PREDICTIVE MAPPING OF AIR POLLUTANTS: A SPATIAL APPROACH

E. Sertel, H. Demirel, S. Kaya

ITU, Faculty of Civil Engineering, Department of Geodesy and Photogrammetry, 34469 Maslak Istanbul, Turkey – (elifsertel, hdemirel, skaya@ins.itu.edu.tr)

KEY WORDS: Statistics, Environment, Pollution, Quality, GIS

ABSTRACT:

Integrated transport, land-use and air quality monitoring are prioritized especially in metropolitan areas. The complexity is straight forward, since data, indicators, variables, methods and approaches vary and isolated. The Spatial Information Sciences (SIS) provides mature solutions for data and policy integration, since the nature of problem is specific to geo-spatial distribution. The interaction between transport and land-use has been studied in several international, national studies; however air quality policies have not fully integrated. For integration, limited number of sample points and sparse spatial observations should be interpolated in order to obtain areal coverage. Geostatistics provides the most probable solution depending upon measurements and other relevant information, where the accuracy of the prediction is known. Within this study, air quality parameters of the European side of the Istanbul Metropolitan area were evaluated by means of geostatistics. For assessing the uncertainty and accuracy; time series plots and comparison of predicted and observed concentrations of parameters for each monitoring station were produced. For mapping the air pollution levels, kriging method was used. Air quality emission parameters, sulfurdioxide (SO_2), carbonmonoxide (CO) and particulates (PM), were collected from six sampling stations between years 2003-2006. Different point combinations were tested for producing SO_2, CO and PM maps, in order to investigate the effect of point distribution on kriging. Research results indicated that the integrated usage of geostatistical methods, remote sensing and spatial analysis can introduce valuable information to identify, visualize and explore the complex relationships between transport, land-use and air quality.

1. INTRODUCTION

The geoinformation sciences currently became the core component of integrated transport, land-use and environmental decision making systems (Geerlings and Stead, 2003; Sperling et. al, 2004). The spatial data are considered to be near perfect representation of the real world, although it involves uncertainties. Spatial data quality affects the reliability of data analysis and presentation and decisions taken in parallel. Clearly, given their important role in assisting policy making, it is important to quantify the uncertainties associated with environmental process models (Leopold et al, 2006). The number of sample points used in environmental studies is limited due to cost and techniques of measurements. Air quality studies are not exceptions, where sample points are sparse observations. This greatly reduces the availability of appropriate interpolation methods for estimating the total air quality. For such analyses geostatistics introduce useful for up-scaling the data on attributes that have been collected at points to provide complete areal coverage (Bierkens et al, 2000).

The integrated usage of geostatistical methods and geo-sciences can introduce valuable information to identify, visualize and explore relationship between transportation, land-use and air quality within a GIS framework. The obtained information is associated with quality, which will aid reliability to taken decisions. The study is organized as follows; firstly air quality and the role of quality is briefly provided. Secondly, the study area is introduced, where data used and methodology is briefly described. Thirdly, geostatistical methods, being time series and kriging combined with spatial analyses are presented at the discussion section. Major findings and proposals are given at the conclusion.

2. AIR QUALITY & GEOSTATISTICS

In both developed and rapidly industrializing countries, the major historic air pollution problem has typically been high levels of smoke and sulphur dioxide (SO_2) arising from the combustion of sulphur-containing fossil fuels such as coal for domestic and industrial purpose. (UK Air Quality Archive, 2007) The major threat to clean air is now posed by traffic emissions. According to the emission inventory analysis, road transportation is the main contributor to local air pollution (Cox, 2005). Petrol and diesel-engined motor vehicles emit a wide variety of pollutants, principally carbon monoxide (CO), oxides of nitrogen (NOx), volatile organic compounds (VOCs) and particulates (PM), which have an increasing impact on urban air quality.

Air quality models are designed to simulate the transport, reaction, emission, and deposition of air pollutants. Typical input to air quality models consists of meteorological conditions and emission inventories that provide information on the spatial distribution and rates of emissions of pollutants from power plants, point sources, area sources, mobile sources, and vegetation. Other inputs relate to surface characteristics such as vegetative refraction and roughness. Typical output from air quality models includes the four-dimensional distribution of pollutant concentrations such as those of ozone, nitrogen dioxide, carbon monoxide and dioxide, and others. The primary resources are defined as source emissions, atmospheric impacts, where secondary resources are topography (3-D spatial information), urbanization and activities (Chandler, 1976).

In environmental studies, where air quality is not an exception, measurements are the most relied part to determine the properties of the events measurements. Geostatistics provide optimal methods of spatial interpolation that include estimates of reliability at unsampled locations. Additionally, it provides the means of estimating the effects of uncertainty in the sampled data on the outcomes of spatial analysis and numerical models and particularly of air quality. If the possibility of increasing the amount of collected data arises, geostatistics provides solutions for data collection time, data collection location and selection of measured parameters.

Geostatistics and multi-variate techniques provide the framework and the tools to build a consistent model for "integrating" information from various sources. The technique of cokriging generalizes kriging to multivariate interpolation, where the relationships between the different variables as well as the spatial structure of the data were explored (Chiles and Delfiner, 1999). However, air quality sample stations generally are not dense enough for applying this technique. Kriging is a widely used technique for air quality data, since data is spatially continuous (Sampson and Guttorp, 1999; Shaddick and Wakeeld, 2002; Diem and Comrie, 2002, Nunnari et al., 2004, Potoglou and Kanaroglou, 2005). Among different types of kriging, ordinary and universal kriging are the most commonly used techniques. The general concept is that the prediction of the value **Y(s)**, such as CO concentration, at any **s** location is obtained as a weighted average of neighboring data (Bailey and Gatrell, 1995):

$$Y(s) = \sum_{i=1}^{n} w_i(s) Y(s_i) \tag{1}$$

In both ordinary and universal kriging, the ultimate goal is to estimate the optimal values of the weights. One of the major advantages of kriging is the statistical evaluation of the results and the estimation of confidence intervals around predicted values. For this purpose, the mean square prediction error or kriging variance is used, where the variance is derived from the known covariance structure. For assessing the uncertainty and accuracy; time series plots comparing hourly predicted and observed concentrations for each monitoring station is recommended by Environmental Protection Agency (EPA) (EPA, 2001).

3. CASE STUDY

3.1 STUDY AREA

The Istanbul Metropolitan Area is the only mega-city in the world that sits astride two continents – Europe and Asia, where the transportation system has been unable to keep pace with the rapid growth and changing urban structure. Rapid motorization is generating serious congestion and air pollution in Istanbul due to the high population density and the lack of supporting infrastructure. According to the State Statistical Institute reports, in 2006, every fifth citizen of Istanbul owns a vehicle. Emissions caused by the road traffic in Istanbul were rapidly increased between years of 1990 and 2000, where the increase in CO, SO_2 and particulates were 50.1%, 55.7% and 82.5%, respectively (Becin, 2002). In order to diminish the congestion problem, since 80's the infrastructure of the town is under constant renovation, new roads, motorways are build, bridges laid, under-ground line is under construction. This unfortunately could not solve the problem of congestion and some negative side effects were observed, such as valuable agricultural lands and forest were cut through new routes and industrial and urban areas expand along the transportation network. The air quality measurements are being performed at ten sampling stations, which automatically measures CO_2, CO, NO_x, SO_2, particulates, VOC and NMVOC parameters and several others on daily bases for 365 days. Location of the sampling stations and emission values are provided at the Istanbul Metropolitan Municipality web-page http://www.ibb.gov.tr/tr-TR/HavaKalitesi/. Emission values have been provided as daily averages and data is available for the period of 2000 to 2007. At the Istanbul Metropolitan area, ten sampling stations are performing measurements, where this number was seventeen till 2000. According to the national environmental reports, a population of 10 million requires approximately 35 sampling stations for SO_2 and particulates, where this study results may aid decision makers for re-evaluating the number of sampling stations. During this study only the European side of the Istanbul Metropolitan area was selected, since the spatial distribution of the sampling stations are more condense (six of them are at the European side).

3.2 DATA & METHODOLOGY

Air quality emission parameters, sulfurdioxide (SO_2), carbonmonoxide (CO) and particulates (PM), retrieved from six sampling stations at the European side of the Istanbul Metropolitan area were used for the period 2003-2006. Among other parameters, these were selected in order to relate transport and land-use to air pollution. Measurements conducted by the municipality are considered as accurate, since relevant tests and calibration were conducted. (IBB, 2007) Additionally the impact of atmospheric parameters such as wind, temperature was considered as constant. Time series of each parameter was plotted to understand the behavior of each parameter in multi-temporal scale and find out the best times to conduct kriging experiments. Since all of the parameters were available for each station in 2005, this year was selected to use for further kriging analysis. January and August 2005 data values were employed to explicitly explore the seasonal impact of transportation on total air quality. Different point combinations were used for producing these kriging maps, and effects of point distribution were also investigated.

Since the aim is to determine the contribution of transport and land-use to air quality, sampling points were related with impervious surface such as urban area, road and industrial areas, soil, vegetation and water classes within a determined impact area. For the spatial analysis, Landsat 5 TM satellite sensor image acquired in 2001 was used. Image has 30m spatial resolution for near and mid infrared bands. 1:25000 scale topographic maps were used to geometrically correct this image within 0.5 pixel root mean square error. The first order polynomial transformation and nearest neighborhood resampling method was utilized for the geometric correction procedure. The Iterative Self Organizing Data Analysis Technique (ISODATA) unsupervised classification algorithm was conducted to extract land cover categories in the study area. The classification accuracy was 82%. In order to determine the impact area a Digital Terrain Model was used. Results are illustrated in Figure 1. When sampling stations impact area was compared to total impact area, %14 of Sarıyer (1), %65 of Besiktas (2), %76 of Alibeyköy (3), %83 of Sarachane (4), %94 of Esenler (5) and %81 of Yenibosna (6) have impervious surface. The vegetation values for the same sampling stations are as follows: %50, %9, %10, %10, %1 and %7.

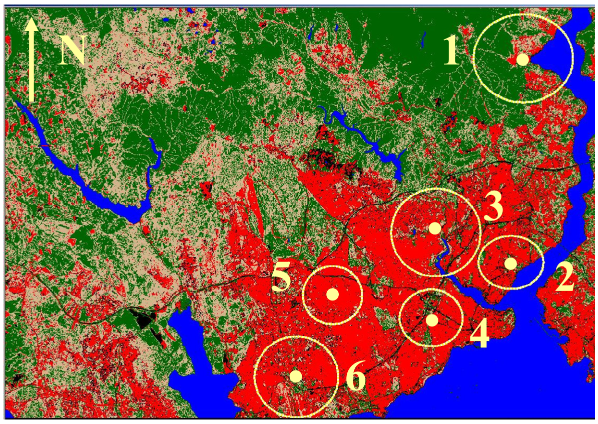


Figure 1. Classification Results and Sampling Stations

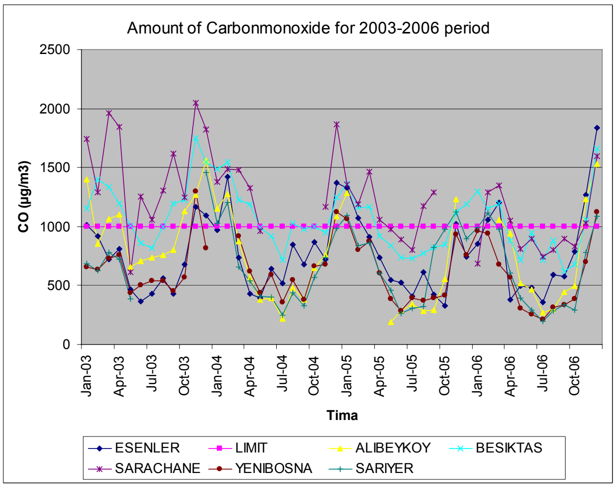
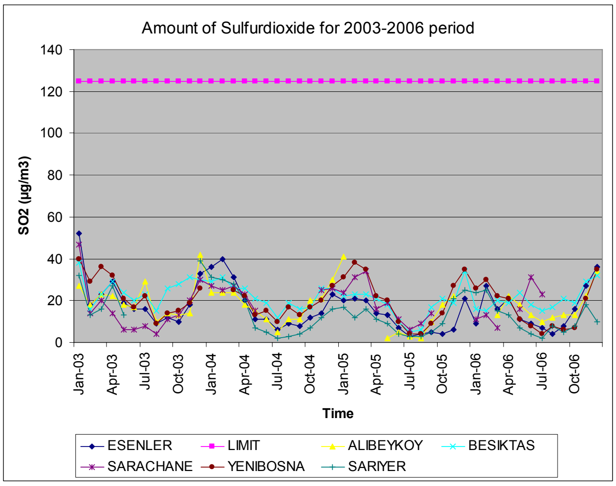


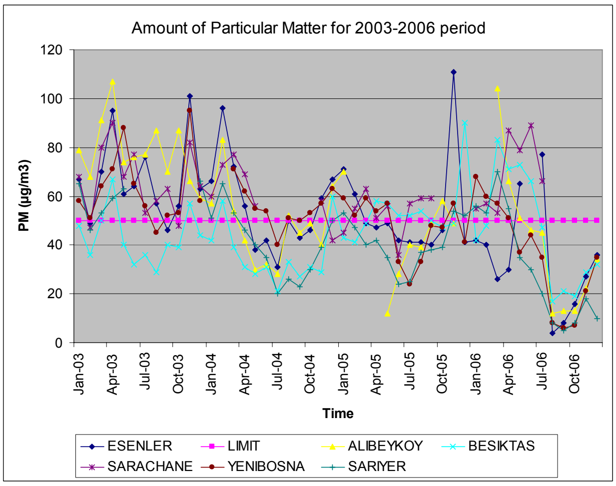
Figure 2. Monthly time series of CO, SO_2 and PM

4. RESULTS&DISCUSSION

For the selected area several statistical and geostatistical methods were applied and results were compared accordingly:

Monthly time series of CO, SO_2 and PM measurements were plotted for January 2003 to December 2006 period, in order to understand general distribution of these data and find out the available seasonality effect (Figure 2).





The obtained results of the CO time series (period of 2003-2006) showed that, the highest values for CO was obtained in winter such as November, December and January months. Comparatively low values can be seen for summer period, such as June and July. For CO, the upper limit value is determined by World Health Organization as 1000 µg/m^3, where limit indicates the reached value is dangerous for human health. For most of the stations the CO limit was exceeded in winter time. In other seasons, the emission values were under this limit.

The time series for SO_2 showed that, the highest value was achieved in winter as 40, where summer values are about 10 especially in June and July. Maximum amount of SO_2 is determined as 125 according to EU standards, where SO_2 amounts belonging to all stations were under the limit. CO and SO_2 values within the period of 2003-2006, showed a sinusoidal trend, where PM has more complex structure. In most cases amount of PM was higher than the maximum value of 50, which was announced by European Union. The most lowest values of CO, SO_2 and PM within the four year period was achieved in Sarıyer (1), where the ratio of impervious surface is %14 and vegetation is %50. The PM values for all stations are higher than the limits in winter time, where the Esenler (5) sampling station reached the highest value in November, 2005. Within this sampling station zone, the hub for regional busses were opened 2005, which can be directly affect the result. For the same sampling station the PM values of June is measured 80 µg/m^3, which is still above the limit.

According to the result of the time series, the air quality emissions for the European side of the Istanbul Metropolitan area increases in winter time and there is a significant decrease in summer. Emission in winter is attributed to house, industrial gases and transport, where in summer season only industrial gases and transport is the source for emissions. Additionally, meteorological conditions of Istanbul in winter time are not mild, heavy fog and snow is frequent. Except Sariyer (1) and Besiktas (2), all sampling stations are located in an highly urbanized part of the city, where dense transportation network, such as Trans-European Network (TEM), Ataturk Airport is located. Descriptive statistics of CO parameter for each station are provided below in Table 1.

	Esenler (5)	Alibeykoy(3)	Besiktas(2)	Sarachane (4)	Yenibosna (6)	Sariyer (1)
Mean	766.54	761.10	1066.35	1241.15	615.61	669.29
Standard Error	48.70	61.30	39.06	58.68	38.30	49.83
Median	722	723.5	1018.5	1251.5	582.5	640.5
Mode	428	1230	1222	1292	760	#N/A
Standard Deviation	337.40	397.25	270.61	371.12	259.77	322.93
Sample Variance	113838.17	157809.21	73227.60	137733.31	67481.35	104283.18
Kurtosis	0.78	-1.09	-0.19	-0.55	-0.13	-0.81
Skewness	0.95	0.34	0.53	0.41	0.67	0.37
Range	1509	1370	1125	1436	1090	1260
Minimum	330	189	623	611	211	199
Maximum	1839	1559	1748	2047	1301	1459
Sum	36794	31966	51185	49646	28318	28110
Count	48	42	48	40	46	42

Table 1. Statistical Results

Kriging maps were created for January 2005 and August 2005. Via kriging maps it is possible to examine the seasonality effect, for the mentioned months since a complete set of CO, SO_2 and PM was available. All following maps are designed as follows; firstly all sampling stations were used in the kriging and then Sarıyer (1), which has the lowest urbanization ratio, was subtracted from the set, which is represented as (b) in all figures. In the third map, named as (c), Yenibosna (6) station is omitted, which was selected randomly. The fourth map was created also from five sampling stations, where the Alibeykoy (3) station was omitted. This station was selected as the fourth option since it has the highest impervious surrounding and data values. The SO_2 maps of the January, 2005 obtained from different point distributions are illustrated in Figure 3. For the kriging procedure, range values obtained from minimum and maximum values of related parameter is very important. When the omitted sampling station value has no significant effect on the range value, such as Sarıyer (1), the obtained new map is very similar to the map which uses all stations data. (illustrated in Figure 3 a/b/c). However, when the omitted value changes the range value, in case of Alibeykoy (3) illustrated in Figure 3d, the obtained new map is different than the maps created. Since the Alibeykoy (3) sampling station has the highest values, omitting this station affects the range value. Hence, a new kriging map is achieved (Figure 3/d). Following this methodology, kriging maps for CO and PM were obtained. Results are very similar of those for SO_2.

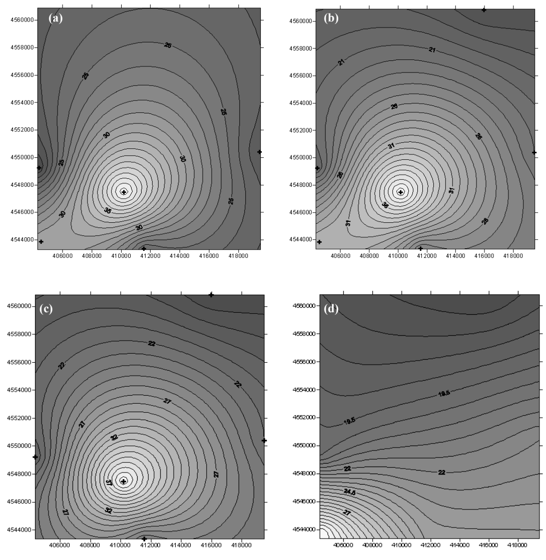


Figure 3. (SO_2) map form the January

These analyze of the resulting differences between observed and predicted values provide the opportunity of evaluating the interpolation results. RMSE of kriging maps obtained for different point distributions for different parameters were illustrated in Table 2 and Table 3. Results showed that, for all parameters, the lowest RMSE values were obtained with the usage of all six sampling stations.

Stations used in Kriging	Parameter	RMSE ($\mu g/m^3$)
All stations	SO_2	14.18
Without Sariyer(1)	SO_2	27.11
Without Yenibosna(6)	SO_2	25.47
Without Alibeykoy(3)	SO_2	20.81

Table 2. Results of Grid Statistics for SO_2

Stations used in Kriging	Parameter	RMSE ($\mu g/m^3$)
All stations	CO	1260.12
Without Sariyer(1)	CO	1331.91
Without Yenibosna(6)	CO	1263.37
Without Alibeykoy(3)	CO	1268.39
All stations	PM	36.32
Without Sariyer(1)	PM	38.04
Without Yenibosna(6)	PM	58.98
Without Alibeykoy(3)	PM	54.96

Table 3. Results of Grid Statistics for CO and PM

To find out the seasonality effect of each variable, for January and August 2005, kriging maps were created using all points and comparisons were performed for these months (Figure 4). As determined from time series, kriging maps of January and August 2005 also illustrated different values for different seasons. The amounts of SO_2, CO and PM in January are considerably higher than the amounts of same parameters in August. SO_2 amount is under the maximum limit for both January and August; however there is approximately 35 $\mu g/m^3$ differences between January and August for some parts of the study area. Amounts of both PM and CO were lower than the maximum limit for August, whereas their amount was significantly higher than the maximum limit in January for the huge part of the study area (Figure 4). In winter season, population of Istanbul is comparatively higher than the summer season as a result of this more vehicles take part in traffic which causes more emission.

In this study, the accuracy of air quality parameter maps was evaluated comparing the results for six sampling stations. The change in spatial support was achieved by interpolating the point observations to the target support using ordinary kriging. The variations of sampling stations were decided using points attributed land-use and vegetation ratios. This study showed that despite large number of observations the interpolation uncertainty was still considerable. System could be used for determining specific impacts of transportation on air quality and suggests location for sampling points, measurement time and duration. Unfortunately, the monitoring data is only available for six sampling stations over the European side of Istanbul and therefore comparison of the spatial pattern between the model and the observations is limited.

This study will be detailed for testing the results of the air quality model, specifically MOBILE5 of EPA, which estimates vehicle emissions depending upon land-use and travel demand. Integrating pollution parameters and measured emissions, new maps can be obtained. However, there exists the risk of insufficiency, particularly when model parameters and inputs have been adjusted so that predictions better match monitoring data. Two other weighting algorithms, being weighting depending upon DTM and depending upon sampling stations and represented areas attribute (e.g. being near to the airport or motorway) can be tested and evaluated.

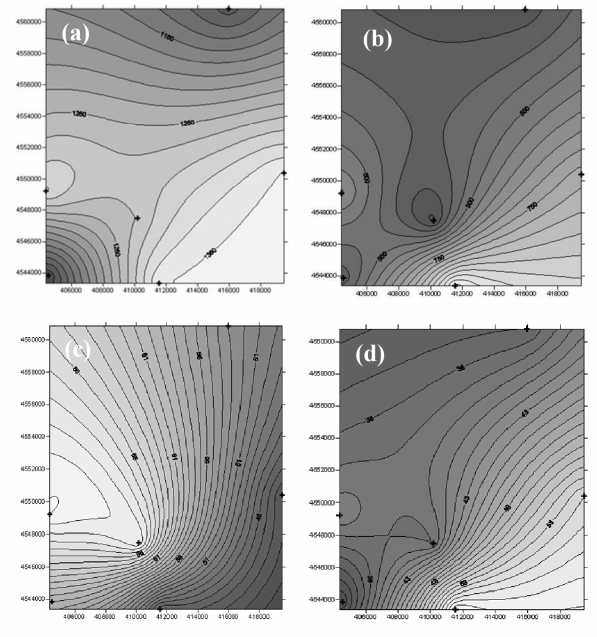


Figure 4. Maps of a) CO in January 2005,
b-) CO in August 2005, c-) PM in January 2005,
d-) PM in August 2005

5. CONCLUSION

Research results indicated that integrated usage of geostatistical methods, remote sensing and spatial analysis can introduce valuable information to identify, visualize and explore the relationship between transportation, land-use and air quality. To obtain more reliable and accurate results, the sampling stations number should be increased, where urbanization and transport network could be represented in a better manner. For illustration purposes, several emission parameters related with transportation activities with varying spatial distribution was estimated and mapped. The integrated approach provided in this study will enlighten the complex task of decision makers and lead to more reliable decisions, since the common approach has been to consider these activities as exogenous information that has loose or no connection.

REFERENCES

Bailey, T.C., Gatrell, A.C., 1995, Interactive Spatial Data Analysis, Longman, Colchester

Becin, A., 2002. Environmental impacts of urban transport, MSc. thesis, Istanbul Technical University, Istanbul, Turkey.

Bierkens, M.F.P., Finke, P.A., Willigen, P., 2000, Upscaling and Downscaling methods for environmental research. Developments in Plant and Soil Sciences, Vol 88, Kluwer Academic Publishers, 190 ppt.

Chandler, T.J., 1976. Urban Climatology and its relevance to urban design, World Meteorological Organisation, Technical Note No: 149, Geneva.

Chilés, J.P., Delfiner,P., 1999. Geostatistics – Modeling Spatial Uncertainty, Wiley Series in Probability and Statistics, pp.5

Cox, J.A.E., Hoff, R.M., 2005. Science–policy data compact: use of environmental monitoring data for air quality policy. *Environmental Science & Policy,* 8, Elsevier, pp. 115-131.

Diem, J.E., Comrie, A.C., 2002. Predictive mapping of air pollution involving sparse spatial observations. Environmental Pollution, 119, pp. 99-117.

Geerlings, H., Stead, D., 2003. The integration of land use planning, transport and environment in European policy and research. *Transport Policy*, Elsevier Ltd, pp.187-196.

Istanbul Metropolitan Municipality (IBB), 2006, retrieved on July 10, 2006 from http://www.ibb.gov.tr/tr-TR/HavaKalitesi/Olcum/ (accessed 14 Feb. 2007)

Leopold, U., Heuvelink, G.B.M., Tiktak, A., Fink, P.A., Schoumans, O., 2006.Accounting for change of support in spatial accuracy assessment of modelled soil mineral phosphorous concentration, *Geoderma*, Elsevier Ltd, pp. 368-386

Nunnari, G., Dorling, S., Schlink, U., Cawley, G., Foxall, R., Chatterton, T., 2004. Modelling SO2 concentration at a point with statistical approaches. Environmental Modelling & Software 19,pp. 887-905.

Potoglou, D., Kanaroglou, P.S., 2005. Carbon monoxide emissions from passenger vehicles: predictive mapping with an application to Hamilton, Canada, Transportation Research Part D 10, pp. 97-109

Sampson, P.D., Guttorp, P., 1999. Operational Evaluation of Air Quality Models, NRCSE Technical Report Series, NRCSE-TRS No:018, pp. 1-22, (http://www.stat.washington.edu/ NRCSE/pdf/trs18_aqmodels.pdf (accessed 14 Feb. 2007)

Shaddick, G., Wakeeld, J., 2002. Modelling daily multivariate pollutant data at multiple sites, Journal of the Royal Statistical Society: Series C (Applied Statistics) 51, pp. 509- 511.

Sperling, D., Hanaoka, S., Okada, A., Okazaki, M., Shade, W., Yagishita, M., 2004. Environmental impacts due to urban transport. In *Urban Transport and the Environment, An International Perspective*, Nakamura, H., Hayashi, Y., May, A. Elsevier Ltd., ISBN: 0-08-044512-8, pp.100-188.

UK Air Quality Archive, 2007, http://www.airquality.co.uk/ (accessed 14 Feb. 2007)

United States Environmental Protection Agency (EPA) (2001). Our Built and Natural Environments, A technical review of the interaction between land-use, transportation and environmental quality, EPA 231_R-01-002, retrieved on July 10, 2006 from http://www.smartgrowth.org/library/built.html (accessed 14 Feb. 2007)