Spatial interpolation techniques for stimating levels of pollutant concentrations in the atmosphere

D. Rojas-Avellaneda

Centro de Investigación en Geografía y Geomática, "Ing. Jorge L. Tamayo",
Contoy No.137, Lomas de Padierna, Tlalpan, 14740 México, D.F.,
e-mail: dariorojas@centrogeo.org.mx,
Tel. 52(55)2615-2289 ext. 109

The inverse distance-weighting method (IDW) and kriging techniques are the most commonly used spatial interpolation techniques for estimating levels of pollutant concentrations in regions that contain a number of monitoring stations. The measured ozone pollution peaks in a period, in the atmosphere of the Mexico City region, are considered to be a sampled data set with a non-stationary mean. In order to study the effect of a non-stationary mean in the performance of interpolation methods IDW and kriging, the data set is transformed by removing the data trend of the sampled data set. The residuals obtained are considered to be a set of stationary random variables. This work initially considers the residuals obtained from measured ozone concentration data at 20 stations at 15:00 hours for a set of 21 days in December, 2001. This set of 420 data is considered to be the training set. To determine the parameter values that define the statistical weights for each of the IDW and kriging methods that are analyzed in this work, a cross-validation method is considered. This method assumes initial parameter values, which are fitted by minimizing the root mean squared error, RMSE, between the observed and estimated values in each of the stations. This process takes the training set in consideration for calculation. Once the parameter values that define the statistical weights for each IDW and kriging methods are obtained, by the process described above, these methods are used to interpolate its corresponding values at the stations at 15:00 hours for the days (3rd, 6th, 9th, ... 27th, 30th) of December, 2001, which are considered to be the testing sets. The RMSE between interpolated and measured values at monitoring stations is also evaluated for these testing values and is shown as a percentage in Table I. These values and the defined generalization parameter G can be used to evaluate the performance and the ability of the models to predict and reproduce the peak of ozone concentrations when the residuals or the sampled data are considered. Scatter plots for testing data are presented for each interpolation method. An interpretation of the ozone pollution levels obtained at 15:00 hours on December 21st was given using the wind field that prevailed in the region at 14:00 hours on the same day.

Keywords: Spatial interpolation; statistical modeling; pollutant concentrations.

Las técnicas de interpolación espacial, peso inverso con la distancia (IDW) y *kriging*, son las más comúnmente usadas para la estimación de niveles de contaminante en regiones que tienen un limitado número de estaciones de monitoreo. Los valores del pico de contaminación por ozono, medidos en la atmósfera de la región de la Ciudad de México, se consideran como un conjunto de datos muestreados cuya media no es estacionaria. Con el fin de estudiar el efecto de una media no estacionaria sobre el desempeño de los métodos de interpolación IDW y kriging, se transforma el conjunto de datos al remover de cada uno de ellos el valor de su tendencia. El conjunto residual obtenido se considera como un conjunto de variables aleatorias estacionarias. Para este caso se considera inicialmente el conjunto residual obtenido de los datos medidos en las 20 estaciones para concentración de ozono a las 15 horas por un periodo de 21 días del mes de Diciembre del 2001. Este conjunto de 420 datos constituye el conjunto de entrenamiento. Para determinar el valor de los parámetros que definen los pesos en cada uno de los métodos IDW y kriging que se analizan en este trabajo, se considera un método de validación cruzada mediante el cual se suponen para los parámetros valores iniciales, que se van ajustando iterativamente hasta obtener el valor que produce el mínimo error cuadrático medio entre los datos medidos y los estimados en cada una de las estaciones, para lo cual hacemos uso de los datos que constituyen el conjunto de entrenamiento. Una vez determinados, por el procedimiento anterior, los valores de los parámetros que definen los pesos de cada uno de los métodos IDW o kriging, se usan estos métodos para hacer estimaciones de los valores de las concentraciones de ozono, a las 15 horas en las estaciones para los 10 días de Diciembre de 2001 no considerados en el conjunto de entrenamiento. El error cuadrático medio entre datos medidos y estimados es calculado para este conjunto de prueba y se muestra en porcentaje en la Tabla I. Estos valores y el parámetro de generalización G pueden ser usados para medir el desempeño y habilidad de los modelos para predecir y reproducir el pico de ozono tanto para los residuales como para los datos originalmente muestreados sin ninguna transformación. Se muestran gráficas de dispersión de los datos de prueba para cada método de interpolación. Se da una interpretación de los niveles de contaminación de ozono obtenidos para Diciembre 21 de 2001 a las 15 horas usando el campo de vientos preexistente en la región, a las 14:00 horas del mismo día.

Descriptores: Interpolación espacial; modelamiento estadístico; contaminantes en el aire.

PACS: 92.60.Sz; 02.50.Tt; 02.50.Sk

1. Introduction

The measurement of pollutant concentrations is important in the study of urban and regional smog problems to determine zones where a high enough concentration may affect human, animal and vegetation health.

Many cities exhibit photochemical smog including Mexico City. Photochemical smog is the air pollution resulting when sunlight intensity and emissions from fossil-fuel combustion sources are high.

Ozone concentrations are strongly linked to meteorological conditions and are also influenced by winds due to the transportation of both ozone and its precursors. Ozone, after suspended particulate matter, is the second most important pollutant found in Mexico City's atmosphere. In fact, the maximum allowed ozone concentration level (0.11 ppm, one hour per year) was exceeded in 2000 on approximately 84% of the days. This percentage of days has been continually reduced thanks to enacted control norms and to improvement in fuel quality, gasoline particularly. In 2005, the maximum ozone concentration level permitted was exceeded 60.5% of the days (Secretaria del Medio Ambiente, Gobierno del Distrito Federal, March, 2006). Although these pollution levels are far from ideal, these regulations have noticeably improved the air quality in the Mexico City region.

Unfortunately, due to high costs, it is only possible to measure pollutant concentrations at some few monitoring sites regarding the extension of the region of interest.

2. Spatial interpolation methods

Spatial interpolation techniques are commonly used for estimating levels of pollutant concentrations in regions that contain a number of monitoring stations. The interpolation techniques are based primarily on distance-weighting methods (De Leeuw *et al.* 1997, Phillips *et al.* 1997) and kriging (Mulholland et al 1998, Phillips *et al.* 1997).

2.1. IDW

The inverse distance-weighting method (IDW) is based on the intuitive idea that nearer observations must have more influence on the estimated value than farther ones. This is a local method for the estimation of Z on x_0 with the following expression:

$$\hat{Z}(x_0) = \frac{\sum_{i=1}^{N} w(d_i^{-1}) z(x_i)}{\sum_{i=1}^{N} w(d_i^{-1})} \qquad (1)$$

where $w\,(\cdot)$ is the weighting function of the inverse of the distance d_i between the observation at x_i and the interpolation point x_0. Equation (1) is referred to as the standard IDW interpolation for the simplest weight definition $w_i(d_i^{-1}) = d_i^{-\alpha}$s with $1 \leq \alpha g \leq 4$. This weight comprises a monotonically decreasing function that vanishes as the distance tends to infinity.

2.2. Optimized IDW

In practice, it is desirable for the method to be flexible enough to optimize the datasets by limiting the radius of influence for the weighting function and exploring with different decay exponents. The optimized IDW is an attempt to provide this flexibility, and it has the advantage that these two parameters are chosen optimally according to a minimum root mean

square error (RMSE) criterion. The interpolated Zvalue for the optimized IDW can be obtained through the following expression,

$$\hat{Z} = \frac{\sum_{n=1}^{N} w_n Z_n}{\sum_{n=1}^{N} w_n}. \qquad (2)$$

The weights are given by;

$$w_n = \begin{cases} \frac{K}{1+(K-1)(d_n/r)^\alpha}, d_n \leq r \\ 0, \ d_n > r \end{cases}, \qquad (3)$$

The parameters α and r can be estimated by minimizing the square root of the mean-square differences between the measured and the estimated value, and d_n is the discrete distance variable. The parameter K is a scaling constant that makes the weight at $d= 0$ finite rather than infinite, as in the standard case.

2.3. Kriging

Kriging is a regression-based technique that estimates values at non-sampled locations using weights that reflect the correlation between data at two sampled locations or between a sample location and the location to be estimated (Wackernagel 2003). Environmental sciences have recently started to use geostatistics as a means of interpolating data and of exploring forms of spatial variation. Pollution can be considered to be a regionalized random variable $z(x)$ defined throughout a domain D,such that kriging can be used to explore the way that pollutants vary in space.

In order to make statistical inferences possible from a single occurrence, the sample data z (x_i) must be considered to be a realization of a stationary random function Z(x). That is, the expected value (or mean) of Z(x) must be constant for all points x, *i.e.* E (Z(x)) = m(x) = m, and the covariance function between any two points x and x + h in the domain D depends on the vector h but not on the point x, *i.e.* C (h) = E [(Z(x +h)-m) (Z(x)-m)]= E[Z(x)Z(x+h)]- m².

In practice, it often happens that these assumptions are not satisfied. On both theoretical and practical grounds it is convenient to be able to weaken this hypothesis. In terms of increments of the function Z(x), these two conditions can be expressed in a weaker and more generalized form.

It assumes that the increments {Z(x +h)-Z(x)} are weakly stationary, that is, the mean and variance of the increments exist and are independent of the point x. If E{Z(x)-Z(x +h)} =m(h), that is, the random function has a linear drift m(h), in order for Z(x) to have a weak stationary mean, i) m(h) must be equal zero and ii) the variance of the increments must exit and be independent of the observation point x, that is, Var $\{Z(x) - Z(x+h)\} = 2\gamma(h)\ \forall x, x + h \in D$

Function γ is called the semivariogram, which is said to be isotropic if it depends only on the modulus of h, and

anisotropic if it depends on both the modulus and the direction of h. The anisotropy of the contamination data would require the use of an anisotropic semivariogram model. The sparse ozone-sampled locations, and the inappropriate spatial distributions of monitors, do not permit the evaluation of an anisotropic semivariogram model and consequently in this work we consider an isotropic model, approach. The random variables $Z(x)$ and $Z(x + h)$ relate to the same attribute z, *i.e.* the concentration of a given contaminant, at two different locations x and x +h.

2.4. Simple kriging estimation

If m (h) =0, that is, if the random function has a constant mean and this mean is known we must consider the simple kriging (SK) model. In order to estimate $Z_0 = Z(\mathrm{x}_0)$, (*i.e.*, the contaminant concentration value at x_0), a linear estimator $Z_0^* = \sum_i \lambda_i Z_i$ is used. The weights λ_i are determined using two conditions:

1. The estimator Z_0* must be unbiased,

$$E[Z_0 * - Z_0] = 0$$

2. The variance of the estimation error must be a minimum:

$$\mathrm{Var}[Z_0 * - Z_0] \text{ minimum,}$$

where Z_0 represents the exact but unknown value of Z at x$_0$.

These two conditions lead to the linear system of N equations, known as the simple kriging system (SK):

$$\sum_{j=1}^{N} = \lambda_j C(x_i, x_j) = C(x_i, x_0) i = 1, 2, \ldots, \mathrm{N}$$

(Chilès and Delfiner 1999).

The λ_i are solutions to this system of equations. C (x_i, x $_j$) is the covariance between the data Z (x_i) and Z (x $_j$) and, C (x_i, x_0) the covariance between the data and the target.

$$C(x_i, x_j) = E[(Z(x_i) - m)(Z(x_j) - m)].$$

2.5. Data with a non-stationary mean

The sample data z(x$_i$) that represent the ozone concentration measured at meteorological station i at 15:00 hours is considered to be a realization of a non-stationary random function Z(x). This means that the expectation and the covariance are not translation invariant over the domain D, *i.e.* for a vector h linking any two points (x) and (x +h) in the domain E [Z(x +h)] $\neq$ E [(x)] and Cov [Z(x +h), Z(x)] = C(x, h). In other words, the expected value (mean value m) E [Z(x)] = m(x) depends on the point x in the domain, and covariance is a function of both the point x and the vector h.

In order to study the effect of a non-stationary mean in the performance of interpolation methods IDW and kriging, the data set is transformed by removing the data mean m(x) of the sampled data set.

The evaluation of m(x) is done through a linear regression process of the sample data on the coordinates of the stations to estimate the mean as a function of the position (Rojas-Avellaneda *et al.* 2006). It is assumed that m(x) can be written as a finite expansion:

$$m(x_0) = \sum_{l=0}^{L} a_l f^l(x_0), \qquad (4)$$

where the f ^{l} (x$_0$) are known basis functions and a$_l$ are fixed but unknown coefficients. The first basis function (case l=0) is the constant function equal to 1, which guarantees that the constant-mean case is included in the model. The other functions are considered monomials in the coordinates (x), so that f ^{l} (x$_0$) = x$_0^l$

In this work the first two terms for the basis functions are considered, the case l=0, corresponding to a constant mean, and l=1 a linearly varying mean. The coefficients a$_l$ in Eq. (4), are determined through a linear regression process of the sample data on the coordinates of the sample locations. The residual function R(x), defined as the difference between Z(x) and the estimated mean, is a random function with a constant mean equal to zero. For this function we must consider the simple kriging model.

The weighting factors λ_i' for the estimator

$$R_0* = \sum_i \lambda_i' R_i$$

are solutions to the Simple Kriging system (SK system)

$$\sum_{j=1}^{N} \lambda_j' C(x_i, x_j) = C(x_i, x_0) i = 1, 2, \ldots, N \qquad (5)$$

where C(x_i, x_j) is the covariance between R(x$_i$) and R(x$_j$), R(x$_i$) = Z(x$_i$) – m(x$_i$), is the residual at sample point x$_i$ and x$_0$ is the interpolation point. The covariance is obtained from the variogram model and its parameter sill and range using the basic relation between the variogram and the corresponding covariance C (h):

$$\gamma(h) = C(0) - C(h) \qquad (6)$$

[Chilès and Delfiner 1999; Eq. (2.3)] which is a relation valid for stationary random functions and where C (0) = σ^2 is the variance.

The value of Z$_0$* is then given by:

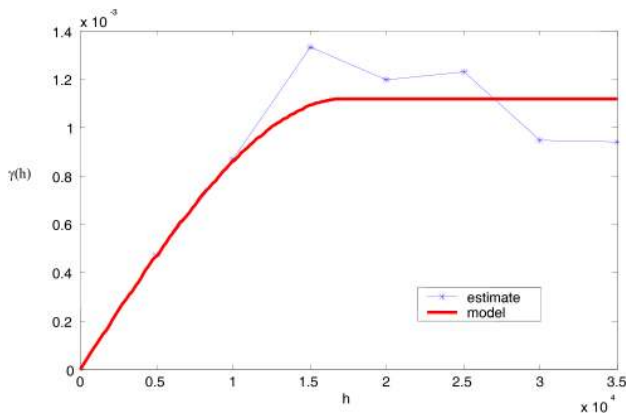$$Z * (x_0) = \sum_i \lambda_i'(Z(x_i) - m(x_i)) + m(x_0). \qquad (7)$$

FIGURE 1. Spherical variogram fitting for the ozone data at 15:00 hours of Dec. 21, 2001.

2.6. Estimation of the variogram model

In practice, the variogram model is estimated from the sample variogram $\gamma*$. We consider in this work the sample variogram proposed by Cressie and Hawkins, [Cressie 1991; Eq. (2.4.12)]

$$2\gamma*(h) \equiv \frac{|N(h)|}{(0.457\,|N(h)| + 0.494)}$$

$$\times \left\{ \frac{1}{|N(h)|} \sum_{N(h)} |Z(x_i) - Z(x_j)|^{\frac{1}{2}} \right\}^4 \quad (8)$$

In this equation, a sample design consisting of n data locations $\{x_j, j=1, \ldots, n\}$ is considered, with $|N(h)|$ distinct data pairs so that $N(h) = \{ (x_i, x_i) : x_i$- x_i=h; i, j =1,..., n$\}$ and where $Z(x_j)$, j=1,...,n are the specific observed values, sampled at the n locations $\{x_j\}$. If the estimated $\gamma*$ is calculated for a number of values of h, the resulting experimental variogram is then used to provide the sill value for the variogram models.

In order to ensure that the variance of any linear combination never becomes negative, only certain functions can be used as models for variograms and covariances. We considered here the most common isotropic variogram models: spherical and exponential. Other variogram models considered here are the rational quadratic, cubic and Gaussian (Cressie 1991, pp 61-63) (Chilès and Delfiner, pp 82-85). All these models are bounded, which means that the variogram reaches an actually or practically limiting value (its sill) at a critical distance r, called the range. As an example Fig. 1 presents the spherical variogram fitting for the ozone data at 15:00 hours on Dec. 21, 2001.

Because the linear variogram model is not bounded by a finite value $\gamma(\infty)$ a covariance function can-not be found for this model and consequently it will not be considered in this work.

For all other bounded models, the covariance function can be obtained from the equation:

$$C(h) = \gamma(\infty) - \gamma(h) \quad (9)$$

[Wackernagel 2003, Eq. (7.26)], where $\gamma(\infty)$ = C (0)=σ^2 is the variance obtained from the sample variogram.

Bounded variograms correspond to stationary variables, and for these variograms the sill is approximately equal to the variance σ^2. The spherical model, for example, is written as

$$\gamma(h) = b + \frac{s}{2} \left[\frac{3\,|h|}{r} - \frac{|h|^3}{r^3} \right] \quad, \quad (10)$$

where b, s and r are the parameters nugget, sill and range respectively to be estimated by the fitting process. While this model reaches its sill at its range, the others reach theirs asymptotically; for these cases, a practical range r is then defined as the distance at which the model value equals 95% of the sill. The estimated sample variogram provides the sill value, s= s_0, and the range's initial value, r_0 for the variogram model. The nugget parameter b is assumed to be equal to zero and b=0 was used throughout this work.

TABLE I.

KRIGING MODEL	PARAMETERS			DRIFTORD	RMSE (1) % TRAINIG	RMSE (2) % TESTING	GENERALI-ZATION G
	SILL	RANGE(Km.)	NUGGET				
EXPONENTIAL	1.50E-03	29.9974	0	0	14.93618	17.17482	0.87
SPHERICAL	1.50E-03	19.4294	0	0	15.83164	17.85536	0.89
RAT. QUAD.	1.50E-03	6.06073	0	0	16.08236	18.07027	0.89
CUBIC	1.50E-03	13.6496	0	0	17.21064	18.78664	0.92
GAUSSIAN	1.50E-03	5.87688	0	0	17.15691	18.89409	0.91
GAUSSIAN	6.00E-04	3.85489	0	1	13.79	14.59591	0.94
EXPONENTIAL	6.00E-04	3.93634	0	1	13.66464	14.68545	0.93
CUBIC	6.00E-04	9.49933	0	1	13.86164	14.79291	0.94
RAT. QUAD.	6.00E-04	3.15153	0	1	13.79	14.79291	0.93
SPHERICAL	6.00E-04	9.9889	0	1	13.64673	15.11527	0.90

Because interpolation values do not depend on the sill parameter, we chose for this parameter, for all models used, the same value $\gamma^*(\infty)$ obtained from the experimental variogram. For this value we found the optimal range r, *i.e.*, the r value that gave the lowest RMSE between the evaluated data and data measured at stations.

3.   Study area and data used

In order to evaluate the performance of each of the two approaches, IDW and kriging, as spatial interpolation techniques, this work considers the data set of measurements of pollutant concentrations in the atmosphere of Mexico City provided by the "Red Automática de Monitoreo Atmosférico de la Ciudad de Mexico, RAMA" (http://www.sma.df.gob.mx/simat/homecontam.php).

The RAMA network provides hourly values for the main air pollutants: carbon monoxide CO, sulfur dioxide SO_2, nitrogen oxides NO_2 and NO_x, ozone O_3 and suspended particulate matter PM10. RAMA uses the ultraviolet photometric method to determine ozone concentrations at monitoring stations (Sistema de monitoreo atmosférico de la Ciudad de México). RAMA also provides hourly values for the magnitude and direction of wind velocity which are measured in 15 stations.

In order to compare the techniques used in this analysis, data monitored at twenty stations for December, 2001, were used since this month showed many days in which the maximum allowed ozone concentration level was exceeded and the measurement data were almost completely recorded. The 31 recorded data at each station that correspond to 15:00 hours provide a sample of 620 data which were used for spatial analysis and interpolation purposes. 15:00 hours is the time at which the highest mean of pollutant concentration is observed, and at which the official Mexican norm for ozone is regularly exceeded.

4.   Estimation process

4.1.   Training process

In the first step, we find the best parameters that define the best weights for the models used in this work for the interpolation. This first step is the training or learning process. In this step we fit the interpolated values using the models IDW or kriging to the sampled data in order to obtain the optimal value for their parameters.

For the kriging models, for fixed values of s and b variogram parameters , the optimal value of the r parameter is obtained, *i.e.*, the r value that gives the lowest mean RMSE between the interpolated values and the data measured at stations for twenty-one days in December, 2001, which we chose as the training set. The estimated sample variogram provides the sill value, s= s_0, and the range's initial value, r_0. The nugget parameter b is assumed to be equal to zero.

Using the variogram model $\gamma(0, s_0, r_0)$ and the first day data at the twenty stations, a cross validation process is applied by removing one ozone monitoring site from this data set, interpolating the remaining sites and evaluating the difference value between the measured concentration at the removed site with the respective interpolated value. The interpolation is done using the SK system equations [Eq. (5)] with the covariance function given in Eq. (9).

The interpolated value $z_1^*(x_0)$ is obtained from Eq. (7), where $z_1(x_i)$- $m_1(x_i)= r_1^*(x_i)$ is the interpolated value for the first day at (x_i) evaluated in the previous step.

The RMSE is calculated for all differences at stations for the first training day and is denoted by S_1:

$$S_1 = \left\{ \frac{1}{20} \sum_{i=1}^{20} \left[z_1^*(x_i) - z_1(x_i) \right]^2 \right\}^{1/2}$$

In the same way, the RMSE is evaluated for all training days. The mean RMSE, S for the training process is evaluated through

$$S = \frac{1}{21} \sum_{j=1}^{21} S_j.$$

The r_0 parameter's optimized value, $r_{opt,}$ is obtained through a minimization process of the S value. The minimum S value is denoted by S_m.

The RMSE values expressed in percentage listed in Table I, are calculated taking as reference value the maximum allowed ozone concentration according to the Mexican norm (0.11 ppm). The following expression was used:

$$\text{RMSE (1) } \% = (S_m/0.11) \times 100.$$

The optimal value r_{opt} and the corresponding RMSE percentage, RMSE (1) %, are listed in Table I for the different kriging models considered here.

4.2.   Testing process

For the variogram model $\gamma(0, s_0, r_{opt})$ and testing data at the twenty stations for a testing day j, a cross validation process is applied by removing one ozone monitoring site from this data set, interpolating the remaining sites and evaluating the difference value between measured concentration at the removed site with the respective interpolated value. The RMSE is calculated for all differences at stations for the chosen testing day j and is denoted by S_j:

$$S_j = \left\{ \frac{1}{20} \sum_{i=1}^{20} \left[z_j^*(x_i) - z_j(x_i) \right]^2 \right\}^{1/2}.$$

The RMSE is evaluated in the same way for all testing days.

The RMSE, S for the testing process is evaluated through

$$S = \frac{1}{10} \sum_{j=1}^{10} S_j.$$

The corresponding RMSE percentages, RMSE (2) %, are calculated taking as reference value the maximum allowed ozone concentration according to the Mexican norm using the expression

$$\text{RMSE (2) \%} = (S/0.11) \times 100. \qquad (11)$$

These percentages are listed in Table I for the different kriging models considered here.

5. Results and discussion

The results obtained for ozone interpolation using IDW and kriging methods are presented graphically and numerically in Fig. 2 and in Tables I and II.

In the first column, these tables list the name of the model considered. The second column lists the parameters that define the model. The next column indicates the drift order, zero or one, which indicates that a constant mean or a linearly varying mean, respectively, is removed from the data set. The RMSE (1) % column corresponds to the minimum RMSE, expressed in percentage, for training data. The RMSE (2)% column are given the RMSE values for the testing data.

In the last column, the generalization parameter G is given. It represents the generalization capacity of the method, a concept that is used in artificial neural networks. We applied this concept to all interpolation methods used in this work, and the generalization G is defined as the relationship between the minimum RMSE value for the training data, S_m, and RMSE value for the corresponding testing data, RMSE (2).

The parameters that define the weights for the IDW method are the exponent α, which determines the degree of smoothing, and the radius r, which defines the neighborhood of interest.

In Table I, the kriging results are shown for the various models considered. For this method the parameters that define the variogram are the sill, nugget and range.

As we can see from Tables I and II for all models of the two methods used, IDW, and kriging (except for the uninteresting cases IDW with $\alpha= 4, 6, 8$), the results obtained under the assumption of a linear drift show a lower RMSE for both training and testing data than results obtained under the assumption of constant drift. This fact shows us the remarkable effect on the accuracy of the interpolation process of the trend removal. Furthermore, the IDW models with drift=1 improve all kriging models with drift =0, except for the uninteresting cases mentioned above.

From the data in Table I, we can see that for all variogram models a constant drift produces results with a relatively low generalization value, which means that the RMSE in the testing procedure is relatively higher than the corresponding RMSE for training data. The generalization constants G have a higher value when linear drifts are considered, corresponding to improved generalization processes.

TABLE II. Results for standard and optimized IDW methods.

IDW MODEL	PARAMETERS		DRIFTORD	RMSE (1) % TRAINING	RMSE (2) % TESTING	GENERALI-ZATION G
	ALPHA	RADIO(Km.)				
STANDARD 1	1	INF	0	18.42845	17.96282	1.03
OPTIMIZED 1	2.1878	23.3294	0	15.92118	16.60173	0.96
STANDARD 2	2	INF	0	16.51218	16.87036	0.98
OPTIMIZED 2	2.83436	INF	0	16.10027	17.01364	0.95
OPTIMIZED 3	2.83436	23.3408	0	15.93909	16.72709	0.95
STANDARD 4	4	INF	0	16.40473	17.65836	0.93
STANDARD 6	6	INF	0	17.10318	18.44636	0.93
STANDARD 8	8	INF	0	17.60464	18.92991	0.93
STANDARD 1	1	INF	1	15.04364	14.84664	1.01
OPTIMIZED 1	1.2902	8.7496	1	14.68545	16.36891	0.90
STANDARD 2	2	INF	1	14.52427	15.366	0.95
OPTIMIZED 2	2.0707	INF	1	14.52427	15.41973	0.81
OPTIMIZED 3	2.0707	8.7496	1	14.75709	16.60173	0.89
STANDARD 4	4	INF	1	15.49136	17.10318	0.91
STANDARD 6	6	INF	1	16.42264	17.96282	0.91
STANDARD 8	8	INF	1	16.95991	18.41055	0.92
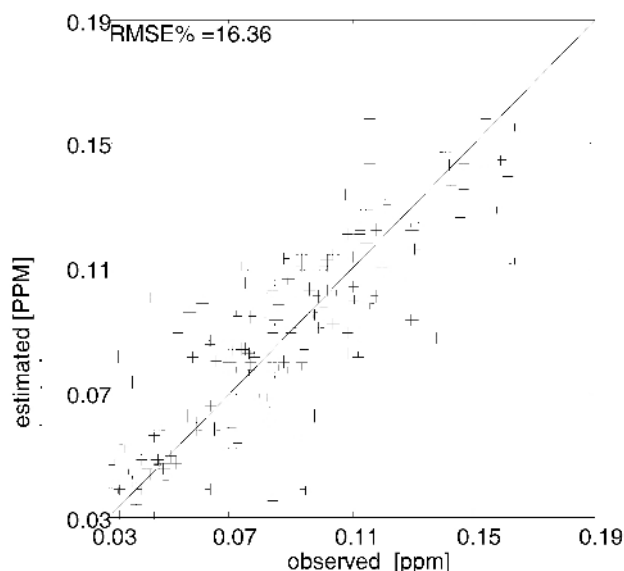
(a)
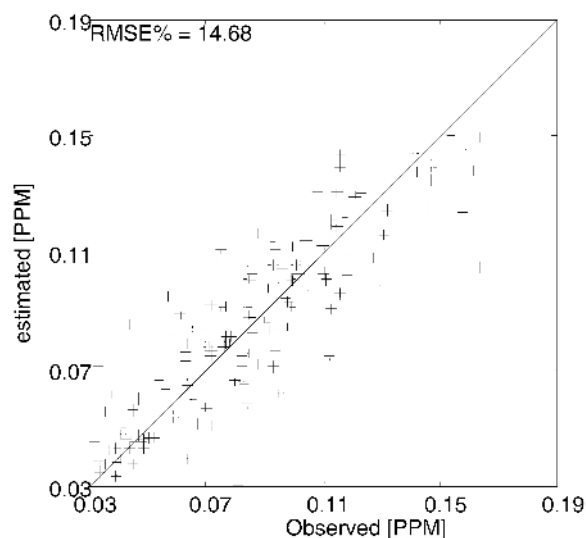
(b)

FIGURE 2. Crossvalidation scattergrams for the two methods: (a) IDW with α=1.2902 and r = 8.7496 km; (b) SK-exponential with nugget=0, sill=6e-4, range=3.9363 km.

In Figs. 2a and 2b, scatter plots are presented for each of the methods as a help for the analysis of the results. These plots confirm the numerical results obtained for RMSE in Tables I and II.

Figures 3a and 3b show the maps for ozone concentration interpolated levels at 15:00 hours on December 21, 2001, for particular parameter values of IDW and kriging models, respectively. The small diamond-shaped figures indicate the station's positions, and their color indicates the measured ozone value; at these points, interpolated and measured values are deliberately the same when kriging or IDW meth-

ods are used for the interpolation, as we can see in Figs. 3a and 3b.

The interpolated wind field is also shown on these maps. The magnitude and direction of wind velocity is measured at 15 monitoring stations in the Mexico City region and its values are provided hourly. Wind field was obtained from values measured at stations at 14:00 hours on Dec. 21, 2001. The highest ozone pollution concentrations are found in the northwest region of Mexico City, as is the wind field convergence direction. Along their trajectories through the city, determined by the wind field, primary pollutants, emitted mainly on the east side of the city, are blown by the wind towards the northwest side, where they arrive as ozone.

6. Conclusions

If one is faced with data that appear to be samples from realizations of non-stationary random functions, one cannot ignore the apparently non-stationary ones when interpolation
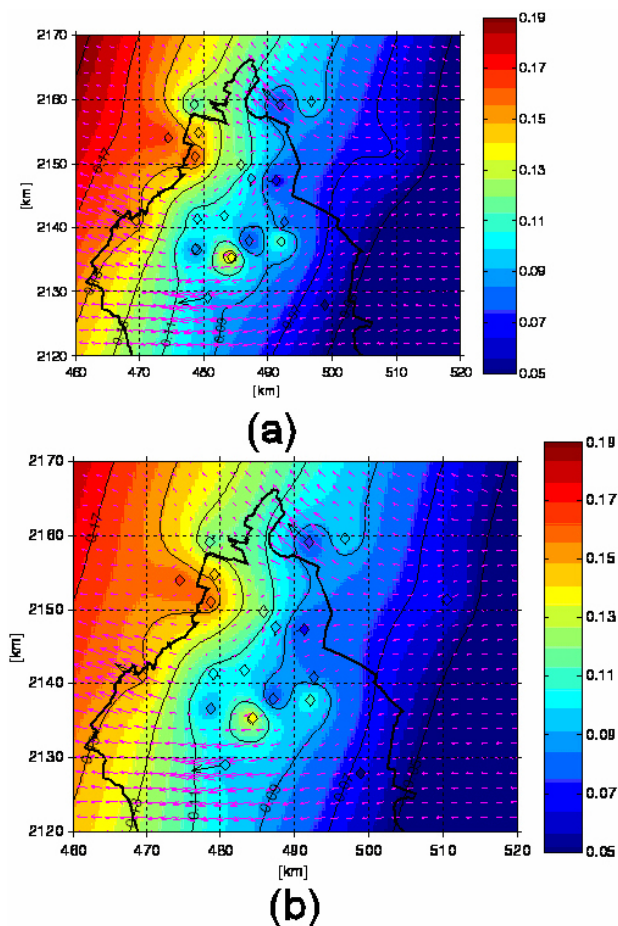

(a)

(b)

FIGURE 3. Ozone concentration maps for December 21: (a) IDW with α = 1.2902, r=Inf; (b) SK-exponential with nugget = 0, sill = 6e-4 and range = 3.9363km In both cases a linear drift was computed through the least square method. The arrows indicate the wind velocity field with their length corresponding to wind displacement in a half hour.

processes are to be considered, because one can obtain unsatisfactory results.

In this case, removal of a trend produces a data set for which interpolation processes obtain noticeably more accurate results, as indicated by a smaller RMSE between measured and interpolated data. The generalization capacity of interpolation methods is also improved when a drift is removed from the data. The parameter G value has a higher value when a linear drift is removed from the data than the G value obtained when a constant drift is removed. A higher G value corresponds to an improved generalization process.

Although linear interpolation processes such as IDW and kriging normally require relatively high sampling densities and uniformly-spaced sample locations, the relative accuracy obtained with the application of these methods for estimating ambient ozone concentrations in Mexico City region was stimulating.

Although kriging and IDW were tested to interpolate ozone pollution values, preliminary calculations indicated that these methods can also be used for other pollutants.

Acknowledgments

1. J.P. Chilès and P. Delfine, *Geostatistics, Modeling Spatial Uncertainty* (John Wiley & Sons., New York, 1999).

2. N. Cressie, *Statistics for Spatial Data* (John Wiley & Sons., New York, 1991).

3. De Leeuw FAAM, *Van Zantvoort EG Environmental Pollution* **96** (1997) 89.

4. J.A. Mulholland, A.J. Butler, J.G. Wilkinson, and A.G. Rusell, *Air and Waste Manage. Assoc.* **48** (1998) 418.

5. D.L. Phillips, D.T. Tingey, E.H. Lee, A.A. Herstrom, and W.E. Hogsett, *Environmetrics* **8** (1997) 43.

6. D. Rojas-Avellaneda and J.L. Silván-Cárdenas, *Stochastic Environmental Research and Risk Assessment (SERRA)* **6** (2006) 455.

7. Secretaria del Medio Ambiente, Gobierno del Distrito Federal. México. March 2006 Informe Ejecutivo de la Calidad del Aire ZMVM

8. Sistema de Monitoreo Atmosférico de la Ciudad de México, http://www.sma.df.gob.mx/simat/mediciones/

9. H. Wackernagel, Multivariate geostatistics. Third edition (Springer-Verlag. Berlin, Heidelberg, New York, 2003).