Computers, Environment and Urban Systems 37 (2013) 95–106

Contents lists available at SciVerse ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

A spatio-temporal geostatistical approach to predicting pollution levels: The case of mono-nitrogen oxides in Madrid

José-María Montero-Lorenzo [a,1], Gema Fernández-Avilés [a,1], José Mondéjar-Jiménez [b,*], Manuel Vargas-Vargas [c,1]

[a] University of Castilla-La Mancha, Faculty of Law and Social Sciences, Cobertizo San Pedro Mártir, s/n, 45.071 Toledo, Spain
[b] University of Castilla-La Mancha, Faculty of Social Sciences, Avda. de los Alfares, 44, 16.071 Cuenca, Spain
[c] University of Castilla-La Mancha, Faculty of Economics, Pza. Universidad, 1, 02.071 Albacete, Spain

ARTICLE INFO

Article history:
Received 25 September 2011
Received in revised form 23 June 2012
Accepted 28 June 2012
Available online 13 August 2012

Keywords:
Spatio-temporal kriging
Functional kriging
Composite weighted likelihood
Pollution
NO_x

ABSTRACT

In spite of the effort made in the last years, NO_x is still one of the main pollution problems in large cities. This is why the literature related to predicting NO_x levels is certainly extensive. However, most of this literature does not take into account the spatio-temporal dependencies of such NO_x levels. As spatio-temporal dependencies are a core aspect of pollution, we propose both a spatio-temporal kriging and a functional kriging strategy to incorporate such dependencies into the prediction procedure. We also use an innovative method for estimating the parameters of the non separable space–time covariance function involved in the spatio-temporal kriging strategy, which significantly reduces the computational burden of traditional likelihood-based methods. The empirical study focuses on Madrid City and is backed by a massive hourly database. Results indicate that the functional strategy outperforms the spatio-temporal procedure at non peripheral sites, which is a remarkable finding due to the high computational requirements of spatio-temporal kriging.

1. Introduction

Air pollution is at the top of the list of citizens' environmental concerns. This is particularly true in large 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 2.5 million people die each year from causes directly attributable to air pollution. Therefore, it is no surprise that studies about air pollution are becoming increasingly popular and that environmental issues have brought atmospheric science to the centre of science and technology, where it now plays a key role in shaping national and international policy. Environmental prediction plays a significant role in the planning of human affairs.

Obviously, in large cities infrastructure for communications plays a core role in the social and economic development of citizens, but is also an important source of problems related to pollution. Therefore, the challenge is to ascertain an optimal trade-off between advantages and disadvantages.

In this paper we focus on NO_x, which is a generic term for mono-nitrogen oxides (NO and NO_2). Both oxides are emitted by high temperature combustion, mainly in high vehicle traffic areas, such as large cities and power stations. The major sources of NO_x formation during combustion processes are thermal NO_x, fuel NO_x and prompt NO_x. Nitrogen oxides can have various damaging effects: acid rain, the greenhouse effect, ozone layer depletion and direct harm to human health when they react with hydrocarbon vapors and sunlight to form photochemical smog.

Unfortunately, in spite of the effort made over the last few years, NO_x (and ground-level ozone, O_3) is still one of the main pollution problems in large cities, especially in cities where anticyclones are frequent. In large metropolitan areas, the presence of so much human activity causes all sorts of negative externalities. Increasing transportation of goods and people leads to higher levels of NO_x and as road traffic is related to human activity and needs, much of it occurs in areas where people live, work, go to school, etc. The latter means that today's urban development will result in NO_x being a more serious problem in the future unless efforts are made to mitigate it.

Due to the adverse effects of pollution and more specifically nitrogen oxides, the literature related to predicting nitrogen oxides is certainly extensive. However, most of this literature uses Gaussian dispersion modeling, time series models, neural networks and to a lesser extent, chaos theory, but in the majority of the cases

* Corresponding author. Tel.: +34 902 204 100; fax: +34 902 204 130.
E-mail addresses: Jose.MLorenzo@uclm.es (J.-M. Montero-Lorenzo), Gema.FAviles@uclm.es (G. Fernández-Avilés), Jose.Mondejar@uclm.es (J. Mondéjar-Jiménez), Manuel.Vargas@uclm.es (M. Vargas-Vargas).
[1] Tel.: +34 902 204 100; fax: +34 902 204 130.

0198-9715/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compenvurbsys.2012.06.005

does not take into account the spatial or spatio-temporal dependencies of the level of pollutants (the Gaussian dispersion model can be considered an exception, because implicitly accounts for spatial correlation through the Gaussian kernel, which changes over time). As spatio-temporal dependencies are a core aspect of pollution, we propose both a spatio-temporal kriging and a functional kriging strategy to incorporate them into the prediction of NO_x levels. The spatio-temporal strategy has been used previously in the study of sulfate deposition processes (Bilonick, 1985; Haas, 1998; Kyriakidis & Journel, 2001; Vyas & Christakos, 1997, among others, are pioneer research) and rainfall acidity (Eynon & Switzer, 1983, is a good example), but considering a reduced number of instants of time. However, for massive databases, it requires high computational power to carry out the Maximum Likelihood (ML) estimation of the parameters of the covariance function and, in spite of the recent developments in approximating the likelihood function, this can be unfeasible when the number of instants of time considered is relatively high. This is the main reason why (i) we use a new method for estimating the parameters of the covariance function when dealing with massive databases, which obtains accurate estimates comparable to those obtained by ML, but with lower computation costs, and (ii) we propose an alternative to spatio-temporal kriging, the recently developed functional kriging, with B-spline bases, which have a high degree of flexibility. Functional kriging deals with curves instead of points and preserves the advantages of spatial kriging once the time series data collected at monitoring stations (MS) are represented as curves. It is less computationally expensive than spatio-temporal kriging and allows for spatially studying long time series of pollution measures. As far as we know, no previous research employs spatio-temporal kriging to predict levels of mono-nitrogen oxides, one of the most harmful pollutants for human health. Functional kriging has never been employed to predict pollution. We use our proprietary codes for (i) including the spatio-temporal covariance values in the spatio-temporal kriging equations, (ii) solutions to the computational effort in the estimation of covariance parameters and (ii) functional kriging equations.

We have focused our empirical analysis on Madrid (Spain). There are several important reasons for choosing Madrid as a study case: (i) the population is highly concerned with the environment in general and air quality in particular; (ii) hourly and annual legal standards are exceeded in most parts of the city; (iii) nitrogen oxides are precursors for a number of harmful secondary air pollutants such as ozone and particulate matter and play a role in the formation of acid rain; and (iv) it can be said that in Madrid there is almost perfect information about air quality all over the city (both excellent monitoring site/population and monitoring site/surface area ratios).

Following this introduction, Section 2 reviews the literature related to the prediction of air pollution and, more specifically, nitrogen oxides, as well as the few papers that take into account the spatial or spatio-temporal dependencies of the level of pollutants. Section 3 is devoted to briefly delineating both the spatio-temporal and functional kriging methodologies. Section 4 focuses on the prediction of nitrogen oxides values in Madrid City as well as the technicalities of using both spatio-temporal kriging and functional kriging, and reports the main results obtained by both strategies. Finally, some concluding remarks are reported in Section 5.

2. Literature review

Due to the importance of air quality in general and especially the pollution caused to a great extent by road traffic, the literature related to the prediction of NO_x is certainly extensive. Without aiming to be exhaustive, Gaussian dispersion modeling is a widely used procedure for the cases where the cause-effect relationships between the source emissions and the levels measured at MS are understood and when information about the emission sources of air pollutants is available (Chelani & Hasan, 2001). De Haan (2009) and Skene et al. (2010) are two recent references worth recommending. However, in the absence of (or due to non-credible) emission inventory information, a frequent circumstance, Gaussian dispersion models are not suitable.

Time series models are another popular procedure for predicting the level of pollutants. Stieb, Judek, and Burnett (2002) conducted a meta-analysis of time-series studies (more than one hundred) of air pollution and mortality placing special emphasis on particulate matter (PM_{10}), carbon monoxide (CO), NO_2, O_3 and sulphur dioxide (SO_2). Two references worth mentioning are Routledge and Ayres (2005) and the references therein and Gilge et al. (2010).

Due to the nonlinear behavior and complexity involved in air pollutant concentrations and that air dynamics encompasses multiple seasonality, long memory and heteroskedasticity, neural networks and the chaos theory are also proposed as efficient tools to capture the behavior of the time series and predict future concentrations. Boznar, Lesjak, and Mlakar (1993) and Gardner and Dorling (1999) are early works using neural networks. Other more recent works focusing on predicting nitrogen oxides in large cities include Dorling, Foxall, Mandic, and Cawley (2003), Kukkonen et al. (2003), Zhang and San (2004) and Brunelli, Piazza, Pignato, Sorbello, and Vitabile (2007). Recently, research using neural networks has turned to specifically predicting vehicle emissions of NO_x (see Kim & Lee, 2010; Manjunatha, Narayama, & Reddy, 2010; Obodeh & Ajuwa, 2009; Vanderhaegen, Deneve, Laget, Faniel, & Mertens, 2010, is recommended literature for NO_x emissions from power plants). However, the neural network models currently available are not applicable for predicting spatial concentration distributions in urban areas. As for the chaos theory, some of the first valuable contributions are from the early nineties (Abarbanel, Brown, Sidorowich, & Tsimring, 1993; Grassberger, Schreiber, & Schaffrath, 1991). Other more recent and interesting research on the topic includes Chelani, Singh, and Devotta (2005), who examine the predictability of chaotic time series of NO_2 concentrations in Kolkata, India, and Jiménez-Hornero, Gutiérrez de Ravé, Ariza-Villarverde, and Giráldez (2009), who focus on the effects of nitrogen dioxide on the seasonal ozone pattern in Cordoba, Spain.

None of the above mentioned research takes into account the spatial or spatio-temporal dependencies of the level of pollutants. Recent literature that considers the spatial argument in the prediction of nitrogen oxides includes the following papers: de Kassteele and Stein (2006) presented a method that combines external drift kriging and a measurement error model and uses Bayesian techniques for inference. They address NO_2 data collected at an urban and a rural site in the Netherlands. de Kassteele, Stein, Dekkers, and Velders (2007) used kriging with external drift including secondary information from a dispersion model to predict atmospheric NO_x concentrations at rural and urban level. Yanosky, Schwartz, and Suh (2008) used census block-group-specific predicted outdoor nitrogen dioxide (a marker of traffic pollution) levels and four census block group socioeconomic position measures to evaluate whether chronic exposures to traffic-related air pollutants were higher in areas with lower socioeconomic position measures, after controlling for spatial autocorrelation in mixed models. NO_2 levels were predicted using a GIS-based spatio-temporal model that was validated with measured NO_2 concentrations. Zarandi, Khajevandi, Damez-Fontaine, and Ardestani (2008) used kriging to draw air pollution maps related to NO_2 and SO_2 measurements in Shiraz city. Janssen, Dumont, Fierens, and Mensink (2008) developed the statistical air pollution interpolation model RIO, an interpolation strategy that can be classified as a de-trended kriging model. They applied their novel method to O_3, NO_2 and PM_{10}.

Research taking into account not only space but also time and space–time interaction when predicting the level of nitrogen dioxides in large cities includes Ferreira, Tente, Torres, Cardoso, and Palma-Oliveira (2000), who presented a methodology that integrates the time–space framework of air quality data to infer the temporal pattern and spatial variability that could be interpreted for environmental decision purposes. Their study focuses on the central part of Lisbon and could be classified as a receptor modeling strategy to identify the type of vehicles responsible for certain pollutant levels, particularly for nitrogen oxides. Dubois, Saisana, Chaloulakou, and Spyrellis (2002) stated that spatial correlation may not only change in time and space but also according to pollution levels. To illustrate this point, they investigate the fractal dimension of the spatial correlation of different levels of annual nitrogen dioxide concentrations in the greater area of Milan, Italy, for the period 1997–1999. Mendes and Turkman (2002) studied the spatio-temporal behavior of NO_2 in the city of Lisbon, the objective being to make short-term daily predictions using a simple procedure based on time series models and kriging methods. De Cesare, Myers, and Posa (2001) and De Iaco, Myers, and Posa (2001a, 2001b, 2002a, 2002b, 2003) developed one of the pioneer works on non-separable spatio-temporal covariograms, the product-sum covariance for space–time modeling. They applied their statistical novelties by using, among others, data on nitrogen dioxide in the Milan district. In the last few years there has been extensive research on spatio-temporal covariance models, but applications to air quality are scarce (two interesting applications to predicting levels of ozone, of which NO_x is a forerunner, are Fuentes, Chen, and Davis (2008) and Reich, Eidsvik, Guindani, Nail, and Schmidt (2010)).

Bayesian hierarchical models (BHM) are a good option when complicated spatio-temporal interactions at different stages of the hierarchy are taken into account. One drawback of this procedure is that implementation could become unfeasible when the number of MS is large and high resolution of predictions is needed. Carslaw, Beevers, and Bell (2007) used this to predict future hourly NO_2 concentrations and the probability of exceeding the EU hourly limit value for NO_2. Yong, Huaicheng, Gouzhu, and Pingjian (2008) applied a BHM to predict urban air quality (SO_2, NO_x and dust fall) to support air quality management under uncertainty. They used the information provided by Huang (1992) for Xiamen City, China. Lapolla (2009) reviewed the main spatial techniques for predicting air pollution and focused on the case of nitrogen oxides in Tuscany, Italy. She dealt with both the spatial bivariate relationship between nitrogen monoxide and nitrogen dioxide and the spatio-temporal evolution of nitrogen oxides as a whole and proposed a dynamical BHM to estimate concentrations at unobserved points in space and time.

Finally, Lee et al. (2011) recently proposed land use regression modeling to examine the distribution of ambient concentrations of NO_2 in Incheon, Korea.

In Spain, the only research focusing on predicting the level of nitrogen dioxides is, as far as we know, Lertxundi-Manterola and Saez (2009), Salcedo-Sanz et al. (2009), Montero and Fernández-Avilés (2009) and Montero, Garcia-Centeno, and Fernandez-Avilés (2011). Lertxundi-Manterola and Saez (2009) used kriging to model the spatial distribution of NO2 and PM10 in Barcelona and Bilbao. Salcedo-Sanz et al. (2009) discussed the performance of Radial Basis Function (RBF) networks in a problem of spatial regression of pollutants in Madrid. Montero and Fernández-Avilés (2009) proposed a novel air pollution index based on six pollutants and computed it for Madrid. And Montero et al. (2011) presented a novel threshold autoregressive asymmetric stochastic volatility strategy to alert an immediate violation of the legal standard of specific pollutants in Madrid.

In summary, on the one hand, when it comes to predicting the value of a particular pollutant, recent research includes spatial dependencies (and is starting to include spatio-temporal dependencies) as a core element of the prediction model. But considering spatio-temporal dependencies for predicting purposes is not an easy task in either theoretical or empirical and computational terms. That is why the literature using spatio-temporal dependencies is scarce and, in addition, sometimes based on unreliable spatio-temporal covariance functions (frequently separable spatio-temporal covariance functions or functions that have been chosen only because they are implemented in well-known packages). On the other hand, functional kriging is a promising alternative to spatio-temporal kriging with less computational requirements. Moreover, as far as we know, no research has used a functional data approach to predicting the level of pollution in general and of NO_x in particular.

3. Methods

3.1. The spatio-temporal kriging approach

Spatio-temporal models arise when data are collected across time as well as space. Until recently, there has not been a theory of spatio-temporal processes other than the already well established theories of spatial statistics and time series analysis. For example, Cressie (1993) devoted only 4 pages of nearly 900 in his book to spatio-temporal models. However, research into spatio-temporal data has grown very rapidly over the last few years and the above mentioned four pages have been turned into a complete text (Cressie & Wikle, 2011).

Let $\{Z(\mathbf{s},t) : \mathbf{s} \in D \subset \mathbb{R}^d, t \in T \subset \mathbb{R}\}$ be a space–time process,[2] typically Gaussian where each of Z, D, and T is possibly random. Typically $Z(\mathbf{s}, t)$ is decomposed into the sum of two components: one deterministic, $\mu(\mathbf{s}, t)$ and the other stochastic, $e(\mathbf{s}, t)$, which models the space–time fluctuations around $\mu(\mathbf{s}, t)$. A limited number of spatially and temporally dispersed observations are a realization of such a process and, on the basis of such a realization, predictions are made at non-observed spatio-temporal points.

For that purpose, we use ordinary spatio-temporal kriging (OSTK) because it is a quite general framework for environmental data; more specifically, we assume the spatio-temporal process to be second-order stationary; that is, (i) the mean of the process is constant (null in our case) for all $\mathbf{s} \in D,\ t \in T$ and (ii) the covariance function, $C(Z(\mathbf{s}+\mathbf{h}, t+\tau) - Z(\mathbf{s}, t))$ is a function only of the spatio-temporal distance $(\mathbf{h}, \tau)$ between the $(\mathbf{s}, t)$ and $(\mathbf{s}+\mathbf{h}, t+\tau)$ points (see Cressie and Wikle (2011) for other types of kriging). This means that OSTK is applied to the stochastic component of $Z(\mathbf{s}, t)$ for which the estimation of the spatio-temporal trend, $\mu(\mathbf{s}, t)$, is a prerequisite. The first core decision to be made in the prediction process is how to perform the spatio temporal de-trending because (i) the spatio-temporal structure of residuals depends on it and (ii) as trend values are model-specific outcomes resulting from the particular algorithm used to determine them, the corresponding residual model is a collective term for all components of variability not included in the trend model and no elusive "true" residual covariance exists. Instead, one should acknowledge the $Z(\mathbf{s}, t)$ modeling decision and accept its consequences, i.e., the specific trend and residual estimates. A review of trend formulations can be found in Dimitrakopoulos and Luo (1997), Carrol et al. (1997) and Kyriakidis and Journel (1999). We have used a spatially extended seasonal-trend decomposition procedure based on local regression (STL) to estimate the trend component due to the marked annual NO_x cycle and its changing over time seasonal

[2] We assume that the observations are not "noisy" versions of true data at the observed points; that is, we do not include an additive error measurement term in the model. Otherwise, see Cressie and Wikle (2011, pp.145–148).

pattern. The STL procedure is a non-parametric regression technique introduced by Cleveland, Cleveland, and Terpenning in 1990, which applies weighted regressions based on neighborhood observations, allows for smooth changes in the seasonal pattern and produces estimates that are not affected by extreme values.

After obtaining the de-trended values, the spatial ordinary kriging predictor is extended to the spatio-temporal case and the spatio-temporal BLUP of the de-trended process, $\widehat{Z}(\mathbf{s},t)$, at the unobserved point $(\mathbf{s}_0,t_0)$, is obtained as a weighted mean of the de-trended values at the neighboring spatio-temporal points:

$$\widehat{Z}^*(\mathbf{s}_0,t_0)=\sum_{i=1}^{n}\lambda_i\widehat{Z}(\mathbf{s}_i,t_i),\quad \lambda_1,\lambda_2,\ldots,\lambda_n,\in\mathbb{R} \tag{1}$$

where the weights are obtained by minimizing the function $V(\widehat{Z}^*(\mathbf{s}_0,t_0)-\widehat{Z}(\mathbf{s}_0,t_0))$ subject to the unbiasedness constraint, $\sum_{i=1}^{n}\lambda_i=1$ (see Sherman (2010) for details). The result of that optimization problem, $\min\varphi(\lambda_i,\alpha)=\min 0.5V[\widehat{Z}^*(\mathbf{s}_0,t_0)-\widehat{Z}(\mathbf{s}_0,t_0)]-\alpha(\sum_{i=1}^{n}\lambda_i-1)$, yields the OSTK equations, which can be expressed in terms of the semivariogram, γ, of the (de-trended) process[3]:

$$\begin{cases}\displaystyle\sum_{i=1}^{n}\lambda_i\gamma(\mathbf{s}_i-\mathbf{s}_j;t_i-t_j)+\alpha=\gamma(\mathbf{s}_i-\mathbf{s}_0;t_i-t_0),\quad \forall i=1,\ldots,n\\ \displaystyle\sum_{i=1}^{n}\lambda_i=1\end{cases} \tag{2}$$

From (2) it can be deduced that the space–time weights $\lambda_1,\lambda_2,\ldots,\lambda_n$ depend on: (i) the spatio-temporal distance between the values of the de-trended process at the observed locations, (ii) the spatio-temporal distance between the values of the de-trended process at the observed and predicted locations and (iii) the structure of the spatio-temporal dependencies. The latter is represented by the covariance function or the semivariogram. The function $(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2)\mapsto C_{st}(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2)$ defined on the product space $\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}\times\mathbb{R}$ is called the spatio-temporal covariance function of the process and, if no further assumptions are made, depends on the space–time coordinates $(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2)$. As is well known, C_{st} is a covariance function if and only if $\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_j C_{st}((\mathbf{s}_i,t_i),(\mathbf{s}_j,t_j))\geqslant 0$ for all finite sets of real coefficients a_i and points $(\mathbf{s}_i,t_i)\in\mathbb{R}^d\times\mathbb{R}, i=1,\ldots,n$. This property is called positive definiteness. In our case, we have made the assumption of stationarity so that the covariance function does not depend on the space–time coordinates but on the spatial and temporal lags, $\mathbf{h}=\mathbf{s}_j-\mathbf{s}_i$ and $\tau=t_2-t_1$, respectively. The stationary semivariogram of the de-trended process is defined as $\gamma_{\widehat{Z}}(\mathbf{h},\tau)=\frac{1}{2}\mathrm{var}(\widehat{Z}(\mathbf{s}+\mathbf{h};t+\tau)-\widehat{Z}(\mathbf{s};t))$ and must satisfy the conditional non-positive definiteness condition, $\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_j\gamma_{st}(\mathbf{h},\tau)\leqslant 0$, for any set of points $(\mathbf{s}_i,t_i)\in\mathbb{R}^d\times\mathbb{R}, i=1,\ldots,n$ and all finite sets of real coefficients a_i satisfying $\sum_{i=1}^{n}\alpha_i=0$.

Substituting $\sum_{i=1}^{n}\lambda_iC(\mathbf{s}_i-\mathbf{s}_j;t_i-t_j)$ with $C(\mathbf{s}_i-\mathbf{s}_0;t_i-t_0)+\alpha$ in $V(\widehat{Z}^*(\mathbf{s}_0,t_0)-\widehat{Z}(\mathbf{s}_0,t_0))$, subject to the unbiasedness constraint, the prediction error variance is easily obtained as:

$$\begin{aligned}V[Z^*(\mathbf{s}_0,t_0)-Z(\mathbf{s}_0,t_0)]&=C(\mathbf{0},0)-\sum_{i=1}^{n}\lambda_iC(\mathbf{s}_i-\mathbf{s}_0;t_i-t_0)+\alpha\\&=\sum_{i=1}^{n}\lambda_i\gamma(\mathbf{s}_i-\mathbf{s}_0;t_i-t_0)+\alpha\end{aligned}$$

The Maximum Likelihood (ML) estimation of the parameters of an authorized covariance function could become computationally unfeasible when the number of MS is large and the number of instants of time considered is relatively high. Several solutions to this problem have been given in the literature: (i) exploiting the special structure of the covariance matrix in order to enhance computation (Kaufman, Schervish, & Nychka, 2007, used reduced versions of the covariance matrix through compactly supported covariance matrices; Fuentes (2007) proposed a procedure based on the spectral representation of the process, in the context of lattice data); and (ii) using likelihood approaches such as the different types of composite likelihood (CL) proposed by Vecchia (1988), Stein (2004), and Careaga and Smith (2006), among others. In this paper, we use a weighted composite likelihood (WCL) approach (a modification of the CL method conceived by Bevilacqua (2008)), which obtains accurate estimates comparable to those obtained by ML, but with lower computation costs.

Taking all possible differences $U_{ij}=Z(\mathbf{s}_i,t_i)-Z(\mathbf{s}_j,t_j),\quad i\neq j$, as a starting point, where $U_{ij}\sim N(0,2\gamma_{ij}(\boldsymbol{\theta}))$, $\boldsymbol{\theta}$ being the vector of parameters of the semivariogram model, the WCL method is composed of the following steps:

(i) Computation of the negative log-likelihood of one of such differences, eliminating all the constants that do not depend on $\boldsymbol{\theta}$:

$$-l(U_{ij};\boldsymbol{\theta})=\ln\gamma_{ij}(\boldsymbol{\theta})+\frac{U_{ij}^2}{2\gamma_{ij}(\boldsymbol{\theta})} \tag{3}$$

which is not but a classical support function with changed sign.

(ii) Computation of CL, extending the Curriero and Lele (1999) approach to the spatio-temporal context:

$$CL(\boldsymbol{\theta})=-\sum_{i=1}^{N}\sum_{j>i}^{N}l(U_{ij};\boldsymbol{\theta}) \tag{4}$$

where it is assumed that the differences U_{ij} are mutually independent, which need not to be the case in practice.

(iii) Maximization of WCL:

$$WCL(\boldsymbol{\theta},\mathbf{d})=-\frac{1}{W_{N,\mathbf{d}}}\sum_{i=1}^{N}\sum_{j>i}^{N}l(U_{ij};\boldsymbol{\theta})w_{ij}(\mathbf{d}) \tag{5}$$

where $w_{ij}(\mathbf{d})$ are non-negative weights which depend on the distance s_i-s_j, t_i-t_j and $w_{N,\mathbf{d}}=\sum_{i=1}^{N}\sum_{j>i}^{N}w_{ij}(\mathbf{d})$. Note that the weights $w_{ij}(\mathbf{d})$ do not depend on $\boldsymbol{\theta}$ to preserve the unbiasedness of the associated estimation function:

$$ee_{WCL(\boldsymbol{\theta},\mathbf{d})}=-\frac{1}{W_{N,\mathbf{d}}}\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{\gamma_{ij}^{(1)}(\boldsymbol{\theta})}{\gamma_{ij}(\boldsymbol{\theta})}\left(1-\frac{U_{ij}^2}{2\gamma_{ij}(\boldsymbol{\theta})}\right)w_{ij}(\mathbf{d}) \tag{6}$$

where $\gamma_{ij}^{(1)}(\boldsymbol{\theta})=\frac{d\gamma_{ij}(\boldsymbol{\theta})}{d\boldsymbol{\theta}}$.

More specifically, $w_{ij}(\mathbf{d})$ are cut-off weights, that is $w_{ij}(\mathbf{d})=1$ if $\|\mathbf{s}_i-\mathbf{s}_j\|\leqslant d_s$, $|t_i-t_j|\leqslant d_t$ and 0 otherwise, d_s and d_t being spatial and temporal distances, respectively, selected accordingly for the chosen purpose. Hence, the WCL estimator depends on $\mathbf{d}=(d_s,d_t)'$ through the weights w_{ij}. Note that instead of looking for an optimal linear combination of weights, $V_{ij}(\boldsymbol{\theta})=1-\frac{U_{ij}^2}{2\gamma_{ij}(\boldsymbol{\theta})}$, we look for an optimal distance, $\mathbf{d}=(d_s,d_t)'$.

(ii) Selection of $\mathbf{d}=(d_s,d_t)'$.

For this purpose, we use the Godambe information matrix, $G_N(\boldsymbol{\theta},\mathbf{d})$ as a measure of efficiency:

$$G_N(\boldsymbol{\theta},\mathbf{d})=\mathbf{H}_N(\boldsymbol{\theta},\mathbf{d})\mathbf{J}_N(\boldsymbol{\theta},\mathbf{d})^{-1}\mathbf{H}_N'(\boldsymbol{\theta},\mathbf{d}) \tag{7}$$

where[4]

[3] Taking into account that $C(\mathbf{h},\tau)=C(\mathbf{0},0)-\gamma(\mathbf{h},\tau)$ where $\mathbf{h}=\mathbf{s}_j-\mathbf{s}_i$ and $\tau=t_2-t_1$., Eq. (2) can be easily rewritten in terms of the covariance function.

[4] (1) and (2) denoting the first and second partial derivatives with respect to θ.

$$\mathbf{H}_N(\theta,\mathbf{d}) = -E[ee^{(1)}_{WCL(\theta,\mathbf{d})}] = -E\left[-\frac{1}{\mathbf{W}_N}\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{\gamma^{(1)}_{ij}(\theta)}{\gamma_{ij}(\theta)}\left(1-\frac{U^2_{ij}}{2\gamma_{ij}(\theta)}\right)w_{ij}(\mathbf{d})\right]^{(1)} = \frac{1}{\mathbf{W}_N}\sum_{i=1}^{N}\sum_{j>i}^{N}\left[\frac{\gamma^{(2)}_{ij}(\theta)\gamma_{ij}(\theta)-\gamma^{(1)}_{ij}(\theta)\gamma^{(1)}_{ij}(\theta)}{\gamma^{(2)}_{ij}(\theta)}\left(1-\frac{U^2_{ij}}{2\gamma_{ij}(\theta)}\right) + \frac{\gamma^{(1)}_{ij}(\theta)}{\gamma_{ij}(\theta)}\frac{2\gamma^{(1)}_{ij}(\theta)'U^2_{ij}}{4\gamma_{ij}(\theta)}\right]w_{ij}(\mathbf{d}) = \frac{1}{\mathbf{W}_N}\sum_{i=1}^{N}\sum_{j>i}^{N}\left[\frac{\gamma^{(1)}_{ij}(\theta)\gamma^{(1)}_{ij}(\theta)'U^2_{ij}}{\gamma^{(2)}_{ij}(\theta)}\right]w_{ij}(\mathbf{d})$$

because if the random function is Gaussian, $E[U^2_{ij}] = 2\gamma_{ij}(\theta)$ and

$$\mathbf{J}_N(\theta,\mathbf{d}) = E[ee_{WCL(\theta,\mathbf{d})}ee_{WCL(\theta,\mathbf{d})'}] = \frac{2}{\mathbf{W}^2_N}\sum_{i=1}^{N}\sum_{j>i}^{N}\sum_{l}^{N}\sum_{k>l}^{N}\left[\frac{\gamma^{(1)}_{ij}(\theta)\gamma^{(1)}_{ij}(\theta)'}{\gamma_{ij}(\theta)\gamma_{kl}(\theta)}\right]Corr(U^2_{ij},U^2_{ij})w_{ij}(\mathbf{d})w_{lk}(\mathbf{d})$$

Note that, if the random function is Gaussian, then:

$$Cov(U^2_{ij},U^2_{ij})w_{ij} = 2(\gamma_{il}(\theta)-\gamma_{jl}(\theta)+\gamma_{jk}(\theta)-\gamma_{ik}(\theta))$$

hence, it is easy to evaluate $\mathbf{J}_N(\boldsymbol{\theta},\mathbf{d})$.

The inverse of $G_N(\boldsymbol{\theta},\mathbf{d})$ is an approximation of the asymptotic variance of the WCL estimator. Hence, in order to enhance statistical efficiency, it seems appropriate to choose the spatio-temporal "lag" $\mathbf{d}$ minimizing the inverse of $G_N(\boldsymbol{\theta},\mathbf{d})$ in the partial order of non-negative definite matrices, or by using an equivalent scalar result (Heyde, 1997), we seek the optimum $\mathbf{d}^*$ such that:

$$\mathbf{d}^* = \arg\min_{\mathbf{d}\in D} tr(\mathbf{G}^{-1}(\theta,\mathbf{d})) \tag{8}$$

where D is a set of lags.

As can be seen, the WCL approach is just a way to drop from the likelihood pairs that are too far away (when far away is specified via the terms in $\mathbf{d}$, i.e., a spatial and a temporal cut-off distance), the objective being to simultaneously estimate the parameters of the variogram model, $\boldsymbol{\theta}$, and the distance cut-offs, $\mathbf{d}$.

3.2. Functional kriging approach

Functional data analysis (FDA) deals with the modeling of random variables that take values in the space of functions. In FDA the basic unit of information is a function instead a series of values. Let $x_i(t)$, $i = 1, 2, \ldots, N$ and $t \in T$ be the ith observation (a real valued function), where T is a real interval ($T \subseteq R^2$ in the case of images) and every x_i is a point in some functions space H (Ramsay & Dalzell, 1991). In the real scenario, an observation can be expressed by a random family $\{X(t_j)\}, j = 1, \ldots, J$. In contrast, in functional analysis, data can be considered as an observation of the continuous family $\boldsymbol{\chi} = \{X(t)\}t \in [t_{\min}; t_{\max}]$ (Ferraty & Vieu, 2006).

A functional variable $\boldsymbol{\chi}$ takes values in an infinite dimensional space (or functional space). An observation χ of $\boldsymbol{\chi}$ is called functional data. A functional data set is given by $\{\chi_1, \chi_2, \ldots, \chi_n\}$, that is to say, an observation of the set of random and equally distributed functional variables $\{\boldsymbol{\chi}_1, \boldsymbol{\chi}_2, \ldots, \boldsymbol{\chi}_n\}$. We go from data to functions by using basis functions (we use B-splines).

Functional kriging (FK) aims to predict a curve at a non-observed location on the basis of the observed curves and taking advantage of the existing spatial correlation among the curves at the observed points.[5] FK is based on the concept of a family of curves, here temporal profiles at monitoring stations. The goal in space–time prediction can be restated as that of reconstructing or predicting the unknown time series at a non-sampled location. The idea behind FK is to predict this unknown profile as a weighted linear combination of observed time profiles. Thus, all the values of a particular time series are multiplied by the same weight, hence the collapse of a space–time prediction problem to a space-only problem. The closer the non-sample location lies to a monitoring station, the more the predicted profile will resemble the profile at the monitoring station.

Adapting the core kriging concepts to the functional case, let $\boldsymbol{\chi}(\mathbf{s}) : s \in D \subset \mathbb{R}^d$ be a second-order stationary random function, so that $\boldsymbol{\chi}_{\mathbf{s}_i}$ is the functional variable corresponding to location $\mathbf{s}_i$ and $\{\chi_{\mathbf{s}_1}, \chi_{\mathbf{s}_2}, \ldots, \chi_{\mathbf{s}_n}\}$ is a realization for the random function at locations $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$. Since $\boldsymbol{\chi}(\mathbf{s}) : s \in D \subset \mathbb{R}^d$ is a second order stationary random function, it can be stated that:

$$\begin{aligned}
&E(\boldsymbol{\chi}_{\mathbf{s}}(t)) = m(t) \quad \forall t \in T, \mathbf{s} \in D\\
&V(\boldsymbol{\chi}_{\mathbf{s}}(t)) = \sigma^2(t) \quad \forall t \in T, \mathbf{s} \in D\\
&C(\boldsymbol{\chi}_{\mathbf{s}_i}(t), \boldsymbol{\chi}_{\mathbf{s}_j}(t)) = C(\mathbf{h}, t) \quad \forall t \in T, \mathbf{s}_i, \mathbf{s}_j \in D, \quad \text{with } \mathbf{h} = \mathbf{s}_i - \mathbf{s}_j\\
&\frac{1}{2}V(\boldsymbol{\chi}_{\mathbf{s}_i}(t), \boldsymbol{\chi}_{\mathbf{s}_j}(t)) = \gamma(\mathbf{h}, t) \quad \forall t \in T, \mathbf{s}_i, \mathbf{s}_j \in D, \quad \text{with } \mathbf{h} = \mathbf{s}_i - \mathbf{s}_j
\end{aligned} \tag{9}$$

Following standard ordinary kriging rationale, the functional BLUP at the unobserved site $\mathbf{s}_0$ will be given by

$$\tilde{\boldsymbol{\chi}}_{\mathbf{s}_0}(t) = \sum_{i=1}^{n}\lambda_i\boldsymbol{\chi}_{\mathbf{s}_i}(t) \quad \text{where } \lambda_1, \lambda_2, \ldots, \lambda_n, \in \mathbb{R} \tag{10}$$

The n weights involved in the BLUP are obtained, in the ordinary functional kriging (OFK) procedure, as the solution of the following optimization problem:

$$\min_{\lambda_1,\lambda_2,\ldots,\lambda_n}\int_T V(\tilde{\boldsymbol{\chi}}_{\mathbf{s}_0}(t) - \boldsymbol{\chi}_{\mathbf{s}_0}(t))dt \quad \text{s.t.} \sum_{i=1}^{n}\lambda_i = 1 \tag{11}$$

where $\sum_{i=1}^{n}\lambda_i = 1$ is the well known unbiasedness constraint.

The objective function derived from the above expression can be rewritten as (see Fernández-Avilés, 2010; Giraldo, 2009):

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\int_T C_{ij}(t)dt + -2\sum_{i=1}^{n}\int_T C_{i0}(t)dt + 2\theta\left(\sum_{i=1}^{n}\lambda_i - 1\right) \tag{12}$$

where θ is the Lagrange multiplier used to take into account the unbiasedness constraint, $\int_T C_{ij}(t)dt$ is the value of the chosen valid covariance function (Gaussian, spherical, exponential, etc.) for the distance corresponding to the spatial separation between the observed curves $\chi_{\mathbf{s}_i}$ and $\chi_{\mathbf{s}_j}$, and $\int_T C_{i0}(t)dt$ the value of such a valid covariance model for the spatial separation between the locations of the observed curve $\chi_{\mathbf{s}_i}$ and the predicted one $\chi_{\mathbf{s}_0}$.

Minimizing (12) with respect to $\lambda_1, \lambda_2, \ldots, \lambda_n$ and θ we obtain the following system of equations (the proof can be found in Delicado, Giraldo, Comas, & Mateu, 2010):

$$\begin{cases}\sum_{j=1}^{n}\lambda_j\int_T C_{ij}(t)dt + \mu = \int_T C_{i0}(t)dt, \quad i = 1, 2, \ldots, n\\ \sum_{i=1}^{n}\lambda_i = 1\end{cases} \tag{13}$$

Or, in terms of the semivariogram,

$$\begin{cases}\sum_{j=1}^{n}\lambda_j\int_T \gamma_{ij}(t)dt - \mu = \int_T \gamma_{i0}(t)dt, \quad i = 1, 2, \ldots, n\\ \sum_{i=1}^{n}\lambda_i = 1\end{cases} \tag{14}$$

[5] Similar approaches have been proposed in the literature, not in the context of non-parametric curves but in the form of parametric temporal profiles whose parameters are estimated in space (Kyriakidis and Journel (2001) is a good example). In this case, the family of curves is described by the spatial variability of the parameters, yet the form of the curves remains the same throughout the study domain.

where $\int_T \gamma_{ij}(t)dt$ and $\int_T \gamma_{i0}(t)dt$ represent the value of a valid semivariogram model for the distance between the observed curves $\chi_{\mathbf{s}_i}$ and $\chi_{\mathbf{s}_j}$, and between the observed curve at $\mathbf{s}_i$, $\chi_{\mathbf{s}_i}$, and the one to be predicted at $\mathbf{s}_0$, $\chi_{\mathbf{s}_0}$.

Then, the prediction trace-variance based on the trace-semivariogram is given by

$$\sigma^2_{\mathbf{s}_0} = \sum_{i=1}^{n} \lambda_i \int_T \gamma_{i,o}(t)dt - \theta \tag{15}$$

which could be interpreted as a measure of the global uncertainty on the prediction.

The trace-semivariogram estimator is the adaptation to the functional case of the classical Method of Moments estimator used in standard kriging:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{i,j \in N(\mathbf{h})} \int_T (\chi_{\mathbf{s}_i}(t) - \chi_{\mathbf{s}_j}(t))^2 dt \tag{16}$$

where $|N(\mathbf{h})|$ is the number of distinct observed sites at a distance h.

4. Case study: predicting the level of NO_x in Madrid City

4.1. NO_x in Madrid City: still an unresolved problem

Madrid (the capital of Spain) is the third most populous city in the European Union after London and Berlin (pop. 6,445,499 in 2010, 3,269,861 of which live in the city). Like other capitals in the world, Madrid is the city where Government institutions, the Parliament, embassies, main museums, central offices of the most relevant companies, etc. are located. This has made Madrid a large city covering 60430.76 ha, together with a large peripheral metropolitan area with more than five million inhabitants that it is closely related to. Obviously, these relations imply movement and a large number of trips and regular flows of both population and also goods, etc., which has led to a complex transportation system.

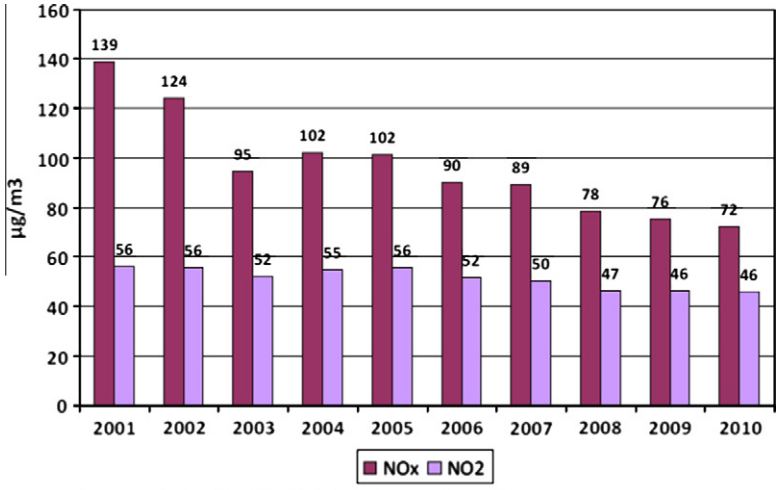
More specifically, Madrid has both a dense ring road network (M-30, M-40, M-45 and M-50) and a dense radial highway network. Both networks have enormously improved accessibility to emerging industrial and high economic activity areas, resulting in competitiveness and dynamism. It is therefore no surprise that the number of vehicles in Madrid has increased by 5.6% over the last decade, amounting in 2010 to a total of 1,917,382. This implies 1202.5 vehicles per km. and 586.4 vehicles per 1000 inhabitants. Two million drivers enter and leave the city on a daily basis. So, car pressure is increasing as well as its negative environmental impacts because road traffic is the main emission source of NO_2. In addition, Madrid has the fourth largest European airport and is the centre for train communications (half a thousand trains enter Madrid from the 10 most important Spanish cities, as well as from Paris and Lisbon). Freight transportation by train is also really important in Madrid. Every day 400 trains enter and leave the city, transporting 150,000 tons of commodities. In fact, Madrid has the largest inland maritime customs centre in Europe. However, as a negative consequence of the above positive factors, transport and specifically road traffic has become the main source of nitrogen dioxides.

According to the Department of Environmental Assessment, Control and Quality of Madrid, the mean values of NO_x and NO_2 in the city in 2010 were 72 and 46 μg/m^3, respectively. The evolution of both pollutants is reported in Fig. 1, which shows that the last decade has witnessed a relevant depletion in the concentrations of both pollutants, especially in the case of NO_x, the reason being the lesser contribution of diesel vehicles, which are predominant in the motor vehicle fleet in Madrid. It must be taken into account that, according to the European Environmental Agency, 34.1% of the total NO_x emissions in Madrid City are due to road transport, 17.2% to other transports, 22.1% to energy industries and 21.3% to industry – energy and processes (38.2%, 14.3%, 20.7% and 16.5%, respectively, in the EU).

However, despite the decreasing trend in NO_x levels, in 2010 the annual legal standard for the protection of human health (40 μg/m^3) was exceeded in most of the monitoring stations (MS). More specifically, all the urban traffic stations exceeded the annual limit and more than half of them also surpassed the limit of 18 h exceeding the hourly limit, which is 200 μg/m^3. Therefore, it is no surprise that in January 2010 the Madrid Council had created both an Air Quality Commission and an action protocol for NO_x pollution episodes which depends on the zone of the city (see zones in Fig. 2).

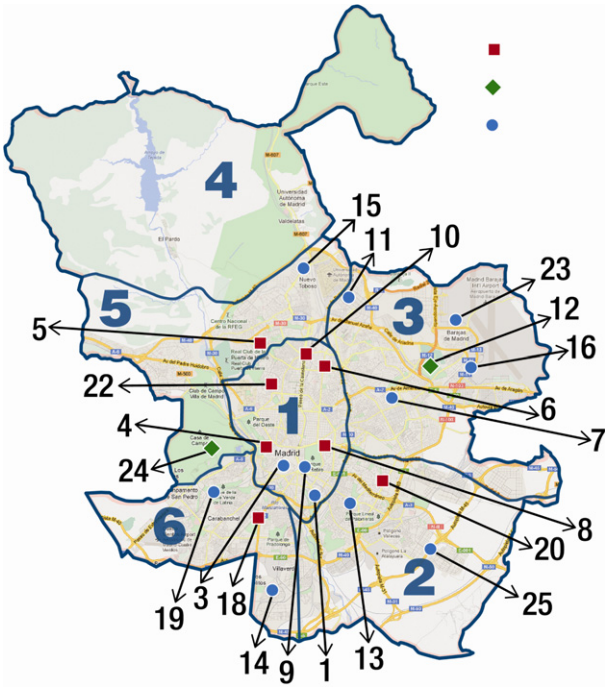
4.2. Database

The data used in this paper have been provided by the Atmosphere Pollution Monitoring System of Madrid. They represent hourly measurements at the 23 fixed operative MS during 2009. Fig. 2 shows the location of the air quality MS. Most of them are



Source: Air Quality (Madrid 2010 Annual Report). Department of Environmental Assessment, Control and Quality. Madrid Municipality.

Fig. 1. Annual mean level of NO_x and NO_2 in Madrid City.



Zone	Population	Area in km^2
1	1,071,003	41
2	593,498	120
3	604,034	100
4	13,498	170
5	300,544	80
6	534,702	89

(*) The monitoring station located in Zone 4 has not been used because it was installed in 2010.
(**) Traffic stations (square), peripheral (circle) and stations in large green areas (diamont).

Fig. 2. Location of the monitoring stations.

located in the urban center and relatively few in the peripheral areas. Note that MS cover the area under study reasonably well, since most of Madrid's population is concentrated in the urban center (zone 1) and zones 2, 3 and 6.

The initial database included hourly values of NO_x during 2009. However, in light of the main sources of NO_x emissions in Madrid City, it seems logical to believe that levels of NO_x will be different on week days and weekends and holidays. This is the reason why a "typical working day" was created for each week. That is, weekends and bank holidays were eliminated and the hourly average for working days was computed. As a result, the final database includes 28,704 observations, the result of considering 23 stations by 1248 time scores, corresponding to the 24 h of the 52 typical working days (one per week) at each station. In other words, the final data consisted of 1248 time measures at 23 spatial locations. As can be appreciated in Fig. 3, these typical days behave the same all year round (regardless of the station). Table 1 reports the main descriptive statistics for the MS.

In order to estimate whether or not the "working day effect" is significant, a dummy binary variable was created. The results showed that the "working day effect" is statistically significant and positive. More specifically, the average level of NO_x estimated on working days is 22.4% higher than at weekends and on bank holidays and this percentage is higher than in the case of other pollutants.

Fig. 4 includes a box-plot for each MS. At first glance, we can see the large amount of outliers there are in the stations, a circumstance that is nevertheless certainly common in data related to air pollution. The fact that there are a large number of extreme values indicates non-normality. Since normality is a frequent requirement when dealing with geostatistical procedures, we proceed to take the Neperian logarithm of the levels of NO_x. Comparing Figs. 4 and 5, it can be seen that the logarithmic transform has managed to eradicate the problem of non-normality. From this point onwards, all the analyses and studies will be performed with the transformed variable.

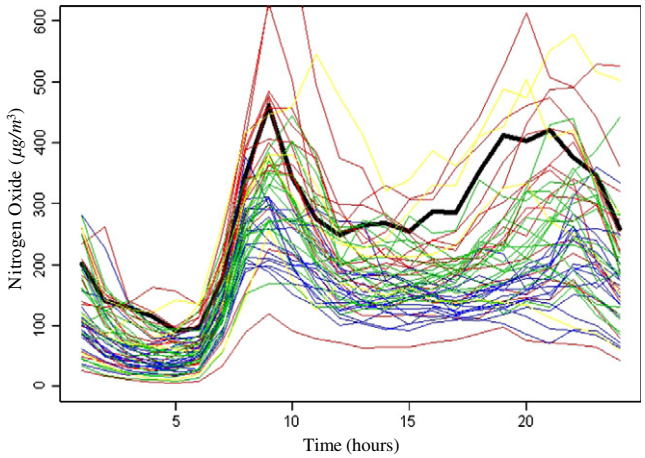


Fig. 3. Hourly NO_x levels for the "working days", 2009.

4.3. Results

4.3.1. OSTK approach

As said in the methodological section, a prerequisite for using OSTK is the estimation of the spatio-temporal trend, $\mu(\mathbf{s}, t)$, and the way of performing the spatio temporal de-trending is the first core decision in the prediction process. For this purpose, we used a spatially extended seasonal-trend decomposition based on local regression (STL) procedure. The reason is that the annual NO_x cycle is very important and its seasonal pattern is not stable, but rather changes over time (otherwise we could have used a Median Polish algorithm including a seasonal effect, a yearly effect and a spatial effect, as in Bruno, Guttorp, Sampson, and Cocchi (2009)).

The second core decision is the selection of the covariance model to represent the spatio-temporal dependencies of the de-trended NO_x series. For this purpose, three competing spatio-temporal

Table 1
Main descriptive statistics of NO_x.

MS	Mean	SD	Median	Trimmed mean	Min	Max	Range	Skew	Kurtosis
1	157.72	96.57	142.87	146.97	24.38	1429.57	1405.19	4.71	47.23
3	97.67	59.31	86.94	89.61	17.03	650.07	633.04	2.72	14.96
4	113.08	93.60	83.71	95.96	13.14	1037.53	1024.39	3.26	18.75
5	89.98	99.16	63.67	72.71	10.86	1511.67	1500.81	6.40	67.92
6	172.26	108.90	165.09	163.54	23.78	1700.67	1676.89	5.76	69.47
7	112.67	78.65	104.75	105.44	7.35	1170.83	1163.48	4.56	48.41
8	115.28	69.73	103.59	107.12	23.51	1030.27	1006.76	4.24	41.46
9	152.54	109.16	127.34	136.11	18.84	1524.00	1505.16	3.46	27.39
10	103.44	84.11	82.30	90.75	9.95	1004.57	994.62	3.70	26.68
11	93.03	76.86	83.28	84.91	9.18	1265.73	1256.55	7.07	89.62
12	94.53	55.92	86.41	88.25	13.53	652.03	638.50	2.24	12.31
13	70.61	51.40	54.56	62.26	8.95	372.22	363.27	1.91	4.80
14	119.70	82.45	98.81	106.60	6.68	800.17	793.49	2.03	6.68
15	117.66	61.62	108.27	112.05	15.60	405.18	389.58	1.00	1.41
16	65.80	50.42	49.37	57.60	10.20	521.57	511.37	2.61	12.43
18	74.23	65.16	53.31	62.30	11.32	794.90	783.58	3.13	18.55
19	115.56	85.89	90.53	101.33	9.57	719.23	709.66	2.08	6.37
20	82.82	57.67	65.48	73.89	13.16	586.07	572.91	2.38	9.65
21	114.88	72.46	102.24	105.04	15.63	534.53	518.90	1.64	3.89
22	116.18	84.32	92.09	102.24	11.52	710.00	698.48	1.99	5.71
23	108.95	87.29	87.10	95.39	15.04	1124.97	1109.93	4.37	35.82
24	43.69	39.83	28.20	36.24	4.13	265.63	261.50	2.16	5.95
25	128.42	89.01	105.80	116.11	13.75	810.03	796.28	1.57	4.07

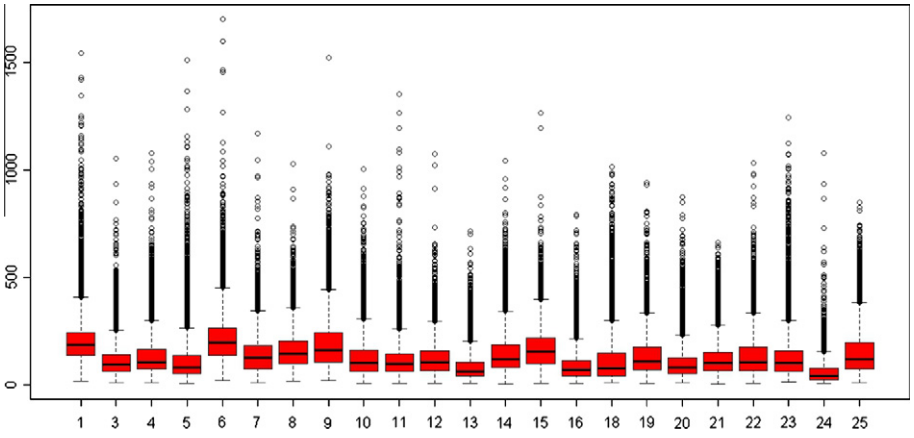


Fig. 4. Box-plots for NO_x by MS.

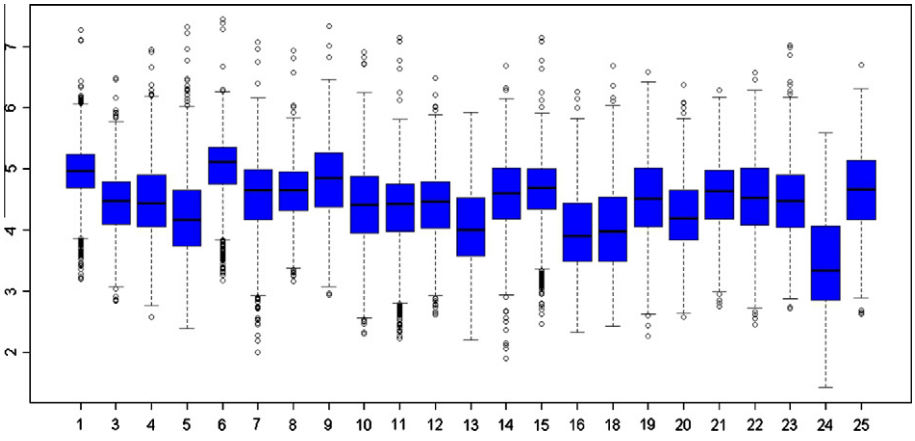


Fig. 5. Box-plots for $Ln(NO_x)$ by MS.

covariance functions are used. The first and second consider space and time separately and have been frequently used in traditional environmental research. The third one is an innovative spatio-temporal covariance function that takes into account the interaction between space and time and which, in our opinion, is highly suitable for most environmental processes. Specifically, they are[6]:

[6] We used proprietary codes of these covariance functions to solve OSTK equations.

Table 2
Empirical spatio-temporal covariance functions for NO_x.

	Separable exponential	Gneiting's separable	Gneiting's non-separable
a	0.00253164	0.00046506	0.00184651
c	0.27291130	0.29018950	0.05710549
δ	0.06301400	0.50246770	0.00038308
α	–	1.37393774	0.98775619
β	–	–	0.90866750
σ^2	0.00222210	0.00242954	0.00294644

(i) The separable exponential spatio-temporal model:

$$C(\mathbf{h},\tau,\boldsymbol{\theta}) = \exp(-c\mathbf{h} - a\tau) \tag{17}$$

(ii) The Gneiting separable spatio-temporal model (Gneiting, 2002):

$$C(\mathbf{h},\tau,\theta) = \frac{1}{(a\tau^{2\alpha}+1)}\exp(-c\mathbf{h}^{2\delta}) \tag{18}$$

(iii) The Gneiting non-separable spatio-temporal model (Gneiting, 2002):

$$C(\mathbf{h},\tau,\boldsymbol{\theta}) = \frac{1}{(a\tau^{2\alpha}+1)}\exp\left[\frac{-c\mathbf{h}^{2\delta}}{(a\tau^{2\alpha}+1)^{\beta\delta}}\right] \tag{19}$$

with a and c being two positive parameters of spatial and temporal scaling respectively and σ^2 the a priori variance of the stochastic process; δ and α are spatial and temporal smoothing parameters, also respectively, which take values in (0, 1]; σ^2 is the variance of the spatio-temporal process and β is a spatio-temporal interaction parameter with a variation range of [0, 1].

The spatio-temporal models (18) and (19) cover a large class of phenomena and if $\tau \geqslant d/2$, the indicating the dimension, a parametric family $C(\mathbf{h}, \tau/\beta)$ is obtained with an easily interpretable space–time interaction parameter β. The case $\beta = 0$ (model (18)) corresponds to a separable model, in which the spatial correlations for different values of the temporal lag τ are proportional to each other. As β increases (model (19), space–time interaction strengthens, and the correlations at nonzero temporal lags fall off less and less rapidly, as compared with the separable model.

The estimation of the parameters of the above models is the third core decision in the prediction process. Maximum Likelihood (ML) estimation could become computationally unfeasible when the number of MS is large and the number of instants of time considered is relatively high. To overcome this limitation, we use a WCL approach. By means of simulation, we have identified the vector of spatial and temporal optimal lags, $\mathbf{d} = (6781; 18)'$, where the spatial distance is expressed in meters and temporal distance in hours.

The estimation results for the parameters of the three competing covariance models are summarized in Table 2. The root mean square error (RMSE) resulting from the use of these covariance models when predicting 1-month ahead on the basis of January to November 2009 data (measured hourly) are reported in Table 3.

4.3.2. The OFK approach

An alternative to the computational problems of OSTK is OFK. For comparative purposes, we also deal with natural logarithms in the functional analysis. Fig. 6 (left panel) displays the level of $Ln(NO_x)$ for each monitoring station on the 1248 "working days" considered. It suggests a strong rate of seasonality, as well as a marked cycle-tendency component (as is logical where environmental variables are concerned). These temporal features make it difficult to find parametric models for this type of curve. As a result, we choose B-splines as basis functions to smooth the original series. The cross-validation process was used to find the optimum number of smoothing parameters: L, number of inner knots, and η, a parameter that controls the trade-trade off between the fit of the observed data and the smoothness of the approximating spline, also know as a roughness penalty. On the basis of a preliminary exploration of the data, two ranges have been chosen: 150, 151, 152, 153,..., 250, and 0, 1, 10, 102, 103 as possible values of L and η, respectively. The values $L^* = 248$ and $\eta^* = 0$ were found as optimal. Fig. 6 (mid panel) shows the set of smoothed functions.

Table 3
RMSE (Log NO_x).

Monitoring stations	Spatial[a]	Temporal[b]	OSTK (a)	OSTK (b)	OSTK (c)	OFK (d)
MS 1	0.437565	0.622622	0.239840	0.231035	0.209881	0.235384
MS 3	0.224087	0.791143	0.182182	0.146736	0.122645	0.345274
MS 4	0.625908	0.994427	0.213238	0.166578	0.144406	0.268091
MS 5	0.331003	0.930160	0.282285	0.250859	0.226962	0.074164
MS 6	0.519234	0.593884	0.232473	0.220146	0.196527	0.385527
MS 7	0.326360	0.634874	0.222233	0.207668	0.189268	0.126622
MS 8	0.216300	0.573495	0.197519	0.199618	0.173327	0.134463
MS 9	0.415965	0.720802	0.195162	0.197229	0.178020	0.380409
MS 10	0.258229	0.770273	0.221172	0.227886	0.213742	0.192049
MS 11	0.341530	0.678682	0.285792	0.278892	0.250933	0.221228
MS 12	0.267184	0.534176	0.194461	0.182709	0.154833	0.025193
MS 13	0.315748	0.631802	0.238563	0.208981	0.205851	0.406905
MS 14	0.232493	0.687727	0.222062	0.211603	0.196491	0.270184
MS 15	0.252818	0.630167	0.249976	0.272746	0.251266	0.115535
MS 16	0.242484	0.519707	0.263937	0.237157	0.210785	0.171820
MS 18	0.470051	0.963018	0.229991	0.210418	0.187051	0.416212
MS 19	0.316526	0.700358	0.207258	0.200165	0.187837	0.647428
MS 20	0.255132	0.650017	0.206842	0.195865	0.170598	0.149414
MS 21	0.256212	0.663647	0.183973	0.156441	0.150892	0.099878
MS 22	0.199899	0.747624	0.215089	0.207973	0.191073	0.111379
MS 23	0.240695	0.611168	0.271248	0.267526	0.233850	0.231858
MS 24	1.037745	1.011580	0.250621	0.220188	0.202067	0.900698
MS 25	0.401745	0.535201	0.254824	0.237020	0.233312	0.358169

[a] Ordinary kriging: Exponential semivariograms were used for 95% of the out-of-sample days.
[b] ARIMA modeling. (a) OSTK with spatio-temporal exponential covariance (b) OSTK with Gneiting separable spatio-temporal covariance (c) OSTK with Gneiting non-separable spatio-temporal covariance and (d) OFK.

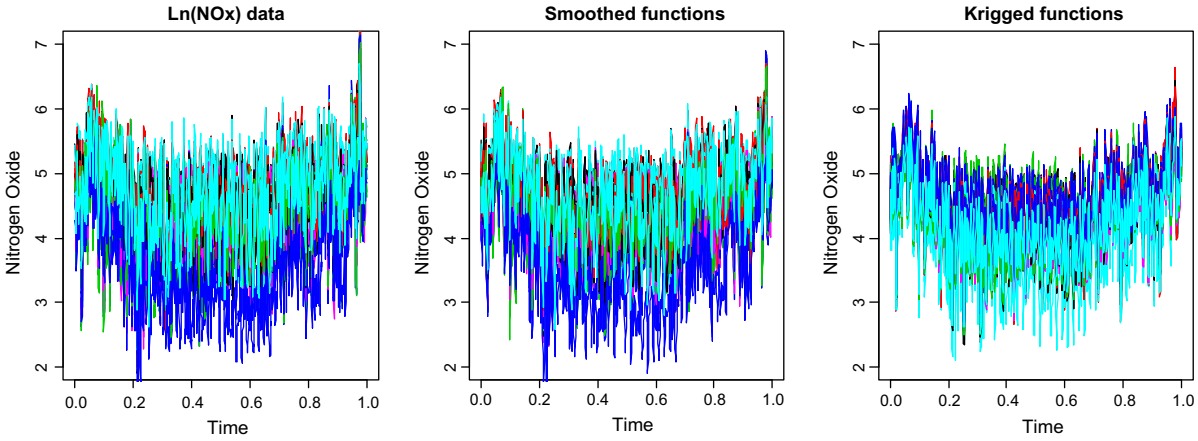


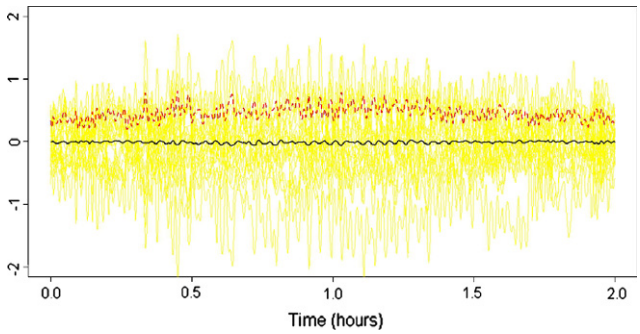
Fig. 6. Original, smoothed and kriged functions.

In light of the empirical semivariogram-trace of the smoothed functions, the spherical semivariogram model was used for prediction tasks. Fig. 6 (right panel) shows the predicted (kriged) curves. Logically, the predicted curves display less variability than the original curves due to both the smoothing process performed previously and the kriging process, which is in itself a smoothing method (leading to a decrease in variance).

As can be seen in Fig. 7, the mean of the residuals (the difference between the data, Fig. 6 (left panel), and the functionally kriged data, Fig. 6 (right panel)), is practically constant. The variance of the residuals is practically constant too.

As in the spatio-temporal case, in order to compare the residuals obtained from OSTK to those obtained with OFK, we obtained hourly predictions for December 2009 on the basis on the hourly measures from January to November. In order to assess and rank the point forecasts, we used RMSE.

For comparison purposes, we have also obtained merely temporal and spatial predictions. As can be seen in Table 3, the merely spatial predictions are clearly better than the purely temporal ones, which indicates the importance of taking into account spatial dependencies for prediction purposes. However, as expected, in all MS, OSTK strategies provide lesser RMSE than the merely spatial or temporal ones, which indicates the importance of considering both space and time in the analysis. More specifically, in 22 out of 23 MS, OSTK with the Gneiting non-separable covariance function provides better results than the two separable approaches. That is, when the interaction between space and time is incorporated into the prediction procedure, results are significantly better than when both arguments are considered separately. For the whole set of MS, the proportional reduction in RMSE obtained by using the non-separable strategy is 15.8% with respect to the separable one with an exponential covariance function and 4.9% when the separable spatio-temporal covariance function is a Gneiting function.



(*) Residual functions in yellow; mean and standard deviation of the residual functions in black and red, respectively.

Fig. 7. Residual functions: mean and standard deviation of the residual functions.

As for OFK, it generally performs worse than OSTK with the non-separable covariance function: the proportional reduction in RMSE obtained by using a non-separable strategy is 6.3% with respect to the OFK procedure. However, it is worth mentioning that OFK predictions are better than the predictions obtained using OSTK with the non-separable space–time Gneiting covariance function in the inner MS. In the peripheral MS, OSTK predictions are more accurate than OFK predictions. This is not surprising, because it must be taken into account that the functional kriging estimator provides weighted means of the neighboring curves as estimates. Obviously this feature makes OFK estimates strongly dependent on where monitoring stations are located and, as a consequence, in peripheral MS, OFK estimates tend to be near to the mean value and their variance is higher than desirable, which leads to large prediction errors in case that the curves in such peripheral locations are clearly below or above the mean curve. In our specific study case, peripheral MS 3, 9, 19 and 25 (all of them in the Southern part of the city, an industrial zone with some high traffic density areas) are, along with the inner MS 6, the most polluted MS in Madrid City. On the contrary, MS 5, 13, 16, 18 and especially 24—all of them sited in or near to green areas and with a low population/surface ratio—report the lowest values of NO_x in the city. This is the reason for which predictions near to the mean in these locations lead to large prediction errors.

Therefore, the OFK strategy not only has computational advantages over the OSTK procedure but provides better estimates at non-peripheral sites. However, in the peripheral MS, the joint consideration of space and time is central to overcoming the drawbacks of the OFK strategy. Finally, data variability does not seem to be a core aspect of deciding which of the two methods should be used.

5. Conclusion

Air pollution is at the top of the list of citizens' environmental concerns. This is particularly true in large cities where more than half the world's population lives. The link between air quality and human health worries many health experts, policy-makers and citizens. The World Health Organization states that almost 2.5 million people die each year from causes directly attributable to air pollution. Therefore, it is no surprise that studies about

pollution are becoming increasingly popular and that environmental issues have brought atmospheric science to the centre of science and technology, where it now plays a key role in shaping national and international policy.

In this paper we focus on NO_x, which in spite of the effort made in recent years is still one of the main pollution problems in large cities. In large metropolitan areas increasing transportation of goods and people means higher levels of NO_x and as road traffic is related to human activity and needs, much of it occurs in areas where people live, work, go to school, etc. The latter means that today's urban development will lead to NO_x being a more serious problem in the future unless efforts are made to mitigate it.

The literature related to the prediction of nitrogen oxides has been based on Gaussian dispersion modeling, time series models, neural networks and, to a lesser extent, the chaos theory, but in most of cases does not take into account the spatial or spatio-temporal dependencies of the level of pollutants. However, spatio-temporal dependencies are a core aspect of pollution and this is the reason why we propose both a spatio-temporal kriging and a functional kriging strategy to incorporate such dependencies into the prediction of NO_x levels.

More specifically, we have compared a OSTK proposal (with separable and non-separable covariance functions) to OFK in Madrid City. For this purpose we have constructed a massive hourly database of "working days". As expected, both spatio-temporal strategies, OSTK and OFK, have noticeable prediction gains with respect to the merely spatial (kriging) or temporal (ARIMA modeling) procedures. This first finding indicates the importance of considering both space and time for prediction purposes. The second finding we wish to highlight is that the joint consideration of space and time notably improves predictions. The third and most important finding is that the functional strategy yields better predictions than the spatio-temporal punctual one (with non-separable spatio-temporal covariance) at non-peripheral sites. This is a relevant finding because the computational difficulties involved in carrying out the ML estimation of the parameters of the covariance function makes punctual spatio-temporal strategies (and kriging in particular) unfeasible when the number of instants of time considered is relatively high (a solution to this computational burden is the innovative composite likelihood method that we use in this research). The fact that the OFK provides worse estimates than OSTK at peripheral MS cannot be considered a worrying problem, because it cannot be expected to obtain good predictions in a spatial context when the number of neighboring sites is scarce. In any case, in our particular study case, the poor performance of the functional procedure at peripheral sites is due to the fact that the lowest and highest values of NO_x are observed at those peripheral locations. Otherwise, the OFK performed better than the OSTK strategy.

Despite the advantages of the OFK strategy we have presented in this article, new functional kriging methods should be studied, this topic being a promising avenue of research. More specifically, we find it challenging (i) to give a functional form to the weights of the observed curves in the functional kriging estimator; (ii) to extend the OFK to a context including auxiliary or secondary information (that is, a functional cokriging approach); and (iii) to combine the functional ordinary cokriging with the complete functional linear model.

Acknowledgment

We would like to thank the anonymous reviewers for their useful, constructive and valuable comments, which have undoubtedly improved the original version of the manuscript.

References

Abarbanel, H. D. I., Brown, R., Sidorowich, J. J., & Tsimring, L. Sh. (1993). The analysis of observed chaotic data in physical systems. *Reviews of Modern Physics, 65*, 1331–1392.

Bevilacqua, M. (2008). *Estimating space and space–time covariance functions: a weighted composite likelihood approach*. Doctoral thesis. University of Padua.

Bilonick, R. A. (1985). The space–time distribution of sulfate deposition in the northeastern United States. *Atmospheric Environment, 19*(11), 1829–1845.

Boznar, M., Lesjak, M., & Mlakar, P. (1993). A neural network-based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain. *Atmospheric Environment, 27B*(2), 221–230.

Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., & Vitabile, S. (2007). Two-days ahead prediction of daily maximum concentrations of SO_2, O_3, PM_{10}, NO_2, NO_x in the urban area of Palermo, Italy. *Atmospheric Environment, 41*(14), 2967–2995.

Bruno, F., Guttorp, P., Sampson, S., & Cocchi, D. (2009). A simple non-separable, non-stationary spatiotemporal model for ozone. *Environmental and Ecological Statistics, 16*, 515–529.

Careaga, P., & Smith, R. (2006). *Approximate likelihoods for spatial processes*. Technical report. Department of Statistics, Iowa State University.

Carrol, R. J., Chen, R., George, E. I., Li, T. H., Newton, H., Schmiediche, H., et al. (1997). Ozone exposure and population density in Harris County, Texas. *Journal of the American Statistical Association, 92*, 392–404.

Carslaw, D. C., Beevers, S. D., & Bell, M. C. (2007). Risks of exceeding the hourly EU limit value for nitrogen dioxide resulting from increased road transport emissions of primary nitrogen dioxide. *Atmospheric Environment, 41*(10), 2073–2082.

Chelani, A. B., & Hasan, M. Z. (2001). Forecasting nitrogen dioxide concentration using artificial neural networks. *International Journal of Environmental Studies, 58*, 487–499.

Chelani, A. B., Singh, R. N., & Devotta, S. (2005). Nonlinear dynamical characterization and prediction of ambient nitrogen dioxide concentration. *Water, Air, & Soil Pollution, 166*, 121–138.

Cleveland, R. B., Cleveland, J. E., & Terpenning, I. (1990). STL: a seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics, 6*, 3–73.

Cressie, N. (1993). *Statistics for spatial data*. Wiley Interscience.

Cressie, N., & Wikle, C. (2011). *Statistics for spatio-temporal data*. Hoboken, NJ: Wiley.

Curriero, F. C., & Lele, S. (1999). A composite likelihood approach to semivariogram estimation. *Journal of Agricultural, Biological and Environmental Statistics, 4*(1), 9–28.

De Cesare, L., Myers, D. E., & Posa, D. (2001). Product-sum covariance for space–time modeling: an environmental application. *Environmetrics, 12*, 11–23.

de Haan, P. (2009). Applied comprehensive NO_2 and particulate matter dispersion modelling for Switzerland. *International Journal of Environment and Pollution, 36*(1–3), 204–223.

De Iaco, S., Myers, D.E., & Posa, D. (2001a). Total air pollution and space–time modelling. In P. Monestiez, et al. (Eds.), *geoENV III – geostatistics for environmental applications* (pp. 45–56). Kluver Academic Publishers.

De Iaco, S., Myers, D. E., & Posa, D. (2001b). Space–time analysis using a general product-sum model. *Statistics and Probability Letters, 52*, 21–28.

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

De Iaco, S., Myers, D. E., & Posa, D. (2002b). Nonseparable space–time covariance models: Some parametric families. *Mathematical Geology, 34*, 23–42.

De Iaco, S., Myers, D. E., & Posa, D. (2003). The linear coregionalization model and the product-sum space–time variogram. *Mathematical Geology, 35*, 25–38.

de Kassteele, J. v., & Stein, A. (2006). A model for external drift kriging with uncertain covariates applied to air quality measurements and dispersion model output. *Environmetrics, 17*, 309–322.

de Kassteele, J. v., Stein, A., Dekkers, A. L. M., & Velders, G. J. M. (2007). External drift kriging of NO_x concentrations with dispersion model ouput in a reduced air quality monitoring network. *Environmental and Ecological Statistics, 16*(3), 321–339.

Delicado, P., Giraldo, R., Comas, C., & Mateu, J. (2010). Statistics for spatial functional data: Some recent contributions. *Environmetrics, 21*, 224–239.

Dimitrakopoulos, R., & Luo, X. (1997). Joint space–time modelling in the presence of trends. In E. Y. Baafi & N. A. Schofield (Eds.). *Geostatistics Wollongong '96* (Vol. 1, pp. 138–149). Kluwer: Dordrecht.

Dorling, S. R., Foxall, R. J., Mandic, D. P., & Cawley, G. C. (2003). Maximum likelihood cost functions for neural networks models of air quality data. *Atmospheric Environment, 37*, 3435–3443.

Dubois, G., Saisana, M., Chaloulakou, A., Spyrellis, N. (2002). Spatial correlation analysis of nitrogen dioxide concentrations in the area of Milan, Italy. In A. E. Rizzoli & A. J. Jakeman (Eds.), *Proceedings of the first biennial meeting of the international modelling and software society* (Vol. 3, pp. 536–541).

Eynon, B. P., & Switzer, P. (1983). The variability of rainfall acidity. *Canadian Journal of Statistics, 11*(1), 11–23.

Fernández-Avilés, G. (2010). *Spatio-temporal modelling of environmental processes derived from economic activity*. Doctoral thesis, University of Castilla-La Mancha, Toledo.

Ferraty, F., & Vieu, P. (2006). *Non parametric functional data analysis. Theory and practice*. New York: Springer.

Ferreira, F., Tente, H., Torres, P., Cardoso, S., & Palma-Oliveira, José. M. (2000). Air quality monitoring and management in Lisbon. *Environmental Monitoring and Assessment, 65*(1–2), 443–450.

Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association, 102*, 321–331.

Fuentes, M., Chen, L., & Davis, J. M. (2008). A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics, 19*(5), 487–597.

Gardner, M. W., & Dorling, S. R. (1999). Neural network modelling and prediction of hourly NO_x and NO_2 concentrations in urban air in London. *Atmospheric Environment, 33*(5), 709–719.

Gilge, S., Plass-Duelmer, C., Fricke, W., Kaiser, A., Ries, L., & Buchmann, B. (2010). Ozone, carbon monoxide and nitrogen oxides time series at four alpine GAW mountain stations in Central Europe. *Atmospheric Chemistry and Physics Discussions, 10*, 19071–19127.

Giraldo, R. (2009). *Geostatistical analysis of functional data*. Doctoral thesis. Universitat Politecnica de Catalunya, Spain.

Gneiting, T. (2002). Stationary covariance functions for space–time data. *Journal of the American Statistical Association, 97*, 590–600.

Grassberger, P., Schreiber, T., & Schaffrath, C. (1991). Non-linear time sequence analysis. *International Journal of Bifurcation and Chaos, 1*, 532–547.

Haas, T. (1998). Statistical assessment of spatio-temporal pollutant trends and meteorological transport models. *Atmospheric Environment, 32*(11), 1865–1879.

Huang, G. H. (1992). A stepwise cluster analysis method for predicting air quality in an urban environment. *Atmospheric Environment, 26*(3), 349–357.

Janssen, S., Dumont, G., Fierens, F., & Mensink, C. (2008). Spatial interpolation of air pollution measurements using CORINE land cover data. *Atmospheric Environment, 42*, 4884–4903.

Jiménez-Hornero, F. J., Gutiérrez de Ravé, E., Ariza-Villarverde, A. B., & Giráldez, J. V. (2009). Description of the seasonal pattern in ozone concentration time series by using the strange attractor multifractal formalism. *Environmental Monitoring and Assessment, 160*, 229–236.

Kaufman, C., Schervish, M. & Nychka, D. (2007). *Covariance tapering for likelihood-based estimation in large spatial datasets*. Technical report. Research Triangle Park, North Carolina: Statistical and Applied Mathematical Sciences Institute.

Kim, D., & Lee, J. (2010). Application of neural network model to vehicle emissions. *International Journal of Urban Sciences, 14*(3), 264–275.

Kukkonen, J., Portanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainenb, M., et al. (2003). Extensive evaluation of neural network models for the prediction of NO_2 and PM10 concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmospheric Environment, 37*(32), 4539–4550.

Kyriakidis, P. C., & Journel, A. G. (1999). Geostatistical space–time models: A review. *Mathematical Geology, 31*(6), 651–684.

Kyriakidis, P. C., & Journel, A. G. (2001). Stochastic modeling of atmospheric pollution: A spatial time-series framework. Part II: Application to monitoring monthly sulfate deposition over Europe. *Atmospheric Environment, 35*(13), 2339–2348.

Lapolla, V. (2009). *Spatio-temporal modeling of environmental processes: Nitrogen oxides concentrations in the Tuscany region*. Unpublished PhD dissertation, RUTA.

Lee, J. Y., Leem, J. H., Kim, H. C., Kim, J. H., Lim, D. H., & Son, B. K. (2011). Ambient nitrogen dioxide prediction by land use regression modeling in Incheon. *Epidemiology, 22*(1), 218.

Lertxundi-Manterola, A., & Saez, M. (2009). Modelling of nitrogen dioxide (NO_2) and fine particulate matter (PM_{10}) air pollution in the metropolitan areas of Barcelona and Bilbao, Spain. *Environmetrics, 20*(5), 477–493.

Manjunatha, R., Narayama, B. P., & Reddy, K. H. C. (2010). application of artificial neural networks for emission modelling of biodiesels for a C.I. engine under varying operating conditions. *Applied Science, 4*(3), 77–89.

Mendes, J. M., & Turkman, K. F. (2002). A simple spatio-temporal procedure for the prediction of air pollution levels. *Journal of Chemometrics, 16*, 223–232.

Montero, J. M., & Fernández-Avilés, G. (2009). How to include environmental quality kriged indexes in hedonic housing price models. *The Polish Journal of Environmental Studies, 18*(5B), 149–158.

Montero, J. M., Garcia-Centeno, M. C., & Fernandez-Avilés, G. (2011). A threshold autoregressive asymmetric stochastic volatility strategy to alert of violations of the air quality standards. *International Journal of Environmental Research, 5*(1), 23–32.

Obodeh, O., & Ajuwa, C. I. (2009). Evaluation of artificial neural network performance in predicting diesel engine NO_x emissions. *European Journal of Scientific Research, 33*(4), 642–653.

Ramsay, J., & Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B, 53*, 539–572.

Reich, B., Eidsvik, J., Guindani, M., Nail, A., & Schmidt, A. (2010). A class of covariate-dependent spatiotemporal covariance functions. In *2010 Joint statistical meetings in Vancouver*, British Columbia.

Routledge, H. C., & Ayres, J. G. (2005). Air pollution and the heart. *Occupational Medicine, 55*, 439–447.

Salcedo-Sanz, S., Portilla-Figueras, J. A., Ortiz-Garcia, E. G., Perez-Bellido, A. M., Garcia-Herrera, R., & Elorrieta, J. I. (2009). Spatial regression analysis of NO_x and O_3 concentrations in Madrid urban area using radial basis function networks. *Chemometrics, 99*(1), 79–90.

Sherman, M. (2010). *Spatial statistics and spatio-temporal data: Covariance functions and directional properties*. John Wiley & Sons, Ltd.

Skene, K. J., Gent, J. F., McKay, L. A., Belanger, K., Leaderer, B. P., & Holford, T. R. (2010). Modeling effects of traffic and landscape characteristics on ambient nitrogen dioxide levels in Connecticut. *Atmospheric Environment, 44*, 5156–5164.

Stein, M. L. (2004). *Statistical methods for regular monitoring data*. University of Chicago. Center for Integrating Statistical and Environmental Science. Technical Report, 15.

Stieb, D. M., Judek, S., & Burnett, R. T. (2002). Meta-analysis of time-series studies of air pollution and mortality: Effects of gases and particles and the influence of cause of death, age, and season. *Journal of the Air & Waste Management Association, 52*, 470–484.

Vanderhaegen, E., Deneve, M., Laget, H., Faniel, N. & Mertens, J. (2010). Predictive emissions monitoring using a continuously updating neural network, Parts A and B, combustion, fuels and emissions. In *ASME conf. proc./year 2010* (Vol. 2, pp. 769–775) Paper no. GT2010-22899.

Vecchia, A. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society, 50*, 297–312.

Vyas, V. M., & Christakos, G. (1997). Spatiotemporal analysis and mapping of sulfate deposition data over Eastern USA. *Atmospheric Environment, 31*(21), 3623–3633.

Yanosky, J. D., Schwartz, J., & Suh, H. H. (2008). Associations between measures of socioeconomic position and chronic nitrogen dioxide exposure in Worcester, Massachusetts. *Journal of Toxicology and Environmental Health, Part A: Current Issues, 71*(24), 1593–1602.

Yong, L., Huaicheng, G., Gouzhu, M., & Pingjian, Y. (2008). A Bayesian hierarchical model for urban air quality prediction. *Atmospheric Environment, 42*, 8464–8469.

Zarandi, M., Khajevandi, S., Damez-Fontaine, M., & Ardestani, M. (2008). Determination of air pollution monitoring stations. *International Journal of Environmental Research, 2*(3), 313–318.

Zhang, Z. & San, Y. (2004). Adaptative wavelet neural network for predition of hourly NO_x and NO_2 concentrations. In R.G. Ingalls, M. D. Rossetti, J. S. Smith, & B. A. Peters (Eds.), *Proceedings of the 2004 winter simulation conference* (pp. 1170–1778).