University of Bergamo

Department of Management, Economics and Quantitative Methods

School of Doctoral Studies in Analytics for Economics and Business

XXIX cycle

Spatio-temporal processes for functional data with application in climate monitoring

A THESIS

presented by

Ferdinand Bertrand Ndongo

Spatio-temporal processes for functional data with application in climate monitoring

A THESIS

presented by
Ferdinand Bertrand Ndongo

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

SSD: SECS-S/02

Prof. Alessandro Fassò Thesis Advisor	Date
Prof. Emilio Porcu Thesis co-Advisor	Date
Dott. Francesco Finazzi Thesis co-Advisor	Date

Academic year 2015/2016

Contents

Abstract

The underlying idea of this thesis is to move from a spatio-temporal process to a functional spatio-temporal process. In order to do that it is important firstly to specify the spatio-temporal framework of our interest. For our purposes, we are interested in processes evolving in continuous space and discrete time. Formally that means the spatial processes of interest will be limited on geostatistical processes and the temporal processes of interest will be time series. Since the rise of functional data analysis as an important tool for statistical analysis of data observed as curves, many classical statistical tools have been extended to their functional versions in order to handle data observed as curves, namely functional data. The new concept here is that a record observed as a vector of finite number of points is considered as a single entity. This single entity is built as linear combination of coefficients and basis functions through non parametric statistical methods. The main question we have to answer is: how can we put together a geostatistical spatio-temporal process and the new concept of functional data in order to build a geostatistical spatio-temporal process for functional data. In other words, the motivation of this thesis is to build a dynamical spatio-temporal model for data observed as functions or curves and furthermore extend an already existing tool for mapping to its functional version.

Introduction

The underlying idea of this thesis is to move from a spatio-temporal process to a functional spatio-temporal process. In order to do that it is important firstly to specify the spatio-temporal framework of our interest. For our purposes, we are interested in processes evolving in continuous space and discrete time. Formally that means the spatial processes of interest will be limited on geostatistical processes and the temporal processes of interest will be time series. Since the rise of functional data analysis as an important tool for statistical analysis of data observed as curves, many classical statistical tools have been extended to their functional versions in order to handle data observed as curves, namely functional data. The new concept here is that a record observed as a vector of finite number of points is considered as a single entity. This single entity is built as linear combination of coefficients and basis functions through non parametric statistical methods. The main question we have to answer is: how can we put together a geostatistical spatio-temporal process and the new concept of functional data in order to build a geostatistical spatio-temporal process for functional data. In other words, the motivation of this thesis is to build a dynamic spatio-temporal model for data observed as functions or curves in order to extend an already existing tool for mapping like D-STEM of Finazzi & Fassó (2014) to handle functional data.

Geostatistical spatio-temporal processes and functional data analysis have already been well described by Cressie (1993) and Ramsay (2006), respectively. In literature we can also find some works putting them together as for example Giraldo & Mateu (2013), and Ignaccolo et al. (2014) but using different approaches. Those works consider the functions observed along the time domain and the related observed functional data are just spatially correlated functional data. In that way, the temporal dynamics are hidden by the function and classical geostatistical techniques are computed on the functional data.

In order to take into account the spatio-temporal interaction between functions, we have to consider another dimension for the function domain as for example height or

depth. This dimension will be hide since the data are considered as a single entity along the function domain, and this will lead to spatio-temporally correlated functional data.

Our proposal in this functional spatio-temporal framework can be resumed in a three step procedure:

$$\text{pre-processing} \rightarrow \text{functional modelling} \rightarrow \text{functional prediction.}$$

(i) Pre-processing: a non-parametric technique is used in order to smooth the data. The spatio-temporal functional data are built along the considered function domain as linear combination of coefficients and B-splines basis functions.

(ii) Functional modelling: since the basis functions are the same for all the observations along the function domain, moving from vectors of observed points to their corresponding functional data can be seen as a linear transformation through the B-spline basis functions. Therefore describing the spatio-temporal interaction between the built functional data can be done through the spatio-temporal description of their coefficients in a Gaussian multivariate dynamic coregionalization model.

(iii) Functional prediction: the spatio-temporal spline coefficients are predicted at unobserved spatio-temporal locations by the means of spatio-temporal kriging technique and used for the reconstruction of the predicted spatio-temporal functions through the B-spline basis functions linear transformation.

According to this three step procedure some existing work in literature can be more or less similar to our approach in a step and different in another one. From our point of view, despite the pre-processing technique and the functional modelling are not the same, Temiyasathit et al. (2009) seems to be the approach closer to ours in our knowledge. The similarity stands here in the idea to extract the coefficients in the pre-processing step with a non-parametric method and use the predicted coefficients for the reconstruction of the predicted function.

This thesis is made up of four chapters. The first two are dedicated to the necessary background in spatio-temporal geostatistics and functional data analysis, respectively. Our goal in these two chapters is not to treat the respective topics in an exhaustive way, but to give an overview of their core concepts in order to understand how they will be considered together afterwards. Geostatistics for functional data is treated in the third chapter firstly from a classical approach existing in literature, and then from our proposal approach. The fourth chapter is a case study on atmospheric vertical profile of temperature from global time series of radiosondes given by the RAOB network (www.raob.com). Finally we will conclude with a discussion about the proposed model and some considerations about future perspectives.

Spatio-temporal processes for functional data

with application in climate monitoring

I. Spatio-temporal geostatistics

As a starting point we have the observed geostatistical data y_s identified as spatial data with a continuous variation. Following Cressie (1993), let $s \in \mathbb{R}^d$ be a generic location in a d-dimensional Euclidean space and

$$\{\chi_s, s \in \mathbb{S} \subset \mathbb{R}^d\} \tag{1.1}$$

be a random field (rf) (spatial random function or spatial random process) , χ denoting the attribute of interest.

Formally, geostatistical data y_s supposed to be a realization of process (1.1) arise when, (i) the domain under study is a fixed continuous set $\mathbb{S} \subset \mathbb{R}^d$ and χ_s can be observed at any point of the continuous domain; and (ii) the points in $\mathbb{S}$ are non-stochastic in the sense that once fixed, $\mathbb{S}$ is the same for all the realizations of the *rf*. Depending on the coordinate system of the data, two options are available, namely $s \in \mathbb{S} \subset \mathbb{R}^2$ or $s \in \mathbb{S}^2$, where $\mathbb{S}^2$ is the sphere in $\mathbb{R}^3$.

With y_s being meteorological, environmental and/or monitoring data collected in a set of N fixed monitoring network stations denoted by $s_i = 1, \ldots, N$ and at T time points, we have now continuous space discrete time data better denoted by $y_{s_i,t}$ with $t = 1, \ldots, T$.

In this case we have to consider how the phenomenon of interest evolves in both space and time rather than only considering its spatial distribution at a given time or its evolution over time at a given location. But we cannot consider the spatial and temporal aspects in the same way because there are significant differences between the space and

time axes like: (i) the scales of distance; (ii) the units of distance; (iii) time axis is ordered (past, present and future) while the space axis is not.

In order to understand how geostatisticians handle the space-time interaction incorporated in data, we will present separately the main characteristics and most relevant models of geostatistical processes and temporal processes, respectively. Then we will present some properties of spatio-temporal geostatistics in continuous time and a model of spatio-temporal geostatistics in discrete time.

1.1 Geostatistical processes

A geostatistical process is a spatial random process (1.1) whose parameters are defined directly through first and second moments (mean, variance and covariance). Geostatistics is mostly concerned with building models of spatial dependence captured by the covariance function and with predicting the spatial process optimally. Here we are going to review the ways that spatial dependence has been incorporated into spatial statistical models through a model-based measure of the spatial statistical dependence in a geostatistical process, called *variogram*. Our goal is not to review the whole field in an exhaustive way, but to establish enough spatial methodology so that our subsequent expositions on spatio-temporal statistical methodology can be adequately explained. The approach followed here is taken from Cressie (1993).

1.1.1 Description of spatial dependence.

1.1.1.1 Second order stationarity. The spatial random process (1.1) is said to be *second order stationary* if its mean and variance are constant over the space and its covariance function is only function of h the distance in space. It can be represent by the model

$$\chi_s = \mu + \epsilon_s, \tag{1.2}$$

where

$$\mathbb{E}(\chi_s) = \mu, \text{ for all } s \in \mathbb{S}, \tag{1.3}$$

is the mean of the process and ϵ_s is a random vector with zero mean and covariance

$$\mathbb{C}_h = \mathbb{E}\left[(\chi_s - \mu)(\chi_{s+h} - \mu)\right], \text{ for all } s, s+h \in \mathbb{S}, \tag{1.4}$$

depends on h the separation between samples in distance and direction and it is a function of h.

1.1.1.2 Stationary covariance function. When restrictions (1.3) and (1.4) are satisfied, the quantity $\mathbb{C}$ is called the stationary covariance function, which must satisfy the non-negative definiteness condition,

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \bar{\alpha}_j \gamma_{s_i - s_j} \geq 0, \tag{1.5}$$

for any positive integer N, any set of spatial locations $\{s_i : i = 1, \ldots, N\}$, and any set of complex numbers $\{\alpha_i : i = 1, \ldots, N\}$. Any covariance function that satisfies (1.5) is referred as *valid*. The condition (1.5) guarantees that model-based prediction variances are non negative. Notice that the non-negative definiteness condition (1.5) has to hold for all complex numbers $\{\alpha_i\}$. Hence, there will be some processes with valid stationary variogram that do not have valid stationary covariance functions.

When $\mathbb{C}_h$ can be written as a function of $||h||$, we say that it is isotropic; otherwise it is said to be anisotropic.

1.1.1.3 Intrinsic stationarity. The spatial random process (1.1) is said to be *intrinsic stationary* if the expected differences are zero

$$\mathbb{E}(\chi_s - \chi_{s+h}) = 0. \tag{1.6}$$

The class of intrinsically stationary processes contains the class of second-order stationary processes. Assuming only the existence of the stationary covariance function given by (1.4), the semivariogram (i.e., one half the variogram) exists and is given by

$$\gamma_h = \mathbb{C}_0 - \mathbb{C}_h, \ h \in \mathbb{R}^d.$$

1.1.1.4 Stationary variogram. When the spatial random process (1.1) is intrinsic stationary (1.6), the covariance of the residuals (1.4) is replaced by the variance of the differences to measure spatial variation

$$\mathbb{V}(\chi_s - \chi_{s+h}) = \mathbb{E}(\chi_s - \chi_{s+h})^2 = 2\gamma_h, \text{ for all } s, s+h \in \mathbb{S}. \tag{1.7}$$

The quantity $2\gamma_h$, which is function of only the difference between the spatial locations s and $s+h$, is called the stationary variogram. The variogram must satisfy the conditional non-positive definiteness condition

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \bar{\alpha}_j \gamma_{s_i - s_j} \leq 0, \tag{1.8}$$

for any positive integer N, any set of spatial locations $\{s_i : i = 1, \ldots, N\}$, and any set of complex numbers $\{\alpha_i : i = 1, \ldots, N\}$ satisfying $\sum_{i=1}^{N} \alpha_i = 0$. (If α is a complex number, then recall that $\bar{\alpha}$ denotes its complex conjugate). The condition (1.8) is extremely important, because it guarantees that all model-based variances are non-negative. Any variogram that satisfies (1.8) is referred to as *valid*.

When $2\gamma_h$ can be written as a function of $||h||$, for $h = (h_1, \ldots, h_d)' \in \mathbb{R}^d$, where $||h|| \equiv (h_1^2 + \cdots + h_d^2)^{1/2}$, the variogram is said to be isotropic; otherwise it is said to be anisotropic.

1.1.2 Example of variogram models. Variogram models that depends on only a few parameters, $\boldsymbol{\theta}$, can be used as succinct summaries of the spatial dependence. Variogram is an important component of optimal spatial linear prediction and its half called semivariogram is expressed as

$$\gamma_h(\theta) = \mathbb{C}_0(\theta) - \mathbb{C}_h(\theta); h \in \mathbb{R}^d. \tag{1.9}$$

1.1.2.1 Matérn. The Matérn semivariogram is expressed as (1.9) where

$$\mathbb{C}_h(\theta) = \sigma_0^2 I_{(||h||=0)} + \sigma_1^2 \{2^{\theta_2 - 1}\Gamma(\theta_2)\}^{-1} \{||h||/\theta_1\}^{\theta_2} K_{\theta_2}(||h||/\theta_1), \tag{1.10}$$

is the Matérn covariance function; $\mathbb{C}_0(\theta) = \sigma_0^2 + \sigma_1^2$, K_{θ_2} is a modified Bessel function of the second kind of order θ_2 and $\boldsymbol{\theta} \in \{(\theta_1, \theta_2, \sigma_0^2, \sigma_1^2)' : \theta_1 > 0, \theta_2 > 0, \sigma_0^2 \geq 0, \sigma_1^2 \geq 0\}$.

1.1.2.2 Powered-exponential. Another common isotropic semivariogram used is the *powered-exponential* where in (1.9)

$$\mathbb{C}_h(\theta) = \sigma_0^2 I_{(||h||=0)} + \sigma_1^2 \exp\{-(||h||/\theta_1)^{\theta_2}\}, h \in \mathbb{R}^d, \tag{1.11}$$

and $\boldsymbol{\theta} \in \{(\theta_1, \theta_2, \sigma_0^2, \sigma_1^2)' : \theta_1 > 0, 0 < \theta_2 \leq 2, \sigma_0^2 \geq 0, \sigma_1^2 \geq 0\}$.

1.1.2.3 Exponential. The most familiar member of the class is the exponential covariance function obtained when $\theta_2 = 1$:

$$\mathbb{C}_h(\theta) = \sigma_0^2 I_{(||h||=0)} + \sigma_1^2 \exp\{-(||h||/\theta_1)\},\ h \in \mathbb{R}^d. \tag{1.12}$$

When the variogram is anisotropic, it is no longer a function purely of distance between the two spatial locations. Anisotropies are caused by underlying physical process evolving differentially in space. That is, different (temporal) dynamics in different directions can cause the process to be anisotropic in space. Sometimes the anisotropy can be corrected by an invertible linear transformation of the lag vector h.

1.1.3 Kriging. Kriging is a spatial prediction methodology due to Matheron (1963) which predicts values at unvisited sites from sparse sample data based on a stochastic model of continuous spatial variation as represented in the variogram or covariance function. It is a best linear unbiased predictor in the sense that its prediction error variances are minimized. It is in practice a weighted moving average in which the weights depend on the variogram and the configuration of the sample points within the neighbourhood of its targets. Ordinary kriging is by far the most popular method, partly because it is robust with respect to departures from the underlying assumptions.

Let (1.1) be a spatial random process satisfying $0 < \mathbb{V}(\chi_s) < \infty$, for all $s \in \mathbb{S}$. The data $\mathbf{y}$ are noisy, incomplete observations on χ, with measurement error ϵ that has mean zero and variance $\sigma^2_\epsilon \in [0, \infty)$. It is simplest to assume for the moment that the mean function,

$$\mu_s = \mathbb{E}(\chi_s),\ s \in \mathbb{S},$$

is known and that the covariance function,

$$\mathbb{C}_{s,s'} = cov(\chi_s, \chi_{s'}),\ s, s' \in \mathbb{S},$$

and the measurement-error variance, σ^2_ϵ, are known. Notice that here, $\mathbb{C}_{s,s'}$ is not necessarily a function of $(s-s')$; we are attempting to be as general as possible. The (non-stationary) variogram is derivable as $2\gamma_{s,s'} = \mathbb{V}(\chi_s - \chi_{s'}) = \mathbb{C}_{s,s} + \mathbb{C}_{s',s'} - 2\mathbb{C}_{s,s'}$.

1.1.3.1 Ordinary kriging. In most practical situations the mean is not known. Ordinary kriging is based on the assumption that variation is random and spatially dependent, and that the underlying process is intrinsically stationary with constant mean and a variance that depends only on separation in distance and direction between places and not on absolute position.

We predict χ at a new point, s_0, by

$$\hat{\chi}_{s_0} = \sum_{i=1}^{N} \lambda_i \chi_{s_i}, \tag{1.13}$$

where λ_i are the weights. To ensure that estimate is unbiased the weights are made to sum to 1:

$$\sum_{i=1}^{N} \lambda_i = 1.$$

The expected difference is $\mathbb{E}(\hat{\chi}_{s_0} - \chi_{s_0}) = 0$, and the prediction variance is given by

$$\begin{aligned} \mathbb{V}(\hat{\chi}_{s_0}) &= \mathbb{E}\left[(\hat{\chi}_{s_0} - \chi_{s_0})^2\right] \\ &= 2\sum_{i=1}^{N} \lambda_i \gamma_{s_i - s_0} - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \gamma_{s_i - s_j}, \end{aligned} \tag{1.14}$$

where the quantity $\gamma_{s_i - s_0}$ is the semivariance of χ between the sampling point s_i and the target point s_0; and $\gamma_{s_i - s_j}$ is the semivariance between the ith and jth sampling points. The semivariances are derived from the variogram model, partly because there is no measure of the semivariances between the data points and the target points where we have no observed values and partly because only by doing so we can guarantee that the variances are not negative.

The next step in kriging is to find the weights that minimize the kriging variances subject to the constraint that they sum to 1. Equations (1.13) and (1.14) lead to a set of $N+1$ equations in the $N+1$ unknowns:

$$\begin{aligned} \sum_{i=1}^{N} \lambda_i \gamma_{s_i - s_j} + \psi_{s_0} &= \gamma_{s_j - s_0}, \text{ for all } j \\ \sum_{i=1}^{N} \lambda_i &= 1. \end{aligned} \tag{1.15}$$

The quantity ψ_{s_0} is a Langrange multiplier introduced to achieve minimization.

The solution of the kriging equations provides the weights in (1.13), and the prediction variance can be obtained as

$$\sigma^2_{s_0} = \sum_{i=1}^{N} \lambda_i \gamma_{s_i - s_0} + \psi_{s_0}. \tag{1.16}$$

1.1.3.2 Simple kriging. Simple kriging is base on the assumption of second-order stationary with known constant mean. The kriged predictions are linear sums of data:

$$\hat{\chi}_{s_0} = \sum_{i=1}^{N} \lambda_i \chi_{s_i} + \{1 - \sum_{i=1}^{N} \lambda_i\}\mu. \tag{1.17}$$

The inclusion of the second term on the right-hand side of (1.17) ensures that the predictions are unbiased.

The simple kriging variance is given by

$$\sigma^2_{s_0} = \mathbb{C}_0 - \sum_{i=1}^{N} \lambda_i \mathbb{C}_{s_i, s_0}, \tag{1.18}$$

where $\mathbb{C}_0$ is the variance of the random process.

1.1.3.3 Universal kriging. For universal kriging it is assumed that the mean is unknown but it is assumed that the mean can represented a linear combination of some surface function

$$\mu_s = \mathbb{E}(\chi_s) = \sum_{p=0}^{P} \beta_p f_p, \tag{1.19}$$

where β_p are unknown coefficients to be estimated and f_p are known basis functions. Usually the first basis function (case $p = 0$) is the constant function identically equal to 1, which guarantees that the constant-mean case is included in the model. The other functions are typically monomials of low degree in the coordinate of s (in practice, the degree does not exceed two).

The universal kriging model is the decomposition $\chi_s = \mu_s + \epsilon_s$ of a smooth deterministic function μ_s describing the systematic aspect of the phenomenon, and called drift, and a zero-mean process ϵ_s, called the residual capturing its erratic fluctuations. An important development of the universal kriging model is the use of external variables to model the drift function μ_s.

1.2 Temporal processes

Many environmental processes are inherently dynamic, and understanding the temporal variability is of primary importance. We consider temporal variability due to randomness and that if enough observations are taken, it will average (in some well defined way) to a central tendency or to a temporal trend. We are interested in characterizing and modelling such temporal variability following the approach in Cressie (1993).

A temporal process can be written as

$$\{\chi_t : t \in \mathcal{T}\}, \tag{1.20}$$

where t indexes the time of the possibly multivariate (deterministic or stochastic) process $\chi(\cdot)$ and $\mathcal{T}$ is a subset of $\mathbb{R}$. We can distinguish three types of temporal process according to the nature of $\mathcal{T}$: (i) a *continuous-time process*, $\mathcal{T}$ is fixed and has non zero length in the continuous interval $[0, \infty)$; (ii) a *discrete-time process* (time series), $\mathcal{T}$ is a fixed index set of a finite or countable set of times $\{1, 2, \ldots, T\}$; (iii) a temporal point process $\mathcal{T}$ is a random set made up of randomly occurring points (events) in $\mathbb{R}^1$.

Since we are mainly interested on discrete-time process, in this section we assume $\mathcal{T} = \{1, 2, \ldots, T\}$ and we refer $\{\chi_t : t = 1, 2, \ldots, T\}$ as time series.

1.2.1 Description.

1.2.1.1 Joint and conditional distributions.

Let consider a time series given by $\{\chi_t : t = 1, \ldots, T\}$. We denote its joint distribution by

$$[\chi_1, \ldots, \chi_T]. \tag{1.21}$$

In high dimensions it can be very difficult to fully specify (1.21) because it requires to specify all possible interactions between $\chi_1, \ldots, \chi_T$. Therefore it is useful to think about the joint distribution from a conditional perspective and write it as a product of conditional distributions:

$$[\chi_1, \ldots, \chi_T] = [\chi_T | \chi_{T-1}, \ldots, \chi_1][\chi_{T-1} | \chi_{T-2}, \ldots, \chi_1] \ldots [\chi_2 | \chi_1][\chi_1]. \tag{1.22}$$

For practical modelling, we usually make additional assumptions about the components of (1.22) as for example we might assume that the process can be modelled by a first-order Markov property,

$$[\chi_t | \chi_{t-1}, \ldots, \chi_1] = [\chi_t | \chi_{t-1}], \text{ for all } t = 1, 2, ...T. \tag{1.23}$$

This property suggests a "lack of memory", so that only the most recent past determines the conditional probabilities about the present given the whole past. In this case, (1.23) becomes

$$[\chi_1, \ldots, \chi_T] = [\chi_1] \prod_{t=2}^{T} [\chi_t | \chi_{t-1}]. \tag{1.24}$$

In a stochastic or random process, we might say that the future is only partially determined from the past and model (1.20) as

$$\chi_t = G(\chi_{t-1}) + \eta_t, \ t = 1, 2, \ldots, T, \tag{1.25}$$

where $\{\eta_t : t = 1, 2, \ldots, T\}$ is a mean-zero random process and η_t is statistically independent of χ_{t-1}.

1.2.1.2 First and second moment functions. The mean function of (1.20) is defined as

$$\mu_t = \mathbb{E}(\chi_t), t \in \mathcal{T}. \tag{1.26}$$

It is usually assumed that μ_t is deterministic and the remaining stochastic component $\{\chi_t - \mu_t, t = 1, \ldots, T\}$ is related to the uncertainty. This uncertainty can be described through the autocovariance function or covariogram,

$$\mathbb{C}_{t,t'} = \mathbb{C}(\chi_t, \chi_{t'}) = \mathbb{E}\{(\chi_t - \mu_t)(\chi_{t'} - \mu_{t'})\}, \ t, t' \in \mathcal{T}, \tag{1.27}$$

which summarizes how the process co-varies across different time lags, after accounting for the mean function.

The variance is a special case of the autocovariance in which $\mathbb{C}_{t,t} = \mathbb{V}(\chi_t) = \sigma_t^2$ and one can obtain the autocorrelation through normalization:

$$\rho_{t,t'} = \frac{\mathbb{C}_{t,t'}}{\sqrt{\mathbb{C}_{t,t}\mathbb{C}_{t',t'}}}, \ t, t' \in \mathcal{T}, \tag{1.28}$$

where $\rho_{t,t'} \in [-1, 1]$. A sufficient condition for (1.27) and (1.28) to exist is $\sigma_t^2 < \infty$, for all $\mathcal{T}$.

1.2.1.3 Property of stationarity. A commonly made assumption for time series is that of stationarity.

A temporal process $\{\chi_t\}$ is *strong stationary* if any collection $\{\chi_t, \ldots, \chi_T\}$ of random variables from the time series has exactly the same joint distribution as $\{\chi_{t+\tau}, \ldots, \chi_{T+\tau}\}$, for any $\tau \in \{1, \ldots, T\}$.

A temporal process $\{\chi_t\}$ with finite variance $\mathbb{V}(\chi_t) = \sigma_t^2 < \infty$ for all $t \in \mathcal{T}$, is said to be *second-order stationary* if

(*i*) $\mathbb{E}\{\chi_t\} = \mu$, for all $t \in \mathcal{T}$.

(*ii*) $\mathbb{V}\{\chi_t, \chi_{t'}\} = \mathbb{C}_{t-t'}$, for all $t, t' \in \mathcal{T}$.

Second-order stationarity says that any pair, $\{\chi_t, \chi_{t'}\}$, has exactly the same first and second moments (including the cross moments, which defines the autocovariance function) as the pair $\{\chi_{t+\tau}, \chi_{t'+\tau}\}$ for any $\tau \in \{1, \ldots, T\}$. For processes with finite variance, strong stationarity implies second-order stationarity, but not vice versa.

Even if μ_t given by (1.26) is not constant, we can define the autocovariance function to be stationary if the covariance function given by (1.27) is equal to $\mathbb{C}_{t-t'}$, for all $t, t' \in \mathcal{T}$.

In the context of a multivariate time series $\{\chi_t\}$, we define $\{\boldsymbol{\chi}_t^{(i)}\}$ to be the ith time series at time t.

The cross-covariance function between the ith time series at time t and the jth time series at time t' is defined as

$$\mathbb{C}_{t,t'}^{i,j} = \mathbb{C}(\chi_t^{(i)}, \chi_{t'}^{(j)}),\ t, t' \in \mathcal{T}. \tag{1.29}$$

A second-order stationary assumption on the multivariate time series $\{\boldsymbol{\chi}_t\}$ requires that the respective mean functions do not vary with time and that the cross-covariance functions only depend on the lag between times t and t'. That is,

$$\mathbb{C}_{\tau}^{i,j} = \mathbb{C}(\chi_t^{(i)}, \chi_{t+\tau}^{(j)}),\ \tau = 1, \ldots, T. \tag{1.30}$$

Note that $\mathbb{C}_{-\tau}^{i,j} = \mathbb{C}_{\tau}^{j,i}$.

The cross-correlation function is given by

$$\rho_{\tau}^{i,j} = \frac{\mathbb{C}_{\tau}^{i,j}}{\sqrt{\mathbb{C}_0^{i,i}\mathbb{C}_0^{j,j}}},\ \tau = 1, \ldots, T. \tag{1.31}$$

1.2.1.4 Empirical mean and empirical covariance function. Let $\{\chi_t\}$, denoted by $\chi_1, \ldots, \chi_T$, be a finite sample from a discrete-time stochastic process.

The *empirical mean* of the time series is

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} \chi_t. \tag{1.32}$$

The *empirical autocovariance function* is

$$\hat{C}_{\tau} = \frac{1}{T}\sum_{t=1}^{T-\tau} (\chi_{t+\tau} - \hat{\mu})(\chi_t - \hat{\mu}),\ \tau = 1, \ldots, T-1. \tag{1.33}$$

Note that $\hat{C}_{-\tau} = \hat{C}_{\tau}$ and $\hat{C}_0$ is the empirical variance.

The *empirical autocorrelation function* is

$$\hat{\rho}_\tau = \frac{\hat{C}_\tau}{\hat{C}_0},\ \tau = 0, 1, \ldots, T-1. \tag{1.34}$$

In multivariate context, the observations on the ith time series are denoted by $\chi_1^{(i)}, \ldots, \chi_T^{(i)}$ and the *empirical cross-covariance function* is

$$\hat{C}_\tau^{i,j} = \frac{1}{T}\sum_{t=1}^{T-\tau}(\chi_{t+\tau}^{(i)} - \hat{\mu}_i)(\chi_t^{(j)} - \hat{\mu}_j),\ \tau = 1, \ldots, T-1, \tag{1.35}$$

where $\hat{\mu}_i$ is the empirical mean (1.32) for the ith time series and likewise for $\hat{\mu}_j$. We note that $\hat{C}_{-\tau}^{i,j} = \hat{C}_\tau^{j,i}$.

The empirical cross-correlation function is obtain from (1.31) after substituting the respective empirical covariance and cross-covariance functions from (1.35). That is,

$$\hat{\rho}_\tau^{i,j} = \frac{\hat{C}_\tau^{i,j}}{\sqrt{\hat{C}_0^{i,i}\hat{C}_0^{j,j}}},\ \tau = 0, 1, \ldots, T. \tag{1.36}$$

1.2.2 Time series models. In order to simplify the characterization of a time series, some probabilistic properties are commonly specified.

1.2.2.1 White-noise process. A white-noise process is defined as a discrete-time stochastic process $\{\epsilon_t : t = 1, \ldots, T\}$ whose elements are mutually independent and have a common probability density function (*pdf*). Typically, its mean μ_ϵ is assumed to be zero, and its covariance function is

$$\mathbb{C}_\tau = \begin{cases} \sigma_\epsilon^2,\ \tau = 1, \\ 0,\ \tau = 2, \ldots, T, \end{cases} \tag{1.37}$$

where $\sigma_\epsilon^2 > 0$ is the white-noise variance.

An example of a white-noise process is a Gaussian white-noise $\epsilon_t \sim iid\ Gau(0, \sigma_\epsilon^2)$, for all t.

1.2.2.2 Random-walk process. A time series $\{\chi_t\}$ is said to be a random walk if its value at the current time is simply its value at the previous time plus some noise. Formally, that is:

$$\chi_t = \chi_{t-1} + \epsilon_t,\ t = 2, \ldots, T, \tag{1.38}$$

where $\{\epsilon_t\}$ is a white-noise process with mean μ_ϵ and variance σ_ϵ^2. We can show that in this case, $\mathbb{E}(\chi_t) = t\mu_\epsilon$ and $\mathbb{V}(\chi_t) = t\sigma_\epsilon^2$. That is, both the mean and the variance are functions of time and, thus, the process is non-stationary.

1.2.2.3 Autoregressive process. A time series $\{\chi_t\}$ is said to be an autoregressive process of order p (or $AR(p)$) if its value at time t depends on the previous p values that immediately precede it plus some noise. Formally, that is

$$\chi_t = G_1\chi_{t-1} + G_2\chi_{t-2} + \cdots + G_p\chi_{t-p} + \epsilon_t,\ t = p, \ldots, T, \tag{1.39}$$

where $\{\epsilon_t\}$ is a white-noise process with mean μ_ϵ and variance σ_ϵ^2 and where $\{\alpha_i : i = 1, \ldots, p\}$, are fixed unknown parameters.

For example, consider the $AR(1)$ process,

$$\chi_t = G_1\chi_{t-1} + \epsilon_t,\ t = 1, \ldots, T, \tag{1.40}$$

where $\{\epsilon_t\}$ is a white-noise process with mean μ_ϵ and variance σ_ϵ^2 and we write $G = G_1$ for notational simplicity. By back-substitution, we obtain

$$\begin{aligned}\chi_t &= \epsilon_t + G\epsilon_{t-1} + G^2\epsilon_{t-2} + G^3\epsilon_{t-3} + \dots \\ &= \sum_{k=0}^{\infty} G^k \epsilon_{t-k},\end{aligned} \tag{1.41}$$

where it is assumed that $|G| < 1$. That is, the $AR(1)$ process can be written as an infinite series of white-noise random variables (i.e., an infinite-order moving average process, as will be discussed in the next section). Since $\mathbb{E}(\epsilon_t) = 0$ and $\mathbb{V}(\epsilon_t) = \sigma_\epsilon^2$, it follows that $\mathbb{E}(\chi_t) = 0$ and

$$\mathbb{V}(\chi_t) = \sigma_\epsilon^2(1 + G^2 + G^3 + \dots) = \frac{\sigma_\epsilon^2}{1 - G^2}, \tag{1.42}$$

which does not depend on t. If the condition of $|G| < 1$ is not met, the process is explosive (i.e., the variance of the time series increases exponentially with time). Furthermore, when $G = 1$, the process is a random walk. In the explosive case where $|G| > 1$, one can show that $\chi_t = -\sum_{k=1}^{\infty} G^{-k}\epsilon_{t+k}$ which is stationary. However, such a model requires knowledge of the future and is not helpful. We say that a temporal process that does not depend on the future is *causal*.

It is also straightforward to show that for the $AR(1)$ process with $|G| < 1$, the autocorrelation function is given by

$$\rho_\tau = G^{|\tau|}, \ \tau = 0, \dots, T. \tag{1.43}$$

Thus, for the AR(1) process, the autocorrelation (i.e., the temporal dependence) decreases in magnitude as the time lag τ increases.

1.2.2.4 Moving-average process. The time series $\{\chi_t\}$ is said to be a *moving-average process of order* q (or $MA(q)$), if

$$\chi_t = \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}, \tag{1.44}$$

where $\beta_1, \beta_2, \ldots, \beta_q$ are fixed but unknown parameters, and $\{\epsilon_t\}$ is a white-noise process with zero mean and variance σ_ϵ^2. It follows that, $\mathbb{E}(\chi_t) = 0$ and $\mathbb{V}(\chi_t) = \sigma_\epsilon^2 \sum_{i=1}^{q} \beta_i^2$, which does not depend on t. Furthermore,

$$\rho_\tau = \begin{cases} 1, \ \tau = 1, \\ \dfrac{\sum_{i=0}^{q-\tau} \beta_i \beta_{i+\tau}}{\sum_{i=0}^{q} \beta_i^2}, \ \tau = 2, \ldots, q, \\ 0, \ \tau > q, \end{cases} \tag{1.45}$$

where the process is standardized by setting $\beta_0 = 1$.

An interesting feature of the MA process is illustrated by considering two different formulations of an $MA(1)$ process (e.g., Chatfield (2016)):

$$\chi_t = \epsilon_t + \beta \epsilon_{t-1}, \tag{1.46}$$

$$\chi_t = \epsilon_t + \frac{1}{\beta} \epsilon_{t-1}. \tag{1.47}$$

From (1.45), it is clear that both models have exactly the same autocorrelation function. However, from (1.46), by back-substitution we obtain the following:

$$\epsilon_t = \chi_t - \beta \chi_{t-1} - \beta^2 \chi_{t-2} - \ldots, \tag{1.48}$$

which is an infinite AR series. This infinite sequence converges if $|\beta| < 1$. However, if we do the same back-substitution for (1.47), the series does not converge for $|\beta| < 1$. We

say that the $MA(1)$ given by (1.46) is invertible if it can be written as an infinite-order autoregressive process, which occurs if $|\beta| < 1$.

1.2.2.5 Autoregressive moving-average process. We note the duality between AR and MA processes says that: an $AR(p)$ process can be written as an infinite-order MA process, and an $MA(q)$ process can be written as an infinite-order AR process. This duality suggests that a more parsimonious model for many processes might result from a combination of AR and MA processes. The time series $\{\chi_t\}$ is said to be an *autoregressive moving-average process of order* (p, q) or $ARMA(p, q)$, if

$$\chi_t = G_1\chi_{t-1} + \cdots + G_p\chi_{t-p} + \epsilon_t + \beta_1\epsilon_{t-1} + \cdots + \beta_q\epsilon_{t-q}, \tag{1.49}$$

where $G_1, \ldots, G_p, \beta_1, \ldots, \beta_q$ are fixed but unknown parameters, and $\{\epsilon_t\}$ is a white-noise process with zero mean and variance σ^2_ϵ.

1.3 Spatio-temporal geostatistics in continuous time

The aim of this section is to introduce the most relevant concepts in the analysis of spatio-temporal random fields with spatio-temporal continuity. Following the approach of Montero et al. (2015), let χ_{s_i,t_i} be a generic spatio-temporal *rf* expressed as

$$\{\chi_{s_i,t_i}, s_i \in \mathbb{S}, t_i \in \mathcal{T}\}, \tag{1.50}$$

with $\mathbb{S} \subset \mathbb{R}^d$ and $\mathcal{T} \subset \mathbb{R}$.

The observed data y_{s_i,t_i} are assumed to be realizations of the spatio-temporal rf (1.50) according to the model

$$y_{s_i,t_i} = \chi_{s_i,t_i} + \epsilon_{s_i,t_i}, \tag{1.51}$$

where ϵ_{s_i,t_i} is a white noise process i.i.d $\mathcal{N}(0, \sigma_\epsilon^2)$.

The rf (1.50) is expressed as

$$\chi_{s_i,t_i} = \mu_{s_i,t_i} + \omega_{s_i,t_i}, \tag{1.52}$$

where μ_{s_i,t_i} is a mean process driven by observed covariates x_{s_i,t_i} and ω_{s_i,t_i} is a zero mean spatio-temporal process.

The spatio-temporal description of (1.50) is done through its covariance function defined as

$$\mathbb{C}(\chi_{s,t}, \chi_{s',t'}) = \mathbb{C}(\omega_{s,t}, \omega_{s',t'}) = \mathbb{C}_{s,s';t,t'}. \tag{1.53}$$

The intuitive notion that values tend to be weakly correlated when they are "far apart" can be captured by using covariance functions that tend to zero as their spatio-temporal coordinates move farther apart. Covariance models are often used that are identically zero when separated by more than a fixed distance in space or in time; these are sometimes referred to as compactly supported covariances. Since the structural analysis is about quantifying and modelling spatial variability through covariance, variogram and/or semi-variogram, the next subsection will be dedicated to properties of such functions.

1.3.1 Theoretical properties. A covariance function (1.53) is a *non-negative definite function.*

Definition 1 *A function* $\{f_{u,v} : u, v \in \mathbb{S}\}$ *defined on* $\mathbb{S} \times \mathcal{T}$ *is said to be non-negative definite, if for any complex numbers* $\{a_i : i = 1, \ldots, N\}$*, any* $\{u_i : i = 1, \ldots, N\}$ *in* $\mathbb{S}$*, an any integer* N*, we have*

$$\sum_{i=1}^{N} \sum_{j=1}^{N} a_i \bar{a}_j f_{u_i,u_j} \geq 0, \tag{1.54}$$

where $\bar{a}$ denotes the complex conjugate of a.

Positive definiteness of $f(\cdot,\cdot)$ involves the inequality above being strictly positive whenever $\mathbf{a} = (a_1, \ldots, a_N)'$ is a non-zero vector. In (1.54), we write $u = (s,t)$ and $f_{u_i,u_j} \equiv f_{(s_i,t_i),(s_j,t_j)}$ to obtain a spatio-temporal covariance function.

The zero mean spatio-temporal process $\omega_{s,t}$ is said to be *covariance stationary* if (1.53) satisfies the condition (1.54) and can be written as

$$\mathbb{C}_{s,s';t,t'} = \mathbb{C}_{s-s';t-t'} = \mathbb{C}_{h,\tau},$$

where $h = s - s'$, $\tau = t - t'$ and the spatio-temporal correlation function associated is

$$\rho_{h,\tau} = \mathbb{C}_{h,\tau}/\mathbb{C}_{0,0}. \tag{1.55}$$

It is possible that stationarity of the covariance function be considered separately for space and time. Spatial stationarity of the covariance function (1.53) corresponds to

$$\mathbb{C}_{s,s';t,t'} = \mathbb{C}_{s-s';t,t'}$$

and temporal stationarity of the covariance function (1.53) correspond to

$$\mathbb{C}_{s,s';t,t'} = \mathbb{C}_{s,s';t-t'}.$$

The process is said to be *isotropic* if

$$\mathbb{C}_{h,\tau} = \mathbb{C}_{||h||,|\tau|},$$

that is, the covariance function depends upon the separation vectors only through their lengths $||d||$ and $|\tau|$. The process is said to be anisotropic when it is not isotropic.

The process $\omega(s,t)$ is said to be *separable* if

$$\mathbb{C}_{s,s';t,t'} = \mathbb{C}_{s,s'}\mathbb{C}_{t,t'}, \tag{1.56}$$

where $\mathbb{C}_{s,s'}$ and $\mathbb{C}_{t,t'}$ are spatial and temporal covariance functions, respectively.

As a consequence of (1.56), a simple class of spatio-temporal covariance functions is given by product of individual spatial and temporal covariance functions. When $\mathbb{C}_{s,s'}$ and $\mathbb{C}_{t,t'}$ are, respectively, spatially and temporally stationary, (1.56) becomes

$$\mathbb{C}_{h,\tau} = \mathbb{C}_h\mathbb{C}_\tau.$$

Also, separability implies that the spatio-temporal correlation function ρ given by (1.55) satisfies

$$\rho_{h,\tau} = \rho_{h,0}\rho_{0,\tau}. \tag{1.57}$$

Usually, spatio-temporal models constructed dynamically often have important components of spatio-temporal interaction, for which a separable covariance function is inadequate.

The spatio-temporal covariance function (1.53) is said to be fully symmetric if

$$\mathbb{C}_{s,s';t,t'} = \mathbb{C}_{s,t'}\mathbb{C}_{s',t}.$$

1.3.2 Sums and products of covariance functions. Because the sum of two non-negative definite functions is non-negative definite, the following function is a valid spatio-temporal covariance functions:

$$\mathbb{C}_{h,\tau} = p\mathbb{C}_{h_1}\mathbb{C}_{\tau_1} + q\mathbb{C}_{h_2} + r\mathbb{C}_{\tau_2},$$

where $p > 0, q > 0$, and $r > 0$, $\mathbb{C}_{h_1}$ and $\mathbb{C}_{h_2}$ are spatial covariance functions, and $\mathbb{C}_{\tau_1}$ and $\mathbb{C}_{\tau_2}$ are temporal covariance functions.

1.3.3 Spatio-temporal variogram. Another way to define spatio-temporal dependence of a spatio-temporal process is through its variogram. The spatio-temporal variogram of the process (1.50) is defined to be

$$\mathbb{V}(\chi_{s,t} - \chi_{s',t'}) = 2\gamma_{s,s';t,t'}, \tag{1.58}$$

along with its stationary version. The quantity $\gamma_{s,s';t,t'}$ is called semivariogram. The process (1.50) is said to be intrinsically stationary, if it has constant expectation and its variogram is stationary. If is second-order stationary with (stationary) covariance function $\mathbb{C}_{h,\tau}$, then as in the purely spatial setting (1.50) is intrinsically stationary with semivariogram,

$$\gamma_{h,\tau} = \mathbb{C}_{0,0} - \mathbb{C}_{h,\tau}, \ h \in \mathbb{S}, \tau \in \mathcal{T}.$$

Similarly, we can define separately the notions of spatial or temporal intrinsic stationary, and so on, of the variogram.

1.3.4 Spatio-temporal kriging. The goal of kriging is to predict χ_{s_0,t_0} from incomplete and noisy data. Any unbiased linear predictor, $\hat{\chi}_{s_0,t_0}$ of χ_{s_0,t_0}, has the property that its mean squared prediction error $\mathbb{E}(\hat{\chi}_{s_0,t_0} - \chi_{s_0,t_0})^2$, can be expressed in terms of the variogram. It can also obviously be expressed in terms of the covariance function.

Suppose that the data are

$$y_{s_i,t_{ij}} = \chi_{s_i,t_{ij}} + \epsilon_{s_i,t_{ij}},\ j = 1,\ldots,T_i,\ i = 1,\ldots,N, \tag{1.59}$$

where $\{\epsilon_{s_i,t_{ij}}\}$ is independent of χ and represents the measurement error that is henceforth assumed to be *iid* with mean zero and variance σ^2_ϵ. Write $y^{(i)} = (y_{s_i,t_{ij}}, j = 1,\ldots,T_i)'$, $i = 1,\ldots,N$. Then we wish to predict the hidden value χ_{s_0,t_0}, based on all the data given by (1.59).

1.3.4.1 Spatio-temporal simple kriging. The simple-kriging predictor, $\tilde{\chi}_{s_0,t_0}$, takes the form of a linear combination of the data:

$$\begin{aligned} \tilde{\chi}_{s_0,t_0} &= \sum_{i=1}^{N}\sum_{i=1}^{Ti} \lambda_{ij} y_{s_i,t_i} + c \\ &= \lambda' \mathbf{y} + c \end{aligned} \tag{1.60}$$

where $\mathbf{y} = (y^{(1)},\ldots,y^{(N)})'$ and λ and c are optimized (i.e. the mean squared prediction error is minimized).

The mean function $\mu_{s,t} = E(\chi_{s,t})$, is assumed to be known.

From the point of view of kriging, time is simply another dimension and any spatio-temporal covariance function (or variogram) would have to respect that distances in time would be treated differently form distances in space.

Write $\mathbb{C}_y \equiv \mathbb{V}(Y)$, $c_0 \equiv \mathbb{C}(\chi_{s_0,t_0}, Y)$ and $\mathbb{C}_{0,0} \equiv \mathbb{V}(\chi_{s_0,t_0})$. If both the hidden process χ and the measurement error process ϵ are assumed to be Gaussian processes and we write $\mathbf{Y} = \boldsymbol{\chi} + \boldsymbol{\epsilon}$, then $\mathbf{Y}$ and $\boldsymbol{\epsilon}$ follow Gaussian distributions with covariance matrices, Σ_Y and Σ_ϵ, respectively. From (1.60), we have

$$\begin{bmatrix} \chi_{s_0,t_0} \\ \mathbf{Y} \end{bmatrix} \sim Gau\left(\begin{bmatrix} \mu_{s_0,t_0} \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} \mathbb{C}_{0,0} & c_0' \\ c_0 & \mathbb{C}_y \end{bmatrix} \right) \tag{1.61}$$

where $\boldsymbol{\mu} \equiv (\mu_{s_i,t_{ij}} : j = 1, \ldots, T_i, i = 1, \ldots, N)'$, and $\mathbb{C}_y = \Sigma_Y + \Sigma_\epsilon$.

Thus the posterior distribution is

$$\chi_{s_0,t_0}|\mathbf{Y} \sim Gau\left(\mu_{s_0,t_0} + c_0'\mathbb{C}_y^{-1}(\mathbf{Y} - \mu), \mathbb{C}_{0,0} - c_0'\mathbb{C}_y^{-1}c_0\right). \tag{1.62}$$

Under the joint Gaussian assumption (1.61), the simple-kriging predictor is just the posterior mean of (1.62)

$$\tilde{\chi}_{s_0,t_0} = \mathbb{E}(\chi_{s_0,t_0}|\mathbf{Y}) = \mu_{s_0,t_0} + c_0'\mathbb{C}_y^{-1}(\mathbf{Y} - \mu). \tag{1.63}$$

The simple kriging variance is the minimized mean squared prediction error and under the join-Gaussian assumption (1.61), it is just the posterior variance

$$\tilde{\sigma}^2_{s_0,t_0} = E(\chi_{s_0,t_0} - \tilde{\chi}_{s_0,t_0})^2 = \mathbb{C}_{0,0} - c_0'\mathbb{C}_y^{-1}c_0. \tag{1.64}$$

Notice that the kriging equations involve taking the inverse of the $T_+ \times T_+$ positive-definite matrix $\mathbb{C}_y \equiv \mathbb{V}(\mathbf{Y})$, where $T_+ = \sum_{i=1}^N T_i$.

If the mean function is unknown, different versions of kriging can be developed. If the mean function is assumed constant over space and time, then we can derive the ordinary kriging predictor.

1.3.4.2 Spatio-temporal ordinary kriging. We know consider kriging for the case where χ has unknown mean constant over space and time. Then $\boldsymbol{\mu} = \mu\mathbf{1}$, where $\mathbf{1}$ is a T_+-dimensional vector of ones. Using the generalized least squares estimator,

$$\hat{\mu}_{gls} \equiv (\mathbf{1}'\mathbb{C}_y^{-1}\mathbf{1})\mathbf{1}'\mathbb{C}_y^{-1}\mathbf{1}, \tag{1.65}$$

for $\mu(\cdot) \equiv \mu$ into (1.63) yields the ordinary kriging predictor,

$$\begin{aligned} \tilde{\chi}_{s_0,t_0} &\equiv \hat{\mu}_{gls} + c_0'\mathbb{C}_y^{-1}(\mathbf{Y} - \hat{\mu}_{gls}) \\ &\equiv \boldsymbol{\lambda}'\mathbf{Y}, \end{aligned} \tag{1.66}$$

where $\boldsymbol{\lambda}' \equiv \{c_0 + \mathbf{1}(1 - \mathbf{1}'\mathbb{C}_y^{-1}c_0)/(\mathbf{1}'\mathbb{C}_y^{-1}\mathbf{1})\}'\mathbb{C}_y^{-1}$. The formula (1.66) gives the predictor that minimizes the mean squared prediction error over all linear unbiased predictors. The minimized mean squared prediction error is the ordinary kriging (ok) variance

$$\hat{\sigma}^2_{s_0,t_0} = \mathbb{C}_{0,0} - c_0'\mathbb{C}_y^{-1}c_0 + (1 - \mathbf{1}'\mathbb{C}_y^{-1}c_0)^2/(\mathbf{1}'\mathbb{C}_y^{-1}\mathbf{1}).$$

1.4 Spatio-temporal geostatistics in discrete time

This section is dedicated to the description of a spatio-temporal geostatistics model in discrete time. We will describe the Dynamic Coregionalization Model (DCM) introduced by Fassò & Finazzi (2011) and implemented in D-STEM by Finazzi & Fassó (2014).

D-STEM stands for distributed space time expectation maximization and it is a statistical tool for the analysis and mapping (spatial prediction) of environmental space-time variables. Its particularity is its ability to deal with multiple variables, heterogeneous spatial supports, heterogeneous sampling networks and missing data. Here we are going to present the hierarchical space-time model handled in D-STEM and next the estimation method and spatial prediction method. For further details about all feasible and computable models handled by D-STEM see Finazzi & Fassó (2014).

1.4.1 The model. D-STEM extends the Dynamic Coregionalization Model (DCM) introduced by Fassò & Finazzi (2011) through a parametric statistical model based on latent space-time random variables, space-time varying coefficients, latent spatial random variables modelled as Gaussian random fields with a Matérn correlation function and latent temporal variables with Markovian dynamics.

Let $Y_{s,t} = (y^1_{s,t}, \ldots, y^Q_{s,t})', t \in \{1, \ldots, T\}$ and $s \in \mathbb{S} \subset \mathbb{R}^d$, be a Q-dimensional spatio-temporal process given by

$$Y_{s,t} = \chi_{s,t} + \epsilon_{s,t}, \tag{1.67}$$

where $\epsilon_{s,t}$ is a zero mean Gaussian process of measurement errors uncorrelated over space and time with $Q \times Q$ covariance diagonal matrix σ^2_ϵ.

The underlying process $\chi_{s,t}$ has qth element given by

$$\chi^q_{s,t} = \mu^q_{s,t} + \omega^q_{s,t}, \tag{1.68}$$

for $q = 1, \ldots, Q$, where $\mu^q_{s,t}$ is the fixed effect, $\omega^q_{s,t}$ is the spatio-temporal random effect.

The fixed effect $\mu^q_{s,t}$ in equation (1.68) is modelled as

$$\mu^q_{s,t} = X_{s,t}\beta^q, \tag{1.69}$$

where $X_{s,t}$ is a b-dimensional spatio-temporal field of known covariates and β^q are to be estimated.

The random effect $\omega^q_{s,t}$ in equation (1.68) is modelled as

$$\omega^q_{s,t} = \sum_{p=1}^{c} \alpha^q X^{w_p}_{s,t} w_{s,t} + z^q_t, \tag{1.70}$$

where X^{w_p} is a p-dimensional vector of spatial covariates, α^q is a p-dimensional vector of spatial coefficients to be estimated and z_t^q is a d-dimensional latent temporal.

The temporal latent state at time t is defined as

$$z_t^q = G^q z_{t-1} + \eta_t^q,$$

where G^q is a $d \times d$ transition matrix assumed to have eigenvalues smaller than 1 in absolute value and innovations $\eta_t^q \sim \mathcal{N}_d(0, \Sigma_\eta)$ and $z_0^q =\sim \mathcal{N}_d(\mu_0, \Sigma_{z_0})$.

The spatial covariance matrix function is given by

$$\mathbb{C}(w_s, w_{s'}) = V\rho(h; \theta, \nu),$$

where $h = ||s - s'||$ is the euclidean distance between two sites $s, s' \in \mathbb{S}$, the covariance function is assumed to be isotropic, for each $p = 1, \ldots, c$, V is a positive semi-definite matrix, the variable $w_{s,t}$ are zero-mean and unit-variance independent Gaussian processes uncorrelated over time but correlated over space with Matérn spatial covariance function

$$\rho(h; \theta, \nu) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\sqrt{2\nu}\frac{h}{\theta}\right)^\nu K_\nu \left(\sqrt{2\nu}\frac{h}{\theta}\right), \tag{1.71}$$

where $\theta > 0$ is a parameter to be estimated, ν can be chosen between 1/2, 3/2 and 5/2; Γ is the gamma function, K_ν is the modified Bessel function of the second kind.

A q-dimensional linear coregionalization model (LCM) of c components describes the latent random spatial effect at time t as

$$\bar{w}_{s,t} = \sum_{p=1}^{c} w_{p_{s,t}},$$

where $w_p = \left(w_p^1, \ldots, w_p^q\right)$ is white noise in time but correlated over space with a $Q \times Q$ covariance and cross-covariance matrix function given by

$$\Gamma_p(h,\theta_p) = \mathbb{C}(w^i_{p_s}, w^j_{p_{s'}})_{i,j=1,\ldots,q} = V_p\rho_p(h,\theta_p),$$

where the cross-covariance functions are assumed to be isotropic, V_p is a positive semi-finite $Q \times Q$ matrix and $\rho_{p;h,\theta_p}$ is a Matérn correlation. In addition, the processes w_p are uncorrelated in the sense that, for any $i \neq j$

$$\mathbb{C}(w^i_s, w^j_{s'}) = 0, \forall s, s' \in \mathbb{S}.$$

The multivariate $Q \times Q$ covariance matrix for w is then

$$\Gamma_w(h,\theta_1,\ldots,\theta_c) = \sum_{p=1}^{c} \Gamma_p(h,\theta_p) = \sum_{p=1}^{c} V_p\rho_p(h,\theta_p).$$

The parameter set to be estimated

$$\psi = vec^*(\beta, \sigma^2_\epsilon, \mu_o, \alpha, \theta, V, G, V_\eta) = (\psi_Y, \psi_{z_0}, \psi_z, \psi_w), \tag{1.72}$$

where $\beta = (\beta'_1, \ldots, \beta'_Q)$ is b-dimensional with $b = \sum_{j=1}^{Q} b_j, \alpha = \{\alpha_1, \ldots, \alpha_c\}, \theta = \{\theta_1, \ldots, \theta_c\}, V = \{V_1, \ldots, V_c\}$, V_η is the $p(p+1)/2$ dimensional vector of unique elements of Σ_η and operator vec^* vectorizes all the unique parameters contained in the covariance matrices excluding structural zeros.

1.4.2 Model estimation. We consider heterotopic spatio-temporal model with spatial LCM components and temporal dynamics. The model parameters in D-STEM are estimated through the maximum likelihood by means of the expectation maximization (EM) algorithm. We present here an overview of the estimation procedure for the DCM model implemented in D-STEM through the estimation of the set of parameter ψ

related to the spatio-temporal dynamics. Moreover cross-validation is performed for model assessment.

To do this available dataset Y, which is given by $y_{s,t}$, $s \in \mathbb{S}, t = 1, \ldots, \mathcal{T}$, is partitioned in two randomly extracted parts non overlapping in space, but same calendar, namely $Y^{[i]} = (y_{s,t}, s \in S_i)$, with S_i of size n_i, $S_i \cap S_j = \varnothing$, $i \neq j = 1, 2,$.

In particular, estimation of ψ is based on the randomly extracted sub-dataset $Y^{[1]}$, respectively, and crossvalidation is performed using $Y^{[2]}$.

Estimation of the model parameters in ψ is based on EM algorithm and is implemented by means of the D-STEM software applied to data $Y^{[1]}$. The EM algorithm implemented in D-STEM is discussed in Fassò & Finazzi (2011) and in more details in Fassò & Finazzi (2013).

The parameter set may be written with initial values as

$$\begin{aligned} \psi &= \{\beta, \sigma^2_\epsilon, \alpha, \theta, V, G, V_\eta\} \\ &= \{\psi_0, \psi_1, \ldots, \psi_c\}, \end{aligned}$$

where $\psi_0 = \{\beta, \sigma^2_\epsilon, \alpha\}$ and $\psi_p = \{\theta_p, v_p\}$, $p = 1, \ldots, c$.

The complete loglikelihood may be written as

$$\log L(\psi, \mathbf{X}) = \log L(\psi_0, \mathbf{X}) + \sum_{p=1}^{c} \log L(\psi_j, \mathbf{X}), \tag{1.73}$$

where $X = (Y, Z, W)$ is the set of all data, either observed or latent, for $s \in S_1, t = 1, \ldots, T$. This additive structure is retained in the E-step of the EM algorithm so that the quantity to be iteratively optimized, namely

$$Q(\psi, \psi^{(m)}) = \mathbb{E}_{\psi^{(m)}}(\log L(\psi, \mathbf{X})|\mathbf{Y})$$

may be written as

$$Q(\psi, \psi^{(m)}) = Q_0(\psi_0, \psi^{(m)}) + \sum_{p=1}^{c} Q_c(\psi_c, \psi^{(m)}).$$

It follows that the M-step of the EM algorithm is given by the following updating equations:

$$\psi_p^{(m+1)} = \arg\max_{\psi_p} Q_j(\psi_p, \psi^{(m)}),\ p = 0, 1, \ldots, c. \tag{1.74}$$

1.4.3 Spatial prediction in DCM. Let $\mathbf{Y} = \{\mathbf{y}_{s,t}^1, \ldots, \mathbf{y}_{s,t}^Q\}, t \in \{1, \ldots, T\}$ and $s \in \mathbb{S} \subset \mathbb{R}$, be a Q-variate spatio-temporal dataset. Each $\mathbf{y}_{s,t}^Q$ is a $N_q T$-dimensional vector of the qth variable at spatial locations $\mathbb{S}_q = \{s_1, \ldots, s_{N_q}\}, q = 1, \ldots, Q$. The Q variables are not necessarily observed at the same spatial locations (see Fassò & Finazzi (2011)) in the sense that the sets $\mathbb{S}_q$ may be different and the vectors $\mathbf{y}_{s,t}^Q$ may include missing values.

The minimum mean square error spatial prediction of $y_{s_0}^q$ at point $s_0 \notin \mathbb{S}_q$ is given by

$$\hat{y}_{s_0}^q = E\left[y_{s_0}^q|\mathbf{Y}^*\right]$$

with uncertainty

$$\mathbb{V}\left[\hat{y}_{s_0}^q\right] = \mathbb{V}\left[y_{s_0}^q|\mathbf{Y}^*\right]$$

where $\mathbf{Y}^* = \{\mathbf{y}_{s,t}^{*1}, \ldots, \mathbf{y}_{s,t}^{*Q}\}$ is the dataset restricted to the non-missing data.

II. Functional data analysis

Functional data analysis (*fda*) is about the statistical analysis of data observed (or supposed to be observed) as functions or curves. The observed function is supposed to be a realization of a functional variable $\boldsymbol{\chi}$.

Definition 2 *Ferraty & Vieu (2006) A random variable, $\boldsymbol{\chi}$, is called a functional variable (fv) if it takes values in a infinite-dimensional space (or functional space). An observation of that variable, χ is called a functional observation.*

In the terms of Ramsay (2006) a functional observation for replication i is a set of discrete measured values y_{i1},$y_{i2}, \ldots, y_{iM}$, where M is not necessary equal for all i. The first task is to convert those discrete measured values to a function χ_i with values $\chi_i(u)$ computable for any desired value of the argument u. This implies that we assume the existence of a "smooth" function gives rise to the finite observations available. A function is said to be smooth when it is differentiable to a certain degree, implying that a number of derivatives can be derived or estimated from the data.

If the data are observed without errors, the building process of functions is interpolation while it requires smoothing when the data are observed with errors. The function can be estimated using parametric or non-parametric approach. In a parametric approach, the function $\chi(.;\theta)$ is known and estimation of the parameters $\hat{\theta}$ is the focus of the analysis. It is clear that a parametric approach is limited because in many applications we do not know the function. In that case we use a non-parametric approach which is more flexible and only assumes smoothness.

From a set of discrete measured values y_{i1},$y_{i2}, \ldots, y_{iM}$, $i = 1, \ldots, n$, once built the smooth curves (functional data) χ_i, $i = 1, \ldots, n$, we obtain a functional dataset.

Definition 3 *Ferraty & Vieu (2006) A functional dataset $\chi_1, \chi_2, \ldots, \chi_n$ is the observation of n functional variables $\boldsymbol{\chi}_1, \boldsymbol{\chi}_2, \ldots, \boldsymbol{\chi}_n$ identically distributed as $\boldsymbol{\chi}$*

The basic statistics of classical analysis are also available in the case of functional data. As example, the functional descriptive statistics are defined in Ramsay (2006) as:

- Mean function

$$\bar{\chi}(u) = \frac{\sum_{i=1}^{n} \chi_i(u)}{n}.$$

- Variance function

$$\mathbb{V}_{\chi}(u) = \frac{\sum_{i=1}^{n} \left[\chi_i(u) - \bar{\chi}(u)\right]^2}{n}.$$

- Covariance function which measures the linear dependence of the data across different values of the argument

$$\mathbb{C}\left[\chi(u), \chi(u')\right] = \frac{\sum_{i=1}^{n} \left[\chi_i(u) - \bar{\chi}(u)\right] \left[\chi_i(u') - \bar{\chi}(u')\right]}{n}, \forall u, u' \in U.$$

- Correlation function

$$\rho\left[\chi(u), \chi(u')\right] = \frac{\mathbb{C}\left[\chi(u), \chi(u')\right]}{\sqrt{\mathbb{V}_{\chi}(u)\mathbb{V}_{\chi}(u')}}, \forall u, u' \in U.$$

- Cross-covariance function

$$\mathbb{C}\left[\chi(u), \mathcal{Y}(u')\right] \frac{\sum_{i=1}^{n} \left[\chi_i(u) - \bar{\chi}(u)\right] \left[\mathcal{Y}_i(u') - \bar{\mathcal{Y}}(u')\right]}{n}, \forall u, u' \in U.$$

- Cross-correlation function

$$\rho\left[\chi(u), \mathcal{Y}(u')\right] = \frac{\mathbb{C}\left[\chi(u), \mathcal{Y}(u')\right]}{\sqrt{\mathbb{V}_{\chi}(u)\mathbb{V}_{\mathcal{Y}}(u')}}, \forall u, u' \in U.$$

In this chapter we are mainly interested to building a smooth function from a collection of discrete observations. We are going to present in the following sections a non parametric approach using a set of functional building block of ϕ_Q B-splines through the following steps: the construction of B-splines basis system, the estimation of the coefficients of splines through regression without roughness penalty and regression with roughness penalty in order to impose a degree of smoothness. For a further details in *fda* see Ramsay et al. (2009).

2.1 Building functional data with B-splines basis system

We build functions using a basis function system which is a linear combination of Q basis functions called a *basis function expansion* defined as

$$\chi(u) = \sum_{q}^{Q} c_q \phi_q = \mathbf{c}'\boldsymbol{\phi}(u) = \boldsymbol{\phi}'(u)\mathbf{c} \tag{2.1}$$

where $c_1, \ldots, c_q$ are the coefficients of the expansion and ϕ in the last term of (2.1) is a vector of length Q containing the basis functions.

It is important to distinguish between the concept of function in a general sense as a single entity and the concept of its value at a specific argument value u. Equation (2.1) refers to the basis function expansion of the value of function χ at specific argument value u, but the expansion of χ as a single entity is better written as

$$\chi = \sum_{q}^{Q} c_q \phi_q = \mathbf{c}'\phi. \tag{2.2}$$

As basic functions for non-periodic data we usually use splines which are piecewise polynomials defined by the range of validity, the break points, the knots, and the order.

The *range of validity* is the domain of the definition of the function. It is divided into subintervals with boundaries at points called *break points* while a sequence of *knots* is defined. The number of knots positioned at a break point determined the number of matching derivatives between neighbouring polynomials at that break point. If only one knot is positioned at a break point the number of matching derivatives is two less than the *order* of the function. If smooth and accurate derivatives are needed, it is necessary to increase the order of the spline. A common choice is to fix the order of the spline basis at least two times higher than the highest order derivative to be used. The number of Q basis functions is computed as

$$number\ of\ basis\ functions = order + number\ of\ interior\ knots. \tag{2.3}$$

Interior knots are knots which are not at the beginning or end of the range of validity.

The most popular basis system for constructing spline functions is the *B-spline basis system*. A useful property of B-spline basis functions is that the sum of the B-spline basis function values at any point u is equal to one. Another property is that the number of order of the function determined the number of adjacent intervals (at most) in which the functions are positive. The computational effort is then proportional to Q rather than Q^2 for basis functions not having that property.

The figure 2.1 shows the construction of B-spline functions on the interval $[0, 1]$ keeping fixed the 9 internal break points and putting one knot at each internal break points. We can note the functions are smoother when the order of functions is higher.

2.2 Regression splines

Once the spline basis function system is specified, we have to estimate the coefficients in order to build functional data (smooth curves) by combining the estimated coefficients with the spline basis functions. We can use two approaches to compute the smooth curves

of functional observations. The simplest one is to use a regression without roughness penalty and the second one is to use a penalized regression by imposing a penalty on the roughness.

2.2.1 Smoothing without roughness penalty. If the functional data are observed with errors, the model is defined as

$$y_j = \chi(u_j) + \epsilon_j = \mathbf{c}'\boldsymbol{\phi}(u) + \epsilon_j = \boldsymbol{\phi}'(u_j)\mathbf{c} + \epsilon_j \tag{2.4}$$

where the residuals ϵ_j are statistically independent and normally distributed with mean 0 and constant variance.

We can use the classical equations of regression analysis defining the functional data fitting as the minimization of the sum of squared errors

$$sse(\chi) = \sum_{j=1}^{M} [y_j - \chi(u_j)]^2. \tag{2.5}$$

Replacing the expression of basis function expansion (2.1) in (2.5), the least-squares estimation problem becomes

$$sse(\mathbf{c}) = \sum_{j=1}^{M} \left[y_j - \sum_{q}^{Q} c_q \phi_q(u_j) \right]^2 = \sum_{j=1}^{M} [y_j - \boldsymbol{\phi}(u_j)'\mathbf{c}]. \tag{2.6}$$

The model (2.4) is expressed in matrix notation as

$$\mathbf{y} = \boldsymbol{\phi}\mathbf{c} + \boldsymbol{\epsilon} \tag{2.7}$$

where $\mathbf{y}$ contains the M values to be fit, $\boldsymbol{\epsilon}$ contains the errors and $\boldsymbol{\phi}$ contains the $M \times Q$ basis function values $\phi_q(u_j)$ and $\mathbf{c}$ are the Q coefficients of splines to be estimated. The least-squares estimate of the coefficient $\mathbf{c}$ is computed as usual

$$\hat{\mathbf{c}} = (\boldsymbol{\phi}'\boldsymbol{\phi})^{-1}\boldsymbol{\phi}'\boldsymbol{\chi}. \tag{2.8}$$

From equation (2.8) we can carry out two important mapping matrices. The first one expressed as $(\boldsymbol{\phi}'\boldsymbol{\phi})^{-1}\boldsymbol{\phi}'$ maps the data vector into the vector of estimated coefficients. The second one, often called *hat matrix*, maps the data vector into the vector of fitted values is defined as

$$\mathbf{H} = \boldsymbol{\phi}(\boldsymbol{\phi}'\boldsymbol{\phi})^{-1}\boldsymbol{\phi}' \tag{2.9}$$

so that $\hat{\boldsymbol{\chi}} = \mathbf{H}\boldsymbol{\chi}$.

2.2.2 Smoothing with roughness penalty. In the roughness penalty approach, we have to define a measure of the roughness of the fitted curve. Since the square of second derivative $[D^2(\chi(u)]^2$ of a function χ at argument value u is often called its curvature at u, a common used measure of function's roughness is the *integrated squared second derivative* or *total curvature*

$$PEN_2 = \int \left[D^2\chi(u)\right]^2 du. \tag{2.10}$$

Once defined the roughness penalty (2.10), we add some multiple of it to the error sum of squares to define the fitting criterion to minimize

$$F(\mathbf{c}) = \sum_j \left[\chi_j - \chi(u_j)\right]^2 + \lambda \int \left[D^2\chi(u)\right]^2 du \tag{2.11}$$

where $\chi(u) = \mathbf{c}'\boldsymbol{\phi}(u)$, and λ is the *smoothing parameter* such that $\lambda \longrightarrow 0$ leaves the function χ free to fit the data as closely as possible.

Let $\mathbf{L} = D^m$ with D^m being the m_{th} derivative, the roughness penalized fitting criterion (2.11) is rewritten as

$$F(\mathbf{c}) = \sum_j \left[\chi_j - \chi(u_j)\right]^2 + \lambda \int \left[\mathbf{L}\chi(u)\right]^2 du. \tag{2.12}$$

Using the expression of basis expansion in equation (2.1) into equation (2.12) gives us

$$F(\mathbf{c}) = \sum_j \left[\chi_j - \boldsymbol{\phi}'(u_j)\mathbf{c}\right]^2 + \lambda \mathbf{c}' \int \left[\mathbf{L}\boldsymbol{\phi}(u)\mathbf{L}\boldsymbol{\phi}'(u)du\right] \mathbf{c}du. \tag{2.13}$$

The Q symmetric roughness penalty matrix is then defined as

$$\mathbf{R} = \int \phi(u)\phi'(u)du \tag{2.14}$$

and the counterpart of (2.8) is

$$\hat{\mathbf{c}} = (\boldsymbol{\phi}'\boldsymbol{\phi} + \boldsymbol{\lambda}\mathbf{R})^{-1}\boldsymbol{\phi}'\boldsymbol{\chi}. \tag{2.15}$$

The counterpart of *hat matrix* (2.9) is

$$\mathbf{H} = \boldsymbol{\phi}(\boldsymbol{\phi}'\boldsymbol{\phi} + \boldsymbol{\lambda}\mathbf{R})^{-1}\boldsymbol{\phi}'. \tag{2.16}$$

The relationship between observed and fitted values is the following:

$$\hat{\boldsymbol{\chi}} = \mathbf{H}(\boldsymbol{\lambda})\boldsymbol{\chi} \tag{2.17}$$

with the smoothing matrix $\mathbf{H}(\boldsymbol{\lambda})$ being square, symmetric, of order M, function of λ.

$\mathbf{H}(\boldsymbol{\lambda})$ is often used as a measure of the effective degrees of freedom of the fit defined by λ

$$df(\lambda) = trace[\mathbf{H}(\boldsymbol{\lambda})]$$

and the associated degrees of freedom for error is $n - df(\lambda)$.

As $\lambda \to 0, df(\lambda) \to \min(M, Q)$, and as $\lambda \to \infty, df(\lambda) \to m$, where M is the number of observations and Q is the number of basis functions and m is the order of the highest derivative used to define the roughness penalty.

The best value for the smoothing parameter λ is given by the generalized cross-validation measure GCV developed by Craven and Wahba (1979) and defined as

$$GCV(\lambda) = \left(\frac{M}{M - df(\lambda)}\right)\left(\frac{sse}{M - df(\lambda)}\right). \tag{2.18}$$

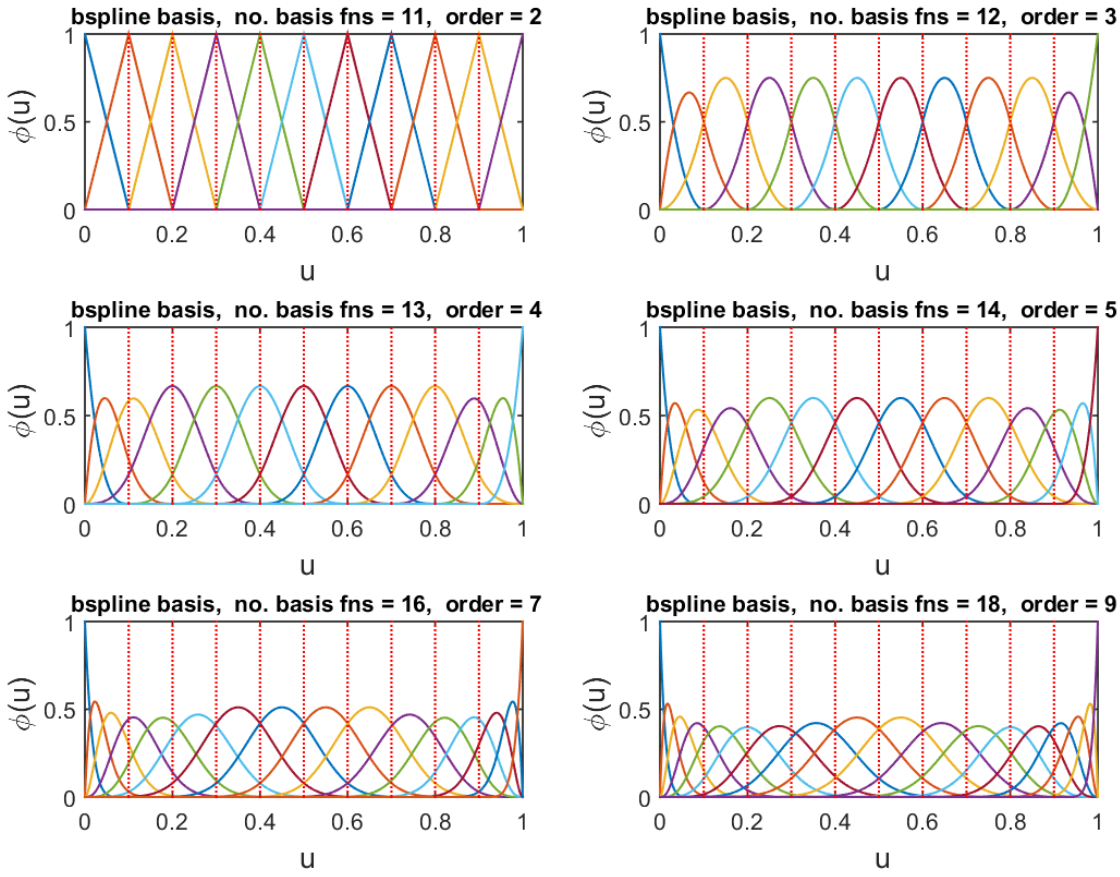

Figure 2.1 B-splines building

III. Geostatistics for functional data

We are interested here to exploit the concept of functional data from *fda* (presented in chapter II) in a continuous space discrete time framework in order to apply geostatistical techniques on curves presenting spatio-temporal dependence. Firstly, we will present the classical functional geostatistical approach found in literature (e.g. Montero et al. (2015), chap.9). Next we will move into a DCM framework presented in section 1.4 and propose our functional approach to dynamic coregionalization model (*F-DCM*).

3.1 Geostatistics for spatially correlated functional data

The functional approach of classical geostatistics for spatio-temporal data is based on the building of spatially indexed curves presenting spatial dependence. The spatial curves are built along the time dimension and since the curves are considered as single entities, the temporal dynamics of the phenomenon under study are hidden and geostatistical techniques can be applied on the spatial entities.

This approach is widely found in literature and applied in several areas. For instance, in climatology see Aguilera-Morillo et al. (2016), Delicado et al. (2010); in oceanology see Monestiez & Nerini (2008), Nerini et al. (2010); in optimal design see Rasekhi et al. (2014); in air pollution monitoring see Montero & Fernández-Avilés (2015), Ignaccolo et al. (2014); in studies on agricultural soils see Cortés-D et al. (2016), in petroleum system modelling and production see Menafoglio et al. (2016). A very interesting review dealing with this approach is Menafoglio & Secchi (2017) where the authors do not treat only the case of spattially correlated functions but the general case of spatially correlated complex data.

Following Cressie (1993) for the definition of spatial process and Ferraty & Vieu (2006) for that of functional random variable, Delicado et al. (2010) define spatial functional process as

$$\{\boldsymbol{\chi}_s(u), u \in U, s \in \mathbb{S} \subseteq \mathbb{R}^d\} \tag{3.1}$$

where $\chi_s(u)$ is a real function from $U = [a, b] \subseteq \mathbb{R}$ to $\mathbb{R}$, s is a generic spatial location in the d-dimensional Euclidean space, the set $\mathbb{S}$ is a fixed subset of $\mathbb{R}^d$ with positive volume and N points $s_1, \ldots, s_N$ in $\mathbb{S}$ are chosen to observe the random functions $\chi_{s_i}(u), i = 1, \ldots, N$.

The spatial functional process (3.1) is said to be second order stationary and isotropic if its mean and variance are constant, and its covariance function depends only on the distance between sampling points. Formally, that is:

$$\mathbb{E}\left[\chi_s(u)\right] = \mu(u), \text{ for all } u \in U, \ s \in \mathbb{S} \tag{3.2}$$

$$\mathbb{V}[\chi_s(u)] = \sigma^2(u), \text{ for all } u \in U, \ s \in \mathbb{S} \tag{3.3}$$

$$\mathbb{C}[\chi_{s_i}(u) - \chi_{s_{i'}}(u')] = \mathbb{C}_h(u, u'), \ s_i, s_{i'} \in \mathbb{S}, \ u, u' \in U, \ h = \|s_i - s_{i'}\|, \tag{3.4}$$

if $u = u'$, $\mathbb{C}[\chi_{s_i}(u) - \chi_{s_j}(u)] = \mathbb{C}_h(u)$

$$\frac{1}{2}\mathbb{V}[\chi_{s_i}(u) - \chi_{s_{i'}}(u')] = \gamma_h(u, u'), s_i, s_{i'} \in D, u, u' \in U, h = \|s_i - s_{i'}\|, \tag{3.5}$$

if $u = u'$,

$$\frac{1}{2}\mathbb{V}[\chi_{s_i}(u) - \chi_{s_{i'}}(u)] = \gamma_h(u) \tag{3.6}$$

which is the variogram.

In this framework the objective is to spatial predict a whole curve. The kriging (see section 1.1.3) has been extended in order to deal with data observed as functions. Thus, the linear unbiased predictor for a curve at an unobserved site of interest is expressed as

$$\hat{\chi}_{s_0}(u) = \sum_{i=1}^{N} \lambda_i \chi_{s_i}(u),\ u \in U = [a, b],\ \lambda_1, \ldots, \lambda_N \in \mathbb{R} \tag{3.7}$$

where the coefficients λ_i are such that:

$$\mathbb{E}(\hat{\chi}_{s_0}(u) - \chi_{s_0}(u)) = 0 \tag{3.8}$$

and

$$\mathbb{E}\left\{\int_a^b [\hat{\chi}_{s_0}(u) - \chi_{s_0}(u)]^2 du\right\} = \int_a^b \mathbb{V}[\hat{\chi}_{s_0}(u) - \chi_{s_0}(u)] du \tag{3.9}$$

is minimized. Thus the optimization problem to be solved is

$$\min_{\lambda_1,\ldots,\lambda_N} \int_a^b \mathbb{V}[\hat{\chi}_{s_0}(u) - \chi_{s_0}(u)] du,\ \ s.t.\ \sum_{i=1}^{N} \lambda_i = 1 \tag{3.10}$$

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

The objective of functional spatial prediction called functional kriging in this framework is faced by keeping or removing the stationary and isotropic assumption of the spatial functional process (3.1).

3.1.1 Functional kriging of stationary spatial processes. Considering a set of curves $\{\chi_{s_i}(u),\ i = 1, \ldots, N\}$ Goulard & Voltz (1993) faced the problem of spatial prediction of curves assuming some restrictive assumptions: (i) the curves are only known only up to a finite set of small number of points $\{u_j, j = 1, \ldots, M\}$; (ii) a parametric model $\chi(.; \theta)$ with a small number of parameter is known. The function χ_{s_i} is then approximated

with the estimated $\chi(.;\hat{\theta}_{s_i})$ and the resulting curve kriging predictor (CKP) can be written as:

$$\hat{\chi}_{s_0}(u) = \sum_{i=1}^{N} \lambda_i \ \chi(.;\hat{\theta}_{s_i}). \tag{3.11}$$

In order to overcome the restrictive assumptions of Goulard & Voltz (1993), some other functional spatial predictor have been proposed for stationary and isotropic processes.

Giraldo et al. (2011) extended the CKP with the ordinary kriging for function-value spatial data ($OKFD$) which approximates the best linear unbiased predictor of $\chi_{s_0}(u)$ in (3.7) with a non-parametric approach through B-spline smoothing. The smoothing parameter is chosen by functional crossvalidation: each functional data location is removed from the dataset and a function is predicted at this location using a functional kriging predictor based on the remaining smoothed function; then the sum of square errors is computed as

$$SSE_{FCV} = \sum_{i=1}^{N} SSE_{FCV}(i) = \sum_{i=1}^{N} \sum_{j=1}^{M} [\hat{\chi}_{s_i}^{(i)}(u_j) - \chi_{s_i}^{(i)}(u_j)]^2, \tag{3.12}$$

where $\hat{\chi}_{s_i}^{(i)}(u_j)$ is the functional kriging prediction on s_i evaluated at u_j, $j = 1, \ldots, M$, by leaving the site s_i temporarily out of the sample. The involved smoothing parameters are chosen by minimization of SSE_{FCV}.

The kriging predictor is given by the solution of the optimization problem (3.10).

Another approach mentioned in Giraldo et al. (2010) considered curves $\chi_{s_i}(u)$ and weights λ_i as functions, both expanded in terms of a set of basis functions (2.1). The coefficients of these basis functions are estimated through the pointwise linear kriging predictor for functional data ($PKFD$) define as

$$\hat{\chi}_{s_0}(u) = \sum_{i=1}^{N} \lambda_i(u)\chi_{s_i}(u),\ i = 1, \ldots, N, \tag{3.13}$$

where the N functional parameters λ_i are given by the solution of the following optimization problem

$$\min_{\lambda_{1(.)},\ldots,\lambda_n(.)} \int_a^b \mathbb{V}[\hat{\chi}_{s_0}(u) - \chi_{s_0}(u)]du,\ s.t.\ \sum_{i=1}^{N} \lambda_i(u) = 1, \text{ for all } u \in [a, b]. \tag{3.14}$$

3.1.2 Functional kriging of non-stationary spatial processes. The case of a non-stationary functional process (3.1) is treated considering the functional mean (3.2) non-constant over space according to the model

$$\chi_{s_i}(u) = \mu_{s_i}(u) + \epsilon_i(u). \tag{3.15}$$

The main difference between the approaches stands in the way the functional mean $\mu_{s_i}(u)$ is estimated.

The commonly used procedure consist to de-trend the estimated mean $\hat{\mu}_{s_i}(u)$ and apply to the functional kriging on the functional residuals expressed as

$$\epsilon_{s_i}(u) = \chi_{s_i}(u) - \hat{\mu}_{s_i}(u), \tag{3.16}$$

where $\mathbb{E}[\epsilon_s(u)] = 0$ and $\mathbb{V}[\epsilon_s(u)] = \sigma^2(u)$ for all $s \in \mathbb{S}$.

And finally predict the curve at the location s_0 as

$$\hat{\chi}_{s_0}(u) = \hat{\mu}_{s_0}(u) + \hat{\epsilon}_{s_0}(u).$$

Reyes et al. (2010) estimate a functional regression model with functional response and scalar covariates using the model (3.15) where $\mu_{s_i}(u) = \alpha(u) + \beta_1(u)x_i + \beta_2(u)y_i$, $\chi_{s_i}(u)$, $i = 1, \ldots, N$ are the functions at visited locations, (x_i, y_i) are geographical scalar coordinates, $\alpha(u), \beta_1(u), \beta_2(u)$ are the functional parameters of interest and $\epsilon(u)$ is a white noise for each $u \in U = [a, b]$. Once estimated the regression model and extracted the residuals, the *OKFD* and *PKFD* presented in the previous section can be applied on the residuals as

$$\hat{\epsilon}_{s_0}(u) = \sum_{i=1}^{N} \lambda_i \epsilon_{s_i}(u) \tag{3.17}$$

$$\hat{\epsilon}_{s_0}(u) = \sum_{i=1}^{N} \lambda_i(u) \epsilon_{s_i}(u). \tag{3.18}$$

Ignaccolo et al. (2014) consider the functional drift into the model (3.15) due to exogenous variables (both scalar and functional) expressed as

$$\mu_{s_i}(u) = \alpha(u) + \sum_p a_p(u) X_{p,i} + \sum_q \beta_q(u) f_{q,i}(u). \tag{3.19}$$

In equation (3.19) $\alpha(u)$ is a functional intercept, $X_{p,i}$ is the p_{th} scalar covariate and $f_{q,i}(u)$ is the q_{th} functional covariate at site s_i, $a_p(u)$ and $\beta_q(u)$ are the covariate coefficients. In the drift, both scalare and functional covariates are included, and the coefficients $a_p(u)$ and $\beta_q(u)$ are also of a functional nature, allowing to estimate non-linear effects of a covariate.

Caballero et al. (2013) proposed the functional universal kriging predictor (*FUKP*) where $\mu_s(u) = \sum_{p=1}^{P} \beta_p(u) f_p(s)$ for all $s \in S$. Consequently, the spatial functional regression model is given by

$$\chi_s(u) = X(u)\beta(u) + \epsilon_s(u). \tag{3.20}$$

The *FUKP* of $\chi_{s_0}(u)$ is then expressed as $\hat{\chi}_{s_0}(u) = \boldsymbol{\lambda}^T \boldsymbol{\chi}_s(u)$ where the components of the weighting vector k are real numbers and are chosen in such a way that *FUKP* is an unbiased predictor of minimum variance.

3.2 Dynamic coregionalization model for spatio-temporally correlated functional data

In this section, our goal is to propose a dynamic coregionalization model for spatio-temporally correlated functional data. We have to put together the concepts of fda and the methodology of DCM in order to build the functional extension of DCM (*FDCM*). The starting point will be the building of spatio-temporally correlated curves from sets of discrete values. Next, we will specify a statistical hierarchical model for the curves, and proceed with model estimation and prediction of curves in the spatio-temporal DCM framework.

3.2.1 From spatio-temporal data to spatio-temporal functional data. As we noted in the previous section, having only space and time dimension in the data do not allow to build spatio-temporally correlated curves. Since the curve is considered as a single entity, the interaction between the dimension along which the function is built and the other one is hidden. Thus, in order to build spatio-temporal functional data we have to consider a third dimension. For our purposes we will consider height as the third dimension but nothing changes in the methodology if the third dimension is depth or something else.

Following Cressie (1993) for the definition of spatial process and Ferraty & Vieu (2006) for that of functional random variable, we can define a spatio-temporal functional process as

$$\{\boldsymbol{\chi}_{s,t}(u), u \in U, s \in \mathbb{S} \subseteq \mathbb{R}^d, t \in \mathcal{T}\} \tag{3.21}$$

where $\chi_{s,t}(u)$ is a spatio-temporal real function from $U = [a, b] \subseteq \mathbb{R}$ to $\mathbb{R}$, s is a generic spatial location in the d-dimensional Euclidean space, t is a generic time step of $\mathcal{T}$ =$1, \ldots, T$. The set $\mathbb{S}$ is a fixed subset of $\mathbb{R}^d$ with positive volume and $(s_1, \ldots, s_N) \times (1, \ldots, T) = N \times T$ points in the spatio-temporal index $\mathbb{S} \times \mathcal{T}$ are chosen to observe the spatio-temporal random functions $\boldsymbol{\chi}_{s_i,t}(u), i = 1, \ldots, N$, $t = 1, \ldots, T$.

Using function basis expansion (2.1), we can write the spatio-temporal function $\chi_{s,t}(u)$ as

$$\chi_{s,t}(u) = \sum_{q}^{Q} c_{s,t}^{q} \phi_q(u) = \mathbf{c}'_{s,t} \boldsymbol{\phi}(u) = \boldsymbol{\phi}'(u) \mathbf{c}_{s,t} \tag{3.22}$$

where $\mathbf{c}_{s,t}$ is the $Q \times 1$ vector of coefficients of the expansion and $\boldsymbol{\phi}$ is a vector of length $Q \times 1$ containing the basis functions.

A spatio-temporal functional datum observed at spatial location s and time t arrives as a set of discrete measured values $y_{s,t,1}, \ldots, y_{s,t,M}$, with M not necessarily equal for all functional data. The first step is to convert these discrete space-time-height measured values to spatio-temporal function $\chi_{s,t}$ built along the height dimension. The observed data are assumed to be noisy observations of the functional variable according to the model

$$y_{s,t,j} = \chi_{s,t}(u_j) + \epsilon_{s,t,j},\ j = 1, \ldots, M. \tag{3.23}$$

The model (3.23) is written better using matrix notation and function basis expansion (3.22) as

$$\mathbf{y}_{s,t} = \chi_{s,t}(u) + \boldsymbol{\epsilon}_{s,t} = \boldsymbol{\phi} \mathbf{c}_{s,t} + \boldsymbol{\epsilon}_{s,t}, \tag{3.24}$$

where $u = u_1, \ldots, u_M$, $\mathbf{y}_{s,t}$ contains de M values to be used in the smoothing process, $\boldsymbol{\phi}$ contains the $M \times Q$ basis function values, $\mathbf{c}_{s,t}$ are the Q coefficients to be estimated (see chapter II, section 3.2), and since the errors $\boldsymbol{\epsilon}_{s,t}$ are assumed to be zero-mean and small enough to be neglectable, we can use the approximation

$$\mathbf{y}_{s,t} \approx \chi_{s,t}(u). \tag{3.25}$$

Hereafter the smoothed curve $\chi_{s,t}(u)$ in (3.25) from the vector of discrete observed points $\mathbf{y}_{s,t}$ will be denoted $\chi^{\mathbf{y}}(u)$ in order to highlight the fact that it is the smoothed curve from $\mathbf{y}$. This specific notation allows us to specify other smoothed curves just changing the subscript as in the case with many functional variables observed.

Note that, with $Q < M$, we are using a dimension reduction approach converting M discrete measured values to a single entity (curve) expressed as a linear combination of Q coefficients of the basis functions used.

3.2.2 Functional dynamic coregionalization model. The smoothed functions in (3.25) are considered as observed functions $\chi^{\mathbf{y}}(u)$, and expressed through basis function expansion

$$\chi^{\mathbf{y}}(u) = \boldsymbol{\phi}'(u)\mathbf{c}^{\mathbf{y}}_{s,t}. \tag{3.26}$$

In the same way, let us consider the fixed effect in (1.69) as coming from b functional variables with basis function expansion given by

$$\chi^{\mu}(u) = X_{s,t}(u)\beta_X = \boldsymbol{\phi}'(u)\mathbf{c}^{X}_{s,t}\beta_X. \tag{3.27}$$

As well, we consider the spatio-temporal random effect in (1.70) as functional with basis function expansion given by

$$\chi^{\omega}(u) = \sum_{p=1}^{c} \alpha_p X_{s,t}^{w_p}(u) w_{p_{s,t}} + X_t^z(u), \tag{3.28}$$

where X^{w_p} are the p functional variables with spatial dependence and X^z are the d functional variables with temporal dependence.

The d temporal functional variables follow Markovian dynamics

$$X_t^z(u) = G X_{t-1}^z(u) + X_t^{\eta}(u),$$

with transition matrix G assumed to have eigenvalues smaller than 1 in absolute value and functional innovations $X_t^{\eta} \sim \mathcal{N}(0, \Sigma_{\eta})$.

The variables w_p are zero-mean and unit-variance independent Gaussian processes uncorrelated over time but correlated over space with Matérn covariance function given by (1.71).

The basis function expansion for χ_{w_p} and χ_z is given by

$$X_{s,t}^{w_p}(u) = \boldsymbol{\phi}'(u) \mathbf{c}_{s,t}^{w_p}$$

$$X_t^z(u) = \boldsymbol{\phi}'(u) \mathbf{c}_t^z.$$

Our proposal F-DCM is then given by

$$\chi^{y}(u) = \chi^{\mu}(u) + \chi^{w}(u) + \chi^{\epsilon}(u) \tag{3.29}$$

which can be write using basis function expansion as

$$\boldsymbol{\phi}'(u) \mathbf{c}_{s,t}^{\mathbf{y}} = \boldsymbol{\phi}'(u) \mathbf{c}_{s,t}^{X} \beta_X + \sum_{p=1}^{c} \alpha_p \boldsymbol{\phi}'(u) \mathbf{c}_{s,t}^{w_p} w_{p_{s,t}} + \boldsymbol{\phi}'(u) \mathbf{c}_t^z + \boldsymbol{\phi}'(u) \mathbf{c}_{s,t}^{\epsilon}. \tag{3.30}$$

We can note that the linear transformation is the same for all the functional variables involved in our functional model. Then the model can be simplified omitting $\boldsymbol{\phi}'(u)$ and only modelling our spatio-temporal DCM on the coefficients

$$\mathbf{c}^{\mathbf{y}}_{s,t} = \mathbf{c}^{X}_{s,t}\beta_X + \sum_{p=1}^{c} \alpha_p \mathbf{c}^{w_p}_{s,t} w_{p_{s,t}} + \mathbf{c}^{z}_{t} + \mathbf{c}^{\epsilon}_{s,t}. \tag{3.31}$$

- $\mathbf{c}^{\mathbf{y}}_{s,t}$ is a Q-variate spatio-temporal Gaussian process of size $N \times T$
- $\mathbf{c}^{X}_{s,t}$ ia a $Q(N \times T \times b)$-dimensional vector for b deterministic functional variables and β_X is the Q-dimensional $(b \times 1)$ vector of parameter to be estimated
- $\mathbf{c}^{w_p}_{s,t}$ is a $Q(N \times T \times p)$-dimensional vector for p functional spatial variables
- $\mathbf{c}^{z}_{t}$ is a $Q(N \times T \times d)$-dimensional vector for d functional temporal variables following a Markovian dynamics.

Furthermore, the model (3.31) can be extended in order to handle functional and non functional variables by inserting the regressor variables in (1.69) and (1.70) into (3.31). The extended F-DCM can be expressed as

$$\mathbf{c}^{y}_{s,t} = X_{s,t}\beta + \mathbf{c}^{X}_{s,t}\beta_X + \sum_{p=1}^{c} \alpha_p \mathbf{c}^{w_p}_{s,t} w_{p_{s,t}} + \sum_{p=1}^{c} \alpha_p X^{w_p}_{s,t} w_{p_{s,t}} + z_t + \mathbf{c}^{z}_{t}, \tag{3.32}$$

where $\mathbf{c}^{\mathbf{y}}_{s,t}, \mathbf{c}^{X}_{s,t}, \mathbf{c}^{w_p}_{s,t}, \mathbf{c}^{z}_{t}$ have the same meaning previously explained and the adding scalar variables keep the same meaning presented in 1.4.

3.2.2.1 Spatial prediction in F-DCM. Let us suppose to have observed $N \times T$ functional data composing our functional dataset

$$\boldsymbol{\chi}(u) = \{\chi_{s_i,t}(u), i = 1, \ldots, N, t = 1, \ldots, T\}. \tag{3.33}$$

The whole dataset can be expressed as

$$\boldsymbol{\chi}(u) = \boldsymbol{\phi}'(u)\mathbf{c}+\epsilon(u),$$

where $\boldsymbol{\chi}(u)$ is the $N \times T$ functional data (3.33), $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_Q)'$ is the vector of cubic B-spline basis functions, $\mathbf{c} = \{c_{s_i,t}, i = 1, \ldots, N, t = 1, \ldots, T\}$ is the $Q(N \times T)$-dimensional vector of spline Gaussian coefficients modelled as (3.31), and $\epsilon(u)$ are the $N \times T$ functional Gaussian errors *iid* supposed to be neglectable for simplicity with $E(\boldsymbol{\epsilon}) = 0$.

The natural functional kriging predictor at unobserved spatio-temporal location $0 = (s_0, t_0) \in \mathbb{S} \times \mathcal{T}$ is given by

$$\hat{\chi}_0(u) = \mathbb{E}\left[\chi_{_0}(u)|\boldsymbol{\chi}(u)\right]. \tag{3.34}$$

The spatial predicted function in FDCM is then

$$\hat{\chi}_0(u) = \mathbb{C}_{(\chi_0,\boldsymbol{\chi})}\mathbb{C}_{\boldsymbol{\chi}}^{-1}\boldsymbol{\chi}(u) \tag{3.35}$$

with uncertainty given by

$$\mathbb{V}\left[\hat{\chi}_0(u)\right] = \mathbb{C}_{(\chi_0,\boldsymbol{\chi})}\mathbb{C}_{\boldsymbol{\chi}}^{-1}\mathbb{C}_{(\boldsymbol{\chi},\chi_0)} \tag{3.36}$$

where $\mathbb{C}_{(\chi_0,\boldsymbol{\chi})} = \mathbb{C}\left[\chi_0(u), \boldsymbol{\chi}(u)\right]$ and $\mathbb{C}_{\boldsymbol{\chi}} = \mathbb{V}\left[\boldsymbol{\chi}(u)\right]$.

Now let us write the left hand side in (3.34) using basis function expansion as

$$\hat{\chi}_0(u) = \sum_{q=1}^{Q} \hat{c}_0^q \phi_q(u) = \hat{\mathbf{c}}_0'\boldsymbol{\phi}(u). \tag{3.37}$$

As already specified, our approach is based on coefficients and moving from coefficients to functional data is just a linear transformation though the matrix of transformation

$\boldsymbol{\phi}'$ made up of B-spline basis functions. Then the functional kriging predictor for in FDCM (3.34) can expressed in terms of coefficients as

$$\hat{\mathbf{c}}_0 = \mathbb{E}(\mathbf{c}_0|\mathbf{c}), \tag{3.38}$$

where $\mathbf{c}_0$ is a Q-dimensional vector of coefficients of the curve to be predicted at spatio-temporal location $(s_0, t_0) \in \mathbb{S} \times \mathcal{T}$.

It follows that the Q spatial predicted coefficients at unobserved spatio-temporal location $0 = (s_0, t_0) \in \mathbb{S} \times \mathcal{T}$ has the form

$$\hat{\mathbf{c}}_0 = \boldsymbol{\Sigma}_{(\mathbf{c}_0,\mathbf{c})}\boldsymbol{\Sigma}_{\mathbf{c}}^{-1}\mathbf{c} \tag{3.39}$$

with uncertainty given by a $Q \times Q$ variance-covariance matrix

$$\hat{\boldsymbol{\Sigma}}_0 = \mathbb{V}(\hat{\mathbf{c}}_0) = \boldsymbol{\Sigma}_{(\mathbf{c}_0,\mathbf{c})}\boldsymbol{\Sigma}_{\mathbf{c}}^{-1}\boldsymbol{\Sigma}_{(\mathbf{c},\mathbf{c}_0)} \tag{3.40}$$

where $\boldsymbol{\Sigma}_{(\mathbf{c}_0,\mathbf{c})} = \mathbb{C}(\mathbf{c}_0, \mathbf{c})$, $\boldsymbol{\Sigma}_{\mathbf{c}} = \mathbb{V}(\mathbf{c})$.

Now using the matrix notation of the basis function expansion of the predicted curve on the right hand side in (3.37) and the variance-covariance matrix of predicted coefficients in (3.40), we can derive the variance-covariance surface of the predicted curve as

$$\mathbb{V}[\hat{\chi}_0(u', u)] = \boldsymbol{\phi}'(u')\hat{\boldsymbol{\Sigma}}_0\boldsymbol{\phi}(u) \tag{3.41}$$

and the variance of the predicted curve when $u' = u$ is given by

$$\mathbb{V}[\hat{\chi}_0(u)] = \boldsymbol{\phi}'(u)\hat{\boldsymbol{\Sigma}}_0^*\boldsymbol{\phi}(u) \tag{3.42}$$

where $\hat{\boldsymbol{\Sigma}}_0^*$ is the main diagonal of $\hat{\boldsymbol{\Sigma}}_0$.

3.2.2.2 Crossvalidation in F-DCM. Crossvalidation in F-DCM takes into account four levels of goodness: the goodness of smoothing, the goodness of prediction, the goodness of prediction between curves and the goodness of simplification. The levels of goodness are assessed at observed height levels $j = 1, \ldots, M_i$, for the observations $i = 1, \ldots, N$ except for the goodness of simplification where the height levels do not have to be necessary the observed ones. The goodness of smoothing assesses the measured points $\mathbf{y}$ and the smoothed curves $\boldsymbol{\chi}$. The goodness of prediction assesses the measured points $\mathbf{y}$ and the predicted curves $\hat{\boldsymbol{\chi}}$. The goodness of prediction between curves assesses how the smoothed curves $\boldsymbol{\chi}$ and the predicted curves $\hat{\boldsymbol{\chi}}$ are close each other at the measured height levels. The goodness of simplification measures the gain or loss of performance in prediction given by moving from infinite dimensional functional data $\boldsymbol{\chi}$ to multivariate finite dimensional data through the linear transformation $\boldsymbol{\phi}\mathbf{c}$. It assesses the impact of simplification in the performance comparing the non simplified predicted curves $\hat{\boldsymbol{\chi}}^*$ with the predicted curves $\hat{\boldsymbol{\chi}}$ computed through the simplification $\boldsymbol{\phi}\hat{\mathbf{c}}$.

There are two different approaches for assessing the goodness of smoothing, prediction and between the curves: the approach by single observation and the approach by height level.

The approach by single observation computes the indexes of goodness for the single functional datum i at its height levels observed M_i. The three levels of goodness in terms of bias, mean square error and R square are expressed as:

$$bias_s^{(i)} = \frac{1}{M_i} \sum_{j=1}^{M_i} \left(y_j^{(i)} - \chi_j^{(i)} \right), \tag{3.43}$$

$$bias_p^{(i)} = \frac{1}{M_i} \sum_{j=1}^{M_i} \left(y_j^{(i)} - \hat{\chi}_j^{(i)} \right), \tag{3.44}$$

$$bias_c^{(i)} = \frac{1}{M_i}\sum_{j=1}^{M_i}\left(\chi_j^{(i)} - \hat{\chi}_j^{(i)}\right), \tag{3.45}$$

$$mse_s^{(i)} = \frac{1}{M_i}\sum_{j=1}^{M_i}\left(y_j^{(i)} - \chi_j^{(i)}\right)^2, \tag{3.46}$$

$$mse_p^{(i)} = \frac{1}{M_i}\sum_{j=1}^{M_i}\left(y_j^{(i)} - \hat{\chi}_j^{(i)}\right)^2, \tag{3.47}$$

$$mse_c^{(i)} = \frac{1}{M_i}\sum_{j=1}^{M_i}\left(\chi_j^{(i)} - \hat{\chi}_j^{(i)}\right)^2, \tag{3.48}$$

$$R_s^{2(i)} = 1 - \frac{mse_s^{(i)}}{\mathbb{V}(\mathbf{y}^{(i)})}, \tag{3.49}$$

$$R_p^{2(i)} = 1 - \frac{mse_p^{(i)}}{\mathbb{V}(\mathbf{y}^{(i)})}, \tag{3.50}$$

$$R_c^{2(i)} = 1 - \frac{mse_c^{(i)}}{\mathbb{V}(\chi^{(i)})}, \tag{3.51}$$

where the subscripts s, p and c stand for smoothing, prediction and prediction between the curves, respectively. The root mean square error is computed as $\sqrt{mse_s^{(i)}}$, $\sqrt{mse_p^{(i)}}$and $\sqrt{mse_c^{(i)}}$. $\mathbb{V}(\mathbf{y}^{(i)})$ is the variance between the M_i measured points of the ith observation and $\mathbb{V}(\chi^{(i)})$ is the variance of the smoothed curve i evaluated at M_i height levels.

The second approach considers a height interval U[1] and computes the indexes taking into account all the observed points and the values of the functions at those points. The

[1] If the height of measurement points are the same for all the observations, then U can be fixed to a specific level.

three levels of goodness in terms of bias, mean square error and R square are expressed in the approach by height level as:

$$bias_s^{(U)} = \frac{1}{\sum_i M_i^{(U)}} \sum_{i=1}^{N} \sum_{j=1}^{M_i^{(U)}} \left(y_{ij}^{(U)} - \chi_{ij}^{(U)} \right), \tag{3.52}$$

$$bias_p^{(U)} = \frac{1}{\sum_i M_i^{(U)}} \sum_{i=1}^{N} \sum_{j=1}^{M_i^{(U)}} \left(y_{ij}^{(U)} - \hat{\chi}_{ij}^{(U)} \right), \tag{3.53}$$

$$bias_c^{(U)} = \frac{1}{\sum_i M_i^{(U)}} \sum_{i=1}^{N} \sum_{j=1}^{M_i^{(U)}} \left(\chi_{ij}^{(U)} - \hat{\chi}_{ij}^{(U)} \right), \tag{3.54}$$

$$mse_s^{(U)} = \frac{1}{\sum_i M_i^{(U)}} \sum_{i=1}^{N} \sum_{j=1}^{M_i^{(U)}} \left(y_{ij}^{(U)} - \chi_{ij}^{(U)} \right)^2, \tag{3.55}$$

$$mse_p^{(U)} = \frac{1}{\sum_i M_i^{(U)}} \sum_{i=1}^{N} \sum_{j=1}^{M_i^{(U)}} \left(y_{ij}^{(U)} - \hat{\chi}_{ij}^{(U)} \right)^2, \tag{3.56}$$

$$mse_c^{(U)} = \frac{1}{\sum_i M_i^{(U)}} \sum_{i=1}^{N} \sum_{j=1}^{M_i^{(U)}} \left(\chi_{ij}^{(U)} - \hat{\chi}_{ij}^{(U)} \right)^2, \tag{3.57}$$

$$R_s^{2(U)} = 1 - \frac{mse_s^{(U)}}{\mathbb{V}(y^{(U)})}, \tag{3.58}$$

$$R_p^{2(U)} = 1 - \frac{mse_p^{(U)}}{\mathbb{V}(y^{(U)})}, \tag{3.59}$$

$$R_c^{2(U)} = 1 - \frac{mse_c^{(U)}}{\mathbb{V}(\chi^{(U)})}, \tag{3.60}$$

where the subscripts s, p and c stand for smoothing, prediction and prediction between the curves, respectively. The root mean square error is computed as $\sqrt{mse_s^{(U)}}$, $\sqrt{mse_p^{(U)}}$and $\sqrt{mse_c^{(U)}}$; $\mathbb{V}(\mathbf{y}^{(U)})$ is the variance between all the observed points within the interval U and $\mathbb{V}(\chi^{(U)})$ is the variance between the values of the smoothed curves at the height levels of the observed points within the interval U.

The two approaches are useful in a different way. The first one can be useful to detect some misspecified curves or outlier profiles. For example we can fix a threshold level for a goodness index so that the functional data with goodness index over the threshold level are considered outliers. The second one has a better interpretation along the overall vertical profile and can allow to detect misspecified interval along the profile. From both approaches we can build synthetic indexes.

From the approach by observation synthetic indexes can be expressed as the averaged indexes for smoothing, prediction and prediction between curves

$$\overline{bias} = \sum_{i=1}^{N} bias^{(i)},$$

$$\overline{mse} = \sum_{i=1}^{N} mse^{(i)},$$

$$\overline{R^2} = \sum_{i=1}^{N} R^{2(i)}.$$

From the approach by height level, it does not have any sense to average the indexes over the height levels but it is more useful to relax the height interval to the whole range considered[2]. In that case we do not have to specify the height in sup-script because all the

[2] As in the case of a specific interval, the range has to be equal for all the observation with same minimum and maximum.

data are taken into account. The resulting bias, mean square error and R square are the same without the supscript U:

$$bias_s = \frac{1}{\sum_i M_i} \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left(y_{ij} - \chi_{ij} \right), \tag{3.61}$$

$$bias_p = \frac{1}{\sum_i M_i} \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left(y_{ij} - \hat{\chi}_{ij} \right), \tag{3.62}$$

$$bias_c = \frac{1}{\sum_i M_i} \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left(\chi_{ij} - \hat{\chi}_{ij} \right), \tag{3.63}$$

$$mse_s = \frac{1}{\sum_i M_i} \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left(y_{ij} - \chi_{ij} \right)^2, \tag{3.64}$$

$$mse_p = \frac{1}{\sum_i M_i} \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left(y_{ij} - \hat{\chi}_{ij} \right)^2, \tag{3.65}$$

$$mse_c = \frac{1}{\sum_i M_i} \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left(\chi_{ij} - \hat{\chi}_{ij} \right)^2, \tag{3.66}$$

$$R_s^2 = 1 - \frac{mse_s}{\mathbb{V}(y)}, \tag{3.67}$$

$$R_p^2 = 1 - \frac{mse_p}{\mathbb{V}(y)}, \tag{3.68}$$

$$R_c^2 = 1 - \frac{mse_c}{\mathbb{V}(\chi)}. \tag{3.69}$$

The goodness of simplification takes into account the values of the predicted curves at prefixed height level $\hat{\boldsymbol{\chi}}(u_j)$ and its counterpart without simplification $\hat{\boldsymbol{\chi}}^*(u_j)$. The resulting bias, mean square error and R square are computed as

$$bias^* = \frac{1}{N}\sum_{i=1}^{N}\left[\hat{\chi}_i(u_j) - \hat{\chi}_i^*(u_j)\right],$$

$$mse^* = \frac{1}{N}\sum_{i=1}^{N}\left[\hat{\chi}_i(u_j) - \hat{\chi}_i^*(u_j)\right]^2,$$

$$R^{2*} = 1 - \frac{mse^*}{\mathbb{V}\left[\boldsymbol{\chi}(u_j)\right]},$$

where $\hat{\chi}(u_j)$ is the value of the predicted curve at height level j given by our approach and computed with the simplification $\phi(u_j)\hat{c}$, $\hat{\chi}^*(u_j)$ is the value of predicted curve computed without simplification but given the values of the observed functions $\boldsymbol{\chi}^*(u_j)$ at height level j. In other words, the goodness of simplification assesses the impact of simplification of modelling the coefficients of the functions instead of modelling the values of functions at fixed heights .

IV. Case study

Here we are interested in the atmospheric vertical profile of an essential climate variable (*ecv*) such as temperature taken from global time series of radiosondes given by the RAOB network (retrieved from www.raob.com). We consider the global time series of 9824 radiosondes launched at night[1] during the year 2015 from 83 monitoring weather stations located within the geographic box $(36.72°N, 67.37°N, -6.77°W, 34.8E)$ as shown in Figure 4.1.

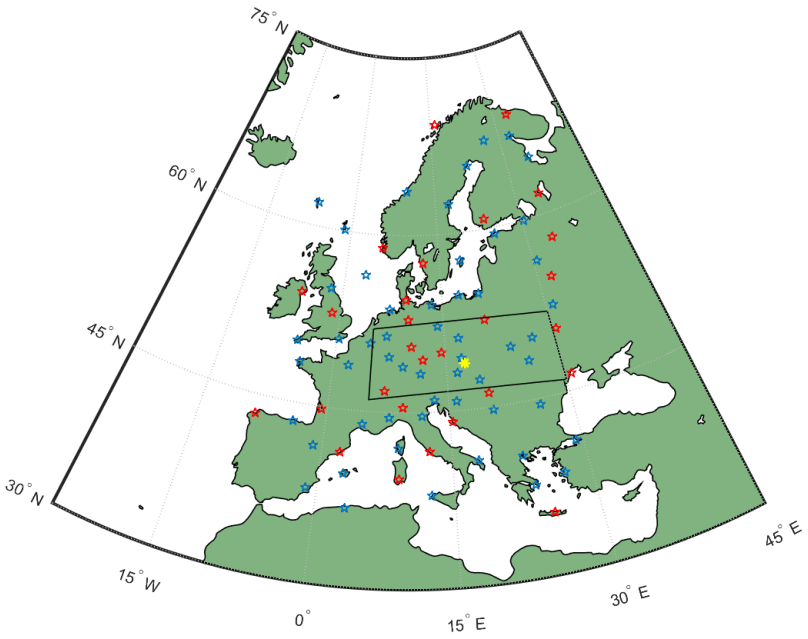

Figure 4.1 Spatial location of the monitoring weather stations.

The radiosonde launches are not equally distributed neither over the weather stations nor over the days[2] as we can see in Figure 4.2. With only one launch considered at night,

[1] Since the considered radiosondes are launched between 22 o'clock and 1 o' clock, with some abuse of language the term night is used to refer that interval of time

[2] Since the time interval of launches covers two day, the related night is assigned to the previous day.

the number of radiosonde launches from a fixed weather station is also its number of days with an available launch, namely its temporal coverage of the year. In the same way, the number of radiosonde launches in a day is also its number of stations with a launched radiosonde, namely its spatial coverage of the region of interest shown in Figure 4.1.

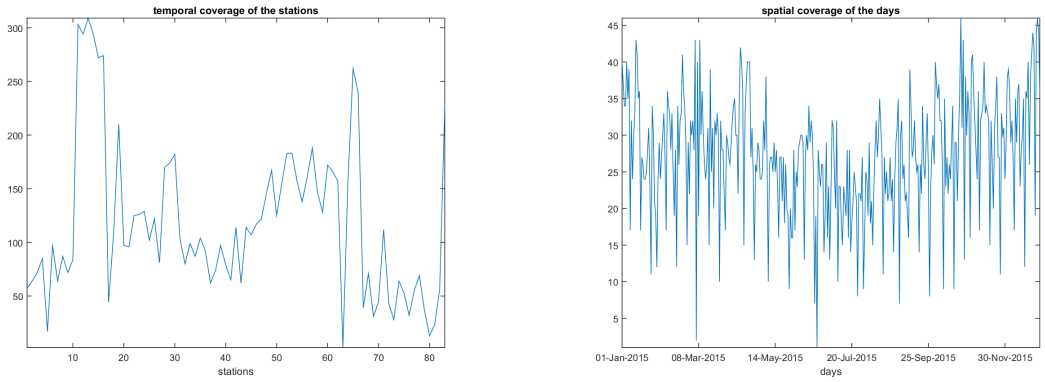

Figure 4.2 Spatio-temporal distribution of radiosondes. Left panel: number of radiosondes launched by station. Right panel: number of radiosondes launched by day.

The temperature, in degree Celsius ($°C$), is measured along the height while the launched radiosondes are going up and neither the number nor the height of measurements is the same for all radiosonde launches. On the left of Figure 4.3 we can see the different maximum heights reached by each radiosonde while in the right we can see the measured vertical points of temperature with respect to the height of measurements.

Our goal is to predict a temperature vertical profile along the height in a spatio-temporal location without any radiosonde launch. We will proceed in this task through a three step procedure: i) the building of vertical profiles of temperature from discrete measured points; ii) the estimation of the F-DCM parameters on the built profiles; iii) the crossvalidation and mapping. The vertical profile building step will be computed for all radiosonde launches available in the dataset. The estimation step will be computed on the

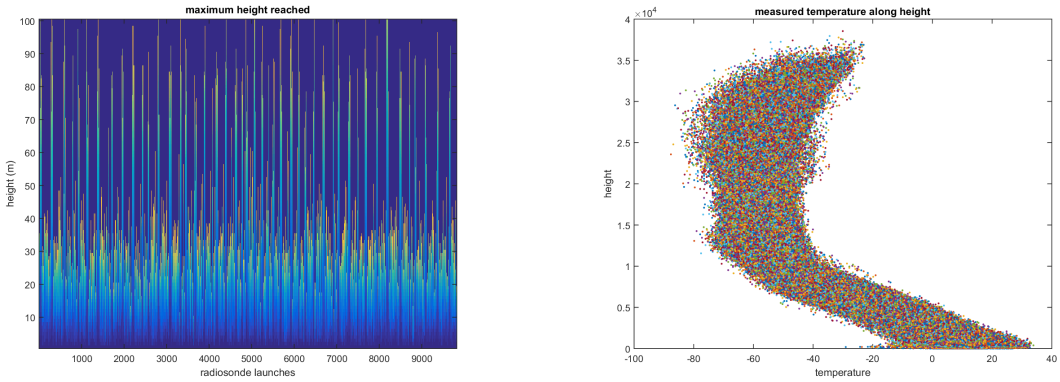

Figure 4.3 Distribution in height of the radiosondes: Left panel: maximum height reached by each radiosonde. Right panel: vertical measured points of temperature.

2/3 of the monitoring weather stations highlighted with blue stars in Figure 4.1 while the remaining 1/3 will be used for crossvalidation highlighted with red stars in Figure 4.1.

4.1 Vertical profile building

From the vertical measured points of temperature shown on the right of Figure 4.3 we have to build the vertical profile of temperature. Since the function is built along the height, it is important to uniform the height domain for all the radiosondes. In order to overcome the differences in the height domain, the minimum and maximum height, used as external knots points, are kept to 600 and 20000 meters, respectively, for all radiosonde launches. Similarly, the internal knot points[3] are kept at 2443, 4199, 6216, 8641, 11736 and 15499 meters for all observations in height.

Once internal and external knots points are fixed for all radiosondes, an interpolation between the observed data points is computed through a cubic Hermite spline in order to generate data within the discrete observed points. There are some benefits from using

[3]For simplicity, the number of internal knot points is kept to 6 in order to assure the computational feasibility.

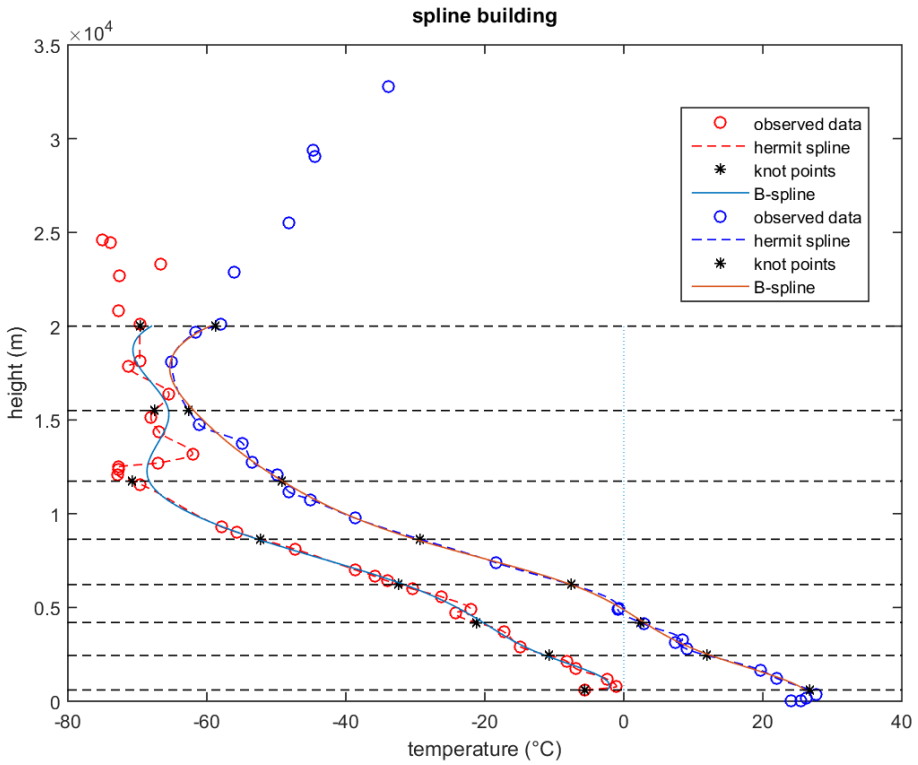

Figure 4.4 Steps of the vertical profile building.

the cubic Hermite interpolator. Firstly, the interpolated points pass through the observed points and that allows us to keep the observed points in the dataset useful for crossvalidation. Secondly, the presence of a datum is ensured at a knot point height. But the cubic Hermite splines match only on the first derivatives through internal knot points while for our purpose we require more smoothness. Therefore from cubic Hermite interpolated points we will build cubic B-splines with second derivatives matching at the internal knot points. The steps of building a vertical profile from discrete measured points are shown in Figure 4.4 with two radiosonde launches as example.

The data are smoothed with roughness penalty and the smoothing parameter is fixed at 0.001 in order to let the function free to fit the data as closely as possible. Since we have

6 internal knot points, the number of B-splines in the basis system is equal to 10 according to (2.3) and the order is 4 since we are using cubic B-splines.

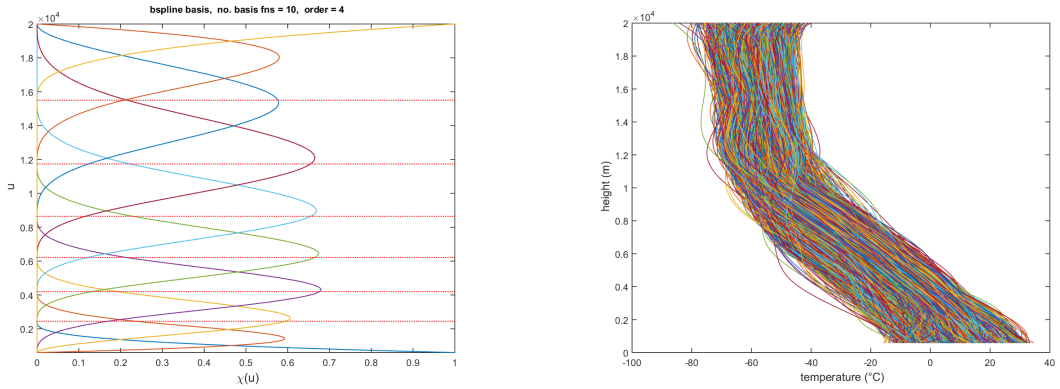

Figure 4.5 B-spline basis and vertical profile of temperature. Left panel: 10 cubic B-spline basis with internal knots knot points highlighted by straight lines. Right panel: vertical profile built.

On the left of Figure 4.5 we see the relationship between the cubic B-spline basis functions and the knot points. When the mode of a basis function is matching a knot point, we can say that the coefficient of this basis function gives the main contribution for the phenomenon under study at the height level of this knot points. On the right of Figure 4.5 we have the result of the building profile procedure with obtained root mean square error equal to 1.21. Since the vertical profiles shown on the right of Figure 4.5 are built from the observed temperature along the height as shown on the right of Figure 4.3, hereafter they are considered as the observed vertical profiles of temperature.

4.2 Estimation

The vertical profiles are modelled through the coefficients of the B-spline functions. In a preliminary analysis we have considered various models. The first one had only stochastic intercept in spatio-temporal random effects. This model presented a seasonality in the temporal random effect which we extracted in order to use it as fixed effect in the

final model. Another models tried various combinations of latitude and longitude in the fixed effect and in the spatial effect resulting in a non significance of longitude in both components. The chosen model is a specific case of (3.32) without functional variables among the regressors. It is represented by the following 10-variate spatio-temporal DCM

$$\mathbf{c}_{s,t} = X_{s,t}\beta + z_t + \alpha X_{s,t}^{w_p} w_{p_{s,t}} + \epsilon_{s,t}, \tag{4.1}$$

where $\mathbf{c}_{s,t}$ are the spatio-temporal coefficients of the B-spline functions, $X_{s,t}$ is a bidimensional covariate vector containing the latitude and the seasonality, β is bidimensional coefficient vector, z_t is a stochastic temporal intercept such that $z_t = Gz_{t-1} + \eta_t$, $X_{s,t}^{w_p}$ is the latitude, α is the spatial coefficient of latitude effect given by a Gaussian LCM, ϵ and η are Gaussian errors *iid* in space and time.

Table 4.1 Estimated parameters with their standard deviations in brackets

	Height (km)									
	0.6	1.2	2.4	4.2	6.2	8.6	11.7	15.5	18	20
$\hat{\beta^1}$	-0.533	-0.533	-0.469	-0.463	-0.418	-0.360	0.059	0.367	0.407	0.290
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)	(0.005)
$\hat{\beta^2}$	0.631	0.507	0.508	0.560	0.620	0.612	0.238	0.387	0.560	0.774
	(0.087)	(0.084)	(0.074)	(0.078)	(0.063)	(0.072)	(0.088)	(0.093)	(0.071)	(0.087)
$\hat{\alpha}$	0.329	0.409	0.487	0.451	0.524	0.528	0.755	0.686	0.587	0.478
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
$\hat{G}$	0.907	0.863	0.823	0.786	0.744	0.715	0.713	0.792	0.805	0.893
	(0.026)	(0.034)	(0.039)	(0.045)	(0.049)	(0.056)	(0.059)	(0.048)	(0.043)	(0.035)
$\hat{\Sigma}_\eta$	0.022	0.041	0.050	0.073	0.067	0.100	0.151	0.105	0.064	0.037
	(0.003)	(0.005)	(0.005)	(0.007)	(0.006)	(0.009)	(0.013)	(0.010)	(0.006)	(0.004)
$\hat{\sigma}_\epsilon^2$	0.170	0.223	0.205	0.147	0.149	0.169	0.241	0.223	0.199	0.128
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)	(0.002)

Table 4.2 Correlation matrix V of spatial random effect expressed as a function of altitude (km)

	Height (km)									
	0.6	1.2	2.4	4.2	6.2	8.6	11.7	15.5	18	20
C_1	1									
C_2	0.676 (0.013)	1								
C_3	0.479 (0.009)	0.249 (0.008)	1							
C_4	0.250 (0.008)	0.559 (0.008)	0.341 (0.006)	1						
C_5	0.233 (0.007)	0.120 (0.006)	0.685 (0.006)	0.482 (0.005)	1					
C_6	0.206 (0.007)	0.355 (0.007)	0.245 (0.005)	0.560 (0.006)	-0.016 (0.004)	1				
C_7	-0.110 (0.006)	-0.200 (0.006)	-0.362 (0.005)	-0.520 (0.005)	-0.446 (0.004)	-0.558 (0.004)	1			
C_8	-0.099 (0.006)	-0.175 (0.006)	-0.138 (0.005)	-0.139 (0.004)	-0.081 (0.004)	-0.142 (0.004)	-0.386 (0.003)	1		
C_9	-0.252 (0.007)	-0.272 (0.004)	-0.264 (0.005)	-0.171 (0.005)	-0.275 (0.004)	-0.098 (0.004)	0.375 (0.004)	-0.285 (0.004)	1	
C_{10}	-0.153 (0.007)	-0.195 (0.004)	-0.016 (0.005)	-0.058 (0.005)	-0.032 (0.004)	-0.065 (0.004)	-0.001 (0.003)	0.566 (0.004)	0.227 (0.004)	1

The estimates of model 4.1 and their standard deviations in brackets are reported in Tables 4.1 and 4.2 with the headers being the heights in kilometres related to the coefficients of the B-splines functions. In Table 4.1, $\hat{\beta}^1$ and $\hat{\beta}^2$ are the estimated fixed effects of latitude and seasonality, $\hat{\alpha}$ is the estimated spatial coefficient of latitude, $\hat{G}$ is the estimated temporal coefficient of a stochastic intercept with Markovian dynamic, $\hat{\Sigma}_\eta$ is the estimated variance of innovation and $\hat{\sigma}^2_\epsilon$ is estimated variance of error.

We can note the change of sign at 11.7 km of the coefficient $\hat{\beta}^1$ related to the latitude in the fixed effects showing in the table 4.1. That means before this height level the temperature decreases when we go north in latitude but this relation is inverted when we reach 11.7 km. The coefficient spatial coefficient $\hat{\alpha}$ of latitude is based on a LCM and a Matèrn correlation with $\nu = 1/2$. The estimated spatial correlation range $\hat{\theta}$ is 1277 km with a standard deviation of 3 km.

In table 4.2 the elements of the correlation matrix $\mathbf{V}$ for the latitude can be interpreted as the cross-correlations between the height levels related to the coefficients of the B-spline functions. We can note again the sign of change at 11.7 km which has the same meaning already explained with the $\hat{\beta}^1$.

4.3 Crossvalidation

The red stars in Figure 4.1 are the remaining 1/3 of monitoring weather stations used for crossvalidation. We consider the indexes presented in section 3.2.2.2 beginning with the two approaches of the goodness of smoothing, prediction and between the curves.

For the approach by height level, it is impossible to analyse the performance of the model for a fixed height because the temperature have not been measured at the same height for all the radiosondes. In order to overcome that limit we have to take into account an interval of height instead of a fixed height. For our purpose we consider the interval of height between the internal knot points. Then we compute the indexes taking into account

all the observed, smoothed and predicted values within the intervals. In Table 4.3 we have the performance of the model interval by interval and it is possible to identify where the predictions are better or worse. We can note the decreasing performance above 11 kilometres also confirmed in Figure 4.6 where in the left we have the trend of functional $RMSE_c$ and in the right the trend of functional R_c^2.

Table 4.3 Crossvalidation table with approach by height level

	$BIAS_s$	$BIAS_c$	$BIAS_p$	$RMSE_s$	$RMSE_c$	$RMSE_p$	R_s^2	R_c^2	R_p^2
$0.6 - 2.4km$	-0.067	-0.144	-0.211	1.005	3.097	3.266	0.982	0.831	0.815
$2.4 - 4.2km$	0.002	-0.243	-0.241	0.804	3.046	3.135	0.989	0.836	0.826
$4.2 - 6.2km$	-0.128	-0.109	-0.237	0.845	3.031	3.117	0.990	0.864	0.861
$6.2 - 8.6km$	0.070	-0.031	0.039	0.934	2.988	3.035	0.989	0.892	0.884
$8.6 - 11.7km$	-0.379	-0.065	-0.444	1.499	2.386	2.984	0.960	0.882	0.840
$11.7 - 15.5km$	0.278	-0.208	0.070	1.844	2.823	3.362	0.897	0.725	0.658
$15.5 - 20km$	-0.079	0.095	0.016	1.483	1.887	2.382	0.919	0.860	0.791

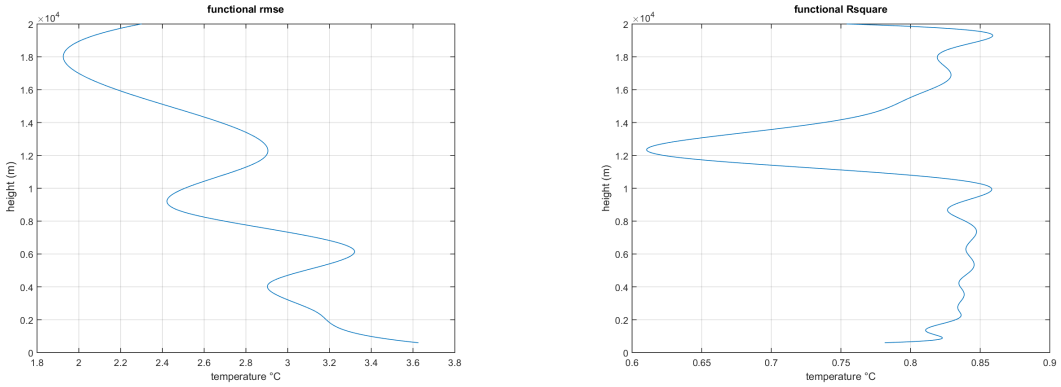

Figure 4.6 Functional performance. Left panel: functional root mean square error between curves. Right panel: functional R square between curves.

The decreasing performance of the model above 11 kilometres can be explained by the instability of the environment above this level of altitude or by the lack of knots points placed from 11 kilometres to 20 kilometres. In fact from 600 meters to 8 kilometres we have 5 knot points and from 8 kilometres to 20 kilometres we only have 3 knot points as we can see in Figure 4.4.

As already said before in the approach by observation the indexes are computed on the single observations. It can be useful to assess how the model performs on a single observation in order to detect outliers or misspecified observations. The results of this approach are difficult to be shown numerically but graphically it is intuitively more interpretable as shown in Figure 4.7 for *bias* and $RMSE$.

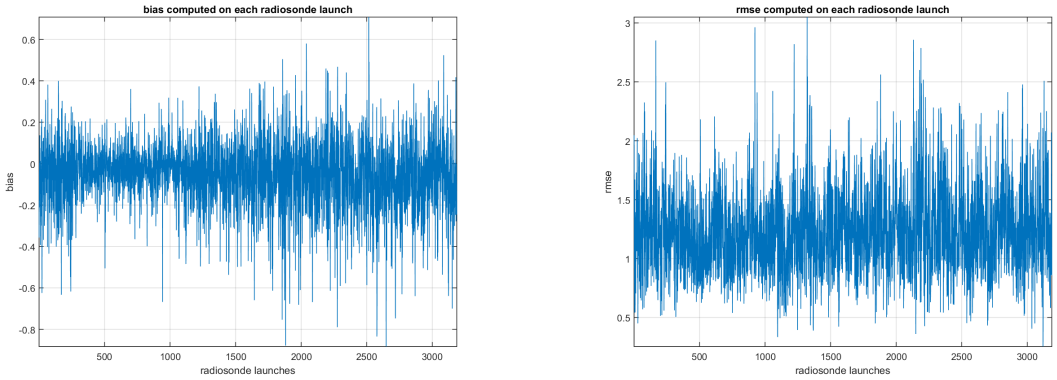

Figure 4.7 Goodness by radiosonde. Left panel: bias by radiosonde. Right panel: rmse by radiosonde.

A possible use of this approach is to fix a threshold level of $RMSE$ and consider as outliers the observations with $RMSE$ higher than the threshold level. However this approach is too observation specific while we need for our purpose to assess the overall performance of the model. A possible solution to have an overall index from this approach is to average the single indexes of the observations. The averaged indexes from this crossvalidation approach are shown on the right of Table 4.4.

Table 4.4 Crossvalidation table. Left side: by the whole height range. Right side: by the mean of indexes computed on the single observation.

	Approach by height			Approach by osservation		
	$BIAS$	$RMSE$	R^2	$\overline{BIAS}$	$\overline{RMSE}$	$\overline{R^2}$
Smoothing	-0.043	1.256	0.997	-0.052	1.225	0.997
Between curves	-0.100	2.786	0.988	-0.175	2.404	0.983
Prediction	-0.143	3.055	0.985	-0.228	2.772	0.980

For the approach by height, an overall index can also be built by relaxing the interval between the knots to the whole selected range. The index is then built between all the observed, smoothed and predicted values and the results are presented on the left of table 4.4.

Another perspective for crossvalidating our model is to assess the implication of simplification in the prediction. The question here is: what is the effect of modelling coefficients instead of modelling the real value of the function in term of performance? The answer is given in Table 4.5 where the indexes of performance have been computed at the knot point heigh levels.

Table 4.5 Implication of simplification in the performance. Left side: $\hat{\chi}$ is predicted through coefficients of splines. Right side: $\hat{\chi}^*$ is predicted through the value of the functions.

	$\hat{\chi}$			$\hat{\chi}^*$		
	$BIAS$	$RMSE$	R^2	$BIAS^*$	$RMSE^*$	R^{2*}
$0.6km$	-0.031	3.724	0.782	0.080	3.726	0.781
$2.4km$	-0.035	3.080	0.836	0.041	3.000	0.839
$4.2km$	-0.031	3.009	0.835	0.046	2.919	0.852
$6.2km$	-0.045	3.244	0.840	0.090	3.181	0.836
$8.6km$	-0.021	2.766	0.827	0.009	2.733	0.853
$11.7km$	0.102	3.051	0.649	0.117	2.983	0.683
$15.5km$	-0.110	2.269	0.799	0.110	2.214	0.795
$20km$	-0.138	2.625	0.754	0.144	2.635	0.754

Modelling the real value of the function gives lightly better performance in terms of R^2 and $RMSE$ as shown in the right side of Table 4.5, but it is computationally unfeasible for the whole function. The performance of our approach shown in the left side of Table 4.5 is better in term of bias and close to the other one in terms of R^2 and $RMSE$. What makes our approach better is its neglectable loss of performance and its computational feasibility for the whole function given by the dimensional reduction of the simplification.

4.4 Mapping

The geographical area for mapping shown in Figure 4.1 as a black rectangle is located within the geographic box $(46°N, 52°N, 5°W, 30E)$ and divided in pixels of 5×5 km. In each pixel s_0 and each time step t_0 the vertical profile of temperature is predicted at spatio-temporal location $0 = (s_0, t_0)$ as (3.37).

The map of the predicted temperatures is computed as $\hat{\chi}_{s_0,t_0}(u_0)$ with t_0 and u_0 fixed. Figure 4.8 is an illustration with t_0 fixed on the 1st July 2015 and u_0 fixed at 8.6 and 11.7 km of altitude.

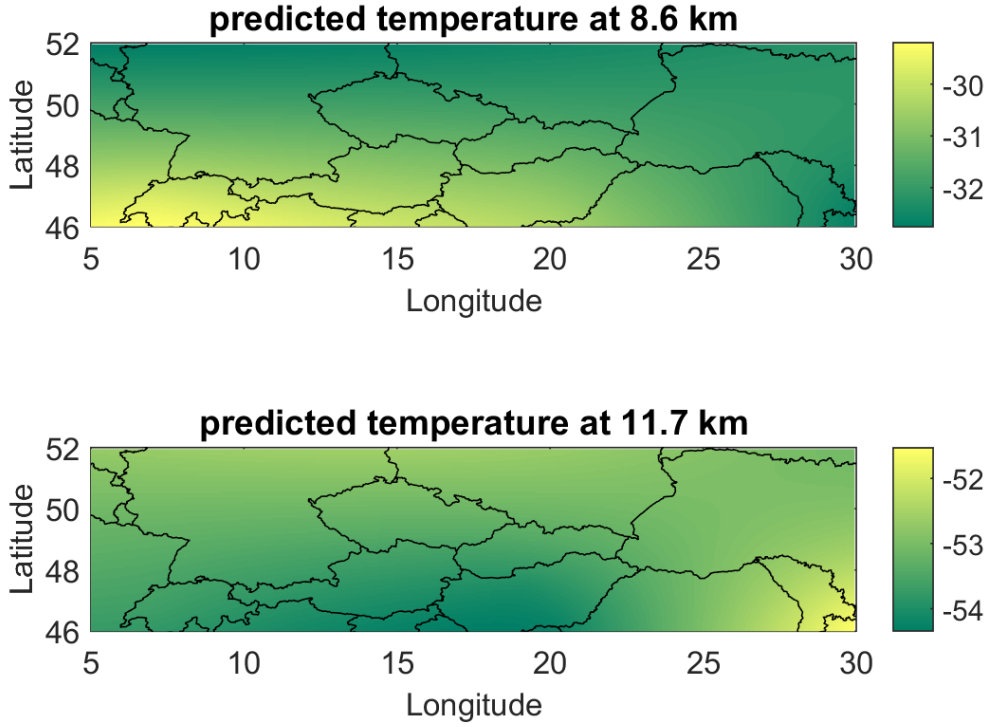

Figure 4.8 Predicted temperature in $°C$ on 1st July 2015 at 8.6 km and 11.7 km of altitude.

As we expected the temperature is lower at 11.7 kilometres of altitude than at 8.6 kilometres of altitude confirming that vertical profile of temperature is a decreasing function of height.

Unfortunately the variance-covariance surface of prediction (3.41) cannot be computed because the software used for computation (DSTEM) only gives up in output the elements $\hat{\mathbf{\Sigma}}_0^{q,q}$ of the main diagonal of $\hat{\mathbf{\Sigma}}_0$ in (3.40) related to the variance of predicted coefficients at each height level q.

In order to compute $\hat{\mathbf{\Sigma}}_0^{q,q'}$, with $q \neq q'$, the related covariance between the predicted coefficients at different height level, we will consider for simplicity that the correlation between the coefficients is given by the estimated correlation matrix $\hat{V}$ of the model already presented in Table 4.2. Hence, the elements out of the main diagonal of the variance-covariance matrix for the predicted coefficients at spatio-temporal location $0 = (s_0, t_0)$ can be expressed as

$$\hat{\mathbf{\Sigma}}_0^{q,q'} = \hat{V}^{q,q'} \sqrt{\hat{\mathbf{\Sigma}}_0^{q,q}} \sqrt{\hat{\mathbf{\Sigma}}_0^{q',q'}}, \text{ for } q \neq q'$$

where $\hat{V}^{q,q'}$ is the element in the qth row and q'th column of the estimated correlation matrix $\hat{V}$ of the model already presented in Table 4.2.

An illustration of variance-covariance surface of prediction (3.41) with its related functional uncertainty is shown in Figure 4.9. The chosen pixel for the illustration is located within the geographic box $(48.95°N, 49°N, 17.45°W, 17.5°E)$ highlighted in Figure 4.1 with a yellow star and the day is kept again on 1st july 2015. The related functional uncertainty of kriging shown in the right panel of Figure 4.9 is computed as the square root of (3.42) and it is the behaviour of uncertainty of prediction with respect to height.

It is clear that the functional uncertainty of kriging shown in the right panel of Figure 4.9 is useful thinking as uncertainty with respect to function domain but do not have any

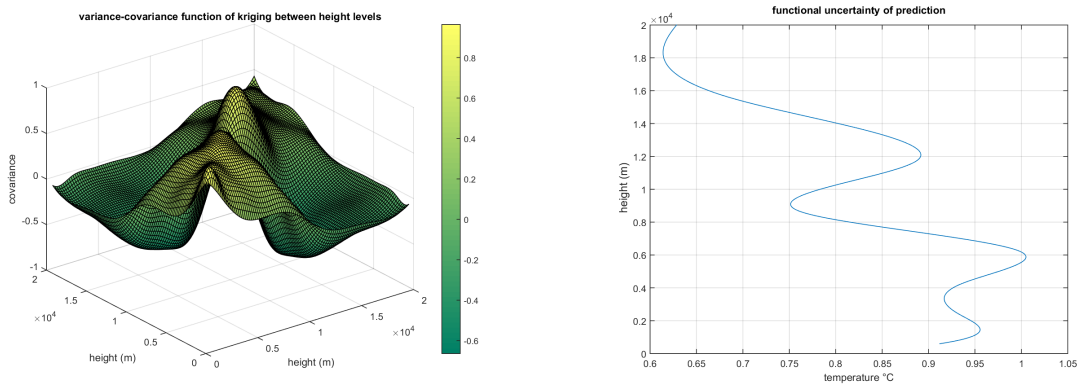

Figure 4.9 Variance-Covariance surface and functional uncertainty of kriging for the day 1st July 2015 in the chosen pixel. Right panel: variance-covariance surface of kriging. Left panel: functional uncertainty of kriging.

sense if we have to assess uncertainty with respect to spatial locations at a fixed height. In this second case we need a map of uncertainty of kriging.

In order to build a map of uncertainty of kriging, we have to specify the desired height and day for which we want to build the map. The functional uncertainty of prediction as already done on the right panel of Figure 4.9 is computed for each pixel inside the kriging area for the specified day and evaluated in the desired height as $\hat{\chi}_{s_0,t_0}(u_0)$. An example is shown in Figure 4.10 where the day is kept again on the 1st July 2015 and the desired heights are fixed again at 8.6 and 11.7 km of altitude.

The map of uncertainty in Figure 4.10 confirms the increasing of uncertainty of kriging from 8.6 km to 11.7 km which can also be observed in the functional uncertainty of kriging shown on the right panel of Figure 4.9. We can also observe that the uncertainty of kriging tends to be smaller when we go north in latitude.

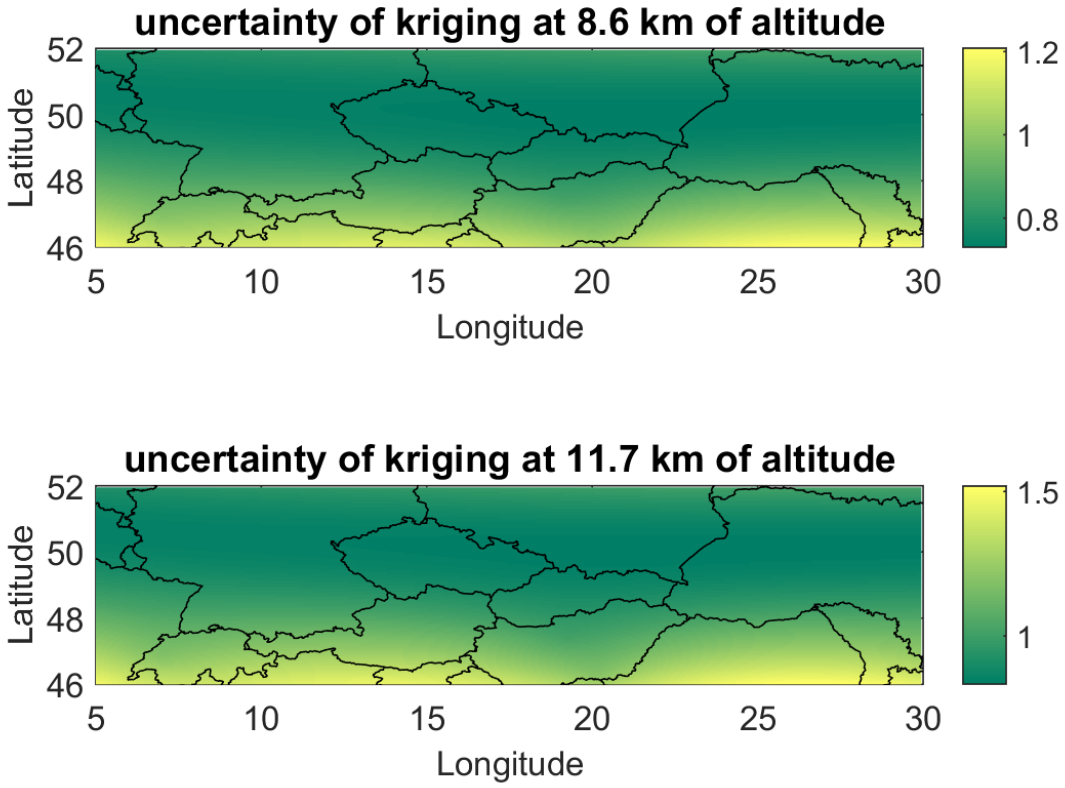

Figure 4.10 Standard deviation for the predicted temperature ($°C$) of the day 1st July 2015 at 8.6 km and 11.7 km.

Conclusion

Our goal was to build a hierarchical spatio-temporal model for functional data and to extend an existing tool for mapping to be able to handle functional data. We did it through the transformation of data from discrete points to a function and then the built function was modelled in a dynamic coregionalization model through its coefficients. The estimated model has been used for spatio-temporal prediction of the function. In crossvalidation we obtained good results but with some considerations. The approach by observation produced better results but it was not possible to understand where the prediction was not good. The approach by height level was good enough and also allowed to understand where the performance of the model was not good. It would be interesting to investigate how the performance of our model can be improved by including some atmospheric variables such as wind speed, wind direction and humidity. It is clear that our approach to geostatistics for spatio-temporally correlated functional data is a starting point with some interesting perspective for the future.

Since the performance was decreasing where there were less knot points, it can be interesting to investigate the optimal use of knot points. We observed in the preliminary analysis that the performance in fitting the curve was an increasing function of the number of knots related to the computational burden given by the dimension. Here dimensional exploration can be computed by the means of some dimension reduction technique such as functional principal component analysis. An interesting issue here is to understand not only how many knot points we need but also where it is needed and where not.

We have also observed a different behaviour of the model above 11 kilometres, specially R^2. An interesting future perspective is the number of spatial components in LCM with respect to the features of the function domain considering that the spatial correlation matrix V is shared or not between the spatial components. An interesting issue here is to understand if the spatial correlation is constant over the function domain

Having assessed the good performance of our approach in the case under study, it would be quite interesting to compare it with the performance of other approaches in the same case under study. Here the issue is to understand if our proposal approach performs better than the existing approaches.

Given the good results obtained in the case under study in terms of performance and considering the future perspectives mentioned above, it can be concluded that our proposal approach represents an interesting alternative in the specific framework of geostatistical spatio-temporal processes for functional data.

Bibliography

Aguilera-Morillo, M. C., Durbán, M., & Aguilera, A. M. (2016). Prediction of functional data with spatial dependence: a penalized approach. *Stochastic Environmental Research and Risk Assessment*, 1–16.

Caballero, W., Giraldo, R., & Mateu, J. (2013). A universal kriging approach for spatial functional data. *Stochastic environmental research and risk assessment*, *27*(7), 1553–1563.

Chatfield, C. (2016). *The analysis of time series: an introduction*. CRC press.

Cortés-D, D. L., Camacho-Tamayo, J. H., & Giraldo, R. (2016). Spatial prediction of soil penetration resistance using functional geostatistics. *Scientia Agricola*, *73*(5), 455–461.

Cressie, N. (1993). Statistics for spatial data: Wiley series in probability and statistics. *WileyInterscience, New York*, *15*, 105–209.

Delicado, P., Giraldo, R., Comas, C., & Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics*, *21*(3-4), 224–239.

Fassò, A., & Finazzi, F. (2011). Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. *Environmetrics*, *22*(6), 735–748.

Fassò, A., & Finazzi, F. (2013). A varying coefficients space-time model for ground and satellite air quality data over europe. *Statistica & Applicazioni, Special Online Issue*, 45–56.

Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media.

Finazzi, F., & Fassó, A. (2014). Dstem: a software for the analysis and mapping of environmental space-time variables. *Journal of Statistical Software. Accepted*.

Giraldo, R., Delicado, P., & 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.

Giraldo, R., Delicado, P., & Mateu, J. (2011). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, *18*(3), 411–426.

Giraldo, R., & Mateu, J. (2013). Kriging for functional data. *Encyclopedia of Environmetrics*.

Goulard, M., & Voltz, M. (1993). Geostatistical interpolation of curves: a case study in soil science. In *Geostatistics tróia'92* (pp. 805–816). Springer.

Ignaccolo, R., Mateu, J., & Giraldo, R. (2014). Kriging with external drift for functional data for air quality monitoring. *Stochastic environmental research and risk assessment*, *28*(5), 1171–1186.

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, *58*(8), 1246–1266.

Menafoglio, A., Grujic, O., & Caers, J. (2016). Universal kriging of functional data: Trace-variography vs cross-variography? application to gas forecasting in unconventional shales. *Spatial Statistics*, *15*, 39–55.

Menafoglio, A., & Secchi, P. (2017). Statistical analysis of complex and spatially dependent data: A review of object oriented spatial statistics. *European Journal of Operational Research*, *258*(2), 401–410.

Monestiez, P., & Nerini, D. (2008). A cokriging method for spatial functional data with applications in oceanology. In *Functional and operatorial statistics* (pp. 237–242). Springer.

Montero, J.-M., & Fernández-Avilés, G. (2015). Functional kriging prediction of pollution series: The geostatistical alternative for spatially-fixed data. *Estudios Econ. Apl*, *33*, 145–174.

Montero, J.-M., Mateu, J., et al. (2015). *Spatial and spatio-temporal geostatistical modeling and kriging* (Vol. 998). John Wiley & Sons.

Nerini, D., Monestiez, P., & Manté, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis*, *101*(2), 409–418.

Ramsay, J. O. (2006). *Functional data analysis*. Wiley Online Library.

Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional data analysis with r and matlab*. Springer Science & Business Media.

Rasekhi, M., Jamshidi, B., & Rivaz, F. (2014). Optimal location design for prediction of spatial correlated environmental functional data. *Journal of Modern Applied Statistical Methods*, *13*(2), 26.

Reyes, A., Giraldo, R., & Mateu, J. (2010). Residual kriging for functional data. application to the spatial prediction of salinity curves.

Temiyasathit, C., Kim, S. B., & Park, S.-K. (2009). Spatial prediction of ozone concentration profiles. *Computational Statistics & Data Analysis*, *53*(11), 3892–3906.