Imputation of spatial air quality data using gis-spline and the index of agreement in sparse urban monitoring networks

Imputación de datos espaciales de calidad del aire usando sig–spline e índice de ajuste en redes urbanas de monitoreo

Libardo Antonio Londoño–Ciro*, **Julio Eduardo Cañón–Barriga**

Grupo de investigación en Gestión y Modelación Ambiental (GAIA), Facultad de Ingeniería, Universidad de Antioquia. Calle 67 # 53-108. A. A. 1226 Medellín, Colombia.

ABSTRACT: This paper presents a procedure to address the lack of spatial air quality data in urban areas, based on the use of Geographic Information Systems (GIS) and spatial interpolation techniques as an alternative to conventional methods of statistical imputation. Two spatial interpolation algorithms are compared: IDW and spline. The procedure considers the spatial interpolation process, the cross validation with the index of agreement (IOA), and the analysis of the effect of sampling density and the coefficient of variation (CVO_i), using different error statistics. The interpolation maps are complemented with gradient and directional gradient maps that may serve as complementary aides in the definition of critical sampling points. The procedure is applied to data imputation of three pollutants NO_2, PM10 (particulate matter of diameter 10 microns) and TSP (total suspended solids) from observed data samples in the city of Medellín (Colombia).

RESUMEN: Este trabajo presenta un procedimiento para abordar la falta de datos espaciales de calidad del aire en zonas urbanas, con base en el uso de Sistemas de Información Geográfica (SIG) y las técnicas de interpolación espacial como una alternativa a los métodos convencionales de imputación estadística. Se comparan dos algoritmos de interpolación espacial: IDW y *spline*. El procedimiento considera el proceso de interpolación espacial, la validación cruzada con el índice de (IOA), y el análisis de la densidad de muestreo y del coeficiente de variación utilizando diferentes estadísticos de error. Los mapas de interpolación se complementan con los mapas de gradiente y de gradiente direccional que pueden servir como complementos en la definición de puntos de muestreo críticos. El procedimiento se aplica a la imputación de datos de tres contaminantes: NO_2, PM10 (partículas de 10 micras de diámetro) y SST (sólidos suspendidos totales) a partir de muestras de datos observados en la ciudad de Medellín (Colombia).

1. Introduction

Models that include scattering algorithms or Geographic Information Systems (GIS) have been commonly used to simulate regional air quality, estimating the concentration of pollutants from fixed and / or mobile sources using data obtained from sampling networks. As this phenomenon is continuous in space and time, it must have a minimum number of sampling points [1-3], and extended periods of observation [4]. Additionally the geometry and location of the points in the network should take into account the effect of variables that can affect the measurement such

as: proximity to the emission source and location of the sampling point in the studied area [5].

The uncertainty in the modeling of this phenomenon increases when data are missing or are not representative [6, 7]. The problem of lacking data has been approached by statistical data imputation techniques, numerical simulation, spatial interpolation, lineal spatial regression, multivariate linear regressions, locally linear reconstruction, spartan random processes, spatial and temporally weighted regression, time series analysis, spatial statistics, econometrics and neural networks [8-15]. One alternative to address this issue is using GIS to implement spatial interpolation algorithms.

Spatial interpolation algorithms are mathematical tools for estimating the unknown values of a variable at different points $Z_0(x_0, y_0)$, based on known values measured at specific locations $Z_i(x_i, y_i)$, within a spatial domain $S(x, y)$

that is defined by a geographic projection system. According to the spatial autocorrelation principle [16], Z values will share similar properties in virtue of their proximity. A way to estimate $Z_0(x_0, y_0)$ is from a linear combination of $Z_i(x_i, y_i)$ (Eq. 1), where λ_i are weighted factors [17]:

$$Z_0(x_0, y_0) = \sum_{i=1}^{n} \lambda_i Z_i(x_i, y_i)$$
(1)

Spatial interpolation algorithms are classified according to three criteria: the way the weighting factors are calculated; the amount of data used for the estimation (global if using all the data and local if using part of the data in a particular neighborhood); and the error of the method (accurate or inaccurate depending on the nature of the phenomenon under study, the quality and quantity of the observed data, the existence of spatial autocorrelation, the sampling method and the spatial distribution of the data observed) [16, 18]. Another way to estimate $Z_0(x_0, y_0)$ is from the implementation of methods such as trend surface analysis, regression models, triangulation, and splines.

This paper analyzes the implementation of two algorithms, IDW and spline, for data imputation of three air quality variables (NO_2, PM10 and TSP) to generate interpolation maps based on a sparse sampling network for the colombian city of Medellín. The algorithms are evaluated with cross validation tests using the index of agreement (IOA), as well as several error statistics as a function of the sampling density (sampling area divided by number of stations) and the coefficient of variation of the observed data (CVO_i). The interpolation maps include directional gradients as a way to represent the spatial patterns of the pollutants.

2. Area of study

The city of Medellín is located in a small valley in the Andean cordillera, between 6.0° - 6.5° N and 75.5° - 75.7° W. The city is 60 km long, with a width that varies between 10 and 20 km with an area of 1157 km². Its average altitude is 1500 m, with hill slopes between 0 and 50% with two inner hills in the center, the result of sedimentation processes which involves a complex topography. Data of the monthly average concentration (in $\mu gm/m^3$) were gathered for the pollutants NO_2, PM10 (particulate matter of diameter 10 microns) and TSP (total suspended solids) in 6, 10 and 15 stations respectively (see Figure 1). Table 1 summarizes the coordinates and the names of the stations (Geographic Coordinate System: GCS MAGNA. Projected Coordinate System: MAGNA Colombia Bogotá, Projection Transverse Mercator). The stations are part of the city of Medellín's air quality network called *RedAire*.
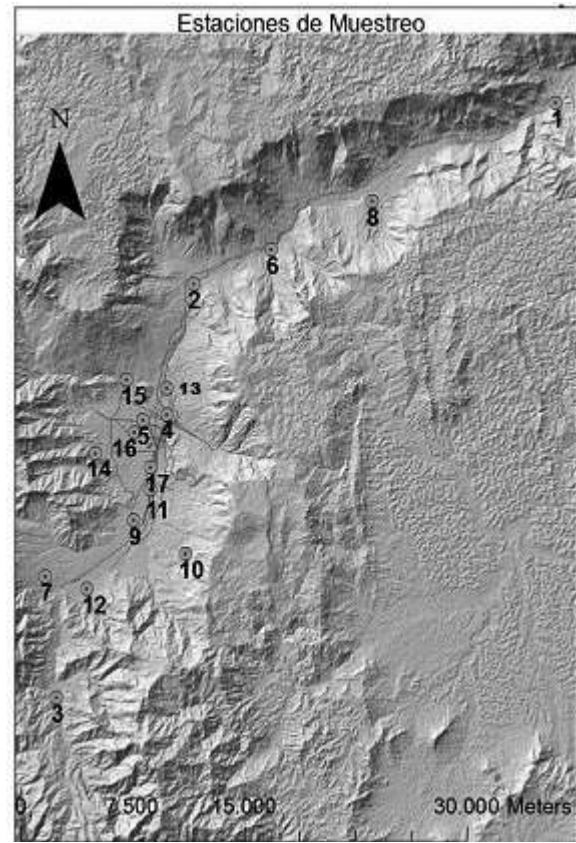

Figure 1 Location of sampling stations for NO_2, PM10 and TSP variables

3. Methodology: imputation of missing spatial data with spatial interpolation algorithms

A spatial data $Zi(xi, yi)$ is defined in a spatial domain $S(x, y)$ with x and y in a geographic projection system. It also contains information about the variable under study (value), an associated geometry (i.e., point, line or polygon), a geographical coordinate system, a structure of data storage (raster or vector) and a color code (RGB) to represent the variation in space. An important property of spatial data is the spatial autocorrelation. Spatial autocorrelation allows disclosing a data value at a site, being able to estimate its value at neighboring positions, which leads to spatial interpolation algorithms.

The algorithms were implemented in ArcMap®, using observed data from the sampling stations to obtain maps of interpolated values for each variable. The maps are validated with cross-validation tests [9] and the IOA. The aim of a cross-validation test is to determine the goodness of fit between observed data and interpolated values in a particular place or control point for spatially distributed data [19-27]. The difference between the observed and the

Table 1 Sampling stations description

ID	X	Y	Station Code	Station Name	NO$_2$	PM10	TSP
1	-75.335767	6.439300	BARBOSA	Municipio de Barbosa			X
2	-75.555467	6.332483	BELLO	Municipio de Bello	X	X	X
3	-75.637967	6.089600	CALD	Municipio de Caldas		X	X
4	-75.571750	6.256017	AGUI	Edificio Miguel de Aguinaga	X	X	X
5	-75.586033	6.252467	CORA	Corporación Autónoma Regional de Antioquia		X	
6	-75.508217	6.352750	COPA	Municipio de Copacabana			X
7	-75.644990	6.160860	ESTRELLA	Municipio de la Estrella			X
8	-75.447300	6.381310	GIRARDOTA	Municipio de Girardota		X	X
9	-75.591250	6.193767	ITSFDO	Planta de Tratamiento de Agua San Fernando		X	X
10	-75.560317	6.174467	ITAGÜÍ	Municipio de Itagüí Colegio el Rosario	X	X	X
11	-75.581133	6.211833	POLITECN	Politécnico Colombiano Jaime Isaza Cadavid	X	X	X
12	-75.619550	6.153500	SABANETA	Municipio de Sabaneta			X
13	-75.571500	6.271250	U DE A	Universidad de Antioquia			X
14	-75.615000	6.233183	U DE M	Universidad de Medellín			X
15	-75.596067	6.276400	UNAL	Universidad Nacional de Colombia	X	X	X
16	-75.591300	6.245317	U P B	Universidad Pontifica Bolivariana	X		X
17	-75.581530	6.224950	SANANT	Parque de San Antonio		X	

interpolated value in the control point is called estimation error or residual at this point. The goodness of fit between the observed data and interpolated values is calculated with IOA using Eq. (2) [10, 11].

where n is the amount of control points, E_i is the value interpolated at the control point i, O_i is the observed value at the control point i, and μ_o is the average of the observed data. An IOA close to one indicates a good fit of the interpolated values to the observed ones [10, 11].

$$IOA = 1 - \frac{\sum_{i=1}^{n}(Ei-Oi)^2}{\sum_{i=1}^{N}(|Ei-\mu_o|+|Oi-\mu_o|)^2} \qquad (2)$$

Except for the stations located in the extreme north (1, 2 and 8) and south (3, 7 and 10) of the study area, which cannot

Table 2 Calculation of the IOA for spline and IDW interpolation

Station Code	NO$_2$			PM10			TSP		
	O_i	$E_{iSpline}$	E_{iIDW}	O_i	$E_{iSpline}$	E_{iIDW}	O_i	$E_{iSpline}$	E_{iIDW}
BELLO	X	X	X	70.0	71.6	80.2	86.1	107.6	129.2
AGUI	49.0	46.9	43.0	76.9	74.9	82.6	162.3	123.2	122.4
CORA	X	X	X	70.1	82.1	84.4	X	X	X
COPA	X	X	X	X	X	X	83.4	92.4	95.4
GIRARDOTA	X	X	X	X	X	X	74.0	93.3	96.3
ITSFDO	X	X	X	69.5	69.0	75.0	113.2	108.8	111.7
ITAGÜÍ	X	X	X	72.6	69.3	68.3	112.2	101.4	114.4
POLITECN.	56.0	45.3	42.0	63.3	82.7	78.2	112.5	111.4	112.8
SABANETA	X	X	X	X	X	X	95.3	99.7	90.2
U DE A	X	X	X	X	X	X	141.9	141.3	139.8
U DE M	X	X	X	X	X	X	88.7	101.9	113.2
UNAL	58.0	43.4	38.9	88.2	71.1	73.5	138.2	109.9	124.0
UPB	40.0	52.2	54.9	X	X	X	86.7	126.7	129.5
SANANT	X	X	X	91.7	67.3	69.3	X	X	X
IOA		0.05	0.06		0.22	0.10		0.71	0.66

be removed in order to generate a interpolation map, all the other stations were removed one by one to calculate a interpolation map and to determine the interpolated value of the variable in the site of the removed station ("leave one out" technique [9]). For NO_2, n-2 control points were taken (except stations 2 and 10). For PM10, n-2 control points (except stations 3 and 8). For TSP, n-3 control points (except stations 1, 3 and 7). Results of IOA calculations are shown in Table 2 and Figure 2.
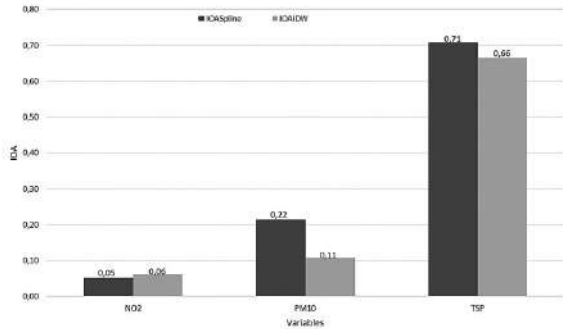

Figure 2 IOA for Spline and IDW interpolation for each variable

The Absolute Relative Error (ARE) and the Mean Absolute Relative Error (MARE) were calculated using Eq. (3) and Eq. (4) respectly, to analyze the performance of the algorithms in interpolating the values for each variable at each station [28, 29]:

$$ARE = \frac{|O_i - E_i|}{O_i} \tag{3}$$

$$MARE = \frac{\sum_{i=1}^{n} \frac{|O_i - E_i|}{O_i}}{n} \tag{4}$$

where n is the amount of control points, E_i and O_i are the interpolated and observed values respectively at control point i. The results are shown in Table 3 and Figures 3 to 5.
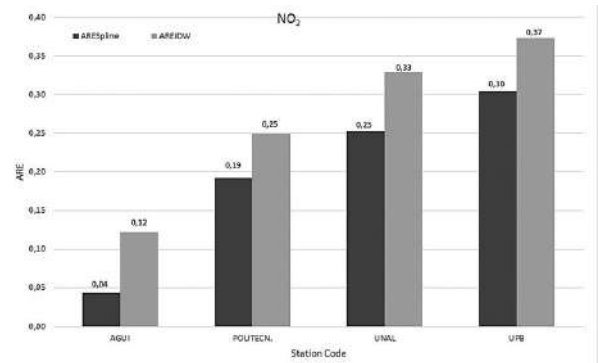

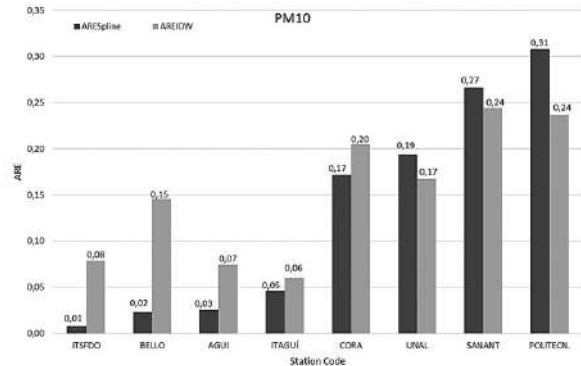**Figure 3** Calculation of ARE for NO_2 at each station

Figure 4 Calculation of ARE for PM10 at each station

Table 3 Calculation of the ARE and MARE

| Station Code	NO₂		PM10		TSP	
	ARE$_{Spline}$	ARE$_{IDW}$	ARE$_{Spline}$	ARE$_{IDW}$	ARE$_{Spline}$	ARE$_{IDW}$
BELLO	X	X	0.02	0.15	0.25	0.50
AGUI	0.04	0.12	0.03	0.07	0.24	0.25
CORA	X	X	0.17	0.20	X	X
COPA	X	X	X	X	0.11	0.14
GIRARDOTA	X	X	X	X	X	X
ITSFDO	X	X	0.01	0.08	0.04	0.01
ITAGUÍ	X	X	0.05	0.06	0.10	0.02
POLITECN.	0.19	0.25	0.31	0.24	0.01	0.00
SABANETA	X	X	X	X	0.05	0.05
U DE A	X	X	X	X	0.00	0.01
U DE M	X	X	X	X	0.15	0.28
UNAL	0.25	0.33	0.19	0.17	0.20	0.10
UPB	0.30	0.37	X	X	0.46	0.49
SANANT	X	X	0,27	0,24	X	X
MARE	0.20	0.27	0.13	0.15	0.16	0.18
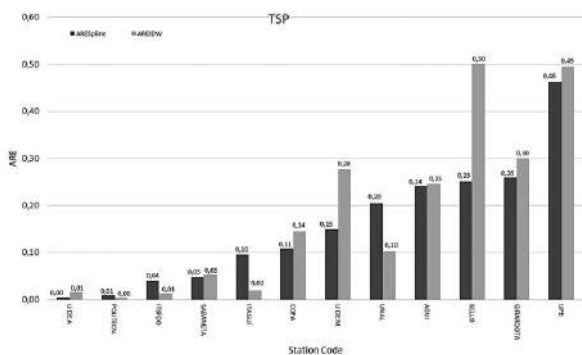
Figure 5 Calculation of ARE for TSP at each station

The sampling density and the coefficient of variation of the observed data (CVO_i) are calculated for each variable and analyzed in terms of the IOA values calculated with IDW and spline [28, 29]. Results are shown in Figures 6 and 7.
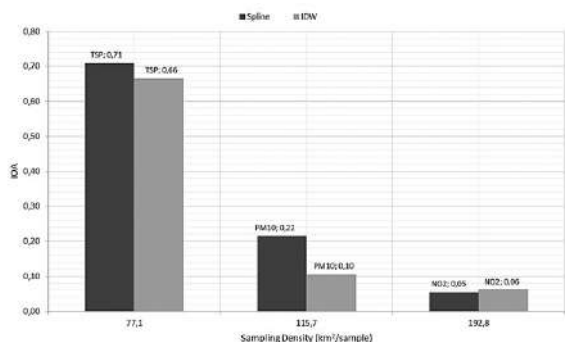

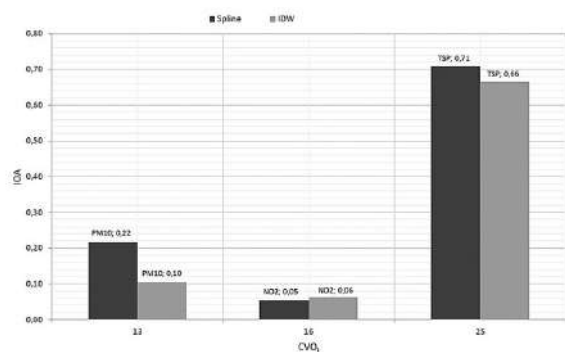**Figure 6** Effect of sampling density on the algorithms

Figure 7 Effect of CVO_i on the algorithms

The relationship between sampling density and IOA is analyzed for TSP at 5 (TSP5), 10 (TSP10) and 15 stations (TSP15), since it is the only variable that is measured in all the 15 stations. Results are shown in Figure 8.
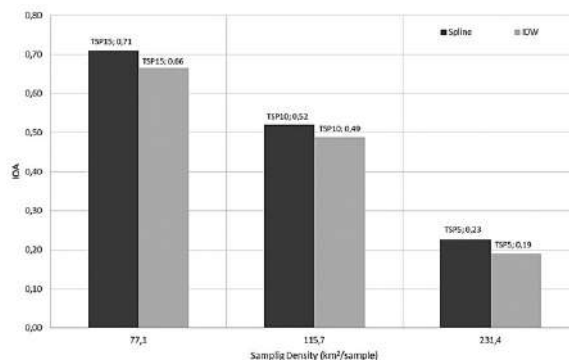

Figure 8 Effect of sampling density on the TSP´s IOA

4. Discussion of results

The use of statistical measurements of error to compare between interpolated and observed data allows determining the reliability of the data interpolated with a particular algorithm and therefore the algorithm's performance in the imputation of data from environmental variables. In this case, the following error statistics were calculated (Table 4): the IOA, the Absolute Relative Error (ARE), the average standard error (ASE), the Root Mean Square Error (RMS) and the Root Mean Square Error Standardized (RMSS) [28]. In the table, IOA and RMSS values close to one and small values of MARE, ASE and RMS represent a good fit. This suggests that the spline algorithm was the most appropriate for data imputation for the variables analyzed.

Table 4 Measures of error for IDW and spline applied to air quality variables

		IOA	RMSS	MARE	ASE	RMS
NO₂	Spline	0.05	2.04	0.20	3.79	10.98
	IDW	0.06	2.90	0.27	7.02	14.29
PM10	Spline	0.22	2.05	0.13	5.95	12.91
	IDW	0.10	2.12	0.15	5.95	13.37
TSP	Spline	0.71	1.45	0.16	14.37	20.77
	IDW	0.66	1.57	0.18	15.17	23.75

The ARE was used to analyze the reliability of the interpolation for each variable in each station according to the implemented algorithm, as shown in Figures 3 to 5. The results not only are different among variables but also for the same variable in different locations. For example, the Politecnico station (critical point station located within 10m of a main road) shows good results in TSP, whereas the UPB station (background station located more than 10 m apart from main roads) does not show good results in TSP. These differences highlight the importance of considering factors such as the type of emission and the time of exposure to the source in the location of the station.

Figures 6 to 8 show the effect of sampling density and the CVO_i on the statistic error. The efficiency of the interpolation depends on factors of sampling density such as the area

of the region under study and the number of sampling data points (i.e., when the amount of data increases within an area, the algorithms tend to be equally efficient) [28, 29]. However, the minimum amount of data required for the results to be acceptably reliable must be determined according to the phenomenon under consideration. For this study, results with 15 data points for TSP significantly improve the interpolation results, compared to the NO_2 (six stations) and the PM10 (10 stations). Figure 7 indicates a slightly major sensitivity of IDW to outliers in the calculation of CVO_i [28].

Figure 8 shows the variation of the IOA for TSP in terms of the sampling density for 5 (TSP5), 10 (TSP10) and 15 (TSP15) data points respectively. While there is an improvement of the IOA by increasing sampling points within the same area, it is also important to determine the coverage, since a low sampling density can be useful at micro-scale measurements but not at meso- or macro-scale levels. Furthermore, the location of the sampling site relative to the emission source must be taken into account, since it would be possible to have a good amount of sampling sites but no significant measurements of the variable. This is an important aspect to consider in the proper definition of locations for monitoring purposes.

In general, the spline algorithm performs better in the interpolation process and therefore in the imputation of missing data. Contrary to methods such as Kriging, for instance, spline does not depend on the statistical distribution of the data and may be used to extrapolate values due to the continuous nature of the polynomial order 3 used to interpolate (this could be beneficial for data imputation on the edges of the study area).

Figure 9 shows the maps of the interpolated values obtained with a spline algorithm for the concentration of NO_2, PM10 and TSP pollutants, in this case for the month of August 2007, which is used as a study case throughout the paper (maps for other months can be viewed in the URL tesislibardolondono.aula.com.co, and files may be accessed upon request to the authors by e-mail). With the maps, it is possible to analyze the spatial patterns of the data and therefore the spatial distribution of air pollutants. Maps allow visualizing areas of greater or lesser concentration (units are given in $\mu gm/m^3$).

As expected, the higher concentrations occur around points that record higher values. In the case of NO_2 and PM10 there is a tendency towards higher concentrations in the center of the study area and in the North-South direction, whereas for TSP the concentration is nearly radial and also high in the center of the study area, which coincides with high vehicular traffic.

Using the spatial pattern maps (Figures 9), it is possible to create a map of spatial gradient that shows areas where the variables have a greater or lesser potential to change with distance. In general, the maps indicate preference in the concentration of pollutants near the city.

Figures 10 represents the potential change of concentration in the area, and can be thought of as a potential for dispersion (maps units are given in percentage change of $\mu gm/m^3$ with distance in m). In the case of NO_2, the highest gradient was 15% on the axis formed by the stations 15, 16 and 11 (see Table 1). In the case of PM10, the highest gradient was 25% on the axis formed by the stations 17 and 11, very close to a main intersection in the study area. In the case of TSP, the largest gradient was 37% between stations 16 and 14, located at the city center towards the west.

Figure 11 exemplifies the maps of the spatial gradient direction. These maps do not show any preferential direction for the change of the concentration gradient, implying that the gradient of concentration (related with dispersion) has the potential to change in any direction. Points located in
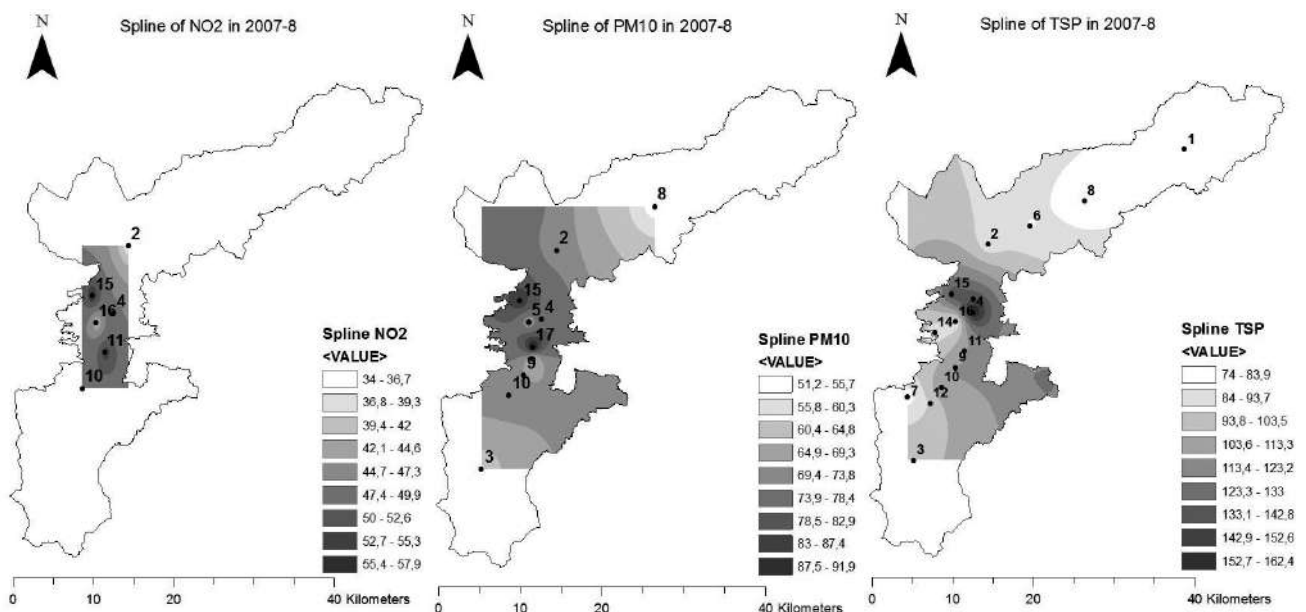

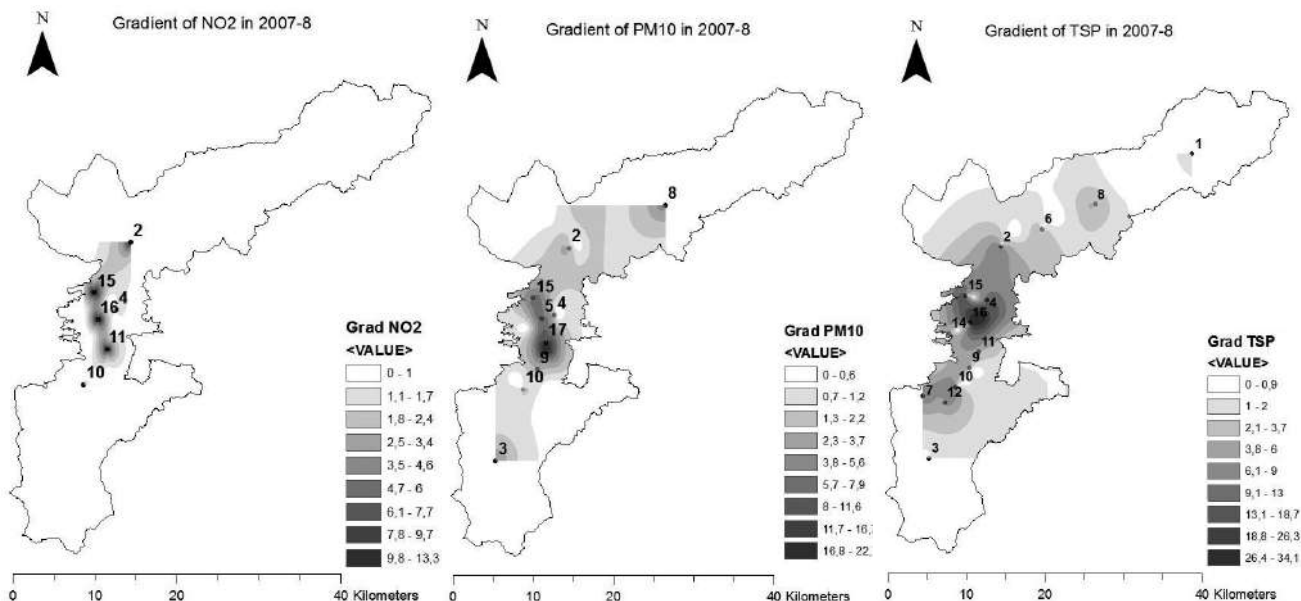**Figure 9** Spatial interpolation for NO_2, PM10 and TSP

Figure 10 Spatial gradient for NO$_2$, PM10 and TSP

the north show a trend to disperse in the North, Northeast and East directions, while at the south points tend to go South and West. The results of directional gradient maps may serve as a complement to analyze the effect of wind patterns in the area.

The combination of spline interpolation, gradient and directional gradient maps allowed a complementary view of the spatial pattern of the concentration, dispersion and direction of dispersion of the variables under study. In addition, this combination of maps may serve to determine
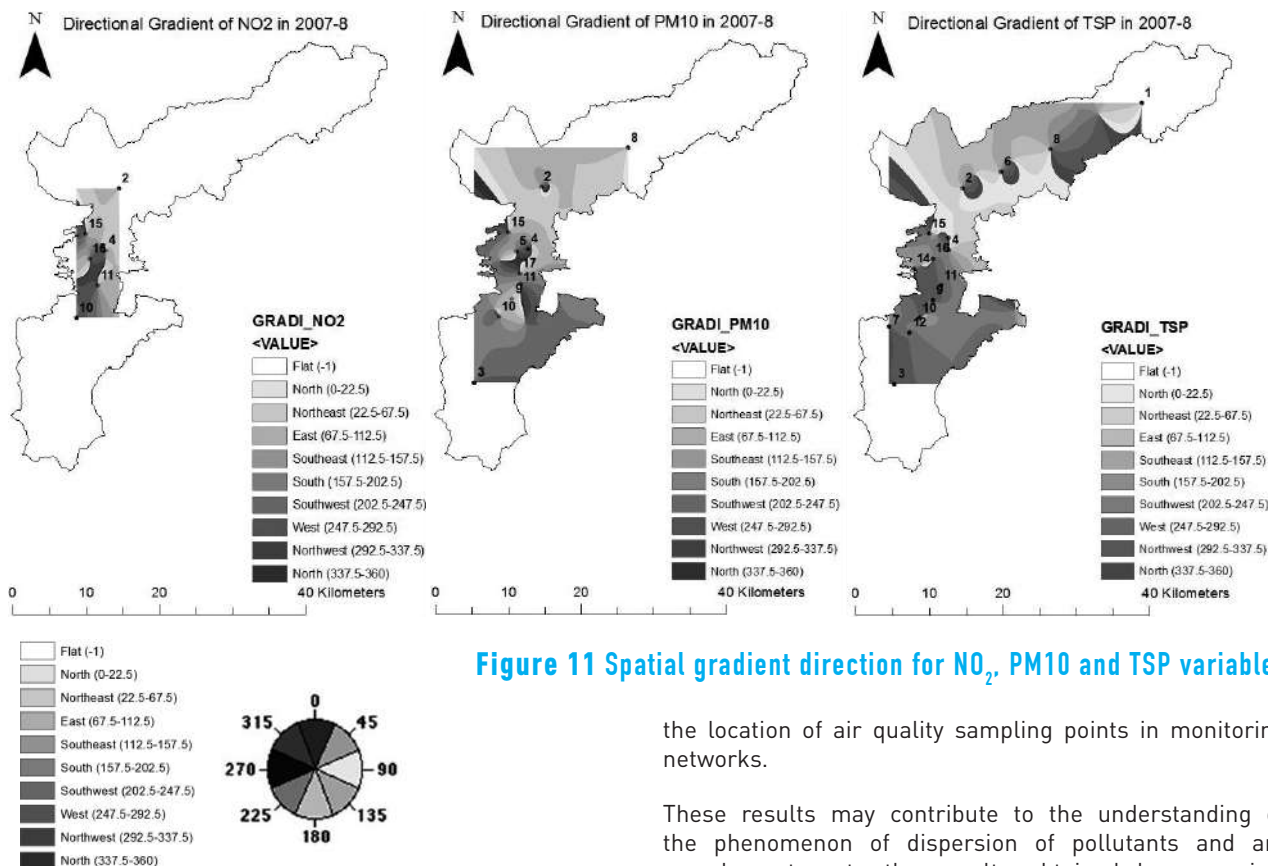

Figure 11 Spatial gradient direction for NO$_2$, PM10 and TSP variables

the location of air quality sampling points in monitoring networks.

These results may contribute to the understanding of the phenomenon of dispersion of pollutants and are complementary to the results obtained by numerical

simulation such as RAMS and WRF models used in other ongoing research in the region.

A different use of the IOA is proposed here to compare directly the observations between months and to determine how similar the pollutant distributions are in time (Table 5). For instance, the comparison of NO_2, PM10 and TSP values between 2006 and 2007, in the months of August and November, show an IOA greater than or equal to 0.80, indicating a good resemblance of values for those months in different years. For NO_2 no similarities were obtained between the observed values in different months. In the case of PM10, a good similarity occurred between the months of August and October of 2006 and 2007. For TSP the similarities were high in the month of November. Not all of the months show the same level of similarities, however, which indicate that the patterns of observed pollution vary both during the year and among years.

Table 5 IOA for different periods

Variable	Ag 06 Vs Ag 07	Sp 06 Vs Sp 07	Oc 06 Vs Oc 07	Nv 06 Vs Nv 07
NO₂	0.36	0.05	0.35	0.10
PM10	0.88	0.64	0.83	0.56
TSP	0.76	0.70	0.44	0.80

5. Conclusive remarks

The IDW and spline algorithms offer an efficient and approximate procedure to address the problem of absence of data in continuous phenomena, whose results can be used as a complement of conventional statistical techniques of data imputation and in support of physical models of environmental air pollution, provided that a good amount of data is available. In our case study, for instance, a 71% reliability was obtained with data available from 15 stations for TSP, according to the calculation of the IOA in cross-validation tests. In general, the spline algorithm showed a better performance than IDW to interpolate the air quality variables within the domain and with the limitations of data points in space. It should be noted that both algorithms are affected by the sampling density. The spline algorithm performed well in the stations located on the edge of the area interpolation which is critical for the imputation of data at the borders of the area considered.

The use of spatial interpolation algorithms for data imputation provides useful maps to understand how the air pollutants may be distributed in space. The combination of maps (showing the spatial patterns of concentration, gradient/dispersion and directional gradient) of NO_2, PM10 and TSP for different months of the year, reveal certain patterns that are important for the study of the spatial variation of air pollution. The interpolation map suggests the spatial variation in the concentration of the pollutant around stations and relative to the emission sources (i.e., roads), and may serve as a basis to determine critical sampling points through spatial proximity models. The map of gradients gives an idea of potential zones were the pollutants would "disperse," helping in the definition of sampling points for long term effects of the pollutants in background stations. Both maps would serve as an aid in the preliminary definition of critical monitoring sites to expand the air quality network. These results may be used in combination with other mapped information relative to variables affecting the dispersion phenomenon, such as monthly wind patterns and topography.

The IOA was employed as a simple indicator of similarity in the observed distribution of the pollutants in time. By replacing the interpolated values E_i in Eq. (2) with the observed values in other month of the year, it is possible to determine the degree of similarity of the compared months. This use of the IOA may be helpful as a complement for the spatiotemporal analysis of air pollutant distributions.

6. Acknowledgments

7. References

1. O. Leal, M. Mendoza, E. Carranza. "Análisis y modelamiento espacial de información climática en la cuenca de Cuitzeo, México". *Invest. Geog*. no. 72. 2010. pp. 49-67.
2. J. Gómez, J. Etchevers, A. Monterroso, C. Gay, J. Campo, M. Martínez. "Spatial estimation of mean temperature and precipitation in areas of scarce meteorological information". *Atmósfera.* Vol. 21. 2008. pp. 35-56.
3. L. Qu, L. Li, Y. Zhang, J. Hu. "PPCA-based missing data imputation for traffic flow volume: A systematical approach". *IEEE Transactions on Intelligent Transportation Systems*. Vol. 10. 2009. pp. 512-522.
4. K. Grønskei, S. Walker, F. Gram. "Evaluation of a model for hourly spatial concentration distributions". *Atmospheric Environment. Part B. Urban Atmosphere.* Vol. 27. 1993. pp. 105-120.
5. M. Rooney, R. Arku, K. Dionisio, C. Paciorek, A. Friedman, H. Carmichael, et al. "Spatial and temporal patterns of particulate matter sources and pollution in four communities in Accra, Ghana". *Science of the Total Environment.* Vol. 435-436. 2012. pp. 107-114.

6. M. Albert, M. Schaap, A. Manders, C. Scannell, C. O'Dowd, G. Leeuw. "Uncertainties in the determination of global sub-micron marine organic matter emissions". *Atmospheric Environment*. Vol. 57. 2012. pp. 289-300.

7. M. Bechle, D. Millet, J. Marshall. "Remote sensing of exposure to NO_2: satellite versus ground based measurement in a large urban area". *Atmospheric Environment*. Vol. 69. 2013. pp. 345-353.

8. M. Žukovič, D. Hristopulos. "Environmental time series interpolation based on Spartan random processes". *Atmospheric Environment*. Vol. 42. 2008. pp. 7669-7678.

9. A. Pollice, G. Jona. "Two Approaches to Imputation and Adjustment of Air Quality Data from a Composite Monitoring Network". *Journal of Data Science*. Vol. 7. 2009. pp. 43-59.

10. C. Willmott, S. Ackleson, R. Davis, J. Feddema, K. Klink, D. Legates, et al. "Statistics for the Evaluation and Comparison of Models". *J. Geophys. Res.* Vol. 90. 1985. pp. 8995-9005.

11. C. Willmott, S. Robeson, K. Matsuura. "A refined index of model performance". *International Journal of Climatology*. Vol. 32. 2012. pp. 2088-2094.

12. J. Urrutia, R. Palomino, H. Salazar. "Metodología para la imputación de datos faltantes en Meteorología". *Scientia et Technica*. no. 46. 2010. pp. 44-49.

13. D. Deligiorgi, K. Philippopoulos. Spatial Interpolation Methodologies in Urban Air *Pollution Modeling: Application for the Greater Area of Metropolitan Athens, Greece*. 2011. Available on: http://cdn.intechopen.com/pdfs-wm/17390.pdf. Accessed: June 01, 2014.

14. Ü. Şahin, C. Bayat, O. Uçan. "Application of cellular neural network (CNN) to the prediction of missing air pollutant data". *Atmospheric Research*. Vol. 101. 2011. pp. 314-326.

15. B. Huang, B. Wu, M. Barry. "Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices". *International Journal of Geographical Information Science*. Vol. 24. 2010. pp. 383-401.

16. W. Tobler. "A computer movie simulating urban growth in the Detroit region". *Economic Geography*. Vol. 46. 1970. pp. 234-240.

17. P. Kang. "Locally linear reconstruction based missing value imputation for supervised learning". *Neurocomputing*. Vol. 118. 2013. pp. 65-78.

18. R. Bilonick. "Risk qualified maps of hydrogen ion concentration for the New York state area for 1966-1978". *Atmos. Environ*. Vol. 17. 1983. pp. 2513-2524.

19. R. Sivacoumar, K. Thanasekaran. "Line source model for vehicular pollution prediction near roadways and model evaluation through statistical analysis". *Environmental Pollution*. Vol. 104. 1999. pp. 389-395.

20. G. Polydoras, J. Anagnostopoulos, G. Bergeles. "Air quality predictions: dispersion model vs. Box–Jenkins stochastic models. An implementation and comparison for Athens, Greece". *Applied Thermal Engineering*. Vol. 18. 1998. pp. 1037-1048.

21. M. Lorber, A. Eschenroeder, R. Robinson. "Testing the USA EPA's ISCST-Version 3 model on dioxins: a comparison of predicted and observed air and soil concentrations". *Atmospheric Environment*. Vol. 34. 2000. pp. 3995-4010.

22. A. Kousa, J. Kukkonen, A. Karppinen, P. Aarnio, T. Koskentalo. "Statistical and diagnostic evaluation of a newgeneration urban dispersion modeling system against an extensive dataset in the Helsinki area". *Atmospheric Environment*. Vol. 35. 2001. pp. 4617-4628.

23. D. Rojas. "Spatial interpolation techniques for estimating levels of pollutant concentrations in the atmosphere". *Rev. mex. de física*. Vol. 53. 2007. pp. 447-454.

24. D. Ibarra. "Distribución espacial del pH de los suelos agrícolas de Zapopan, Jalisco, México". *Agric. Téc. Méx.* Vol. 35. 2009. pp. 267-276.

25. Y. Xie, T. Chen, M. Lei, J. Yang, Q. Guo, B. Song, X. Zhou. "Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: Accuracy and uncertainty analysis". *Chemosphere*. Vol. 82. 2011. pp. 468-476.

26. E. Jabot, I. Zin, T. Lebel, A. Gautheron, C. Obled. "Spatial interpolation of sub-daily air temperatures for snow and hydrologic applications in mesoscale Alpine catchments". *Hydrological Processes*. Vol. 26. 2012. pp. 2618-2630.

27. K. Stahl, R. Moore, J. Floyer, M. Asplin, I. McKendry. "Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density". Agricultural and Forest Meteorology. Vol. 139. 2006. pp. 224-236.

28. J. Li, A. Heap. "A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors". *Ecological Informatics*. Vol. 6. 2011. pp. 228-241.

29. J. Li, A. Heap. "Spatial interpolation methods applied in the environmental sciences: A review". *Environmental Modelling & Software*. Vol. 53. 2014. pp. 173-189.