Building an Environmental Quality Index for a big city: a spatial interpolation approach with DP2

Montero, José María; Larraz, Beatriz and Chasco, Coro
Universidad de Castilla-La Mancha, Universidad Autónoma de Madrid

24. September 2008

Building an Environmental Quality Index for a big city: a spatial interpolation approach with DP2[1]

José Mª Montero, Universidad de Castilla-La Mancha
Beatriz Larraz, Universidad de Castilla-La Mancha
Coro Chasco, Universidad Autónoma de Madrid

ABSTRACT

The elaboration of Environmental Quality Indexes (EQI) for big cities is one of the main topics in regional and environmental economics. One of the usual methodological paths consists of generating a single measure as a linear combination of several air contaminants applying Principal Component Analysis (PCA). Then, as a final step, a spatial interpolation is carried out to determine the level of contamination across the city in order to point out the so-called 'hot points'. In this article, we propose an alternative approach to build an EQI introducing some methodological and practical novelties. From the point of view of the selection of the variables, first we will consider noise -joint to air pollution- as a relevant environmental variable. We also propose to add 'subjective' data -available at the census tracts level- to the group of 'objective' environmental variables, which are only available at a number of environmental monitoring stations. This combination leads to a mixed environmental index (MEQI), which is more complete and adequate in a socioeconomic context. From the point of view of the computation process, we use kriging to match the monitoring stations registers to the Census data. We follow an inverse process as usual, since it leads to better estimates. In a first step, we krige the environmental variables to the complete surface and finally, we elaborate the environmental index. At last, in order to build the final synthetic index, we do not use Principal Components Analysis -as it is usual in this kind of exercises- but a better one, the Pena Distance method (DP2).

Key words: Environmental index, Air pollution, Noise, Subjecive expectations, Kriging, Distance indicators

JEL codes: C21, C43, Q53

1. Introduction

Air pollution is at the top on the list of citizens' environmental concerns. This is particularly true in big cities where more than half the world's population (3.3 billion people) lives. The link between air quality and human health worries many health experts, policy-makers and citizens. The World Health Organization states that almost

[1] Jose-Maria Montero and Beatriz Larraz acknowledge financial support from the FEDER PAI-05-021 Project of the *Junta de Comunidades de Castilla-La Mancha*. Coro Chasco acknowledges financial support from the Spanish Ministry of Education and Science SEJ2006-02328/ECON and SEJ2006-14277-C04-01.

2.5 million people die each year from causes directly attributable to air pollution. In this sense, the elaboration of Environmental Quality Indexes (EQIs) for big cities is one of the main topics in regional and environmental economics. Making EQIs can pursuit several objectives. The main one is to report daily air pollution levels to the public in order to prevent from potential health effects of air pollutants and determine specific actions when alert thresholds are exceeded. Environmental variables are also important as determinants of housing prices. In effect, it is reasonable to assume that pollution enters into the utility function of potential house buyers, since consumers are willing to pay for environmental goods, such as air quality, absence of acoustic pollution, etc. In the two last decades, Smith and Kaoru (1995), Smith and Huang (1993, 1995), Kim et al. (2003), Anselin and Le Gallo (2006) and Anselin and Lozano-Gracia (2008) among others, are good examples of the focus on hedonic property-value models for estimating the marginal willingness of people to pay for a reduction in the local concentration of specified air pollutants.

For all the abovementioned reasons, in this paper we elaborate an EQI for the municipality of Madrid (Spain), at the spatial level of census tracts, since there are no similar measures for this city. In addition, we propose some methodological and practical improvements, which are novel in this kind of analysis. From the point of view of the selection of the variables, first we will consider noise -joint to air pollution- as a relevant environmental variable. We also propose to add ‘subjective’ data -available at the census tracts level- to the group of ‘objective’ environmental variables, which are only available at a number of environmental monitoring stations. This combination leads to a more complete mixed (objective-subjective) environmental index (MEQI), which is more adequate in socioeconomic contexts. From the point of view of the computation process, we will use kriging to match the monitoring stations registers to the Census data -which are available for the much numerous census tracts. We follow an inverse process as usual: in a first step, we krige the environmental variables to the complete surface and finally, we elaborate the environmental index. It can be demonstrated that this process leads to better estimates (less MSE). At last, in order to build the final synthetic index, we do not use Principal Components Analysis -as it is usual in this kind of exercises- but a better one, the Pena Distance method (DP2).

The paper is organized as follows. In the following section, we present the methodological aspects used in the paper. In the third section, we describe the complete construction process of a Mixed Environmental Quality Index (MEQI) for the city of Madrid. The article concludes with a summary of key findings and future research.

2. Methodological questions

2.1. Selection of the variables

As stated before, in order to build a more complete environmental index, we propose on the one hand, the introduction of noise and on the other hand, the consideration of subjective data to the group of objective environmental variables. In fact, though noise policies have been implemented in several developed countries in the recent decades, the proportion of the population that is exposed to noise levels above legal limits is still relatively important. For this reason, in the urban contexts, noise levels have an economic value (e.g. on housing prices) that has been quantified in the empirical literature using different methodologies. The hedonic approach is the more dominant. It infers individual preferences as revealed in the markets (Baranzini and Ramírez, 2005). For example, housing market data can be analyzed in order to assess whether and how much of the house selling price differentials can be explained by different noise levels.

We also recommend joining ‘subjective’ to ‘objective’ environmental variables in the composition of the environmental index. In empirical applications, it is quite common to use data extracted from the monitoring stations as environmental variables. These ones are considered as ‘objective’ information in the sense that they are based on observable phenomena. Alternatively, people’s perceptions of contamination, which are usually available in the Census at the level of census tracts, are considered as ‘subjective’ indicators. It must be said that subjective data are not always correlated with the real air quality or noise pollution.

In the specialized literature on hedonic house price models, where these kind of environmental indexes haven been built as explanatory variables (see Escobar 2006), it is not frequent to find applications using a mixture of objective-subjective variables.

Hedonic specifications typically include air pollutants such as ground-level ozone (Banzhaf 2005, Hartley *et al.* 2005 and Anselin and Le Gallo 2006), or particle matter (Chay and Greenstone 2005, Murthy *et al*. 2003), since these are more visible (like smog) and have the greatest impact on health. Sometimes, they include two pollutants, such as carbon monoxide and particle matter or ground level ozone (Neill at al 2007, Anselin and Lozano-Gracia 2008, respectively). Moreover, as far as we know Baranzini and Ramirez (2005) is the only case that considers jointly air and acoustic pollutants and there are no articles considering both objective/subjective pollutants.

In a socioeconomic context, an EQI is more realistic when contains both kind of information. For example, prospective homebuyers most likely evaluate air quality based on whether or not the air 'appears' to be polluted or what people and the media say about the local air contamination (Delucchi *et al*., 2002). The same can be said in the case of noise (Miedema and Oudshoorn, 2001, Nelson, 2004 and Palmquist, 2004). Therefore, mixed –objective and subjective- indexes (MEQI) are preferable to only objective measures.

2.2. The combination of point-data and area-data with kriging

The elaboration of a MEQI implies the combination of different kind of data available at different spatial supports. The objective variables are registered in a small number of monitoring stations, which produces point-data, whereas the Census always provides information for area-data at the level of much numerous census tracts. We also find that the location of the air quality monitoring stations rarely coincides with the acoustic ones. In effect, the location of environmental monitoring stations is based on regular sampling and unfortunately, they are certainly scarce due to both physical and economic constraints. This is the case of many other similar applications as De Iaco *et al*. (2002), which work with an air pollution data set available at 30 locations in Milan district or Anselin and Le Gallo (2006), Anselin and Lozano-Gracia (2008), which consider 27 and 28 stations in California, respectively.

Matching all these heterogeneous data can lead to a well-known situation called the "change of support problem" (COSP). Kriging is very often the solution to overcome this mismatch of spatial support (Gotway and Young 2002), particularly when dealing with socioeconomic data, since it takes into account spatial dependence. In the specialized literature, the usual solution to the abovementioned problem is to interpolate the environmental variable(s) to obtain their interpolated values in the locations where socioeconomic data are available (Census data, housing prices, etc.). Several interpolative alternatives have been considered in recent research: Thiessen polygons, inverse distance method, splines, kriging and cokriging, though the last two ones are more appropriate when dealing with environmental variables (Anselin and Le Gallo, 2006). When dealing with an only spatial environmental variable, kriging is a good option to get optimal estimates, since it considers its spatial dependence[2].

Kriging is a univariate procedure, which interpolates the values of the target variable at unobserved locations using the available observations of the same variable. This interpolation procedure, which is a minimum mean-squared-error method of spatial estimation, produces the best linear unbiased estimator. In order to obtain the interpolative estimates, it uses the covariance or variogram function, which is the spatial equivalent of the autocorrelation function in time series analysis. Kriging strategy is based on the idea that variables follow a stochastic process over space. It takes into account the multidirectional feature of space in a similar fashion as time series in the unidirectional stochastic process. This approach, which has been applied to a wide range of phenomena (Tzeng et al. 2005, Spence et al. 2007), implies dealing with an infinite family of random variables $X(\mathbf{s})$ constructed at all points $\mathbf{s}$ in a region. Depending on the location and the correlation structure, the variables adopt different values. Each observed datum $x(\mathbf{s})$ is supposed to be a realization of the process.

Observing the set of air quality monitoring sites $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$ as a group of n points in a map, the pollution level of pollutant k (for $k = 1, \ldots, K$), measured at each site, could be regarded as a spatial process $X_k(\mathbf{s}_i)$. The observed values are x_{ki}, i.e. the registered level for pollutant k at the i^{th} site. As the monitoring sites only report data for

[2] In a multivariate approach, cokriging can also be a good option since it not only accounts for the spatial dependence of each variable but also for the inter-variable correlation. However, it is more complex than kriging and, in many occasions, does not provide added benefits. For example, it is the case of the so-called 'isotopic case', i.e. when variables are measured at the same monitoring stations. Cokriging also reduces to kriging in the specific case of autokrigeability (Subramanyam and Pandalai, 2004). Besides, when using cokriging, not only valid variograms are needed to represent the structure of the spatial dependence of the variables but also valid cross-variograms.

a limited number of n locations, we use interpolation to estimate the pollution level for each of the j much more numerous census tracts of the city $j, j \in \{1,\ldots,m\}$. The kriged estimate for pollutant k in site j is computed as a weighted average of the levels of this pollutant in the n sampled sites as follows:

$$X_k^*(\mathbf{s}_j) = \sum_{i=1}^{n} \lambda_i X_k(\mathbf{s}_i) \tag{1}$$

being λ_i the weight assigned to pollutant level X_k in the sample site i.

Depending on the nature of stochastic processes, there are different kinds of kriging: simple kriging (SK), ordinary kriging (OK) and universal kriging (UK). In this work, we will use OK since the stochastic processes are intrinsically stationary with unknown constant means. A spatial intrinsically stationary stochastic process is such that for every vector $\mathbf{h}$ linking two locations in the map, $\mathbf{s}_i$ and $\mathbf{s}_i + \mathbf{h}$, the difference of $X(\mathbf{s}_i + \mathbf{h}) - X(\mathbf{s}_i)$ is a second-order stationary stochastic process.

Hence, requiring the classical conditions of unbiasedness:

$$E\left[X_k^*(\mathbf{s}_j) - X_k(\mathbf{s}_j)\right] = 0 \Leftrightarrow \sum_{i=1}^{n} \lambda_i = 1 \tag{2}$$

and minimum error variance:

$$\min V\left[X_k^*(\mathbf{s}_j) - X_k(\mathbf{s}_j)\right] = \min\left(2\sum_{i=1}^{n} \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_j) - \sum_{i=1}^{n}\sum_{l=1}^{n} \lambda_i \lambda_l\, \gamma(\mathbf{s}_i - \mathbf{s}_l)\right) \tag{3}$$

where $\mathbf{s}_i - \mathbf{s}_l$ represents the vector that links each air monitoring stations i, l.

The weights in expression (1) could be achieved from $\boldsymbol{\lambda} = \boldsymbol{\Gamma}^{-1}\, \boldsymbol{\Gamma}_0$ as follows (see in Montero and Larraz 2006, pp. 207-209, a further explanation):

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma(\mathbf{0}) & \gamma(\mathbf{s}_1 - \mathbf{s}_2) & \cdots & \gamma(\mathbf{s}_1 - \mathbf{s}_n) & 1 \\ \gamma(\mathbf{s}_2 - \mathbf{s}_1) & \gamma(\mathbf{0}) & \cdots & \gamma(\mathbf{s}_2 - \mathbf{s}_n) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_1) & \gamma(\mathbf{s}_n - \mathbf{s}_2) & \cdots & \gamma(\mathbf{0}) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix},\ \boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \alpha \end{pmatrix}\ \boldsymbol{\Gamma}_0 = \begin{pmatrix} \gamma(\mathbf{s}_1 - \mathbf{s}_j) \\ \gamma(\mathbf{s}_2 - \mathbf{s}_j) \\ \vdots \\ \gamma(\mathbf{s}_n - \mathbf{s}_j) \\ 1 \end{pmatrix} \tag{4}$$

In this expression, α is a Lagrange multiplier and $\gamma(\mathbf{h})=\frac{1}{2}V\left[X(\mathbf{s}_i+\mathbf{h})-X(\mathbf{s}_i)\right]$ is the variogram function that shows how the dissimilarity between pairs of observations $\mathbf{s}_i$ and $\mathbf{s}_i+\mathbf{h}$ evolves with separation (or distance).

We have followed a two-step procedure to obtain the variograms. First, we have reached ballpark point estimates of the variograms using the classical variogram estimator based on the method-of-moments (Lark and Papritz 2003). Second, in order to ensure a positive definite model, we have fitted a theoretical variogram function (see, e.g. Emery, 2000, pp. 93-104) to the sequence of average dissimilarities in keeping with the linear model of regionalization (see, e.g. Goovaerts 1997, pp. 108-115)[3].

Once presented the kriging rudiments, we will focus on the reason why kriging the environmental variables and then elaborating an index is a better option than following the inverted process. In effect, the usual procedure in the literature consists of building first an environmental synthetic index that will be kriged afterwards to the whole map, arguing that it is a way to transform a multivariate problem in an univariate one (Preisendorfer 1988; De Iaco *et al.* 2001, 2002). Nevertheless, we think that our option –building a synthetic index first and kriging afterwards- is a better option because it leads to a lower error variance (Myers, 1983)[4].

In effect, let the variables of different pollutants, $X_1, X_2, \ldots, X_K$, be intrinsic stationary stochastic processes of order zero. There are two options to linearly estimating an environmental index:

(i) Elaborating a synthetic index with the *K* environmental variables provided by the *n* monitoring stations, $MEQI\left(\mathbf{s}_i\right)$, and after that computing the kriged estimates of this index for the total number of *m* census tracts:

$$MEQI^*\left(\mathbf{s}_j\right)=\sum_{j=1}^{m}\lambda_i\cdot MEQI\left(\mathbf{s}_i\right),\quad j=1,\ldots,m \tag{5}$$

[3] We have used ISATIS v4.1.1. (2001) to reach the OK estimates.
[4] Another alternative could be the direct estimation of the environmental index including the correction factor and the conditions proposed by Matheron (1979), but it is -in our opinion- much more difficult to implement than our proposal.

for $MEQI(\mathbf{s}_i)=\sum_{k=1}^{K} a_k X_k(\mathbf{s}_i)=\mathbf{A}'\mathbf{X}$, being

$\mathbf{A}'=(a_1,\cdots,a_K)$, $\mathbf{X}=\left[X_1(\mathbf{s}_i),\cdots,X_K(\mathbf{s}_i)\right]$ the vectors of weights and variables, respectively.

(ii) Kriging each original variable $X_1(\mathbf{s}),\cdots,X_K(\mathbf{s})$ for the *m* census tracts, and next compute the synthetic index of the interpolated variables $X_1^*(\mathbf{s}),\cdots,X_K^*(\mathbf{s})$ as follows:

$$\widehat{MEQI}(\mathbf{s}_j)=\mathbf{A}'\mathbf{X}_j^*=\sum_{k=1}^{K} a_k X_k^*(\mathbf{s}_j)=\sum_{k=1}^{K}\sum_{i=1}^{n} a_k \lambda_i\, X_k(\mathbf{s}_i) \tag{6}$$

Following Myers (1983, pp.634), it can be demonstrated that:

$$Var\left[MEQI^*(\mathbf{s}_j)-MEQI(\mathbf{s}_j)\right]>Var\left[\widehat{MEQI}(\mathbf{s}_j)-MEQI(\mathbf{s}_j)\right] \tag{7}$$

2.3. The use of DP2 to build environmental quality indexes

Finally, in order to build the global synthetic index, we opt to use a distance indicator, the Pena Distance or DP2, instead of the more commonly used PCA[5]. DP2 is an iterative procedure that weights partial indicators depending on their correlation with a global index. Its most attractive feature is that it uses all the valuable information contained in the partial indicators eliminating all the redundant variance present in these variables (i.e. avoiding multicollinearity). This method has mainly been used to compute quality of life and other social indicators (Pena 1977, Zarzosa 1996, Royuela et al. 2003). However, we propose its use in other fields -like environmental indexes- due to its good statistical properties; i.e. multidimensionality, comparability and comprehensibility.

First, it is a multidimensional indicator, which is able to aggregate different environmental quality variables expressed in different measurement units. Second, it is

[5] PCA and DP2 are complementary -no substitute- methods (see Zarzosa 1996, p. 194 or Cancelo and Uriz 1994, pp. 177-178). The first is capable of reducing the information of a group of variables eliminating redundant information. Nevertheless, DP2 also allows relative comparisons between different spatial units and/or time periods.

a quantitative distance indicator, which allows comparing the environmental quality in several spatial units, since it is referred to a same base or ‘ideal state’. Third, it is an exhaustive indicator, which is not based on a mere reduction of information as PCA. It uses all the ‘valuable information’ contained in the partial indicators; i.e. it gets the statistical information that is not either false or duplicate, which can be interpreted using ordinal or -better- cardinal scales. This property allows including a great number of variables since all useless redundant variance will be removed by the own process, avoiding multicollinearity. Following Ivanovic (1974), the more data are included in the partial indicators (related to the subject matter) the more complete will be the final synthetic index, since each variable always contain unique and proper information not present in the others. DP2 can eliminate all the superfluous common variance selecting only the part of the information which is original.

These characteristics allow including -in the same synthetic index- several sources of pollution, such as air and noise, as well as subjective information. Although these data are measured in different units and can contain more or less repeated information, DP2 distance method will express all them in abstract comparable units, taking into account only the useful variance, excluding the rest.

DP2 is a relatively complex method, which implies several iterations or matrix rearrangements. The point of departure of the whole process is a matrix V of order (K,m), in which m is the number of census tracts and K is the number of partial indicators (which includes both the interpolated objective variables and the subjective ones). Each element of this matrix, v_{kj}, represents the state of the partial indicator k in the census tract j. In this matrix, those partial indicators negatively connected with environmental quality must change their sign (i.e. all their data must be multiplied by -1). On their side, variables positively linked with environmental quality do not suffer any change. As a result, an increase/decrease of the values of any partial indicator will correspond with an improvement/worsening in environmental quality.

In a second stage, we compute a distance matrix D such that each element, d_k, is defined as follows:

$$d_k = \left| v_{kj} - v_{k*} \right| \qquad (8)$$

where v_{k*} is the k^{th} component of the reference base vector $v_* = \{v_{1*} \quad v_{2*} \quad ... \quad v_{K*}\}$. It is necessary to define a reference value for each partial indicator in order to make comparisons -in terms of environmental quality- between different spatial units (census tracts). In quality-of-life applications, it is quite common to consider the minimum value as the reference (Vicéns and Chasco 2001, Sánchez and Rodríguez 2003, CES Murcia 2003). As a result, a higher value in DP2 (which will always adopt positive values) will imply a higher environment quality level, since it implies a longer distance respect to a theoretical 'non-desired' situation[6]. In addition, this property allows making a ranking between the spatial units in terms of environmental quality. Therefore, d_k measures the distance between the partial indicator k in the census tract j and its reference value.

In a third stage, in order to express all the indicators in abstract comparable units, we compute a first global index, the Frechet Distance (DF), which is defined as:

$$DF(j) = \sum_{k=1}^{K} \frac{d_k}{\sigma_k} = \sum_{k=1}^{K} \frac{|v_{kj} - v_{k*}|}{\sigma_k} \quad ; \quad j = 1, 2, ..., m \tag{9}$$

where σ_k is the standard deviation of partial indicator k. For each partial indicator, the distance between two spatial units d_k is weighted by the inverse of σ_k. That is to say, the contribution of each d_k to the global indicator is inversely proportional to their corresponding indicator standard deviation. This weighting scheme, which is similar to those used in heteroskedastic models, gives less importance to those distances with more variability, and vice versa.

DF is a valid concept of distance only in a theoretical situation of uncorrelated indicators. When there is a direct relationship between the partial indicators (as it is usual), DF will include some duplicated information. Therefore, DF must be corrected in order to eliminate this dependence effect (i.e. the redundant information existent in other variables), which is supposed to be linear. This is why -for each spatial unit *j*- DF is the maximum value that can reach DP2, which is defined as follows:

[6] Some indicators have clear reference values (e.g. those legally established by national or international organizations). This is the case of most air quality variables (SO_2, CO, etc.), for which the EU has fixed limit levels for the protection of human health (Official Journal of the European Union 2008). However, we have opted not to use them due to the complexity and diversity of the measurements, which do not match with the average monthly data available for the city of Madrid.

$$DP2(j)=\sum_{k=1}^{K}\frac{d_k}{\sigma_k}\left(1-R^2_{k\cdot k-1,k-2,\dots,1}\right) \quad ; \quad j=1,2,\dots,m \qquad (10)$$

where $R^2_{k\cdot k-1,k-2,\dots,1}$ is the determination coefficient of the regression of each partial indicator k on the others (k–1, k–2,…,1). It expresses the part of the variance of k that is linearly explained by the rest of partial indicators[7]. As a result, the correction factor $\left(1-R^2_{k\cdot k-1,k-2,\dots,1}\right)$ deducts the part of the variation of the observed values that is explained by the linear dependence[8]. Note that R^2 is an abstract concept, which is unrelated with the measurement units of the indicators.

DP2 implies a decision about the entrance order of the partial indicators in the computation process. That is to say, it must be decided which partial indicator k is the first in contributing its variance to the global index, which one will be the second, etc. In this process, the first indicator (k=1) will contribute all its information to the global index (d_1/σ_1). However, the second indicator (k=2) will only add the part of its variance that is not correlated with the first one: $(d_2/\sigma_2)\left(1-R^2_{2\cdot 1}\right)$. Regarding the third indicator, it will contribute to DP2 the part of its variance that is not correlated with the first and the second one: $(d_3/\sigma_3)\left(1-R^2_{3\cdot 2,1}\right)$. And so forth.

Obviously, depending on the decision DP2 will adopt different values. Thus, it is important to find an objective hierarchical method that leads to a unique entrance order of the partial indicators. If DF is a compendium of all the partial indicators, it seems logical to make the selection taking into account the correlation between each partial indicator and DF. The indicator with the highest correlation with DF will be the leader given that it is the most informative; i.e. the indicator that contributes more variance to the global index.

The whole process is a four-step procedure that can be summarized as follows:

[7] If all the partial indicators are uncorrelated, R^2=0 and DP2=DF.
[8] Ivanovic (1963) proposed the I-Distance, which considered the partial coefficients as a correction factor. However, as stated in Pena (1977), this procedure cannot eliminate the redundant information of the DF.

- First, we compute the DF values for each spatial unit using expression (9); i.e. taking into account the reference base vector v_* of minimum values.
- Second, we calculate the correlation coefficients of the partial indicators and DF to ordering the former in accordance with their degree of dependence with the later.
- Third, we compute DP (expression 10) considering the previously determined entrance order of the partial indicators. This first global index is called DP-1.
- Forth, we make a new ranking with the partial indicators in accordance with their correlation degree with DP-1 with the aim of re-computing DP. We call this second global index as DP-2.
- We repeat this iterative process until a convergence is reached; i.e. the difference between two DP contiguous indexes is null. In the case of non-convergent DP values, we can choose the first DP index (or even the average of the two final ones).

The numeric value of DP index has no real sense but it is useful to compare the state of different spatial units (census tracts) about environmental quality. We can rank census tracts according to this criterion. If we use the same variables and method, we can compare our results for Madrid with those obtained in other cities or even in other moments of time. DP2 lets comparing changes in relative positions and even detecting their causes.

3. Building a Environmental Quality Index (MEQI) for the city of Madrid

3.1. Data set

There are several types of air pollutants. These include the primary pollutants, which are directly emitted from a process, and the secondary ones, which are formed in the air when primary pollutants react or interact together to produce harmful chemicals. Primary pollutants are the ones that cause most damage to ecosystems and human health. They are, among others, sulphur dioxide (SO_2), oxides of nitrogen (NO_x), nitrogen dioxide (NO_2), carbon monoxide (CO) and particulate matter (PM). Regarding secondary pollutants, ground-level ozone (O_3) is considered -joint with PM- the most dangerous pollutant for human health.

a) SO_2 is produced by volcanoes, coal burning (e.g. for home heating), road transport, power stations and other industries. When inhaled at very high levels, it results in panting breathing, coughing and -in some occasions- permanent pulmonary damage. SO_2 causes more damage when particulate and other pollution concentrations are high.

b) PM is a general term used for a mixture of solid particles and liquid droplets found in the air. It comes from many sources, including non-combustion processes (24%), industrial combustion plants (17%), commercial and residential combustion, as domestic heating (16%) and power stations (15%). Fine particles can affect lungs, where they cause inflammation, and heart.

c) NO_x is a generic term for mono-nitrogen oxides (NO and NO_2). It means the sum of the volume mixing ratio (ppbv) of nitrogen monoxide (nitric oxide) and nitrogen dioxide expressed in units of mass concentration of nitrogen dioxide (µg/m3). Both oxides are emitted by elevated temperature combustion, mainly in high vehicle traffic areas, such as large cities, and power stations. As well as SO_2, frequent exposure to high concentrations of these gasses affect specially to children and those who suffer from acute respiratory illness.

d) CO is a very poisonous gas, which comes from the incomplete combustion of fuels (e.g. natural gas, coal or wood) being vehicular exhaust its major source. In sensitive individuals, this gas prevents the normal transport of oxygen in the body, affecting particularly to people suffering from heart diseases.

e) O_3 is formed when NO_x and volatile organic compounds, such as hydrocarbon fuel vapours and solvents, react chemically in the presence of sunlight in the lowest layers of the atmosphere (close to the ground). Most of it is produced in hot sunny weather, being more prevalent in summer. This gas has an irritant effect on the surface tissues of the body, such as eyes, nose and lungs. Irreversible damage to the respiratory tract can occur if ground-level ozone is present in sufficiently high quantities.

'Noise pollution' is the named given to the unwanted sound. Noise is the most pervasive environmental pollutant of the modern world. The excessive noise induce imbalance in a person's mental state, affecting its psychological health. It can cause annoyance, high stress levels as well as noise-induced hearing loss. The source of most acoustic pollution worldwide is transportation systems (motor vehicles, aircrafts, rails), as well as machinery and construction works. It is measured in decibels $(dB(A))$.

Apart from these seven aforementioned pollutants, we suggest to complement the 'objective' information with other 'subjective' variables, such as the population perception of pollution, green areas and noise around their homes. Therefore, we will use ten indicators to elaborate a mixed environmental quality index (MEQI), which can synthesize true pollution values with citizen perceptions of their own residential place welfare. The seven objective variables provide all the necessary scientific information about air and sound pollution in a specific area of the city, whereas the three subjective variables measure the opinion of the people about the contamination levels in their neighborhood.

Table 1. Description of the environmental variables

Variables		Statistical font	Unit	Spatial level	Reference
1. Objective indicators					
1.1. Air quality indicators					
SO_2	Sulphur dioxide	Council of Madrid	$\mu g / m^3$	25 stations	Average (Jan. 2008)
CO	Carbon monoxide	Council of Madrid	mg / m^3	25 stations	Average (Jan. 2008)
NO_x	Oxides of nitrogen	Council of Madrid	$\mu g / m^3$	25 stations	Average (Jan. 2008)
NO_2	Nitrogen dioxide	Council of Madrid	$\mu g / m^3$	25 stations	Average (Jan. 2008)
PM	PM_{10} particulate matter (fraction of suspended particles < 10 $\mu g / m^3$ in diameter)	Council of Madrid	$\mu g / m^3$	25 stations	Average (Jan. 2008)
O_3	Ground-level ozone	Council of Madrid	$\mu g / m^3$	25 stations	Average (Jan. 2008)
1.2. Noise pollution indicators:					
L_{Aeq}	Equivalent continuous noise, dB(A)	Council of Madrid	dB(A)	28 stations	Average (Jan. 2008)
2. Subjective indicators					
pollut	Proportion of houses with air-pollution problems in the neighborhood	Census	%	2,358 cen. tracts	October 1, 2001
ngreen	Proportion of houses with scarcity of green areas in the neighborhood	Census	%	2,358 cen. tracts	October 1, 2001
noise	Proportion of houses with noise in the neighborhood	Census	%	2,358 cen. tracts	October 1, 2001

The data used in this paper come from two different sources (Table 1). On the one hand, the environmental 'objective' measures are published in the 'Atmosphere

Pollution Monitoring System' (Council of Madrid)[9]. The six air pollution variables are measured at 25 fixed operative monitoring stations as monthly averages of hourly readings in January 2008. The noise measure comes from 28 fixed operative monitoring stations, which include the above mentioned[10]. It indicates the equivalent continuous noise level in January 2008 (according to the standardized curve *A*). On the other hand, the three 'subjective' variables, which report the opinion of the people about pollution and noise in their own neighborhood, are available in the 2001 Spanish Census of Population at the level of census tracts.

Figure 1 Location of the active monitoring stations in the districts of Madrid

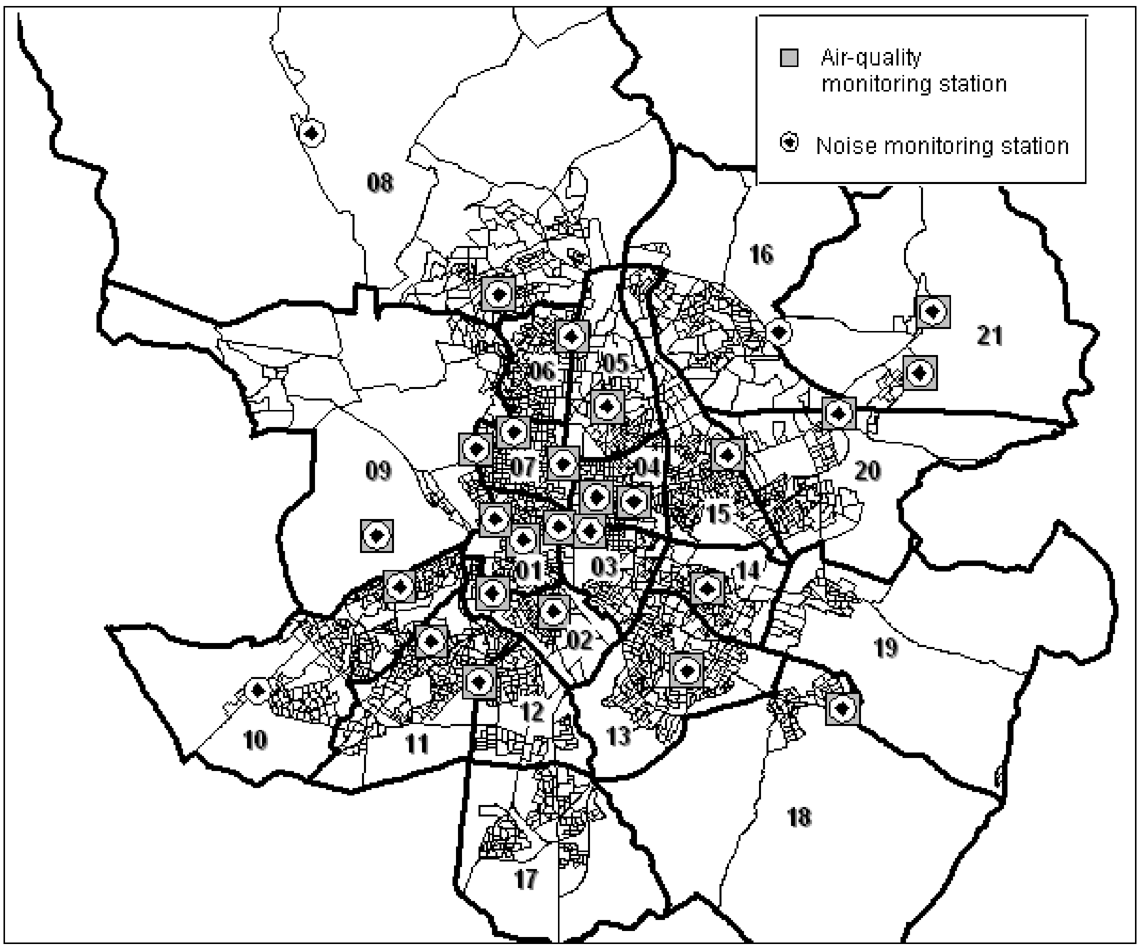


Figure 1 shows the locations of the operative air quality and noise monitoring stations. As it can be seen, most of them are located in the central districts and only a relatively small number can be found in the periphery. Note the reasonable coverage of

[9] These data can be downloaded from the Municipality of Madrid's web page (www.munimadrid.es).
[10] The three noise monitoring stations that do not register pollution are *Cuatro Vientos* (district 10), *El Pardo* (district 8) and *Campo de las Naciones* (district 21).

the domain under study by the monitoring stations since every district has one o more stations or, in the case of the peripheral less densely populated ones, share a station with their neighbors.

3.2. Kriging process

As it was pointed out in the introduction, there is a mismatch between the spatial level of the environmental measured 'objective' variables and the support for 2001 Spanish Census of Population (at the census tract level). This disparity lead us to interpolate the values at the monitoring stations to the locations of every 2,358 census tracts using kriging in order to homogenize the support of the variables considered in the MEQI.

Table 2. Descriptive statistics of the environmental variables

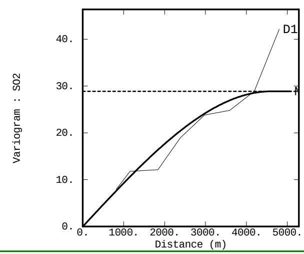
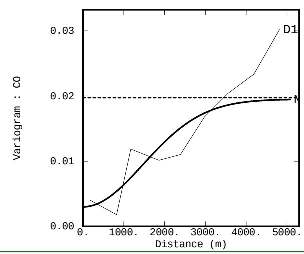
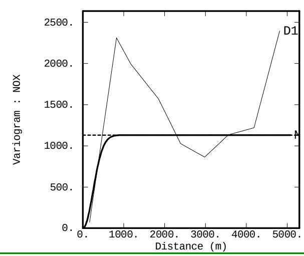
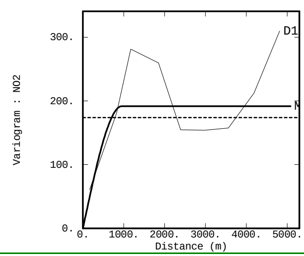
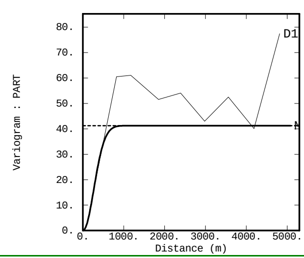
Variables		Min	Max	Mean	PVC*
1. Objective indicators: 25 stations					
1.1. Air quality indicators					
SO_2	Sulphur dioxide ($\mu g / m^3$)	8.00	28.00	16.36	0.33
CO	Carbon monoxide (mg / m^3)	0.23	0.81	0.58	0.23
NO_x	Oxides of nitrogen ($\mu g / m^3$)	84.00	221.00	143.64	0.23
NO_2	Nitrogen dioxide ($\mu g / m^3$)	38.00	93.00	65.76	0.20
PM	PM_{10} particulate matter ($\mu g / m^3$)	25.00	50.00	33.08	0.19
O_3	Ground-level ozone ($\mu g / m^3$)	11.00	23.00	16.68	0.19
1.2. Noise pollution indicators:					
L_{Aeq}	Equivalent continuous noise, dB(A)	50.10	71.30	65.63	0.06
2. Subjective indicators: 2,358 tracts (%)					
pollut	Houses with air-pollution problems	0.00	82.21	25.49	0.51
ngreen	Houses with scarce green areas	0.00	89.44	31.66	0.68
noise	Houses with noise outside	1.15	92.33	38.59	0.31

* Pearson Variation Coefficient (std. dev. / mean)

As a first step, in order to show the structure of the spatial dependence of each variable, we have computed their experimental and fitted theoretical variograms (Table 3). In order to calculate the experimental variograms, we have considered 10 lags with a lag size of 600 meters.

Note that the variogram fitting shows a short range for NO_x, NO_2 and PM. In the rest of the cases, the sill is reached between 2.5 and 5.0 km. Once the variogram models have been chosen, we next proceed to the kriged estimation. (ordinary kriging, OK) of the monthly averages of each pollutant in the total area of Madrid (and thus, in the 2,358 census tracts of this city).

Table 3. Variogram fitting

Environmental pollution variables		Variogram model			Experimental and fitted variogram
			Sill	Range	
Sulphur dioxide ($\mu g/m^3$)	SO_2	Spherical	28.8704	4600	
Carbon monoxide (mg/m^3)	CO	Gaussian Nugget	0.0195 0.0005	3600	
Oxides of nitrogen ($\mu g/m^3$)	NO_x	Gaussian	1130.4704	600	
Nitrogen dioxide ($\mu g/m^3$)	NO_2	Spherical	174	950	
Particulate matter ($\mu g/m^3$)	PM	Gaussian	41.2736	650	

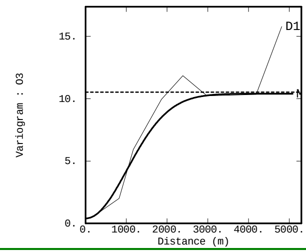
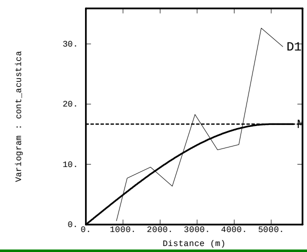
Environmental pollution variables		Variogram model	Sill	Range	Experimental and fitted variogram
Ground-level ozone ($\mu g/m^3$)	O_3	Gaussian Nugget	10 0.4000	2500	
Noise (dB(A))	L_{Aeq}	Spherical	16.6664	5000	

Figure 2.a to 2.f show the resulting kriged values of each objective contaminant in a regular grid (25 meters mesh) and its corresponding standard deviation error map. In general, the precision of the kriged values varies across the sample, becoming worse for locations further away from monitoring sites.

As can be seen in Figure 2, the city center (mainly districts 1 and 7) is the most contaminated place of Madrid, due to the significant emissions of road vehicles. This is why it registers the highest score in all the pollutants, with the exception of ground-level ozone because of the peculiar behavior of this variable. In effect, when O_3 finds NO_x in the air, it reacts and transforms itself into NO_2. For this reason, it usually registers higher values in places with lower concentrations of NO_x -mainly green areas and rural places-, and vice versa. Another sensitive area is the Northeast, in the surroundings of the International Airport, where SO_2 and CO are also intensively active. However, there has not been found an elevated level of noise pollution there, maybe because the corresponding monitoring site is not located just in the airport but in the nearby residence area. The North and West are the least contaminated neighborhoods of the city, coinciding with the most extensive green areas of the city: *El Pardo* and *Casa de Campo*, respectively.

Figures 2.c to 2.d show the resulting estimation map of NO_2, NO_x and PM, respectively, as well as their corresponding standard deviation error maps. These figures clearly show a small range of the variograms, which represent the spatial autocorrelation of such contaminants. Because of this fact, estimates are very accurate nearby the monitoring stations, whereas pollutant levels far from them are approached by the mean. This circumstance must be taken into account when interpreting the kriged values obtained for these three pollutants.

Figure 2 Estimation (left) and standard deviation error (right) maps of the pollution variables

Fig. 2.a. Sulphur dioxide (SO_2)

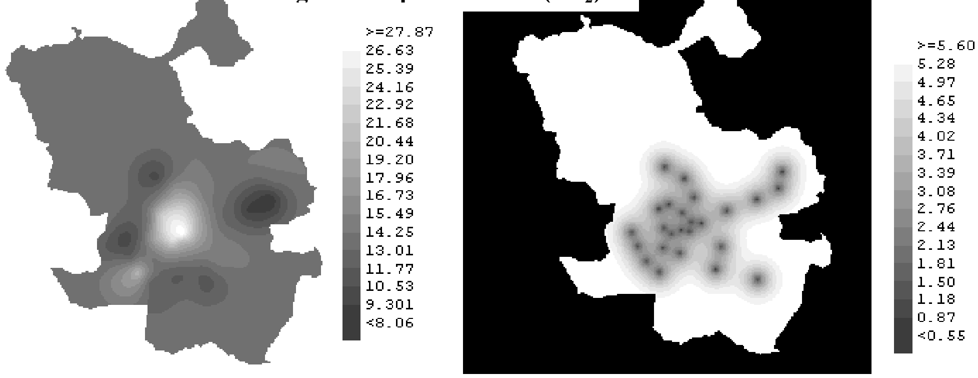


Fig. 2.b. Carbon monoxide (CO)

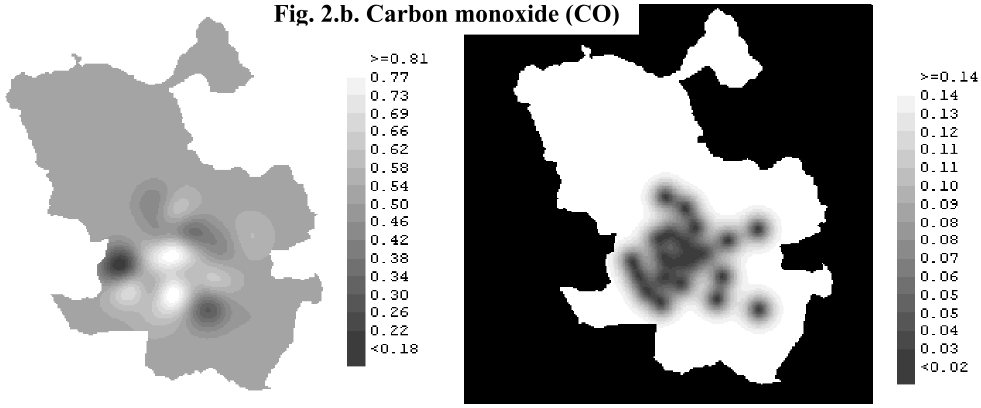


Fig. 2.c. Oxides of nitrogen (NO_x)

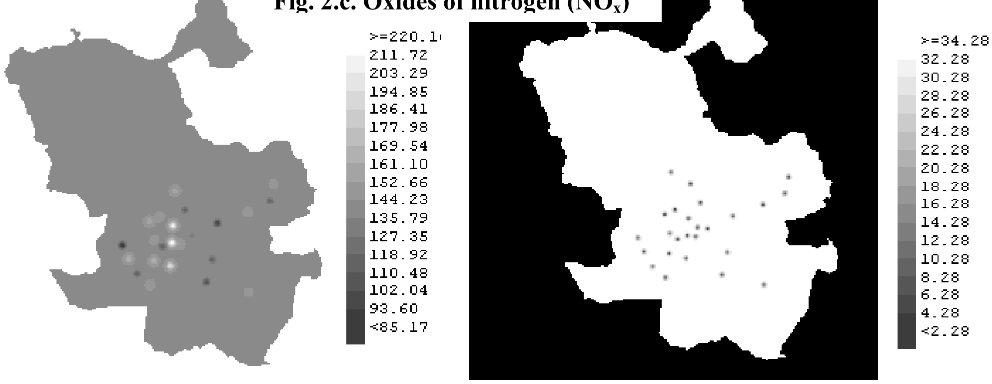


Fig. 2.d. Nitrogen dioxide (NO_2)

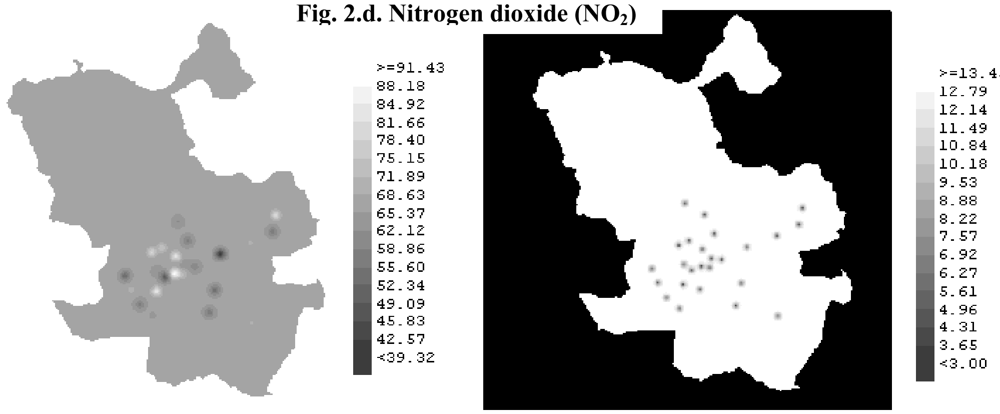


Fig. 2.e. Particulate matter (PM)

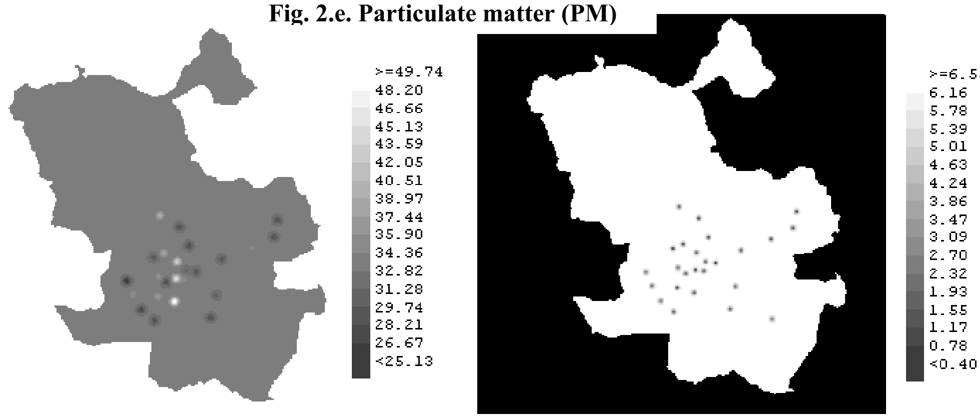


Fig. 2.f. Ground-level ozone (O_3)

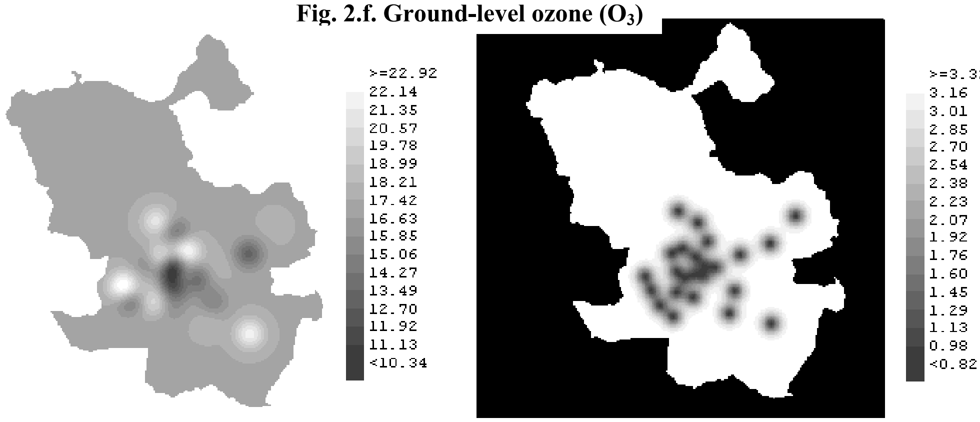
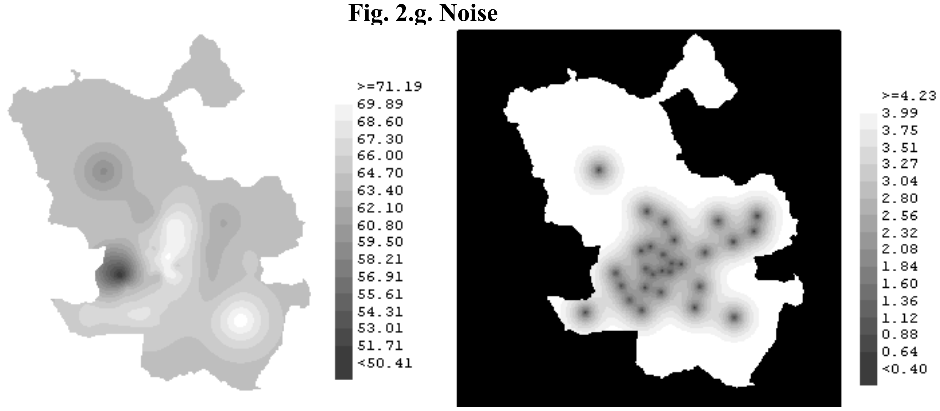


Fig. 2.g. Noise



Regarding noise, this variable highlights other critical parts of the city not sufficiently detected by the other objective indicators. They are mainly the Southeastern industrial districts of the city (18 and 19) and (though to a lesser extent) the East and Northwestern segments of the M-40 bypass, which are the most congested ones.

3.3. Computing the MEQI with DP2 method

In the previous sections, we have presented our research variables and we have kriged the objective indicators with the intention of building the matrix V of partial indicators. Note that we have not followed the same process as usual in this kind of applications. Firstly, we have kriged the objective indicators to secondly building a synthetic index, what allows including other subjective variables. As shown above, this procedure leads to estimators that are more accurate.

In order to compare the results of objective measures (EQI) and mixed objective-subjective indexes (MEQI), we have computed several indexes. Firstly, we have applied DP2 to three matrixes: V_1, which is of order 6 air-pollution objective indicators by 2,358 census tracts, V_2, which is of order 7 objective indicators (air-pollution plus noise) by 2,358 census tracts and V_3, which is of order 10 objective-subjective indicators by 2,358 census tracts. Secondly, we have also applied PCA to the total 10 indicators with the purpose of comparing the MEQI results obtained with both DP2 and PCA.

Before starting the DP2 computation process, we must determine how the partial indicators contribute to the global one; i.e. if they have a positive or negative impact on environmental quality. As it was stated before, all the indicators must have a positive contribution to the phenomenon we are measuring, which is -in our case- environmental quality. In our case, the whole set of variables is negatively related to environmental quality. Hence, we decided not to change the sign of all the variables so as instead of measuring environmental quality, we will measure pollution; i.e. an increase/decrease of the values of any partial indicator will correspond with an improvement/worsening in pollution.

Next, for each matrix V_1, V_2 and V_3, we have computed their corresponding DF considering the minimum value of every partial indicator as the reference base ($v_* = \min\{v_{kj}\}$). As a result, a higher value in the global indicator will imply a higher pollution level, since it implies a longer distance respect to a theoretical 'most-desired' situation. After that, we have calculated the correlation coefficients of each variable and their corresponding DF what lead to a first arrangement of the variables, in order to compute a first estimation of DP2. The final DP2 indexes reached the convergence after 2, 3 and 2 iterations for MEQI, EQI7 and EQI6, respectively. In Table 4, we show the main results of DP2 computations for the three environmental indexes. Additionally, we have also included the coefficients of the first component (F1) computed after the application of PCA. From a total number of 4 components, we have selected the first one, which is an 80% of the total variance (10 environmental indicators).

Table 4. Main results of DP2 and PCA computations for MEQI and EQI

		Statistics for the last iteration of DP2							PCA Comp. 1
		Rank	Correlation coefficient			Correction factor			
		MEQI DP2	MEQI DP2	EQI7	EQI6	MEQI	EQI7	EQI6	MEQI PCA
1. Objective kriged indicators:									
SO_2	Sulphur dioxide ($\mu g/m^3$)	9	0.46	0.56	0.53	0.56	0.62	0.74	0.03
CO	Carbon monoxide (mg/m^3)	5	0.59	0.68	0.71	0.46	0.46	0.45	0.56
NO_x	Oxides of nitrogen ($\mu g/m^3$)	1	0.78	0.90	0.92	1.00	1.00	1.00	0.90
NO_2	Nitrogen dioxide ($\mu g/m^3$)	2	0.72	0.82	0.84	0.21	0.21	0.21	0.86
PM	PM_{10} particulate matter ($\mu g/m^3$)	8	0.52	0.60	0.73	0.37	0.38	0.49	0.88
O_3	Ground-level ozone ($\mu g/m^3$)	10	-0.10	-0.10	-0.10	0.55	0.56	0.57	-0.03
L_{Aeq}	Noise, dB(A)	3	0.66	0.72	-	0.77	0.77	-	0.24
2. Subjective indicators (%):									
pollut	Houses with air-pollution	4	0.66	-	-	0.87	-	-	0.15
ngreen	Houses with scarce green areas	6	0.58	-	-	0.83	-	-	0.06
noise	Houses with noise outside	7	0.56	-	-	0.51	-	-	0.08

Note: *MEQI*: Mixed Environmental Quality Index, *EQI*: Environmental Quality Index, *Rank*: entrance order of partial indicators in the final DP2, *Correlation coefficient*: Pearson correlation coefficient of each indicator with the final DP2, *Correction factor*: $\left(1-R^2_{k\cdot k-1,k-2,\dots,1}\right)$ or the part of the variance that is not explained by the previously introduced indicators, *PCA*: Prinicpal Components Analysis, *Comp. 1*: first component.

Concerning DP2 results, the correlation coefficients of the indicators and the final index -produced by the last iteration of DP2- are quite high and significant for the three indexes. Only in the case of ground-level ozone (O_3) the correlation is low and even negative (-0.10). As already shown, this variable has a peculiar behavior since it experiences an opposite performance than the oxides of nitrogen (NO_x), which is the most influent variable. In effect, NO_x registers the highest correlation with the final DP2 in both indexes. This is why in the computation of DP2, it enters the first contributing all its variance to the final DP2 (correction factor=1). While a primary gaseous pollutant -NO_x- is the most important variable, the second contributor to DP2 is a secondary pollutant (NO_2). Nevertheless, it only donates to the final DP2 a 21% of its variance (correction factor=0.21), since the remaining 79% is already present in NO_x.

In the third place (only in the case of MEQI-DP2 and EQI7), the variable of 'objective' noise (L_{Aeq}) is also highly correlated with DP2 to which it donates a 77% of its variance. For this reason, noise will have a relevant role in those indexes that include this variable. It must be noted that though O_3 is the least important indicator in both global indexes, the rest of partial indicators collects less than the 50% of its variance. This is why it gives to the final DP2 a 55-57% of its information. It must also be highlighted the high level of contribution of the subjective indicators in MEQI-DP2, mainly 'pollut' (proportion of houses with air pollution) and 'ngreen' (proportion of houses with scarce green areas), with a correction factor above 0.80 in both indexes. It can be due to the originality of this richer information (originally available for the complete set of 2,358 census tracts), which is based on citizens' perceptions.

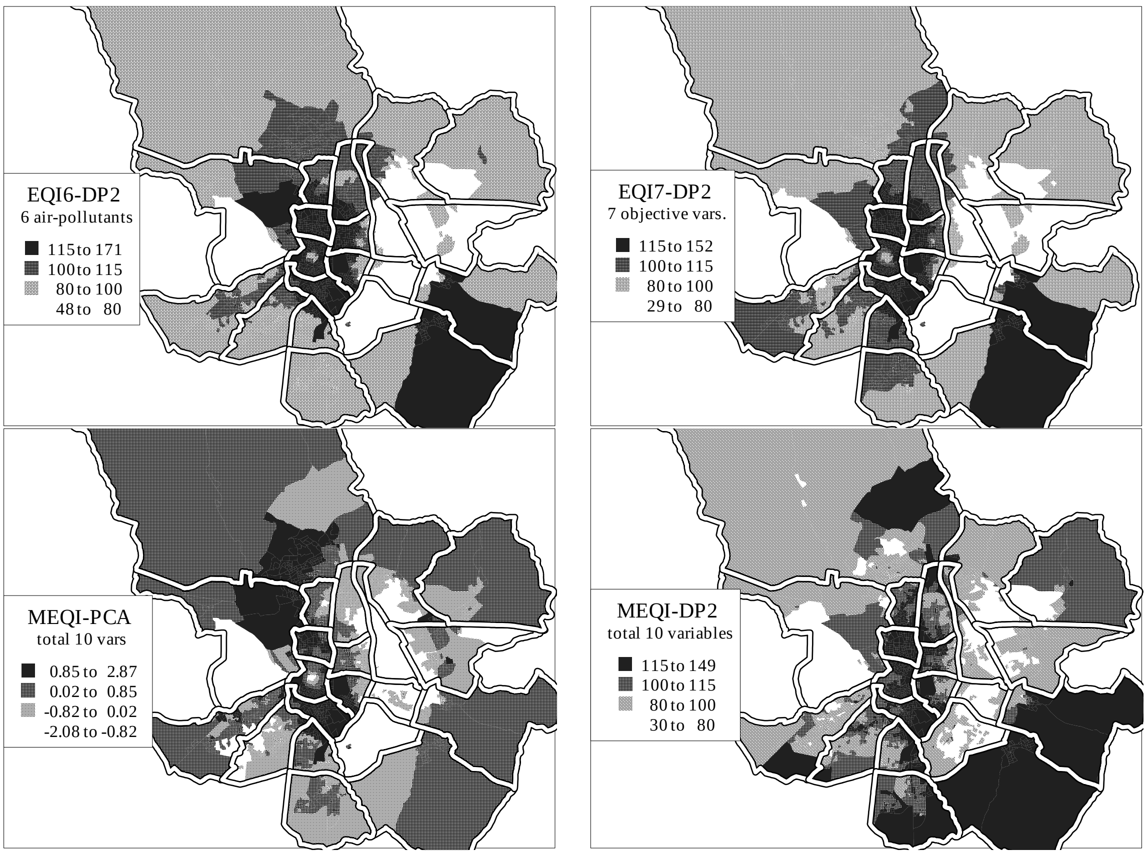
The decisive importance of NO_x and NO_2 in the composition of EQIs (above 0.90 for NO_x and 0.80 for NO_2 in both EQI7 and EQI6) can produce less accurate estimates in the locations faraway the monitoring stations. In effect, as stated before the lowest degree of spatial autocorrelation exhibited by these pollutants (joint with PM) produces kriged estimates only accurate nearby the monitoring stations, whereas the locations far from them are approached by the mean. Even in the case of F1 (first component in PCA), the highest component coefficients correspond to NO_x (0.90), PM (0.88) and NO_2 (0.86)[11]. This is another reason that supports our preference for MEQI (calculated with DP2), in which the important role of these pollutants is shared with other variables.

The computation results are apparently quite similar for the four indexes (EQI6, EQI7, MEQI-PCA and MEQI-DP2), though some interesting differences can be

[11] It must been remarked that the three subjective indicators are basically present in the second component; i.e. the first component cannot conveniently include all the relevant information. All the computations are available upon request from the authors.

detected in their spatial distribution. In Figure 3, we have represented these indexes. However, we have previously standardized the three DP2 variables to facilitate their interpretation. In effect, though the original DP2 values are nonsensical in real terms, it is possible to compute the deviation to the mean value (multiplied by 100). Therefore, a value of 100 will correspond to the DP2 city average and values above/below 100 mean pollution levels better/worse than the city average.

Figure 3 Distribution of the environmental quality indexes in the city of Madrid



Notes: The classification method is "natural breaks" (Jenks and Caspall 1971)

A first analysis of the maps can conclude that they reach to the same conclusion: the highest levels of pollution are concentrated in the 'Central Almond' (the 7 central districts surrounded by the M-30 first belt) and the industrial northern and southeastern peripheries. The lowest levels of pollution seem to be located in some eastern/western districts. Nevertheless, some interesting differences can be appreciated when comparing these results. Actually, EQI7 seems to estimate better those districts in which an extra noise monitoring station is places, particularly districts 8 and 10. Besides, MEQI-DP2, which also includes subjective information, when compared with the objective indexes, penalizes some peripheral neighborhoods affected by the main radial highways and the M-40 second belt; i.e. people seem to be particularly sensitive to traffic congestion and its consequent noise. On the other side, northwestern neighborhoods are better perceived possibly due to their proximity to big green areas (*El Pardo* and *Casa de Campo*), as well as the existence of several groups of high-class residences. Regarding MEQI-PCA, the main function played by the oxides of nitrogen and particulate matter in the final index is possibly biasing the results, since they benefit the southeastern districts but penalize the whole north periphery area. One interesting result is the higher level of pollution detected by both MEQI in the census tracts closed to the International Airport. In effect, as stated before, while the monitoring stations nearby are not located in the same airport, people's perceptions worsen the kriged objective estimates.

4. Main conclusions

As it is well known, the elaboration of Environmental Quality Indexes for big cities is one of the main topics in regional and environmental economics. However, research in this topic is in his early stages and there is a vast field for new insights. In this paper, we have contributed to the development of the topic with several practical and methodological novelties. Concerning the first, we build a Mixed Environmental Quality Index with both objective and subjective environmental indicators. The inclusion of subjective indicators must be regarded because people (e.g. prospective homebuyers) most likely evaluate air quality based on whether or not the air 'appears' to be polluted or what the media say about the local air or noise contamination. In addition, while in the literature it is difficult to find environmental indexes with more than tree partial indicators, we have considered seven objective air-pollution variables (SO_2, CO, NO_x, NO_2, PM and O_3) as well as a noise indicator.

The elaboration of Mixed Environmental Quality Indexes can lead to the well-known 'change of support' problem. In effect, the subjective indicators are commonly available for much more locations than the objective ones. Kriging is the solution we propose to overcome this mismatch of spatial support since it takes into account spatial dependence, which is a usual effect in the environmental variables. Although this scope is not new in the literature, we propose -as an innovation- a change of order in the procedure, since it leads to lower estimation errors. Firstly, we obtain the kriged estimates of the partial objective indicators for the desired locations, and secondly we compute the global index. Furthermore, we also recommend using a distance indicator - the Pena Distance or DP2- instead of other synthesis methods, such as PCA. On the one hand, PCA is based on a mere reduction of information, while DP2 uses all the valuable information contained in the partial indicators, eliminating all the redundant variance present in these variables. On the other hand, DP2 has good statistical properties; i.e multidimensionality, comparability and comprehensibility.

The abovementioned practical and methodological novelties have empirically been checked in a study case: the elaboration of a Mixed Environmental Quality Index for the city of Madrid. Results have been certainly satisfactory and some interesting differences can be detected in their spatial distribution. For example, since the proposed MEQI includes subjective information, when compared with the objective indexes, it penalizes some peripheral neighborhoods affected by the main radial highways and belts. On the other side, it favors those neighborhoods that are close to big green areas and high-class residences. Besides, the PCA estimation is not always capable of including all the relevant information in the first component. In our case, this first component is mainly determined by the oxides of nitrogen and particulate matter, which kriging estimators are less accurate, and seems to underestimate the subjective indicators.

Once shown the main concluding remarks, new future lines of research immediately arise. For instance, in certain situations, cokriging could overcome better than kriging the ‘change of support’ problem or even extending this framework to a spatio-temporal context. Besides, the use of observation networks could reduce the estimation errors in the interpolative stage of the elaboration of the index. At last, in other empirical context, Mixed Environmental Quality Indexes could be used as explanatory variables in hedonic housing price models.

REFERENCES

Anselin L, Le Gallo J (2006) Interpolation of air quality measures in hedonic house price models: Spatial Aspects, *Spatial Economic Analysis* 1-1, pp. 31-52.

Anselin L, Lozano-Gracia N (2008) Errors in variables and spatial effects in hedonic house price models of ambient air quality, *Empirical Economics* 34, pp. 5-34.

Banzhaf HS (2005) Green price indices, *Journal of Environmental Economics and Management* 49(2), pp. 262-280.

Baranzini A, Ramírez JV (2005) Paying for quietness: the impact of noise on Geneva rents, *Urban Studies* 42-4, pp. 633-646.

Cancelo JR, Uriz P (1994) Una metodología general para la elaboración de índices complejos de dotación de infraestructuras, *Estudios Regionales* 40, pp. 167-188.

CES Murcia (2003), La renta familiar disponible bruta y el índice de bienestar de los municipios de la Región de Murcia durante el periodo 1995-2000: estimación y análisis. Available from www.cesmurcia.org.

Chay KY, Greenstone, M (2005) Does air quality matter? Evidence from the housing market, *Journal of Political Economy*, 113(2), pp. 376-424.

Delucchi MA, Murphy JJ, McCubbin DR (2002) The health and visibility cost of air pollution: a comparison of estimation methods, *Journal of Environmental* Management 64, pp. 139-152.

De Iaco S, Myers DE, Posa D (2001) Total air pollution and space–time modelling. In: Monestiez P, Allard D, Froidevaux R (eds.), *GeoEnv III—Geostatistics for Environmental Applications*, Kluwer Academic Publishers, Dordrecht, pp. 45–56.

De Iaco S, Myers DE, Posa D (2002) Space-time variograms and a functional form for total air pollution measurements, *Computational statistics & Data Analysis* 41, pp. 311-328.

Emery X (2000) *Geoestadística lineal*, Departamento de Ingeniería de Minas, Facultad de CC. Físicas y Matemáticas, Universidad de Chile.

Escobar L (2006) Indicadores sintéticos de calidad ambiental: un modelo general para grandes zonas urbanas, *Eure* XXXII-96, pp. 73-98.

Goovaerts P (1997) *Geostatistics for natural resources evaluation*, Oxford University Press, New York.

Gotway CA, Young L (2002) Combining incompatible spatial data, *Journal of the American Statistical Association* 97 (458), pp. 632-648.

Hartley PR, Hendrix ME, Osherson D (2005) *Real estate values and air pollution: measured levels and subjective expectations*, Discussion Paper, Rice University.

ISATIS (2001) *Isatis v4.1.1. Software Manual*, Avon, France: Geovariances and Ecole des Mines de Paris.

Ivanovic B (1963) Classification of underdeveloped areas according to level of economic development, *Eastern European Economics* II, pp. 1-2.

Ivanovic B (1974) Comment ètablir une liste des indicateurs de development, *Revue de Statistique Apliquée* 22-2, pp. 37-50.

Jenks GF, Caspall FC (1971) Error on choroplethic maps: definition, measurement, reduction, *Annals of the Association of American Geographers* 61-2, pp. 217-244.

Kim C-W, Phipps TT, Anselin L (2003) Measuring the benefits o fair quality improvement: a spatial hedonic approach, *Journal of Environmental Economics and Management*, 45, pp. 24-39.

Lark, RM, Papritz A (2003) Fitting a linear model of corregionalization for soil properties using simulate annealing, *Geoderma* 115, pp. 245-260.

Matheron G. (1979) *Recherche de simplification dans un problème de cokrigeage*. Publication nº 628, Centre de Géostatistique, Ecole des Mines de Paris, Fontainebleau.

Miedema HME, Oudshoorn CGM (2001) Annoyance from transportation noise: relationships with exposure metrics DNL and DENL and their confidence interval, *Environmental Health Perspectives*, 109(4), pp. 409-416.

Montero JM, Larraz B (2006) Estimación espacial del precio de la vivienda mediante métodos de krigeado, *Estadística Española* 48, pp. 62-108.

Murthy MN, Gulati SC, Banerjee A (2003) *Hedonic property prices and valuation of benefits from reducing urban air pollution in India*, Delhi Discussion Papers 61, Institute of Economic Growth, Delhi, India. Available from http://ideas.repec.org/p/ind/iegddp/61.html.

Myers DE (1983) Estimation of linear combinations and cokriging, *Mathematical Geology* 15, pp. 633-637.

Neill HR, Hassenzahl DM, Assane DD (2007) Estimating the effect of air quality: spatial versus traditional hedonic price models, *Southern Economic Journal* 73 (4), pp. 1088-1111.

Nelson JP (2004) Meta-analysis of airport noise and hedonic property values, *Journal of Transport Economics and Policy*, 38(1), pp. 1-28.

Official Journal of the European Union (2008) *Directive 2008/50/Ec of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe*. Available from http://eur-lex.europa.eu/JOIndex.do.

Palmquist RB (2004) Property values models. In Mäler K-G, Vincent J (eds) *Handbook of Environmental Economics* vol. 2. North Holland, Amsterdam.

Pena JB (1977) *Problemas de la medición del bienestar y conceptos afines (Una aplicación al caso español)*, Presidencia del Gobierno, Instituto Nacional de Estadística, Madrid.

Preisendorfer RW (1988) *Principal Component Analysis in meteorology and oceanography*. Elsevier, Amsterdam.

Royuela V, Suriñach J, Reyes M (2003) Measuring quality of life in small areas over different periods of time. Analysis of the province of Barcelona, *Social Indicators Research* 64, pp. 51-74.

Sánchez MA, Rodríguez N (2003) El bienestar social en los municipios andaluces en 1999, *Revista Asturiana de Economía* 27, pp. 99-119.

Smith VK, Huang, J-C (1993) Hedonic models and air pollution: twenty-five years and counting, *Environmental and Resource Economics* 36-1, pp. 23-36.

Smith VK, Huang JC (1995) Can markets value air quality? A meta-analysis of hedonic property value models, *Journal of Political Economy* 103, pp. 209-227.

Smith VK, Kaoru Y (1995) Signals or noise—explaining the variation in recreation benefit estimates, *American Journal of Agricultural Economics* 72**,** pp. 419-433.

Spence JS, Carmack PS, Gunst RF, Schucany WR, Woodward WA, Haley W (2007) Accounting for spatial dependence in the analysis of SPECT Brain Imaging Data, *Journal of the American Statistical Association* 102, pp. 464-473.

Subramanyam A, Pandalai HS (2004) On the equivalence of the cokriging and kriging systems, *Mathematical Geology* 36-4, pp. 507-523.

Tzeng S, Huang H-C, Cressie N (2005) A fast, optimal spatial-prediction method for massive datasets, *Journal of the American Statistical Association* 100, pp. 1343-1357.

Vicéns J, Chasco C (2001) *Estimación del indicador sintético de bienestar social publicado en el Anuario Social 2000*, Colección Estudios Sociales. Working Papers 1.

Zarzosa P (1996) *Aproximación a la medición del bienestar social*, Secretariado de Publicaciones e Intercambio Científico, Universidad de Valladolid.