Stochastic Environmental Research and Risk Assessment (2022) 36:4087–4102
https://doi.org/10.1007/s00477-022-02252-9

ORIGINAL PAPER

A Kronecker-based covariance specification for spatially continuous multivariate data

Angélica Maria Tortola Ribeiro[1,3] · Paulo Justiniano Ribeiro Junior[2,3] · Wagner Hugo Bonat[2,3]

Accepted: 8 May 2022 / Published online: 17 June 2022

Abstract
We propose a covariance specification for modeling spatially continuous multivariate data. This model is based on a reformulation of Kronecker's product of covariance matrices for Gaussian random fields. The structure holds for different choices of covariance functions with parameters varying in their usual domains. In comparison with classical models from the literature, we used the Matérn correlation function to specify the marginal covariances. We also assess the reparametrized generalized Wendland model as an option for efficient calculation of the Cholesky decomposition, improving the model's ability to deal with large data sets. The reduced computational time and flexible generalization for increasing number of variables, make it an attractive alternative for modelling spatially continuous data. The proposed model is fitted to a soil chemistry properties dataset, and adequacy measures, forecast errors and estimation times are compared with the ones obtained based on classical models. In addition, the model is fitted to a North African temperature dataset to illustrate the model's flexibility in dealing with large data. A simulation study is performed considering different parametric scenarios to evaluate the properties of the maximum likelihood estimators. The simple structure and reduced estimation time make the proposed model a candidate approach for multivariate analysis of spatial data.

Keywords Gaussian random fields · Covariance functions · Geostatistics · Matérn correlation model · Multivariate spatial data

1 Introduction

Multivariate random fields have been of interest from the very early days of the geostatistical literature, with an increasing number of proposed approaches as data sets became richer and the ever-increasing computational power. The specification of the covariance structure is central in the estimation and prediction process. Recent contributions include asymmetric models (Qadir et al. 2021; Alegría et al. 2018), modeling on spheres (Bevilacqua et al. 2020; Emery et al. 2019; Alegría et al. 2019; Emery and Porcu 2019), on mapping disease (Martinez-Beneito 2020; MacNab 2018, 2016), on multinary problems (Teichmann et al. 2021), to name a few.

We are interested in multivariate random fields analysis, in the specific context of spatially continuous data. Possible applications cover a wide range of disciplines, such as climatology, meteorology, geophysics, among others, where spatially referenced data is usually of interest. We consider two illustrative examples, one on chemical soil properties relevant for agriculture and another on North African temperatures, important for climatology and environmental sciences.

Let $\mathbf{Y}(\mathbf{s}) = \{Y_1(\mathbf{s}), \ldots, Y_p(\mathbf{s})\}^\top$, on $\mathbb{R}^d$, $d \geq 1$, a p-dimensional multivariate Gaussian random field with mean

✉ Angélica Maria Tortola Ribeiro
angelicaribeiro@utfpr.edu.br

Paulo Justiniano Ribeiro Junior
paulojus@ufpr.br

Wagner Hugo Bonat
wbonat@gmail.com

[1] Academic Department of Statistics, Federal Technological University of Paraná, Curitiba, Paraná 80230-901, Brazil

[2] Department of Statistics, Federal University of Paraná, Curitiba, Paraná 81530-000, Brazil

[3] Post-Graduate Program in Numerical Methods in Engineering, Federal University of Paraná, Curitiba, Paraná 81530-000, Brazil

Table 8 Elapsed estimation time (Elps.Time), number of iterations required (N.Iter) and average time per iteration (Time.by.Iter) for each model, considering simulated data from the MatConstr model for different sample sizes

Models	Size	Elps.Time	N.Iter	Time.by.Iter
MatSimpler	100	5.356	770	0.007
	225	29.018	768	0.038
	400	74.989	564	0.133
	625	380.968	912	0.418
	900	684.169	634	1.079
MatConstr	100	90.341	7826	0.012
	225	130.974	1830	0.072
	400	1112.987	4328	0.257
	625	2937.484	4130	0.711
	900	5700.465	3050	1.869
MatSep	100	6.386	446	0.014
	225	57.690	806	0.072
	400	215.376	754	0.286
	625	496.981	624	0.796
	900	998.202	470	2.124
LMC	100	4.271	755	0.006
	225	27.107	827	0.033
	400	141.529	1035	0.137
	625	481.601	1067	0.451
	900	1939.578	1539	1.260

vector $\mu(\mathbf{s}) = E[\mathbf{Y}(\mathbf{s})]$ and matrix-valued covariance function:

$$\mathbf{\Sigma}(\mathbf{h}) = \text{cov}\{\mathbf{Y}(\mathbf{s_1}), \mathbf{Y}(\mathbf{s_2})\} = [\Sigma_{ij}(\mathbf{h})]_{i,j=1}^{p}, \quad (1)$$

where $\mathbf{h} = \mathbf{s_1} - \mathbf{s_2} \in \mathbb{R}^d$ is the spatial separation vector. We consider a stationary and isotropic process (Chilès and Delfiner 2012; Diggle and Ribeiro Jr 2007; Gneiting 1999) where $i = j$, the functions $\Sigma_{ii}(\mathbf{h})$ in Eq. (1) describe the spatial variability of the ith process $Y_i(\mathbf{s})$, for $i = 1, \ldots, p$, and are referred as the direct- or marginal-covariance functions (Genton and Kleiber 2015) and, if $i \neq j$, the functions $\Sigma_{ij}(\mathbf{h})$ in Eq. (1) describe the spatial variability between the process $Y_i(\mathbf{s})$ and $Y_j(\mathbf{s})$ and are called as cross-covariance functions. An important condition for (1) is that it must meet the positive definite condition, that is, $\mathbf{a}^\top \mathbf{\Sigma a} > 0$, for any vector $\mathbf{a} \neq \mathbf{0}$.

The main goal is to propose a valid covariance specification for (1) in the case of spatially continuous multivariate data. The model, presented in Sect. 2, is based on the Kronecker products and it is quite flexible to handle with two or more variables. Furthermore, we present the conditions for positive definiteness of the proposed model, perform comparisons in terms of computational times and adequacy measures with classical models, and consider compactly supported covariance functions as an efficient approach to compute the Cholesky decomposition.

The literature on covariance functions for multivariate random fields is extensive. A careful review of the main works in the area can be found in Genton and Kleiber (2015) and Salvaña and Genton (2020). An intuitive early proposal and possibly the most traditional model is the linear model of corregionalization (LMC) (Goulard and Voltz 1992; Bourgault and Marcotte 1991; Wackernagel 2003). The key idea for the LMC is the overlap of spatial processes in order to induce a multivariate field. This approach is widely explored, including under the Bayesian approach for inference and prediction. Finley et al. (2015), Banerjee et al. (2003), Gelfand et al. (2004), Schmidt and Gelfand (2003) and Cecconi et al. (2016) are examples where the LMC structure underlies the models.

Another popular structure considers the class of Matérn correlation functions (Matérn 1986; Guttorp and Gneiting 2006). For the univariate case, the Matérn class covariance model is defined as $\sigma^2 M(\mathbf{h}|\nu, \phi)$, where $M(\mathbf{h}|\nu, \phi) = \frac{2^{1-\nu}}{\Gamma(\nu)}(|\mathbf{h}|/\phi)^\nu K_\nu(|\mathbf{h}|/\phi)$, is the Matérn spatial correlation at distance $|\mathbf{h}|$, K_ν is the modified Bessel function, $\sigma^2, \nu, \phi > 0$ are the variance, smoothness and scale parameters, respectively. When $\nu = 0.5$, the Matérn model reduces to the exponential covariance function. Gneiting et al. (2010) elegantly extended this class for multivariate case considering the Matérn family for the marginal and cross-covariance functions. The authors present conditions for the parameters that lead to a valid covariance structure, the full bivariate Matérn model. For more than two variables the authors presented the parsimonious multivariate Matérn model, which considers common scale and constrained smoothness parameters. Another important specification are the separable models, which considers that the components of the multivariate random field share the same correlation structure (Bevilacqua et al. 2016a; Vallejos et al. 2020) and appears as a parsimonious modeling alternative because it allows a simplification of the more complex models. Bevilacqua and Morales-Oñate (2018) and Vallejos et al. (2020) present two simplified structures with Matérn correlation function for bivariate data, the bivariate separable Matérn model and the bivariate Matérn model with constraints. Both models can be estimated by the `Geomodels` package (Bevilacqua and Morales-Oñate 2018).

The aforementioned models are widely assessed in the geostatistical literature. Cressie (1993), Gneiting et al. (2010), Goovaerts et al. (1997), Porcu et al. (2013), Bevilacqua et al. (2016b) noticed some difficulties to handling with the LMC due, for example, to its lack of flexibility and difficulty in recovering the smoothness of latent processes. Separable models are not capable to

capture the different scales and smoothness for the variables under study (Bevilacqua et al. 2016a, 2015). The bivariate Matérn model presents some restrictions in the parametric space. Vallejos et al. (2020) note that variation of the colocated correlation parameter is constrained by the values of the scale and smoothness parameters, resulting in difficulties for the estimation process and parameter interpretation.

The covariance specification presented here emerges as an additional modeling alternative for multivariate random fields that can be more flexible to deal with two or more variables, as it allows the model parameters to vary freely in their usual parametric domains. With a simple construction, the computational implementation has no major difficulties for a number of variables and sampling location points, with a parsimonious estimation computational time. Furthermore, unlike the separable models, our proposal allows different marginal correlation structures, making it able to capture the structure of each variable.

The article is organized as follows. In Sect. 2 we present our covariance specification for multivariate spatial data and discuss some results. Section 3 provides an efficient approach for calculating the Cholesky decomposition using compactly supported covariance functions. The dataset analyses are presented in Sect. 4. In Sect. 5, through a simulation study, we evaluate the properties of the proposed model estimators. Finally, the main conclusions are summarized in Sect. 6. The model implementation and reported analysis are performed using the computational statistical software R (R Core Team 2021).

2 Model specification

This section presents our proposed covariance specification for multivariate Gaussian random fields, which is based upon Martinez-Beneito (2013).We present the proof of its validity and discuss how to obtain the maximum likelihood estimates of the model parameters. We also present computational time estimation results comparing it with classical approaches.

In Martinez's proposal, the results are presented for modeling multivariate mapping diseases problems based on Gaussian Markov random fields (GMRF), which are discretely indexed, following a Gaussian multivariate distribution with the additional restriction of conditional independence (Rue and Held 2005).

Our proposal extends Martinez's approach to construct a covariance function for Gaussian random fields that are continuously indexed, with several applications in geostatistical problems.

We present a simple construction that allows its generalization to larger dimensions more easily. The idea is to write the cross-covariance matrix as a product of matrices that induce variability within processes and between processes, and it is built upon the Kronecker products reformulation of covariance matrices. The resulting construction will be always positive definite for any parameter values in their usual domains.

Consider a symmetric correlation matrix Σ_b, with dimension $p \times p$, induces a correlation between spatial processes, while the marginal-covariance functions Σ_{ii}, for $i = 1, \ldots, p$, model the variability within each process. We specify the covariance matrix for the $\mathbf{Y}$ process considering the generalized Kronecker product, presented in Martinez-Beneito (2013). Thus, for the Gaussian random fields continuously indexed, the matrix-valued covariance function is defined by:

$$\begin{aligned}\boldsymbol{\Sigma}(\mathbf{h}) = \text{Bdiag}\left(\tilde{\boldsymbol{\Sigma}}_{11}, \tilde{\boldsymbol{\Sigma}}_{22}, \ldots, \tilde{\boldsymbol{\Sigma}}_{pp}\right)(\boldsymbol{\Sigma}_b \otimes \mathbf{I}) \\ \text{Bdiag}\left(\tilde{\boldsymbol{\Sigma}}_{11}^{\top}, \tilde{\boldsymbol{\Sigma}}_{22}^{\top}, \ldots, \tilde{\boldsymbol{\Sigma}}_{pp}^{\top}\right),\end{aligned} \tag{2}$$

where, $\tilde{\boldsymbol{\Sigma}}_{ii}$ is the lower triangular matrix of the Cholesky decomposition of the matrix Σ_{ii}, Bdiag represents the matrix in diagonal blocks of the matrices $\tilde{\boldsymbol{\Sigma}}_{11}$, $\tilde{\boldsymbol{\Sigma}}_{22}$, ..., $\tilde{\boldsymbol{\Sigma}}_{pp}$ and $\mathbf{I}$ is the identity matrix. The structure defined in (2) is very flexible, allowing different marginal-covariance functions for Σ_{ii} and different correlation structures for $\boldsymbol{\Sigma}_b$.

Without loss of generality, we will consider that the correlation between the processes will be induced by the matrix:

$$\boldsymbol{\Sigma}_b = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{pmatrix}, \tag{3}$$

where ρ_{ij}, $ij = 1, \ldots, p$, is the correlation parameter between the variables i and j.

To quantify the variability within each process, different marginal-covariance structures could be used in (2). In a general way, we can write:

$$\boldsymbol{\Sigma}_{ii}(\mathbf{h}) = \sigma_i^2 R(\mathbf{h}|\Psi_i), \tag{4}$$

where $R(\mathbf{h}|\Psi_i)$ is a valid correlation function, with Ψ_i denoting the parameters vector that model the spatial dependence structure of the i-th component, for $i = 1, \ldots, p$. For simplification and without loss of generality, we can consider for $R(\mathbf{h}|\Psi_i)$, the Matérn correlation function. Thus, the marginal covariance function takes the form:

$$\boldsymbol{\Sigma}_{ii}(\mathbf{h}) = \sigma_i^2 M(\mathbf{h}|\nu_i, \phi_i), \quad \text{for} \quad i = 1, 2, \ldots, p. \tag{5}$$

The structure specified by (2), (3) and (5) will be called as *simpler multivariate Matérn* (MatSimpler) model. It accepts different marginal behaviors and it is able to handle

with different smoothness and scale parameters for each variable. The *simpler multivariate exponential* (ExpSimpler) model is a particular case when the exponential correlation function is used. (Ribeiro et al. 2021) illustrates the ExpSimpler model in bivariate analysis of meteorological data.

In Theorem 2.1, below, we prove the validity of our covariance specification for modeling multivariate spatial data.

Theorem 2.1 *Let* Σ_{ii}, *for* $i = 1, \ldots, p$, *the marginal covariance functions of dimension* $n \times n$, $\mathbf{\Sigma}_b$ *a valid spatial correlation function of dimension* $p \times p$ *and* $\mathbf{I}$ *the identity matrix of dimension* $n \times n$, *then the covariance function defined in* (2) *is a valid and full rank* np *specification for multivariate spatial data modeling.*

Proof Since the marginal covariance functions, Σ_{ii}, for $i = 1, \ldots, p$, are symmetric positive definite matrix, the matrices $\tilde{\mathbf{\Sigma}}_{ii}$, resulting from the Cholesky decomposition, are lower triangular with positive diagonal elements and therefore, full rank (Banerjee and Roy 2014). Thus, $\text{rank}(\tilde{\mathbf{\Sigma}}_{ii}) = n$, for all i, and from the rank properties of block-diagonal matrix, the rank of any block-diagonal matrix is the sum of the ranks of its diagonal blocks (Banerjee and Roy 2014), that is:

$$\text{rank}\left[\text{Bdiag}\left(\tilde{\mathbf{\Sigma}}_{11}^{\top}, \tilde{\mathbf{\Sigma}}_{22}^{\top}, \ldots, \tilde{\mathbf{\Sigma}}_{pp}^{\top}\right)\right] = \sum_{i=1}^{p} \text{rank}(\tilde{\mathbf{\Sigma}}_{ii}) = np,$$

therefore, $\text{Bdiag}\left(\tilde{\mathbf{\Sigma}}_{11}^{\top}, \tilde{\mathbf{\Sigma}}_{22}^{\top}, ..., \tilde{\mathbf{\Sigma}}_{pp}^{\top}\right)$ is a full np rank matrix.

On the other hand, since $\mathbf{\Sigma}_b$ and $\mathbf{I}$ are positive definite matrices, it follows by the kronecker product properties that $(\mathbf{\Sigma}_b \otimes \mathbf{I})$ is also a positive definite matrix (Hardy and Steeb 2019). With $\mathbf{\Sigma}_b$ of dimension p and $\mathbf{I}$ of dimension n, the resulting kronecker product between them will be a positive definite matrix of dimension np.

Now, for simplicity of notation, let's denote by $\mathbf{A}$ and $\mathbf{B}$ the respective $(\mathbf{\Sigma}_b \otimes \mathbf{I})$ and $\text{Bdiag}\left(\tilde{\mathbf{\Sigma}}_{11}^{\top}, \tilde{\mathbf{\Sigma}}_{22}^{\top}, \ldots, \tilde{\mathbf{\Sigma}}_{pp}^{\top}\right)$ matrices. Since $\mathbf{A}$ is a positive definite matrix and $\mathbf{B}$ is a full rank matrix, follows that $\mathbf{B}^{\top}\mathbf{A}\mathbf{B}$ preserves not only the rank but also the positive definiteness (Gentle 2017; Petersen et al. 2008).

To visualize this, let $\mathbf{a} \neq \mathbf{0}$, any vector of dimension np, and let $\mathbf{z} = \mathbf{B}\mathbf{a}$, where $\mathbf{z} \neq \mathbf{0}$, because $\mathbf{B}$ is a full rank matrix. Using the positive definite matrix definition, follows:

$$\begin{aligned}\mathbf{a}^{\top}(\mathbf{B}^{\top}\mathbf{A}\mathbf{B})\mathbf{a} &= (\mathbf{a}\mathbf{B})^{\top}\mathbf{A}(\mathbf{B}\mathbf{a}) \\ &= \mathbf{z}^{\top}\mathbf{A}\mathbf{z} \\ &> 0\end{aligned}$$

□

The result holds if variables are observed in different numbers of sample locations, since the incomplete data can be treated as missing information and this does not imply any additional complexity and the proof of theorem 2.1 remains valid.

The dimensions of the resulting covariance matrix will depend on the number of sample locations for each variable considered in the analysis. If we consider marginal covariance functions with dimension n_i, for $i = 1, \ldots, p$, the resulting covariance matrix will have dimension $N = \sum_{i=1}^{p} n_i$.

Considering that $\mathbf{\Sigma}$ is a valid covariance specification for any valid choice of marginal-covariance and correlation functions, the proposed model allows its parameters to vary in their usual domains, favoring the inferential process and allowing the model parameters to be more easily interpreted.

In our covariance specification, considering the correlation structure defined in (3), the matrix-valued covariance function $\mathbf{\Sigma}(\mathbf{h})$, can be written in a more compact form, in terms of the cross-covariance matrices between the process. Thus, for the ij-th component, with $1 \leq i \neq j \leq p$, the cross-covariance function takes the form:

$$\mathbf{\Sigma}_{ij}(\mathbf{h}) = \rho_{ij}\tilde{\mathbf{\Sigma}}_{ii}(\mathbf{h})\tilde{\mathbf{\Sigma}}_{jj}(\mathbf{h})^{\top}. \tag{6}$$

Clearly, when $i = j$, $\rho_{ii} = 1$ and we achieve $\rho_{ii}\tilde{\mathbf{\Sigma}}_{ii}(\mathbf{h})\tilde{\mathbf{\Sigma}}_{ii}(\mathbf{h})^{\top} = \mathbf{\Sigma}_{ii}(\mathbf{h})$, the spatial covariance matrix of the i-th component.

The Theorem 2.2 shows that the separable model (Vallejos et al. 2020; Bevilacqua et al. 2016a) is a particular case of the simpler covariance model class.

Theorem 2.2 *Let* $Y_1(\mathbf{s}), Y_2(\mathbf{s}), \ldots, Y_p(\mathbf{s})$ *a p-dimensional random field with the same spatial dependence structure for all* $Y_i(\mathbf{s})$, $i = 1, \ldots, p$, *then the simpler covariance model specified by* (2), (3) *and* (4) *is reduced to the class of separable models.*

Proof By (6) and considering the general structure of the marginal covariance matrices, defined in (4), we see that in the particular case where the process components share the same spatial dependency structure, that is, $\Psi_i = \Psi_j = \Psi$, for all $i, j = 1, \ldots, p$, the simpler covariance specification reduces to the class of separable models, ie,

$$\begin{aligned}\mathbf{\Sigma}_{ij}(\mathbf{h}) &= \rho_{ij}\tilde{\mathbf{\Sigma}}_{ii}(\mathbf{h})\tilde{\mathbf{\Sigma}}_{jj}(\mathbf{h})^{\top} \\ &= \rho_{ij}\sigma_i\tilde{R}(\mathbf{h}|\Psi)\sigma_j\tilde{R}(\mathbf{h}|\Psi)^{\top} \\ &= \rho_{ij}\sigma_i\sigma_j\tilde{R}(\mathbf{h}|\Psi)\tilde{R}(\mathbf{h}|\Psi)^{\top} \\ &= \rho_{ij}\sigma_i\sigma_j R(\mathbf{h}|\Psi), \text{ for } i,j = 1,\ldots,p \text{ and } \rho_{ii} = 1.\end{aligned}$$

Here, $\tilde{R}(\mathbf{h}|\Psi)$ is the lower triangular Cholesky decomposition of the correlation matrix, $R(\mathbf{h}|\Psi)$. □

2.1 Estimation and inference

For the estimation process, let $N = np$ and $\mathbf{Y} = \{\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_p^\top\}^\top$, be the $N \times 1$ stacked vector of response variables, with $N \times 1$ mean-vector $\mu = \{\mu_1^\top, \ldots, \mu_p^\top\}^\top$, where $\mu_i = \mathbf{X}_i\beta_i$ denotes the $n \times 1$ vector of expected values for the response variable $\mathbf{Y}_i$, $i = 1, \ldots, p$, with $\mathbf{X}_i$ being a $n \times k_i$ design matrix composed of k_i covariates and, consequently, β_i denotes a $k_i \times 1$ regression parameter vector. We suppress the spatial indexes for convenience.

We will denote the set of parameters to be estimated by $\theta = (\beta^\top, \lambda^\top)^\top$, where $\beta = (\beta_1^\top, \ldots, \beta_p^\top)^\top$ denotes the regression parameters vector and $\lambda = (\rho_1, \ldots, \rho_{p(p-1)/2}, \sigma_1, \ldots, \sigma_p, \nu_1, \ldots, \nu_p, \phi_1, \ldots, \phi_p)^\top$ is the covariance specification parameters vector. Considering $\mathbf{y}$ the stacked vector of observed values, the log-likelihood function for θ is given by:

$$\mathcal{L}(\theta; \mathbf{y}) = -\frac{1}{2}\Big[N\ln(2\pi) + \ln|\mathbf{\Sigma}(\lambda)| + (\mathbf{y} - \mu(\beta))^\top \mathbf{\Sigma}(\lambda)^{-1}(\mathbf{y} - \mu(\beta))\Big]. \tag{7}$$

The covariance matrix proposed in (2) involves block-diagonal matrices and a Kronecker product. Thereby, the calculation of its determinant and inverse can be expressed, respectively, by $|\mathbf{\Sigma}| = \left(\prod_{i=1}^p |\tilde{\mathbf{\Sigma}}_{ii}|\right)^2 |\mathbf{\Sigma}_b|^n$ and $\mathbf{\Sigma}^{-1} = \mathbf{Q}^\top\left(\mathbf{\Sigma}_b^{-1} \otimes \mathbf{I}\right)\mathbf{Q}$, where $\mathbf{Q} = \text{Bdiag}\left(\tilde{\mathbf{\Sigma}}_{11}^{-1}, \ldots, \tilde{\mathbf{\Sigma}}_{pp}^{-1}\right)$. It is worth mentioning that the calculation of the inverse of the covariance matrix in the log-likelihood function will not involve the complete matrix but only the Cholesky decompositions of the marginal covariance matrices, which allows computational advantages and reduction of the estimation time. Therefore, the log-likelihood function at (7) can be rewritten as,

$$\mathcal{L}(\theta; \mathbf{y}) = -\frac{1}{2}\left[N\ln(2\pi) + 2\sum_{i=1}^{p}\sum_{j=1}^{n}\ln\left(k_{jj}^{(i)}\right) + n\ln|\mathbf{\Sigma}_b|\right] - \frac{1}{2}\left[(\mathbf{y} - \mu(\beta))^\top \mathbf{Q}^\top\left(\mathbf{\Sigma}_b^{-1} \otimes \mathbf{I}\right)\mathbf{Q}(\mathbf{y} - \mu(\beta)\right], \tag{8}$$

where $k_{jj}^{(i)}$ is the jth diagonal element of $\tilde{\mathbf{\Sigma}}_{ii}$.

The vector of expected values, $\mu(\beta)$, depends on the regression parameters while the covariance specification $\mathbf{\Sigma}(\lambda)$ depends on the covariance parameters vector, λ. We obtain the maximum likelihood estimators by maximizing the function in (8) with respect to the θ parameter vector.

The proposed covariance specification, in addition to its flexibility and interpretability of its parameters, also reduces the number of parameters to be estimated when compared to the multivariate Matérn model (Gneiting et al. 2010), which is quite useful from the point of view of the estimation process. Considering the Matérn correlation function, the number of parameters involved in the MatSimpler model is $p(p+1)/2 + 2p$, where p is the number of variables. In contrast, the number of parameters for the multivariate Matérn model is $p(p+1)/2 + 2(p(p+1)/2)$. Table 1 summarizes how the proposed covariance specification reduces the number of parameters as p increases. For $p > 7$, the number of parameters for multivariate Matérn model is more than twice the number of parameters for the MatSimpler model.

Table 1 Number of parameters involved in the specification of the covariance function, when the number of variables p increases, for the MatSimpler and the multivariate Matérn models

Models	p								
	2	3	4	5	6	7	8	9	10
MatSimpler	7	12	18	25	33	42	52	63	75
Multivariate Matérn	9	18	30	45	63	84	108	135	165

In Sect. 5 through a simulation study we evaluate some properties of the maximum likelihood estimators. In Appendix A we describe the score function, the Newton scoring iterative algorithm and we find the Fisher information matrix associated with the proposed model.

2.2 Prediction

For the case of multivariate spatial data, spatial prediction is a generalization of the univariate case that consists of predicting $\mathbf{Y}$ at some unknown location, $\mathbf{s}_0$, based on other sample information, $\mathbf{s}_i$, for $i = 1, \ldots, n$ (Ver Hoef and Cressie 1993; Bivand et al. 2008; Pebesma 2004).

Let $\mathbf{\Sigma}_{\mathbf{Y}_1}$ be the covariance matrix of $\mathbf{Y}_1 = \mathbf{Y}(\mathbf{s}_i)$, for $i = 1, \ldots, n$, $\mathbf{\Sigma}_{\mathbf{Y}_0}$ be the covariance matrix of $\mathbf{Y}_0 = \mathbf{Y}(\mathbf{s}_0)$, $\mathbf{\Sigma}_{\mathbf{Y}_1\mathbf{Y}_0}$ be the covariance matrix between $\mathbf{Y}_1$ and $\mathbf{Y}_0$ and $\mathbf{d}_0 = \text{Bdiag}(\mathbf{x}_1(\mathbf{s}_0), \ldots, \mathbf{x}_p(\mathbf{s}_0))$. Then, the best linear unbiased predictor for $\mathbf{Y}_0$ is:

$$E(\mathbf{Y}_0|\mathbf{Y}_1) = E(\mathbf{Y}_0) + \mathbf{\Sigma}_{\mathbf{Y}_1\mathbf{Y}_0}^\top \mathbf{\Sigma}_{\mathbf{Y}_1}^{-1}(\mathbf{Y}_1 - E(\mathbf{Y}_1)),$$

with prediction covariance matrix:

$$Cov(\mathbf{Y}_0|\mathbf{Y}_1) = \mathbf{\Sigma}_{\mathbf{Y}_0} - \mathbf{\Sigma}_{\mathbf{Y}_1\mathbf{Y}_0}^\top \mathbf{\Sigma}_{\mathbf{Y}_1}^{-1}\mathbf{\Sigma}_{\mathbf{Y}_1\mathbf{Y}_0}.$$

We can replace the unknown parameters of the model by their respective maximum likelihood estimators (Martins et al. 2016) and, with this, we obtain the stacked vector prediction for the p variables in the $\mathbf{s}_0$ unobserved locations.

2.3 Computational resources

In our computational implementation, we used the `R` statistical software (R Core Team 2021) for the estimation, simulation and prediction procedures. For the implementation of the proposed covariance matrix specification, we used the `kronecker` function, basic to `R`, which enables the calculation of the kronecker product between the matrices $\mathbf{\Sigma}_b$ and $\mathbf{I}$. We use the `Matrix` (Bates and Maechler 2021) package to perform matrix operations more efficiently, such as the Cholesky decomposition of the marginal-covariance matrices, through the `chol` function, and the construction of the diagonal block matrix $\text{Bdiag}\left(\tilde{\mathbf{\Sigma}}_{11}, \tilde{\mathbf{\Sigma}}_{22}, \ldots, \tilde{\mathbf{\Sigma}}_{pp}\right)$, through the `bdiag` function. Furthermore, the `crossprod` and `tcrossprod` functions allowed a more efficient calculation of products between matrices. When working with the Matérn correlation function we use the `matern` function from `geoR` (Ribeiro Jr et al. 2020) package. For the efficient calculation of the Cholesky factor, when working with sparse correlation functions, we use functions from the `spam` package (Furrer and Sain 2010).

For simulations we use the `rmvn` function from `mvnfast` package (Fasiolo 2016) that provides computationally efficient methods related to the multivariate normal distribution. For the log-likelihood function optimization, we use `optim` function. R codes are available in on-line supplementary material.

2.4 Computational results

As already mentioned, the Simpler covariance model can be extended for more than two variables with relative ease. The computational time estimation will depend on the number of variables and sample locations considered in the analysis. To illustrate the computational time spent on estimation, we implement the generic model for p variables and n sample locations and simulate scenarios of the MatSimpler model considering different sample sizes and variable numbers. To make the simulation process easier, we set $\phi_i = 0.2$, $\nu_i = 0.5$, $\sigma_i = 0.3$, for all $i = 1, \ldots, p$. The correlation parameters were chosen between -0.7 to 0.7 such that the resulting $\mathbf{\Sigma}_b$ was a valid structure. We consider the number of variables, p, ranging from 2 to 6 and the number of sample locations, $n = (100, 225, 400, 625, 900)$, taken in a unit square grid.

The results, presented in Appendix B (Fig. 7) shows that the estimation computational time increases with the number of variables and sample locations, which is due to the Cholesky decompositions of the marginal-covariance matrices. In the Sect. 3 we present an approach to make the Cholesky decomposition calculation more efficient, especially for large data.

We also compare the estimation computational times for the bivariate case of the MatSimpler model with three other literature models, the *bivariate Matérn model with constraints* (MatConstr), the *bivariate separable Matérn model* (MatSep) and the LMC model. The data were simulated from the MatConstr model. We set $\phi_i = 0.2$, $\nu_i = 0.5$, $\sigma_i = 0.3$, for $i = 1, 2$, and $\rho_{12} = 0.8$. The MatConstr, MatSep and LMC models were estimated by `GeoModels` package, in which we consider the standard likelihood function. For all models we used the Nelder-Mead optimizer and convergence was successful in all scenarios.

The Table 8 in Appendix B presents the results for all models with respect to the elapsed estimation time (Elps.Time), number of iterations required (N.Iter) and average time per iteration (Time.by.Iter) and Fig. 8 illustrates the Time.by.Iter of each model. The Elps.Time was calculated using the function `system.time` of the R software.

Based on the simulations performed, we see that the MatSimpler model, with 7 parameters, presents lower Time.by.Iter values compared to the other models with the Matérn correlation function for all sample sizes. Compared to the LMC, the MatSimpler model presents lower Time.by.Iter values for sample sizes greater than 400. The model also presents a much lower Elps.Time and N.Iter when compared to the MatConstr model which have the same number of parameters. The R codes and the simulation results presented in Figs. 7 and 8 are available in the supplementary material.

3 An efficient way to calculate the Cholesky factor

The Matérn covariance model is a globally supported model that has been widely used in spatial statistics due to its flexibility and for its well-discussed theoretical justifications in the area of spatial statistics. It has interesting particular cases, such as the exponential and Gaussian models. However, from a computational point of view, this model presents the restriction of generating dense matrices and in this case, the calculation of the Cholesky decomposition becomes impractical when the sample size n increases.

Recent approaches (Bevilacqua et al. 2019, 2022) suggest that working with compactly supported covariance matrices has a computational advantage over globally supported covariance models, such as the Matérn model, for example, since they favor the use of algorithms for sparse matrices, reducing the computational complexity and consequently the estimation time.

As previously described, our Simpler model specification has the flexibility to accommodate different marginal covariance structures, unlike the multivariate Matérn model and its simplifications, which consider only the

Matérn correlation function in its marginal specifications. Therefore, in this section, we will show how the use of compactly supported covariance functions can be used to reduce the computational complexity, and consequently the estimation time for large samples. In particular, we will work with the Generalized Wendland model.

The Generalized Wendland family represents a class of isotropic correlation functions defined by:

$$GW_{v,\mu,\phi}(\mathbf{h}) = \begin{cases} \frac{1}{B(2v,\mu+1)} \int_{\frac{\mathbf{h}}{\phi}}^{1} x(x^2 - (\mathbf{h}/\phi)^2)^{v-1}(1-x)^{\mu} dx, & 0 \le \mathbf{h} \le \phi \\ 0, & \mathbf{h} \ge \phi \end{cases} \tag{9}$$

where $B(.)$ denotes the beta function. The function in (9) is positive definite in $\mathbb{R}^d$ for $\mu \ge (d+1)/2 + v$, $v \ge 0$ and ϕ is a positive compact support (Zastavnyi and Trigub 2002).

Bevilacqua et al. (2022) showed that:

$$\psi_{v,\mu,\phi}(\mathbf{h}) = GW_{v,\mu,\tilde{\phi}_{v,\mu,\phi}}(\mathbf{h}), \tag{10}$$

where $\tilde{\phi}_{v,\mu,\phi} = \phi\left(\frac{\Gamma(\mu+2v+1)}{\Gamma(\mu)}\right)^{\frac{1}{1+2v}}$, is a positive definite reparametrization of the Generalized Wendland family for $\mu \ge (d+1)/2 + v$, $v \ge 0$ and $\phi > 0$, whose compact support is jointly specified by the parameters v, μ and ϕ. This reparametrization allows considering, from the same point of view, correlation functions of compact and global support. In particular, the Matérn model, $M(\mathbf{h}|v+\frac{1}{2},\phi)$, is a special limit case of the model $\psi_{v,\mu,\phi}(\mathbf{h})$ (when $\mu \to \infty$).

To illustrate the reparameterized Wendland compactly supported model in reducing computational complexity and estimating time, we simulate a zero-mean Gaussian random process with MatSimpler covariance model. We set $\sigma_1 = \sigma_2 = 0.3$ and $\phi_1 = \phi_2 = 0.1$, so the practical range is approximately 0.3. We estimate the parameters considering the MatSimpler and the RGW-Simpler models, where the notation RGW-Simpler will be used to denote the model specified by the equations by (2), (3) and (4), considering the reparametrized Generalized Wendland model, in (10), for the correlation function, $R(\mathbf{h}|\Psi_i)$. Regarding the RGW-Simpler model estimation, we set $\mu = 1.5$, resulting in a sparse covariance matrix with percentage of null entries around 85%. The vector of parameters to be estimated is $\theta = (\phi_1, \phi_2, \sigma_1, \sigma_2, \rho_{12})^{\top}$.

The Table 2, presents the Elps.Time, N.Iter and the Time.by.Iter (in seconds). In addition, it provides information about the log-likelihood value and parameter estimates.

Although the log-likelihood values were smaller for the RGW-Simpler model, the reduction in estimation computational time is notable when considering large sample sizes. This shows that considering compactly supported marginal-covariance structures can additionally reduce the computational complexity of the model.

4 Data analysis

This section illustrates the application of the proposed model in two datasets. The first one, the soil dataset, aims to model the relationship between hydrogen content and the catium

Table 2 Elapsed estimation time (Elps.Time), number of iterations required (N.Iter), average time per iteration (Time.by.Iter), log-likelihood value (LL) and parameters estimates for each model, considering simulated data from the MatSimpler model for different sample sizes

Models									
Size	Elps.Time	N.Iter	Time.by.Iter	LL	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$
MatSimpler									
900	494.96	436	1.13	835.02	0.085	0.121	0.282	0.336	0.727
					(0.009)	(0.018)	(0.015)	(0.024)	(0.016)
1600	2222.48	400	5.55	1929.64	0.103	0.099	0.309	0.295	0.706
					(0.013)	(0.012)	(0.019)	(0.017)	(0.012)
2500	6354.83	350	18.15	3590.12	0.105	0.110	0.305	0.311	0.702
					(0.013)	(0.015)	(0.019)	(0.020)	(0.010)
3600	21197.48	398	53.25	5765.22	0.115	0.095	0.319	0.295	0.707
					(0.016)	(0.010)	(0.022)	(0.016)	(0.008)
RGW-Simpler									
900	265.28	490	0.54	779.35	0.112	0.112	0.326	0.333	0.723
					(0.001)	(0.001)	(0.007)	(0.008)	(0.016)
1600	1099.18	494	2.22	1872.65	0.117	0.117	0.337	0.326	0.707
					(0.0008)	(0.0007)	(0.006)	(0.005)	(0.012)
2500	2863.52	456	6.28	3439.73	0.095	0.094	0.302	0.299	0.699
					(0.0003)	(0.0004)	(0.004)	(0.004)	(0.010)
3600	5782.80	341	16.96	5529.29	0.100	0.099	0.312	0.317	0.706
					(0.0003)	(0.0003)	(0.003)	(0.003)	(0.008)

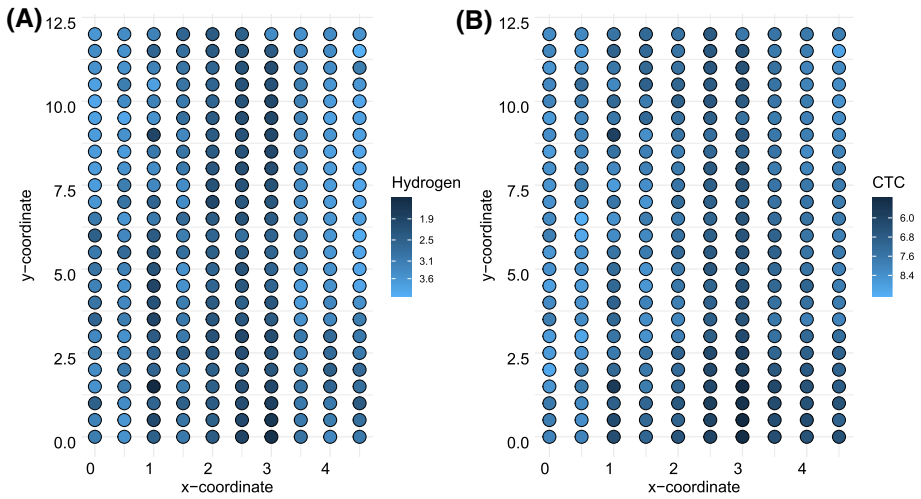

Fig. 1 Circle plot of the **a** Hydrogen content and **b** CTC for soil250 data

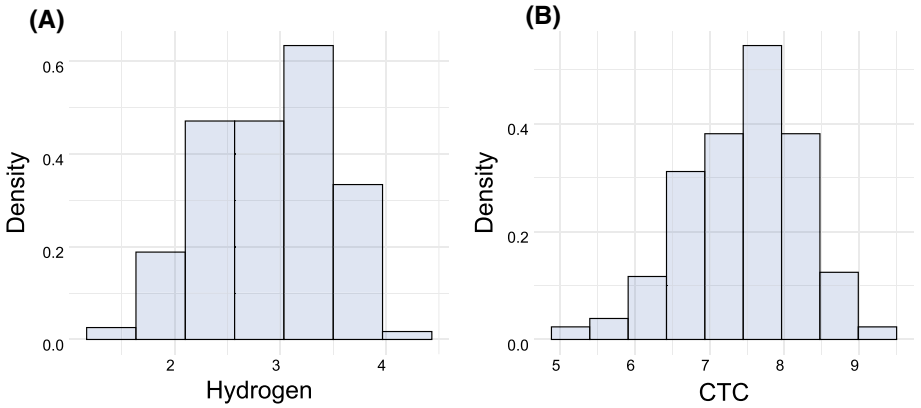

Fig. 2 Histogram of **a** Hydrogen content and **b** CTC for soil250 data

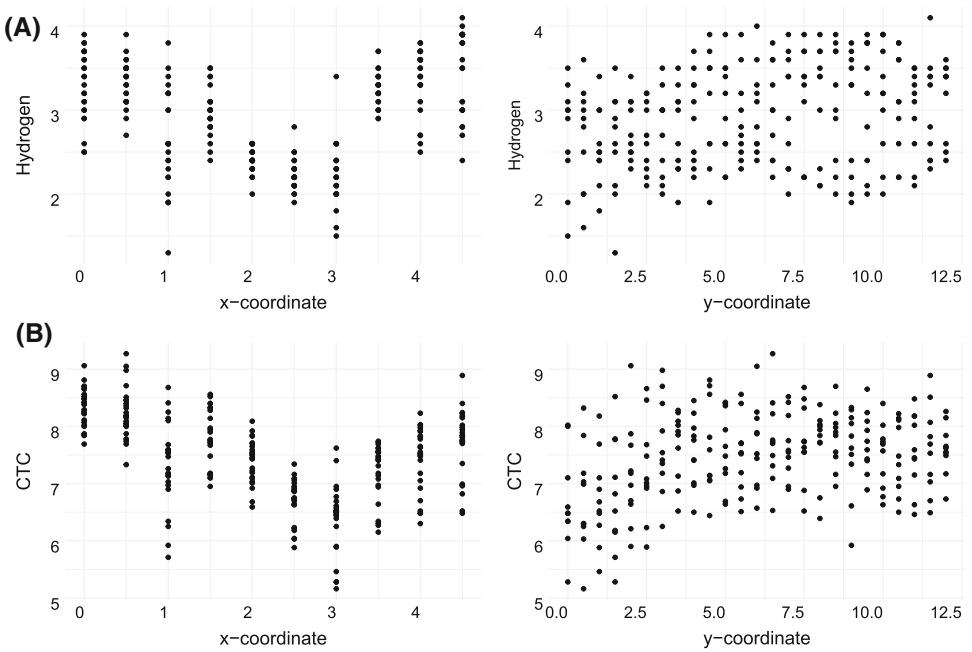

Fig. 3 Scatterplots of **a** Hydrogen content and **b** CTC against the coordinates for soil250 data

exchange capability (CTC). We compare results obtained fitting the MatSimpler and other models. The second one, a North African temperature data is used to illustrate the model's flexibility in dealing with large datasets.

4.1 Example 1: Soil data

The *soil250* dataset from `geoR` package (Ribeiro Jr et al. 2020) contains some soil chemistry properties measured on a regular grid with 10x25 points spaced by 5 meters. We study the relation between the hydrogen content and the catium exchange capability (CTC). These data illustrate a scenario with strong correlation between the variables under study. Figure 1 shows circle plots of the hydrogen (left panel) and CTC (right panel) data separately. The coordinates are divided by a constant to easy the visualizations.

Figure 2 shows the histograms of each variable. Scatterplots of each variable against each spatial coordinate in Fig. 3 are used to check for spatial trends. We proceed with the residuals of a linear regression with constant trend as a realization of a zero-mean bivariate isotropic stationary Gaussian random field. The sample correlation between the variables was 0.68. Standard deviations are 0.60 for hydrogen content and 0.77 for the CTC.

The models considered in the estimation step are the MatSimpler, LMC, MatConstr, MatInd and MatSep. For MatSimpler and MatConstr the vector of parameters to be estimated is $\lambda = (\phi_1, \phi_2, \nu_1, \nu_2, \sigma_1, \sigma_2, \rho_{12})$, for the separable model, MatSep, we have $\phi_1 = \phi_2$ and $\nu_1 = \nu_2$ and for the independent model we have $\rho_{12} = 0$. The LMC model is parameterized by $\lambda = (a_{11}, a_{12}, a_{22}, a_{21}, \phi_1, \phi_2)$, where a_{11}, a_{12}, a_{22} and a_{21} are elements of the corregionalization matrix and ϕ_1 and ϕ_2 are parameters of the exponential correlation model. Table 3 presents the parameter estimates, the maximized log-likelihood (LL), the Akaike information criterion (AIC) and Bayesian information criterion (BIC) values. The standard errors for all models (in parentheses) were calculated using the Hessian matrix approximation. The MatConstr, MatInd, LMC and MatSep models are estimated using the `GeoFit` function from the `GeoModels` package (Bevilacqua and Morales-Oñate 2018) for which the standard likelihood function and the Euclidean distance were considered. The Nelder-Mead optimizer is the algorithm of choice for all model fits. Table 4 presents the Elps.Time, N.Iter and the Time.by.Iter (in seconds), considering each model. The R codes are available in the supplementary material.

The predictive behavior was assessed by a random training selection of 200 locations (80% of the data), from which we estimate the models under study and compute the mean absolute error (MAE), the root mean square error (RMSE) and the normalized mean square error (NMSE) for each model using cokriging predictor for the 50 remaining locations (20% of the data). These measures are defined as,

Table 4 Elapsed estimation time (Elps.Time), number of iterations required (N.Iter) and average time per iteration (Time.by.Iter) for each model for soil250 data

Estimates	Models				
	MatSimpler	MatConstr	MatInd	LMC	MatSep
Elps.Time	39.387	420.807	176.267	93.72	110.315
N.Iter	822	4842	2013	2685	1264
Time.by.Iter	0.0479	0.0869	0.0875	0.0349	0.0872
N. Param	7	7	6	6	5

Table 3 Parameter estimates of each model for soil250 data

Estimates	Models				
	MatSimpler	MatConstr	MatInd	LMC	MatSep
$\hat{a}_{11}$	–	–	–	0.615 (0.072)	–
$\hat{a}_{12}$	–	–	–	− 0.067 (0.236)	–
$\hat{a}_{22}$	–	–	–	0.544 (0.263)	–
$\hat{a}_{21}$	–	–	–	0.608 (0.119)	–
$\hat{\phi}_1$	1.077 (0.315)	1.415 (0.460)	1.849 (1.339)	1.156 (0.308)	1.978 (0.818)
$\hat{\phi}_2$	1.994 (0.689)	1.483 (0.472)	2.340 (1.877)	2.816 (1.287)	
$\hat{\nu}_1$	0.543 (0.095)	0.509 (0.082)	0.398 (0.102)	–	0.495 (0.070)
$\hat{\nu}_2$	0.513 (0.070)	0.562 (0.087)	0.422 (0.102)	–	
$\hat{\sigma_1}$	0.627 (0.054)	0.671 (0.006)	0.648 (0.015)	–	0.774 (0.016)
$\hat{\sigma_2}$	0.920 (0.115)	0.850 (0.014)	0.848 (0.051)	–	0.891 (0.024)
$\hat{\rho}_{12}$	0.823 (0.021)	0.811 (0.022)	–	–	0.816 (0.021)
LL	− 161.551	− 166.143	− 301.813	− 164.688	− 166.298
AIC	337.103	346.285	615.626	341.376	342.595
BIC	361.753	375.787	640.914	366.663	363.668

Table 5 Mean and standard deviation (sd) for prediction errors considering 150 splits of data into training (80%) and test (20%) for each model and each variable for soil250 data

Models		MAE_H	$RMSE_H$	$NMSE_H$	MAE_{CTC}	$RMSE_{CTC}$	$NMSE_{CTC}$
MatSimpler	Mean	0.277	0.367	0.215	0.314	0.416	0.151
	SD	0.032	0.045	0.032	0.032	0.049	0.023
MatConstr	Mean	0.271	0.364	0.204	0.307	0.410	0.150
	SD	0.032	0.046	0.030	0.034	0.052	0.024
MatInd	Mean	0.271	0.363	0.210	0.308	0.411	0.156
	SD	0.032	0.045	0.029	0.034	0.051	0.024
LMC	Mean	0.271	0.363	0.208	0.307	0.409	0.153
	SD	0.032	0.046	0.030	0.034	0.051	0.025
MatSep	Mean	0.270	0.363	0.202	0.307	0.409	0.152
	SD	0.033	0.047	0.030	0.034	0.052	0.024

$$\text{MAE}_i = \frac{1}{50}\sum_{k=1}^{50} |Y_i(\mathbf{s}_k) - \hat{Y}_i(\mathbf{s}_k)|,$$

$$\text{RMSE}_i = \sqrt{\frac{1}{50}\sum_{k=1}^{50} (Y_i(\mathbf{s}_k) - \hat{Y}_i(\mathbf{s}_k))^2},$$

$$\text{NMSE}_i = \frac{\text{RMSE}_i}{\max(\hat{Y}_i(\mathbf{s}_k)) - \min(\hat{Y}_i(\mathbf{s}_k))},$$

where $\hat{Y}_i(\mathbf{s}_k)$ is the cokriging predictor of the variable $Y_i(\mathbf{s}_k)$, with $i = \text{H}, \text{CTC}$, representing hydrogen content and CTC, respectively. We repeated the same process 150 times, calculating the values of the MAE_i, RMSE_i and NMSE_i, for each variable each time. Table 5 presents the summary of this measures. In general, all models presented equivalent results in terms of predictive capacity.

The MatSimpler model showed a better fit when compared to the other models in terms of log-likelihood, AIC and BIC values. The Elps.Time and the N.Iter for the MatSimpler model were much lower than the other models. Thus, we observe that the proposed MatSimpler model is a competitive model with the classical models in the literature.

4.2 Example 2: North Africa's minimum and maximum temperature

In this section we illustrate the application of the proposed model to a large dataset. We considered data of minimum and maximum average temperatures for the years 1970-2000. In particular, a region of North Africa in the period of September observed at 3061 locations. Data were obtained from WorldClim (www.worldclim.org), a high spatial resolution climate database (Fick and Hijmans 2017).

Figure 4 shows the spatial locations of the temperatures along the region under study. We detrend the data to remove patterns along the coordinates and consider the resulting residuals as a realization of a bivariate zero-mean Gaussian random field. We transform the coordinate system to UTM using the `spTransform` function from the `sp` package (Pebesma and Bivand 2005) and divide the coordinates by a constant to facilitate the estimation process. We consider the Euclidean distance. The empirical variograms of the residuals are presented in Fig. 5.

In the estimation step, aiming to demonstrate the computational efficiency of our proposal for large data sets, we consider the RGW-Simpler model and the MatConstr classical model. To make the estimation process more agile we set $\nu_1 = \nu_2$ at 0.5, for the Mátern model, and at 0, for the RGW model. Regarding the RGW model we also set $\mu = 1.5$ in order to obtain sparse matrices. The parameters vector to be estimated is $\theta = (\phi_1, \phi_2, \sigma_1, \sigma_2, \rho_{12})^\top$, where the indexes 1 and 2 represent the minimum and maximum temperature parameters, respectively. The N.Iter, Time.by.Iter (in seconds), log-likelihood value and parameter estimates are presented in Table 6.

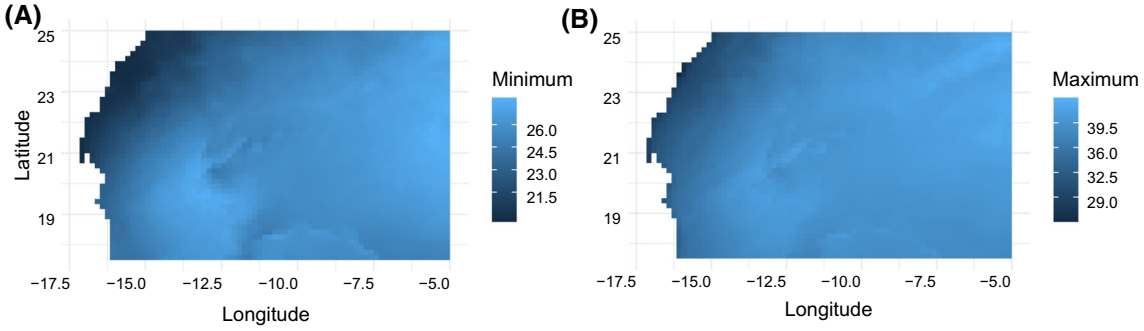

Fig. 4 Spatial locations for **a** minimum and **b** maximum temperatures, considering a region of North Africa in the period of September

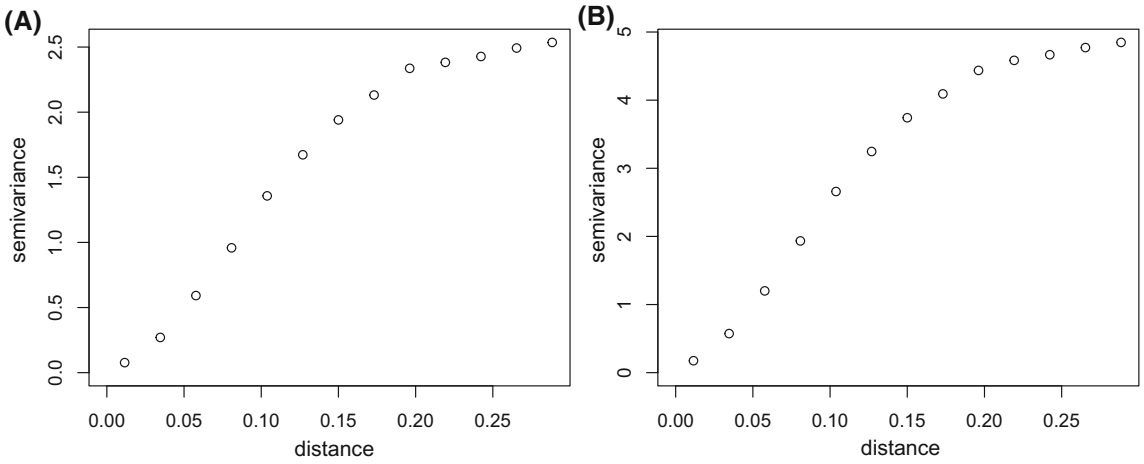

Fig. 5 Empirical semivariogram of residuals for **(a)** minimum and **(b)** maximum temperature after removing tendencies, considering a region of North Africa in the period of September

Table 6 Number of iterations required (N.Iter), average time per iteration (Time.by.Iter), log-likelihood value (LL) and parameters estimates for each model, considering a region of North Africa in the period of September

Models	Estimates							
	N.Iter	Time.by.Iter	LL	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\rho}_{12}$
RGW-Simpler	906	24.409	3799.687	2.538	9.119	2.264	6.609	0.584
				(2.915)	(10.652)	(1.300)	(3.859)	(0.012)
MatConstr	2244	44.303	3799.950	2.895	19.593	2.418	9.688	0.872
				–	–	–	–	–

For the MatConstr model, the convergence was not successful, so it was not able to estimate the standard errors of the estimators. In contrast, for the RGW model, the convergence was successful and the standard errors are shown in parentheses in Table 6. The number of iterations required for estimation and the average time per iteration were also notably lower considering our covariance specification. The RGW-Simpler model proved to be more advantageous in terms of estimation time and computational complexity, when compared to the MatConstr model.

5 Simulation study

This section presents a simulation study to evaluate the behavior of the maximum likelihood estimators. We consider the MatSimpler model and explored a bivariate random field in three different scenarios to illustrate different situations that could occur in practice, exemplifying the flexibility of the proposed model in each case. Table 7 summarizes the simulated scenarios.

Scenario 1 considers the variables have smaller variability, for which we consider relatively small values for the variance parameters: $\sigma_1^2 = 0.25$ and $\sigma_2^2 = 1.0$. Also in this scenario, we consider smaller smoothness for the variables, with values equal to $\nu_1 = 0.3$ and $\nu_2 = 0.4$. When $\nu = 0.5$ the Matérn correlation function reduces to the exponential correlation function. Thus, this scenario represents lesser smoothness marginal behavior for the variables when compared to the exponential correlation model.

Scenario 2 keeps $\nu_1 = 0.3$ and $\nu_2 = 0.4$, and considers a greater variability with variances values equal to $\sigma_1^2 = 2.25$ and $\sigma_2^2 = 4.00$.

Scenario 3 considers a situation of smaller variability, with variance values fixed at $\sigma_1^2 = 0.25$ and $\sigma_2^2 = 1.0$, and considers higher smoothness values for variables: $\nu_1 = 0.7$ and $\nu_2 = 1.0$. In this scenario, we have marginal behaviors that are smoother in comparison with the exponential correlation model.

In all scenarios, we set the scale parameters values in $\phi_1 = 0.05$ and $\phi_2 = 0.1$. For the correlation parameter between the variables the values considered are $\rho_{12} = (-0.7, -0.4, 0.0, 0.4, 0.7)$, aiming to illustrate different correlation structures that could occur in practice between the variables, with values ranging from a strong negative correlation ($\rho_{12} = -0.7$) to a strong positive correlation ($\rho_{12} = 0.7$) and including the no correlation case where $\rho_{12} = 0$.

For each scenario, 500 samples of a bivariate stationary isotropic Gaussian random field, of sizes 100, 225, 400 and 625, were simulated in a regular unit grid. Fig. 6 illustrates the results of the simulated scenarios showing the expected

Table 7 Parameter values for each simulated scenario

Scenarios		Parameters						
	Situation	ϕ_1	ϕ_2	ν_1	ν_2	σ_1	σ_2	ρ_{12}
1	Less smoothness and less variability	0.05	0.1	0.3	0.4	0.5	1.0	− 0.7 −0.4
2	Less smoothness and greater variability	0.05	0.1	0.3	0.4	1.5	2.0	0.0 0.4
3	Greater smoothness and less variability	0.05	0.1	0.7	1.0	0.5	1.0	0.7

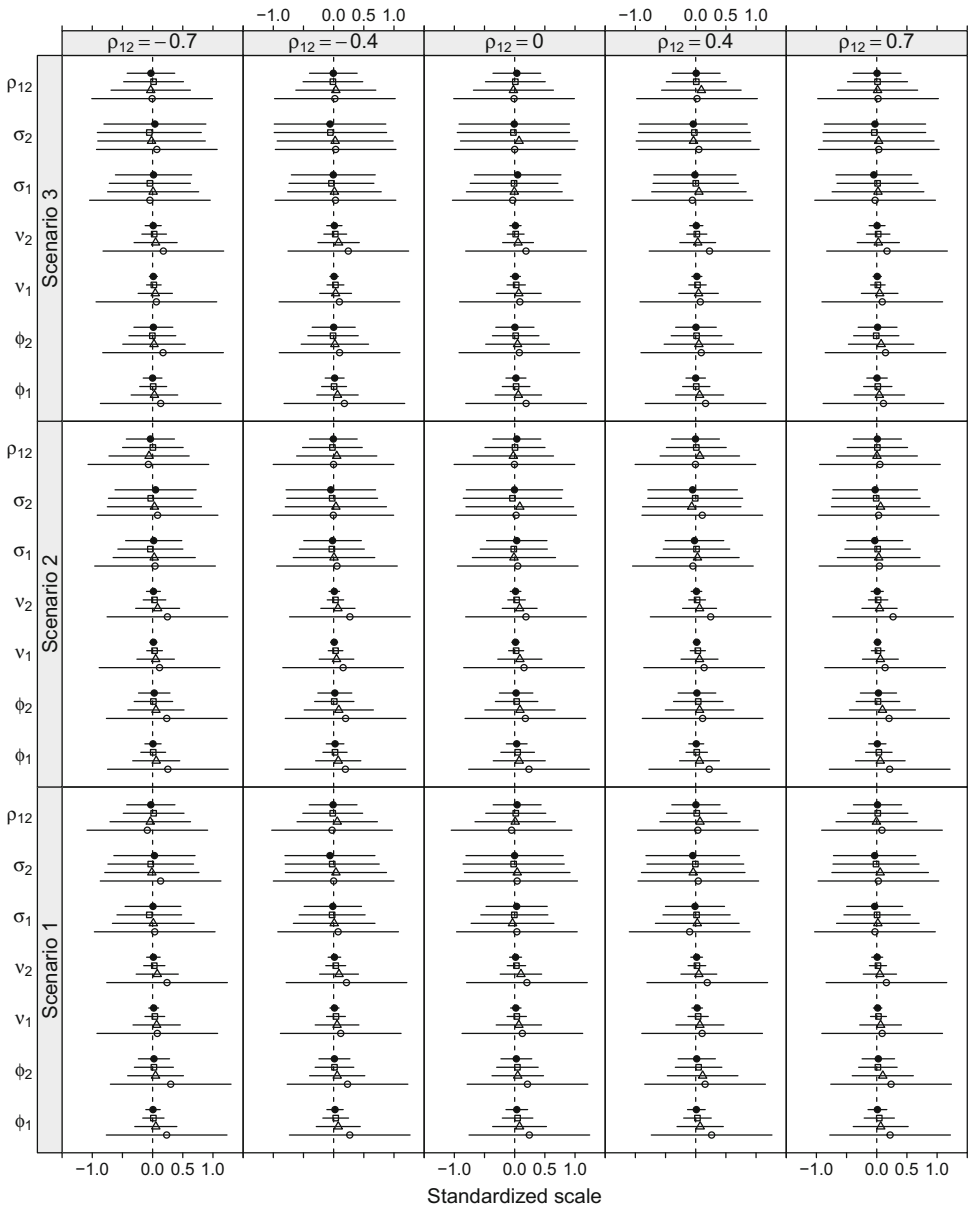

Fig. 6 Expected bias and confidence interval on a standardized scale for each scenario and sample size *Open circle*, 100; Triangle, 225; Square, 400; Black circle, 625; for the parameters of the MatSimpler model

bias plus and minus the expected standard error for estimators of the model for each scenario.

To facilitate the visualization we follow Bonat and Jørgensen (2016) and Petterle et al. (2019), considering, for each parameter, standardized scales with respect to the standard error of the sample size 100, that is, for each parameter, the expected bias and the limits of the confidence intervals are divided by the standard error obtained on the sample of size 100. Standard errors and biases gets closer to zero as the sample size increases for the

considered scenarios. In all scenarios, there appears to be a small overestimate for the smoothness and scale parameters, especially for smaller samples.

6 Discussion

We presented a covariance specification for multivariate Gaussian random fields given by the product of matrices for continuously indexed data. The model is simple whilst flexible, allowing for different correlation structures and marginal-covariance functions. Its structure facilitates the specification, estimation, computation and generalization for more than two variables.

As it is a flexible structure allowing different marginal-covariance specifications, we considered the Matérn correlation function, for being flexible, widely discussed in the geostatistical literature and for having a closer relationship with the traditionally discussed models. We also consider the reparameterized Generalized Wendland model to illustrate the flexibility of our specification in allowing compactly supported covariance structures that bring computational advantages for large datasets.

We illustrate the computational time growth for the MatSimpler model for up to six response variables and 900 sample locations. In this scenario, the estimate time was approximately two hours (Fig. 7). Precise times will be hardware dependent. However, comparing it with some literature models for the bivariate case (Fig. 8), our proposal was competitive, presenting a lower average time per iteration and number of iterations required, specially when compared to the MatConstr model, which has the same number of parameters.

The analysis of two data-sets illustrate the ability to deal with different covariance structures. The use of compactly supported covariance functions made it possible to deal with large data sets due to the efficient calculation of Cholesky factors for sparse matrices, showing that the proposed model has a good balance between flexibility and computational complexity, with reduced computational estimation times when compared to other competing classical approaches.

Future directions include the organization and construction of an R package that involves the proposed approach. Furthermore, the presented proposal opens options for future research in the context of non-Gaussian modeling and asymmetric data for multivariate spatial problems.

Appendix: Derivatives of the covariance matrix

Based on matrix properties (Wand 2002; Bonat et al. 2020), the score functions with respect to the β and λ parameters, are given, respectively, by:

$$\begin{aligned}\mathcal{U}_\beta &= \mathbf{D}^\top \boldsymbol{\Sigma}(\lambda)^{-1}\mathbf{r}(\beta),\\ \mathcal{U}_\lambda &= -\frac{1}{2}\left\{\boldsymbol{\Sigma}(\lambda)^{-1} - \boldsymbol{\Sigma}(\lambda)^{-1}\mathbf{r}(\beta)\mathbf{r}(\beta)^\top\boldsymbol{\Sigma}(\lambda)^{-1}\right\}\frac{\partial\boldsymbol{\Sigma}(\lambda)}{\partial\lambda},\end{aligned} \tag{11}$$

with $\mathbf{r}(\beta) = (\mathbf{y} - \mu(\beta))$, $\mathbf{D} = \frac{\partial\mu(\beta)}{\partial\beta} = \text{Bdiag}(\mathbf{X}_1, \ldots, \mathbf{X}_p)$ and the inverse calculation was described earlier.

We achieve the maximum likelihood estimator of β by solving $\mathcal{U}_\beta$, which results in:

$$\hat{\beta} = \left(\mathbf{D}^\top\boldsymbol{\Sigma}(\lambda)^{-1}\mathbf{D}\right)^{-1}\left(\mathbf{D}^\top\boldsymbol{\Sigma}(\lambda)^{-1}\mathbf{y}\right).$$

Making similar calculations, we find the Fisher information matrix which, for β, is given by:

$$\mathcal{F}_\beta = \mathbf{D}^\top\boldsymbol{\Sigma}(\lambda)^{-1}\mathbf{D}.$$

For λ, the $(i,j)^{th}$ entry of the Fisher information matrix, is given by:

$$[\mathcal{F}_\lambda]_{ij} = \frac{1}{2}\text{tr}\left[\mathbf{W}_{\lambda_i}\boldsymbol{\Sigma}(\lambda)\mathbf{W}_{\lambda_j}\boldsymbol{\Sigma}(\lambda)\right],$$

where $\mathbf{W}_{\lambda_i} = -\partial\boldsymbol{\Sigma}(\lambda)^{-1}/\partial\lambda_i$.

Considering $\hat{\theta} = (\hat{\beta}^\top, \hat{\lambda}^\top)^\top$ the maximum likelihood estimator of θ parameter, the asymptotic distribution of $\hat{\theta}$ is $\hat{\theta} \sim N(\theta, \mathcal{F}_\theta^{-1})$, where $\mathcal{F}_\theta = \begin{pmatrix}\mathcal{F}_\beta & \mathbf{0}\\ \mathbf{0} & \mathcal{F}_\lambda\end{pmatrix}$ denotes the Fisher information matrix of θ. This result is compatible with the increasing domain regime (Cressie 1993).

The maximum likelihood estimates of λ can be found through Newton's scoring iterative algorithm (Bonat et al. 2020):

$$\lambda^{i+1} = \lambda^i - \alpha\mathcal{F}_\lambda^{-1}\mathcal{U}_\lambda(\tilde{\theta}; \mathbf{y}),$$

where $\tilde{\theta} = (\hat{\beta}^\top, \lambda^\top)^\top$ and α controls the step length.

Now, let ρ_r, for $r = 1, \ldots, p(p-1)/2$, denoting the correlation parameters of $\boldsymbol{\Sigma}_b$, σ_i^2, ϕ_i and ν_i, denoting the variance, scale and smoothness parameters of the marginal-covariance matrix, Σ_{ii}, for $i = 1, \ldots, p$.

The partial derivative of the matrix-valued covariance function $\boldsymbol{\Sigma}$, with respect to each correlation parameter, ρ_r, is given by:

$$\frac{\partial \boldsymbol{\Sigma}}{\partial \rho_r} = \text{Bdiag}\left(\tilde{\boldsymbol{\Sigma}}_{11}, \tilde{\boldsymbol{\Sigma}}_{22}, \ldots, \tilde{\boldsymbol{\Sigma}}_{pp}\right)\left(\frac{\partial \boldsymbol{\Sigma}_b}{\partial \rho_r} \otimes \mathbf{I}\right) \text{Bdiag}\left(\tilde{\boldsymbol{\Sigma}}_{11}^{\top}, \tilde{\boldsymbol{\Sigma}}_{22}^{\top}, \ldots, \tilde{\boldsymbol{\Sigma}}_{pp}^{\top}\right).$$

To obtain the partial derivative with respect to variance parameter, σ_i^2, we will use matrix properties, that is,

$$\begin{aligned}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_i^2} = &\text{Bdiag}\left(\mathbf{0}, \ldots, \frac{\partial \tilde{\boldsymbol{\Sigma}}_{ii}}{\partial \sigma_i^2}, \ldots, \mathbf{0}\right)(\boldsymbol{\Sigma}_b \otimes \mathbf{I}) \\ &\text{Bdiag}\left(\tilde{\boldsymbol{\Sigma}}_{11}^{\top}, \tilde{\boldsymbol{\Sigma}}_{22}^{\top}, \ldots, \tilde{\boldsymbol{\Sigma}}_{pp}^{\top}\right) \\ &+ \text{Bdiag}\left(\tilde{\boldsymbol{\Sigma}}_{11}, \tilde{\boldsymbol{\Sigma}}_{22}, \ldots, \tilde{\boldsymbol{\Sigma}}_{pp}\right)(\boldsymbol{\Sigma}_b \otimes \mathbf{I}) \\ &\text{Bdiag}\left(\mathbf{0}, \ldots, \frac{\partial \tilde{\boldsymbol{\Sigma}}_{ii}^{\top}}{\partial \sigma_i^2}, \ldots, \mathbf{0}\right).\end{aligned} \tag{12}$$

An analogous procedure to the Eq. (12) can be used to obtain the derivatives with respect to ϕ_i and v_i. Thus, to obtain the derivatives of $\boldsymbol{\Sigma}$ with respect to each parameter, we must calculate the parcial derivatives in (12). Using the result of partial derivatives of Cholesky's factorization (Särkkä 2013; Bonat and Jørgensen 2016), follows:

$$\frac{\partial \tilde{\boldsymbol{\Sigma}}_{ii}}{\partial \sigma_i^2} = \tilde{\boldsymbol{\Sigma}}_{ii}\Phi\left(\tilde{\boldsymbol{\Sigma}}_{ii}^{-1}\frac{\partial \boldsymbol{\Sigma}_{ii}}{\partial \sigma_i^2}\tilde{\boldsymbol{\Sigma}}_{ii}^{-1}\right),$$

$$\frac{\partial \tilde{\boldsymbol{\Sigma}}_{ii}}{\partial \phi_i} = \tilde{\boldsymbol{\Sigma}}_{ii}\Phi\left(\tilde{\boldsymbol{\Sigma}}_{ii}^{-1}\frac{\partial \boldsymbol{\Sigma}_{ii}}{\partial \phi_i}\tilde{\boldsymbol{\Sigma}}_{ii}^{-1}\right),$$

$$\frac{\partial \tilde{\boldsymbol{\Sigma}}_{ii}}{\partial v_i} = \tilde{\boldsymbol{\Sigma}}_{ii}\Phi\left(\tilde{\boldsymbol{\Sigma}}_{ii}^{-1}\frac{\partial \boldsymbol{\Sigma}_{ii}}{\partial v_i}\tilde{\boldsymbol{\Sigma}}_{ii}^{-1}\right),$$

where $\Phi(.)$ is the strictly lower triangular part of the argument and half of its diagonal.

Appendix: Computational results

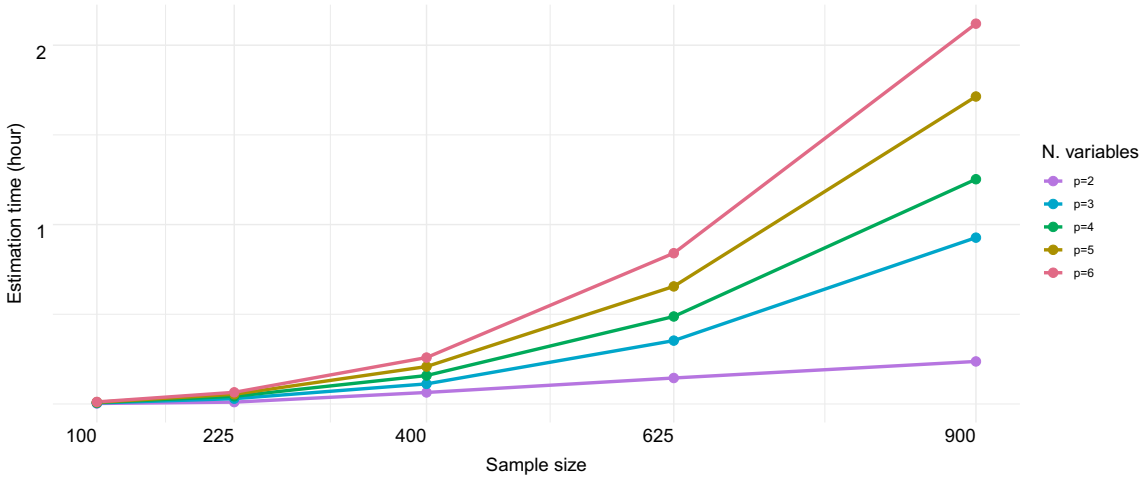

Fig. 7 Estimation time for MatSimpler model considering different sample sizes and number of variables

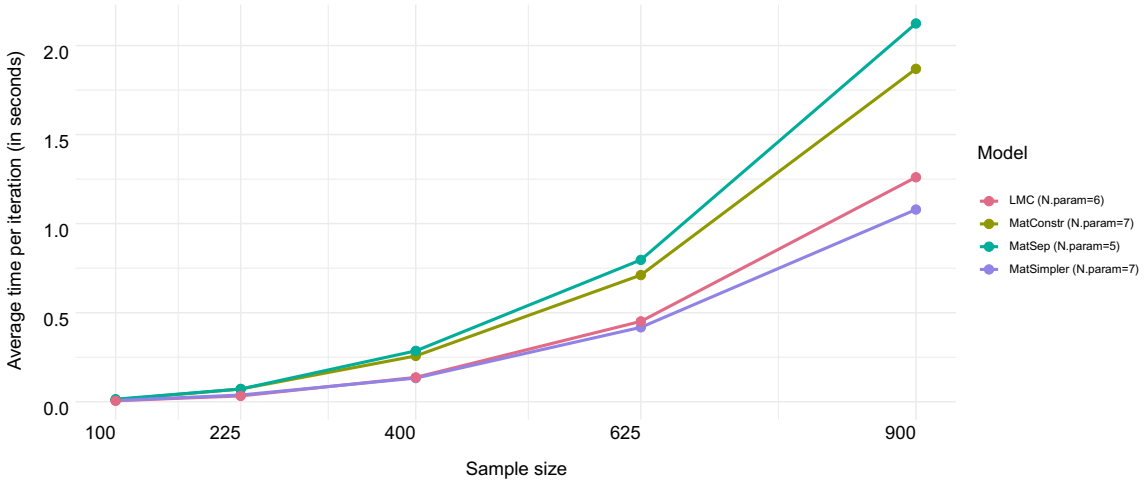

Fig. 8 Average time per iteration (Time.by.Iter) for each model, considering simulated data from the MatConstr model for different sample sizes

Author Contributions All authors contributed equally to the design and preparation of the material, with comments and suggestions for improvements. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Alegría A, Porcu E, Furrer R (2018) Asymmetric matrix-valued covariances for multivariate random fields on spheres. J Stat Comput Simul 88(10):1850–1862. https://doi.org/10.1080/00949655.2017.1406488

Alegría A, Porcu E, Furrer R et al (2019) Covariance functions for multivariate gaussian fields evolving temporally over planet earth. Stoch Env Res Risk Assess 33(8):1593–1608. https://doi.org/10.1007/s00477-019-01707-w

Banerjee S, Roy A (2014) Linear algebra and matrix analysis for statistics. CRC Press, Boca Raton. https://doi.org/10.1201/b17040

Banerjee S, Carlin BP, Gelfand AE (2003) Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, Boca Raton. https://doi.org/10.1201/9780203487808

Bates D, Maechler M (2021) Matrix: Sparse and Dense Matrix Classes and Methods. https://CRAN.R-project.org/package=Matrix, R package version 1.3-4

Bevilacqua M, Morales-Oñate V (2018) GeoModels: A Package for Geostatistical Gaussian and non Gaussian Data Analysis. https://vmoprojs.github.io/GeoModels-page/, R package version 1.0.3-4

Bevilacqua M, Vallejos R, Velandia D (2015) Assessing the significance of the correlation between the components of a bivariate gaussian random field. Environmetrics 26(8):545–556. https://doi.org/10.1002/env.2367

Bevilacqua M, Alegría A, Velandia D et al (2016) Composite likelihood inference for multivariate gaussian random fields. J Agric Biol Environ Stat 21(3):448–469. https://doi.org/10.1007/s13253-016-0256-3

Bevilacqua M, Fassò A, Gaetan C et al (2016) Covariance tapering for multivariate gaussian random fields estimation. Stat Methods & Appl 25(1):21–37. https://doi.org/10.1007/s10260-015-0338-3

Bevilacqua M, Faouzi T, Furrer R et al (2019) Estimation and prediction using generalized wendland covariance functions under fixed domain asymptotics. Ann Stat 47(2):828–856

Bevilacqua M, Diggle P, Porcu E (2020) Families of covariance functions for bivariate random fields on spheres. Spatial Stat 40(100448):1–29. https://doi.org/10.1016/j.spasta.2020.100448

Bevilacqua M, Caamaño-Carrillo C, Porcu E (2022) Unifying compactly supported and matern covariance functions in spatial statistics. J Multivariate Anal 104949

Bivand RS, Pebesma EJ, Gómez-Rubio V et al (2008) Applied spatial data analysis with R. Springer, New York. https://doi.org/10.1007/978-0-387-78171-6

Bonat WH, Jørgensen B (2016) Multivariate covariance generalized linear models. J Roy Stat Soc: Ser C (Appl Stat) 65(5):649–675. https://doi.org/10.1111/rssc.12145

Bonat WH, Petterle RR, Balbinot P et al (2020) Modelling multiple outcomes in repeated measures studies: comparing aesthetic eyelid surgery techniques. Stat Model 21:564–582. https://doi.org/10.1177/1471082X20943312

Bourgault G, Marcotte D (1991) Multivariable variogram and its application to the linear model of coregionalization. Math Geol 23(7):899–928. https://doi.org/10.1007/BF02066732

Cecconi L, Grisotto L, Catelan D et al (2016) Preferential sampling and bayesian geostatistics: Statistical modeling and examples. Stat Methods Med Res 25(4):1224–1243. https://doi.org/10.1177/0962280216660409

Chilès JP, Delfiner P (2012) Geostatistics: Modeling Spatial Uncertainty. Wiley, New York. https://doi.org/10.1007/s11004-012-9429-y

Cressie N (1993) Statistics for spatial data. Wiley, New York. https://doi.org/10.1002/9781119115151

Diggle P, Ribeiro PJ Jr (2007) Model-based Geostatistics. Springer, New York. https://doi.org/10.1007/978-0-387-48536-2

Emery X, Porcu E (2019) Simulating isotropic vector-valued gaussian random fields on the sphere through finite harmonics approximations. Stoch Env Res Risk Assess 33(8):1659–1667. https://doi.org/10.1007/s00477-019-01717-8

Emery X, Porcu E, Bissiri PG (2019) A semiparametric class of axially symmetric random fields on the sphere. Stoch Env Res Risk Assess 33(10):1863–1874. https://doi.org/10.1007/s00477-019-01725-8

Fasiolo M (2016) An introduction to mvnfast. https://CRAN.R-project.org/package=mvnfast, R package version 0.1.6

Fick SE, Hijmans RJ (2017) Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol 37(12):4302–4315

Finley AO, Banerjee S, Gelfand AE (2015) spBayes for large univariate and multivariate point-referenced spatio-temporal data models. J Stat Softw, Art 63(13):1–28. https://doi.org/10.18637/jss.v063.i13

Furrer R, Sain SR (2010) spam: a sparse matrix r package with emphasis on mcmc methods for gaussian markov random fields. J Stat Softw 36(10):1–25

Gelfand AE, Schmidt AM, Banerjee S et al (2004) Nonstationary multivariate process modeling through spatially varying coregionalization. TEST 13(2):263–312. https://doi.org/10.1007/BF02595775

Gentle JE (2017) Matrix algebra: theory, computations, and applications in statistics. Springer, New York. https://doi.org/10.1007/978-3-319-64867-5

Genton MG, Kleiber W (2015) Cross-covariance functions for multivariate geostatistics. Stat Sci 30(2):147–163. https://doi.org/10.1214/14-STS487

Gneiting T (1999) Correlation functions for atmospheric data analysis. Q J R Meteorol Soc 125(559):2449–2464. https://doi.org/10.1002/qj.49712555906

Gneiting T, Kleiber W, Schlather M (2010) Matérn cross-covariance functions for multivariate random fields. J Am Stat Assoc 105(491):1167–1177. https://doi.org/10.1198/jasa.2010.tm09420

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

Goulard M, Voltz M (1992) Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. Math Geol 24(3):269–286. https://doi.org/10.1007/BF00893750

Guttorp P, Gneiting T (2006) Studies in the history of probability and statistics XLIX on the matérn correlation family. Biometrika 93(4):989–995

Hardy Y, Steeb WH (2019) Matrix Calculus, Kronecker Product and Tensor Product: A Practical Approach to Linear Algebra, Multilinear Algebra and Tensor Calculus with Software Implementations. World Scientific, Singapore, https://doi.org/10.1142/11338

MacNab YC (2016) Linear models of coregionalization for multivariate lattice data: Order-dependent and order-free cMCARs. Stat Methods Med Res 25(4):1118–1144. https://doi.org/10.1177/0962280216660419

MacNab YC (2018) Some recent work on multivariate gaussian markov random fields. TEST 27(3):497–541. https://doi.org/10.1007/s11749-018-0605-3

Martinez-Beneito MA (2013) A general modelling framework for multivariate disease mapping. Biometrika 100(3):539–553. https://doi.org/10.1093/biomet/ast023

Martinez-Beneito MA (2020) Some links between conditional and coregionalized multivariate gaussian markov random fields. Spatial Stat 40(100383):1–17. https://doi.org/10.1016/j.spasta.2019.100383

Martins ABT, Bonat WH, Ribeiro PJ Jr (2016) Likelihood analysis for a class of spatial geostatistical compositional models. Spatial Stat 17:121–130. https://doi.org/10.1016/j.spasta.2016.06.008

Matérn B (1986) Spatial variation. Springer, Berlin. https://doi.org/10.1002/bimj.4710300514

Pebesma E, Bivand RS (2005) Classes and methods for spatial data: the sp package. R News 5(2):9–13

Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. Comp & Geosci 30:683–691. https://doi.org/10.1016/j.cageo.2004.03.012

Petersen KB, Pedersen MS, et al. (2008) The matrix cookbook. Technical University of Denmark

Petterle RR, Bonat WH, Scarpin CT (2019) Quasi-beta longitudinal regression model applied to water quality index data. J Agric Biol Environ Stat 24(2):346–368. https://doi.org/10.1007/s13253-019-00360-8

Porcu E, Daley DJ, Buhmann M et al (2013) Radial basis functions with compact support for multivariate geostatistics. Stoch Env Res Risk Assess 27(4):909–922. https://doi.org/10.1007/s00477-012-0656-z

Qadir GA, Euán C, Sun Y (2021) Flexible modeling of variable asymmetries in cross-covariance functions for multivariate random fields. J Agric Biol Environ Stat 26(1):1–22. https://doi.org/10.1007/s13253-020-00414-2

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/

Ribeiro AMT, Ribeiro PJ Jr, Bonat WH (2021) Comparison of exponential covariance functions for bivariate geostatistical data. Revista Brasileira de Biometria 39(1):89–102. https://doi.org/10.28951/rbb.v39i1.558

Ribeiro Jr PJ, Diggle PJ, Schlather M, et al. (2020) geoR: Analysis of Geostatistical Data. https://CRAN.R-project.org/package=geoR, R package version 1.8-1

Rue H, Held L (2005) Gaussian Markov random fields: theory and applications. CRC Press, Boca Raton. https://doi.org/10.1201/9780203492024

Salvaña MLO, Genton MG (2020) Nonstationary cross-covariance functions for multivariate spatio-temporal random fields. Spatial Stat 37(100411):1–24. https://doi.org/10.1016/j.spasta.2020.100411

Särkkä S (2013) Bayesian filtering and smoothing. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139344203

Schmidt AM, Gelfand AE (2003) A bayesian coregionalization approach for multivariate pollutant data. J Geophys Res 108(D24):1–9. https://doi.org/10.1029/2002JD002905

Teichmann J, Menzel P, Heinig T et al (2021) Modeling and fitting of three-dimensional mineral microstructures by multinary random fields. Math Geosci 53(5):877–904. https://doi.org/10.1007/s11004-020-09871-4

Vallejos R, Osorio F, Bevilacqua M (2020) Spatial relationships between two georeferenced variables: With applications in R. Springer, New York. https://doi.org/10.1007/978-3-030-56681-4

Ver Hoef JM, Cressie N (1993) Multivariable spatial prediction. Math Geol 25(2):219–240. https://doi.org/10.1007/BF00893273

Wackernagel H (2003) Multivariate geostatistics: an introduction with applications. Springer, Berlin. https://doi.org/10.1007/978-3-662-05294-5

Wand M (2002) Vector differential calculus in statistics. Am Stat 56(1):55–62. https://doi.org/10.1198/000313002753631376

Zastavnyi VP, Trigub RM (2002) Positive-definite splines of special form. Sbornik: Math 193(12):1771