Statistica Sinica Preprint No: SS-13-240wR3	
Title	Likelihood approximations for big nonstationary spatial temporal lattice data
Manuscript ID	SS-13-240wR3
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.2012.240w
Complete List of Authors	Joseph Guinness and Montserrat Fuentes
Corresponding Author	Joseph Guinness
E-mail	jsguinne@ncsu.edu

LIKELIHOOD APPROXIMATIONS FOR BIG NONSTATIONARY SPATIAL TEMPORAL LATTICE DATA

Joseph Guinness and Montserrat Fuentes

NC State University, Department of Statistics

Abstract: We propose a nonstationary Gaussian likelihood approximation for the class of evolutionary spectral models for data on a regular lattice. Lattice data include many important environmental data sources such as weather model output or gridded data products derived from satellite observations. The likelihood approximation is an extension of the Whittle likelihood and is computationally efficient to evaluate when the evolutionary transfer function can be expressed in a flexible low-dimensional form. The low-dimensional form for the evolutionary transfer function is an attractive modeling framework since it allows the practitioner to build nonstationary models in a sequential manner and choose the appropriate dimension based on changes in approximate loglikelihood. While the transfer functions are low-dimensional, the resulting covariance matrices are generally full rank, and thus no rank reduction is required for the computational efficiency of the methods. We study the covariance matrix implied by the likelihood approximation and give its asymptotic rate of approximation to the exact covariance matrix. We evaluate the likelihood approximation in a simulation study and show that it can produce asymptotically efficient parameter estimates when an operation similar to tapering is applied. We introduce an algorithm based on the Ising model to partition the domain into stationary subregions and show in a simulation that the methods can reliably recover an unknown partition. We apply our modeling and estimation framework to analyze spatial-temporal output from a regional weather model comprised of 151,200 wind speed values, and we demonstrate that the fitted covariances are consistent with local empirical variograms.

Key words and phrases: Fast Fourier Transform, Ising model, locally stationary, spectral analysis.

1. Introduction

In many environmental datasets, when the temporal or spatial domain is large, the assumption of stationarity is often obviously violated. Environmental processes depend on diverse geography, seasonal variation, and a wide array of other complex influences that can rarely be considered constant over a large region or an extended length of time. An incorrect specification of the covariance function may be detrimental for prediction, especially for assessing the uncertainty in prediction. Further, when the spatial-temporal domain is large, the number of observations in the dataset is often large as well, so when the data call for nonstationary models, they also call for computationally efficient methods for

fitting those models.

We consider the case of specifying and fitting nonstationary Gaussian process models to data observed on a regular spatial-temporal lattice. Lattice data include many important environmental data sources, such as numerical climate or weather model output, and many processed data products, such as Level 3 satellite data, which are usually interpolated to a grid. Here, we consider output from the HIRLAM weather model that consists of 151,200 spatial-temporal wind speed values. If the model is stationary, and the data locations are on a regular lattice, the Whittle likelihood (Whittle, 1954) can be used in place of the exact likelihood in order to perform approximate inference, but the Whittle likelihood is not valid when the model is nonstationary.

We propose a new computationally efficient Gaussian likelihood approximation for nonstationary d-dimensional lattice data that is a generalization of the Whittle likelihood. This work extends to higher dimensions a likelihood approximation for nonstationary time series introduced in Guinness and Stein (2013) and addresses edge effects that are negligible in the $d = 1$ time series case but are nonnegligible when $d > 1$. The work here also makes use of the Ising model to partition the domain into stationary subregions, whereas the previous work in Guinness and Stein (2013) employed a genetic algorithm. The nonstationary models we study use the idea of evolutionary spectra originally proposed by Priestley (1965) for nonstationary time series data. Evolutionary spectral models allow us to specify the local covariance properties flexibly and guarantee that the resulting full covariance matrices are always positive definite. To gain a better understanding of the asymptotic properties of the model that our approximation implies, we adopt the locally stationary framework advanced by Dahlhaus (1996), who studied nonstationary time series with evolutionary spectra. In $d > 1$ dimensions, the notation is slightly more burdensome, but the concept is essentially the same as it is in the $d = 1$ time series case.

Specifically, let $\mathbb{N}^d$ be the set of d-vectors of nonnegative integers. For any $\boldsymbol{n} = (n_1, \ldots, n_d) \in \mathbb{N}^d$, define $\mathbb{J}_{\boldsymbol{n}}$ to be integer lattice of size $\boldsymbol{n}$, that is, the set of vectors $\boldsymbol{x} = (x_1, \ldots, x_d)$ with $1 \leq x_j \leq n_j$ for every $j = 1, \ldots, d$. Define $\mathbb{T}$ to be the unit circle and $A(\boldsymbol{u}, \boldsymbol{\omega}) : [0, 1]^d \times \mathbb{T}^d \to \mathbb{C}$ to be a complex-valued transfer function with $\int_{\mathbb{T}^d} |A(\boldsymbol{u}, \boldsymbol{\omega})|^2 d\boldsymbol{\omega} < \infty$ for every $\boldsymbol{u}$. Let $\boldsymbol{x}/\boldsymbol{n} = (x_1/n_1, \ldots, x_d/n_d)$ and $Z(\boldsymbol{\omega})$ be a d-dimensional orthogonal increment process. For $\boldsymbol{x} \in \mathbb{J}^d$, the sequence of processes $Y_{\boldsymbol{n}}$ defined by

$$Y_{\boldsymbol{n}}(\boldsymbol{x}) = \int_{\mathbb{T}^d} A(\boldsymbol{x}/\boldsymbol{n}, \boldsymbol{\omega}) \exp(i\boldsymbol{\omega}'\boldsymbol{x}) dZ(\boldsymbol{\omega}) \tag{1.1}$$

forms locally stationary sequence. In Section 3, we require the components of $\boldsymbol{n}$ to grow at the same rate. Let * denote complex conjugation. If $A(\boldsymbol{u}, -\boldsymbol{\omega}) =$

$A(\boldsymbol{u},\boldsymbol{\omega})^*$ for every $\boldsymbol{u}$ and $\boldsymbol{\omega}$, and $Z(-\boldsymbol{\omega}) = Z(\boldsymbol{\omega})^*$ for every $\boldsymbol{\omega}$, then the processes are real-valued, and if Z is Gaussian, then $Y_{\boldsymbol{n}}$ is Gaussian as well.

The locally stationary model is convenient since for any permissible choice of A the covariance function for $Y_{\boldsymbol{n}}$ is given by

$$K_{\boldsymbol{n}}(\boldsymbol{x},\boldsymbol{y}) = \mathrm{Cov}(Y_{\boldsymbol{n}}(\boldsymbol{x}), Y_{\boldsymbol{n}}(\boldsymbol{y})) = \int_{\mathbb{T}^d} A(\boldsymbol{x}/\boldsymbol{n},\boldsymbol{\omega})A(\boldsymbol{y}/\boldsymbol{n},\boldsymbol{\omega})^* \exp(i\boldsymbol{\omega}'(\boldsymbol{x}-\boldsymbol{y})d\boldsymbol{\omega}, \tag{1.2}$$

and we denote by $K_{\boldsymbol{n}}(A)$ the covariance matrix for the vector $\boldsymbol{Y}_{\boldsymbol{n}}$ containing the observations at all locations $\boldsymbol{x} \in \mathbb{J}_{\boldsymbol{n}}$. Then defining $n = \prod_{j=1}^d n_j$, we can write down the Gaussian loglikelihood for A based on $\boldsymbol{Y}_{\boldsymbol{n}}$ as

$$-\ell_0(A;\boldsymbol{Y}_{\boldsymbol{n}}) = \frac{n}{2}\log(2\pi) + \frac{1}{2}\log\det K_{\boldsymbol{n}}(A) + \frac{1}{2}\boldsymbol{Y}_{\boldsymbol{n}}^T K_{\boldsymbol{n}}(A)^{-1}\boldsymbol{Y}_{\boldsymbol{n}}. \tag{1.3}$$

As discussed before, when the model is not stationary, that is when $A(\boldsymbol{u},\boldsymbol{\omega})$ is not constant as a function of $\boldsymbol{u}$, computing the likelihood generally requires the $O(n^2)$ storage of $K_{\boldsymbol{n}}(A)$ and the $O(n^3)$ operations required to compute the Cholesky decomposition of $K_{\boldsymbol{n}}(A)$.

We propose in Section 2 an approximation to ℓ_0 that is based on the exact likelihood for a process that approximates $Y_{\boldsymbol{n}}$. The computational efficiency of our likelihood approximation relies on expressing A in the form

$$A(\boldsymbol{x}/\boldsymbol{n},\boldsymbol{\omega}) = \sum_{m=1}^{M} w_m(\boldsymbol{x}/\boldsymbol{n})A_m(\boldsymbol{\omega}), \tag{1.4}$$

where M is a small integer. The representation in (1.4) is useful for facilitating a sequential approach to building nonstationary models. For example, when $M = 1$ the model is what Priestley (1965) termed uniformly modulated, in which the correlation structure is stationary, but the variance of the process is allowed to vary across the domain. Uniformly modulated models include stationary models as a special case when $w_1(\boldsymbol{x})$ is constant. If uniform modulation is not sufficient to capture the nonstationary nature of the process, the modeler may increase M to 2, adding a second component to A. This procedure may continue until an appropriate value of M is chosen. In Section 6, we apply this sequential model-fitting approach to a spatial-temporal field of wind speed output from a regional weather model. Based on the results of the data analysis, we recommend using changes in relative loglikelihood to choose M.

A common method of constructing nonstationary spatial covariance functions is through convolving a stationary process–usually white noise–with kernels that vary across the spatial domain (Higdon, 1998; Calder, 2007). Fuentes (2002a) studied connections between kernel convolution models and models with

evolutionary spectra. Kernel convolution models are very flexible, but when the number of observations is large, the computational burden required for constructing the covariance matrices and evaluating the likelihood call for the convolution to be discretized into a small number of components, which results in a reduced rank covariance matrix for the observations. The low-dimensional representation for A in (1.4) should not be confused with a low-rank approximation for a covariance matrix. Even in the lowest dimensional stationary case where $M = 1$, any covariance matrices generated from the transfer function will generally be full rank when A is bounded away from 0.

Another computationally efficient approach for modeling and estimating nonstationary processes has been through the specification of stochastic partial differential equations with parameters that vary across the spatial domain (Lindgren et al., 2011). Efficient likelihood computations with these models are possible when the process model can be approximated by a Markov random field, which induces sparsity in the inverse covariance matrix. Our methods do not require sparsity in either the covariance matrix or the inverse covariance matrix, and we show in a simulation study in Section 5 that the approximate likelihoods are capable of producing efficient parameter estimates.

The low-dimensional form for the transfer function implies that the nonstationary process can be written as a linear combination of stationary processes, which is also true of the approach taken in Fuentes (2001) and Fuentes (2002b), where the nonstationary process is written as a weighted sum of independent stationary processes. Here, however, we do not assume that the stationary processes are independent; they have dependence structure implied by the covariance function in (1.2), so even if the weight functions are indicators, the covariance between the process at any two locations will generally be nonzero.

2. The Likelihood Approximation

The likelihood approximation arises from the exact likelihood for a process that approximates a locally stationary process. Defining $\boldsymbol{\omega_j} = 2\pi \boldsymbol{j}/\boldsymbol{n}$ to be a Fourier frequency for $\boldsymbol{j} \in \mathbb{J}_{\boldsymbol{n}}$, we approximate the stochastic integral in (1.1) with the sum

$$\widetilde{Y}_{\boldsymbol{n}}(\boldsymbol{x}) = \frac{(2\pi)^{d/2}}{n^{1/2}} \sum_{\boldsymbol{j} \in \mathbb{J}_{\boldsymbol{n}}} A(\boldsymbol{x}/\boldsymbol{n}, \boldsymbol{\omega_j}) \exp(i\boldsymbol{\omega_j}'\boldsymbol{x}) \widehat{Z}(\boldsymbol{\omega_j}), \tag{2.1}$$

where $\widehat{Z}(\boldsymbol{\omega_j})$ are uncorrelated, mean-zero, unit variance complex normal random variables satisfying $\widehat{Z}(-\boldsymbol{\omega_j}) = \widehat{Z}(\boldsymbol{\omega_j})^*$. Then we can write the approximate process vector as a linear transformation of the vector $\widehat{\boldsymbol{Z}}$, as in $\widetilde{\boldsymbol{Y}}_{\boldsymbol{n}} = C_{\boldsymbol{n}}(A)\widehat{\boldsymbol{Z}}$, where $C_{\boldsymbol{n}}(A)$ is the $n \times n$ matrix that performs the transformation in (2.1).

The approximate process vector has covariance matrix $\Delta_{\boldsymbol{n}}(A) = C_{\boldsymbol{n}}(A)C_{\boldsymbol{n}}(A)^H$, where H is the conjugate transpose. In Section 3, we study the approximation $\Delta_{\boldsymbol{n}}(A)$ to the covariance matrix $K_{\boldsymbol{n}}(A)$. Then the loglikelihood for A based on $\widetilde{\boldsymbol{Y}}_{\boldsymbol{n}}$ is

$$-\ell(A;\widetilde{\boldsymbol{Y}}_{\boldsymbol{n}}) = \frac{n}{2}\log(2\pi) + \log|\det C_{\boldsymbol{n}}(A)| + \frac{1}{2}\|C_{\boldsymbol{n}}(A)^{-1}\widetilde{\boldsymbol{Y}}_{\boldsymbol{n}}\|^2,$$

and we use $\ell(A;\boldsymbol{Y}_{\boldsymbol{n}})$ to approximate $\ell_0(A;\boldsymbol{Y}_{\boldsymbol{n}})$.

In the stationary case, that is, when A is constant as a function of $\boldsymbol{u}$ and thus can be written $A(\boldsymbol{u},\boldsymbol{\omega}) = S(\boldsymbol{\omega})$, $\Delta_{\boldsymbol{n}}(A)$ has a block circulant structure and is therefore diagonalizable by the d-dimensional discrete Fourier transform (DFT). Thus the log determinant term is

$$\log|\det C_{\boldsymbol{n}}(S)| = \frac{nd}{2}\log(2\pi) + \sum_{\boldsymbol{j}\in\mathbb{J}_{\boldsymbol{n}}}\log|S(\boldsymbol{\omega}_{\boldsymbol{j}})|, \tag{2.2}$$

and the inverse transformation appearing in the quadratic form is

$$\left[C_{\boldsymbol{n}}(S)^{-1}\boldsymbol{Y}_{\boldsymbol{n}}\right]_{\boldsymbol{j}} = \frac{1}{(2\pi)^{d/2}n^{1/2}}\frac{1}{S(\boldsymbol{\omega}_{\boldsymbol{j}})}\sum_{\boldsymbol{x}\in\mathbb{J}_{\boldsymbol{n}}} Y_{\boldsymbol{n}}(\boldsymbol{x})\exp(-i\boldsymbol{\omega}_{\boldsymbol{j}}'\boldsymbol{x}), \tag{2.3}$$

which is simply the d-dimensional DFT of $\boldsymbol{Y}_{\boldsymbol{n}}$ scaled by the reciprocal of the transfer function. Therefore our approximation reduces to the Whittle likelihood approximation in the stationary case.

In the nonstationary case, the computation of $C_{\boldsymbol{n}}(A)^{-1}\boldsymbol{Y}_{\boldsymbol{n}}$ can be made efficient when A has the form given in (1.4). In this case, the transformation $C_{\boldsymbol{n}}(A)\widehat{\boldsymbol{Z}}$ can be written as

$$Y_{\boldsymbol{n}}(\boldsymbol{x}) = \frac{(2\pi)^{d/2}}{(n)^{1/2}}\sum_{m=1}^{M} w_m(\boldsymbol{x}/\boldsymbol{n})\sum_{\boldsymbol{j}\in\mathbb{J}_{\boldsymbol{n}}} A_m(\boldsymbol{\omega}_{\boldsymbol{j}})\exp(i\boldsymbol{\omega}_{\boldsymbol{j}}'\boldsymbol{x})\widehat{Z}(\boldsymbol{\omega}_{\boldsymbol{j}}), \tag{2.4}$$

which is simply a weighted sum of M d-dimensional inverse DFTs. This allows us to solve the system $\boldsymbol{Y}_{\boldsymbol{n}} = C_{\boldsymbol{n}}(A)\widehat{\boldsymbol{Z}}$ efficiently using iterative methods that rely on fast forward multiplication $C_{\boldsymbol{n}}(A)\widehat{\boldsymbol{Z}}$.

Throughout the rest of the paper, we assume further that A has the form

$$A(\boldsymbol{x}/\boldsymbol{n},\boldsymbol{\omega}) = \sum_{m=1}^{M} I_m(\boldsymbol{x}/\boldsymbol{n})A_m(\boldsymbol{\omega}), \tag{2.5}$$

where $I_m(\boldsymbol{x}/\boldsymbol{n})$ is an indicator function, and $\sum_{m=1}^{M} I_m(\boldsymbol{x}/\boldsymbol{n}) = 1$ for every $\boldsymbol{x}$, so that the set of indicator functions defines a partition of the space-time domain. This assumption is useful both computationally and conceptually, since if

$I_m(\boldsymbol{x}/\boldsymbol{n}) = 1$, the local second-order properties of the process nearby location $\boldsymbol{x}$ are determined by the transfer function $A_m(\boldsymbol{\omega})$. This assumption is also useful for computing a simple preconditioner for the iterative matrix algorithm. Specifically, we precondition with $C(1/A)^H \boldsymbol{Y_n}$, where $(1/A)(\boldsymbol{\omega}) = A(\boldsymbol{\omega})$, which in the case of (2.5) can be expressed as

$$\left[C(1/A)^H \boldsymbol{Y_n}\right]_{\boldsymbol{j}} = \frac{1}{\sqrt{2\pi n}} \sum_{m=1}^{M} \frac{1}{A_m(\boldsymbol{\omega_j})} \sum_{\boldsymbol{x} \in \mathbb{J}_{\boldsymbol{n}}} I_m(\boldsymbol{x}/\boldsymbol{n}) \exp(-i\boldsymbol{\omega}'\boldsymbol{x}) Y_{\boldsymbol{n}}(\boldsymbol{x}), \quad (2.6)$$

which again can be computed efficiently with M applications of an FFT algorithm.

The log determinant of $C_{\boldsymbol{n}}(A)$ is more difficult to compute, but accurate approximations exist. In this paper, we consider an extension of an approximation suggested by Dahlhaus (2000), namely

$$\log|\widetilde{\det} C_{\boldsymbol{n}}(A)| = \frac{nd}{2}\log(2\pi) + \frac{1}{n}\sum_{\boldsymbol{x} \in \mathbb{J}_{\boldsymbol{n}}} \sum_{\boldsymbol{j} \in \mathbb{J}_{\boldsymbol{n}}} \log|A(\boldsymbol{x}/\boldsymbol{n}, \boldsymbol{\omega_j})|.$$

It is also easy to see that this approximation also reduces to the log determinant in the Whittle likelihood in the stationary case. The evaluation of this log determinant approximation is efficient when A can be expressed as in (2.5).

In Section 3, we study the approximate covariance function implied by $\ell(A; \boldsymbol{Y_n})$, and in Section 5, we evaluate the likelihood approximation

$$\widetilde{\ell}(A; \boldsymbol{Y_n}) = \frac{n}{2}\log(2\pi) + \log|\widetilde{\det} C_{\boldsymbol{n}}(A)| + \frac{1}{2}\|C_{\boldsymbol{n}}(A)^{-1}\boldsymbol{Y_n}\|^2. \quad (2.7)$$

in simulation studies and show an example in which it can produce efficient parameter estimates.

3. Nonstationary Covariance Matrix Approximation

The likelihood approximation $\ell(A; \boldsymbol{Y_n})$ arises from assuming that the process has covariance matrix $\Delta_{\boldsymbol{n}}(A) = C_{\boldsymbol{n}}(A)C_{\boldsymbol{n}}(A)^H$ instead of $K_{\boldsymbol{n}}(A)$. Here we study this matrix approximation, particularly the dependence on the dimension d. We also require each dimension to grow at the same rate, that is, $n_j = a_j n^{1/d}$ with $0 < a_j < \infty$ for every j. The following theorem establishes the rate at which the matrix error in Frobenius norm grows with n for dimension d and requires A to be uniformly smooth in frequency:

Theorem 1. *If $A(\boldsymbol{u}, \boldsymbol{\omega})$ is $d+1$ times continuously differentiable in $\boldsymbol{\omega}$ for every $\boldsymbol{u}$, then*

$$\|\Delta_{\boldsymbol{n}}(A) - K_{\boldsymbol{n}}(A)\|_F^2 = O(n^{1-1/d}).$$

The proof is given in Section S1 of the supplementary material. Since our likelihood approximation $\ell(A; \boldsymbol{Y_n})$ reduces to the Whittle likelihood in the stationary case, the same is true of the covariance matrix assumed by Whittle likelihood. Guyon (1982) showed further that the Whittle likelihood itself has an approximation rate of $O(n^{1-1/d})$ to the exact Gaussian loglikelihood. Although we do not have a proof for the rate of approximation for the likelihood in the nonstationary case, the result in Theorem 1 and the simulation results in Section 5 to follow support the efficacy of this likelihood approximation for parameter estimation.

4. Edge Effects and Searching for Stationary Subregions

Just as is the case with the Whittle likelihood approximation, our approximation essentially assumes that the process is periodic in every dimension since the process approximation involves only Fourier frequencies. To mitigate the effect of this assumption on parameter estimation in the stationary case, Guyon (1982) suggested estimating the spectral density via the discrete Fourier transform of an unbiased estimator of the covariance function. More relevant to our study is the work of Dahlhaus and Künsch (1987), who proposed tapering the observations on the edges of the region before computing the Whittle likelihood and proved that tapering can produce asymptotically efficient parameter estimates when $d \leq 3$.

We now describe our methods for mitigating the edge effects. When we compute the approximate likelihood, we assume instead that A has the form

$$A(\boldsymbol{x}/\boldsymbol{n}, \boldsymbol{\omega}) = \sum_{j=1}^{m} \widetilde{I}_j(\boldsymbol{x}/\boldsymbol{n}) A_j(\boldsymbol{\omega}) + I_{m+1}(\boldsymbol{x}/\boldsymbol{n}) A_{m+1}(\boldsymbol{\omega}),$$

where $I_{m+1}(\boldsymbol{x}/\boldsymbol{n})$ is an indicator for a buffer zone around the boundary of the spatial region. Then for $j = 1, \ldots, m$, $\widetilde{I}_j(\boldsymbol{x}/\boldsymbol{n}) = 0$ if $\boldsymbol{x}/\boldsymbol{n}$ is in the buffer zone and equals $I_j(\boldsymbol{x}/\boldsymbol{n})$ if $\boldsymbol{x}/\boldsymbol{n}$ is not in the buffer zone. An illustration of the definitions of the various indicator functions is given in Figure 5.1.

In the two-dimensional simulation in Section 5, we set the buffer size to be $\sqrt{n_1}/3$, so that the buffer size grows with the sample size, but the overall proportion of buffered observations shrinks to zero as $n \to \infty$. In practice, the buffer should be set manually, and in general it should be larger when the spatial correlation is stronger, but the size of the buffer should be balanced against the size of the dimensions–small dimensions cannot afford very large buffers.

We may be interested in determining whether the process has some specific nonstationary structure. Perhaps the nonstationarity covariance can be explained by some geographic or political boundaries. For example, an atmospheric

process may behave differently over land versus over ocean, or maps of incidence of infectious diseases may vary based on different countries' control strategies. In these cases, the partition functions $I_j(\boldsymbol{x}/\boldsymbol{n})$ may be assumed to be known and correspond to coastal or political borders. In many cases, though, we may not be able to assume that the partitions are known, so it is necessary to propose methods for searching over the space of partitions.

Estimating the indicator functions requires searching over the space of M-component partitions of the n observation locations. With even moderate sample sizes, exhaustive searches are not feasible since the number of possible partitions is M^n, so we cannot hope to always find the globally optimal partition. To help reduce the size of the search, it seems reasonable to assume that the "best" partitions will be composed of mostly contiguous regions, and we use a random algorithm to search over this space.

A natural choice for generating mostly contiguous random partitions of a lattice is the Ising model (Ising, 1925). The Ising model is a probability distribution on spin lattices, $S(\mathbb{J}_{\boldsymbol{n}}) = \{s(\boldsymbol{x})\}_{\boldsymbol{x}\in\mathbb{J}_{\boldsymbol{n}}}$ with $s(\boldsymbol{x}) = \pm 1$. The probability mass function for the simplest case of this distribution is

$$p(S;T) = \frac{1}{c(T)}\exp(-H(L)/T),$$

where T is a temperature parameter, $c(T)$ is a normalizing constant, and

$$H(S) = -\sum_{\boldsymbol{x}\sim\boldsymbol{y}} s(\boldsymbol{x})s(\boldsymbol{y}),$$

where the sum is over all pairs of adjacent locations of the lattice, and each pair is counted once. For any positive temperature parameter T, the Ising distribution we consider places more probability mass on spin lattices with fewer spin transitions between adjacent locations. Thus, draws from the Ising model will produce mostly contiguous partitions of the lattice.

For even the simplest Ising model, the normalizing constants are not known, and thus a Metropolis-Hastings algorithm is commonly used to sample from the distribution. The Metropolis-Hastings algorithm is convenient for this application since we wish to randomly search around the partition space. We describe the Metropolis-Hastings algorithm for sampling from the partitions. The spin lattice S_0 assigns ± 1 independently to each spatial location with probability 1/2. Our proposal distribution selects a location uniformly at random and switches the spin at that location. The proposal S^* is accepted with probability $\min\{1, \exp((H(S_k) - H(S^*))/T)\}$.

We start by running the Metropolis-Hastings algorithm for many iterations at a very low temperature. The low temperature helps ensure that the algorithm

converges to a local minimum. The resulting spin lattice defines our starting partition, and we maximize our approximate likelihood over any transfer function parameters to obtain a starting likelihood. Then we increase the temperature and at each iteration we initialize the Metropolis-Hastings algorithm with the current partition and run it for several steps to obtain a candidate partition. If the maximum approximate likelihood for this candidate partition exceeds the current maximum likelihood, we accept the candidate partition as the current partition and store the maximum approximate likelihood as the current likelihood. The updates continue for 10 iterations, and if the approximate likelihood has increased sufficiently, we pursue additional iterations. This allows us to quickly discard bad starting partitions and explore more thoroughly the best candidates.

Our method for searching the partition space embeds the maximization over the transfer function parameters inside of a search over the space of partitions that makes use of Metropolis-Hastings updates of a spin lattice under the Ising Model. To find partitions into more than two blocks, we use a "greedy" search that requires our best $(M+1)$-block partition to be a subpartition of our best M-block partition. This procedure guarantees that our search will return an $(M+1)$-block partition with a higher approximate likelihood than our best M-block partition. We note that there is no guarantee that the best $(M+1)$-block partition will be a subpartition of the best M-block partition. However, we have found that unconstrained searches of this type into more than two blocks are not feasible for these data and a spatial domain of this size.

5. Simulations

We studied in Section 3 the covariance model implied by the likelihood approximation $\ell(A; \boldsymbol{Y_n})$. However, to use the likelihood approximation in practice, we usually require an approximation of the log determinant of $C_{\boldsymbol{n}}(A)$, which led to the approximation $\widetilde{\ell}(A; \boldsymbol{Y_n})$ in (2.7). In the following simulations, we study the behavior of parameter estimates found by maximizing this approximation to the likelihood and show that with the use of buffering, the parameter estimates can be asymptotically efficient and nearly unbiased.

The first simulation considers the case of $d = 2$ and $M = 2$ and studies the behavior of parameter estimates as the size of the region increases. We assume that the evolutionary spectral density is

$$A(\boldsymbol{x}/\boldsymbol{n}, \boldsymbol{\omega}) = I_1(\boldsymbol{x}/\boldsymbol{n})A_1(\boldsymbol{\omega}) + I_2(\boldsymbol{x}/\boldsymbol{n})A_2(\boldsymbol{\omega}),$$

where

$$A_j(\boldsymbol{\omega}) = \sigma_j \left(1 + \alpha_j^2(\sin^2(\omega_1/2) + \sin^2(\omega_2/2))\right)^{-2}$$

The spatial indicator functions partition the domain with a diagonal line as

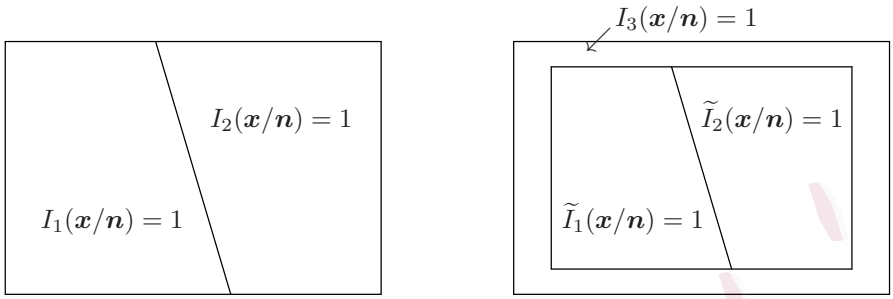

Figure 5.1: Indicator Functions

seen in Figure 5.1 and will be assumed to be known. We also assume that σ_1 and σ_2 are known, and the focus will be on the estimation of the parameters $\alpha_1, \alpha_2 > 0$, which can be interpreted as range parameters since increasing the range parameters produces stronger spatial correlation. We set $\alpha_1 = 1$ and $\alpha_2 = 2$ for all the simulations, so observations on the right side of the region generally exhibit stronger spatial correlation, and we set $\sigma_1 = 2.7379$ and $\sigma_2 = 5.9131$ so that the process has constant variance 1. An example of a simulation with $\boldsymbol{n} = (30, 60)$ is plotted in Figure 5.2.

Computationally efficient and nearly exact simulations from these models is possible since the stochastic integrals can be closely approximated by discretizing on a very fine evenly-spaced grid in frequency and computing the inverse discrete Fourier transform with an FFT. The details of the simulation of the processes are given in Section S2 of the supplementary material.

To obtain results on asymptotic efficiency of parameter estimates with the Whittle likelihood, Dahlhaus and Künsch (1987) required the size of the data taper to grow slower than the rate of growth of the sample size. In the simulations here, we set the size of the buffer on each edge to be $\sqrt{n_1}/3$, rounding to the nearest integer. The buffer essentially discards the observations near the boundary of the region, so it will no doubt lead to some loss of efficiency relative to the maximum likelihood parameter estimates. However, the ratio of the number of observations in the buffer to the total number of observations is proportional to $n^{-1/4}$ and thus goes to zero as $n \to \infty$.

We take $n_2 = 2n_1$, and for each $n = n_1 n_2$, we generate 1000 simulations, each time finding the unbuffered approximate likelihood parameter estimates, the buffered approximate likelihood parameter estimates, and when n is less than 1000, the exact maximum likelihood parameter estimates. The computa-

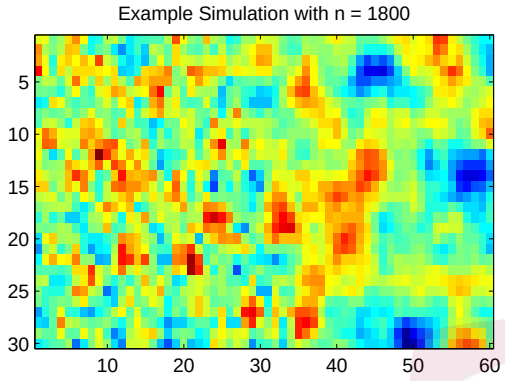

Figure 5.2: A single realization of the nonstationary lattice process with $\boldsymbol{n} = (30, 60)$.

tion of the exact likelihood would be too expensive with the largest sample sizes. To evaluate the estimates, we report the average bias and compare the root mean squared errors to the asymptotic standard deviations obtained by inverting the Fisher information matrices. The Fisher information can be computed exactly for all cases except $n = 20000$ and $n = 45000$, in which case we use a stochastic Hutchinson trace estimator (Hutchinson, 1990) to approximate the Fisher information, which produces an unbiased estimate of the Fisher information. The trace estimator is repeated with 50 different random probing vectors to produce a highly accurate estimate. We report the results in Tables 5.1 and 5.2.

The results in Table 1 indicate that the buffered approximate likelihood can produce asymptotically efficient parameter estimates for this nonstationary model. The buffered approximate likelihood is always an improvement over the unbuffered approximate likelihood, which is expected due to the connections between our approximate likelihood and the Whittle likelihood and the fact that the Whittle likelihood requires tapering to produce efficient parameter estimates when $d = 2$ or $d = 3$. When using the buffered approximate likelihood, the root mean squared errors for both parameter estimates approach the asymptotic error derived by inverting the Fisher information matrix. The results on the average bias in Table 2 show that the buffered approximate likelihood appears to be producing estimates that are nearly unbiased, with the size of the bias generally decreasing with the sample size. It is interesting to note that the biases for both of the approximate likelihoods are almost always negative, which is a consequence of the fact that the approximate likelihood assumes periodic correlation when it does not exist.

In the second simulation, we take $d = 3$ and $M = 2$ but do not assume that

			Root Mean Squared Error					
	Asymp. SE		Exact		Unbuffered		Buffered	
n	α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2
200	3.744	4.761	3.651	4.610	6.213	16.493	4.283	5.212
512	2.310	2.943	2.341	2.941	4.002	11.128	2.505	3.244
800	1.843	2.342	1.889	2.370	3.247	9.270	2.041	2.552
1800	1.221	1.557	–	–	2.173	6.270	1.378	1.709
3200	0.913	1.164	–	–	1.622	4.752	0.989	1.276
5000	0.730	0.930	–	–	1.342	3.836	0.748	0.982
20000	0.363	0.463	–	–	0.637	1.895	0.369	0.489
45000	0.242	0.309	–	–	0.434	1.304	0.252	0.350

Table 5.1: Asymptotic standard error (based on the Fisher information) and root mean squared error (all multiplied by 100) for estimating α_1 and α_2 using the exact likelihood, the unbuffered approximate likelihood, and the buffered approximate likelihood. The root mean squared errors are based on 1000 simulations for each sample size.

	Average Bias					
	Exact		Unbuffered		Buffered	
n	α_1	α_2	α_1	α_2	α_1	α_2
200	0.002	0.575	−4.788	−14.627	−0.069	−0.872
512	0.076	0.215	−3.136	−10.093	0.031	−0.580
800	0.030	0.105	−2.569	−8.478	−0.042	−0.478
1800	–	–	−1.729	−5.871	−0.024	−0.511
3200	–	–	−1.306	−4.471	−0.023	−0.410
5000	–	–	−1.109	−3.616	−0.072	−0.297
20000	–	–	−0.523	−1.806	−0.010	−0.114
45000	–	–	−0.355	−1.251	−0.013	−0.110

Table 5.2: Average bias (multiplied by 100) for estimating α_1 and α_2 using the exact likelihood, the unbuffered approximate likelihood, and the buffered approximate likelihood. The average biases are based on 1000 simulations for each sample size.

the partition is known. Our goal is to study how well the methods are are able to recover an unknown partition in three dimensions. We set $n_1 = 20$, $n_2 = 20$, and $n_3 = 30$ ($n = 12000$). The component transfer functions are

$$A_m(\boldsymbol{\omega}) = (1 + \alpha_m^2(\sin^2(\omega_1/2) + \sin^2(\omega_2/2) + \sin^2(\omega_3/2)))^{-1},$$

and we set $\alpha_1 = 4$ and $\alpha_2 = 8$. To define the partition, we set $I_1(\boldsymbol{x}/\boldsymbol{n}) = 1$ if $x_1/n_1 < -2x_3/n_3 + 1.5$. If we consider the first two smaller dimensions to be spatial dimensions and the third larger dimension to be a temporal dimension, then there is a spatial-temporal interaction in the nonstationarity. The partition changes near the middle of the time domain, with the bottom of the spatial region changing before the top of the spatial region. Figure 5.3 includes an illustration

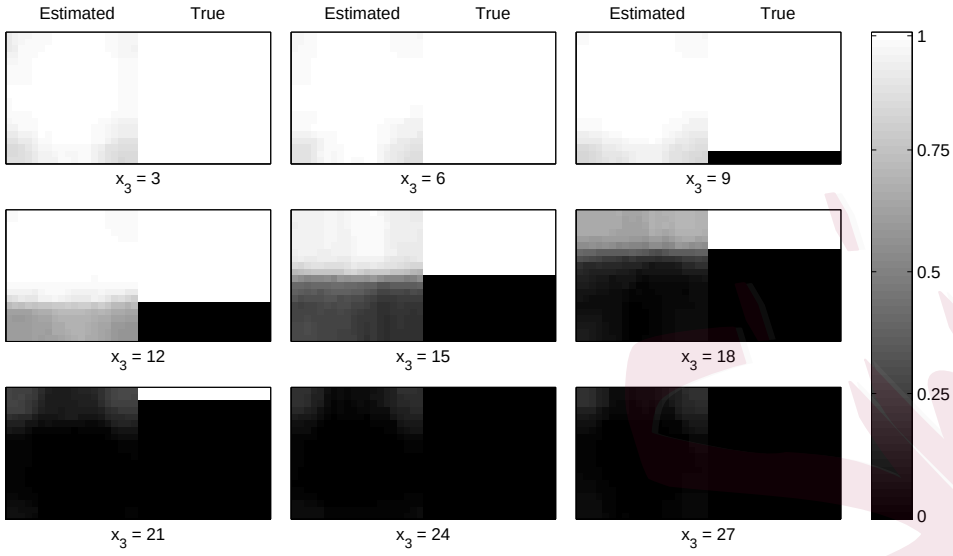

Figure 5.3: Proportion of fitted models out of 100 simulations that assign each location to the first block. The true model is plotted as well.

of the partition. Such a scenario may occur as a weather front moves across a region. We consider this boundary to be irregular since it does not divide the three-dimensional lattice into three-dimensional rectangles, in which case the Whittle likelihood could be applied separately to each rectangular subregion.

We simulate from this model 100 times, and for each simulation, we use the Ising model as described in Section 4 to search over the space of partitions with a buffer size of 2 observations in each dimension. For each simulation, we run the search algorithm separately with 200 random starting partitions, and we keep the final partition with the highest likelihood as the best-fitting model for each simulation. In Figure 5.3, we plot the proportion of best-fitting models (out of 100) that assigned each spatial location to the first block of the partition. Here we define the first block to be the one with shorter estimated spatial range since the labeling of the blocks is otherwise arbitrary. We plot these maps at several values in the third dimension. We see that the method is generally capturing the nonstationary structure present in the true model.

To investigate our methods' ability to select an appropriate number of components– $M = 2$ in this case–we fit models with $M = 3$ and $M = 4$ components using the greedy search algorithm described in Section 4. In Figure 5.4, we plot histograms of twice the change in the approximate loglikelihood among the 100 simulations as we move from an $M = j - 1$ model to an $M = j$ model. Increasing M by one

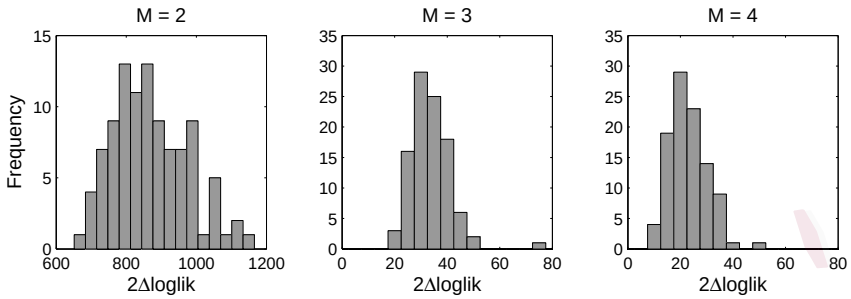

Figure 5.4: Histograms of twice the change in approximate loglikelihood from the $M = j - 1$ to the $M = j$ model.

introduces four additional parameters, which means that selection criteria based on AIC select $M > 2$ in every simulation, so stronger loglikelihood penalties are necessary. The BIC criterion selects $M > 2$ in 27 simulations if the penalty $4\log(n)$ is used and in 8 simulations if the penalty $4\log(2\pi n)$ is used. In Section 6, we advocate using changes in loglikelihood proportional to n for selection of M is less idealized situations, where practically minor deviations from stationarity can result in large changes in approximate loglikelihood when the number of observations is very large. This criterion is greater than 5% for the change from $M = 1$ to $M = 2$ in all 100 simulations and is less than 1% for the change from $M = 2$ to $M = 3$ in all 100 simulations.

6. Regional Weather Model Data

In the previous section, we showed that our approximate likelihood can produce nearly efficient parameter estimates compared with the exact likelihood, and we showed that the Ising model search algorithm and the approximate likelihood are able to detect stationary subregions when they exist. In this section, we apply our methods to a set of wind speed output from the HIRLAM regional weather model.

The model region contains the strait of Gibraltar separating southern Spain from northern Morocco. In Figure 6.5, we plot a map of the region and see that it contains parts of the Atlantic Ocean, the Mediterranean Sea, Africa, and Europe, so the geography of the region is diverse, especially for the study of surface winds, which depend heavily on the roughness of the Earth's surface. The output from the model is at a resolution of 0.05 degrees in both latitude and longitude and 3 hours in time. The size of the region is $\boldsymbol{n} = (21, 30, 240)$ in (latitude, longitude, time), which gives a total of $n = 151,200$ observations. Of course, the grid boxes do not form an exact lattice since the surface of the Earth is curved, but the region is small enough and close enough to the equator that the lattice

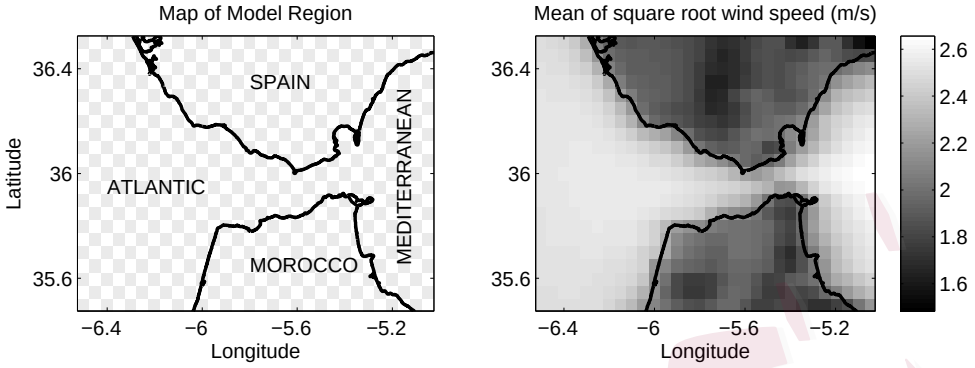

Figure 6.5: Map of the model region and mean of square root wind speed over 30 days. The squares indicate the size of the spatial grid boxes.

assumption is a good approximation. Square-root-transformed wind speeds appear to satisfy the Gaussian assumption more closely than untransformed wind speeds do. Histograms supporting this transformation can be found in Section S3 of the supplementary material.

The assumption of stationarity in both the mean and the covariance structure does appear to be violated. We plot in Figure 6.5 the mean of square root wind speed for each spatial grid box, where the mean is taken over time. It is clear that the model produces stronger winds over water than over land, which is not surprising since the sea surface is generally smoother than the land surface is. The covariance structure is also not stationary, which can be illustrated by plotting average local empirical variograms. The local empirical variograms are empirical variograms from subsets of data consisting of a 5×5 grid of observations at a single time point. The empirical variograms are computed for each time point and averaged over time. This is repeated for seven different spatial subsets, and the results are plotted in Figure 6.6. The local variograms indicate that the spatial correlation of wind speed is strongest in the Atlantic Ocean subregion and in the western part of the Spanish subregion.

Since the data are spatial-temporal, we write $\boldsymbol{x} = (\boldsymbol{s}, t)$ to denote spatial location $\boldsymbol{s}$ and temporal location t, and we write $\boldsymbol{\omega} = (\boldsymbol{\nu}, \lambda) = (\nu_1, \nu_2, \lambda)$ to denote spatial frequencies $\boldsymbol{\nu} = (\nu_1, \nu_2)$ and temporal frequency λ. We model the data as

$$Y(\boldsymbol{s}, t) = \mu_1(\boldsymbol{s}) + \mu_2(t) + Y_0(\boldsymbol{s}, t),$$

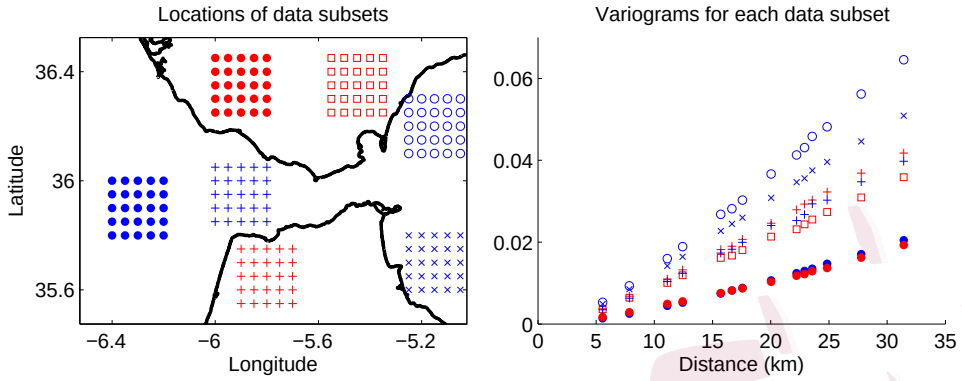

Figure 6.6: Local variograms in seven subregions indicating nonstationarity covariance.

where Y_0 is a mean-zero process with evolutionary spectrum

$$A(\boldsymbol{x}/\boldsymbol{n}, \boldsymbol{\omega}) = \sum_{m=1}^{M} I_m(\boldsymbol{s}) A_m(\boldsymbol{\omega}),$$

where I_m are indicator functions depending on space alone, and A_m has parametric form

$$A_m(\boldsymbol{\omega}) = \sigma_m^2 \left(1 + \alpha_m^2(\sin^2(\nu_1/2) + \sin^2(\nu_2/2)) + \beta_m^2 \sin^2(\lambda/2)\right)^{-2}, \qquad (6.1)$$

so that the spatial range parameter for component m is α_m, and the temporal range parameter for component m is β_m, as opposed to three-dimensional model in the simulation, which assumed a constant range among the three dimensions within a stationary subregion. Each component also has a scale parameter σ_m^2.

Although we do not pursue them for the wind data here, it is possible to define more flexible parametric forms for the spatial-temporal transfer functions. One may assign a parameter for the exponent in (6.1), which is fixed there at -2, giving

$$A_m(\boldsymbol{\omega}) = \sigma_m^2 \left(1 + \alpha_m^2(\sin^2(\nu_1/2) + \sin^2(\nu_2/2)) + \beta_m^2 \sin^2(\lambda/2)\right)^{-\nu_m}.$$

We may also desire a model that more flexibly distinguishes between the spatial and temporal domains. For example, we could use

$$A_m(\boldsymbol{\omega}) = \sigma_m^2 \left[\left(\frac{1}{2} + \alpha_m^2(\sin^2(\nu_1/2) + \sin^2(\nu_2/2)) \right)^{\gamma_m} + \left(\frac{1}{2} + \beta_m^2 \sin^2(\lambda/2) \right)^{\delta_m} \right]^{-\nu_m},$$

which reduces to the previous model if $\gamma_m = \delta_m = 1$.

As in the simulation study, we will use a buffer to mitigate the edge effects, so the model-fitting procedure will assume instead that

$$A(\boldsymbol{x}/\boldsymbol{n}, \boldsymbol{\omega}) = \sum_{m=1}^{M} \widetilde{I}_m(\boldsymbol{s}) A_m(\boldsymbol{\omega}) + I_{M+1}(\boldsymbol{s}) A_{M+1}(\boldsymbol{\omega}),$$

where $I_{M+1}(\boldsymbol{s})$ indicates a buffer of two observations from the spatial border, and $\widetilde{I}_m(\boldsymbol{s}) = I_m(\boldsymbol{s})$ if $I_{M+1}(\boldsymbol{s}) = 0$ and equals zero otherwise. We do not buffer in the time dimension since there are many more observations in time than there are in either of the two spatial dimensions.

Since the indicator functions depend on space alone, the process is assumed to be nonstationary in space but stationary in time. In other applications, one may remove this constraint or, alternatively, constrain the indicator functions to depend on time alone, which would specify a model that is stationary in space but not in time. The time span for the simulation is 30 days in January, so we do not expect a significant seasonal component to the nonstationarity. There is a possibility of a diurnal cycle to the covariance nonstationarity, but we assume that the nonstationarity associated with the diurnal cycle is captured in the mean function $\mu_2(t)$. We estimate $\mu_1(\boldsymbol{s})$ by averaging over time the square root wind speed for each spatial grid box, and we estimate $\mu_2(t)$ by averaging the spatial map of square root wind speeds at each time point and subtracting a grand mean. The mean functions are assumed to be known and subtracted from the data so that we analyze the anomalies that are assumed to have a mean of 0.

We are interested in whether the nonstationary covariance may be attributed to different atmospheric behavior over the land versus over the sea. To explore this possibility, we fit a model with $M = 2$ and a partition defined by the coastal boundaries, as seen in Figure 6.7. We find that a land/ocean model increases the approximate loglikelihood by 89.99 units over the best stationary model. The fitted variograms over land and ocean do not differ much, suggesting that coastal boundaries do not explain the nonstationary covariance very well.

The empircal variograms indicate that the data do exhibit nonstationary covariance, but the model fitted above suggests that this nonstationary covariance model cannot be simply attributed to land/ocean effects. We use the Ising model algorithm to search over the space of possible two-block partitions. We ran the algorithm with 100 different starting partitions, and we plot the partition that gives the highest approximate likelihood in Figure 6.8. This partition divides the spatial domain into one region that contains the Atlantic and the western part of the Spanish subregion, and one region that contains the rest of the spatial domain. We also plot in Figure 6.8 the fitted variograms associated with the two blocks of the best partition and see that they qualitatively match the empirical variograms from Figure 6.6. The maximum approximate likelihood associated

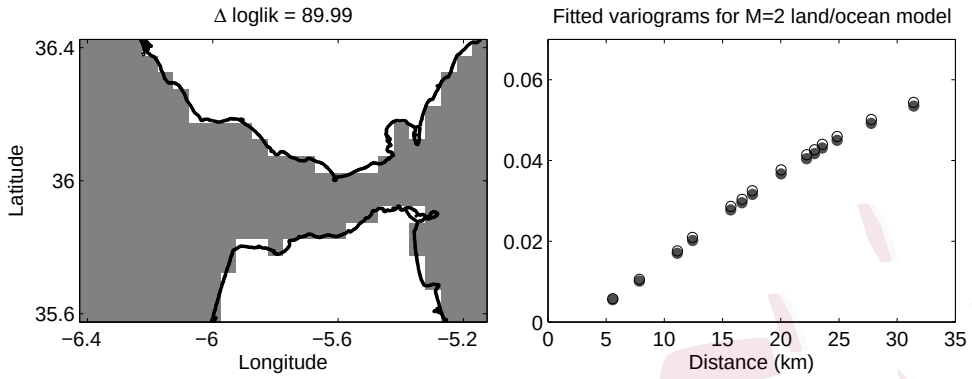

Figure 6.7: A model in which the partition is determined by the coastal boundaries. The number above the map gives the increase in approximate loglikelihood above the best stationary model. The buffered region is left out of the map. The variograms associated with each subregion are plotted on the right.

with this partition increased the approximate likelihood by 4235.53 units over a stationary model. This is a significantly larger increase compared with the increase of 89.99 loglikelihood units achieved with the land/ocean model.

To find our best 3-block partition, we search over partitions of the spatial domain that are formed by keeping the first block as it is in Figure 6.8, and then partitioning the second block into two smaller blocks, for three blocks total. Then we search over partitions that are formed by keeping the second block as it is, partitioning the first block into two smaller blocks. We conducted the search with 25 different starting subpartitions of each of the two blocks. We plot the best three-block partition we found in Figure 6.9, along with the fitted variograms associated with the three different blocks. The three-block partition increases the approximate likelihood by 5188.67 units over a stationary model.

The fitted variograms for both the two-block and the three-block models agree qualitatively with the empirical variograms plotted in Figure 6.6. The three-block model implies that the spatial correlation is strongest over the Atlantic and the western Spanish subregion and weakest over the eastern land regions and the Mediterranean. The model also finds that there is a small region of weaker spatial correlation along the western coast of Morocco, suggesting that the winds may be more variable and turbulent in this area due to the geographic boundary between land and sea.

To explore the need for more terms in the low dimensional form for the evolutionary transfer function, we continued our sequential model fitting procedure with $M = 4$ and $M = 5$. Each additional term comes with three additional parameters. We denote the maximum approximate loglikelihood values for the

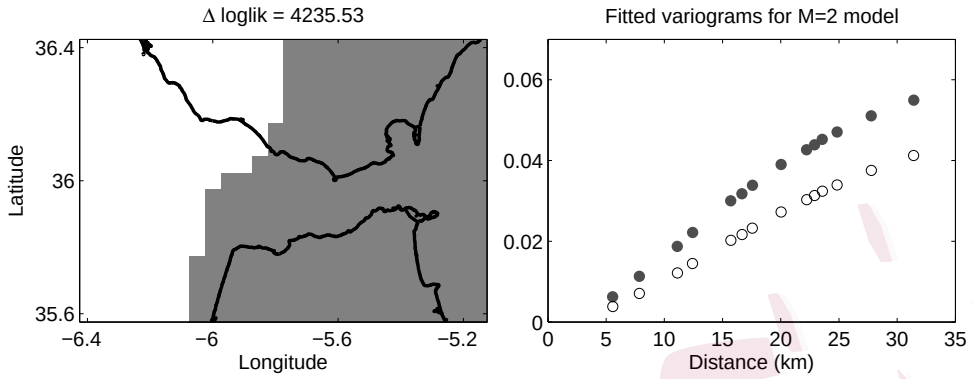

Figure 6.8: Best 2-block partition. The number above the map gives the increase in approximate loglikelihood above the best stationary model. The buffered region is left out of the map. The variograms associated with each subregion are plotted on the right.

M-component model by L_M, and in Table 6.3, we provide the resulting values. We consider a comparison of the twice the change in loglikelihood to the number of observations to be a useful measure of the practical value of the increased model complexity. As we increase M from 1 to 2, twice the change in loglikelihood is 5.6% of the total number of observations but less than 1% as we increase M from 3 to 4 and 4 to 5. For these data, the $M = 2$ or $M = 3$ models are a sufficiently complex description of the data, and we do not see much practical value in increasing the model complexity beyond $M = 3$.

M	$2(L_M - L_{M-1})$	$2(L_M - L_{M-1})/n$
2	8471	0.056
3	1906	0.013
4	1153	0.008
5	988	0.007

Table 6.3: Changes in approximate loglikelihood with increasing model complexity. L_M gives the approximate loglikelihood for the best-fitting M-component model.

7. Discussion

We propose a new Gaussian likelihood approximation for nonstationary and nonseparable spatial-temporal lattice models and conduct a careful study of its properties. The approximation is computationally efficient since it relies on preconditioned linear solvers and an FFT algorithm. Through simulation, we show an example in which the likelihood approximation produces efficient parameter estimates when a buffering operation is applied. The fact that buffering improves the parameter estimates is not surprising since it is well-known that tapering re-

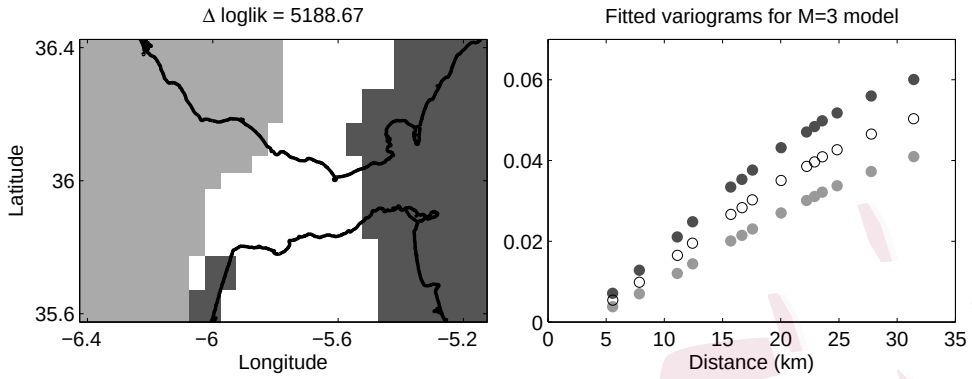

Figure 6.9: Best 3-block partition. The number above the map gives the increase in approximate loglikelihood above the best stationary model. The buffered region is left out of the map. The variograms associated with each subregion are plotted on the right.

duces edge effects and improves parameter estimates when the Whittle likelihood is used, and our likelihood approximation is an extension of the Whittle likelihood. In this work, we consider weight functions that are indicators since an effective and efficiently computable preconditioner is available. In principle, our methods can be extended to handle smoother weight functions as long as one can find good preconditioners.

We advocate for a sequential procedure for building nonstationary spatial-temporal models that is well-suited for the evolutionary spectrum models that we study. Our methods employ parametric models for the evolutionary spectra. The sequential procedure is implemented to estimate spatial heterogeneity in the covariance structure of wind speed output from a regional weather model. We employ random algorithms based on the Ising model to search over the space of partitions of the spatial domain that separate the domain into contiguous regions of stationarity. Our results agree with an exploratory variogram analysis and indicate that our methods are capable of recovering the underlying spatial structure of nonstationarity. We recommend using changes loglikelihood relative to the number of observations to choose the number of components, M.

Acknowledgment

We acknowledge Pilar Muñoz of the Laboratory of Information Analysis and Modelling at Universitat Politècnica de Catalunya for supplying the HIRLAM weather model output data. The authors are supported by the National Science Foundation's Research Network for Statistical Methods for Atmospheric and Oceanic Sciences (STATMOS), award number 1107046.

References

Calder, C. A. (2007). Dynamic factor process convolution models for multivariate space–time data with application to air quality assessment. *Environmental and Ecological Statistics*, 14(3):229–247.

Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. *Stoch. Proc. and their Applications*, 62(1):139–168.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *The Annals of Statistics*, 28(6):1762–1794.

Dahlhaus, R. and Künsch, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, 74(4):877–882.

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

Fuentes, M. (2002a). Interpolation of nonstationary air pollution processes: a spatial spectral approach. *Statistical Modelling*, 2(4):281–298.

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

Guinness, J. and Stein, M. L. (2013). Transformation to approximate independence for locally stationary Gaussian processes. *J. of Time Series Analysis*.

Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika*, 69(1):95–105.

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

Hutchinson, M. F. (1990). A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Communications in Statistics-Simulation and Computation*, 19(2):433–450.

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1):253–258.

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

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *J. of the Royal Stat. Soc., Series B*, 27(2):204–237.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41(4):434–449.