Boletín de la Asociación de Geógrafos Españoles N.º 65 - 2014, págs. 495-500
I.S.S.N.: 0212-9426

URBAN AIR QUALITY MODELLING. A METHODOLOGICAL EXPERIMENT BASED ON SPATIAL INTERPOLATION TECHNIQUES

María Rosa Cañada Torrecilla, Antonio Moreno Jiménez & Heizel González Lorenzo
Departamento de Geografía. Universidad Autónoma de Madrid
rosa.canada@uam.es, antonio.moreno@uam.es, heizelglez@gmail.com

I. INTRODUCTION

Air is a vital resource and air pollution can be harmful to human health and the environment. Therefore, for years now, the quality of the atmosphere and its protection has become an essential goal in environmental policy. Industrialization and urban land development processes have led to the degradation of air quality, making atmospheric pollution in cities a universal public health problem and one of the world's first priorities according to the World Health Organization (WHO).

It is patently clear that knowledge of local pollution levels, in particular in the interior of cities, is insufficient owing to the limitations deriving from the number and the spatial distribution of monitoring stations. Overcoming this situation requires spatial estimation instruments, which constitute a major branch of scientific development in the field of earth sciences. It is envisaged that, with these, and from a sample of empirical observations, estimated values can be obtained for the locations lacking such measurements. In this regard, however, several problems arise when applying conventional spatial interpolation. These may originate from causes such as a particularly biased distribution, an insufficient number of sample points, an inadequate sampling scale, i.e. the spatial variability of the phenomenon occurs at a considerably higher spatial scale (that is to say, at smaller distances) than the sampling scale (that is to say, with greater distances between sampling points), etc. All these are liable to give – as, in fact, has been shown in some cases – an insufficient basis from which to reach successful and credible estimates. Spatial interpolation techniques can be considered as simple, but at the same time sophisticated, tools. They are simple from a conceptual point of view, if we consider the number of variables or factors that make up the interpolation formula. For example, in the case of air pollution, we can mention those referring to time and the associated atmospheric stability/instability, winds, spatial dispersion of emissions mechanisms, etc. This places certain limits on their capability. At the same time they are sophisticated, partly on account of their mathematical formulation of uneven complexity, and partly, and more importantly, because they require meditated (non-automatic) decisions to specify a great many of the parameters in the formulas, whose influence on the results appear to be decisive. Therefore, intelligent handling of these techniques is essential, since many implicit factors must be subsumed which impact the spatial phenomenon being estimated. Spatial estimates are difficult to validate, except at the sparse sampling points.

Processes for the generation of spatial patterns are complex, in addition to which the mechanisms operating in each type of process are highly specific. To refine the estimating capacity, therefore, a thorough empirical knowledge of the phenomenon and of the relevant geographical environment is needed, so that the results can also be evaluated with contextual or qualitative criteria.

In view of the above, it seems advisable as a general aim to work toward establishing procedural protocols to improve estimates, given the adverse circumstances at the outset. This paper describes a strategy designed to estimate the concentration of nitrogen dioxide (NO_2) in the atmosphere at two Spanish cities with very different environmental conditions: Barcelona, with serious pollution problems, and Santa Cruz de Tenerife, with a good quality environment.

II. BACKGROUND AND CURRENT STATUS

This article forms part of a line of research on air quality in cities and the socio-environmental injustices deriving therefrom. Within this framework a number of methodological issues are studied, such as the comparison of results from interpolation methods, analysing the differences and similarities arising from the spatial patterns obtained with different techniques, assessing their degree of reliability and determining the most appropriate among them, applying both statistical and geographical criteria.

Many authors have compared different interpolation/spatial estimation methods, taking into consideration the most influential factors: sampling density, spatial distribution of samples, grouping, surface type, data variance, normality, target spatial resolution, etc. Other authors have recorded the diversity of distance radii adopted to generate estimates at new points, based on data from monitoring stations, and the variable number of stations included in the calculation formula or issues related to the exponent affecting distance.

One crucial difference in interpolation methods is the manner of determining sample point weighting. If determinist methods are used, such as for instance inverse distance weighting (IDW), weighting will depend on the distance between monitoring stations, raised to an exponent that is arbitrary and independent with regard to the data, although in reality no such lack of definition exists since this value depends on factors such as orography, prevailing winds, etc. that condition the concentration of pollutants. Conversely, Kriging, an optimal interpolator, uses the semivariogram to determine weights which depend on the spatial autocorrelation statistics of the sampled data set; the weights, therefore, depend on the values of the actual data.

The work plan consisted of the execution of a series of experiments, according to a selective trial-and-error strategy which aims to achieve, after a limited number of operations, acceptable technical and environmental results from the available data.

III. DATA AND METHODS

3.1. Data sources

The environmental data for Barcelona were provided by the *Xarxa de Vigilància i Previsió de la Contaminació Atmosfèrica* (XVPCA) [Air pollution surveillance and prediction network], which is dependent on the *Direcció General de Qualitat Ambiental, Departament de Territori i Sostenibilitat de la Generalitat de Catalunya* [Directorate General for Environmental Quality, Department of Territorial Affairs and Sustainability, Regional Government of Catalonia], while the data for Santa Cruz de Tenerife were provided by the *Consejería de Educación, Universidades y Sostenibilidad del Gobierno de Canarias* [Department of Education, Universities and Sustainability of the Regional Government of the Canary Islands].

The selected environmental variable was the mean annual concentration of NO_2 for the year 2010 in Barcelona, and for 2012 in Santa Cruz de Tenerife. We note that the mean annual rate permissible under Spanish, European and WHO regulation is 40 $\mu g/m^3$. The basic digital cartography used consists of the spatial delimitations established by the INE [Spanish National Institute of Statistics] for districts and municipalities. The analysis, however, is restricted to the areas described as «urban populated zone», excluding manifestly non-residential and undeveloped urban territory.

3.2. Methodology for air pollution spatial interpolation

To meet the aims of this study, a procedure has been designed comprising several stages. In the first stage, an exploratory spatial analysis of the pollution sample data is carried out to obtain measures of location, spread and shape.

In the second stage, the data are subjected to a structural analysis by means of building a semivariogram, to reveal whether semivariance changes only with distance or also as a function of direction. The objective at this point lies in obtaining a series of statistical results such as range, sill and nugget, and to identify the directions of maximum and minimum spatial continuity. In this second stage we have also used, as an exploration tool, a global polynomial interpolation (GPI), which provides orientation on the general spatial pattern of the variable.

In the third stage in this analysis, both of the selected interpolation methods are applied. The first of these, deterministic, precise and local, applies inverse distance weighting (IDW). The second method, geostatistical and analytical, applies ordinary Kriging, taking into account the spatial autocorrelation of the interpolated variable. The aim at this point is to compare the results and determine their reliability, adopting conventional statistical criteria, such as goodness of fit, and other more qualitative geographical criteria.

These techniques were implemented by means of the ArcGIS Geostatistical Analyst 10.1 extension. The final goal was to elucidate which interpolation method gave fewer errors and, at the same time, presented a spatial distribution of the pollutant most compatible and coherent with the territorial framework and local urban structure.

The results of interpolation is a geostatistical surface that must be converted to a raster image, having selected a resolution of 50 m. Subsequently, and using map algebra, the raster layer is fitted to the layer representing the populated urban zone. The same statistical data describing this fitted interpolated layer will be used in the next section to check and evaluate the degree to which the predicted NO_2 values match those actually observed.

IV. ANALYSIS OF RESULTS

4.1. Case study: Barcelona

The univariate description of mean annual nitrogen dioxide levels in Barcelona show a high concentration of values observed close to the average value, which explains the high kurtosis value. Its coefficient of skewness is positive, although with two distinct tails representing the lowest and highest values in the series. Likewise, data distribution is not normal and shows a trend toward greater values near the central part of the municipality, which decrease from the centre outwards due north, south, east and west.

The semivariogram surface confirms that the semivariance of nitrogen dioxide data in Barcelona presents spatial differences, both as a function of distance and as a function of direction. It can be observed that levels are higher in the central area and decrease toward the northeast and southwest. The use of GPI allows the anisotropic behaviour of nitrogen dioxide in Barcelona to be spatially determined, on the basis of the data registered at the monitoring stations.

The implementation of preliminary exploration techniques has corroborated the anisotropic behaviour of NO_2 in Barcelona: in other words, that in a given direction, generally NNE-SSW, the variable demonstrates greater spatial continuity, while in the perpendicular direction, WNW-ESE, spatial continuity is smaller. This means that the differences in semivariance in the second case are considerably sharper at shorter distances. The reasons for this behaviour in the case of Barcelona appear linked to the topographical features of the city, located on a lightly sloping platform enclosed between the sea and the Sierra de Collserola mountains to the northeast, with prevailing winds from the north aiding the dispersion of pollutants further along the axis of greatest spatial continuity.

Implementation of the Kriging technique showed that anisotropic Kriging yielded better results, not only in validating the model, but also the spatial configuration of the pollutant. None the less, the greater homogeneity produced led to an increase in the distance between the estimated and the observed mean values.

The IDW technique revealed that the anisotropic model yielded results that were closer to the local geographical situation, as well as estimates that match empirical values more closely.

Lastly, mean squared prediction errors were fewer with IDW than with Kriging, which further endorses the former interpolation method.

An examination of the most representative maps resulting from interpolation, using IDW and anisotropic Kriging techniques, shows a great extension covering most of the centre of the city and its port area containing the highest concentration levels of NO_2, exceeding 46 $\mu g/m^3$ and reaching 63 $\mu g/m^3$. This area includes the quarters of Gràcia and part of St. Gervasi, traversed by major avenues carrying heavy traffic, such as Ronda General Mitre, Vía Augusta, Travessera de Dalt or Avenida de la Diagonal. We must also mention the pollution produced by the industrial activity located in the city's port facilities. In this sector, pollution levels diminish gradually, with values in the range of 40 to 46 $\mu g/m^3$, both toward the NE

(Poblenou) and the NW. The least polluted areas (values below 40 µg/m^3) are Vall d'Hebron and Torre Girona, a buffer zone comprising residential areas and large green spaces between the city and Sierra de Collserola.

4.2. Case study: Santa Cruz de Tenerife

The NO_2 data present very little variability. Their distribution is practically symmetrical, platykurtic and closely resembles normal distribution. The semivariogram surface confirms the spatial differences in semivariance in function of distance and direction. The lowest semivariance values lie on a NW-SE axis, while from the centre of the municipality these values increase toward the NE and the SW. First-degree GPI and the theoretical adjustment of the semivariogram confirm the anisotropy behaviour of mean nitrogen dioxide levels in the municipality of Santa Cruz de Tenerife.

The anisotropy Kriging method showed better results, both in validation of the model and in spatial configuration. However, it smoothed the predicted values, increasing the difference between the observed and the estimated values.

The results obtained using anisotropic IDW techniques presented a slightly higher rate of mean squared errors than those obtained with the Kriging techniques. The average of the data calculated by all models was very similar and close to the observed mean value. The highest discrepancy rate occurred in the maximum and minimum estimated values, while IDW produced a better adjustment between observed and estimated values, and Kriging reduced the maximum values and increased the minimum values, producing a greater smoothing effect.

The spatial configuration generated by anisotropic Kriging and IDW was fairly similar, although the surface deriving from IDW presented a slightly higher frequency of spatial incoherencies than from Kriging, which together with higher RMSE values makes it difficult to give it precedence over Kriging. The main drawback to Kriging is that it presents greater smoothing effect. In summary, the criteria do not warrant a clear choice.

In both cases, the estimated values indicate a convincing overall NE-SW orientation, confirmed on calculating the anisotropy. NO_2 values decrease from the coast toward the NW, and continue to decrease the greater the distance from the city centre and the greater the altitude.

V. CONCLUSIONS

This paper examines a crucial aspect of environmental studies, namely the estimation of pollution values at locations for which no measurements are available, from a sample of empirical observations performed on a series of points or stations, using spatial interpolation techniques. In the study, we have tested a procedural protocol to assist technical decision-making in order to improve such estimates on urban pollution on the basis of insufficient data, at affordable costs.

From a methodological viewpoint, the work has been structured into three stages, each contributing toward the common goal. In the first stage, univariate distribution of the sample data was explored with several tools, and research was conducted on the existence of anisotropic behaviour in the spatial distribution of NO_2 by means of the semivariogram, which assumption was confirmed.

In the second stage, sample data was subjected to structural analysis through the use of two complementary techniques: on one hand, the adjustment of a function over the empirical semivariogram, to calculate objectively the spatial anisotropy parameters for NO_2 pollution. In particular, the techniques yielding the most satisfactory results for Barcelona were: ellipse dimensions 9000 m by 5000 m, orientation 30º, between 6 and 4 neighbours to include, and the ellipse divided into four 45º sectors. For the city of Tenerife, the best parameters were: 15000 m and 3000 m for the axes of the ellipse, orientation 43º, between 7 and 4 neighbours to include, and the ellipse, as for Barcelona, divided into four 45º sectors.

To the above we added the use of GPI (or trend surface analysis) to reveal the overall latent trend in the spatial distribution of the pollutant. Both intermediate results had considerable weight in the methodological decisions taken in the following stage.

In the third and final stage two distinct interpolation techniques were employed: IDW and ordinary Kriging. The resulting statistics made it possible to validate the resulting models and to compare the spatial patterns of NO_2 pollution.

The spatial patterns generated presented some notable differences, although the most similar were those yielded by anisotropic IDW and Kriging. In choosing between these two options, on the basis of statistical and qualitative criteria, the anisotropic IDW solution would be the more favourable choice for the city of Barcelona and, less clearly, the anisotropic Kriging solution for Santa Cruz de Tenerife. Both solutions were obtained by means of systematic trials, in which acceptable (but not necessarily the best) goodness of fit indicators were combined with estimated values that were more coherent with the urban structure and the territorial framework.

To sum up, it is worth pointing out that the difficulty in generating estimates for pollution levels over a complete urban space, when data are sparse and limited in their territorial representativeness, can be offset by making intelligent use of a set of analytical and spatial modelling tools. With a procedure such as the one adopted in this study it is possible to progress, on a more solid ground, toward a more comprehensive knowledge of local pollution patterns, which subsequently facilitates their insertion in studies and policies dealing with quality and environmental justice, morbidity and environmental health, socio-environmental discomfort, sustainability, etc.

ACKNOWLEDGEMENTS

This paper forms part of the research project «Urban Air Pollution and Environmental Justice: Evaluation Methods and Case Studies with Geographical Information Systems», funded by the Spanish Ministry of Economy and Competitiveness, reference CSO2011-26177.