Kriging prediction for manifold-valued random fields

Davide Pigoli[a,*], Alessandra Menafoglio[b], Piercesare Secchi[b]

[a]*Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, UK*
[b]*MOX-Department of Mathematics, Politecnico di Milano, Milan, Italy*

Abstract

The statistical analysis of data belonging to Riemannian manifolds is becoming increasingly important in many applications, such as shape analysis, diffusion tensor imaging and the analysis of covariance matrices. In many cases, data are spatially distributed but it is not trivial to take into account spatial dependence in the analysis because of the non linear geometry of the manifold. This work proposes a solution to the problem of spatial prediction for manifold valued data, with a particular focus on the case of positive definite symmetric matrices. Under the hypothesis that the dispersion of the observations on the manifold is not too large, data can be projected on a suitably chosen tangent space, where an additive model can be used to describe the relationship between response variable and covariates. Thus, we generalize classical kriging prediction, dealing with the spatial dependence in this tangent space, where well established Euclidean methods can be used. The proposed kriging prediction is applied to the matrix field of covariances between temperature and precipitation in Quebec, Canada.

Keywords: Non Euclidean data, Residual kriging, Positive definite symmetric matrices

Introduction

This work is part of a line of research which deals with the statistical analysis of data belonging to Riemannian manifolds. Studies in this field have been motivated by many applications: for example Shape Analysis (see, e.g, Jung et al., 2011), Diffusion Tensor Imaging (see Dryden et al., 2009, and references therein)

*Corresponding address: Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, United Kingdom. Email address: dp497@cam.ac.uk

and estimation of covariance structures (Pigoli et al., 2014). Using the terminology of Object Oriented Data Analysis (Wang and Marron, 2007), in all these applications the atom of the statistical analysis belongs to a Riemannian manifold and therefore its geometrical properties should be taken into account in the statistical analysis.

We develop here a kriging procedure for Riemannian data. Spatial statistics for complex data has recently received much attention within the field of functional data analysis (see Nerini et al., 2010; Delicado et al., 2012; Gromenko et al., 2012; Menafoglio et al., 2013, 2014; Menafoglio and Petris, 2015) but the extension to non Euclidean data is even a greater challenge because they do not belong to a vector space.

Many works have considered the problem of dealing with manifold-valued response variables. Some of them propose non parametric (see Yuan et al., 2012, and references therein) or semi-parametric (see Shi et al., 2012) approaches but this implies a lack of interpretability or the reduction of multivariate predictors to univariate features. In particular, these approaches do not allow to introduce the spatial information in the prediction procedure.

A different line of research is followed in the present work, along the lines of those who try to extend to manifold-valued data parametric (generalized) linear models (see, e.g. Fletcher, 2013). We propose a linear regression model for Riemannian data based on a tangent space approximation. Even though the latter model is here developed in view of kriging prediction for manifold data, it may be used in general to address parametric regression in the context of Riemannian data, since it allows to consider multiple predictors in models with manifold-valued response variables. The central idea of this work consists in using the local geometry of the manifold to find a data-driven linearization, i.e. looking for the tangent space where the parametric model provides the best possible fitting for the available data.

The proposed method is illustrated here for the special case of positive definite symmetric matrices and in view of a meteorological application to covariance matrices. More in general, this approach is valid every time a Hilbert tangent space and correspondent logarithmic and exponential map can be defined, as we show in the Appendix.

1. Statistical analysis of positive definite symmetric matrices

Positive definite symmetric matrices are an important instance of data belonging to a Riemannian manifold. In this section, we introduce notation and a few

metrics, together with their properties, that we deem useful when dealing with data that are positive definite symmetric matrices. A broad introduction to the statistical analysis of this kind of data can be found, e.g., in Pennec et al. (2006) or Dryden et al. (2009).

Let $PD(p)$ indicate the Riemannian manifold of positive definite symmetric matrices of dimension p. It is a convex subset of $\mathbb{R}^{p(p+1)/2}$ but it is not a linear space: in general, a linear combination of elements of $PD(p)$ does not belong to $PD(p)$. Moreover, the Euclidean distance in $\mathbb{R}^{p(p+1)/2}$ is not suitable to compare positive definite symmetric matrices (see Moakher, 2005, for details). Thus, more appropriate metrics need to be used for statistical analysis. A good choice could be a Riemannian distance: the shortest path between two points on a manifold, once this has been equipped with a Riemannian metric, as we illustrate below. A description of the properties of Riemannian manifolds in general, and of $PD(p)$ in particular, can be found in Moakher and Zéraï (2011) and references therein.

Let $Sym(p)$ be the space of symmetric matrices of dimension p. The tangent space $T_\Sigma PD(p)$ to the manifold of positive definite symmetric matrices of dimension p in the point $\Sigma \in PD(p)$ can be identified with the space $Sym(p)$, which is linear and can be equipped with an inner product. A Riemannian metric in $PD(p)$ is then induced by the inner product in $Sym(p)$. Indeed, the choice of the inner product in the tangent space determines the form of the geodesic (i.e. the shortest path between two elements on the manifold) and thus the expression of the geodesic distance between two positive definite symmetric matrices. A possible choice for the Riemannian metric is generated by the scaled Frobenius inner product in $Sym(p)$, which is defined as $\langle A, B\rangle_\Sigma = \mathrm{trace}(\Sigma^{-\frac{1}{2}} A^T \Sigma^{-1} B \Sigma^{-\frac{1}{2}})$, where $A, B \in Sym(p)$. This choice is very popular for covariance matrices, because it generates a distance which is invariant under affine transformation of the random variables.

For every pair $(\Sigma, A) \in PD(p) \times Sym(p)$, there is a unique geodesic curve $g(t)$ such that

$$g(0) = \Sigma,$$
$$g'(0) = A.$$

When the Riemannian metric is generated by the scaled Frobenius inner product, the expression of the geodesic becomes

$$g(t) = \Sigma^{\frac{1}{2}} \exp(t\Sigma^{-\frac{1}{2}} A \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}},$$

where $\exp(C)$ indicates the exponential matrix of $C \in Sym(p)$. The exponential

map of $PD(p)$ in Σ is defined as the point at $t = 1$ of this geodesic:

$$\exp_\Sigma(A) = \Sigma^{\frac{1}{2}} \exp(\Sigma^{-\frac{1}{2}} A \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}}.$$

Thus, the exponential map takes the geodesic passing through Σ with "direction" A and follows it until $t = 1$. The exponential map has an inverse which is called logarithmic map and is defined as

$$\log_\Sigma(P) = \Sigma^{\frac{1}{2}} \log(\Sigma^{-\frac{1}{2}} P \Sigma^{-\frac{1}{2}}) \Sigma^{\frac{1}{2}},$$

where $\log(D)$ is the logarithmic matrix of $D \in PD(p)$. The logarithmic map returns the tangent element A that allows the corresponding geodesic to reach P at $t = 1$.

The *Riemannian distance* between elements $P_1, P_2 \in PD(p)$ is the length of the geodesic connecting P_1 and P_2, i.e.

$$d_R(P_1, P_2) = ||\log(P_1^{-1/2} P_2 P_1^{-1/2})||_F = \sqrt{\sum_{i=1}^{p} (\log r_i)^2},$$

where the r_i are the eigenvalues of the matrix $P_1^{-1} P_2$ and $||.||_F$ is the Frobenius norm for matrices, defined as

$$||A||_F = \sqrt{\text{trace}(A^T A)}.$$

This distance is called affine invariant Riemannian metric or *trace metric*, for instance in Yuan et al. (2012).

Other distances have been proposed in the literature to compare two positive definite symmetric matrices, both for computational reasons (Pennec et al., 2006) and for convenience in specific problems (Dryden et al., 2009). For example, we may consider the Cholesky decomposition of the positive definite symmetric matrix P, i.e. the lower triangular matrix with positive entries $L = \text{chol}(P)$ such that $P = LL^T$. Then, Wang et al. (2004) defined a *Cholesky distance* between two positive definite symmetric matrices as

$$d_C(P_1, P_2) = ||\text{chol}(P_1) - \text{chol}(P_2)||_F,$$

while Pennec (2006) proposed the Log-Euclidean distance

$$d_L(P_1, P_2) = ||\log(P_1) - \log(P_2)||_F,$$

based on the matrix logarithm. Another possibility is to resort to the *square root distance* (Dryden et al., 2009):

$$d_S(P_1, P_2) = ||P_1^{\frac{1}{2}} - P_2^{\frac{1}{2}}||_F,$$

where $P^{\frac{1}{2}}$ is the matrix square root of P. It is worth to notice that the square root distance is also defined for non negative definite matrices. Thus, it is to be preferred in applications where matrix data may have zero eigenvalues, or very small eigenvalues which lead to instability in the computation of the affine invariant Riemannian distance, the Cholesky decomposition or the matrix logarithm.

Each definition of distance has a corresponding geodesic and thus exponential and logarithmic map. For example, the geodesic curve associated with the square root metric in $\Sigma \in PD(p)$ with tangent vector $A \in Sym(p)$ is

$$g(t) = (\Sigma^{\frac{1}{2}} + tA)^T(\Sigma^{\frac{1}{2}} + tA),\ t \geq 0,$$

and the corresponding exponential and logarithmic map are therefore

$$\exp_\Sigma(A) = (\Sigma^{\frac{1}{2}} + A)^T(\Sigma^{\frac{1}{2}} + A),$$

$$\log_\Sigma(P) = P^{\frac{1}{2}} - \Sigma^{\frac{1}{2}}.$$

The expression of the logarithmic map for the Cholesky and logarithmic distance is similar, the square root transformation being substituted by $chol(\Sigma)$ and $\log(\Sigma)$ respectively.

Once a metric d has been introduced in $PD(p)$, we can address the problem of estimating the mean given a sample of positive definite symmetric matrices. In recent years, many authors (Fletcher et al., 2004; Pennec et al., 2006; Dryden et al., 2009) proposed to use the Fréchet mean for a more coherent approach in dealing with data belonging to a Riemannian manifold. Note that the Fréchet mean with respect to a Riemannian distance is often called intrinsic mean in the literature (see, e.g., Bhattacharya and Patrangenaru, 2003, 2005). The Fréchet mean of a random element S, with probability distribution μ on a Riemannian manifold $\mathcal{M}$, is defined as

$$M = \operatorname*{arginf}_{P \in \mathcal{M}} \int d(S, P)^2 \mu(dS)$$

and it can be estimated with the sample Fréchet mean

$$\overline{S} = \operatorname*{arginf}_{P \in \mathcal{M}} \sum_{i=1}^{n} d(S_i, P)^2,$$

if $(S_1, ..., S_n)$ is a sample from μ. In case $\mathcal{M} = PD(p)$, both the Fréchet mean and the sample Fréchet mean exist and are unique (see, e.g, Osborne et al., 2013, and reference therein). Analogously, the variance of S can be defined as

$$\sigma^2 = \mathrm{Var}(S) = \mathbb{E}[d(S, M)^2]$$

and estimated with the sample variance

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} d(S_i, \overline{S})^2.$$

By means of extensive comparisons, Dryden et al. (2009) show that using estimators based on a Riemannian distance gives better results than the estimator based on Euclidean metric. In particular, the affine-invariant Riemannian metric d_R has desirable properties such as avoiding the *swelling effect* (i.e. the fact that the average has larger determinants than the observation) and providing positive definite matrices also when extrapolating. For these reasons, we are going to prefer the affine-invariant Riemannian metric for the application of the Kriging procedure. However, these advantages come with a relevant computational cost that may be too heavy in some applications, for example in Diffusion Tensor Imaging (DTI). The Cholesky distance d_C and the Log-Euclidean distance d_L provide then more efficient alternatives. In particular, the Log-Euclidean distance shares most of the good properties of the affine-invariant Riemannian distance (see Pennec et al., 2006, for more details) with the main difference being in producing more anisotropic estimates. Finally, when data can include also non-negative definite matrices, we need to use the square root distance d_S or some of its generalization such as the power Euclidean distance or the Procrustes size-and-shape distance (see Dryden et al., 2009).

In the following, we will need to endow the linear space $Sym(p)$ of symmetric matrices with a metric; we will consider the metric induced by the Frobenius inner product $\langle \cdot, \cdot \rangle_F$. Based on this, we define the covariance between two random matrices $A, B \in Sym(p)$ as $\mathrm{Cov}(A, B) = \mathbb{E}[\langle A - \mathbb{E}[A], B - \mathbb{E}[B] \rangle_F]$.

Finally, we need some notation for dealing with array of matrices, being these our data. Let $\mathbf{T} \in Sym(p)^n$ a $p \times p \times n$ array, where $T_{..i} \in Sym(p)$ for $i = 1, \ldots, n$. Then, we can define a norm for this array as $||\mathbf{T}||^2_{Sym(p)^n} = \sum_{i=1}^{n} ||T_{..i}||^2_F$ and a scalar product $\langle \mathbf{T}, \mathbf{U} \rangle_{Sym(p)^n} = \sum_{i=1}^{n} \langle T_{..i}, U_{..i} \rangle_F$. Let us consider a matrix $G \in Sym(n)$, the product of this matrix for an array $\mathbf{T} \in Sym(p)^n$ can be defined as $G\mathbf{T} \in Sym(p)^n$, where $(G\mathbf{T})_{..i} = \sum_j G_{ij} T_{..j}$. A property we need in

the following is that

$$\langle G\mathbf{T}, \mathbf{U}\rangle_{Sym(p)^n} = \langle \mathbf{T}, G\mathbf{U}\rangle_{Sym(p)^n},$$

because of the symmetry of G. Finally, we define the covariance matrix for an array $\mathbf{T} \in Sym(p)^n$ as the matrix $\mathrm{Cov}(\mathbf{T}) \in Sym(n)$ such that $[\mathrm{Cov}(\mathbf{T})]_{ij} = \mathrm{Cov}(T_{..i}, T_{..j})$.

2. A tangent space model for Riemannian data

Under the hypothesis that the dispersion of the observations on the manifold is not too large, a tangent space can be used to approximate data in a linear space, where an additive model can be employed to describe the relationship between response variable and covariates. This allows to extend well established statistical methods for regression models to the context of manifold valued response variable. Let us consider the model

$$S(\mathbf{x}; \boldsymbol{\beta}, \Sigma) = \exp_{\Sigma}(A(\mathbf{x}; \boldsymbol{\beta}) + \Delta) \tag{1}$$

where $\Sigma \in PD(p)$ and $A(\mathbf{x}; \boldsymbol{\beta}) \in Sym(p)$ depends on the parameters collected in the array of matrices $\boldsymbol{\beta} = (\beta_{..0}, ..., \beta_{..r}) \in Sym(p)^{r+1}$, r is the number of predictors and $\mathbf{x} \in \mathbb{R}^r$ is the vector of covariates. $\Delta \in Sym(p)$ is a random matrix such that its mean is the null matrix and $\mathrm{Var}(\Delta) = \sigma^2$. Thus, this manifold valued random variable is generated following the geodesic passing through Σ with tangent vector $A(\mathbf{x}; \boldsymbol{\beta}) + \Delta$: the geodesic to be followed to obtain the realization of the variable is controlled by the covariate vector $\mathbf{x}$ but an additive error is also present. The model for the symmetric matrix A has to be specified and in this work we deal with the linear model

$$A(\mathbf{x}; \boldsymbol{\beta}) = \sum_{k=0}^{r} \beta_k Z_k, \tag{2}$$

where the Z_k's are the components of the vector $\mathbf{Z} = (1 \ \mathbf{x}^T)^T \in \mathbb{R}^{r+1}$. This strategy, involving a tangent space projection, can be followed also to generalize to manifold valued response variables more complex models, such as non linear or generalized regression models. This model follows the general idea proposed in Fletcher (2013) in using the local Euclidean properties of the manifold to define a regression model. The model we propose differs in the fact that the errors are defined in a common tangent space, thus allowing for the extension to the case

of correlated errors we consider in Section 3. Moreover, we use a more flexible model for the symmetric matrix A and this is helpful when modeling the drift effect.

Let us consider a sample $((\mathbf{x}_1, S_1), \ldots, (\mathbf{x}_n, S_n))$ from model (1) with uncorrelated errors. The goal is to fit a tangent plane approximation which best models the relationship between $\mathbf{x}$ and S, i.e. to find

$$(\widehat{\Sigma}, \widehat{\boldsymbol{\beta}}) = \underset{\Sigma \in PD(p), \boldsymbol{\beta} \in Sym(p)^{r+1}}{\operatorname{argmin}} \sum_{i=1}^{n} || \sum_{k=0}^{r} \beta_{..k} Z_{ik} - \log_{\Sigma}(S_i)||_F^2, \tag{3}$$

where $\log_{\Sigma}$ indicates the logarithmic map, which is the inverse of the exponential map and projects each element of $PD(p)$ on the tangent space of $PD(p)$ in Σ, $Z = [Z_{ik}]$ is the $n \times (r+1)$ design matrix and $\boldsymbol{\beta} \in Sym(p)^{r+1}$ is the array of matrix coefficients. Thus, our estimator looks for the linear model in the tangent space which minimizes the error sum of squares.

For a given known Σ, minimization of (3) is an ordinary least square problem in the tangent space, since

$$\begin{aligned}\sum_{i=1}^{n} || \sum_{k=0}^{r} \beta_{..k} Z_{ik} - \log_{\Sigma}(S_i)||_F^2 &= \sum_{i=1}^{n} \sum_{l=1}^{p} \sum_{q=1}^{p} (\sum_{k=0}^{r} \beta_{lqk} Z_{ik} - (\log_{\Sigma}(S_i))_{lq})^2 = \\ &= \sum_{l=1}^{p} \sum_{q=1}^{p} ||Z\beta_{lq.} - Y_{lq}(\Sigma)||_{\mathbb{R}^n}^2,\end{aligned}$$

where $Y_{lq}(\Sigma) \in \mathbb{R}^n$ is the vector $((\log_{\Sigma}(S_1))_{lq}, \ldots, (\log_{\Sigma}(S_n))_{lq})^T$ of elements in position (l, q) of the projected observations. Thus, each term of the previous sum is minimized with respect to β's by the ordinary least square solution

$$\widehat{\beta}_{lq.}(\Sigma) = (Z^T Z)^{-1} Z^T Y_{lq}(\Sigma), \tag{4}$$

and $\widehat{\beta}_{..k}(\Sigma) \in Sym(p)$ is the estimate for the k-th matrix parameter $\beta_{..k}$. The global solution should thus be in the form $(\Sigma, \widehat{\boldsymbol{\beta}}(\Sigma))$ and the problem

$$\min_{\Sigma \in PD(p)} \sum_{i=1}^{n} || \sum_{k=0}^{r} \widehat{\beta}_{..k}(\Sigma) Z_{ik} - \log_{\Sigma}(S_i)||_F^2 \tag{5}$$

has to be numerically solved. Here we resort to Nelder-Mead algorithm implemented in the `optim()` function in the R software (R Development Core Team,

2009), with constraints to ensure the matrix to be positive definite. This works well for two or three dimensional matrices, while in case of higher dimensional matrices more efficient optimization tools would be needed, for example gradient descend or Newton methods on the manifold (see, e.g, Dedieu et al., 2003). In this way, we obtain an estimate for the parameters:

$$\begin{cases} \widehat{\Sigma} = \underset{\Sigma \in PD(p)}{\operatorname{argmin}} \sum_{i=1}^{n} || \sum_{k=0}^{r} \widehat{\beta}_{..k}(\Sigma) Z_{ik} - \log_{\Sigma}(S_i)||_F^2, \\ \\ \widehat{\beta}_{lq.} = (Z^T Z)^{-1} Z^T Y_{lq}(\widehat{\Sigma}) \quad \text{for} \quad 1 \leq l \leq p, 1 \leq q \leq p \end{cases} \tag{6}$$

This method asks to specify the model (2) for the matrix A. Since this model is defined on the tangent space, well established techniques can be used for model selection, for example cross-validation.

3. Kriging prediction

In this section we apply the tangent space regression described above to non-stationary manifold-valued random fields. The main idea is to use a tangent space model to approximate the geometry of the manifold and to refer to the tangent space to deal with spatial dependence. Thus, the proposed model is the following: for $\mathbf{s} \in D \subset \mathbb{R}^d$,

$$S(\mathbf{s}; \boldsymbol{\beta}, \Sigma) = \exp_{\Sigma}(A(\mathbf{x}(\mathbf{s}); \boldsymbol{\beta}) + \Delta(\mathbf{s})), \tag{7}$$

where Δ is a zero-mean, globally second-order stationary and isotropic random field taking values in $Sym(p)$, the Euclidean space of symmetric matrices of order p. This means that $\mathbb{E}[\Delta(\mathbf{s})]$ is the null matrix for all $\mathbf{s}$ in D and, for $\mathbf{s}_i, \mathbf{s}_j \in D$, the covariance between $\Delta(\mathbf{s}_i)$ and $\Delta(\mathbf{s}_j)$ depends only on the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$, i.e., $\operatorname{Cov}(\Delta(\mathbf{s}_i), \Delta(\mathbf{s}_j)) = C(||\mathbf{s}_i - \mathbf{s}_j||)$. This definition of the covariogram C follows the approach detailed in (Menafoglio et al., 2013) for data belonging to Hilbert spaces. Equivalently, under the previous assumptions, for $\mathbf{s}_i, \mathbf{s}_j \in D$, the spatial dependence of the field can be represented by the semivariogram defined as $\gamma(||\mathbf{s}_i - \mathbf{s}_j||) = \frac{1}{2} \operatorname{Var}(\Delta(\mathbf{s}_i) - \Delta(\mathbf{s}_j)) = \frac{1}{2}\mathbb{E}[||\Delta(\mathbf{s}_i) - \Delta(\mathbf{s}_j)||_F^2]$. The assumption of isotropy can be relaxed but it brings additional complication in the estimation of the spatial dependence and this is outside of the scope of this work.

Let $\mathbf{s}_1, ..., \mathbf{s}_n$ be distinct locations in the domain D and assume that, in each location $\mathbf{s}_i$, the covariate vector $\mathbf{x}(\mathbf{s}_i)$ is observed together with a realization S_i of the random field (7). If the spatial dependence structure was known, the parameters of (7) could be estimated following a generalized least square approach.

Indeed, let Γ be the covariance matrix of the random errors in the observed locations, i.e. $\Gamma_{ij} = C(||\mathbf{s}_i - \mathbf{s}_j||)$ and $\mathbf{R} \in Sym(p)^n$ the array of matrices such that $R_{..i} = \sum_{k=0}^{r} \beta_{..k} Z_{ik} - \log_\Sigma(S_i)$. Thus, $\Gamma^{-1/2}\mathbf{R} \in Sym(p)^n$ has zero mean and identity covariance matrix, since

$$\begin{aligned}
[\mathrm{Cov}(\Gamma^{-1/2}\mathbf{R})]_{ij} &= E[\langle(\Gamma^{-1/2}\mathbf{R})_{..i}, (\Gamma^{-1/2}\mathbf{R})_{..j}\rangle_F] = \\
&= E[\langle\sum_l \Gamma_{il}^{-1/2} R_{..l}, \sum_m \Gamma_{jm}^{-1/2} R_{..m}\rangle_F] = \\
&= E[\sum_l \sum_m \Gamma_{il}^{-1/2}\Gamma_{jm}^{-1/2}\langle R_{..l}, R_{..m}\rangle_F] = \\
&= \sum_l \sum_m \Gamma_{il}^{-1/2}\Gamma_{jm}^{-1/2} E[\langle R_{..l}, R_{..m}\rangle_F] = \\
&= \sum_l \sum_m \Gamma_{il}^{-1/2}\Gamma_{jm}^{-1/2}\Gamma_{lm} = (\Gamma^{-1/2}\Gamma\Gamma^{-1/2})_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases},
\end{aligned}$$

where $\Gamma^{-1/2}$ is the inverse of the matrix square root of Γ. Hence, the generalized least square problem is

$$(\widehat{\Sigma}, \widehat{\boldsymbol{\beta}}) = \underset{\Sigma \in PD(p), \boldsymbol{\beta} \in Sym(p)^{r+1}}{\operatorname{argmin}} \sum_{i=1}^{n} ||(\Gamma^{-1/2}\mathbf{R})_{..i}||_F^2 = \tag{8}$$

$$= \underset{\Sigma, \boldsymbol{\beta}}{\operatorname{argmin}} \sum_i \sum_j (\Gamma^{-1})_{ij} \langle \sum_{k=0}^{r} \beta_{..k}(\Sigma) Z_{ik} - \log_\Sigma(S_i), \sum_{k=0}^{r} \beta_{..k}(\Sigma) Z_{jk} - \log_\Sigma(S_j)\rangle_F.$$

Likewise in the case of uncorrelated errors, the minimizer with respect to $\boldsymbol{\beta}$ – given Σ – is the generalized least square estimator in the tangent space,

$$\begin{aligned}
\widehat{\beta}_{lq.}^{GLS}(\Sigma) &= \underset{\beta_{lq.} \in \mathbb{R}^{r+1}}{\operatorname{argmin}} (Z\beta_{lq.} - Y_{lq}(\Sigma))^T \Gamma^{-1} (Z\beta_{lq.} - Y_{lq}(\Sigma)) = \\
&= (Z^T\Gamma^{-1}Z)^{-1} Z^T \Gamma^{-1} Y_{lq}(\Sigma)
\end{aligned} \tag{9}$$

where $Y_{lq}(\Sigma)$ is the vector $((\log_\Sigma(S_1))_{lq}, \ldots, (\log_\Sigma(S_n))_{lq})^T$ of elements in position (l, q) of the projected observations. For finding Σ, one thus has to solve

$$\widehat{\Sigma} = \underset{\Sigma \in PD(p)}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{n} (\Gamma^{-1})_{ij} \langle \sum_{k=0}^{r} \widehat{\beta}_{..k}^{GLS}(\Sigma) Z_{ik} - \log_\Sigma(S_i), \sum_{k=0}^{r} \widehat{\beta}_{..k}^{GLS}(\Sigma) Z_{jk} - \log_\Sigma(S_j)\rangle_F. \tag{10}$$

If Σ and $\boldsymbol{\beta}$ were known, the spatial dependence of the random field Δ on the tangent space could be estimated in $Sym(p)$ by considering the residuals $\Delta(\mathbf{s}_i) = A(\mathbf{x}(\mathbf{s}_i); \boldsymbol{\beta}) - \log_{\Sigma}(S_i)$ as an incomplete realization of the random field Δ. Indeed, the empirical semivariogram in $Sym(p)$ could thus be estimated with the method of moments estimator proposed in Cressie (1993),

$$\widehat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j) \in N(h)} ||\Delta(\mathbf{s}_i) - \Delta(\mathbf{s}_j)||_F^2,$$

where $N(h) = \{(\mathbf{s}_i, \mathbf{s}_j) \in D : h - \Delta h < ||\mathbf{s}_i - \mathbf{s}_j|| < h + \Delta h; i, j = 1, \ldots, n\}$, Δh is a positive (small) quantity acting as a smoothing parameter, $h > 0$ and $|N(h)|$ is the number of couples $(\mathbf{s}_i, \mathbf{s}_j)$ belonging to $N(h)$. A model semivariogram $\widehat{\gamma}_m(h)$ can be fitted to the empirical semivariogram, for example via weighted least squares (see Section 2.6.2 of Cressie, 1993). Cressie (1993) advocates the use of the semivariogram to estimate the spatial dependence. However, an estimate of the covariogram is also needed and this can be obtained as $\widehat{C}(||\mathbf{s}_i - \mathbf{s}_j||) = \lim_{h \to \infty} \widehat{\gamma}_m(h) - \widehat{\gamma}_m(||\mathbf{s}_i - \mathbf{s}_j||)$. In practice a good estimate of the spatial dependence (including the choice of the model semivariogram) is crucial for the analysis. The space of symmetric matrices being linear, all the existing methods of geostatistics can be used (see Diggle and Ribeiro, 2007; Chilès and Delfiner, 2009, for more recent developments).

Since we aim to estimate both the spatial dependence and the linear model in the tangent space, we resort to a nested iterative algorithm to solve problem (8). We initialize the algorithm with the estimates of equation (6). Then, we apply the Nelder-Mead algorithm to problem (10) where at each evaluation of the objective function the following iterative procedure is used to estimate both the spatial dependence structure and the parameters $\boldsymbol{\beta}$.

Evaluation of the objective function. Let $\widetilde{\Sigma}$ be the argument in which we want to evaluate the objective function to be minimized in problem (8). Let $\widehat{\boldsymbol{\beta}}^{(0)}$ be the estimate obtained via ordinary least squares as in equation (4). For $m \geq 1$

1. The tangent space residuals are computed $\widehat{\Delta}^{(m)}(\mathbf{s}_i) = A(\mathbf{x}(\mathbf{s}_i); \widehat{\boldsymbol{\beta}}^{(m-1)}) - \log_{\widetilde{\Sigma}}(S_i)$, $i = 1, \ldots, n$.
2. The empirical variogram $\widehat{\gamma}^{(m)}(h)$ is estimated from $\widehat{\Delta}^{(m)}(\mathbf{s}_1), \ldots, \widehat{\Delta}^{(m)}(\mathbf{s}_n)$, the model semivariogram $\widehat{\gamma}_m^{(m)}(h)$ is fitted via weighted least squares. The corresponding covariogram $\widehat{C}^{(m)}$ and the covariance matrix $\widehat{\Gamma}_{ij}^{(m)} = \widehat{C}^{(m)}(||\mathbf{s}_i - \mathbf{s}_j||)$ are computed.

3. A new estimate for $\widehat{\boldsymbol{\beta}}^{(m)}$ is obtained with a plug-in estimator $\left[\widehat{\boldsymbol{\beta}}^{(m)}(\widetilde{\Sigma})\right]_{\Gamma=\widehat{\Gamma}^{(m)}}$ using equation (9).
4. Steps 1-3 are iterated until convergence.

This is repeated at each evaluation of Σ in the Nelder-Mead algorithm.

Once $(\Sigma, \boldsymbol{\beta})$ and γ have been estimated, a kriging interpolation of the residuals provides an estimate for the field S in the unobserved location $\mathbf{s}_0$. Indeed, the problem of kriging is well defined when working in the tangent space $T_{\widehat{\Sigma}}PD(p)$, because this is a Hilbert space with respect to the Frobenius inner product and the covariogram C has been coherently defined. Therefore, by applying the kriging theory in Hilbert spaces (Menafoglio et al., 2013), the simple kriging predictor for $\Delta(\mathbf{s}_0)$ is derived as $\sum_{i=1}^{n} \lambda_i^0 \widehat{\Delta}(\mathbf{s}_i)$, where the vector of weights $\boldsymbol{\lambda}_0 = (\lambda_1^0, \ldots, \lambda_n^0)^T$ solves

$$\boldsymbol{\lambda_0} = \widehat{\Gamma}^{-1}\mathbf{c}$$

with $\mathbf{c} = (\widehat{C}(||\mathbf{s}_1 - \mathbf{s}_0||), \ldots, \widehat{C}(||\mathbf{s}_n - \mathbf{s}_0||))^T$. Given the vector of covariates $\mathbf{x}(\mathbf{s}_0)$ observed at location $\mathbf{s}_0$, the prediction of S_0 – the unobserved value of the field S at $\mathbf{s}_0$ – can be eventually obtained through the plug-in estimator:

$$\begin{aligned}\widehat{S}_0 &= \widehat{S}_0(\mathbf{x}(\mathbf{s}_0), (\mathbf{x}(\mathbf{s}_1), S_1), \ldots, (\mathbf{x}(\mathbf{s}_n), S_n)) = \\ &= \exp_{\widehat{\Sigma}}(\widehat{\beta}_{..0}^{GLS}(\widehat{\Sigma}) + \sum_{k=1}^{r} \widehat{\beta}_{..k}^{GLS}(\widehat{\Sigma}) x_k(\mathbf{s}_0) + \sum_{i=1}^{n} \lambda_i^0 \widehat{\Delta}(\mathbf{s}_i)). \qquad (11)\end{aligned}$$

It is worth to notice that this residual approach includes both the cases of stationary and non stationary random field, differently from what happens with Euclidean data. This depends on the need to estimate the parameter Σ and compute the tangent space projections even if the random field is indeed stationary, while in linear spaces the same problem can be addressed with a weighted average of the observed values, without the need to estimate the common mean. In general, the proposed estimation procedure asks for the knowledge of the model in the tangent space. As mentioned above, cross-validation techniques can be used to choose the model among a starting set of candidate models (for furhter details, see Algorithm 4 of Menafoglio et al., 2013). However, it is also important to mention that the chosen model needs tangent space residuals with a variogram compatible with stationarity. Therefore, a good strategy could be starting from a simple (even constant) model and adding covariates until the residuals variogram satisfies the stationarity hypothesis.

Prediction error. Correctly assessing the expected prediction error is an important property of the kriging procedure for Euclidean data. In general, this is not an easy task in non Euclidean spaces, given the non linearity of the space and of the operations involved in the prediction. The quantity we would like to control is $p(\mathbf{s}_0) = \mathbb{E}[d(S_0, \widehat{S}_0)^2]$. Here we propose a semiparametric bootstrap to estimate $p(\mathbf{s}_0)$. This means that we build a bootstrap procedure using empirical residuals, as proposed initially in Freedman and Peters (1984) for linear models. Moreover, we address the problem of the spatial dependence of the residuals using the estimated covariogram, as common practice when applying bootstrap to spatial observations (see, e.g., Solow, 1985; Iranpanah et al., 2011). Thus, we assume that the semivariogram γ, the related covariogram C, together with the parameters $(\Sigma, \boldsymbol{\beta})$ have been estimated by $\widehat{\gamma}, \widehat{C}$ and $(\widehat{\Sigma}, \widehat{\boldsymbol{\beta}})$ respectively, devised as in Section 3.

Let $\widehat{\Delta}(\mathbf{s}_1), \ldots, \widehat{\Delta}(\mathbf{s}_n)$ be the residuals of the estimated model in the known locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$. We do not want to bootstrap directly from these residuals because of spatial dependence. Hence, we aim at representing uncorrelated errors through the linear transformation $\mathbf{E} = \widehat{\Gamma}^{-1/2}\mathbf{R}$, where $\mathbf{R} \in Sym(p)^n$ is the array of the residuals such that $R_{..i} = \widehat{\Delta}(\mathbf{s}_i)$, for $i = 1, \ldots, n$, and $\widehat{\Gamma}^{-1/2}$ is the inverse of the square root matrix of $\widehat{\Gamma}$, which estimates the covariance matrix of errors in the observed locations. For $b = 1, \ldots, B$, we sample with replacement $n+1$ elements from $(E_{..1}, \ldots, E_{..n})$ and obtain a bootstrap array $\mathbf{E}^{(b)} \in Sym(p)^{n+1}$. We then compute the bootstrap residuals in the observed locations and in the unobserved location $\mathbf{s}_0$ as $\mathbf{R}^{(b)} = \widehat{\Gamma}_0^{1/2}\mathbf{E}^{(b)}$. Here $\widehat{\Gamma}_0^{1/2}$ is the matrix square root of the estimate of the covariance matrix of errors in the locations $\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_n$. Finally, we obtain the bootstrap realizations of the random field in the locations $\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_n$,

$$S_i^{(b)} = \exp_{\widehat{\Sigma}}(A(\mathbf{x}(\mathbf{s}_i); \widehat{\boldsymbol{\beta}}) + R_{..i}^{(b)}), \quad i = 0, \ldots, n,$$

and the kriging prediction in $\mathbf{s}_0$,

$$\widehat{S}_0^{(b)} = \widehat{S}_0^{(b)}(\mathbf{x}(\mathbf{s}_0), (\mathbf{x}(\mathbf{s}_1), S_1^{(b)}), \ldots, (\mathbf{x}(\mathbf{s}_n), S_n^{(b)}))$$

achieved by the estimator (11), given the bootstrap realizations $S_1^{(b)}, \ldots, S_n^{(b)}$. We can now estimate the expected prediction error as

$$\widehat{p}(\mathbf{s}_0) = \frac{1}{B}\sum_{b=1}^{B} d(S_0^{(b)}, \widehat{S}_0^{(b)})^2. \tag{12}$$

It is worth noting that the same resampling strategy can be applied to the estimation of the standard error for the regression parameters β. For example, an

estimate of the element-wise variance of $\widehat{\beta}$ is provided by $\frac{1}{B}\sum_{b=1}^{B}(\widehat{\beta}_{lqk} - \widehat{\beta}_{lqk}^{(b)})^2$, $k, l = 1, ..., p$, where $\widehat{\beta}_{lqk}$ is the estimate from the original sample and $\widehat{\beta}_{lqk}^{(b)}$ for $b = 1, \ldots, B$ are the estimates from the bootstrapped samples. To provide a global measure of uncertainty, one can alternatively consider the squared Frobenius distance in $Sym(p)$ between the estimate $\widehat{\beta}_{\cdot\cdot k}$ and the bootstrap estimate $\widehat{\beta}_{\cdot\cdot k}^{(b)}$ for $b = 1, \ldots, B$, i.e., $\frac{1}{B}\sum_{b=1}^{B}\|\widehat{\beta}_{\cdot\cdot k} - \widehat{\beta}_{\cdot\cdot k}^{(b)}\|^2$. This approach can be used to make inference about the regression parameters. For instance, one can employ the Chebychev inequality to provide a bound on the squared Frobenius distance between the estimated parameters $\widehat{\beta}_{\cdot\cdot k}$, $k = 0, ..., r$ and the underlying parameters. However, one should pay close attention when interpreting element-wise changes, since they need to be considered jointly with the other elements of the coefficients matrix $\widehat{\beta}_{\cdot\cdot k}$ and with Σ.

4. Simulation studies

In this section we illustrate the simulation studies performed to evaluate the estimators proposed in the previous sections. Two main goals drive the section: (a) evaluate the performance of the kriging predictor and the semiparametric bootstrap estimator of the expected prediction error when data are generated according to model (7); (b) evaluate the performance of the same estimators when applied to data which actually do not follow model (7).

Throughout this section, we set the true tangent point to be

$$\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix},$$

and the tangent vector to be

$$A(\mathbf{x}(\mathbf{s}); \boldsymbol{\beta}) = 0.2 s_1,$$

where $\mathbf{s} = (s_1, s_2)^T \in D$, are coordinates on a bidimensional spatial domain D.

Let us first consider the performance of the proposed method for what concerns parameter estimation when data are generated from model (1), that is when errors are independent. We follow the proposal in Fletcher (2013) to evaluate the performance using the mean squared error (MSE) of the estimator $\widehat{\Sigma}$ and $A(\mathbf{x}; \widehat{\beta})$, for a fixed predictor vector $\mathbf{x}$. The mean squared error for $\widehat{\Sigma}$ can be defined with respect to the affine invariant Riemannian distance as $\mathbb{E}[d_R(\widehat{\Sigma}, \Sigma)^2]$. The MSE for $A(\mathbf{x}; \widehat{\beta})$ is defined by the Frobenius distance between $A(\mathbf{x}; \beta)$ and the projection

of the action of $A(\mathbf{x};\widehat{\beta})$ in the tangent space centered in Σ, which is where $A(\mathbf{x};\beta)$ is defined, i.e. $\mathrm{MSE}(A(\mathbf{x};\widehat{\beta})) = \mathbb{E}[||\log_{\Sigma}(\exp_{\widehat{\Sigma}}(A(\mathbf{x};\widehat{\beta}))) - A(\mathbf{x};\beta)||_F^2]$.

We consider the different sample sizes $N = 5, 15, 30, 40, 60, 80, 100$ and we simulate $M = 100$ replicates for each value of the sample size. We estimate the mean squared error as $\widehat{\mathrm{MSE}}_N(\widehat{\Sigma}) = \frac{1}{M}\sum_{i=1}^{M} d_R(\widehat{\Sigma}_{N,i},\Sigma)^2$ and $\widehat{\mathrm{MSE}}(A((1,1);,\widehat{\beta})) = \frac{1}{M}\sum_{i=1}^{M} ||\log_{\Sigma}(\exp_{\widehat{\Sigma}_{N,i}}(A((1,1);\widehat{\beta}_{N,i}))) - A((1,1);\beta)||_F^2$, where $\widehat{\Sigma}_{N,i}$ and $\widehat{\beta}_{N,i}$ are the estimates for the i-th replicate with sample size N. The estimated mean squared errors can be seen in Fig. 1 , alongside the computation times. The MSEs approach zero when the sample size increases, thus suggesting empirically that the proposed estimator are consistent. We report in Fig. 1 also the boxplots of the bias $\widehat{\beta}_{N,i} - \beta$ for the regression coefficients. Their estimates appear to be unbiased for large N. This is not surprising, since the estimators for β would be unbiased if Σ was known and we have seen that $\widehat{\Sigma}_N$ becomes closer and closer to Σ when N increases. However, for finite sample sizes we expect the need to estimate Σ to add some bias on the estimators for the regression coefficients. Finally, we can see that the computation times (obtained running an R code on an Intel i5-4690K CPU 3.5GHz machine) for the estimator increase with the sample size but only linearly.

We want now to assess the performance of the proposed method for kriging prediction, which is the our final goal. We independently simulate $M = 100$ realizations from a 2×2 positive definite matrix random field defined on the bidimensional spatial domain D and consistent with model (7), where errors are spatially correlated. The two diagonal elements and one of the off-diagonal elements of the error matrix in the tangent space are independent realizations of a real valued, zero-mean Gaussian field with Gaussian variogram of decay 0.1, sill 0.25 and zero nugget. The remaining off-diagonal element is determined by symmetry.

We generate the realizations of the random field on a 10×10 regular grid of locations and we observe the field in 15 randomly selected grid points $\mathbf{s}_1, ..., \mathbf{s}_{15}$; these 15 locations are kept fixed across the M simulations.

Fig. 2 shows the results obtained for the first simulation: the reference realization is shown in panel (a), the predicted field in panel (b) and the expected prediction error assessed through (12) in panel (c). In Fig. 2a and 2b, each 2×2 positive definite matrix S observed or predicted at a location $\mathbf{s}$ is represented as an ellipse that is centered in $\mathbf{s}$ and has axis $\sqrt{\sigma_j}\mathbf{e}_j$, where $S\mathbf{e}_j = \sigma_j\mathbf{e}_j$, for $j = 1, 2$. Overall, the pattern of the reference realization appears to be fairly captured by the kriged field. Only the bottom right corner of the spatial domain seems to be affected by a significant prediction error, due to the absence of data near this

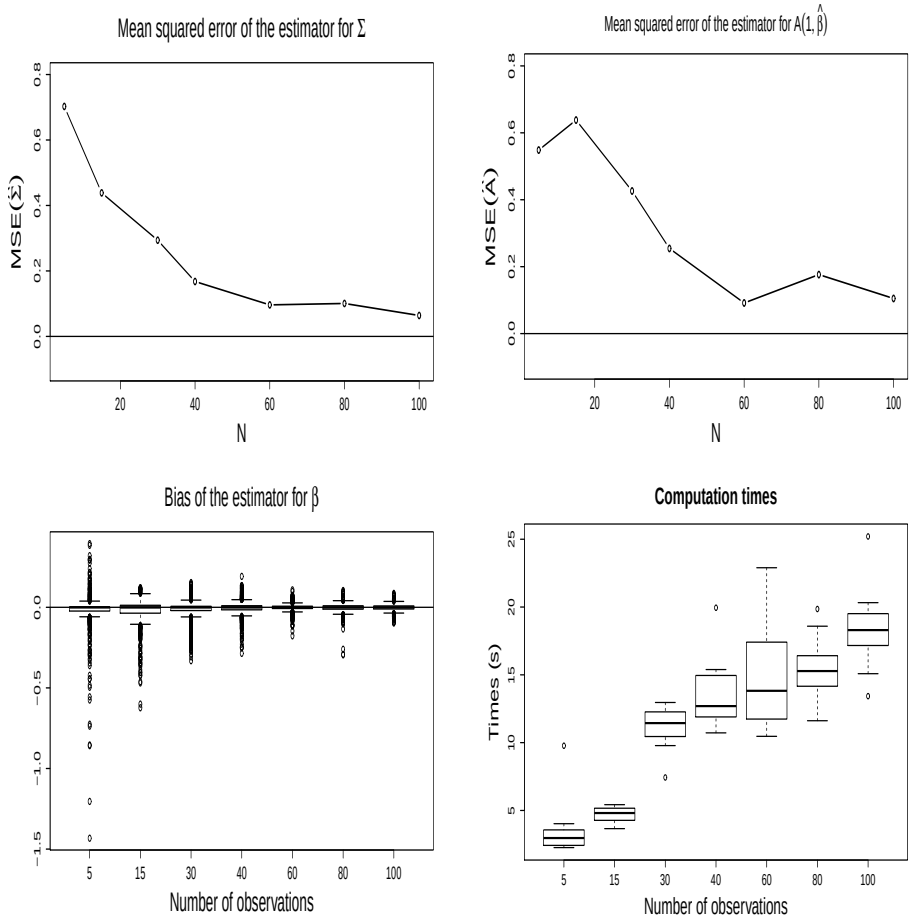
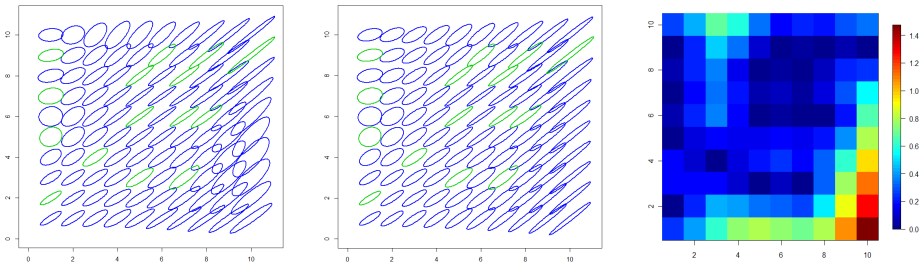


Figure 1: MSEs of the estimators $\widehat{\Sigma}$ (top left) and $A((1,1);,\widehat{\beta})$ (top right) and element-wise bias on the regression parameters $\widehat{\beta}-\beta$ (bottom left) as a function of the sample size and corresponding computation times (bottom right), obtained running an R code on an Intel i5-4690K CPU (3.5GHz) machine.



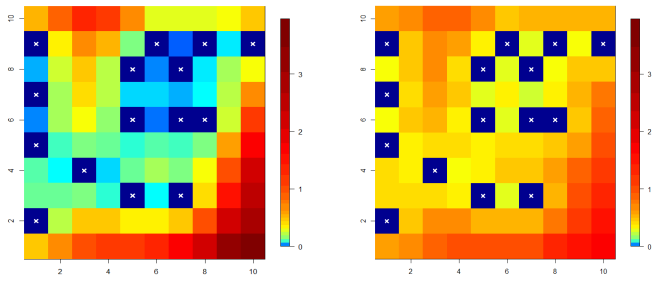
(a) Reference realization (b) Predicted field (c) Bootstrapped prediction error

Figure 2: Panels (a) and (b): Comparison between the reference realization of the matrix field and the predicted field. Each positive definite matrix is represented by an ellipse centered in the respective location. Observed data are represented in green. Panel (c): Estimated expected prediction error.

boundary of the domain. This is reflected by the estimated prediction error (Fig. 2c).

To evaluate the goodness of the estimator (12) for the expected prediction error, we use the $M = 100$ independently simulated field realizations as follows. For every simulation, in each of the 85 grid locations different from $\mathbf{s}_1, ..., \mathbf{s}_{15}$, we compute both the kriging prediction secured by the estimator (11) and the estimate of the expected prediction error achieved by (12), having set $B = 100$ to be the number of bootstrap replicates. Fig. 3 compares the maps of: (a) the average prediction error across the $M = 100$ simulations, i.e. the average of the square distance between the known field realization and its kriging prediction, and (b) the average, across the same $M = 100$ simulations, of the estimates of the expected prediction error supplied by the estimator (12). It can be noticed that the semi-parametric bootstrap estimator provides a good evaluation of the prediction error, with only a slight overestimation near the observed locations.

In the remaining part of the section, we illustrate a second simulation study aiming at evaluating the performance of the kriging predictor when data are generated from a model different from (7). In particular, we follow Pigoli and Secchi (2012) to generate a non stationary matrix field according to a probabilistic model with mean $P(\mathbf{s}) = \exp_\Sigma(A(\mathbf{x}(\mathbf{s}); \boldsymbol{\beta}))$, where Σ and $A(\mathbf{x}, \boldsymbol{\beta})$ are set to be the same as in the previous simulation study. This random matrix field is obtained through



(a) Average prediction error across 100 simulations

(b) Average estimate of the expected prediction error across the same 100 simulations

Figure 3: Comparison between the average prediction error, computed across 100 simulations, and the average estimate of the expected prediction error. Crosses indicate the position of the observed locations.

the sample covariance matrices generated by the realizations of a Gaussian random vector field $\mathbf{v}$.

Let $D \subset \mathbb{R}^2$ indicate the common spatial domain of two independent gaussian random fields $w(\mathbf{s}), y(\mathbf{s}),\ \mathbf{s} \in D$. Both random fields w and y have zero mean and Gaussian spatial covariance with decay $\phi = 0.1$, sill equal to 1 and zero nugget, i.e.

$$\mathrm{Cov}(w(\mathbf{s}_i), w(\mathbf{s}_j)) = \mathrm{Cov}(y(\mathbf{s}_i), y(\mathbf{s}_j)) = \begin{cases} \exp(-\phi\|\mathbf{s}_i - \mathbf{s}_j\|^2) & \|\mathbf{s}_i - \mathbf{s}_j\|^2 > 0, \\ 1 & \|\mathbf{s}_i - \mathbf{s}_j\|^2 = 0, \end{cases} \tag{13}$$

for $\mathbf{s}_i, \mathbf{s}_j \in D$. For $\mathbf{s} \in D$, the covariance matrix (between components) of the random vector field $\mathbf{v}(\mathbf{s}) = (P(\mathbf{s}))^{\frac{1}{2}}(w(\mathbf{s}), y(\mathbf{s}))^T$ is equal to $P(\mathbf{s})$. We generate N independent realizations of the random vector field $\mathbf{v}$ and, for $\mathbf{s} \in D$, we compute the sample covariance matrix

$$S(\mathbf{s}) = \frac{1}{N-1}\sum_{k=1}^{N}(\mathbf{v}_k(\mathbf{s}) - \bar{\mathbf{v}}(\mathbf{s}))(\mathbf{v}_k(\mathbf{s}) - \bar{\mathbf{v}}(\mathbf{s}))^T \sim \mathrm{Wishart}_2\left(\frac{1}{N-1}P(\mathbf{s}), N-1\right), \tag{14}$$

$\bar{\mathbf{v}}(\mathbf{s})$ being the sample mean in $\mathbf{s} \in D$.

For $N = 4, 5, 6$, we replicate the simulation $M = 25$ times on the same 10×10 regular grid of the first simulation study described in this section. For

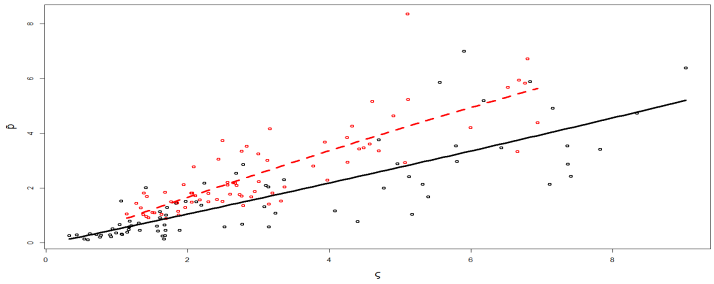
each simulation, we observe the realizations of the field S in the same 15 locations $\mathbf{s}_1, ..., \mathbf{s}_{15}$ selected in the first simulation study and we then predict the field realizations in the remaining 85 locations through the kriging estimator (11). Note that the parameter N controls the marginal variability of the matrix random field S.

We evaluate the performance of the kriging procedure when applied to these simulated fields by comparison with the case when data are generated by model (7). Thus, we generate a second set of $M = 25$ simulations using model (7) and setting the value of $\mathrm{Var}(\Delta) = \sigma^2 = 0.35, 1, 1.9$. This provides field marginal variabilities comparable to those of the random fields generated by model (14) when $N = 6, 5, 4$, respectively. Indeed, it is fair to compare prediction performances when the marginal variabilities of the random fields which generate the data are close. However, it is not trivial to obtain an explicit relationship between the parameters controlling the field marginal variability in the two models, i.e., N and σ^2.

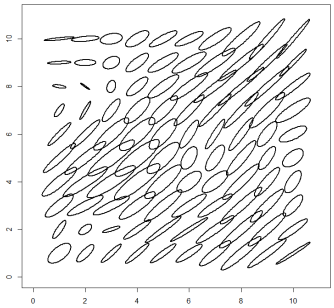
Thus, for both models we measure the empirical mean square error with respect to the mean of model (14), expressed by $\varsigma = \frac{1}{100}\sum_{i=1}^{100} d(S(\mathbf{s}_i), P(\mathbf{s}_i))^2$, and the average prediction error expressed by $\bar{p} = \frac{1}{85}\sum_{i=1}^{85} d(S(\mathbf{s}_i), \widehat{S}(\mathbf{s}_i))^2$. Each point in Fig. 4a represents the joint values of ς and $\bar{p}$ relative to one of the 150 simulations considered in this study: red points are relative to the 75 simulations - 25 simulations for each of the three values chosen for the parameter N - when data are generated by model (14), while black points are relative to the 75 simulations - 25 simulations for each of the three values chosen for the parameter σ^2 - when data are generated by model (7). Polynomial smoothing lines are added to help visual comparison. Inspection of Fig. 4a suggests that the performance of the kriging predictor (11) when model (7) is violated is worse than its performance when model (7) is correct. Moreover, the higher the value of ς the worse is the relative performance of the kriging predictor. This is to be expected because a high dispersion on the manifold means that no tangent space can accurately describe the data. However, for low values of ς the performance of the kriging predictor in the two situations is comparable, supporting its robustness to the violation of the model provided that the tangent space approximation is able to describe in a fairly accurate way the observations.

By way of example, Fig. 4b and 4c represent two realizations of the matrix field generated from (14) for high and low values of N, respectively, i.e. low or high values of the field empirical marginal variability ς.

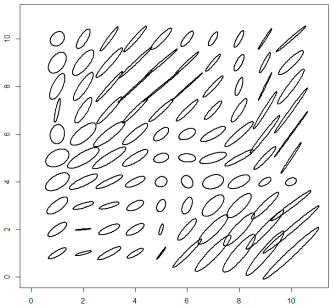
In all the above simulations we use a Gaussian covariance function, since this



(a) Prediction error and variability



(b) Field realization with $\varsigma = 2.9$



(c) Field realization with $\varsigma = 6.7$

Figure 4: Panel (a): Prediction error as a function of ς , with a local polynomial smoothing added to help visual comparison, for data generated from the tangent space model (7) (black points and solid black line) and from procedure (14) (red points and dashed red line), both with Gaussian covariance function. Panel (b) and (c): Examples of simulated fields from model (14) for $N = 6$ (b) and $N = 4$ (c) and Gaussian covariance function, with the respective values of ς.

appears to be the most suitable model for spatial dependence in the case study we analyze in Section 5. However, the Gaussian model has a particularly smooth behaviour that can be sometimes misleading as a test case, see Stein (1999, Section 2.7). For this reason, we also replicate the same simulations using an exponential covariance function, both in the tangent space model and in equation (13). The results can be found in the Supplementary material and they are very close to the ones obtained in the Gaussian case.

5. Kriging for Quebec meteorological data

In this section a kriging interpolation is proposed for the covariance matrices between temperature and precipitation in Quebec. The co-variability of temperature and precipitation is of great interest for meteorological purposes, since a good understanding of their relationship can improve weather forecasting methods. Moreover, relative behavior of temperature and precipitation affects agricultural production (see Lobell and Burke, 2008). For a detailed description of the temperature - precipitation relationship and its estimate see, e.g., Trenberth and Shea (2005) and references therein.

We focus on the Quebec province, Canada. Data from Canadian meteorological stations are made available by Environment Canada on the website `http://climate.weatheroffice.gc.ca`. We restrict to the 7 meteorological stations where all monthly temperature and precipitation data are available from 1983 to 1992. For each station and for each month from January to December, we use these 10-year measures to estimate the 2×2 covariance matrix between temperature and precipitation. We thus obtain and separately analyse 12 datasets, each composed by $n = 7$ spatially dependent sample covariance matrices (with the previous notation, $n = 7$ and $p = 2$). These datasets have been previously considered in Pigoli and Secchi (2012) with the aim to estimate a regional covariance matrix starting from the observations coming from different meteorological stations. Here, the aim of the analysis is different, since we want to predict the covariance matrix in unobserved locations, starting from the incomplete realization of the field.

For the analysis, we choose the affine invariant Riemannian metric as the distance between covariance matrices because of its invariance property to changes in the measurement units. At first, we make the hypothesis of stationarity for the random field (i.e., that the matrix random field has a constant mean), thus including only a constant term in the tangent space model. A Gaussian semivariogram model with nugget seems appropriate to fit the empirical semivariogram in the

tangent space and estimate the structure of spatial dependence. Notice that, since we are working on different tangent spaces for the different months, a separate semivariogram needs to be estimated for each month. Having estimated the semivariogram, the simple kriging interpolation of the residuals is performed, eventually estimating the matrix field as in (11). Fig. 5 shows kriging results for the months of January and July, as well as the correspondent semivariogram in the tangent space.

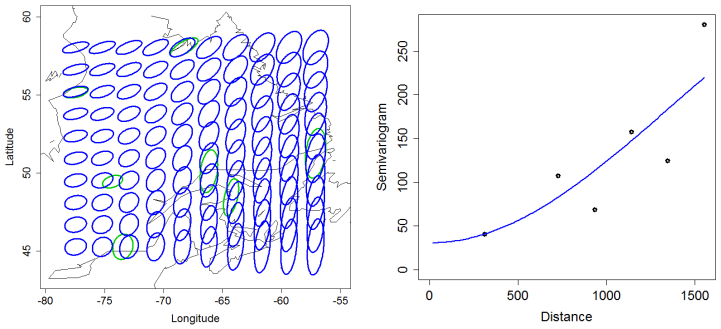
As suggested by the graphical inspection of Fig. 5, the hypothesis of stationarity is unlikely to be adequate for all the months, meaning that the covariation between temperature and precipitation is influenced by seasonal effects. In the following, we focus on the results relative to January and July, which are representative of the two most different meteorological behaviors. From a geostatistical point of view this reflects on the estimated semivariograms: while the July semivariogram does look stationary, the January semivariogram suggests to move toward a non-stationary model, by introducing a space dependent drift term.

In order to choose an appropriate model for the drift in January, we investigated linear and quadratic models with respect to longitude and latitude, including an interaction term. We found that the choice which seems to balance the most the complexity of the drift model with the residuals stationarity assumption is the following linear model depending on longitude:

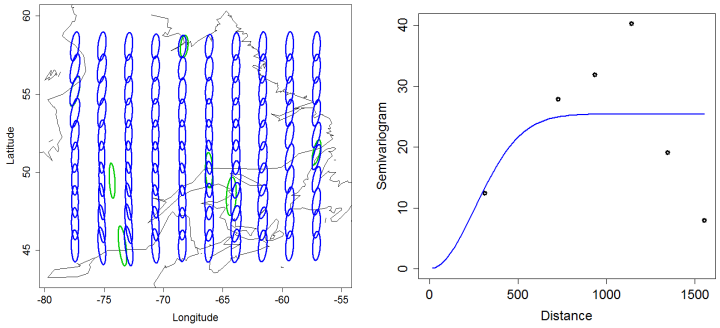
$$A(\mathrm{Longitude}_i, \mathrm{Latitude}_i) = \beta_0 + \beta_{Long}\mathrm{Longitude}_i. \tag{15}$$

The dependence of the field on the longitude seems to be suggested also by the stationary kriging interpolation in Fig. 5a. A possible meteorological interpretation relies in the exposition of the region toward the sea. Indeed, model (15) accounts for the distance between the location of interest and the Atlantic Ocean, which is likely to influence temperatures, precipitations and their covariability. Such an influence has been also observed by Menafoglio et al. (2013) when analyzing temperature curves recorded in the Maritimes Provinces of Canada, located SE of Quebec.

Fig. 6 shows the kriging estimates of the matrix field, the estimated drift and the semivariogram of the residuals for January. The semivariogram estimated from the residuals (Fig. 6c) substantiates the residuals stationarity assumption. Moreover, the residual spatial variability looks similar to the one characterizing July (ranges: 490 km and 625 km; sills: 36.83 and 25.41; nuggets: 0.01 and 0 for January and July respectively). The comparison of the January semivariograms estimated from data (Fig. 5a) and from residuals (Fig. 6c) suggests that the spatial

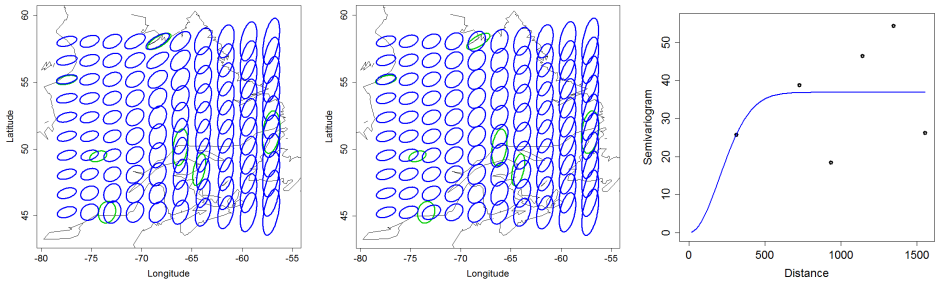


(a) January



(b) July

Figure 5: On the left: Ordinary kriging for the (temperature, precipitation) covariance matrix field; green ellipses indicate original data. A covariance matrix S at location $\mathbf{s}$ is represented as an ellipse centered in $\mathbf{s}$ and with axis $\sqrt{\sigma_j}\mathbf{e}_j$, where $S\mathbf{e}_j = \sigma_j \mathbf{e}_j$ for $j = 1, 2$. Horizontal and vertical axes of the ellipses represent temperature and precipitation respectively. On the right: empirical semivariogram (symbols) and fitted exponential model (solid line). Distances are measured in km.



(a) Kriging prediction (b) Estimated drift (c) Residuals semivariogram

Figure 6: Kriging of the (temperature, precipitation) covariance matrix field during January, with a drift term depending on longitude. A covariance matrix S at location $\mathbf{s}$ is represented as an ellipse centered in $\mathbf{s}$ and with axis $\sqrt{\sigma_j}\mathbf{e}_j$, where $S\mathbf{e}_j = \sigma_j \mathbf{e}_j$ for $j = 1, 2$. Horizontal and vertical axes of the ellipses represent temperature and precipitation respectively. In subfigure (a) and (b) green ellipses indicate the data, blue ellipses the field and drift estimates respectively. In subfigure (c) the residual empirical semivariogram (symbols) and the fitted exponential model (solid line) are reported. In subfigure (c) distances are measured in km.

variability is mostly explained by the drift term (i.e., by the distance from the Atlantic Ocean), with a low variability left to the stochastic component. This turns in a spatial prediction strongly driven by the drift term, as noticed by the comparison of Fig. 6a and 6b. The very opposite happens during July, when the spatial variability of the phenomenon seems to be mostly due to a stochastic fluctuation. This is not completely unexpected, since Menafoglio et al. (2013) report similar results when analyzing daily mean temperature curves observed during 1980.

From a meteorological point of view, the kriged map in Fig. 6b shows that the temperature - precipitation relationship appears to significantly vary when moving from the Ocean toward the internal regions, precipitation being affected by a higher variability along the coastline than in the western zone. Moreover, this prediction is overall characterized by a positive correlation between temperature and precipitation, in accordance with the results by Trenberth and Shea (2005).

Finally, Fig. 7 reports the estimated maps of prediction error, obtained via the estimator (12) computed over a grid of new locations equally spaced. The two panels show similar pattern of prediction error. As expected, the peripheral areas of the Eastern Quebec are those associated with the higher uncertainty due to the absence of observations. However, the prediction appears generally very accurate, with a low estimated prediction error.

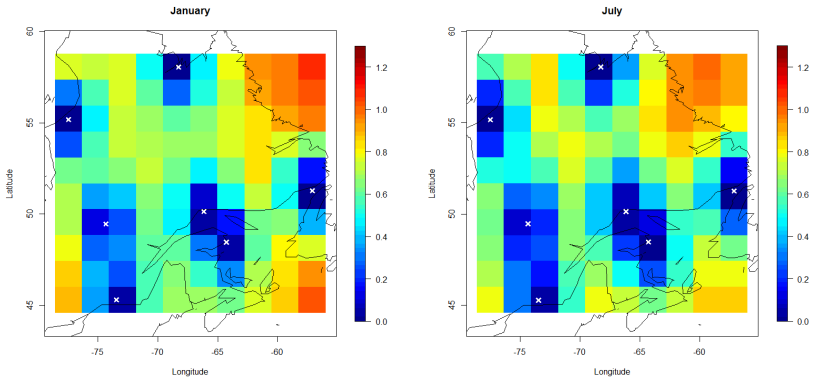


Figure 7: Estimated prediction error for January (left) and July (right) for the final model chosen for each month. Crosses indicate the position of the meteorological stations.

6. Conclusions and further development

In this work we illustrate a possible way to deal with non stationary manifold-valued random fields and we introduce a kriging predictor based on an additive model on the tangent space. We focus here on the case of positive definite matrices, in view of the prediction of the covariance matrix between temperature and precipitation in Quebec. However, this general strategy can be applied to other examples of manifold-valued data, such as shapes, M-reps or directional data.

The proposed method can be also generalized in many different directions. For example, a more complex model can be used for the deterministic drift $A(\mathbf{x}(\mathbf{s}); \boldsymbol{\beta})$. This opens non trivial computational challenges in solving the least square problem. Moreover, kriging prediction can be extended to data belonging to stratified space, such as non negative definite matrices or infinite dimensional covariance operators. Here a tangent space is not well defined in every point; however, it is still possible to define transformations that map the observations in a Hilbert space where the additive model can be used. A simple example in the case of non negative definite matrices is to use the square root transformation.

Other approaches are also possible to model the non stationarity of the manifold-valued random field. An example of this is the probabilistic model on the manifold illustrated for comparison in the second simulation study. This is a special case of a broad family of models whose full distribution is yet to be characterized. Further investigation in this direction could open new perspectives for the statistical analysis of spatially dependent manifold-valued data, for instance through maximum

likelihood methods.

Appendix: dealing with general manifold data

Even though we mainly considered manifold data belonging to $PD(p)$, our approach is of general validity. Indeed, any Riemmanian manifold admits an approximation based on a Hilbert tangent space, where linear (geo)statistical methods can be developed. In this Appendix we rephrase the methodology introduced in this work, in the framework of a general Riemmanian manifold $\mathcal{M}$.

Given a point Σ in $\mathcal{M}$, we call H the Hilbert tangent space at the point Σ, $H = T_\Sigma\mathcal{M}$, and denote by $\langle\cdot,\cdot\rangle_H$ the inner product in H. Given a distance in $\mathcal{M}$, we indicate with $\exp_\Sigma$ the corresponding exponential map in Σ, and with $\log_\Sigma$ its inverse. We call H^n the cartesian product space $H \times ... \times H$, whose elements are arrays of elements of H, that is $H^n \ni \boldsymbol{T} = (T_1, ..., T_n), T_i \in H$. The space H^n is a Hilbert space when endowed with the inner product $\langle \boldsymbol{T}_1, \boldsymbol{T}_2\rangle_{H^n} = \sum_{i=1}^n \langle T_{1i}, T_{2i}\rangle_H$, for $\boldsymbol{T}_1, \boldsymbol{T}_2 \in H^n$. Hereafter, we will use the following matrix notation on H^n: for a matrix $G \in \mathbb{R}^{q\times n}$ and $\boldsymbol{T} \in H^n$, we call $G\boldsymbol{T}$ the element of H^q such that $(G\boldsymbol{T})_i = \sum_{j=1}^n G_{ij}T_j$.

We here focus on the extension to $\mathcal{M}$ of the geostatistical model and the related methods developed in Section 3. Note that the regression model for uncorrelated errors can be obtained as a particular case. For a location $\mathbf{s}$ in the spatial domain D, we model the random element $S(\mathbf{s})$ in $\mathcal{M}$ as

$$S(\mathbf{s}; \boldsymbol{\beta}, \Sigma) = \exp_\Sigma(A(\mathbf{x}(\mathbf{s}); \boldsymbol{\beta}) + \Delta(\mathbf{s})). \tag{16}$$

Here, $A(\mathbf{x}(\mathbf{s}); \boldsymbol{\beta})$ is a drift term in H, described by a linear model with coefficients $\boldsymbol{\beta} = (\beta_0, ..., \beta_r)$ in H and scalar regressors $\mathbf{x}(\mathbf{s})$, collected in the design vector $\mathbf{Z} = (1, \mathbf{x}(\mathbf{s}))^T$:

$$A(\mathbf{x}(\mathbf{s}); \boldsymbol{\beta}) = \sum_{k=0}^{r} \beta_k Z_k.$$

Instead, $\{\Delta(\mathbf{s}), \mathbf{s} \in D\}$ denotes a zero-mean globally second-order stationary and isotropic random field in the Hilbert space H, with covariogram C and semivariogram γ, i.e., for $\mathbf{s}_i, \mathbf{s}_j$ in D, (Menafoglio et al., 2013)

$$\begin{aligned} C(\|\mathbf{s}_i - \mathbf{s}_j\|) &= \mathbb{E}[\langle\Delta(\mathbf{s}_i), \Delta(\mathbf{s}_j)\rangle_H^2] \\ \gamma(\|\mathbf{s}_i - \mathbf{s}_j\|) &= \frac{1}{2}\mathbb{E}[\|\Delta(\mathbf{s}_i) - \Delta(\mathbf{s}_j)\|_H^2]. \end{aligned}$$

Let $\mathbf{s}_1, ..., \mathbf{s}_n$ be n locations in D, and let $S_1, ..., S_n$ the observations of (16) at these locations. As in Section 3, we assume that the regressors $\mathbf{x}(\mathbf{s}_i)$, $i = 1, ..., n$ are known, or observed together with the S_i's. We denote by $\Gamma \in \mathbb{R}^{n\times n}$ the covariance matrix of the array $\mathbf{\Delta} = (\Delta(\mathbf{s}_1), ..., \Delta(\mathbf{s}_n))^T$ in H^n, that is $\Gamma_{ij} = C(\|\mathbf{s}_i - \mathbf{s}_j\|^2)$, and call $\mathbf{R}$ in H^n the array of residuals $R_i = A(\mathbf{x}(\mathbf{s}); \boldsymbol{\beta}) - \log_\Sigma(S_i)$. To estimate $(\Sigma, \boldsymbol{\beta})$ accounting for the spatial dependence, we optimize the generalize least square (GLS) functional

$$(\widehat{\Sigma}, \widehat{\boldsymbol{\beta}}) = \underset{\Sigma\in\mathcal{M}, \boldsymbol{\beta}\in H^{r+1}}{\operatorname{argmin}} ||\Gamma^{-1/2}\mathbf{R}||^2_{H^n}. \tag{17}$$

When Γ is known, problem (17) can be solved iteratively, by alternatively minimizing the GLS functional in (17) with respect to Σ given $\boldsymbol{\beta}$ and to $\boldsymbol{\beta}$ given Σ. The minimizer in $\boldsymbol{\beta}$ given Σ can be always explicitly determined as

$$\widehat{\beta}^{GLS}(\Sigma) = (\mathbb{Z}^T\Gamma^{-1}\mathbb{Z})^{-1}\mathbb{Z}^T\Gamma^{-1}\boldsymbol{Y}(\Sigma), \tag{18}$$

where $\mathbb{Z} \in \mathbb{R}^{n\times(r+1)}$ is the design matrix, $\mathbb{Z}_{ik} = Z_{ik}$, and $\boldsymbol{Y}(\Sigma)$ is the array $\boldsymbol{Y}(\Sigma) = (\log_\Sigma(S_1), ..., \log_\Sigma(S_n))^T \in H^n$. Instead, an expression for the minimizer in Σ given $\boldsymbol{\beta}$ is not available, in general. The complexity of such minimization is problem dependent, and may require the development of specific optimization techniques.

Whenever the spatial dependence is unknown, one can estimate the variogram similarly as proposed in Section 3. Indeed, given $(\Sigma, \boldsymbol{\beta})$ and in the notation of Section 3, the empirical estimate of the semivariogram $\gamma(h)$ can be computed as

$$\widehat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(\mathbf{s}_i, \mathbf{s}_j)\in N(h)} ||\Delta(\mathbf{s}_i) - \Delta(\mathbf{s}_j)||^2_H.$$

A parametric model can be then fitted to the empirical estimate, to obtain a valid model. A nested iterative algorithm analogue to that detailed in Section 3 can be employed for the joint estimation of the spatial dependence and of the model parameters Σ and β.

Estimated the parameters of model (16) as $(\widehat{\Sigma}, \widehat{\boldsymbol{\beta}}, \widehat{\Gamma})$, the kriging prediction can be performed likewise in $PD(p)$. In the Hilbert space H the simple kriging predictor for $\Delta(\mathbf{s})$ is well-defined. It is obtained as $\sum_{i=1}^n \lambda_i^0 \widehat{\Delta}(\mathbf{s}_i)$, where $\widehat{\Delta}(\mathbf{s}_i)$ stands for the estimated residual at $\mathbf{s}_i$, and the vector of kriging weights $\boldsymbol{\lambda_0} = (\lambda_1^0, ..., \lambda_n^0)$ is found as $\boldsymbol{\lambda_0} = \widehat{\Gamma}^{-1}\mathbf{c}$, with $\mathbf{c} = (\widehat{C}(||\mathbf{s}_1 - \mathbf{s}_0||), \ldots, \widehat{C}(||\mathbf{s}_n - \mathbf{s}_0||))^T$. The spatial prediction of S at the target location $\mathbf{s}_0$ is then

$$\widehat{S}_0 = \exp_{\widehat{\Sigma}}(\widehat{\beta}_0^{GLS}(\widehat{\Sigma}) + \sum_{k=1}^r \widehat{\beta}_k^{GLS}(\widehat{\Sigma}) x_k(\mathbf{s}_0) + \sum_{i=1}^n \lambda_i^0 \widehat{\Delta}(\mathbf{s}_i)),$$

where $\mathbf{x}(\mathbf{s}_0)$ is the vector of covariates given at the location $\mathbf{s}_0$.

Finally, the assessment of the prediction error can be performed through semi-parametric bootstrap, by following the same strategy introduced in Section 3.

References

Bhattacharya, R. and Patrangenaru, V. (2003) "Large sample theory of intrinsic and extrinsic sample means on manifolds", *Ann. Stat.*, **31**:1–29.

Bhattacharya, R. and Patrangenaru, V. (2005) "Large sample theory of intrinsic and extrinsic sample means on manifolds – II", *Ann. Stat.*, **33**:1225–1259.

Chilès, J. P. and Delfiner, P. (2009) *Geostatistics: modeling spatial uncertainty*. Wiley, New York.

Cressie, N.A.C. (1993) Statistics for spatial data, Revised edition. Wiley, New York.

Dedieu, J., Priouret, P. and Malajovich, G. (2003) "Newton's method on Riemannian manifolds: convariant alpha theory", *IMA J. Numer. Anal.*, **23**:395–419.

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

Diggle, P. and Ribeiro, P. J. (2007). *Model-based geostatistics*. Springer Series in Statistics.

Dryden, I.L., Koloydenko, A., Zhou, D. (2009), "Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging", *Ann. Appl. Stat.*, **3**:1102–1123.

Fletcher, P.T. (2013) "Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds", *Int. J. Comput. Vision*, **2**:171–185.

Fletcher, P.T., Conglin Lu, Pizer, S.M., Sarang Joshi (2004) "Principal geodesic analysis for the study of nonlinear statistics of shape". *IEEE T. Med. Imaging*, **23**:995–1005.

Freedman, D. A. and Peters, S. C. (1984) "Bootstrapping a regression equation: Some empirical results", *J. Am. Stat. Assoc.*, **79**:97–106.

Gromenko, O., Kokoszka, P., Zhu, L., Sojka, J. (2012) "Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends", *Ann. Appl. Stat.*, **6**:669–696.

Iranpanah, N., Mohammadzadeh, M. and Taylor, C. C. (2011) "A comparison of block and semi-parametric bootstrap methods for variance estimation in spatial statistics", *Comput. Stat. Data An.*, **55**:578–587.

Isaac, G.A., Stuart R.A. (1991) "Temperature-Precipitation Relationships for Canadian Stations", *J. Climate*, **5**:822–830.

Jung, S., Foskey, M., Marron J.S. (2011) "Principal Arc Analysis on direct product manifolds", *Ann. Appl. Stat.*, **5**:578–603.

Lobell, D.B., Burke, M.B. (2008) "Why are agricultural impacts of climate so uncertain? The importance of temperature relative to precipitation", *Environ. Res. Lett.*, **3**, 034007.

Menafoglio, A., Secchi, P. and Dalla Rosa, M. (2013) "A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space", *Electron. J. Stat.*, **7**:2209–2240.

Menafoglio, A., Guadagnini, A. and Secchi, P. (2014) "A Kriging Approach based on Aitchison Geometry for the Characterization of Particle-Size Curves in Heterogeneous Aquifers", *Stoch. Env. Res. Risk A.*, **28**:1835–1851.

Menafoglio, A., and Petris, G. (2015): "Kriging for Hilbert-space valued random fields: the Operatorial point of view", *J. Multivariate Anal.*, DOI: 10.1016/j.jmva.2015.06.012.

Moakher, M. (2005) "On the Averaging of Symmetric Positive-Definite Tensors". *J. Elasticity*, **82**:273–296.

Moakher, M., Zéraï, M. (2011) "The Riemannian Geometry of the Space of Positive-Definite Matrices and Its Application to the Regularization of Positive-Definite Matrix-Valued Data", *J. Math. Imaging Vis.*, **40**:171–187.

Nerini, D., Monestiez, P. and Manté, C. (2010) "Cokriging for spatial functional data", *J. Multivariate Anal.*, **101**:409–418.

Osborne, D., Patrangenaru, V., Ellingson, L., Groisser, D. and Schwartzman, A. (2013) "Nonparametric two-sample tests on homogeneous Riemannian manifolds, Cholesky decompositions and Diffusion Tensor Image analysis", *J. Multivariate Anal.*, **119**:163-175.

Pennec, X. (2006) "Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements". *J. Math. Imaging Vis.*, **25**:127–154.

Pennec, X., Fillard, P., Ayache, N. (2006) "A Riemannian framework for tensor computing", *Int. J. Comput. Vision*, **6**:41–66.

Pigoli, D., Aston, J.A.D., Dryden, I.L. and Secchi, P. (2014) "Distances and Inference for Covariance Operators", *Biometrika*, **101**:409–422.

Pigoli, D. and Secchi P. (2012) "Estimation of the mean for spatially dependent data belonging to a Riemannian manifold", *Electron. J. Stat.*, **6**:1926–1942.

R Development Core Team (2009) "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria, `http://www.R-project.org`.

Shi, X., Zhu, H., Ibrahim, J.G., Liang, F., Lieberman, J. and Styner, M. (2012) "Intrinsic Regression Models for Medial Representation of Subcortical Structures", *J. Am. Stat. Assoc.*, **107**:12–23.

Solow, A. R. (1985) "Bootstrapping correlated data", *Math. Geo.*, **17**:769–775.

Stein, M. L. (1999) Interpolation of spatial data: some theory for kriging. Springer-Verlag, New York.

Trenberth, K.E. and Shea, D.J. (2005) "Relationship between precipitation and surface temperature", *Geophys. Res. Lett.*, **32**, L14703.

Wang, H. and Marron, J.S. (2007) "Object oriented data analysis: Sets of trees", *Ann. Stat.*, **35**:1849–1873.

Wang, Z., Vemuri, B., Chen, Y. And Mareci, T. (2004) "A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI", *IEEE T. Med. Imaging*, **23**:930–939.

Yuan, Y., Zhu, H., Lin, W. and Marron, J.S. (2012) "Local polynomial regression for symmetric positive definite matrices", *J. Roy. Stat. Soc. B*, **74**:697–719.