CRRAO Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS)

Research Report

Author (s): TATA SUBBA RAO, SOURAV DAS AND GEORGI N. BOSHNAKOV

Title of the Report: A FREQUENCY DOMAIN APPROACH FOR THE ESTIMATION OF PARAMETERS OF SPATIO-TEMPORAL STATIONARY RANDOM PROCESSES

Research Report No.: RR2014-18

Date: July 7, 2014

Prof. C R Rao Road, University of Hyderabad Campus,
Gachibowli, Hyderabad-500046, INDIA.
www.crraoaimscs.org

JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **35:** 357–377 (2014)
Published online 28 March 2014 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12069

ORIGINAL ARTICLE

A FREQUENCY DOMAIN APPROACH FOR THE ESTIMATION OF PARAMETERS OF SPATIO-TEMPORAL STATIONARY RANDOM PROCESSES†

TATA SUBBA RAO[a,b] SOURAV DAS[c]* AND GEORGI N. BOSHNAKOV[a]

[a] *University of Manchester, Manchester, UK*
[b] *CR Rao AIMSCS, Hyderabad University, Hyderabad, India*
[c] *National University of Singapore, Singapore*

A frequency domain methodology is proposed for estimating parameters of covariance functions of stationary spatio-temporal processes. Finite Fourier transforms of the processes are defined at each location. Based on the joint distribution of these complex valued random variables, an approximate likelihood function is constructed. The sampling properties of the estimators are investigated. It is observed that the expectation of these transforms can be considered to be a frequency domain analogue of the classical variogram. We call this measure frequency variogram. The method is applied to simulated data and also to Pacific wind speed data considered earlier by Cressie and Huang (1999). The proposed method does not depend on the distributional assumptions about the process.

Received 24 September 2012; Revised 14 January 2014; Accepted 27 January 2014

Keywords: Frequency variogram; discrete fourier transform; composite likelihood.

1. INTRODUCTION AND NOTATION

Spatio-temporal data arise in many areas such as epidemiology, environmental sciences (in particular weather sciences), marine biology, agriculture, geology and finance to name a few. It is therefore necessary to develop suitable statistical methods for analysis of such data. There is a vast literature devoted to the analysis of spatial data (i.e. data that are a function of spatial coordinates only). Once an extra dimension, like time, is introduced, the available methodology is no longer applicable, and any method developed should take into account not only spatial and temporal dependencies but also their interaction. The literature on spatio-temporal processes is a bit sparse compared to the literature on spatial processes. Recent books by Cressie and Wikle (2011) and Sherman (2010) should help to fill in this gap. In the following, we briefly introduce the notation and summarize the contents of the paper.

Let the spatio-temporal process be denoted by $Z(\mathbf{s}, t)$, where $\{(\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{Z}\}$. Assume that the process is observed at m different spatial locations and n equally spaced time points. So, we have a total of $m.n = M_1$ observations of the process $\{Z(\mathbf{s}_i, t) : i = 1, 2, \ldots, m; t = 1, 2, \ldots, n\}$. To ensure that the random process has finite second-order moments, we assume that $\text{Var}[Z(\mathbf{s}, t)]$ is finite. The mean and covariance functions of the process are defined as follows:

$$\begin{aligned} \mu(\mathbf{s}, t) &= \text{E}[Z(\mathbf{s}, t)] \\ C(\mathbf{s}_i, \mathbf{s}_j; t+u, t) &= \text{Cov}[Z(\mathbf{s}_i; t+u), Z(\mathbf{s}_j; t)]\,\{i = 1, 2, \ldots, m; t = 1, 2, \ldots, n\}. \end{aligned} \tag{1}$$

* Correspondence to: Sourav Das, National University of Singapore, Singapore. E-mail: sourav2.das@googlemail.com
†This article is dedicated to the memory of Professor M. B. Priestley who passed away on 15 June 2013.

We assume that the random process is second-order spatially and temporally stationary, that is,

$$\begin{aligned} \mu &= \mathrm{E}[Z(\mathbf{s}, t)] \\ C(\mathbf{s}_i - \mathbf{s}_j; u) &= \mathrm{Cov}[Z(\mathbf{s}_i; t+u), Z(\mathbf{s}_j; t)]. \end{aligned} \tag{2}$$

We note that $C(\mathbf{s}_i - \mathbf{s}_j; 0)$ and $C(\mathbf{0}; u)$ correspond to the purely spatial and purely temporal covariances of the process respectively. A further, stronger, assumption is isotropy. The random process $Z(\mathbf{s}_i; t_j)$ is said to be spatially isotropic if

$$\mathrm{Cov}[Z(\mathbf{s}_i; t+u), Z(\mathbf{s}_j; t)] = C(\|\mathbf{s}_i - \mathbf{s}_j\|; u). \tag{3}$$

The variogram for the aforementioned spatio-temporal process can be defined similarly to that of spatial processes as

$$2\gamma(\mathbf{h}; u|\boldsymbol{\theta}) = \mathrm{Var}\left\{Z(\mathbf{s}+\mathbf{h}; t+u) - Z(\mathbf{s}; t)\right\}. \tag{4}$$

In the case of second-order stationarity, the variogram reduces to

$$2\gamma(\mathbf{h}; u|\boldsymbol{\theta}) = 2\{C(\mathbf{0}; 0|\boldsymbol{\theta}) - C(\mathbf{h}; u|\boldsymbol{\theta})\}, \ \mathbf{h} \in \mathbb{R}^d, \, u \in \mathbb{Z}. \tag{5}$$

The knowledge of the covariance function $C(\mathbf{h}; u|\boldsymbol{\theta})$ (vis-è-vis variogram) is essential for the linear prediction of an unknown observation at a known location. We briefly outline the approach. Denoting the observed sample by $\mathbf{Z}(\mathbf{s}) = (Z(\mathbf{s}_1; t_1), \ldots, Z(\mathbf{s}_{M_1}; t_{M_1}))'$, one is often interested in predicting the process at a specified location and time point, say $Z(\mathbf{s}_0, t_0)$, based on the observation vector. The minimum mean square (optimal) linear predictor is well known to be

$$Z(\mathbf{s}_0, t_0) = \mu(\mathbf{s}_0, t_0) + \mathbf{c}(\mathbf{s}_0, t_0)' \Sigma^{-1}(\mathbf{Z}(\mathbf{s}) - \boldsymbol{\mu}(\mathbf{s})), \tag{6}$$

where $\mathbf{c}(\mathbf{s}_0; t_0) = \mathrm{Cov}[Z(\mathbf{s}_0; t_0), \mathbf{Z}(\mathbf{s})]$, $\Sigma = \{\mathrm{Cov}[Z(\mathbf{s}_i; t_i), Z(\mathbf{s}_j; t_j)]\}$ and $\boldsymbol{\mu}(\mathbf{s}) = \mathrm{E}(\mathbf{Z}(\mathbf{s}))$. When the mean $\boldsymbol{\mu}$ and the dispersion matrix Σ are known, the aforementioned predictor is called the *simple kriging predictor* (see, e.g. Cressie, 1993, Ch. 3). The *mean square prediction error* is given by $\sigma^2 - \mathbf{c}(\mathbf{s}_0, t_0)' \Sigma^{-1} \mathbf{c}(\mathbf{s}_0, t_0)$.

To evaluate (6), we need the inverse of Σ, and as the number of spatial locations and length of time series increase, the inversion becomes complicated. In many practical situations, we need to estimate Σ and $\mathbf{c}(\mathbf{s}_0, t_0)$. Though in theory Σ can be estimated, the estimation of the elements of $\mathbf{c}(\mathbf{s}_0, t_0)$ is not possible, as we do not have observations at the location $\mathbf{s}_0$. So as to circumvent this, it is often assumed that a parametric covariance function can be specified. The covariance function will be a function of some unknown parameters, which may need to be estimated from the data. Any covariance function defined and used must be positive definite (see Cressie and Wikle, 2011). Once a covariance function is decided, an important problem is the estimation of the parameters of this function using the data. Though there is a substantial literature for the estimation of the parameters of the covariance functions and their limitations in the case of purely spatial processes, there has been limited research in the case of spatio-temporal processes.

In this article, our objective is to consider the estimation of the parameters of spatio-temporal covariance functions using the frequency domain approach. Cressie and Huang (1999), Gneiting (2002) and Ma (2002), among others, have constructed non-separable spatio-temporal covariance functions and have considered the estimation of the parameters using similar methods as were used for spatial data. There are several limitations of their approaches, and it is our objective here to provide a more useful and satisfactory approach.

We propose a frequency domain method for the estimation of a given spatio-temporal covariance function (or, equivalently, its spectral density function).

The approach proposed is akin to the likelihood approach proposed by Whittle (1953, 1954) and is often used in time-series modelling. Our approach takes into account spatial correlation, temporal correlation and spatio-temporal interaction. In Section 2, we briefly outline the earlier time domain approaches for the estimation of

DOI: 10.1111/jtsa.12069

parameters and their limitations. In Sections 3 and 4, we describe the frequency domain approach and study the asymptotic sampling properties of the estimators thus obtained. Simulation results are discussed in Section 5. Using the methods proposed here, we consider the estimation of the parameters of three well-defined spatio-temporal covariance functions and use them to model the covariance function of the Pacific wind speed data earlier considered by Cressie and Huang (1999).

2. NON-SEPARABLE CLASS OF COVARIANCES AND THE ESTIMATION

We briefly describe the class of covariance functions proposed by Cressie and Huang (1999) and extended by Gneiting (2002) and the methods of their estimation. As pointed out by several authors, it is non-trivial to construct non-separable class of covariances, which are positive semi-definite. Important developments in the construction of spatio-temporal covariances are given by Brown *et al.* (2000), Cressie and Huang (1999), Gneiting (2002) and Ma (2003). Gneiting (2002) and Gneiting *et al.* (2007) have proposed a flexible class of non-separable spatio-temporal covariance functions based on generalization of the ideas of Cressie and Huang (1999). It is given by

$$C(\mathbf{h}; u) = \frac{\sigma^2}{\psi(|u|^2)^{d/2}} \phi\left(\frac{\| \mathbf{h} \|^2}{\psi(|u|^2)}\right), \ (\mathbf{h}; u) \in \mathbb{R}^d \times \mathbb{R}, \tag{7}$$

where it is assumed that

- $\phi(z)$ is a completely monotone function of $z \in (0, \infty)$ with $\lim_{z \to 0} \phi(z) = \lim_{z \to \infty} \phi(z) = 0$.
- $\psi(w)$ is a positive function of $w \in (0, \infty)$ with completely monotone derivative. w in turn is a function of parameters, which control the range of the covariance.
- $\sigma^2 > 0$ and $\delta \geq d/2$ are scalar parameters.

For details, we refer to Gneiting (2002) and to the recent article Kent *et al.* (2011). The authors of the latter article point out that in certain circumstances, covariances defined by Gneiting (2002) possess a counter-intuitive dimple, and in some cases, the magnitude of the dimple can be non-trivial. Since we are assuming spatial and spatio-temporal stationarity, we expect that the covariances tend to zero monotonically as the spatial and temporal lags increase. So, one should be careful in the choice of covariance functions. However, in this article, we are primarily interested in the estimation of parameters of *a given* class of covariance functions (or, equivalently, its spectral density function) and not in the choice of these functions.

We now consider the estimation of the spatio-temporal variogram defined by (4). Given the sample $\{Z(\mathbf{s}_i, t), i = 1, 2, \ldots, m, t = 1, 2, \ldots, n\}$ from $\{Z(\mathbf{s}, t)\}$, we define the estimator $2\hat{\gamma}(\mathbf{h}(l); u)$ as

$$2\hat{\gamma}(\mathbf{h}(l); u) = \frac{1}{|\mathrm{N}(\mathbf{h}(l); u)|} \sum\sum_{(i,j,t,t') \in \mathrm{N}(\mathbf{h}(l); u)} \left(Z(\mathbf{s}_i, t) - Z(\mathbf{s}_j, t')\right)^2, \tag{8}$$

where

$$\mathrm{N}(\mathbf{h}(l); u) \equiv \left\{(i, j, t, t') : \mathbf{s}_i - \mathbf{s}_j \in \mathbf{h}(l); |t - t'| = u; (i, j) = 1, 2, \ldots, m\right\}.$$

$|\mathrm{N}(\mathbf{h}(l); u)|$ is the number of distinct elements in the set $\mathrm{N}(\mathbf{h}(l); u); l = 1, 2, \ldots, L; u = 0, 1, \ldots, U$. This follows from the spatial case where the aforementioned estimator, due to Matheron (1963), is known as the *classical variogram estimator*. The study of sampling properties of the aforementioned estimator (such as its variance, sampling distribution, etc.) becomes more difficult even with additional assumption of Gaussianity of the process, for the simple reason that we have to take into account not only spatial dependence but also its temporal dependence. Even if we assume that the process is Gaussian, which implies that the spatial and temporal squared differences

DOI: 10.1111/jtsa.12069

$\big(Z(\mathbf{s}_i, t) - Z(\mathbf{s}_j, t')\big)^2$ are proportional to chi-square variables, the sum in (8) is no longer the sum of independent chi-squares, and as such, the assumptions on which the sampling properties of the aforementioned estimator are based are unrealistic.

We may point out here that Li *et al.* (2007) have derived the asymptotic distribution of the variogram estimate and also given asymptotic expressions for the variances and covariances in terms of the fourth-order moments. These expressions are not easily computable. However, Cressie and Huang (1999) have adapted assumptions extending from Cressie (1985) in case of spatial processes to propose a least squares based parameter fitting criterion. Their proposal can be summarized as follows.

If $2\gamma(\mathbf{h}(l); u|\boldsymbol{\theta})$ is the representative parametric variogram of the spatio-temporal process $Z(\mathbf{s}, t)$, with unknown parameter vector $\boldsymbol{\theta}$, for a chosen spatial distance $\mathbf{h}(l)$, $l = 1, 2, \ldots, L$, they approximate

$$\mathrm{Var}[2\hat{\gamma}(\mathbf{h}(l); u)] \simeq \frac{2(2\gamma(\mathbf{h}(l); u|\boldsymbol{\theta}))^2}{|\mathrm{N}(\mathbf{h}(l); u)|} \tag{9}$$

and hence propose that the parameter vector $\boldsymbol{\theta}$ be estimated by minimizing the following weighted least squares criterion:

$$W(\boldsymbol{\theta}) = \sum_{l=1}^{L}\sum_{u=0}^{U} |\mathrm{N}(\mathbf{h}(l); u)| \left\{ \frac{\hat{\gamma}(\mathbf{h}(l); u)}{\gamma(\mathbf{h}(l); u|\boldsymbol{\theta})} - 1 \right\}^2 . \tag{10}$$

In the following sections, we propose a frequency domain method that circumvents the outlined dependency problems and study the asymptotic properties of the estimators obtained by this method.

3. ESTIMATION OF SPACE–TIME PARAMETERS-FREQUENCY VARIOGRAM APPROACH

Consider the stationary spatio-temporal random process $\{Z(\mathbf{s}_i, t), i = 1, 2, \ldots, m, t = 1, 2, \ldots, n\}$. We further assume that the process is isotropic, and without loss of generality, we assume that the mean is zero with variance and covariances given by

$$\begin{aligned} \mathrm{E}[Z(\mathbf{s}_i, t)] &= 0, \ \text{for all } i \text{ and } t, \\ \mathrm{Var}[Z(\mathbf{s}_i, t)] &= C(0, 0) = \sigma^2 < \infty, \\ \mathrm{E}[Z(\mathbf{s}_i + \mathbf{h}, t + u) Z(\mathbf{s}_i, t)] &= C(\|\mathbf{h}\|, |u|) \quad \text{for all } i, |u| \geq 0. \end{aligned} \tag{11}$$

We may point out that the assumption of spatio-temporal isotropy along with second-order stationarity implies the following symmetry relation (see Gneiting, 2002),

$$C(\|\mathbf{h}\|, |u|) = C(\|-\mathbf{h}\|, |u|) = C(\|\mathbf{h}\|, |-u|) = C(\|-\mathbf{h}\|, |-u|). \tag{12}$$

Note that in the literature, the aforementioned second-order property (12) is referred to as space–time symmetry (see Li *et al.*, 2007; Gneiting *et al.*, 2007). We now define a new spatio-temporal random process $Y_{ij}(t)$,

$$Y_{ij}(t) = Z(\mathbf{s}_i, t) - Z(\mathbf{s}_j, t), \ \text{for each } t = 1, 2, \ldots, n, \tag{13}$$

and for all locations $\mathbf{s}_i, \mathbf{s}_j$ where $\mathbf{s}_i$ and $\mathbf{s}_j$ are pairs that belong to the set $\mathrm{N}(\mathbf{h}_l) = \{\mathbf{s}_i, \mathbf{s}_j; \|\mathbf{s}_i - \mathbf{s}_j\| = \|\mathbf{h}_l\|\}, l = 1, 2, \ldots, L$. Note that if there is any common trend in both series $\{Z(\mathbf{s}_i, t)\}$ and $\{Z(\mathbf{s}_j, t)\}$, the differenced series will be free from trend. Now, we define the finite Fourier transform (FT) of $\{Y_{ij}(t),\ i \neq j\}$ at the frequencies $\omega_k = 2\pi k/n,\ k = 0, 1, \ldots, [n/2]$ as

DOI: 10.1111/jtsa.12069

$$J_{\mathbf{s}_i\mathbf{s}_j}(\omega_k) = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n} Y_{ij}(t)e^{-\mathrm{it}\omega_k}, \tag{14}$$

the second-order periodogram of $\{Y_{ij}(t)\}$ as

$$I_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k) = |J_{\mathbf{s}_i\mathbf{s}_j}(\omega_k)|^2 = \frac{1}{2\pi}\sum_{u=-(n-1)}^{n-1} \hat{c}_{y,ij}(u)e^{-iu\omega_k}, \tag{15}$$

where $\omega_k = \frac{2\pi k}{n}$, $k = 0, 1, 2, \ldots, [n/2]$, are the Fourier frequencies, and $\hat{c}_{y,ij}(u)$ is the sample autocovariance function of time lag u of the stationary series $\{Y_{ij}(t), i \neq j\}$, defined by

$$\hat{c}_{y,ij}(u) = \frac{1}{n}\sum_{t=1}^{n-|u|}(Y_{ij}(t+u) - \bar{Y})(Y_{ij}(t) - \bar{Y}), \qquad \text{for } |u| \le n-1,$$

and $\bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_{ij}(t)$ is the sample mean for the time series $\{Y_{ij}(t)$.

The sampling properties of the finite FT and the periodogram in the case of a second-order stationary process have been thoroughly investigated and reported. For details, see Brillinger (2001), Priestley (1981) and the recent article of Dwivedi and Subba Rao (2011).

From (13) and (14), we obtain

$$\begin{aligned} J_{\mathbf{s}_i\mathbf{s}_j}(\omega_k) &= J_{\mathbf{s}_i}(\omega_k) - J_{\mathbf{s}_j}(\omega_k) \quad \text{and hence} \\ I_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k) &= I_{\mathbf{s}_i}(\omega_k) + I_{\mathbf{s}_j}(\omega_k) - 2Re[I_{\mathbf{s}_i\mathbf{s}_j}(\omega_k)], \end{aligned} \tag{16}$$

where $J_{\mathbf{s}_i}(\omega_k)$ and $J_{\mathbf{s}_j}(\omega_k)$ are finite FTs of the individual series $\{Z(\mathbf{s}_i, t)\}$ and $\{Z(\mathbf{s}_j, t)\}$. The corresponding periodograms are respectively $I_{\mathbf{s}_i}(\omega_k)$ and $I_{\mathbf{s}_j}(\omega_k)$, while $I_{\mathbf{s}_i\mathbf{s}_j}(\omega_k)$ is the cross-periodogram between $\{Z(\mathbf{s}_i, t)\}$ and $\{Z(\mathbf{s}_j, t)\}$. Note that in the preceding equation, we have denoted the cross-periodogram of $\{Z(\mathbf{s}_i, t)\}$ and $\{Z(\mathbf{s}_j, t)\}$ by $I_{\mathbf{s}_i\mathbf{s}_j}(\omega_k)$ (without any comma between $\mathbf{s}_i$ and $\mathbf{s}_j$), while the real valued periodogram of the single series $Y_{ij}(t)$ is denoted by $I_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k)$. In other words, we expressed the periodogram of the univariate time series $\{Y_{ij}(t)\}$ in terms of the periodograms of the individual series $\{Z(\mathbf{s}_i, t)\}$, $\{Z(\mathbf{s}_j, t)\}$ and also their cross-periodogram.

From (16), we obtain

$$\mathrm{E}[I_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k)] = \mathrm{E}[I_{\mathbf{s}_i}(\omega_k)] + \mathrm{E}[I_{\mathbf{s}_j}(\omega_k)] - 2Re\mathrm{E}[J_{\mathbf{s}_i}(\omega_k)J^*_{\mathbf{s}_j}(\omega_k)]. \tag{17}$$

Let $g_{\mathbf{s}_i,\mathbf{s}_j}(\omega, \boldsymbol{\theta})$ denote the second-order spectral density function of the series $\{Y_{ij}(t), i \neq j\}$, which is a function of the parameter vector $\boldsymbol{\theta}$. Then for large n, the expectation in the left-hand side of (17) can be approximated by

$$g_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k, \boldsymbol{\theta}) = 2f(\omega_k, \boldsymbol{\theta}) - 2f_{\|\mathbf{h}\|}(\omega_k, \boldsymbol{\theta}), \tag{18}$$

where $f(\omega_k, \boldsymbol{\theta})$ is the second-order spectral density function of stationary spatial processes $\{Z(\mathbf{s}_i, t); i = 1, 2, \ldots, m\}$ and $f_{\|\mathbf{h}\|}(\omega_k, \boldsymbol{\theta})$ is the cross-spectral density function of the process at locations $\{Z(\mathbf{s}_i, t)\}$ and $\{Z(\mathbf{s}_j, t)\}$, given by

$$f_h(\omega_k, \boldsymbol{\theta}) = \frac{1}{2\pi}\sum_{u=-\infty}^{\infty} c(\mathbf{s}_i - \mathbf{s}_j, u)e^{-iu\omega_k}.$$

DOI: 10.1111/jtsa.12069

In view of the assumption of isotropy, the aforementioned cross-spectral density function reduces to

$$f_h(\omega_k, \boldsymbol{\theta}) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} c(\|\mathbf{s}_i - \mathbf{s}_j\|, u) e^{-iu\omega_k} = f_{\|\mathbf{h}\|}(\omega_k, \boldsymbol{\theta}). \tag{19}$$

As mentioned earlier, the cross-spectral density function $f_h(\omega_k, \boldsymbol{\theta})$ is usually a complex valued function, but the assumption of stationarity and isotropy, $c(\mathbf{s}_i - \mathbf{s}_j, u) = c(\|\mathbf{s}_i - \mathbf{s}_j\|, u)$ and $c(\|\mathbf{s}_i - \mathbf{s}_j\|, u) = c(\|\mathbf{s}_i - \mathbf{s}_j\|, -u)$, implies that $f_{\|\mathbf{h}\|}(\omega_k, \boldsymbol{\theta}_2)$ is a real valued function.

Note that an interesting consequence of (18) is that the spectral density function of $\{Y_{ij}(t)\}$, $g_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k, \boldsymbol{\theta})$ defined in (18) can be interpreted as the frequency domain analogue of the classical semi-variogram (of a spatial process), namely,

$$\begin{aligned} \frac{1}{2} g_{\|\mathbf{h}\|}(\omega_k, \boldsymbol{\theta}) &= f(\omega_k, \boldsymbol{\theta}) - f_{\|\mathbf{h}\|}(\omega_k, \boldsymbol{\theta}) \\ &= f(\omega_k, \boldsymbol{\theta}) \left[1 - f_{\|\mathbf{h}\|}(\omega_k, \boldsymbol{\theta}) / f(\omega_k, \boldsymbol{\theta})\right]. \end{aligned} \tag{20}$$

We call $g_{\|\mathbf{h}\|}(\omega_k, \boldsymbol{\theta})$ (20), frequency variogram (FV). Also note that $I_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k)$, defined in (16), is the corresponding asymptotically unbiased estimator. Let us define

$$W_{\|\mathbf{h}\|}(\omega_k, \boldsymbol{\theta}) = f_{\|\mathbf{h}\|}(\omega_k, \boldsymbol{\theta}) / f(\omega_k, \boldsymbol{\theta}), \tag{21}$$

which lies between [0, 1] for all k and all $\|\mathbf{h}\|$. This measure is similar to the coherency measure used in signal processing and multi-variate time series to study the linear dependence between two series. If they are strongly linearly dependent, of course, the spatial coherency will be close to 1. If $\|\mathbf{h}\| = 0$ obviously, it is equal to one. We may point out here that when the process is separable, the aforementioned ratio $W_{\|\mathbf{h}\|}(\omega_k, \boldsymbol{\theta})$ will be a function of the spatial component only and does not depend on the temporal spectral function. We may also point out that Fuentes (2006) defined coherency function and constructed tests for separability.

By plotting the measure $W_{\|\mathbf{h}\|}(\omega_k, \boldsymbol{\theta})$ for a given $\mathbf{h}$ and at each frequency, one can have an idea in which frequency bands the two spatial series are strongly correlated. Also, averaging over all frequencies (suitably normalized) and plotting these values against Euclidean distances $\|\mathbf{h}\|$, one can have an idea of the spatial distance over which the processes are correlated. Such plots may be helpful in modelling. This measure can be used as part of exploratory data analysis to provide an idea about the range parameter of the spatio-temporal covariance, similar to the use of a variogram plot for modelling spatial processes. These ideas need to be further investigated. Of course, the measure of spatial coherency proposed here needs to be studied further. As we pointed out earlier, our objective in this article is the estimation of the parameters only.

It is well known (see, for example, the books of Priestley (1981), Brillinger (2001) and Brockwell and Davis (1991), and the recent article of Dwivedi and Subba Rao (2011)) that the discrete finite FTs of a stationary process are asymptotically uncorrelated over distinct canonical frequencies and have complex normal distribution (see Brillinger, 2001, Thm 4.4.1), and further, the FTs at distinct frequencies of two random processes are also asymptotically independent, and each have complex normal distribution (see Brillinger, 1983, Ch. 2; Calder and Davis, 1997). In view of this asymptotic property, the vector $\mathbf{J}_{\|\mathbf{h}\|}' = [J_{\|\mathbf{h}\|}(\omega_1), J_{\|\mathbf{h}\|}(\omega_2), \ldots, J_{\|\mathbf{h}\|}(\omega_M)]'$, where $J_{\|\mathbf{h}\|}(\omega_k) = J_{\mathbf{s}_i\mathbf{s}_j}(\omega_k)$ and $M = [n/2]$ are as defined by (14), is distributed as asymptotically multi-variate complex normal with mean zero and variance–covariance matrix with diagonal $[g_{\|\mathbf{h}\|}(\omega_1, \boldsymbol{\theta}), g_{\|\mathbf{h}\|}(\omega_2, \boldsymbol{\theta}), \ldots, g_{\|\mathbf{h}\|}(\omega_M, \boldsymbol{\theta})]'$.

We note that because of the asymptotic independence, the off-diagonal elements are zero. In view of the aforementioned distributional properties, following Whittle (1954, 1953) and Walker (1964) (see also Calder and Davis, 1997), the log likelihood of $\mathbf{J}_{\|\mathbf{h}\|}$ can be shown to be proportional to

  J. Time Ser. Anal. 35: 357–377 (2014)
DOI: 10.1111/jtsa.12069

$$Q_{n,\mathrm{N}(\mathbf{h})}(\boldsymbol{\theta}) = \frac{1}{|\mathrm{N}(\mathbf{h})|} \sum_{(\mathbf{s}_i,\mathbf{s}_j)\in \mathrm{N}(\mathbf{h})} \sum_{k=1}^{M} \left[\log(g_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k,\boldsymbol{\theta})) + \frac{I_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k)}{g_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k,\boldsymbol{\theta})} \right]. \tag{22}$$

Here, $\mathrm{N}(\mathbf{h})$ is the collection of all distinct location pairs $\mathbf{s}_i$ and $\mathbf{s}_j$ such that $\mathrm{N}(\mathbf{h}) = \{\mathbf{s}_i, \mathbf{s}_j : \|\mathbf{s}_i - \mathbf{s}_j\| = \|\mathbf{h}\|\}$. However, $Q_{n,\mathrm{N}(\mathbf{h})}(\boldsymbol{\theta})$ depends on spatial lag $\mathbf{h}$. So we propose a weighted sum of $Q_{n,\mathrm{N}(\mathbf{h})}(\boldsymbol{\theta})$ over all possible Euclidean distance lags as our estimating function. One can choose those lags for which $|\mathrm{N}(\mathbf{h}_l)| \geq 30$, as recommended by Chilès and Delfiner (1999). Assuming that the m spatial locations lead to a finite number L of lag bins, such that each bin has a cardinality $|\mathrm{N}(\mathbf{h}_l)|$, we propose the following criterion as the estimation criterion:

$$Q_n(\boldsymbol{\theta}) = 1/L \sum_{l=1}^{L} Q_{n,\mathrm{N}(\mathbf{h}_l)}(\boldsymbol{\theta}). \tag{23}$$

For estimation, we minimize $Q_n(\boldsymbol{\theta})$ with respect to parameters $\boldsymbol{\theta}$. Here, we have used equal weights $1/L$ for each lag, since $Q_{n,\mathrm{N}(\mathbf{h}_l)}(\boldsymbol{\theta})$ includes weights, $\frac{1}{|\mathrm{N}(\mathbf{h}_l)|}$, proportional to the cardinality of each lag. However, one can use other spatial weighting criteria.

In the next section, we state the asymptotic distributional properties of the spatio-temporal parameter estimators derived from equation (23) for the process $\{Z(\mathbf{s},t); \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{Z}\}$. In defining the aforementioned criterion, (23), we have assumed that the correlation between two distinct location pairs, belonging to the same spatial lag $\mathrm{N}(\mathbf{h})$, say $(\mathbf{s}_i, \mathbf{s}_j)$ and $(\mathbf{s}'_i, \mathbf{s}'_j)$, is negligible. This would certainly result in some loss of efficiency, but the computational gains are substantial, since we avoid the formidable challenge of inverting a huge spatio-temporal covariance matrix.

The ideas involved in proposing (23) are analogous to a composite likelihood function. An early proposal for such an approximate likelihood criterion was suggested by Subba Rao (1970) in the context of non-stationary time-series analysis. Subba Rao (1970) defined this likelihood function as the weighted likelihood function. Several authors have since studied such criteria. In particular, the work of Lindsay (1988) brought this method to widespread attention. Fuentes (2007) has proposed an approximate likelihood method for irregularly based spatial data. For review of the literature on composite likelihood, we refer to the recent article Varin *et al.* (2011) on composite likelihood estimation.

We may point out that a similar method was recently proposed (a time domain approach) for spatio-temporal processes by Bevilacqua *et al.* (2012), which is a generalization of the approach proposed by Curriero and Lele (1999), for spatial processes. They arrived at the likelihood under Gaussianity assumption of the original spatio-temporal process. Further, their time domain criterion (see Bevilacqua *et al.*, 2012, eq 5) depends on weights that rely on spatial distances and temporal differences, which have to be chosen. When working with spatio-temporal data, one usually observes large datasets of time series at a relatively small (and fixed) number of spatial locations, which are restricted because of various technical, geographic and economic reasons. That is, in reality, the spatial domain is generally fixed and cannot be increased. In the present frequency domain approach, we do not have truncation over time and hence avoid the choice of time threshold.

In the following section, we discuss sampling properties of the estimates of the parameters.

4. ASYMPTOTIC CONVERGENCE OF PARAMETER ESTIMATES

Let us denote the true parameter vector by $\boldsymbol{\theta}_0$. We now show that the parameter estimator $\hat{\boldsymbol{\theta}}_n$ obtained by minimizing (23) with respect to the unknown parameter vector $\boldsymbol{\theta}$ converges in probability to the original parameter vector $\boldsymbol{\theta}_0$ as $n \to \infty$. For our proof, we use the well-known lemma based on the Arzela–Ascoli theorem (see, for example, Billingsley, 1968, p. 221). For convenience, state the result in Theorem 1 for any sequence of random functions $Q_n^*(\boldsymbol{\theta})$. Throughout our discussion, we assume that $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$, where Θ is a compact set and $Q_n(\boldsymbol{\theta})$ has a unique minimum.

DOI: 10.1111/jtsa.12069

Theorem 1. Let $\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} Q_n^*(\boldsymbol{\theta})$ and $\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}} Q^*(\boldsymbol{\theta})$, where $Q^*(\boldsymbol{\theta}) = E\left[Q_n^*(\boldsymbol{\theta})\right]$. Suppose that $Q^*(\boldsymbol{\theta})$ has a unique minimum and

1. for every $\boldsymbol{\theta} \in \Theta$, we have $Q_n^*(\boldsymbol{\theta}) \stackrel{\text{a.s.}}{\to} Q^*(\boldsymbol{\theta})$ (pointwise convergence),
2. the parameter space Θ is compact,
3. $Q_n^*(\boldsymbol{\theta})$ is stochastic equicontinuous.

Then $\hat{\boldsymbol{\theta}}_n \stackrel{\text{a.s.}}{\to} \boldsymbol{\theta}_0$ as $n \to \infty$. The same result holds if we replace almost sure convergence by convergence in probability.

For the details of the proof, we refer to Billingsley (1968). In Lemma 3, we use the preceding theorem to show consistency of the estimators. We now introduce some assumptions that are needed for obtaining the asymptotic properties of the estimator $\hat{\boldsymbol{\theta}}_n$.

Assumption 1.

(i) For any $n \in \mathbb{Z}^+$ and $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n \in \mathbb{R}^d$, we have the following α-mixing assumption:

$$\sup_{\substack{\{A \in \sigma[\mathbf{Z}(\mathbf{s},0),\mathbf{Z}(\mathbf{s},-1),\ldots]\} \\ \{B \in \sigma[\mathbf{Z}(\mathbf{s},t),\mathbf{Z}(\mathbf{s},t+1),\ldots]\}}} |P(A \cap B) - P(A)P(B)| \le C|t|^{-\alpha}$$

for some $\alpha > 0$ – which will need to be determined later.

(ii) We assume that all fourth-order moments of $\{Z(\mathbf{s}, t)\}$ exist.

(iii) The covariance and fourth-order cumulants satisfy

$$\sup_{\mathbf{s}_1,\mathbf{s}_2} \sum_r |r||\mathrm{Cov}\{Z(\mathbf{s}_1, 0), Z(\mathbf{s}_2, r)\}| < \infty,$$

$$\sup_{\mathbf{s}_1,\mathbf{s}_2,\mathbf{s}_3,\mathbf{s}_4} \sum_{t_1,t_2,t_3} |t_i||\mathrm{Cum}\{Z(\mathbf{s}_1, 0), Z(\mathbf{s}_2, t_1), Z(\mathbf{s}_3, t_2), Z(\mathbf{s}_4, t_3)\}| < \infty.$$

Lemma 1. Suppose Assumption 1 holds. Then it also holds for the differenced series $Y_{ij}(t) = Z(\mathbf{s}_i, t) - Z(\mathbf{s}_j, t)$.

Let us define the set $S = \{\mathbf{u} = (\mathbf{s}_1, \mathbf{s}_2) : \|\mathbf{s}_1 - \mathbf{s}_2\| = h\}$. Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ and $\mathbf{u}_4$ belong to S, and let $\{Y_{\mathbf{u}_1}(t)\}$, $\{Y_{\mathbf{u}_2}(t)\}$, $\{Y_{\mathbf{u}_3}(t)\}$ and $\{Y_{\mathbf{u}_4}(t)\}$ be the corresponding differenced time series. Let us denote the cross-covariance between $\{Y_{\mathbf{u}_1}(t)\}$ and $\{Y_{\mathbf{u}_2}(t)\}$ by $c_{\mathbf{u}_1,\mathbf{u}_2}(v) = \mathrm{Cov}\{Y_{\mathbf{u}_1}(t), Y_{\mathbf{u}_2}(t+v)\}$ and the fourth-order cumulants of the series $\{Y_{\mathbf{s}_i}(t); i = 1, 2, 3, 4\}$ by $\mathrm{Cum}(Y_{\mathbf{u}_1}(t), Y_{\mathbf{u}_2}(t+v_1), Y_{\mathbf{u}_3}(t+v_2), Y_{\mathbf{u}_4}(t+v_3)) = C_{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\mathbf{u}_4}(v_1, v_2, v_3)$. Let $f_{\mathbf{u},\mathbf{u}}(\omega)$ and $f_{\mathbf{u}_1,\mathbf{u}_2}(\omega)$ and $f_{\mathbf{u}_1,\mathbf{u}_2\mathbf{u}_1,\mathbf{u}_4}(\omega_1, \omega_2, \omega_3)$ denote the corresponding second-order spectra, cross-spectra and cumulant spectra respectively, of the process $\{Y_{\mathbf{s}_i}(t); i = 1, 2, 3, 4\}$ (for definitions, see Brillinger, 2001).

Lemma 2. Suppose Assumption 1 holds. Then define

$$W_n = \sum_l \sum_{\mathbf{u}_l \in S} \sum_{k=1}^{\lfloor n/2 \rfloor} K_{\mathbf{u}_l}(\omega_k) I_{\mathbf{u}_l}(\omega_k), \tag{24}$$

where $K_{\mathbf{u}_l}(.)$ is a bounded continuous function and $I_{\mathbf{u}_l}(.)$ is the periodogram of $Y_{\mathbf{s}_1,\mathbf{s}_2}(t)$. Then we have

(i)

$$\mathrm{E}\left\{\frac{1}{n} W_n\right\} \to \sum_l \sum_{\mathbf{u}_l \in S} \frac{1}{2\pi} \int_0^{\pi} K_{\mathbf{u}_l}(\omega) f_{\mathbf{u}_l,\mathbf{u}_l}(\omega) \mathrm{d}\omega.$$

DOI: 10.1111/jtsa.12069

(ii)

$$\mathrm{Var}\left\{\frac{1}{\sqrt{n}}W_n\right\} \xrightarrow{p} 2\sum_{l}\sum_{l'}\sum_{\mathbf{u}_{1l},\mathbf{u}_{2l}\in S}\frac{1}{2\pi}\int_0^{\pi} K_{\mathbf{u}_{1l}}(\omega)h_{\mathbf{u}_{2l}}(\omega)|f_{\mathbf{u}_l,\mathbf{u}_l}(\omega)|^2\mathrm{d}\omega$$
$$+\sum_{l}\sum_{\mathbf{u}_{1l},\mathbf{u}_{2l}\in S}\left(\frac{1}{2\pi}\right)^2\int_0^{\pi}\int_0^{\pi} K_{\mathbf{u}_{1l}}(\omega_1)K_{\mathbf{u}_{2l}}(\omega_2)f_{\mathbf{u}_{1l},\mathbf{u}_{1l},\mathbf{u}_{2l},\mathbf{u}_{2l}}(\omega_1,-\omega_1,\omega_2)\mathrm{d}\omega_1\mathrm{d}\omega_2$$
$$+\sum_{l}\sum_{l'\neq l}\sum_{\mathbf{u}_l,\mathbf{u}_{l'}}\left(\frac{1}{2\pi}\right)^2\int_0^{\pi}\int_0^{\pi} K_{\mathbf{u}_l}(\omega_1)K_{\mathbf{u}_{l'}}(\omega_2)f_{\mathbf{u}_l,\mathbf{u}_l,\mathbf{u}_{l'},\mathbf{u}_{l'}}(\omega_1,-\omega_1,\omega_2)\mathrm{d}\omega_1\mathrm{d}\omega_2.$$

Proof
By taking expectation on both sides of (24), we obtain

$$\begin{aligned}\mathrm{E}\left\{\frac{1}{n}W_n\right\} &= \sum_{l}\sum_{\mathbf{u}_l\in S}\frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor} K_{\mathbf{u}_l}(\omega_k)\mathrm{E}\{I_{\mathbf{u}_l}(\omega_k)\} \qquad \text{(for large n)}\\ &\simeq \sum_{l}\sum_{\mathbf{u}_l\in S}\left[\frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor} K_{\mathbf{u}_l}(\omega_k)f_{\mathbf{u}_l,\mathbf{u}_l}(\omega_k)\right]\\ &\simeq \sum_{l}\sum_{\mathbf{u}_l\in S}\frac{1}{2\pi}\int_0^{\pi} h_{\mathbf{u}_l}(\omega)f_{\mathbf{u}_{1l},\mathbf{u}_{2l}}(\omega)\mathrm{d}\omega.\end{aligned} \tag{25}$$

The last approximation is obtained by using results on discrete FTs (see Briggs and Henson, 1995, Ch. 2). To obtain an expression for the asymptotic variance, we use the following well-known results (see Brillinger, 2001, Ch. 2 and 3).

(i)

$$\begin{aligned}\mathrm{Cov}\left[|J_{\mathbf{u}_1}(\omega_{k_1})|^2,|J_{\mathbf{u}_2}(\omega_{k_2})|^2\right] =&\mathrm{Cov}(J_{\mathbf{u}_1}(\omega_{k_1}),J_{\mathbf{u}_2}(\omega_{k_2}))\mathrm{Cov}(\overline{J_{\mathbf{u}_1}(\omega_{k_1})},\overline{J_{\mathbf{u}_2}(\omega_{k_2})})\\ &+\mathrm{Cov}(J_{\mathbf{u}_1}(\omega_{k_1}),\overline{J_{\mathbf{u}_2}(\omega_{k_2})})\mathrm{Cov}(\overline{J_{\mathbf{u}_1}(\omega_{k_1})},J_{\mathbf{u}_2}(\omega_{k_2}))\\ &+\mathrm{Cum}(J_{\mathbf{u}_1}(\omega_{k_1}),\overline{J_{\mathbf{u}_1}(\omega_{k_1})},J_{\mathbf{u}_2}(\omega_{k_2}),\overline{J_{\mathbf{u}_2}(\omega_{k_2})}).\end{aligned}$$

(ii)

$$\overline{J_{\mathbf{u}}(\omega_k)} = J_{\mathbf{u}}(\omega_{n-k}).$$

(iii)

$$\begin{aligned}\mathrm{Cov}(J_{\mathbf{u}_1}(\omega_{k_1}),J_{\mathbf{u}_2}(\omega_{k_2})) &= f_{\mathbf{u}_1,\mathbf{u}_2}(\omega_{k_1})\delta_n(k_1-k_2)+O\left(\frac{1}{n}\right),\\ \text{where}\quad \delta_n(k_1-K_2) &= \frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}e^{-i(k_1-k_2)w_k}.\end{aligned} \tag{26}$$

(iv)

$$\begin{aligned}&\mathrm{Cum}(J_{\mathbf{u}_1}(\omega_{k_1}),\overline{J_{\mathbf{u}_1}(\omega_{k_1})},J_{\mathbf{u}_2}(\omega_{k_2}),\overline{J_{\mathbf{u}_2}(\omega_{k_2})})\\ &\qquad=\frac{(2\pi)^2}{n}f_{\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_2}(\omega_{k_1},-\omega_{k_1},\omega_{k_2})+O\left(\frac{1}{n^2}\right).\end{aligned}$$

DOI: 10.1111/jtsa.12069

Using the aforementioned results in the evaluation of $\mathrm{Var}\{\frac{1}{\sqrt{n}}W_n\}$, we obtain the result (ii) of the lemma. □

We now consider the asymptotic properties of the parameter estimator $\hat{\boldsymbol{\theta}}_n$.

Assumption 2. (iv) The parameter space Θ is compact and is such that for all $\boldsymbol{\theta} \in \Theta$, $f_{\mathbf{u}_1,\mathbf{u}_2}(\omega;\boldsymbol{\theta})$ is a well-defined spectral density and $f_{\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_2}(\omega_1,\omega_2,\omega_3;\boldsymbol{\theta})$ a well-defined tri-spectrum.

(v) The true parameter vector $\boldsymbol{\theta_0}$ lies in the interior of Θ.

(vi) $g_{\mathbf{u}_l}(\omega_k,\boldsymbol{\theta})$ is bounded away from zero and infinity.

Define the criterion

$$Q_n(\boldsymbol{\theta}) = \frac{1}{L}\sum_{l=1}^{L}\frac{1}{|\mathrm{N}(\mathbf{h}_l)|}\sum_{\mathbf{u}_l\in\mathrm{N}(\mathbf{h}_l)}\sum_{k=1}^{\lfloor n/2\rfloor}\left\{\log g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}) + \frac{I_{\mathbf{u}_l}(\omega_k)}{g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta})}\right\}. \tag{27}$$

Let $\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}}\ \frac{1}{n}Q_n(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}}\ Q_n^*(\boldsymbol{\theta})$. Also let $\nabla^2 Q_n(\boldsymbol{\theta})$ denote the matrix of second derivatives of $Q_n(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ and $(\nabla^2 Q_n(\boldsymbol{\theta}))^{-1}$ be the corresponding inverse matrix.

Lemma 3. Suppose Assumptions 1 and 2 hold. Let $\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}} Q^*(\boldsymbol{\theta})$, where $Q^*(\boldsymbol{\theta}) = E\left[Q_n^*(\boldsymbol{\theta})\right]$. Then

1. for every $\boldsymbol{\theta} \in \Theta$, we have $Q_n^*(\boldsymbol{\theta}) \xrightarrow{p} Q^*(\boldsymbol{\theta})$ (pointwise convergence),
2. $Q_n^*(\boldsymbol{\theta})$ is stochastic equicontinuous.

Then $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$ as $n \to \infty$. The proof is given in Appendix A.

Now, we can prove asymptotic normality of $\hat{\boldsymbol{\theta}}_n$. For the proof, we also need that the second-order derivative $\nabla^2 Q_n(\boldsymbol{\theta})$ converges uniformly. We omit the proof here, since it is very similar to the preceding proof under the additional assumption that the derivative of $\frac{1}{n}\sum_{k=0}^{[n/2]} g_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k,\boldsymbol{\theta})$, with respect to $\boldsymbol{\theta}$, denoted by $g'_{1n}(\boldsymbol{\theta})$, exists for all n and converges uniformly, say, to $g(\boldsymbol{\theta})$.

Theorem 2. Let Assumptions 1 and 2 be true so that Lemmas 2 and 3 hold. Then we have

$$\begin{gathered}\sqrt{n}(\hat{\boldsymbol{\theta}_n} - \boldsymbol{\theta}_0) \xrightarrow{D} \mathrm{N}(\mathbf{0}, (\nabla^2 Q_n(\boldsymbol{\theta}_0))^{-1} V \nabla^2 Q_n(\boldsymbol{\theta}_0)), \text{ where} \\ V = \lim_{n\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{n}}\nabla Q_n(\boldsymbol{\theta}_0)\right).\end{gathered} \tag{28}$$

Proof

See Appendix B. □

In the following section, we use the method described earlier for the estimation of unknown parameters of a parametric covariance function of simulated spatio-temporal random processes. We consider Gaussian and non-Gaussian processes. The estimates are compared with maximum likelihood (ML) estimates.

5. SIMULATION

For our simulation study, to ascertain the performance of the proposed FV approach, we generated two spatio-temporal datasets with the same covariance function (29). The first series we consider is a multi-variate linear Gaussian AR(1) model, which has the mentioned covariance structure. The second series is non-Gaussian, generated by multiplying two stationary Gaussian random processes with the same covariance structure as before.

DOI: 10.1111/jtsa.12069

The covariance function used for generating the aforementioned two series corresponds to a separable process, and it is given by

$$C(\mathbf{h}, u) = \sigma^2 \frac{\phi^{|u|}}{1-\phi^2} \exp(-\|\mathbf{h}\|/\alpha). \tag{29}$$

We note that the aforementioned covariance function has parameters $\boldsymbol{\theta} = \{\sigma, \phi, \alpha\}'$.

Example 1. Let us the define the vectors

$$\begin{aligned}\mathbf{Z}(t) &= \{Z(\mathbf{s}_1; t), Z(\mathbf{s}_2; t), \ldots, Z(\mathbf{s}_m; t)\}' \qquad \text{for } t = 1, 2, \ldots, n,\\ \mathbf{e}(t) &= \{e(\mathbf{s}_1; t), e(\mathbf{s}_2; t), \ldots, e(\mathbf{s}_m; t)\}',\end{aligned}$$

where $\mathbf{e}(t)$ is a sequence of i.i.d. Gaussian random vectors with mean zero and variance–covariance matrix, $\Sigma_{m\times m} = \{\sigma^2 e^{-\|\mathbf{s}_i - \mathbf{s}_j\|/\alpha}\}$, where $i, j = 1, 2, \ldots, m$. Using these Gaussian vectors, the spatio-temporal series $Z(\mathbf{s}, t)$ is now generated from the AR(1) model

$$\mathbf{Z}(t) = \phi \mathbf{Z}(t-1) + \mathbf{e}(t), \tag{30}$$

where ϕ is a scalar and $|\phi| < 1$. It can easily be shown that the series $Z(\mathbf{s}, t)$ thus generated will have the spatio-temporal covariance function given by the expression (29). For our simulation, we have chosen $\sigma = 1$, $\alpha = 5.38$, $\phi = 0.5$, $m = 289$, $n = 480$ and $L = 5$ in the summation given in (23). So as to assess the sampling properties of the estimates obtained by both the ML and FV methods, we simulated 1000 realizations. The bias and the mean square errors are computed using the following formulae.

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \frac{1}{1000}\sum_{i=1}^{1000}\hat{\boldsymbol{\theta}}_i\\ \text{Bias } (\hat{\boldsymbol{\theta}}) &= \frac{1}{1000}\sum_{i=1}^{1000}\{\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\}\\ \text{MSE } (\hat{\boldsymbol{\theta}}) &= \text{diagonal of } \left\{\frac{1}{999}\sum_{i=1}^{999}\{\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\}\{\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\}'\right\}.\end{aligned} \tag{31}$$

The ML estimates are obtained using the function `optim()` in `R`, and FV estimates obtained by minimizing Q given by (23) (with $L = 5$) using the `nlm` package Pinheiro *et al.* (2012) for the statistical system `R` (R Development Core Team, 2012).

The estimates, their biases and mean squares computed using the preceding formulae are summarized in Table I. The mean square errors of the estimates obtained from maximizing the Gaussian likelihood are smaller than FV estimates as one would expect, and we also note that no assumption of Gaussianity is made in computing FV

Table I. Parameter estimates: linear Gaussian process

Parameter	True value	FV estimates	FV MSE	FV bias	ML estimates	ML MSE	ML bias
σ	1	1.1467	0.0215	0.1467	0.9989	3.5×10^{-5}	−0.001
α	5.38	5.2887	0.0083	−0.0913	5.3787	0.0045	−0.0013
ϕ	0.5	0.4989	9.5×10^{-6}	−0.0011	0.4998	5.8×10^{-6}	−0.0002

DOI: 10.1111/jtsa.12069

estimates. Later, we will also remark on the computational times required. In our computation of ML estimates, we made use of the fact that the underlying model is linear satisfying multi-variate AR(1) model; this became possible because of the separable covariance structure we assumed. It must be pointed that the assumption of separability is very restrictive and unrealistic in real situations, and covariance functions are usually more complex than the ones we assumed here. In our second example, we consider non-Gaussian but separable process.

Example 2. For our second illustration, we generated the spatio-temporal process $Z(\mathbf{s}, t)$ by multiplying two independent Gaussian processes, namely, $Z(\mathbf{s}, t) = X(\mathbf{s})Y(t)$, where spatially dependent random process $X(\mathbf{s})$ is assumed to have mean zero and covariance function.

$\text{Cov}(X(\mathbf{s}_i), X(\mathbf{s}_j)) = \sigma^2.\exp(-\alpha_1 \|\mathbf{s}_j - \mathbf{s}_j\|)$. Thus, the covariance function here is the reparametrized form of (29) with the range parameter α_1 in the exponent. The time series $Y(t)$ is generated from AR(1) model

$$Y(t) = \phi Y(t-1) + e(t), \tag{32}$$

where $e(t)$ is a sequence of i.i.d. Gaussian variables with mean zero mean and variance one. For our simulations, we have chosen $m = 49$, $n = 400$, $\sigma = 1$, $\alpha_1 = 3$ and $\phi = 0.5$. We simulated 1000 realizations as before, and estimates and their mean square errors are computed using the formulae given earlier. The estimates obtained by the ML and FV methods are given in Table II. The variance–covariance matrices of these estimates are given in Table III. We note that if we define the vector $\mathbf{Z}(t)$ by stacking all n time series observed at all the m locations, the dimension of the covariance matrix will be $mn \times mn$ (here $mn = 19,600$). The inversion of this huge matrix may cause problems to the calculation of the ML estimates. This matrix may be really huge when the number of spatial locations and/or time points is large. However, in view of our assumption that the process is separable, we can achieve significant simplifications by using the fact that the covariance matrix Σ can be written as a Kronecker product, $\Sigma = S(\theta) \otimes T(\theta)$, of two smaller dimensional matrices.

We note from Table III of the variance–covariance matrices that the trace of the matrix of the FV method, if taken as our measure, is smaller than the trace of the ML method. We note that the ML estimates are computed on the basis of the likelihood obtained under Gaussianity assumption even though the process we generated is not Gaussian. Therefore, it is not surprising that the estimates obtained by the ML method are not that efficient. The estimation based on FV, similar to Example 1, is robust against any departure from Gaussianity. The following further remarks are in order.

Table II. Estimates of the parameters: non-Gaussian processes

		ML estimates			FV estimates		
Parameters	Original	Estimate	Bias	MSE	Estimate	Bias	MSE
σ	1	0.9995	−0.0006	0.0003	0.9957	−0.0043	0.0145
α_1	3	3.4999	0.4999	5.5040	3	1.9×10^{-07}	1.81×10^{-11}
ϕ	0.5	0.4968	−0.0032	0.0019	0.4953	−0.0047	0.0019

Table III. Variance–covariance matrices: non-Gaussian processes

	ML estimates			FV estimates		
Parameters	σ	α_1	ϕ	σ	α_1	ϕ
σ	0.0003	0.0026	-1.7×10^{-05}	0.0145	-8.3×10^{-08}	0.0002
α_1	0.0026	5.2539	−0.0004	-8.3×10^{-08}	1.8×10^{-11}	-1.9×10^{-09}
ϕ	-1.7×10^{-05}	−0.0004	0.0019	0.0002	-1.9×10^{-09}	0.0019

DOI: 10.1111/jtsa.12069

Remark 1. The simulations were run on a Windows 7 operating system with 32-GB RAM and Xeon processor (Microsoft, Redmond, WA). To compute ML estimates, it took more than 4 days, and it had to run in four parallel sessions to enhance the speed. The simulation times were computed using the function Sys.time () in R. These timings could be improved by programming the computation of the likelihood in a lower level language, but the critical issue is that the Gaussian likelihood involves a general covariance matrix of size $mn \times mn$, which is bound to become impractical for large numbers of locations and/or time points.

Remark 2. Further, in this particular simulation study, the separable covariance matrix is a Kronecker product of spatial and temporal covariances. Thus, we were able to exploit the algebra of Kronecker products to enhance the computational speed of the Gaussian likelihood. In absence of such a structure (as would be the case for non-separable processes), the huge computational cost of the likelihood makes the ML estimation prohibitively slow. The computations were carried out in `R` using the general optimization functions `optim(.)` and `nlm(.)`.

6. APPLICATION TO WIND SPEED DATA

The data provide the record of east–west wind speed on a 17×17 rectangular lattice at grid spacings of 210 km, every 6 h from November 1992 to February 1993. So the process is observed at 289 locations and 480 time points (that is, $m = 289$ and $n = 480$).

Before parameter estimation, we check if the data are weakly stationary. In Figure 1, the spatial and temporal 'mean against standard deviation' plots for the wind speed data are given to check the assumption of heteroscedasticity. The figures do not indicate any particular pattern, and thus an assumption of homoscedasticity may be justified as noted by Cressie and Huang (1999). We observe the plots of spatial and temporal means to look for the presence of any deterministic trend. Note that the spatial and temporal sample averages are defined as follows:

$$\begin{aligned} \bar{Z}(\mathbf{s}_i;.) &= \frac{1}{n}\sum_{t=1}^{n} Z(\mathbf{s}_i,t), \text{ for } i = 1,2,\ldots,m, \\ \bar{Z}(.;t) &= \frac{1}{m}\sum_{i=1}^{m} Z(\mathbf{s}_i,t), \text{ for } t = 1,2,\ldots,n. \end{aligned} \quad (33)$$

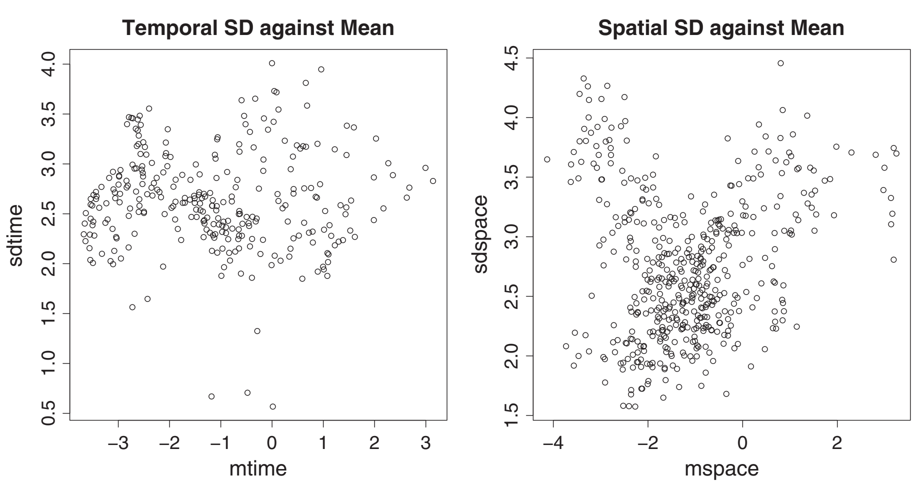

Figure 1. Mean against standard deviation plots. The left panel plots the temporal standard deviation against mean. The right panel plots the spatial standard deviation against mean

DOI: 10.1111/jtsa.12069

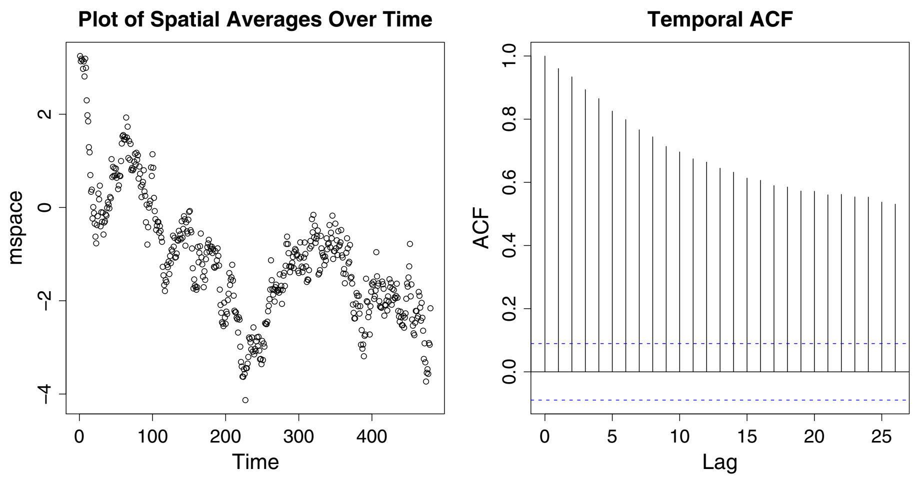

Figure 2. The spatial averages of the original wind speed data, $\bar{Z}(.;t)$, against time and plot of sample ACF $\hat{\rho}_{(u)}$

In Figure 2, the spatial averages (over all locations) displayed against time points and the corresponding temporal sample autocorrelation [autocorrelation function (ACF)] plot for the series are displayed. Here, we define the sample ACF at lag u by $\hat{\rho}_t(u)$ as

$$\hat{\rho}(u) = \frac{1}{n}\sum_{t=1}^{n-u}(\bar{Z}(.;t+u) - \bar{\bar{Z}})(\bar{Z}(.;t) - \bar{\bar{Z}})/V, \qquad u \geq 0, \tag{34}$$

where

$$\bar{\bar{Z}} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n} Z(\mathbf{s}_i, t) \qquad \text{and}$$

$$V = \frac{1}{n-1}\sum_{t=1}^{n}\left\{\bar{Z}(.;t+u) - \frac{1}{n}\sum_{t=1}^{n}\bar{Z}(.;t+u)\right\}^2.$$

The temporal averages are displayed on their corresponding locations in the 3D image of Figure 3.

Cressie and Huang (1999) assumed spatial and temporal second-order stationarity for the Pacific wind data. However, from the mean and ACF plots of Figure 2, it is clear that there is a long-term temporal deterministic trend in the wind speed data. The 3D spatial plot of Figure 3 has a cascading shape with the height decreasing from the west to east direction of the observation domain, which indicates the presence of a spatial deterministic trend as well. The plots show that there may be a long-term trend present in the data.

To remove the trend, we subtract the temporal averages of each location from the respective time series. We denote the adjusted data by $Z^*(\mathbf{s}_i, t)$ defined as

$$Z^*(\mathbf{s}_i, t) = Z(\mathbf{s}_i, t) - \bar{Z}(\mathbf{s}_i;.); \qquad \text{for } i = 1, 2, \ldots, m. \tag{35}$$

The respective adjusted means are denoted by $\bar{Z}^*(\mathbf{s}_i, .)$ and $\bar{Z}^*(., t)$. The mean plots are given in Figures 2 and 3.

From Figure 4, we observe that the deterministic temporal trend has been removed from the observations $Z^*(\mathbf{s}_i, t)$. The 3D plot of temporal averages in Figure 5 shows that the cascading effect has been removed. From now on, we treat $Z^*(\mathbf{s}, t)$ as a second-order stationary spatio-temporal process and denote it by $Z(\mathbf{s}; t)$.

DOI: 10.1111/jtsa.12069

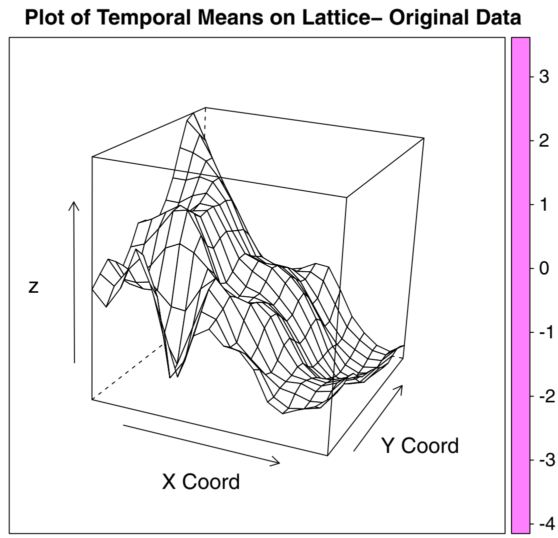

Figure 3. 3D image of the temporal averages of the original wind speed data at the corresponding locations on the lattice grid points

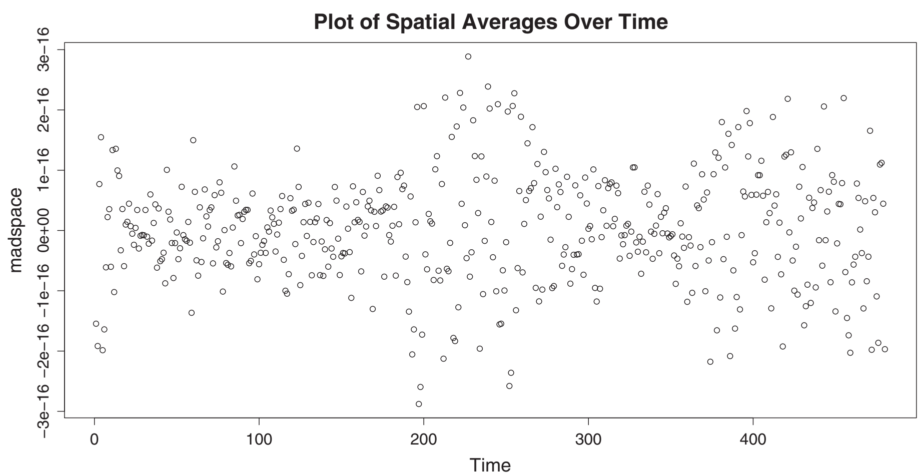

Figure 4. The spatial averages of the wind speed data, $\bar{Z}^*(., t)$, are plotted against time

We now fit three covariance models, given later, to the spatio-temporal process $Z(\mathbf{s}; t)$. The first two models are non-separable spatio-temporal second-order stationary covariance functions chosen from Cressie and Huang (1999), while the third is a generalized version of the first model obtained by Gneiting (2002). All the three covariance functions are convex functions, chosen based on the spatio-temporal sample variogram (see Cressie and Huang, 1999). In all these models, 'a' is the temporal scale parameter and b^2 is the spatial range parameter. Parameter 'g' in Model-3 is the non-separability parameter. Cressie and Huang (1999) have discussed that variogram of the wind speed data may have discontinuity at origin. To incorporate this 'nugget' (see Cressie, 1993, Ch. 2) effect, we have also included a parameter τ. Following Cressie and Huang (1999), a purely spatial covariance is also incorporated to address the fact that the empirical spatial variogram does not change shape at larger

DOI: 10.1111/jtsa.12069

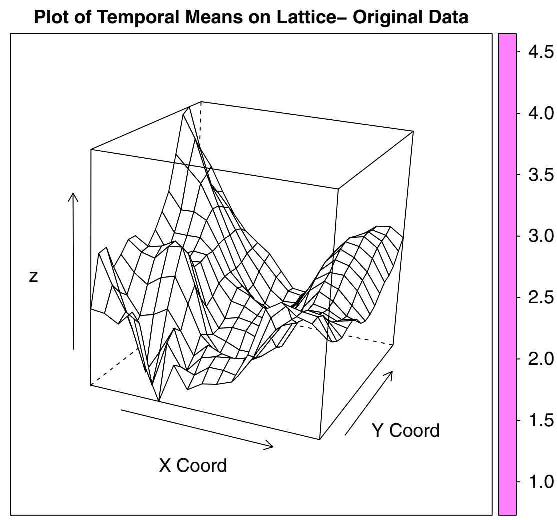

Figure 5. 3D image of the temporal averages, $\bar{Z}^*(\mathbf{s}_i, .)$, at various locations

Table IV. Estimates of the parameters

Models	Parameter estimates							
	σ	a	b	τ	a_1	b_1	g	Min $Q_n(\boldsymbol{\theta})$
Model-1	2.406	0.388	7.623	0.100	0.012	0.000	0.000	78,539.704
Model-2	2.404	0.383	5.149	0.100	0.006	0.000	0.000	110,423.484
Model-3	2.405	0.390	5.954	0.100	0.010	0.998	1.005	78,665.974

temporal lags. Since we have carried out a trend adjustment on the original data, we choose the second-order stationary exponential variogram, instead of the power variogram used by Cressie and Huang (1999). The parameter estimates are given in Table IV.

$$\text{Model-1: } C(\|\mathbf{h}\|; |u|) = \tau + \sigma^2 \frac{1}{a|u|+1} e^{-\frac{b^2\|\mathbf{h}\|^2}{a|u|+1}} + e^{-a_1\|\mathbf{h}\|} \tag{36}$$

$$\text{Model-2: } C(\|\mathbf{h}\|; |u|) = \tau + \sigma^2 \frac{a|u|+1}{\{(a|u|+1)^2 + b^2\|\mathbf{h}\|^2\}^{3/2}} + e^{-a_1\|\mathbf{h}\|} \tag{37}$$

$$\text{Model-3: } C(\|\mathbf{h}\|; |u|) = \tau + \sigma^2 \frac{1}{a|u|+1} e^{-\frac{b^2\|\mathbf{h}\|^{2g}}{(a|u|+1)^{b_1 g}}} + e^{-a_1\|\mathbf{h}\|}. \tag{38}$$

We noted that the Pacific wind speed data were observed on a 17×17 rectangular lattice. We obtain estimates by minimizing (23) for the first 13 vertical lag distances. The reported estimates are the sample means of these estimates. For more details on computation of lags for the data, see Cressie and Huang (1999). Note that the aforementioned covariance functions have finite spectral density function but do not have closed form expressions. Therefore, we have used the finite FT. We used the `fft(.)` routine in the `stats` package in `R` (R Core Team, 2002).

Based on the minimum values of $Q_n(\boldsymbol{\theta})$, we recommend the use of the covariance Model-1 for the transformed data among the three models, though there is not much significant difference between 1 and 3. A further analysis, such as cross-validation, may be further necessary to differentiate between these two. Since we used

DOI: 10.1111/jtsa.12069

transformed data, comparing our estimates or minimum values with those obtained by Cressie and Huang (1999) is not appropriate.

7. CONCLUSION

The method of estimation proposed here is based on discrete FTs of the stationary processes. We exploited the interesting properties of these transforms evaluated at canonical frequencies so as to obtain a likelihood function for the maximization as is often performed in time series. As we noticed, the advantage of these transforms is that they are approximately uncorrelated (in the case of Gaussian processes, they are independent) even though the original processes are non-Gaussian. Since the estimation is dependent on discrete FTs, calculated at each location, our analysis is based on these complex valued realizations at each location. In doing so, our methodology depends on spatial parameters only (though is an implicit function of time). As we have seen, the asymptotic properties of the estimates can be obtained under fairly general conditions, and this was not possible (at least not easy) if we use the spatio-temporal domain approaches suggested earlier. The practical estimation method depends on obtaining the FTs, and the minimization can easily be performed using routines readily available in standard software (such as R) as is used in the time-series literature (may need minor changes).

With an extensive literature, ML estimation in conjunction with the assumption of Gaussianity remains perhaps the most used statistical estimation method in applications. However, because of the rapid development of data storage facilities, massive covariance matrices, as in the aforementioned examples, are becoming a common feature of many scientific investigations. Such complex and big data call to question the classical assumptions of Gaussianity and ML as a method of estimation both from distribution and computational points of view. In this article, we have proposed a Fourier domain method of estimation for large spatio-temporal datasets, which is comparable in terms of efficiency and faster computationally and does not need specification of distributions.

ACKNOWLEDGEMENTS

The research of the second author (Das) was pursued at the University of Manchester as part of his PhD thesis with sponsorship from British Council under the UKIERI research grant. The authors thank two referees and the associate editors whose comments have been very helpful in improving the manuscript. Researchers Das and Subba Rao were also partially supported by a grant from the Department of Science and Technology, Government of India (no. SR/S4/516/07) while visiting the CRRAO AIMSCS, Hyderabad, India. The second author (Das) received partial funding for the project from the J C Bose Fellowship of Prof. Arup Bose at the Indian Statistical institute as a visiting scientist. Part of the research reported in this article was carried out when one of the authors (Subba Rao) was visiting the University of Debrecen, Debrecen, Hungary. The visit was partially supported by a grant by the European Union, co-financed by the European Social Fund with project number TÁMOP-4.2.2.C-11/1/KONV-2012-0001. The authors are thankful to Dr Jyotishman Bhowmick of University of Oslo, Norway, and Dr Suhasini Subbarao of Texas A&M University, USA.

REFERENCES

Bevilacqua M, Gaetan C, Mateu J, Porcu E. 2012. Estimating space and space–time covariance functions for large data sets: A weighted composite likelihood approach. *Journal of the American Statistical Association* **107**(497): 268–280.

Billingsley P. 1968. *Convergence of Probability Measures*. New York: Wiley.

Briggs WL, Henson VE. 1995. *The DFT: An Owner's Manual for the Discrete Fourier Transform. 45*. Philadelphia, PA: Society for Industrial Mathematics.

Brillinger DR. 1983. *Time series in the frequency domain*. In *Handbook of Statistics*, Brillinger DR, Krishnaiah PR (eds.) Amsterdam: North-Holland, pp. 1 –485.

Brillinger DR. 2001. *Time Series: Data Analysis and Theory*. Philadelphia: Society for Industrial Mathematics.

DOI: 10.1111/jtsa.12069

Brockwell PJ, Davis RA. 1991. *Time Series: Theory and Methods*. New York: Springer-Verlag.
Brown P.E., Roberts G.O., Kåresen K.F., Tonellato S. 2000. Blur-generated non-separable space–time models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**(4): 847–860.
Calder M, Davis RA. 1997. Introduction to Whittle (1953). The analysis of multiple stationary time series. *Breakthroughs in Statistics* **3**: 141–148.
Chilès JP, Delfiner P. 1999. *Modeling Spatial Uncertainty*, Geostatistics, Wiley Series in Probability and Statistics. New York: Wiley Interscience.
Cressie N. 1985. Fitting variogram models by weighted least squares. *Mathematical Geology* **17**(5): 563–586.
Cressie N, Huang HC. 1999. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association* **94**(448): 1330–1339.
Cressie NAC. 1993. *Statistics for Spatial Data*. New York: John Wiley & Sons.
Cressie NAC, Wikle CK. 2011. *Statistics for Spatio-temporal Data*. New York: John Wiley & Sons.
Curriero FC, Lele S. 1999. A composite likelihood approach to semivariogram estimation. *Journal of Agricultural, Biological and Environmental Statistics* **4**(1): 9–28.
Dwivedi Y, Subba Rao S. 2011. A test for second-order stationarity of a time series based on the discrete fourier transform. *Journal of Time Series Analysis* **32**(1): 68–91.
Fuentes M. 2006. Testing for separability of spatial–temporal covariance functions. *Journal of Statistical Planning and Inference* **136**(2): 447–466.
Fuentes M. 2007. Approximate likelihood for large irregularly spaced spatial data. *Journal of the American Statistical Association* **102**(477): 321–331.
Gneiting T. 2002. Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association* **97**(458): 590–600.
Gneiting T, Genton MG, Guttorp P. 2007. Geostatistical space–time models, stationarity, separability, and full symmetry. *Monographs On Statistics and Applied Probability* **107**: 151–175.
Kent JT, Mohammadzadeh M, Mosammam AM. 2011. The dimple in Gneiting's spatial–temporal covariance model. *Biometrika* **98**(2): 489–494.
Lee J, Subba Rao S. 2011. A note on general quadratic forms of nonstationary time series. *Technical Report*, *Department of Statistics, Texas A and M University College Station*, U.S.A.
Li B, Genton MG, Sherman M. 2007. A nonparametric assessment of properties of space–time covariance functions. *Journal of the American Statistical Association* **102**(478): 736–744.
Lindsay BG. 1988. Composite likelihood methods. *Contemporary Mathematics* **80**(1): 221–39.
Ma C. 2002. Spatio-temporal covariance functions generated by mixtures. *Mathematical Geology* **34**(8): 965–975.
Ma C. 2003. Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference* **116**(2): 489–501.
Matheron G. 1963. Principles of geostatistics. *Economic Geology* **58**(8): 1246–1266.
Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. 2012. *nlme: Linear and Nonlinear Mixed Effects Models.* R package version 3.1-104. Vienna, Austria: R Foundation for Statistical Computing.
Priestley MB. 1981. *Spectral Analysis and Time Series*. London: Academic press.
R Core Team. 2002. *The R Stats Package*. Vienna, Austria: R Foundation for Statistical Computing. Available from: http://www.R-project.org/.
R Development Core Team. 2012. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: http://www.R-project.org/.
Sherman M. 2010. *Spatial Statistics and Spatio-temporal Data: Covariance Functions and Directional Properties*. New York: John Wiley & Sons.
Subba Rao T. 1970. The fitting of non-stationary time-series models with time-dependent parameters. *Journal of the Royal Statistical Society. Series B (Methodological)* **32**(2): 312–322.
Varin C, Reid N, Firth D. 2011. An overview of composite likelihood methods. *Statistica Sinica* **21**(1): 5–42.
Walker AM. 1964. Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *Journal of the Australian Mathematical Society* **4**(03): 363–384.
Whittle P. 1953. Estimation and information in stationary time series. *Arkiv för matematik* **2**(5): 423–434.
Whittle P. 1954. On stationary processes in the plane. *Biometrika* **41**(3-4): 434–449.

APPENDIX A: PROOF OF LEMMA 3

Proof

Because of boundedness of $g(\omega;\boldsymbol{\theta})$ (Assumption 2) for all $\boldsymbol{\theta}$ and ω, we have pointwise convergence of $Q_n(\boldsymbol{\theta})$ (by Lemma 2), that is,

  J. Time Ser. Anal. **35:** 357–377 (2014)
DOI: 10.1111/jtsa.12069

$$\frac{1}{n}Q_n(\boldsymbol{\theta}) \xrightarrow{p} \sum_l \sum_{\mathbf{u}_l \in \mathrm{N}(\mathbf{h})} \int \left[\log g_{\mathbf{u}_l}(\omega;\boldsymbol{\theta}) + \frac{g_{\mathbf{u}_l}(\omega;\boldsymbol{\theta}_0)}{g_{\mathbf{u}_l}(\omega;\boldsymbol{\theta})} \right] \mathrm{d}\omega.$$

We have earlier assumed that the parameter space Θ is compact. Proving conditions (1) and (2) is equivalent to proving equicontinuity in probability. To prove that, we note that from mean value theorem, we have

$$\left| \frac{1}{n}Q_n(\boldsymbol{\theta}_1) - \frac{1}{n}Q_n(\boldsymbol{\theta}_2) \right| = \left| \frac{1}{n}\nabla Q_n(\check{\boldsymbol{\theta}})(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \right|, \tag{A1}$$

where $\check{\boldsymbol{\theta}}$ lies in the interval $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. We observe that

$$\frac{1}{n}\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \theta_i} = \sum_l \sum_{\mathbf{u}_l} \frac{1}{n} \sum_{k=1}^{\lfloor n/2 \rfloor} \left[\frac{1}{g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta})} \frac{\partial g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta})}{\partial \theta_i} - \frac{I_{\mathbf{u}_l}(\omega_k)}{(g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}))^2} \frac{\partial g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta})}{\partial \theta_i} \right]. \tag{A2}$$

Now, under the assumption that $g_{\mathbf{u}_l}(.)$ is bounded, from the preceding equation, we obtain

$$\frac{1}{n}\frac{\partial Q_n(\boldsymbol{\theta})}{\partial \theta_i} \leq \sum_l \sum_{\mathbf{u}_l} \left[C + \frac{C}{n} \sum_{k=1}^{\lfloor n/2 \rfloor} I_{\mathbf{u}_l}(\omega_k) \right] = M_n, \text{ for some } C > 0. \tag{A3}$$

From Lemma 2, it follows that the expectation of M_n tends to (as $n \to \infty$)

$$\mathrm{E}[M_n] \to C \sum_l \sum_{\mathbf{u}} \left[1 + \frac{1}{2\pi} \int f_{\mathbf{u}_l,\mathbf{u}_l}(\omega)\mathrm{d}\omega \right]$$

and that

$$\mathrm{Var}(M_n) \to 0. \text{In fact it can be shown to be} \mathrm{Var}(M_n) = O\left(\frac{1}{n}\right).$$

This implies that

$$\left| \frac{1}{n}Q_n(\boldsymbol{\theta}_1) - \frac{1}{n}Q_n(\boldsymbol{\theta}_2) \right| \leq [\mathrm{E}(M_n) + o_p(1)]\, |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| \tag{A4}$$

and hence equicontinuity in probability. Thus, by Theorem 1, we have convergence in probability. □

APPENDIX B: PROOF OF THEOREM 2

Proof

Since $\nabla Q_n(\boldsymbol{\theta})$ is a vector, we consider a Taylor expansion pointwise on $\nabla Q_n(\boldsymbol{\theta})$. Thus, pointwise, we have by the mean value theorem

$$\frac{1}{n}\nabla Q_n(\hat{\boldsymbol{\theta}}) \,|_i = \frac{1}{n}\nabla Q_n(\boldsymbol{\theta}_0) \,|_i + \left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \frac{1}{n}\nabla^2 Q_n\left(\check{\boldsymbol{\theta}}_n\right) |_i, \tag{B1}$$

where $\check{\boldsymbol{\theta}}_n$ lies in $(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0)$. Note that the preceding expression is a scalar. Now, using the uniform convergence of $\nabla^2 Q_n(\boldsymbol{\theta})$, we have

DOI: 10.1111/jtsa.12069

$$\sup_{\boldsymbol{\theta}\in\Theta} \mid \frac{1}{n}\nabla^2 Q_n\left(\hat{\boldsymbol{\theta}}\right) - \nabla^2 Q(\boldsymbol{\theta}) \mid\to 0, \text{ where}$$
$$\nabla^2 Q(\boldsymbol{\theta}) = \lim_{n\to\infty} \mathrm{E}\left(\frac{1}{n}\nabla^2 Q_n(\hat{\boldsymbol{\theta}})\right). \tag{B2}$$

It is easy to calculate this limit using Lemma 2. This implies that $\frac{1}{n}\nabla^2 Q_n(\check{\boldsymbol{\theta}}_n) \to \nabla^2 Q(\boldsymbol{\theta}_0)$. Hence,

$$\frac{1}{n}\nabla Q_n(\hat{\boldsymbol{\theta}}) \mid_i = \frac{1}{n}\nabla Q_n\left(\boldsymbol{\theta}_0\right) \mid_i + \left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right)\nabla^2 Q(\boldsymbol{\theta}_0) \mid_i + o_p\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \mid_i,$$

and this in turn implies

$$\frac{1}{n}\nabla Q_n(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\nabla Q_n(\boldsymbol{\theta}_0) + \nabla^2 Q(\boldsymbol{\theta}_0)'\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) + o_p\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right). \tag{B3}$$

We note that the left-hand side of (B3) is zero for the optimum value $\hat{\boldsymbol{\theta}}$. Also note that $\nabla^2 Q(\boldsymbol{\theta}_0)$ is a deterministic quantity. So for proving asymptotic normality of $\hat{\boldsymbol{\theta}}_n$, we need to show asymptotic normality of $\frac{1}{\sqrt{n}}\nabla Q_n(\boldsymbol{\theta}_0)$. Recall from equation (A3) that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\nabla Q_n(\boldsymbol{\theta}_0)\\
&= \sum_l \sum_{\mathbf{u}_l} \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor n/2\rfloor} \left[\frac{1}{g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)}\nabla g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0) - \frac{I_{\mathbf{u}_l}(\omega_k)}{g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)^2}\nabla g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)\right]\\
&= \sum_l \sum_{\mathbf{u}_l} \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor n/2\rfloor} \left[\mathrm{E}\left[I_{\mathbf{u}_l}(\omega_k)\right] - I_{\mathbf{u}_l}(\omega_k)\right] \frac{\nabla g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)}{g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)^2}\\
&- \sum_l \sum_{\mathbf{u}_l} \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor n/2\rfloor} \left[\mathrm{E}\left[I_{\mathbf{u}_l}(\omega_k)\right] - g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)\right] \frac{\nabla g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)}{g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)^2},\\
&= I + II.
\end{aligned} \tag{B4}$$

The term II is deterministic, which is the bias. It is known that $|I_{\mathbf{u}}(\omega_k) - g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)| \le \frac{K}{n}$. Thus, $I_{\mathbf{u}}(\omega_k)$ and is $O(1/\sqrt{n})$. Thus, the preceding equation reduces to

$$\frac{1}{\sqrt{n}}\nabla Q_n(\boldsymbol{\theta}_0) = \sum_l \sum_{\mathbf{u}_l} \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor n/2\rfloor} \left\{\mathrm{E}\left[I_{\mathbf{u}_l}(\omega_k)\right] - I_{\mathbf{u}_l}(\omega_k)\right\} \frac{\nabla g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)}{g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)^2} + O\left(\frac{1}{\sqrt{n}}\right). \tag{B5}$$

Hence, we only need to show the asymptotic normality of I. We have

$$I = \sum_l \sum_{\mathbf{u}_l} \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor n/2\rfloor} \left[\mathrm{E}\left[I_{\mathbf{u}_l}(\omega_k)\right] - I_{\mathbf{u}_l}(\omega_k)\right] H_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0),$$

where

$$H_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0) = \frac{\nabla g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)}{g_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)^2}. \tag{B6}$$

  J. Time Ser. Anal. **35:** 357–377 (2014)
DOI: 10.1111/jtsa.12069

We assume $H_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0)$ is smooth and twice differentiable with respect to ω; therefore, we can write

$$\begin{aligned} I &= \sum_l \sum_{\mathbf{u}_l} \frac{1}{\sqrt{n}} \sum_t \sum_\tau \left[Y_{\mathbf{u}_l}(t)Y_{\mathbf{u}_l}(\tau) - \mathrm{E}(Y_{\mathbf{u}_l}(t)Y_{\mathbf{u}_l}(\tau))\right] \frac{1}{n} \sum_{k=1}^{n} H_{\mathbf{u}_l}(\omega_k;\boldsymbol{\theta}_0) e^{i\omega_k(t-\tau)} \\ &= \sum_{\mathbf{u}_l} \frac{1}{\sqrt{n}} \sum_t \sum_\tau \left[Y_{\mathbf{u}_l}(t)Y_{\mathbf{u}_l}(\tau) - \mathrm{E}(Y_{\mathbf{u}_l}(t)Y_{\mathbf{u}_l}(\tau))\right] K_{\mathbf{u}_l}(t-\tau) + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{B7}$$

where $K_{\mathbf{u}}(\tau) = \int_{-\infty}^{\infty} H_{\mathbf{u}}(\omega;\boldsymbol{\theta}_0) e^{i\omega(\tau)} \mathrm{d}\omega$. In view of the assumption that $H_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)$ is differentiable twice with respect to ω, it follows that the impulse response sequences $\{K_{\mathbf{u}_l}(\tau)\}$ must decay to zero at the rate $\frac{1}{|\tau|^2}$ (see Briggs and Henson, 1995, Ch. 6). Note that $\{Y_{\mathbf{u}_l}(t)\}$ are α mixing at the rate specified in Assumption 1. Then by Theorem 2.2 of Lee and Subba Rao (2011), we can establish the asymptotic normality of $\frac{1}{\sqrt{n}} \nabla Q_n(\boldsymbol{\theta}_0)$. That is,

$$\begin{aligned} &\frac{1}{\sqrt{n}} \nabla Q_n(\boldsymbol{\theta}_0) \xrightarrow{D} \mathrm{N}(0, V), \text{ where} \\ &V = \lim_{n\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{n}} \nabla Q_n(\boldsymbol{\theta}_0)\right). \end{aligned} \tag{B8}$$

An expression for V can be deduced from Lemma 2. The aforementioned result together with equation (B3) gives

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{D} \mathrm{N}\left(\mathbf{0}, \nabla^2 Q_n^{-1}(\boldsymbol{\theta}_0) V \nabla^2 Q_n(\boldsymbol{\theta}_0)\right). \tag{B9}$$

□

DOI: 10.1111/jtsa.12069