollutant was estimated, considering that the confidence level (LC) has a significance level of 5% of the mean value (Eq. 1).

$$LC=\frac{s}{\sqrt{n}}t_{0.975} <=> n=\left(\frac{s}{LC}\,t_{0.975}\right)^2 \text{(Eq. 1)}$$

This approach resulted in an estimate of about six observation points for CO and two for NO_2 required for a representative description of the study area.





ii)

Figure 2: Air concentrations (µg.m^{-3}) estimated by TAPM model in the Asprela area: i) CO; ii) and NO_2.

Based on these results two measuring periods were chosen (19 November – 16 December 2009 in winter and 28 June – 20 July 2010 in summer) using a diffusive sampler technique to monitor pollutants (PASSAM, 2010). In order to control deviations and diffusive samplers lost due to vandalism, 12 observation points were used. The diffusive samplers were placed on poles at a height of 3 m and evenly distributed within the area (Figure 3). Additionally, and in order to monitor the background concentration, measurements were also done at the top of four buildings of the city (height: 40–50 m): Antas tower, Burgos building, JN building and CMP building. At these locations replicas were used to control eventual deviations. Due to vandalism point number 4 of C_6H_6 was lost during the measuring period.

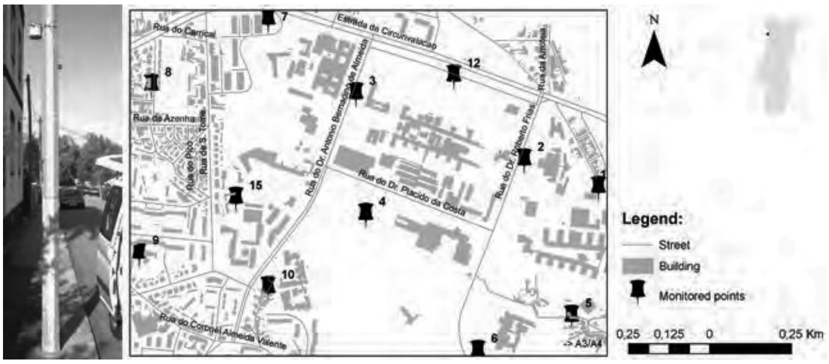


Figure 3: Location of the diffusive measurement points in the Asprela area (Porto).

To estimate the spatial distribution of C_6H_6 and NO_2 during the winter period, two interpolation methods were used: (i) a deterministic method, the Inverse Distance Weighting (IDW), (ii) and a geostatistical method, the ordinary kriging. These methods were implemented using the ArcGIS 9.3 software.

The IDW is a method often used to interpolate data from air quality, given its simplicity (Brigs et al., 2005; Keuken et al., 2005; Lindley and Walch, 2005). To predict a value for any unmeasured location, this method uses the measured values surrounding the prediction location. Closest values have more influence on the predicted value than those farther away, hence the name inverse distance weighted. The surface calculated depends on the selection of a power value and the neighbourhood search strategy. IDW is an exact interpolator, where the maximum and minimum values in the interpolated surface can only occur at sample points. The output surface is sensitive to clustering and the presence of outliers. IDW assumes that the surface is being driven by the local variation, which can be captured through the neighbourhood.

$$f(x,y)=\left[\frac{\Sigma^{n}_{i=1} w(d_i) z_i}{\Sigma^{n}_{i=1} w(d_i)}\right] \quad \text{(Eq. 2)}$$

Where z_i is an observed value i, d_i is the distance between the estimated point and the observed point i; $w(d_i)=1/(d_i)^p$ is the ponderation of the observation i; and p is the power function.

The IDW predictions were performed varying the number of power (0.5 and 3) and using different radiuses and neighbours.

The kriging method is similar to the IDW for considering the measured values in the neighbourhood to predict the concentrations in an unmeasured location. In ordinary kriging the weights depends on the model fitted to the measurement points of the local distance estimate, and the spatial relationships between the measured values around the local forecast (Johnston et al., 2001). Ordinary Kriging assumes the model:

$$Z(s)=\mu+\varepsilon(s) \quad \text{(Eq. 3)}$$

Where μ is an unknown constant.

The performance of each evaluation technique was assessed comparing the deviation of estimates using the cross-validation method. To use this technique, a point was excluded and then the model was applied to estimate the concentrations at this removed point. Therefore, the comparison of the performance between the different interpolation techniques was achieved using the average error (*ME*)(Eq. 4), the Mean Absolute Error (*MAE*) (Eq. 5), the Root Mean Squared Error (*RMSE*) (Eq. 6) and de Normalized Root Mean Squared Error (*NRMSE*) (Eq. 7), and the coefficient of determination (R^2) (Eq. 8) (Mesquita, 2010):

$$ME=1/n\ \Sigma^{n}_{i=1}E_i\text{-}O_i \quad \text{(Eq. 4)}$$

$$ME=1/n\ \Sigma^{n}_{i=1}\left|E_i\text{-}O_i\right| \quad \text{(Eq. 5)}$$

$$RMSE=\sqrt{1/n\ \Sigma^{n}_{i=1}\ (E_i\text{-}O_i)^2} \quad \text{(Eq. 6)}$$

$$NRMSE=\frac{RMSE}{\bar{O}_i} \quad \text{(Eq. 7)}$$

$$R^2=1\text{-}\ \frac{\Sigma^{n}_{i=1}(O_i\text{-}E_i)^2}{\Sigma^{n}_{i=1}(O_i\text{-}O_i)^2} \quad \text{(Eq. 8)}$$

Where: E = Estimated value; O = Observed value; n = Number of cases; O = Mean of observed values.

All of these parameters have to be equal or close to zero, except the R^2 where the optimal value is 1.

To verify if the selected map can be considered from the legal point of view, the value of uncertainty was also calculated according to Directive 2008/50/EC:

$$Uncertanty=\frac{\text{Observed value - Modelled value}}{\text{Limit value}} \quad \text{(Eq. 9)}$$

The results were compared with the uncertainty of estimation, for both pollutants, defined by Directive 2008/50/EC, 100% to C_6H_6 and 75% to NO_2.

3. RESULTS AND DISCUSSION

Table 1 presents the results of the statistical analysis using cross-validation to the best interpolation map, by pollutant and interpolation method. The specifications are presented in Table 2. The comparison between the two methods of spatial distribution, IDW and ordinary kriging, shows that ordinary kriging is the best method to simulate the C_6H_6 but the differences to the IDW method are small. In the case of the NO_2 concentrations the IDW gives better results. This might be the case because of the different patterns of field concentrations for the study pollutants (different emission and reactivity pattern) and/ or due to the lack of one of the C_6H_6 control points (which was lost due to vandalism).

		C_6H_6		NO_2	
		IDW	Ordinary kriging	IDW	Ordinary kriging
ME		-0.0247	-0.0001	0.0164	6.8370
MAE		0.22	0.20	0.09	2.74
RMSE		0.30	0.29	0.11	3.69
NRMSE		0.24	0.23	0.002	0.078
R^2	All values	≈ 0.00	0.09	0.99	0.17
	Excluding the best and the worse case	0.55	0.32	0.99	0.78
Uncertanty		≈ 4.36%	4.00%	0.22%	6.85%

Table 1: Results of statistical analysis using cross-validation.

			C_6H_6		NO_2	
			IDW	Ordinary kriging	IDW	Ordinary kriging
Specifications	IDW	Output	5	-	5	-
		Power	3	-	1	-
		Search radius	Variable	-	Variable	-
Specifications	Ordinary kriging	Method	-	Spherical	-	Gaussian
		Lag size	-	500	-	500
		Major range	-	1000	-	1000
		Partial still	-	15	-	15
		Nugget	-	30	-	30
	Radius number of points		6	12	16	12

Table 2: Selected specifications by pollutant and method.

Figure 4 and 5 show the two best maps resulting from different simulations for C_6H_6 and NO_2 respectively using the IDW method and ordinary kriging for the winter measuring period. Figure 6 and 7 show the same results for the summer measuring period.

Although for all measuring periods the C_6H_6 maps record always low concentrations, the NO_2 map shows some areas with some critical points (hotspots). This is the case of sample 5 and 6 during the winter season (Figure 5) and sample 2 and 6 during the summer season (Figure 7) where the concentration values are high and are the main concern of the area. Due to this situation the IDW method can pick up these fluctuations better than the ordinary kriging method. Although the whole area is characterized by the presence of roads and car parks, this critical point (sample 4) was located in an area surrounded by vegetation and a sports park with low direct emission of pollutants. Due to the one measuring point that was destroyed by vandalism the C_6H_6 concentrations for this point were lost, but the records of NO_2 concentrations confirm that this is an area with less pollution. This fact can be decisive to show, in future measuring campaigns, that the IDW method may be the best interpolated method for the Asprela area.

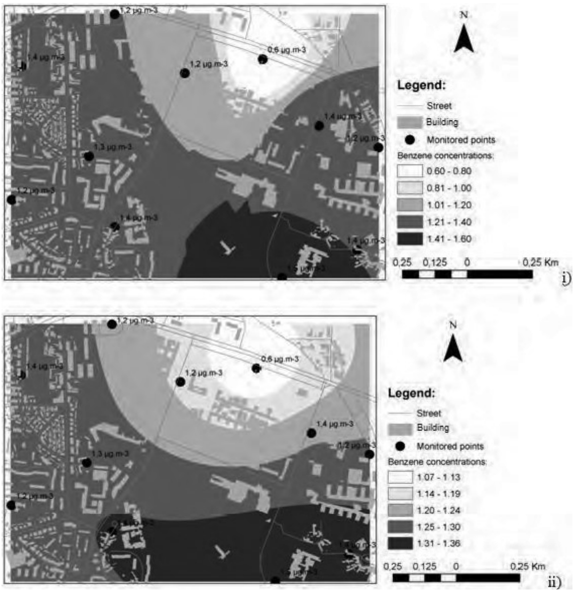


Figure 4: Average C_6H_6 concentrations (µg.m-$^{3)}$ in the Asprela area from 19/11-16/12/2009 using different interpolation methods: i) IDW, ii) Ordinary kriging.

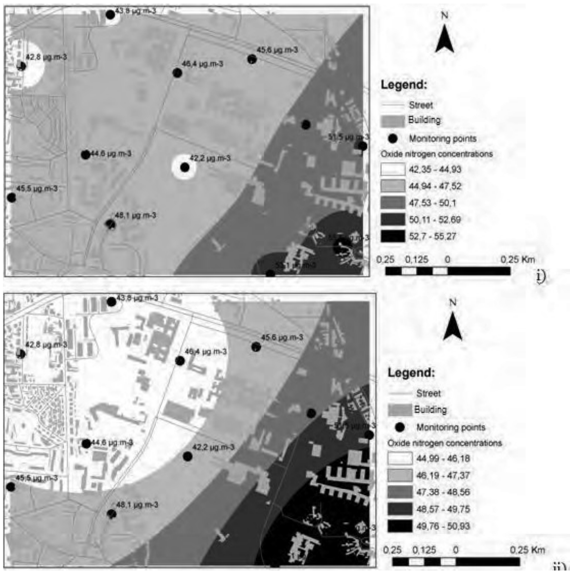


Figure 5: Average NO$_2$ concentrations (µg.m^{-3}) in the Asprela area from 19/11-16/12/2009 using different interpolation methods: i) IDW, ii) Ordinary kriging.

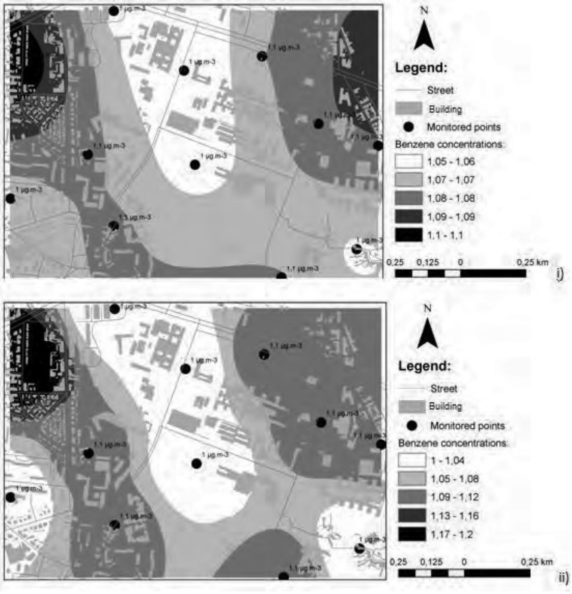


Figure 6: Average C$_6$H$_6$ of air concentrations (µg.m^{-3}) in the Asprela area from 28/06-21/04/2010 using different interpolation methods: i) IDW, ii) Ordinary kriging.

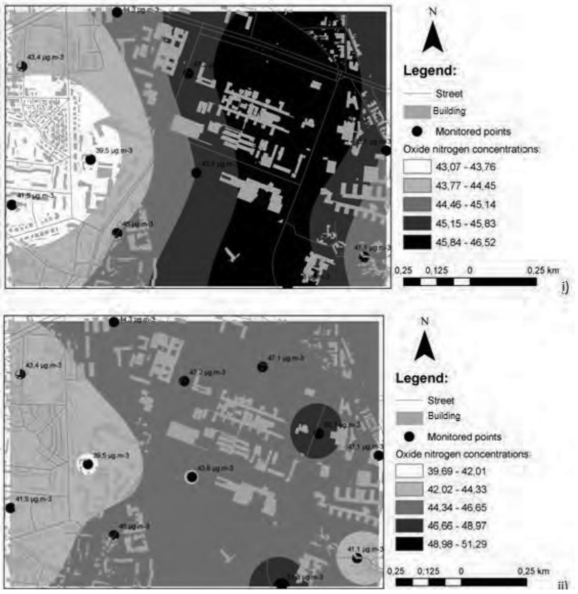


Figure 7: Average NO_2 of air concentrations ($\mu g.m^{-3}$) in the Asprela area from 28/06-21/07/2010 using different interpolation methods: i) IDW, ii) Ordinary kriging.

Once the best map for each pollutant had been validated and selected, the spatial analysis of the results took place as summarised in the next chapter.

4. DATA ANALYSIS

At a height of 3 m, the observed average concentrations of C_6H_6 are 1.25 ± 0.23 $\mu g.m^{-3}$ and 47.22 ± 4.18 $\mu g.m^{-3}$ of NO_2 in the winter period and 1.08 ± 0.06 $\mu g.m^{-3}$ of C_6H_6 and 44.98 ± 3.66 $\mu g.m^{-3}$ of NO_2 in the summer period. For the two pollutants the minimum values were recorded close to the “Estrada da Circunvalação” and the maximum values were recorded in the south close to the A3/ A4 motorway where the volume of traffic increases and traffic speed decreases causing an emissions increase. In the winter period the background average concentrations (observations in the top of the buildings) are 1.02 ± 0.11 $\mu g.m^{-3}$ of C_6H_6 and 34.86 ± 1.45 $\mu g.m^{-3}$ of NO_2, 18.4% and 26.2% lower than the average values recorded at 3 m respectively. Otherwise, in the summer period the background average concentrations (observations in the top of the buildings) are 1.07 ± 0.10 $\mu g.m^{-3}$ of C_6H_6 and 41.70 ± 2.02 $\mu g.m^{-}$

^{3} of NO_2, 0.6% and 7.3% lower than the average values recorded at 3 m respectively, which may show a stronger atmospheric stratification situation during the winter measuring period when compared with the measuring period in summer.

The uncertainty of the estimates for both pollutants in winter is below the limits defined in Directive 2008/50/EC, 100% for C_6H_6 and 75% for NO_2 (Table 1). Thus, these maps are representative for the study area and are a good indicative tool to evaluate the results from the legal point of view. The comparison of C_6H_6 results with the human health protection value of 5 µg.m^{-3} defined by the Directive 2008/50/EC shows that, during the study periods, at all monitoring points, the C_6H_6 concentrations are lower than this limit value. Even considering the 23.0% of uncertainty of the C_6H_6 measurement method (PASSAM, 2010) the measured concentrations never exceed the average annual limit value for human health protection. On the other hand, the average NO2 concentrations are in the whole study area higher than the average annual limit value for human health protection defined by the Directive 2008/50/EC (40 µg.m^{-3}). However, considering the 18.7% of uncertainty of measurement method (PASSAM, 2010) for NO_2, 25% of these control points may not exceed the mentioned limit value.

5. CONCLUSIONS

The methods of spatial analysis used in this study do not render possible the definition of an optimal interpolation method for the winter and summer measuring period in the Asprela area. However, the analysis indicates that, in general, the IDW method should be the best method to be used for the study area. In order to confirm this, other tests have to be done for other periods and pollutants. Furthermore, other methods such as ordinary cokriging or multiple linear regression, using auxiliary variables to adjust the spatial interpolation, could be tested in order to minimize errors.

The results for both measuring periods show that, when compared with the Directive 2008/50/EC average annual limit values for the human health protection, the C_6H_6 concentrations are very low (lower during the summer when compared with the winter values), while the NO_2 concentrations present some values above that limit, in particular during the winter period. Even considering the uncertainties of the measurement methods, the C_6H_6 recorded concentrations are never above the annual average limit value for human health protection. Nevertheless, the NO_2 recorded concentrations are above the annual average limit value for human health protection at most of the control points of the analysed area during both measuring periods.

REFERENCES

Alpopi, Christina and Sofia Elena Colesca. 2010. Urban air quality, a comparative study of major European Capitals. *Theoretical and Empirical Researches in Urban Management* 6(15): 92–107.

Aranoff, Stan. 1995. *Geographic information Systems: A management perspective*. Ottawa, Canada: WDL Publications.

Briggs, David J., Hans Martin Seip, Chris Dore, Gerard Hoek, Michael Petrakis and Gavin Shaddick. 2005. *Air pollution modelling for support to policy on health and environmental risks in Europe. Final Report Section 6*. London: Imperial College.

Boavida, Filomena, Ana Teresa Perez, Pedro Torres, Teresa Costa Pereira, Vitor Góis, Tiago Seabra and Hugo Maciel. 2008. *Alocação espacial de emissões em 2005*. Lisboa: Agência Portuguesa do Ambiente.

Cressie, Noel. 1993. *Statistics for Spatial Data Revised Edition*. New York: John Wiley & Sons, Inc.

DeMers, Michael M. 2005. *Fundamentals of geographic information systems*. EUA: John Wiley & Sons. Inc.

European Commission. *Directive 2008/50/EC. Define the ambient air quality guidelines to avoid, prevent or reduce harmful effects on human health and the environment*.

Fontes, Tânia. 2010. *Impacto da qualidade do ar na saúde pública: O caso da cidade do Porto*. Thesis presented to the University of Aveiro to obtain a PhD on Environmental Sciences.

Horálek, Jan, Pavel Kurfurst, Bruce Denby, Peter de Smet, Frank de Leeuw, Marek Brabec and Jaroslav Fiala. 2005. *Interpolation and assimilation methods for European scale air quality assessment and maping Part II: Development and testing new methodologies*. Final draft, ECT/ACC Techical Paper 2005/8.

Hurley, Peter J.. 2008. *TAPM V4. CSIRO Marine and Atmospheric Research Paper N°25*.

Hurley, Peter J., William L. Physick and Ashok K. Luhar. 2005. TAPM – A practical approach to prognostic meteorological and air pollution modelling. *Env. Modelling and Software* 20: 737–752.

Johnston, Kevin, Jay M. Ver Hoef, Konstantin Krivoruchko and Neil Lucas. 2001. *Using ArcGIS Geostatistical Analyst*. GISby ESRIS.

Keuken, Menno, Eric Sanderson, Roel van Aalst, Jens Borken and Jürgen Schneider. 2004. *Health effects of transport-related air pollution, contribution of traffic to levels of ambient air pollution in Europe*. Available at: http://www.bvsde.paho.org/bvsacd/cd63/e86650/cap2.pdf, 16. 4. 2010.

Lindley, Sarah and Tanya Walsh. 2005. Inter-comparation of interpolated background nitrogen dioxide concentrations across Greater Manchester. *Atm. Env*. 39(5): 2709–2724.

Mesquita, Sandra Maria Pereira. 2010. *Modelação da distribuição espacial da qualidade do ar em Lisboa usando sistemas de informação geográfica*. Instituto Superior de Estatística e Gestão de Informação da Universidade Nova de Lisboa. Master thesis in Geographical Information Systems.

PASSAM. 2010. Available at: http://www.passam.ch/, 16. 4. 2010.

Sánchez, Hermes U. Ramirez, Maria D. Andrare García, Rubén Bejaran, Mario E. García Guadalupe, Antonio Vallo Vázquez, Ana C. Pompa Toledano and Odila de la Torre Villasenor. 2009. The spatial–temporal distribution of the atmospheric polluting agents during the period 2000–2005 in the Urban Area of Guadalajara, Jalisco, Mexico. *J. Hazardous Materials* 165: 1128–1141.

Wong, David W., Lester Yuan and Susan E. Perlin. 2004. Comparison of spatial interpolation methods for the estimation of air quality data. *J. Exp. Analysis and Env. Epid*. 14: 404–415

ACKNOWLEDGEMENTS

To *Fundação para a Ciência e Tecnologia* (FCT) for grant support to Tânia Fontes (Ref.º SFRH/BD/19027/2004).