

Asymptotic Theory of Cepstral Random Fields

Tucker S. McElroy* and Scott H. Holan^{†‡}

Abstract

Random fields play a central role in the analysis of spatially correlated data and, as a result, have a significant impact on a broad array of scientific applications. Given the importance of this topic, there has been a substantial amount of research devoted to this area. However, the cepstral random field model remains largely underdeveloped outside the engineering literature. We provide a comprehensive treatment of the asymptotic theory for two-dimensional random field models. In particular, we provide recursive formulas that connect the spatial cepstral coefficients to an equivalent moving-average random field, which facilitates easy computation of the necessary autocovariance matrix. Additionally, we establish asymptotic consistency results for Bayesian, maximum likelihood, and quasi-maximum likelihood estimation of random field parameters and regression parameters. Further, in both the maximum and quasi-maximum likelihood frameworks, we derive the asymptotic distribution of our estimator. The theoretical results are presented generally and are of independent interest, pertaining to a wide class of random field models. The results for the cepstral model facilitate model-building: because the cepstral coefficients are unconstrained in practice, numerical optimization is greatly simplified, and we are always guaranteed a positive definite covariance matrix. We show that inference for individual coefficients is possible, and one can refine models in a disciplined manner. Finally, our results are illustrated through simulation and the analysis of straw yield data in an agricultural field experiment.

Keywords: Bayesian estimation; Cepstrum; Exponential spectral representation; Lattice data; Spatial statistics; Spectral density.

*Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, tucker.s.mcelroy@census.gov

[†](to whom correspondence should be addressed) Department of Statistics, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO, 65211-6100, holans@missouri.edu

[‡]Disclaimer: This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

1 Introduction

Spatial data feature heavily in many scientific disciplines including ecology, environmental science, epidemiology, geography, geology, small area estimation, and socio-demographics. Although spatial data can be broadly placed into three categories: *geostatistical data*, *lattice data*, and *spatial patterns* (Cressie, 1993), our focus mainly resides in the development of cepstral random field models for spatial lattice data. That is, we consider random fields where the index set for the variables is $\mathbb{Z} \times \mathbb{Z}$ – appropriate for image processing, for example.

Research on spatial random fields dates back over half a century, e.g., see (Whittle, 1954). Other references on spatial random fields include Besag (1972, 1974), Guyon (1982), Rosenblatt (1985), Besag and Green (1993), and Rosenblatt (2000), among others. Comprehensive overviews can be found in Cressie (1993), Stein (1999), Banerjee et al. (2004), Cressie and Wikle (2011), and the references therein.

For a stationary Gaussian random field, it is natural to impose a Markov structure, as described in Rue and Held (2010), in order to obtain an inverse covariance matrix (i.e., a precision matrix) that has a sparse structure, because this will ensure speedy computation of maximum likelihood estimates. Rue and Held (2010) show how careful specification of conditional distributions generates a well-defined random field. However, this technique relies upon imposing *a priori* a sparse structure on the precision matrix, i.e., demanding that many conditional precisions be zero. In contrast, the cepstral random field does not generate a sparse covariance (or precision) matrix, and yet always yields a well-defined spatial random field. It is actually a general way of specifying a non-isotropic random field when we do not have prior notions about conditional variances or precisions, and in particular for fields that are not Markovian.

The cepstral random field allows for unconstrained optimization of the objective function, i.e., each model coefficient can be any real number independently of the others; this appealing property is in marked contrast to other models and approaches, such as moving averages or Markov random fields (these require constraints on parameters to achieve identifiability and/or a well-defined process). Despite these advantages, the cepstral random field model, first developed by Solo (1986), has only played a minor role. In the development of this model, Solo (1986) presents estimation approaches by both log periodogram regression and Whittle maximum likelihood. Additionally, based on information criterion, Mallows' C_p , and hypothesis testing, the author briefly describes methods for model selection. Broadly speaking, Solo (1986) presents an approach to modeling cepstral random fields, based on regression and Whittle maximum likelihood, and leaves the asymptotic properties of the estimators untreated.

This paper provides a different approach to model fitting and develops the first comprehensive treatment of the theory for cepstral random field models. In particular, we establish recursive formulas for connecting cepstral random fields to moving average random fields, thus facilitating

efficient computation of the spatial autocovariances. Critically, the resulting autocovariance matrix is guaranteed to be positive-definite; note that if we were to work with a moving average (MA) field instead, it would not be identifiable without imposing further complicated parameter restrictions. Thus, given these autocovariances, the exact likelihood function for the cepstral random field model can be easily expressed as a function of the unconstrained parameters.

Additionally, we develop asymptotic results for Bayesian, maximum likelihood, and quasi-maximum likelihood estimation of field parameters and regression parameters. In particular, we establish asymptotic consistency in both the Bayesian and likelihood settings and provide Central Limit Theorems for the frequentist estimators we propose. We discuss the computational advantages of the cepstral model, and propose an exact Whittle likelihood that avoids the burdensome inversion of the autocovariance matrix. Although our primary focus is on cepstral models, the theoretical developments are presented for general random field models with regression effects. Our results are of independent interest and extend the existing results of Mardia and Marshall (1984), providing a rigorous framework for conducting model building and inference.

As discussed in Sections 2 and 3, the proposed cepstral models are computationally advantageous over many current models (e.g., spatial autoregressive models), because no constraints need to be imposed on the parameters to insure the resulting autocovariance matrix remains positive definite. In fact, given the recursive formulas of Section 2, one can model the two-dimensional cepstral coefficients (i.e., the Fourier coefficients of the two-dimensional log spectrum) and arrive at the autocovariances without the need for direct Fourier inversion.

Since the model's first inception (Solo, 1986), the cepstral random field literature has remained sparse, with relatively few examples to date. For example, Cressie (1993, p. 448) makes brief mention of the model. In a different context, Noh and Solo (2007) use cepstral random fields to test for space-time separability. Sandgren and Stoica (2006) use two-dimensional cepstrum thresholding models to estimate the two-dimensional spectral density. However, this work does not treat the random field case. Related to our work, Kizilkaya and Kayran (2005) derive an algorithm for computing cepstral coefficients from a known ARMA random field, whereas Kizilkaya (2007) provides a recursive formula for obtaining a nonsymmetric half plane MA random field models for a given cepstral specification. In contrast, our recursive formulas provide unrestricted MA random fields as well as the necessary autocovariances for expressing the Gaussian likelihood.

This paper proceeds as follows. Section 2 describes the cepstral model and its computation. Specifically, this section lays out the recursive formulas that are needed to estimate the autocovariances given the cepstral coefficients. Section 3 details the different model fitting methods, including Bayesian, maximum likelihood, quasi-maximum likelihood, and exact Whittle likelihood. In addition, this section describes extensions to missing data, imputation, and signal extraction. Our theoretical results are provided in Section 4. Here, we establish consistency and asymptotic normality of the proposed estimators. Section 5 illustrates the models effectiveness through sim-

ulation and an application to straw yield data from an agricultural field experiment. Section 6 contains concluding discussion. For convenience of exposition all proofs are left to the Appendix.

2 The Cepstral Model and its Computation

We begin by introducing some basic concepts about spatial random fields, and then we specialize to the cepstral random field, with a focus on computation of autocovariances. References on spatial random fields include Whittle (1954), Besag (1972), Rosenblatt (1985, 2000), Solo (1986), Cressie (1993), Kedem and Fokianos (2002), and Rue and Held (2010). A random field $\mathbb{Y} = \{\mathbb{Y}_{r,s}\}$ is a process with indices on a lattice, which in this paper we take to be $\mathbb{Z} \times \mathbb{Z}$. Typically a random field has a mean function $\mu_{r,s} = \mathbb{E}\mathbb{Y}_{r,s}$, which may be modeled through regression variables (Cressie, 1993). The mean-corrected field $\mathbb{Y} - \{\mu_{r,s}\}$ will be denoted by \mathbb{W} .

Interest focuses upon stationary random fields – which in practice is often adequate once mean effects are identified and accounted for – whose joint distribution functions are shift invariant. When all moments are defined, this is equivalent to the higher-order cumulants (Brillinger, 2001) being dependent only on lags between the mean-centered variables. The second-order cumulant function, or autocovariance function (acf), is defined via $\text{Cov}(\mathbb{Y}_{r,s}, \mathbb{Y}_{a,b}) = \mathbb{E}[\mathbb{W}_{r,s}\mathbb{W}_{a,b}] = \gamma_{a-r,b-s}$ for all $a, b, r, s \in \mathbb{Z}$. It is convenient to summarize this second-order structure through the spectral density F defined on $[-\pi, \pi]^2$, which depends on two frequencies. Letting $Z_j = e^{-i\lambda_j}$ for $j = 1, 2$, the spectral density is related to the acf via the formula

$$F(\lambda_1, \lambda_2) = \sum_{h,k \in \mathbb{Z}} \gamma_{h,k}(F) Z_1^h Z_2^k. \quad (1)$$

Here we write $\gamma(F)$ for the acf associated with the spectrum F , and it in turn is expressed in terms of F via Fourier inversion as

$$\gamma_{h,k}(F) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} F(\lambda_1, \lambda_2) Z_1^{-h} Z_2^{-k} d\lambda_1 d\lambda_2. \quad (2)$$

As a general notation, let the normalized double integral over both frequencies be abbreviated by the expression $\langle \cdot \rangle$, so that $\gamma_{h,k}(F) = \langle F Z_1^{-h} Z_2^{-k} \rangle$ is compactly expressed. Now it follows elementarily from the commutativity of the field \mathbb{Y} variables that $\gamma_{h,k} = \gamma_{-h,-k}$, and hence the corresponding F in (1) must have mirror reflectional symmetry through both axes, i.e., $F(\lambda_1, \lambda_2) = F(-\lambda_1, -\lambda_2)$. Furthermore, the acf of a random field is always positive-definite (Cressie, 1993) and the corresponding spectrum is non-negative (Bochner, 1955).

With these preliminaries, let us proceed to consider models of lattice data. A spatial model for continuous-valued random variables should – at a minimum – capture second order structure in the data, which is summarized through the acf. However, a putative acf may or may not have non-negative Fourier Transform (FT) (1), whereas any valid acf of a stationary field *must* have

non-negative spectrum F . One way to ensure our model has such a valid acf is to model F – utilizing some class of non-negative functions – and determine the corresponding covariances via (2). This is the philosophy behind the versatile exponential time series model of Bloomfield (1973). The idea there was to expand the log spectrum in the complex exponential basis functions, with a truncation of the expansion corresponding to a postulated model.

The same idea is readily adapted to the spatial context; Solo (1986) seems to be the first formal presentation of this idea. So expanding log F in each frequency concurrently yields

$$\log F(\lambda_1, \lambda_2) = \sum_{j,k \in \mathbb{Z}} \Theta_{j,k} Z_1^j Z_2^k.$$

The coefficients $\{\Theta_{j,k} = \langle \log F Z_1^{-j} Z_2^{-k} \rangle\}$ are called the cepstral coefficients; see also the recent treatment of Kizilkaya and Kayran (2005). A pleasing feature of this representation is that F^{-1} has cepstral coefficients $\{-\Theta_{j,k}\}$. By truncating the summation, we obtain a parametric model that can approximate the second-order structure of *any* random field with bounded spectrum. So we obtain the cepstral model of order (p, q) given by

$$F(\lambda_1, \lambda_2) = \exp \left\{ \sum_{j=-p}^p \sum_{k=-q}^q \Theta_{j,k} Z_1^j Z_2^k \right\}. \quad (3)$$

Note that the cepstral coefficient $\Theta_{0,0}$ has no sinusoidal function multiplying it, and hence $\exp \Theta_{0,0}$ quantifies the scale of the data. In one dimension, this would be called the innovation variance; note that $\Theta_{0,0} = \langle \log F \rangle$. Because the complex exponentials form a complete orthonormal basis set, it is impossible for two distinct values of Θ to produce an identical function F ; hence the model is identifiable.

Further special cases of the general cepstral field model are considered in Solo (1986). Because F has mirror reflectional symmetry, the cepstral coefficients do as well, i.e., $\Theta_{j,k} = \Theta_{-j,-k}$. We might consider a model completely generated by the positive orthant of the Θ matrix, so that $\Theta_{j,k} = \Theta_{-j,k} = \Theta_{j,-k} = \Theta_{-j,-k}$. Alternatively, we might have a positive quadrant model, where $\Theta_{j,k}$ is only nonzero when $j, k \geq 0$. Zeroing out other batches of coefficients produces special cases; e.g., a separable random field is obtained by imposing that only the axes of Θ are non-zero. Also an approximately isotropic random field structure can be imposed, as described in Solo (1986).

In order to fit this model to Gaussian data, it is necessary to compute the acf from a given specification of cepstral coefficients. We next describe two approaches to this: one is approximate, and the other is exact. Both differ from the fitting techniques in Solo (1986), who advocates an asymptotic likelihood (or Whittle) calculation. The first approach involves a straightforward discretization of (2) – together with (3) – utilizing the Riemann approximation. So long as the spectrum is a bounded function, this method is arbitrarily accurate (since the practitioner controls the mesh size). In order to accomplish the computation, without loss of generality let $q = p$, so

that the cepstral coefficients are given by a $(2p + 1) \times (2p + 1)$ grid Θ (if $q < p$, just fill in some entries of Θ with zeroes).

Now we refer to the entries of Θ via $\Theta_{j,k}$ with $-p \leq j, k \leq p$, which is a Cartesian mode of indexing; this differs from the style of indexing pertinent to matrices. We can map this grid to a matrix $[\Theta]$ (and back), with the following rule:

$$[\Theta]_{jk} = \Theta_{k-p-1, p+1-j} \quad \Theta_{mn} = [\Theta]_{p+1-n, m+p+1} \quad (4)$$

for $1 \leq j, k \leq 2p + 1$ and $-p \leq m, n \leq p$. We will consider a set of frequencies $\{j\pi/M, k\pi/M\}$ for $-M \leq j, k \leq M$, which is an order M discretization of $[-\pi, \pi]^2$. Consider a complex-valued $2p+1 \times 2M+1$ matrix E with entries $E_{jk} = \exp\{\pi i(p+1-j)(M-k+1)/M\}$ for $j = 1, 2, \dots, 2p+1$ and $k = 1, 2, \dots, 2M+1$. Then

$$\log F \left(\pi \frac{M+1-s}{M}, \pi \frac{M+1-r}{M} \right) = \left\{ \overline{E}' [\Theta] E \right\}_{rs} \quad (5)$$

for $1 \leq r, s \leq 2M+1$ holds (proved in the Appendix). So we can evaluate F on the desired grid of frequencies by exponentiating each entry of the matrix $\overline{E}'[\Theta]E$, according to (5).

Next, consider the grid of autocovariances given by $\Gamma = \{\gamma_{h,k}\}_{h,k=-H}^H$ for some maximal lag H . Discretizing equation (2) yields

$$\begin{aligned} \gamma_{h,k} &\cong (2M+1)^{-2} \sum_{r,s=1}^{2M+1} F \left(\pi \frac{M+1-s}{M}, \pi \frac{M+1-r}{M} \right) \exp \left\{ \pi i h \frac{M+1-s}{M} \right\} \exp \left\{ \pi i k \frac{M+1-r}{M} \right\} \\ &= (2M+1)^{-2} \sum_{r,s=1}^{2M+1} \exp \left\{ \left(\overline{E}' [\Theta] E \right)_{rs} \right\} \overline{G}_{h+H+1,s} G_{H+1-k,r}, \\ G_{jk} &= \exp \left\{ \pi i \frac{(H+1-j)(M+1-k)}{M} \right\} \end{aligned}$$

for $1 \leq j \leq 2H+1$ and $1 \leq k \leq 2M+1$. Here G is a $2H+1 \times 2M+1$ dimensional matrix similar to E . For applications, it is convenient to compute $[\Gamma]$ directly as follows:

$$[\Gamma]_{\ell,m} = \gamma_{m-H-1, H+1-\ell} \cong (2M+1)^{-2} \sum_{r,s=1}^{2M+1} \exp \left\{ \left(\overline{E}' [\Theta] E \right)_{rs} \right\} \overline{G}_{ms} G_{\ell r},$$

or in matrix form is given by

$$[\Gamma] \cong (2M+1)^{-2} G \exp \{ \overline{E}' [\Theta] E \} \overline{G}'. \quad (6)$$

In this formula we have written the exponential of a matrix, which here is *not* the “matrix exponential,” but rather just consists of exponentiating each entry of the matrix (this is clear from the antecedent formula). So (6) produces an arbitrarily fine approximation to the acf (taking M as large as desired). The algorithm takes a given Θ , produces $[\Theta]$ via (4), computes E and G (ahead of time, as they do not depend upon the parameters), and determines $[\Gamma]$ via (6).

Now we present the exact method for computing the acf from the cepstral matrix. Our approach is similar to that of Section 3 of Kizilkaya and Kayran (2005), though with one important difference. They present an algorithm for computing cepstral coefficients from known coefficients of an ARMA random field. Instead, we take the cepstral coefficients as given, compute coefficients of certain corresponding MA random fields, and from there obtain the acf. In order to fit the Gaussian likelihood, we need to compute the acf from the cepstral matrix, not the reverse.

First, consider decomposing the cepstral matrix Θ into four quadrants, plus the four semi-axes, plus the origin (Kizilkaya (2007) has a similar decomposition). In the following treatment, we can let $p = \infty$ for a general decomposition, or take $p < \infty$ for the purposes of modeling:

$$\begin{aligned} \sum_{j,k=-p}^p \Theta_{j,k} Z_1^j Z_2^k &= \Theta_{0,0} + \sum_{j,k=1}^p \Theta_{j,k} Z_1^j Z_2^k + \sum_{j,k=1}^p \Theta_{j,k} Z_1^{-j} Z_2^{-k} \\ &+ \sum_{j,k=1}^p \Theta_{-j,k} Z_1^{-j} Z_2^k + \sum_{j,k=1}^p \Theta_{-j,k} Z_1^j Z_2^{-k} \\ &+ \sum_{j \neq 0} \Theta_{j,0} Z_1^j + \sum_{k \neq 0} \Theta_{0,k} Z_2^k. \end{aligned} \quad (7)$$

We have used some of the simple reflection properties of the cepstral matrix for this decomposition. Also terms are grouped appropriately. We introduce the device of a “causal” field and a “skew” field as follows. The causal field is a MA field that only involves coefficients with indices in the positive quadrant, whereas the skew field essentially is defined over the second quadrant. More precisely, we have

$$\begin{aligned} \gamma_{r,s}(\Psi) &= \sum_{m,n \geq 0} \psi_{r+m,s+n} \psi_{m,n} \\ \left| \sum_{j,k \geq 0} \psi_{j,k} Z_1^j Z_2^k \right|^2 &= \sum_{r,s \in \mathbb{Z}} \gamma_{r,s}(\Psi) Z_1^r Z_2^s \end{aligned} \quad (8)$$

for the causal field. The causal field may be written formally (in terms of backshift operators B_1, B_2) as $\Psi(B_1, B_2) = \sum_{j,k \geq 0} \psi_{j,k} B_1^j B_2^k$. That is, the $\psi_{j,k}$ coefficients define the moving average representation of the causal field, and $\{\gamma_{r,s}(\Psi)\}$ is its acf. It is important that we set $\psi_{0,0} = 1$. Similarly, let $\Phi(B_1, B_2) = \sum_{j,k \geq 0} \phi_{j,k} B_1^{-j} B_2^k$ for the skew-field, which in the first index depends on the forward shift operator B_1^{-1} , but on the backshift operator B_2 in the second index. Also,

$$\begin{aligned} \gamma_{r,s}(\Phi) &= \sum_{m,n \geq 0} \phi_{r+m,s+n} \phi_{m,n} \\ \left| \sum_{j,k \geq 0} \phi_{j,k} Z_1^{-j} Z_2^k \right|^2 &= \sum_{r,s \in \mathbb{Z}} \gamma_{r,s}(\Phi) Z_1^{-r} Z_2^s. \end{aligned} \quad (9)$$

We also have two one-dimensional random fields, corresponding to the axes of the cepstral matrix, given by $\Xi(B_1) = \sum_{k \geq 0} \xi_k B_1^k$ and $\Omega(B_2) = \sum_{k \geq 0} \omega_k B_2^k$, which have acfs $\gamma_h(\Xi) = \sum_{k \geq 0} \xi_k \xi_{k+h}$ and

$\gamma_h(\Omega) = \sum_{k \geq 0} \omega_k \omega_{k+h}$, respectively. Now each of these MA random fields has a natural cepstral representation, each of which can be made to correspond to a term of (7):

$$\sum_{j,k \geq 0} \psi_{j,k} Z_1^j Z_2^k = \exp \left\{ \sum_{j,k=1}^p \Theta_{j,k} Z_1^j Z_2^k \right\} \quad (10)$$

$$\sum_{j,k \geq 0} \phi_{j,k} Z_1^{-j} Z_2^k = \exp \left\{ \sum_{j,k=1}^p \Theta_{-j,k} Z_1^{-j} Z_2^k \right\} \quad (11)$$

$$\sum_{j \geq 0} \xi_j Z_1^j = \exp \left\{ \sum_{j=1}^p \Theta_{j,0} Z_1^j \right\} \quad (12)$$

$$\sum_{j \geq 0} \omega_j Z_2^j = \exp \left\{ \sum_{j=1}^p \Theta_{0,j} Z_2^j \right\}. \quad (13)$$

By expanding the exponentials as power series on the right hand side of (10), we see that Z_1 and Z_2 both occur at least once in every term; hence we must have $\psi_{0,0} = 1$, $\psi_{0,1} = 0$, and $\psi_{1,0} = 0$. The other coefficients $\psi_{j,k}$ have no constraints, but are determined by the cepstral coefficients. Similarly $\phi_{0,0} = 1$, $\phi_{0,1} = 0$, and $\phi_{1,0} = 0$, while $\xi_0 = 1 = \omega_0$.

Comparing these associations with (7), we see at once that the spectrum is equal to the magnitude squared of the product of the above factors (10), (11), (12), and (13), along with the constant $e^{\Theta_{0,0}}$. Taking the product of the corresponding squared magnitudes yields

$$\begin{aligned} & e^{\Theta_{0,0}} \sum_{r,s,a,b,m,n \in \mathbb{Z}} \gamma_{r,s}(\Psi) \gamma_{a,b}(\Phi) \gamma_m(\Xi) \gamma_n(\Omega) Z_1^{r-a+m} Z_2^{s+b+n} \\ &= e^{\Theta_{0,0}} \sum_{j,k \in \mathbb{Z}} \left(\sum_{a,b \in \mathbb{Z}} \gamma_{a,b}(\Phi) \left[\sum_{m,n \in \mathbb{Z}} \gamma_{j+a-m,k-b-n}(\Psi) \gamma_m(\Xi) \gamma_n(\Omega) \right] \right) Z_1^j Z_2^k. \end{aligned}$$

Matching coefficients (i.e., computing the inverse FT) then produces the acf of the cepstral model:

$$\gamma_{j,k}(F) = e^{\Theta_{0,0}} \sum_{a,b \in \mathbb{Z}} \gamma_{a,b}(\Phi) \left[\sum_{m,n \in \mathbb{Z}} \gamma_{j+a-m,k-b-n}(\Psi) \gamma_m(\Xi) \gamma_n(\Omega) \right]. \quad (14)$$

This produces the model acf in terms of the intermediary acfs $\gamma(\Psi)$, $\gamma(\Phi)$, $\gamma(\Xi)$, and $\gamma(\Omega)$. To make the algorithm complete, we must obtain the field coefficients in terms of known cepstral matrix coefficients. Differentiating (10) with respect to Z_1 yields

$$\sum_{j,k \geq 0} \psi_{j,k} j Z_1^{j-1} Z_2^k = \sum_{h,l \geq 1} \left(\sum_{m,n \geq 1} \psi_{l-m,h-n} \Theta_{m,n} m \right) Z_1^{l-1} Z_2^h$$

as a formal expression, after recollecting terms. Then matching terms, we find a recursive relation between $\psi_{j,k}$ (where both j and k are positive) and $\Theta_{m,n}$, given below. Similar calculations for the

other associations (11), (12), and (13) yield

$$\psi_{j,k} = \frac{1}{j} \sum_{m=1}^p m \left(\sum_{n=1}^k \psi_{j-m,k-n} \Theta_{m,n} \right) \quad (15)$$

$$\phi_{j,k} = \frac{1}{j} \sum_{m=1}^p m \left(\sum_{n=1}^k \phi_{j-m,k-n} \Theta_{-m,n} \right) \quad (16)$$

$$\xi_j = \frac{1}{j} \sum_{m=1}^p m \Theta_{m,0} \xi_{j-m} \quad (17)$$

$$\omega_j = \frac{1}{j} \sum_{m=1}^p m \Theta_{0,m} \omega_{j-m}, \quad (18)$$

for $j \geq 1$ and $k \geq 1$. The initializations of the recursions for indices $j = 0$ or $k = 0$ are described above. These are recursive formulas. In the causal case, one would compute $\psi_{1,1}, \psi_{2,1}, \dots, \psi_{p,1}, \psi_{1,2}, \psi_{2,2}, \dots$, etc. Alternative computational patterns could be utilized, noting that $\psi_{j,k}$ only requires knowledge of $\psi_{\ell,h}$ with $\ell < j$ and $h < k$. The full algorithm for the acf is then:

1. Use the recursions (15), (16), (17), and (18) to obtain the various moving average random field coefficients, to the desired level of approximation.
2. Compute acfs for the causal, skew, and one-dimensional random fields, via (8) and (9), etc.
3. Compute the final acf via (14).

When $p = \infty$, this gives the precise mapping of cepstral coefficients to various MA coefficients, and ultimately to the autocovariance function. If $p < \infty$, it provides an algorithm for determining autocovariances for a given cepstral model. These formulas are already much more complicated than in the time series case (see Pourahmadi, 1984), and for higher dimensional fields become intractable. When p is small these exact recursions may be less burdensome; otherwise the approximate method based on discretizing (2) may be preferable.

3 Model Fitting Methods

In this section we give additional details on various methods for fitting cepstral random field models, and present some tools for refining specified models. Once a model is specified, we can estimate the parameters via exact maximum likelihood, Bayesian posterior simulation, an approximate Whittle likelihood, or an exact Whittle likelihood (described below). Whereas other methods of estimation can be conceived, we focus on these four techniques due to their mixture of being flexible and possessing good statistical properties. Most of our treatment holds for general stationary random fields, but when a result only holds for the cepstral random field we will make explicit mention of this.

The first two approaches require the exact Gaussian likelihood, whereas the third approach uses an approximation. Once model parameters have been estimated, in the frequentist paradigm we may be interested in refining our model, which we might approach by determining whether certain cepstral coefficients are significantly different from zero. If they are not, we can easily fix these coefficients to zero (it is trivial to do this sort of constrained estimation with the cepstral model) and rerun our estimation.

In this section, we first define Kullback-Leibler (KL) discrepancy, the exact Whittle likelihood, and the quasi-maximum likelihood estimate (QMLE). Then we proceed to describe Central Limit Theorems for the maximum likelihood estimates (MLEs) and QMLEs, which are novel results in the random field literature, under an expanding domain asymptotic theory. These results, proved for fairly general linear random fields with regression effects, are then specialized to the case of the cepstral field, and the backward deletion model refinement procedure is described.

While the MLE procedure is discussed in Cressie (1993) and other works, there are few precise mathematical results that provide Central Limit Theorems for estimates. In the prior literature, we observe that Mardia and Marshall (1984) do indeed establish efficiency of the MLE when the random field is Gaussian, under a condition on the trace of the covariance matrix. In fact, we provide (see Section 4) sufficient conditions under which our Lemma 1 yields the validity of condition (iii) of Theorem 2 of Mardia and Marshall (1984). Although Solo (1986) advocates the approximate QMLE method in practice, asymptotic results are not proved in that paper. Therefore we believe that the theoretical results of Section 4 are novel and useful, filling a gap in the spatial statistics literature and allowing a disciplined approach to model building and inference.

There is also an existing literature on Discrete FTs for random fields (see Bandyopadhyay and Lahiri, 2010) under various asymptotic schedules, including in-filling asymptotics, expanding (template-based) asymptotics, and mixtures of both. Whereas it would be intriguing to extend our theoretical results to these frameworks, we have not attempted to prove such results in this paper. For one thing, in-fill asymptotics require a continuous-parameter random field, and the cepstral approach requires non-trivial modifications from the lattice situation (the complex exponentials will not span the space of log spectra). For another, the results given here are sufficient to guide inference for regular lattice data, so that only exotic shapes and irregular sampling are excluded.

Now we proceed to discuss spatial modeling, adapting the vector time series treatment in Taniguchi and Kakizawa (2000). Again, we do not assume a cepstral formulation in what follows. Since we want to formulate asymptotic results for parameter estimates, we will speak of our data in somewhat more idealistic terms for the purposes of formulating precise mathematical statements. So we will suppose that our data comes to us in gridded form, corresponding to a $N \times N$ matrix \mathbb{Y}^N (with \mathbb{W}^N denoting the de-meaned version). This presumes a square domain, which however is only necessary for formulating the asymptotic results below. Rectangular domains can be handled by taking the smallest circumscribing square, and utilizing the asymptotic results for this super-

square. In other words, we might let N be the maximum of the length and width of the rectangle, and apply the asymptotic results (this is reasonable if the dimensions are roughly equal).

Now \mathbb{Y}^N and \mathbb{W}^N can be vectorized into length N^2 vectors Y and W via the so-called lexicographical rule

$$Y_k = \mathbb{Y}_{r,s}^N \quad W_k = \mathbb{W}_{r,s}^N, \quad k = N(r-1) + s.$$

Here, $Y = \text{vec}(\mathbb{Y}^{N'})$, where vec stands for the vector operation on a matrix. Note that $r-1 = k \text{ div } N$ and $s = k \text{ mod } N$. Also let $\mu = \mathbb{E}Y$, so that $\mu_k = \mathbb{E}Y_k = \mathbb{E}\mathbb{Y}_{r,s}^N = \mu_{r,s}$. In the simplest scenario the mean matrix $\{\mu_{r,s}\}$ is constant with respect to the indices r, s . More generally, we might model the mean through regressor functions defined upon the grid, i.e., $\mu_{r,s} = \sum_{\ell=1}^L \beta_\ell X_\ell(r, s)$ for some specified lattice functions $\{X_\ell\}_{\ell=1}^L$. Then

$$\mu_k = \sum_{\ell=1}^L \beta_\ell X_\ell(k \text{ div } N + 1, k \text{ mod } N) = \sum_{\ell=1}^L \beta_\ell \tilde{X}_\ell(k)$$

maps each X_ℓ from a lattice function to a function \tilde{X}_ℓ of the natural numbers. The parameters $\beta_1, \beta_2, \dots, \beta_L$ then enter the regression linearly, and we can express things compactly via $\mu = \tilde{X}\beta$, where \tilde{X} is the regression matrix with columns given by the various \tilde{X}_ℓ .

The spectral density of a mean zero random field \mathbb{W} has already been defined in (1), and the discrete FT of the field is now defined as

$$\tilde{\mathbb{W}}(\lambda_1, \lambda_2) = \sum_{t_1=1}^N \sum_{t_2=1}^N \mathbb{W}_{t_1, t_2}^N e^{-i\lambda_1 t_1} e^{-i\lambda_2 t_2} = \sum_{t_1=1}^N \sum_{t_2=1}^N W_{N(t_1-1)+t_2} Z_1^{t_1} Z_2^{t_2}$$

for $\lambda_1, \lambda_2 \in [-\pi, \pi]$. Note that we define this FT over all pairs of frequencies, not just at the so-called Fourier frequencies $\{\pi j/N\}$. Also the FT depends on β through the mean-centering; if we center the data \mathbb{Y}^N by any regression parameter other than the true β , some bias will be introduced. The periodogram will be defined at all frequencies, and is proportional to the squared magnitude of the FT:

$$I_\beta(\lambda_1, \lambda_2) = N^{-2} |\tilde{\mathbb{W}}(\lambda_1, \lambda_2)|^2 = \sum_{|h_1| < N, |h_2| < N} \gamma_{h_1, h_2}(I) Z_1^{h_1} Z_2^{h_2}.$$

Here $\gamma_{h_1, h_2}(I)$ is the sample acf. Note that our notation $\gamma_{h_1, h_2}(I)$ is consistent with (2), i.e., the sample autocovariance is the inverse FT of the periodogram I_β . We also will consider an unbiased acf estimate given by

$$\hat{\gamma}_{h_1, h_2} = \frac{N^2}{(N - |h_1|)(N - |h_2|)} \gamma_{h_1, h_2}(I).$$

We emphasize that the computation of this periodogram requires a choice of β , and so is written I_β . This can be used to assess the frequency domain information in the random field along any row or column; the periodogram can also be viewed as a crude estimate of the spectral density F (Cressie, 1993).

Model fitting can be performed and assessed through the Kullback-Leibler (KL) discrepancy, just as with time series. Although KL is mentioned in Solo (1986) and Cressie (1993), we provide an in-depth treatment here – see Lemma 2, for example. If F and G are two (mean zero) random field spectral densities, their KL discrepancy is defined to be

$$KL(F, G) = \langle \log F + G/F \rangle .$$

This is a convenient mechanism, since KL is convex in F . As β parametrizes mean effects, we let θ be a parameter vector describing the second order structure. If the true data process has spectrum \tilde{F} , and we utilize a model with spectrum F_θ , then $KL(F_\theta, \tilde{F})$ can be used to assess proximity of the model to truth. The convexity of KL guarantees that when the model is correctly specified, the true parameter $\tilde{\theta}$ minimizes the discrepancy. When the model is misspecified, the minima $\tilde{\theta}$ are called pseudo-true values (cf. Taniguchi and Kakizawa, 2000). For the cepstral model, the parameter vector is $\theta = \text{vec}[\Theta]$. The full parameter vector is written ϕ , where $\phi' = [\theta', \beta']$.

It is natural to use KL to fit models as well. For this, consider $KL(F_\theta, I_\beta)$ – which is called the exact Whittle likelihood – and minimize with respect to θ , which produces by definition the estimate $\hat{\theta}_{QMLE}$. Then using (2) we obtain the practical expression

$$KL(F_\theta, I_\beta) = \langle \log F_\theta \rangle + N^{-2} \sum_{t_1, t_2=1}^N \sum_{s_1, s_2=1}^N W_{N(t_1-1)+t_2} W_{N(s_1-1)+s_2} \gamma_{s_1-t_1, s_2-t_2}(F_\theta^{-1}).$$

This assumes that the correct regression parameters have been specified. In the case that F_θ is a cepstral spectrum (3), the above expression is even easier to compute: $\langle \log F_\theta \rangle = \Theta_{0,0}$ and $\gamma(F_\theta^{-1}) = \gamma(F_{-\theta})$, i.e., multiply each cepstral coefficient by -1 to obtain the acf of F_θ^{-1} from the acf of F_θ .

The KL discrepancy can also be written as

$$\langle \log F_\theta \rangle + \sum_{|h_1| < N, |h_2| < N} \gamma_{h_1, h_2}(I) \gamma_{h_1, h_2}(F_\theta^{-1}).$$

Unfortunately, $\gamma_{h_1, h_2}(I)$ is biased as an estimate of $\gamma_{h_1, h_2}(\tilde{F})$, and this has a nontrivial impact for spatial data, though not for time series. Essentially, the presence of “corners” in the observed data set reduces the number of data points that are separated by a given lag (h_1, h_2) ; if either of $|h_1|$ or $|h_2|$ is large, we have a very biased estimate. Note, the impact of corners can be visualized by comparing the volume of a d -dimensional cube with that of an inscribed ball; the ratio is $\pi/(2d)$ for $d \geq 2$, which tends to zero as d increases. Thus, corners increasingly dominate the region as d increases, which interferes with one’s ability to measure correlation as a function of lag. This effect is more pronounced as dimension increases. For this reason, we propose using $\hat{\gamma}_{h_1, h_2}$ instead of $\gamma_{h_1, h_2}(I)$, because $\mathbb{E}\hat{\gamma}_{h_1, h_2} = \gamma_{h_1, h_2}(\tilde{F})$. Let us call the modified $KL(F_\theta, I_\beta)$ by $\widehat{KL}(F_\theta)$:

$$\widehat{KL}(F_\theta) = \langle \log F_\theta \rangle + \sum_{|h_1| < N, |h_2| < N} \hat{\gamma}_{h_1, h_2} \gamma_{h_1, h_2}(F_\theta^{-1}).$$

Using this criterion instead will produce asymptotically normal cepstral parameter estimates, and therefore is to be preferred.

If even faster computation of the objective function is desired, we may discretize \widehat{KL} and utilize values of F directly, without having to compute the inverse FT $\gamma(F_\theta^{-1})$. The result is the approximate Whittle likelihood, denoted \widehat{KL}_N , and is obtained by discretizing the integrals in $KL(F_\theta, I_\beta)$ with a mesh corresponding to Fourier frequencies, but replacing I_β with the FT of the $\widehat{\gamma}_{h_1, h_2}$ sequence, denoted by \widehat{I}_β . Note that \widehat{I}_β may not be positive at all frequencies, because the unbiased autocovariance estimates are no longer guaranteed to be a positive definite sequence. This is the price of adjusting for bias. So the discrepancy is

$$\widehat{KL}_N(F_\theta) = N^{-2} \sum_{j_1=-N}^N \sum_{j_2=-N}^N \left\{ \log F_\theta(\pi j_1 N^{-1}, \pi j_2 N^{-1}) + \frac{\widehat{I}_\beta(\pi j_1 N^{-1}, \pi j_2 N^{-1})}{F_\theta(\pi j_1 N^{-1}, \pi j_2 N^{-1})} \right\}$$

which can be minimized with respect to θ . The resulting estimate has asymptotic properties identical to the QMLE, and in practice one may use either \widehat{KL} or \widehat{KL}_N according to computational convenience. It will be convenient to present a notation for this double discrete sum, which is a Fourier approximation to $\langle \cdot \rangle$, denoted by $\langle \cdot \rangle_N$; then $\widehat{KL}_N(F_\theta) = \langle \log F_\theta + \widehat{I}_\beta / F_\theta \rangle_N$.

On the other hand, we can also compute the exact Gaussian likelihood for the field. Let the covariance matrix of \mathbb{W}^N be denoted $\Sigma(\widetilde{F})$, which is defined via $\Sigma(\widetilde{F}) = \mathbb{E}WW'$; the resulting block-Toeplitz structure of this matrix is analyzed in Section 4. The entries of this matrix can be determined from \widetilde{F} via the algorithms of Section 2, along with careful bookkeeping. A model for the data involves a spectrum F_θ – let the associated block-Toeplitz covariance matrix be denoted $\Sigma(F_\theta)$ – which is hoped to be a suitable approximation to $\Sigma(\widetilde{F})$. Then the log Gaussian likelihood is equal (up to constants) to

$$\mathcal{L}(\theta, \beta) = -.5 \log |\Sigma(F_\theta)| - .5(Y - \widetilde{X}\beta)' \Sigma^{-1}(F_\theta)(Y - \widetilde{X}\beta). \quad (19)$$

Maximizing this function with respect to θ yields the MLE $\widehat{\theta}_{MLE}$; also $\widehat{\beta}_{MLE}$ is given by the Generalized Least Squares (GLS) estimate by standard arguments (see Mardia and Marshall, 1984):

$$\widehat{\beta}_{MLE} = \left[\widetilde{X}' \Sigma^{-1}(F_{\widehat{\theta}_{MLE}}) \widetilde{X} \right]^{-1} \widetilde{X}' \Sigma^{-1}(F_{\widehat{\theta}_{MLE}}) Y, \quad (20)$$

which expresses the regression parameter in terms of $\widehat{\theta}_{MLE}$. For the computation of (19) we must calculate the acf corresponding to F_θ , which can be done using the algorithms of Section 2.

Given these notations, we can also re-express the formula for KL more compactly:

$$KL(F_\theta, I_\beta) = \langle \log F_\theta \rangle + N^{-2} (Y - \widetilde{X}\beta)' \Sigma(F_{-\theta})(Y - \widetilde{X}\beta). \quad (21)$$

Contrast this expression with (19); they are similar, the main difference being the replacement of the inverse of $\Sigma(F_\theta)$ by $\Sigma(F_\theta^{-1}) = \Sigma(F_{-\theta})$. We propose using (21) to estimate regression parameters,

but θ is to be determined by \widehat{KL} . The formula for the regression QMLE is then

$$\widehat{\beta}_{QMLE} = \left[\widetilde{X}' \Sigma(F_{\widehat{\theta}_{QMLE}}^{-1}) \widetilde{X} \right]^{-1} \widetilde{X}' \Sigma(F_{\widehat{\theta}_{QMLE}}^{-1}) Y, \quad (22)$$

where $\widehat{\theta}_{QMLE}$ minimizes $\widehat{KL}(F_\theta)$, which in turn depends upon $\widehat{\beta}_{QMLE}$ through \widehat{I}_β . These formulas do not apply when we use the approximate Whittle, although the same asymptotic properties will hold as for the exact Whittle.

Most prior literature on random fields seems to utilize approximate Whittle estimation, or QMLE, since the objective function is quite simple to write down. However, there is some bias in the model parameters QMLEs when $\gamma_h(I)$ are used as acf estimates, which precludes a central limit theorem and thereby interferes with model-building; this is a dimension-dependent bias that is asymptotically irrelevant for time series, but pertinent for spatial data. On the other hand, the parameter MLEs do not have this problem, but require more effort to compute. We can use the approximate algorithm given by equation (6), together with (19), to compute the MLEs. The QMLEs based on unbiased $\widehat{\gamma}_h$ acf estimates are faster to compute than MLEs (no matrix inversions required), and enjoy the same asymptotic normality and efficiency.

However, if one prefers a Bayesian estimation of θ (and β), it is necessary to compute $\exp \mathcal{L}(\theta, \beta)$, which is proportional to the data likelihood $p(Y|\theta, \beta)$. The posterior for θ is proportional to the likelihood times the prior, and one can use Markov chain Monte Carlo (MCMC) methods to approximate $p(\theta|Y)$ (Geweke, 2005). Recall that the mean of this distribution, which is the conditional expectation of θ given Y , is called the posterior mean, and will be denoted $\widehat{\theta}_B$.

Note that equations (19) and (20) can each be used in an iterative estimation scheme. To determine the MLE, minimize (19) to obtain an estimate of θ for a given β computed via (20); then update $\widehat{\beta}$ by plugging into (20), and iterate. For the QMLE, demean the data by computing $Y - X\beta$ and determining \widehat{I}_β , for a given β , and then minimize either \widehat{KL} or \widehat{KL}_N to obtain θ estimates (either exact or approximate); then update β by plugging into (22) and iterate. From now on, we refer to these estimates as the exact/approximate QMLEs (if using biased acf estimates $\gamma_h(I)$, only consistency holds, and not asymptotic normality).

We can now provide a description of the asymptotics for the various estimates; a rigorous treatment is given in Section 4, with formal statements of sufficient conditions and auxiliary results. Firstly, the Bayesian estimates $\widehat{\theta}_B$ and $\widehat{\beta}_B$ are consistent when the data is a Gaussian random field that satisfies suitable regularity conditions (Theorem 2). Secondly, in the Frequentist case we can say a bit more about the asymptotic distribution. Recall that the full sample size is N^2 , so a Central Limit Theorem result requires scaling by $N = \sqrt{N^2}$. Let the Hessian of the KL be denoted $H(\theta) = \nabla \nabla' KL(F_\theta, \widetilde{F})$, which will be invertible at the unique pseudo-true value $\widetilde{\theta}$ by assumption. Then the exact QMLE, approximate QMLE, and MLE for θ are all consistent, and are also asymptotically normal at rate N with mean $\widetilde{\theta}$ and variance $H^{-1}(\widetilde{\theta}) V(\widetilde{\theta}) H^\dagger(\widetilde{\theta})$, where \dagger denotes inverse transpose and $V(\theta) = 2 \langle \widetilde{F}^2 \nabla F_\theta^{-1} \nabla' F_\theta^{-1} \rangle$. (This assumes that the fourth-order

cumulants are zero; otherwise a more complicated expression for V results, involving the fourth-order spectral density.) The estimates of the regression parameters are asymptotically normal and independent of the θ estimates (when the third-order cumulants are zero), for all three types of estimates.

These theoretical results can be used to refine models. Typically, one uses these types of asymptotic results under the null hypothesis that the model is correctly specified, so that $\tilde{\theta}$ is the true parameter and $V = 2 \langle \nabla \log F_{\theta} \nabla' \log F_{\theta} \rangle$, which equals twice H . See McElroy and Holan (2009) and McElroy and Findley (2010) for more exposition on model misspecification in the frequency domain. Thus, the asymptotic variance is twice the inverse Hessian, or the inverse of $H/2$. Note that the Fisher information matrix is the Hessian of the asymptotic form of the Whittle likelihood, and hence is equal to one half of the Hessian of KL, i.e., $H/2$. Therefore when the model is correctly specified, parameter estimation is efficient.

Furthermore, the Fisher information matrix has a particularly elegant form in the case of a cepstral model. The gradient of the log spectrum is in this case just the various Z_1^j or Z_2^k , so that as in the time series case the Hessian equals twice the identity matrix (because of mirror reflectional symmetry in Θ , there is a doubling that occurs), except for the case of the entry corresponding to $\Theta_{0,0}$ – in this case the derivative of the log spectrum with respect to $\Theta_{0,0}$ equals one. Thus the Fisher information matrix for all the parameters except $\Theta_{0,0}$ is equal to the identity matrix, and hence the asymptotic variance of any cepstral coefficient estimate is N^{-2} (or $2N^{-2}$ in the case of $\Theta_{0,0}$). The lack of cross-correlation in the parameter estimates asymptotically indicates there is no redundancy in the information they convey, which is a type of “maximal efficiency” in the cepstral model.

In terms of model-building with cepstral random fields, one procedure is the following: postulate a low order cepstral field model (e.g., order $p = 1$) and jointly test for whether any coefficients (estimated via MLE or QMLE) are zero. We might consider expanding the model – in the direction of one spatial axis or another as appropriate – if coefficients are significantly different from zero. This type of forward addition strategy would stop once all additional coefficients are negligible. Alternatively, one could start with a somewhat larger cepstral model, and iteratively delete insignificant coefficients.

Gaussian Likelihood Ratio test statistics can be utilized for nested cepstral models, along the lines given in Taniguchi and Kakizawa (2000) – which ultimately just depend on the asymptotic normality of the parameter estimates – in order to handle batches of parameters concurrently. Model selection and assessment can also be assisted by examination of spatial residuals, which are defined by applying the inverse square root of the estimated data covariance matrix $\Sigma(F_{\hat{\theta}})$ to the vectorized centered data W – the result is a vectorized residual sequence, which should behave like white noise if the model has extracted all correlation structure. Note that examining whiteness of the vectorized residuals is equivalent to looking at all spatial correlations of the spatial residuals

defined by undoing the vec operation. In the context of lattice data, one popular method for testing the null hypothesis of the absence of spatial autocorrelation is through the use of Moran's I statistic (Moran, 1950). See Cressie and Wikle (2011) for further details. In our case (Section 5.2), we will evaluate goodness-of-fit by applying Moran's I statistic to the spatial residuals obtained from the estimated model.

Extensions to missing data, imputation, and signal extraction are possible, as discussed below. Suppose that there is missing data in our sample of the field. We can always express this situation as $Z = JY$, where Z is now our observed data vector of dimension $N^2 - T$, where T is the number of missing observations in the field, and Y is the vectorized random field, as described above. J is a selection matrix with $N^2 - T$ rows and N^2 columns, which eliminates any entries of Y – corresponding to omitted observations in \mathbb{Y} – for which no measurements exists. Then the covariance matrix of Z is $J\Sigma(\tilde{F})J'$, which is easily computed from any given parameter specification. The quadratic form in our Gaussian likelihood is then $(Z - J\mu)'[J\Sigma(\tilde{F})J']^{-1}(Z - J\mu)$; a matrix of this form is always invertible. It is obtained by crossing out rows and corresponding columns of $\Sigma(\tilde{F})$, and contracting the result.

Signal extraction for random fields can be handled as follows. We suppose that our random field \mathbb{Y} consists additively of signal and noise, i.e.,

$$\mathbb{Y} = \mathbb{S} + \mathbb{N}.$$

The signal random field \mathbb{S} and noise random field \mathbb{N} may each be given by cepstral field models, but are assumed to be completely independent of one another. Each may also have their own regression mean effects. Let the cepstral coefficient grids of each be denoted $\Theta_{\mathbb{S}}$ and $\Theta_{\mathbb{N}}$. We can compute the acf for each component via the methods outlined in the previous section, obtaining $\gamma_{\mathbb{S}}$ and $\gamma_{\mathbb{N}}$. In turn we can determine the vectorized versions of signal and noise, denoted by S and N , with covariance matrices Σ_S and Σ_N . With $Y = S + N$ equal to the vectorized data, the quadratic form in the Gaussian likelihood is $(Y - \mu)'(\Sigma_S + \Sigma_N)^{-1}(Y - \mu)$. So this is easy to calculate given the parameters $\Theta_{\mathbb{S}}$ and $\Theta_{\mathbb{N}}$. The distribution of the signal conditional on the data is Gaussian with mean and variance given by

$$\mathbb{E}[S] + \Sigma_S(\Sigma_S + \Sigma_N)^{-1}(Y - \mu) \quad \text{and} \quad \Sigma_S - \Sigma_S(\Sigma_S + \Sigma_N)^{-1}\Sigma_S.$$

Of course, the conditional mean is for the vectorized field S ; in order to get our estimate of \mathbb{S} , we reverse the vectorization mapping. Also, if there is missing data and $Z = JY$ is our observation vector, the likelihood uses the quadratic form $(Z - J\mu)'(J\Sigma_S J' + J\Sigma_N J')^{-1}(Z - J\mu)$, and the conditional mean will now be

$$J\mathbb{E}[S] + \Sigma_S J'(J\Sigma_S J' + J\Sigma_N J')^{-1}(Z - J\mu).$$

The mean $\mathbb{E}[S]$ is determined in practice by assigning some of the fixed effects in \tilde{X} to the signal. For example, if the signal is defined to be a smooth “trend” effect and the noise is idiosyncratic noise, then all mean effects might be assigned to the signal so that $\mathbb{E}[S] = \mu$.

4 Theory of Inference

This section provides rigorous mathematical results regarding the inference problems delineated in Section 3. We do not assume a cepstral random field process, retaining greater generality, but assume a fair amount of regularity on the higher moments of the field through the Brillinger-type cumulant conditions (Brillinger, 2001).

We begin with an important Lemma that extends Lemma 4.1.2 of Taniguchi and Kakizawa (2000) to the spatial context. First we define a block-Toeplitz matrix $\Sigma(F)$ associated with spectral density F to be $N^2 \times N^2$ -dimensional with jk th block given by the $N \times N$ -dimensional matrix $\Sigma(F_{j-k})$, which is defined as follows. If we integrate over the second variable of F we obtain a function of the first frequency:

$$F_h(\lambda_1) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\lambda_1, \lambda_2) e^{ih\lambda_2} d\lambda_2.$$

Then $\Sigma(F_h)$ is the matrix of inverse FTs of F_h , namely with jk th entry given by

$$\gamma_{h,j-k}(F) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} F(\lambda_1, \lambda_2) e^{ih\lambda_2} d\lambda_2 e^{i(j-k)\lambda_1} d\lambda_1 = \langle F Z_1^{k-j} Z_2^{-h} \rangle.$$

Based on how we have defined \mathbb{W}^N and W , it follows that $\Sigma(\tilde{F}) = \mathbb{E}[W W']$, where \tilde{F} corresponds to the Data Generating Process (DGP). That is, lexicographical ordering of a stationary field produces this structure in the covariance matrix. Also let \mathcal{F} denote the set of admissible spectra for two-dimensional random fields, defined as follows. For any spatial autocovariance function $\{\gamma_{h,k}\}$, consider the sums $S_h^+ = \sum_k |k| |\gamma_{h,k}|$, $S_k^- = \sum_h |h| |\gamma_{h,k}|$, and $S^* = \sum_{h,k} |k| |h| |\gamma_{h,k}|$. Then

$$\mathcal{F} = \{F : [-\pi, \pi]^2 \rightarrow \mathbb{R}^+, F(\lambda_1, \lambda_2) = \sum_{h,k} \gamma_{h,k} Z_1^h Z_2^k, S_h^+ < \infty \forall h, S_k^- < \infty \forall k, S^* < \infty\}.$$

Note that this class excludes spectra with zeroes, which is a minor imposition in practice.

Lemma 1 *Let $\Sigma(F)$ and $\Sigma(G)$ be block-Toeplitz matrices with $F, G \in \mathcal{F}$. Then*

$$N^{-2} \text{tr}\{\Pi_{(F,G) \in \mathcal{F} \times \mathcal{F}} \Sigma(F) \Sigma^{-1}(G)\} = \langle \Pi_{(F,G) \in \mathcal{F} \times \mathcal{F}} F G^{-1} \rangle + O(N^{-2}).$$

Next, we discuss a Lemma that provides a Central Limit Theorem for weighted averages of the spatial periodogram, which is a natural extension of Lemma 3.1.1 of Taniguchi and Kakizawa (2000). We take Brillinger’s approach, stipulating summability conditions on higher order cumulants of the spatial field. Let us denote an integer-valued bi-variate index by $t \in \mathbb{Z}^2$, which has integer

coordinates (t_1, t_2) . Then a collection of spatial variables can be written $\{\mathbb{W}_{t^{(1)}}, \mathbb{W}_{t^{(2)}}, \dots\}$. The strict stationarity condition stipulates that joint distributions of such variables only depend upon differences between indices: $t^{(1)} - t^{(2)} = (t_1^{(1)} - t_1^{(2)}, t_2^{(1)} - t_2^{(2)})$, etc. If we sum a function with respect to $t \in \mathbb{Z}^2$, the notation refers to a double sum over t_1 and t_2 . A similar notation is used for frequencies $\lambda \in [-\pi, \pi]^2$, in that $\lambda = (\lambda_1, \lambda_2)$.

The next result presumes that spatial data is sampled from a true spatial field with spectrum \widetilde{F} , and that we have a collection of continuous weighting functions $G_j : [-\pi, \pi]^2 \mapsto \mathbb{R}^+$. The second-order cumulant function of the spatial field is the autocovariance function γ_h with $h \in \mathbb{Z}^2$, whereas the fourth-order cumulant function is denoted

$$\gamma_{h,k,\ell} = \text{cum} [\mathbb{W}_t, \mathbb{W}_{t+h}, \mathbb{W}_{t+k}, \mathbb{W}_{t+\ell}].$$

We require absolute summability of these second and fourth order cumulant functions. Then the fourth order spectrum is well-defined via

$$\widetilde{FF}(\lambda, \omega, \xi) = \sum_{h,k,\ell} \gamma_{h,k,\ell} \exp\{-i\lambda \cdot h - i\omega \cdot k - i\xi \cdot \ell\},$$

with \cdot denoting the dot product of bi-variate vectors. In order to establish the Lemma, we impose more regularity via

$$\sum_{h^{(1)}, h^{(2)}, \dots, h^{(k)}} \left(1 + \overline{|h^{(1)}|} \overline{|h^{(2)}|} \dots \overline{|h^{(k)}|}\right) |\gamma_{h^{(1)}, h^{(2)}, \dots, h^{(k)}}| < \infty, \quad (23)$$

where \bar{t} denotes the product of the components of t . This will be referred to as Condition B_k , for any $k \geq 1$. Finally, recall that the periodogram is computed from a sample of size N^2 . When the regressors are correctly specified, we will write $\widetilde{\beta}$ for the true parameter. Then $I_{\widetilde{\beta}}$ denotes the periodogram of the data Y correctly adjusted for mean effects; equivalently, it is the periodogram of \mathbb{W}^N .

Lemma 2 *Suppose that Condition B_k (23) holds for all $k \geq 1$, and that G_j for $j = 1, 2, \dots, J$ are continuous functions. Then both $\langle G_j \widehat{I}_{\widetilde{\beta}} \rangle$ and $\langle G_j \widehat{I}_{\widetilde{\beta}} \rangle_N$ have the same asymptotic behavior, and both converge in probability to $\langle G_j \widetilde{F} \rangle$. Moreover,*

$$N \{ \langle G_j (\widehat{I}_{\widetilde{\beta}} - \widetilde{F}) \rangle \}_{j=1}^J \xrightarrow{\mathcal{L}} \mathcal{N}(0, V)$$

as $N \rightarrow \infty$. The covariance matrix V has jk th entry

$$\frac{1}{(2\pi)^4} \int_{[-\pi, \pi]^4} G_j(\lambda) G_k(\omega) \widetilde{FF}(\lambda, -\omega, \omega) d\lambda d\omega + \langle (G_j G_k(-\cdot) + G_j G_k) \widetilde{F}^2 \rangle.$$

Also, $\langle GI_{\widetilde{\beta}} \rangle$ and $\langle GI_{\widetilde{\beta}} \rangle_N$ both converge in probability to $\langle G\widetilde{F} \rangle$, but $N \langle G_j(\mathbb{E}I_{\beta} - \widetilde{F}) \rangle$ does not converge to zero in probability.

The last assertion of Lemma 2 means that the asymptotic normality result for $\langle G_j(I_{\tilde{\beta}} - \tilde{F}) \rangle$ cannot be established; as pointed out in Guyon (1982), there remains a degree of asymptotic bias that precludes the central limit theorem from holding. However, $\langle G_j(\hat{I}_{\tilde{\beta}} - \tilde{F}) \rangle$ does not suffer from this bias problem. Both lemmas are important preliminary results for our main theorems, but also are of interest in their own right, extending known time series results to the spatial context. Although generalizations to dimensions higher than two seem feasible, the actual mechanics become considerably more tricky. Now the key assumptions that we require for our stationary random field are the following:

A1: $\tilde{F} \in \mathcal{F}$.

A2: The spectral density F_θ is twice continuously differentiable and uniformly bounded above and away from zero, and moreover all components of $F_\theta, \nabla F_\theta, \nabla \nabla' F_\theta$ are in \mathcal{F} .

A3: The Brillinger conditions B_k – given by equation (23) – hold for all $k \geq 1$.

A4: The pseudo-true value $\tilde{\theta}$ exists uniquely in the interior of the parameter space.

A5: $H(\theta) = \nabla \nabla' KL(F_\theta, \tilde{F})$ is invertible at $\tilde{\theta}$.

Conditions A1, A3, and A5 cannot be verified from data, but some assumptions of this nature must be made to obtain asymptotic formulas. Condition A2 will hold for cepstral models (and other random field models as well) by the following argument. The coefficients of the causal and skew fields will have exponential decay in either index argument, by extensions of the classical time series argument (see for example Hurvich (2002)) applied to (15) and (16). (The time series argument can be directly applied to (17) and (18) as well.) Combining these results using (14), the acf of the cepstral field will also have exponential decay so that $F_\theta \in \mathcal{F}$. Of course, another way to verify this condition is to examine the boundedness of partial derivatives of the spectrum; at once we see that A2 holds for the cepstral model.

Although condition A4 may be problematic for certain moving average models (which may have complicated constraints on coefficients), the cepstral model uses no constraints on the entries of Θ , i.e., any entry of the matrix can be any real number, independently of all other entries. Euclidean space is open, so any pseudo-true value is necessarily contained in the interior. Also, existence of a pseudo-true value is guaranteed by convexity of the KL discrepancy. We now state the limit theorems for our parameter estimates. For the QMLE estimates, we suppose that they are either exact or approximate Whittle estimates defined using the unbiased acf estimates.

Theorem 1 *Assume that conditions A1 through A5 hold, and that the third and fourth order order cumulants of the field \mathbb{W} are zero. Assume the regressors are correctly specified – where $\tilde{\beta}$ the true regression parameter vector – and the regression matrix satisfies $(\tilde{X}'\tilde{X})^{-1} \rightarrow 0$, the zero matrix, as $N \rightarrow \infty$. Then in the case of MLE or the QMLE, both $\hat{\beta}$ and $\hat{\theta}$ are jointly asymptotically normal*

and independent, with distributions given by

$$\begin{aligned}
N(\widehat{\beta} - \widetilde{\beta}) &\stackrel{\mathcal{L}}{\Rightarrow} \mathcal{N}\left(0, M_X^{-1}(\widetilde{\theta}) \left[\widetilde{X}' \Sigma(F_{-\widetilde{\theta}}) \Sigma(\widetilde{F}) \Sigma(F_{-\widetilde{\theta}}) \widetilde{X} \right] M_X^\dagger(\widetilde{\theta})\right) \\
N(\widehat{\theta} - \widetilde{\theta}) &\stackrel{\mathcal{L}}{\Rightarrow} \mathcal{N}\left(0, H^{-1}(\widetilde{\theta}) V(\widetilde{\theta}) H^\dagger(\widetilde{\theta})\right) \\
M_X(\theta) &= \left[\widetilde{X}' \Sigma(F_{-\theta}) \widetilde{X} \right] \\
H(\theta) &= \nabla \nabla' KL(F_\theta, \widetilde{F}) \\
V(\theta) &= 2 \langle \widetilde{F}^2 \nabla F_\theta^{-1} \nabla' F_\theta^{-1} \rangle,
\end{aligned}$$

where \dagger denotes an inverse transpose.

Remark 1 The condition on the fourth order cumulants can be relaxed, in which case the expression for V is more complicated, involving the fourth order spectrum $\widetilde{F}\widetilde{F}$. We state our theorem with the simpler result, as this form tends to be used in practice in the time series case. If the third order cumulants are nonzero, there can be asymptotic correlation between $\widehat{\beta}$ and $\widehat{\theta}$.

Remark 2 Application of the same techniques in the case of a one-dimensional random field, or time series, yields asymptotic normality of regression and time series parameters under Brillinger's conditions. To our knowledge, the only other results of this flavor for time series with regression effects is the work of Pierce (1971), which focuses on ARIMA models but allows for skewed non-Gaussian distributions.

For the result on Bayesian estimation, we will assume that the model is correctly specified, which is fairly typical in the literature on Bayesian consistency. As the data likelihood is Gaussian, it is consistent to assume that the data is also Gaussian. The model must be identifiable, i.e., $F_{\theta_1} = F_{\theta_2}$ implies $\theta_1 = \theta_2$, which helps ensure asymptotic concentration of the likelihood. We assume the parameters belong to some compact subset of Euclidean space, and the true parameter vector lies in the interior. This assumption can often be accomplished by prior transformation (and is easily accomplished for the cepstral coefficients in the cepstral model).

Theorem 2 *Assume that the data process is Gaussian and satisfies A2, and the model is correctly specified and is identifiable. Suppose that the true parameters lie in the interior of the compact parameter space. Also suppose that the regressors are correctly specified – where $\widetilde{\beta}$ is the true regression parameter vector – and the regression matrix satisfies $N^{-2} \widetilde{X}' \Sigma^{-1}(F_\theta) \widetilde{X} \rightarrow M(\theta)$, a positive definite matrix that has two-norm uniformly bounded in θ away from zero and infinity. Then $\widehat{\beta}_B \xrightarrow{P} \widetilde{\beta}$ and $\widehat{\theta}_B \xrightarrow{P} \widetilde{\theta}$.*

It is worth comparing the conditions of the two theorems. In Theorem 2 the assumption of a correct model makes A1 automatic, and the Gaussian assumption makes A3 automatic. Furthermore, the assumption in Theorem 2 on the parameters – together with the assumption of a correct

model – automatically entails A4 as well. Theorem 1 also assumes A5, which is chiefly needed to establish asymptotic normality of the Frequentist estimates. The Bayesian result requires a slightly stronger assumption on the regression matrix in order to get asymptotic concentration of the likelihood. For example, if we seek to estimate a constant mean by taking \tilde{X} to be a column vector of all ones, then $M(\theta)$ exists and is just the scalar $F_{-\theta}(0, 0)$; this will be bounded away from zero and infinity in the cepstral model if all the cepstral coefficients are restricted to a range of values.

5 Applications

5.1 Simulation Study

To demonstrate the effectiveness of our approach, we conducted a small simulation study using maximum likelihood estimation as outlined in Sections 2 and 3. The model autocovariances were calculated according (6), with $M = 1000$ and $p = 2$. The exact parameter values for the simulation were calibrated to the straw yield data analysis presented in Section 5.2.

For this simulation, we generated 200 datasets with parameters $\theta = \text{vec}[\Theta]$ and $\beta = (\beta_0, \beta_1, \beta_2)'$, as defined in Table 1. In this case, β_1 and β_2 correspond to “row” and “column” effects, respectively, in the agricultural experiment considered. Here, the row and column effects are obtained by regressing the vectorized response on the corresponding row and column indices (since rows and columns are equally spaced). The X matrix used in this simulation consisted of a column of ones followed by columns associated with the row and column effects and was taken from the analysis presented in Section 5.2. Analogous to (19), the following log Gaussian likelihood (up to constants) was numerically maximized for each simulated dataset using the *optim* function in R (R Development Core Team, 2012)

$$\mathcal{L}(\theta, \beta) = -.5 \log |\Sigma(F_\theta)| - .5(Y - X\beta)' \Sigma^{-1}(F_\theta)(Y - X\beta).$$

As demonstrated in Table 1, the model parameters can be estimated with a high degree of precision. In most cases the mean maximum likelihood estimates across the 200 simulations were remarkably close to the true value. This can be quantified through an assessment of the mean squared error (MSE) across the 200 simulations, where we see that the MSE is extremely small - especially in the case of the mean parameters, β .

5.2 Straw Yield Data

To illustrate the utility of our approach, we analyzed the Mercer and Hall straw yield dataset (Mercer and Hall, 1911) considered by Solo (1986), among others. This dataset can be obtained through the R contributed package *agridat* (R Development Core Team, 2012). One important distinction between our analysis and the analysis of Solo (1986) is that the lattice in Solo (1986)

is taken as 19×25 , whereas, similar to Cressie (1993) and the dataset as provided in the R contributed package *agridat*, our lattice is 20×25 . The reason for the smaller lattice in Solo (1986) remains unclear. Also, similar to Cressie (1993), we considered the data as aggregations over 3.30×2.51 meter plots. The observed data are displayed in Figure 1 and comprehensive details of the experiment, as it relates to yields of wheat grain, can be found in Cressie (1993).

Exploratory analyses consisting of estimating linear models with independent errors revealed potentially significant “row” and “column” effects. Unlike the model formulation of Solo (1986), these trend effects could be readily accounted for within our modeling framework and their statistical significance evaluated. Additionally, further exploratory analysis using the empirical variogram seemed to indicate a negligible nugget effect (measurement error). Therefore, we considered an analysis with no nugget effect and note that if one had an empirical estimate of this effect, it could be easily incorporated into our model. An initial assessment of the spatial correlation was assessed using Moran’s I statistic with a nearest neighbor weighting matrix. The p -value for testing the null hypothesis of the absence of spatial autocorrelation was 2.2×10^{-16} , indicating a high-degree of spatial autocorrelation.

We considered several models in our analysis, including models with trend (i.e., row and column effects) removed and not removed. To remove the trend, we regressed the vectorized responses on a design matrix, X , consisting of a column of ones followed by columns containing the associated row and column indices, respectively. In the case where no trend was removed, we included a constant mean. Lastly, we considered models in which no trend was removed and no constant-mean was included. Not surprisingly, the latter model performed poorly relative to the other models considered and, therefore, is not discussed further.

Similar to the simulation study in Section 5.1, the model autocovariances were calculated according to (6), with $M = 1000$. Here, we evaluated models with $p = 1, 2$, and 3 . Due to the mirror reflectional symmetry, the models with $p = 1$ had 6 parameters when no trend was removed and 8 parameters when the trend was accounted for. Similarly, for $p = 2$ and $p = 3$ the no-trend model had 14 and 26 parameters, while the model accounting for the trend had 16 and 28 parameters, respectively.

Model selection was conducted using Akaike information criterion (AIC), Bayesian information criterion (BIC), and the Hannan-Quinn information criterion (HQ), defined as $2k - 2 \log(L)$, $k \log(n) - 2 \log(L)$, and $2k \log\{\log(n)\} - 2 \log(L)$, respectively, where k , $n = 500$, and L denote the number of model parameters, the sample size, and the likelihood function. For the trend and no-trend models, the values of $-\log(L)$, AIC, BIC, and HQ are given in Table 2. Using the information criteria, it appeared that a model with $p = 2$ or $p = 3$, with the trend accounted for, was preferred. For reasons of model parsimony and because AIC is known to over-fit, we chose the $p = 2$ model, with the trend accounted for.

Table 3 provides the estimated model parameters along with their standard errors (as deter-

mined by the square root of the diagonal elements of the inverse observed Hessian matrix). Judging by the 95% confidence intervals for the estimated parameters (e.g., $\hat{\theta} \pm 2 \times \text{se}(\hat{\theta})$ and $\hat{\beta} \pm 2 \times \text{se}(\hat{\beta})$) the trend parameters, β , are clearly statistically significant. Applying Moran’s I statistic with a nearest neighbor weighting matrix to the residuals (i.e., $Z = \Sigma(F_{\hat{\theta}})^{-1/2}(Y - \mu)$) yielded a p -value of 0.999, which indicates that the spatial autocorrelation has been removed.

It should be noted that fitting a model without accounting for trend (i.e., constant mean only; $p = 2$) also removes the spatial autocorrelation (Moran’s I statistic has p -value=1). However, since the row and column effects are clearly statistically significant and the model having the trend accounted for is preferred according to the information criteria, we ultimately selected the $p = 2$ trend model. It is also of worth noting that the mean parameters, as specified here, do not suffer from the “spatial confounding” problem recently described in the spatial statistics literature; e.g., see Hodges and Reich (2010), Paciorek (2010), and the references therein. This lack of spatial confounding is clearly illustrated in Section 5.1, where we can accurately estimate the true trend parameters. Finally, the approach proposed herein provides a natural framework for inclusion of trend parameters within the model, which is not the case with the models proposed in Solo (1986). Even in the constant mean case, the approach of Solo (1986) would need to estimate the mean nonparametrically (e.g., using the sample mean) and remove it prior to model estimation.

6 Conclusion

The general modeling approach and asymptotic theory we propose extends the spatial random field literature in several directions. By providing recursive formulas for calculating autocovariances, from a given cepstral random field model, we have facilitated usage of these models in a both Bayesian and likelihood settings. This is extremely notable as many models suffer from a constrained parameter space, whereas the cepstral random field model imposes no constraints on the parameter values. More specifically, the autocovariance matrix obtained from our approach is guaranteed to be positive definite.

Further, we provide extensions to accommodate missing data and to conduct signal extraction. In doing so we provide a comprehensive approach for modeling spatial lattice data. Also, we provide several approximate methods for speeding up computation. Importantly, this makes our approach increasingly practicable in high-dimensional settings.

In addition, we establish results on consistency and asymptotic normality. This provides a rigorous platform for conducting model selection and statistical inference. The asymptotic results are proven generally and can be viewed as an independent contribution to the random field literature, expanding on the results of Mardia and Marshall (1984) and others, such as Guyon (1982). It also expands known asymptotic results for time series estimated with regression effects, as alluded to in Remark 2.

Owing to the computational advantages and flexibility of the cepstral random field model, coupled with the asymptotic results presented here, we believe the model provides a natural alternative to those currently used in practice. The primary contribution of this paper resides in firmly establishing the theoretical aspects of this model. Through simulation, we have demonstrated the effectiveness and practicality of our approach. Further, we illustrate the performance of our model through an application to straw yield data from an agricultural field experiment. In this setting, it is readily seen that our model is easily able to characterize the underlying spatial dependence structure. A topic of future research is to apply the models across an expanded and diverse set of applications. These studies are expected to further demonstrate the utility and flexibility of the cepstral random field model.

Acknowledgments

This research was partially supported by the U.S. National Science Foundation (NSF) and the U.S. Census Bureau under NSF grant SES-1132031, funded through the NSF-Census Research Network (NCRN) program and by a University of Missouri Research Board grant.

Appendix

Derivation of (5). $\left\{ \overline{E}'[\Theta]E \right\}_{rs}$ equals

$$\begin{aligned} &= \sum_{j,k=1}^{2p+1} \exp \left\{ -\pi i \frac{(p+1-j)(M+1-r)}{M} \right\} \Theta_{k-p-1,p+1-j} \exp \left\{ \pi i \frac{(p+1-k)(M+1-s)}{M} \right\} \\ &= \sum_{\ell,h=-p}^p \Theta_{\ell,h} \left[\exp \left\{ -\pi i \frac{(M+1-r)}{M} \right\} \right]^h \left[\exp \left\{ -\pi i \frac{(M+1-s)}{M} \right\} \right]^\ell, \end{aligned}$$

which produces the stated result using (3).

Proof of Lemma 1. For notational convenience in the proofs, we will consider the frequencies in the interval $[0, 2\pi]$ rather than $[-\pi, \pi]$. So let the Fourier frequencies be denoted $\lambda_{s,N} = 2\pi s N^{-1}$, and define $\overline{\gamma}_{h,k}(F) = N^{-2} \sum_{s=1}^N \sum_{t=1}^N F(\lambda_{s,N}, \lambda_{t,N}) \exp\{ih\lambda_{s,N} + ik\lambda_{t,N}\}$, which is a discrete approximation to the relation $\gamma_{h,k}(F) = \langle F \exp\{ih \cdot + ik \cdot\} \rangle$. Fixing h , we can consider $\overline{\Sigma}(F_h)$ to be a $N \times N$ matrix with jk th entry $\overline{\gamma}_{h,j-k}(F)$. The notation suggests that this is an approximation to $\Sigma(F_h)$ defined above. By univariate results (Taniguchi and Kakizawa, 2000, p.491) we can write

$$\overline{\Sigma}(F_h) = U_N^* D_N(F_h) U_N \tag{A.1}$$

with $[U_N]_{jk} = N^{-1/2} \exp\{ij\lambda_{k,N}\}$ and $D_N(F_h)$ defined to be a diagonal matrix with ℓ, ℓ th entry given by $N^{-1} \sum_{s=1}^N F(\lambda_{s,N}, \lambda_{\ell,N}) \exp\{ih\lambda_{s,N}\}$. The $*$ denotes the complex conjugate transpose.

Next, we construct $\overline{\Sigma}(F)$ as an approximation to $\Sigma(F)$, consisting of Toeplitz blocks $\overline{\Sigma}(F_{j-k})$ in the jk th slot. Since the U_N matrices in (A.1) are common to each $\overline{\Sigma}(F_h)$, we obtain

$$\overline{\Sigma}(F) = \text{diag}(U_N^*) [D_N(F_{j-k})]_{j,k=1}^N \text{diag}(U_N).$$

Here $[D_N(F_{j-k})]_{j,k=1}^N$ denotes a block-Toeplitz matrix consisting of the diagonal matrices $D_N(F_{j-k})$ in each jk th block. This block-Toeplitz matrix (with diagonal blocks) can be reformulated as a block-diagonal matrix (with Toeplitz blocks) by application of a permutation matrix P . To formalize this, let P be a $N^2 \times N^2$ dimensional permutation matrix that for each $1 \leq k \leq N$ takes the k th row of each block matrix and stacks them in order in the k th “mega-row”, i.e., rows $(k-1)N+1$ through kN . Then $P[D_N(F_{j-k})]_{j,k=1}^N P'$ is equal to a block-diagonal matrix consisting of Toeplitz matrices $[D_N(F_{j-k})]_{\ell\ell}$ in the ℓ th block. We denote these matrices by $\Sigma[F(\lambda_{\ell,N})]$, which have jk th entry $N^{-1} \sum_{s=1}^N F(\lambda_{s,N}, \lambda_{\ell,N}) \exp\{i(j-k)\lambda_{s,N}\}$. It then follows that

$$\overline{F} = \text{diag}(U_N^*) P' \text{diag}\{\text{diag}\{F(\lambda_{j,N}, \lambda_{\ell,N})\}_{j=1}^N\}_{\ell=1}^N P \text{diag}(U_N).$$

Now the middle matrix is a doubly-diagonal matrix consisting of all the values of F at Fourier frequencies, in lexicographical order. Let us abbreviate this matrix by $D_N(F)$. Then this approximation $\overline{\Sigma}(F)$ is particularly convenient, since

$$\overline{\Sigma}(F)\overline{\Sigma}^{-1}(G) = \text{diag}(U_N^*) P' D_N(F/G) P \text{diag}(U_N) = \overline{\Sigma}(F/G)$$

for bounded spectra. At once we find that the trace of this expression is

$$\sum_{s=1}^N \sum_{t=1}^N F(\lambda_{s,N}, \lambda_{t,N})/G(\lambda_{s,N}, \lambda_{t,N}),$$

which when divided by N^2 will converge to $\langle F/G \rangle$. The extension to multiple F s and G s uses similar arguments. So the proof of the lemma depends on establishing that we can swap $\overline{\Sigma}(F)$ for $\Sigma(F)$.

Consider the base case of the trace of $\Sigma(F)\Sigma^{-1}(G)$, since the same techniques can be generalized to multiple terms. Then

$$\begin{aligned} & N^{-2} \text{tr} \{\Sigma(F)\Sigma^{-1}(G)\} - N^{-2} \text{tr} \{\overline{\Sigma}(F)\overline{\Sigma}^{-1}(G)\} \\ &= N^{-2} \text{tr} \{[\Sigma(F) - \overline{\Sigma}(F)] \Sigma^{-1}(G)\} + N^{-2} \text{tr} \{\overline{\Sigma}(F)\Sigma^{-1}(G) [\overline{\Sigma}(G) - \Sigma(G)] \overline{\Sigma}^{-1}(G)\}. \end{aligned}$$

Note that by Lemma A.1 of Dahlhaus and Wefelmeyer (1996) $|\text{tr}(AB)| \leq |A||B|$ where $|\cdot|$ denotes the Frobenius norm. Also we use the fact that the Frobenius norm of a product can be broken down in terms of 2-norms $\|\cdot\|$, so that

$$|\text{tr} \{\overline{\Sigma}(F)\Sigma^{-1}(G) [\overline{\Sigma}(G) - \Sigma(G)] \overline{\Sigma}^{-1}(G)\}| \leq \|\overline{\Sigma}(F)\Sigma^{-1}(G)\| \cdot |\overline{\Sigma}(G) - \Sigma(G)| \cdot \|\overline{\Sigma}^{-1}(G)\|.$$

With these results, it will suffice to show the boundedness of

$$|\Sigma(F) - \bar{\Sigma}(F)|^2 = \sum_{j,k,h} |\gamma_{h,j-k} - \bar{\gamma}_{h,j-k}|^2,$$

or equivalently that $\sum_{j,k,h} |\gamma_{h,j-k} - \bar{\gamma}_{h,j-k}|$ is bounded. The analysis requires a generalization of the Wahba (1968) result from time series to spatial random fields. First note that $\bar{\gamma}_{h,k} = \sum_{m,\ell} \gamma_{h+mN,k+\ell N}$ by properties of Fourier frequencies, and thus

$$\gamma_{h,j-k} - \bar{\gamma}_{h,j-k} = \sum_{m,\ell \neq (0,0)} \gamma_{h+mN,j-k+\ell N}.$$

Then the sum of the absolute difference can be bounded as follows:

$$\begin{aligned} & \sum_{j,k,h} |\gamma_{h,j-k} - \bar{\gamma}_{h,j-k}| \\ & \leq \sum_{j,k,h} \left| \sum_{m,\ell \neq (0,0)} \gamma_{h+mN,j-k+\ell N} \right| \\ & \leq \sum_{|i| < N} (N - |i|) \sum_{|h| < N} \sum_{m,\ell \neq (0,0)} |\gamma_{h+mN,i+\ell N}| \\ & = \sum_{|m|=1}^{\infty} \sum_{\ell} \sum_{|h| < N} \sum_{|i| < N} (N - |i|) |\gamma_{h+mN,i+\ell N}| + \sum_{|l|=1}^{\infty} \sum_{|h| < N} \sum_{|i| < N} (N - |i|) |\gamma_{h,i+\ell N}| \\ & = 2 \sum_{|m|=2}^{\infty} \sum_{\ell} \sum_{|h| < N} \sum_{|i| < N} (N - |i|) |\gamma_{h+mN,i+\ell N}| + \sum_{\ell} \sum_{|h| < N} \sum_{|i| < N} (N - |i|) (|\gamma_{h+N,i+\ell N}| + |\gamma_{h-N,i+\ell N}|) \\ & + 2 \sum_{|l|=2}^{\infty} \sum_{|h| < N} \sum_{|i| < N} (N - |i|) |\gamma_{h,i+\ell N}| + \sum_{|h| < N} \sum_{|i| < N} (N - |i|) (|\gamma_{h,i+N}| + |\gamma_{h,i-N}|) \\ & \leq 2 \sum_{|m| > N} \sum_{\ell} \sum_{|i| < N} (N - |i|) |\gamma_{m,i+\ell N}| + 2 \sum_{h,\ell} \sum_{|i| < N} (N - |i|) |\gamma_{h,i+\ell N}| \\ & + 2N \sum_{|l| > N} \sum_{|h| < N} |\gamma_{h,\ell}| + 2 \sum_h \sum_{|i| < N} |i| |\gamma_{h,i}| \\ & \leq 2C \sum_{|m| > N} \sum_{|i| < N} |i| |\gamma_{m,i}| + 2N \sum_{|m| > N} \sum_{|l| > N} |\gamma_{m,\ell}| \\ & + 2C \sum_h \sum_{|i| < N} |i| |\gamma_{h,i}| + 2N \sum_h \sum_{|l| > N} |\gamma_{h,\ell}| \\ & + 2N \sum_{|l| > N} \sum_{|h| > N} |\gamma_{h,\ell}| + 2 \sum_h \sum_{|i| < N} |i| |\gamma_{h,i}|, \end{aligned}$$

where $C > 0$ denotes a generic constant. At this point we utilize our definition of \mathcal{F} to conclude that the entire sum is bounded in N . As a result, the trace approximation has error of order N^{-2} , which is also the approximation error in passing to the Riemann integral. \square

Proof of Lemma 2. First note that $I_{\tilde{\beta}}$ is the periodogram of the random field \mathbb{W}^N . To establish joint convergence, we use the Cramer-Wold device and linearity of the forms; so without loss of generality consider the problem with a single weighting function G . First note that the expectation of the spatial periodogram is

$$\mathbb{E}I_{\tilde{\beta}}(\lambda) = \sum_{t,s} \mathbb{E}\gamma_{t-s}(I) \exp\{-i\lambda \cdot (t-s)\} = \sum_{|h_1|<N, |h_2|<N} \gamma_h \exp\{-i\lambda \cdot h\} (1 - |h_1|/N)(1 - |h_2|/N), \quad (\text{A.2})$$

where the first sums have integer indices ranging from 1 to N . Thus under (23) we have $\mathbb{E}I_{\tilde{\beta}}(\lambda) - \tilde{F}(\lambda) = O(N^{-1})$ as in Guyon (1982) (note that this is weaker than the $O(N^{-2})$ result analogous to the time series case). Further results, which are needed to prove the Lemma's claims, involve generalizations of Theorems 4.3.1 and 4.3.2 of Brillinger (2001) to the spatial field case. Consider $N^{-1} \sum_{\lambda} G(\lambda) I_{\tilde{\beta}}(\lambda)$ (where the sum is a double sum over pairs of Fourier frequencies), which has mean $N^{-1} \sum_{\lambda} G(\lambda) \mathbb{E}[I_{\tilde{\beta}}(\lambda)] \sim N \langle G\tilde{F} \rangle$. That is, $N^{-2} \sum_{\lambda} G(\lambda) \mathbb{E}[I_{\tilde{\beta}}(\lambda)] \rightarrow \langle G\tilde{F} \rangle$, which does not vanish when we normalize by multiplying by N , as happens in the one-dimensional (i.e., time series) case. This arises from the fact that in (A.2)

$$(1 - h_1/N)(1 - h_2/N) = 1 - (h_1 + h_2)/N + h_1 h_2/N^2,$$

and the $O(N^{-1})$ terms do not vanish; this establishes the last claim of the Lemma. However, the results change when we redefine the periodogram using unbiased sample acfs, i.e., $\mathbb{E}\hat{\gamma}_h = \gamma_h(\tilde{F})$. Also

$$\mathbb{E}\hat{I}_{\tilde{\beta}}(\lambda) - \tilde{F}(\lambda) = \sum_{|h_1| \geq N \cup |h_2| \geq N} \gamma_h(\tilde{F}) \exp\{-i\lambda \cdot h\},$$

so that $N^{-1} \sum_{\lambda} G(\lambda) (\mathbb{E}[\hat{I}_{\tilde{\beta}}(\lambda)] - \tilde{F}(\lambda)) \rightarrow 0$ by condition B_1 . The variance of $N^{-1} \sum_{\lambda} G(\lambda) \hat{I}_{\tilde{\beta}}(\lambda)$ is given by an expansion involving cumulants. Let $\vartheta_{\ell m}$ and $\varphi_{\ell m}$ for $\ell, m = 1, 2$ be defined as various signed combinations of frequencies $\lambda_1, \lambda_2, \omega_1, \omega_2$, given as follows:

$$\begin{bmatrix} (-\lambda_1, -\lambda_2) & (\lambda_1, \lambda_2) \\ (-\omega_1, -\omega_2) & (\omega_1, \omega_2) \end{bmatrix} = \begin{bmatrix} (\vartheta_{11}, \varphi_{11}) & (\vartheta_{12}, \varphi_{12}) \\ (\vartheta_{21}, \varphi_{21}) & (\vartheta_{22}, \varphi_{22}) \end{bmatrix}.$$

For the variance calculations, we can work with $I_{\tilde{\beta}}(\lambda)$ instead of $\hat{I}_{\tilde{\beta}}(\lambda)$, using condition B_1 to control the error. The variance is then

$$\begin{aligned} & N^{-2} \sum_{\lambda, \omega} G(\lambda) G(\omega) \text{cum} \left(I_{\tilde{\beta}}(\lambda), I_{\tilde{\beta}}(\omega) \right) \\ &= N^{-2} N^{-4} \sum_{\nu} \sum_{\lambda, \omega} G(\lambda) G(\omega) \text{cum} \left(\widetilde{\mathbb{W}}(\vartheta_{\ell m}, \varphi_{\ell m}) : (\ell, m) \in \nu_1, \dots, \widetilde{\mathbb{W}}(\vartheta_{\ell m}, \varphi_{\ell m}) : (\ell, m) \in \nu_q \right), \end{aligned}$$

where the sum is over all indecomposable partitions ν of the table

$$\begin{bmatrix} (1, 1) & (1, 2) \\ (2, 1) & (2, 2) \end{bmatrix}.$$

Therefore the only proper partitions that need be considered are $\{(1, 1), (2, 1)\} \cup \{(1, 2), (2, 2)\}$ and $\{(1, 1), (2, 2)\} \cup \{(1, 2), (2, 1)\}$. We proceed to calculate the relevant asymptotic cumulants:

$$\begin{aligned} & \text{cum} \left(\widetilde{\mathbb{W}}(-\lambda_1, -\lambda_2), \widetilde{\mathbb{W}}(-\omega_1, -\omega_2) \right) \\ &= \sum_{t,s} \gamma_{t_2-t_1, s_2-s_1} \exp\{i(\lambda_1 t_1 + \lambda_2 s_1 + \omega_1 t_2 + \omega_2 s_2)\} \\ &= \sum_h \gamma_h \exp\{i\omega_1 h_1 + i\omega_2 h_2\} \left(\sum_{t_1=1}^{N-h_1} \sum_{s_1=1}^{N-h_2} \exp\{i(\lambda_1 + \omega_1)t_1 + i(\lambda_2 + \omega_2)s_1\} \right), \end{aligned}$$

which is order N^2 when $\lambda_1 = -\omega_1$ and $\lambda_2 = -\omega_2$, but is lower order otherwise. In the former case, the cumulant is asymptotic to $N^2 \widetilde{F}(-\omega_1, -\omega_2)$. The next cumulant is that for $\widetilde{\mathbb{W}}(\lambda_1, \lambda_2)$ times $\widetilde{\mathbb{W}}(\omega_1, \omega_2)$, but the analysis is just like the previous case (only with minus signs deleted), and the asymptotic in the case $\lambda_1 = -\omega_1$ and $\lambda_2 = -\omega_2$ is $N^2 \widetilde{F}(\omega_1, \omega_2)$. Next, we have

$$\begin{aligned} & \text{cum} \left(\widetilde{\mathbb{W}}(-\lambda_1, -\lambda_2), \widetilde{\mathbb{W}}(\omega_1, \omega_2) \right) \\ &= \sum_{t,s} \gamma_{t_2-t_1, s_2-s_1} \exp\{i(\lambda_1 t_1 + \lambda_2 s_1 - \omega_1 t_2 - \omega_2 s_2)\} \\ &= \sum_h \gamma_h \exp\{-i\omega_1 h_1 - i\omega_2 h_2\} \left(\sum_{t_1=1}^{N-h_1} \sum_{s_1=1}^{N-h_2} \exp\{i(\lambda_1 - \omega_1)t_1 + i(\lambda_2 - \omega_2)s_1\} \right), \end{aligned}$$

which is order N^2 when $\lambda_1 = \omega_1$ and $\lambda_2 = \omega_2$, but is lower order otherwise, and thus is asymptotic to $N^2 \widetilde{F}(\omega_1, \omega_2)$. Finally, the cumulant of $\widetilde{\mathbb{W}}(\lambda_1, \lambda_2)$ and $\widetilde{\mathbb{W}}(-\omega_1, -\omega_2)$ is similarly computed, with the result of $N^2 \widetilde{F}(-\omega_1, -\omega_2)$.

There is also the cumulant involving all four discrete Fourier Transforms, namely

$$\begin{aligned} & \text{cum} \left(\widetilde{\mathbb{W}}(-\lambda_1, -\lambda_2), \widetilde{\mathbb{W}}(\lambda_1, \lambda_2), \widetilde{\mathbb{W}}(-\omega_1, -\omega_2), \widetilde{\mathbb{W}}(\omega_1, \omega_2) \right) \\ &= \sum_{t,s,m,n} \text{cum} (W_t, W_s, W_m, W_n) \exp\{i(\lambda_1 t_1 + \lambda_2 t_2 - \lambda_1 s_1 - \lambda_2 s_2 + \omega_1 m_1 + \omega_2 m_2 - \omega_1 n_1 - \omega_2 n_2)\} \\ &\sim N^2 \sum_{h,k,\ell} \gamma_{h,k,\ell} \exp\{i(-\lambda_1 h_1 - \lambda_2 h_2 + \omega_1 k_1 + \omega_2 k_2 - \omega_1 \ell_1 - \omega_2 \ell_2)\} \\ &= N^2 \widetilde{F}F(\lambda_1, \lambda_2, -\omega_1, -\omega_2, \omega_1, \omega_2). \end{aligned}$$

Then consolidating, the variance of $N^{-1} \sum_{\lambda} G(\lambda) I_{\widetilde{\beta}}(\lambda)$ is asymptotic to

$$\begin{aligned} & N^{-4} \sum_{\lambda} \sum_{\omega} G(\lambda) G(\omega) \widetilde{F}F(\lambda_1, \lambda_2, -\omega_1, -\omega_2, \omega_1, \omega_2) \\ &+ N^{-2} \sum_{\lambda} (G(\lambda) G(-\lambda) + G^2(\lambda)) \widetilde{F}^2(\lambda) \\ &\sim (2\pi)^{-4} \int_{[-\pi, \pi]^4} G(\lambda) G(\omega) \widetilde{F}F(\lambda, -\omega, \omega) d\lambda d\omega + (2\pi)^{-2} \int_{[-\pi, \pi]^2} (G(\lambda) G(-\lambda) + G^2(\lambda)) \widetilde{F}^2(\lambda) d\lambda. \end{aligned}$$

This is the main calculation, but some details remain. The above expression provides the asymptotic variance, and once generalized to a joint result over several functions G_j yields the stated matrix V . Our results have been derived for $N^{-1} \langle G_j(\widehat{I}_{\tilde{\beta}} - \tilde{F}) \rangle_N$; for $N^{-1} \langle G_j(\widehat{I}_{\tilde{\beta}} - \tilde{F}) \rangle$ we can proceed as in Theorem 5.10.2 of Brillinger (2001), using the boundedness of the functions G_j and the above discrete Fourier Transform results to bound the difference of $I_{\tilde{\beta}}(2\pi t N^{-1}, 2\pi s N^{-1})$ and $I_{\tilde{\beta}}(\lambda)$ over the region $\{\lambda \in [2\pi t N^{-1}, 2\pi(t+1)N^{-1}] \times [2\pi s N^{-1}, 2\pi(s+1)N^{-1}]\}$ by $O_P(N^{-1})$.

Finally, higher order moments of $N^{-1} \langle G(\widehat{I}_{\tilde{\beta}} - \tilde{F}) \rangle$ will be asymptotically negligible due to the higher order ($k > 3$ in Condition B_k) cumulant conditions, which is proved along the lines of the combinatorial analysis of Theorem 1 in McElroy and Holan (2009). This asymptotic structure for the cumulants indicates a limiting normal distribution, as stated. \square

Proof of Theorem 1. We begin with the analysis of $\widehat{\beta}$. From (22) and (20) we obtain

$$\begin{aligned}\widehat{\beta}_{QMLE} - \tilde{\beta} &= \left[\tilde{X}' \Sigma(F_{\widehat{\theta}_{QMLE}}^{-1}) \tilde{X} \right]^{-1} \tilde{X}' \Sigma(F_{\widehat{\theta}_{QMLE}}^{-1}) (Y - \tilde{X} \tilde{\beta}) \\ \widehat{\beta}_{MLE} - \tilde{\beta} &= \left[\tilde{X}' \Sigma^{-1}(F_{\widehat{\theta}_{MLE}}) \tilde{X} \right]^{-1} \tilde{X}' \Sigma^{-1}(F_{\widehat{\theta}_{MLE}}) (Y - \tilde{X} \tilde{\beta}).\end{aligned}$$

So the regression errors in both cases are linear functions of W ; consistency follows from A2 and the condition that $(\tilde{X}' \tilde{X})^{-1}$ tends to the zero matrix. By the cumulant conditions that we assume (A3), the random field \mathbb{W}^N satisfies a central limit theorem. Moreover, we can replace the estimates $\widehat{\theta}$ by their limits $\tilde{\theta}$ (their consistency is proved below) because the error in doing so is of lesser order by Lemma 1. Moreover by the same Lemma, $N^{-1} \tilde{X}' \Sigma^{-1}(F_{\tilde{\theta}}) W$ has the same asymptotics as $N^{-1} \tilde{X}' \Sigma(F_{-\tilde{\theta}}) W$, which is asymptotically normal with mean zero and covariance

$$N^{-2} \tilde{X}' \Sigma(F_{-\tilde{\theta}}) \Sigma(\tilde{F}) \Sigma(F_{-\tilde{\theta}}) \tilde{X}.$$

This result follows from the argument used in Theorem 2 of Mardia and Marshall (1984), together with A2 and the assumption that $(\tilde{X}' \tilde{X})^{-1}$ tends to a zero matrix as $N \rightarrow \infty$ (plus the central limit theorem for \mathbb{W}^N by A3, as our data process may be non-Gaussian). Now applying the matrix $M_X(\tilde{\theta})$ to the asymptotic normality result yields the limit theorems for the regression QMLE and MLE. This argument holds for the exact Whittle estimates, but as in the proof of Lemma 2, can be extended to the case of approximate Whittle estimates as well (see next paragraph). Below, we will show that the MLE and QMLE for θ can be expressed as a quadratic form in the data (asymptotically), and hence is uncorrelated with the regression estimates when the third order cumulants of the data process are zero.

Now we consider the asymptotics for the parameter θ . First we examine the exact and approximate QMLEs. One may develop an error expression along the lines given for Lemma 3.1.1 of

Taniguchi and Kakizawa (2000), with $\hat{\theta}$ denoting the exact QMLE and $\hat{\theta}_N$ the approximate QMLE:

$$\begin{aligned} (\hat{\theta} - \tilde{\theta}) &= o_P(1) - \left[\nabla \nabla' KL(F_{\tilde{\theta}}, \tilde{F}) \right]^{-1} \langle \nabla F_{\tilde{\theta}}^{-1} (\hat{I}_{\tilde{\beta}} - \tilde{F}) \rangle \\ (\hat{\theta}_N - \tilde{\theta}) &= o_P(1) - \left[\nabla \nabla' KL_N(F_{\tilde{\theta}}, \tilde{F}) \right]^{-1} \langle \nabla F_{\tilde{\theta}}^{-1} (\hat{I}_{\tilde{\beta}} - \tilde{F}) \rangle_N, \end{aligned}$$

which are developed by Taylor Series expansion in each case. In order to write down this expansion, we must replace $\hat{I}_{\hat{\beta}}$ by $\hat{I}_{\tilde{\beta}}$, which uses the consistency of $\hat{\beta}$. Now in the proof of Lemma 2 it is established that KL and KL_N differ in their stochastic terms of order $O_P(N^{-1})$, whereas the log determinant terms differ by $O(N^{-2})$ via the definition of the Riemann integral. These types of results also hold for the difference of the gradients of the KL functions, since $\nabla F_{\tilde{\theta}}^{-1}$ is continuous by assumption. Then apply Lemma 2 to the above convergences, as $\nabla F_{\tilde{\theta}}^{-1} = -\nabla F_{\tilde{\theta}} \cdot F_{\tilde{\theta}}^{-2}$, which is continuous by assumption A2. Because the fourth order cumulants are zero, V has the form stated in the theorem.

Next consider the MLE case, where the scaled log Gaussian likelihood is \mathcal{L} given in (19). The method of proof follows the treatment in Section 4.1 of Taniguchi and Kakizawa (2000), but with their Lemma 4.1.2 replaced by our Lemma 1. Also note that many of the calculations are identical with those of Mardia and Marshall (1984), though our regularity conditions together with Lemma 1 essentially verify condition (iii) of their Theorem 2 (also those authors assume a Gaussian spatial process, whereas we allow small departures from normality). First

$$0 = \nabla \log \mathcal{L}(\hat{\theta}_{MLE}) = \nabla \log \mathcal{L}(\tilde{\theta}) + \nabla \nabla' \log \mathcal{L}(\tilde{\theta})(\hat{\theta} - \tilde{\theta}) + \frac{1}{2} J(\theta^*, \hat{\theta} - \tilde{\theta}) \quad (\text{A.3})$$

for some θ^* such that $\|\theta^* - \tilde{\theta}\| \leq \|\hat{\theta} - \tilde{\theta}\|$, with $\|\cdot\|$ the Euclidean norm. Here J is a 3-tensor that linearly depends on $\hat{\theta} - \tilde{\theta}$; if $\hat{\theta} - \tilde{\theta} = O_P(N^{-1})$, then this term will be $O_P(N^{-3})$. As in Taniguchi and Kakizawa (2000, p.180) set $U_N = N(\hat{\theta} - \tilde{\theta})$ and $Z_N(\theta) = N^{-1} \nabla \log L(\theta)$, which equals $X' \Sigma^{-1}(F_{\theta}) \nabla \Sigma(F_{\theta}) \Sigma^{-1}(F_{\theta}) X / (2N) - \text{tr}(\Sigma^{-1}(F_{\theta}) \nabla \Sigma(F_{\theta})) / (2N)$. Here $\nabla \Sigma(F_{\theta})$ is a shorthand for an array of matrices, each of which is the derivative of $\Sigma(F_{\theta})$ with respect to one of the θ_j . Then from (A.3) we have $U_N = -[N^{-2} H_N(\tilde{\theta})]^{-1} Z_N(\tilde{\theta})$ plus terms of order N^{-2} , with H_N equal to the Hessian of \mathcal{L} . The mean of $Z_N(\theta)$ is given by

$$\frac{1}{2N} \text{tr} \left\{ \Sigma^{-1}(F_{\theta}) \nabla \Sigma(F_{\theta}) \Sigma^{-1}(F_{\theta}) \left[\Sigma(\tilde{F}) - \Sigma(F_{\theta}) \right] \right\} = O(N^{-1}) + \frac{N}{2} \langle \nabla F_{\theta} F_{\theta}^{-2} [\tilde{F} - F_{\theta}] \rangle,$$

where the second equality follows from Lemma 1. The final integral is actually equal to the gradient of $KL(F_{\theta}, \tilde{F})$, and thus is identically zero when θ equals the pseudo-true value, no matter whether the model is correctly specified or not. Also since the fourth cumulants are assumed to be zero,

$$\text{Var} [Z_N(\theta)] = \frac{1}{2N^2} \text{tr} \left(\Sigma^{-1}(F_{\theta}) \nabla \Sigma(F_{\theta}) \Sigma^{-1}(F_{\theta}) \Sigma(\tilde{F}) \Sigma^{-1}(F_{\theta}) \nabla' \Sigma(F_{\theta}) \Sigma^{-1}(F_{\theta}) \Sigma(\tilde{F}) \right),$$

which by Lemma 1 converges to the matrix $V(\theta)/4$. Higher order cumulants of Z_N are analyzed along the lines offered in Taniguchi and Kakizawa (2000), only replacing their use of Lemma 4.1.2

(e.g., in their equation (4.1.24)) by our Lemma 1. Hence $Z_N(\theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V(\theta)/4)$. Also we have

$$\begin{aligned} 2\partial_j\partial_k \log \mathcal{L}(\theta) &= -X'\Sigma^{-1}(F_\theta)\partial_j\Sigma(F_\theta)\Sigma^{-1}(F_\theta)\partial_k\Sigma(F_\theta)\Sigma^{-1}(F_\theta)X \\ &\quad - X'\Sigma^{-1}(F_\theta)\partial_k\Sigma(F_\theta)\Sigma^{-1}(F_\theta)\partial_j\Sigma(F_\theta)\Sigma^{-1}(F_\theta)X \\ &\quad + X'\Sigma^{-1}(F_\theta)\partial_j\partial_k\Sigma(F_\theta)\Sigma^{-1}(F_\theta)X \\ &\quad - \text{tr} \left\{ \Sigma^{-1}(F_\theta)\partial_j\partial_k\Sigma(F_\theta) - \Sigma^{-1}(F_\theta)\partial_j\Sigma(F_\theta)\Sigma^{-1}(F_\theta)\partial_k\Sigma(F_\theta) \right\}, \end{aligned}$$

which is twice the Hessian matrix $H_N(\theta)$. The expectation is

$$\begin{aligned} \mathbb{E}H_N(\theta) &= -\text{tr} \left\{ \Sigma^{-1}(F_\theta)\nabla\Sigma(F_\theta)\Sigma^{-1}(F_\theta)\nabla'\Sigma(F_\theta)\sigma^{-1}(F_\theta)\Sigma(\tilde{F}) \right\} \\ &\quad + \text{tr} \left\{ \Sigma^{-1}(F_\theta)\nabla\nabla'\Sigma(F_\theta)\Sigma^{-1}(F_\theta)\Sigma(\tilde{F}) \right\} / 2 \\ &\quad - \text{tr} \left\{ \Sigma^{-1}(F_\theta)\nabla\nabla'\Sigma(F_\theta) - \Sigma^{-1}(F_\theta)\nabla\Sigma(F_\theta)\Sigma^{-1}(F_\theta)\nabla'\Sigma(F_\theta) \right\} / 2, \end{aligned}$$

and applying Lemma 1 we have $N^{-2}\mathbb{E}H_N(\theta)$ tends to

$$- \langle F_\theta^{-3}\nabla F_\theta\nabla'F_\theta\tilde{F} \rangle + \langle F_\theta^{-2}\nabla\nabla'F_\theta\tilde{F} \rangle / 2 - \langle F_\theta^{-1}\nabla\nabla'F_\theta \rangle / 2 + \langle F_\theta^{-2}\nabla F_\theta\nabla'F_\theta \rangle / 2,$$

which is half the Hessian of $KL(F_\theta, \tilde{F})$, i.e., $H(\theta)/2$. Also by Lemma 1 we obtain

$$\text{Var} (N^{-2}\nabla\nabla' \log L(\theta)) \sim \frac{1}{2N^2} \langle \left(2\nabla F_\theta\nabla'F_\theta\tilde{F}^{-2} - \nabla\nabla'F_\theta\tilde{F}^{-1} \right)^2 \rangle$$

as $N \rightarrow \infty$, and hence $N^{-2}H_N(\tilde{\theta}) \xrightarrow{P} H(\tilde{\theta})/2$. Then $H(\tilde{\theta})U_N \xrightarrow{\mathcal{L}} \mathcal{N}(0, V(\tilde{\theta}))$, and the stated result follows. \square

Proof of Theorem 2. All integrals are multi-dimensional. Let $\|\cdot\|$ denote the Euclidean norm. Let $p(\theta|X)$ denote the posterior for $\phi = [\theta', \beta']'$, which is proportional to the Gaussian likelihood $p(X|\phi)$ times $p(\phi)$, the prior for the parameter vector. The posterior estimate is the vector quantity $\hat{\phi} = \int \phi p(\phi|X) d\phi$; we seek to show that it converges in probability to $\tilde{\phi}$, the true parameter value. We further develop the likelihood as

$$\begin{aligned} p(X|\phi) &= \exp \left[-\frac{N^2}{2} \{ \log(2\pi) + \mathcal{L}_N(\theta, \beta) \} \right] \\ \mathcal{L}_N(\theta, \beta) &= N^{-2} \log |\Sigma(F_\theta)| + N^{-2} (Y - \tilde{X}\beta)' \Sigma^{-1}(F_\theta) (Y - \tilde{X}\beta). \end{aligned}$$

We next expand the quadratic form so as to relate it to the true $\tilde{\beta}$ parameter:

$$\begin{aligned} (Y - \tilde{X}\beta)' \Sigma^{-1}(F_\theta) (Y - \tilde{X}\beta) &= (Y - \tilde{X}\tilde{\beta})' \Sigma^{-1}(F_\theta) (Y - \tilde{X}\tilde{\beta}) \\ &\quad + 2(\tilde{\beta} - \beta)' \tilde{X}' \Sigma^{-1}(F_\theta) (Y - \tilde{X}\tilde{\beta}) \\ &\quad + (\tilde{\beta} - \beta)' \tilde{X}' \Sigma^{-1}(F_\theta) \tilde{X} (\tilde{\beta} - \beta) \end{aligned}$$

The first term on the right hand represents the same quadratic form, but with $\tilde{\beta}$ replacing β ; therefore it is a quadratic form in W . The second term is linear in W , so as in the proof of Theorem 1 will obey a central limit theorem and is $O_P(N)$. The final term is deterministic, and under the conditions of the theorem is of order N^2 . As a result,

$$\mathcal{L}_N(\theta, \beta) = \mathcal{L}_N(\theta, \tilde{\beta}) + E_N + Q(\theta, \beta) \quad (\text{A.4})$$

where E_N is an error term tending to zero in probability and $Q(\theta, \beta) = (\tilde{\beta} - \beta)' M(\theta) (\tilde{\beta} - \beta)$. Note that by our assumptions on $M(\theta)$, the quadratic form $Q(\theta, \beta)$ is positive whenever $\beta \neq \tilde{\beta}$, for any θ , and moreover is bounded above by $\bar{\Lambda} \|\tilde{\beta} - \beta\|^2$ for all θ , where $\bar{\Lambda}$ bounds all the eigenvalues of $M(\theta)$ for all θ . Also $\underline{\Lambda}$ is the lower (positive) bound on the eigenvalues of $M(\theta)$.

So for any $\epsilon > 0$, define neighborhoods $N_\epsilon(\tilde{\theta}) = \{\theta : \|\theta - \tilde{\theta}\| < \epsilon\}$ and similarly $N_\epsilon(\tilde{\beta})$. We can then decompose the parameter space (because all parameters are contained in a compact subset of Euclidean space) into $N_\epsilon(\tilde{\theta}) \times N_\epsilon(\tilde{\beta})$ and its complement. This complement can be broken into two overlapping regions, R_1 and R_2 , where $R_1 = N_\epsilon^c(\tilde{\theta})$ and $R_2 = N_\epsilon^c(\tilde{\beta})$. The complements are taken in the entire product parameter space, so R_1 corresponds to $\|\tilde{\theta} - \theta\| > \epsilon$ but β is unconstrained, while R_2 corresponds to $\|\tilde{\beta} - \beta\| > \epsilon$ but θ is unconstrained. The intersection region $R_1 \cap R_2$ is characterized by both constraints being true. In order to get disjoint regions, let $\bar{R}_1 = R_1 \setminus (R_1 \cap R_2)$ and $\bar{R}_2 = R_2 \setminus (R_1 \cap R_2)$. Then we decompose the vector parameter error as

$$\begin{aligned} \hat{\phi} - \tilde{\phi} &= \frac{\int_{N_\epsilon(\tilde{\theta}) \times N_\epsilon(\tilde{\beta})} (\phi - \tilde{\phi}) \exp\left\{-\frac{N^2}{2} \mathcal{L}_N(\theta, \beta)\right\} p(\phi) d\phi}{\int \exp\left\{-\frac{N^2}{2} \mathcal{L}_N(\theta, \beta)\right\} p(\phi) d\phi} \\ &+ \frac{\int_{\bar{R}_1 \cup \bar{R}_2 \cup (R_1 \cap R_2)} (\phi - \tilde{\phi}) \exp\left\{-\frac{N^2}{2} \mathcal{L}_N(\theta, \beta)\right\} p(\phi) d\phi}{\int \exp\left\{-\frac{N^2}{2} \mathcal{L}_N(\theta, \beta)\right\} p(\phi) d\phi}. \end{aligned}$$

The integrands $\phi - \tilde{\phi}$ are vectors, and the integrals are interpreted as integrating each component and stacking the result. The first summand has Euclidean norm bounded by ϵ with probability one. Focusing on the second term, we proceed to develop two key bounds (see Liseo et al., 2001; Holan et al., 2009, for a similar approach):

$$\mathcal{L}_N(\theta, \tilde{\beta}) - \mathcal{L}_N(\tilde{\theta}, \tilde{\beta}) \geq \bar{\mathcal{C}}(\theta) \quad (\text{A.5})$$

for all $\theta \in N_\epsilon^c(\tilde{\theta})$ with probability tending to one. Note that this bound will pertain to the region R_1 , as β is not involved. The bounding function $\bar{\mathcal{C}}$ will be shown to be convex. We also will show

$$\mathcal{L}_N(\theta, \tilde{\beta}) - \mathcal{L}_N(\tilde{\theta}, \tilde{\beta}) \leq \underline{\mathcal{C}}(\theta) \quad (\text{A.6})$$

for all $\theta \in N_\delta^c(\tilde{\theta})$, for a convex function $\underline{\mathcal{C}}$ and another radius $0 < \delta < \epsilon$. It must be chosen such that $\delta^2 \bar{\Lambda} < \epsilon^2 \underline{\Lambda}$. In the above error decomposition $\epsilon > 0$ was arbitrary, but it will be chosen sufficiently

small so as to guarantee that

$$\sup_{\theta \in N_\delta(\tilde{\theta})} \underline{C}(\theta) + \bar{\Lambda}\delta^2 < \inf_{\theta \in N_\epsilon^c(\tilde{\theta})} \overline{C}(\theta). \quad (\text{A.7})$$

The proofs of (A.5) and (A.6) follow the strategy in Theorem 3.1 of Dahlhaus (1989), although we have recourse to Lemma 1 for the spatial random field case. The expectation of $\mathcal{L}_N(\theta, \tilde{\beta}) - \mathcal{L}_N(\tilde{\theta}, \tilde{\beta})$ equals

$$-N^{-2} \log |\Sigma^{-1}(F_\theta)\Sigma(F_{\tilde{\theta}})| + N^{-2} \text{tr}(\Sigma^{-1}(F_\theta)\Sigma(F_{\tilde{\theta}}) - 1_{N^2}).$$

Of course 1_{N^2} is the identity matrix. Using a Taylor Series expansion, as in Theorem 3.1 of Dahlhaus (1989), produces the expectation being equal to

$$\text{tr} \left((1_{N^2} + \tau A)^{-1} A (1_{N^2} + \tau A)^{-1} A \right), \quad (\text{A.8})$$

where $A = \Sigma^{-1}(F_\theta)\Sigma(F_{\tilde{\theta}}) - 1_{N^2}$ and τ is some number between 0 and 1; note that τ will depend upon the entries of A . The above trace expression can be re-written as a sum of traces of block-Toeplitz matrices, to which Lemma 1 can be applied. Observing that $A = \Sigma^{-1}(F_\theta)[\Sigma(F_{\tilde{\theta}}) - \Sigma(F_\theta)]$ and $1_{N^2} + \tau A = \Sigma^{-1}(F_\theta)[(1 - \tau)\Sigma(F_\theta) + \tau\Sigma(F_{\tilde{\theta}})]$, then $(1_{N^2} + \tau A)^{-1} = \Sigma^{-1}(G)\Sigma(F_\theta)$ with $G = (1 - \tau)F_\theta + \tau F_{\tilde{\theta}}$. The sum of the block-Toeplitz matrices is again the block-Toeplitz matrix of G , because the summands are positive functions. The dependency of G on τ , θ , and $\tilde{\theta}$ will be suppressed in the notation. The trace expression (A.8) can then be expanded using these derivations, which yields the trace of

$$\begin{aligned} & \Sigma^{-1}(G)\Sigma(F_{\tilde{\theta}})\Sigma^{-1}(G)\Sigma(F_{\tilde{\theta}}) \\ & - \Sigma^{-1}(G)\Sigma(F_\theta)\Sigma^{-1}(G)\Sigma(F_{\tilde{\theta}}) \\ & - \Sigma^{-1}(G)\Sigma(F_{\tilde{\theta}})\Sigma^{-1}(G)\Sigma(F_\theta) \\ & + \Sigma^{-1}(G)\Sigma(F_\theta)\Sigma^{-1}(G)\Sigma(F_\theta). \end{aligned}$$

For any fixed τ , the trace of the above expression, divided by N^2 , converges to $\langle G^{-2}(F_{\tilde{\theta}} - F_\theta)^2 \rangle$ as $N \rightarrow \infty$; therefore the infimum of this quantity over $\tau \in [0, 1]$ yields a lower bound, which is a function of θ . Furthermore, utilizing the bound $0 < F_{\tilde{\theta}}/F_\theta < K$ for some constant K , for all frequencies, we have $G/F_\theta < K$ uniformly (and τ becomes irrelevant). Then let

$$\overline{C}(\theta) = K^{-2} \langle (F_{\tilde{\theta}}/F_\theta - 1)^2 \rangle.$$

So $\overline{C}(\theta)$ is clearly non-negative, and is zero only when $F_{\tilde{\theta}} = F_\theta$. By the identifiability restriction, this will only occur when $\theta = \tilde{\theta}$ (the cepstral model class satisfies this condition, due to the basis expansion of log spectra implicit in the model's construction). $\overline{C}(\theta)$ is also concave down, since $x \mapsto (x - 1)^2$ is (cf., Proposition 2.15, Vajda, 1989).

Moreover, the variance of $\mathcal{L}_N(\theta, \tilde{\beta}) - \mathcal{L}_N(\tilde{\theta}, \tilde{\beta})$ equals $2N^{-4}\text{tr}[A^2]$ under a Gaussian DGP, and again by Lemma 1 – after an expansion of A^2 – we obtain $O(N^{-2})$ growth. This establishes (A.5)

for any θ . By taking the infimum of $\overline{\mathcal{C}}(\theta)$ over $N_\epsilon^c(\tilde{\theta})$, we can ensure a non-zero value as well, utilizing the convexity of the function. For the other bound, note that for any θ_1, θ_2 we have

$$\mathcal{L}_N(\theta_2, \tilde{\beta}) - \mathcal{L}_N(\theta_1, \tilde{\beta}) = N^{-2}(\theta_2 - \theta_1)' \text{tr}\{\Sigma^{-1}(F_\theta)\Sigma(\nabla F_\theta)\} + N^{-2}W'(\Sigma^{-1}(F_{\theta_2}) - \Sigma^{-1}(F_{\theta_1}))W,$$

where θ is in-between θ_1 and θ_2 , using the Mean Value Theorem and the assumed smoothness on spectra. The notation $\Sigma(\nabla F_\theta)$ denotes an array of matrices, indexed by the various derivatives of F_θ ; after the trace has been computed, the resulting vector is multiplied by $[\theta_2 - \theta_1]'$. Using the smoothness conditions on spectra, we may apply Lemma 5.5 of Dahlhaus (1989) and our Lemma 1 to bound the absolute value of the above difference by a constant times $\|\theta_2 - \theta_1\|$, with probability tending to one. Letting θ_1 be $\tilde{\theta}$ and $\underline{\mathcal{C}}(\theta) = K\|\theta - \tilde{\theta}\|$ for some constant $K > 0$ now yields (A.6).

Because of the condition on $M(\theta)$, its eigenvalues are uniformly (in θ) bounded between $0 < \underline{\Lambda}$ and $\overline{\Lambda} < \infty$, and the quadratic form satisfies

$$Q(\theta, \beta) \geq \underline{\Lambda}\epsilon^2 \tag{A.9}$$

for all $\beta \in N_\epsilon^c(\tilde{\beta})$ and any θ . Furthermore we also have

$$Q(\theta, \beta) \leq \overline{\Lambda}\delta^2 \tag{A.10}$$

for all $\beta \in N_\delta(\tilde{\beta})$ and any θ . Below, it is shown that we must select δ for a given ϵ such that $\delta^2\overline{\Lambda} < \epsilon^2\underline{\Lambda}$. Next, using the convexity of $\underline{\mathcal{C}}$ and $\overline{\mathcal{C}}$, adjusting their constants if necessary, we can ensure that (A.7) holds for some ϵ , making δ smaller if necessary. Now the second term in the parameter error can be rewritten as

$$\frac{\int_{\overline{R}_1 \cup \overline{R}_2 \cup (R_1 \cap R_2)} (\phi - \tilde{\phi}) \exp\left\{-\frac{N^2}{2} [\mathcal{L}_N(\theta, \tilde{\beta}) - \mathcal{L}_N(\tilde{\theta}, \tilde{\beta}) + E_N + Q(\theta, \beta)]\right\} p(\phi) d\phi}{\int \exp\left\{-\frac{N^2}{2} [\mathcal{L}_N(\theta, \tilde{\beta}) - \mathcal{L}_N(\tilde{\theta}, \tilde{\beta}) + E_N + Q(\theta, \beta)]\right\} p(\phi) d\phi},$$

by multiplying top and bottom by the constant $\mathcal{L}_N(\tilde{\theta}, \tilde{\beta})$ and utilizing (A.4). Furthermore, the denominator (which is a scalar) can be made smaller by restricting the region of integration, which only increases the absolute error of each component of the parameter error vector. We will restrict the integration to the set $N_\delta(\tilde{\theta}) \times N_\delta(\tilde{\beta})$. At this point, we break the analysis up by considering the numerator integral (still as a vector quantity) over the various disjoint regions. We present an analysis for integration over R_1 and R_2 , and note that either analysis will then pertain to $R_1 \cap R_2$.

First consider the numerator restricted to integration over R_1 , and take the Euclidean norm of this portion of the error vector; then we obtain the upper bound

$$\begin{aligned} & \frac{\int_{R_1} \|\phi - \tilde{\phi}\| \exp\left\{-\frac{N^2}{2} [\overline{\mathcal{C}}(\theta)]\right\} p(\phi) d\phi}{\int_{N_\delta(\tilde{\theta}) \times N_\delta(\tilde{\beta})} \exp\left\{-\frac{N^2}{2} [\underline{\mathcal{C}}(\theta) + \underline{\Lambda}\delta^2]\right\} p(\phi) d\phi} \\ & \leq \frac{\int_{R_1} \|\phi - \tilde{\phi}\| \exp\left\{-\frac{N^2}{2} \left[\inf_{\theta \in N_\epsilon^c(\tilde{\theta})} \overline{\mathcal{C}}(\theta) - \sup_{\theta \in N_\delta(\tilde{\theta})} \underline{\mathcal{C}}(\theta) - \overline{\Lambda}\delta^2\right]\right\} p(\phi) d\phi}{\int_{N_\delta(\tilde{\theta}) \times N_\delta(\tilde{\beta})} p(\phi) d\phi} \end{aligned}$$

with probability tending to one. The second expression follows from the first by taking supremums and infimums. The first expression uses the triangle inequality for integrals along with (A.5), (A.6), and (A.10). Note that these bounds hold with probability tending to one, and therefore they also hold with probability tending to one when the term E_N (which tends to zero in probability as $N \rightarrow \infty$) is inserted. In the numerator, (A.5) applies because we are in the set R_1 , whereas the other bounds hold on the set $N_\delta(\tilde{\theta}) \times N_\delta(\tilde{\beta})$. The overall bound on this term tends to zero as $N \rightarrow \infty$, because the numerator integral involves the exponential of $-N^2$ times a positive quantity – it is positive by (A.7).

Next, we consider the numerator restricted to R_2 , and mimic the same line of argument, obtaining the upper bound

$$\begin{aligned} & \frac{\int_{R_2} \left\| \phi - \tilde{\phi} \right\| \exp \left\{ -\frac{N^2}{2} [\underline{\Lambda} \epsilon^2] \right\} p(\phi) d\phi}{\int_{N_\delta(\tilde{\theta}) \times N_\delta(\tilde{\beta})} \exp \left\{ -\frac{N^2}{2} [\underline{C}(\theta) + \underline{\Lambda} \delta^2] \right\} p(\phi) d\phi} \\ & \leq \frac{\int_{R_2} \left\| \phi - \tilde{\phi} \right\| \exp \left\{ -\frac{N^2}{2} [\underline{\Lambda} \epsilon^2 - \sup_{\theta \in N_\delta(\tilde{\theta})} \underline{C}(\theta) - \overline{\Lambda} \delta^2] \right\} p(\phi) d\phi}{\int_{N_\delta(\tilde{\theta}) \times N_\delta(\tilde{\beta})} p(\phi) d\phi} \end{aligned}$$

with probability tending to one, using (A.9). Now we need the condition that $\sup_{\theta \in N_\delta(\tilde{\theta})} \underline{C}(\theta) < \epsilon^2 \underline{\Lambda} - \delta^2 \overline{\Lambda}$, which is assured by taking δ even smaller as needed. Then this term also tends to zero as $N \rightarrow \infty$, which concludes the proof. \square

References

- Bandyopadhyay, S. and Lahiri, S. (2010). “Asymptotic Properties of Discrete Fourier Transforms for Spatial Data.” *Sankhya, Series A*, 71-A, 2, 221–259.
- Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, vol. 101. Chapman & Hall.
- Besag, J. (1972). “On the Correlation Structure of Some Two-Dimensional Stationary Processes.” *Biometrika*, 59, 1, 43–48.
- (1974). “Spatial Interaction and the Statistical Analysis of Lattice Systems.” *Journal of the Royal Statistical Society – Series B (Methodological)*, 192–236.
- Besag, J. and Green, P. (1993). “Spatial Statistics and Bayesian Computation.” *Journal of the Royal Statistical Society – Series B (Methodological)*, 25–37.
- Bloomfield, P. (1973). “An Exponential Model for the Spectrum of a Scalar Time Series.” *Biometrika*, 60, 217–226.
- Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. Berkeley: University of California Press.
- Brillinger, D. (2001). *Time Series: Data Analysis and Theory*, vol. 36. Philadelphia: SIAM.
- Cressie, N. (1993). *Statistics for Spatial Data*. New York: John Wiley and Sons.
- Cressie, N. and Wikle, C. (2011). *Statistics for Spatio-Temporal Data*. John Wiley and Sons.
- Dahlhaus, R. (1989). “Efficient Parameter Estimation for Self-Similar Processes.” *The Annals of Statistics*, 17, 4, 1749–1766.
- Dahlhaus, R. and Wefelmeyer, W. (1996). “Asymptotically Optimal Estimation in Misspecified Time Series Models.” *The Annals of Statistics*, 24, 3, 952–974.
- Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*, vol. 537. Wiley-Interscience.
- Guyon, X. (1982). “Parameter Estimation for a Stationary Process on a d-Dimensional Lattice.” *Biometrika*, 69, 1, 95–105.
- Hodges, J. S. and Reich, B. J. (2010). “Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love.” *The American Statistician*, 64, 4, 325–334.
- Holan, S., McElroy, T., and Chakraborty, S. (2009). “A Bayesian Approach to Estimating the Long Memory Parameter.” *Bayesian Analysis*, 4, 1, 159–190.
- Hurvich, C. M. (2002). “Multistep Forecasting of Long Memory Series Using Fractional Exponential Models.” *International Journal of Forecasting*, 18, 2, 167–179.

- Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*, vol. 323. John Wiley and Sons.
- Kizilkaya, A. (2007). “On the Parameter Estimation of 2-D Moving Average Random Fields.” *Circuits and Systems II: Express Briefs, IEEE Transactions on*, 54, 11, 989–993.
- Kizilkaya, A. and Kayran, A. (2005). “ARMA-Cepstrum Recursion Algorithm for the Estimation of the MA Parameters of 2-D ARMA Models.” *Multidimensional Systems and Signal Processing*, 16, 4, 397–415.
- Liseo, B., Marinucci, D., and Petrella, L. (2001). “Bayesian Semiparametric Inference on Long-Range Dependence.” *Biometrika*, 88, 4, 1089–1104.
- Mardia, K. and Marshall, R. (1984). “Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression.” *Biometrika*, 71, 1, 135–146.
- McElroy, T. and Findley, D. (2010). “Selection Between Models Through Multi-Step-Ahead Forecasting.” *Journal of Statistical Planning and Inference*, 140, 12, 3655–3675.
- McElroy, T. and Holan, S. (2009). “A Local Spectral Approach for Assessing Time Series Model Misspecification.” *Journal of Multivariate Analysis*, 100, 4, 604–621.
- Mercer, W. and Hall, A. (1911). “The Experimental Error of Field Trials.” *Journal of Agricultural Science*, 4, 107–132.
- Moran, P. A. (1950). “Notes on Continuous Stochastic Phenomena.” *Biometrika*, 17–23.
- Noh, J. and Solo, V. (2007). “A True Spatio-Temporal Test Statistic for Activation Detection in fMRI by Parametric Cepstrum.” In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 1, 1–321. IEEE.
- Paciorek, C. J. (2010). “The Importance of Scale for Spatial-Confounding Bias and Precision of Spatial Regression Estimators.” *Statistical Science*, 25, 1, 107–125.
- Pierce, D. (1971). “Least Squares Estimation in the Regression Model with Autoregressive-Moving Average Errors.” *Biometrika*, 58, 2, 299–312.
- Pourahmadi, M. (1984). “Taylor Expansion of $\exp\{\sum_{k=0}^{\infty} a_k z^k\}$ and Some Applications.” *American Mathematical Monthly*, 91, 5, 303–307.
- R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rosenblatt, M. (1985). *Stationary Sequences and Random Fields*. Birkhauser.
- (2000). *Gaussian and Non-Gaussian Linear Time Series and Random Fields*. Springer Verlag.
- Rue, H. and Held, L. (2010). “Discrete Spatial Variation.” In *Handbook of Spatial Statistics*, eds. M. F. A. Gelfand, P. Diggle and P. Guttorp. London: Chapman and Hall.
- Sandgren, N. and Stoica, P. (2006). “On Nonparametric Estimation of 2-D Smooth Spectra.” *Signal Processing Letters, IEEE*, 13, 10, 632–635.

- Solo, V. (1986). “Modeling of Two-Dimensional Random Fields by Parametric Cepstrum.” *IEEE Transactions on Information Theory*, 32, 6, 743–750.
- Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Verlag.
- Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. New York: Springer.
- Vajda, I. (1989). *Theory of Statistical Inference and Information*. Kluwer Academic Publishers Boston.
- Wahba, G. (1968). “On the Distribution of Some Statistics Useful in the Analysis of Jointly Stationary Time Series.” *The Annals of Mathematical Statistics*, 1849–1862.
- Whittle, P. (1954). “On Stationary Processes in the Plane.” *Biometrika*, 434–449.

Parameter	True Value	Mean	Standard Deviation	Mean Squared Error
θ_1	0.009	0.002	0.055	0.003
θ_2	-0.028	-0.040	0.047	0.002
θ_3	0.132	0.123	0.046	0.002
θ_4	0.067	0.063	0.041	0.002
θ_5	0.271	0.256	0.054	0.003
θ_6	0.383	0.379	0.043	0.002
θ_7	0.001	-0.017	0.048	0.003
θ_8	-0.017	-0.036	0.048	0.003
θ_9	-0.003	-0.010	0.056	0.003
θ_{10}	-0.055	-0.056	0.044	0.002
θ_{11}	-0.015	-0.018	0.042	0.002
θ_{12}	0.144	0.138	0.049	0.002
θ_{13}	-0.871	-0.906	0.061	0.005
β_0	7.656	7.646	0.191	0.036
β_1	-0.035	-0.036	0.010	0.000
β_2	-0.059	-0.057	0.009	0.000

Table 1: Simulation results for the simulation presented in Section 5.1. Note, there were 200 simulated datasets and the mean squared error for β_1 and β_2 are 9.54×10^{-5} and 8.62×10^{-5} , respectively. The values in the above table are only reported to three decimal places and the elements of $\theta = \text{vec}[\Theta]$ and β correspond to the vectorized data.

	$p = 1$	$p = 2$	$p = 3$
$-\log(L)_T$	64.906	36.985	21.195
$-\log(L)_{NT}$	85.245	51.113	30.499
AIC_T	145.813	105.970	98.388
AIC_{NT}	182.490	130.226	112.998
BIC_T	179.530	173.404	216.399
BIC_{NT}	207.778	189.230	222.578
HQ_T	159.043	132.431	144.696
HQ_{NT}	192.413	153.379	155.997

Table 2: Log likelihood ($\log(L)$), Akaike information criterion (AIC), Bayesian information criterion (BIC), and Hannan-Quinn information criterion (HQ) for the models evaluated in Section 5.2. T and NT denote models with the trend (row and column) effect accounted for and no-trend effect accounted for (i.e., constant-mean only), respectively. The bold entries are the minimum AIC, BIC and HQ.

Parameter	Estimate	Standard Error
θ_1	0.009	0.052
θ_2	-0.028	0.049
θ_3	0.132	0.052
θ_4	0.067	0.049
θ_5	0.271	0.047
θ_6	0.383	0.046
θ_7	0.001	0.055
θ_8	-0.017	0.048
θ_9	-0.003	0.051
θ_{10}	-0.055	0.049
θ_{11}	-0.015	0.047
θ_{12}	0.144	0.047
θ_{13}	-0.871	0.063
β_0	7.646	0.176
β_1	-0.035	0.010
β_2	-0.059	0.009

Table 3: Straw yield data analysis results (Section 5.2). Note that the estimated standard error is obtained as the square root of the diagonal of the inverse observed Hessian matrix and that elements of $\theta = \text{vec}[\Theta]$ and β correspond to the vectorized data.

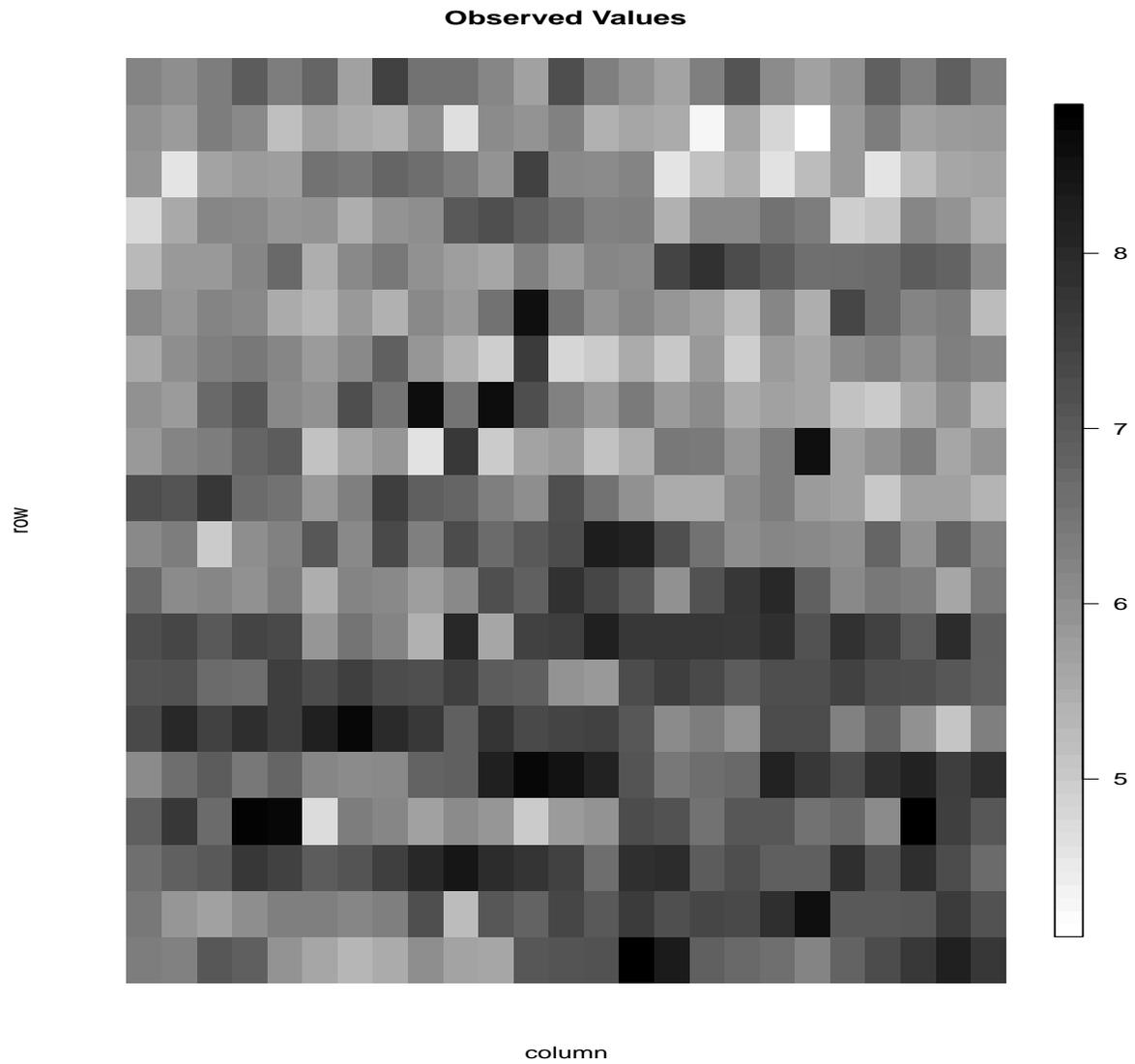


Figure 1: Observed Mercer and Hall straw yield data (Mercer and Hall, 1911). The observed lattice is 20×25 . Full details of the agricultural experiment are described in Section 5.