Multivariate Nearest-Neighbors Gaussian Processes with Random Covariance Matrices

Isabelle Grenier[a]*, Bruno Sansó[a] and Jessica L. Matthews[b†a] Department of Statistics, University of (

September 22, 2022

Abstract

We propose a non-stationary spatial model based on a normal-inverse-Wishart framework, conditioning on a set of nearest-neighbors. The model, called Nearest-Neighbor Gaussian Process with Random Covariance matrices (NN-RCM) is developed for both univariate and multivariate spatial settings and allows for fully flexible covariance structures that impose no stationarity or isotropic restrictions. In addition, the model can handle duplicate observations and missing data. We consider an approach based on integrating out the spatial random effects that allows fast inference for the model parameters. We also consider a full hierarchical approach that leverages the sparse structures induced by the model to perform fast Monte Carlo computations. Strong computational efficiency is achieved by leveraging the adaptive localized structure of the model that allows for a high level of parallelization. We illustrate the performance of the model with univariate and bivariate simulations, as well as with observations from two stationary satellites consisting of albedo measurements.

1 Introduction

The land surface albedo is the ratio between the upward and downward reflected solar radiation at the Earth's surface (NOAA, 2018). Quantifying the amount of light that hits the surface of the Earth without being reflected is essential for understanding climate change and its potential impact on human health. Missions such as the National Oceanic and Atmospheric Administration's (NOAA) Geostationary Operational Environmental Satellites (GOES) East and West have been designed to target specifically the forecasting of extreme weather events (e.g floods) and the monitoring of land process analysis (NOAA, 2016). In light of being a sensitive indicator of environmental changes, surface albedo has been classified an Essential Climate Variable (ECV) by the Global Climate Observing System (GCOS).

Surface albedo can be measured by a variety of instruments including ground-based stations and satellites. Of particular importance are the observations gathered from agency participants in an international network called SCOPE-CM (Sustained and COordinated Processing of Environmental satellite data for Climate Monitoring). The near-global coverage of participating geostationary missions is illustrated in Figure 1. This effort has so far produced a land surface albedo dataset that contains more than 12.5 million files at a total size of more than 270 TB. In addition to albedo,

---

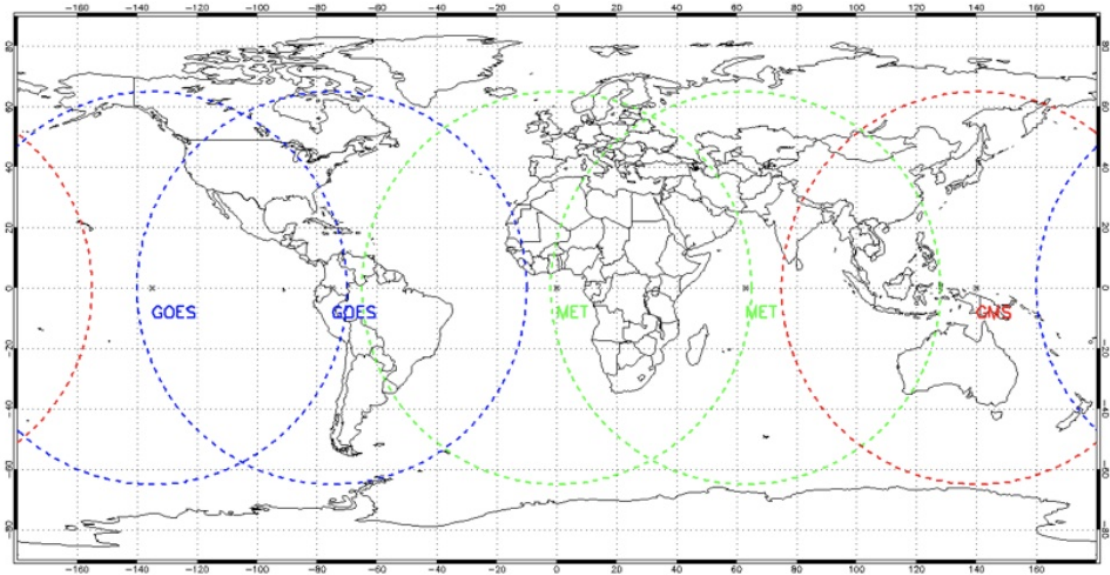*CONTACT Isabelle Grenier. Email: igrenier@ucsc.edu

Figure 1: Illustration of the °70 viewing angle limit of the five world satellites. Blue: GOES from NOAA. Green: Meteosat satellites from EUMETSAT. Red: Geostationary Meteorological Satellite (GMS) from JMA. Figure adapted from Govaerts et al. (2008).

other essential climate variables currently observed from the geostationary platform include: aerosol properties, cloud characteristics, wind speed and direction, radiation budget, wildfires, temperature profiles, precipitation, and snow cover. The analysis of these observations presents a number of statistical challenges, including the need to handle very large spatial domains with massive numbers of observations corresponding to heterogeneous spatial fields and different sources of information. Those challenges motivate the development of the methods presented in this paper. Here we propose a model-based approach for multivariate non-stationary spatial surfaces that scales to large amounts of data.

The last couple of decades have seen an explosion in the development of geostatistical methods producing a solid body of literature and software, see, for example, the books by Cressie (1993); Gelfand et al. (2010); Cressie and Wikle (2011); Banerjee et al. (2014). Modern geostatistical approaches provide flexible probabilistic models, coupled with learning methods, that are used to investigate inferential questions related to geographically-referenced data. Traditionally, model-based spatial models have relied on the Gaussian process (GP). GPs capture the dependence due to proximity through a covariance function. For a likelihood-based approach to GPs, the bottleneck lies in the computation of the determinant and the inverse of the covariance matrix induced by the locations of the available observations. To tackle this problem most current methods take one of two approaches: exploit sparsity in the structure of the covariance matrix (e.g. Furrer et al., 2006; Kaufman et al., 2008; Du et al., 2009; Shaby and Ruppert, 2012) or reduce the dimensionality of the problem by seeking representations of GPs on lower dimensional subspaces (e.g. Higdon, 1998; Banerjee et al., 2008; Lemos and Sansó, 2009; Cressie and Johannesson, 2008; Katzfuss and

2

Cressie, 2011). In both cases the goal is to speed up calculations, as well as reduce the size of the objects that need to be handled and stored in memory when performing computations. For further information about the state of the art model-based geostatistics methods suitable for large data sets see Banerjee (2017) and Heaton et al. (2019). Nearest-neighbor GP (NNGP) Datta et al. (2016) are a particularly intriguing class of models, as they blend features of both the dimension reduction and the sparsity approaches, formalizing to a Gaussian process framework the popular likelihood approximation proposed in Vecchia (1988).

In addition to the challenge of dealing with very large datasets, variables such as land surface albedo, observed over very large regions, exhibit a behavior that is incompatible with the common assumption of stationarity of the underlying spatial field. Many of the approaches for large spatial fields result in non-stationary processes, even though they are not built specifically to deal with such a property. One of the first attempts to deal explicitly with non-stationary spatial fields is the deformation approach in Sampson and Guttorp (1992); Schmidt and O'Hagan (2003). In Brown et al. (1994b), non-stationarity occurs from the choice of using an inverse Wishart prior for the covariance matrix within the general multivariate normal framework. Using a different approach, Paciorek and Schervish (2006) create a new general class of non-stationary covariance functions. Partition models like the ones introduced in Gramacy and Lee (2008) and Kim et al. (2005) construct non-stationary models by averaging over locally stationary processes. Further examples include Fuentes (2001) and Fuentes and Smith (2003) where the non-stationary process is a discrete or continuous weighted average of stationary models for subregions of the space. Also, like any reduced rank representation of a Gaussian process, process convolutions result in non-stationary processes. An explicit model to capture the heterogeneous nature of the spatial field is developed in Higdon (1998) and Lemos and Sansó (2009), where the shape of the kernels is modeled as spatially-varying. A different approach is taken in Kirsner and Sansó (2020), that build a multi-resolution model with spatially-varying resolution. Finally, more recently, a nonparametric and non-stationary approach that looks at inferring the sparse elements of the Cholesky factor of an inverse covariance matrix has been presented in Kidd and Katzfuss (2022).

In this paper, we start by considering a multivariate normal framework for the observations. We assume a fully unknown covariance and leverage the inverse Wishart prior to set our focus on developing a non-isotropic covariance structure. Then, we extend the model to a spatial process that is suitable for large spatial datasets by including a nearest-neighbors idea. We denote the resulting multivariate process as Nearest-Neighbors Gaussian Processes with Random Covariance Matrices (NN-RCM). We develop two inferential approaches. The first version leverages the marginal posterior distribution of the model and allows to obtain posterior inference of the covariance parameters efficiently. The second version builds a hierarchical structure which includes fixed effects and an observational error. Computations for both versions are naturally geared to leverage parallelization and multi-threading.

The remainder of the paper is organized as follows: In Section 2 we present the methodological framework of the NN-RCM model. In Section 3, we discuss the different features of the model through a number of simulated examples and comparisons with existing methods. We first consider univariate cases. We then detail the implementation of the bivariate NN-RCM models and discuss the many questions regarding multivariate nearest-neighbor spatial processes that were unanswered by current methods. In particular, we first describe the creation of multivariate neighborhoods, and then touch on the issue of misaligned sources of information. Finally, in Section 4, we apply the developed methodology to the land surface albedo dataset. We focus on the BHRiso, which stands for bihemispherical reflectance, or *white sky albedo*. We look at datasets of various sizes and

corresponding to different areas within the continental United States (CONUS), fitting a bivariate hierarchical NN-RCM model to the two dimensional vectors obtained from the retrievals of the GOES-East and GOES-West satellites. The goal is to extract the common surface between the two satellites and quantify the discrepancy surface in the process. We then illustrate the ability of our proposed model to handle a massive bivariate dataset by considering the two dimensional vectors where the first components correspond to BHRiso and the second to DHR30, which stands for directional hemispherical reflectance, or *black sky albedo*. This analysis covers data over the whole CONUS and amounts to about 800,000 observations for each component.

2 Nearest-Neighbor GP with Random Covariance Matrices

We define a nearest-neighbors GP with random covariance matrix (NN-RCM) based on the normal inverse-Wishart (NIW) framework. Let $\mathcal{S} = \{s_1, ..., s_k\}$ be a set of locations in a spatial domain $\mathcal{D}$ for which we have spatially dependent q-variate observations $\boldsymbol{y}(s_i), i = 1, \ldots, k$ stacked in the vector $\boldsymbol{y}_\mathcal{S} \in \mathbb{R}^{kq}$. Assume that

$$\boldsymbol{y}_\mathcal{S}|\Sigma \sim N_{kq}(0, \Sigma), \quad \Sigma \sim IW_{kq}(\alpha, (\alpha - kq - 1)C_{\boldsymbol{\theta}}), \tag{1}$$

where $N_n(\boldsymbol{a}, \boldsymbol{V})$ denotes a n-dimensional normal distribution with mean vector $\boldsymbol{a}$ and covariance matrix $\boldsymbol{V}$, $IW_n(\alpha, \boldsymbol{V})$ denotes an $n \times n$-dimensional inverse-Wishart distribution with shape parameter α and scale parameter $\boldsymbol{V}$, and $C_{\boldsymbol{\theta}}$ is a $kq \times kq$ covariance matrix obtained from a valid cross-covariance function with parameters $\boldsymbol{\theta}$. Recall that a cross-covariance function $K_{\boldsymbol{\theta}}$ takes two locations, s_i and s_j and returns a $q \times q$ covariance matrix where each element is the covariance between the respective components:

$$K_{\boldsymbol{\theta}}(s_i, s_j) = \begin{pmatrix} cov(y_1(s_i), y_1(s_j)) & \cdots & cov(y_1(s_i), y_q(s_j)) \\ \vdots & \ddots & \vdots \\ cov(y_q(s_i), y_1(s_j)) & \cdots & cov(y_q(s_i), y_q(s_j)) \end{pmatrix}.$$

Notice that that $E(\Sigma) = C_{\boldsymbol{\theta}}$ and that, as $\alpha \to \infty$, Σ concentrates around its mean. Thus, in the limit, the model corresponds to a regular GP. Our approach leverages the inverse-Wishart distribution of Σ to obtain a non-stationary, locally adaptive model.

To introduce sparsity and perform conditional inference we consider an ordering of the locations based on their index. We can then write the joint multivariate normal density as the product of conditional normal densities, thus

$$p(\boldsymbol{y}_\mathcal{S}) = \prod_{i=1}^k p(\boldsymbol{y}(s_i)|\boldsymbol{y}(s_1), ..., \boldsymbol{y}(s_{i-1})).$$

Normality implies that the recursive conditional representation of $\boldsymbol{y}(s_i)$ can be obtained as a linear combination of $\boldsymbol{y}(s_1), ..., \boldsymbol{y}(s_{i-1})$. Thus for a lower triangular matrix L and a diagonal matrix Λ, both of size $kq \times kq$, we have that $(I_{kq} - L)\boldsymbol{y}_\mathcal{S} = e, \quad e \sim N_{kq}(0, \Lambda)$, where I_{kq} is the identity matrix and $\Sigma = (I_{kq} - L)^{-1}\Lambda(I_{kq} - L)^{-T}$. To induce sparsity most of the elements of L can be set to zero. Vecchia's approximation is an increasingly popular approach (Vecchia, 1988; Datta et al., 2016) that consists of reducing the size of the conditioning set to a small number, say m, of variables. Denote $N(s_i)$ as the m closest neighbors of s_i in $\{s_1, ..., s_{i-1}\} \in \mathcal{S}$ when $i \geq m$, and as the set

$\{s_1, ..., s_{i-1}\}$ when $i < m$. Denote as $\boldsymbol{y}_{N(s_i)}$ the vector obtained by stacking all the observations corresponding to $N(s_i)$. Then

$$p(\boldsymbol{y}_{\mathcal{S}}) \approx \tilde{p}(\boldsymbol{y}_{\mathcal{S}}) = \prod_{i=1}^{k} p(\boldsymbol{y}(s_i)|\boldsymbol{y}_{N(s_i)}),$$

which corresponds to setting the elements of the matrix L outside of $N(s)$ to zero. For small sample sizes, the impact of the ordering of the locations can be relevant and depend on the choice of covariance function and on the specific data at hand. Guinness (2018) proposes optimal ordinal strategies based on KullbackLeibler divergence and Datta et al. (2016) consider root mean square predictive error to score models with different orderings. For large datasets the common choice of fully randomizing the locations is a reasonable default option.

2.1 Linearization of the NIW model

The fundamental component of our proposed model is the linearization of the the distribution of $\boldsymbol{y}(s_i), i = 1, \ldots, k$ given its m neighbors. Let $(\boldsymbol{y}(s_i), \boldsymbol{y}_{N(s_i)})$ be the vector obtained by stacking the vector of observations at location s_i with those at the neighboring locations within $\mathcal{S}$. Its covariance is denoted as $\Sigma_{\{s_i, N(s_i)\}}$ and is equal to

$$\begin{pmatrix} \Sigma_{s_i,s_i} & \Sigma_{s_i,N(s_i)} \\ \Sigma_{N(s_i),s_i} & \Sigma_{N(s_i),N(s_i)} \end{pmatrix} = \begin{pmatrix} \Phi(s_i) + \Gamma'(s_i)\Sigma_{N(s_i),N(s_i)}\Gamma(s_i) & \Gamma(s_i)'\Sigma_{N(s_i),N(s_i)} \\ \Sigma_{N(s_i),N(s_i)}\Gamma(s_i) & \Sigma_{N(s_i),N(s_i)} \end{pmatrix},$$

where the subindexes indicate the corresponding subvectors, $\Gamma(s_i) = \Sigma_{s_i,N(s_i)}\Sigma_{N(s_i),N(s_i)}^{-1}$, and $\Phi(s_i) = \Sigma_{s_i|N(s_i)} = \Sigma_{s_i,s_i} - \Sigma_{s_i,N(s_i)}\Sigma_{N(s_i),N(s_i)}^{-1}\Sigma_{N(s_i),s_i}$. This equality is known as the Bartlett decomposition (Le and Zidek, 2006). Denote $N_{h,l}(M, A, B)$ as the distribution of a $k \times l$ matrix of normal elements with mean M, and such that within any column the covariance matrix is $A \in \mathbb{R}^{h \times h}$, and within any row the covariance matrix is $B \in \mathbb{R}^{l \times l}$. Using the results in Brown et al. (1994a) we have the following lemma:

Lemma 2.1. *For $i = 1, \ldots, k$, suppose that $\Sigma_{\{s_i, N(s_i)\}}$ is a matrix of size $(m+1)q \times (m+1)q$ that follows an inverse-Wishart distribution $IW_{(m+1)q}(\delta, V)$. Then: $\Sigma_{N(s_i),N(s_i)}$ is independent of $\Gamma(s_i)$ and $\Phi(s_i)$; $\Sigma_{N(s_i),N(s_i)} \sim IW_{mq}(\delta, V_{N(s_i),N(s_i)})$; $\Phi(s_i) \sim IW_q(\delta + m, V_{s_i|N(s_i)})$; and $\Gamma(s_i)|\Phi(s_i) \sim N_{m,q}(V_{N(s_i),N(s_i)}^{-1}V_{N(s_i),s_i}, V_{N(s_i),N(s_i)}^{-1}, \Phi(s_i))$.*

From equation (1), and using the properties of the normal and the inverse Wishart distributions, we have that $(\boldsymbol{y}(s_i), \boldsymbol{y}_{N(s_i)})|\Sigma_{\mathcal{S}} \sim N_{(m+1)q}(0, \Sigma_{\{s_i, N(s_i)\}})$ with $\Sigma_{\{s_i, N(s_i)\}} \sim IW_{(m+1)q}(\alpha - kq + (m+1)q, (\alpha - kq - 1)C_{\boldsymbol{\theta},\{s_i, N(s_i)\}})$, as a sub-vector of a multivariate normal is normal and a sub-matrix of an inverse-Wishart is inverse-Wishart. Thus, the following corollary is obtained by using the definition of the conditional distribution of two normal vectors together with Lemma 2.1.

Corollary 1. *For $i = 1, \ldots k$,*

$$\begin{aligned} \boldsymbol{y}(s_i)|(\boldsymbol{y}_{N(s_i)}, \Gamma(s_i), \Phi(s_i)) &\sim N_q(\Gamma'(s_i)\boldsymbol{y}_{N(s_i)}, \Phi(s_i)) \\ \Gamma(s_i)|\Phi(s_i) &\sim N_{mq,q}\left(C_{\boldsymbol{\theta},N(s_i)}^{-1}C_{\boldsymbol{\theta},N(s_i),s_i}, \frac{1}{(\alpha - kq - 1)}C_{\boldsymbol{\theta},N(s_i)}^{-1}, \Phi(s_i)\right) \\ \Phi(s_i) &\sim IW_{mq}\left(\alpha - kq + (m+1)q, (\alpha - kq - 1)C_{\boldsymbol{\theta},s_i|N(s_i)}\right), \end{aligned}$$

where $C_{\boldsymbol{\theta},s_i|N(s_i)} = C_{\boldsymbol{\theta},s_i} - C_{\boldsymbol{\theta},s_i,N(s_i)}C_{\boldsymbol{\theta},N(s_i)}^{-1}C_{\boldsymbol{\theta},N(s_i),s_i}$.

Corollary 1 allows for the transformation of a model based on a covariance matrix likelihood into a collection of spatially localized hierarchical linear models. This series of local models provides adaptive weights to associate the vector of observations at each location with its neareast neighbors within $\mathcal{S}$. It also provides locally varying covariance matrices. This structure is coherent with a global model for all locations featuring a fully unstructured covariance matrix. In addition, it intrinsically suggests a methodology to parallelize the calculations needed for posterior inference of the model, as the distributions of the matrix of weights $\Gamma(s_i)$, and the covariance matrix $\Phi(s_i)$, depend only on location s_i. This offers strong computational efficiency potential.

2.1.1 Extension to a stochastic process

The model introduced in Corollary 1 can be extended to a valid stochastic process over the whole spatial domain by showing that it produces properly defined multivariate distributions for any collection of locations in $\mathcal{D}$. To this end we first consider an arbitrary set $\mathcal{U} = \{u_1, ..., u_b\} \subset \mathcal{D}\backslash\mathcal{S}$ and let $N(u_i)$ denote the neighborhood of u_i in $\mathcal{S}$. Then

$$\tilde{p}(\boldsymbol{y}_{\mathcal{U}}|\boldsymbol{y}_{\mathcal{S}}, \Sigma_{\mathcal{U},\mathcal{S}}) = \prod_{i=1}^{n} p(\boldsymbol{y}(u_i)|\boldsymbol{y}_{N(u_i)}, \Sigma_{u_i,N(u_i)}) = \prod_{i=1}^{n} N(\Gamma(u_i)\boldsymbol{y}_{N(u_i)}, \Phi(u_i)).$$

This expression is multiplied by $p(\boldsymbol{y}_{\mathcal{S}}|\Sigma_{\mathcal{S}})$, which is a multivariate normal. Integrating $\boldsymbol{y}_{\mathcal{S}}$ out of the resulting product of normals, yields $\tilde{p}(\boldsymbol{y}_{\mathcal{U}}|\Sigma_{\mathcal{U},\mathcal{S}})$. For a fully general extension, consider $\mathcal{V} = \{v_1, ..., v_b\} \subset \mathcal{D}$ and let $\mathcal{U} = \mathcal{V}\backslash\mathcal{S}$. The joint conditional density for the vector $\boldsymbol{y}(\mathcal{V})$ is

$$\tilde{p}(\boldsymbol{y}_{\mathcal{V}}|\Sigma_{\mathcal{V},\mathcal{S}}) = \int \tilde{p}(\boldsymbol{y}_{\mathcal{U}}|\boldsymbol{y}_{\mathcal{S}}, \Sigma_{\mathcal{U},\mathcal{S}})\tilde{p}(\boldsymbol{y}_{\mathcal{S}}|\Sigma_{\mathcal{S}}) \prod_{s \in \mathcal{S}\backslash\mathcal{V}} d(\boldsymbol{y}(s)),$$

where we integrate out the locations in $\mathcal{S}$ that do not appear in $\mathcal{V}$. Since $\tilde{p}(\boldsymbol{y}_{\mathcal{U}}|\boldsymbol{y}_{\mathcal{S}}, \Sigma_{\mathcal{U},\mathcal{S}})$ and $\tilde{p}(\boldsymbol{y}_{\mathcal{S}}|\Sigma_{\mathcal{S}})$ are multivariate Normal distributions, the resulting distribution $\tilde{p}(\boldsymbol{y}_{\mathcal{V}}|\Sigma_{\mathcal{V}})$ is also a multivariate Normal. The covariance for two locations $v_1, v_2 \in \mathcal{V}$, conditional on $\Sigma_{\mathcal{V},\mathcal{S}}$, is,

$$\Sigma_{v_1,v_2} = \begin{cases} \Sigma_{v_1,v_2}, & \text{if } v_1, v_2 \in \mathcal{S} \\ \Gamma(v_1)\Sigma_{N(v_1),v_2}, & \text{if } v_1 \notin \mathcal{S}, \ v_2 \in \mathcal{S}, \\ \Gamma(v_1)\Sigma_{N(v_1),N(v_2)}\Gamma^{'}(v_2) + 1_{(v_1=v_2)}\Phi(v_1), & \text{if } v_1, v_2 \notin \mathcal{S}. \end{cases}$$

In summary, we have developed a multivariate nearest-neighbor non-stationary stochastic process with a tremendous potential for parallel computations. The NN-RCM model creates a dedicated non-isotropic framework that leverages the advantages of multivariate linear regression in terms of computational efficiency. As will be demonstrated in the next section, the possibility to marginalize the spatial random effects furthers even more the ability of obtaining fast posterior inference.

2.2 Marginal NN-RCM model

To facilitate the estimation of $\boldsymbol{\theta}$, the parameters that define the prior cross-covariance function, we obtain a marginal model where the random coefficients $\Gamma(s)$ and random covariances $\Phi(s)$ are integrated out. The result is obtained in the following lemma.

Lemma 2.2. *The marginal distribution of $\boldsymbol{y}(s_i), i = 1, \ldots, k$ conditional on its m neighbors in $\mathcal{S}$ and $\boldsymbol{\theta}$ is*

$$m(\boldsymbol{y}(s_i)|\boldsymbol{y}_{N(s_i)}, \boldsymbol{\theta}) = \sqrt{\frac{1}{\pi}} \frac{\Gamma\left(\frac{\alpha-kq+mq+2}{2}\right)}{\Gamma\left(\frac{\alpha-kq+mq+1}{2}\right)} \frac{|V_{\boldsymbol{\theta},\{s_i,N(s_i)\}}|^{\frac{\alpha-kq+mq+1}{2}}}{|(V_{\boldsymbol{\theta}}+S)_{\{s_i,N(s_i)\}}|^{\frac{\alpha-kq+mq+2}{2}}}$$

$$\times \frac{|(V_{\boldsymbol{\theta}}+S)_{N(s_i)}|^{\frac{\alpha-kq+mq+1}{2}}}{|V_{\boldsymbol{\theta},N(s_i)}|^{\frac{\alpha-kq+m}{2}}},$$

where $V_{\boldsymbol{\theta}} = (\alpha - kq - 1)C_{\boldsymbol{\theta}}$ and $S = \boldsymbol{y}(\mathcal{S})\boldsymbol{y}(\mathcal{S})'$.

Proof. Recall that $m(\boldsymbol{y}(s_i)|\boldsymbol{y}_{N(s_i)}, \boldsymbol{\theta}) = m(\boldsymbol{y}(s_i), \boldsymbol{y}_{N(s_i)}|\boldsymbol{\theta})/m(\boldsymbol{y}_{N(s_i)}|\boldsymbol{\theta})$. As $(\boldsymbol{y}(s_i), \boldsymbol{y}_{N(s_i)})|\Sigma, \boldsymbol{\theta} \sim N_{(m+1)}(0, \Sigma_{\{s_i,N(s_i)\}})$ and $\Sigma_{\{s_i,N(s_i)\}}$ follows an inverse-Wishart distribution, the result can be obtained using the properties of NIW densities. $\square$

Further simplification of the expression in Lemma 2.2 can be obtained with the approximation

$$\frac{\Gamma\left(\frac{\alpha-kq+mq+2}{2}\right)}{\Gamma\left(\frac{\alpha-kq+mq+1}{2}\right)} \approx \sqrt{\frac{\alpha - kq + mq + 2}{2} - 1}.$$

that is important for numerical stability for large k. On the other hand, To obtain an approximation to the marginal distribution for all the available observations we let

$$m(\boldsymbol{y}_{\mathcal{S}}|\boldsymbol{\theta}) \approx \tilde{m}(\boldsymbol{y}_{\mathcal{S}}|\boldsymbol{\theta}) = \prod_{i=1}^{k} m(\boldsymbol{y}(s_i)|\boldsymbol{y}_{N(s_i)}, \boldsymbol{\theta}).$$

$\tilde{m}(\boldsymbol{y}(\mathcal{S})|\boldsymbol{\theta})$ can be used as a likelihood for $\boldsymbol{\theta}$. Multiplication by a prior density $\pi(\boldsymbol{\theta})$ yields an approximate posterior distribution for $\boldsymbol{\theta}$ that can be explored through sampling or maximization. Notice that computing the marginal involves, for each location, the determinants of four matrices of dimensions no larger than $(m+1)q \times (m+1)q$. Such local computations can be performed concurrently, enabling the possibility of important speed gains through straightforward parallelization.

2.2.1 Prior distribution of α and $\boldsymbol{\theta}$

The parameters that constitute $\boldsymbol{\theta}$ depend on the choice of cross-correlation function used for the model. We delay the discussion of the prior cross-covariance model to Section 3.2. In the univariate setting the Matérn family has become the most popular choice of spatial covariance function due to its flexibility. The Matérn family depends on two parameters, the range parameter ν and the smoothness parameter κ:

$$C_{\nu,\kappa}(d) = \frac{2^{\kappa-1}}{\Gamma(\kappa)} \left(\sqrt{2\kappa}\frac{d}{\nu}\right)^{\kappa} K_{\kappa}\left(\sqrt{2\kappa}\frac{d}{\nu}\right),$$

where Γ is the gamma function and K_{κ} is the modified Bessel function of the second kind. In addition, we also include a nugget ξ^2 to account for the initial lag at the zero distance. Finally, we multiply by the partial sill σ^2 to obtain the prior covariance function $C_{\boldsymbol{\theta}}(d) = \sigma^2(C_{\nu,\kappa}(d) + \xi^2)$, where $\boldsymbol{\theta} = (\nu, \kappa, \xi, \sigma^2)$.

7

The difficulty in imposing valid priors for the range parameter ν in a GP setting is discussed in Berger et al. (2001) and Kazianka and Pilz (2012) whose focus is to use priors that produce a proper posterior distribution. The maximization aspect of the marginal model also requires attention, as there is no guarantee that the maximum of the marginal will not occur with ν tending to zero or ∞. Following a similar argument as in Gu and Berger (2016), we take $p(\nu)$ such that $p(\nu) \to 0$ when $\nu \to 0$ or $\nu \to \infty$. In fact, our default choice of prior for ν a is a gamma distribution with mean equal to the smallest observed distance between two datapoints and shape parameter greater than 1. For ξ^2 and σ^2 we consider independent diffuse inverse gamma distributions. Finally, a natural choice for the prior distribution of the degrees of freedom α is the Pareto(x_m, p) distribution which has a truncated domain and has very thick tails. By setting the scale parameter x_m at the minimum possible value for the degrees of freedom, we ensure that the distribution is only valid on the proper interval. For the shape parameter p, a value of $p = 1$ is recommended as it results in a distribution with infinite variance.

2.2.2   Posterior predictive distribution

The approximate posterior marginal likelihood model results in samples or point estimates of the cross-covariance parameters, say $\boldsymbol{\theta}^*$. Inference for the random field $\boldsymbol{y}(s)$, is performed by drawing samples from the posterior predictive distribution for any location, say s^* in $\mathcal{D}$. To obtain a sample of the posterior predictive distribution of $\boldsymbol{y}(s^*)$ recall the results in Corollary 1 and sample $\Phi_b(s^*)$ as

$$\Phi_b(s^*) \sim IW_{mq}\left(\alpha^* - kq + (1+m)q, (\alpha^* - kq - 1)C_{\boldsymbol{\theta}^*, s^*|N(s^*)}\right)$$

where $b = 1, ..., B$, denotes the sample replicate. We then sample the matrix $\Gamma_b(s^*)$ as

$$\Gamma_b(s^*)|\Phi_b(s^*) \sim N_{mq,q}\left(C^{-1}_{\boldsymbol{\theta}^*, N(s^*)}C_{\boldsymbol{\theta}^*, N(s^*), s^*}, \frac{1}{(\alpha^* - kq - 1)}C^{-1}_{\boldsymbol{\theta}^*, N(s^*)}, \Phi_b(s^*)\right),$$

for $b = 1, \ldots, B$. Finally, using the two former sets of samples, together with the neighboring observations, we generate a set of predictions at location s^* as,

$$\boldsymbol{y}_b(s^*)|(\Phi_b(s^*), \Gamma'(s^*)) = \Gamma'(s^*)\boldsymbol{y}_{N(s^*)} + \epsilon(s^*), \;\; \epsilon(s^*) \sim N_q(0, \Phi_b(s^*)).$$

Consequently, the samples $\boldsymbol{y}_1(s^*), ..., \boldsymbol{y}_B(s^*)$ provide inference for the random field of interest, together with quantification of the predictive variability. Notice that, thanks to the definition of $N(s^*)$, the predictive distribution does not depend on the ordering of $\mathcal{S}$.

2.3   Hierarchical NN-RCM model with covariates

The NN-RCM model considered so far assumes that the spatial observations have zero mean. To model the complexities of realistic multivariate random fields we consider a hierarchical extension of the NN-RCM that incorporates fixed effects, linear combinations of spatial random effects and observational errors. Thus we formulate the model

$$\boldsymbol{y}(s) = X(s)\boldsymbol{\beta} + A\boldsymbol{w}(s) + \boldsymbol{\epsilon}(s)$$

where $X(s)$ is a $q \times p$ matrix of spatially varying covariates, $\boldsymbol{\beta}$ is the p-dimensional vector of fixed effects coefficients, A is a $q \times q$ known nonsingular matrix and $\boldsymbol{w}(s)$ is a latent q-variate NN-RCM

process, and $\boldsymbol{\epsilon}(s)$ is a q-variate vector of observational errors. A hierarchical formulation of the model is given as:

$$\boldsymbol{y}(s)|\boldsymbol{\beta},\boldsymbol{w}(s),\tau^2 \sim N_q(X(s)\boldsymbol{\beta} + A\boldsymbol{w}(s),\boldsymbol{\tau}^2 I_q)$$
$$\boldsymbol{w}(s)|\Gamma(s),\Phi(s) \sim N_q(\Gamma(s)\boldsymbol{w}_{N(s)},\Phi(s))$$
$$\Gamma(s)|\Phi(s) \sim N_{mq,q}\left(C_{\boldsymbol{\theta},N(s)}^{-1}C_{\boldsymbol{\theta},N(s),s}, \frac{1}{\alpha-kq-1}C_{\boldsymbol{\theta},N(s)}^{-1},\Phi(s)\right)$$
$$\Phi(s) \sim IW_{mq}(\alpha - kq + (1+m)q, (\alpha - kq - 1)C_{\boldsymbol{\theta},s|N(s)}),$$

With priors $\boldsymbol{\beta} \sim N_p(0, s_\beta^2 I)$ and $\boldsymbol{\tau}^2 \sim \prod_{i=1}^q IG(\tau_i^2|a_\tau, b_\tau)$.

2.4 Posterior inference for α and $\boldsymbol{\theta}$

To sample α and $\boldsymbol{\theta}$, we leverage the marginal likelihood obtained for the marginal model. By following the same derivation, the marginal likelihood for $w(s)$ can be used in the Metropolis step. Therefore, by removing the random effects, the likelihood involving α and θ is reduced to

$$m(w(s)|w_{N(s)},\boldsymbol{\theta},\alpha) \propto \frac{\Gamma\left(\frac{\alpha-k+m+2}{2}\right)}{\Gamma\left(\frac{\alpha-k+m+1}{2}\right)}\frac{|V_{\boldsymbol{\theta},\{s_i,N(s_i)\}}|^{\frac{\alpha-k+m+1}{2}}}{|(V_{\boldsymbol{\theta}}+S)_{\{s_i,N(s_i)\}}|^{\frac{\alpha-k+m+2}{2}}}\frac{|(V_{\boldsymbol{\theta}}+S)_{N(s_i)}|^{\frac{\alpha-k+m+1}{2}}}{|V_{\boldsymbol{\theta},N(s_i)}|^{\frac{\alpha-k+m}{2}}},$$

where $V_{\boldsymbol{\theta}} = (\alpha - k - 1)C_{\boldsymbol{\theta}}$ and $S = \boldsymbol{w}\boldsymbol{w}'$.

It is worth noting that sampling α and $\boldsymbol{\theta}$ can be subject to slow mixing and result in a significant increase in computation time. A possible strategy to accelerate the inference is to sample α and $\boldsymbol{\theta}$ in only a fraction of the total number of iterations. Taking this strategy to the extreme, this parameters can be held fixed at the point estimates obtained from the marginal approach for the duration of the algorithm.

2.4.1 Misalignment of observations

In many practical problems involving multivariate observations we have to deal with incomplete or missing values at some locations. An extreme case example is given in Section 4, where the observations from two different satellites are slightly misaligned, implying that at most locations only one of the two components is observed. Missing observations can readily be handled within the model, as it relies on a valid cross covariance structure. Thus, proximity of observations from any component can be used for prediction of any other component. To achieve this is it is important to make sure that a number of observations from all components are included in every neighborhood.

Our approach consists of introducing an ordering of the components of $\boldsymbol{y}(s)$. Let $J(s_i)$ be the set of indexes corresponding to components that are observed at location s_i. Let q_i indicate the number of elements in $J(s_i)$. We define neighbors as we did previously based on the m closest locations for each variate. We denote the corresponding set for the j-th component as $N_j(s_i)$. Thus,

$$\tilde{p}(\boldsymbol{y}_S) = \prod_{i=1}^k \prod_{j\in J(s_i)} p(y_j(s_i)|\boldsymbol{y}_{N_j(s_i)}).$$

Notice that the m neighbors from each q-variate need not come from the same set of m locations. Figure 2 illustrates how the neighborhoods are created using a small example. Assuming that we

9

have 5 observed locations with two components shown in orange and purple. The ordering is from left to right and the components are ordered sequentially. Therefore, to create a neighborhood of size 6 for the second component of location s_4, which is denoted in black in the right plot of Figure 2, we would select the three closest previous observations from component 1 and component 2 separately. We denote these in red and together, they form the neighborhood of the observation. It is important to highlight that the neighborhood does include the first component of the observed location.
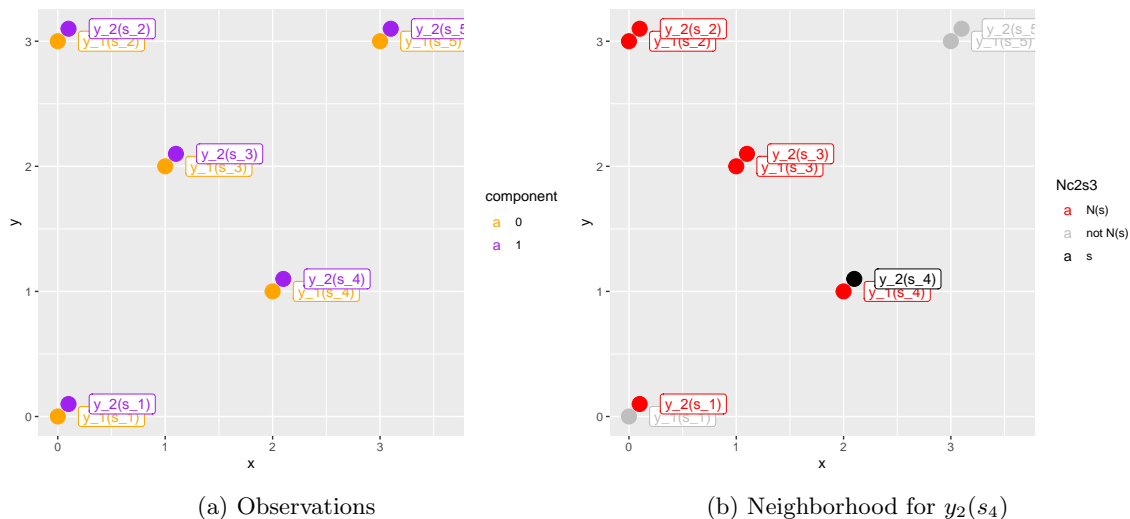

(a) Observations (b) Neighborhood for $y_2(s_4)$

Figure 2: Creation of a bivariate neighborhood using the same number of neighbors from each component.

3 Simulation Study

We present results of a simulation study in order to explore the performance of the marginal and the hierarchical version of the proposed NN-RCM model. We obtain an assessment of the model fit by splitting the sample into a training set and a validation set, and computing four different scores calculated for each component of the field.

The *Predictive Mean Squared Error* (PMSE) is defined as

$$PMSE = \frac{1}{v} \sum_{i=1}^{v} (y_j(s_i) - \hat{y}_j(s_i))^2,$$

where $\hat{y}_j(s_i)$ corresponds to a sample of the predictive posterior distribution at the validation set location $s_i, i = 1, \ldots, v$, for the j-th component. The *Continuous Rank Probability Score* (CRPS) (Gneiting and Raftery, 2007) is defined as

$$CRPS(s) = \frac{1}{B} \sum_{i=1}^{B} |\hat{y}_j^{(i)}(s) - y_j(s)| - \frac{1}{2B^2} \sum_{i=1}^{B} \sum_{l=1}^{B} |\hat{y}_l^{(i)}(s) - \hat{y}_j^{(l)}(s)|,$$

10

where $\hat{y}_j^{(i)}(s)$ and $\hat{y}_j^{(l)}(s)$ $i, l = 1, \ldots, B$ are samples from the posterior predictive for the j-th component at loction s. The CRPS is computed for all the locations in the validation set and then averaged. The *Posterior posterior predictive loss criterion* (PPLC) (Gelfand and Ghosh, 1998) is defined as

$$D_p = \sum_{i=1}^{v} var(\hat{y}_j(s_i)) + \frac{p}{p+1} \sum_{i=1}^{v} (y_j(s_i) - E(\hat{y}_j(s_i)))^2 \ .$$

Finally, we consider the 95% predicted interval coverage *(Coverage)*, defined as.

$$Coverage = \frac{1}{v} \sum_{i=1}^{v} 1_{\{y_j(s_i) \geq y_j^{(0.025v)}(s_i)\}} 1_{\{y_j(s_i) \leq y_j^{(0.975v)}(s_i)\}},$$

where the boundaries of the interval are given by the 2.5% and 97.5% quantiles of the predictive sample at location s_j for the j-th component.

3.1   Univariate simulation

We generate a dataset from a Gaussian Process with a Matérn covariance function with smoothness parameter $\kappa = 1/2$. We obtain $n = 2500$ observations $y(s_i)$ in a 10×10 square using the covariance parameters $\boldsymbol{\theta} = (\sigma^2, \nu) = (1, 1)$, the partial sill and the range. We add white noise according to a $N(0, 0.5^2)$ distribution. Finally, we split the sample evenly into a training set and a testing set.

We fitted the simulated data using the marginal model proposed in Section 2.2, as well as the hierarchical model presented in Section 2.3. We used a neighborhood of size $m = 10$ and an inverse gamma prior with the partial sill σ^2. As suggested in Section 2.2.1 we use a gamma prior with hyperaramaters based on the smallest observed distance of the training dataset and a Pareto distribution with low information level is used on the degrees of freedom α. Using the base R optim function, we obtain posterior estimates for θ and α. Starting the optimization at some default values where all parameters are equal to 0.5, the process converges after 50 iterations. We compare our results to the ones obtained using the R packages spNNGP and SpConjNNGP that implement Bayesian versions of the nearest neighbor Gaussian process using, respectively, MCMC and cross-validation using a marginal model (Finley et al., 2017). The resulting posterior predictive means are shown in Figure 3 for the marginal models, and Figure 4 for the models with a full MCMC implementation. We observe that, graphically, all four predictions are very similar, all of them corresponding to an accurate recovery of the original field. We observe that the panels corresponding to the NNGP approaches show slightly more localized variability than the ones corresponding to NN-RCM. A quantitative comparison of the four different results is presented in Table 1. Using the four proposed scores we observe very small differences between the four predictive methods.

Model	PMSE	CRPS	PPLC	Coverage
marginal NN-RCM	0.2627	0.3005	821.47	0.9992
hierarchical NN-RCM	0.2749	0.3160	1041.29	1.0000
conjugate NNGP	0.2646	0.3021	841.71	0.9984
MCMC NNGP	0.2619	0.3028	872.04	0.9984

Table 1: Scores comparison for the univariate simulation under the four studied models.

Inference for the full MCMC NNGP model took 3.45 seconds and predictions on the testing set took 8.55 seconds. The analysis was performed using a Razer Blade 15 with Intel(R) Core(TM)
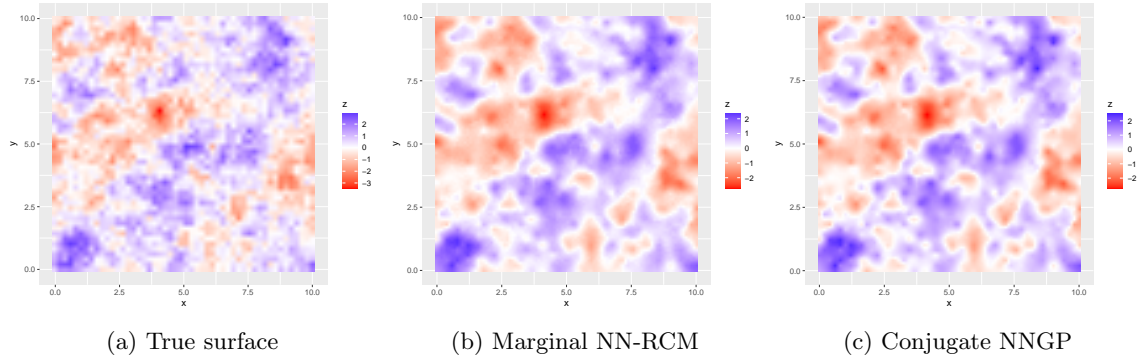
(a) True surface   (b) Marginal NN-RCM   (c) Conjugate NNGP

Figure 3: Comparison of true surface (no noise) and predicted surfaces for the marginal NN-RCM model and conjugate NNGP model.

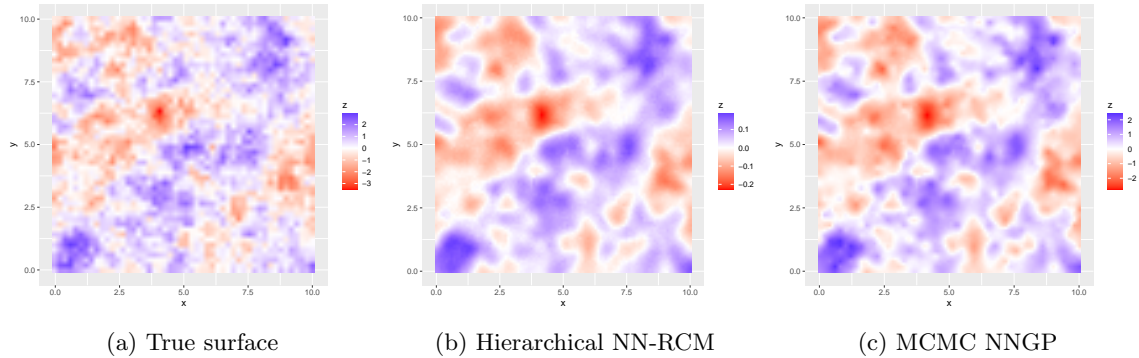(a) True surface   (b) Hierarchical NN-RCM   (c) MCMC NNGP

Figure 4: Comparison of true surface (no noise) and predicted surfaces for the hierarchical NN-RCM model and full MCMC NNGP model.

i7-10750H Processor with 16.0GB of RAM. The hierarchical NN-RCM model took 4.05 seconds for posterior inference but produced samples of the predictions 4 times faster, taking only 2.21 seconds. Obtaining the high resolution predictions (10,000 locations) took 70 seconds for the NNGP `predict` which is exactly 4 times more than the NN-RCM model at 17.5 seconds. For the conjugate NNGP model the runtime was 50 seconds. In comparison, the marginal NN-RCM model optimized the covariance parameters in under 5 seconds with no guidance on the starting values. We notice that the runtime is affected by the stopping rule of the optimization process. Using a very large convergence tolerance or limiting the number of iterations can decrease the runtime artificially. For the results presented here we rely on the default stopping rule of the R package `optim`, which uses a convergence tolerance of $1e - 08$ and a total number of iterations of 100. Finally, we note that the starting values can also have a small impact on the runtime of the marginal NN-RCM model. For the simulation, we made an honest attempt at keeping the starting values within a reasonable guess, similar to the grid of values given to the `spConjNNGP` function. In conclusion, the exploration of the univariate simulated dataset reveals that both implementations of NN-RCM are competitive with state of the art NNGP models, with some advantages in terms of showing less local variability

and faster runtime to obtain predictive surfaces.

3.2 Bivariate Simulation

The formulation of a NN-RCM requires the specification of a cross-covariance function. Genton and Kleiber (2015) gives a thorough overview of multiple parametric models for spatial modeling. We consider three methods in detail and discuss their specific implementation. The first choice is a separable model where the cross-covariance matrix is the result of the Kronecker product between an inter-component covariance matrix T and a spatial covariance matrix generated from a covariance function C. That is, the covariance between the variate i and j for two locations s_1 and s_2 can be written as $C_{ij}(s_1, s_2) = t_{ij} * C(s_1, s_2)$. The second option corresponds to the linear model of coregionalization, consisting of the combination of univariate random fields using a linear representation. Assuming the linear combinations can be represented by the nonsingular matrix A, the cross-covariance is then defined as $A'C^*A$, where C^* is a block-matrix where each block represents the covariance matrix obtained from the respective univariate covariance function C_k. Any element of the cross-covariance matrix follows the equation below,

$$C_{ij}(s_1, s_2) = \sum_{k=1}^{q} C_k(s_1, s_2) a_{ik} a_{jk}.$$

Finally, we consider a multivariate extension of the Matérn covariance function. Gneiting et al. (2010) define a flexible family of Matérn cross covariance functions that depend on correlation coefficients ρ_{ij}, on smoothness parameters $\kappa_i j$ and on range parameters ν_{ij}. These are subject to constraints that present inferential challenges. A parsimonious option is given by the case where the range parameter ν is the same for all components. This offers a largely simplified implementation of the function that results in the cross-covariance being

$$C_{ij}(s_1, s_2) = \rho_{ij} \sigma_i \sigma_j M(s_1, s_2 | (\kappa_i + \kappa_j)/2, \nu),$$

where $M(\cdot | \cdot)$ is the univariate Matérn covariance function for two locations s_1 and s_2, $\rho_{i1} = 1$. The only condition, that is both sufficient and necessary, for the validity of this cross-covariance function is

$$|\rho_{ij}| \leq \left(\frac{\Gamma(\kappa_i + \frac{d}{2})}{\Gamma(\kappa_i)} \right)^{1/2} \left(\frac{\Gamma(\kappa_j + \frac{d}{2})}{\Gamma(\kappa_j)} \right)^{1/2} \frac{\Gamma(\frac{1}{2}(\kappa_i + \kappa_j))}{\Gamma(\frac{1}{2}(\kappa_i + \kappa_j) + \frac{d}{2})},$$

where d is the dimension of the space. On the Euclidean plane, where $d = 2$, this condition reduces to

$$|\rho_{ij}| \leq \frac{(\kappa_i \kappa_j)^{1/2}}{\frac{1}{2}(\kappa_i + \kappa_j)}.$$

A slightly more general version, specific to the bivariate case, allows different range parameters per component, as long as $\nu_1 + \nu_2 \geq 2\nu_{12}$. Our default implementation assumes $\nu_1 + \nu_2 = 2\nu_{12}$.

3.2.1 Simulation 1

The first bivariate example that we discuss consists of two surfaces generated using a cross-covariance Matérn function. We use the package `RandomFields` to generate the datasets (Schlather

Model	Runtime	Component	PMSE	CRPS	PPLC	Coverage
Coregionalization	49.01s	Component 1	0.0503	0.1674	115.66	1.0000
		Component 2	0.6838	0.4658	418.68	0.9754
Parsimonious	6.98s	Component 1	0.0522	0.1680	116.42	1.0000
		Component 2	0.6884	0.4657	415.16	0.9754
Flexible	8.05s	Component 1	0.0507	0.1673	116.77	1.0000
		Component 2	0.6749	0.4617	417.37	0.9754
Separable	6.58s	Component 1	0.0643	0.1772	121.98	1.0000
		Component 2	0.7063	0.4706	419.81	0.9815

Table 2: Scores comparison between the four possible cross-covariance function choices for the bivariate marginal NN-RCM model. The scores are computed for each simulated surface individually.

et al., 2021, 2015). We use range and smoothness parameters that vary greatly between the two surfaces. Moreover, we include a correlation of -0.5 between the two surfaces and spatially-independent noise to each observation. For the cross-validation, we use a 50-50 split on the 2,500 simulated observations. We used four different cross-covariance functions to fit a marginal NN-RCM to the simulated data: separable, co-regionalization, parsimonious Matérn, and flexible Matérn. As was the case for the univariate simulation, we use the package `optim` to estimate the parameters of the cross-covariance functions. The size of the neighborhood was chosen to be $m = 20$, where each observation has an equal number of neighbors (10) from each component. The priors for the parameters also include an inverse gamma distribution for the partial sills and cross-covariance terms. We use gamma distributions for the range parameters based on the smallest distance observed for each component. Finally, we use a Pareto distribution for the degrees of freedom with scale parameter 1 which results in a prior with infinite variance.

A quantitative comparison of the different approaches using the same scoring functions calculated for the univariate simulation is reported in Table 2. We observe that the scores are very similar for all four approaches. We notice that the coregionalization approach requires substantially longer time to converge than any of the other three approaches. The runtimes reported in Table 2 only account for the optimization process to obtain the posterior estimates. A graphical comparison of the posterior predictive means for the parsimonious model is presented in Figures 5 and 6. These, along with the comparable figures produced by the other three methods, seem to accurately depict the original datasets. Note that overly smooth posterior predictive mean for the second component is expected as the underlying true surface presents a high degree of roughness.

Since the `spNNGP` package does not offer a multivariate implementation of the NNGP model, we resort to comparing our results to the `spBayes` package (Finley et al., 2015). For the surface of the second component, the limitations of the knots was apparent as the predictive process struggled to fully recover the true surface. In terms of runtime, obtaining the posterior inference, the `spBayes` package took 2.35 minutes compared to the bivariate NN-RCM model which took 6.98 seconds.

3.2.2 Simulation 2

Our second simulation consists of transforming the data previously simulated using the linear equation

$$\boldsymbol{y}(s) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \boldsymbol{w}(s).$$
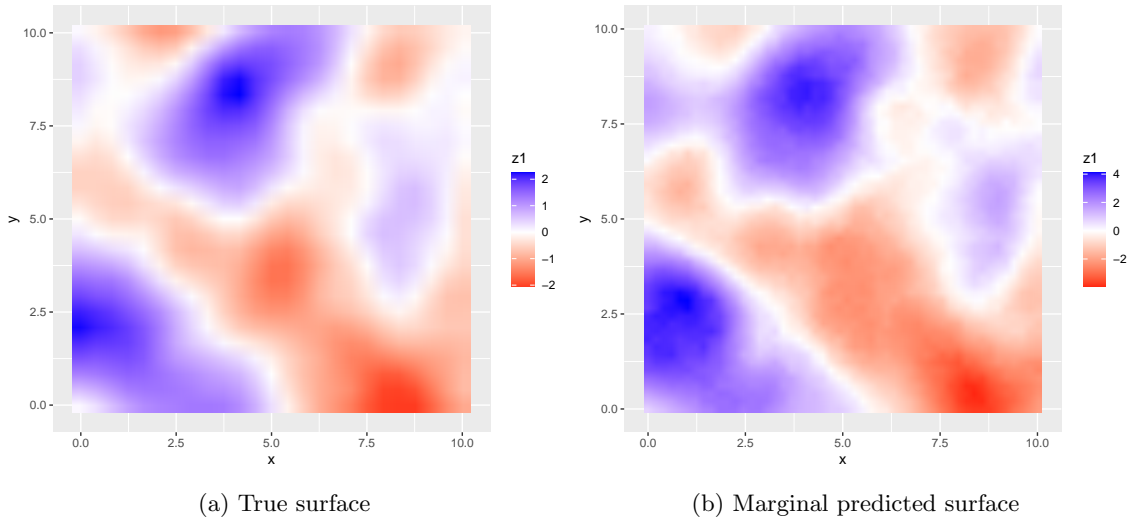
14

(a) True surface

(b) Marginal predicted surface

Figure 5: Comparison of the first true surface (no noise) and predicted surface from the bivariate marginal NN-RCM model.
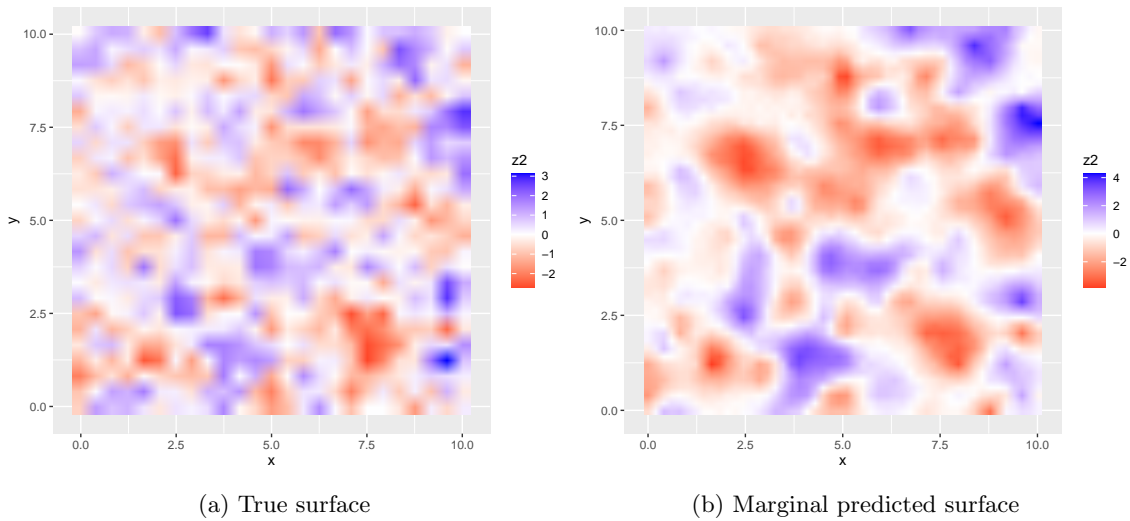

(a) True surface

(b) Marginal predicted surface

Figure 6: Comparison of the second true surface (no noise) and predicted surface from the bivariate marginal NN-RCM model.

The objective of this second simulation is to demonstrate the ability of the hierarchical NN-RCM model to recover the spatial random effects denoted by $w(s)$. Using the same priors described for the first simulation and the posterior estimates as starting values, we implement the hierarchical NN-RCM model with the parsimonious Matérn cross-covariance function. For computational efficiency, we hold the degrees of freedom α and the range parameters fixed. Figure 7 and 8 show the posterior

mean of $w_1(s)$ and $w_2(s)$. This confirms that we are able to recover the spatial random effects accurately given the two transformed surfaces $y_1(s)$ and $y_2(s)$.
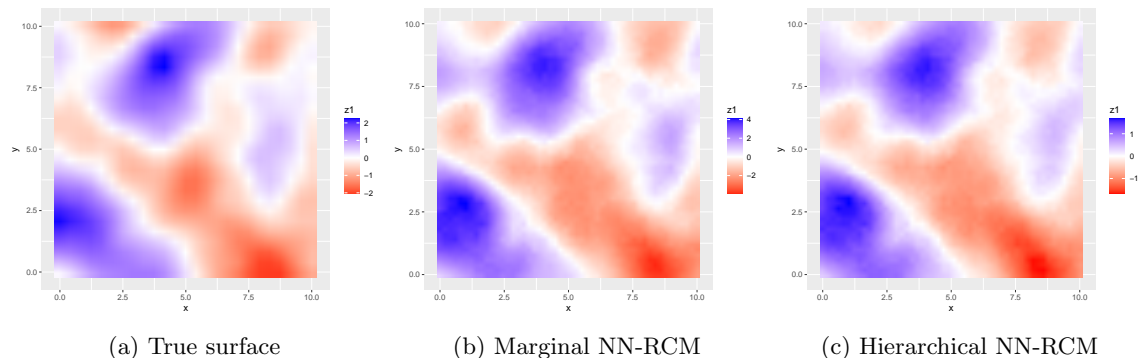

(a) True surface (b) Marginal NN-RCM (c) Hierarchical NN-RCM

Figure 7: Comparison of the first true surface (no noise) and predicted surfaces from the bivariate NN-RCM models.

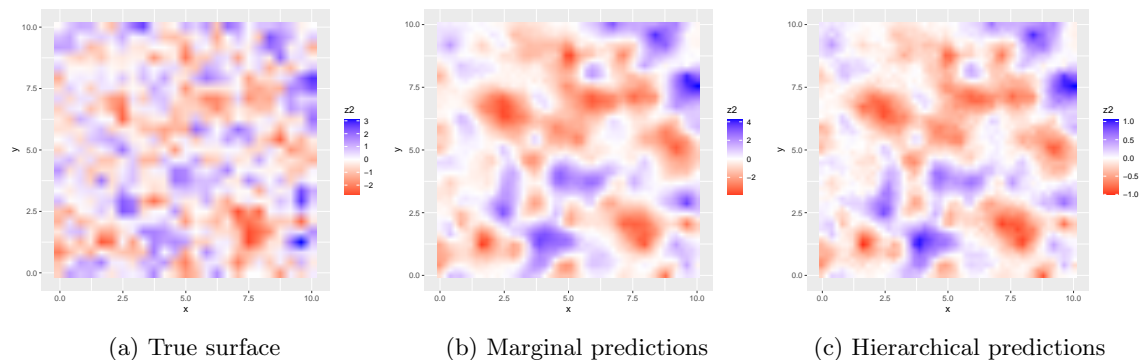(a) True surface (b) Marginal predictions (c) Hierarchical predictions

Figure 8: Comparison of the second true surface (no noise) and predicted surfaces from the bivariate NN-RCM models.

4   Albedo Application

To illustrate the behavior of the NN-RCM model, we selected data from the GOES-East and GOES-West satellites for the first day of the month of July 2000. The data are retrieved at a resolution of 4km by 4km. By restricting our analysis to the continental United States (CONUS), the result is a dataset with approximately 665 000 observations per day. As previously noted, the albedo product is captured by a percentage of the weight between upward and downward radiant influxes. In order to transform the range to be applicable for the GP model, we use a logit transformation on the data, after truncation of values above 0.5. Figure 9 shows a subset of size 10,000 for the July 1st, 2000 data. Points colored bright yellow are likely to correspond to areas with dense cloud covering.
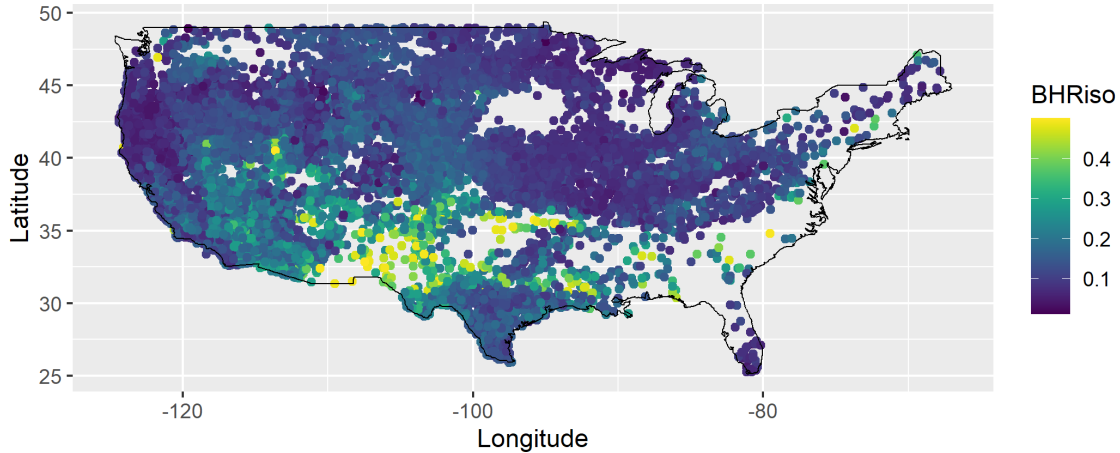
Figure 9: Subset of surface albedo observations from July 1, 2000.

As a first demonstration we apply the hierarchical bivariate spatial model developed in section 2.3 to an area restricted to the state of Colorado. The hierarchical bivariate model allows us to merge the observations from the two satellites using a common surface $w(s)$ and a differential surface $d(s)$. Then, in light of the results obtained, we continue our analysis by applying the univariate hierarchical NNRCM model to the whole CONUS by stacking the information obtained from both satellites. Finally, we conclude with an example of the application of the marginal bivariate NNRCM model. We again apply it to the whole CONUS area but focus on the joint learning of BHRiso and DHR30.

4.1 Bivariate surface albedo (Colorado)

The model that we seek to use to analyse the albedo observations relies on two spatial processes, a common surface and a differential surface. Let s be any location in our space of interest, we denote the two possible observations from the GOES-East and the GOES-West satellites as $y_E(s)$ and $y_W(s)$. We attempt to reconcile them to obtain a common surface albedo represented by the spatial process $w(s)$ and a discrepancy quantification from the process $d(s)$. The model we use is therefore denoted as

$$\begin{pmatrix} f(y_E(s)) \\ f(y_W(s)) \end{pmatrix} = X(s)\boldsymbol{\beta} + \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} w(s) \\ d(s) \end{pmatrix} + \begin{pmatrix} \epsilon_1(s) \\ \epsilon_2(s) \end{pmatrix},$$

where f is any appropriate transformation to the real line, $X(s)$ and $\boldsymbol{\beta}$ represent the covariates and fixed effects to be included, and $\epsilon_i \sim N(0, \tau_i^2)$, for $i = 1, 2$. The covariates included in $X(s)$ are an

17

Covariate	Mean	St. dev.
Intercept	-2.4466	0.0820
Longitude	-0.0017	0.0005
Latitude	-0.0032	0.0015

Table 3: Posterior mean of the fixed effects of the stacked albedo linear model for the state of Colorado.

intercept and the longitude and the latitude of the location s. The choice of having common fixed effects $\boldsymbol{\beta}$ for both sources of information reflects our assumption that the effects of the longitude and latitude on the albedo assessments are not impacted by the satellite retrieval process.

As an initial exploration, we focus our analysis on the state of Colorado where the satellites GOES-East and GOES-West contain around 18,000 and 11,000 observations respectively. To capture the predictive fit of the model, we first split the data into two equal subsets which we refer to as the training set and the testing set. We will first fit the model on the training set only to evaluate the predictive power. Then, we will fit the model once more on the full 29,000 datapoints to obtain high resolution posterior predictive surfaces.

To fit this model as efficiently as possible, we proceed with a multi-tier approach. First, we discuss the transformation f and the fixed effects. Second, we look at the univariate marginal nearest-neighbor Gaussian processes with random covariance matrices (NN-RCM) models used to obtain point estimates for the covariance parameters and the degrees of freedom. Since the bivariate marginal model does not output point estimates for linear combinations of the observations, we use the univariate marginal NN-RCM model in a two-step process. Finally, we use the bivariate hierarchical NN-RCM model to obtain the posterior inference on $w(s)$ and $d(s)$. Using these posterior samples, we obtain posterior predictive samples for both $w(s)$ and $d(s)$ which help us understand the uncertainty of our predictions and the difference between the two GOES satellites.

4.2 Univariate marginal model tiered approach

As we previously explained, the albedo observations are percentages and first need to be transformed. We used a logit transformation to transform the percentages to the real line. We then use a univariate linear model on the stacked training observations to remove the average fixed effects. Table 3 summarizes the posterior mean and standard deviation for the fixed effects obtained using the `lm` function in `R`. From these, we can remove the predicted expected value from our transformed observations and continue with our spatial analysis.

Denote the transformed and centered observations as $y_E^*(s)$ and $y_W^*(s)$. The bivariate marginal NN-RCM model is fitted on two spatial processes that correspond to each source of information. Therefore, in this scenario, the point estimates that the bivariate marginal NN-RCM model obtains do not correspond to the covariance parameters for $w(s)$ and $d(s)$. Instead, we suggest using a two-step method to obtain an approximation of the those covariance parameters.

First, we fit the univariate marginal model using the observations from only the first satellite, GOES-East. To maintain an acceptable runtime for each of the increasingly larger domain, we chose to fix the number of neighbors to 10. The runtime for this first step was 19.75 seconds. Using the posterior inference obtained, we extract predictions $y_W^{E(b)}(s)$, for $b = 1, ..., 1000$, for each the observed locations s of the second satellite, GOES-West. This second step took 3.08 seconds.

18

Parameter	Notation	$w(s)$	$d(s)$	Bivariate estimate
Degrees of freedom	$\alpha_w, \alpha_d, \alpha$	9028	5751	29557
Range	ν_w, ν_d, ν	0.2274	0.3081	0.2678
Partial Sill	σ_w^2, σ_d^2	0.1832	0.1054	-
Correlation	ρ_{wd}	-	-	-0.0430
Nugget	ξ_w^2, ξ_d^2	0.1726	0.9117	-

Table 4: Point estimates for the covariance parameters obtained from the multi-tiered approach using the univariate marginal NN-RCM models for the state of Colorado. The last column shows the aggregated values which are used in the bivariate model.

Finally, we fit a second univariate marginal NN-RCM model on the difference between the observed value and the average predicted value for the second source:

$$y_W(s) - \sum_{b=1}^{1000} \frac{y_W^{E(b)}(s)}{1000} \sim NN - RCM(0, C_\theta).$$

This last step took 9.89 seconds for an overall runtime of 33 seconds. Together, the two models supply us with approximates for $\xi_w^2, \xi_d^2, \sigma_w^2, \sigma_d^2, \nu_w, \nu_d, \alpha_w, \alpha_d$. We can therefore use the proxy nugget estimates and the partial sill estimates directly. For the range parameter, since the bivariate model uses a common range, we have multiple options. We decided to use the average of the two estimates, but another choice was to use the smallest of the two ranges. Finally, for the degrees of freedom, we can obtain the point estimate for α by adding the two estimates. We therefore have $\alpha = 2*(\alpha_w + \alpha_d)$, where the estimate must be multiplied by 2 to account for the misalignment of the satellites. More precisely, the two satellites do not share the same assessment grid which results in a bivariate dataset with exactly 50% missing data.

This methodology does not provide us with an estimate for the cross-covariance term of the Matérn covariance function, ρ_{wd}. A good approximation of the parameter is obtained by computing the empirical correlation between the posterior predictive averages for the second satellite and the difference between the albedo assessments and the posterior predictive means. Carrying this calculation and adjusting for the smoothness parameters, we obtain an estimate of $\rho_{wd} = -0.0430$. See Table 4 for the complete list of point estimates obtained from the univariate marginal NN-RCM models approach.

4.3 Bivariate hierarchical model

The final step of our spatial inference analysis consists of fitting the bivariate hierarchical NN-RCM model to $y_E^*(s)$ and $y_W^*(s)$. The first few steps have allowed us to reduce the model to the following,

$$\begin{pmatrix} y_E^*(s) \\ y_W^*(s) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} w(s) \\ d(s) \end{pmatrix} + \begin{pmatrix} \epsilon_1(s) \\ \epsilon_2(s) \end{pmatrix},$$

where $\epsilon_i \sim N(0, \tau_i^2)$, for $i = 1, 2$. Therefore, we seek to sample the spatial random effects $w(s)$ and $d(s)$ and the observational errors τ_i^2 using the bivariate hierarchical NN-RCM model with the point estimates as outlined in Table 4 for the parsimonious Matérn cross-correlation function. Overall,

19

Satellite	Model	PMSE	CRPS	PPLC	Coverage
GOES-E	Bivariate NN-RCM	0.0424	0.1067	814.89	0.9826
	Univariate NN-RCM	0.0365	0.1651	3567.11	1.0000
GOES-W	Bivariate NN-RCM	0.0496	0.1506	1712.83	0.9967
	Univariate NN-RCM	0.03401	0.1623	2205.35	1.0000

Table 5: Scores comparison for the hierarchical NN-RCM model and the univariate marginal NN-RCM model for the state of Colorado. The scores for the GOES-East and GOES-West testing sets are reported individually.

the posterior inference took 2.53 minutes to complete. Using these posterior samples, we complete the final step of the cross-validation analysis by sampling the posterior predictive distribution of the testing set locations. With those, we can compute the four scoring values and compare them to using a univariate NN-RCM model on the stacked dataset. Table 5 summarizes the results and shows that the bivariate model does indeed outperform its univariate counterpart. While the predictive mean squared error is marginally smaller for the univariate model, we reduce the predictive variability by using both sources of information.

4.4 High resolution predicted surface

Repeating the same steps but for the full state of Colorado dataset, we again obtain posterior samples for the bivariate hierarchical NN-RCM model. Using these samples we generate posterior predictive samples for a high resolution grid. Figure 10 shows the posterior mean for the spatial random effect $w(s)$ and the albedo predictions on the original scale.
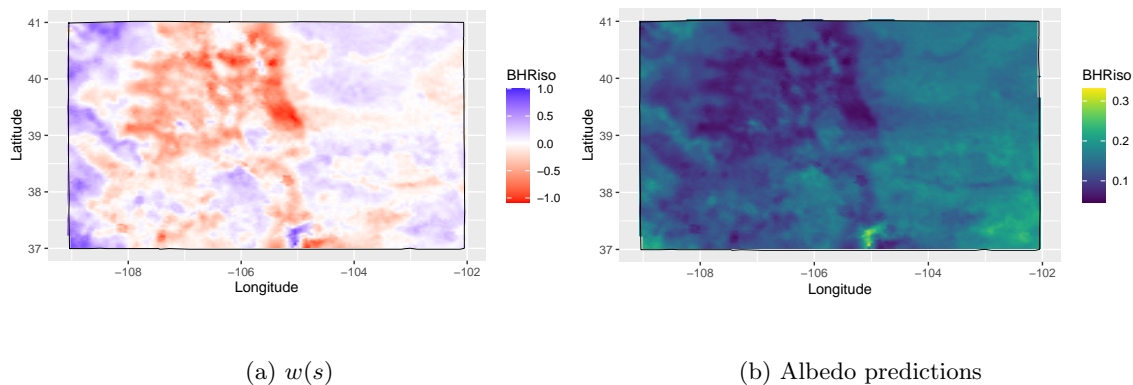

(a) $w(s)$ (b) Albedo predictions

Figure 10: Spatial random effects $w(s)$ and albedo predictions from the bivariate hierarchical NN-RCM model for the state of Colorado.

We are particularly interested in the spatial surface $d(s)$ to learn about the differences between the two satellites GOES-East and GOES-West. Figure 11 shows the posterior predictive mean for $d(s)$ and the posterior predictive difference between $y_E(s)$ and $y_W(s)$ on the original scale. That

is, for a location s, we plot the average of the following quantity:

$$d_b^*(s) = f^{-1}(w_b(s) + d_b(s) + X(s)\boldsymbol{\beta}) - f^{-1}(w_b(s) + X(s)\boldsymbol{\beta}),$$

where $b = 1, ..., 1000$.

We immediately notice that the discrepancy between the two satellites presents irregular spatial patterns that correspond to the Colorado Rocky Mountains. These can not be attributed to simple effects of longitude/latitude or the satellite view angle. By looking at the 90% posterior predictive intervals at each location s, we note that, for over 99% of the locations, the estimated discrepancy field overlaps with 0. This indicates that any significant differences between the two satellites are very localized. At a regional or global scale, those differences are dominated by the variability of the albedo fields.

In the next section, taking into consideration the small global significance of the field $d(s)$, we instead look at the full CONUS area by stacking the information from both satellites. This will allow us to obtain a unified surface and demonstrate the computational efficiency of the presented method.
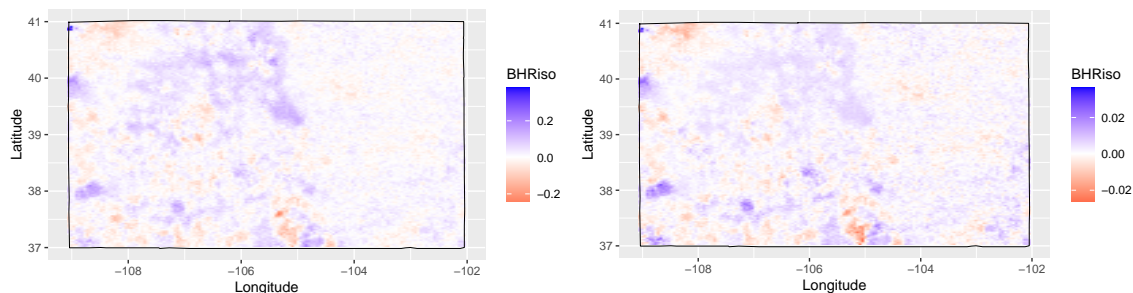
(a) $d(s)$ (b) $y_E(s) - y_W(s)$

Figure 11: Spatial random effects $d(s)$ and posterior predictive mean of $y_E(s) - y_W(s)$ from the bivariate hierarchical NN-RCM model for the state of Colorado.

4.5 Univariate Hierarchical Model application

By stacking the albedo information from both satellites, we obtain a dataset with approximately 800,000 datapoints. Our approach to obtain the posterior inference is to proceed in two steps. First, we will use the marginal NN-RCM model to recover point estimates for the degrees of freedom and the spatial parameters of the model. Second, we apply the hierarchical model to obtain the full posterior inference for the spatial random effects. The objective in using the two steps is to reduce the overall computational burden of handling this large dataset. The total runtime for both steps was 17 minutes. The optimization of the marginal likelihood took 14 minutes and the posterior inference for the spatial random effects took 3 minutes. In Figure 12, we show the posterior predictive mean obtained from fitting the univariate hierarchical NN-RCM model.
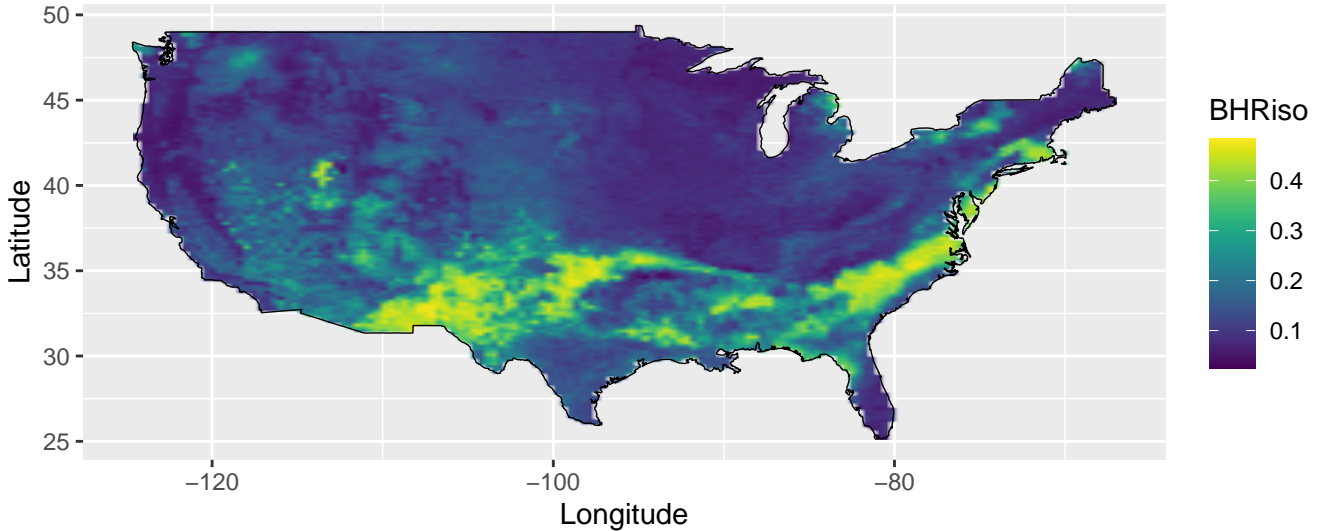
Figure 12: Posterior predictive average over CONUS obtained from the univariate hierarchical NN-RCM model.

4.6   Bivariate Marginal Model application

For the final application, we use the bivariate marginal model to reconcile two albedo measures: BHRiso and DHR30. In this scenario, we are stacking the information from both the GOES-E and GOES-W satellites for each variable. Therefore, we have no misaligned observations as both albedo assessments are available for the same set of locations for each source of data. In total, we have about 800,000 locations, each with the BHRiso and DHR30 observations. The objective for this case is to obtain the posterior predictive surfaces for each component while leveraging the additional knowledge of jointly modeling both sources of highly correlated information.

Proceeding similarly as what was done in the example for the state of Colorado, we initially fit and subtract linear fixed effects based on longitude and latitude. Next, the bivariate marginal model was implemented for the parsimonious Matérn cross-covariance function using the available optimization routine in R. Starting with default values where each parameter is 0.5, the runtime to obtain the point estimates was 1.65 hours. Table 6 summarizes the point estimates obtained from the bivariate marginal optimization routine.

After obtaining the inference for the cross-covariance function, we are able to sample the posterior predictive distribution for both components. The runtime for the predictive step was around 1.5 hours. It is worth noting that a non-negligible portion of the runtime is spent building the neighborhoods for the high resolution predictive grid. Figures 13 and Figure 14 show the posterior predictive means of each component after fitting the bivariate marginal model. As noted by the high correlation between the components, on a global scale, the two albedo assessments are very similar.

The two quantities we have studied, BHRiso and DHR30, are integral to defining how light is reflected by an opaque surface. BHRiso (white sky albedo) is the reflectance of the surface under

Parameter	Notation	$w(s)$	$d(s)$	Bivariate estimate
Degrees of freedom	α			1661086
Common Range	ν			1.2000
Partial Sill	σ_w^2, σ_d^2	0.2417	0.2256	-
Correlation	ρ_{wd}	-	-	0.9865
Nugget	ξ_w^2, ξ_d^2	0.0456	0.0280	-

Table 6: Point estimates for the covariance parameters obtained from bivariate marginal NN-RCM models
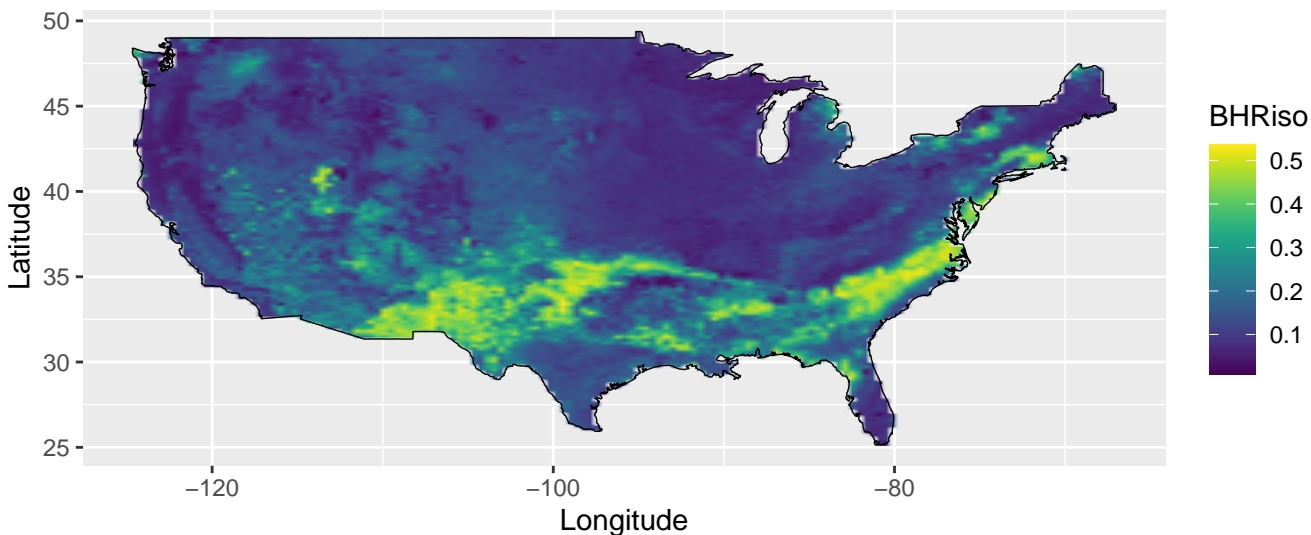

Figure 13: Predictions of BHRiso for CONUS from Bivariate Marginal NN-RCM model

diffuse illumination while DHR30 (black sky albedo) is the reflectance of the surface under direct illumination at a 30 degree angle. However, both of these quantities are merely theoretical, they do not quantify the true reflection of the land surface. The true blue sky albedo, α, may be calculated by a simple linear relationship of these terms as $\alpha = (1\text{-}D) \cdot \text{DHR30} + D \cdot \text{BHRiso}$ where the coefficient D is the proportion of diffuse illumination. D may be estimated on a pixel by pixel basis utilizing a radiative transfer simulation model driven by the bidirectional reflectance distribution function defined by BHRiso and DHR30 as well as covariates such as solar angles determined by pixel time and location and aerosol load characteristics. The output from the bivariate marginal model illustrated in this section will contribute to the accuracy of the downstream blue sky albedo assessment.
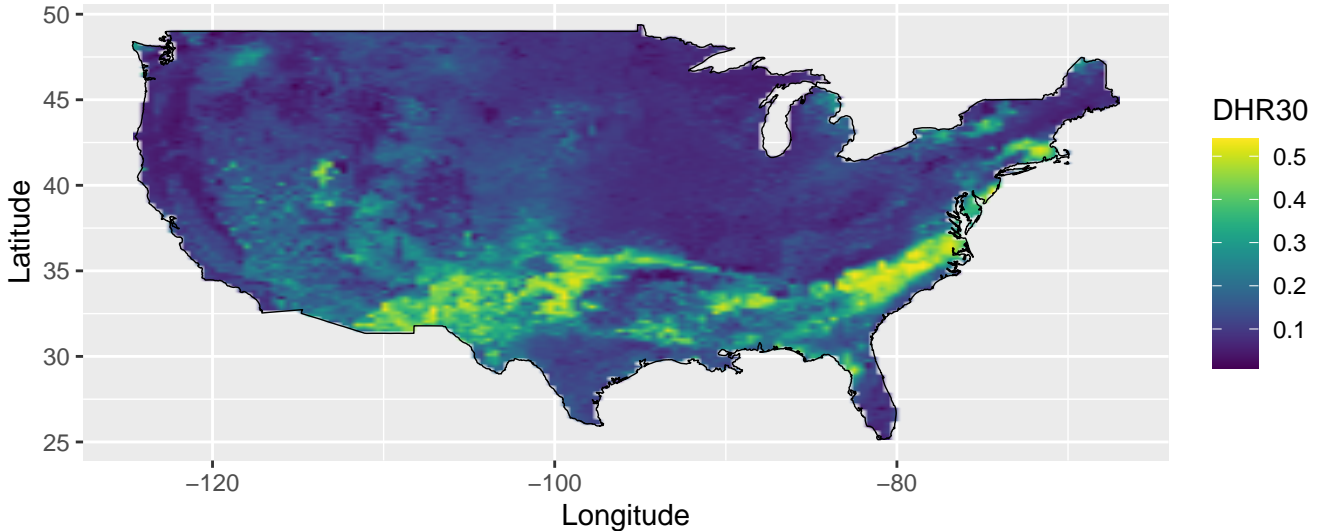
Figure 14: Predictions of DHR30 for CONUS from Bivariate Marginal NN-RCM model

5 Conclusions and future work

We have demonstrated the pertinence of a non-isotropic nearest-neighbor hierarchical spatial model and discussed its implementation for both univariate and multivariate observations. We presented simulated examples and real life cases to illustrate how the model was implemented.

Our main application of the NN-RCM considered surface albedo observations from NOAA's GOES East and West. The importance of surface albedo stems from its potential to display the rate of climate change. The objective of the study was to quantify the discrepancy between GOES-East and GOES-West and obtain a unified albedo field for the whole CONUS. While the two satellites seem similar on a global level, for smaller areas, subtle differences are visible. We used the bivariate hierarchical NN-RCM model to quantify such differences, using a common albedo surface $w(s)$ and a discrepancy surface denoted $d(s)$. We found that $d(s)$ is significant only for very localized areas. The conclusion is that the global variability of the albedo observations is too large to effectively quantify the discrepancy between the two satellites.

The next stage of our analysis will be to expand the model to the entirety of the data available from both satellites. This includes most of the Americas totalling over two millions observations each day. It is also of interest to understand if the time of year has an impact on the discrepancy between the satellites. While we focused on the month of July for CONUS, it will be interesting to see how surface albedo is impacted in the winter months.

The implementation of the bivariate model generates an important discussion on the handling of missing data and neighborhoods in multivariate nearest-neighbor methods. The misalignment of multivariate observations is a key issue for many datasets that had not yet been addressed. Similarly, the question of neighborhoods in multivariate models can be answered with multiple different options. Our implementation for the bivariate model concatenates two neighborhoods each based on one source of data.

Further advancement for the NN-RCM model include improving the distributed implementations. While the current development relies on parallel methods, divide-and-conquer algorithms present stronger opportunities for computational efficiency compared the one currently implemented. Distributed methods can not only reduce the computing time linearly, but also avoid redundancies when it comes to recurring data collection.

Acknowledgement

References

Banerjee, S. (2017). High-dimensional Bayesian geostatistics. *Bayesian Analysis 12*(2), 583–614.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2014). *Hierarchical Modeling and Analysis of Spatial Data* (second ed.). New York: Chapman and Hall.

Banerjee, S., A. E. Gelfand, A. O. Finley, and S. Huiyan (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(4), 825–848.

Berger, J. O., V. de Oliveira, and B. Sans (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association 96*(456), 1361–1374.

Brown, P. J., N. D. Le, and J. V. Zidek (1994a). Inference for a covariance matrix. In P. Freeman and A. Smith (Eds.), *Aspects of Uncertainty: A Tribute to D.V. Lindley*, pp. 77–92. Wiley.

Brown, P. J., N. D. Le, and J. V. Zidek (1994b). Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique 22*(4), 489–509.

Cressie, N. A. C. (1993). *Statistics for Spatial Data, Revised Edition*. New York: John Wiley and Sons.

Cressie, N. A. C. and G. Johannesson (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(1), 209–226.

Cressie, N. A. C. and C. K. Wikle (2011). *Statistics for Spatio-Temporal Data*. Hoboken, NJ: Wiley.

Datta, A., S. Banerjee, A. O. Finley, and A. E. Gelfand (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association 111*(514), 800–812.

Du, J., H. Zhang, and V. S. Mandrekar (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *The Annals of Statistics 37*(6A), 3330 – 3361.

Finley, A. O., S. Banerjee, and A. E. Gelfand (2015). spBayes for large univariate and multivariate point-referenced spatio-temporal data models. *Journal of Statistical Software 63*(13), 128.

Finley, A. O., A. Datta, and S. Banerjee (2017). *spNNGP: Spatial Regression Models for Large Datasets using Nearest Neighbor Gaussian Processes*. R package version 0.1.1.

Fuentes, M. (2001). A high frequency kriging approach for non-stationary environmental processes. *Environmetrics 12*(5), 469–483.

Fuentes, M. and R. Smith (2003, 02). A new class of nonstationary spatial models. *Journal of the American Statistical Association*.

Furrer, R., M. G. Genton, and D. Nychka (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics 15*(3), 502–523.

Gelfand, A. E., P. J. Diggle, M. Fuentes, and P. Guttorp (2010). *Handbook of Spatial Statistics*. Boca Raton, USA: Chapman and Hall.

Gelfand, A. E. and S. K. Ghosh (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika 85*(1), 1–11.

Genton, M. G. and W. Kleiber (2015). Cross-covariance functions for multivariate geostatistics. *Statistical Science 30*(2), 147 – 163.

Gneiting, T., W. Kleiber, and M. Schlather (2010). Matrn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association 105*(491), 1167–1177.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Govaerts, Y., A. Lattanzio, M. Taberner, and B. Pinty (2008, 06). Generating global surface albedo products from multiple geostationary satellites. *Remote Sensing of Environment - REMOTE SENS ENVIRON 112*, 2804–2816.

Gramacy, R. B. and H. K. H. Lee (2008). Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association 103*(483), 1119–1130.

Gu, M. and J. O. Berger (2016). Parallel partial Gaussian process emulation for computer models with massive output. *The Annals of Applied Statistics 10*(3), 1317 – 1347.

Guinness, J. (2018). Permutation and grouping methods for sharpening Gaussian process approximations. *Technometrics 60*(4), 415–429.

Heaton, M. J., A. Datta, A. O. Finley, R. Furrer, J. Guinness, R. Guhaniyogi, F. Gerber, R. B. Gramacy, D. Hammerling, M. Katzfuss, F. Lindgren, D. W. Nychka, F. Sun, and A. Zammit-Mangion (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics 24*(3), 398–425.

Higdon, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics 5*(2), 173–190.

Katzfuss, M. and N. A. C. Cressie (2011, 07). Spatiotemporal smoothing and EM estimation for massive remotesensing data sets. *Journal of Time Series Analysis 32*, 430–446.

Kaufman, C. G., M. J. Schervish, and D. W. Nychka (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association 103*(484), 1545–1555.

Kazianka, H. and J. Pilz (2012). Objective Bayesian analysis of spatial data with uncertain nugget and range parameters. *Canadian Journal of Statistics-revue Canadienne De Statistique 40*, 304–327.

Kidd, B. and M. Katzfuss (2022). Bayesian Nonstationary and Nonparametric Covariance Estimation for Large Spatial Data (with Discussion). *Bayesian Analysis 17*(1), 291 – 351.

Kim, H.-M., B. K. Mallick, and C. C. Holmes (2005). Analyzing nonstationary spatial data using piecewise Gaussian processes. *Journal of the American Statistical Association 100*(470), 653–668.

Kirsner, D. and B. Sansó (2020). Multi-scale shotgun stochastic search for large spatial datasets. *Comput. Stat. Data Anal. 146*, 106931.

Le, N. and J. Zidek (2006). *Statisticsal Analysis of Environmental Space-Time Processes.* Springer.

Lemos, R. T. and B. Sansó (2009). A spatio-temporal model for mean, anomaly, and trend fields of North Atlantic sea surface temperature. *Journal of the American Statistical Association 104*(485), 5–18.

NOAA (2016). GOES-R Series Mission. *https://www.nesdis.noaa.gov/GOES-R-Mission.*

NOAA (2018). Geostationary Satellite Server. *https://www.goes.noaa.gov/index.html.*

Paciorek, C. J. and M. J. Schervish (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics 17*(5), 483–506.

Sampson, P. D. and P. Guttorp (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association 87*(417), 108–119.

Schlather, M., A. Malinowski, P. J. Menck, M. Oesting, and K. Strokorb (2015). Analysis, simulation and prediction of multivariate random fields with package RandomFields. *Journal of Statistical Software 63*(8), 1–25.

Schlather, M., A. Malinowski, M. Oesting, D. Boecker, K. Strokorb, S. Engelke, J. Martini, F. Ballani, O. Moreva, J. Auel, P. J. Menck, S. Gross, U. Ober, P. Ribeiro, B. D. Ripley, R. Singleton, B. Pfaff, and R Core Team (2021). *RandomFields: Simulation and Analysis of Random Fields.* R package version 3.3.10.

Schmidt, A. M. and A. O'Hagan (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(3), 743–758.

Shaby, B. and D. Ruppert (2012). Tapered covariance: Bayesian estimation and asymptotics. *Journal of Computational and Graphical Statistics 21*(2), 433–452.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B (Methodological) 50*(2), 297–312.

Appendix A

As previously mentioned, the structure, reminiscent of multivariate linear regression models, allows us to derive the full conditionals for each observed location s_i independently. Therefore, in the implementation of the MCMC algorithm, the spatial field $w(s_i)$ can be sampled in a fully parallel fashion. The full conditional posterior distribution for s_i, where $i = 1, ..., k$ are obtained using the Bayesian normal linear model posterior inference formulas,

$$\boldsymbol{w}(s_i)|- \sim N_q(\boldsymbol{u}(s_i), S(s_i)),$$

where

$$\boldsymbol{u}(s_i) \;=\; S(s_i)\left(\Gamma'(s_i)\Phi^{-1}(s_i)\boldsymbol{w}_{N(s_i)} + A'\left(\boldsymbol{\tau}^2 I_q\right)^{-1}\boldsymbol{y}(s_i) + \sum_{u \in N^{-1}(s_i)} \Gamma'_{u,s_i}\Phi^{-1}(u)b_{u,s_i}\right)$$

$$S(s_i) \;=\; \left(\Phi^{-1}(s_i) + A'\left(\boldsymbol{\tau}^2 I_q\right)^{-1} A + \sum_{u \in N^{-1}(s_i)} \Gamma'_{u,s_i}\Phi^{-1}(u)\Gamma_{u,s_i}\right)^{-1},$$

where the set $u \in N^{-1}(s_i)$ is the group of locations which has s_i as a neighbor. The notation Γ_{u,s_i} indicates the rows of $\Gamma(u)$ that corresponds to the neighbors located at s_i and $b_{u,s_i} = \boldsymbol{w}(u) - \sum_{r \in N(u)\backslash s_i} \Gamma'(u,r)\boldsymbol{w}_r$.

The full conditionals for the spatial random effects $\Gamma(s_i)$ and $\Phi(s_i)$ can also be obtained for each location separately. Similar to the spatial field, we leverage the Bayesian linear model posterior inference equations to obtained the closed-form full conditional distributions.

$$\Gamma(s_i)|- \;\sim\; N_{mq,q}(U(s_i), V(s_i), \Phi(s_i)),$$
$$U(s_i) \;=\; V(s_i)\left(\boldsymbol{w}_{N(s_i)}\boldsymbol{w}(s_i) + (\alpha - kq - 1)\mathcal{C}_{\boldsymbol{\theta},N(s_i),s_i}\right)$$
$$V(s_i) \;=\; \left(\boldsymbol{w}_{N(s_i)}\boldsymbol{w}'_{N(s_i)} + (\alpha - kq - 1)\mathcal{C}_{\boldsymbol{\theta},N(s_i)}\right)^{-1}$$
$$\Phi(s_i)|- \;\sim\; IW(\alpha - kq + 2q(1 + m), (\alpha - kq - 1)\mathcal{C}_{\boldsymbol{\theta},s_i|N(s_i)} + Q(s_i) + R(s_i)$$
$$Q(s_i) \;=\; (\boldsymbol{w}(s_i) - \Gamma'(s_i)\boldsymbol{w}_N(s_i))(\boldsymbol{w}(s_i) - \Gamma'(s_i)\boldsymbol{w}_N(s_i))'$$
$$R(s_i) \;=\; (\alpha - kq - 1)(\Gamma(s_i) - \mathcal{C}^{-1}_{\boldsymbol{\theta},N(s_i)}\mathcal{C}_{\boldsymbol{\theta},N(s_i),s})'\mathcal{C}_{\boldsymbol{\theta},N(s_i)}(\Gamma(s_i) - \mathcal{C}^{-1}_{\boldsymbol{\theta},N(s_i)}\mathcal{C}_{\boldsymbol{\theta},N(s_i),s}).$$

Finally, for each sampling iteration of the MCMC algorithm, we have global full conditionals for each element τ^2_q of the observational error $\boldsymbol{\tau}^2$. The likelihood and the choice of prior for the observational errors in the hierarchical structure leads to conditional conjugacy. Therefore, the full conditional of τ^2_q is also an inverse gamma distribution,

$$\tau^2_q|- \sim IG\left(a_\tau + \frac{k}{2}, b_\tau + \sum_{s_i \in S} \frac{(w_q(s_i) - y_q(s_i))^2}{2}\right).$$

Special case: univariate observations full conditionals

We again call attention to the special case of univariate spatial observations. Setting $q = 1$, the model simplifies to ,

$$y(s) = (s)\boldsymbol{\beta} + w(s) + \epsilon(s),$$

where

$$y(s)|\boldsymbol{\beta}, w(s), \tau^2 \sim N((s)\boldsymbol{\beta} + w(s), \tau^2)$$
$$\boldsymbol{\beta} \sim N_p(0, s_\beta^2 I)$$
$$w(s)|\gamma(s), \phi(s), \sigma^2 \sim N(\gamma(s)\boldsymbol{w}_{N(s)}, \sigma^2\phi(s))$$
$$\gamma(s)|\phi(s) \sim N_m(\mathcal{C}_{\boldsymbol{\theta},N(s)}^{-1} C_{\boldsymbol{\theta},N(s),s}, \frac{\phi(s)}{\alpha - k - 1}\mathcal{C}_{\boldsymbol{\theta},N(s)}^{-1})$$
$$\phi(s) \sim IG(\alpha - k + (1 + m), (\alpha - k - 1)\mathcal{C}_{\boldsymbol{\theta},s|N(s)}),$$
$$\pi(\sigma^2, \tau^2) = IG(\tau^2|a_\tau, b_\tau) \times IG(\sigma^2|a_\sigma, b_\sigma),$$

where p is the dimension of the fixed effects and m is the number of neighbors. We recognize the framework of a multivariate linear regression model in the prior structure of the spatial random effects $w(s)$. By using the posterior equations from the Normal-Normal model, we obtain a closed-form solution for the full conditional distribution of $w(s)$:

$$w(s_i)|- \sim N(\mu(s_i), \sigma^2(s_i)),$$

where

$$\mu(s_i) = \left(\frac{1}{\sigma^2\phi(s_i)} + \frac{1}{\tau^2}\right)^{-1}\left(\frac{\boldsymbol{\gamma}'(s_i)\boldsymbol{w}_{N(s_i)}}{\sigma^2\phi(s_i)} + \frac{y(s_i)}{\tau^2} + \sum_{u \in N^{-1}(s_i)}\frac{\gamma_{u,s_i}}{\sigma^2\phi(u)}b_{u,s_i}\right)$$

$$\sigma^2(s_i) = \left(\frac{1}{\sigma^2\phi(s_i)} + \frac{1}{\tau^2} + \sum_{u \in N^{-1}(s_i)}\frac{\gamma_{u,s_i}^2}{\sigma^2\phi(u)}\right)^{-1},$$

where the notation γ_{u,s_i} indicates the element of $\gamma(u)$ that corresponds to the neighbors located at s_i and $b_{u,s_i} = w(u) - \sum_{r \in N(u)\setminus s_i}\gamma_{u,r}w(r)$.

To obtain the full conditional distributions of the random effects $\boldsymbol{\gamma}(s_i)$ and $\phi(s_i)$, we use the normal equations from the linear models posterior inference.

$$\boldsymbol{\gamma}(s_i)|- \sim N_m(\boldsymbol{\mu}_{\boldsymbol{\gamma}(s_i)}, h_{\boldsymbol{\gamma}(s_i)}^2),$$
$$\boldsymbol{\mu}_{\boldsymbol{\gamma}(s_i)} = \frac{h_{\boldsymbol{\gamma}(s_i)}^2}{\phi(s_i)}\left(\frac{\boldsymbol{w}_{N(s_i)}w(s_i)}{\sigma^2} + (\alpha - k - 1)C_{\boldsymbol{\theta},N(s_i),s_i}\right)$$
$$h_{\boldsymbol{\gamma}(s_i)}^2 = \phi(s_i)\left(\frac{\boldsymbol{w}_{N(s_i)}\boldsymbol{w}'_{N(s_i)}}{\sigma^2} + (\alpha - k - 1)C_{\boldsymbol{\theta},N(s_i)}\right)^{-1}$$
$$\phi(s_i)|- \sim IG\left(\alpha - k + 1 + m + \frac{m+1}{2}, (\alpha - k - 1)C_{\boldsymbol{\theta},s_i|N(s_i)} + \frac{q(s_i)}{2\sigma^2} + \frac{r(s_i)}{2}\right)$$
$$q(s_i) = (w(s_i) - \boldsymbol{\gamma}'(s_i)\boldsymbol{w}_N(s_i))^2$$
$$r(s_i) = (\alpha - k - 1)(\boldsymbol{\gamma}(s_i) - C_{\boldsymbol{\theta},N(s_i)}^{-1}C_{\boldsymbol{\theta},N(s_i),s})'C_{\boldsymbol{\theta},N(s_i)}(\boldsymbol{\gamma}(s_i) - C_{\boldsymbol{\theta},N(s_i)}^{-1}C_{\boldsymbol{\theta},N(s_i),s})$$

Finally, we are left with the posterior inference of the observational error τ^2 and the partial sill σ^2. In contrast to the multivariate model, the partial sill is factored out of the covariance matrix. In the previously defined model, we therefore have

$$\sigma^2 C_{\boldsymbol{\theta}} = \sigma^2(C_\nu + \xi^2 I).$$

This allows us to learn the full posterior distribution for the partial sill under the hierarchical NN-RCM model. The choice of prior for the observational error and partial sill both lead to closed-form full conditional distributions:

$$\sigma^2|- \sim IG\left(a_\sigma + \frac{k}{2}, b_\sigma + \sum_{s_i \in S} \frac{(w(s_i) - \boldsymbol{\gamma}'(s_i)\boldsymbol{w}_N(s_i))^2}{2\phi(s_i)}\right)$$

$$\tau^2|- \sim IG\left(a_\tau + \frac{k}{2}, b_\tau + \sum_{s_i \in S} \frac{(w(s_i) - y(s_i))^2}{2}\right).$$