EVALUATION OF OPTIMUM METHODS FOR PREDICTING POLLUTION CONCENTRATION IN GIS ENVIRONMENT

R. Shad [a], H Ashoori [b], N. Afshari [b]

[a] Faculty of Geodesy and Geomatics Eng. K.N.Toosi University of Technology, Roozbeh_Shad@yahoo.com
[b] Dept. of Geomatic Engineering, Shahid Rajaee Teacher Training University, Hamed_Ashoori@yahoo.com , nafshari@yahoo.com

KEY WORDS: Geostatistic, Pollution, Spatial Correlation, GIS

ABSTRACT:

Air pollution is overflowing in big cities, especially in areas where pollution sources and the human population are concentrated. Economic growth and industrialization caused the increasing emissions of air polluting. Then, the quantities of polluting have increased dramatically; the evaluation of a suitable method for predicting and monitoring the pollution will be very important. To prevent or minimize damages of atmospheric pollution, optimum predicting methods are urgently needed which can rapidly and reliably detect and quantify air quality. One of the important spatial analyses for this application in GIS environment is surface simulation using Gostatistical methods. These lead us through creating a statistically valid surface which subsequently is used in GIS models for optimum decision making. Analysis create predicted surface for unmeasured points (Which we have not enough information on them) in study area. For This purpose, Ground stations and MODIS image of Tehran are used for collecting online air pollution information. Then, different geostatistical methods have been used for finding out the optimum prediction method for air pollution, based on received observations. These methods are performed based on spatial relationships (spatial similarity) among the measured points. In here, we use fractal and simple semivariogram for calculating correlation between points and determining which one of them is better for our application. We tested that the fractal dimension which measured by the spatial correlation length is more reliable based on autocorrelation and structural analysis. After that, we proved co-Kriging interpolation is more accurate by producing and evaluating prediction standard error maps.

1. INTRODUCTION

While urban aspects and urban planning often reveal a limited and fixed idea of the concept of the "environment" in our minds, the "environment", is in fact a much wider concept. Tehran, capital of Iran, is one of the polluted cities in the world. Due to its geographical location, the ensnared condition as surrounded by mountain ranges, and also lack of perennial winds, the smoke and other particulate matters produced from daily life do not go far in the air. Therefore, usually there is a thick layer of aerosols and other particulate matters in the nearby atmosphere. Also Economic growth and industrialization in Tehran caused the increasing emissions of air polluting. Then, the quantities of polluting have increased dramatically and the evaluation of a suitable method for predicting and monitoring the pollution is very important. For solving this problem, we use decision support system together with ground station data to evaluate different models of geostatistical air pollutants. This system generally encompass air quality monitoring and air quality modelling of various strategies in support of evaluation of different geostatistical methods by using information to the public about present air quality levels. Examples of researches have been done in major European cities such as Berlin, Geneva, Vienna, Oslo, Athens, Austrian AirWare (Fedra and Haurie, 1999), the Norwegian AirQUIS (Bohler et al., 2002) and the Swedish EnviMan (Tarodo, 2003). A few more decision support systems for specific air quality management studies are in operation in the world (Fine and Ambrosiano, 1996; Schmidt and Schafer, 1998; Jensen et al., 2001; and Lin, 2002).

There are other attempts for air quality monitoring using MODIS data in (Engel-Cox et. al., 2004) and (Tatem, Goetz, and Hay., 2004). In these cases MODIS data represented the collective mass of pollutants and aerosols; while, ground stations collected the amounts of each pollutant separately. In these researches, there is no effort to seek an optimum method for analysis based on collection source data.

Our goal in this research is to evaluate the capabilities of different geostatistical methods to study air pollutants using both satellite and ground stations' data. Then the overall structure of this paper is as follow. In section 2, our study area and its physical properties is described for air pollution application. In section 3, system architecture and different steps for system development is presented. The basic concepts of analysis to support Geostatistical modelling in air quality management are defined in section 3. Finally different methods are performed using simple and fractal semivariogram and the results will be compared.

2. STUDY AREA

Tehran is located at 51° to 51°40′ east longitudes and 35°30′ to 35°51′ north latitudes. This city with about 12 million inhabitants is located at the north side of Iran in the Middle East. The city is located in a basin surrounded by Alborz mountain range in the north. This natural barrier has a strong influence in the meteorological conditions determining the air pollution situation. Industry and traffic is the most air-polluting sector in this city.
There are 13 ground stations across Tehran that record daily amounts of pollutant matters in the city Figure. 1. These records include aerosols, CO, O_3, NO_2 and SO_2. These data are recorded daily and archived in each station. Communication tools and Global Positioning System (GPS) are used for

collecting online information in each monitoring station. Ground pollution survey stations collect daily amounts of pollutants in different parts of the city. However, these ground stations do not have equal distribution across the city and are not sufficient to get a complete figure of sophisticated patterns of pollution in the area.

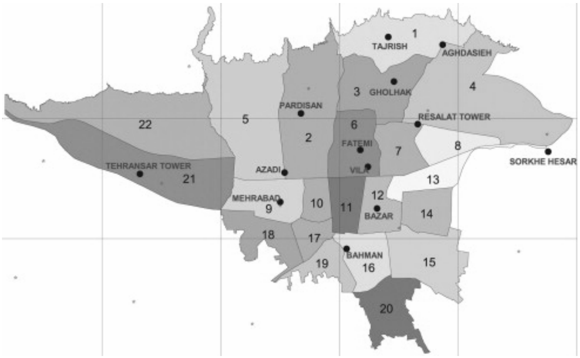

Figure 1. Ground station across Tehran

3. ARCHITECTURE

The diagram in Fig. 3 represents the major steps of the system developed. First, MODIS data (Figure 2) used to obtain information on atmospheric temperatures and humidity (NASA, 2007). MODIS collects its data in 36 spectral bands ranging from 0.4 to 14.5 µm. The spatial resolution of MODIS range from 250 m to 1 km; bands 1 and 2 have 250 m, 3 to 7, 500 m and 8 to 36 have 1 km spatial resolution at nadir. These characteristics are common in both versions of MODIS on Terra and Aqua; however, there are some technical differences between them (NASA, 2007). Bands 1 and 3 cover optical region of electromagnetic spectrum, so were used to collect information on aerosols and particulate matters while band 7 covers infra red region (2105–2155 nm) and is used for calibration purposes only (Martins and et. al., 2002).In Table. 1

Band	Application
1,2	Land/cloud/aerosols boundaries
3-7	Land/cloud/aerosols properties
8-16	Ocean color/phytoplankton/biogeochemistry
17-19	Atmospheric water vapour
20-23	Surface/cloud temperature
24,25	Atmospheric temperature
26-28	Cirrus clouds water vapour
29	Cloud properties
30	Ozone
31,32	Surface/cloud temperature
33-36	Cloud top altitude

Table 1. MODIS bands

MODIS bands are presented for studying air pollution (NASA, 2007).

Quantitative information extraction from MODIS data needs tremendous processing operations. This job has been done by NASA for the users. For this study, Tehran MODIS was downloaded form internet.

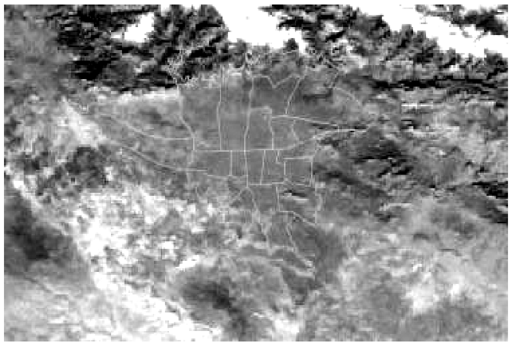

Figure 2. MODIS Image of Tehran

Also, It is necessary to gather the industrial sources specific information, raw materials used, manufacturing processes, fuel consumptions, population data; and the traffic source information on number of vehicles for completion of MODIS data.

Emission sources were broadly categorized as point, line and area sources covering industrial, vehicular and domestic sources, respectively. Amounts of four major pollutants, namely the PM, SO2, NOx (in NO2 units) and CO emitted from these sources were estimated by using fuel consumption data and appropriate emission factors to be multiplied by them. Emission factors used were taken from European CORINAIR database (CITEPA, 1992) and US Environmental Protection Agency (USEPA) emission factors catalogue (USEPA, 1995). Whenever European emission factors were insufficient to indicate the industrial subcategories, EPA emission factors were used. The more information about emission inventory and the emission factors used in the study can be found in a recent study (Elbir and Muezzinoglu, 2004).

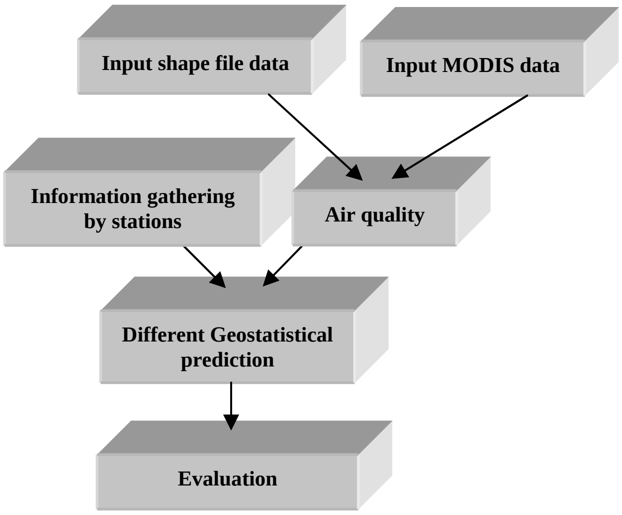

Figure 3. Major steps of system developing

CALPUFF which is a modelling system recommended by the Environmental Protection Agency (EPA) for simulating long-range transport (USEPA, 1998) is designed to be used in air quality applications. The CALMET meteorological model in its basic form produces hourly fields of three dimensional winds and various micrometeorological variables based on the input of routinely available surface and upper air meteorological observations. As input for the CALPUFF modelling system one needs to describe each source according to the emission inventory: stack dimensions, output stack temperature, emission flow and velocity, etc.

Then advanced system is developed using MODIS and emission data on the existing framework process basis which is designed for air quality aided monitoring and prediction in CALLPUFF modeling system. Then in every one of 13 stations, sample data are gathered and prepared for interpolation.

As well, the ground pollution survey stations, which gathered by Communication tools, can be interpolated to produce surface maps of different pollutants.

Different geostatistical methods are used to interpolate the spatial distribution of air pollution samples. The ArcGIS application developed by ESRI was selected for analysing different collected data because of its dynamic Geostatistical environmental models. In this software localization of information extracted of MODIS image, industries, emission patterns, etc can be displayed together.

Finally, different Geostatistical methods are applied on different sources of data and the results are compared and evaluated.

4. USING GEOSTATISTICAL METHODS

Geostatistics is concerned with a variety of techniques aimed at understanding and modeling spatial variability through prediction and simulation. The primary aim of geostatistics is to estimate the spatial relationship between sample values. This estimate is used to make spatial prediction of unobserved values from neighbouring samples and to give an estimate of the variance of the prediction error.

In geostatistics (Cressie, 1993 and Diggle et al., 1998) it is assumed that the data Yi sampled at the locations si, i=1,…,n, are partial realizations of a Gaussian random process $\{Y(s) \mid s \in D \subset Rd\}$ such that, $\forall s \in D$:

$$E[Y(s_0)] = \mu(s) \qquad (1)$$

And the variance

$$\operatorname{var}[Y(s_i) - Y(s_j)] = 2\gamma(s_i, s_j) \qquad (2)$$

Exists. The process is called intrinsic stationary if its semi-variogram γ(si,sj) depends only on the distance between si and sj (and eventually on the direction of the vector h=si−sj) but not on their locations.

The goal of kriging is to predict in an optimal way the value of the process Y(·), at an unsampled location s0 from a linear combination of the observed values Yi. $\hat{Y}(s_0) = \sum_{1-l}^{n} \omega_i Y_i$. The weights ωi are chosen in order to minimize the mean square prediction error

$$MSE = E[(Y(s_0) - \hat{Y}(s_0))^2] \qquad (3)$$

Subject to the unbiased constraints $E[(\hat{Y}(s_0)] = E[(Y(s_0)]$. The resulting predictor has minimum variance, the so called kriging variance, within the class of unbiased predictors.

Fractal dimension predicted using the local variogram. The procedure is described elsewhere. For completeness, the procedure was as follows: (i) experimental variograms were obtained locally within a local moving window, (ii) the variograms were logged on both axes to produce a near-linear plot and (iii) the slope of the line was predicted and this was used to predict the fractal dimension (Klinkenberg and Goodchild, 1992). It was important to base the segmentation on some aspect of the variogram because it is variation in the variogram that affects the sampling strategy. The rationale for using the fractal dimension (rather than, say, any of the coefficients of a fitted variogram model) was that the fractal dimension compressed the essentially two-dimensional (semivariance, lag) information in the variogram into a single variable suitable for processing via simple segmentation algorithms.

Kriging is a Geostatistical interpolation which, initiated by Krige (1951), is the basic statistical methodology for predicting values at un sampled locations based on the indices sampled spatially surrounding the un sampled one. The kriging procedure is optimal in the sense that it gives optimal predictions when the covariance structure is known. The conventional kriging approach consists in plug in estimated parameter values and proceeds as if the estimates were the true values. In practice, the semi-variogram is estimated from the data using the empirical semi-variogram given by

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Y(s_i + h) - Y(s_i)]^2 \qquad (4)$$

Where N(h) is the number of pairs of observations in D which are at distance h. Ordinary kriging refers to spatial prediction under the assumptions that μ(s)=μ is constant and

$$\sum_{i=1}^{n} \omega_i = 1 \qquad (5)$$

(this last assumption guarantees unbiasedness).

Universal kriging refers to spatial prediction under the assumptions that there is a trend in the sample values $\mu(s)=\sum \beta_i f_i(s)$, where the β_i are fixed unknown parameters and the f_i are known functions of the spatial locations chosen to model the trend**.**

If we want to consider several response variables simultaneously co-kriging can be used. It accounts for the spatial cross correlation between primary and secondary variables. Besides fitting a semi-variogram model to the response variable, co-kriging requires to fit a semi-variogram model to the secondary variable and a cross-semivariance model**.**

To estimate the prediction error reliably cross-validation can be used to perform model validation (Davison and Hinkley, 1997). The method which leads to the smallest estimated prediction error is preferred.

5. ANALYSIS

Air pollution which is local phenomena modelled by mentioned methods in section 4 and determined which geostatiscal analyse presents better results in this application. For identifying the optimum model, statistics from the MODIS data, manufacturing processes, fuel consumptions, population data; and the traffic source which combined by the CALPUFF models in GIS, compared by the practical results of interpolation of the ground stations. Sample data in 13 stations gathered based on processed MODIS and online stations collected data and repair for interpolation. Different Kriging procedures were applied by simple and fractal semi-variogram. In order to reveal the possible spatial structure of the response variable, we have computed its semi-variogram. The semi-variograms is fitted using an exponential model.

The results of the mean squared prediction errors with the various methods were estimated using leave-one-out cross-validation and are displayed in Table 2. This procedure removes a single observation at a time from the data set and the model is fitted to the remaining observations. Then the actual outcome Y_i is compared with the predicted outcome $\hat{Y}_i$ using the model based on the remaining $n-1$ observations. The process is repeated n times to obtain an average accuracy that can be expressed by the mean squared prediction error $\sum(Y_i - \hat{Y}_i)^2$ at each location All the kriging methods reduce the mean squared prediction error. Ordinary kriging yields only a small improvement compared to the model with independent errors. Including more than one variable in the kriging process reduced the error to 0.07. As expected Universal Kriging does not improve the prediction with respect to Ordinary kriging.

Ground stations	**Ordinary Kriging**	**Universal Kriging**	**Co-Kriging**
(MSE) Simple Semivariogram	0.091	0.092	0.070
(RMSE) Fractal Semivariogram	0.082	0.090	0.068
MODIS&Shape Files	**Ordinary Kriging**	**Universal Kriging**	**Co-Kriging**
(MSE) Simple Semivariogram	0.079	0.088	0.071
(MSE) Fractal Semivariogram	0.074	0.083	0.066

Table 2. MSE calculation based on observations

The final results show Co-Kriging by fractal semi variogram is present minimum MSE for air pollution quality in Tehran

6. CONCLUSIONS

GIS is one of the important technologies which help us to monitor, analyze and decide on the air pollution quality in the urban area. Obtained data from monitoring stations in Tehran show the pollution levels are going higher. This problem leads us through creating a statistically valid surface which subsequently used in GIS for optimum decision making based on the air quality factors which can be collected as maps, satellite images, and ground stations data. Geostatistical methods are urgently needed for the amount of pollution in everywhere. Then we set ourselves the goal of identifying the best prediction and interpolation method; and identified a procedure which allows the investigation of causal relationships at the same time. Our research shows the fractal semi variogram with co-kriging has shown good performance to use in our spatial decision making system.

REFERENCES

Bohler, T., Karatzas, K., Peinel, G., Rose, T., San Jose, R., 2002. Providing multi-modal access to environmental data–customizable information services for disseminating urban air quality information in APNEE. Computers, Environment and Urban Systems. 26, 39–61.

Engel-Cox J.A., Holloman C.H., Coutant B.W. and Hoff R.M., 2004. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality, Atmospheric Environment. 38 pp. 2495–2509.

Fedra, K., Haurie, A., 1999. A decision support system for air quality management combining GIS and optimization techniques. International Journal of Environment and Pollution 12, 125–146.-Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metallurgicall and Mining Society of South Africa, 52, 119–139.

Fine, S.S., Ambrosiano, J., 1996. The environmental decision support system: overview and air quality application. Environmental Applications, 152–157.

Krige, D. G. ., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metallurgicall and Mining Society of South Africa, 52, 119–139.

Klinkenberg B. and Goodchild M.F., 1992. The fractal properties of topography: a comparison of methods, Earth Surface Processes and Landforms 17 pp. 217–234.

Jensen, S.S., Berkowicz, R., Hansen, H.S., Hertel, O., 2001. A Danish decision–support GIS tool for management of urban air quality and human exposures. Transportation Research Part D: Transport and Environment 6, 229–241.

Lin, M.D., Lin, Y.C., 2002. The application of GIS to air quality analysis in Taichung city, Taiwan, ROC. Environmental Modelling and Software 17, 11–19.

Schmidt, M., Schafer, R.P., 1998. An integrated simulation system for traffic induced air pollution. Environmental Modelling and Software 13, 295–303.

Tarodo, J., 2003. Continuous emission monitoring. World Cement 34, 67–72.

Tatem A.J., Goetz S.J., Hay S.I., 2004. Terra and Aqua: new data for epidemiology and public health, International Journal of Applied Earth Observation and Geoinformation. 6 pp. 33–46.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and Their Application. Cambridge University Press.

Cressie, N.A.C., 1993. Statistics for Spatial Data, revised ed. Wiley

CITEPA (Centre Inter professional Technique de la Pollution Atmospherique), 1992. Corinair Inventory–Default Emission Factors Handbook, Paris, France.

Elbir, T., Muezzinoglu, A., 2004. Estimation of emission strenghts of primary air pollutants in the city of Izmir, Turkey. Atmospheric Environment 38, 1851–1857.

J.V. Martins, D. Tanre, L. Remer, Y. Kaufman, S. Mattoo and R. Levy, 2002. MODIS Cloud screening for remote sensing of aerosols over oceans using spatial variability, Geophysical Research Letters 29 (12) .

USEPA (The UnitedStates Environmental Protection Agency), 1995. Compilation of Air Pollutant Emission Factors-Stationary Point and Area Sources. Research Triangle Park, NC, AP–42.

] Cressie, N.A.C., 1993. Statistics for Spatial Data, revised ed. Wiley

USEPA (The UnitedStates Environmental Protection Agency), 1995. Compilation of Air Pollutant Emission Factors-Stationary Point and Area Sources. Research Triangle Park, NC, AP–42

Elbir, T., Muezzinoglu, A., 2004. Estimation of emission strenghts of primary air pollutants in the city of Izmir, Turkey. Atmospheric Environment 38, 1851–1857

B. Klinkenberg and M.F. Goodchild, 1992., The fractal properties of topography: a comparison of methods, Earth Surface Processes and Landforms 17 , pp. 217–234.

J.A. Engel-Cox, C.H. Holloman, B.W. Coutant and R.M. Hoff, 2004. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality, Atmospheric Environment. 38 pp. 2495–2509

NASA,2007.
(http://terra.nasa.gov/factsheets/Aerosols),accessed: June, 2007.

USEPA (The UnitedStates Environmental Protection Agency), 1998. Interagency Workgroup on Air Quality Modeling (IWAQM) Phase 2 Summary Report andRecommend ations for Modeling Long Range Transport Impacts. (http:// www.epa.gov/ttn/scram/7thconf /calpuff/phase2.pdf.)