

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

## Kriging with External Drift for functional data for air quality monitoring

### **This is the author's manuscript**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/137791> since 2016-02-06T11:57:38Z

*Published version:*

DOI:10.1007/s00477-013-0806-y

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

## Kriging with External Drift for functional data for air quality monitoring

Rosaria Ignaccolo · Jorge Mateu · Ramon  
Giraldo

Received: date / Accepted: date

**Abstract** Functional data featured by a spatial dependence structure occur in many environmental sciences when curves are observed, for example, along time or along depth. Recently, some methods allowing for the prediction of a curve at an unmonitored site have been developed. However, the existing methods do not allow to include in a model exogenous variables that, for example, bring meteorology information in modeling air pollutant concentrations. In order to introduce exogenous variables, potentially observed as curves as well, we propose to extend the so-called kriging with external drift - or regression kriging - to the case of functional data by means of a three-step procedure involving functional modeling for the trend and spatial interpolation of functional residuals. A cross-validation analysis allows to choose smoothing parameters and a preferable kriging predictor for the functional residuals. Our case study considers daily  $PM_{10}$  concentrations measured from October 2005 to March 2006 by the monitoring network of Piemonte region (Italy), with the trend defined by meteorological time-varying covariates and orographical constant-in-time variables. The performance of the proposed methodology is evaluated by predicting  $PM_{10}$  concentration curves on 10 validation sites, even with simulated realistic datasets on a larger number of spatial sites. In this application the proposed methodology represents an alternative to spatio-temporal modeling but it can be applied more generally to spatially dependent functional data whose domain is not a time interval.

---

Rosaria Ignaccolo's work was partially supported by Regione Piemonte and FIRB 2012 grant (project no. RBFR12URQJ) provided by the Italian Ministry of Education, Universities and Research. This research was also supported in part by the Spanish Ministry of Education and Science and Bancaja through grants MTM2010-14961 and P1-1B2012-52.

---

R. Ignaccolo  
Università degli Studi di Torino, Italy  
Tel.: +39-011-6703873  
Fax: +39-011-6703895  
E-mail: rosaria.ignaccolo@unito.it

J. Mateu  
Universitat Jaume I, Castello, Spain

R. Giraldo  
Universidad Nacional de Colombia, Bogota, Colombia

**Keywords** Functional data modeling · Functional linear model · Residual kriging · Particulate matter · Spatial dependence

## 1 Introduction

In recent years there has been an increasing interest in modeling functional data that, in environmental studies, often arise when dense sets of measurements are recorded over a period of time or over some domain (depth or height for instance). Statistical methods for analyzing this type of data are enclosed in a new branch of statistics called *Functional Data Analysis (FDA)* [26], [25], [8]. Over the last few years the analysis of functional data has been the focal attention of the statistical community. Functional data models provide a suitable framework for the statistical analysis of several environmental phenomena involving continuous time evolution and/or spatial variation. The functional autoregressive model in Salmerón and Ruiz-Medina (2009) [33] extends the classical autoregressive model to the infinite-dimensional space context. In Ruiz-Medina and Salmerón (2010) [30] the problem of functional filtering of an autoregressive Hilbertian (ARH) process, affected by additive Hilbertian noise, is addressed when the functional parameters defining the ARH(p) equation are unknown. Pseudodifferential evolution models have been widely used in the description of biological, geophysical and environmental systems. As an interesting case, Ruiz-Medina and Fernández-Pascual (2010) [29] consider the case where functional sample information is available from such systems. Despite these references, the scarce amount of contributions in the FDA inferential area has stood out. In particular, more recently, attention has been directed to spatially dependent functional data so that the term *Spatial Functional Statistics* has been introduced; Delicado *et al.* (2010) [5], Ruiz-Medina (2012) [31], Horváth and Kokoszka (2012) [16](Ch. 17-8), and Kokoszka (2012) [18] give a review of recent contributions and open challenges in this field. Specifically, for geostatistical data, the problem to predict a curve at a specified location using the curves at available locations has been addressed in [13], [11], [10], [21] by generalizing univariate and multivariate geostatistical techniques to the functional context and giving rise to the so-called *functional kriging*. Nevertheless, these recently developed geostatistical techniques for functional data consider ordinary kriging models such that the mean function of the process is supposed to be constant. In many applicative contexts this assumption is not realistic, hence there is need for methodologies suitable for non-stationary functional data. In this context, Caballero *et al.* (2013) [1] provide a solution to the problem of the spatial prediction of functional data in the absence of stationarity when the spatial trend is modeled as a function of the coordinates only. Actually Gromenko and Kokoszka (2013) [14] develop a methodology to estimate a mean function that is a linear combination of known covariate functions, but this mean function does not depend on space since it is defined to represent the mean temporal evolution of spatially distributed curves. Instead Temiyasathit *et al.* (2009) [34] propose a procedure to obtain spatial prediction of ozone concentration profiles using meteorological variables. They use multiple linear regression to model the wavelet coefficients of the ozone profile as a function of the wavelet coefficients of the profiles of meteorological variables. In a later stage, kriging is used to interpolate the regression coefficients at unsampled locations.

In classical geostatistics, spatial prediction for nonstationary processes is performed by taking into account a spatial trend (also called “drift”) that is modeled as a function of the coordinates only - in the Universal kriging - or defined “externally” through some auxiliary/exogenous variables - in the Kriging with External Drift (KED) model. Being mathematically equivalent, also the term Regression Kriging is used to specify that the drift and residuals are fitted separately and then summed up afterwards [15]. In this paper, we propose to extend the KED model to the case of functional data by means of a three-step procedure allowing for the introduction of exogenous variables, both scalar and functional, in a functional drift. We focus on the air quality monitoring, and in particular on particulate matter that is the most problematic pollutant for health and has only decreased slightly over the last decade in Europe<sup>1</sup>, despite primary PM emissions from transport have been reduced (but emissions of primary PM from commercial, institutional and households fuel combustion have increased). [6]. In the Po Valley the limit values fixed by the European and national directives (EU Council Directive 1999/30/EC) are usually not met, especially in urban areas and during the winter season. In this context, particulate matter concentration spatial prediction is useful for assessing air quality and health risk also where no monitoring stations are displaced, and an effective prediction can be reached only by taking into account meteorological and orographical covariates. Since air quality pollutants and meteorological variables are gathered along time, with a certain frequency, we consider their underlying functional form and treat the observed time series as functional data. Then, by including meteorological and orographical information in the external drift, we carry out the proposed Functional Kriging with External Drift (FKED) to predict curves of particulate matter concentration. The dataset of our case study has already been considered by Cameletti *et al.* (2011 [2] and 2012 [3]). In [2] six hierarchical models are compared: they share a common large-scale component whereas the residual detrended process is modeled by specifying certain spatio-temporal covariance functions with increasing complexity. The authors conclude that the model named “A1”, whose residual process has a purely spatial covariance function, is preferable because of good performance at a reasonable computational cost (models are fitted by MCMC); while the model called “C” has a slightly better prediction capability but a larger computational cost (since a spatial process evolving in time according to an AR(1) equation is involved). Such computational cost can be reduced by implementing a Stochastic Partial Differential Equations (SPDE) approach with the INLA algorithm, as done in Cameletti *et al.* [3] where predictions at the same validation sites are obtained with slightly worst performance but quite smaller computational time.

Here, with the same dataset, we adopt a functional data approach such that the temporal component is hidden in the data domain. Since the issue of which approach - between functional and space-time geostatistics - should be used is still open, we aim to compare the spatial prediction capability of our proposal at the same validation sites by means of some performance indexes (also used in [2]). However, when many observations per functional data are collected and there is interest in prediction of a whole function - and not a single value - at an unvisited

---

<sup>1</sup> The European Environmental Agency has designated 2013 as the Year of Air (see <http://www.eea.europa.eu/highlights/2013-kicking-off-the-2018year>).

site, geostatistics for functional data should be the natural approach. Even more if the functional data domain is not temporal.

The paper is organized as follows. In Section 2 we propose a methodology to carry out Kriging with External Drift for functional data and suggest a criterion to choose the smoothing parameters and find a preferable kriging predictor among three different alternatives. Then we apply Functional KED to predict curves of particulate matter concentration in Section 3, where results of a cross-validation analysis are presented and an application on a validation dataset allows to assess its spatial prediction capability. In Section 4, we use the real dataset as a basis to simulate curves spatially correlated and explore the large sample behavior of the proposed procedure. A discussion with insights on further developments follows in Section 5.

## 2 Kriging with External Drift for functional data

### 2.1 Notations and assumptions

Let  $\mathcal{Y}_s = \{Y_s(t); t \in T\}$  be a *functional random variable* (f.r.v.) observed at location  $s \in D \subseteq \mathbb{R}^d$ , whose realization is a function of  $t \in T$  - that is a functional data - where  $T$  is a compact subset of  $\mathbb{R}$  [8]. Assume that we observe a sample of curves  $\mathcal{Y}_{s_i}$ , for  $s_i \in D$ ,  $i = 1, \dots, n$ , that take values in a separable Hilbert space of square integrable functions. The set  $\{\mathcal{Y}_s, s \in D\}$  constitutes a *functional random field* or a *spatial functional process* as defined in [5], that can be non-stationary and whose elements are supposed to follow the model

$$\mathcal{Y}_s = \mu_s + \epsilon_s. \quad (1)$$

The term  $\mu_s$  is interpreted as a drift describing a spatial trend while  $\epsilon_s$  represents a residual random field that is zero-mean, second-order stationary and isotropic, so that

- i)  $\mathbb{E}(\mathcal{Y}_s) = \mu_s, s \in D;$
- ii)  $\mathbb{E}(\epsilon_s) = 0, s \in D;$
- iii)  $Cov(\epsilon_{s_i}, \epsilon_{s_j}) = C(h), \quad \forall s_i, s_j \in D$  with  $h = \|s_i - s_j\|.$

At the generic site  $s_i$ ,  $i = 1, \dots, n$ , and at point  $t$  model (1) can be rewritten as a functional concurrent linear model [26]

$$Y_{s_i}(t) = \mu_{s_i}(t) + \epsilon_{s_i}(t) \quad (2)$$

with the drift

$$\mu_{s_i}(t) = \alpha(t) + \sum_p \gamma_p(t) C_{p,i} + \sum_q \beta_q(t) X_{q,i}(t) \quad (3)$$

where  $\alpha(t)$  is a functional intercept,  $C_{p,i}$  is the  $p$ -th scalar covariate and  $X_{q,i}(t)$  is the  $q$ -th functional covariate at site  $s_i$ ,  $\gamma_p(t)$  and  $\beta_q(t)$  are the covariate coefficients. In the drift (3) both scalar and functional covariates are included, and the coefficients  $\gamma_p(t)$  and  $\beta_q(t)$  are also of a functional nature, allowing to estimate nonlinear effects of a covariate.

## 2.2 Functional kriging

In order to predict a curve at an unmonitored site  $s_0$ , taking into account exogenous variables in the drift, we propose a three-step procedure. At the **first step** the functional regression model (2) with functional response and scalar and functional covariates is fitted by Generalized Cross-Validation (GCV) as specified in Section 2.4, in order to estimate the drift coefficients and obtain the functional residuals

$$e_{s_i}(t) = Y_{s_i}(t) - \hat{\mu}_{s_i}(t) = Y_{s_i}(t) - \left[ \hat{\alpha}(t) + \sum_p \hat{\gamma}_p(t) C_{p,i} + \sum_q \hat{\beta}_q(t) X_{q,i}(t) \right].$$

At the **second step** the residual curve prediction at the unmonitored site  $s_0$  can be obtained by *ordinary kriging for functional data (OKFD)* [13], according to which

$$\hat{e}_{s_0}(t) = \sum_{i=1}^n \lambda_i e_{s_i}(t)$$

where the kriging coefficients  $\lambda_i \in \mathbb{R}$  are constant, so that the predicted curve is a linear combination of data residual curves. The weights  $\lambda_i$  are determined as the solution of a linear system written to solve the optimization problem

$$\min_{\lambda_1, \dots, \lambda_n} \int_T \text{Var}(\hat{e}_{s_0}(t) - e_{s_0}(t)) dt, \quad \text{s.t.} \quad \sum_{i=1}^n \lambda_i = 1,$$

where, similarly to classical geostatistics, some semivariogram values need to be known. In particular, the kriging coefficients  $\lambda_i$  depend on the so-called *trace-semivariogram* defined for a zero-mean weakly-stationary process as  $v(h) = \int_T v_h(t) dt$  where  $v_h(t) = \frac{1}{2} \text{Var}(e_{s_i}(t) - e_{s_j}(t))$  with the Euclidean distance  $h = \|s_i - s_j\|$ . This is estimated by (for further details see [13])

$$\hat{v}(h) = \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} \int_T (e_{s_i}(t) - e_{s_j}(t))^2 dt$$

where  $N(h) = \{(s_i, s_j) : \|s_i - s_j\| = h\}$  and  $|N(h)|$  is the number of distinct elements in  $N(h)$ . Once the trace-semivariogram is estimated for a sequence of  $G$  values  $h_g$ , a classical parametric model (exponential or Matérn for example) can be fitted to the points  $(h_g, \hat{v}(h_g))$ ,  $g = 1, \dots, G$ , as in classical geostatistics.

As an alternative to the ordinary kriging, we can consider “*continuous time-varying kriging*” for functional data (CTKFD) [11], providing

$$\hat{e}_{s_0}(t) = \sum_{i=1}^n \lambda_i(t) e_{s_i}(t),$$

where the kriging coefficients depend on  $t$ . This predictor is a hybrid between ordinary kriging and the functional linear concurrent (*point-wise*) model such as shown in [26]. The estimation of the functional parameters  $\lambda_i(t)$ ,  $i = 1, \dots, n$ , is carried out by using an approach based on the use of  $Nb$  basis functions. The

curves and the functional parameters are represented in terms of basis functions, that is

$$e_{s_i}(t) = \sum_{l=1}^{Nb} a_{il} B_l(t) \quad \text{and} \quad \lambda_i(t) = \sum_{l=1}^{Nb} b_{il} B_l(t). \quad (4)$$

The coefficients  $a_{il}$  are assumed to form a multivariable random field  $\{\alpha_l\}_{l=1}^{Nb}$  with  $\alpha_l = (a_{1l}, \dots, a_{nl})$  and multivariate geostatistics, such as a Linear Model of Coregionalization (LMC) [37], is applied to estimate cross-covariances. Then an optimization problem is solved to have a BLUP so that  $b_{il}$  are estimated (for further details see [11]).

Another option is considering the so-called “*functional kriging total model*” (FKTM) [10], [21] where the kriging coefficients are defined on  $T \times T$  and the prediction at  $t$  is obtained integrating over  $T$ , that is

$$\hat{e}_{s_0}(t) = \sum_{i=1}^n \int_T \lambda_i(\tau, t) e_{s_i}(\tau) d\tau.$$

Again the curves and the functional parameters  $\lambda_i(\tau, t)$   $i = 1, \dots, n$  are expanded in terms of  $Nb$  basis functions, as

$$e_{s_i}(t) = \sum_{l=1}^{Nb} a_{il} B_l(t) \quad \text{and} \quad \lambda_i(\tau, t) = \sum_{j=1}^{Nb} \sum_{l=1}^{Nb} c_{jl}^i B_j(\tau) B_l(t)$$

and again a LMC is fitted in order to give a solution to the problem of estimating the functional parameters  $\lambda_i(\tau, t)$  through the estimation of  $c_{jl}^i$ .

Finally, at the **third step** we get the prediction at the unmonitored site  $s_0$  by adding - as in the classical regression kriging - the two terms, that is

$$\hat{Y}_{s_0}(t) = \hat{\mu}_{s_0}(t) + \hat{e}_{s_0}(t)$$

where  $\hat{\mu}_{s_0}(t) = \hat{\alpha}(t) + \sum_p \hat{\gamma}_p(t) C_{p,0} + \sum_q \hat{\beta}_q(t) X_{q,0}(t)$  depends on the covariate values  $C_{p,0}$  and  $X_{q,0}(\cdot)$  at the site  $s_0$ .

### 2.3 Choosing the smoothing parameters and evaluating kriging predictors

Despite the functional framework, data are usually gathered as a finite discrete set of pairs  $(t_j, y_{ij})$ ,  $t_j \in T, j = 1, \dots, M, i = 1, \dots, n$ , and (unless there is no observational noise)

$$y_{ij} = Y_{s_i}(t_j) + \delta_{ij}$$

where  $\delta_{ij}$  represents a measurement error and  $Y_{s_i}(\cdot)$  is a continuous function that corresponds to a realization of the functional random field  $\{Y_s, s \in D\}$  at the site  $s_i$ . The set of points  $\{t_j\}_{j=1}^M \subset T$  can be considered the same for all functions in a functional dataset, and often these points are evenly spaced in  $T$ . We make here these assumptions but in general the set of points where a curve is observed could vary from site to site. In the latter case, more attention needs to be paid when converting discrete data to functional data (by using for example free knots for the splines smoothing introduced below).

The conversion from discrete data to curves involves smoothing, and linear combinations of B-spline functions  $B_l(t)$  (spline functions are constructed by joining

polynomials of degree  $d$  together at points called “knots”) are used since they are flexible and appropriate for use in general environmental variables, while Fourier basis functions are appropriate when in presence of periodic data [26].

Then, for all  $t$  the curve  $Y_{s_i}(t)$  is estimated by

$$\tilde{Y}_{s_i}(t) = \sum_{l=1}^{K+d+1} \hat{c}_l^i B_l(t) \quad (5)$$

where  $d$  is the spline degree (in the following  $d = 3$  so that we have cubic splines) and  $K$  the number of interior knots in the domain  $T = (a, b)$  of the function  $Y_{s_i}$ . Note that the number of basis functions in (5) is now  $Nb = K + d + 1$  and smoothed data  $\tilde{Y}_{s_i}(t_j)$  will be used to fit Model (2) in the case study presented in Section 3, since we assume to have observational noise.

Spline coefficients  $\mathbf{c}^i = (c_1^i, \dots, c_{K+d+1}^i)$  are estimated, for each  $i$ , by adopting a penalized least squares criterion

$$PENSSSE(\mathbf{c}^i) = \sum_j (y_{ij} - Y_{s_i}(t_j))^2 + \eta \int_T [D^2 Y_{s_i}(t)]^2 dt \quad (6)$$

where  $Y_{s_i}(t) = \sum_{l=1}^{K+d+1} c_l^i B_l(t)$ ,  $\eta$  is the *penalty parameter*,  $D^2$  denotes the second derivative and the penalty term is chosen as the integrated square of the second derivative that quantifies the total curvature of the function (and hence its roughness). The parameter  $\eta$  controls the trade-off between the fit to the observed data and the smoothness of the fitted curve, so that when  $\eta$  is large the fitted curve is smoother but the data fits worse. In order to estimate spline coefficients by minimizing  $PENSSSE$  we need to fix a value for  $\eta$  and, in addition, we need to choose a sequence of  $K$  interior knots  $\xi = (\xi_1, \xi_2, \dots, \xi_K)$ . By taking equally spaced knots in  $T = (a, b)$  we need only to fix the number of knots  $K$ , besides  $\eta$ , while  $d = 3$  is fixed.

Since the aim here is to predict a whole function  $Y_{s_0}(t)$  at an unmonitored site  $s_0$ , we adopt a choice criterion based on leave-one-out cross-validation (as introduced in [13] and [11]). Data gathered at each site  $s_i$  is temporarily removed from the dataset and a smoothed function is predicted at this location using a functional kriging predictor based on the remaining smoothed functions, for  $K \in [K_{min}, K_{max}]$  and  $\eta \in [\eta_{min}, \eta_{max}]$ . Then the optimal values  $(K^*, \eta^*)$  are those minimizing the function

$$FCVr(K, \eta) = \sum_{i=1}^n SSE(i) = \sum_{i=1}^n \sum_{j=1}^M \left( y_{ij} - \hat{Y}_{s_i}(t_j)^{(i)} \right)^2 \quad (7)$$

where  $\hat{Y}_{s_i}(t_j)^{(i)}$  is the prediction at  $s_i$  evaluated at  $t_j$ , by leaving  $s_i$  data out of the sample. In [13] and [11] a similar procedure has been called *functional cross-validation* and the objective function  $FCV$  contains squared differences between smoothed data and predictions. Here we call the function  $FCVr$  - where  $r$  stands for *raw* - to stress the fact that raw data  $y_{ij}$  are compared to the predicted values  $\hat{Y}_{s_i}(t_j)^{(i)}$  when evaluating  $SSE$  at each site. This is motivated by the fact that using smoothed data  $\tilde{Y}_{s_i}(t_j)$  instead of raw data  $y_{ij}$  in (7) would provide a minimum value when  $K$  is small and  $\eta$  large (we have observed this in our case study); in that case the smoothed functions  $\tilde{Y}_{s_i}(t)$  would be close to the overall

mean function and the predictions would be easier to obtain - with  $SSE$  smaller - but they would be too smooth in the air quality context, where we want to predict concentration peaks that people breaths (removing only the observational noise).

Besides the choice of the smoothing parameters, the function  $FCVr$  allows to compare the three alternatives OKFD, CTKFD and FKTM at the second step of the procedure proposed in this paper (also by looking in detail  $SSE(i)$  values). Moreover, in order to compare our proposal to competitors, such as Bayesian hierarchical models in [2] (for which cross-validation is computationally unfeasible), we use 10 validation sites and some prediction capability indexes in Section 3.3.

## 2.4 Implementation details

To implement our proposal all computations are coded in R [24]. First of all, the conversion to functional data is realized by using the `fda` package [28] that also contains the function `fRegress` for fitting a concurrent functional linear model, as shown in [27] (Chapter 10). However, only univariate independent variables are currently allowed and indeed the case of a functional response prediction in practice remains largely unexplored. Then to implement the *first step* of our procedure we need to figure out how to estimate the covariate coefficients in the drift (3) in an alternative way. In classical geostatistics,  $\mu_s$  is seen as a deterministic large scale component and when it is supposed to depend on exogenous variables the related coefficients are estimated by least squares criteria. In the functional framework, where the covariate coefficients are also functional, we follow Ivanescu et al. [17] who propose a general framework for smooth regression of a functional response on one or multiple functional predictors by re-writing a functional linear model as a standard additive model. In particular in our model the functional coefficients in (3) are assumed to be expandable as

$$\alpha(t) = \sum_{l=1}^{k_0} A_{0,l}(t)c_{0,l}, \quad \gamma_p(t) = \sum_{l=1}^{k_p} a_{p,l}(t)c_{p,l} \quad \text{and} \quad \beta_q(t) = \sum_{l=1}^{k_q} a_{q,l}(t)c_{q,l}$$

where  $A_{0,l}(t)$ ,  $a_{p,l}(t)$  and  $a_{q,l}(t)$  are known basis functions, while  $c_{0,l}$ ,  $c_{p,l}$  and  $c_{q,l}$  are the related coefficients (to be estimated). Then we can write

$$\gamma_p(t)C_{p,i} = \sum_{l=1}^{k_p} a_{p,l}(t)C_{p,i}c_{p,l} = \sum_{l=1}^{k_p} A_{p,l,i}(t)c_{p,l},$$

and

$$\beta_q(t)X_{q,i}(t) = \sum_{l=1}^{k_q} a_{q,l}(t)X_{q,i}(t)c_{q,l} = \sum_{l=1}^{k_q} A_{q,l,i}(t)c_{q,l}$$

so that the functional linear model (2) can be re-written as a standard additive model

$$Y_{s_i}(t) = \sum_{l=1}^{k_0} A_{0,l}(t)c_{0,l} + \sum_p \sum_{l=1}^{k_p} A_{p,l,i}(t)c_{p,l} + \sum_q \sum_{l=1}^{k_q} A_{q,l,i}(t)c_{q,l} + \epsilon_{s_i}(t)$$

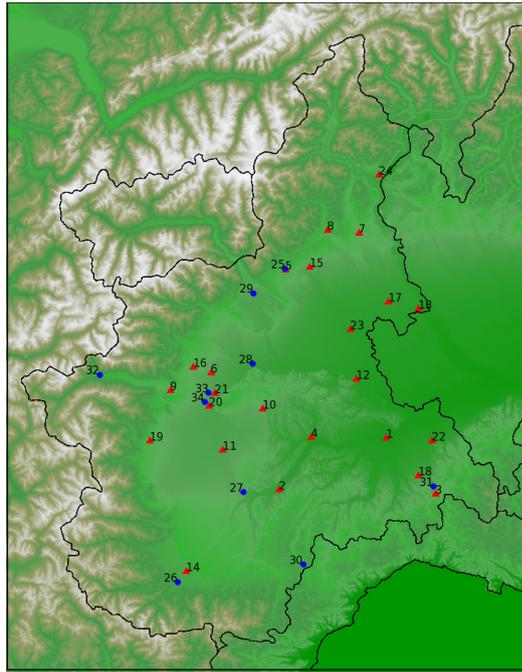
where  $A_{p,l,i}(t) = a_{p,l}(t)C_{p,i}$  and  $A_{q,l,i}(t) = a_{q,l}(t)X_{q,i}(t)$  are known because  $C_{p,i}$  and  $X_{q,i}(t)$  are “observed” without noise. This representation makes it possible to fit a concurrent functional linear model by means of the robust `mgcv` package [40], where the underlying representation and estimation of the models is based on a penalized regression spline approach, with automatic smoothness selection by using the Generalized Cross Validation (GCV) criterion (see [38] and [39]). This entails that  $k_0$ ,  $k_p$  and  $k_q$  are chosen as very large inside the GCV procedure and penalties are designed to suppress excessive roughness of the functional parameters.

Then, at the *second step*, we apply kriging for functional data in three alternative ways to the functional residuals in order to get predictions at the temporarily out of sample site in cross-validation or at the 10 validation sites. The first method OKFD is implemented by means of the `geofd` package [12] that includes the automatic choice of a model (in a list) for a trace-variogram, by minimizing the SSE between the theoretical variogram and the empirical one. Instead CTKFD and FKTM are carried out by our own R code (available upon request) that takes advantage of the package `gstat` [22] when fitting a linear coregionalization model. By adding predicted drifts and residuals, at the *third step*, we obtain predicted curves of the response variable that are compared with the observed time series later.

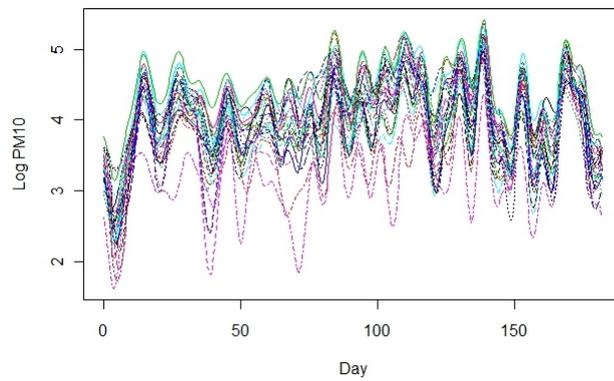
### 3 Spatial prediction of PM<sub>10</sub> curves

#### 3.1 Data

Our case study considers daily PM<sub>10</sub> concentrations (in  $\mu\text{g}/\text{m}^3$ ) measured from October 2005 to March 2006 (so that  $M = 182$ ) by the monitoring network of Piemonte region (Italy) in 24 sites (red triangles in Figure 1). For model performance assessing we have 10 extra validation stations (blue dots in Figure 1) as in [2] and [3], where the same dataset is considered. We carry out a cross-validation analysis in the following section, but the set of validation sites allows us to compare our results with those in [2]. The exogenous variables in the trend term are: (i) coordinates ( $UTMX$  and  $UTMY$ , in  $km$ ) and altitude ( $A$ , in  $m$ ), that are scalar; (ii) daily maximum mixing height ( $HMIX$ , in  $m$ ), daily total precipitation ( $PREC$ , in  $mm$ ), daily mean wind speed ( $WS$ ,  $m/s$ ), daily mean temperature ( $TEMP$ , in  $^\circ K$ ) and daily emission rates of primary aerosols ( $EMI$ , in  $g/s$ ), that are functional. Note that the time-varying variables are obtained from a nested system of deterministic computer-based models implemented by the environmental agency *ARPA Piemonte* [9]. For a complete description and preliminary analysis of the data we refer to [2]. In order to stabilize the within-station variances and making the marginal distribution of PM<sub>10</sub> data approximately normal, we also choose to transform data by the logarithm; Figure 2 shows the curves of  $\log(PM_{10})$ , obtained by using cubic splines, at the 24 monitoring sites. Moreover, since the ranges of the covariates are quite different, a standardization procedure is applied subtracting the mean and dividing by the standard error computed considering the 24 monitoring stations (note that functional covariates are standardized after the smoothing step).



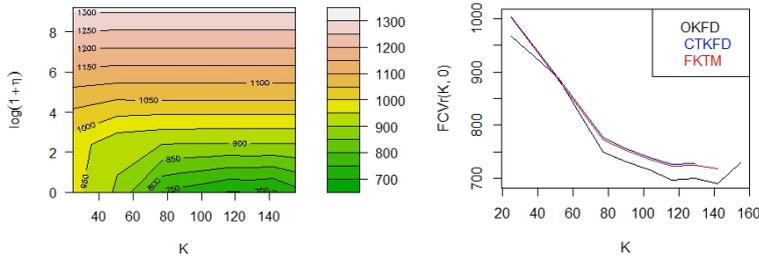
**Fig. 1** Locations of the 24  $PM_{10}$  monitoring sites - red triangles - and 10 validation stations - blue dots.



**Fig. 2** Smoothed time series of  $\log(PM_{10})$  at the 24 monitoring sites (color changes with sites).

### 3.2 Cross-validation

The leave-one-out cross-validation procedure described in Section 2.3 is applied to find the optimal smoothing parameters ( $K^*$  and  $\eta^*$ ) and to compare the three alternatives for kriging functional residuals. For the number of interior knots  $K$  we fix the set of possible values  $\{25, 51, 77, 90, 103, 116, 129, 142, 155\}$ , such that



**Fig. 3**  $FCVr$  surface for OKFD (left) and  $FCVr(K, 0)$  profiles in the three kriging cases (right).

there are from 1 to 6 knots per week (in the dataset we have 26 weeks). Note that  $K + 4 = Nb$  is the number of spline coefficients to be estimated - see (5) - that is also equal to the number of considered basis functions. Instead for the penalty parameter  $\eta$  we take - since it is usual to explore  $\eta$  values for a few log units [27] - the set of possible values  $\{0, 10, 10^2, 10^3, 10^4\}$ , defined as a power of 10 and visualized on a logarithmic scale. Clearly, when  $\eta = 0$  the criterion (6) is not penalized and the fitted curve could be very close to data, as much as possible with the selected basis functions (cubic splines here). When OKFD is applied, an automated algorithm chooses the exponential model for the trace-variogram almost always (sometimes a spherical one is chosen during the cross-validation). For CTKFD and FKTM, based on a preliminary exploration of the dataset, an exponential model was used for all direct variograms and cross-variograms for the LMC.

Figure 3 (left) shows the contour plot of the function  $FCVr(K, \eta)$  in case of OKFD at the second step, using a logarithmic scale for  $\eta$ . The  $FCVr$  surfaces for the alternatives CTKFD and FKTM - not reported here - are very similar although their values are generally slightly larger, while for a few values of  $(K, \eta)$  - as for example  $(155, 0)$  in the CTKFD case -  $FCVr$  becomes very large or impossible to evaluate for singularity problems; i.e. for some sites SSE is missing or too large (and so  $FCVr$  is not plotted in the profiles in Figure 3, right). This fact occurs when the number of interior knots  $K$  is large, making large the number  $Nb$  of “variables” in the LMC to fit in CTKFD and FKTM procedures and small the functional residuals to kriging.<sup>2</sup> Anyway, in all the three cases,  $FCVr$  is smaller when  $\eta = 0$ . Hence, to compare the cases of minimum  $FCVr$ , the profiles  $FCVr(K, 0)$  are shown in Figure 3 (right). It is evident that  $FCVr$  decreases sharply until  $K = 77$ , then the rate of decrease is smaller but in the OKFD case we can observe a small increasing for the largest  $K$ . It is also clear that, among the three methods, OKFD is preferable in terms of  $FCVr$  and Table 1 shows values of the 24  $SSE(i)$  for the three cases (OKFD, CTKFD and FKTM) when  $K = 116$  and  $\eta = 0$ . Overall,  $SSE(i)$  values are very close in the CTKFD and FKTM cases. In a few cases (see numbers in italic font) OKFD has slightly larger values but there is only one case

<sup>2</sup> Note that when we apply CTKFD and FKTM to the whole dataset (24 sites for fitting and 10 sites for validation) in Section 3.3 this kind of numerical problems does not occur.

**Table 1**  $SSE(i)$  over 24 fitting sites (red triangles in Figure 1) for FKED with the three alternative second steps: OKFD\_116\_0, CTKFD\_116\_0 and FKTM\_116\_0 corresponds to OKFD, CTKFD and FKTM respectively with  $K = 116$  and  $\eta = 0$ .

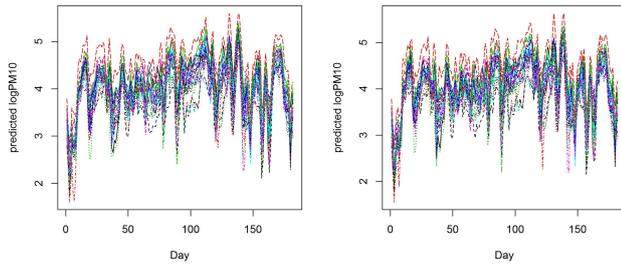
	Site	OKFD_116_0	CTKFD_116_0	FKTM_116_0
1	AL - Piazza D'Annunzio	23.694	27.047	27.176
2	Alba	35.044	41.409	42.075
3	Arquata Scrivia	25.578	29.711	30.387
4	AT - Scuola D'Acquisto	23.885	25.055	25.028
5	BI - Via Don Sturzo	<i>14.476</i>	<i>13.831</i>	<i>13.853</i>
6	Borgaro	<i>19.306</i>	<i>19.162</i>	<i>18.931</i>
7	Borgomanero	17.488	17.928	17.916
8	Borgosesia	9.866	10.238	10.333
9	Buttigliera Alta	17.932	20.806	20.936
10	Buttigliera d'Asti	19.431	20.555	20.688
11	Carmagnola	24.827	30.201	30.385
12	Casale Monferrato	50.689	53.424	53.690
13	Cerano	25.436	28.552	28.857
14	CN - Piazza II Regg. Alpini	<i>93.106</i>	<i>71.544</i>	<i>70.611</i>
15	Cossato	9.625	10.960	11.074
16	Druento La Mandria	24.846	26.701	27.024
17	NO - Viale Verdi	16.778	18.217	18.321
18	Novi Ligure	20.868	21.143	21.247
19	Pinerolo	17.068	18.997	19.109
20	TO - Lingotto	30.988	31.367	31.525
21	TO - Via Consolata	<i>14.614</i>	<i>14.374</i>	<i>14.183</i>
22	Tortona	22.645	23.191	23.568
23	VC - Corso Gastaldi	19.185	21.048	21.072
24	Verbania	118.944	127.802	129.053
	<b>FCVr</b>	<b>696.319</b>	<b>723.263</b>	<b>727.042</b>

that worths to be noted, that is site *14 CN - Piazza II Regg. Alpini*, where SSE is quite larger for OKFD. However this site is outside the spatial domain when it is left out in cross-validation, and the same happens to site *24 Verbania* where the worst performance occurs.

Moreover OKFD is preferable because CTKFD and FKTM could be numerically instable and involve the fitting of a Linear Coregionalization Model that is highly computational time demanding (while OKFD is fast). To apply OKFD in our case study, we observe in Figure 3 (right) that the values minimizing  $FCVr$  are  $K^* = 142$  and  $\eta^* = 0$ , although  $(116, 0)$  and  $(129, 0)$  give very close results. In fact Figure 4 shows very similar predicted curves at the 24 sites (when left out one at a time) obtained by OKFD at the second step and the couple  $(K, \eta)$  equal to  $(116, 0)$  and  $(142, 0)$ .

### 3.3 Prediction at the validation sites

Following the results of the cross-validation, we apply our proposal with OKFD at the second step and  $(K, \eta) = (142, 0)$ . Let us note that the number of interior knots  $K$ , related to the number of basis functions  $Nb = K + 4$ , has to be fixed at the beginning of the application (for the smoothing step and for fitting a FLM to the smoothed data) so that the first step results change if  $K$  changes.

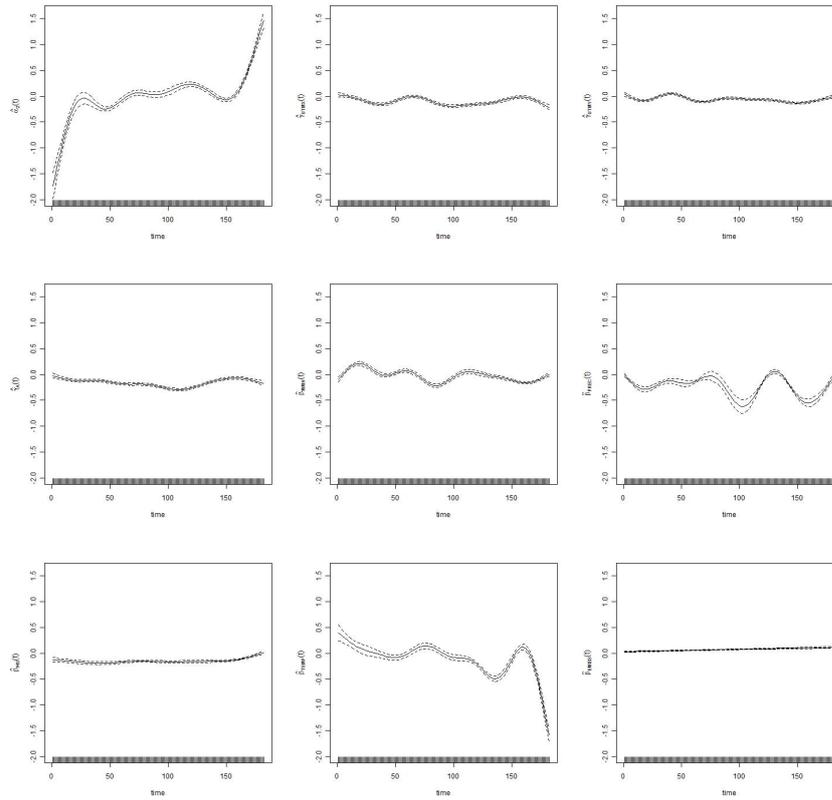


**Fig. 4** Cross-validation results: 24 predicted curves when  $K = 116$  (left) and  $K = 142$  (right),  $\eta = 0$ , with OKFD at the second step.

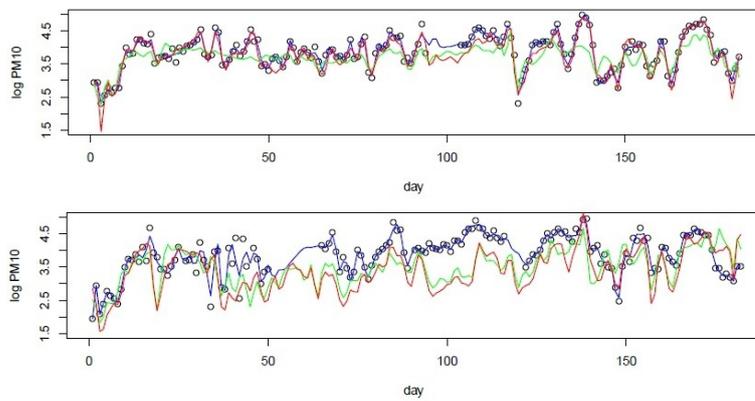
The functional intercept in (3) is decomposed as  $\alpha(t) = \alpha_1 + \alpha_2(t)$ , since in fitting an additive model the smooth terms are subject to sum-to-zero identifiability constraints [38]. The estimated scalar intercept is  $\hat{\alpha}_1 = 3.93$  on the log scale, that corresponds to an average pollution level of about  $50.9 \mu\text{g}/\text{m}^3$  while  $\hat{\alpha}_2(t)$  is shown in Figure 5, where we can observe that it changes in time mostly at the beginning and end of the considered winter semester. The estimated functional coefficients  $\hat{\gamma}_p(t)$  of the scalar covariates UTMX, UTMY and A, and  $\hat{\beta}_q(t)$  of the functional covariates HMIX, PREC, WS, TEMP and EMI are also shown in Figure 5. All covariates are significant (p-values for the smooth terms are obtained as discussed in [41]) as they were in [2], but now we can observe how they vary with time. In fact, for all the six models considered in [2] the estimated covariate coefficients - that are scalar - are negative except the one for the emissions (EMI). Instead our results show estimated functional coefficients varying with time with different behaviours, although they are negative in most of the  $t$  domain except for HMIX and TEMP - that change sign in time - and EMI that has a positive relationship with  $\text{PM}_{10}$ , as expected. Hence the importance of meteorological variables on air quality is confirmed, as well as the significant effect of altitude (A) in reducing  $\text{PM}_{10}$  concentration.

By applying the second and third step of the proposed FKED, we get predictions in the 10 validation sites (blue dots in Figure 1). To have a detailed example, Figure 6 shows raw data  $y_{ij}$ , smoothed data  $\hat{Y}_{s_i}(t)$ , the predicted drift  $\hat{\mu}_{s_i}(t)$  and the predicted curve  $\hat{Y}_{s_i}(t)$  at the sites 25 *Biella - Largo Lamarmora* and 30 *Saliceto*. The contribution of kriging residuals is clear at site 25, where the predicted curve has a local (in time) variability that the predicted drift does not reach. Instead at site 30, where we have the worst results, the predicted drift is - for the most part of time - not close to the observed data and the contribute of kriging does not compensate enough because this site is the farthest from the 24 sites used to fit the model.

In order to assess the spatial prediction capability of FKED, we compare observed and predicted data at the 10 validation sites, by evaluating performance indexes as in [2] (a simple residual analysis is also performed in [3]). Since we evaluate different predictors, we take as observed data the raw data  $y_{ij}$ , despite their observational noise, also because smoothed data change with  $(K, \eta)$ . We consider four indicators based on the differences between predicted and observed data: the



**Fig. 5** Estimated functional coefficients. First row:  $\hat{\alpha}_2(t)$ ,  $\hat{\gamma}_{UTMX}(t)$ ,  $\hat{\gamma}_{UTMY}(t)$ . Second row:  $\hat{\gamma}_A(t)$ ,  $\hat{\beta}_{HMIX}(t)$ ,  $\hat{\beta}_{PREC}(t)$ . Third row:  $\hat{\beta}_{WS}(t)$ ,  $\hat{\beta}_{TEMP}(t)$ ,  $\hat{\beta}_{EMI}(t)$ .



**Fig. 6** Prediction at 25 Biella - Largo Lamarmora (top) and 30 Saliceto (bottom). Black: raw data  $y_{ij}$ , blue: smoothed data  $\hat{Y}_{s_i}(t)$ , green: predicted drift  $\hat{\mu}_{s_i}(t)$  and red: predicted curve  $\hat{Y}_{s_i}(t)$ .

**Table 2** Performance indexes over 10 validation sites (blue dots in Figure 1) of FKED with OKFD at the second step and  $(K, \eta) = (142, 0)$ , called “OKFD\_142.0”.

	Site	NMBF	RMSE	WNNR	$\rho$
25	BI - Largo Lamarmora	-0.020	0.196	0.003	0.943
26	Borgo San Dalmazzo	-0.048	0.460	0.018	0.762
27	Bra	-0.059	0.313	0.006	0.922
28	Chivasso - Edipower	0.004	0.250	0.004	0.885
29	Ivrea	-0.058	0.331	0.008	0.914
30	Saliceto	-0.130	0.719	0.055	0.612
31	Serravalle Scrivia	0.038	0.463	0.012	0.793
32	Susa	0.005	0.481	0.016	0.785
33	TO - Piazza Rivoli	-0.005	0.199	0.002	0.940
34	TO - Via Gaidano	-0.015	0.267	0.004	0.903

usual root mean square error (RMSE) and the correlation coefficient  $\rho$ , together with the Normalized Mean Bias Factor (NMBF) [42] and the Weighted Normalized mean square error of the Normalized Ratios [23]. For a fixed location  $s_i$ , let  $z_j$  and  $\hat{z}_j$  be the observed and predicted time series (in our case  $y_{ij}$  and  $\hat{Y}_{s_i}(t_j)$ ) respectively, with  $j = 1, \dots, M$  and let  $\bar{z}$  and  $\bar{\hat{z}}$  be the corresponding mean values. The Normalized Mean Bias Factor is defined on  $\mathbb{R}$  by

$$\text{NMBF} = \begin{cases} \frac{\sum_j \hat{z}_j}{\sum_j z_j} - 1 & \text{if } \bar{\hat{z}} \geq \bar{z} \\ 1 - \frac{\sum_j z_j}{\sum_j \hat{z}_j} & \text{if } \bar{\hat{z}} < \bar{z} \end{cases}.$$

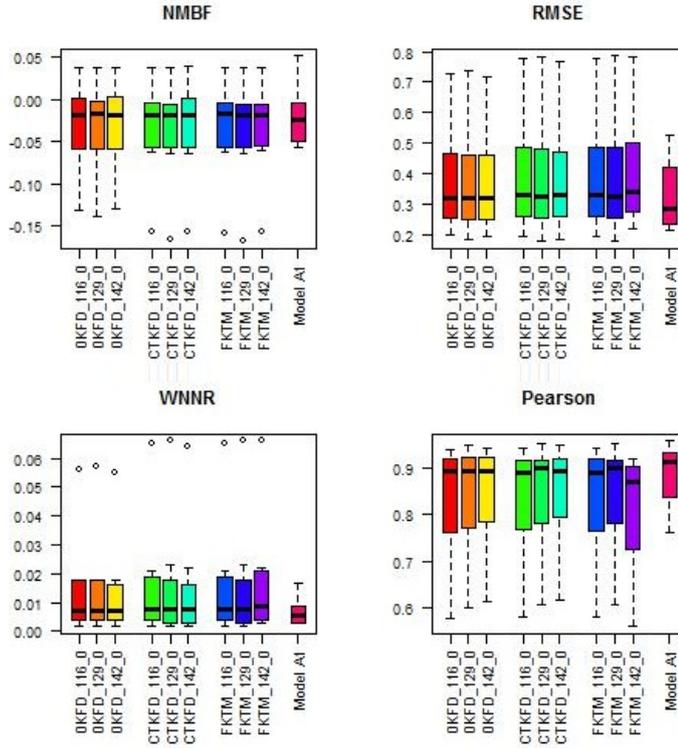
and has the advantage of both avoiding inflation due to low values of observations and overcoming the asymmetry problem between overestimation and underestimation, as discussed in [42]. The Weighted Normalized mean square error of the Normalized Ratios is defined by

$$\text{WNNR} = \frac{\sum_j s_j^2 (1 - k_j)^2}{\sum_j s_j k_j},$$

where  $s_j = z_j/\bar{z}$  is the weight and  $k_j = \exp(-|\ln(\hat{z}_j/z_j)|)$  is the Normalized ratio. WNNR is positive and has the advantage of taking properly into account the peaks of observed data (see the discussion in [23]).

Table 2 shows values of the four indexes in the case of FKED with OKFD at the second step and  $(K, \eta) = (142, 0)$ , called “OKFD\_142.0” in the following (and in Figure 7). Predictions are generally good, with a slight underestimation in 7 out of 10 sites (see NMBF). Correlations between predicted and observed data is above 0.75, except for the site *30-Saliceto* where we get the worst results in terms of prediction performance for all the four indexes, as expected by looking at Figure 6 (bottom). That figure also shows one of the sites where FKED performs better, that are *25* and *33*: for both of them the kriging predictor takes advantage of information given by very close neighbours.

A summary of Table 2 can be seen in Figure 7 by looking at the indexes distribution boxplots for OKFD\_142.0. Figure 7 also shows boxplots synthetizing the four indexes at the 10 validation sites for OKFD with  $(K, \eta) = (116, 0)$  and  $(K, \eta) = (129, 0)$ , as well as for CTKFD and FKTM with  $K = 116, 129, 142$  and  $\eta = 0$  (called by “name\_K.η”). We apply the proposed FKED to obtain predictions



**Fig. 7** Boxplots of the performance measure distributions computed for each model over the 10 validation stations (blue dots in Figure 1). Models are FKED with the three alternative second steps (OKFD, CTKFD and FKTM),  $K = 116, 129, 142$  and  $\eta = 0$ , and Model A1 in [2]. OKFD\_116.0 corresponds to OKFD with  $(K, \eta) = (116, 0)$  and so on.

at the 10 validation sites also with CTKFD and FKTM and some values of  $K$  in order to confirm the cross-validation results in comparing the three alternatives. Indeed, leave-one-out cross-validation is sometimes criticized in spatial modeling because the estimated spatial structure could change every time that a site is left out and, in addition, we experienced numerical problems during cross-validation with CTKFD and FKTM. By considering the whole dataset and predicting at the 10 validation sites, we obtain performance index values that allow us to compare OKFD, CTKFD and FKTM.

Moreover, a table similar to Table 2 for Model A1 in [2] is summarized through the last boxplots, where Model A1 is a spatio-temporal hierarchical model with a purely spatial covariance function (the residual spatio-temporal process is serially independent) fitted by MCMC on the same dataset. Note that cross-validation requires re-fitting the model for each left-out datum and becomes practically unfeasible when a model is fitted by MCMC; this is why here we compare FKED models and Model A1 on the 10 validation sites.

Boxplots in Figure 7 confirm that results with a number of knots equal to 116, 129 and 142 (and so 120, 133 and 146 basis functions) are very similar for

OKFD, with a light better performance when  $K = 142$  as known from the cross-validation results in Section 3.2. Note that RMSE is just a transformation of  $FCVr$ , so that they are equivalent preference criteria. Also in the case of CTKFD the number  $K = 142$  seems to be preferable, whereas it does not for FKTM (see e.g. Pearson boxplots). In comparing the three second step alternatives, this analysis with 10 validation sites confirms that OKFD is generally preferable, as seen by means of  $FCVr$  in Section 3.2. Only NMBF seems to say that it is overall worst but the only difference is at site 30 with  $NMBF \cong -0.13$  not highlighted as outlier, whereas  $NMBF \cong -0.16$  for CTKFD and FKTM.

With the same data, and hence the same 10 validation sites, six different hierarchical models are compared in [2]: Model A1 is suggested as preferable for its good performance at a reasonable cost because Model C has a better prediction performance but requires additional computational costs, since it includes an autoregressive component (as already said in Section 1). By looking at boxplots in Figure 7 we can compare FKED models with Model A1 and see that performances are not so different. In fact, if we exclude the “critical” site 30 *Saliceto* model OKFD\_142.0 has RMSE ranging from 0.196 to 0.481 while for Model A1 RMSE’s range is (0.215, 0.527). Analogously, without site 30 Pearson correlation ranges are (0.762, 0.943) and (0.779, 0.958) for OKFD\_142.0 and A1, respectively; as well as WNNR ranges are (0.002, 0.018) and (0.003, 0.013). Moreover, for both models  $|NMBF| > 0.05$  in 3 out of 10 sites. Hence overall OKFD\_142.0 performance is comparable to A1’s one except that for site 30 and for its ability to predict observed peaks. However a worst WNNR was expected since observed raw data are smoothed when converted to functional data, so that a part of observed peaks is smoothed away as observational noise.

To give another bit of comparison, let us note that with the same dataset Cameletti *et al.* (2012) [3] consider a spatio-temporal model with an autoregressive component (very similar to Model C in [2]), approximated by a Gaussian Markov Random Field and fitted by adopting the INLA algorithm, and obtain a global RMSE equal to 0.5328 and the correlation coefficient equal to 0.7015 when the 10 validation sites are taken altogether.

#### 4 Large sample behaviour

The asymptotic theory for spatial prediction has been object of study during the last decades. Prediction based on kriging has been scrutinized in a number of papers. In the mid eighties, Molnar (1985) [20] shows some mathematical properties of the classical and universal Kriging method investigating certain conditions for the convergence of these methods as the number of observation points tends to infinite. Lahiri *et al.* (2002) [19] consider the least-squares approach for estimating parameters of a spatial variogram and establish consistency and asymptotic normality of these estimators under general conditions. Large-sample distributions are also established under a spatial regression model where the sampling design possibly has an infill sampling component. Crujeiras and Van Keilegom (2010) [4] study the asymptotic and finite sample properties of an estimator of a nonlinear regression function when errors are spatially correlated, and when the spatial dependence structure is unknown. Vazquez and Bect (2010) [35] and Sakata *et al.* (2010) [32] deal with several issues related to the pointwise consistency of the

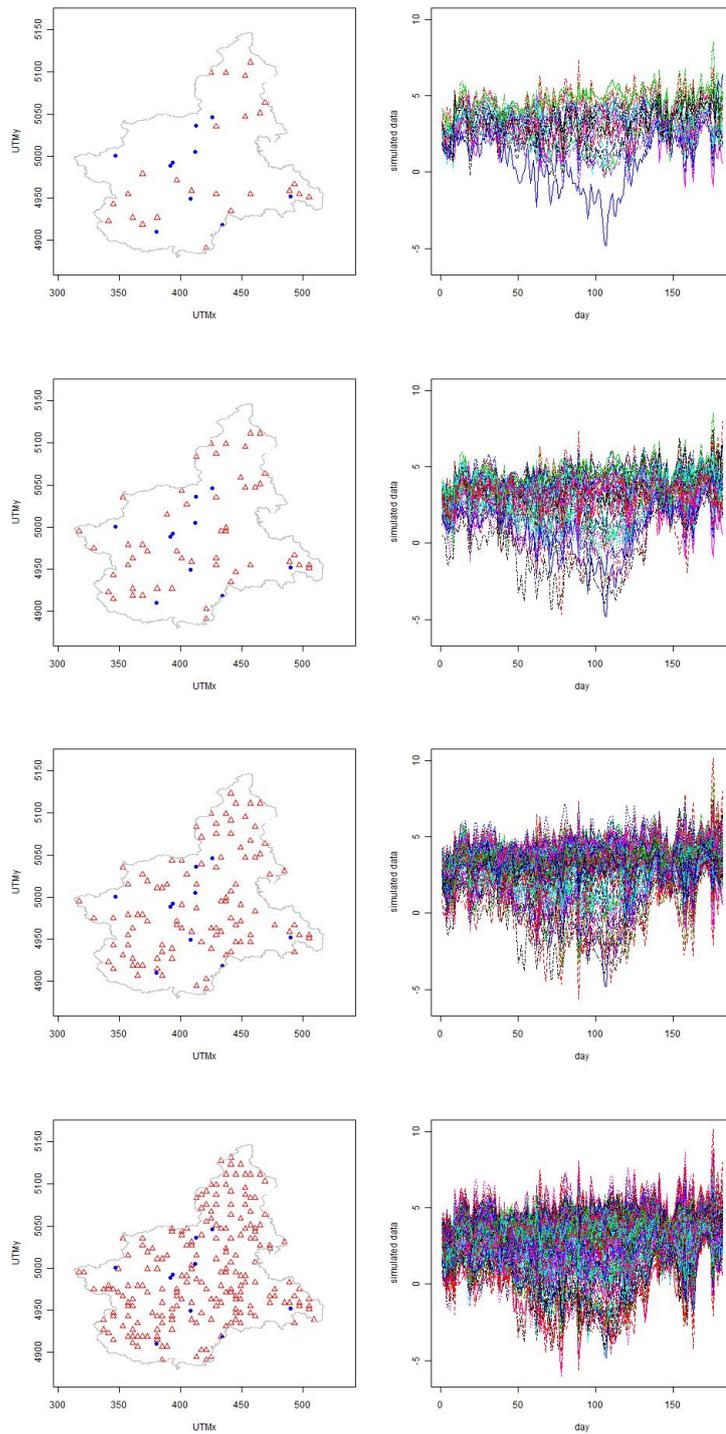
kriging predictor when the mean and the covariance functions are known. These questions are of general importance in the context of computer experiments. The analysis is based on the properties of approximations in reproducing kernel Hilbert spaces.

When we look at the more complete picture of the spatio-temporal context, we can hardly find information on kriging asymptotics. For example, Zhang and Zheng (2012) [43] study the asymptotic properties of maximum likelihood estimates under a general asymptotic framework for spatial-temporal linear models. Finally, we should add that the mixed field of functional and spatio-temporal prediction is completely open to asymptotic properties. At the best of our knowledge nothing can be found in this line. This would be object of a completely new paper focused on this topic. Thus we have preferred to show some light in this regard based on simulations.

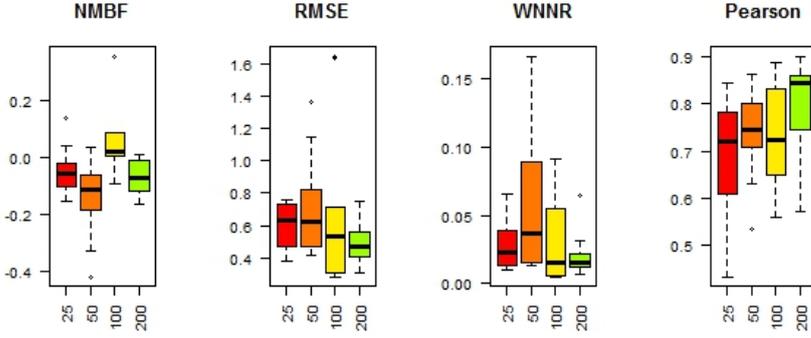
Hence, in order to explore the large sample behavior of the proposed three-steps predictor, we carry out a simulation study where the real dataset - described in Section 3 - is used as a basis for generating realistic simulated data. First of all, we consider Piemonte region as the spatial domain where we randomly select spatial sites whose coordinates belong to a regular grid with resolution  $4km \times 4km$  that covers Piemonte, neighbor Italian regions and parts of foreign countries (4032 grid points). Such a regular grid defines the spatial support of the output of a numerical model implemented by the environmental agency *ARPA Piemonte* (as already said in Section 3.1) that provides us with time-varying, meteorological and emission, covariates. The grid points on Piemonte region are 1587 and among them we randomly select 200 sites, in total, and we then consider different sample sizes: 25, 50, 100 and 200. Figure 8 (left column) shows the selected sites, increasing from 25 to 200, and the 10 validation sites from the real dataset in the previous section.

To obtain realistic data from a non-stationary spatial functional process, following the model (1), we simulate separately a drift and a stationary residual process. The drift functions are created by means of Equation (3) with functional coefficients equal to those estimated from the real dataset and shown in Figure 5, whereas the scalar and functional covariates are known. Instead, to obtain realizations of a functional random field we apply Equation (4)-left, with the spline coefficients simulated as realizations of a multivariable random field. This last one is simulated by means of `predict.gstat` in `gstat` package [22] and its variograms and cross-variograms are those estimated for the real dataset, when a LMC is fitted to apply CTKFD or FKTM, with  $K = 142$ . The four (nested) functional dataset are shown in Figure 8 (right column).

Therefore we apply our proposal with OKFD at the second step (since in Section 3 we have seen that it has a better performance at a lower computational cost) and we obtain predictions at the 10 validation sites (blue points in Figures 1 and 8). When comparing predicted functions with the observed raw data at these 10 sites, we get values of the four performance indexes described in Section 3.3 that are summarized by boxplots in Figure 9 for all the considered sample sizes. It is evident that the larger the sample size, the lower the variability of the four measures reflected in the green boxplots (for  $n = 200$ ), except for the Pearson case that is not that clear. Indeed, the median values of RMSE and WNNR obtained with  $n = 200$  are lower than those obtained from other sample sizes, while the Pearson median for  $n = 200$  is the highest one.



**Fig. 8** Left: locations of  $n$  randomly selected sites - red triangles - and 10 real validation stations - blue dots. Right: simulated functional dataset at  $n$  sites. Sample size  $n = 25, 50, 100, 200$  from top to bottom.



**Fig. 9** Boxplots of the performance measure distributions computed for each sample size over the 10 validation stations (blue dots in Figure 1 and 8). Predictions are obtained with OKFD at the second step and  $(K, \eta) = (142, 0)$ .

## 5 Discussion

We can obtain in practice realizations of a *multivariate functional random field*  $\{\mathbf{Y}_s(t), s \in D \subset \mathbb{R}^d, t \in T\}$ . Given  $s_i \in D, i = 1, \dots, n$ , and  $p$  functional variables, we have the realization

$$\begin{aligned} \mathbf{Y}_s(t) &= [(Y_{s_1,1}(t), \dots, Y_{s_n,1}(t)), \dots, (Y_{s_1,p}(t), \dots, Y_{s_n,p}(t))]^T \\ &= [\mathbf{Y}_{s_i,1}(t), \dots, \mathbf{Y}_{s_i,p}(t)]^T. \end{aligned}$$

In this case we could be interested in predicting simultaneously the vector of random functions  $\mathbf{Y}_{s_0}(t) = (Y_{s_0,1}(t), \dots, Y_{s_0,p}(t))$ , based on all the information of  $\mathbf{Y}_s(t)$  available. For example suppose that we have several pollutants curves at each sampling site  $s_i, i = 1, \dots, n$  and we want to predict them at unmonitored sites. This scenario is the natural extension of multivariate geostatistics [36] to the multivariate functional geostatistics. In addition, if we consider scalar and functional covariates we would have the multivariate functional kriging with external drift. By extending (2) to the multivariate functional context, we have the model

$$\mathbf{Y}_{s_i}(t) = \mu_{s_i}(t) + \varepsilon_{s_i}(t),$$

with  $\mu_{s_i}(t) = (\mu_{s_i,1}(t), \dots, \mu_{s_i,p}(t))^T$ , and  $\varepsilon_{s_i}(t) = (\varepsilon_{s_i,1}(t), \dots, \varepsilon_{s_i,p}(t))^T, i = 1, \dots, n$ . Following the same idea proposed in this work, if we want to predict simultaneously a random vector of functional variables at the unmonitored site  $s_0$ , we need to estimate initially the vector of functional residuals  $\mathbf{e}_s(t) = \mathbf{Y}_s(t) - \mu_s(t)$  and then, in a second step, carry out prediction of a vector of functional residuals  $\hat{\mathbf{e}}_{s_0}(t) = (\hat{e}_{s_0,1}(t), \dots, \hat{e}_{s_0,p}(t))$ . Two problems must be solved to fulfill these tasks. First a *multivariate functional regression model* must be estimated and posteriorly a *multivariate functional kriging predictor* used for predicting the vector of functional residuals. At the best of our knowledge these topics have not been studied and are open research problems. We think that a possible solution could be obtained by using basis functions for smoothing the set of curves  $\mathbf{Y}_{s_i}(t)$ .

Thus, we can propose a classical multivariate regression model with the responses corresponding to the coefficients estimated from the smoothing process. In this case we would obtain a matrix of residuals which could be predicted by using multivariate geostatistics. This approach looks reasonable from a technical point of view. However the estimation of the covariance structure by means of a LMC (or any other method) is restrictive in practice even with a small number of responses and basis functions. This alternative deserves special attention.

In this work, we propose Kriging with External Drift for functional data that are curves along time. Thus, we have an alternative to spatio-temporal modeling capable to predict a whole curve and providing covariate nonlinear effects' estimates straightforwardly. Moreover covariates - and response too - can be observed with different time frequency, so that treating time series data as functional data can be advantageous because a possible time misalignment problem can be avoided.

Our proposal is not necessarily an alternative to spatio-temporal modeling and indeed it can be applied to functional data that do not vary in time, as for example PM vertical profiles measured along height by an instrument deployed on a tethered balloon or climate variables measured by means of high technology radiosondes launched in the atmosphere. It is relevant to mention that in these cases a multivariate kriging approach could be also considered for predicting  $e_{s_0}(t)$ . With this approach the vector  $(e_{s_i}(t_1), \dots, e_{s_i}(t_M))$ , with  $t_1, \dots, t_M$  corresponding to discrete values of the domain  $T$ , is the observation of a  $M$  dimensional random variable at site  $s_i, i = 1, \dots, n$ . Then a cokriging predictor [36] could be applied to predict the random vector  $(\hat{e}_{s_0}(t_1), \dots, \hat{e}_{s_0}(t_M))$  at the unmonitored site  $s_0$ . Then a parametric or nonparametric model could be fitted to these values for reconstructing a whole function  $\hat{e}_{s_0}(t)$ . This approach is restrictive when  $M$  is large (the common situation in functional data analysis) and the methods based on OKFD, CTKFD, and FKTM, which involve the use of basis functions for smoothing the data, are a better option from a practical point of view.

For estimating the functional spatial trend we propose a functional regression model but other alternatives such as functional nonparametric models [7] could be considered. On the other hand, we take advantage of additive models theory to carry out variable selection (if the need arises, in our case study covariates were previously selected [2]). Gromenko and Kokoszka (2013) [14] derive a test to determine the significance of the regression coefficients when the mean function is a linear combination of known covariate functions and depends only on time. Further research to develop inference for drift model selection for spatially correlated functional data is necessary.

For kriging functional residuals we use and compare three alternatives, namely OKFD, CTKFD and FKTM. In our case study, both cross-validation results and good predictions on the validation sites suggest to choose the simplest version where kriging coefficient are constant. Nevertheless, the desirable application of these methods to other datasets could reveal different performances.

Finally, it would be convenient to provide confidence bands for the predicted curves and, to this goal, a resampling procedure to evaluate the prediction uncertainty is part of our ongoing research.

## Acknowledgments

The authors are grateful to the Associate Editor and two anonymous referees whose comments and suggestions improved the reading and quality of the manuscript.

## References

1. Caballero, W., Giraldo, R. and Mateu, J. (2013). A universal kriging approach for spatial functional data. *Stochastic Environmental Research and Risk Assessment* DOI: 10.1007/s00477-013-0691-4.
2. Cameletti, M., Ignaccolo, R., and Bande, S. (2011). Comparing spatio-temporal models for particulate matter in Piemonte. *Environmetrics* **22**, 985-996.
3. Cameletti, M., Lindgren, F., Simpson, D. and Rue, H. (2012). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, **97(2)**, 109-131.
4. Crujeiras, R.M., and Van Keilegom, I. (2010). Least squares estimation of nonlinear spatial trends. *Computational Statistics and Data Analysis*, **54**, 452-465.
5. Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics* **21**, 224-239.
6. EEA (2012). Air quality in Europe 2012 report. Report No 4/2012, European Environment Agency.
7. Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2011). Kernel regression with functional response. *Electronic Journal of Statistics* **5**, 150-171.
8. Ferraty, F., and Vieu, P. (2006). Nonparametric functional data analysis: theory and practice, Springer-Verlag, New York.
9. Finardi, S., DeMaria, R., D'Allura, A., Cascone, C., Calori, G., Lollobrigida, F. (2008). A deterministic air quality forecasting system for Torino urban area, Italy. *Environ. Model. Softw* **23(3)**, 344-355.
10. Giraldo, R., Delicado, P., and Mateu, J. (2009). Geostatistics with infinite dimensional data: A generalization of cokriging and multivariable spatial prediction. *Reporte Interno de Investigacion No. 14*, Universidad Nacional de Colombia
11. Giraldo, R., Delicado, P., and Mateu, J. (2010). Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological, and Environmental Statistics* **15(1)**, 66-82.
12. Giraldo, R., Mateu, J., and Delicado, P. (2012). geofd: An R package for function-valued geostatistical prediction. *Revista Colombiana de Estadística* **35 (3)**, 383-405.
13. Giraldo, R., Delicado, P., and Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics* **18(3)**, 411-426.
14. Gromenko, O., and Kokoszka, P. (2013). Nonparametric inference in small data sets of spatially indexed curves with application to ionospheric trend determination. *Computational Statistics and Data Analysis* **59**, 82-94.
15. Hengl, T., Heuvelink, G. B. M., and Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. *Computers & Geosciences*, **33 (10)**, 1301-1315.
16. Horváth, L., and Kokoszka, P. (2012). Inference for Functional Data with Applications, Springer, New York.
17. Ivanescu, A.E., Staicu, A.M., Greven, S., Scheipl, F., and Crainiceanu, C. M. (2012). Penalized function-on-function regression (April 2012). Johns Hopkins University, Dept. of Biostatistics Working Papers. *Working Paper 240*. Available at <http://biostats.bepress.com/jhubiostat/paper240>
18. Kokoszka P. (2012). Dependent functional data. *ISRN Probability and Statistics*, doi:10.5402/2012/958254, 1-30.
19. Lahiri, S.N., Leea, Y., and Cressie, N. (2002). On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. *Journal of Statistical Planning and Inference*, **103**, 65-85.
20. Molnar, S. (1985). On the convergence of the kriging method. *Annales Univ. Sci. Budapest., Sect. Comp*, **6**, 81-90.
21. Nerini, D., Monestiez, P., and Mantè, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis*, **101**, 409-418.

22. Pebesma, E.J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* **30**, 683-691.
23. Poli, A. and Cirillo, M. (1993). On the use of the normalized mean square error in evaluating dispersion model performance. *Atmospheric Environment* **27**, 2427-2434.
24. R Development Core Team (2012). R: a Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, www.R-project.org.
25. Ramsay, J. O., and Silverman, B. W. (2002). Applied Functional Data Analysis, Springer-Verlag, New York.
26. Ramsay, J. O., and Silverman, B. W. (2005). Functional Data Analysis, Springer-Verlag, New York.
27. Ramsay, J. O., Hooker G., and Graves S. (2009). Functional Data Analysis with R and Matlab, Springer-Verlag, New York.
28. Ramsay, J. O., Wickham H., Graves S., and Hooker G. (2012). fda: Functional Data Analysis, R package version 2.3.2.
29. Ruiz-Medina, M. D. and R. Fernández-Pascual (2010). Spatiotemporal filtering from fractal spatial functional data sequence. *Stochastic Environmental Research and Risk Assessment* **24**, 527538.
30. Ruiz-Medina, M. D. and R. Salmerón (2010). Functional maximum-likelihood estimation of arh(p) models. *Stochastic Environmental Research and Risk Assessment* **24**, 131146.
31. Ruiz-Medina, M.D. (2012). New challenges in spatial and spatiotemporal functional statistics for high-dimensional data. *Spatial Statistics*, **1**, 82-91.
32. Sakata, S., Ashida, F., and Tanaka, H. (2010). Stabilization of parameter estimation for Kriging-based approximation with empirical semivariogram. *Computer Methods in Applied Mechanics and Engineering*, **199**, 1710-1721.
33. Salmerón, R. and M. D. Ruiz-Medina (2009). Multi-spectral decomposition of functional autoregressive models. *Stochastic Environmental Research and Risk Assessment* **23(3)**, 289297.
34. Temiyasathit, C., Kim, S.B., and Park, S.-K. (2009). Spatial prediction of ozone concentration profiles. *Computational Statistics and Data Analysis*, **53**, 3892-3906.
35. Vazquez, E., and Bect, J. (2010). Pointwise Consistency of the Kriging Predictor with Known Mean and Covariance Functions. In *mODa 9 Advances in Model-Oriented Design and Analysis*. Contributions to Statistics 2010, 221-228.
36. Ver Hoef, J. and Cressie, N. (1993). Multivariable spatial prediction *Mathematical Geology* **25(2)**, 219-240.
37. Wackernagel, H. (1995). Multivariable geostatistics: An introduction with applications. Berlin: Springer-Verlag.
38. Wood, S.N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *J. Amer. Statist. Ass.* **99**, 673-686.
39. Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* **73(1)**, 3-36.
40. Wood, S. N. (2012). mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation, R package version 1.7-22.
41. Wood, S. N. (2012). On p-values for smooth components of an extended generalized additive model. *Biometrika* doi: **10.1093/biomet/ass048**, 1-8.
42. Yu, S., Eder, B., Dennis, R., Chu, S., and Schwartz, S. (2006). New unbiased symmetric metrics for evaluation of air quality models. *Atmospheric Science Letters* **7**, 26-34.
43. Zhang, X., and Zheng, Y. (2012). A note on spatial-temporal lattice modeling and maximum likelihood estimation. *Statistics and Probability Letters*, **82**, 2145-2155.