A space-time stochastic model and a covariance function for the stationary spatio-temporal random process and spatio-temporal prediction (kriging)

Journal:	*Journal of Time Series Analysis*
Manuscript ID	JTSA-3963.R4
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Subba Rao, Tata; University of Manchester, Mathematics; Terdik, Gyorgy; University of Debrecen, IT
Keywords:	Complex Stochastic Partial Di¤erential Equations, Covariance functions, Space-Time Processes, Prediction(kriging), Frequency Variogram, Whittle likelihood
Abstract:	Consider a stationary spatio-temporal random process with a sample Our object here is to estimate values at a given locations using the frequency domain approach. To obtain an esti- mator, we de fine a sequence of Discrete Fourier transforms at the Fourier frequencies using the time series observed at the locations si; (i = 1; 2; 3:::m), and use these complex random variables as our observations. Assuming the complex valued process satisfi es a complex stochastic partial di¤erential equation of the Laplacian type, and using the properties of Fourier transforms of stationary processes, we obtain an expression for the spatio-temporal covariance function and for the spectral density function. The covariance function of the Dis- crete Fourier transforms has shown to be a function of the Euclidean distance and the temporal frequency and its second order spectrum corresponds to non separable class of ran- dom process. We show further that the model defi ned here includes as special cases the spatio-temporal models de ned by Jones and Zhang [1997], Lindgren et al. [2011] and Sigrist et al. [2015]. The estimation of the parameters of the spatio-temporal covariance function has also been considered. The data considered is Air Pollution data (Particulate Matter PM2:5).

A space-time stochastic model and a covariance function for the stationary spatio-temporal random process and spatio-temporal prediction (kriging)

T. Subba Rao
University of Manchester, U.K. and
C. R. Rao AIMSCS, University of Hyderabad Campus, India,
tata.subbarao@gmail.com

Gyorgy Terdik
University of Debrecen, Hungary,
Terdik.Gyorgy@inf.unideb.hu

November 22, 2016

Abstract

Consider a stationary spatio-temporal random process $\{Y_t(\mathbf{s})|\mathbf{s}\in\mathbb{R}^d,\ t\in\mathbb{Z}\}$ and let $\{Y_t(\mathbf{s}_i)|i=1,2,...m;\ t=1,\dots,n\ \}$ be a sample from the random process $\{Y_t(\mathbf{s})\}$. Our object here is to predict $\{Y_t(\mathbf{s}_o)\}$ for all t at the location $\mathbf{s}_o$ given a sample $\{Y_t(\mathbf{s}_i)|\ i=1,2,...m;\ t=1,\dots,n\}$. To obtain the predictors, we define a sequence of Discrete Fourier transforms $\{J_{\mathbf{s}_i}(\omega_j);\ i=1,2,...m\}$ using the time series observed at the locations $\mathbf{s}_i$ $(i=1,2,3...m)$. We use these complex valued random variables as a sample from the complex valued random process $\{J_{\mathbf{s}}(\omega)\}$. The Fourier transforms are now functions of the spatial coordinates only. Assuming that the Discrete Fourier Transforms satisfy a complex stochastic partial differential equation (CSPDE) of the Laplacian type with a scaling function which is a polynomial in the temporal spectral frequency ω, we obtain, in a close form, expressions for the spatio-temporal covariance function and for the spectral density function. These expressions are used in obtaining the optimal predictors. The spectral density function obtained corresponds to a non-separable random process. The optimal predictor of the DFT $J_{\mathbf{s}_o}(\omega)$ is in terms of the covariance function of the DFT's. The covariance function of the Discrete Fourier transforms at two distinct locations,under isotropy condition, has been shown to be a function of the Euclidean distance$||\mathbf{h}||$ and the temporal frequency ω. We show further that the CSPDE model defined here includes as special case the spatio-temporal models defined by Jones and Zhang [1997], Lindgren et al. [2011] and Sigrist et al. [2015]. The estimation of the parameters of the spatio-temporal covariance function has also been considered. The method of estimation is based on Frequency Variogram(FV) approach recently introduced and the method does not involve inversion of large dimensional matrices and also found to be robust against departure from Gaussianity. The Discrete Fourier Transforms, can be evaluated using Fast Fourier Transform algorithms, and, therefore, the estimation methods are quick compared to other methods. The methods are illustrated with a real data. The data considered is Air Pollution data (Particulate Matter $PM_{2.5}$) recorded at 15 locations in New York city observed over a period of 9 months. The frequency domain methods given above for prediction and estimation can be extended to situations where the observations are corrupted by independent white noise process.

Keywords: Complex Stochastic Partial Differential Equations, Covariance Functions, Discrete Fourier Transforms, Measurement Errors, Spatio-Temporal Processes, Prediction (Kriging), Frequency Variogram.

1 Introduction and Summary

In recent years it has become necessary to develop statistical methods for the analysis of data coming from diverse areas such as, environment, marine biology, agriculture, finance etc. The data which comes from these areas, are usually, functions of both space and time. Any statistical method developed must take into account both spatial dependence, temporal dependence and any interaction between space and time. There is a vast literature on statistical analysis of stationary spatial data (for example refer to the books of Cressie [1993], Stein [1999]) but not to the same extent in the case of stationary spatio-temporal data. The inclusion of an extra temporal dimension, which cannot be embedded into spatial dimension gives raise to many problems. One such problem is finding a suitable covariance function which is positive semi-definite and depends on spatial lag difference and temporal lag. In recent years several authors (see Cressie and Huang [1999], Gneiting [2002], Diggle and Ribeiro [2007], Stein [2005a], Craigmile and Guttorp [2011], Jones and Zhang [1997], Ma [2002], Ma [2003], Lindgren et al. [2011], and Sigrist et al. [2015]) have proposed various covariance functions, and majority of them are Matern type of functions. Jones and Zhang [1997], Lindgren et al. [2011] and Sigrist et al. [2015] have considered transport-diffusion Stochastic Partial Differential (SPDE) equations for modelling of stationary spatio-temporal random process. The models defined by these authors are stochastic versions of the classical Heat equation, which is in fact a dynamic form of the Laplace equation, and the temporal correlation is explained by the inclusion of a first order time derivative in the operator. Lindgren et al. [2011] considered the approximation of Gaussian Field (GF) models by the Gaussian Markov Random Field (GMRF) models and considered modelling the data by GMRF models. Sigrist et al. [2015] have approximated the solution of the SPDE models by a linear combination of deterministic spatial functions (Fourier functions in terms of spatial wave numbers) with random coefficients that evolve dynamically and requires discretization in time for application to discrete time series.

Though for the processes defined by the above SPDE models one can obtain explicit expressions for the space-time spectrum, no such expressions in a closed form for the space-time covariance function are available (see Sigrist et al. [2015]).Such expressions are required for prediction. An alternative to the covariance based approach for prediction is to model the spatio-temporal processes dynamically and use Bayesian methodology for obtaining the predictive distribution. Such models are called Hierarchical Dynamic Spatio-temporal (HDST) models (see Cressie and Wikle [2011]). In Hierarchical modelling, models are specified at various levels together with parameters along with their prior distributions, and using these stated distributions one obtains the predictive distribution (see Banerjee et al. [2014]). This method requires evaluation of multiple integrals, analytic evaluation of such integrals are not feasible. Therefore, often MCMC techniques are used in such evaluations.

The novel features of our present paper are as follows. We consider the discrete Fourier transforms of the given spatio-temporal data and treat these complex Gaussian variables as our data. We define the Fourier transforms which are complex valued through a stochastic partial differential equation in spatial coordinates of the Laplacian type with a scaling function which is a polynomial in the temporal frequency. By defining in this way and operating on the Fourier transforms, the dependency of the operator on time is removed. By analyzing the DFTs and using the defined Complex Stochastic Partial Differential Equations (CSPDE), we are reducing the number of computations required for evaluating the predictors, the estimation of the parameters of the covariance function etc. Under the assumption of isotropy, we obtain an expression for the covariance of the Discrete Fourier Transforms at different locations, and it is shown to be in terms of the modified Bessel function (Matern Class). The expression for the covariance function given here is fundamentally different from the other covariance functions defined and used by other authors. We further show that the second order spectral density function of the spatio-temporal random process defined here through the above operator on the discrete Fourier transforms includes all the second order spectra of the processes so far defined through the stochastic versions of the Laplace equation such as, Heat equation (transport-diffusion equation of Jones and Zhang [1997], Lindgren et al. [2011] and Sigrist et al. [2015]), Wave equation and Helmholtz equation as special cases. The second order spectral density function obtained here belongs to nonseparable random process.

We summarize the contents of each section. In section 2, the notation and the spectral representation of the spatio-temporal random process are introduced. The properties of the Discrete Fourier transforms of second order stationary processes are discussed in the appendix 2 given at the end. Expressions for the spatio-temporal covariance function and for the spectral density function when the Discrete Fourier Transforms satisfy a Complex Stochastic Partial Differential equation (CSPDE) are obtained in section 3. We also show in this section that the second order spectra of the processes satisfying the models defined by Jones and Zhang [1997], and Sigrist et al. [2015] can be obtained as special cases of the CSPDE model defined here. The main results related to the process satisfying CSPDE are stated in theorems 1 and 2. The prediction of the entire data at a known location given the data in the neighborhood using the Discrete Fourier transforms is considered in section 4. The estimation of the parameters, using Frequency Variogram method, of the spatio-temporal covariance function of the DFT's is considered in section 4. 1. In section 5, prediction and estimation of the parameters when the observations are corrupted by white noise are considered. In section 6,we give an algorithm for generating a spatio-temporal time series with a given second order space-time spectral density and also discuss estimation and prediction for the simulated data.In section 7, the analysis of the Air Pollution data (Particulate Matter $PM_{2.5}$) collected at 15 locations in New York City is considered. In section 8 (appendix) some well known results related to Discrete Fourier Transforms are included.

2 Notation and Preliminaries

Let $Y_t(\mathbf{s})$, where $\left\{\mathbf{s} \in \mathbb{R}^d, t \in \mathbb{Z}\right\}$, denote the spatio-temporal random process. We assume that the random process is spatially and temporally second order stationary, i. e.

$$E\left[Y_t(\mathbf{s})\right] = \mu,$$
$$Var\left[Y_t(\mathbf{s})\right] = \sigma_Y^2 < \infty,$$
$$Cov\left[Y_t(\mathbf{s}), Y_{t+u}(\mathbf{s}+\mathbf{h})\right] = c(\mathbf{h}, u), \quad \mathbf{h} \in \mathbb{R}^d,\, u \in \mathbb{Z}.$$

We note that $c(\mathbf{h}, 0)$ and $c(\mathbf{0}, u)$ correspond to the purely spatial and purely temporal covariances of the process respectively. A further common stronger assumption that is often made is that the process is isotropic. The assumption of isotropy is a stronger assumption. The process is said to be isotropic if

$$c(\mathbf{h}, u) = c(\|\mathbf{h}\|, u), \quad \|\mathbf{h}\| \geq 0,\, u \in \mathbb{Z},$$

where $\|\mathbf{h}\|$ is the Euclidean distance. Without loss of generality, we set μ equal to zero. As in the case of spatial process, one can define the spatio-temporal variogram for $\{Y_t(\mathbf{s})\}$ as

$$2\gamma(\mathbf{h}, u) = Var\left[Y_{t+u}(\mathbf{s}+\mathbf{h}) - Y_t(\mathbf{s})\right]. \tag{1}$$

If the random process $\{Y_t(\mathbf{s})\}$ is spatially and temporally stationary, then we can rewrite the above as

$$2\gamma(\mathbf{h}, u) = 2\left[c(\mathbf{0}, 0) - c(\mathbf{h}, u)\right], \tag{2}$$

and for an isotropic process, $\gamma(\mathbf{h}, u) = \gamma(\|\mathbf{h}\|, u)$. We note that $\gamma(\mathbf{h}, u)$ is defined as the semi-variogram.

In view of our assumption that the zero mean random process $\{Y_t(\mathbf{s})\}$ is second order spatially and temporally stationary, we have the spectral representation

$$Y_t(\mathbf{s}) = \int_{-\infty}^{\infty}\int_{-\pi}^{\pi} e^{i(\mathbf{s}\cdot\underline{\lambda}+t\omega)} dZ_Y(\underline{\lambda}, \omega), \tag{3}$$

where $\mathbf{s}\cdot\underline{\lambda} = \sum_{i=1}^{d} \mathbf{s}_i\lambda_i$ and $\int_{-\infty}^{\infty}$ represents $d-$fold multiple integral. We note that $Z_Y(\underline{\lambda}, \omega)$ is a zero mean complex valued random process with orthogonal increments with

$$E\left[dZ_Y(\underline{\lambda}, \omega)\right] = 0,$$
$$E\left|dZ_Y(\underline{\lambda}, \omega)\right|^2 = dF_Y(\underline{\lambda}, \omega), \tag{4}$$

where $dF_Y(\underline{\lambda}, \omega)$ is a spectral measure. If we assume further that $dF(\underline{\lambda}, \omega)$ is absolutely continuous with respect to Lebesgue measure then $dF(\underline{\lambda}, \omega) = f(\underline{\lambda}, \omega)\, d\underline{\lambda}\, d\omega$. Here $f(\underline{\lambda}, \omega)$ which is strictly positive and real valued, is defined as the spatio-temporal spectral density function of the random process $\{Y_t(\mathbf{s})\}$, and $-\infty < \lambda_1, \lambda_2, \ldots, \lambda_d < \infty$, $-\pi \leq \omega \leq \pi$. In view of the orthogonality of the function $Z_Y(\underline{\lambda}, \omega)$, we can show that the positive definite covariance function $c(\mathbf{h}, u)$ has the representation

$$c(\mathbf{h}, u) = \int_{-\infty}^{\infty}\int_{-\pi}^{\pi} e^{i(\mathbf{h}\cdot\underline{\lambda}+u\omega)} f(\underline{\lambda}, \omega)\, d\underline{\lambda}\, d\omega, \tag{5}$$

and by Fourier inversion, we have

$$f(\underline{\lambda}, \omega) = \frac{1}{(2\pi)^{d+1}} \sum_{u} \int_{-\infty}^{\infty} e^{-i(\mathbf{h}\cdot\underline{\lambda}+u\omega)} c(\mathbf{h}, u)\, d\mathbf{h}, \tag{6}$$

where $d\mathbf{h} = \prod_{i=1}^{d} dh_i$. From equation (5) we obtain

$$c(0,u) = \int_{-\infty}^{\infty}\int_{-\pi}^{\pi} e^{iu\omega} f(\underline{\lambda},\omega)\, d\underline{\lambda}\, d\omega = \int_{-\pi}^{\pi} e^{iu\omega} g_0(\omega)\, d\omega,$$

where $g_0(\omega) = \int f(\underline{\lambda},\omega)\, d\underline{\lambda}$ is the temporal spectrum of the spatio-temporal random process $\{Y_t(\mathbf{s})\}$. In view of our assumption of spatial stationarity $g_0(\omega)$ is same for all $\mathbf{s}$. It may be pointed out here that a study of the properties of the second order temporal spectrum in the spatio-temporal context can be of considerable interest in several scientific fields, for example in neurosciences as shown by Ombao et al. [2008]. Ombao et al. [2008] considered the estimation of temporal spectrum at a given location assuming that spatio-temporal process is slowly spatially changing using a methodology similar to Priestley [1965] for estimating the evolutionary spectra Here, we consider the estimation of $g_0(\omega)$ assuming that the process is spatially and temporally stationary. From the above relation, we obtain by inverting $g_0(\omega) = \frac{1}{2\pi}\sum e^{-iu\omega}c(0,u)$. We further note that if the process is fully symmetric (see Gneiting [2002]) then $c(\mathbf{h},u) = c(-\mathbf{h},-u)$ and $f(\underline{\lambda},\omega) = f(-\underline{\lambda},-\omega)$ and $f(\underline{\lambda},\omega) > 0$ for all $\underline{\lambda}$ and ω. Here $\underline{\lambda}$ is the frequency associated with spatial coordinates (usually called the wave number) and ω is the temporal frequency. In the following we define the discrete Fourier Transforms of the stationary process and summarise in the Appendix their well known properties which will be used (for details refer to Brillinger [2001], Giraitis et al. [2012]).

Let $\{Y_t(\mathbf{s}_i) \,|\, i = 1,2,\ldots m;\ t = 1,\ldots,n\}$ be a sample from the zero mean spatio-temporal stationary process $\{Y_t(\mathbf{s})\}$. Consider the time series data $\{Y_t(\mathbf{s}_i) \,|\ t = 1,\ldots,n\}$ at the location $\mathbf{s}_i$, and define the Discrete Fourier transform (DFT)

$$J_{\mathbf{s}_i}(\omega_k) = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n} Y_t(\mathbf{s}_i)\, e^{-it\omega_k}, \tag{7}$$

where $\omega_k = \frac{2\pi k}{n}$, $k = 0, \pm 1, \ldots, \pm\left[\frac{n}{2}\right]$. In practice one uses Fast Fourier Transform algorithm to compute the DFT. From the above, by inversion we get

$$Y_t(\mathbf{s}) = \sqrt{\frac{n}{2\pi}}\int_{-\pi}^{\pi} J_{\mathbf{s}}(\omega)\, e^{it\omega} d\omega. \tag{8}$$

The above integral representation shows that the process can be decomposed into various sine and cosine terms and complex valued DFT's as the amplitudes. We also see from the above that there is a one to one correspondence between the DFT's and the data, a property we use later for prediction.

We may point out that one may define the m dimensional vector $\mathbf{Y}_t' = (Y_t(\mathbf{s}_1), Y_t(\mathbf{s}_2), \ldots Y_t(\mathbf{s}_m))$, (for $t = 1,2,\ldots n$) and analyze the data by using the classical multivariate time series methodology. The usual assumption here is that the elements of the vector temporally jointly stationary. The spatial dependence is not taken into account if analyzed in this way. Also,modelling this way may lead to over parametrisation (see Antunes and Subba Rao [2006]). In view of this, Pfeifer and Deutrch [1980] suggested Space Time ARMA (STARMA)models as an alternative to the multivariate ARMA models where known weighing matrices with elements which are functions of spatial dependence (for example Euclidean distances) are included as coefficients. The choice of weights, definition of neighborhood are arbitrary in this approach.We refer to Stein [2005a], Stein [2005b] for further discussion on the advantages of modelling spatio-temporal data compared to modelling using the classical multivariate time series approach.

3 A Complex Stochastic Partial Differential Equation (CSPDE) and an Expression for the Spectrum $g_{||\mathbf{h}||}(\omega)$

In the following we consider the spatial model proposed by Whittle [1954], spatio-temporal models proposed by Jones and Zhang [1997], Lindgren et al. [2011], and Sigrist et al. [2015] and derive the spectral properties of the processes satisfying these models. Our object here is to define an alternative frequency domain based complex stochastic partial differential equation for the discrete Fourier transforms and derive expressions for the spectrum and the covariance function of the process satisfying the CSPDE model. We show that the spectra obtained from the stochastic partial differential equations defined by Whittle [1954], Jones and Zhang [1997], Lindgren et al. [2011], Sigrist et al. [2015] can be derived as special cases of the CSPDE model defined here.

It is well known that to study turbulence, dissipation of heat or fluid, equations such as Laplace equation, Heat equation, Wave equation and Helmholtz equation are often used. Laplace equation is used to describe the static behavior of the material (say fluid) where as Heat and Wave equations are used to describe the dynamic behavior and are usually called Diffusion equations. Stochastic version of the Laplace equation was used by Whittle [1954] to study the correlation pattern of soil fertility in agricultural uniformity trials at different locations and by studying the solution of the Stochastic Laplace equation, Whittle [1954] has shown that the correlation of the yields at points at 's' units apart falls off as a power of 's^{-1}', a property observed by agricultural scientists. We briefly discuss the models proposed by the above researchers.

3.1 Stochastic version of the Laplace Equation (Whittle [1954])

For illustration purposes, let us assume $d = 2$. Let $Y(\mathbf{s})$ (here the spatial coordinates are denoted by $\mathbf{s} = (s_1,\ s_2)$) denote the stationary spatial random field. In the case of the example considered by Whittle [1954], $Y(\mathbf{s})$ denotes the yield at the location s. Let $\bigtriangledown = \frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}$ be the Laplace operator. Whittle [1954] defined the model $(\bigtriangledown - \gamma^2)Y(\mathbf{s}) = e(\mathbf{s})$, where $e(\mathbf{s})$ is defined as spatial Gaussian white noise and γ is a scale parameter. We can obtain an expression for the spectrum of the process satisfying the above model by considering the spectral representation of the process given by $Y(\mathbf{s}) = \int e^{i\mathbf{s}.\underline{\lambda}}\ dZ_y(\underline{\lambda})$, where $Z_Y(\underline{\lambda})$ is an orthogonal function with $E(dZ_Y(\underline{\lambda})) = 0$, and $E|dZ_Y(\underline{\lambda})|^2 = f_Y(\underline{\lambda})d\underline{\lambda}$. Here $\underline{\lambda} = (\lambda_1,\ \lambda_2)$ corresponds to the spatial frequency (known as wave number), $f_Y(\underline{\lambda})$ is defined as the spatial spectrum. We can define a similar spectral representation for Gaussian white noise process $e(\mathbf{s})$ with $Z_e(\underline{\lambda})$ denoting the orthogonal random set function of the process. By substituting the spectral representations for $Y(\mathbf{s})$ and $e(\mathbf{s})$ and equating the integrands and taking expectations of the modulus squares both sides, we can show that the spectral density function of $Y(\mathbf{s})$ is given by $f_Y(\underline{\lambda}) \propto 1/(\lambda_1^2 + \lambda_2^2 + \gamma^2)^2$.

If the process is isotropic (see Stein and Weiss [1971], Ch. IV. Theorem 1. 1), we can show by inversion,that the corresponding spatial covariance function at Euclidean distance $||\mathbf{h}||$ is given by $(||\mathbf{h}||/\gamma)K_1(\gamma||\mathbf{h}||)$, where $K_1(.)$ is the modified Bessel function of the second kind of the first order. The covariance function obtained belongs to Matern Class of covariance functions. The above model is a static version of the dynamic model considered below.

3.2 Stochastic version of the Heat Equation (Jones and Zhang [1997])

Now consider the model $[\frac{\partial}{\partial t} + (\bigtriangledown - \gamma^2)]\ Y_t(\mathbf{s}) = e_t(\mathbf{s})$, where the white noise process $\{e_t(\mathbf{s})\}$ is defined as above. The models considered by Lindgren et al. [2011] and recently by Sigrist et al. [2015] are variations of the above diffusion model. If we set transport direction vector (see Sigrist et al. [2015]) zero and diffusion matrix to identity matrix in the models by Lindgren et al. [2011] and Sigrist et al. [2015], we get the model defined by Jones and Zhang [1997]. By substituting the spectral representations for the processes $\{Y_t(\mathbf{s})\}$ and $\{e_t(\mathbf{s})\}$ and equating the integrands, and after taking expectations we can show that the spatio-temporal spectral density function of the process $\{Y_t(\mathbf{s})\}$ is given by $f_Y(\underline{\lambda}, \omega) \propto 1/[((\lambda_1^2 + \lambda_2^2 + \gamma^2)^2 + \omega^2]$. We note that no closed form for the covariance function is available in this case (see Sigrist et al. [2015]).

3.3 Stochastic version of the Wave Equation

This is a dynamic stochastic version of the classical Wave equation used to describe sound waves, water waves, light waves arising in fields like acoustics, fluid dynamics etc. We are interested in the statistical properties of the process. Consider the model $[\frac{\partial^2}{\partial t^2} + (\bigtriangledown - \gamma^2)]\ Y_t(s) = e_t(\mathbf{s})$. By substituting the spectral representations, and taking expectations, we can show that the spatio-temporal spectrum is given by $f_Y(\underline{\lambda}, \omega) \propto 1/[(\lambda_1^2 + \lambda_2^2 + \gamma^2) + \omega^2]^2$. As in the case of the Heat equation, no closed form for the space time covariance function is available.

3.4 CSPDE and an expression for the spectrum $g_{||\mathbf{h}||}(\omega)$

We note that the above equations considered by Jones and Zhang [1997], Lindgren et al. [2011], and Sigrist et al. [2015], include a first order time derivative $\frac{\partial}{\partial t}$ in the operators. This is equivalent to assuming that the temporal dynamics in the spatio-temporal process can be explained by an autoregressive model of order one and in a similar way the inclusion of the second order time derivative in the wave equation is equivalent to assuming that the temporal dynamics can be explained by an autoregressive model of order two. These specific assumptions can be unrealistic in some situations. In view of this, we propose a model which includes a nonparametric function which is a polynomial in $e^{-i\omega}$, and by including this function in our operator, we can derive the spectra defined by the processes satisfying the above models as special cases. To arrive at the model, let us consider once again the Laplace operator $(\bigtriangledown - \gamma^2)$ operating on the process $\{Y_t(\mathbf{s})\}$. We have shown (see equation (8),Section 2)

$$Y_t(\mathbf{s}) = \sqrt{\frac{n}{2\pi}} \int_{-\pi}^{\pi} J_{\mathbf{s}}(\omega) e^{it\omega} d\omega.$$

Multiplying both sides of the above equation by the operator $(\bigtriangledown - \gamma^2)$, we get

$$(\bigtriangledown - \gamma^2) Y_t(\mathbf{s}) = \sqrt{\frac{n}{2\pi}} \int_{-\pi}^{\pi} (\bigtriangledown - \gamma^2) J_{\mathbf{s}}(\omega) e^{it\omega} d\omega.$$

This relation shows that operating on the process $Y_t(\mathbf{s})$ is equivalent to operating on the complex valued DFT $J_{\mathbf{s}}(\omega)$ at the fixed frequency ω and then integrating over all the frequencies. In other words, just like the interpretation we have for the spectral representation which is a frequency decomposition of the process in terms of sine and cosine functions and the contribution of each frequency is measured by the corresponding amplitude $J_{\mathbf{s}}(\omega)$, the above frequency domain modelling is equivalent to modelling the complex valued process (DFT) for each frequency ω. Heuristically, one can interpret the Heat equation and Wave equation etc. to describe what happens to the process locally in space during a small time interval. In a similar way one can interpret the CSPDE to describe what happens to the DFT in space during a small frequency interval. We see from the above, that the two representations are mathematically equivalent. Later we will obtain an expression for the covariance function of the complex valued process $\{J_{\mathbf{s}}(\omega)\}$ satisfying CSPDE which will be in terms of the temporal spectrum $g_0(\omega)$ and the spatial distance $||h||$. We will show this covariance function is useful for spatio-temporal prediction considered in this paper (see section 4). In the course of the derivation of this result, we will also obtain an expression for the second order spectrum which is a function of the spatial frequency (wave number) and the temporal frequency. The obtained spectrum is strictly greater than zero implying that the corresponding spatio-temporal covariance function is positive definite. Also the spectrum obtained is non separable.

We now show that the above models can be considered as special cases of the following CSPDE model.

Lemma 1 *Let d=2. Consider the complex stochastic partial differential equation*

$$\left[\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2} - \gamma(\omega)\right] J_{\mathbf{s}}(\omega) = J_{\mathbf{s},e}(\omega), \tag{9}$$

where $\gamma(\omega)$ is a complex valued function. Let $\gamma(\omega) = c(\omega) + i\, b(\omega)$. Then the spectral density function of the process $\{Y_t(\mathbf{s})\}$ satisfying the above model is given by

$$f_Y(\underline{\lambda}, \omega) \propto \frac{1}{[((\lambda_1^2 + \lambda_2^2 + c(\omega))^2 + b^2(\omega)]}, \tag{10}$$

Proof. We can proceed as before to obtain the above expression and hence the proof is omitted. ■

Let us now consider the following special cases.

3.5 Special Cases:

1. Let $c(\omega) = \gamma^2$, $b(\omega) = \omega$. Substitute these in the equation $f_Y(\underset{\sim}{\lambda}, \omega)$ given above, we obtain the spectrum of the process satisfying the Heat equation considered by Jones and Zhang [1997], Lindgren et al. [2011] and Sigrist et al. [2015].

2. To obtain the spectrum of the Wave equation, let $\gamma(\omega)$ be real valued, and let $\gamma(\omega) = c(\omega) = \gamma^2 + \omega^2$. Substitution of this gives us the spectrum corresponding to the Wave equation.

Through the above examples we have shown that by defining the stochastic version of the Laplacian model in terms of the frequency dependent non-parametric function $\gamma(\omega)$, the second order properties of the classical equations can be obtained as special cases.

We will now state the main model and derive expressions for the spectrum and for the covariance function $g_{\|\mathbf{h}\|}(\omega)$ of the DFT's at two different locations which are functions of the Euclidean distance $\|\mathbf{h}\|$ and temporal spectral frequency ω. We will state the results for $d = 2$ and later consider its generalization for all d. Stein [2005a], Stein [2005b] defines the covariance function $g_{\|\mathbf{h}\|}(\omega)$ as spectral in time but not in space.

Theorem 1 *Let $J_{\mathbf{s}}(\omega)$ be the discrete Fourier transform of the data $\{Y_t(\mathbf{s}) | t = 1, 2, \ldots n\}$ at the location* $\mathbf{s}$. *Let $\upsilon > 0$, and let $\{J_{\mathbf{s}}(\omega)\}$ satisfy the model*

$$\left[\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2} - |c(\omega)|^2\right]^{\nu} J_{\mathbf{s}}(\omega) = J_{\mathbf{s},e}(\omega), \tag{11}$$

where $J_{\mathbf{s}}(\omega)$ and $J_{\mathbf{s},e}(\omega)$ are given by (51) and (53). Then the second order spectral density function $f_Y(\underset{\sim}{\lambda}, \omega)$ of the process$\{Y_t(\mathbf{s})\}$ is given by

$$f_Y(\underset{\sim}{\lambda}, \omega) = \frac{\sigma_e^2}{(2\pi)^3 \left(\lambda_1^2 + \lambda_2^2 + |c(\omega)|^2\right)^{2\nu}}. \tag{12}$$

If the stationary spatio-temporal process is isotropic, then the covariance function between the discrete Fourier Transforms $J_{\mathbf{s}}(\omega)$ and $J_{\mathbf{s}+\mathbf{h}}(\omega)$ is given by

$$g_{\|\mathbf{h}\|}(\omega) = Cov(J_{\mathbf{s}}(\omega), J_{\mathbf{s}+\mathbf{h}}(\omega)) = \frac{\sigma_e^2}{(2\pi)^2} \left(\frac{\|\mathbf{h}\|}{2|c(\omega)|}\right)^{2\nu-1} \frac{K_{2\nu-1}(|c(\omega)| \|\mathbf{h}\|)}{\Gamma(2\nu)}, \tag{13}$$

where $\|\mathbf{h}\| = \left(h_1^2 + h_2^2\right)^{\frac{1}{2}}$ and $K_\nu(x)$ is the modified Bessel function of the second kind of order ν.

Later we will see the significance of inclusion of the frequency dependent function $c(\omega)$ in the above model. We see that by defining the model(11) in terms of Discrete Fourier Transforms , we embedded one component ,namely,$c(\omega)$ which is related to the second order temporal spectrum of the process $\{Y_t(\mathbf{s})\}$into the Laplacian model. The model thus defined takes into acccount spatial correlation ,temporal correlation and their interaction.

Proof. Substitute the spectral representations for $J_{\mathbf{s}}(\omega)$ and $J_{\mathbf{s},e}(\omega)$ given by (51) and (53) respectively, and taking the operators inside the integrands and equating the integrands both sides of the equation (this is valid because of the uniqueness of the Fourier transforms), we obtain

$$\left(-\lambda_1^2 - \lambda_2^2 - |c(\omega)|^2\right)^{\nu} dZ_Y(\underset{\sim}{\lambda}, \omega) = dZ_e(\underset{\sim}{\lambda}, \omega), \tag{14}$$

where $\underset{\sim}{\lambda} = (\lambda_1, \lambda_2)$. Taking the modulus squares, and taking expectations both sides of the modulus squares we obtain the spatio-temporal spectral density function of the spatio-temporal process $Y_t(\mathbf{s})$ satisfying the above model (11) and it is given by

$$f_Y(\underset{\sim}{\lambda}, \omega) = \frac{\sigma_e^2}{(2\pi)^3 \left(\lambda_1^2 + \lambda_2^2 + |c(\omega)|^2\right)^{2\nu}}, \tag{15}$$

which is the stated result.
We note that the above spectral density is real and strictly positive, and this implies that the associated spatio-temporal covariance function is positive definite. Further the spectral density function given above

belongs to a nonseparable class of processes. Since the spectrum depends on the distance of the wave numbers $\lambda_1^2 + \lambda_2^2$ its Fourier transform will depend on the distance between locations as well (see Stein and Weiss [1971] Ch. IV. Theorem 1. 1.) , hence the process is isotropic. Now, to obtain the covariance function $g_{\|\mathbf{h}\|}(\omega)$, we need to take its inverse Fourier transform. We use the result used by Whittle [1954] (equation (65) of the paper)

$$\frac{1}{(2\pi)^2}\int\int \frac{e^{i(x\omega_1+y\omega_2)}}{(\omega_1^2+\omega_2^2+\alpha^2)^{\mu+1}}d\omega_1 d\omega_2 = \frac{1}{2\pi}\left(\frac{r}{2\alpha}\right)^{\mu}\frac{K_\mu(\alpha r)}{\Gamma(\mu+1)}, \tag{16}$$

where $r = (x^2+y^2)^{\frac{1}{2}}$, $K_\mu(x)$ is the modified Bessel function of the second kind of order μ. We use the above result to obtain the inverse transform of $f_Y(\underline{\lambda}, \omega)$, given by (15). Taking the inverse transform over the wave number $\underline{\lambda}$ only (for fixed temporal frequency ω), we obtain

$$\begin{aligned} g_{\|\mathbf{h}\|}(\omega) &= \frac{\sigma_e^2}{(2\pi)^3}\int\int \frac{e^{i(h_1\lambda_1+h_2\lambda_2)}}{\left(\lambda_1^2+\lambda_2^2+|c(\omega)|^2\right)^{2\nu}}d\lambda_1 d\lambda_2 \\ &= \frac{\sigma_e^2}{(2\pi)^2}\left(\frac{\|\mathbf{h}\|}{2|c(\omega)|}\right)^{2\nu-1}\frac{K_{2\nu-1}(|c(\omega)|\,\|\mathbf{h}\|)}{\Gamma(2\nu)}. \end{aligned} \tag{17}$$

■

The above interesting expression shows that the covariance function between two discrete Fourier Transforms of the spatial process separated by the spatial distance $\|\mathbf{h}\|$ again can be written in terms Matern - Whittle class of covariance functions.

However, the most important and fundamental difference between the above given expression and the covariance expressions given by other authors mentioned before is that the argument of the Bessel function derived above is not only a function of the spatial distance, but also a function of the frequency dependent scaling function $|c(\omega)|$ which is related to the second order temporal spectral density function. This will be shown in the following lemma.

To see the significance of inclusion of $|c(\omega)|$ in the model (11), we consider the limiting behavior of $g_{\|\mathbf{h}\|}(\omega)$ as $\|\mathbf{h}\| \to 0$. We have noted earlier that $Var(J_{\mathbf{s}}(\omega))$, is proportional to the spectral density function $g_0(\omega)$ of the random process for all s. So it is interesting to examine the behavior of $g_{\|\mathbf{h}\|}(\omega)$ when $\|\mathbf{h}\| \to 0$, as the limit must tend to the second order spectral density function $g_0(\omega)$ of the process $\{Y_t(\mathbf{s})\}$. We state the result in the following Lemma.

Lemma 2 *For the above isotropic process, and under the conditions stated above, as $\|\mathbf{h}\| \to 0$, $g_{\|\mathbf{h}\|}(\omega)$ tends to*

$$g_0(\omega) = \frac{\sigma_e^2}{2(2\pi)^2\left(|c(\omega)|^2\right)^{2\nu-1}(2\nu-1)}. \tag{18}$$

Proof. It is well known that, for all $\nu > 0$,

$$\lim_{x\to 0}\frac{x^\nu K_\nu(x)}{2^{\nu-1}\Gamma(\nu)} = 1. \tag{19}$$

Therefore, if we take the limit of $g_{\|\mathbf{h}\|}(\omega)$ given by (17) as $\|\mathbf{h}\| \to 0$, we get the stated result (18). ■

3.5.1 Special Case:

Let us consider the case $\nu = 1$. Then from the equation (17) we have

$$g_{\|\mathbf{h}\|}(\omega) = \frac{\sigma_e^2}{(2\pi)^2}\left(\frac{\|\mathbf{h}\|}{2|c(\omega)|}\right)K_1(|c(\omega)|\,\|\mathbf{h}\|), \tag{20}$$

and from the equation (18) we have

$$g_0(\omega) = \frac{\sigma_e^2}{2(2\pi)^2|c(\omega)|^2}, \tag{21}$$

which implies that $|c(\omega)|^2$ is proportional to $g_0^{-1}(\omega)$, which is defined as the inverse second order spectral density function of the process. Let us assume that $g_0^{-1}(\omega)$ is absolutely integrable, then $g_0^{-1}(\omega)$ can be expanded in Fourier series

$$g_0^{-1}(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} ci(k) \cos k\omega, |\omega| \leq \pi,$$

where we used the fact that $g_0^{-1}(\omega) = g_0^{-1}(-\omega)$. The coefficients $\{ci(k)\}$ are usually known as inverse autocovariances, and sometimes are used to estimate the orders of the linear time series models. For example, if the time series $\{Y_t(\mathbf{s})\}$ satisfies (for a given s) an autoregressive model of order p, say, then it can easily be shown that $ci(k) = 0$ for all $k > p$. In view of this interesting property one can use the inverse auto-covariances to determine the order of the linear AR models. We note further that the covariance function $g_{\|\mathbf{h}\|}(\omega)$ given above is in terms of the modified Bessel function, the argument of the Bessel function is a product of the spatial distance $\|\mathbf{h}\|$ and the inverse temporal spectrum $g_0^{-1}(\omega)$. Therefore the rate of convergence of the covariance function to tend to zero as $\|\mathbf{h}\| \to \infty$ depends on the second order temporal spectrum of the process at the frequency ω.

From (20) and (21) we can also obtain an expression for the auto-correlation function. We have the auto-correlation function when $d = 2$, and for all $\upsilon > 0$,

$$\begin{aligned} \rho(\|\mathbf{h}\|, \omega) &= \frac{g_{\|\mathbf{h}\|}(\omega)}{g_0(\omega)} \\ &= \frac{(\|\mathbf{h}\| |c(\omega)|)^{2\nu-1}}{2^{2\nu-2}\Gamma(2\nu-1)} K_{2\nu-1}(|c(\omega)| \|\mathbf{h}\|). \end{aligned} \tag{22}$$

It is interesting to note that $\rho(\|\mathbf{h}\|, \omega)$ is in fact the coherency coefficient between two Discrete Fourier Transforms separated by the spatial distance $\|\mathbf{h}\|$ at the frequency ω. We now consider the generalization of Theorem 1.

Theorem 2 *Let $\nu > 0$ and $d \geq 2$. Let the Discrete Fourier Transform $J_{\mathbf{s}}(\omega)$ satisfy the equation*

$$\left[\sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2} - |c(\omega)|^2\right]^{\nu} J_{\mathbf{s}}(\omega) = J_{\mathbf{s},e}(\omega), \tag{23}$$

Then the second order spectral density function is given by

$$f_Y(\underline{\lambda}, \omega) = \frac{\sigma_e^2}{(2\pi)^{d+1}} \frac{1}{\left(\sum \lambda_i^2 + |c(\omega)|^2\right)^{2\nu}}. \tag{24}$$

If the process is isotropic then the covariance function is given by

$$g_{\|\mathbf{h}\|}(\omega) = Cov(J_{\mathbf{s}}(\omega), J_{\mathbf{s}+\mathbf{h}}(\omega)) = \frac{\sigma_e^2}{(2\pi)^{\frac{d}{2}+1} 2^{2\nu-1}\Gamma(2\nu)} \left(\frac{\|\mathbf{h}\|}{|c(\omega)|}\right)^{2\nu-\frac{d}{2}} K_{2\nu-\frac{d}{2}}(\|\mathbf{h}\| |c(\omega)|), \tag{25}$$

and the autocorrelation function is given by

$$\rho(\|\mathbf{h}\|, \omega) = \frac{g_{\|\mathbf{h}\|}(\omega)}{g_0(\omega)} = \frac{(\|\mathbf{h}\| |c(\omega)|)^{2\nu-\frac{d}{2}}}{2^{2\nu-\frac{d}{2}-1}\Gamma\left(2\nu-\frac{d}{2}\right)} K_{2\nu-\frac{d}{2}}(\|\mathbf{h}\| |c(\omega)|). \tag{26}$$

Proof. See the Appendix 1. ■

In the following we consider prediction of the data, and the optimal predictive function is given in terms of the Discrete Fourier Transforms and the covariance function $g_{\|\mathbf{h}\|}(\omega)$.

4 Spatio-temporal Prediction

Our object in this section is to obtain an optimal predictor for $\{Y_t(\mathbf{s}); t = 1, 2, \ldots, n\}$ at the location $\mathbf{s}_0$ given the m spatial time series $\{Y_t(\mathbf{s}_i) \mid i = 1, 2, \ldots, m; t = 1, 2, \ldots, n\}$ from a spatio-temporal stationary, isotropic process $\{Y_t(\mathbf{s})\}$. In other words, we are predicting the entire data set at the location $\mathbf{s}_0$. Using the predicted data at the location $\mathbf{s}_0$, we can obtain the optimal predictors for the future values$\{Y_t(\mathbf{s});t=n+\nu,\nu\geqslant 0\}$ at the location $\mathbf{s}_0$. As in the case of the observed data $\{Y_t(\mathbf{s}_i)\}$, we define the discrete Fourier transform $J_{\mathbf{s}_0}(\omega)$ of $\{Y_t(\mathbf{s}_0)\}$, and predict the Fourier transform $J_{\mathbf{s}_0}(\omega)$ for all ω. Using the inverse Fourier Transform, we compute the predicted values of $Y_t(\mathbf{s}_0)$ for all $1\leq t\leq n$. We pointed out earlier that there is a one to one correspondence between the discrete Fourier Transforms and the data. We have shown that if

$$J_{\mathbf{s}_0}(\omega)=\frac{1}{\sqrt{(2\pi n)}}\sum_{t=1}^{n}Y_t(\mathbf{s}_0)\,e^{-it\omega}, \tag{27}$$

then we have

$$Y_t(\mathbf{s}_0)=\sqrt{\frac{n}{2\pi}}\int_{-\pi}^{\pi}J_{\mathbf{s}_0}(\omega)\,e^{it\omega}d\omega. \tag{28}$$

Consider the vector of the discrete Fourier transforms obtained from all m locations at the frequency ω,

$$\underline{J}'_m(\omega)=[J_{\mathbf{s}_1}(\omega),J_{\mathbf{s}_2}(\omega),\ldots,J_{\mathbf{s}_m}(\omega)].$$

We note that

$$\begin{aligned}E[\underline{J}_m(\omega)]&=0,\\ E[\underline{J}_m(\omega)\,\underline{J}^*_m(\omega)]&=F_m(\omega),\end{aligned} \tag{29}$$

where the real,$m\times m$ dimensional symmetric, positive definite square matrix $F_m(\omega)=\left(g_{\|s_i-s_j\|}(\omega);\ i,j=1,2,\ldots,m\right)$, and each element $g_{\|s_i-s_j\|}(\omega)$ of the matrix $F_m(\omega)$ is given by (17). The complex random vector $\underline{J}_m(\omega)$ has a multivariate complex Gaussian distribution with mean zero and variance covariance matrix $F_m(\omega)$. Consider now the $(m+1)$ dimensional complex valued random vector,

$$\underline{J}'_{m+1}(\omega)=\left[J_{\mathbf{s}_o}(\omega),\underline{J}'_m(\omega)\right].$$

It can be shown that the mean of the vector is zero, and the variance covariance matrix is given by

$$\begin{aligned}E\left[\underline{J}_{m+1}(\omega)\,\underline{J}^*_{m+1}(\omega)\right]&=\begin{bmatrix}E(J_{s_o}(\omega)\,J^*_{s_o}(\omega)) & E\left(J_{s_0}(\omega)\,\underline{J}^{*\prime}_m(\omega)\right)\\ E\left(\underline{J}_m(\omega)\,J^*_{s_0}(\omega)\right) & E\left(\underline{J}_m(\omega)\,\underline{J}^*_m(\omega)\right)\end{bmatrix}\\ &=\begin{bmatrix}g_0(\omega) & \underline{G}'_0(\omega)\\ \underline{G}_0(\omega) & F_m(\omega)\end{bmatrix},\end{aligned}$$

where $g_0(\omega)$ is the second order spectral density function of the spatial process $\{Y_t(\mathbf{s}_0)\}$ and the row vector $\underline{G}'_0(\omega)$ is given by

$$\begin{aligned}\underline{G}'_0(\omega)&=E\left[J_{\mathbf{s}_o}(\omega)\,J^{*\prime}_m(\omega)\right]\\ &=[g_{\|s_0-s_1\|}(\omega),g_{\|s_0-s_2\|}(\omega),\ldots,g_{\|s_0-s_n\|}(\omega)],\end{aligned}$$

and $F_m(\omega)$ is defined above. Therefore, the optimal linear least squares predictor of $J_{s_0}(\omega)$ given the vector $\underline{J}_m(\omega)$, is given by the conditional expectation

$$E\left[J_{\mathbf{s}_o}(\omega)\,|\,\underline{J}_m(\omega)\right]=\underline{G}'_0(\omega)\,F_m^{-1}(\omega)\,\underline{J}_m(\omega), \tag{30}$$

and the minimum mean square prediction error is given by

$$\sigma^2_m(\omega)=g_0(\omega)-\underline{G}'_0(\omega)\,F_m^{-1}(\omega)\,\underline{G}_0(\omega). \tag{31}$$

To predict the data $Y_t(\mathbf{s}_0)$ for all t, we use the inverse transform (28). In computing the predictor of $J_{\mathbf{s}_0}(\omega)$ using the expression (30) one usually replaces the elements of the matrices $\underline{G}_0(\omega)$ and $F_m(\omega)$ by their corresponding estimates and obtain

$$\widehat{Y}_t(\mathbf{s}_0)=\sqrt{\frac{n}{2\pi}}\int_{-\pi}^{\pi}e^{it\omega}\widehat{\underline{G}}'_0(\omega)\,\widehat{F}_m^{-1}(\omega)\,\underline{J}_m(\omega)\,d\omega. \tag{32}$$

It is interesting to compare the formulae (30) and (31) with the corresponding expressions obtained using the time domain approach (see Cressie and Wikle [2011], equations 6.49 and 6.50; Banerjee et al.

[2014], equations 11.21 and 11.22). They are similar, but it is important to note that for evaluating the expressions given in Cressie and Wikle [2011], Banerjee et al. [2014], one requires the inversion of matrices of order $mn \times mn$, where as the evaluation of (ref:eq.5.3) given above only requires the inversion of $m \times m$ order matrices. Of course, if one assumes the spatio-temporal process is separable, the dimensions of the variance covariance matrices reduce considerably in the time domain case. But the assumption of separability can be a severe restriction and may not always be feasible. Further, we note that the predictor of $J_{s_0}(\omega)$ given by (31) is optimal in view of the fact the DFT's are asymptotically distributed as complex Gaussian, where as the time domain predictors are optimal only under the assumption that the process is Gaussian. We also note that to evaluate the predictor $\widehat{Y}_t(\mathbf{s}_0)$,we need an expression for the covariance function $g_{\|s_i - s_j\|}(\omega)$. Any valid expression can be used for the evaluation purposes. In section 3, we obtained such an expression when the process $Y_t(s)$ (or its DFT) satisfies a specific embedded Laplacian model.

We note $E(Y_t(\mathbf{s}_0)) = 0$ and $Var(Y_t(\mathbf{s}_0)) \simeq \int_{-\pi}^{\pi} G_0'(\omega)\, F_m^{-1}(\omega)\, G_0(\omega)\, d\omega$. We can show by an application of Parseval's Theorem

$$E\left(Y_t(\mathbf{s}_0) - \widehat{Y}_t(\mathbf{s}_0)\right)^2 = E\left|\mathcal{F}^{-1}\left(J_{\mathbf{s}_o}(\omega) - \widehat{J}_{\mathbf{s}_o}(\omega)\right)\right|^2 = \int_{-\pi}^{\pi} \sigma_m^2(\omega)\, d\omega. \quad (33)$$

In practice, the above integrals are approximated by finite sums of the form

$$\widehat{Y}_t(\mathbf{s}_0) = \sqrt{\frac{2\pi}{n}} \sum_{j=0}^{n-1} e^{it\omega_j} \underline{\widehat{G}}_0'(\omega_j)\, \widehat{F}_m^{-1}(\omega_j)\, \underline{J}_m(\omega_j). \quad (34)$$

for all $t = 1,\ 2, \ldots\ n$, where the estimates $\underline{\widehat{G}}_0(\omega_j)$ and $\widehat{F}_m(\omega)$ are substituted for $\underline{G}_0(\omega)$ and $F_m(\omega)$ respectively. As noted earlier, the vector $\underline{G}_0(\omega)$ and the matrix $F_m(\omega)$ have covariance functions $g_{\|s_i - s_j\|}(\omega)$ as their elements. The covariance functions are functions of some unknown parameters which are related to the spatial correlation and temporal correlation. From the expression of the covariance function (17), we see that the parameters to be estimated are σ_e^2 (the variance of the white noise process $e_t(\mathbf{s})$), and the parameters of the spatio-temporal spectrum $g_0(\omega)$. Let us denote the parameter vector which characterizes $g_0(\omega)$ by ϑ_1 and the entire parameter vector by $\vartheta = (\sigma_e^2, \vartheta_1)$. The parameter ν is related to the smoothness of the process. In practice one considers several possible choices for $\nu(>0)$ a priory. The widely used choice is $\nu = 1$. The estimation of the parameter vector ϑ of the spatio-temporal covariance function $g_{\|\mathbf{h}\|}(\omega)$ is extremely important and this will be considered in the following section. We now make some comments on computational aspects.

As pointed out earlier,in this frequency domain approach described here, the evaluation of the conditional expectation and the calculation of the minimum mean square error require inversion of $m \times m$ dimensional matrices (where m corresponds to the number of locations)only, unlike in the case of time domain approach for prediction where one needs to invert $mn \times mn$ dimensional matrices. In many real data analysis usually the number of time points n will be very large (and m can be large too). Besides, there is no data ordering problem involved here (see Cressie and Wikle [2011] p. 324). Once we have an expression for the covariance function $g_{\|\mathbf{h}\|}(\omega)$, all the elements of the column vector $G_0(\omega)$ and the elements of $F_m(\omega)$ are known. By substituting the relevant expressions (or their estimates), we can evaluate (30) and (31).

It may be pointed out that there are other approaches for obtaining predictors in the context of spatio-temporal data. Sahu and Mardia [2005] used Bayesian approach based on MCMC, Ruiz-Medina [2012] and Giraldo et al. [2010] based their methodology on the assumption that the spatio-temporal data is of functional data type. Giraldo et al. [2010] assumed that the process can be expanded in terms of some chosen deterministic basis functions with random coefficients and the predictor can also be written as a linear combination of the same basis functions and the same number of terms. The solution depends on inversion of matrices whose dimensions depend not only on the number of locations and also on the number of Basis functions included in the expansion of the process and the estimator proposed. As pointed out earlier, an alternative is to model the process dynamically. In the dynamic spatio-temporal models approach,models are specified and the parameters are assumed to be either random with known prior probability distributions or fixed but unknown. In the case of fixed parameters,the parameters are estimated and substituted in the models, and the predictive distributions are obtained, and this approach is now known as Empirical Hierarchical modeling approach. For more details, we refer to the books by Cressie and Wikle [2011], Banerjee et al. [2014] and Blangiardo and Cameletti [2015]. All the above approaches are time domain approaches, and we refer to their papers and papers there in for more details.

As pointed out earlier, the computation of the above predictor depends on the knowledge of $\underline{G}_0(\omega)$and $F_m(\omega)$which in turn depends on the parameters of the covariance function $g_{\|\mathbf{h}\|}(\omega)$. The parameters usually are unknown and has to be estimated efficiently. In the following section we will consider the estimation of the parameters. The estimation is based on Frequency Variogram (FV) approach recently proposed by Subba Rao et al. [2014]. In their paper, Subba Rao et al. [2014] considered the estimation of the parameters of the covariance function, their asymptotic properties and their efficiency compared to Gaussian likelihood approach. To avoid repetition, we refer to Subba Rao et al. [2014] for full details

4.1 Estimation of the Parameters of the Covariance function $g_{\|\mathbf{h}\|}(\omega, \vartheta)$ by Frequency Variogram (FV) Method.

Here we discuss briefly the FV methodology, and for details, we refer to Subba Rao et al. [2014]. We may point out here that the FV method depends on taking the differences of DFTs (for a given frequency) evaluated at two distinct locations spaced apart by $||h||$ units and then considering the joint distribution of these differences calculated for all frequencies. It is similar to the approaches proposed by Bevilacqua et al. [2012], Hall and Keilegom [2003] and Bliznyuk et al. [2012] which are based on the differences of the processes. The method here is based on the differences of DFTs.

Let $g_{\|\mathbf{h}\|}(\omega) = Cov(J_{\mathbf{s}}(\omega), J_{\mathbf{s}+\mathbf{h}}(\omega))$ be the covariance function and, for example, let $g_{\|\mathbf{h}\|}(\omega)$ be of the form given by (17). Assume the function $g_{\|\mathbf{h}\|}(\omega)$ is characterized by the parameter vector ϑ. For convenience, we denote this covariance function by $g_{\|\mathbf{h}\|}(\omega, \vartheta)$. Our object here is to estimate $\underline{\vartheta}$. We note that ω is the temporal spectral frequency, and $\|\mathbf{h}\|$ is the spatial Euclidean distance. The estimation of the parameters of the covariance function have also been considered by other authors (see for example, Cressie and Huang [1999], Gneiting [2002], Ma [2002], Ma [2003], Stein et al. [2004], Stein [2005b]), using either variogram method or likelihood method. All these methods can be described as Time Domain based approaches.

We note that in the case of purely spatial processes, the parameters are estimated either by minimizing the differences between the estimated variograms and the theoretical variograms evaluated for spatial distances $\|\mathbf{h}\|$ (weighted least squares approach) or by maximizing the Gaussian likelihood function. Because of the inclusion of temporal dimension, and if one uses time domain approach, the observations vector to use will be of order $mn \times 1$, and the variance covariance matrix of the observation vector will be of dimension $mn \times mn$. The number of computational operations required for inversion of such large dimensional matrices can be formidable. For example, it is well known that the calculation of Gaussian likelihood from such vectors requires $(nm)^3$ operations. In view of this, Stein et al. [2004], Stein [2005a], Stein [2005b] suggested using the restricted likelihood approach, an extension of the method proposed by Vecchia [1988] to reduce the number of computations. In FV approach proposed here, one does not need inversion of such high dimensional matrices as the likelihood function calculated is based on complex Gaussianity of Discrete Fourier Transforms evaluated at several distinct frequencies. It is well known that at these distinct frequencies, the Discrete Fourier Transforms of a stationary process are asymptotically uncorrelated and have complex Gaussian distribution (see Brillinger [2001], Theorem 4. 4. 1,see also Giraitis et al. [2012]). Under these assumptions on the frequencies, the variance covariance matrix of the vector of DFT's evaluated at distinct frequencies will be diagonal. Further, the Discrete Fourier Transforms can be calculated using the Fast Fourier transform algorithms. It has been shown in Subba Rao et al. [2014] that the FV estimates are robust against departure from Gaussianity and are as efficient as Gaussian estimates, if the process happens to be Gaussian and require less computational time. Briefly the details are as follows. We define a new spatio-temporal random process based on differences of the observed process $\{Y_t(\mathbf{s})\}$. Calculate the differences

$$X_{ij}(t) = Y_t(\mathbf{s}_i) - Y_t(\mathbf{s}_j), \quad \text{for each } t = 1, 2, \ldots, n$$

and for all locations $\mathbf{s}_i$, $\mathbf{s}_j$ where $\mathbf{s}_i$ and $\mathbf{s}_j$, $(i \neq j)$ are the pairs that belong to the set $N(\mathbf{h}_l) = \{\mathbf{s}_i, \mathbf{s}_j | \; \|\mathbf{s}_i - \mathbf{s}_j\| = \|\mathbf{h}_l\|, \; l = 1, 2, \ldots, L\}$. Define the Finite Fourier transform of the new time series $\{X_{ij}(t) | \; i \neq j\}$ at the frequencies $\omega_k(n) = 2\pi s_k(n)/n, \; k = 1, \ldots, K$, where $s_k(n), k = 1, 2, \ldots K$ are set of integers and $\omega_j(n) \pm \omega_k(n) \neq 0 (\bmod 2\pi)$ for $1 \leq j < k \leq K$ (see Brillinger [2001], Giraitis et al. [2012]). Let $J_{X_{ij}}(\omega_k(n)) = \frac{1}{\sqrt{(2\pi n)}} \sum_{t=1}^{n} X_{ij}(t) e^{-it\omega_k(n)} = J_{\mathbf{s}_i}(\omega_k(n)) - J_{\mathbf{s}_j}(\omega_k(n))$. Let $I_{X_{ij}}(\omega_k(n))$ be the second order periodogram of the time series $\{X_{ij}(t)\}$ given by

$$I_{X_{ij}}(\omega_k(n)) = \left|J_{X_{ij}}(\omega_k(n))\right|^2 . \quad (35)$$

Let $G_{X_{ij}}(\omega_k(n), \vartheta) = E(I_{X_{ij}}(\omega_k(n))$. The function $G_{X_{ij}}(\omega_k(n), \vartheta)$ is defined as the Frequency Variogram by Subba Rao et al. [2014]. It is very similar to the classical definition of spatio-temporal variogram $2\gamma(\mathbf{h}, u)$ defined in section 2 of the present paper (set $u = 0$) in (1). The usefulness of FV as a measure of dissimilarity between two spatial processes and its further properties were discussed in a recent paper by the authors (Subba Rao and Terdik [2016])

From (35), we obtain

$$\begin{aligned} E\left[I_{X_{ij}}(\omega_k(n))\right] &= G_{X_{ij}}(\omega_k(n), \vartheta) \\ &= E\left[I_{\mathbf{s}_i}(\omega_k(n))\right] + E\left[I_{\mathbf{s}_j}(\omega_k(n))\right] - 2\ Real\ E\left[I_{\mathbf{s}_i\mathbf{s}_j}(\omega_k(n))\right], \end{aligned} \quad (36)$$

where $I_{\mathbf{s}_i\mathbf{s}_j}(\omega_k(n))$ is the cross periodogram between the processes $\{Y_t(\mathbf{s}_i)\}$ and $\{Y_t(\mathbf{s}_j)\}$ and $I_{\mathbf{s}_i}(\omega_k(n))$ is the periodogram of the series $Y_t(s_i)$. For large n, it can be shown that for a stationary process $E\left[I_{\mathbf{s}_i}(\omega_k(n))\right] = E\left[I_{\mathbf{s}_j}(\omega_k(n))\right] = g_0(\omega_k(n); \underline{\vartheta})$ and for a stationary and an isotropic process $E\left[I_{\mathbf{s}_i\mathbf{s}_j}(\omega_k(n))\right] = g_{\|s_i - s_j\|}(\omega_k(n); \underline{\vartheta})$ which is real. Therefore, the expectation of (36) is given by

$$G_{(s_i,s_j)}(\omega_k(n); \underline{\vartheta}) = 2\left[g_0(\omega_k(n); \underline{\vartheta}) - g_{\|s_i - s_j\|}(\omega_k(n); \underline{\vartheta})\right], \quad (37)$$

It is interesting to compare $G_{(s_i,s_j)}(\omega_k(n); \underline{\vartheta})$ with spatio-temporal variogram $2\gamma(\mathbf{h}, u)$ given by equation (2). The similarity between these two functions shows that one can use the Frequency variogram which is a frequency domain version of spatio-temporal variogram for estimating the effective range $||\mathbf{h}||$, and also the parameters etc.

Now for the estimation of the parameter vector $\underline{\vartheta}$ we proceed as in Subba Rao et al. [2014]. Consider the K-dimensional complex valued random vector,

$$\underline{\chi}_{\|s_i - s_j\|}(\omega) = \left[J_{X_{ij}}(\omega_1(n)), J_{X_{ij}}(\omega_2(n)), \ldots, J_{X_{ij}}(\omega_K(n))\right],$$

which is distributed asymptotically as complex normal with mean zero and with variance covariance matrix with diagonal elements
$[g_{\|\mathbf{h}\|}(\omega_1(n), \vartheta),\ g_{\|\mathbf{h}\|}(\omega_2(n), \vartheta),\ \ldots,\ g_{\|\mathbf{h}\|}(\omega_K(n), \vartheta)]$, where $||h|| = ||s_i - s_j||$. We note that because of asymptotic independence of Fourier transforms at distinct frequencies chosen above, the off diagonal elements of the variance covariance matrix of the complex Gaussian random vector $\underline{\chi}_{\|\mathbf{h}\|}(\omega)$ are zero. Therefore, the minus of log likelihood function can be shown to be proportional to

$$Q_{n,N(\mathbf{h})}(\underline{\vartheta}) = \frac{1}{|N(\mathbf{h})|} \sum_{(\mathbf{s}_i,\mathbf{s}_j)\in N(\mathbf{h})} \sum_{k=1}^{K} \left[\ln G_{(\mathbf{s}_i,\mathbf{s}_j)}(\omega_k(n); \underline{\vartheta}) + \frac{I_{Xij}(\omega_k(n))}{G_{(\mathbf{s}_i,\mathbf{s}_j)}(\omega_k(n); \underline{\vartheta})}\right]. \quad (38)$$

Here $|N(\mathbf{h})|$ is the total number of all distinct pairs $\mathbf{s}_i$ and $\mathbf{s}_j$ such that $N(\mathbf{h}) = \{(\mathbf{s}_i, \mathbf{s}_j) \mid \ \|\mathbf{s}_i - \mathbf{s}_j\| = ||\mathbf{h}||\}$. The above criterion (38) is defined only for one distance $\|\mathbf{h}\|$. Suppose we now define L spatial distances from the observed data. We can now define an over all criterion for the minimization

$$Q_n(\vartheta) = \frac{1}{L}\sum_{l=1}^{L} Q_{n,N(\mathbf{h}_l)}(\underline{\vartheta}), \quad (39)$$

We minimize (39) with respect to $\underline{\vartheta}$ (for details refer to Subba Rao et al. [2014]). The asymptotic normality of the estimator $\underline{\vartheta}$ obtained by minimizing (39) has been proved in Theorem 2 of the paper of Subba Rao et al. [2014]. To avoid repetition, we refer to their paper for details. We state the asymptotic distribution of the estimates. It has been shown in Subba Rao et al. [2014] that under certain conditions, and for large n,

$$\sqrt{n}\left(\underline{\vartheta}_n - \vartheta_0\right) \xrightarrow{D} N\left(\underline{0}, (\nabla^2 Q_n(\underline{\vartheta}_0))^{-1}\ V\ (\nabla^2 Q_n(\underline{\vartheta}_0))\right),$$

where $V = \lim_{n\to\infty} Var\left[\frac{1}{\sqrt{n}}\nabla Q_n(\vartheta_0)\right]$, $\nabla Q_n(\vartheta_0)$ is a vector of first order partial derivatives, $\nabla^2 Q_n(\vartheta_0)$ is a matrix of second order partial derivatives.

We may point out here that Whittle approximation to the Gaussian likelihood of random processes defined on d-dimensional lattices is well known (for example refer to Dahlhaus and Künsch [1987], Guyon [1982], Guinness and Fuentes [2015]) . Here we have considered the spatio-temporal processes defined on irregular space and observed at equally spaced time points and the likelihood is based on the DFT's of the differenced series The method of estimation proposed here is valid even under the weaker assumption of intrinsic stationarity (for details to Subba Rao and Terdik [2016]).

4.2 Spatio-Temporal Prediction and Estimation of the Parameters: Measurement Errors Case.

In section 4, we considered the prediction of $\{Y_t(\mathbf{s}_0)\}$ under the assumption that the data we observe $\{Y_t(\mathbf{s})\}$ has no measurement errors. This assumption may be unrealistic in some situations. In this section we show how the methods given earlier can be modified when the observations are corrupted by White noise. Instead of observing the true observations $\{Y_t(\mathbf{s}_i) \mid i = 1, 2, \ldots, m; t = 1, 2, \ldots, n\}$ we assume that we observe the corrupted observations $\left\{\widetilde{Y}_t(\mathbf{s}_i) \mid i = 1, 2, \ldots, m; t = 1, 2, \ldots, n\right\}$ where for each t and $\mathbf{s}_i$, $(i = 1, 2, \ldots, m)$ we have

$$\widetilde{Y}_t(\mathbf{s}_i) = Y_t(\mathbf{s}_i) + \eta_t(\mathbf{s}_i); i = 1, 2, \ldots, m; t = 1, 2, \ldots, n. \tag{40}$$

In the terminology used by Cressie and Wikle [2011], (chapter 6), the above model can be defined as the Data model, which is written in terms of the process model for$\{Y_t(\mathbf{s}_i)\}$,the uncorrupted process. In order to draw inference on the process $\{Y_t(\mathbf{s}_0)\}$,the process is further decomposed into various components and the likelihood function is built using various components of the decomposition. In our approach, we do not make any assumptions on the process and base our inference on DFT's of the process, which are asymptotically complex Gaussian.

We assume $\{Y_t(\mathbf{s})\}$ is a zero mean second order spatially, temporally stationary process and is independent of the noise process $\eta_t(\mathbf{s})$. We assume further that the noise process $\eta_t(\mathbf{s})$ is spatially and temporally a white noise process with

$$E\eta_t(\mathbf{s}) = 0,$$

$$Cov\left(\eta_t(\mathbf{s}_i), \eta_{t'}(\mathbf{s}_j)\right) = \begin{cases} 0 & \text{if} \quad t \neq t' \\ 0 & \text{if} \quad i \neq j \\ \sigma_\eta^2 & \text{if} \quad t = t', i = j \end{cases}$$

$$Cov\left(Y_t(\mathbf{s}_i), \eta_{t'}(\mathbf{s}_j)\right) = 0, \text{ for all } t, t', i \text{ and } j.$$

Let $\widetilde{J}_{\mathbf{s}}(\omega)$, $J_{\mathbf{s}}(\omega)$, $J_{\eta,\mathbf{s}}(\omega)$ be the Discrete Fourier Transforms calculated from the time series data $\left\{\widetilde{Y}_t(\mathbf{s}_i)\right\}$, $\{Y_t(\mathbf{s}_i)\}$, and $\{\eta_t(\mathbf{s}_i)\}$ respectively. Then from the relation (40), we have

$$\widetilde{J}_{\mathbf{s}}(\omega) = J_{\mathbf{s}}(\omega) + J_{\eta,\mathbf{s}}(\omega). \tag{41}$$

Let

$$\widetilde{c}(\mathbf{h}, u) = Cov\left[\widetilde{Y}_t(\mathbf{s}), \widetilde{Y}_{t+u}(\mathbf{s}+\mathbf{h})\right] = \widetilde{c}(\|\mathbf{h}\|, u),$$
$$c(\mathbf{h}, u) = Cov\left[Y_t(\mathbf{s}), Y_{t+u}(\mathbf{s}+\mathbf{h})\right] = c(\|\mathbf{h}\|, u),$$
$$c_\eta(\mathbf{h}, u) = Cov\left[\eta_t(\mathbf{s}), \eta_{t+u}(\mathbf{s}+\mathbf{h})\right] = \begin{cases} 0 & \text{if} \quad \mathbf{h} \neq 0, u \neq 0 \\ \sigma_\eta^2 & \text{if} \quad \mathbf{h} = 0, u = 0 \end{cases}$$

We note that under the above assumptions, we have

$$Cov(\widetilde{J}_{\mathbf{s}}(\omega), \widetilde{J}_{\mathbf{s}+\mathbf{h}}(\omega)) = Cov\left(J_{\mathbf{s}}(\omega), J_{\mathbf{s}+\mathbf{h}}(\omega)\right) + Cov\left(J_{\eta,\mathbf{s}}(\omega), J_{\eta,\mathbf{s}+\mathbf{h}}(\omega)\right), \tag{42}$$

and under isotropy condition,we have

$$\widetilde{g}_{\|\mathbf{h}\|}(\omega) = Cov(\widetilde{J}_{\mathbf{s}}(\omega), \widetilde{J}_{\mathbf{s}+\mathbf{h}}(\omega))$$
$$= \begin{cases} g_{\|\mathbf{h}\|}(\omega) & \text{if} \quad \mathbf{h} \neq 0, \\ g_{\|\mathbf{h}\|}(\omega) + \frac{\sigma_\eta^2}{2\pi} & \text{if} \quad \mathbf{h} = 0, \end{cases}$$

where

$$g_{\|\mathbf{h}\|}(\omega) = Cov(J_{\mathbf{s}}(\omega), J_{\mathbf{s}+\mathbf{h}}(\omega))$$
$$Cov\left(J_{\eta,\mathbf{s}}(\omega), J_{\eta,\mathbf{s}+\mathbf{h}}(\omega)\right) = \frac{\sigma_\eta^2}{2\pi}, \text{if } \mathbf{h} = 0,$$
$$= 0 \text{ if } \mathbf{h} \neq \mathbf{0}.$$

Further, we note that $\widetilde{c}(\|\mathbf{h}\|, u) = c(\|\mathbf{h}\|, u)$, if $\mathbf{h} \neq 0$. We note that $g_{\|\mathbf{0}\|}(\omega)$ is the second order temporal spectrum of the uncorrupted process $\{Y_t(\mathbf{s})\}$, for all $\mathbf{s}$.

Our object here is to predict the uncorrupted time series $\{Y_t(\mathbf{s}_0)\}$ given the corrupted spatio-temporal data $\left\{\widetilde{Y}_t(\mathbf{s}_i) \mid i = 1, 2, \ldots, m; t = 1, 2, \ldots, n\right\}$.

As before we predict the Discrete Fourier Transform $J_{\mathbf{s}_0}(\omega)$ of the true data $\{Y_t(\mathbf{s}_0)\}$ for all ω, given the Discrete Fourier Transforms of the corrupted data and then invert the Discrete Fourier transform $J_{\mathbf{s}_0}(\omega)$ using (34) to obtain the predicted values of the data $\{Y_t(\mathbf{s}_0)\}$ at $t = 1, 2, \ldots, n$. We use the predicted data thus obtained to predict the future values of $Y_{n+m}(\mathbf{s}_0)$, for all $m > 0$.

Consider the $(m+1)$ dimensional complex valued random vector

$$\underline{\widetilde{J}}'_{m+1}(\omega) = \left[J_{\mathbf{s}_o}(\omega), \underline{\widetilde{J}}'_m(\omega)\right].$$

where

$$\underline{\widetilde{J}}'_m(\omega) = \left[\widetilde{J}_{\mathbf{s}_1}(\omega), \widetilde{J}_{\mathbf{s}_2}(\omega), \ldots, \widetilde{J}_{\mathbf{s}_m}(\omega)\right].$$

which has a complex Gaussian distribution with mean zero and and variance -covariance matrix

$$\begin{aligned} E\left[\underline{\widetilde{J}}_{m+1}(\omega)\underline{\widetilde{J}}^*_{m+1}(\omega)\right] &= \begin{bmatrix} E(J_{s_o}(\omega)\widetilde{J}^*_{s_o}(\omega)) & E\left(J_0(\omega)\underline{\widetilde{J}}^{*\prime}_m(\omega)\right) \\ E\left(\underline{\widetilde{J}}_m(\omega)J^*_0(\omega)\right) & E\left(\underline{\widetilde{J}}_m(\omega)\underline{\widetilde{J}}^*_m(\omega)\right) \end{bmatrix} \\ &= \begin{bmatrix} g_0(\omega) & \underline{G}'_0(\omega) \\ \underline{G}_0(\omega) & F_m(\omega) \end{bmatrix}, \end{aligned}$$

where

$$\begin{aligned} E\left|J_{\mathbf{s}_0}(\omega)\right|^2 &= g_0(\omega), \\ \underline{\widetilde{G}}'_0(\omega) &= E\left(J_{s_o}(\omega)\underline{\widetilde{J}}^{*\prime}_m(\omega)\right) \\ &= [g_{\|s_0-s_1\|}(\omega), g_{\|s_0-s_2\|}(\omega), \ldots, g_{\|s_0-s_n\|}(\omega)], \\ E\left(\underline{\widetilde{J}}_m(\omega)\underline{\widetilde{J}}^*_m(\omega)\right) &= \widetilde{F}_m(\omega) = \left[g_{\|s_i-s_j\|}(\omega)\right] + \frac{\sigma^2_\eta}{2\pi}I \\ &= F_m(\omega) + \frac{\sigma^2_\eta}{2\pi}I \end{aligned}$$

and the matrix $F_m(\omega)$ is defined earlier in section 4. We note that the (i,j) th element of the matrix $F_m(\omega)$ is $g_{\|s_i-s_j\|}(\omega)$, which is the covariance between the DFT's of the uncorrupted processes $\{Y_t(\mathbf{s}_i)$ $\}$and $\{Y_t(\mathbf{s}_j)\}$ evaluated under the spatial isotropy condition. In view of the complex Gaussianity of the DFT's, the optimal predictor of the $J_{\mathbf{s}_0}(\omega)$ is given by the conditional expectation

$$E\left[J_{\mathbf{s}_o}(\omega) \mid \underline{\widetilde{J}}_m(\omega)\right] = \underline{\widetilde{G}}'_0(\omega)\widetilde{F}^{-1}_m(\omega)\underline{\widetilde{J}}_m(\omega), \tag{43}$$

and the minimum mean square prediction error is given by

$$\widetilde{\sigma}^2_m(\omega) = g_0(\omega) - \underline{\widetilde{G}}'_0(\omega)\widetilde{F}^{-1}_m(\omega)\underline{\widetilde{G}}_0(\omega). \tag{44}$$

We can now obtain the predictors of $Y_t(\mathbf{s}_0)$ from the DFT using the relation (34). The methodology is similar to the methodology described in Section 4, hence details are omitted.

4.3 Estimation of the parameters.

Here we briefly describe the estimation procedure in the measurement errors case. The computation of the predictors given above depends on the knowledge of $g_0(\omega)$, $\widetilde{F}_m(\omega)$, $\underline{\widetilde{G}}_0(\omega)$ and σ^2_η. and these in turn depend on the parameters ϑ and also on the variance of the measurement errors σ^2_η. Let $\Psi = \left(\vartheta, \sigma^2_\eta\right)$.

Consider the differences, for $s_i \neq s_j$, for each t,

$$\widetilde{X}_{ij}(t) = \widetilde{Y}_t(\mathbf{s}_i) - \widetilde{Y}_t(\mathbf{s}_j) = X_{ij}(t) + e_{ij}(t), \tag{45}$$

where

$$\begin{aligned} X_{ij}(t) &= Y_t(\mathbf{s}_i) - Y_t(\mathbf{s}_j), \\ e_{ij}(t) &= \eta_t(\mathbf{s}_i) - \eta_t(\mathbf{s}_j) \end{aligned}$$

Let $\widetilde{J}_{X_{ij}}(\omega_k)$, $J_{X_{ij}}(\omega_k)$ and $J_{e_{ij}}(\omega_k)$ be the Discrete Fourier Transforms of the time series data $\left\{\widetilde{X}_{ij}(t)\right\}$, $\{X_{ij}(t)\}$, and $\{e_{ij}(t)\}$ respectively. From (45), we have

$$\widetilde{J}_{X_{ij}}(\omega_k) = J_{X_{ij}}(\omega_k) + J_{e_{ij}}(\omega_k),$$

where

$$J_{X_{ij}}(\omega) = J_{\mathbf{s}_i}(\omega) - J_{\mathbf{s}_j}(\omega),$$
$$J_{e_{ij}}(\omega_k) = J_{\eta_i}(\omega) - J_{\eta_j}(\omega).$$

We note

$$E(\widetilde{J}_{X_{ij}}(\omega_k)) = 0,$$
$$E\left|\widetilde{J}_{X_{ij}}(\omega_k)\right|^2 = E\left|J_{X_{ij}}(\omega_k)\right|^2 + E\left|J_{e_{ij}}(\omega_k)\right|^2,$$

where

$$E\left|J_{X_{ij}}(\omega_k)\right|^2 = 2\left[g_0(\omega,\vartheta) - g_{\|s_i-s_j\|}(\omega,\vartheta)\right]$$
$$E\left|J_{e_{ij}}(\omega_k)\right|^2 = \frac{\sigma_\eta^2}{\pi}.$$

Let

$$\widetilde{G}_{\|s_i-s_j\|}(\omega,\Psi) = 2\left[g_0(\omega,\vartheta) - g_{\|s_i-s_j\|}(\omega,\vartheta)\right] + \frac{\sigma_\eta^2}{\pi}.$$

Consider the frequencies $\omega_k(n), k = 1,2,\ldots K$ defined earlier and consider the K dimensional complex valued random vector

$$\underline{\widetilde{X}}`_{\|s_i-s_j\|}(\omega) = \left[\widetilde{J}_{X_{ij}}(\omega_1(n)), \widetilde{J}_{X_{ij}}(\omega_2(n)), \ldots, \widetilde{J}_{X_{ij}}(\omega_K(n))\right]$$

which is distributed asymptotically as complex normal with mean zero and variance covariance matrix with diagonal elements

$$\left[\widetilde{G}_{\|s_i-s_j\|}(\omega_1(n),\Psi), \widetilde{G}_{\|s_i-s_j\|}(\omega_2(n),\Psi), \ldots, \widetilde{G}_{\|s_i-s_j\|}(\omega_K(n),\Psi)\right],$$

The off diagonal elements of the matrix are zero. By proceeding as in Section 5, we arrive at the minimisation criterion

$$\widetilde{Q}_n(\vartheta) = \frac{1}{L}\sum_{l=1}^{L} \widetilde{Q}_{n,N(\mathbf{h}_l)}(\Psi),$$

where

$$\widetilde{Q}_{n,N(\mathbf{h})}(\Psi) = \frac{1}{|N(\mathbf{h})|}\sum_{(\mathbf{s}_i,\mathbf{s}_j)\in N(\mathbf{h})}\sum_{k=1}^{K}\left[\ln \widetilde{G}_{(||\mathbf{s}_{i-},\mathbf{s}_j||)}(\omega_k(n);\Psi) + \frac{\left|\widetilde{J}_{X_{ij}}(\omega_k(n))\right|^2}{\widetilde{G}_{(||\mathbf{s}_i-\mathbf{s}_j||)}(\omega_k(n);\Psi)}\right]'$$

The minimisation is done with respect to ϑ and σ_η^2.. Proceeding as in Subba Rao et al. [2014], we can show that for large n,

$$\sqrt{n}\left(\widehat{\Psi}_n - \Psi_0\right) \xrightarrow{D} N\left(\underline{0}, (\nabla^2\widetilde{Q}_n(\underline{\vartheta}_0))^{-1}\ \widetilde{V}\ (\nabla^2\widetilde{Q}_n(\Psi_0))\right),$$

where $\widetilde{V} = \lim_{n\to\infty} Var\left[\frac{1}{\sqrt{n}}\nabla\widetilde{Q}_n(\vartheta_0)\right]$, $\nabla^2\widetilde{Q}_n(\vartheta_0)$ is the matrix of second order partial derivatives.

5 Simulation Study

In the following we briefly describe a method for generating a stationary spatio-temporal random process $\{Y_t(\mathbf{s}_i);\ i = 1,2,\ldots,m; t = 1,2,\ldots,n\}$ at the locations $\{s_i\ ; i = 1,2,3,\ldots m\}$ with a given second order spatio-temporal spectral density function $f_Y(\underline{\lambda},\omega)$. Equivalently, we generate the DFT's $\{J_{\mathbf{s}_i}(\omega)\}$ for

$i = 1, 2, \ldots m$; and for all $|\omega| \leq \pi$ with the given covariance function $g_{\|\mathbf{h}\|}(\omega,\ \vartheta)$ and by inverting the DFT's we obtain the spatio-temporal time series with the specified spectrum.. For illustration purposes we assume $d = 2, \upsilon = 1$ and the second order spectrum is given by (12), and $g_{\|\mathbf{h}\|}(\omega,\ \vartheta)$, $g_0(\omega,\ \vartheta)$ are respectively given by (20) and (21). Let $c_{i,j}(k) = g_{||\mathbf{s}_i - \mathbf{s}_j||}(\omega_k,\ \vartheta)$.

Several methods are now available for simulating stationary spatial data (see the books by Schabenberger and Gotway [2005]; Cressie [1993]), and as far as we are aware not many methods are available for simulating stationary spatio-temporal data(Subba Rao et al ,2004). We describe below an algorithm for simulating such data.using the Discrete Fourier Transforms.

We first simulate the Discrete Fourier Transforms $\{J_{\mathbf{s}}(\omega_k),\ k = 0, 1, \ldots, n-1\}$, and then by inversion we obtain the data. We briefly outline the steps.

1. The locations $\mathbf{s}_i;\ i = 1, 2, \ldots, m$ are chosen randomly from the unit square.

2. Let n be even. Let $g_0(\omega,\ \vartheta) = \left(\sigma^2 / (2\pi)^2\right) \left|\vartheta_q\left(e^{-2i\pi\omega}\right) / \varphi_p\left(e^{-2i\pi\omega}\right)\right|^2$, be the second order temporal spectrum of an ARMA(p,q) model with innovation variance $\sigma^2/2\pi$. We set $\sigma^2 = \sigma_e^2$ for our simulation purposes. Define the matrices $\mathbb{C}_k = (c_{i,j}(k))$ where $c_{i,j}(k) = g_{||\mathbf{s}_i - \mathbf{s}_j||}(\omega_k,\ \vartheta)$, $k = 0, 1, \ldots, \frac{n}{2}$. We note $g_{||\mathbf{s}_i - \mathbf{s}_j||}(\omega_k,\ \vartheta)$ is given by (21) where $|c(\omega)| = \sigma / \left(2\pi\sqrt{2g_0(\omega,\ \vartheta)}\right)$.

3. Generate a series of independent complex Gaussian, zero mean random vectors $\mathbf{U}_k$ each of order $m \times 1$, $k = 0, 1, \ldots, n/2$, such that $Var(\mathbf{U}_k) = I$, an identity matrix. We note that $\mathbf{U}_0$ and $\mathbf{U}_{n/2}$ are real valued random vectors. Define the m dimensional complex valued column vector $\mathbf{J}(\omega_k) = [J_{\mathbf{s}_i}(\omega_k)]_{i=1}^m = \sqrt{\mathbb{C}_k}\mathbf{U}_k$ We note $E(\mathbf{J}(\omega_k)) = 0, E(\mathbf{J}(\omega_k)\mathbf{J}^*(\omega_k)) = \mathbb{C}_k$. The DFT's generated thus will have the given covariance function $g_{||\mathbf{s}_i - \mathbf{s}_j||}(\omega_k,\ \vartheta)$.

We now consider a specific example.

4. We assume the spatio-temporal process $Y_t(\mathbf{s})$, for each $\mathbf{s}$, satisfies an ARMA(2,1) model of the form $\varphi_2(B)Y_t(\mathbf{s}) = \vartheta_1(B)e_t(\mathbf{s})$, where $\sigma^2 = 4$.
$$\varphi_2(B) = 1 + 4/17B + 4/17B^2,$$
$$\vartheta_1(B) = 1 - 2/3B.$$

5. The inverse Fourier transforms of $\mathbf{J}(\omega_k)$ gives us the spatio-temporal realizations $\{Y_t(s_i)|i = 1 \ldots m\}$ which has the required spatio-temporal covariance function.

5.1 Estimation of the parameters and the Prediction of $\widehat{Y}_t(\mathbf{s}_0)$

We briefly describe the steps required to estimate the parameters ϑ and also the steps required for prediction. We now assume that we have the spatio-temporal data $\{Y_t(\mathbf{s}_i);\ i = 1, 2, \ldots, m; t = 1, 2, \ldots, n\}$. Here we have chosen $m = 9.n = 2^{11}$.We considered the prediction of the data at the location $\mathbf{s}_0 = 10$.

.Let $X_{ij}(t) = Y_t(\mathbf{s}_i) - Y_t(\mathbf{s}_j)$ as defined in section 4. We compute the Discrete Fourier Transforms of the differenced series $X_{ij}(t)$ which is the difference of the DFT's of the individual series $\{Y_t(\mathbf{s}_i)\}$ and $\{Y_t(\mathbf{s}_j)\}$. The parameters of the covariance function are now estimated by minimizing the criterion (38).

1. The parameters of the polynomials $\varphi_2(z)$, $\vartheta_1(z)$ and variance σ^2 are estimated as described in section 4. The estimated value are found to be $\widehat{\sigma} = 2.0744$ and
$$\widehat{\varphi}(z) = 1 + 0.2404z + 0.2356z^2,$$
$$\widehat{\vartheta}(z) = 1 - 0.6406z.$$

2. We note from the above that the estimates of the coefficients of the ARMA (2,1) are very close to the true values given above. The parameters are estimated using the spatio-temporal data generated as described above. The matrices $\{\mathbb{C}_k\}$ and $\underline{G}_0(\omega_k)$ given in the formula above are calculated using the estimated parameters.

3. The predicted values of $\{\widehat{Y}_t(\mathbf{s}_0)\}$ are obtained using the relation (34) and the plots of the predicted data and the true data are not given due to space considerations.

6 Real Data Analysis

For our illustration, we consider the Air Pollution data analyzed by Sahu and Mardia [2005], and we refer to their paper for full details. The data analyzed corresponds to atmospheric particulate matter that is less than 2. 5 ωm in size (usually known as $PM_{2.5}$) which is one of six primary air pollutants and is a mixture of fine particles and gaseous compounds such as sulphur dioxide (SO_2) and nitrogen oxides. The data was observed at 15 monitoring stations in New York city during the first 9 months of the year 2002. The data was observed once in every 3 days, thus giving 91 equally spaced time series for each monitoring station. The total number of observations are $1365 = 15 \times 91$. The data can be obtained from the website http://www. blackwellpublishing. com/rss. We use the data given at the 15 locations along with their spatial coordinates. The spatial coordinates of 625 nearby locations are also known and can be found at the website..

Out of 1365 data points, 126 were missing and the majority of values which are missing are at the location 11 (with coordinates: latitude −73. 84, longitude 40. 77). We note that our predictor depends on the DFT's of the data and these can be computed even when the data is missing and we note that the time series data with missing values can be considered as time series observed at unequally spaced time points (see Scargle [1989]). Let $X(\mathbf{s}_i, t)$ $(i = 1,\ 2,\ \ldots 15; t = 1,\ 2,\ \ldots 91)$ denote the $PM_{2.5}$ observation at the location $\mathbf{s}_i$ and at time t. We have chosen to estimate all 91 observations at this location 11 using the data from the other 14 locations.

We used the first differences $\{Y_t(\mathbf{s}) = \Delta_t X(\mathbf{s}, t) : \mathbf{s} \in \mathbb{R}^2, \quad t \in \mathbb{Z}\}$ to remove the linear trend as suggested by Sahu and Mardia [2005], and used the detrended data for our analysis. We use the detrended data only for the entire analysis including estimation and prediction. The Preliminary time series analysis carried out on all the14 time series suggests that AR(1) model may be adequate to explain the temporal dependence. The second order spectrum for a AR(1) process is given by $g_0(\omega, \vartheta) = \sigma^2 / |1 + \varphi z|^2$ where $z = \exp(i\omega)$ and the vector of the parameters to be estimated using the entire spatio temporal detrended data are $\vartheta = (\sigma^2, \phi)$ and these are estimated by minimizing the criterion (39) with $L = 91$. We note that we scaled the equation (11) such that $\sigma_e^2 = \sigma^2$, see (21). We may point out here that one can do correlation analysis and model fitting when the data is missing and there is a huge literature on time series analysis with missing values, and we refer to Dunsmuir [1983], Dunsmuir and Robinson [1981] and also the proceedings edited by Parzen [1984].

The final estimates obtained by minimizing the criterion (38) for the AR(1) parameters are found to be $\widehat{\varphi} = 0.\ 4659$; $\widehat{\sigma} = 5.\ 9264$. Using these estimated values, all the elements of the vector $G_0(\omega)$ and the elements of the square matrix $F_m(\omega)$ (which is of order 14×14) are evaluated. The vectors $J_0(\omega)$ at the Fourier frequencies $\omega_k = 2\pi k / 2^6$ are estimated using the equation $\underline{\widehat{G}}_0^{\prime}(\omega_k)\, \widehat{F}_m^{-1}(\omega_k)\, \underline{J}_m(\omega_k)$. The data at the location 11 is predicted using the equation (34). The plot of the 91 predicted values (with (+) sign), plot of the first 23 given observations (with o sign), corresponding 95% confidence bands using (46) are given in Figure 2. We see a good agreement between the predicted values and the observed values, suggesting that the prediction method given here is useful. Also we find that strong spatial correlation and the temporal correlation present in the data can satisfactorily be explained by the spatio-temporal covariance function defined here..

In order to check the overall performance, we computed the leave-one-out cross-validation (Giraldo et al. [2010]) criterion. Here we estimated the data at one location, taken one at a time, using the data given at other 13 locations. The Mean Square Error calculated for all the 14 locations is

$$MSSE = \sum SSE(j) / (14 * 91) = 15.3050,$$

where

$$SSE(j) = \sum_{t=1}^{91} \left(Y_t(\mathbf{s}_j) - \widehat{Y}_t(\mathbf{s}_j | \mathbf{s}_k \neq \mathbf{s}_j) \right)^2,$$

and $\widehat{Y}_t(\mathbf{s}_j | \mathbf{s}_k \neq \mathbf{s}_j)$ is the estimator of the data at time t at location $\mathbf{s}_j$ conditional on the data at the locations $\{\mathbf{s}_1, \mathbf{s}_2, \ldots \mathbf{s}_{j-1}, \mathbf{s}_{j+1}, \ldots \mathbf{s}_{13}\}$. We have also estimated the prediction error variance using the equation (31) for each location.

We estimated the $PM_{2.5}$ values for all $t = 1, 2, \ldots 91$ and for all the 625 locations (including 14 locations where the data is available). The plot of the these predicted values at $t = 22$ are given in Figure 3 (the black dots correspond to the locations where we have observed data)..

In our analysis of the pollution data, we considered the prediction of entire data (all the 91 observations) at the location 11 using the other 13 locations. We repeated the procedure at other locations as

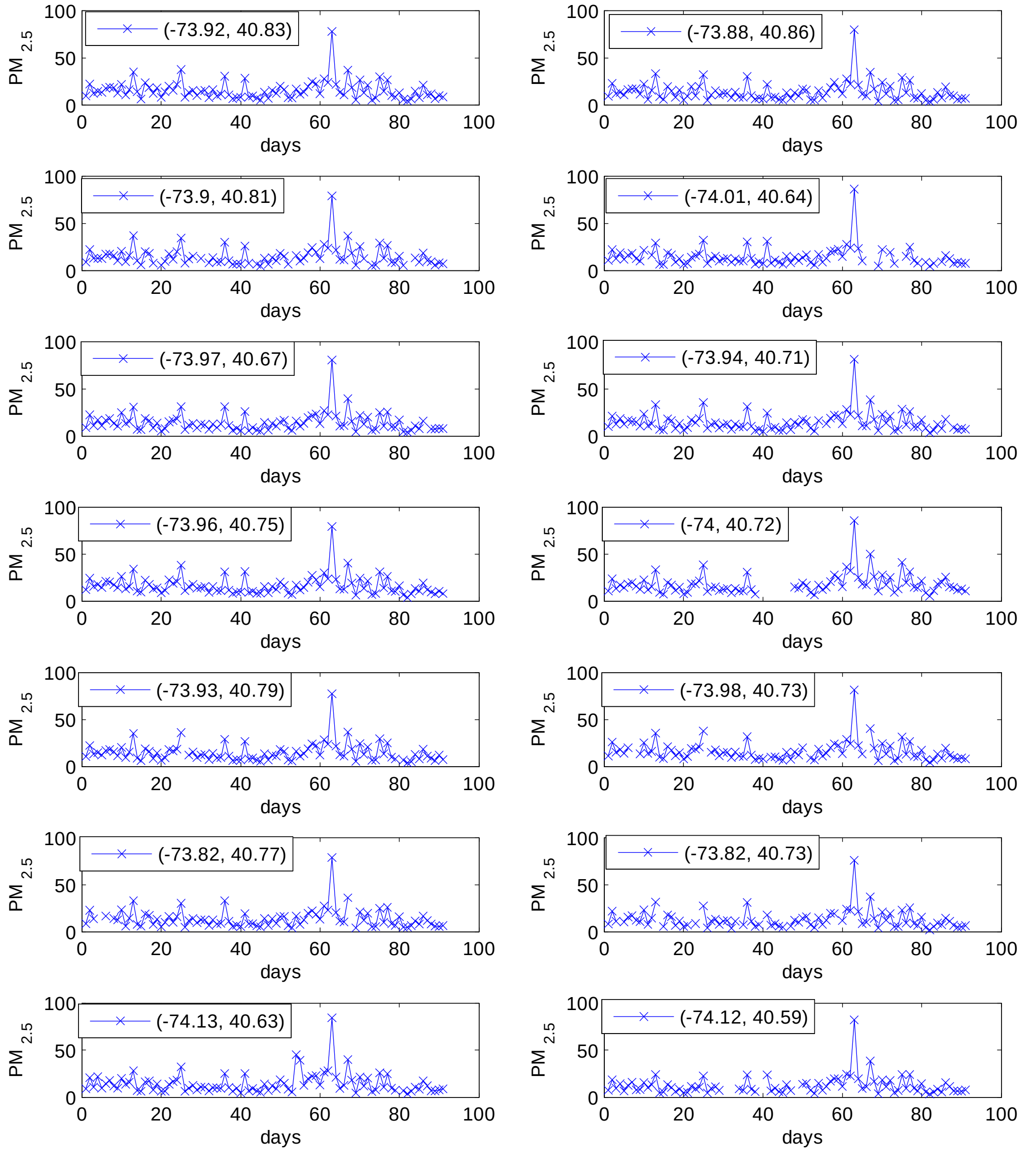

Figure 1: Time Series Data at the 14 locations with their spatial coordinates

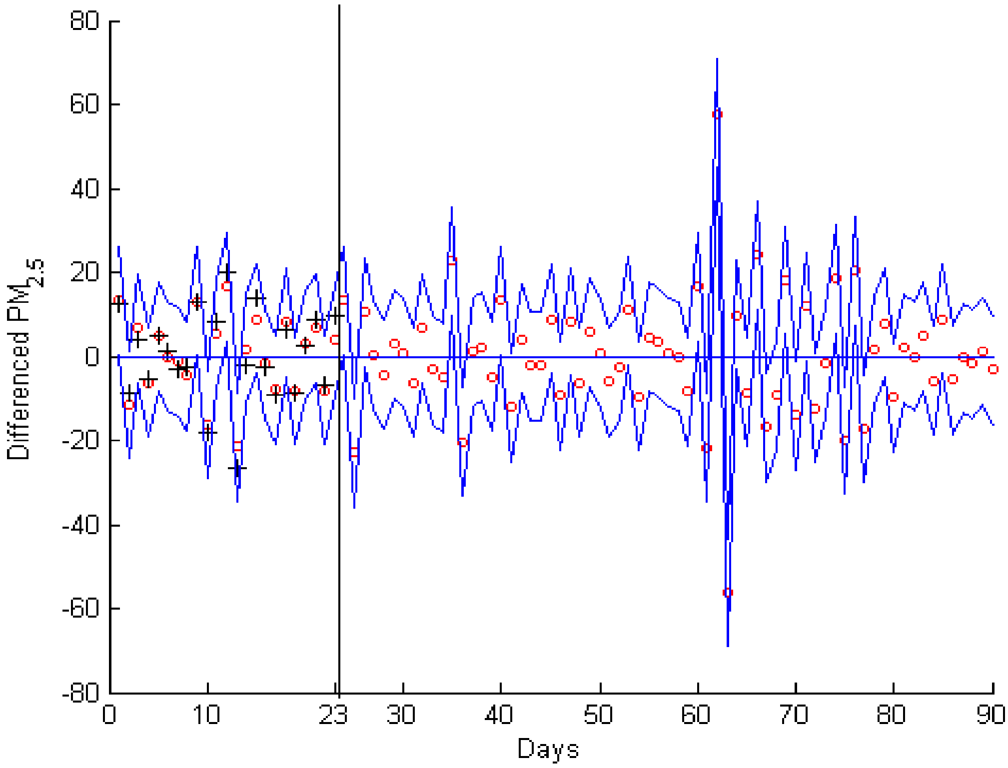

Figure 2: Location 11. Plot of the observed values (denoted by + sign), predicted values (denoted by 'o' sign), 95% prediction intervals (denoted by – sign)

well. It is also possible to predict the future values at a given location using the data we have at that location and other locations as well.

For our illustration purposes we consider the prediction at the location #8. We considered the data, $t = 1, 2, \ldots 83$ as given and predicted the rest of the values $(84, 83, \ldots 91)$ using the AR(1) model fitted. We note that AR(1) model fitted using the time series data at all 14 locations. The forecasting methodology is well known, and therefore, we briefly summarise the method (details can be found in any time series book). We note that we are illustrating only for AR(1) model, and the method can be used for any linear or nonlinear time series models (see for example the book by Brockwell and Davis [1987]). In practice the parameters are unknown, and the estimates are substituted for the true values.

Let

$$\widehat{Y}_{n+h}\left(\mathbf{s}_0 | t = 1, 2, \ldots, n; \mathbf{s}_j; j = 1, 2, \ldots m\right) = E\left(Y_{n+h}\left(\mathbf{s}_0\right) | Y_t\left(\mathbf{s}_j\right); t = 1, 2, \ldots n, ; j = 1, 2, \ldots m\right)$$

For the AR(1) model considered here, we have

$$\widehat{Y}_{n+h}\left(\mathbf{s}_0 | t = 1, 2, \ldots n, \mathbf{s}_j; j = 1, 2, \ldots m\right) = \varphi^h \widehat{Y}_n\left(\mathbf{s}_0\right),$$

hence

$$Y_{n+h}\left(\mathbf{s}_0\right) - \widehat{Y}_{n+h}\left(\mathbf{s}_0\right) = \varphi^h \left(Y_{n+h}\left(\mathbf{s}_0\right) - \widehat{Y}_n\left(\mathbf{s}_0\right)\right) + \sum_{k=1}^{h} \varphi^{k-1} \varepsilon_{n+k}\left(\mathbf{s}_0\right),$$

where $\widehat{Y}_n\left(\mathbf{s}_0\right)$ denotes the spatial prediction and $\varepsilon_t\left(\mathbf{s}_0\right)$ is the innovation error at $\mathbf{s}_0$ and at time t. Therefore, we obtain

$$\begin{aligned} E\left(Y_{n+h}\left(\mathbf{s}_0\right) - \widehat{Y}_{n+h}\left(\mathbf{s}_0\right)\right)^2 &= \varphi^{2h} E\left(Y_n\left(\mathbf{s}_0\right) - \widehat{Y}_n\left(\mathbf{s}_0\right)\right)^2 + \sigma^2 \frac{\varphi^{2h} - 1}{\varphi^2 - 1} \\ &= \varphi^{2h} \int_{-\pi}^{\pi} \sigma_m^2\left(\omega\right) d\omega + \sigma^2 \frac{\varphi^{2h} - 1}{\varphi^2 - 1}. \end{aligned} \tag{46}$$

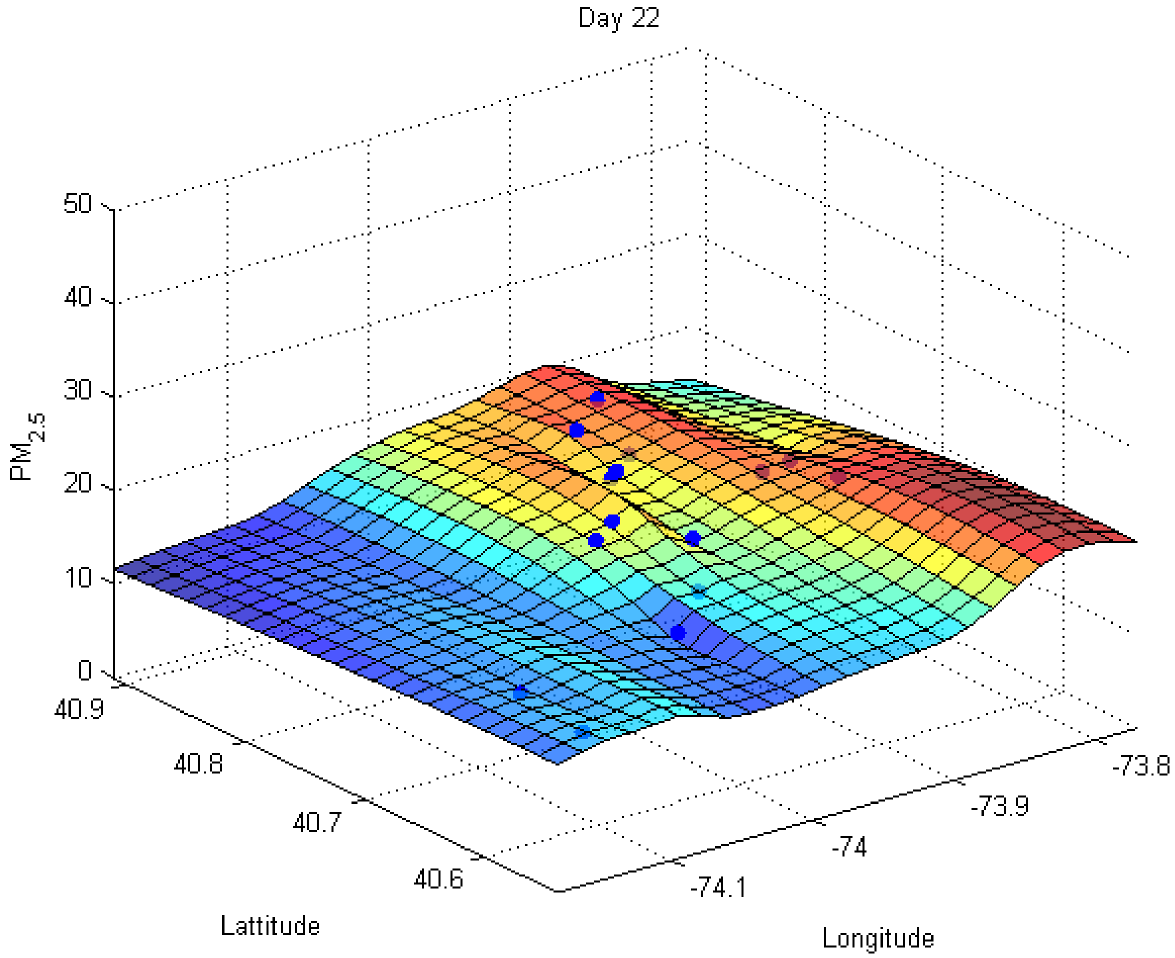

Figure 3: Estimated $PM_{2.5}$ for all the locations at $t = 22$.

We note that usually the parameters ϕ and σ^2 are unknown and the estimates are substituted. We note that since we are considering embedded Laplacian models for describing the spatio-temporal data , and used ARMA spectrum inside the operator , we can measure the efficiency of the models by the minimum mean square prediction errors computed from the ARMA models

.

7 Discussion:

Prediction of time series data at a known location $\mathbf{s}_0$ given the time series at other locations $\{\mathbf{s}_i; i = 1, 2, \ldots m\}$ is one of our objectives in this paper. Instead of predicting the data, we consider the prediction of its DFT, and then inversion of the DFT to obtain predictions of the data. The minimum mean square error predictor is in terms of the covariances of the DFT's $g_{||h||}(\omega)$, where $||h|| = ||\mathbf{s}_i - \mathbf{s}_j||$. Any valid space-time covariance function $c(||h||, n)$ can be used to obtain $g_{||h||}(\omega)$. In this paper we propose an embedded Laplacian model for the DFT's which we used to derive expressions for the covariances. The embedded Laplacian model proposed here takes into account both spatial correlation and temporal correlation .An analytic expression for $g_{||h||}(\omega)$ is derived. The space-time spectrum derived from this model using the proposed model is non-separable. Since the predictor depends on the covariances of the DFT's, one need to invert $m \times m$ dimensional matrices only and not $mn \times mn$ dimensional matrices as required in time domain approaches. Further, to evaluate the DFT's, one can use Fast Fourier Transform (FFT) Algorithms and it is well known that these require less number of computational operations. Besides the computational convenience, we may also note that DFT's are asymptotically uncorrelated and are asymptotically distributed as complex Gaussian. This property has been exploited in the estimation of the parameters of $g_{||h||}(\omega)$. The method of estimation proposed is based on the evaluation of the likelihood function of DFT's obtained from the differences of the spatial time series. The method is valid even when the original spatio-temporal data is not spatially, temporally stationary. The estimation methodology proposed here is still valid under the weaker assumption of Intrinsic stationarity (see Subba Rao and Terdik [2016]) We have shown that the methods proposed can also be used in situations when the data is corrupted by white noise. No distributional assumption of the process is necessary. Also it may be pointed out that the assumption here is that the time series data we observe is equally spaced and there are no missing values. It will be interesting to see how these methods can be modified in situations where such assumptions cannot be made..

Appendix

1. Proof of Theorem 2.

By proceeding as in Theorem 1, we can show that the spectral density function is given by

$$f_Y\left(\underline{\lambda}, \omega\right) = \frac{\sigma_e^2}{(2\pi)^{d+1}} \frac{1}{\left(\sum \lambda_i^2 + |c(\omega)|^2\right)^{2\nu}}.$$

To obtain the inverse transform we proceed as follows. Let $\rho = \|\underline{\lambda}\|$. We have,

$$\begin{aligned} g_{\|\mathbf{h}\|}(\omega) &= \frac{\sigma_e^2}{(2\pi)^{d+1}} \int_{\mathbb{R}^d} \frac{e^{-i\mathbf{h}\cdot\underline{\lambda}}}{\left(\|\underline{\lambda}\|^2 + |c(\omega)|^2\right)^{2\nu}} d\underline{\lambda} \\ &= \frac{\sigma_e^2}{(2\pi)^{d+1}} \int_0^\infty \frac{\rho^{d-1}}{\left(\rho^2 + |c(\omega)|^2\right)^{2\nu}} \int_{\mathbb{S}_{d-1}} e^{-i\rho\|\mathbf{h}\|\cos\alpha} d\Omega d\rho, \end{aligned}$$

where $\mathbb{S}_{d-1}$ is the unit sphere in $\mathbb{R}^d$ and Ω is Lebesgue element of surface area on $\mathbb{S}_{d-1}$.

We know further

$$\int_{\mathbb{S}_{d-1}} e^{-i\rho\|\mathbf{h}\|\cos\alpha} d\Omega = (2\pi)^{\frac{d}{2}} \left(\rho\|\mathbf{h}\|\right)^{-\frac{d}{2}+1} \mathcal{J}_{\frac{d}{2}-1}\left(\rho\|\mathbf{h}\|\right),$$

where $\mathcal{J}_{\frac{d}{2}-1}$ denotes the Bessel function of the first kind, see Stein and Weiss [1971], p. 176. Now we use Hankel-Nicholson Type Integral, see Abramowitz and Stegun [1992], 11. 4. 44, if $d < 4\nu + 3$, then

$$\int_0^\infty \frac{\mathcal{J}_{\frac{d}{2}-1}(r\rho)}{\left(\rho^2 + |c(\omega)|^2\right)^{2\nu}} \rho^{\frac{d}{2}} d\rho = \frac{r^{2\nu-1} |c(\omega)|^{\frac{d}{2}-2\nu}}{2^{2\nu-1}\Gamma(2\nu)} K_{\frac{d}{2}-2\nu}(r|c(\omega)|).$$

Using the above integrals and noting $K_{\frac{d}{2}-2\nu} = K_{2\nu-\frac{d}{2}}$, for all d, the covariance function can be shown to be

$$g_{\|\mathbf{h}\|}(\omega) = \frac{\sigma_e^2}{(2\pi)^{\frac{d}{2}+1} 2^{2\nu-1}\Gamma(2\nu)} \left(\frac{\|\mathbf{h}\|}{|c(\omega)|}\right)^{2\nu-\frac{d}{2}} K_{2\nu-\frac{d}{2}}(\|\mathbf{h}\| |c(\omega)|),$$

and the auto-correlation function is

$$\rho(\|\mathbf{h}\|, \omega) = \frac{g_{\|\mathbf{h}\|}(\omega)}{g_{\mathbf{0}}(\omega)} = \frac{(\|\mathbf{h}\| |c(\omega)|)^{2\nu-\frac{d}{2}}}{2^{2\nu-\frac{d}{2}-1}\Gamma\left(2\nu - \frac{d}{2}\right)} K_{2\nu-\frac{d}{2}}(\|\mathbf{h}\| |c(\omega)|),$$

since

$$g_0(\omega) = \frac{\sigma_e^2}{(2\pi)^{\frac{d}{2}+1} 2^{\frac{d}{2}} \left(|c(\omega)|^2\right)^{2\nu-\frac{d}{2}}} \frac{\Gamma\left(2\nu - \frac{d}{2}\right)}{\Gamma(2\nu)}.$$

2. Discrete Fourier Transforms and their properties

Let us assume that we have time series data from m locations spatially distributed.
Let $\{Y_t(\mathbf{s}_i) | i = 1, 2, \ldots m;\ t = 1, \ldots, n\}$, be a sample from the zero mean spatio-temporal stationary process $\{Y_t(\mathbf{s})\}$. Consider the time series data $\{Y_t(\mathbf{s}_i) | t = 1, \ldots, n\}$ at the location $\mathbf{s}_i$, and define the Discrete Fourier transform (DFT)

$$J_{\mathbf{s}_i}(\omega_k) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} Y_t(\mathbf{s}_i) e^{-it\omega_k}, \quad (47)$$

where $\omega_k = \frac{2\pi k}{n}$, $k = 0, \pm 1, \ldots, \pm\left[\frac{n}{2}\right]$. In practice one uses Fast Fourier Transform algorithm to compute the DFT. It is well known that the number of operations required to calculate n discrete Fourier transforms from n dimensional data is of order $n \log n$. The corresponding second order periodogram is defined as

$$I_{\mathbf{s}_i}(\omega_k) = |J_{\mathbf{s}_i}(\omega_k)|^2,$$

and the cross periodogram between the time series $\{Y_t(\mathbf{s}_i)\}$ and $\{Y_t(\mathbf{s}_j)\}$ is given by $I_{\mathbf{s}_i s_j}(\omega_k) = J_{\mathbf{s}_i}(\omega_k) J^*_{\mathbf{s}_j}(\omega_k)$.

It is well known that the periodogram is asymptotically an unbiased estimator of the second order spectral density function, but it is not mean square consistent and hence to obtain a consistent estimate the periodograms are smoothed using various kernels (see Priestley [1981]). It is well known that (see for example Priestley [1981])

$$\begin{aligned} &E(J_{\mathbf{s}_i}(\omega)) = 0, \\ &Var(J_{\mathbf{s}_i}(\omega)) = E(I_{\mathbf{s}_i}(\omega)) \simeq g_{\mathbf{s}_i}(\omega) = g_0(\omega), \end{aligned} \quad (48)$$

where $g_0(\omega)$ is the second order spectral density function of the random process. We further note that $\sigma_Y^2 = \int g_0(\omega) d\omega$.

.

In the following propositions we summarise the well known properties of the Fourier Transforms of stationary process. For details refer to Brillinger [2001], Giraitis et al. [2012], Priestley [1981], Dwivedi and Subba Rao [2011]. We refer to Lahiri [2003b], Yajima [1989], Robinson [1995] for further results regarding the DFTs.

Proposition 1 *Let $J_{\mathbf{s}_i}(\omega_k)$ and $J_{\mathbf{s}_j}(\omega_k)$ be the discrete Fourier Transforms of the spatio-temporal stationary processes $\{Y_t(\mathbf{s}_i)\}$, $\{Y_t(\mathbf{s}_j)\}$ respectively. For large n,*

$$Cov(J_{\mathbf{s}_i}(\omega_k), J_{\mathbf{s}_i}(\omega_{k'})) \simeq 0, \quad k \neq k',$$

$$Cov\left(J_{\mathbf{s}_i}(\omega_k), J_{\mathbf{s}_j}(\omega_k)\right) = E\left[I_{\mathbf{s}_i\mathbf{s}_j}(\omega_k)\right] \simeq \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} c(\mathbf{s}_i - \mathbf{s}_j, n)\, e^{-in\omega_k} = g_{\mathbf{s}_i - \mathbf{s}_j}(\omega_k), \tag{49}$$

where $g_{(\mathbf{s}_i-\mathbf{s}_j)}(\omega)$ is defined as the cross spectrum between the two processes and it is usually a complex valued function. If the process is isotropic, then

$$c(\mathbf{s}_i - \mathbf{s}_j, n) = c(\|\mathbf{s}_i - \mathbf{s}_j\|, n) = c(\|\mathbf{s}_i - \mathbf{s}_j\|, -n).$$

and under the isotropy assumption the cross spectrum $g_{\mathbf{s}_i-\mathbf{s}_j}(\omega)$ between the two processes reduces to

$$g_{\|\mathbf{h}\|}(\omega) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} c(\|\mathbf{h}\|, n) e^{-in\omega}, \quad |\omega| \le \pi, \tag{50}$$

and the spectral function $g_{\|\mathbf{h}\|}(\omega)$, where $\|\mathbf{h}\| = \|\mathbf{s}_i - \mathbf{s}_j\|$, is real and symmetric in ω.

Proof. The above results are well known and hence details are omitted. ■

Consider the DFTs $\{J_{\mathbf{s}_i}(\omega_k(n))\}$, where $\omega_k(n) = 2\pi s_k(n)/n, k = 1, 2, \ldots K$ are distinct and $\{s_k(n)\}$ are integers, and as $n \longrightarrow \infty$, $\omega_k(n) \rightharpoonup \omega_k$,. Let $\omega_j(n) \pm \omega_k(n) \neq 0 \pmod{2\pi}$ for $1 \le j < k < K$.

Proposition 2 *Under the above conditions of the frequencies, The DFT's $\{J_{\mathbf{s}_i}(\omega_k(n)); k = 1, 2, 3 \ldots K\}$ are asymptotically independent and has a multivariate Complex Gaussian distribution.*

Proof. For Proof, we refer to Brillinger [2001], Lahiri [2003a], Robinson [1995], Giraitis et al. [2012], Yajima [1989]. ■

As pointed out earlier the second order periodogram $I_{\mathbf{s}_i}(\omega)$ defined above is always real, where as the cross periodogram $I_{\mathbf{s}_i,\mathbf{s}_j}(\omega_k)$ defined above between the spatial processes $\{Y_t(\mathbf{s}_i)\}$ and $\{Y_t(\mathbf{s}_j)\}$ is usually a complex valued function. Under the isotropy assumption, however, the cross spectrum is a function of the Euclidean distance $\|\mathbf{h}\| = \|\mathbf{s}_i - \mathbf{s}_j\|$ and the temporal frequency ω (i. e., spectral in time, but not in space) and, therefore, it is real. It is interesting to see the similarity between the above function and the spectral density functions defined by Cressie and Huang [1999] and Stein [2005a], Stein [2005b], which are spectral in time but not in space.

In the following proposition we show that the Discrete Fourier Transform of $Y_t(\mathbf{s})$ can be written in terms of orthogonal set function $Z_Y(\underline{\lambda}, \omega)$.

Proposition 3 *Let $J_{\mathbf{s}}(\omega)$ be the Discrete Fourier Transform of $\{Y_t(s)\}$ and let the spectral representation of $Y_t(s)$ be given by (3). Then*

$$J_{\mathbf{s}}(\omega) \simeq \int e^{i\underline{s}\,\underline{\lambda}}\sqrt{\frac{n}{2\pi}}\, dZ_Y(\underline{\lambda}, \omega). \tag{51}$$

Proof. Substitute the spectral representation (3) for $Y_t(\mathbf{s})$ in (47), and after some simplification, we obtain

$$J_{\mathbf{s}}(\omega) = \int\int e^{i\underline{s}\,\underline{\lambda}}\left[e^{i(n+1)\frac{\varphi}{2}} F_n^{\frac{1}{2}}(\varphi)\right] dZ_Y(\underline{\lambda}, \omega), \tag{52}$$

where $\varphi = \mu - \omega$, $\int$ is a d dimensional multiple integral, (see Priestley [1981], p. 419) and in obtaining the above, we used the result

$$\sum_{t=1}^{n} e^{it\varphi} = e^{i(n+1)\frac{\varphi}{2}}\left[\frac{\sin n\frac{\varphi}{2}}{\sin\frac{\varphi}{2}}\right] = e^{i(n+1)\frac{\varphi}{2}}\sqrt{2\pi n}F_n^{\frac{1}{2}}(\varphi),$$

where the Fejér kernel $F_n(\varphi)$ is given by

$$F_n(\varphi) = \frac{1}{2\pi n}\frac{\sin^2 n\frac{\varphi}{2}}{\sin^2\frac{\varphi}{2}}.$$

It is well known that the Fejér kernel behaves like a Dirac Delta function as $n \to \infty$ and as $\varphi \to 0$, $F_n(\varphi) = O(n)$. As pointed out by Priestley [1981], p. 419), that $F_n^{\frac{1}{2}}(\varphi)$ does not strictly tend to a Dirac Delta $\delta-$function as $n \to \infty$, nevertheless, behaves in a similar manner to a $\delta-$function. In particular as $n \to \infty$ and for all $\varphi \neq 0$, $F_n^{\frac{1}{2}}(\varphi) \to 0$, and as $\varphi \to 0$, $F_n^{\frac{1}{2}}(\varphi) \to \sqrt{n/2\pi}$. Therefore, as $n \to \infty$, $F_n^{\frac{1}{2}}(\varphi)$ vanishes everywhere except at the origin. In view of this, for large n, we have the result

$$J_{\mathbf{s}}(\omega) \simeq \int e^{i s \underline{\lambda}} \sqrt{\frac{n}{2\pi}} dZ_Y(\underline{\lambda}, \omega).$$

We note that the above integral is over the wave number space $\underline{\lambda}$ only. ■

Proposition 4 *Let $\{e_t(\mathbf{s}) | \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{Z}\}$ be a white noise process in space and time, that is, it is a generalized process with constant spectrum, (see Yaglom [1987], 24, p. 411) it satisfies the following conditions.*

$$E(e_t(\mathbf{s})) = 0,$$
$$Var(e_t(\mathbf{s})) = \sigma_e^2, \textit{ does not depend on } \mathbf{s} \textit{ or } t,$$
$$Cov(e_t(\mathbf{s}), e_{t'}(\mathbf{s})) = \sigma_e^2 \delta(\mathbf{s} - \mathbf{s}') \delta_{t-t'},$$

where $\delta(\mathbf{s} - \mathbf{s}')$ denotes the Dirac Delta function and

$$\delta_{t-t'} = \begin{cases} 1 & \textit{if } t = t', \\ 0 & \textit{otherwise.} \end{cases},$$

is the Kronecker delta. Then the Discrete Fourier transform of the white noise process, formally

$$J_{\mathbf{s},e}(\omega) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} e_t(\mathbf{s}) e^{-it\omega},$$

can be approximated by

$$J_{\mathbf{s},e}(\omega) \simeq \int e^{i\mathbf{s}\cdot\underline{\lambda}} \left[\sqrt{\frac{n}{2\pi}} \right] dZ_e(\underline{\lambda}, \omega), \tag{53}$$

where the orthogonal random process $Z_e(\underline{\lambda}, \omega)$ satisfies

$$EdZ_e(\underline{\lambda}, \omega) = 0,$$
$$E|dZ_e(\underline{\lambda}, \omega)|^2 = \frac{\sigma_e^2}{(2\pi)^{d+1}} d\underline{\lambda} d\omega.$$

Proof. It is similar to Proposition 3 and hence omitted. ■

Acknowledgement. *The visit of Subba Rao to the CRRAO AIMSCS was supported by a grant from the Department of Science and Technology, Government of India, grant number SR/S4/516/07. The authors are thankful to the editor, co-editor and the referees for their valuable comments and suggestions. They are also thankful to Professor L. Giraitis, Queen Mary University,London and Dr Suhasini Subba Rao, Texas A&M University, USA for their suggestions for improvement of the paper.*

References

M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications Inc., New York, 1992. ISBN 0-486-61272-4. Reprint of the 1972 edition.

A. M. C Antunes and T. Subba Rao. On hypotheses testing for the selection of spatio-temporal models. *Journal of Time Series Analysis*, 27(5):767–791, 2006.

S. Banerjee, B. P Carlin, and A. E Gelfand. *Hierarchical modeling and analysis for spatial data*. Crc Press, 2014.

M. Bevilacqua, C. Gaetan, J. Mateu, and E. Porcu. 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, 2012.

M Blangiardo and M. Cameletti. *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons, 2015.

N. Bliznyuk, R. J Carroll, M. G Genton, and Y. Wang. Variogram estimation in the presence of trend. *Statistics and its interface*, 5(2):159, 2012.

D. R. Brillinger. *Time Series; Data Analysis and Theory*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. ISBN 0-89871-501-6. Reprint of the 1981 edition.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, New York, 1987.

P. F. Craigmile and P. Guttorp. Space-time modelling of trends in temperature series. *Journal of Time Series Analysis*, 32(4):378–395, 2011.

N. Cressie. *Statistics for Spatial Data*. Wiley Series in Probability & Mathematical Statistics, 1993.

N. Cressie and H.-C. Huang. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448):1330–1339, 1999.

N. Cressie and C. K. Wikle. *Statistics for Spatio-Temporal Data*. Wiley Series in Probability and Statistics, 2011.

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

P. J. Diggle and P. J. Ribeiro. *Model-based geostatistics*. Springer, 2007.

W. Dunsmuir. A central limit theorem for estimation in gaussian stationary time series observed at unequally spaced times. *Stochastic Processes and their Applications*, 14(3):279–295, 1983.

W. Dunsmuir and P. M. Robinson. Asymptotic theory for time series containing missing and amplitude modulated observations. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 260–281, 1981.

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(1):68–91, 2011.

L. Giraitis, H. L. Koul, and D. Surgailis. *Large sample inference for long memory processes*, volume 201/2. World Scientific, 2012.

R. Giraldo, P. Delicado, and J. Mateu. Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of agricultural, biological, and environmental statistics*, 15(1):66–82, 2010.

T. Gneiting. Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, 97(458):590–600, 2002.

J. Guinness and M. Fuentes. Likelihood approximations for big nonstationary spatial temporal lattice data. *Statistica Sinica*, 25(1):329–349, 2015.

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

P. Hall and I. V. Keilegom. Using difference-based methods for inference in nonparametric regression with time series errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65 (2):443–456, 2003.

R. H. Jones and Y. Zhang. Models for continuous stationary space-time processes. In *Modelling longitudinal and spatially correlated data*, pages 289–298. Springer, 1997.

S. N. Lahiri. Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhyā: The Indian Journal of Statistics*, pages 356–388, 2003a.

SN Lahiri. A necessary and sufficient condition for asymptotic independence of discrete fourier transforms under short-and long-range dependence. *Annals of Statistics*, pages 613–641, 2003b.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

C. Ma. Spatio-temporal covariance functions generated by mixtures. *Mathematical geology*, 34(8):965–975, 2002.

C. Ma. Spatio-temporal stationary covariance models. *Journal of Multivariate Analysis*, 86(1):97–107, 2003.

H. Ombao, X. Shao, E. Rykhlevskaia, M. Fabiani, and G. Gratton. Spatio-spectral analysis of brain signals. *Statistica Sinica*, 18(4):1465, 2008.

E. Parzen, editor. *Time Series Analysis of Irregularly Observed Data: Proceedings of a Symposium Held at Texas a and M University, College Station, February 10-13, 1983*, 1984. Springer-Verlag.

Phillip E Pfeifer and Stuart Jay Deutrch. A three-stage iterative procedure for space-time modeling phillip. *Technometrics*, 22(1):35–47, 1980.

M. B. Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 204–237, 1965.

M. B. Priestley. *Spectral Analysis and Time Series.* Academic Press, New York, 1981.

P. M. Robinson. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23(3):1048 – 1072, 1995. URL http://www.jstor.org/stable/2242436.

M. D. Ruiz-Medina. New challenges in spatial and spatiotemporal functional statistics for high-dimensional data. *Spatial Statistics*, 1:82–91, 2012.

S. K. Sahu and K. V. Mardia. A bayesian kriged kalman model for short-term forecasting of air pollution levels. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):223–244, 2005.

J. D. Scargle. Studies in astronomical time series analysis III. fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data. *The Astrophysical Journal*, 343:874–887, 1989.

O. Schabenberger and C. A. Gotway. *Statistical methods for spatial data analysis.* CRC Press, 2005.

F. Sigrist, H. R Künsch, and W. A Stahel. Stochastic partial differential equation based modelling of large space–time data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):3–33, 2015.

E. M. Stein and G. Weiss. *Introduction to Fourier analysis on Euclidean spaces.* Princeton University Press, Princeton, N.J., 1971. Princeton Mathematical Series, No. 32.

M. L. Stein. *Interpolation of spatial data: some theory for kriging.* Springer, 1999.

M. L. Stein. Space–time covariance functions. *Journal of the American Statistical Association*, 100(469): 310–321, 2005a.

M. L. Stein. Statistical methods for regular monitoring data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(5):667–687, 2005b.

M. L. Stein, Z. Chi, and L. J. Welty. Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2):275–296, 2004.

T. Subba Rao and Gy. Terdik. On the frequency variogram and on frequency domain methods for the analysis of spatio-temporal data. *ArXiv e-prints, 1610.05891 [math.ST]*, 2016. (submited for publication).

T. Subba Rao, S. Das, and G. Boshnakov. A frequency domain approach for the estimation of parameters of spatio-temporal random processes. *Journal of Time Series Analysis*, 35:357–377, 2014.

A. V. Vecchia. Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50:297–312, 1988.

P. Whittle. On stationary processes in the plane. *Biometrika*, 41(3/4):434–449, 1954.

A. M. Yaglom. *Correlation Theory of Stationary Related Random Functions,*, volume I. Springer-Verlag, New York, 1987.

Y. Yajima. A central limit theorem of fourier transforms of strongly dependent stationary processes. *Journal of time series analysis*, 10(4):375–383, 1989.

JTSA 3963 R3
Reviewer 1.

Dear Reviewer,

Thank you for your comments. We are sorry that you still feel that things are not clear. Below, we describe briefly the steps involved which may help.

GENERAL METHODOLOGY

1. Instead of considering the data $\{Y_t\ (\mathbf{s}_i);\}t = 1, 2, ..n; i = 1, 2, ...m\}$ we consider their DFt's $\{\ J_{\mathbf{s}_i}(\omega); i = 1, 2...m\}$.as our data. We note that there is a one-to-one correspondence between the DATA and the DFT's. It is well known that DFT's for a stationary process are asymptotically uncorrelated (over distinct frequencies) and Gaussian. (The properties are briefly discussed in the Appendix of the paper).Therefore the periodograms (properly scaled) are approximately distributed as Chi squares with two degrees of freedom(ie exponential) at all frequencies except at $\omega \neq 0, \pm\pi$ This property is used to obtain the likelihood function based on DFT's and used for the estimation purposes in section 4.1(Pl. see Brillinger's book for full discussion on asymptotic properties of DFT's).

2. Since we consider the DFT's as our data, we predict the DFT of the original data at the location $\mathbf{s}_0$given the DFT's at other locations. By inversion of the DFT's we can obtain the data.

3. The optimal predictor of the DFT at $\mathbf{s}_0$.depends on the covariance between the DFTs at various locations. In order to evaluate this predictor, we can substitute any valid space-time covariance function. We note that various covariance functions have been suggested. in the literature No close form for the covariance functions are available if we assume that the random process satisfies the standard diffusion models such as described by Jones and Zhang (1997), Sigrist et al (2015), Lindgren et al (2011) etc, even though one can obtain space time spectrum. This can be considered as a serious set back if one wants to follow a covariance based approach for prediction. If one wants to use these models the covariance functions need to be evaluated by numerical integration.

4. In view of this we defined a new set of models (11) and (23) (see Th 1 for $d = 2.Th\ 2$ for$d > 2$) for the DFT's which are of Laplacian type with scaling function $|c(\omega)|^2$.

5. We have shown that this scaling function, which is a polynomial in temporal frequency, is in fact related to the Inverse second order temporal spectrum. In other words the temporal spectrum is embedded into Laplacian model, resulting in a covariance function of Matern-Whitle type covariance function which takes into account temporal dependence, spatial dependence and their interaction, if any. The spectral density function of the process corresponding to the model defined is non-separable.

6. For the embedded model defined above, we obtained a close form for the covariance function which we used in obtaining the optimal predictors. The

methods proposed are extended to situations where the observations are corrupted by noise.

7. The covariance function defined here depends on various parameters and the scaling function $|c(\omega)|^2$ which is shown to be proportional inverse spectrum, and this needs to be estimated from the data.

8. We assumed that the second order spectral density function can be modelled by an ARMA (p,q) spectrum

9. The ARMA parameters and the other parameters associated with spatial dependence are estimated using the "entire" sample. We considered the Frequency variogram approach(defined in earlier papers) for the estimation.

10. The efficiency of the Embedded model defined can be assessed by the predictive performance of the ARMA (p,q) model. using the existing forecasting techniques (see the books of Box and Jenkins, Brockwell and Davis for details). We choose that order p and q for which the mean square error is minimum. For the real data we considered, we found AR(1) is appropriate, and using the estimated model, we predicted the data at the location 11(see Fig 2), and at the location 8, we considered forecasting of the data corresponding to t=84,85...91.

we believe the method described can be used for analysing any spatio-temporal data

GENERAL COMMENTS OF THE REVIEWER 1

1.we are sorry that you find the simulation study difficult to understand. we explain below the purpose behind simulation study.

At the suggestion of the Reviewer of the previous submissions, we added a section on Simulation of spatio-Temporal data(pl. see section 5). We have rewritten this section and divided the section into two separate parts. We thank the reviewer for his/her comments which helped us to rewrite. Generating time series data with a known spectral density function is interesting and such generated data are often used for testing any new methodology.

There are several methods available for simulating spatial data with a given covariance function,. for example one can refer to Cressie (1993) and Schabenberger and Gotway (2005) for details. As far as we are aware not many such methods are available for generating spatio-temporal data with a given space-time covariance function or equivalently space-time spectrum. In this paper we present a method(section 5, please see the steps described in the paper) for simulating the DFT's and then generating the data from the DFT's. We described the method for generating the spatio-temporal data (when d=2,m=9, n=2^{11}). We assumed that the time series model at each location can be described by an ARMA (2,1) model and thus the ARMA spectrum $g_o(\omega)$ is related to$|c(\omega)|^2$ in the model defined in Theorems 1and 2. Using the simulated data and the method described in section 4.1 we estimated the parameters of ARMA (2,1)model and also considered the prediction of the data at the location $\mathbf{s}_o$.We included this numerical illustration to high light the steps to help the readers. .

2. About our Data Analysis :. Our object here is to illustrate the methodology with some real data. We apologise to the reviewer if he/she thinks the

data is not interesting and exciting. We are sure that the reviewer knows that it is quite common amongst time series analysts and spatial analysts to use repeatedly the same data for different purposes. For example the two sets of data, Canadian Lynx data, Wolfer's annual sunspot numbers have been used several times (the authors of this paper used several time series mentioned. Very likely the reviewers too) by time series analysts and Irish Wind data (first analysed by Haslett and Raftery) is now widely analysed by spatio-temporal analysts. We are only illustrating our methods of prediction and estimation. Since we are analysing the data by the classical time series methods, accuracy of the forecasts are assessed by Prediction intervals and prediction bands and these are shown in Fig 2. We are not claiming the analysis is exhaustive.

SPECIFIC COMMENTS OF THE REVIEWER 1

1.We included a reference to the paper by Fuentes and Guinness(2015). We added one para at the end of section 4.1. Please note that we approached the problem of estimation somewhat differently. We considered the likelihood function of the DFT's of the intrinsic stationary process, and then maximized likelihood obtained from the dFT's from the differences of the processes for a given distance $||h_l||$,then pooled all likelihoods over all possible distances. The efficiency of the method and sampling properties of the estimates were discussed in an earlier by Subba Rao et al (2014) and hence details are omitted to avoid repetition and to save space.

2. We included a ' Discussion' section as requested by the Reviewer

3. We removed some Figures and as suggested we made legends more clear. We also pointed out that we considered through out detrended series only. We considered the prediction of $PM_{2.5}$at all the 625 locations on a specific day t=22 (pl. see Fig 3). The observed data at 14 locations are shown in bold circles.

JTSA 3963 R3

REVIEWER 2.

Dear Reviewer,

We sincerely apologise to you for giving the impression that we are wearing you down. It was never our intention to criticize other authors or wear you down. Our intentions are very simple. We are proposing an alternative approach using DFT's instead of the original data. Because the DFT's are complex valued, related to the periodogram (thus the spectrum), we modified the model considered (Th 1 and Th2) and prediction in terms of DFT's. The likelihood function of the DFT's of the intrinsic processes is considered for prediction.

GENERAL METHODOLOGY.

Below we summarise the methods suggested which we hope will make things clear.

1. Instead of considering the data $\{Y_t\ (\mathbf{s}_i);\}t = 1, 2, ..n; i = 1, 2, ...m\}$ we consider their DFt's $\{\ J_{\mathbf{s}_i}(\omega); i = 1, 2...m\}$.as our data. We note that there is a one-to-one correspondence between the DATA and the DFT's. It is well known that DFT's for a stationary process are asymptotically uncorrelated (over distinct frequencies) and Gaussian. (The properties are briefly discussed in the Appendix of the paper).Therefore the periodograms (properly scaled) are approximately distributed as Chi squares with two degrees of freedom(ie exponential) at all frequencies except at $\omega \neq 0, \pm\pi$ This property is used to obtain the likelihood function based on DFT's and used for the estimation purposes in section 4.1(Pl. see Brillinger's book for full discussion on asymptotic properties of DFT's).

2. Since we consider the DFT's as our data, we predict the DFT of the original data at the location $\mathbf{s}_0$given the DFT's at other locations. By inversion of the DFT's we can obtain the data.

3. The optimal predictor of the DFT at $\mathbf{s}_0$.depends on the covariance between the DFTs at various locations. In order to evaluate this predictor, we can substitute any valid space-time covariance function. We note that various covariance functions have been suggested. in the literature No close form for the covariance functions are available if we assume that the random process satisfies the standard diffusion models such as described by Jones and Zhang (1997), Sigrist et al (2015), Lindgren et al (2011) etc, even though one can obtain space time spectrum. This can be considered as a serious set back if one wants to follow a covariance based approach for prediction. If one wants to use these models the covariance functions need to be evaluated by numerical integration.

4. In view of this we defined a new set of models (11) and (23) (see Th 1 for $d = 2.Th\ 2$ for$d > 2$) for the DFT's which are of Laplacian type with scaling function $|c(\omega)|^2$

5. We have shown that this scaling function, which is a polynomial in temporal frequency, is in fact related to the Inverse second order temporal spectrum. In other words the temporal spectrum is embedded into Laplacian model, resulting in a covariance function of Matern-Whitle type covariance function which takes into account temporal dependence, spatial dependence and their interaction, if any. The spectral density function of the process corresponding to the model defined is non-separable.

6. For the embedded model defined above, we obtained a close form for the covariance function which we used in obtaining the optimal predictors. The methods proposed are extended to situations where the observations are corrupted by noise.

7. The covariance function defined here depends on various parameters and the scaling function $|c(\omega)|^2$ which is shown to be proportional inverse spectrum, and this needs to be estimated from the data.

8. We assumed that the second order spectral density function can be modelled by an ARMA (p,q) spectrum

9. The ARMA parameters and the other parameters associated with spatial dependence are estimated using the "entire" sample. We considered the Frequency variogram approach(defined in earlier papers) for the estimation.

10. The efficiency of the Embedded model defined can be assessed by the predictive performance of the ARMA (p,q) model. using the existing forecasting techniques (see the books of Box and Jenkins, Brockwell and Davis for details). We choose that order p and q for which the mean square error is minimum. For the real data we considered, we found AR(1) is appropriate, and using the estimated model, we predicted the data at the location 11(see Fig 2), and at the location 8, we considered forecasting of the data corresponding to t=84,85...91.

GENERAL COMMENTS of REVIEWER 2

1. The reviewer 2 commented on our lack of understanding the implications of MV AR(1) and spatio-temmporal modelling. We are sorry that the Reviewer felt this way. We added a paragraph (last paragraph) of section 2 pointing out the differences between MV AR modelling and space time modelling. We refer to the paper by Antunes and Subba Rao (2006) where the authors examined the differences between MV AR models and STARMA models. Also we refer to Stein (2005b) section 1, where he devoted one whole section on the advantages of space time modelling compared to Multivariate modelling where the spatial correlation is not taken into account. We hope this is what the reviewer 2 wants.

2. We cannot comment on the reviewer 2's subjective opinion that the method would not have much utility in real world spatio-temporal problems where the number of spatial locations can be huge. We can only request the co- editor and the editor to give the methods proposed a chance and only time will tell. We believe that the practitioners should have several techniques at their disposable and they can choose what they want. It is like a Free market economy in a democratic world. We are not forecasting at the moment.

3. We included a 'Discussion' section.

4. We included a reference to the paper by Fuentes and Guinness(2015). We added one para at the end of section 4.1. Please note that we approached the problem of estimation somewhat differently. We considered the likelihood function of the DFT's of the intrinsic stationary process, and then maximized likelihood obtained from the dFT's from the differences of the processes for a given distance $||h_l||$, then pooled all likelihoods over all possible distances. The efficiency of the method and sampling properties of the estimates were discussed in an earlier by Subba Rao et al (2014) and hence details are omitted to avoid repetition and to save space

SPECIFIC COMMENT
.About Simulation. Please see the Reply to the Reviewer 1.above.,

Thank you

T Subba Rao
Gy.Terdik