

Modeling nonstationary spatial data using local likelihood estimation and Matérn-SAR covariance translation

Ashton Wiens^{a,*}, Douglas Nychka^b, William Kleiber^a

Colorado, United States

^a*Department of Applied Mathematics, University of Colorado, Boulder, Colorado, USA*

^b*Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, Colorado, USA*

Abstract

Modeling data with a nonstationary covariance structure is important to represent heterogeneity in geophysical and other environmental spatial fields. However, spatial data sets with large numbers of locations are difficult to handle using standard methods. A multistage approach to modeling nonstationary covariances is presented that is efficient for large data sets. The key idea is to divide the modeling into two steps. First, we estimate spatial varying covariance parameters using a local likelihood applied to a moving spatial window. These parameters are then translated into an equivalent model as a spatial autoregression (SAR) that is global and sparse. The global feature of the SAR is a solution to the problem of how to combine the local estimates of the covariance into a single coherent description of a Gaussian process. Moreover, This strategy combines the ease and interpretability of the Matérn family of covariances and the efficient computation afforded by spatial models where the precision matrix is sparse. A main result is establishing the accuracy of approximating the Matérn family of covariances with a SAR. We also demonstrate through a Monte Carlo study that local likelihood estimates are accurate even for correlation ranges much larger than the window size provided a sufficient number of replicate fields are available. This has important practical implications for keeping window sizes moderate. The method is applied to an important suite of climate model simulations where replicated fields are available and the covariance varies significantly as the result of land/ocean effects. Although this

*This document is a collaborative effort. Declarations of interest for all authors: none

*Corresponding author: Department of Applied Mathematics, University of Colorado, 1111 Engineering Dr, Boulder, CO 80309

Email addresses: ashton.wiens@colorado.edu (Ashton Wiens), nychka@mines.edu (Douglas Nychka), william.kleiber@colorado.edu (William Kleiber)

example benefits from spatial observations on a regular grid the methods are easily extended to irregular data using a basis expansion.

Keywords: nonstationary Gaussian process, local likelihood, Gaussian Markov random field, spatial autoregression, process convolution

1. Introduction

This work is motivated by a climatological application where the goal is to emulate the variability of an ensemble of spatial fields generated by a climate forecast model. Accordingly, we are lead to model spatial data consisting of independent replicated spatial fields that exhibit a nonstationary covariance structure using a Gaussian process. To accurately emulate these fields, modeling the nonstationarity in the second-order structure of the data is essential. To avoid the computational limitations of representing and simulating nonstationary Gaussian processes, we investigate a two-step approach similar to the methodology in [23]. First, assuming the field is approximately locally stationary, we perform moving window local likelihood estimation to infer spatially varying Matérn covariance parameters. These parameter fields are translated into the parameters of a spatial autoregressive model which best reproduces the behavior of the Matérn correlations locally. Finally, the spatially varying parameters are encoded into the nonstationary SAR covariance model, specifying all of the data jointly.

Local estimation is not a new idea in spatial statistics [11, 12, 32, 28]. A local estimation approach circumvents the $\mathcal{O}(n^3)$ computational burden, where n is the number of observations. Instead, the task becomes n embarrassingly parallel subproblems on the order of the window size used in the local estimation.

We assume that local estimation is a data-driven approach: when the data consists of densely observed independent replicates, robust local estimation of covariance parameters is possible. In practice, there is often no clear indication of which parameters in the model should be allowed to vary spatially [10] and what spatial scales are appropriate for the parameter surfaces. This difficult modeling choice is avoided when using local estimation: we can allow all parameters to vary initially, and the local estimates will indicate whether the parameters are constant or vary over space. Furthermore, with local estimation, we do not have to decompose the parameter functions into some prespecified low-dimensional representation [9, 27], which can influence the estimation.

Weighted local likelihoods have been studied to accommodate irregularly spaced observations [2], but in this work we use a simple moving window applied to data on a lattice. Here, we focus on estimation of Matérn parameters, primarily because of their interpretability and in order to study the relationship between the Matérn and SAR covariance models, detailed below. To establish

local estimation as a reliable technique, we use a Monte Carlo experiment to study the robustness of local estimation of the correlation range parameter.

With locally estimated covariance parameters in hand, some care is required to combine these into a valid global nonstationary covariance model. A simple option is to use the estimates to construct local covariance functions and perform local simulation. However, a global covariance specifying the relationships among all of the data is desirable for efficient simulation and necessary for prediction. A global representation also avoids potential artifacts and ad hoc choices in synthesizing the spatial analysis across local windows.

There are several general classes of nonstationary models, such as deformation methods [30, 3], basis function methods [5, 17, 22, 24], process-convolution construction [14, 16, 15, 25, 8, 7, 34], and the SPDE approach [21, 20, 31, 29]. See [27] for a review of nonstationary models and [13] for a review of methods for large spatial data sets. Unfortunately, only a few of these methods are easily implemented due to the complexity of the models. Here, we investigate two existing nonstationary models from the process convolution and GMRF families of methods which are amenable to plug-in local estimates.

In this work, we study the nonstationary spatial autoregressive (SAR) model, related to the Gaussian Markov random field (GMRF) approach to approximating GPs. The idea is to identify members of the Matérn family of spatial processes as solutions to a stochastic partial differential equation. The SPDE is then discretized to a lattice and this motivates the form of the SAR [21]. The correspondence between the Matérn/SPDE form and a SAR was presented in [21], and an analytical formula was proposed to connect the parameters between the continuous and discrete cases. We have found that the analytical formula is inaccurate for large correlation ranges and one contribution of this work is to sharpen this relationship using numerical results. The advantage is that if one can successfully translate the Matérn formulation into a SAR framework, one can exploit sparse matrix algorithms for fast computation.

Finally, we apply this multistage modeling framework to analyze a nonstationary climate model output data set consisting of 30 temperature anomaly fields from the NCAR CESM project. First, we locally estimate stationary, anisotropic Matérn parameters. We then translate these local Matérn parameters into the SAR parameters which yield the best numerical approximation between the stationary Matérn and the approximately stationary SAR model. Finally, we encode the spatially varying SAR parameters into the nonstationary SAR model. This model convincingly captures many of the nonstationary features of the climate distribution in simulations.

The paper is organized as follows. In section 2, the Matérn family of correlation functions and the process convolution model are introduced, along with the approximately stationary and nonstationary SAR models from the SPDE ap-

proach. We also include a numerical experiment investigating the link between the stationary Matérn and SAR covariance models. In section 3, we develop the local likelihood framework we employ, and we conduct a local estimation
80 simulation study. In section 4, we apply the multistage modeling framework to analyze a nonstationary climatological data set. We conclude with a summary of the method and a discussion of some of the relevant practical aspects.

2. Nonstationary covariance models

In this section, we introduce the Matérn family of covariance models, as well
85 as the process convolution and the SAR/SPDE approaches to constructing a Gaussian process/GMRF. We then state the connection between the Matérn and SPDE models, and explore this relationship in a numerical study.

2.1. The Matérn covariance model

Let f be a Gaussian process with mean zero and covariance function $k(\mathbf{x}, \mathbf{x}')$.
90 The Matérn family of stationary covariance models is important because of its flexibility and the interpretability of its parameters. The Matérn covariance function with a unit range (scale) parameter is

$$\omega(d) = C(d | \nu, \sigma^2) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (d)^\nu \mathcal{K}_\nu(d),$$

where d is the Euclidean distance between \mathbf{x} and \mathbf{x}' , $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν , and $\Gamma(\cdot)$ is the gamma function.
95 σ^2 is the spatial process variance (sill), and ν is the smoothness parameter which controls the mean square differentiability of the process. The isotropic covariance function with range κ is given by

$$k(\mathbf{x}, \mathbf{x}') = \omega(\kappa d)$$

This model can be extended to include geometric anisotropy through the definition of the distance that is a linear scaling and rotation of the coordinates.
100 Let $A = D^{-1}U^T$ be a 2×2 matrix where U is a rotation matrix parameterized by angle θ

$$U = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix},$$

and

$$D = \begin{bmatrix} \xi_x & 0 \\ 0 & \xi_y \end{bmatrix}$$

is a diagonal matrix scaling the x and y coordinate axes separately. Then the pairwise Mahalanobis distances among the observation locations are defined as

$d = \|As - As'\|$ and these distances are then used as the argument to the isotropic Matérn covariance function. The interpretation is that if one transforms the coordinates according the linear transformation A then the resulting field will be isotropic.

2.2. Process convolution construction

Process convolution is a useful method for constructing valid nonstationary GPs using a spatially varying kernel. Let $\Psi(\mathbf{u}, \mathbf{v})$ be a kernel such that

$$\int_{G \times G} |\Psi(\mathbf{u}, \mathbf{v})|^2 d\mathbf{u} d\mathbf{v} < \infty,$$

then we have the process representation

$$Y(\mathbf{s}) = \int_G \Psi(\mathbf{s}, \mathbf{u}) dW(\mathbf{u}),$$

where W is d-variate Brownian motion. It follows that Y has covariance function

$$k(\mathbf{s}, \mathbf{s}') = \int_G \Psi(\mathbf{s}, \mathbf{u}) \Psi(\mathbf{s}', \mathbf{u}) d\mathbf{u} \quad (1)$$

The advantage of the convolution representation is that any choice of Ψ will yield a valid covariance function, and this provides flexibility in devising dependence that varies over the spatial domain. However, except in some special cases, the covariance in (1) is difficult to compute in closed form from. To our knowledge only a Gaussian kernel is amenable to a closed form covariance function [16]. Given that parameter estimation and simulation require many evaluations of the covariance at pairs of locations the direct use of this model is problematic for large data sets. A key idea in this work is that shifting to a SAR process model gives an approximation to this convolution form but is also efficient to evaluate. This is due to a SAR giving rise to sparse precision matrices that approximate the inverses of the covariance matrices.

We note that Paciorek and Schervish [25] have also proposed a family of nonstationary covariance functions. However, this construction deviates from the process convolution approach and since it is in terms of covariances does not result in sparse matrices. It is an open question whether the SAR will approximate the form proposed in [25].

2.3. The SAR model

In contrast to modeling a continuous covariance function, the SAR model parameterizes the precision matrix for the process on a discrete lattice. In our case, we have observations of a spatial field \mathbf{y} located on a regular rectangular grid. This is the setup used throughout the paper and in the data analysis

¹³⁰ section. We do not believe that this assumption is overly restrictive for development of this model and in the discussion we explain how to generalize this model to irregularly spaced observations and other linear functionals of the process.

¹³⁵ In two dimensions assume the spatial process is indexed as $\mathbf{Y}_{i,j}$ and without loss of generality take the integers $1 \leq i \leq m$ and $1 \leq j \leq n$ to be the lattice locations. The isotropic SAR model can be written using graphical notation as

$$\begin{array}{c|cc|c} 0 & -1 & 0 \\ \hline -1 & 4 + \kappa_S^2 & -1 \\ \hline 0 & -1 & 0 \end{array} \quad (2)$$

Given a white noise vector $\mathbf{e}_{i,j}$ distributed independent $\mathcal{N}(0, 1)$, we interpret the lattice process to satisfy

$$(4 + \kappa_S^2)\mathbf{Y}_{ij} - (\mathbf{Y}_{i-1,j} + \mathbf{Y}_{i+1,j} + \mathbf{Y}_{i,j-1} + \mathbf{Y}_{i,j+1}) = \mathbf{e}_{i,j}, \quad (3)$$

with lattice values outside the range of the indices set to zero.

¹⁴⁰ Here $\kappa_S > 0$ is suggestive of a range parameter controlling the dependence of the field and is similar but not identical to κ for the continuous case. For this finite lattice one can identify a matrix B using 3 such that $B\mathbf{y} = \mathbf{e}$. Furthermore, it is clear that if $\kappa_S > 0$, the diagonal dominance of B guarantees its invertibility. With $\mathbf{y} = B^{-1}\mathbf{e}$ the covariance matrix for \mathbf{y} is $B^{-1}B^{-T}$ and thus \mathbf{y} has precision matrix $Q = B^T B$. The precision matrix implied by the SAR ¹⁴⁵ model is sparse, but in two (or more) dimensions is not banded. The sparsity property makes the SAR model amenable to modeling large data sets because the precision matrix can be used instead of a dense covariance matrix for likelihood estimation and simulation. Following the ideas from [21] one can iterate the spatial autoregressive weights to obtain higher order models. For example, if $BB\mathbf{y} = \mathbf{e}$, this implies a SAR model extending to second-order nearest neighbors and gives the precision matrix: $Q_2 = (BB)^T(BB)$. The SAR model detailed here is a special case of a Gaussian Markov random field. For a given row of the precision matrix, the nonzero, off-diagonal entries index the neighbors that will determine the Markov property. Conditioning on these neighbors, the field at the lattice location (corresponding to the given row) will be independent of the remaining values of the field. For the first-order SAR described above the nonzero elements in Q will include second-order neighbors. Thus this first-order SAR will be a GMRF based on second-order neighbors and the weights will depend on B .

¹⁵⁰ Two additional points should be mentioned about the SAR model. First, this stencil should be modified at the boundaries of the domain. The center value of the stencil should be the κ^2 plus the sum of the weights of its non-zero neighbors. Second, the value of κ affects the marginal variance of the process,

so σ^2 is a parameter that allows for modulation of the variance, but it is not the marginal variance of the process itself.

A version of the SAR model that exhibits approximate stationarity and also geometric anisotropy will be detailed in the next section.

2.4. Connection to Matern family

Lindgren et al. [21] suggested that the SAR covariance model can be thought of as a discrete approximation to the Matérn covariance model. The idea is that a Gaussian Markov random field with SAR covariance can provide an approximation to a Gaussian field with Matérn covariance. The connection is established through an SPDE formulation. In particular, it is known that a Gaussian field $u(\mathbf{s})$ with stationary Matérn covariance is a solution to the SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} u(\mathbf{s}) = \mathcal{W}(\mathbf{s})$$

where $\alpha = \nu + \frac{d}{2}$, $\kappa > 0$, $\nu > 0$, $\mathbf{s} \in \Omega = \mathbb{R}^d$, $d = 1$ or 2 , and $\mathcal{W}(s)$ is identified as white noise with zero mean and variance σ^2 . As in the Matérn model, ν controls the smoothness of the Gaussian field. Fixing $\nu = 1$ and $d = 2$, the authors showed that the SAR covariance structure obtained by discretizing the pseudodifferential operator $(\kappa^2 - \Delta)$ approximates a Matérn covariance structure with range $\kappa \approx \kappa_S$. This relationship is approximate.

Similar results can be obtained for different smoothness parameters ν by convolving the finite difference stencil in (2) with itself ν times, as detailed in the previous section for $\nu = 1$ and $\nu = 2$.

The SAR model can also be extended to incorporate geometric anisotropy. Let H denote an anisotropy matrix and modify the Laplacian in the pseudodifferential operator as follows

$$(\kappa^2 - \nabla \cdot H \nabla)^{\alpha/2} u(\mathbf{s}) = \mathcal{W}(\mathbf{s}) \quad (4)$$

Here H is assumed to be a symmetric positive-definite matrix. To avoid potential ambiguity we also identify the Laplacian operator above for two dimensions in an expanded form as

$$\nabla \cdot H \nabla \equiv H_{1,1} \frac{\partial^2}{\partial^2 s_1} + 2H_{2,1} \frac{\partial^2}{\partial s_1 \partial s_2} + H_{2,2} \frac{\partial^2}{\partial^2 s_2}.$$

From this, the first-order finite difference discretization of the anisotropic

SPDE 4 gives the following stencil for filling the rows of the B matrix

$$\begin{array}{c|c|c} \frac{2H_{12}}{h_x h_y} & -\frac{H_{22}}{h_y^2} & -\frac{2H_{12}}{h_x h_y} \\ \hline -\frac{H_{11}}{h_x^2} & \kappa^2 + \frac{2H_{11}}{h_x^2} + \frac{2H_{22}}{h_y^2} & -\frac{H_{11}}{h_x^2} \\ \hline -\frac{2H_{12}}{h_x h_y} & -\frac{H_{22}}{h_y^2} & \frac{2H_{12}}{h_x h_y} \end{array} \quad (5)$$

where h_x and h_y are the grid spacings along the x-axis and y-axis. This is just a reparameterization of the results in Appendix A of [21] which facilitates the practical translation of these models. Note that setting $h_x = h_y = 1$, $H_{12} = H_{21} = 0$, and $H_{11} = H_{22} = 1$ yields the first-order isotropic model from (2).

Finally, we connect the role of H in the SPDE formulation to the anisotropic model for the Matérn. Under the linear transformation $A = D^{-1}U^T$ from Section 2.1, let $\mathbf{s}^* = A^{-1}\mathbf{s}$, and let u be an isotropic field solution to the SPDE with Laplacian $\nabla \cdot \nabla$. Furthermore, set $u^*(\mathbf{s}) = u(A^{-1}\mathbf{s}) = u(\mathbf{s}^*)$. Then from elementary properties of the gradient

$$\nabla u^*(\mathbf{s}) = \nabla(u(A^{-1}\mathbf{s})) = A^{-1}\nabla u(A^{-1}\mathbf{s})|_{\mathbf{s}=A\mathbf{s}^*} = A^{-1}\nabla u(\mathbf{s}^*)|_{\mathbf{s}^*=A^{-1}\mathbf{s}}$$

and so we have

$$\nabla \cdot \nabla u^* = (A^{-1}\nabla) \cdot A^{-1}\nabla u(\mathbf{s}^*) = \nabla \cdot A^{-T}A^{-1}\nabla u(\mathbf{s}^*)|_{\mathbf{s}^*=A^{-1}\mathbf{s}}.$$

From this expression we identify $H = A^{-T}A^{-1}$. From Section 2.1, if u is an isotropic field then u^* will be anisotropic with coordinates transformed by A^{-1} . Moreover, u^* will also be the solution to the SPDE with $H = A^{-T}A^{-1}$. This connection provides guidance how to interpret H . Note that if A is a pure rotation then $H = I$ and isotropy is preserved.

2.5. Numerical translation of range parameters between the Matérn and SAR models

The connection between the anisotropic Matérn family and a SAR relies on the approximation of a discretized Laplacian operator with finite differences of the fields on a lattice. To provide an accurate statistical model, it is important to quantify this approximation and improve its calibration over the limiting expression suggested in [21]. In this section we provide numerical evidence to show that an accurate calibration is possible if restricted to ranges of the covariance parameters.

The computational setup is as follows. Given a Matérn range parameter a , we estimate the value of κ in the SAR model which gives the best approximation

²¹⁰ to the Matérn correlation function. We conducted this experiment with the smoothness of the Matérn model fixed at $\nu = 1$ and $\nu = 2$, and with unit marginal variance for all models. The first step is to fix the Matérn range parameter and evaluate a Matérn correlation matrix on the grid. Then, we perform an optimization over κ by encoding it into the SAR precision matrix using (2), inverting and normalizing it to give the implied SAR correlation matrix, and minimizing the distance between these matrices by some measure.

²¹⁵ It is known that the SAR covariance model suffers from edge effects. To avoid the interference of edge effects in this optimization , we quantify the difference between the two correlation matrices by only comparing the correlation of the lattice point in the center of the grid for both models. For an $N \times N$ lattice of locations, with N odd, let σ_a denote the vector of correlations between the center point in this lattice and all other locations based on the Matérn covariance function. Let σ_κ be the same correlation vector for the SAR model with parameter κ . We then find

$$\min_{\kappa} \|\sigma_a - \sigma_\kappa\|_2$$

Denote this minimizer $\hat{\kappa}(a)$ and so this value will be a mapping from the range parameter of the Matérn family into the SAR model.

²²⁰ The approximation results are summarized in Figure 1. Here a is varied over the interval $[1, 20]$, $N = 51$ and the lattice points have unit spacing. In 1(a) $\hat{\kappa}(a)$ is plotting as a function of a . Orange corresponds to the $\nu = 1$ case, cyan corresponds to $\nu = 2$, and the solid black line shows the theoretical relationship, $\frac{1}{a} = \frac{1}{\kappa}$ from [21]. From these computations we conclude that at this level of discretization it is important not to rely on the analytic formula to translate between a and κ parameters.

²²⁵ The relative error of using the SAR correlation with κ value derived from the numerical experiment is shown in 1(d). The ℓ_2 distance measure used in the optimization of the model correlation matrices is used to quantify the resulting model error, normalized by $\|\sigma_a\|$. The relative error incurred when using $\frac{1}{\kappa} = \frac{1}{a}$ is shown by the dashed lines. Note that there is tenfold decrease in the relative error in many cases.

²³⁰ In the data analysis below, we found it necessary to include geometric anisotropy in the covariance model. For this reason, we also investigated how the presence of geometric anisotropy affects the numerical correspondence established for the isotropic case above. Overall, the behavior of the approximation was similar to the isotropic case: the estimated eigenvalues of the SAR anisotropy matrix H were smaller than the eigenvalues fixed in the Matérn anisotropy D^2 matrix. In this experiment, we encoded fixed values for ξ_x and ξ_y in a Matérn correlation matrix such that the length scale ratio $\xi_x : \xi_y = 4:1$, which was consistent with the anisotropic estimates in the data analysis. In

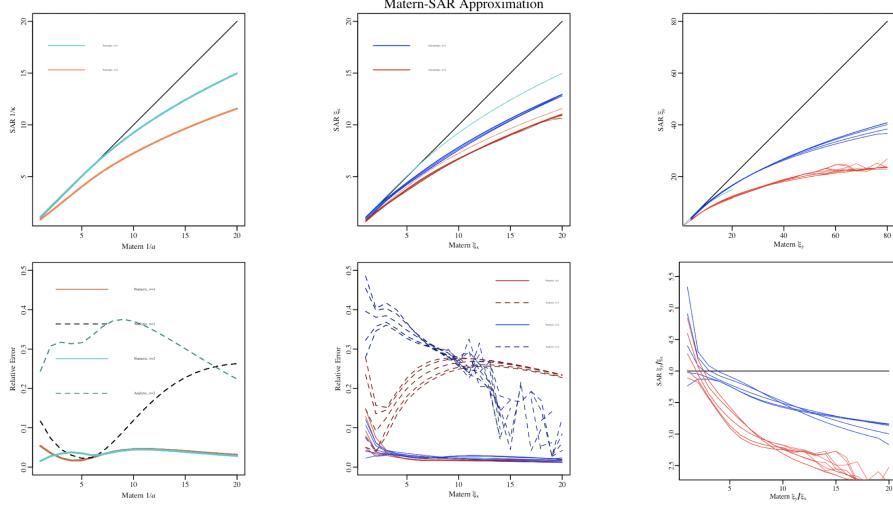


Figure 1: For the isotropic case, the optimal $1/\kappa$ parameter for a given Matérn inverse range $1/a$ is plotted in (a). The relative error incurred by using the SAR model with optimal κ as an approximation to the Matérn model is shown in (d). For the anisotropic case, the optimal eigenvalues of H are plotted against the fixed diagonal values of D^2 in panels (b) and (c), and the relative error is shown in (e). Finally, the ratio of eigenvalues is shown in (f)

particular, we let $\xi_x = 1, \dots, 20$ and $\xi_y = 4\xi_x$. Then, the optimal eigenvalues of the SAR anisotropy matrix H were found. The experiment was repeated with rotation angles $\theta = 0^\circ, 10^\circ, \dots, 90^\circ$, and the rotation angle was assumed to be known and fixed in the eigendecomposition of the H matrix. The anisotropic parameter translation results for ξ_x and ξ_y are shown in panels (b) and (c) of Fig 1, respectively, and the relative error of approximation is shown in panel (e). Finally, the ratio of the estimated eigenvalues of H is shown in panel (f); compare to the line $y = 4$.

The approximation results may be slightly affected by the rotation angle and oblateness of the geometric anisotropy, but the effect is negligible in practice. From these results, we have ascertained a numerical translation among the anisotropy parameters. We can use these results to translate locally estimated Matérn range parameters into SAR parameters with better accuracy than the conjectured analytic relationship.

2.6. The nonstationary SAR model

The nonstationary SAR model can be constructed by allowing the parameters κ , H , and σ^2 in the generating SPDE to vary over space. Let

$$\mathcal{L}(\mathbf{s}) = H_{1,1}(\mathbf{s}) \frac{\partial^2}{\partial s_1^2} + 2H_{2,1}(\mathbf{s}) \frac{\partial^2}{\partial s_1 \partial s_2} + H_{2,2}(\mathbf{s}) \frac{\partial^2}{\partial s_2^2}.$$

The SPDE can then be written as

$$(\kappa^2(\mathbf{s}) - \mathcal{L}(\mathbf{s}))^{\alpha/2} u(\mathbf{s}) = \mathcal{W}(\mathbf{s})$$

where $\kappa(\mathbf{s}) > 0$, $\mathcal{W}(s) \sim \text{WN}(0, \sigma^2(\mathbf{s}))$, and $\sigma^2(\mathbf{s}) > 0$. Furthermore, we specialize to a spatially varying linear transformation of the coordinates, $A(\mathbf{s})$, and so $H(\mathbf{s}) = A^{-T}(\mathbf{s})A^{-1}(\mathbf{s})$. Note that $A(\mathbf{s})$ varying in space is equivalent to specifying spatial fields for θ , ξ_x and ξ_y within U and D .

260 Discretizing this equation results in a valid GMRF that is nonstationary. In particular, the autoregressive B matrix from Section 2.3 could have different elements in each row based on the variation in $H(\mathbf{s})$ or $\kappa(\mathbf{s})$. However, B will still be a sparse matrix and Q will be positive-definite. One can connect the matrix B^{-1} as a discretization of a kernel Ψ to the lattice points, as well as the covariance matrix $B^{-1}B^{-T}$ as a discrete approximation to the integral in (1).

270 The process variance can also be allowed to vary in the same way as with the nonstationary Matérn model, but this must be done balancing the identifiability of κ and H and the fact that edge effects may introduce spurious variation in the GMRF variance. Our approach is to first construct the precision matrix and then use sparse matrix methods to solve for the diagonal elements of the covariance matrix. The rows of B are then weighted so that this new version gives a GMRF with constant marginal variance. With this normalization of the SAR model $\sigma(\mathbf{s})^2$ can be introduced to capture explicit spatial variation in the process marginal variance.

275 **3. Local moving window likelihood estimation**

3.1. Local estimation strategy

280 Estimating a nonstationarity model can be challenging due to the increased number of covariance parameters. When enough data is available, however, local estimation can give insight into what type of nonstationarity is present. Moreover, we show in this section that a modest number of spatial replicate fields results in stable local covariance estimates.

285 Local estimation is usually accompanied by the assumption of approximate local stationarity. For this work, we define local stationarity and the local likelihood estimation technique for a Gaussian process with stationary Matérn covariance as follows. First, divide the region of interest \mathcal{D} into M possibly overlapping subregions, or *windows*, $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_M$. Then under the assumption of approximate local stationarity, we can model the data \mathbf{y}_i within the subregion \mathcal{D}_i using a Gaussian process Y_i defined using the following specification:

$$\mathbf{y}_i(\mathbf{s}) = \mu_i(\mathbf{s}) + Z_i(\mathbf{s}) + \epsilon_i(\mathbf{s}) \quad (6)$$

where $\epsilon_i \sim \text{WN}(0, \tau_i^2)$ is spatial white noise, and $Z_i \sim \text{GP}(\mathbf{0}, F_i)$ is a spatially correlated Gaussian process with covariance matrix F_i parameterized by an anisotropic but stationary Matérn covariance function. Let $G_i = F_i + \tau_i^2 I$. The approximate Gaussian process likelihood L based on p replicates \mathbf{y}_i is

$$\log L(\nu, \boldsymbol{\xi}_i, \theta_i, \sigma_i) = -\frac{np}{2} \log 2\pi + \frac{1}{2} \log |G|^{-1} - \sum_{i=1}^p \frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_i)^T G_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \quad (7)$$

where $\boldsymbol{\mu}_i$ is the mean function μ_i evaluated at the locations of \mathbf{y}_i , and G_i is the covariance matrix for \mathbf{y}_i and depends implicitly on the parameters enumerated in the likelihood. The likelihood is approximate because we are assuming stationarity within each data window.

After partitioning the data, finding each local likelihood estimate is an embarrassingly parallel task, which makes it a viable strategy for large data sets by using many computational cores. In fact, in our application the parallelization is efficient to the point that we take the subregions to be an exhaustive set of moving windows centered at every grid point. We assign these estimated parameters to the location of the center of the subregion \mathcal{D}_i , and after translating into the SAR parameterization these specify the row of B , the SAR matrix, at this location. This assignment is, of course, predicated on the assumption that over the region there is little variation in these parameters. This issue will be discussed in more detail in the last section.

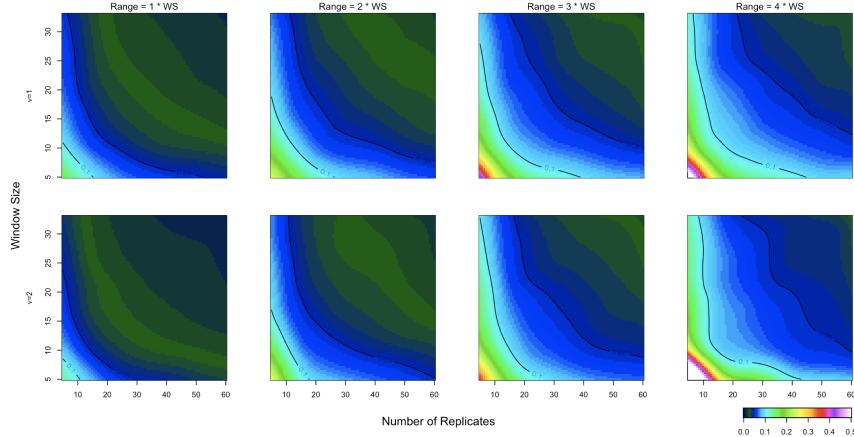
Given that the SAR model also gives a specification of the covariance it may seem indirect that the local estimates focus on the Matérn model, and then the estimates are transformed into the SAR representation. An alternative would be to estimate the SAR version directly from local likelihood windowing. There are several reasons for the two-step procedure. Fitting the covariance model directly avoids any boundary effects that would come about by applying the SAR to a small window. Also, the local fitting is by definition small in size and sparse matrix methods associated with the SAR will not be as efficient as the direct ML estimation using dense covariance matrices. Finally, the Matérn parameters are easier to interpret and will be more simple to model in a hierarchical statistical framework.

3.2. Estimation accuracy of the Matérn range parameter

A practical issue for a local approach, especially in the context of determining covariance parameters, is whether the number of replicates and the size of the window are adequate for robust estimation of parameters. Although choosing a data adaptive window is beyond the scope of this work, it is important to identify the conditions under which parameter estimates will be accurate. As a side issue, it is also useful to understand the benefits of replicate spatial fields in estimating a covariance model. In particular, the hope is that replication

325 makes it possible to estimate correlation ranges that are much larger than the local window size. Here we focus on the Matérn covariance family in estimation because of its prevalence, flexibility, interpretability, and the useful theoretical guidance concerning estimation of the range and variance parameters [18].

330 The Monte Carlo experiment is organized with four factors: window size ranging between 5×5 grid and a 33×33 grid, a Matérn range parameter being multiples of 1, 2, 3, and 4 times the window size, the Matérn smoothness parameter taking on values 1 and 2, and the number of replicates ranging between 5 and 60. Thus the full factorial design is $11 \times 4 \times 2 \times 9$ (window size \times range parameter \times smoothness parameter \times replicates). For each combination, 335 replicates with given range and smoothness were simulated with given window size, and the Matérn range was estimated using maximum likelihood. This was done 100 times for each combination, and statistics were assembled from the 100 independent MLEs for the range parameter. The main quantity of interest is the percent error of the estimate, and so percent error surfaces as a function of replicate number and window size are summarized in Figure 2. In this experiment, the range parameter was varied based on the window size which may seem unusual. However, the motivation was to address the computational requirements of the problem: given a computational budget to accommodate windows of a specific size, what size range parameter can be accurately estimated? Note that with constraints on the window size, accuracy can also be improved by increasing the number of replicates.



340 Figure 2: Each panel displays the absolute percent error from estimating the Matérn range parameter given a certain number of replicates and a window size (size of grid). Fixed Matérn range parameters one, two, three, and four times the size of the grid were tested, faceted in columns (a)-(d). The top row corresponds to $\nu = 1$ and bottom to $\nu = 2$. Thin plate splines were fit using the 100 repeated optimization results, performed at each grid location. The splines were used to predict the surfaces shown. Note that white indicates $> 50\%$ error

These surfaces can be used as guidelines to decide how many replicates are necessary and what window size should be used to achieve a specific estimation error tolerance, given something is known about the size of the range to be estimated. The results are encouraging; e.g. only a small number of replicates (> 10) are needed with a window size of > 10 to estimate a range of 10. In the extreme case, a Matérn range four times the size of the window might be estimated to within 10% error if 30 replicates are available and using a window size of 10 or greater. Using these guidelines, we can be more confident that local moving window likelihood estimation is a viable technique if enough data is used.

4. NCAR LENS Data

The data set from the NCAR CESM Large Ensemble project [19] is comprised of 30 spatial fields that can be assumed to be independent replicates from the same distribution. This feature is based on the nature of climate model experiments run over a long period and started with different initial conditions. Nychka et al. [23] first analyzed these data using the LatticeKrig model, and the original article details the climate science application. The data locations are on a 288×192 grid with approximately one degree resolution, covering the entire globe. Details about the pattern scaling approach to statistical emulation can also be found in [1]. Briefly, each field is a measure of how the local temperature average is affected by a global temperature average increase of one degree Celsius. Generating this ensemble involved a large amount of supercomputer resources. The statistical task is represent these spatial fields with a probability distribution where is more efficient to generate additional fields (e.g. several hundred or thousands) that track the original 30 member model results.

To streamline the example, we focus on the subregion including the Americas and surrounding oceans containing 13,052 observations on a 102×128 grid. The top row of Figure 4 shows the first four sample fields from the data set we analyze. The one data modification from [23] is that, in addition to subtracting the ensemble mean from each grid box, we have also standardized the fields by dividing by the ensemble standard deviation of each grid box.

4.1. Covariance parameter estimates

Moving window local MLEs were found using window sizes between 8×8 and 15×15 . Among these choices there was little change in the estimates and subsequent analysis used a 9×9 window. This window size is consistent with the long range correlations over the ocean and also the estimate from Section 3.2. The estimation was performed on the NCAR Cheyenne supercomputer [4] using the R programming language [26] with the `Rmpi` [33] and `fields` packages

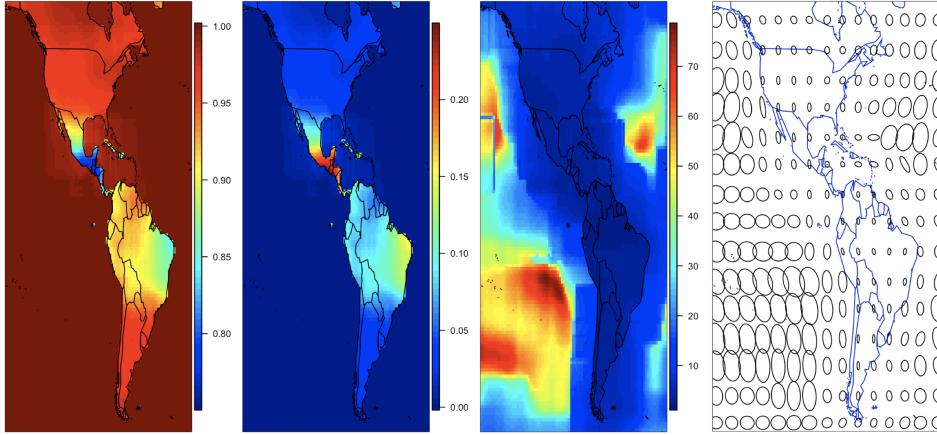


Figure 3: The results of the moving window likelihood estimation. The sill $\sigma^2(\mathbf{s})$ (a), nugget $\tau^2(\mathbf{s})$ (b), geometric average range $\sqrt{\xi_x(\mathbf{s})\xi_y(\mathbf{s})}$ (c), and anisotropy ellipses defined by $A^T(\mathbf{s})A(\mathbf{s})$ (d)

[6]. The details of the parallel implementation are the same as in [23]. Since the fields were standardized, the constraint $\sigma^2 = 1 - \tau^2$ was included.

The estimates for the spatially varying parameters are shown in Figure 3.
 385 The sill and nugget variances are shown in (a) and (b). Panel (c) shows the
 geometric mean of ξ_x and ξ_y as a measure of the average correlation range, and
 this also agrees with the range in the isotropic case. Finally in panel (d), the
 estimated anisotropy matrix $A(\mathbf{s}_i)$ is depicted by glyphs indicating the range
 390 and departure from isotropy. The large signal to noise ratio σ^2/τ^2 (not shown)
 and the evident transition in the covariance structure between land and ocean
 suggests that the nonstationarity in the second-order structure of the data is
 being accurately estimated. Based on the coastlines in some regions, we think
 that the addition of a land/ocean covariate may be useful.

4.2. Model checking

395 The nonstationary SAR model is convenient for plugging in locally estimated
 parameters and capable of modeling large data sets. For this reason, we chose
 to translate the local Matérn parameters into their approximate SAR parameter
 equivalents. The translation was done using the numerical relationship derived
 400 in this 2.5. Then, the local SAR parameters were encoded into the nonstationary
 SAR model. Simulations from this covariance are shown in the bottom row of
 Figure 4. The simulations do a reasonable job emulating the data, but are
 lacking some of the long range anisotropy over the ocean.

To illustrate how the nonstationarity estimated for this model is related to
 land/ocean boundaries, Figure 5 shows several different locations in the spatial

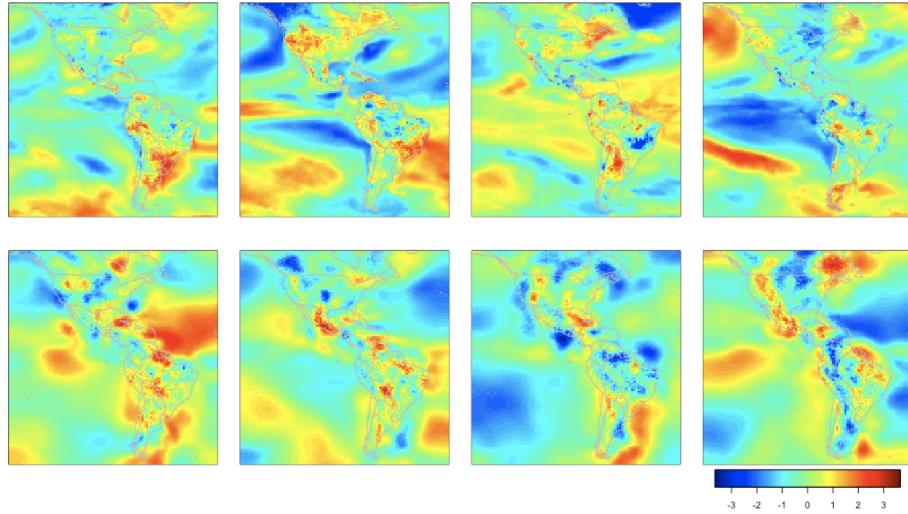


Figure 4: The top row consists of the first four ensemble members from the NCAR CESM data set. The bottom row shows four unconditional simulations from the nonstationary SAR model

405 domain and plots the correlations implied by the nonstationary SAR model. Both anisotropy and nonstationarity are evident in Figure 5.

The SAR representation as a global model for the spatial field provides a convenient way to check the model fit. Under the assumption that the nugget variance is small relative to the smooth Gaussian process, the SAR matrix should decorrelate the observed field. Let \mathbf{y} be the observed field with covariance matrix Σ . The simple idea is to factor Σ as $A^{-1}A^{-T}$ and then check that $A\mathbf{y}$ is a white noise field, or at least a spatial process with greatly reduced spatial dependence. Note that the choice of A is not unique, but it makes sense to choose a version of the square root that has weights that are localized around each observation location. In this analysis

$$\Sigma = \sigma^2 B^{-1} B^{-T} + \tau^2 I = \sigma B^{-1} (I + \tau^2 Q) \sigma B^{-T},$$

where Q is the precision matrix. If the $\tau = 0$ then the SAR matrix provides a transformation to white noise that is justified, and its approximate will be due just to the local stationary assumption in the fitting and the translation from the Matern to the SAR, which are both negligible as we have shown. then the SAR matrix gives a version of A that is localized. MORE EXPLANATION If τ^2 is small relative to σ^2 , then $B\mathbf{y}$ will have covariance $(I + \tau^2 Q)$ and will approximate a white noise field. Small τ^2 is a reasonable assumption in practice because one is often interested in simulation and prediction of spatial data that has strong spatial coherence. Note that Q is sparse and when viewed as a

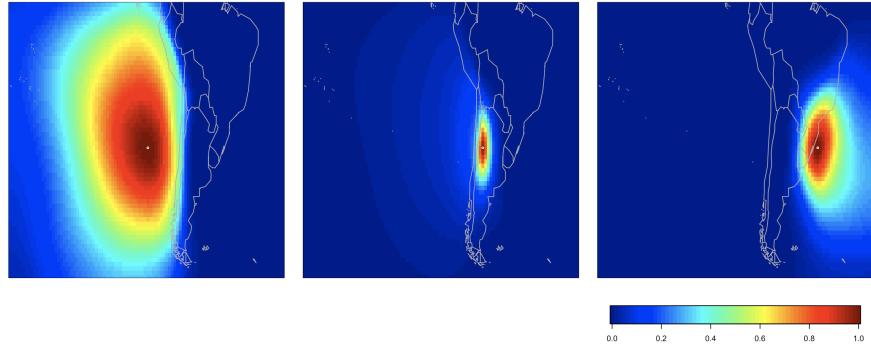


Figure 5: Correlations corresponding to three observation locations at the same latitude implied by the nonstationary SAR model

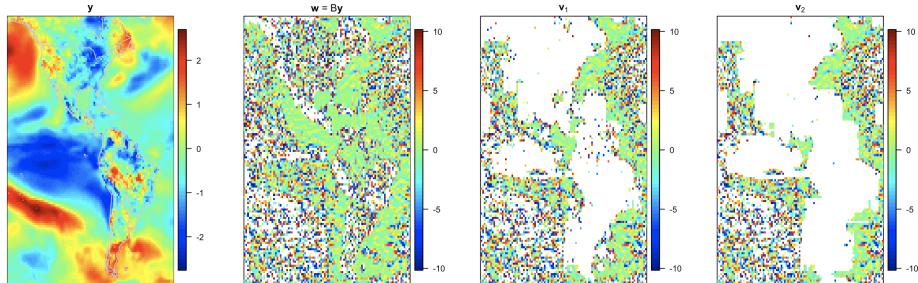


Figure 6: Three decorrelated fields corresponding to successive applications of matrices to one of the spatial replicates in the data. In (a), the symmetric square root of the precision matrix B has been applied to a data replicate \mathbf{y} yielding $\mathbf{w} = \mathbf{B}\mathbf{y}$. The first-order approximation is then applied in (b) giving $\mathbf{v}_1 = (I - \frac{1}{2}\tau^2 Q)\mathbf{w}$, and the second-order approximation $\mathbf{v}_2 = (I - \frac{1}{2}\tau^2 Q - \frac{3}{8}\tau^4 Q^2)\mathbf{w}$ is shown in (c).

covariance matrix will have localized, finitely supported correlations. Avoiding explicitly calculating A (which requires dense matrix computation), we propose a refinement of just using B as a decorrelation transformation: in addition, approximate the inverse square root of $(I + \tau^2 Q)$ with a series expansion to find a more accurate approximation to A . From elementary power series expansions we have

$$1/\sqrt{1+u} = 1 - \frac{1}{2}u + \frac{1 \cdot 3}{2 \cdot 4}u^2 - \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6}u^3 + \dots,$$

and provided that $\|\tau^2 Q\| < 1$ one can use this power series to approximate $(I + \tau^2 Q)^{-1/2}$ as a symmetric matrix. For example, a first-order correction would be $I - (1/2)\tau^2 Q$ and it easily applied since Q is formed from the sparse SAR matrix. Thus, for transforming a limited number of fields one would never construct the matrix explicitly, but rather just multiply the fields of interest by successive power of Q .

Figure 6(b) depicts the result of B applied to one of the replicates (shown in (a)) to which the model was fitted. To carry out this matrix multiplication, the spatial field is flattened into a vector in the same order specified by the covariance matrix. In panels (c) and (d), the first- and second-order corrections were applied to the field in (b). As a diagnostic tool one can visually assess the goodness-of-fit of the model covariance matrix to the spatial distribution of the data using these techniques. If the spatial distribution of the data is fitted accurately, this process should result in a decorrelated field of white noise. Excluding the slight heteroskedasticity present near coastal regions, Figure 6 indicates that the vast majority of the correlation in the data has been captured in the model, and therefore has been removed from the data via these matrix transformations. This success is encouraging given the long range correlations over the ocean that have been identified from local estimation, and as a SAR only operates on second-order neighbors. A formal test of independence was not implemented on the decorrelated fields, although this could be used as a more general goodness-of-fit test in covariance modeling.

5. Conclusion

In this paper, we have investigated a two-step framework of local estimation and global encoding to represent large and nonstationary covariance functions. We have shown that when independent replicates of spatial data are available where locally stationarity holds, local ML estimation is a robust technique for estimating the nonstationarity in the covariance parameters. in particular, the Monte Carlo results indicate the climate model example falls within this context.

We also explored the stationary Matérn-SAR covariance model approximation, conducting a numerical experiment to compare against existing results.

The analytic approximation between the models is not reliable for long correlation ranges; however, we can use a numerical approximation to translate parameters between the Matérn and SAR models more accurately. To our knowledge 440 this is the first time detailed numerical mappings have been made between the anisotropic SAR model and an anisotropic Matérn covariance function.

An important contribution of this work is showing nonstationary data can 445 be modeled by combining local maximum likelihood estimation with a simple global nonstationary covariance model that is straightforward to implement. We focused on encoding the locally estimated parameters in the nonstationary SAR model. In addition, the multistage approach is computationally efficient and can 450 be applied to very large spatial data sets: local estimation avoids the big n problem of global estimation, and encoding local estimates in a SAR model allows us to exploit sparsity for prediction and simulation. Another major advantage 455 of this method is that it can be applied to both continuously indexed and lattice data. To reduce the scope of this work we have focused on lattice data. Although this restricted format will continue to be standard for climate models, the SAR model can also be extended to irregularly spaced data. One approach for non-lattice spatial data is the LatticeKrig model that imposes the SAR and lattice structure on coefficients in a basis function expansion rather than directly on the field. We believe the anisotropic models developed here will carry over 460 for more general models such as basis expansions, and the inverse square root transformation will be an important diagnostic tool for nonstationary modeling.

460 DISCUSS: how to generalize this model to irregularly spaced observations and other linear functionals of the process.

AND ADD THIS TO INTRODUCTION

References

- [1] Alexeeff, S.E., Nychka, D., Sain, S.R., Tebaldi, C., 2018. Emulating mean 465 patterns and variability of temperature across and within scenarios in anthropogenic climate change experiments. Climatic Change 146, 319–333.
- [2] Anderes, E.B., Stein, M.L., 2011. Local likelihood estimation for nonstationary random fields. Journal of Multivariate Analysis 102, 506–520.
- [3] Anderes, E.B., Stein, M.L., et al., 2008. Estimating deformations of 470 isotropic gaussian random fields on the plane. The Annals of Statistics 36, 719–741.
- [4] Computational and Information Systems Laboratory, 2017. Cheyenne: HPE/SGI ICE XA System (University Community Computing). doi:doi: 10.5065/D6RX99HX.

- 475 [5] Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial
 data sets. *Journal of the Royal Statistical Society: Series B (Statistical
 Methodology)* 70, 209–226.
- 480 [6] Douglas Nychka, Reinhard Furrer, John Paige, Stephan Sain, 2017. *fields:*
 Tools for spatial data. URL: www.image.ucar.edu/~nychka/Fields,
 doi:10.5065/D6W957CT. R package version 9.6.
- [7] Fuentes, M., 2002. Spectral methods for nonstationary spatial processes.
Biometrika 89, 197–210.
- 485 [8] Fuentes, M., Smith, R.L., 2001. A new class of nonstationary spatial models.
 Technical Report. Technical report, North Carolina State University,
 Raleigh, NC.
- [9] Fuglstad, G.A., Lindgren, F., Simpson, D., Rue, H., 2015a. Exploring a
 new class of non-stationary spatial gaussian random fields with varying
 local anisotropy. *Statistica Sinica* , 115–133.
- 490 [10] Fuglstad, G.A., Simpson, D., Lindgren, F., Rue, H., 2015b. Does non-
 stationary spatial data always require non-stationary random fields? *Spatial
 Statistics* 14, 505–531.
- [11] Haas, T.C., 1990a. Kriging and automated variogram modeling within a
 moving window. *Atmospheric Environment. Part A. General Topics* 24,
 1759–1769.
- 495 [12] Haas, T.C., 1990b. Lognormal and moving window methods of estimating
 acid deposition. *Journal of the American Statistical Association* 85, 950–
 963.
- 500 [13] Heaton, M.J., Datta, A., Finley, A., Furrer, R., Guhaniyogi, R., Gerber, F.,
 Gramacy, R.B., Hammerling, D., Katzfuss, M., Lindgren, F., et al., 2017.
 Methods for analyzing large spatial data: A review and comparison. *arXiv
 preprint arXiv:1710.05013* .
- [14] Higdon, D., 1998. A process-convolution approach to modelling tempera-
 tures in the north atlantic ocean. *Environmental and Ecological Statistics*
 5, 173–190.
- 505 [15] Higdon, D., 2002. Space and space-time modeling using process convolu-
 tions, in: Quantitative methods for current environmental issues. Springer,
 pp. 37–56.
- [16] Higdon, D., Swall, J., Kern, J., 1999. Non-stationary spatial modeling.
Bayesian statistics 6, 761–768.

- 510 [17] Katzfuss, M., Cressie, N., 2011. Spatio-temporal smoothing and em estimation for massive remote-sensing data sets. *Journal of Time Series Analysis* 32, 430–446.
- [18] Kaufman, C., Shaby, B., 2013. The role of the range parameter for estimation and prediction in geostatistics. *Biometrika* 100, 473–484.
- 515 [19] Kay, J., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J., Bates, S., Danabasoglu, G., Edwards, J., et al., 2015. The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society* 96, 1333–1349.
- 520 [20] Lindgren, F., Rue, H., 2007. Explicit construction of gmrf approximations to generalised matérn fields on irregular grids. *Preprints in Mathematical Sciences* 5.
- [21] Lindgren, F., Rue, H., 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, 423–498.
- 525 [22] Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., Sain, S., 2015. A multiresolution gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics* 24, 579–599.
- 530 [23] Nychka, D., Hammerling, D., Krock, M., Wiens, A., 2018. Modeling and emulation of nonstationary gaussian fields. *Spatial Statistics* .
- [24] Nychka, D., Wikle, C., Royle, J.A., 2002. Multiresolution models for non-stationary spatial covariance functions. *Statistical Modelling* 2, 315–331.
- 535 [25] Paciorek, C.J., Schervish, M.J., 2004. Nonstationary covariance functions for gaussian process regression, in: *Advances in neural information processing systems*, pp. 273–280.
- [26] R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- 540 [27] Risser, M.D., 2016. Nonstationary spatial modeling, with emphasis on process convolution and covariate-driven approaches. *arXiv preprint arXiv:1610.02447* .

- 545 [28] Risser, M.D., Calder, C.A., 2015. Local likelihood estimation for covariance
functions with spatially-varying parameters: the convospat package for r.
arXiv preprint arXiv:1507.08613 .
- 550 [29] Rue, H., Held, L., 2005. Gaussian Markov random fields: theory and
applications. CRC press.
- [30] Sampson, P.D., Guttorp, P., 1992. Nonparametric estimation of nonsta-
tionary spatial covariance structure. Journal of the American Statistical
Association 87, 108–119. URL: <http://www.jstor.org/stable/2290458>.
- 555 [31] Simpson, D., Lindgren, F., Rue, H., 2012. Think continuous: Markovian
gaussian models in spatial statistics. Spatial Statistics 1, 16–29.
- [32] Ver Hoef, J.M., Cressie, N., Barry, R.P., 2004. Flexible spatial models for
kriging and cokriging using moving averages and the fast fourier transform
(fft). Journal of Computational and Graphical Statistics 13, 265–282.
- 560 [33] Yu, H., 2002. Rmpi: parallel statistical computing in R. R News 2, 10–14.
- [34] Zhu, Z., Wu, Y., 2010. Estimation and prediction of a class of convolution-
based spatial nonstationary models for large spatial data. Journal of Com-
putational and Graphical Statistics 19, 74–95.