A Frequency domain approach for the estimation of parameters of spatio-temporal stationary random processes

Subba Rao, Tata and Das, Sourav and Boshnakov, Georgi N.

2012

MIMS EPrint: **2012.90**

ISSN 1749-9097

A Frequency domain approach for the estimation of parameters of spatio-temporal stationary random processes

Tata Subba Rao*, Sourav Das†and Georgi Boshnakov‡

September 24, 2012

Abstract

In this paper we consider the estimation of the parameters of the spatio-temporal covariances of spatio-temporal stationary random processes. We define Finite Fourier Transforms of the processes at each location and based on joint distribution of these complex valued random variables we define an approximate likelihood function and consider the maximization. Ideas are similar to Whittle likelihood function considered in time series. The sampling properties of the estimators are investigated. The method is applied to simulated data and also to pacific wind speed data considered earlier by Cressie and Huang [6].

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 which is 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 Once an extra dimension, like time, is introduced the available methodology is no longer applicable and any method developed should not only take into account spatial and temporal dependencies but also their interaction. The literature on spatio-temporal processes is is a bit sparse compared to the literature on spatial processes. Recent books by Cressie and Wikle [8] and Sherman[19] should help to fill in this gap. In the following we briefly introduce the notation and summarise 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 $\{Z(\mathbf{s}_i, t_j) : i = 1, 2, ..., m; j = 1, 2, ..., n\}$. Let S denote the set of ordered pairs $S = \{(i, j) : i = 1, 2, ..., m; j = 1, 2, ..., n\}$. For convenience of notation let us define a one-one mapping from $S \to \mathbb{N}$ enumerating the elements of the set S from $(1, 1), (1, 2), ..., (m, n)$ as $1, 2, ..., M_1$. To ensure that the random process has finite second order moments we assume that $Var[Z(\mathbf{s}, t)] < \infty$. We define the mean and covariance functions based on the above defined mapping as

$$\begin{aligned} E[Z(\mathbf{s},t)] &= \mu(\mathbf{s},t) \\ Cov[Z(\mathbf{s}_i;t_i), Z(\mathbf{s}_j;t_j)] &= C(\mathbf{s}_i,\mathbf{s}_j;t_i,t_j)\ \{i,j = 1,2,...,M_1\} \end{aligned} \tag{1}$$

*University of Manchester, UK and CR Rao AIMSCS, Hyderabad, India
†Corresponding author. University of Manchester,UK and CR Rao AIMSCS, Hyderabad, India
‡University of Manchester, UK

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

$$\begin{aligned} E\left[Z(\mathbf{s},t)\right] &= \mu \\ Cov[Z(\mathbf{s}_i;t_i),Z(\mathbf{s}_j;t_j)] &= C(\mathbf{s}_i-\mathbf{s}_j;t_i-t_j) \end{aligned} \tag{2}$$

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

$$Cov[Z(\mathbf{s}_i;t_i),Z(\mathbf{s}_j;t_j)] = C(\|\mathbf{s}_i-\mathbf{s}_j\|;|t_i-t_j|) \tag{3}$$

As in the case of spatial process one can also define the variogram for the above spatio-temporal process 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-a-vis variogram) is essential in the prediction of an unknown observation at a known location which is an integral part of the spatial analysis and we briefly outline the approach. Given the above set up, 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 $\mathbf{Z} = (Z(\mathbf{s}_1; t_1), ..., Z(\mathbf{s}_{M_1}; t_{M_1}))^{'}$. The minimum mean square (optimal) linear predictor is well known to be

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

where $\mathbf{c}(\mathbf{s}_0; t_0) = Cov[Z(\mathbf{s}_0; t_0), \mathbf{Z}]$,$\Sigma = \{Cov[Z(\mathbf{s}_i; t_i), Z(\mathbf{s}_j; t_j)]\}$ and $\boldsymbol{\mu} = E(\mathbf{Z})$. When the mean $\boldsymbol{\mu}$ and the dispersion matrix Σ are known, the above predictor is called the *simple kriging predictor* (see e.g [7, Ch. 3]). The *mean square prediction error*(MSPE) is given by $\mathbf{c}(\mathbf{s}_0, t_0)^{'}\Sigma^{-1}\mathbf{c}(\mathbf{s}_0, t_0)$.

To evaluate (6), we need the inversion 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 $\mathbf{s}_0$. In order 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 [8]). Once a covariance function is decided an important problem is the estimation of the parameters of this function using the data. Though there is substantial literature in the case of spatial data for the estimation of parameters using the variogram ,not much literature exists in the case of spatio-temporal data.

In this paper our objective is to consider the estimation of the parameters of non-separable spatio-temporal covariance functions using the frequency domain approach. Cressie and Huang [6], Gneiting [10] and Ma[14] 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 object 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 Whittle likelihood approach often used in time series modelling and our approach takes into account spatial correlation, temporal correlation and interaction as well.

In section 2 we briefly outline the earlier time domain approaches for the estimation of 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. We consider the Pacific wind speed data earlier considered by Cressie and Huang [6] and estimate the parameters of three covariance functions using the method proposed.

2 Non-separable class of covariances and the estimation

In this section we briefly describe the class of covariance functions proposed and the method of estimation suggested by Cressie and Huang ([6]) and Gneiting ([10]). As pointed out by several authors it is non-trivial to construct non-separable class of covariances which are positive semi-definite. One class of spatially isotropic covariances proposed by Gneiting ([10]) and Gneiting et al. [11] based on generalizing the ideas of Cressie and Huang [6] has the following form

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

where it is assumed that

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

For details we refer to Gneiting [10] and to a recent paper by Kent et al. [12]. Kent et al. [12] point out in their paper that the class of covariances defined by Gneiting [10], in certain circumstances, possess a counter intuitive dimple and in some circumstances the magnitude of the dimple can be non trivial. Since we assumed 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, for our present estimation purposes we are primarily interested in the estimation of parameters of a given class of covariance functions (or equivalently its spectral density function) and not about the choice of these functions.

Given a sample from the stationary spatio-temporal process, Cressie and Huang [6] and others defined its estimator in terms of its corresponding variogram as

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

where,

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

$|N(\mathbf{h}(l);u)|$ is the number of distinct elements in the set $N(\mathbf{h}(l);u); l = 1,2,...,L; u = 0,1,...U$. This follows from the spatial case where the above estimator, due to Matheron [15], is termed as the classical variogram estimator. The study of sampling properties of the above estimator (such as its variance, sampling distribution etc.) gets more difficult even with additional assumption of the 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 $\left(Z(\mathbf{s}_i,t) - Z(\mathbf{s}_j,t')\right)^2$ is proportional to a chi-square, the sum in (8), is no longer the sum of independent chi-squares and as such the assumption on which the above estimator is based, is unrealistic. However Cressie and Huang [6] have used these assumptions extending from Cressie [5] in case of spatial processes. More formally, for a chosen spatial distance $\mathbf{h}(l), l = 1, 2, ..., L$ they approximate

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

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

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

As far as we are aware, no sampling properties of the estimators are known. In the following sections we propose a frequency domain method which upto an extent circumvents those dependency problems outlined. Further we can also study the asymptotic properties of the estimators thus obtained using some classical results available in time series.

3 Frequency Method of Estimation of Covariance Parameters

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

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

We now define a new spatio-temporal random process $Y_{ij}(t)$,

$$Y_{ij}(t) = Z(\mathbf{s}_i, t) - Z(\mathbf{s}_j, t), \text{ for each } t = 1, 2, ..., n \tag{12}$$

and for locations $\mathbf{s}_i, \mathbf{s}_j$, where $\mathbf{s}_i$ and $\mathbf{s}_j$ are defined in the set $\{N(\mathbf{h}_l) = \{\mathbf{s}_i, \mathbf{s}_j; \|\mathbf{s}_i - \mathbf{s}_j\| = \|\mathbf{h}_l\}\|,\ l = 1, 2, ..., L\}$. 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 of $\{Y_{ij}(t),\ i$ is not equal to $j\}$ at the frequencies $\omega_k = 2\pi k/n,\ k = 0, 1, ..., [n/2]$ as

$$J_{\mathbf{s}_i\mathbf{s}_j}(\omega_k) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} Y_{ij}(t)e^{-it\omega_k} \tag{13}$$

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

$$\begin{aligned} I_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k) &= |J_{\mathbf{s}_i\mathbf{s}_j}(\omega_k)|^2 \\ &= \frac{1}{2\pi} \sum_{u=-(n-1)}^{n-1} \hat{c}_{y,ij}(u)e^{-iu\omega_k} \end{aligned} \tag{14}$$

where $\hat{c}_{y,ij}(u)$ is the sample autocovariance function of time lag u of the stationary series $\{Y_{ij}(t), i$ is not equal to$j\}$.

For the sampling properties of Finite Fourier Transforms and periodograms, we refer to Brillinger [3], Priestley [17] and a recent paper of Dwivedi and Subba Rao [9]

From (12) and (13), we obtain

$$J_{\mathbf{s}_i\mathbf{s}_j}(\omega_k) = J_{\mathbf{s}_i}(\omega_k) - J_{\mathbf{s}_j}(\omega_k) \text{ and hence} \tag{15}$$

$$I_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k) = I_{\mathbf{s}_i}(\omega_k) + I_{\mathbf{s}_j}(\omega_k) - 2Re[I_{\mathbf{s}_i\mathbf{s}_j}(\omega_k)] \tag{16}$$

where $J_{\mathbf{s}_i}(\omega_k)$ and $J_{\mathbf{s}_j}(\omega_k)$ are Finite Fourier Transforms of the individual series $\{Z(\mathbf{s}_i,t)\}$ and $\{Z(\mathbf{s}_j,t)\}$, $I_{\mathbf{s}_i}(\omega_k)$ and $I_{\mathbf{s}_j}(\omega_k)$ are their corresponding periodograms and $I_{\mathbf{s}_i\mathbf{s}_j}(\omega_k)$ is the cross periodogram. Note that in the above 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 $I_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k)$

Let $g_{\mathbf{s}_i\mathbf{s}_j}(\omega,\boldsymbol{\theta})$ denote the second order spectral density function of the series $\{Y_{ij}(t), i$ is not equal to $j\}$ which we assume is function of the parameter vector $\boldsymbol{\theta}$. Then from (15) we obtain

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

and for large n, the above can be approximated by

$$g_{\mathbf{s}_i\mathbf{s}_j}(\omega_k)(\boldsymbol{\theta}) = 2f(\omega_k,\boldsymbol{\theta}_1) - 2f_{\|\mathbf{h}\|}(\omega_k,\boldsymbol{\theta}_2) \tag{18}$$

where $f(\omega_k,\boldsymbol{\theta}_1)$ is the second order spectral density function of the stationary spatial process $\{Z(\mathbf{s}_i,t); i = 1,2,...,m\}$ and $f_{\|\mathbf{h}\|}(\omega_k,\boldsymbol{\theta}_2)$ is the cross spectral density function of the process $\{Z(\mathbf{s}_i,t)$ and $\{Z(\mathbf{s}_i,t)$ given by

$$f_h(\omega_k,\boldsymbol{\theta}_2) = \frac{1}{2\pi}\sum_{u=-\infty}^{\infty} c(\mathbf{s}_i - \mathbf{s}_j, u)e^{-iu\omega}$$

In view of isotropy the above cross spectral density function reduces to

$$f_h(\omega_k,\boldsymbol{\theta}_2) = \frac{1}{2\pi}\sum_{u=-\infty}^{\infty} c(\|\mathbf{s}_i - \mathbf{s}_j\|, u)e^{-iu\omega} \quad = f_{\|\mathbf{h}\|}(\omega_k,\boldsymbol{\theta}_2) \tag{19}$$

The cross spectral density function $f_h(\omega_k,\boldsymbol{\theta}_2)$ is usually a complex valued function, but in view of the assumption of stationarity and isotropy, $c(\mathbf{s}_i - \mathbf{s}_j, u) = c(\|\mathbf{s}_i - \mathbf{s}_j\|, u)$ and $c(\|\mathbf{s}_i - \mathbf{s}_j\|, u) = c(\|\mathbf{s}_i - \mathbf{s}_j\|, -u)$ which implies that $f_{\|\mathbf{h}\|}(\omega_k,\boldsymbol{\theta}_2)$ is a real valued function. Using the analogy of the analysis of spatial process, we can define $g_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k,\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ as frequency domain version of the variogram. Now consider $g_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k,\boldsymbol{\theta})$ defined in (18), the frequency domain version of the classical semi-variogram namely,

$$\begin{aligned}\frac{1}{2}g_{\|\mathbf{h}\|}(\omega_k) &= f(\omega_k,\boldsymbol{\theta}_1) - f_h(\omega_k,\boldsymbol{\theta}_2)\\ &= f(\omega_k,\boldsymbol{\theta}_1)\,[1 - f_h(\omega_k,\boldsymbol{\theta}_2)/f(\omega_k,\boldsymbol{\theta}_1)]\end{aligned}$$

Let us denote $f_h(\omega_k,\boldsymbol{\theta}_2)/f(\omega_k,\boldsymbol{\theta}_1)$ by $W_h(\omega_k,\boldsymbol{\theta})$, which lies between $[0,1]$ for all k and all $\|\mathbf{h}\|$. This measure is similar to coherency measure used in signal processing and multivariate 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 one. If $\|\mathbf{h}\| = 0$ obviously it is equal to one.

By plotting this measure, individually for each frequency we can have an idea in which frequency bands the two spatial series are strongly correlated or if we take an average over all the frequencies (suitably normalised) and plot these values against Eucledean distance $\|\mathbf{h}\|$, we can get an idea of the spatial distance over which the processes are correlated and thus help us in modelling the processes. This measure, in a way, can be used to give an idea about the range parameter similar to that of variogram in the spatial situations. These ideas need to be further investigated. Of course, the measure of spatial coherency proposed here need to be studied further. As we pointed out earlier, our object in this paper is the estimation of the parameters only.

It is well known (see for example the books by Priestley [17], Brillinger [3], Brockwell and Davis [4] etc. and a recent paper by Dwivedi and Subba Rao [9]) that discrete Finite Fourier transforms of a stationary process are asymptotically uncorrelated over distinct canonical frequencies, and have complex normal distribution (see Brillinger [3, Theorem 4.4.1]) (thus independent) and further the Fourier transforms at distinct frequencies of two spatial processes are also asymptotically independent and each have complex normal distribution if the random process is Gaussian. In view of this asymptotic property, we can consider the vector $\mathbf{J}_{\|\mathbf{h}\|}^{'} = [J_{\|\mathbf{h}\|}(\omega_1), J_{\|\mathbf{h}\|}(\omega_2), ..., J_{\|\mathbf{h}\|}(\omega_M)]^{'}$, where $J_{\|\mathbf{h}\|}(\omega_k) = J_{\mathbf{s}_i\mathbf{s}_j}(\omega_k)$ as defined by (13) and $M = [n/2]$, is distributed as asymptotically multivariate complex normal with mean zero and variance covariance matrix with diagonal $[g_{\|\mathbf{h}\|}(\omega_1), g_{\|\mathbf{h}\|}(\omega_2), ..., g_{\|\mathbf{h}\|}(\omega_M)]^{'}$. We note that because of the asymptotic independence, the off diagonal elements are zero. In view of this asymptotic distribution, the log likelihood can be shown to be proportional to

$$Q_{n,ij}(\boldsymbol{\theta}) = \sum_{k=1}^{M}\left[\ln(g_{\mathbf{s}_i\mathbf{s}_j}(\omega_k,\boldsymbol{\theta})) + \frac{I_{\mathbf{s}_i\mathbf{s}_j}(\omega_k)}{g_{\mathbf{s}_i\mathbf{s}_j}(\omega_k,\boldsymbol{\theta})}\right] \tag{20}$$

Since the above likelihood function is evaluated for only one Euclidean distance $\|\mathbf{h}_l\|, l = 1, 2, ..., L$ we can use for the minimisation purposes the pooled (and scaled) sum calculated over all L distances and consider $Q_n(\boldsymbol{\theta})$.

$$Q_n(\boldsymbol{\theta}) = \frac{1}{N}\sum_i\sum_j Q_{n,ij}(\boldsymbol{\theta}), \text{ here } N = M\sum_{l=1}^{L}|N(\mathbf{h}_l)| \tag{21}$$

We minimize $Q_n(\boldsymbol{\theta})$ with respect to parameters of $\boldsymbol{\theta}$

Note: In defining the above we have assumed that $Q_{n,ij}(\boldsymbol{\theta})$ and $Q_{n,i'j'}(\boldsymbol{\theta})$ are independent. Though this assumption seems to be a bit unrealistic at the moment, we see from the simulations that the estimates of $\boldsymbol{\theta}$ are very close to the true values and are consistent.

In the next section we discuss strong consistency properties of the estimates of the unknown parameters obtained from the above minimization.

4 Asymptotic Convergence of Parameter Estimates

Using a well known lemma based on the Arzela-Ascoli theorem (see for example [1, p. 221]) here we show that the parameter estimator $\hat{\boldsymbol{\theta}}_n$ obtained by minimizing (21) with respect to the unknown parameter vector $\boldsymbol{\theta}$ converges in probability to the original parameter vector $\boldsymbol{\theta}_0$ as $n \to \infty$. Throughout our discussion we assume that $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ where Θ is a compact set and $Q_n(\boldsymbol{\theta})$ has a unique minimum. We now state the stochastic Ascoli lemma for any sequence of random functions $Q_n^*(\boldsymbol{\theta})$, which is an extension of the Arzela-Ascoli convergence theorem of sequence of functions on the probability space of sequences of random functions.

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* $Q_n^*(\boldsymbol{\theta}) \overset{a.s}{\to} Q^*(\boldsymbol{\theta})$, *(pointwise convergence)*

2. *the parameter space* Θ *is compact,*

3. $Q_n^*(\boldsymbol{\theta})$ is stochastic equicontinuous.

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

We skip the proof here but the interested reader may refer Billingsley [1].We later use the above theorem (see Lemma 3) to show consistency of the estimators. We now introduce some assumptions which are needed for obtaining asymptotic properties of the parameter estimator $\hat{\boldsymbol{\theta}}_n$.

Assumption 1. *(i) For any $n \in \mathbb{Z}^+$. and $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n \in \mathbb{R}^2$ we have the following $\alpha-$ mixing assumption:*

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

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 for the difference series $Y_{ij}(t) = Z(\mathbf{s}_i,t) - Z(\mathbf{s}_j,t)$ the above assumptions also hold.*

Let us define the set $S = \{\mathbf{u} = (\mathbf{s}_1, \mathbf{s}_2) : \|\mathbf{s}_1 - \mathbf{s}_2\| = h\}$. Let $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4)$ 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 $c_{\mathbf{u}_1,\mathbf{u}_1}(s) = \mathrm{Cov}\{Y_{\mathbf{u}_1}(t), Y_{\mathbf{u}_2}(t+s)\}$ and the fourth order cumulant of the series $\{Y_{\mathbf{s}_i}(t); i = 1,2,3,4\}$ by $\mathrm{Cum}(Y_{\mathbf{u}_1}(t), Y_{\mathbf{u}_2}(t+s_1), Y_{\mathbf{u}_3}(t+s_2), Y_{\mathbf{u}_4}(t+s_3)) = C_{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\mathbf{u}_4}(s_1,s_2,s_3)$. 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 (see Brillinger[3]) respectively, of the process $\{Y_{\mathbf{s}_i}(t); i = 1,2,3,4\}$.

Lemma 2. *Suppose Assumption 1 holds. Then define*

$$W_n = \sum_{\mathbf{u} \in S} \sum_{k=1}^{\lfloor n/2 \rfloor} h_{\mathbf{u}}(\omega_k) I_{\mathbf{u}}(\omega_k), \tag{22}$$

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

(i)

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

(ii)

$$\mathrm{Var}\{\frac{1}{\sqrt{n}}W_n\} \xrightarrow{p} 2\sum_{\mathbf{u}_1,\mathbf{u}_2\in S}\frac{1}{2\pi}\int_0^\pi h_{\mathbf{u}_1}(\omega)h_{\mathbf{u}_2}(\omega)|f_{\mathbf{u},\mathbf{u}}(\omega)|^2 d\omega$$
$$+\sum_{\mathbf{u}_1,\mathbf{u}_2\in S}\left(\frac{1}{2\pi}\right)^2\int_0^\pi\int_0^\pi h_{\mathbf{u}_1}(\omega_1)h_{\mathbf{u}_2}(\omega_2)f_{\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_2}(\omega_1,-\omega_1,\omega_2)d\omega_1 d\omega_2$$

Proof.

$$\begin{aligned}&\mathrm{E}\{\frac{1}{n}W_n\}\\ &=\sum_{\mathbf{u}\in S}\frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}h_{\mathbf{u}}(\omega_k)\,\mathrm{E}\{I_{\mathbf{u}}(\omega_k)\}\\ &\approxeq\sum_{\mathbf{u}\in S}\left[\frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}h_{\mathbf{u}}(\omega_k)f_{\mathbf{u},\mathbf{u}}(\omega_k)\right]\\ &\approxeq\sum_{\mathbf{u}\in S}\frac{1}{2\pi}\int_0^\pi h_{\mathbf{u}}(\omega)f_{\mathbf{u}_1,\mathbf{u}_2}(\omega)d\omega\text{see [2, Ch 2]}\end{aligned}\tag{23}$$

To obtain an expression for the asymptotic variance, we first note the well known results (see Brillinger [3, Ch 2 and 3])

$$\begin{aligned}&\text{(i) Cov}\left[|J_{\mathbf{u}_1}(\omega_{k_1})|^2,|J_{\mathbf{u}_2}(\omega_{k_2})|^2\right]=\left[\mathrm{Cov}(J_{\mathbf{u}_1}(\omega_{k_1}),J_{\mathbf{u}_2}(\omega_{k_2}))\,\mathrm{Cov}(\overline{J_{\mathbf{u}_1}(\omega_{k_1})},\overline{J_{\mathbf{u}_2}(\omega_{k_2})})\right]\\ &+\left[\mathrm{Cov}(J_{\mathbf{u}_1}(\omega_{k_1}),\overline{J_{\mathbf{u}_2}(\omega_{k_2})})\,\mathrm{Cov}(\overline{J_{\mathbf{u}_1}(\omega_{k_1})},J_{\mathbf{u}_2}(\omega_{k_2}))\right]\\ &+\left[\mathrm{Cum}(J_{\mathbf{u}_1}(\omega_{k_1}),\overline{J_{\mathbf{u}_1}(\omega_{k_1})},J_{\mathbf{u}_2}(\omega_{k_2}),\overline{J_{\mathbf{u}_2}(\omega_{k_2})})\right]\\ &\text{(ii)}\overline{J_{\mathbf{u}}(\omega_k)}=J_{\mathbf{u}}(\omega_{n-k})\\ &\text{(iii) Cov}(J_{\mathbf{u}_1}(\omega_{k_1}),J_{\mathbf{u}_2}(\omega_{k_2}))=\left[f_{\mathbf{u}_1,\mathbf{u}_2}(\omega_{k_1})\frac{1}{n}\sum_{k=1}^{\lfloor n/2\rfloor}e^{-i(k_1-k_2)w_k}+O\left(\frac{1}{n}\right)\right]\\ &\text{(iv) Cum}(J_{\mathbf{u}_1}(\omega_{k_1}),\overline{J_{\mathbf{u}_1}(\omega_{k_1})},J_{\mathbf{u}_2}(\omega_{k_2}),\overline{J_{\mathbf{u}_2}(\omega_{k_2})})=\frac{(2\pi)^2}{n}f_{\mathbf{u}_1,\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_2}(\omega_{k_1},-\omega_{k_1},\omega_{k_2})+O\left(\frac{1}{n^2}\right)\end{aligned}\tag{24}$$

Now expanding $\mathrm{Var}\{\frac{1}{\sqrt{n}}W_n\}$ and using the above equations; additionally using the Fourier-Stieltjes integral approximation of limit of discrete Fourier transforms (as in the above proof of Expectation) we have the desired expression. □

We now study 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* $\theta\in\Theta$, $f_{\mathbf{u}_1,\mathbf{u}_2}(\omega;\theta)$ *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) And the original parameter vector $\boldsymbol{\theta_0}$ *lies in the interior of* Θ.

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

Define the criterion

$$Q_n(\theta) = \sum_{\mathbf{u}\in S} \sum_{k=1}^{\lfloor n/2 \rfloor} \left\{ \log g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}) + \frac{I_{\mathbf{u}}(\omega_k)}{g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta})} \right\} \tag{25}$$

Let $\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} \frac{1}{n} Q_n(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} Q_n^*(\boldsymbol{\theta})$

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 $n \to \infty$. .

Proof. Due to boundedness of $g(\omega;\boldsymbol{\theta})$ (Assumption 2) for all $\boldsymbol{\theta}$ and ω we have point wise convergence of $Q_n(\boldsymbol{\theta})$ (by Lemma 2), that is

$$\frac{1}{n} Q_n(\boldsymbol{\theta}) \to \sum_{\mathbf{u}\in S} \int \left[\log g_{\mathbf{u}}(\omega;\boldsymbol{\theta}) + \frac{g_{\mathbf{u}}(\omega;\boldsymbol{\theta}_0)}{g_{\mathbf{u}}(\omega;\boldsymbol{\theta})} \right] d\omega$$

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

$$\left| \frac{1}{n} Q_n(\boldsymbol{\theta}_1) - \frac{1}{n} Q_n(\boldsymbol{\theta}_2) \right| = \left| \frac{1}{n} \nabla Q_n(\check{\boldsymbol{\theta}})(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \right| \tag{26}$$

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

$$\frac{1}{n} \frac{\partial Q_n(\boldsymbol{\theta})}{\partial \theta_i} = \sum_{\mathbf{u}} \frac{1}{n} \sum_{k=1}^{\lfloor n/2 \rfloor} \left[\frac{1}{g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta})} \frac{\partial g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta})}{\partial \theta_i} - \frac{I_{\mathbf{u}}(\omega_k)}{(g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}))^2} \frac{\partial g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta})}{\partial \theta_i} \right] \tag{27}$$

Now under the assumption that $g_{\mathbf{u}}(.)$ is bounded, from the above, we get

$$\frac{1}{n} \frac{\partial Q_n(\boldsymbol{\theta})}{\partial \theta_i} \leq \sum_{\mathbf{u}} \left[C + \frac{C}{n} \sum_{k=1}^{\lfloor n/2 \rfloor} I_{\mathbf{u}}(\omega_k) \right] = M_n, \text{ for some } C > 0 \tag{28}$$

From lemma 2 it follows that the expectation of M_n tends to

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

and that

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

This implies that

$$\left| \frac{1}{n} Q_n(\boldsymbol{\theta}_1) - \frac{1}{n} Q_n(\boldsymbol{\theta}_2) \right| \leq \left[\mathrm{E}(M_n) + o_p(1)\right] |\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| \tag{29}$$

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

Now we can prove asymptotic normality of $\hat{\boldsymbol{\theta}}_n$. For the proof we also need that the second derivative of $\nabla^2 Q_n(.)$ converges uniformly. We skip the proof here which is very similar to the above proof under the additional assumption that the derivative of $\frac{1}{n}\sum_{k=0}^{[n/2]} g_{\mathbf{s}_i\mathbf{s}_j}(\omega_k,\boldsymbol{\theta})$, w.r.t to $\boldsymbol{\theta}$, denoted by $g'_{1n}(\boldsymbol{\theta})$ exists for all n and converges uniformly to function $g(\boldsymbol{\theta})$.

Theorem 2. *Let Assumptions 1 and 2 be true so that Lemma 2 and 3 hold. Then*

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0) &\xrightarrow{D} N(\mathbf{0},\nabla^2 Q_n^{-1}(\boldsymbol{\theta}_0)V\nabla^2 Q_n(\boldsymbol{\theta}_0)),\ \textit{where}\\ V &= \lim_{n\to\infty}\operatorname{Var}\left(\frac{1}{\sqrt{n}}\nabla Q_n(\boldsymbol{\theta}_0)\right)\end{aligned}\tag{30}$$

Proof. Since $\nabla Q_n(\boldsymbol{\theta})$ is a vector we can only make a Taylor expansion point wise on $\nabla Q_n(\boldsymbol{\theta})$. Thus point wise we have by the mean value theorem

$$\frac{1}{n}\nabla Q_n(\hat{\boldsymbol{\theta}})\mid_i=\frac{1}{n}\nabla Q_n(\boldsymbol{\theta}_0)\mid_i+(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\frac{1}{n}\nabla^2 Q_n(\check{\boldsymbol{\theta}}_n)\mid_i\tag{31}$$

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

$$\begin{aligned}&\sup_{\boldsymbol{\theta}\in\Theta}\mid\frac{1}{n}\nabla^2 Q_n(\hat{\boldsymbol{\theta}})-\nabla^2 Q(\boldsymbol{\theta})\mid\to 0,\text{ where}\\ &\nabla^2 Q(\boldsymbol{\theta})=\lim_{n\to\infty}\mathrm{E}\left(\frac{1}{n}\nabla^2 Q_n(\hat{\boldsymbol{\theta}})\right)\end{aligned}\tag{32}$$

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

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

and this implies

$$\frac{1}{n}\nabla Q_n(\hat{\boldsymbol{\theta}})=\frac{1}{n}\nabla Q_n(\boldsymbol{\theta}_0)+\nabla^2 Q(\boldsymbol{\theta}_0)'(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)+o_p(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)\tag{33}$$

We note that the left hand side 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 (28) that

$$\begin{aligned}&\frac{1}{\sqrt{n}}\nabla Q_n(\boldsymbol{\theta}_0)\\ &=\sum_{\mathbf{u}}\frac{1}{\sqrt{n}}\sum_{k=1}^{\lfloor n/2\rfloor}\left[\frac{1}{g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)}\nabla g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)-\frac{I_{\mathbf{u}}(\omega_k)}{g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)^2}\nabla g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)\right]\\ &=\sum_{\mathbf{u}}\frac{1}{\sqrt{n}}\sum_{k=1}^{\lfloor n/2\rfloor}\left[\mathrm{E}\left[I_{\mathbf{u}}(\omega_k)\right]-I_{\mathbf{u}}(\omega_k)\right]\frac{\nabla g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)}{g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)^2}\\ &-\sum_{\mathbf{u}}\frac{1}{\sqrt{n}}\sum_{k=1}^{\lfloor n/2\rfloor}\left[\mathrm{E}\left[I_{\mathbf{u}}(\omega_k)\right]-g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)\right]\frac{\nabla g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)}{g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)^2}\\ &=I+II\end{aligned}\tag{34}$$

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

$$\frac{1}{\sqrt{n}}\nabla Q_n(\boldsymbol{\theta}_0) = \sum_{\mathbf{u}} \frac{1}{\sqrt{n}} \sum_{k=1}^{\lfloor n/2 \rfloor} \{\mathrm{E}\left[I_{\mathbf{u}}(\omega_k)\right] - I_{\mathbf{u}}(\omega_k)\} \frac{\nabla g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)}{g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)^2} + O(\frac{1}{\sqrt{n}}) \tag{35}$$

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

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

where,

$$H_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0) = \frac{\nabla g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)}{g_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)^2} \tag{36}$$

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

$$\begin{aligned} I &= \sum_{\mathbf{u}} \frac{1}{\sqrt{n}} \sum_{t}\sum_{\tau} \left[Y_{\mathbf{u}}(t)Y_{\mathbf{u}}(\tau) - \mathrm{E}(Y_{\mathbf{u}}(t)Y_{\mathbf{u}}(\tau))\right] \frac{1}{n}\sum_{k=1}^{n} H_{\mathbf{u}}(\omega_k;\boldsymbol{\theta}_0)e^{i\omega_k(t-\tau)} \\ &= \sum_{\mathbf{u}} \frac{1}{\sqrt{n}} \sum_{t}\sum_{\tau} \left[Y_{\mathbf{u}}(t)Y_{\mathbf{u}}(\tau) - \mathrm{E}(Y_{\mathbf{u}}(t)Y_{\mathbf{u}}(\tau))\right] h_{\mathbf{u}}(t-\tau) + o_p(\frac{1}{\sqrt{n}}) \end{aligned} \tag{37}$$

Where $h_{\mathbf{u}}(\tau) = \int_{-\infty}^{\infty} H_{\mathbf{u}}(\omega;\boldsymbol{\theta}_0)e^{i\omega(\tau)}d\omega$. 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 $\{h_{\mathbf{u}}(\tau)\}$ must decay to zero at the rate $\frac{1}{|\tau|^2}$ (see Briggs and Henson [2, Ch 6]). Note that $\{Y_{\mathbf{u}}(t)\}$ are α-mixing at the rate specified in Assumption 1. Then by Theorem 2.2 of Lee and Subba Rao [13] we can show the asymptotic normality of $\frac{1}{\sqrt{n}}\nabla Q_n(\boldsymbol{\theta}_0)$. That is

$$\begin{aligned} &\frac{1}{\sqrt{n}}\nabla Q_n(\boldsymbol{\theta}_0) \xrightarrow{D} N(0,V) \\ &V = \lim_{n\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{n}}\nabla Q_n(\boldsymbol{\theta}_0)\right) \end{aligned} \tag{38}$$

An expression for V can be deduced from Lemma 2. The above result together with equation (33) gives

$$\sqrt{n}(\hat{\boldsymbol{\theta}_n} - \boldsymbol{\theta}_0) \xrightarrow{D} N(\mathbf{0}, \nabla^2 Q_n^{-1}(\boldsymbol{\theta}_0)V\nabla^2 Q_n(\boldsymbol{\theta}_0)) \tag{39}$$

□

In the next section we apply the method described so far to estimation of unknown parameters of a parametric spatio-temporal covariance function of a simulated spatio temporal random process.

5 Numerical Example - Simulated Data

In the following we consider the analysis of the data generated from a specific model with the main interest of illustrating the methodology proposed earlier. We use the lattice $\mathbb{Z}^2$ as the spatial domain for sampling locations. We consider a rectangular lattice of size 17×17 and at each of the 289 locations we generate a time series of length 480. The spatio-temporal stochastic process we consider is

$$Z(\mathbf{s}_i; t) = X(\mathbf{s}_i).Y_t; \ \ (i = 1, 2, ..., 289; t = 1, 2, ..., 480)$$

Where the spatial process $\{X(\mathbf{s}_i)\}$ and temporal process $\{Y_t\}$ are assumed to be independent, each with mean zero. We note that the above is a separable spatio-temporal process with

$$\operatorname{Cov}\{Z(\mathbf{s}+\mathbf{h}; t+u), Z(\mathbf{s}; t)\} = \operatorname{Cov}\{X(\mathbf{s}+\mathbf{h}), X(\mathbf{s})\}. \operatorname{Cov}\{Y_{t+u}, Y_t\} \tag{40}$$

we assume, for generation of the process, the spatial covariance is of the form

$$\operatorname{Cov}\{X(\mathbf{s}+\mathbf{h}), X(\mathbf{s})\} = \sigma^2 \exp\left(-\|\mathbf{h}\|/\alpha\right), \alpha > 0 \tag{41}$$

and the times series Y_t satisfies and AR(1) model

$$Y_t = \phi Y_{t-1} + \epsilon_t, |\phi| < 1 \tag{42}$$

and $\{\epsilon_t\}$ is a Gaussian White noise with zero mean and variance ν^2. It is easy to show that $\operatorname{Var}(Y_t) = \frac{\nu^2}{1-\phi^2}$, $\operatorname{Cov}\{Y_{t+u}, Y_t\} = \phi^{|u|}\frac{\nu^2}{1-\phi^2}$ and the first order correlation coefficient as ϕ which can be estimated from a single realization of the time series data. For simulation of the data, we have chosen $\sigma = 7, \alpha = 5.38, \phi = .5, \nu^2 = 1$.

We note that for the above separable spatio-temporal process, the second order spectral density function is given by

$$\begin{aligned} f_z(\omega, \lambda) &= f_x(\omega) f_y(\lambda) \\ \text{where } f_x(\omega) &= \frac{\sigma^2\alpha}{\pi(1+\omega^2\alpha^2)} \\ \text{and } f_y(\lambda) &= \frac{\nu^2}{2\pi|1-\phi e^{-i\lambda}|^2}, |\lambda| \leq \pi. \end{aligned} \tag{43}$$

The parameters we are interested in estimating are $(\sigma, \alpha, \phi, \nu)$, and we minimize the objective function $Q(\boldsymbol{\theta})$ defined in (21). The minimization is done using the nlm package of the software **R** developed by Jose Pinheiro et al. [16]. In order to assess the sampling properties, 1000 realizations have been generated using the above model, of which the minimization did not converge for 13 realizations. The parameters are estimated using 987 realizations. The parameter estimates and the mean square errors are calculated as follows

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \frac{1}{987}\sum_{i=1}^{987} \hat{\boldsymbol{\theta}}_i \\ MSE(\hat{\boldsymbol{\theta}}) &= \text{digonal of } \left\{ \frac{1}{986}\sum_{i=1}^{987}\{\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}\}\{\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}\}' \right\} \end{aligned} \tag{44}$$

In Table 1 the estimates and their mean square errors are given From the above table, we observe that the estimates obtained are very close to the true values. In Fig 1, we have plotted the histograms and Q-Q plots of the estimates are given in Fig 2.

Table 1: Simulation Results of Parameter Estimates

Parameters	Original Values	Mean Estimates	MSE	Bias
σ	2.645	2.6376	0.0004	0.0074
α	5.38	5.3805	.00002	-0.0005
ϕ	0.5	0.4951	0.0015	0.0049
ν	1	0.9991	0.0026	0.001

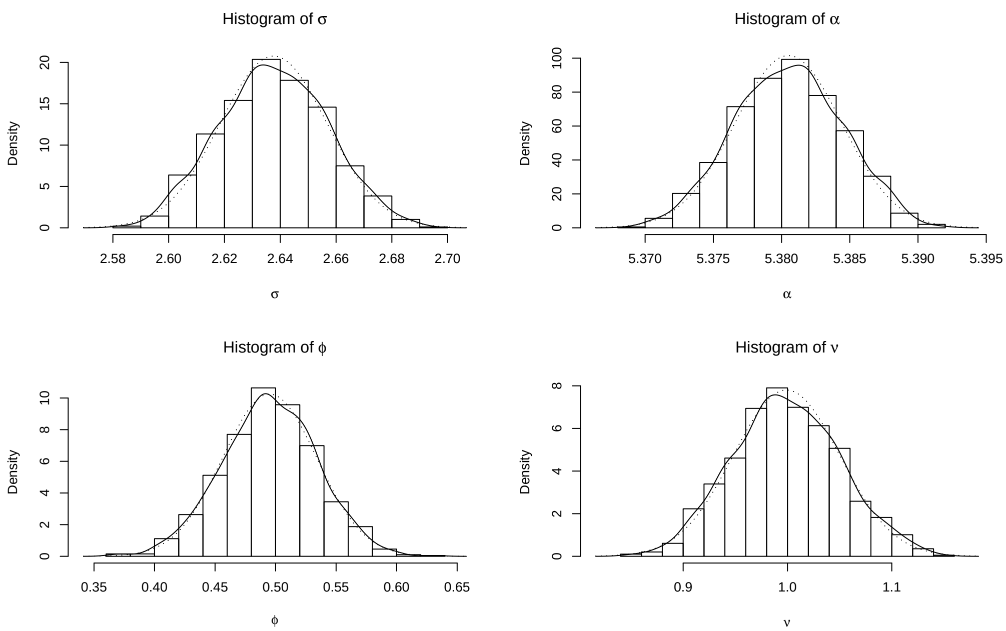

Figure 1: Histogram Plots of Estimates for spatio-temporal parameters σ, α, ϕ and ν obtained by minimizing $Q_n(\boldsymbol{\theta})$.

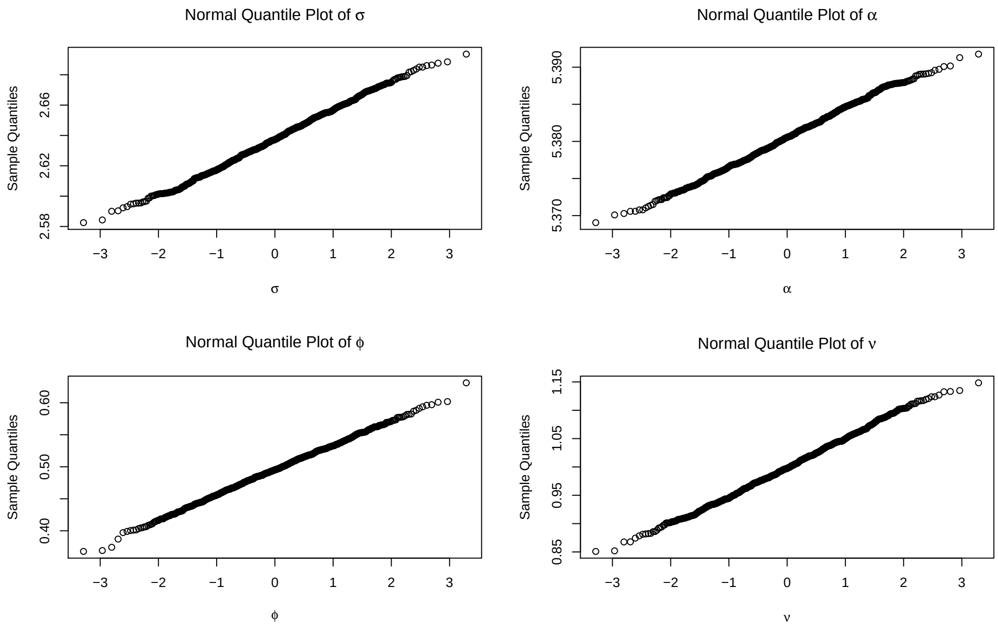

Figure 2: Q-Q Plots of Estimates for spatio-temporal parameters σ, α, ϕ and ν obtained by minimizing $Q_n(\boldsymbol{\theta})$.

From these two figures, we see that the marginal distributions of the estimators are very close to Gaussian distribution.

6 Application to Wind Speed Data

The data provides the record of east-west wind speed on a 17×17 rectangular lattice at grid spacings of 210 km, every six hours from November 1992 to February 1993. So the process is observed at 289 locations and 480 time points.

Before parameter estimation we check if the data is weakly stationary. Figure 3 depicts the spatial and temporal 'mean against standard deviation' plots for the wind speed data to check for evidence of heteroscedasticity. The figures do not indicate any particular pattern and thus an assumption of homoscedasticity is justified as noted in [6]. Next 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-

$$\begin{aligned} \bar{Z}(\mathbf{s}_i;.) &= \frac{1}{n}\sum_{t=1}^{n} Z(\mathbf{s}_i,t) \text{ for } i=1,2,...,m \\ \bar{Z}(.;t) &= \frac{1}{m}\sum_{i=1}^{m} Z(\mathbf{s}_i,t) \text{ for } t=1,2,...,n \end{aligned} \tag{45}$$

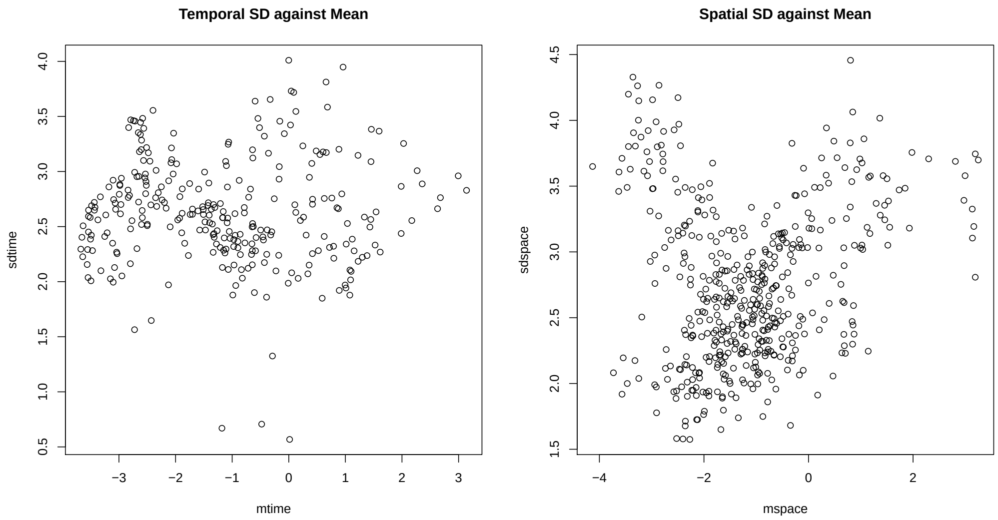

Figure 3: Mean against standard deviation plots. The left panel plots the temporal standard deviation against mean . The right panel plots the spatial standard deviation against mean.

In Figure 4 we observe that the spatial averages (over all locations) displayed against time points and the corresponding temporal sample auto correlation (ACF) plot for the series. Here we define the sample ACF at lag u by $\hat{\rho}_t(u)$ as

$$\begin{aligned} \hat{\rho}_t(u) &= \frac{1}{n}\sum_{t=1}^{n}\bar{Z}(.;t+u)\bar{Z}(.;t)/\sqrt{V_{t+u}V_t} \\ \text{where, } V_{t+u} &= \frac{1}{n-1}\sum_{t=1}^{n}\left\{\bar{Z}(.;t+u)-\frac{1}{n}\sum_{t=1}^{n}\bar{Z}(.;t+u)\right\}^2 \end{aligned} \tag{46}$$

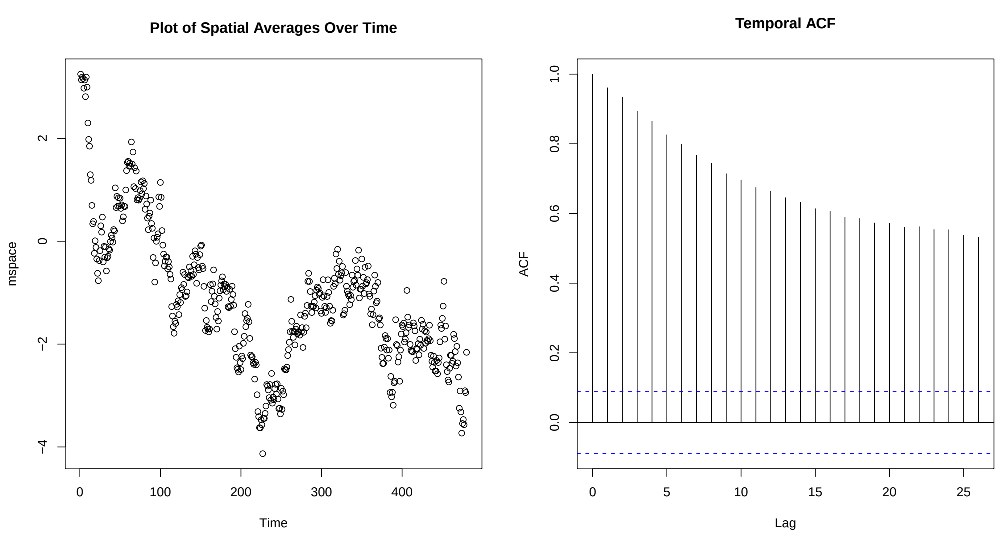

Figure 4: The spatial averages of the original wind speed data, $\bar{Z}(.;t)$, against time and plot of sample ACF $\hat{\rho}_t(u)$.

The temporal averages (at each spatial location) are displayed on the lattice grid in the 3D image of Figure 5.

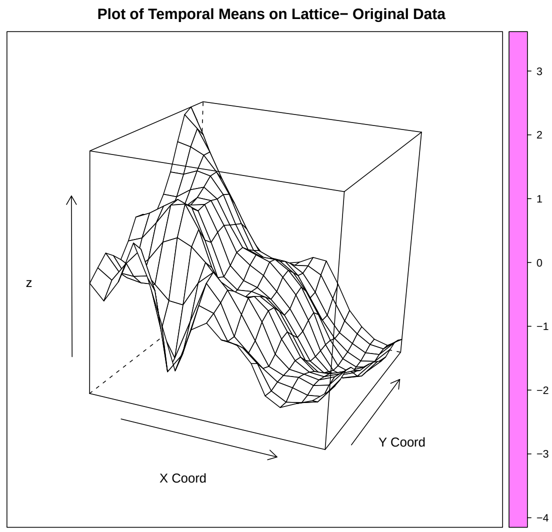

Figure 5: 3D image of the temporal averages of the original wind speed data at the corresponding locations on the lattice grid points.

In [6] Cressie and Huang have assumed spatial and temporal second order stationarity. But from the mean and ACF plots of Figure 4 it is clear that there is a long term temporal deterministic trend in the wind speed data. The 3D spatial plot of Figure 5 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 indicate that data is not second order stationary.

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

$$Z^*(\mathbf{s}_i, t) = Z(\mathbf{s}_i, t) - \bar{Z}(\mathbf{s}_i; .); \text{ for } i = 1, 2, ..., m \tag{47}$$

The respective adjusted means are denoted by $\bar{Z}^*(\mathbf{s}_i, .)$ and $\bar{Z}^*(., t)$. The adjusted mean plots are given in Figures 6 and 7.

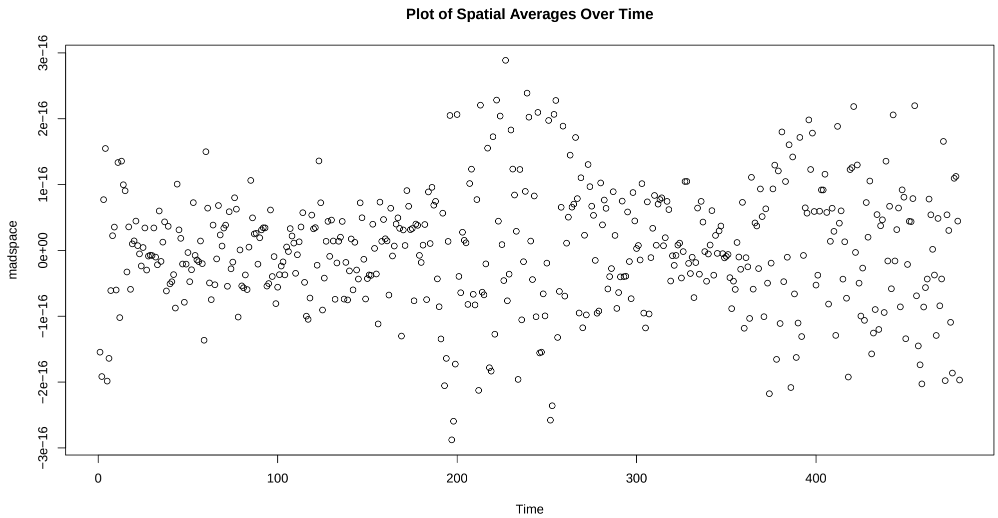

Figure 6: The spatial averages of the adjusted wind speed data, $\bar{Z}^*(.,t).$, at each time point are plotted against time.

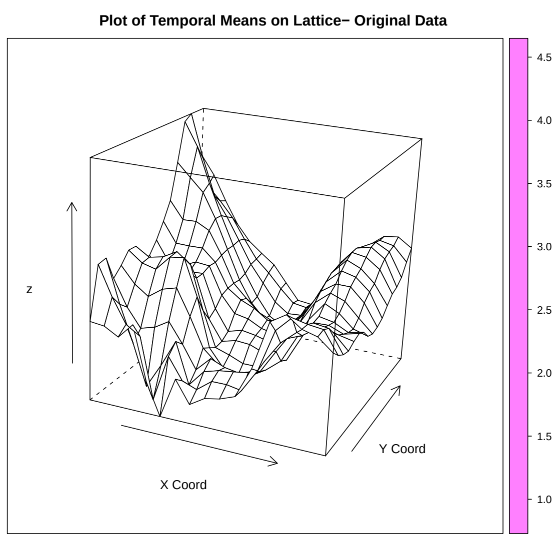

Figure 7: 3D image of the temporal averages of the adjusted wind speed data, $\bar{Z}^*(\mathbf{s}_i,.)$, at the corresponding locations on the lattice grid points.

From Figure 6 we observe that the deterministic temporal trend has been removed from the adjusted observations $Z^*(\mathbf{s}_i,t)$. The 3D plot of temporal averages in Figure 7 shows that the cascading effect has been removed. For the rest of our analyses we treat the adjusted data $Z^*(\mathbf{s}_i,t)$ as the second order stationary spatio-temporal process and denote it by $Z(\mathbf{s};t)$.

We now fit three covariance models, given below, to the spatio-temporal procecss $Z(\mathbf{s};t)$. The first two models are non-separable spatio-temporal isotropic second order stationary covariance functions chosen from [6] while the third model is a generalized version of the first model obtained by Gneiting ([10]). All the three covariance functions are convex functions, chosen based on the spatio-temporal sample variogram (see [6]). In all these models a is the temporal scale parameter, b^2

is the spatial range parameter. Parameter g in the third model is the non-separability parameter. Cressie and Huang have discussed the discontinuity of the data at origin. To incorporate this "nugget" (see[7, Ch 2]) effect we have also included a nugget parameter τ. Following [6] a purely spatial covariance is also incorporated to address the fact that the empirical spatial variogram doesn't change shape at larger temporal lags. But contrary to Cressie and Huang we opt for the second order stationary exponential variogram. We don't see any advantage to choose an intrinsic stationary power variogram when the spatio-temporal process $Z(\mathbf{s};t)$ is assumed to be second order stationary.

$$\text{Model-1: } C(\|\mathbf{h}\|;|u|) = \tau + \sigma^2 \frac{1}{a|u|+1} e^{-\frac{b^2\|\mathbf{h}\|^2}{a|u|+1}} + e^{-a_1\|\mathbf{h}\|} \tag{48}$$

$$\text{Model-2: } C(\|\mathbf{h}\|;|u|) = \tau + \sigma^2 \frac{a|u|+1}{\{(a|u|+1)^2 + b^2\|\mathbf{h}\|^2\}^{3/2}} + e^{-a_1\|\mathbf{h}\|} \tag{49}$$

$$\text{Model-3: } C(\|\mathbf{h}\|;|u|) = \tau + \sigma^2 \frac{1}{a|u|+1} e^{-\frac{b^2\|\mathbf{h}\|^{2g}}{(a|u|+1)^{b_1 g}}} + e^{-a_1\|\mathbf{h}\|} \tag{50}$$

Recall that the pacific wind speed data was observed on a 17×17 rectangular lattice. We obtain estimates of the above parameters for each of the first thirteen vertical spatial lags, $\|\mathbf{h}\|$, of the lattice of locations (for more details see [6]) and report the average of the observations below. Note that the above covariance functions have finite spectral density function but do not have closed form expressions. In such situations we can use their Finite Fourier transforms. We have used the fft(.) routine in **R** (see [18]).

Table 2: Whittle Likelihood Parameter estimates

Covariances	σ	a	b	τ	a_1	b_1	g	Function Value
Model-1	2.406	0.388	7.623	0.100	0.012	0.000	0.000	78539.704
Model-2	2.404	0.383	5.149	0.100	0.006	0.000	0.000	110423.484
Model-3	2.405	0.390	5.954	0.100	0.010	0.998	1.005	78665.974

Based on the minimum values of Q, we recommend the use of the covariance Model 1 for the transformed data amongst 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 transformed data, comparing our estimates or minimum values with that obtained by Cressie and Huang [6] is not appropriate.

7 Conclusion

The method of estimation proposed here are based on Discrete Fourier Transforms of the stationary processes. We exploited the interesting properties of these transforms evaluated at canonical frequencies in order to obtain a likelihood function for the maximisation as is often done in time series. As we noticed, the advantage of these transforms are that they are approximately uncorrelated (in the case of Gaussian processes, they are independent) even though the original processes are stationary ,but highly correlated. In doing so, our methodology depends on spatial parameter only (though implicit function of time). As we have seen the asymptotic properties of the estimates can be obtained under fairly general assumptions and this was not possible (at least not easy) if we use the spatio-temporal domain approaches suggested earlier. The practical estimation method depend on obtaining the Transforms and the minimization can be easily performed using routines readily available in standard software (say using R) as is recommended in the time series literature (may need minor changes). In our illustrations, we found fast convergence of the estimates.

8 Acknowledgements

The research of the second author (Das) was pursued at the University of Manchester, with sponsorship from British Council under the UKIERI research grant. The researchers Das and Subba Rao were also partially supported by a grant from Department of Science and Technology, Govt. of Indian, New Delhi, while visiting the CRRAO AIMSCS, Hyderabad(India). The second author (Das) is grateful to Prof. Arup Bose, Indian Statistical Institute, Kolkata, India, for providing financial support during his stay at ISI. The authors thankful to Dr. Jyotishman Bhowmick of University of Oslo and Dr. Suhasini Subbarao of Texas A & M University, Texas, USA, for invaluable suggestions.

References

[1] P. Billingsley, *Convergence of probability measures*, Wiley New York, 1968.

[2] W.L. Briggs et al., *The dft: an owner's manual for the discrete fourier transform*, no. 45, Society for Industrial Mathematics, 1995.

[3] D.R. Brillinger, *Time series: data analysis and theory*, Society for Industrial Mathematics, 2001.

[4] P.J. Brockwell and R.A. Davis, *Time series: theory and methods*, Springer, 1991.

[5] N. Cressie, *Fitting variogram models by weighted least squares*, Mathematical geology **17** (1985), no. 5, 563–586.

[6] N. Cressie and H.C. Huang, *Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions.*, Journal of the American Statistical Association **94** (1999), no. 448.

[7] N.A.C. Cressie, *Statistics for spatial data*, John Wiley & Sons, New York, 1993.

[8] N.A.C. Cressie and C.K Wikle, *Statistics for Spatio-Temporal Data*, John Wiley & Sons, New York, 2011.

[9] Y. Dwivedi and S. Subba Rao, *A test for second-order stationarity of a time series based on the discrete fourier transform*, Journal of Time Series Analysis **32** (2011), no. 1, 68–91.

[10] T. Gneiting, *Nonseparable, stationary covariance functions for space-time data*, Journal of the American Statistical Association **97** (2002), no. 458, 590–600.

[11] T. Gneiting, M.G. Genton, and P. Guttorp, *Geostatistical space-time models, stationarity, separability, and full symmetry*, 2007, pp. 151–175.

[12] J.T. Kent, M. Mohammadzadeh, and A.M. Mosammam, *The dimple in gneiting's spatial-temporal covariance model*, Biometrika **98** (2011), no. 2, 489–494.

[13] J. Lee and S. Subba Rao, *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 (2011).

[14] C. Ma, *Spatio-temporal covariance functions generated by mixtures*, Mathematical geology **34** (2002), no. 8, 965–975.

[15] G. Matheron, *Principles of geostatistics*, Economic geology **58** (1963), no. 8, 1246.

[16] Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team, *nlme: Linear and nonlinear mixed effects models*, 2012, R package version 3.1-104.

[17] M.B. Priestley, *Spectral analysis and time series*, (1981).

[18] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012, ISBN 3-900051-07-0.

[19] M. Sherman, *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*, John Wiley & Sons, New York, 2010.