Statistica Sinica Preprint No: SS-2017-0536	
Title	Efficient Estimation of Non-stationary Spatial Covariance Functions with Application to High-resolution Climate Model Emulation
Manuscript ID	SS-2017-0536
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0536
Complete List of Authors	Yuxiao Li and Ying Sun
Corresponding Author	Ying Sun
E-mail	ying.sun@kaust.edu.sa
Notice: Accepted version subject to English editing.	

Efficient Estimation of Non-stationary Spatial Covariance Functions with Application to High-resolution Climate Model Emulation

Yuxiao Li and Ying Sun

King Abdullah University of Science and Technology

Abstract: Spatial processes exhibit non-stationarity in many climate and environmental applications. Convolution-based approaches are often used to construct non-stationary covariance functions in Gaussian processes. Although convolution-based models are flexible, their computation is extremely expensive when the dataset is large. Most existing methods rely on fitting an anisotropic but stationary model locally and reconstructing the spatially varying parameters. In this study, we propose a new estimation procedure to approximate a class of non-stationary Matérn covariance functions by local-polynomial fitting of the covariance parameters. The proposed method allows for efficient estimation of a richer class of non-stationary covariance functions with the local stationary model as a special case. We also develop an approach for fast high-resolution simulation with non-stationary features on the small scale and apply it to precipitation data in climate model.

Key words and phrases: Climate model runs, conditional simulation, large datasets, local likelihood estimation, non-stationary Matérn covariance function, polynomial approximation

1. Introduction

Gaussian random fields (GRFs) or Gaussian processes (GPs) for spatial data have been called "the most valuable tools in the toolkit for geostatistical modeling" (Gelfand and Schliep, 2016). These tools are important because the probabilistic distribution of a GRF can be fully determined by its first and second moments. For convenience, GRFs are typically assumed to be stationary or isotropic, which implies that the second moment is finite, the mean function is constant and certain property of the covariance function is invariant. Nevertheless, non-stationarity often exists in spatial processes, such as between the land and the ocean, or between mountains and plains. Generally, the non-stationarity exists in the mean and the covariance function. In this study, we assume that the mean function is constant in order to focus on non-stationarity in the covariance function.

The existing literature has provided various approaches to modeling non-stationary covariance functions. We classify approaches into six groups: 1) basis function expansions (Nychka et al., 2002); 2) deformation approaches (Sampson and Guttorp, 1992; Anderes and Stein, 2008); 3) differential operator approaches (Jun and Stein, 2008; Lindgren et al., 2011); 4) process convolution approaches (Higdon, 1998; Paciorek and Schervish, 2006); 5) predictive processes (Gramacy and Lee, 2008); and 6) treed Gaus-

sian processes (Banerjee et al., 2008). Risser (2016) reviewed most of these methodologies with emphasis on convolution-based methods. The convolution-based model has become one of the most popular methods in spatial non-stationary modeling due to its high flexibility. In particular, Paciorek and Schervish (2006) derived a class of valid non-stationary Matérn covariance functions with convolution models. They characterized non-stationarity by spatially varying parameters, which could be viewed as a function over space.

However, if we specify a non-stationary type using spatially varying parameters, the number of parameters to be estimated will be proportional to the size of the locations. Accuracy and efficiency in such an estimation are extremely difficult to achieve. To address this problem, state-of-art methods assume that the covariance function is local stationary (Paciorek and Schervish, 2006; Anderes and Stein, 2011) or weighted local stationary (Risser and Calder, 2015a; Fouedjio et al., 2016). Although these approximations simplified the model fitting in many applications, all of them rely on the assumption that the spatially varying parameters are locally constant, or the process is stationary in each subregion. Therefore, the approximation quality is largely affected by the choice of region partition, which is not an easy task. In general, an accurate local fitting requires a reasonably

large size of observations which are often collected over a large, and thus, possibly non-stationary spatial region. A smaller size of the subregion tends to be more stationary, however, there might not be enough data to fit the model locally.

To avoid this dilemma, we propose a local estimation approach for non-stationary covariance functions by higher-order polynomial approximation of the spatially varying parameters. Compared to the local stationary model, our higher-order polynomial approximation allows for local non-stationarity and incorporates local stationary and weighted local stationary models as special cases. We also develop an efficient likelihood-based estimation method for model fitting. Additionally, we show that each polynomial parameter measures the degree of a certain type of non-stationarity.

Fitted non-stationary models have many potential applications. We apply the proposed method to precipitation data in North America from relatively coarse-resolution climate model runs. With the fitted model, we perform high-resolution simulations and generate non-stationary precipitation fields. High-resolution non-stationary simulation is also challenging. Kleiber (2016) proposed to combine the approaches of circular embedding and deformation for exact simulation. We propose a different approach using sequentially conditional simulation. Using a personal computer, it

takes approximately 15 minutes in total to fit our model using 5 runs of the general circulation models (GCM) data from $13 \times 30 = 390$ locations and perform one simulation at the scale of a regional climate model (RCM) at $62 \times 210 = 13,020$ locations.

The remainder of our paper is organized as follows. Section 2 introduces the convolution-based non-stationary spatial covariance modeling and reviews existing estimation methods. Section 3 proposes our new estimation approaches. Sections 4 describes the interpretations, computational issues and implied simulation algorithms. Section 5 presents simulation studies to fit non-stationary Matérn covariance functions. Section 6 applies our estimation and high-resolution emulation approaches to precipitation data from climate model runs. Section 7 summarizes our main results and suggests some directions for future work.

2. Non-stationary Covariance Function and Its Estimation

2.1. Non-stationary Spatial Covariance Function

A univariate Gaussian random field (GRF), $\{Z(\mathbf{s}), \mathbf{s} \in D\}$, defined on $D \subset \mathbb{R}^d$, can be specified as

$$Z(\mathbf{s}) = m(\mathbf{s}) + Y(\mathbf{s}) + \varepsilon(\mathbf{s}), \ \mathbf{s} \in D, \tag{2.1}$$

where $m(\cdot)$ is the mean function, $Y(\cdot)$ is a spatially dependent and zero-mean GRF with covariance function $C(\cdot, \cdot)$, and $\varepsilon(\cdot) \sim N(0, \tau^2(\cdot))$ is the

nugget effect caused by measurement inaccuracy and environmental variability. Moreover, $m(\cdot)$ is assumed to be a constant for simplicity, $C(\cdot,\cdot) = C(\cdot,\cdot;\boldsymbol{\theta}_0)$ has a parametric form with $\boldsymbol{\theta}_0$ in $\mathbb{R}^d$, and $\varepsilon(\cdot)$ and $Y(\cdot)$ are independent.

Equation (2.1) is a general representation that allows for non-stationarity. Let $\boldsymbol{\theta} = \{\boldsymbol{\theta}_0, \boldsymbol{\tau}\}$ be the vector of all unknown parameters to be estimated and $C(\cdot,\cdot;\boldsymbol{\theta})$ be the covariance function that incorporates the nugget effect. When assuming an isotropic covariance function, $C^I(\cdot;\boldsymbol{\theta})$, various classes (exponential, Gaussian, Matérn, Cauchy, etc) of covariance models are available in spatial statistics (Cressie, 2015). The Matérn covariance function is the most popular among these covariance functions. However, $C(\cdot,\cdot;\boldsymbol{\theta})$, as a valid non-stationary covariance function, may not have a closed form in general and may involve many spatially varying parameters that require estimation. Based on kernel convolution (Higdon, 1998), Paciorek and Schervish (2004), Stein (2005) and Paciorek and Schervish (2006) provided a rich class of valid parametric non-stationary Matérn covariance functions on $\mathbb{R}^d$:

$$C^{\mathrm{NS}}(\mathbf{s}_i,\mathbf{s}_j;\boldsymbol{\theta}) = \tau(\mathbf{s}_i)\mathbb{1}_{[i=j]}(\mathbf{s}_i,\mathbf{s}_j) + \sigma(\mathbf{s}_i)\sigma(\mathbf{s}_j)|\Sigma(\mathbf{s}_i)|^{\frac{1}{4}}|\Sigma(\mathbf{s}_j)|^{\frac{1}{4}}\times \left|\frac{\Sigma(\mathbf{s}_i)+\Sigma(\mathbf{s}_j)}{2}\right|^{-\frac{1}{2}} \left(2\sqrt{\nu Q_{ij}}\right)^{\nu} K_\nu\left(2\sqrt{\nu Q_{ij}}\right), \tag{2.2}$$

where $\boldsymbol{\theta}$ is the vector of unknown parameters, $\sigma(\mathbf{s}_i)$ is the spatially vary-

ing standard deviation (squared root of the partial sill), $\Sigma(\mathbf{s}_i)$ is the $d \times d$ kernel matrix at $\mathbf{s}_i$, which controls the spatially varying local anisotropy (including the spatial range and the direction of anisotropy), $K_\nu(\cdot)$ is the modified Bessel function of the second kind, ν is the smoothness parameter, and $Q_{ij} = (\mathbf{s}_i - \mathbf{s}_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1} (\mathbf{s}_i - \mathbf{s}_j)$ is the Mahalanobis distance between $\mathbf{s}_i$ and $\mathbf{s}_j$. The smoothness parameter can be spatially varying as well. For example, the model proposed by Stein (2005) allows for non-stationarity in the smoothness by letting $\nu_{ij} = \frac{\nu(\mathbf{s}_i)+\nu(\mathbf{s}_j)}{2}$. The covariance functions in Equation (2.2) were derived from kernel convolution models (Higdon, 1998). The non-stationarity is controlled by the spatially varying parameters, $\boldsymbol{\theta}(\mathbf{s}_i) = \{\Sigma(\mathbf{s}_i), \sigma(\mathbf{s}_i), \tau(\mathbf{s}_i), \nu(\mathbf{s}_i)\}$. Usually, kernel matrices are obtained through spectral decomposition. For example, in the case of $d = 2$,

$$\Sigma(\mathbf{s}_i) = \Sigma_i = \begin{bmatrix} \cos(\phi_i) & -\sin(\phi_i) \\ \sin(\phi_i) & \cos(\phi_i) \end{bmatrix} \begin{bmatrix} \lambda_{1i} & 0 \\ 0 & \lambda_{2i} \end{bmatrix} \begin{bmatrix} \cos(\phi_i) & \sin(\phi_i) \\ -\sin(\phi_i) & \cos(\phi_i) \end{bmatrix}, \quad (2.3)$$

where $\lambda_{1i},\ \lambda_{2i} > 0$ are the eigenvalues that represent spatial ranges and $\phi_i \in (0, \pi/2)$ represents the angle of rotation.

Specifically, a GRF is stationary if and only if all of the spatially varying parameters in Equation (2.2) are constant, and the isotropic assumption holds if further the kernel matrix is a scalar matrix. In principle, all of the

parameters could vary spatially, suggesting that there are different types of non-stationarity. However, allowing too many types of non-stationarity is not feasible in practice due to possible optimization and model identifiability issues (Anderes and Stein, 2011).

2.2. Likelihood Based Covariance Estimation

The maximum likelihood method is appropriate for estimation of a GRF model with a specified parametric non-stationary covariance function. Let $Z(\mathbf{s})$ be a GRF with mean $m(\mathbf{s})$ and covariance function $C^{\text{NS}}(\cdot,\cdot;\boldsymbol{\theta})$ and assume that $m(\mathbf{s}) = 0$ for simplicity. Suppose that we observe $Z(\mathbf{s})$ at n locations, $\mathbf{s}_1,\ldots,\mathbf{s}_n$ (the locations could be regularly or irregularly spaced). The random vector $\mathbf{Z} = \{Z(\mathbf{s}_1),\ldots,Z(\mathbf{s}_n)\}^T$ follows an n-variate Gaussian distribution, i.e. $\mathbf{Z} \sim N_n(\mathbf{0}, \Sigma^{\text{NS}}_{n\times n})$, where the (i,j)-th element of the non-stationary covariance matrix, Σ^{NS}, is $C^{\text{NS}}(\mathbf{s}_i,\mathbf{s}_j;\boldsymbol{\theta}(i,j)), i,j = 1,\ldots,n$. As a result, the corresponding log-likelihood function is

$$\ell(\boldsymbol{\theta};\mathbf{Z}) = -\frac{1}{2}\log|\Sigma^{\text{NS}}| - \frac{1}{2}\mathbf{Z}^T(\Sigma^{\text{NS}})^{-1}\mathbf{Z} - \frac{n}{2}\log(2\pi). \tag{2.4}$$

By maximizing the likelihood function in Equation (2.4), we can get the maximum likelihood estimates (MLEs), $\hat{\boldsymbol{\theta}}$. Note that the mean function is assumed to be zero. If not, a restricted maximum likelihood (REML) method is needed (Patterson and Thompson, 1975; Risser and Calder, 2015a). In most of environmental studies, only one replicate is available.

When independent replicates are obtained, Equation (2.4) can be simply modified by summing the likelihoods for the replicates. However, the main difficulty is that the parameter vector is spatially varying. Paciorek and Schervish (2006) first provided a solution by assuming that each non-stationary parameter varies slowly and regularly over space, i.e., $\theta_i \approx \theta_{R(i)}$, where $R(i)$ indicates that the location $\mathbf{s}_i$ falls into the subregion $R(i)$, and $\theta_i = \theta(\mathbf{s}_i)$, where $\theta(\mathbf{s}_i)$ indicates any non-stationary parameter. By fitting a stationary model in each of m subregions, typically $m \ll n$, they obtain $\hat{\theta}_{R_{(1)}}, \ldots, \hat{\theta}_{R_{(m)}}$. Although the computational burden of this estimation is significantly reduced, this result is not desirable, since both $\theta(\cdot)$ and the fitted surface are discontinuous when they fit spatial models to each subregion.

To overcome this problem, several groups recently proposed methods (Risser and Calder, 2015a; Fouedjio et al., 2016; Li and Zhu, 2016) to smooth the local stationary estimates. For instance, Risser and Calder (2015a) and Fouedjio et al. (2016) made $\theta(\cdot)$ a discrete mixture of the parameters at representative locations, or anchor locations as Fouedjio et al. (2016) called them, and the final estimate of $\theta(\mathbf{s}_i)$ is a smoothly weighted

estimate. That is,

$$\widehat{\theta}(\mathbf{s}_i) = \sum_{k=1}^{m} w(\mathbf{s}_i, \mathbf{s}_k)\widehat{\theta}_k, \qquad w(\mathbf{s}_i, \mathbf{s}_k) = \frac{K(\mathbf{s}_i, \mathbf{s}_k)}{\sum_{k=1}^{m} K(\mathbf{s}_i, \mathbf{s}_k)}, \tag{2.5}$$

where m is the number of subregions, $\widehat{\theta}_k$ is the local stationary estimator in the k-th subregion with anchor location $\mathbf{s}_k$, $w(\mathbf{s}_i, \mathbf{s}_k)$ is the weight function, $K(\cdot)$ denotes a d-variate kernel function. The Gaussian kernel, $K(\mathbf{s}_i, \mathbf{s}_k) = \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_k\|^2}{2h}\right)$, is the most commonly used kernel function, where $h > 0$ is the bandwidth parameter.

Another popular approach to estimate θ_i is by moving-window: fit a stationary model locally in a small neighborhood of one location, $\mathbf{s}_i$, and then move to the next location in a new neighborhood. In this approach, only locations in the neighborhood of $\mathbf{s}_i$ are used to estimate θ_i, whereas the information from locations far away from $\mathbf{s}_i$ is ignored. Anderes and Stein (2011) then improved the moving-window method by maximizing a weighted local likelihood function that smoothly down-weights faraway locations. Although moving-window methods improve the estimation in some sense, they are extremely difficult to fit when the dataset is large.

These approaches are all based on a local stationary or weighted local stationary assumption, implying that parameter estimates are constant in predefined subregions or neighborhoods. For the sake of both model flexibility and inference efficiency, we therefore extend the local stationary

approach and propose a model that allows for local non-stationarity, by approximating the spatially varying parameters with a local polynomial.

3. New Covariance Estimation Approaches

3.1. Local Polynomial Approximation of Covariance Parameters

Suppose that $\theta(\mathbf{s})$ is continuously differentiable with the α-th differentiation denoted by $D^\alpha\theta$. For a given location, $\mathbf{s_0} = (s_{01}, \ldots, s_{0d}) \in D$, Taylor expansion allows $\theta(\mathbf{s})$ to be expanded at $\mathbf{s_0}$ as follows (Königsberger, 2013):

$$\theta(\mathbf{s}) = \sum_{|\alpha|\leq k} \frac{D^\alpha\theta(\mathbf{s_0})}{\alpha!}(\mathbf{s}-\mathbf{s_0})^\alpha + \sum_{|\alpha|=k} h_\alpha(\mathbf{s})(\mathbf{s}-\mathbf{s_0})^\alpha, \tag{3.1}$$

where $\lim_{\mathbf{s}\to\mathbf{s_0}} h_\alpha(\mathbf{s}) = 0$. In particular, for any location $\mathbf{s} = (s_1, s_2) \in \mathbb{R}^2$, the first-order Taylor approximation of $\theta(\cdot)$ at $\mathbf{s}_0 = (s_{01}, s_{02})$ is

$$\theta(s_1, s_2) \approx \beta_0 + \beta_1(s_1 - s_{01}) + \beta_2(s_2 - s_{02}), \tag{3.2}$$

where $\beta_0 = \theta(s_{01}, s_{02})$, $\beta_1 = \frac{\partial}{\partial s_1}\theta(s_{01}, s_{02})$, $\beta_2 = \frac{\partial}{\partial s_2}\theta(s_{01}, s_{02})$, and the remainder is bounded by $\frac{M}{2}(|\mathbf{s}_1 - \mathbf{s}_{01}| + |\mathbf{s}_2 - \mathbf{s}_{02}|)^2$, where M is the upper bound of all second-order partial derivatives of $\theta(\cdot)$.

Equation (3.2) includes local stationarity as a special case. If $\beta_1 = \beta_2 = 0$, then $\theta(s_1, s_2) \approx \theta(s_{01}, s_{02}) = \beta_0$ is a constant in a small neighborhood centered at $\mathbf{s}_0$, which implies local stationarity and the estimates of β_0 can be obtained by using the information in the neighborhood of $\mathbf{s}_0$. This type of estimator defined in (2.5) is called a locally constant estimator or a Nadaraya-Watson estimator for nonparametric fitting. It can be viewed as

an estimator associated with order $\alpha = 0$ in Equation (3.1). If we assume that $\alpha = 1$, the local linear approximation implies that $\theta(\cdot)$ varies linearly. Therefore, by choosing a higher-order, we can model more complex non-stationarity in $\theta(\cdot)$.

Therefore, based on Equation (3.2), we propose a natural non-stationary estimator of $\theta(\mathbf{s}_i)$ at a location, $\mathbf{s}_i = (s_{i_1}, \ldots, s_{i_d})$, as an extension of Equation (2.5). For $\theta(\mathbf{s}_i)$ of the form

$$\theta(\mathbf{s}_i) = \sum_{k=1}^{m} w(\mathbf{s}_i, \mathbf{s}_k) \left\{ \beta_{0k} + \sum_{q=1}^{d} \beta_{qk}(s_{i_q} - s_{k_q}) \right\}, \tag{3.3}$$

we estimate it by plugging in the MLEs of all the β_{qk}'s, $q = 0, \ldots, d$, $k = 1, \ldots, m$.

The final estimation of $\theta(\mathbf{s}_i)$ is approximated by kernel smoothing, in order to smooth the estimations with m different linear trends. The smoothing method is not restricted, thus many common approaches are viable, such as smoothing spline and Kriging.

In principle, this model can be fitted to any order polynomial approximation with other covariates included. However, for many real-world environmental applications, the first-order model is usually sufficient since it generally requires a large amount of local data to make inference on more complicated non-stationarity. Consequently, we consider only the first-order non-stationarity and develop statistical inference methods.

3.2. Estimation Procedures for Non-stationary Kernel Matrix

In this subsection, we describe an algorithm for estimating the two dimensional (2D) spatially varying kernel matrix, $\Sigma(\mathbf{s}_i)$, characterized by three parameters, $\boldsymbol{\theta}_{ns}(\mathbf{s}) = \{\lambda_1(\mathbf{s}), \lambda_2(\mathbf{s}), \phi(\mathbf{s})\}$, as in Equation (2.3). Other parameters are set to be stationary (constant) and denoted by $\boldsymbol{\theta}_s$.

To estimate $\boldsymbol{\theta}_{ns}(\mathbf{s})$, we need to maximize the full likelihood as in Equation (2.4) with Σ^{NS} specified via Equations (2.2) and (3.3). To improve the computational efficiency, we propose an independent likelihood estimation method by assuming subregions are independent with each other, where $\boldsymbol{\beta}_k, k = 1, \ldots, m$, for each subregion is estimated independently. The independent log-likelihood is of the form

$$\tilde{\ell}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m; \mathbf{Z}) = -\frac{1}{2}\sum_{k=1}^{m} \log|\Sigma_k^{\text{NS}}| - \frac{1}{2}\sum_{k=1}^{m} \mathbf{Z}_k^T (\Sigma_k^{\text{NS}})^{-1} \mathbf{Z}_k - \frac{n}{2}\log(2\pi), \quad (3.4)$$

where Σ_k^{NS} parameterized by $\boldsymbol{\beta}_k$ is the covariance matrix and $\mathbf{Z}_k$ is the data vector corresponding to the locations in the k-th subregion, respectively.

The independent likelihood method significantly improves the computational efficiency, because the local fitting involve fewer parameters and the computation can be easily parallelized. However, the independent likelihood method may be less accurate at boundary locations.

For comparisons with other methods, we call the local stationary and weighted local stationary estimators S0 and WS0 estimators, respectively,

since they correspond to constant (zero-order) approximations, and we call our estimator the first-order non-stationary (NS1) estimator. The estimation procedure is as follows:

1) Divide the region into m overlapping or non-overlapping subregions. Denote the partial random vector including all locations in the k-th subregion to be $\mathbf{Z}_k, k = 1, \ldots, m$.

2) Select the anchor locations, $\mathbf{s}_1, \ldots, \mathbf{s}_m$, corresponding to the m subregions, and fit a local stationary model to obtain the stationary and local stationary (S0) estimators $\boldsymbol{\theta}_s$ and $\boldsymbol{\beta}_{0k} = (\beta_{0k}^{\lambda_1}, \beta_{0k}^{\lambda_2}, \beta_{0k}^{\phi})$, $k = 1, \ldots, m$, by maximizing the independent likelihood $\tilde{\ell}(\boldsymbol{\beta}_{01}, \ldots, \boldsymbol{\beta}_{0m}, \boldsymbol{\theta}_s; \mathbf{Z})$.

3) Plug the local stationary (S0) estimators into Equation (3.3) for each spatially varying parameter. For example, denote $\lambda_{1k}(\mathbf{s}_i) = \hat{\beta}_{0k}^{\lambda_1} + \sum_{q=1}^{2} \beta_{qk}^{\lambda_1}(s_{i_q} - s_{k_q})$. Then estimate the slope parameters, $\boldsymbol{\beta}_{1k} = (\beta_{1k}^{\lambda_1}, \beta_{1k}^{\lambda_2}, \beta_{1k}^{\phi})$ and $\boldsymbol{\beta}_{2k} = (\beta_{2k}^{\lambda_1}, \beta_{2k}^{\lambda_2}, \beta_{2k}^{\phi})$, $k = 1, \ldots, m$, by maximizing the independent likelihood $\tilde{\ell}(\boldsymbol{\beta}_{11}, \ldots, \boldsymbol{\beta}_{1m}, \boldsymbol{\beta}_{21}, \ldots, \boldsymbol{\beta}_{2m}; \mathbf{Z}, \hat{\boldsymbol{\beta}}_{01}, \ldots, \hat{\boldsymbol{\beta}}_{0m}, \hat{\boldsymbol{\theta}}_s)$.

4) Smooth the estimators in steps 2 and 3 based on Equation (3.3). For example, $\hat{\lambda}_1(\mathbf{s}_i) = \sum_{k=1}^{m} w(\mathbf{s}_i, \mathbf{s}_k)\{\hat{\beta}_{0k}^{\lambda_1} + \sum_{q=1}^{2} \hat{\beta}_{qk}^{\lambda_1}(s_{i_q} - s_{k_q})\}$. Finally, the first order non-stationary (NS1) estimator of kernel matrix is the approximate estimations of all nonstationary parameters $\hat{\boldsymbol{\theta}}_{ns}(\mathbf{s}_i)$ at each location $\mathbf{s}_i$.

4. Interpretations and Computations

4.1. Computational issues

With 2D data, six parameters can exhibit non-stationarity at a given anchor location: three in the kernel matrix, one in the partial sill, one in the smoothness, and one in the nugget effect. For both full likelihood and independent likelihood methods, if we allow all of the six parameters to have first-order non-stationarity, then the number of parameters to be estimated is $3 \times 6 \times m$. Usually, $m \ll n$. The number of parameters to be estimated therefore increases with the size of anchor locations m and is less than n.

Obviously, the full likelihood method has better accuracy than independent likelihood method, especially near the boundary between subregions. However, this could lead to optimization issues since too many parameters need to be simultaneously optimized. Using independent likelihood approach, the parameters can be optimized in parallel and locally. Instead of solving a high-dimensional problem, we solve several lower-dimensional problems. Although our method may not reach the global maxima, the optimization is more stable.

Step 2 involves the estimation of both stationary and nonstationary parameters. The maximization is not trivial when the stationary parameters are unknown, but it can be well fitted by `NSconvofit` function in the

R package `convoSPAT`, and the details of their algorithms can be found in Risser and Calder (2015a).

The weight function depends on the bandwidth, h. As our method relies on the `NSconvofit` function, we therefore choose the same default bandwidth as in the function, which is the square of half of the distance between the two closest anchor locations, i.e., $h = \{\|\mathbf{s}_1 - \mathbf{s}_2\|/2\}^2$. As a result, the local estimates depend only on the data in the subregions with range $2\sqrt{h}$, indicating the range of high density points in a Gaussian distribution. However, more choices can be made by cross validation, which are discussed in the Supplement (Section S1.1).

The optimization are made by `optim` function in R software. With the approximated Hessian matrix, we quantify the uncertainty of our estimations by asymptotic standard errors and corresponding confidence intervals.

4.2. Covariates and degree of nonstationarity

When the same kernel function, $K(\cdot)$, and bandwidth parameter, h, as defined in Equation (2.5) are selected, it is easy to see that $\widehat{\theta}(\mathbf{s}_i)$ in Equation (3.3) is a weighted local stationary estimator when $\beta_{1k}, \ldots, \beta_{dk} = 0$. Moreover, in the k-th subregion, $\boldsymbol{\beta}$ can be viewed as the regression parameters of the covariate 1 for $q = 0$ and covariates $s_{i_q} - s_{k_q}$ for $q = 1, \ldots, d$. Hence, our model can be further extended to include more covariates in a similar

fashion as in the regression-based non-stationary model proposed by Risser and Calder (2015b), with all parameters identifiable.

Compared with the local stationary model, the additional parameters in Equation (3.3), $\{\beta_{qk}\}_{q=1}^{d}$ for $k = 1, \ldots, m$, have useful interpretations as measurements of non-stationarity. To illustrate the idea, let us first define the degree of non-stationarity.

Definition 1. Let $\theta(\mathbf{s}), \mathbf{s} = (s_1, \ldots, s_d) \in \mathbb{R}^d$ be a spatially varying parameter. Define the r-th order non-stationarity index of $\theta(\mathbf{s})$ denoted by D_r as

$$D_r = \frac{1}{d} \sum_{q=1}^{d} \left| \int_{s_q} \frac{\partial^r \theta(\mathbf{s})}{\partial s_q^r} ds_q \right|. \tag{4.1}$$

In particular, we call D_1 the trend-nonstationary index and D_2 the wiggliness-nonstationary index.

In Definition 1, large values of D_r are associated with stronger non-stationarity. For example, under stationarity, $\theta(\mathbf{s})$ is a constant at any location, $\mathbf{s}$. As a result, we obtain that $D_r = 0$, which indicates the smallest non-stationarity. In contrast, if D_r is large and far from zero, then $\theta(\mathbf{s})$ deviates from the stationary model along at least one direction of $\mathbf{s}$. This definition gives an appropriate and convenient measure for the degree of the r-th order non-stationarity.

Specifically, when $r = 1$ and $d = 1$, D_1 is an overall measure of the slope. When D_1 increases, the spatially varying parameter, $\theta(s)$, changes

at a faster rate and we expect a larger difference between $\theta(s_1)$ and $\theta(s_2)$ at two given locations s_1 and s_2. To avoid confusion with the trend of $\theta(s)$ in s, we call D_1 the trend-nonstationarity index. Similarly, the wiggliness-nonstationary index, D_2, is associated with how far these differences change. A related idea for time-series modeling can be found in Das and Nason (2016).

Now, it is easy to see that the parameters of interest, $\{\beta_{qk}\}_{q=1}^{d}$, $k = 1, \ldots, m$, directly define the trend-nonstationarity index, D_1. In our model, when $s \in [0, 1]$ and $m = 1$, we have $\theta(s) \approx \beta_{01} + \beta_{11}(s - s_1)$ and $D_1 = \left| \int_0^1 \frac{\mathrm{d}\theta(s)}{\mathrm{d}s} ds \right| \approx |\beta_{11}|$.

In general, in the k-th subregion and assuming that $\mathbf{s} \in [0, 1]^d$, we propose to use an empirical estimator of D_1, defined as $\hat{D}_{1k}$, to measure the non-stationarity. That is,

$$\hat{D}_{1k} = \frac{1}{d} \sum_{q=1}^{d} |\hat{\beta}_{qk}|, k = 1, \ldots, m. \tag{4.2}$$

4.3. Implication to high-resolution emulation

Estimation and emulation (simulation) are closely linked. Using the estimated parameters, we can directly generate realizations at N new locations, where $N \gg n$, after fitting our model at n observed locations following the method in Section 3.2. In the 2D situation, let $\{\hat{\beta}_{0k}^{\theta}\}_{k=1}^{m}$, $\{\hat{\beta}_{1k}^{\theta}\}_{k=1}^{m}$, and $\{\hat{\beta}_{2k}^{\theta}\}_{k=1}^{m}$ be the MLEs for a spatially varying parameter, $\theta(\mathbf{s})$. Then,

$\theta(\mathbf{s})$ at an unknown location, $\mathbf{s}^{\text{new}} = \{s_1^{\text{new}}, s_2^{\text{new}}\}$, is simply specified as

$$\hat{\theta}(\mathbf{s}^{\text{new}}) = \sum_{k=1}^{m} w(\mathbf{s}^{\text{new}}, \mathbf{s}_k) \left\{ \hat{\beta}_{0k}^{\theta} + \sum_{q=1}^{2} \hat{\beta}_{qk}^{\theta} (s_q^{\text{new}} - s_{k_q}) \right\}, \tag{4.3}$$

where $\mathbf{s}_k = (s_{k_1}, s_{k_2})$ is the anchor location at the k-th subregion. Hence, the new covariance matrix for $\mathbf{s}_i^{\text{new}}$, $i = 1, \ldots, N$, formed by the non-stationary Matérn in Equation (2.2), will be of size $N \times N$.

Statistical simulation has a much lower computational burden than estimation has. However, simulation is also challenging if we need to simulate high-resolution realizations where N is very large. Early studies of high-resolution simulation (Gneiting et al., 2006; Wood and Chan, 1994) typically focused on stationary or isotropic situations. Only a few studies considered the fast simulation of non-stationary GRFs. Nychka et al. (2015) introduced an idea based on Gaussian Markov random fields (GMRFs) and spatial autoregressive (SAR) models, for non-stationary processes. Kleiber (2016) proposed an efficient non-stationary simulation method based on spatial deformation. However, it is computationally expensive to estimate the deformation function before performing the simulation.

We propose to use an efficient algorithm based on sequentially conditional simulation. The sequential Gaussian simulation has been widely used in geostatistical simulations (Gotway and Rutherford, 1994; Fredericks and Newman, 1998). The idea is to simulate realizations only on a subset at a

time. In our application, we simulate one dataset for a subregion, then the next subregion is simulated conditioning only on all or a part of previous simulated ones. The details can be found in the Supplement (Section S1.2).

5. Simulation Study

To investigate the performance of our estimation approach, we consider a simple example where the non-stationarity exists only in the variance (partial sill), $\sigma^2(\mathbf{s})$. Let $W(\mathbf{s})$ be a stationary GRF with a known covariance function and $\sigma(\mathbf{s}) : \mathbb{R}^d \to \mathbb{R}^+$ be an unknown function. Then, $Z(\mathbf{s}) = \sigma(\mathbf{s})W(\mathbf{s})$ defines a GRF with a non-stationary variance. For simplicity, we consider the one-dimensional (1D) case with gridded data, $\sigma(s) : \mathbb{R} \to \mathbb{R}^+$. A similar example was also employed by Anderes and Stein (2011).

In our example, we set $(\tau, \nu, \lambda) = (0, 1, 0.2)$ in Equation (2.2) and let $\sigma(s) = 2\sin(s/0.15)+2.8$, $s \in [0, 1]$. Figure S8 in the Supplement shows one realization of $Z(s)$ and its increments. The pattern matches with the true curve of $\sigma(s)$ in Figure 1. Using the algorithm discussed in Section 3.2, we evenly divide the region into four subregions and choose the anchor location as the central point of each subregion. We generate 200 observations at equally spaced locations with 50 observations in each subregion. Here, only one parameter, $\sigma(s)$, is non-stationary.

As expected, in Figure 1, the estimation using our proposed model is

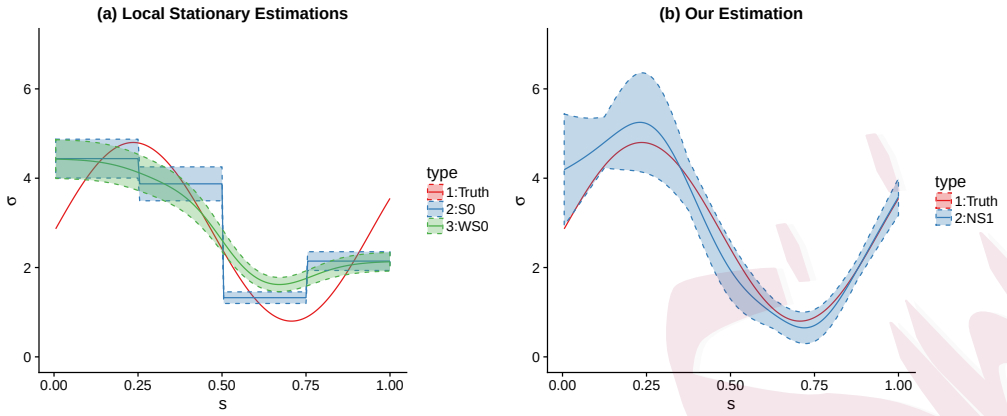

Figure 1: Estimations of $\sigma(s) = 2\sin(s/0.015) + 2.8$ with 95% confidence intervals using (a) local stationary model (S0) and weighted local stationary (WS0) model and (b) our first-order non-stationary (NS1) model

the closest to the truth of all models and the performance is much better than those of the S0 and WS0 models. The corresponding mean squared errors (MSEs) for S0, WS0, and NS1 are 0.54, 0.47, and 0.12, respectively. As the cost of more accurate estimation, the uncertainty of our method is the largest, which is especially significant at the boundaries. In contrast, at the center of each subregion, the uncertainty reaches the minimum.

Approximated confidence intervals give us a way to quantify the uncertainty with one realization. To measure the uncertainties better, we repeated the simulation 100 times. In Table 1, we show the the mean, standard error, and trend-nonstationary index, D_1, from 100 local stationary and slope estimators, $\hat{\beta}_0$ and $\hat{\beta}_1$. We also calculate the MSEs of $\hat{\sigma}(s)$ from

three models based on the mean of 100 $\hat{\beta}_0$'s and $\hat{\beta}_1$'s.

To visualize 100 estimated curves of $\sigma(s)$, we employ functional boxplots (Sun and Genton, 2011) as shown in Figure 2. Here, only WS0 and NS1 models are compared because the S0 model performs the worst. In the

Table 1: Summary of results from 1000 estimations. Definitions of columns: (1)-(2) The mean (standard deviation) of the S0 estimators $\hat{\beta}_0$ and slope estimators $\hat{\beta}_1$, respectively; (3) the empirical trend-nonstationarity index in the k-th subregion based on the mean of $\hat{\beta}_0$s, $\hat{D}_{1k}$; (4) the mean squared error of $\hat{\sigma}(s)$, MSE_{NS1}, calculated from the mean of $\hat{\beta}_0$s and $\hat{\beta}_1$s from the NS1 model with the corresponding ratios of MSE_{S0} and MSE_{WS0} to MSE_{NS1}; (5) $\text{MSE}_{\text{S0}}/\text{MSE}_{\text{NS1}}$; (6) $\text{MSE}_{\text{WS0}}/\text{MSE}_{\text{NS1}}$.

Subregion	1	2	3	4
$\sigma(s) = 2\sin(s/0.015) + 2.8$ (non-stationary case)				
$\hat{\beta}_0$	4.19 (0.43)	3.92 (0.43)	1.38 (0.19)	2.24 (0.28)
$\hat{\beta}_1$	9.08 (6.18)	-11.37 (5.49)	-5.21 (1.97)	11.41 (2.98)
$\hat{D}_{1k}$	9.08	11.37	5.21	11.41
MSE_{NS1}		0.050		
$\text{MSE}_{\text{S0}}/\text{MSE}_{\text{NS1}}$		9.397		
$\text{MSE}_{\text{WS0}}/\text{MSE}_{\text{NS1}}$		8.065		
$\sigma(s) = 2$ (stationary case)				
$\hat{\beta}_0$	2.01 (0.20)	2.01 (0.20)	1.99 (0.20)	2.00 (0.20)
$\hat{\beta}_1$	0.00 (1.96)	-0.02 (1.99)	-0.02 (1.42)	-0.02 (1.69)
$\hat{D}_1$	0.00	0.02	0.02	0.02
MSE_{NS1}		$6.463e^{-5}$		
$\text{MSE}_{\text{S0}}/\text{MSE}_{\text{NS1}}$		0.942		
$\text{MSE}_{\text{WS0}}/\text{MSE}_{\text{NS1}}$		1.199		

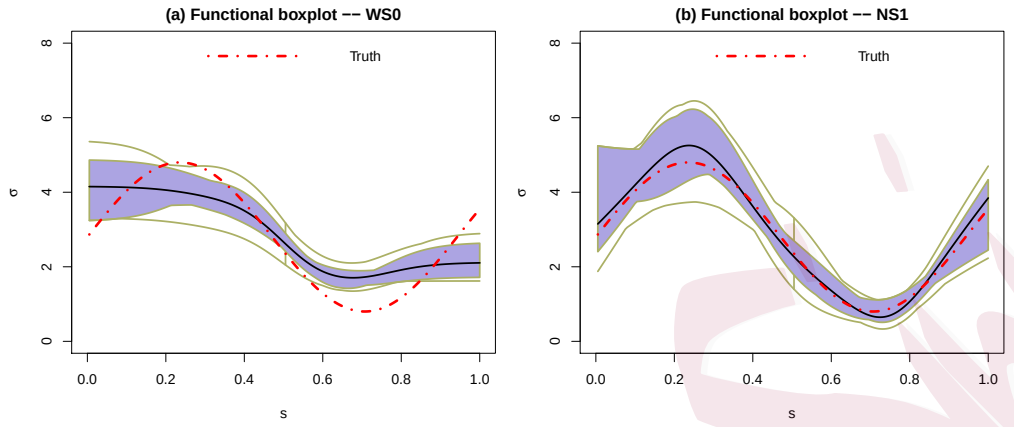

Figure 2: Functional boxplots of the 100 estimated curves of $\sigma(s)$ based on (a) WS0 model and (b) NS1 model.

functional boxplot, the black line indicates the most representative estimate of $\sigma(s)$ among the 100 simulations, and the box (shaded area) shows the variability. We can see that the NS1 model provides the most accurate estimates and outperforms the WS0 model.

The result can be understood in the following way. Generally, $\sigma(s)$ increases in the first subregion and decreases in the second one, but this change cannot be captured by local constant estimators. If we divide the region into finer subregions, then all the methods perform better. When estimating $\sigma(s)$, both WS0 and NS1 estimate $\sigma(s)$ as a weighted average of $\hat{\sigma}_k$'s. For the central location in each subregion, the estimate is dominated by the corresponding $\hat{\sigma}_k$, but not for the boundary locations. For example, when s falls in subregion k but close to subregion k', the estimation will be

dominated by both $\hat{\sigma}_k$ and $\hat{\sigma}_{k'}$, although the observation at s contributes to $\hat{\sigma}_k$ only.

When the true underlying GRF is stationary, we expect that $\hat{\beta}_1$ is close to 0. Therefore, with the same setup, we simulate data from a stationary GRF with $\sigma(s) = 2$. The estimation results in Table 1 clearly show that under stationarity, the estimated trend-nonstationarity index, $\hat{D}_{1k}$, and MSEs from all models in each subregion are close to zero. S0 model in this case can give the best result, but our model provides similar mean squared error. Hence, even if our NS1 model is designed for estimating complex non-stationary GRFs, it can also provide satisfactory results when the true model is stationary.

More simulation studies are conducted to investigate the sensitivity of estimation to region division and the computational time, as well as the 2D estimation with more non-stationary parameters. Detailed results can be found in the Supplement (Section S2). The simulation results show that our model is less sensitive to region division than WS0 and S0 models. However, an appropriate region division is still crucial to achieve a good estimation. In general, too few subregions lead to a poor estimation while too many subregions will increase the estimation uncertainty. For example, in the 1D simulation, two subregions ($m = 2$) are not enough, while $m = 4, 8, 10$ are

helpful. However, the uncertainty increases too much when $m = 8, 10$, since there are not enough data in local fitting. In terms of the computations, more subregions and fewer nonstationary parameters generally reduce the running time. For the same number of subregions, since our model has more parameters to estimate, it requires more computational time than the other two rough approximation methods, but much faster than the conventional moving window method when the window size is the same as the size of the subregion. In the 2D simulation study, our model still performs the best, although the result is not as significant as in the 1D case.

6. Application to Climate Model Emulation

Our application focuses on estimating and simulating data products obtained from general circulation model (GCM) outputs. GCMs are useful in forecasting weather and climate changes, but their resolutions are too coarse to characterize local patterns. The famous technique to produce high resolution realizations is downscaling (Wilby and Wigley, 1997), such as regional climate model (RCM). Driven by GCM, RCM simulations can involve more physical dynamics using local information, such as local humidity, wind speed, and other atmospheric variables. However, the local information is not always available. In addition, the process to produce the downscaling data is based on physical approaches. They often involve solv-

ing complicated partial differential equations, and thus are time consuming and cannot be completed on personal computers.

Different from RCM, we are aiming at using our efficient simulation methods to generate high-resolution outputs at RCM scale, in the hope that some of the fine-scale statistical properties we observed from the RCM output, the local non-stationarity in particular, can be reproduced. We call the downscaling outputs runs and our simulated outputs emulations. Without knowing the additional local information, we cannot easily reproduce the complicated local and regional features from RCM runs. However, we are able to reproduce some types of covariance non-stationarity, such as spatial range and variance.

In this study, we fit third-generation coupled global climate model runs from the Canadian Centre for Climate Modeling and Analysis (CCCma CGCM3 T47, Scinocca and McFarlane 2004) and consider the corresponding RCM it drives: Canadian RCM (CRCM Version 4.2.3, Shrestha et al. 2014). The spatial resolution for the GCM runs is about 3.75° for both latitude and longitude, whereas the RCM runs have much finer resolutions with a 45-km horizontal mesh (less than 1°).

The data of interest are seasonal average precipitation rates for the time period from 1971 to 2000 in a rectangular region in which the longitude

ranges from 157.71°W to 35.72°W and the latitude ranges from 19.45°N to 70.26°N (see Figure S9 in the Supplement). Spatially, there are 390 locations (longitude $\times$ latitude $= 30 \times 13$) for each GCM run. We first transform the data using square-root transformations to reduce the non-Gaussianity. Then, we detrend the transformed data by removing the 30-year average. Here, to look at the different non-stationarities along time, we choose four reference periods including two years, a non El Niño year 1971 and an El Niño year 1998, with two seasons, summer (JJA) and winter (DJF), for each year. Figure S10 in the Supplement shows the preprocessed precipitation rate residuals in summer 1971 from five GCM runs as our observations and one RCM run as our benchmark. Figure 3 shows the preprocessed precipitation rate residuals from the RCM runs in four percoids.

In this estimation procedure, we scale the coordinates down to $[0, 1]^2$ and equally divide the region into four subregions, as shown in Figure S4, with four anchor locations chosen at the center of each subregion. We model $(\lambda_1, \lambda_2, \phi)$ in the kernel matrix as spatially varying parameters and estimate these parameters along with others in the Matérn covariance function treating the five GCM runs as independent replicates. The estimation results of the two spatial ranges (λ_1, λ_2) from the WS0 and NS1 models for the four periods of interest are shown in Figure 4. To see similarities with the

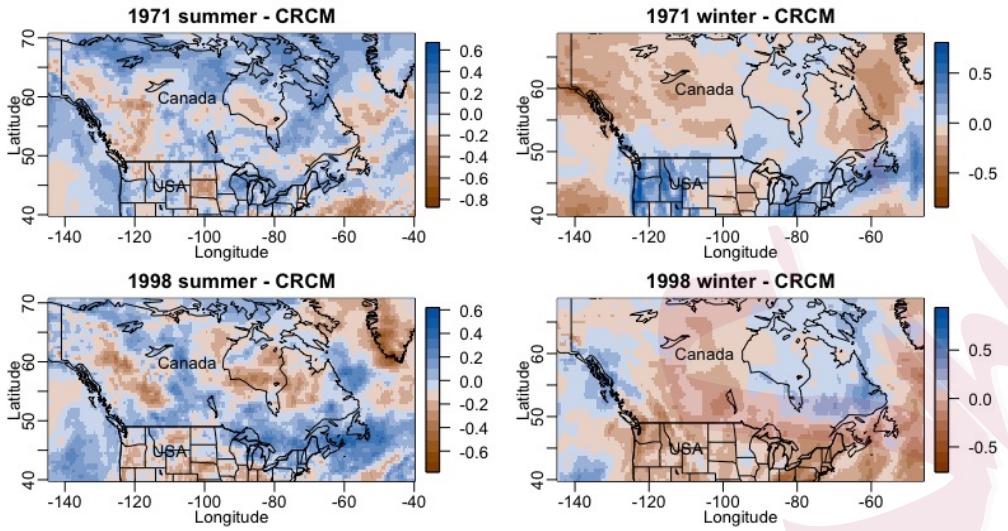

Figure 3: Preprocessed precipitation rate residuals (mm/day): four RCM runs in four reference periods.

RCM runs, we interpolate the spatially varying parameters to the RCM scale using Equation (4.3).

Visually, a larger spatial range is associated with a more homogenous spatial pattern of the observations. We observe that in the RCM run (Summer 1971), as shown in Figure 3, the north and southeast regions are more homogenous. This agrees with our estimation in Figure 4 where both regions show larger estimated spatial range parameters.

By looking at the non-stationarity along time from the NS1 results in Figure 4, we see clear non-homogeneity among seasons. In general, the spatial range is greater during winter, indicating fewer rapid fluctuations

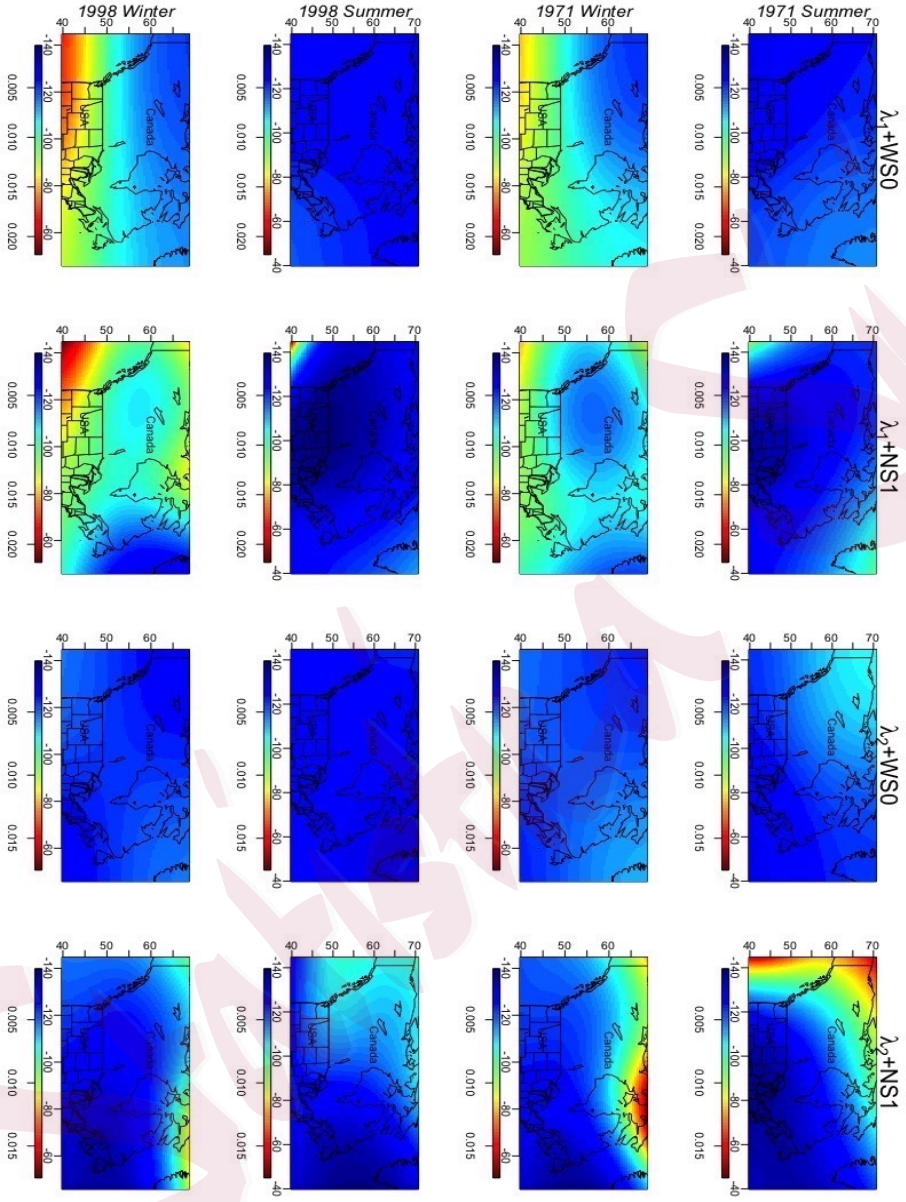

Figure 4: Plots of the estimated spatial ranges, λ_1, λ_2, based on two models, WS0 and NS1, at four reference periods, 1971 summer, 1971 winter, 1998 summer, and 1998 winter.

over space, which is comparable with the RCM runs. However, the non-stationarity of spatial ranges is less sensitive among different years for a given season. We can only see the larger spatial ranges out to sea along the Pacific Ocean in the US and in Greenland in Figure 4, indicating that there is non-stationarity in the land-ocean areas. Although the WS0 model also captures some of these characteristics, the estimation is too rough to observe changes in the spatial range at a finer scale.

We also estimate the trend-nonstationarity index, D_1, defined in (4.2). Table 2 shows the values of $D_{1k}, k = 1, \ldots, 4$, which describe the changes in the spatially varying parameters for each subregion in a given season. In terms of both spatial ranges, λ_1 and λ_2, $\hat{D}_{11}$ tends to be higher in the summers and $\hat{D}_{14}$ is higher in the winters.

We then implement the high-resolution simulation method to simulate the precipitation data at the RCM scale with a total of 13,020 locations (longitude $\times$ latitude=210$\times$62) during summer and $11,484$ locations (longitude $\times$ latitude=198$\times$58) during winter. From west to east, we sequentially simulate a subset including around 10% of the locations for each step conditioning on around 20% of the simulated data in the previous step on the eastern boundary. Our six independent emulations for the residuals of the winter precipitation rate in 1998 are displayed in Figure 5.

Table 2: Estimated trend-nonstationarity index, $\hat{D}_{1k}, k = 1, \ldots, 4$

Season	1971 Summer	1998 Summer
$\hat{D}_{1k}^{\lambda_1}$	(4.598, 0.855, 0.258, 1.156)	(11.232, 1.570, 0.148, 1.185)
$\hat{D}_{1k}^{\lambda_2}$	(4.122, 0.332, 1.622, 2.525)	(3.616, 0.056, 1.322, 3.382)
Season	1971 Winter	1998 Winter
$\hat{D}_{1k}^{\lambda_1}$	(0.696 , 0.647, 2.658, 1.891)	(1.037, 0.155, 2.493, 4.156)
$\hat{D}_{1k}^{\lambda_2}$	(0.813 , 0.048, 2.601, 6.221)	(0.777, 0.262, 2.279, 4.164)

Visually, from the Figure 5, we can see that our emulations share common variance and spatial ranges with the RCM run for winter 1998. For instance, we see similar variabilities overall and clusters in the coastal areas along the Pacific Ocean and Greenland. Furthermore, we select four local regions in the corners and show the empirical variograms based on the RCM runs and our emulations in Figure 6. As we can see, the variograms are significantly different in the local regions, which is evidence of spatial nonstationarity. In addition, the variograms of the RCM run can be viewed as one of the realization from our emulated ones. These results suggest that our model can provide similar statistical properties to RCM, although the RCM run exhibits more local and regional features.

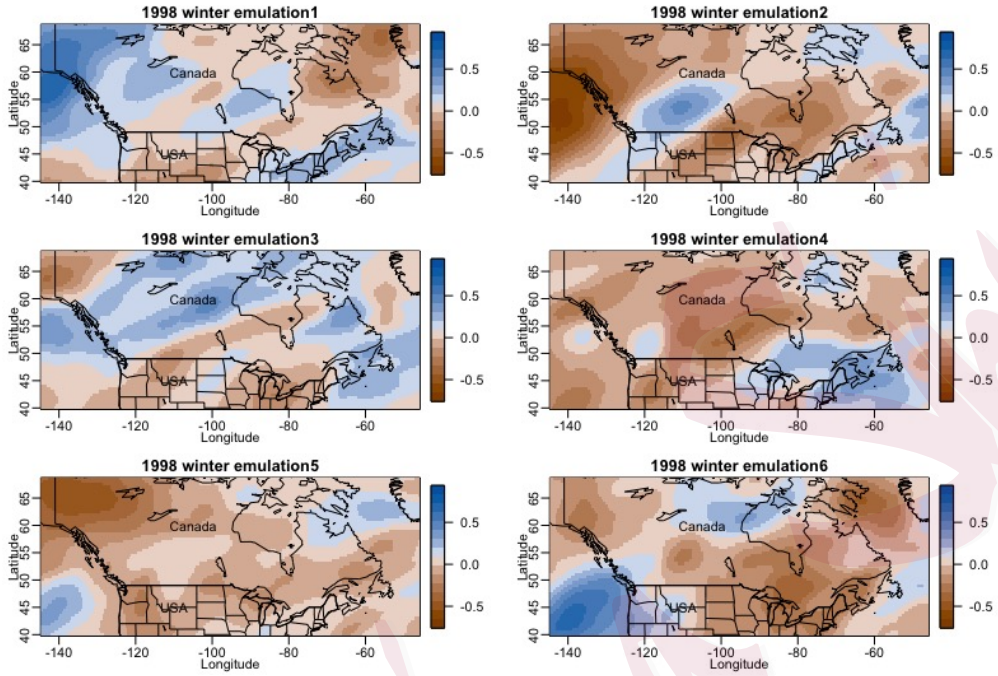

Figure 5: Six independent emulations of the residuals of winter precipitation rate in 1998 from our high-resolution simulation method.

7. Discussion

In this study, we have proposed a flexible method to model the spatially non-stationary covariance function, including the local stationary and weighted local stationary models as special cases. The proposed model allows more complex non-stationary features with interpretable parameterization that characterizes the degree of non-stationarity. We have also developed an efficient estimation approach and validated its performance by simulation studies. Motivated by the computational issue of climate

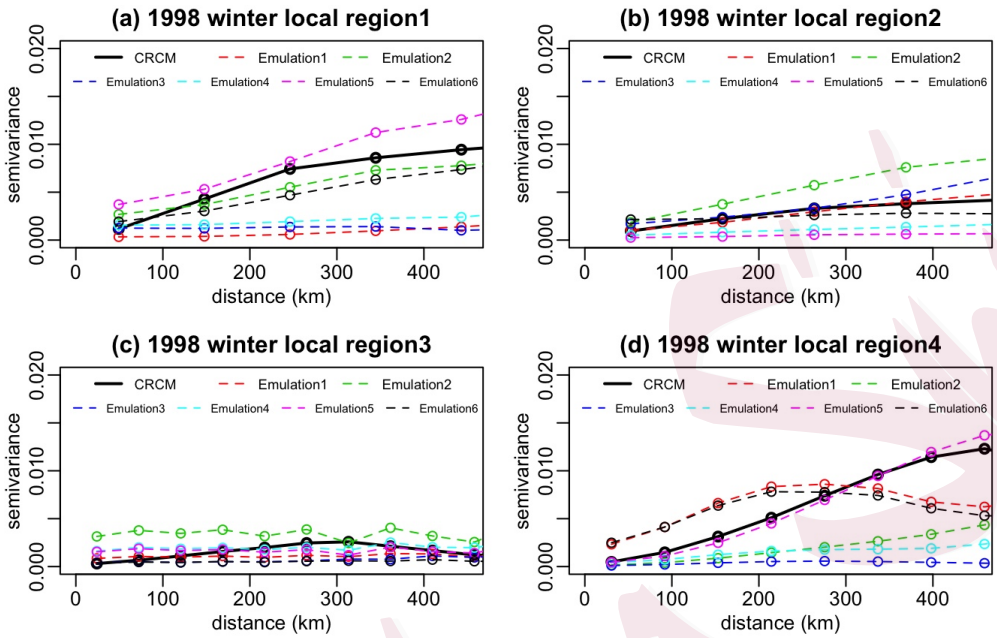

Figure 6: Empirical variograms of four local regions of RCM runs and our emulations. The ranges of the four local regions are (a) from (19.45°N,130.00°W) to (43.00°N, 157.71°W), (b) from (62.00°N,130.00°W) to (70.26°N, 157.71°W), (c) from (19.45°N, 35.72°W) to (43.00°N, 53.00°W), and (d) from (62.00°N, 53.00°W) to (35.72°W, 70.26°N).

model downscaling, we have developed a fast high-resolution simulation method. Compared with RCM runs from traditional downscaling methods as the benchmark, our model can capture similar spatial non-stationarity in a more efficient way.

Different from the full likelihood method that estimates parameters directly, our methodology is to approximate global non-stationary covariance

parameters by smoothing independently-estimated local covariance parameters. In our work, the fundamental theory is the Taylor approximation. Similar to nonparametric function estimation, we use a local polynomial representation for the spatially varying parameter. This is also different from the usual local polynomial estimator in a nonparametric regression model, where the approximation is about the mean function itself. After estimating all the unknown coefficients, we are able to plot and visualize the estimated spatially varying parameter as a smooth function in space.

Our methodology can be also embedded into the Bayesian hierarchical framework, where the polynomial terms can be viewed as hyper-parameters of the spatially varying parameters. However, these parameters, which are in the covariance function, are usually very sensitive to the priors, and sampling from a multi-layer hierarchical model is computationally expensive.

In terms of high-resolution simulation, the method in Kleiber (2016) is more efficient by using the circulate embedding technique. For example, in Section 3.3 of Kleiber (2016), simulating a realization at a grid of $120 \times 98 = 11,760$ locations only needs approximately 1 second, whereas around 15 minutes are required for our methods. However, this technique needs to estimate a spatial deformation function and then simulate data at regularly spaced locations. Therefore, it only works for regularly spaced spatial data,

while ours is suitable for both regularly and irregularly spaced data.

For our future work, the non-stationarity index mentioned in Section 4.2 can be further extended to build a non-stationarity test. The null hypothesis is just whether the slope parameters are zeros. It can be achieved either by likelihood ratio test based on the maximized likelihood ratio between WS0 and NS1 or the Wald test based on the estimations of slope parameters. Future research is also needed for developing automatic and adaptive partitioning criteria. Although our models are less sensitive to region partitioning than the weighted local stationary model, making use of some validation tools or the information from covariates (see Risser et al. 2016) could further help optimize the region partition.

Supplementary Materials

Section S1 illustrates more details with regards to computations. Section S2 mentions more simulation studies. Section S3 includes some additional figures. Section S4 gives the R source codes for the simulation studies and applications. Section S5 provides the GCM and RCM data sources including instructions on how to obtain related data.

Acknowledgements

This research was supported by funding from King Abdullah University

of Science and Technology (KAUST). We would like to thank the editor, associate editor, and two anonymous reviewers for their valuable comments.

References

Anderes, E. B. and Stein, M. L. (2008). Estimating deformations of isotropic Gaussian random fields on the plane. *The Annals of Statistics*, 36(2):719–741.

Anderes, E. B. and Stein, M. L. (2011). Local likelihood estimation for nonstationary random fields. *Journal of Multivariate Analysis*, 102(3):506–520.

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

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Das, S. and Nason, G. P. (2016). Measuring the degree of non-stationarity of a time series. *Stat*, 5(1):295–305.

Fouedjio, F., Desassis, N., and Rivoirard, J. (2016). A generalized convolution model and estimation for non-stationary random functions. *Spatial Statistics*, 16:35–52.

Fredericks, A. K. and Newman, K. B. (1998). A comparison of the sequential gaussian and markov-bayes simulation methods for small samples. *Mathematical Geology*, 30(8):1011–1032.

Gelfand, A. E. and Schliep, E. M. (2016). Spatial statistics and Gaussian processes: a beautiful

marriage. *Spatial Statistics*, 18:86–104.

Gneiting, T., Ševčíková, H., Percival, D. B., Schlather, M., and Jiang, Y. (2006). Fast and exact simulation of large Gaussian lattice systems in $\mathbb{R}^2$: exploring the limits. *Journal of Computational and Graphical Statistics*, 15(3):483–501.

Gotway, C. and Rutherford, B. (1994). Stochastic simultation for imaging spatial uncertainty: Comparison and evaluation of available algorithms. In *Geostatistical simulations*, pages 1–21. Springer.

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

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

Jun, M. and Stein, M. L. (2008). Nonstationary covariance models for global data. *The Annals of Applied Statistics*, 2(4):1271–1289.

Kleiber, W. (2016). High resolution simulation of nonstationary Gaussian random fields. *Computational Statistics & Data Analysis*, 101:277–288.

Königsberger, K. (2013). *Analysis 2*. Springer-Verlag.

Li, Y. and Zhu, Z. (2016). Modeling nonstationary covariance function with convolution on sphere. *Computational Statistics & Data Analysis*, 104:233–246.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., and Sain, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, 24(2):579–599.

Nychka, D., Wikle, C., and Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, 2(4):315–331.

Paciorek, C. J. and Schervish, M. J. (2004). Nonstationary covariance functions for Gaussian process regression. In *Advances in neural information processing systems*, pages 273–280.

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

Patterson, H. and Thompson, R. (1975). Maximum likelihood estimation of components of variance. In *Proceedings of the 8th international biometric conference*, pages 197–207.

Risser, M. D. (2016). Review: Nonstationary spatial modeling, with emphasis on process convolution and covariate-driven approaches. *arXiv preprint arXiv:1610.02447*.

Risser, M. D. and Calder, C. A. (2015a). Local likelihood estimation for covariance functions with spatially-varying parameters: the convoSPAT package for R. *arXiv preprint arXiv:1507.08613*.

Risser, M. D. and Calder, C. A. (2015b). Regression-based covariance functions for nonstationary spatial modeling. *Environmetrics*, 26(4):284–297.

Risser, M. D., Calder, C. A., Berrocal, V. J., and Berrett, C. (2016). Nonstationary spatial process modeling via treed covariate segmentation, with application to soil organic carbon stock assessment. *arXiv preprint arXiv:1608.05655*.

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

Scinocca, J. F. and McFarlane, N. A. (2004). The variability of modeled tropical precipitation. *Journal of the Atmospheric Sciences*, 61(16):1993–2015.

Shrestha, R. R., Schnorbus, M. A., Werner, A. T., and Zwiers, F. W. (2014). Evaluating hydroclimatic change signals from statistically and dynamically downscaled GCMs and hydrologic models. *Journal of Hydrometeorology*, 15(2):844–860.

Stein, M. L. (2005). Nonstationary spatial covariance functions. *Unpublished report*.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334.

Wilby, R. L. and Wigley, T. (1997). Downscaling general circulation model output: a review of methods and limitations. *Progress in physical geography*, 21(4):530–548.

Wood, A. T. and Chan, G. (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics*, 3(4):409–432.

King Abdullah University of Science and Technology, CEMSE Division, Thuwal 23955-6900, Saudi Arabia

E-mail: yuxiao.li@kaust.edu.sa

E-mail: ying.sun@kaust.edu.sa