Received: 23 December 2019 | Revised: 28 July 2020 | Accepted: 3 August 2020

DOI: 10.1002/env.2652

SPECIAL INVITED ARTICLE

WILEY

Modeling spatial data using local likelihood estimation and a Matérn to spatial autoregressive translation

Ashton Wiens | **Douglas Nychka** | **William Kleiber**

Department of Applied Mathematics Statistics, University of Colorado, Boulder, Colorado

Correspondence
Douglas Nychka, Department of Applied Mathematics Statistics, University of Colorado, 1500 Illinois St, Golden, CO 80401.
Email: douglasnychka@gmail.com

Funding information
NSF Awards, Grant/Award Numbers: DMS-1811294, DMS-1923062

Modeling data with nonstationary covariance structure is important to represent heterogeneity in geophysical and other environmental spatial processes. In this work, we investigate a two-stage approach to modeling nonstationary covariances that is efficient for large data sets. First, maximum likelihood estimation is used in local, moving windows to infer spatially varying covariance parameters. These surfaces of covariance parameters are then encoded into a global covariance model specifying the second-order structure for the complete spatial domain. From this second step, the resulting global model allows for efficient simulation and prediction. This work uses a nonstationary spatial autoregressive (SAR) model, related to Gaussian Markov random field methods, as the global model which is amenable to plug in local estimates and practical for large datasets. A simulation study is used to establish the accuracy of local Matérn parameter estimation as a reliable technique for small window sizes and a modest number of replicated fields. This modeling approach is implemented on a nonstationary climate model dataset with the goal of emulating the variation in the numerical model ensemble using a Gaussian process.

KEYWORDS

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

1 | INTRODUCTION

Climate models produce high-dimensional spatial fields of variables related to various processes that comprise the Earth system. To quantify uncertainty, ensembles are generated typically by perturbing initial conditions to the climate model; however the ensemble size is usually limited to a handful of members due to the extreme computational demands of such codes. An alternative approach is to emulate the climate model output using a statistical model from which uncertainty can be readily derived. Many temperature fields appear to be well approximated by a Gaussian process but the covariance structure is distinctly nonstationary. This paper focuses on statistical emulation of high-dimensional climate model spatial output that exhibit substantial nonstationarity in the spatial covariance. The major challenges are in specifying a flexible nonstationary model that is amenable to estimation for large datasets, but also allows for computationally efficient simulation.

We investigate a two-stage approach to estimating and modeling nonstationary covariances, similar to the methodology in Nychka, Hammerling, Krock, and Wiens (2018). First, assuming the field is approximately locally stationary,

This article is based on the President's invited Lecture from the 28th Annual Conference of the International Environmetrics Society (TIES), presented 16 July 2018.

we perform moving window local likelihood estimation to infer spatially varying Matérn covariance parameters. In the approach of Nychka et al. (2018) these parameters are mapped into those of the LatticeKrig model (Nychka, Bandyopadhyay, Hammerling, Lindgren, & Sain, 2015). A more direct approach for regularly spaced observations is to exploit a relationship between the Matérn parameters and those of a spatial autoregressive (SAR) random field, to reproduce local Matérn correlations. Finally, the spatially varying parameters are encoded into a global SAR precision matrix, specifying the global field's dependence structure simultaneously. As this work is at the intersection of nonstationary modeling, local estimation and Matérn-SAR connections, we begin with a brief review on these distinct topics.

There are many general classes of nonstationary models, such as deformation methods (Anderes & Stein, 2008; Sampson & Guttorp, 1992), basis function methods (Cressie & Johannesson, 2008; Kang & Cressie, 2013; Kang, Cressie, & Sain, 2012; Katzfuss & Cressie, 2011; Nychka et al., 2015; Nychka, Wikle, & Royle, 2002), process-convolutions (Fuentes, 2002; Fuentes & Smith, 2001; Higdon, 1998, 2002; Higdon, Swall, & Kern, 1999; Paciorek & Schervish, 2003; Zhu & Wu, 2010), and the stochastic partial differential equation (SPDE) approach (Lindgren & Rue, 2007; Lindgren, Rue, & Lindström, 2011; Rue & Held, 2005; Simpson, Lindgren, & Rue, 2012).

Both the spatial random effects model (SRE) (Cressie & Johannesson, 2008; Kang et al., 2012; Kang & Cressie, 2013) and LatticeKrig (Nychka et al., 2015) use spatial basis functions at multiple resolutions to capture nonstationarity. The inference for these spatial models includes estimating the distributions of the weights, or coefficients, on these basis functions. Thus, computation reduces from the order of the number of observations in the dataset to the order of the number of basis functions. In the SRE model, the weights are multivariate normal with a general covariance matrix. In the LatticeKrig model, the weights are multivariate normal with a Gaussian Markov random field (GMRF) structure imposed that increases the computational efficiency even further.

The setup of both of these models using randomly weighted basis functions significantly decreases the number of parameters to estimate in the model compared to the method we develop here. However, the local estimation technique in our method provides an exploratory tool for assessing nonstationarity. Additionally, the variance reduction gained by windowing the data can yield fields of covariance parameters that are useful as an end in themselves.

Here, we investigate two existing nonstationary models from the GMRF families, and focus our attention on computationally efficient estimation using localized moving windows. Risser (2016) provides a review of nonstationary models and Heaton, Datta, and Finley (2019) give a review of methods for large spatial datasets. Local estimation is a well-established idea in spatial statistics (Haas, 1990a, 1990b; Risser & Calder, 2017; Ver Hoef, Cressie, & Barry, 2004), and is popular in that full likelihood-based calculations, which become prohibitive for large sample sizes, can be circumvented. Moreover, this strategy is an easily parallelizable problem which can lead to further computational improvements. In practice, there is often no clear indication of which parameters in the model should vary spatially (Fuglstad, Simpson, Lindgren, & Rue, 2015) 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 lead to a diagnostic as to whether the parameters should be constrained to be constant or vary over space. Furthermore, with local estimation, we do not have to decompose the covariance fields into some prespecified low-dimensional representation (Fuglstad et al., 2015a; Marques, Klein, & Kneib, 2019; Risser, 2016). Weighted local likelihoods have been studied to accommodate irregularly spaced observations (Anderes & Stein, 2011), but in this work we use a simple moving window applied to data on a lattice. We focus on estimation of locally varying Matérn parameters, primarily because of their interpretability and to exploit a relationship between Matérn and SAR covariance models. To justify local estimation as a reliable technique, we use a Monte Carlo experiment to study the robustness of local estimation of the correlation range parameter.

A drawback in our approach is the need for spatial replicates to estimate nonstationary structure. Although this may seem overly restrictive, the availability of independent replicates are becoming the norm in climate simulations. Moreover, many space-time datasets can be approximated as replicates in space when the time dependence is weak or explained by fixed effects. Finally, we note that nonstationary covariance models that are estimated from replicates reduce the problem of separating the variation in the covariance function from the variation in the spatial field itself. We believe that working with a single realization of a spatial field is much more difficult.

For single realizations of a spatial field, basis function expansions, developed by Kang et al. (2012), for the individual responses of regional climate models, and the recent work by Marques et al. (2019), for regional rainfall, are useful ways to capture nonstationarity with a global model and so provide an alternative to local fitting. One can fit these models

based on the reduced degrees of freedom that is implied by the basis representation. This basis representation will not be effective, however, if the field has more spatial detail than can be represented by a reduced basis.

With locally estimated covariance parameters in hand, some care is required to combine these into a valid global nonstationary covariance model. In this work, we use a nonstationary SAR model, related to the GMRF approximation to GPs. The idea is to identify members of the Matérn family of spatial processes as solutions to a SPDE. The SPDE is then discretized to a lattice and this motivates the form of the SAR (Lindgren et al., 2011). The correspondence between the Matérn/SPDE form and a SAR was presented in Lindgren et al. (2011), 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 we can successfully translate the Matérn formulation into a SAR framework, we can exploit sparse matrix algorithms for fast computation.

An important contribution of this work is showing nonstationary spatial processes can be modeled by combining local maximum likelihood estimation with a simple global nonstationary covariance model that is straightforward to implement. As an illustration, we apply this two-stage modeling framework to analyze climate fields related to surface temperature and simulated from the Community Earth System Modeling's Large Ensemble (LENS) project (Kay et al., 2015). These data comprise about 13,000 spatial locations over the Americas, and exhibit strong nonstationarities that challenge the construction of statistical emulators.

2 | STATIONARY COVARIANCE MODELS

In this section, we discuss the connection between the isotropic Matérn family of covariance models for Gaussian processes with the spatial autoregression (SAR) construction for GMRFs, and quantify this relationship in a numerical study.

2.1 | The Matérn covariance model

Let $f(\boldsymbol{x})$ be a mean zero Gaussian process on $\boldsymbol{x} \in \mathbb{R}^2$ with covariance function $k(\boldsymbol{x}, \boldsymbol{x}') \in \mathbb{R}$. 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 parameter is

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

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

$$k(\boldsymbol{x}, \boldsymbol{x}') = C(\kappa d \mid \nu, \sigma^2).$$

2.2 | The SAR model

In contrast to modeling a covariance function for a process of continuous spatial variation, the SAR model parameterizes the precision matrix for the process on a discrete lattice. For the following development we denote by $\mathbf{y}$ a Gaussian process on an infinite regular lattice in $\mathbb{R}^2$.

Denote by $y_{i,j}$ the element of $\boldsymbol{y}$ at lattice location (i,j). An homogeneous, approximately isotropic SAR model can be written using lattice notation (Lindgren et al., 2011) as

$$\begin{array}{|c|c|c|}\hline 0 & -1 & 0 \\ \hline -1 & 4+\kappa_S^2 & -1 \\ \hline 0 & -1 & 0 \\ \hline \end{array}, \tag{1}$$

which visually illustrates a set of decorrelating weights on the random vector $\boldsymbol{y}$. To be precise, we interpret (1) as implying that the following equation holds

$$(4+\kappa_S^2)y_{ij} - (y_{i-1,j} + y_{i+1,j} + y_{i,j-1} + y_{i,j+1}) = e_{ij}, \tag{2}$$

for a mean zero unit variance normally distributed white noise vector $\boldsymbol{e} = \{e_{ij}\}$.

Here $\kappa_S > 0$ is suggestive of a range parameter controlling the correlation length scale of the field and is similar, but not identical, to κ for the continuous Matérn case above. For the model in (1), one can populate a matrix B using (2) such that $B\boldsymbol{y} = \boldsymbol{e}$. Furthermore, it is clear that if $\kappa_S > 0$, the diagonal dominance of B guarantees its invertibility. With $\boldsymbol{y} = B^{-1}\boldsymbol{e}$, the covariance matrix for $\boldsymbol{y}$ is $B^{-1}B^{-T} = (B^TB)^{-1}$ and thus $\boldsymbol{y}$ has precision matrix $Q = B^TB$. Moreover, the precision matrix implied by the SAR model is sparse. The SAR model is amenable to modeling large datasets because the sparse precision matrix can be used in place of a dense covariance matrix for likelihood estimation and simulation. Following the ideas from Lindgren et al. (2011) one can iterate the SAR weights to obtain higher-order models that approximate smoother processes. For example, if $BB\boldsymbol{y} = \boldsymbol{e}$, this implies a precision matrix $Q_2 = (BB)^T(BB)$. The SAR model detailed here is a special case of a GMRF. For a given row of the precision matrix, the nonzero, off-diagonal entries index the neighbors that determine the Markov property. Citing the *order* of neighborhoods is ambiguous depending on whether one is referring to the SAR weight matrix, the precision matrix of the process, or the weight matrix in the related CAR model. In particular, the first-order SAR described above corresponds with a third-order CAR (Ver Hoef, Hanks, & Hooten, 2018), and the nonzero elements in Q will include second-order neighbors. Two additional points in applying the SAR model are important. For a given dataset $\boldsymbol{y}$ which is not defined on an infinite lattice, but rather a (typically) rectangular one, the stencil in (1) should be modified at the boundaries of the domain. The center value of the stencil should be the κ_S^2 plus the sum of the weights of its nonzero neighbors, although a simpler option is to artificially extend the lattice a few nodes outward to reduce boundary effects. Second, the value of κ_S has a one-to-one relationship with the marginal variance of the process. It is also important to normalize so that the SAR has variance 1 at all grid points.

A version of the SAR model that exhibits approximate stationarity and also geometric anisotropy will be detailed in Section 3.

2.3 | The Matérn-SAR link

Lindgren and Rue (Lindgren et al., 2011) developed an approximation of Gaussian random fields with Matérn covariance functions using GMRFs with particular SAR structures. The connection is established through a SPDE formulation in 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 $\alpha = \nu + \frac{d}{2}$, $\kappa > 0, \nu > 0$, $\mathbf{s} \in \mathbb{R}^d$, $d = 1$ or 2, and $\mathcal{W}(s)$ is a Gaussian white noise process with zero mean and variance σ^2. As in the Matérn model, ν controls the smoothness of realizations of the Gaussian field. Fixing $\nu = 1$ and $d = 2$, Lindgren et al. (2011) first proposed 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, however, is approximate. Similar results can be obtained for different smoothness parameters ν by convolving the finite difference stencil in (1) with itself ν times, as detailed in the previous section for $\nu = 1$ and $\nu = 2$.

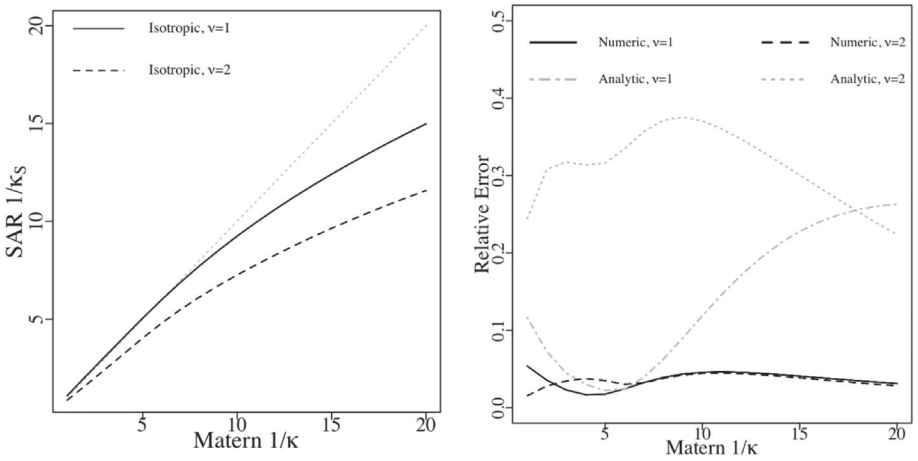

FIGURE 1 For the isotropic case, the optimal $1/\kappa_S$ parameter for a given Matérn inverse range $1/\kappa$ is plotted in (a). The relative error incurred by using the spatial autoregression model with optimal κ_S as an approximation to the Matérn model is shown in (b) (estimated relationship in black lines, Lindgren et al., 2011 relationship of $\kappa_S = \kappa$ in gray lines)

2.4 | Numerical translation of range parameters between the Matérn and SAR models

The connection between the isotropic Matern family and a SAR relies on the approximation of a discretized Laplacian operator with finite differences of the field on a lattice. To develop an accurate statistical model, it is important to quantify the error in such an approximation and improve its calibration beyond the limiting expression suggested in Lindgren et al. (2011). In this section we provide numerical evidence to show that an accurate calibration is possible if restricted to specific ranges of the covariance parameters.

Our calibration setup is as follows: given a Matérn range parameter κ, we estimate the value of κ_S in the SAR model which gives the best approximation to the Matérn correlation function. We consider 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 for the process evaluated on the lattice grid. Then, we find the optimal κ_S from the SAR model by computing its equivalent correlation matrix (the standardized inverse precision matrix) and minimizing distance between the Matérn and SAR correlation matrices.

It is known that the SAR covariance model suffers from edge effects. To avoid the interference of edge effects, we quantify the difference between the two correlation matrices by only comparing the correlations from the central lattice point under both models. For an $N \times N$ lattice of locations, with N odd, let $\boldsymbol{\omega}_\kappa$ denote the vector of correlations between the center point in this lattice and all other locations based on the Matérn covariance function with range κ. Let $\boldsymbol{\omega}_{\kappa S}$ be the analogous correlation vector for the SAR model with range parameter κ_S. Note that for the SAR correlations we normalize so the marginal variance is equal to 1. We then find

$$\hat{\kappa}_S(\kappa) = \underset{\kappa_S}{\operatorname{argmin}} \|\boldsymbol{\omega}_\kappa - \boldsymbol{\omega}_{\kappa S}\|_2.$$

If the relationship proposed by Lindgren et al. (2011) holds we would expect $\hat{\kappa}_S(\kappa) \approx \kappa$.

κ^{-1} is varied over the interval [1,20] with $N = 51$ and lattice points having unit spacing. The approximation results are summarized in Figure 1. In Figure 1a $\hat{\kappa}_S(\kappa)^{-1}$ is plotted as a function of κ^{-1}. The solid line corresponds to the $\nu = 1$ case, the dashed line corresponds to $\nu = 2$, and the dotted line shows the theoretical relationship, $\kappa_S^{-1} = \kappa^{-1}$ from Lindgren et al. (2011). From this experiment we conclude that, at least over this distance scale, it is important not to rely on the analytic formula to translate between κ and κ_S parameters. Accordingly, we use these numerical results to calibrate the value of κ_S which achieves the best approximation to the desired Matérn covariance with range parameter κ.

The relative error of approximating Matérn correlation with range κ using the SAR correlation with κ_S value derived from the numerical experiment is given by $\|\boldsymbol{\omega}_\kappa - \boldsymbol{\omega}_{\kappa S}\|_2 / \|\boldsymbol{\omega}_\kappa\|_2$ and is shown in Figure 1b. The ℓ_2 distance measure used in the optimization of the model correlation matrices is used to quantify the resulting model error, normalized by $\|\boldsymbol{\omega}_\kappa\|_2$. The relative error incurred when using $\frac{1}{\kappa_S} = \frac{1}{\kappa}$ as shown by the gray lines can be on the order of 20% or worse for higher correlation ranges. However, using the calibrated range (black lines) gives relative errors on the order of 5%.

3 | ANISOTROPIC AND NONSTATIONARY PROCESSES

In this section, we extend the Matérn and SAR models to include geometric anisotropy, verify the accuracy of the approximation, and finally discuss some related nonstationary models.

3.1 | The anisotropic Matérn covariance model

The Matérn family can be extended to include geometric anisotropy (Delfiner & Chilès, 2012) by defining a distance measure based on a linear scaling and rotation of the coordinates. 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

$$D = \begin{bmatrix} \xi_{s_1} & 0 \\ 0 & \xi_{s_2} \end{bmatrix},$$

is a diagonal matrix scaling the s_1 and s_2 coordinate axes separately. Then the pairwise Mahalanobis distance between two locations $\boldsymbol{s},\boldsymbol{s}'$ is defined as $d = \|A\boldsymbol{s} - A\boldsymbol{s}'\| = [(\boldsymbol{s}-\boldsymbol{s}')^T A^T A(\boldsymbol{s}-\boldsymbol{s}')]^{\frac{1}{2}}$, which is used as the argument to the isotropic Matérn covariance function. A useful interpretation of this form is that if one transforms the coordinates according the linear transformation A then the resulting field will be isotropic.

3.2 | The anisotropic Matérn-SAR link

The SAR model can also be extended to incorporate geometric anisotropy. Let H denote a 2×2 symmetric positive definite anisotropy matrix and modify the Laplacian in the pseudodifferential operator as follows

$$(\kappa^2 - \nabla\cdot H\nabla)^{\alpha/2} u(\mathbf{s}) = \mathcal{W}(\mathbf{s}). \tag{3}$$

To avoid potential ambiguity it is helpful to identify the Laplacian operator above for two dimensions in its 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 expression, a first-order finite difference discretization of the anisotropic SPDE at (3) gives the following stencil for filling the rows of the B matrix,

$$\begin{array}{c|c|c} \frac{2H_{12}}{h_{s_1}h_{s_2}} & -\frac{H_{22}}{h_{s_2}^2} & -\frac{2H_{12}}{h_{s_1}h_{s_2}} \\ \hline -\frac{H_{11}}{h_{s_1}^2} & \kappa^2 + \frac{2H_{11}}{h_{s_1}^2} + \frac{2H_{22}}{h_{s_2}^2} & -\frac{H_{11}}{h_{s_1}^2} \\ \hline -\frac{2H_{12}}{h_{s_1}h_{s_2}} & -\frac{H_{22}}{h_{s_2}^2} & \frac{2H_{12}}{h_{s_1}h_{s_2}} \end{array}, \tag{4}$$

where h_{s_1} and h_{s_2} are the grid spacings along the x-axis and y-axis. This is just a reparameterization of the results in Appendix A of Lindgren et al. (2011) but facilitates the practical translation between these models. Moreover, setting $h_{s_1} = h_{s_2} = 1$, $H_{12} = H_{21} = 0$, and $H_{11} = H_{22} = 1$ yields the first-order isotropic model from (1).

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 3, let $\boldsymbol{s}^* = A^{-1}\boldsymbol{s}$, and let u be an isotropic field solution to the SPDE with Laplacian

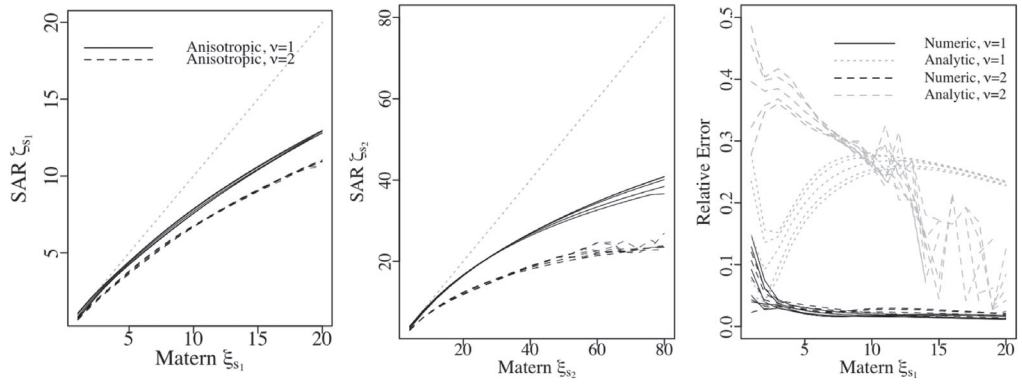

FIGURE 2 For the anisotropic case, the optimal eigenvalues of H, ζ_{s_1} and ζ_{s_2}, are plotted against the ξ_{s_1} and ξ_{s_2}, in panels (a) and (b), and the relative error is shown in (c), analogous to Figure 1

$\nabla \cdot \nabla$. Furthermore, set $u^*(\boldsymbol{s}) = u(A^{-1}\boldsymbol{s}) = u(\boldsymbol{s}^*)$. Then from elementary properties of the gradient

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

and so we have

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

From this expression we identify $H = A^{-T}A^{-1} = U^T D^2 U$. 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. Finally, note that if A is a pure rotation then $H = I$ and isotropy is preserved.

3.3 | Numerical translation of anisotropic range parameters between the Matérn and SAR models

In the climate data analysis below, we find it necessary to include geometric anisotropy in the covariance model. For this reason, we also investigate how the presence of geometric anisotropy affects the numerical correspondence established in Section 2.4. We repeat the experiment from Section 2.4, modified by adding the presence of geometric anisotropy. Let ξ_{s_1} and ξ_{s_2} be the eigenvalues in the Matérn anisotropy matrix, encoded as the diagonal values of D^2, and recall $A = D^{-1}U^T$. Through the connection of the Matérn covariance model to the SPDE formulation, we also have $H = A^{-T}A^{-1} = U^T D^2 U$, so these are the eigenvalues of H. To avoid ambiguity, let H_D be the anisotropy matrix encoded into the finite difference stencil 4, the discretization defining the SAR model. Let the eigenvalues of H_D be denoted ζ_{s_1} and ζ_{s_2}. In essence, in this experiment we fix ξ_{s_1} and ξ_{s_2} in the anisotropic Matérn model and estimate the optimal ζ_{s_1} and ζ_{s_2} in the anisotropic SAR model which best reproduces the correlations in the Matérn model. We do not optimize over the rotation θ and just use the same orthogonal matrix. We follow the same optimization strategy for the isotropic case; noting that if the discretization had no approximation error, we would expect $(\zeta_{s_1}, \zeta_{s_2}) \approx (\xi_{s_1}, \xi_{s_2})$.

Specifically, we encode fixed values for ξ_{s_1} and ξ_{s_2} such that the length scale ratio is $\xi_{s_1} : \xi_{s_2} = 4 : 1$, which is consistent with anisotropic estimates in the data analysis. In particular, we let $\xi_{s_1} = 1, \ldots, 20$ and $\xi_{s_2} = 4\xi_{s_1}$. Then, the optimal values of ζ_{s_1} and ζ_{s_2} are found. The experiment is repeated with fixed rotation angles $\theta = 0°, 10°, \ldots, 90°$ in U, and the rotation angle is assumed to be known and fixed in the eigendecomposition of the H_D matrix. The anisotropic parameter translation results for the optimal ζ_{s_1}, ζ_{s_2} values given ξ_{s_1}, ξ_{s_2} are shown in panels (a) and (b) of Figure 2, respectively. The relative error of approximation $\|\boldsymbol{\omega}_\xi - \boldsymbol{\omega}_\zeta\|_2 / \|\boldsymbol{\omega}_\xi\|_2$ is shown in panel (c). Overall, the behavior of the approximation is similar to the isotropic case: the optimized eigenvalues ζ_{s_1} and ζ_{s_2} of the SAR anisotropy matrix H_D are smaller than the eigenvalues ξ_{s_1} and ξ_{s_2} of the Matérn anisotropy matrix H.

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. Overall, we see an error of 5% or less when using the optimized values of ζ_{s_1} and ζ_{s_2} for $\xi_{s_1} \geq 3$.

3.4 | A nonstationary SAR model

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

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

The SPDE can then be written as

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

where $\kappa(\mathbf{s}) > 0$, $\mathcal{W}(s) \sim \text{WN}(0, \sigma^2(\mathbf{s}))$, $\sigma^2(\mathbf{s}) > 0$, and we assume $H(\boldsymbol{s})$ is everywhere positive definite. Furthermore, we specialize to a spatially varying linear transformation of the coordinates, $A(\boldsymbol{s})$, and so $H(\boldsymbol{s}) = A^{-T}(\boldsymbol{s})A^{-1}(\boldsymbol{s})$. Note that $A(\boldsymbol{s})$ varying in space is equivalent to specifying spatial fields for θ, ξ_{s_1} and ξ_{s_2} within U and D.

Discretizing this equation results in a valid GMRF that is nonstationary. In particular, the autoregressive B matrix from Section 2.2 could have different elements in each row based on the variation in $H(\boldsymbol{s})$ or $\kappa(\boldsymbol{s})$. However, B will still be a sparse matrix and $Q = B^T B$ will always be positive-definite. 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 a normalized precision matrix and then weight each row of B so that this new version gives a GMRF with constant marginal variance. With this normalization of the SAR model $\sigma(\boldsymbol{s})^2$ can be introduced to capture explicit spatial variation in the process marginal variance.

4 | LOCAL MOVING WINDOW LIKELIHOOD ESTIMATION

4.1 | Local estimation strategy

Estimating a nonstationary model can be challenging due to the increased number of covariance parameters. The modeling framework we develop in this work depends on spatial data with independent replicates, and we also work under this assumption in the simulation study. We illustrate in the simulation study, however, that a modest number of replicated fields result in stable local covariance estimates.

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*, $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_M$. Then the assumption of approximate local stationarity is that we can model the data $\boldsymbol{y}_i$ within the subregion $\mathcal{D}_i$ using a stationary Gaussian process Y_i defined using the following specification:

$$Y_i(\mathbf{s}) = \mu_i(\mathbf{s}) + Z_i(\mathbf{s}) + \varepsilon_i(\mathbf{s}), \tag{5}$$

where ε_i is mean zero spatial white noise with variance τ_i^2, and Z_i is a mean zero Gaussian process with anisotropic but stationary Matérn covariance function, and $\mu_i(\mathbf{s})$ is a fixed mean function. Let $G_i = \text{Var}\,\boldsymbol{y}_i$ be the spatial covariance matrix of $\boldsymbol{y}_i$ for the ith region. Then the local log likelihood based on p independent replicates $\{\boldsymbol{y}_{ij}\}_{j=1}^p$ for the ith region is, up to a constant,

$$\frac{p}{2}\log|G_i|^{-1} - \sum_{j=1}^{p}\frac{1}{2}(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_i)^T G_i^{-1}(\boldsymbol{y}_{ij} - \boldsymbol{\mu}_i), \tag{6}$$

where $\boldsymbol{\mu}_i$ is the mean function $\mu_i(\mathbf{s})$ evaluated at the locations of $\boldsymbol{y}_i$. 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 practical strategy for large datasets using many computational cores. 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 parameters 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. Furthermore, the Matérn parameters are easier to interpret and will be simpler to model in a hierarchical statistical framework.

4.2 | A numerical study of local Matérn estimation

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. To our knowledge there are no results in the literature that give simple rules-of-thumb for window size. Moreover, it is also useful to understand the extent that replication makes it possible to estimate correlation ranges that are much larger than the window size.

We perform a Monte Carlo experiment with four factors: window size ranging between a 5×5 grid and a 33×33 grid, the Matérn range parameter being multiples of one, two, three, and four 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, 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 maximum likelihood estimates (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 3. This design may seem unusual because the range parameter is varied based on the window size. 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.

The contours in Figure 3 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. The results are encouraging. For example, for a Matern covariance with smoothness equal to 1 and with only 10 replicates one can estimate the range parameter to 10% accuracy provided it is no larger than the window size. In a more extreme the extreme case, a Matérn range four times the size of the window can be estimated to within 10% error if 30 replicates are available and provide the window size is 10 or greater. However, substantially larger window sizes are required, greater than 25 units in width, to achieve a 5% error with 30 replicated fields. Using these guidelines, we can be more confident that local moving window likelihood estimation is a viable technique if enough data is used and if we expect an accuracy of about 10% in estimating the range parameter. Such results give confidence to analyze the 30 member (replicate) ensemble from the LENS experiment in the next section.

5 | CLIMATE DATA APPLICATION

The data set from the CESM LENS project (Kay et al., 2015) is comprised of 30 spatial fields that can be assumed to be independent replicates from the same distribution. This well-known feature is based on the highly nonlinear properties of a global climate model run over a long period and started with different initial conditions. This feature is also termed

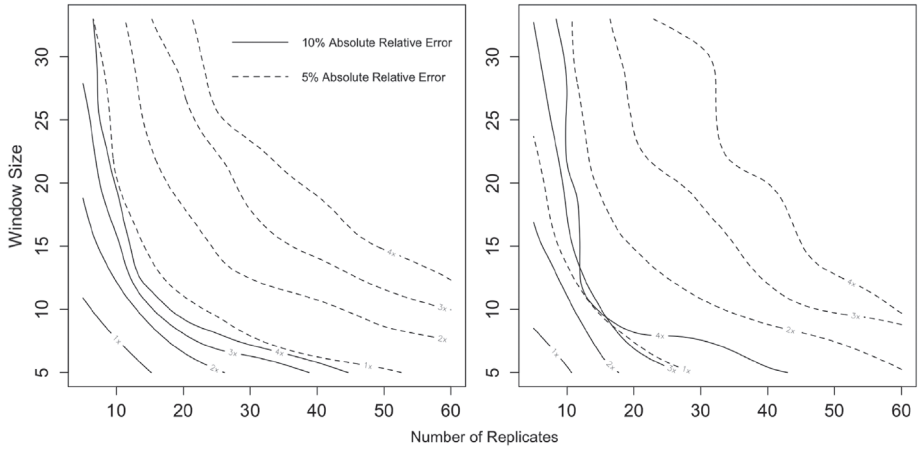

FIGURE 3 Both panels show estimated contours indicating the 5% and 10% absolute percent error from estimating the Matérn range parameter given a certain number of replicates and a window size (size of grid). Note that the number of spatial observations used is the square of the window size multiplied by the number of replicates. To obtain these curves 100 Monte Carlo simulations were done on a grid of different replicate and window sizes and Matérn range parameters one, two, three, and four times the size of the window. The left panel corresponds to $\nu = 1$ and the right to $\nu = 2$ for the Matern smoothness parameters

sensitive dependence on initial conditions and is a hallmark of chaotic dynamical systems. Nychka et al. (2018) first analyzed these data using the LatticeKrig model, and this article details the climate science application. The data locations are on a 288×192 grid with approximately one degree resolution, covering the entire globe. The specific dataset used in this application is publically available in R binary format from the LatticeKrig github repository [1] Details about the pattern scaling approach to statistical emulation can also be found in Alexeeff, Nychka, Sain, and Tebaldi (2018). Briefly, each field is a measure of how the local temperature average is affected by a global temperature average increase of 1°C. Generating this ensemble of only 30 members was an expensive computational campaign, requiring dedicated super computing resources over a period of months. The statistical task is to represent these spatial fields with a probability distribution where it is more efficient to generate additional fields (e.g., several hundred or thousands) that track the original 30 member model results.

We focus on the subregion including the Americas and surrounding oceans containing 13,052 spatial locations on a 102×128 grid. The top row of Figure 5 shows the first four sample fields from the data. The one data modification from Nychka et al. (2018) is that, in addition to subtracting the ensemble mean from each grid box, we have also standardized the fields by dividing by the ensemble *SD* of each grid box.

5.1 | Covariance parameter estimates

We experiment with moving window local MLEs using window sizes between 8×8 and 15×15. Among these choices there was little change in the estimates and subsequent analysis uses a 9×9 window. The estimation was performed on the NCAR Cheyenne supercomputer (Computational and Information Systems Laboratory, 2019) using the R programming language (R Core Team, 2018) with the `Rmpi` (Yu, 2002) and `fields` packages (Nychka, Furrer, Paige, & Sain, 2017). The details of the parallel implementation are the same as in Nychka et al. (2018). Since the fields were standardized, the constraint $\sigma^2 = 1 - \tau^2$ was included in this model.

The estimates for the spatially varying parameters are shown in Figure 4. The process variance is shown in panel (a) and the nugget variance τ^2 is $1 - \sigma^2$. Panel (b) shows the geometric mean of ξ_{s_1} and ξ_{s_2} as a measure of the average correlation range, and this also agrees with the range in the isotropic case. Finally in panel (c), the estimated anisotropy matrix $A(\mathbf{s}_i)$ is depicted by glyphs indicating the range and departure from isotropy. 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

[1] See `github.com/NCAR/LatticeKrig/Datasets/LENS/BRACEUfields.rda` but also refer to the README file in this folder for more background in using these data.

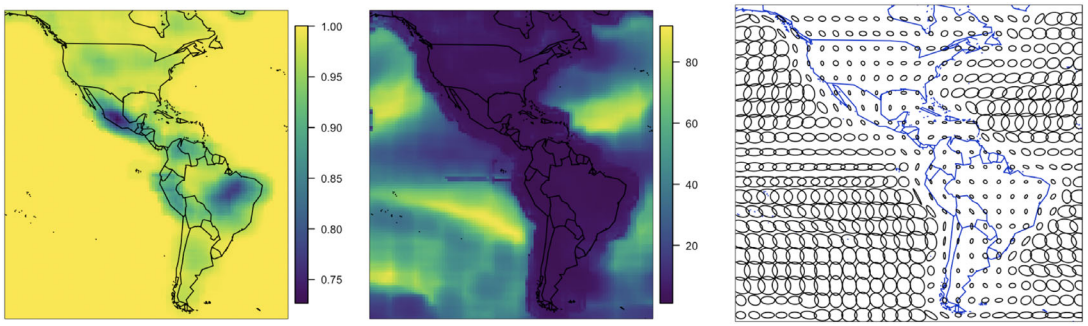

FIGURE 4 Fitted parameter fields based on the moving window likelihood estimation. The first panel shows the process variance $\sigma^2(\mathbf{s})$. The middle panel shows the geometric average range $\sqrt{\xi_{s_1}(\mathbf{s})\xi_{s_2}(\mathbf{s})}$, and the right panel shows the anisotropy ellipses, defined by $A^T(\mathbf{s})A(\mathbf{s})$

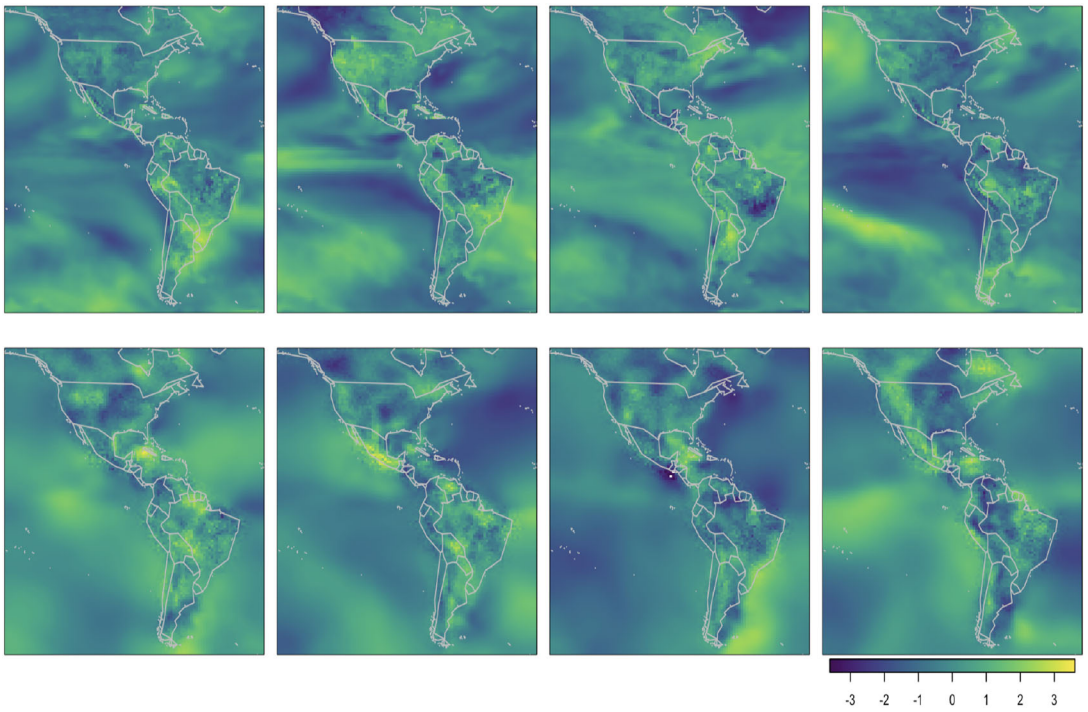

FIGURE 5 The top row consists of the first four ensemble members from the NCAR CESM dataset. The bottom row shows four unconditional simulations from the estimated nonstationary spatial autoregressive model

accurately estimated. There is a larger nugget variance over the land compared to the ocean, which is expected and adds to the nonstationarity. Based on the coastlines in some regions, we hypothesize that the addition of a land/ocean covariate may also be useful. However, using just this covariate would not be sufficient to model these boundary regions. See Figure 6.

5.2 | Model checking

5.2.1 | A visual comparison

The nonstationary SAR model is convenient for simulating high-dimensional fields using plug-in estimates of locally varying parameters. For this reason, we translate the local Matérn parameters into their approximate SAR parameter

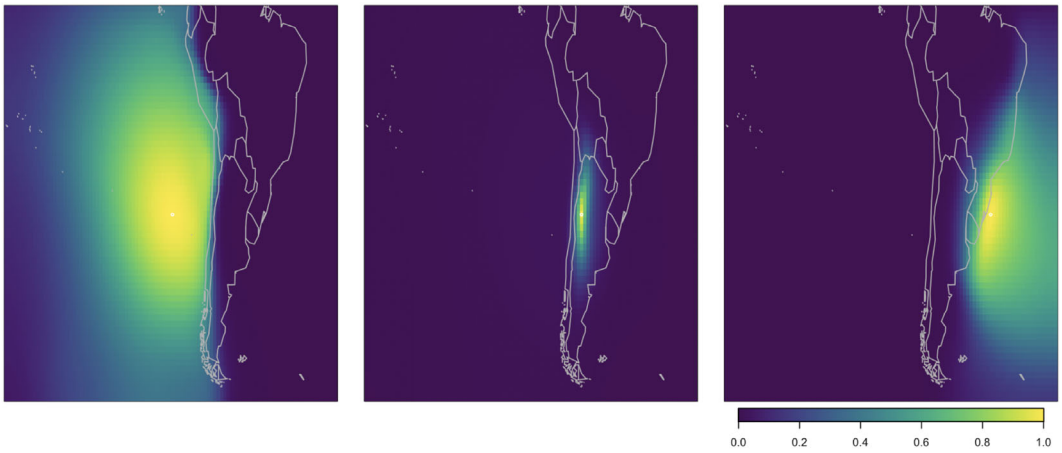

FIGURE 6 Estimated correlation functions centered at three locations along the same latitude implied by the nonstationary spatial autoregressive model

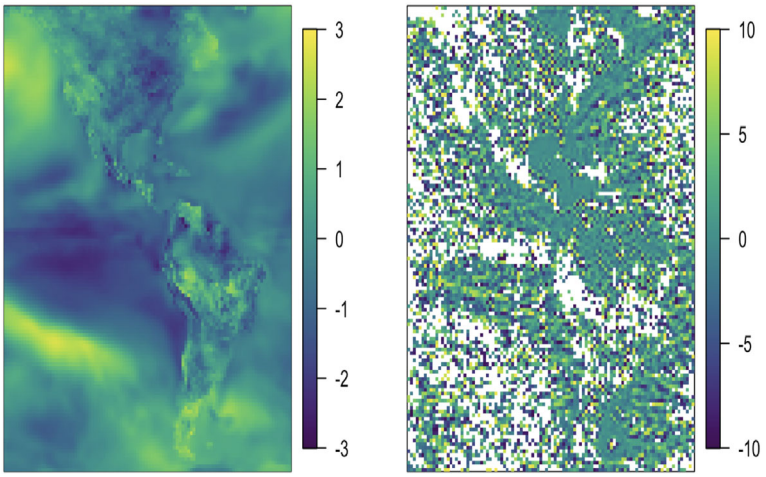

FIGURE 7 Panel (a) shows the a data replicate $\mathbf{y}$, the fourth in the ensemble, and panel (b) shows the symmetric square root of the precision matrix B applied to the data replicate yielding $\mathbf{w} = B\mathbf{y}$

equivalents. The translation is done using the numerical relationship derived in Section 2.4. Then, the local SAR parameters are encoded into the nonstationary global SAR model. Simulations from this covariance are shown in the bottom row of Figure 5. Exploiting sparse linear algebra, a field can be simulated on the order of tens of seconds using R on a standard notebook computer. 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 6 shows several different locations in the spatial domain and plots the correlations implied by the estimated nonstationary model. Both anisotropy and nonstationarity are evident in Figure 6.

5.2.2 | Transformation to white noise

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 linear transformation defined by the SAR weight matrix should decorrelate the spatial field. In particular, let $\boldsymbol{y}$ be the observed field with covariance matrix Σ. The simple idea is to factor Σ as $\Omega^{-1}\Omega^{-T}$ and then check that $\Omega\boldsymbol{y}$ is a white noise field, or at least a spatial process with greatly reduced spatial dependence. Note that the choice of Ω 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,

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

If $\tau = 0$ then the SAR matrix provides a transformation to white noise. If τ^2 is small relative to σ^2, then $B\boldsymbol{y}$ will have covariance $(I + \tau^2 Q)$ and will approximate a white noise field. Small τ^2 is often 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 covariance matrix will have localized, finitely supported correlations.

Figure 7b shows the result of B applied to one of the replicates (shown in (a)) to which the model was fitted.

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 7 indicates that the vast majority of the correlation in the data has been captured in the model, and therefore suggests the model fit is adequate. This success is encouraging given the long range correlations over the ocean that have been identified from local estimation, and represented by only second-order neighbors. A formal test of independence could be implemented on the decorrelated fields, and the model could be refined to eliminate residual dependence.

6 | CONCLUSION

In this paper, we investigate a two-step framework of local estimation and global encoding to represent large and nonstationary spatial datasets. We have shown that when independent replicates of spatial processes are available and local stationarity holds, local maximum likelihood estimation is a robust technique for estimating spatially varying 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 this is the first time detailed numerical mappings have been made between the anisotropic SAR model and an anisotropic Matérn covariance function.

A major contribution of this work is in connecting the local likelihood estimation techniques to a flexible and computationally efficient spatial statistical model. We focus on encoding the locally estimated parameters in the nonstationary SAR model. In addition, the two-stage approach is computationally efficient and can 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 sparse linear algebra for prediction and simulation. Although other methods such as Kang et al. (2012) can be effective in representing a nonstationary spatial field, they are not as readily interpretable in terms of local covariance functions.

Another advantage of this method is that it can be applied to both continuously indexed and lattice data. To reduce the scope of this work, however, 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 nonlattice 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. The local estimation technique in the method developed here gives insight into the nature of the nonstationarity present in the data, and the local estimates can suggest more restricted models as one level in a hierarchical Bayesian model.

We believe the anisotropic models developed here will carry over for more general models such as basis expansions, and the inverse square root transformation will be an important diagnostic tool for nonstationary modeling.

ACKNOWLEDGEMENTS

This research was supported in part by NSF Awards `DMS-1811294` and `DMS-1923062`. Douglas Nychka also thanks *The International Environmetrics Society (TIES)* for the generous support to attend the 28th Annual TIES Conference, July, 2018 in Guanajuato, Mexico and to give the President's Invited Lecture. This article constitutes some of the core

research material from that lecture. Finally, the authors acknowledge Stacey Alexeeff and Claudia Tebaldi for formulating the climate scaling application and creating the original dataset from LENS.

DATA AVAILABILITY STATEMENT

The specific dataset used in this application is publically available in R binary format from the LatticeKrig github repository. See github.com/NCAR/LatticeKrig/Datasets/LENNS and the files BRACEUfelds.rda and README in this folder for more background in using these data.

ORCID

Ashton Wiens https://orcid.org/0000-0002-7030-0602
Douglas Nychka https://orcid.org/0000-0003-1387-3356

REFERENCES

Alexeeff, S. E., Nychka, D., Sain, S. R., & Tebaldi, C. (2018). Emulating mean patterns and variability of temperature across and within scenarios in anthropogenic climate change experiments. *Climatic Change*, *146*(3-4), 319–333.

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

Computational and Information Systems Laboratory (2019). *Cheyenne: HPE/SGI ICE XA System (Climate Simulation Laboratory)*. Boulder, CO: National Center for Atmospheric Research. https://doi.org/10.5065/D6RX99HX.

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*(1), 209–226.

Delfiner, P., & Chilès, J.-P. (2012). *Geostatistics: Modeling spatial uncertainty*. Hoboken, NJ: John Wiley & Sons.

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, *89*(1), 197–210.

Fuentes, M., & Smith, R. L. (2001). *A new class of non-stationary spatial models. Technical report*. Raleigh, NC: North Carolina State University.

Fuglstad, G.-A., Lindgren, F., Simpson, D., & Rue, H. (2015). Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. *Statistica Sinica*, *25*, 115–133.

Fuglstad, G.-A., Simpson, D., Lindgren, F., & Rue, H. (2015). Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics*, *14*, 505–531.

Haas, T. C. (1990a). Kriging and automated variogram modeling within a moving window. *Atmospheric Environment Part A. General Topics*, *24*(7), 1759–1769.

Haas, T. C. (1990b). Lognormal and moving window methods of estimating acid deposition. *Journal of the American Statistical Association*, *85*(412), 950–963.

Heaton, M., Datta, A., & Finley, A. E. A. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, *24*, 398–425.

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

Higdon, D. (2002). *Space and space-time modeling using process convolutions*. In *Quantitative methods for current environmental issues* (pp. 37–56). London, UK: Springer.

Higdon, D., Swall, J., & Kern, J. (1999). Non-stationary spatial modeling. *Bayesian Statistics*, *6*(1), 761–768.

Kang, E. L., & Cressie, N. (2013). Bayesian hierarchical Anova of regional climate-change projections from Narccap Phase II. *International Journal of Applied Earth Observation and Geoinformation*, *22*, 3–15.

Kang, E. L., Cressie, N., & Sain, S. R. (2012). Combining outputs from the North American regional climate change assessment program by using a Bayesian hierarchical model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *61*(2), 291–313.

Katzfuss, M., & Cressie, N. (2011). Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *Journal of Time Series Analysis*, *32*(4), 430–446.

Kay, J., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., 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*(8), 1333–1349.

Lindgren, F., & Rue, H. (2007). Explicit construction of GMRF approximations to generalised Matérn fields on irregular grids. *Preprints in Mathematical Sciences*, *12*, 5.

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*(4), 423–498.

Marques, I., Klein, N., & Kneib, T. (2019). Non-stationary spatial regression for modelling monthly precipitation in Germany. *Spatial Statistics*, 100386. https://doi.org/10.1016/j.spasta.2019.100386.

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*(2), 579–599.

Nychka, D., Furrer, R., Paige, J., and Sain, S. (2017). *Fields: Tools for spatial data. R package version 9.6*. Cambridge, Mass: MIT Press.

Nychka, D., Hammerling, D., Krock, M., & Wiens, A. (2018). Modeling and emulation of nonstationary Gaussian fields. *Spatial Statistics*, *28*, 21–38.

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

Paciorek, C. J., & Schervish, M. J. (2003). *Nonstationary covariance functions for Gaussian process regression*. Proceedings of the 16th International Conference on Neural Information Processing Systems, NIPS'03 (pp. 273-280). Cambridge, MA: MIT Press.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

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

Risser, M. D., & Calder, C. A. (2017). Local likelihood estimation for covariance functions with spatially-varying parameters: The convoSPAT package for R. *Journal of Statistical Software*, *81*, 1–32.

Rue, H., & Held, L. (2005). *Gaussian Markov random fields: Theory and applications*. Boca Raton, FL: Chapman & Hall/CRC Press.

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

Simpson, D., Lindgren, F., & Rue, H. (2012). Think continuous: Markovian Gaussian models in spatial statistics. *Spatial Statistics*, *1*, 16–29.

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*(2), 265–282.

Ver Hoef, J. M., Hanks, E. M., & Hooten, M. B. (2018). On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. *Spatial Statistics*, *25*, 68–85.

Yu, H. (2002). Rmpi: Parallel statistical computing in R. *R News*, *2*(2), 10–14.

Zhu, Z., & Wu, Y. (2010). Estimation and prediction of a class of convolution-based spatial nonstationary models for large spatial data. *Journal of Computational and Graphical Statistics*, *19*(1), 74–95.

How to cite this article: Wiens A, Nychka D, Kleiber W. Modeling spatial data using local likelihood estimation and a Matérn to spatial autoregressive translation. *Environmetrics*. 2020;31:e2652. https://doi.org/10.1002/env.2652