J Geograph Syst (2002) 4:193–208

Journal of Geographical Systems

Improvement of satellite derived pollution maps with the use of a geostatistical interpolation method

P.S. Kanaroglou[1], **N.A. Soulakellis**[1], **N.I. Sifakis**[2]

[1] University of the Aegean, Department of Geography, GR - 81100 Mytilini, Greece
(e-mail: Kanaroglou@aegean.gr)
[2] National Observatory of Athens, Institute for Space Applications and Remote Sensing, Greece

Received: 17 July 2001 / Accepted: 11 December 2001

Abstract. The small number of ground stations for the assessment of the spatial distribution of air pollutants motivates the search for methods that make use of satellite images. One such method, known as Differential Texture Analysis (DTA), is used to measure the Aerosol Optical Thickness in the Visible (AOTV), which correlates highly with air quality. With this method, the presence of clouds and/or land cover changes produce patches of missing values. In this paper we demonstrate that universal kriging can be used to obtain reasonable estimates for these missing values. The methodology was applied to a satellite derived AOTV map of the city of Brescia (Italy).

Key words: Satellite, air pollution, kriging, image processing

JEL classification: cøø

1 Introduction

Increasing urban air pollution and its effects are priority issues for environmental scientists and policy makers (EEA 1995). The development of strategies focusing on urban pollution reduction requires reliable information on a variety of gaseous and particulate pollutants. Traditionally, spatial information on the distribution of these pollutants is poor, derived from measurements acquired by a small number of ground point monitoring stations. Such measurements, however, are hardly enough to allow a reliable assessment of the spatial variability of pollutants over extended urban areas. Furthermore, most local governments are forced to balance their budgets by minimizing the operating costs. Such efforts have led to drastic reductions in the number of urban air pollution monitoring stations. A plausible alternative source of information that may allow an evaluation of the spatial distribution of pollution is the use of satellite imagery, also known as Earth observation or remote sensing data. Such

images may provide spatial information at local, regional and global scales (Foody and Curran 1994).

Over the last decade several methodologies and algorithms have been developed that allow the use of remote sensing data for the development of urban air pollution indicators (e.g., Holben et al. 1992, Sifakis and Soulakellis 1996). More specifically, satellite data from high-resolution sensors, such as, Landsat and SPOT, have been used for the calculation of atmospheric transparency indicators. The Aerosol Optical Thickness in the Visible (AOTV), a variable linked to the aerosol mass loading (Cachorro and Tanre 1997), is such an indicator and can be used for the assessment of air quality.

The specific method used in this study to assess AOTV was developed by Sifakis and Deschamps (1992) and is known as Differential Texture Analysis (DTA). For several reasons AOTV cannot be assessed over the entire study area. Most important among them are the presence of clouds and land cover changes over time. As a result, it is very often the case that the obtained map of AOTV missing values over parts of the study area. In addition, problems of validation of the AOTV results arise when ground measurement stations happen to fall into a patch of missing values. To overcome such problems this study investigates the use of geostatistical interpolation methods, in particular of kriging, to obtain reasonable estimates for missing values. In the remainder of this paper we first provide a brief overview of the DTA method and of kriging. We then discuss the data and the method of analysis we use. This is followed by a discussion of the results and the last section offers concluding comments.

2 Theoretical background

Satellite sensors of high spatial resolution (or ground sampling distance) are proven to be useful in monitoring environmental parameters at the geographic scale of an urban area. With respect to air pollution, sensors of this kind cannot provide us with analytical information with regards to a pollutant's physical or chemical nature. They do allow us, however, to estimate the atmosphere's optical parameters, which are known to be highly correlated with certain pollutant loads. An atmospheric parameter, commonly quantified via satellite data, is the aerosol optical thickness (AOT). For a specific point on the Earth's surface, AOT is defined as the integral of the particulate scattering extinction coefficient (K^{p}_{scat}), a dimensionless number expressing the level of optical interactions between light and the atmospheric particles at a specific height from the Earth's surface. The physical interpretation of AOT for a specific point on the Earth viewed by a satellite sensor is the added values of K^{p}_{scat} over the column of the atmosphere, perpendicular to the Earth's surface at the point of interest. It is then that the smaller the content of particles is in the atmosphere, the lower is the AOT value. If the AOT is measured in the visible spectrum then the particles involved are small in size (i.e., aerosols). A pollution-free atmosphere produces negligible AOTV values.

Sifakis and Deschamps (1992) used satellite images to obtain estimates of AOTV through the DTA methodology. Application of this method requires two image sets of the study area. The first image is a ''reference image'' which

should ideally be acquired under exceptionally clear atmospheric conditions. In reality, however, hardly ever the atmosphere over an urban area would be pollution free. Thus, some in situ AOTV measurements can be used to benchmark the quantitative results of mapping. The second image, referred to as the "pollution image", will be acquired during a typical pollution episode over the study area. With respect to a location of interest **s** on the Earth's surface, the AOTV value is approximated by the following formula:

$$\mathrm{AOTV}(\mathbf{s}) = \ln(\mathrm{SD_r}(\mathbf{s})/\mathrm{SD_p}(\mathbf{s})) = \ln(\mathrm{SDr}(\mathbf{s})) - \ln(\mathrm{SD_p}(\mathbf{s})) \quad (1)$$

Where $\mathrm{SD_r}(\mathbf{s})$, $\mathrm{SD_p}(\mathbf{s})$ are the standard deviations of the radiometric values for the reference and pollution images for pixels around point **s**. The basic premise for this formula is that a polluted image is "blurred", rendering similar the radiometric values of nearby pixels. It is thus expected that the standard deviation ($\mathrm{SD_p}$) of these values will be small, and certainly smaller than the standard deviation ($\mathrm{SD_r}$) for values of the corresponding pixels in the reference image.

In practice, however, two kinds of problems may appear. The first is that the satellite images used for estimating the AOTV are usually not completely cloud or haze free. This is especially true in areas that are notorious for the presence of haze. Clouds interfering between a satellite and the study area can create patches of pixels for which no radiometric values are recorded. The second problem is land cover changes in the time interval between the acquisitions of the reference and of the pollution images. Human interventions, especially in the surroundings of urban areas, is a usual phenomenon. It normally involves the conversion of agricultural or forested land to urban land uses. Such changes alter sun light reflectance on the Earth's surface and the resulting radiometric values recorded by the satellite sensors. Because urban land uses produce more uneven sun light reflectance than agricultural land, the tendency is for the standard deviation of radiometric values over a set of pixels to increase. Thus, other things being equal, an artificial increase through land cover changes in the standard deviation of the pollution image relative to the reference image will cause problems to the estimation of the AOTV values.

The DTA algorithm for the calculation of AOT values, implemented into a computer program by Sifakis and Deschamps (1992), makes use of SPOT satellite data. Estimation involves the following general stages:

- Radiometric correction of SPOT-HRV images [Bands-1 (wavelength: green 0.50 – 0.59 μm)]
- Geometric correction of SPOT-HRV images [Bands-1 (wavelength: green 0.50 – 0.59 μm)]
- Calculation of standard deviations of the radiometric values for both the reference and the pollution image. For each site in the study area a window of 30 × 30 pixels is created around it. The standard deviation corresponding to the site is calculated over the resulting 900 pixels.
- Equation (1) is applied to calculate the value of AOT for all sites of the study area.
- For every site that the calculated value of $\mathrm{SD_p}$ is greater than $\mathrm{SD_r}$, the corresponding value of AOT is recorded as missing.

The presence of missing AOT values leads to patches of various sizes in the final AOT map. Since the distribution of air pollution is a spatially

continuous phenomenon, missing AOT values do not afford any physical interpretation. In addition, cross validation of the obtained AOT values with ground based measurements cannot be performed when the ground station falls into a patch of missing values. It is therefore very important to obtain reasonable estimates for these missing values with the help of spatial interpolation methods. Bennett et al. (1984) classify the existing methods into three main categories. 'Ad hoc' encompass the simplest set of methods, while the second category, known as 'cartographic techniques', include methods used in computerized map drawing from sample data. The third category, referred to as 'distribution based techniques', include variations of the geostatistical method known as kriging. As Haining (1990) points out, an advantage of the latter methods is that they allow the estimation of interpolation errors. For this reason we focus exclusively on kriging in this paper.

Several types of kriging methods have been developed. Those used most often are ordinary and universal kriging. Both are methods of optimal local spatial prediction. This means that, given a random variable Y with known or measured values at sites $\mathbf{s_1}$, $\mathbf{s_2}$,. . .$\mathbf{s_n}$, the methods focus on associating a site **s** in the study area with the best possible value for Y on the basis of the available information. Following the notation of Bailey and Gatrell (1995), in both cases the value to be predicted is expressed as the weighted linear combination of the known values:

$$\hat{Y}(s) = \sum_{i=1}^{n} w_i(s) Y(s_i) \tag{2}$$

The objective then in both cases is to estimate the optimal values of the weights $\mathbf{w_i(s)}$, i = 1,2,…,n. This is accomplished by selecting the weights so that the *expected mean square error* becomes as small as possible. The minimization problem leads to the choice of weights, which in matrix notation are:

$$w_+(s) = C_+^{-1} c_+(s) \tag{3}$$

The problem formulation in equation (2), and equation (3), which provides the solution to the problem, are the same for both ordinary and universal kriging. The difference between the two methods is that universal kriging takes into account the global trend of the data, while ordinary kriging assumes the global trend is known or that the data have been detrended. This difference requires that the definition of matrices in the above formula is different. For ordinary kriging $\mathbf{w_+(s)}$ is a column vector with n + 1 elements. The first n elements are the weights to be estimated, while the $(n + 1)^{th}$ element is a Lagrange multiplier that enters because of the minimization of the mean square error. $\mathbf{c_+(s)}$ is also a column vector with n + 1 elements. The first n of these elements are the covariances $\mathbf{C(s, s_i)}$ between the prediction point **s** and each of the n sample sites, while the last element is simply 1. $\mathbf{C_+}$ for ordinary kriging is the augmented variance-covariance matrix of the observations, as follows:

$$C_+ = \begin{bmatrix} C(s_1,s_1) & \cdots & C(s_1,s_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C(s_n,s_1) & \cdots & C(s_n,s_n) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}$$

For universal kriging the matrices are augmented by the independent variables that are used to estimate the global trend.

Ordinary kriging is often used when there is no global trend. If exploratory analysis of the data indicates a strong trend, then universal kriging is preferred. Computational problems arise when the number of observations n is large. With satellite images this is normally the case. A SPOT image, for example, is 3,000 × 3,000; or 9,000,000 pixels. In such cases matrix $\mathbf{C}_+$ with dimension n × n becomes large and its inversion is computationally challenging. Once the inverse of the matrix has been obtained, however, then it can be used for all the required prediction points. Another reason for voluminous calculations in kriging is that the number of prediction sites is large and the exercise is repeated for all the sites. In many instances it is advantageous to establish a neighbourhood around a prediction point and perform ordinary kriging within this area. One has to make sure that the area is large enough to contain a statistically substantial number of observations, but at the same time, small enough in order to avoid significant computational problems. The recent literature indicates that the battle to improve computing efficiency in kriging continues (Hessami et al. 2001).

Equation (3) indicates that the estimation is dependent on covariances. Estimates of those covariances are obtained with the help of a semivariogram or variogram, for short. Thus, a computer program that performs kriging requires a variogram among its other inputs. No matter what strategy one is to follow in kriging, it is in general advisable that the variogram be estimated for the whole study area. It is well known (Isaaks and Srivastava 1989) that good kriging estimates are very sensitive to the choice of an appropriate variogram. Thus, one needs to experiment significantly before a final variogram is selected for the analysis. Another approach, proposed by Barry and Pace (1997), is to base kriging on the covariogram rather than the variogram. The advantage of a latter is that it possesses a statistically robust estimator, but as the authors point out, kriging based on covariograms can be computationally efficient.

For statistical evaluation of the obtained results and creation of confidence intervals around predicted values the *mean square prediction error* or *kriging variance* is used:

$$\sigma_e^2 = \sigma^2 - c_+^T(s) C_+^{-1} c_+(s) \tag{4}$$

where the variance σ^2 is derived from the known covariance structure. The mean square prediction error is basically the minimized expected mean square error when the estimated weight values are substituted in it. Another way of evaluating the results of kriging is through *cross-validation*. According to this method each observation in turn is deleted and its value is predicted using the same kriging methodology. This results in a set of n error values, which are the differences between the predicted and observed values. The errors are then studied in order to evaluate the predictions.

3 Data sets and methodology

Brescia is a city of approximately 200,000 inhabitants located in the Po Valley, a heavily industrialised area of Northern Italy. This area is renown for a high frequency of smog episodes associated to low visibility, especially during winter months. The SPOT satellite images covering the greater Brescia area were selected and purchased as part of the ICAROS Project (funded by the European Commission). It was specified that the images fulfill the following criteria:

- Both images should be cloud free.
- The image with the lower pollutant concentrations, according to ground monitoring station measurements, is selected as the "reference" image. The other image then should be taken as the "pollution" image.

The orbital and other technical parameters of the selected images are presented in Table 1. The two images that were used as input to the DTA computer program are shown in Fig. 1.

The software used to process the two SPOT images was ERDAS Imagine v.3.3.1 for Windows NT. On a Pentium III, 733Mhz computer with 128 Mb RAM the approximate processing time of the DTA code application was 5 hours. The main remote sensing techniques involved for the AOTV calculation and mapping were:

– The radiometric correction of the satellite data, resulting to the transformation of the images' digital numbers into the corresponding reflectance values.
– Geometric correction of the satellite data, resulting to obtain a georeferenced pair of satellite images.

Table 1. Orbital and technical parameters of the two SPOT satellite images used for AOT mapping over the city of Brescia, Italy

Image parameters	Acquisition date	
	04 January 1990 "Pollution"	6 August 1998 "Reference"
K, J	059–258	059–258
Time	10:03	10:19
Instrument	HRV 1	HRVIR
Processing level	1B	1B
Spectral mode	XS	XI
Number of spectral bands	3	4
Orientation angle [deg]	9.1	11.7
Sun Azimuth angle [deg]	160.6	149.6
Sun Elevation angle [deg]	19.5	58.6
Absolute calibration gains XS1 [$W/m^2/sr/\mu m$]	1.00107	1.93500
Absolute calibration gains XS2 [$W/m^2/sr/\mu m$]	0.94591	2.28786
Absolute calibration gains XS3 [$W/m^2/sr/\mu m$]	0.90668	1.62000
Number of lines	2987	2999
Number of pixels per line	3924	3180
Latitude of scene center	N0453604	N0453257
Longitude of scene center	E0102950	E0103342

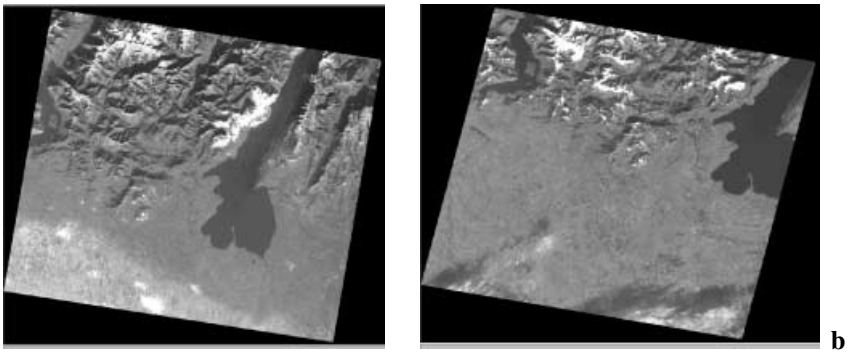

Fig. 1. SPOT satellite images selected for the calculation of AOT over Brescia. **a** ''Reference'' image and **b** ''Pollution'' image

– Application of the DTA computer code resulting in the calculated AOTV values over the study area.

The resulting image is shown in Fig. 2. For the estimation of missing values S-PLUS and the associated Spatial Statistics module were used.

4 Results and discussion

As described in the theoretical section, the missing values of the AOT map, shown in Fig. 2, are estimated through spatial interpolation. Our first task is to produce a variogram as good as possible. An omnidirectional variogram of the data, as shown in Fig. 3, indicates that the variogram values increase

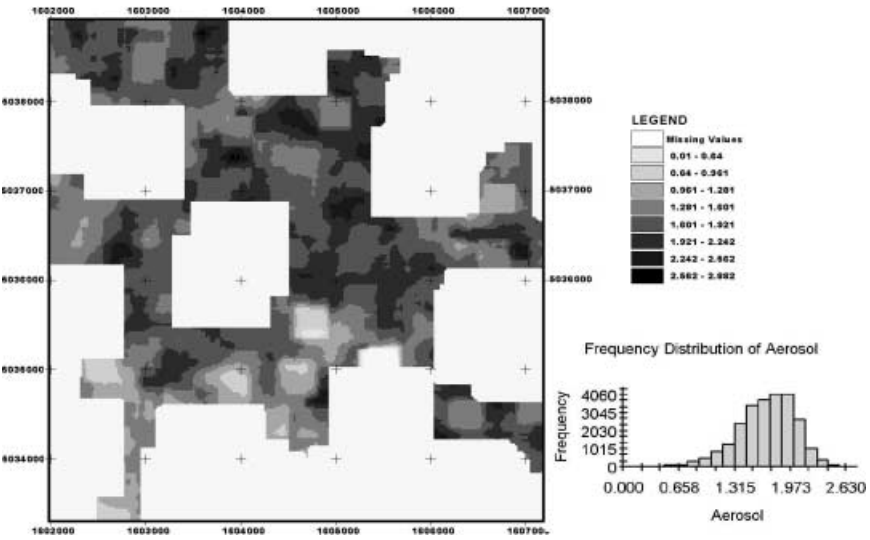

Fig. 2. AOT map over the city of Brescia, Italy

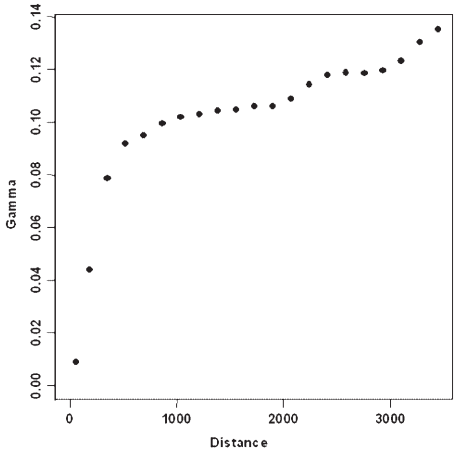

Fig. 3. Omnidirectional variogram of AOT levels

with distance, failing to reach an upper level or *sill*, as one would expect from theory. This indicates that probably first order effects are present and that it would be necessary to remove the global trend. To verify this assertion a directional variogram was produced as shown in Fig. 4. It is obvious that for certain directions no sill is reached. Furthermore, tests of anisotropy, shown in Fig. 5, indicate the presence of geometric as well as zonal anisotropy.

To remove the trend we fit a second degree trend surface model. A third degree model did not represent any noticeable improvement in fit over the second degree model. A histogram of the residuals from the trend surface model, as shown in Fig. 6, exhibits good properties of normality. An omnidirectional variogram of the residuals, shown in Fig. 7, indicates that

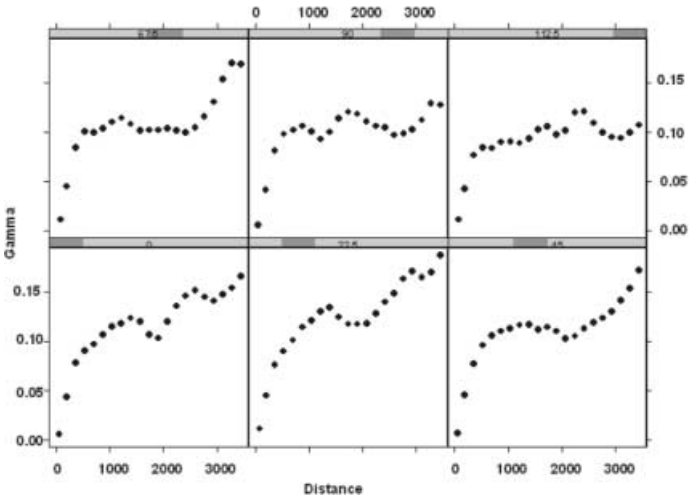

Fig. 4. Directional variogram of AOT levels

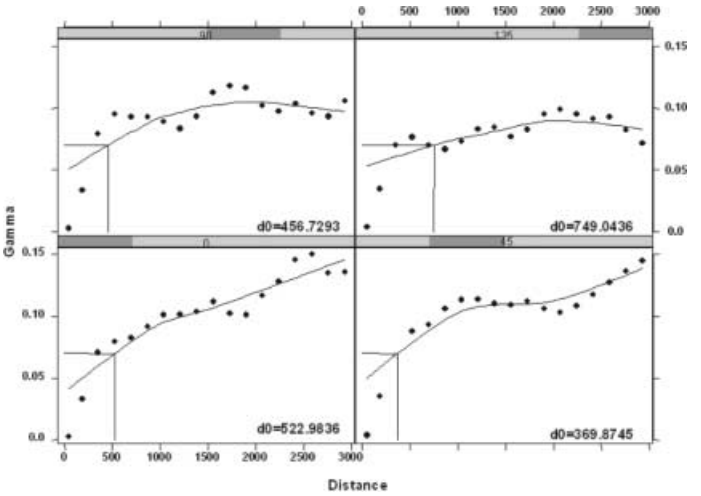

Fig. 5. Anisotropy test for AOT data

the trend is no longer present as a sill is now obviously reached. Directional variograms of residuals, shown in Fig. 8, indicate that the shape of the variogram is preserved in all four directions examined. Thus, in the remaining analysis we work with the omnidirectional variogram.

One potential numerical problem for the analysis is the large number of observations. This is because kriging requires inversion of the variance – covariance matrix of observations, the dimension of which is the square of the number of observations. To be able to tackle the problem with a conventional desktop microcomputer we resort to working with a random sample from the total number of observations. We tried various sample sizes. It appears that with a sample size of less than 1,500 observations the characteristics of the variogram deviate from those of a variogram that is estimated from the population of the observations. We establish a sample

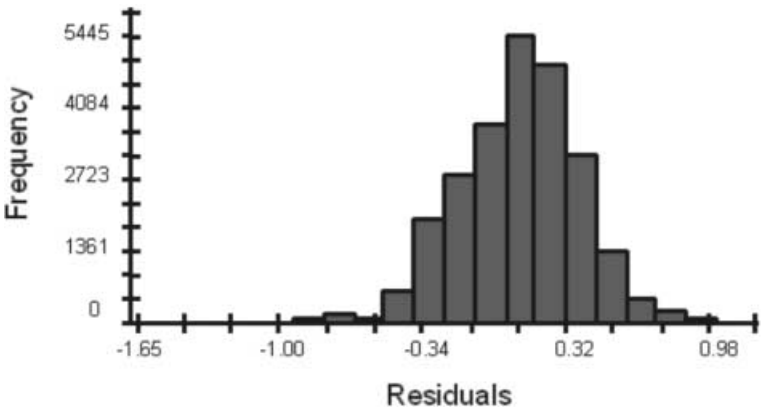

Fig. 6. Histogram of residuals from a trend surface model

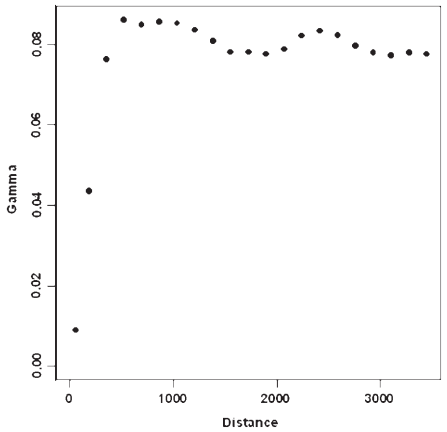

Fig. 7. Variogram of the residuals

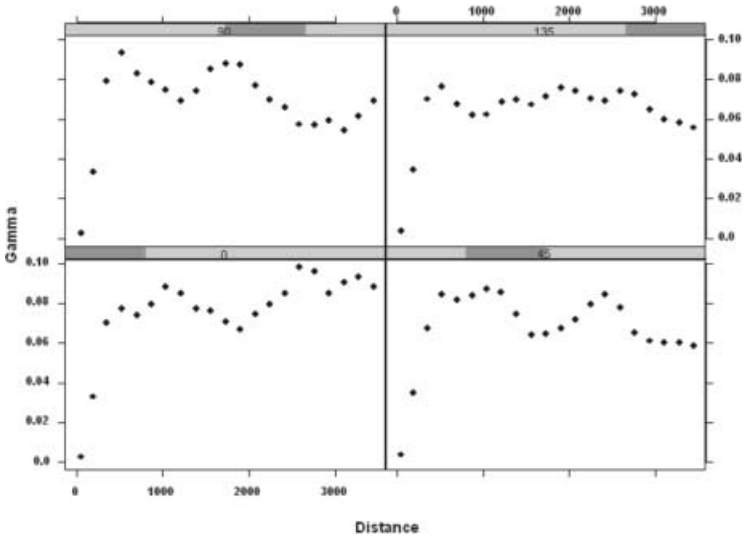

Fig. 8. Directional variograms based on the residuals

size of 3,000 that is small enough to work comfortably, but in the same time it is large enough to inspire confidence in the obtained results. Thus, the variogram in Fig. 9 obtained from a random sample of 3,000 observations is almost identical to that of Fig. 7.

We fit several model variograms to the empirical variogram. The results of this effort are shown in Table 2. Obviously, the spherical model provides the best fit and as such it is adopted for the ensuing analysis. The results of universal kriging with the help of the spherical variogram obtained earlier are shown on Fig. 10. The *mean square prediction error*, otherwise known as *kriging variance*, provides the means for deriving confidence intervals for the predictions. The square root of the kriging variance is the *standard error of*

Table 2. Least squares estimate of various variogram models

Model	Nugget	Sill	Range	Slope	Alpha	Standard error	Log (like)	AIC	AICC	Schwartz
Spherical	0.020	0.796	330.067	–	0.253	0.282	−39.135	84.269	85.869	87.102
Gaussian	0.020	0.796	262.816	–	0.254	0.282	−39.281	84.562	86.162	97.365
Exponential	0.054	0.080	196.300	–	0.067	0.282	−57.232	120.464	122.064	123.297
Linear	0.077	–	–	1.00E−06	–	–	−45.991	95.981	96.731	97.870

Alpha = proportion of nugget effect

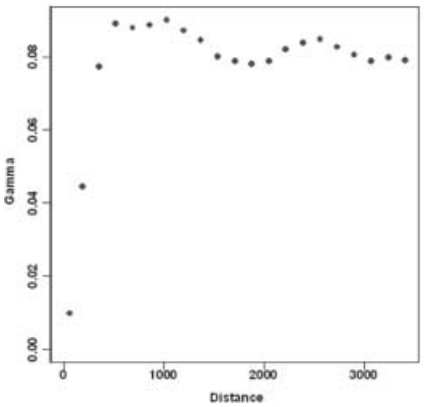

Fig. 9. Variogram based on random subset of residuals

kriging predictions and is shown in Fig. 11. One can see that within the larger patches of missing values the standard error is reasonably small, about 0.3. Thus, with a 5% confidence interval the true value is within ± 1.96 * $0.3 = \pm$ 0.59 of the estimated value in the worst cases.

Cross-validation of the kriging predictions demonstrate that the results are reasonably good. Predictions made with the chosen scheme depend heavily on the variance – covariance structure chosen, which in turn depends on the estimated variogram. Furthermore, the trend surface estimated automatically with the universal kriging scheme is also a significant factor for good predictions. To evaluate such issues we make use of cross-validation, which involves the deletion of one observation and its estimation through kriging from the remaining observations. The difference between an observed and its predicted value provides a residual. This process is repeated for all the observations of the data set. The resulting residuals are analyzed with conventional statistical methods. Here the case prediction of the values was carried out using the spherical variogram that we estimated earlier and with the help of ordinary kriging. This type of kriging assumes that the trend surface is known. For this reason kriging is normally performed in the neighbourhood of the point to be predicted and on the assumption that the trend within this neighbourhood is constant. The trend of a neighbourhood in this case is defined by the number of nearest neighbour observations to the point of interest.

This analysis was carried out for 4, 6, 8 and 12 neighbours, as shown in Table 3. It appears that the distribution of errors for all cases is very similar. In all cases the mean error is close to zero, as expected. The histogram of residuals for 4 and 8 neighbours shows a symmetric distribution that is close to normal. The plot of residuals vs predicted values shows that there is over and under prediction for isolated cases, as shown in Fig. 12.

5 Conclusions

Aerosol optical thickness in the visible (AOTV) is an indicator of atmospheric pollution. AOTV values can be obtained with the help of two

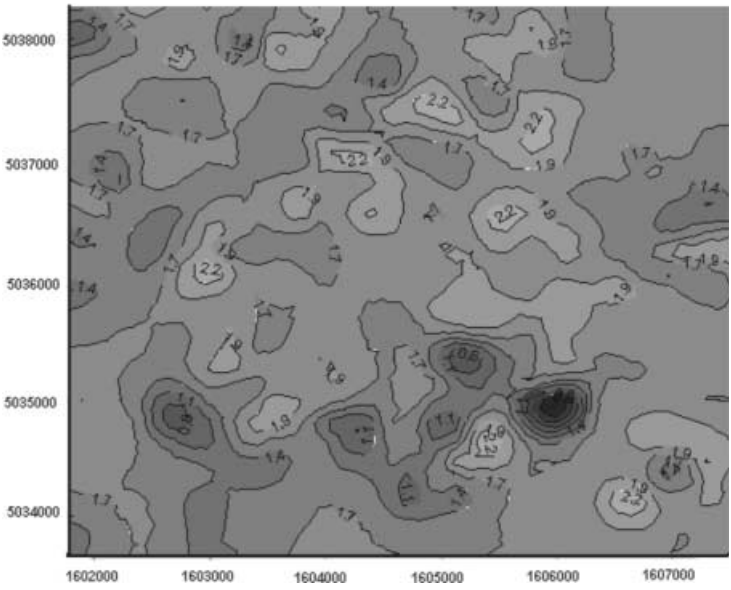

Fig. 10. Kriging prediction of Aerosols from random subset

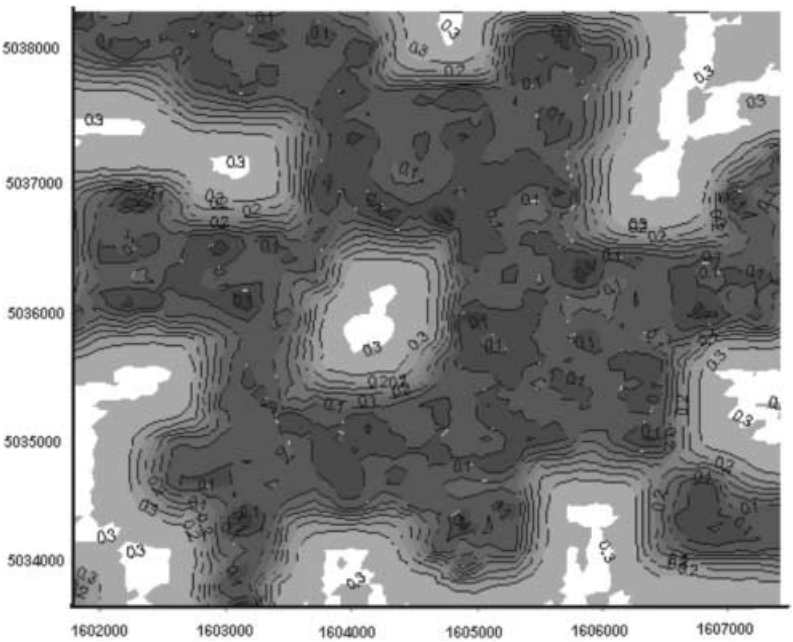

Fig. 11. Standard errors of Kriging prediction of Aerosols from random subset

Table 3. Summary of residuals

Neighbours	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
4	−0.65920	−0.03619	−0.00288	0.00004	0.03276	0.52530
6	−0.67420	−0.03782	−0.00356	0.00125	0.03341	4.20700
8	−0.64190	−0.03334	−0.00247	−0.00034	0.03002	0.54830
12	−0.65780	−0.03443	−0.00296	0.00007	0.03200	0.52460

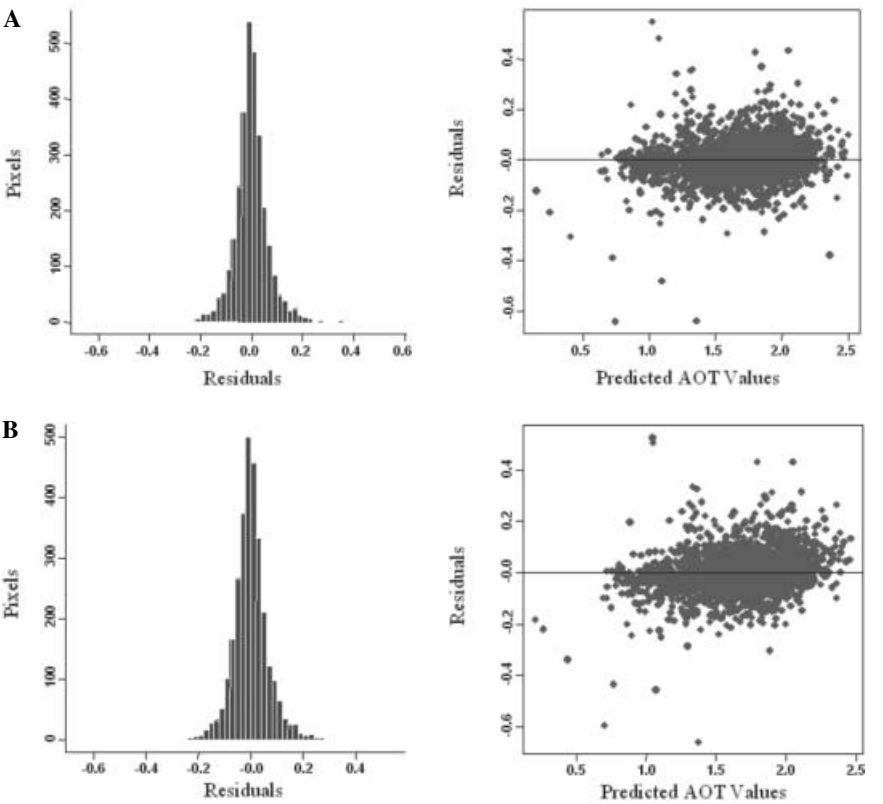

Fig. 12. Histograms and scatter plots of residuals for (**A**) 4 and (**B**) 8 neighbour

satellite images one that is used as ''reference'' and one used as ''pollution'' image according to the image processing DTA code proposed by Sifakis et al. (1992). When applying the code, problems with land cover changes or clouds that obscure parts of the input images force the computer program to produce patches of missing values on the final AOTV map. To produce estimates of the missing values this study examined the use of kriging. This method was applied to relatively small images apportioned from larger SPOT satellite images that correspond to the city of Brescia (Italy). It was shown that through kriging reasonable estimates of the missing values could be obtained.

Since a full SPOT satellite image consists of 3,000 × 3,000 pixels numerical challenges lie ahead for the method to be fully operational. However, it was already shown here that a good result could be achieved with a random sample of the total number of pixels. An alternative methodology would be to establish a neighbourhood for each point of interest and perform the prediction with observations from within the neighbourhood. It would be reasonable to consider, in this case, that because of the small size of the neighbourhood a trend is absent. Therefore, ordinary kriging could be applied, which is also computationally less demanding than universal kriging.

A fruitful area of future research would be the extension of the existing DTA algorithm to substitute automatically missing AOTV values with estimated ones through kriging. This can be achieved with an additional computer program, which would undertake control automatically after the first program has finished estimating AOTV values. Input to this second program would be a variogram that has been estimated for the whole area of study. The program will perform a pass of the AOTV maps and for every site that is associated to a missing value, a square window of pixels will be established around it. Ordinary kriging will be applied to estimate the missing value at the site of interest based on known values of AOTV within the window.

References

Bailey TC, Gatrell AC (1995) *Interactive spatial data analysis*, Longman Scientific and Technical, Essex England

Barry RP, Pace RK (1997) Kriging with large data sets using sparse matrix techniques, *Communications in Statistics, Simulation and Computation*, Marcel Dekker Inc. NY, 26: 619–629

Bennett RJ, Haining RP, Griffith DA (1984) The problem of missing data on spatial surfaces, *Annals, Association of American Geographers*, 74:138–156

Cachorro VE, Tanre D (1997) The correlation between particle mass loading and extinction. Application to Desert Dust Aerosol Content Estimation, *Remote Sensing of Environment* 60:187–194

EEA (European Environment Agency) (1995) Europe's environment: The dobris assessment, Chapter 10: The Urban Environment Copenhagen (eds) D. Stanners and Ph. Bourdeau

Foody G, Curran P (1994) *Environmental remote sensing from regional to global scales*. Wiley publications, England

Haining RP (1990) *Spatial data analysis in the social and environmental sciences*, Cambridge University Press, Cambridge, England

Hessami M, Anctil F, Viau AA (2001) Delaunay implementation to improve kriging computing efficiency. *Computers and Geosciences*, 27:237–240

Holben B, Vermote E, Kaufman YJ, Tanre D, Kalb V (1992) Aerosol retrieval over land from AVHRR data – Application for atmospheric correction. *IEEE Transactions on Geoscience and Remote Sensing*, 30:212–232

Isaaks EH, Srivastava RM (1989) *An introduction to applied geostatistics*, Oxford University Press, Oxford, England

Sifakis N, Deschamps PY (1992) Mapping of air pollution using SPOT satellite data. *Photogrammetric Engineering and Remote Sensing,* LVIII:1433–1437

Sifakis N, Soulakellis N (1996) Air-pollution observations and mapping from Earth-observation satellites: Potentials and limitations, *Proceedings of the International Conference on Protection and Restoration of the Environment*, Technical University of Crete, Chania, Greece, pp. 370–378

Sifakis N, Soulakellis N, Paronis D (1998) Quantitative mapping of air pollution density using EO. A new processing method and application to an urban area, *International Journal of Remote Sensing*, 19:3289–3300

Sifakis N, Soulakellis N (2000) Satellite image processing for haze and aerosol mapping (SIPHA): Code description and presentation of results. *Proceedings of IGARSS-2000 Conference*, Hawai, 1:222–224