Multivariate Spatio-temporal modeling and simulation

by

Rigele Te

B.S., Minzu University of China, 2016

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

Doctor of Philosophy

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Abstract

In view of multivariate nature of general spatio-temporal data sets observed in various disciplines such as meteorology, engineering and medical science, we are interested in gaining some scientific insight into the underlying stochastic processes, which incorporate the complex dependence structure among different variables at different locations over time. To this end, this dissertation studies the methods and applications of the multivariate modelling, simulation and missing value imputation of data sets in space and time under different settings.

Chapter I, Space-time data sets are often multivariate and collected at monitored discrete time lags, which are usually viewed as a component of time series in environmental science and related areas. Valid and practical covariance models are needed to characterize geostatistical formulations of these types of data sets in a wide range of applications. We propose several classes of multivariate spatio-temporal functions to characterize underlying random fields whose discrete temporal margins are some celebrated autoregressive and moving average (ARMA) models, and obtain sufficient and/or necessary conditions for them to be valid covariance matrix functions. The possibility of taking advantage of well-established time series and spatial statistics tools makes it relatively straightforward to identify and fit the proposed models in practice. Finally, applications of the proposed multivariate covariance matrix functions are illustrated on Kansas weather data in terms of co-kriging, compared with some traditional space-time models for prediction.

Chapter II, We propose an efficient method for simulating multivariate spatio-temporal data on a compact two-point homogeneous space with sphere as a special case. These large scale global data sets are obtained based on truncating the series expansion of multivariate spatio-temporal random fields on this space. The algorithm can be boiled down to simply simulate a uniformly distributed random vector on a sphere, on which the great circle

distance is defined. Multiple covariance models are compared to fit the simulated multivariate space-time data including the model proposed in the Chapter I. The simulation studies suggest some guideline for choosing appropriate models and parameterizations for different multivariate data in space and time.

Chapter III, Motivated by dealing with incompleteness of energy data to study network situational awareness, we proposed a spatial Gaussian process variational autoencoders (GP-VAE) to impute the corrupted or missing information based on multivariate Gaussian processes and Bayesian deep learning. The missing data is learned by projecting the multivariate data space including both space and time into a lower dimensional latent space without missingness, where the low dimensional dynamics is modeled with vector Gaussian time series. Model comparison has been made with traditional data imputation method in literature on a simulated energy data set from smart meters, solar inverters, grid automation/SCADA sensors and micro PMU on a geographical lattice over time.

Multivariate Spatio-temporal modeling and simulation

by

Rigele Te

B.S., Minzu University of China, 2016

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

Doctor of Philosophy

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Juan Du

Copyright

Abstract

In view of multivariate nature of general spatio-temporal data sets observed in various disciplines such as meteorology, engineering and medical science, we are interested in gaining some scientific insight into the underlying stochastic processes, which incorporate the complex dependence structure among different variables at different locations over time. To this end, this dissertation studies the methods and applications of the multivariate modelling, simulation and missing value imputation of data sets in space and time under different settings.

Chapter I, Space-time data sets are often multivariate and collected at monitored discrete time lags, which are usually viewed as a component of time series in environmental science and related areas. Valid and practical covariance models are needed to characterize geostatistical formulations of these types of data sets in a wide range of applications. We propose several classes of multivariate spatio-temporal functions to characterize underlying random fields whose discrete temporal margins are some celebrated autoregressive and moving average (ARMA) models, and obtain sufficient and/or necessary conditions for them to be valid covariance matrix functions. The possibility of taking advantage of well-established time series and spatial statistics tools makes it relatively straightforward to identify and fit the proposed models in practice. Finally, applications of the proposed multivariate covariance matrix functions are illustrated on Kansas weather data in terms of co-kriging, compared with some traditional space-time models for prediction.

Chapter II, We propose an efficient method for simulating multivariate spatio-temporal data on a compact two-point homogeneous space with sphere as a special case. These large scale global data sets are obtained based on truncating the series expansion of multivariate spatio-temporal random fields on this space. The algorithm can be boiled down to simply simulate a uniformly distributed random vector on a sphere, on which the great circle

distance is defined. Multiple covariance models are compared to fit the simulated multivariate space-time data including the model proposed in the Chapter I. The simulation studies suggest some guideline for choosing appropriate models and parameterizations for different multivariate data in space and time.

Chapter III, Motivated by dealing with incompleteness of energy data to study network situational awareness, we proposed a spatial Gaussian process variational autoencoders (GP-VAE) to impute the corrupted or missing information based on multivariate Gaussian processes and Bayesian deep learning. The missing data is learned by projecting the multivariate data space including both space and time into a lower dimensional latent space without missingness, where the low dimensional dynamics is modeled with vector Gaussian time series. Model comparison has been made with traditional data imputation method in literature on a simulated energy data set from smart meters, solar inverters, grid automation/SCADA sensors and micro PMU on a geographical lattice over time.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to acknowledge and extend my greatest gratitude to my Major Professor Dr. Juan Du for her invaluable advice, continuous support, and patience during my PhD study. During these three and half years, I faced numerous challenges in my research study. I appreciate her always on my side, cheering me on and helping me stay on the right track. Thanks for Dr. Du helping me revise my thesis until late night even though she is extremely tired. I would not have been able to complete this dissertation without her generous sharing of her knowledge and expertise.

I also would like to thank my committee members, Dr. Vahl, Dr. Song and Dr. Wang. Dr. Vahl made significant suggestions on this dissertation that helped me produce more rigorous results. I appreciate Dr. Wang's assistance in helping me understand the combination of statistical theorems and real-world Geography data. I also want to appreciate Dr. Song for his Probability as well as Time Series Analysis classes. His classes helped me to solidify my skills in statistical theorem proof and to learn critical time series knowledge systematically, which helped me greatly in my research study.

In addition, I would like to thank Dr. Bala Natarajan and his group members, particularly Shweta Dahale. They provided us the original data set in Chapter 3. Based on the collaboration with them, we had the opportunity to learn about this interesting data inputting topic in engineering, which led to the proposed spatial GP-VAE data imputation method. They also gave me a lot of advice on modeling and coding in Chapter 3.

In closing, I would like to express my gratitude to my family and friends who have supported me throughout my academic pursuits.

Dedication

This is dedicated to my grand mother. For raising me up from a little girl. For her unconditional love and dedication to me more than 20 years. For her kindness to everyone that deeply affected me. And to my parents Getu Chao and Yuhua Zhao. For their mental and financial support. For believing me, encouraging me to accomplish more than I had. I remember whenever I'm feeling worn out, their encourage and love is always a cure that help me encounter every difficulties. And to my husband, Yuwen Liu. For his consideration. For always cooking delicious dinner for me. For understanding me. For laughing, singing, crying with me. For long-term company. I'm not able to complete this dissertation without them.

And to my uncles, my cousins and my friends. Love makes me strong and I dedicate this work to them. Finally, I'd like to thanks to myself. Thank you for being brave, and never give up!

Chapter 1

Multivariate Modeling of Some Datasets in Continuous Space and Discrete time

1.1 Introduction

Studies in environmental science, meteorology, and geophysics and many other areas often need to deal with multivariate type of data in space and time, such as studying the impact of soil greenhouse gas fluxes on global warming potential, learning the temperature precipitation relationships in climate change, etc. (see Gaspari and Cohn (1999), Sain et al. (2011), Tebaldi and Lobell (2008), among others). Comparing with the irregular spatial locations like whether stations, the temporal points are often collected in a much more evenly spaced manner. It is of ever-increasing importance to efficiently model the complex dependence and correlated structure exhibited in these datasets.

In this chapter, we focus on constructing valid covariance matrix structure that can integrate marginal space and time information together for multivariate modeling of underlying random fields. In spatial statistics literature, limited types of multivariate spatial models incorporating time information can be found, while time plays an important role in most

of environmental and geophysical processes. Reasonable geostatistical formulations capturing the temporal information that fluctuates regularly over time can inevitably improve understanding and analysis of these types of data. Traditionally, due to the computational convenience and lacking of non-separable model, separable spatial-temporal model have been popular by simply timing the purely spatial correlation function and purely temporal correlation together to formulate the spatio-temporal covariance structure, in which space-time interaction is ignored, even when space-time interaction plays an important role in most of environmental and geophysical process. Until recently nonseparability of spatio-temporal models have been emphasized and addressed. For example, Cressie and Huang (1999) provides a method to derive nonseparable, stationary univariate spatio-temporal covariance functions, and other closely related generalizations for stationary or nonstationary cases are developed by Ma (2003), Castruccio and Stein (2013), among others. Recently Wan et al. (2021) and Xu et al. (2019) proposed several spatio-temporal model for particulate matter and gaseous pollutants in Beijing, China. Medeiros et al. (2019) shows a novel way of modeling the precipitation trend component by using an inflated gamma distribution of zeros. However, most of these methods mainly focus on modelling in the continuous time scales. In practice, most of the data are recorded in a discrete regular time lines, as opposed to spatial domains distributed more randomly. In literature, there are some spatial models with discrete temporal assumed, most of them are either based on spectrum for discrete-time signals or stochastic differential equation (see, e.g. Ma (2003), Demel and Du (2015)). There is no closed form for the covariance structure. In this work, we will give the explicit expression for multivariate covariance matrix with discrete temporal domain. The discrete temporal margins focused are some celebrated autoregressive and moving average (ARMA) models, which are well-established and have been widely used for time series modeling. With a wealth of knowledge in the data exploration of this ARMA processes, constructing a model that takes advantage of this as well as well-studied spatial process is expected to enhance analysis efficiently.

In addition, in modern scientific study, such as geosciences, environmental study, economics, it is common to have a large number of variables observed simultaneously. They

are often coherently related and it is believed that borrowing information from secondary variables can improve the predication of primary variable, which shows scarceness in some location. For simplicity, these spatial variables are often modelled separately, which means cross correlations among variables are ignored. The critical step in this chapter is to construct a multivariate spatial covariance structure which will not only capture the covariance structure within each variable but also cross covariances between each variables to enhance the prediction accuracy, so called co-kriging in spatial statistics. There have been effort made towards multivariate spatial modeling, however, they are either in spatial only, continuous time or through Bayesian approach. Classical work includes Gneiting et al. (2010), Sain and Cressie (2007). Recently Zhu et al. (2020) studies the application of multivariate spatial autoregressive model on social network data. Multivariate Poisson lognormal spatial model shows the improvement of model fits in predicting crash frequency for multi-vehicle collision types Hosseinpour et al. (2018). Somayasa et al. (2021) established universal kriging formula for multivariate spatial data observed over a second order stationary random field. Krupskii and Genton (2019) proposed a new copula model for replicated multivariate spatial data. Song et al. (2006) considered several Bayesian multivariate conditional autoregressive spatial models for estimating the crash rates from different kinds of crashes. Apanasovich et al. (2012) introduced a valid parametric family of cross-covariance functions for multivariate spatial random fields where each component has a covariance function from a well-celebrated Matérn class. We will try to make the marginal spatial modeling part flexible to incorporate the interpretability of parameters of these existing spatial models in the overall space-time models to ease model identification.

As we know, in global scale, large amount of data are always recorded in spherical coordinates. The Euclidean distances and covariances can be easily distorted on the sphere especially for large distances, hence the models that are dedicated to fit data on sphere can critical in geophysical and atmospheric sciences. Recently, several spherical spatial models have been developed. For instance, Gneiting (2013) reviews and develops characterizations and constructions of isotropic positive definite functions on spheres, and they applied them to provide rich parametric classes of such functions. Ma (2012) present the characterization

of the covariance matrix function of a Gaussian or second-order elliptically contoured vector random field on the sphere which is stationary, isotropic, and mean square continuous. Du et al. (2013a) derive characterization of the continuous and isotropic variogram matrix function on a sphere, in terms of an infinite sum of the products of positive definite matrices and ultraspherical polynomials. Porcu et al. (2016) propose stationary covariance functions for processes that evolve temporally over a sphere, as well as cross-covariance functions for multivariate random fields defined over a sphere. Guella et al. (2018) study the strict positive definiteness of matrix-valued covariance functions associated to multivariate random fields defined over dimensional spheres. Jeong et al. (2017) illustrate the realizations obtained from Gaussian processes with different covariance structures and the use of isotropic and nonstationary covariance models through deformations and geographical indicators for global surface temperature data. Motivated by some techniques on models on spheres, we will also extend our discrete multivariate spatial-temporal model into different spaces, ensuring the validity of the covariance matrix functions on both Euclidean space and spherical space.

We are aiming to build a type of multivariate spatio-temporal model that can combine all the information of discrete time, Euclidean and/or Spherical space locations and other highly correlated variables together. Now suppose we have an p-variate space-time random field $\{Z(\mathbf{s},t) = (Z_1(\mathbf{s},t), \ldots . Z_p(\mathbf{s},t))', \mathbf{s} \in \mathbb{S}^d \text{ or } \mathbb{R}^d, t \in \mathbb{Z}\}$, $\mathbb{S}^d$ is the spherical shell of radius 1 and center at $0_{(d+1)\times 1}$ in $\mathbb{R}^{d+1}$, with its covariance matrix function defined by:

$$\mathbf{C}(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2) = \begin{pmatrix} C_{1,1}(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2) & \cdots & C_{1,p}(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2) \\ \vdots & \ddots & \vdots \\ C_{p,1}(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2) & \cdots & C_{p,p}(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2) \end{pmatrix}$$

where each

$$C_{i,j}(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2) = Cov(Z_i(\mathbf{s}_1,t_1), Z_j(\mathbf{s}_2,t_2)),$$

$(i, j = 1, 2 \ldots p)$. Furthermore, $\text{Cov}(Z_i(\mathbf{s}_1, t), Z_i(\mathbf{s}_2, t))$ is the covariance function of $\{Z_i(\mathbf{s}, t), \mathbf{s} \in \mathbb{S}^d \text{ or } \mathbb{R}^d, t \in \mathbb{Z}\}$, $\text{Cov}(Z_i(\mathbf{s}_1, t), Z_j(\mathbf{s}_2, t)), i \neq j$ is the cross covariance function of $\{Z_i(\mathbf{s}, t), \mathbf{s} \in \mathbb{S}^d \text{ or } \mathbb{R}^d, t \in \mathbb{Z}\}$, and $\{Z_j(\mathbf{s}, t), \mathbf{s} \in \mathbb{S}^d \text{ or } \mathbb{R}^d,\ t \in \mathbb{Z}\}$. Stationary in both space and time if $E(Z(\mathbf{s}; t))$ is constant for all $(\mathbf{s}; t)$ and $C(\mathbf{s}_0, \mathbf{s}_0 + \mathbf{s}; t_0, t_0 + t)$ depends only on space lag $\mathbf{s}$ and time lag t for all $(\mathbf{s}_0; t_0) \in S \times T$. We simply write $C(\mathbf{s}, t)$. Following Ma (2005), the spatial margin of spatio-temporal covariance is defined as $C(\mathbf{s}_1, \mathbf{s}_2, t, t)$ and the temporal margin in a fixed location is given by $C(\mathbf{s}, \mathbf{s}, t_1, t_2)$. In practical applications, analyzing multivariate spatial temporal type of data often starts with marginal exploration, series and multivariate spatial analysis. Time series models can reveal the characteristics of the data through time and multivariate spatial data analysis is a good tool to utilizes information from correlated variables in order to improve the estimation of all variables. Since there are lots of research achievements in both topics, we can absorb those benefits of each method for conducting model selection and fitting our own model.

The paper is organized as follows: In section 2 we proposed several classes of multivariate spatial-temporal matrix functions whose discrete temporal margins are some celebrated autoregressive and moving average (ARMA) models. We then generated necessary and sufficient conditions for these matrix functions to be valid covariance functions which are more convenient to utilize in the real data analysis. In section 3, the model will be extended to include some autoregressive and moving average margins. Finally in section 4, we program and fit proposed spatio-temporal model with moving average type of temporal margin to the cleaned average weekly minimum and maximum temperature in Kansas to demonstrate the application of the proposed model in terms of spatio-temporal prediction compared with some classical spatio-temporal models.

1.2 Moving-average-type Temporal Margin

We start building the foundation of whole picture by exploring the covariance structure for first order of moving average discrete temporal margin. It's obvious that (1.1) is meets the condition of MA(1) model at fixed location. This sturcture doesn't require the stationary

assumption. The difficulty of proving (1.1) is that it is a discrete space-time matrix function varies in different time scale, comparing to simply proofing a matrix to be a valid covariance matrix. Theorem 8 of Du and Ma (2011) provided some clues for the proof of theorem 1(See Appendix A).

Theorem 1. Let $\mathbf{G}_0(\mathbf{s}_1, \mathbf{s}_2)$ and $\mathbf{G}_1(\mathbf{s}_1, \mathbf{s}_2), \mathbf{s}_1, \mathbf{s}_2 \in \mathbb{D}$, $\mathbb{D} \subset \mathbb{R}^d$ or $\mathbb{S}^d, d \geq 1$ be $p \times p$ matrix functions, and let $\mathbf{G}_0(\mathbf{s}_1, \mathbf{s}_2)$ be symmetric, i.e., $\mathbf{G}_0(\mathbf{s}_1, \mathbf{s}_2)' = \mathbf{G}_0(\mathbf{s}_1, \mathbf{s}_2)$. Then the $p \times p$ function

$$\mathbf{C}(\mathbf{s}_1, \mathbf{s}_2; t) = \begin{cases} \mathbf{G}_0(\mathbf{s}_1, \mathbf{s}_2), & t = 0, \\ \mathbf{G}_1(\mathbf{s}_1, \mathbf{s}_2), & t = 1, \\ \mathbf{G}_1(\mathbf{s}_2, \mathbf{s}_1)', & t = -1, \\ \mathbf{0}, & t \neq 0, \pm 1, t \in \mathbb{Z}, \ \mathbf{s}_1, \mathbf{s}_2 \in \mathbb{D}. \end{cases} \tag{1.1}$$

is a covariance matrix function on $\mathbb{D} \times \mathbb{Z}$ if and only if the following two conditions are satisfied:

(i) $\mathbf{G}_0(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{G}_1(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{G}_1(\mathbf{s}_2, \mathbf{s}_1)'$ is a covariance matrix function on $\mathbb{D}$,

(ii) $\mathbf{G}_0(\mathbf{s}_1, \mathbf{s}_2) - \mathbf{G}_1(\mathbf{s}_1, \mathbf{s}_2) - \mathbf{G}_1(\mathbf{s}_2, \mathbf{s}_1)'$ is a covariance matrix function on $\mathbb{D}$.

Based on basic structure we built, now we are good to impose different kinds of spatial covariance margin to expand the reasonable class of model. We merge the most commonly used Matérn type of spatial margin into the model and then provide the full conditions for the parameters. In the theorem 2, we will begin with the simple Matérn covariance structure where all the α in $\mathbf{M}(h|\mathbf{v}, \alpha)$ are the same, i.e., (1.4) holds. By Cramér's theorem, a covariance matrix is valid for $p = 2$ if and only if $f_{11}(t) \geq 0, f_{22}(t) \geq 0, |f_{12}(t)|^2 \leq f_{11}(t) f_{22}(t)$, $t = ||\omega|| \geq 0$, where $f_{11}(t)$, $f_{22}(t)$, $f_{12}(t) = f_{21}(t)$ are Fourier transforms of C_{11}, C_{22}, $C_{12} = C_{21}$, i.e., the four components in the $2 * 2$ matrix. Theorem 3 in Gneiting et al. (2010) can provide us some conditions in different cases for (4) to be a valid covariance matrix, which gives us the hint for proving the following theorem:

Theorem 2. Let $\mathbf{v} = (v_1, v_2, \ldots, v_p)$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, $\boldsymbol{\beta} = (\beta_1, \beta_2)$,be constant vectors. $v_k \geq 0, \alpha_k \geq 0, -1/2 \leq \beta_k \leq 1/2$,and let $v_{ij} = (v_i + v_j)/2$, $\mathbb{D} \subset \mathbb{R}^d$. The sufficient condition for

the $p \times p$ matrix function

$$\mathbf{C}(h;t) = \begin{cases} c\mathbf{M}(h|\mathbf{v},\alpha_1) + (1-c)\mathbf{M}(h|\mathbf{v},\alpha_2), & t = 0, \\ c\mathbf{M}(h|\mathbf{v},\alpha_1)\beta_1 + (1-c)\mathbf{M}(h|\mathbf{v},\alpha_2)\beta_2, & t = \pm 1, h \in \mathbb{D} \\ \mathbf{0}, & otherwise, \end{cases} \tag{1.2}$$

to be a correlation matrix function on $\mathbb{D} \times \mathbb{Z}$ is that the constant c satisfies

$$0 \leq c \leq 1. \tag{1.3}$$

And if $p \leq 2$, (1.3) is also necessary.

Where

$$\mathbf{M}(h|\mathbf{v},\alpha) = ((\rho_{ij} m(h|v_{ij},\alpha))_{1\leq i,j\leq p}, \tag{1.4}$$

$m(h|v_{ij},\alpha) = \frac{2^{1-v_{ij}}}{\Gamma(v_{ij})}(\alpha h)^{v_{ij}} K_{v_{ij}}(\alpha h)$, $i,j = 1,2$, $\rho_{ij} = \frac{\Gamma(v_i+\frac{d}{2})^{\frac{1}{2}}}{\Gamma(v_i)^{\frac{1}{2}}} \frac{\Gamma(v_j+\frac{d}{2})^{\frac{1}{2}}}{\Gamma(v_j)^{\frac{1}{2}}} \frac{\Gamma(v_{ij})}{\Gamma(v_{ij}+\frac{d}{2})}$.

Where $c_v = \pi^{-d/2}\Gamma(v+d/2)/\Gamma(v)$. The following theorem is under these two assumptions. This theorem expands the simple Matérn covariance structure to a more complex one by removing requirement of all α needs to be equal in $\mathbf{M}(h|\mathbf{v},\alpha)$, i.e., (1.7).

Theorem 3. Let $\mathbf{v} = (v_1, v_2, v_{12})$,$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_{12})$, $\boldsymbol{\alpha'} = (\alpha'_1, \alpha'_2, \alpha'_{12})$, $\boldsymbol{\beta} = (\beta_1, \beta_2)$ be constant vectors. $v_k \geq 0, \alpha_k \geq 0, \alpha'_k \geq 0, -1/2 \leq \beta_k \leq 1/2$, $\mathbb{D} \subset \mathbb{R}^d$ or $\mathbb{S}^d$. A sufficient and necessary condition for the $p \times p$ matrix function, $p \leq 2$

$$\mathbf{C}(h;t) = \begin{cases} c\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha},\rho_{12}) + (1-c)\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha}',\rho'_{12}), & t = 0, \\ c\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha},\rho_{12})\beta_1 + (1-c)\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha}',\rho'_{12})\beta_2, & t = \pm 1, h \in \mathbb{D} \\ \mathbf{0}, & otherwise, \end{cases} \tag{1.5}$$

to be a correlation matrix function on $\mathbb{D} \times \mathbb{Z}$ is that the constant c satisfies

$$\inf_{h\geq 0, D(h)>0} \frac{c^2(1\pm 2\beta_1)^2 H(h) + (1-c)^2(1\pm 2\beta_2)^2 \tilde{H}(h)}{(1\pm 2\beta_1)(1\pm 2\beta_2)D(h)} \geq c(c-1). \tag{1.6}$$

Where

$$\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha},\rho_{12})=\begin{bmatrix} m_{11}(h|v_1,\alpha_1) & \rho_{12}m_{12}(h|v_{12},\alpha_{12}) \\ \rho_{12}m_{12}(h|v_{12},\alpha_{12}) & m_{22}(h|v_2,\alpha_2) \end{bmatrix}, \tag{1.7}$$

$m_{ij}(h|v_k,\alpha_k)=\frac{2^{1-v_k}}{\Gamma(v_k)}(\alpha_k h)^{v_k}K_{v_k}(\alpha_k h)$, $i,j=1,2$, $k=1,2,12$.

$$H(h)=\frac{\alpha_1^{2v_1}\alpha_2^{2v_2}c_{v_1}c_{v_2}}{(\alpha_1^2+h^2)^{v_1+d/2}(\alpha_2^2+h^2)^{v_2+d/2}}-\frac{\rho_{12}^2\alpha_{12}^{4v_{12}}c_{v_{12}}^2}{(\alpha_{12}^2+h^2)^{2v_{12}+d}},$$

$\widetilde{H}(h)$ is defined like $H(h)$ with α_i replaced with α_i', $i=1,2,12$ and

$$D(h)=\frac{\alpha_1^{2v_1}\alpha_2'^{2v_2}c_{v_1}c_{v_2}}{(\alpha_1^2+h^2)^{v_1+d/2}(\alpha_2'^2+h^2)^{v_2+d/2}}+\frac{\alpha_1'^{2v_1}\alpha_2^{2v_2}c_{v_1}c_{v_2}}{(\alpha_1'^2+h^2)^{v_1+d/2}(\alpha_2^2+h^2)^{v_2+d/2}}-\frac{2\rho_{12}\rho_{12}'\alpha_{12}^{2v_{12}}\alpha_{12}'^{2v_1 2}c_{v_{12}}^2}{((\alpha_{12}^2+h^2)(\alpha_{12}'^2+h^2))^{v_{12}+d/2}}.$$

From Gneiting(2010), $\mathbf{M}(h|\mathbf{v},\alpha,\rho_{12})$ is a valid covariance matrix if and only if

$$\rho_{12}^2\leq\frac{c_{v1}c_{v2}}{c_{v_{12}}^2}\frac{\alpha_1^{2v_1}\alpha_2^{2v_2}}{\alpha_{12}^{2v_1+2v_2}}\underset{h\geq 0}{inf}\frac{(\alpha_{12}^2+h^2)^{v_1+v_2+d}}{(\alpha_1^2+h^2)^{v_1+d/2}(\alpha_2^2+t^2)^{v_2+d/2}}, \tag{1.8}$$

$$\rho_{12}'^2\leq\frac{c_{v1}c_{v2}}{c_{v_{12}}^2}\frac{\alpha_1'^{2v_1}\alpha_2'^{2v_2}}{\alpha_{12}'^{2v_2+2v_2}}\underset{h\geq 0}{inf}\frac{(\alpha_{12}'^2+h^2)^{v_1+v_2+d}}{(\alpha_1'^2+h^2)^{v_1+d/2}(\alpha_2'^2+t^2)^{v_2+d/2}}. \tag{1.9}$$

We can show $H(h)\geq 0$, $\widetilde{H}(h)\geq 0$, $D(h)\geq 0$, when $D(h)=0$, the inequality holed automatically. The minimum of left hand side of inequality (1.8) can be equal to zero under certain conditions, so the following corollary comes.

Corollary 3.1. The sufficient and necessary condition for (1.7) to be a correlation matrix function can be reduced to $0\leq c\leq 1$ in the following cases:

(a) When $\alpha_{12} \leq min(\alpha_1, \alpha_2)$, $\alpha'_{12} \leq min(\alpha'_1, \alpha'_2)$, $v_{12} = \frac{v_1+v_2}{2}$,

$$\rho_{12}^2 = \frac{c_{v_1}c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha_{12}^2}{\alpha_1\alpha_2})^d, \rho_{12}'^2 = \frac{c_{v_1}c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha_{12}'^2}{\alpha'_1\alpha'_2})^d$$

(b) When $\alpha_{12} \geq max(\alpha_1, \alpha_2)$, $\alpha'_{12} \geq max(\alpha'_1, \alpha'_2)$, $v_{12} = \frac{v_1+v_2}{2}$,

$$\rho_{12}^2 = \frac{c_{v_1}c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha_1}{\alpha_{12}})^{2v_1}(\frac{\alpha_2}{\alpha_{12}})^{2v_2}, \rho_{12}'^2 = \frac{c_{v_1}c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha'_1}{\alpha'_{12}})^{2v_1}(\frac{\alpha'_2}{\alpha'_{12}})^{2v_2}$$

The proof of theorem 3 and Corollary 3.1 is deferred to the Appendix. As we know, taking $v = 1/2$ in Matérn function can yield the exponential type of function, which can be result in the following example:

Example 1. Let $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}'$, ρ_{12}, ρ'_{12} and $\beta_k (k = 1, 2)$ be assumed as in Theorem 3, and take $v_1 = v_2 = v_{12} = \frac{1}{2}$, then the matrix function

$$\mathbf{C}(h;t) = \begin{cases} c\mathbf{E}(h|\boldsymbol{\alpha}, \rho_{12}) + (1-c)\mathbf{E}(h|\boldsymbol{\alpha'}, \rho'_{12}), & t = 0, \\ c\mathbf{E}(h|\boldsymbol{\alpha}, \rho_{12})\beta_1 + (1-c)\mathbf{E}(h|\boldsymbol{\alpha'}, \rho'_{12})\beta_2, & t = \pm 1, h \in \mathbb{D} \\ \mathbf{0}, & otherwise, \end{cases} \tag{1.10}$$

to be a stationary correlation matrix function on $\mathbb{D} \times \mathbb{Z}$ is that the constant c satisfies inequality (1.6). Where

$$\mathbf{E}(h|\boldsymbol{\alpha}, \rho_{12}) = \begin{bmatrix} e_{11}(h|\alpha_1) & \rho_{12}e_{12}(h|\alpha_{12}) \\ \rho_{12}e_{12}(h|\alpha_{12}) & e_{22}(h|\alpha_2) \end{bmatrix}.$$

$e_{ij}(h|\alpha_k) = exp(-\alpha_k h)$, $i, j = 1, 2$, $k = 1, 2, 12$.

1.3 ARMA type Temporal Margin

Last section we only consider spatial temporal covariance structure for MA(1) type of temporal margin. However, it's not enough to catch up the complexity of real world environment in this case, more complex temporal margin need to be taken into consider in the model. Thus, in this section, we will extend our covariance matrix to more general cases: Autoregressive and moving average temporal (ARMA) margin. The following model provides the sufficient and necessary conditions for the valid spatial temporal covariance matrix with (ARMA) type of temporal margins. Again, this theorem works for the uniform α in $\mathbf{M}(h|\mathbf{v},\alpha)$.

Theorem 4. Let $\mathbf{v}=(v_1,v_2,v_{12})$, $\boldsymbol{\beta}=(\beta_1,\beta_2)$, be constant vectors. $v_k\geq 0,\alpha_k\geq 0,-1\leq\beta_k\leq 1$, and let $v_{12}=(v_1+v_2)/2$, $\mathbb{D}\subset\mathbb{R}^d$ or $\mathbb{S}^d$. A sufficient condition for the $p\times p$ matrix function

$$C(h;t)=c\mathbf{M}(h|\mathbf{v},\alpha_1)\beta_1^{|t|}+(1-c)\mathbf{M}(h|\mathbf{v},\alpha_2)\beta_2^{|t|},\ t\in\mathbb{Z},h\in\mathbb{D} \tag{1.11}$$

to be a correlation matrix function on $\mathbb{D}\times\mathbb{Z}$ is that the constant c satisfies:

$$0\leq c\leq 1. \tag{1.12}$$

And if $p\leq 2$, (1.12) is also necessary.

Where

$$\mathbf{M}(h|\mathbf{v},\alpha)=((\rho_{ij}m(h|v_{ij},\alpha))_{1\leq i,j\leq p}, \tag{1.13}$$

$m(h|v_k,\alpha)=\frac{2^{1-v_k}}{\Gamma(v_k)}(\alpha h)^{v_k}K_{v_k}(\alpha h)$, $i,j=1,2$, $k=1,2,12$, $\rho_{12}=\frac{\Gamma(v_1+\frac{d}{2})^{\frac{1}{2}}}{\Gamma(v_1)^{\frac{1}{2}}}\frac{\Gamma(v_2+\frac{d}{2})^{\frac{1}{2}}}{\Gamma(v_2)^{\frac{1}{2}}}\frac{\Gamma(v_{12})}{\Gamma(v_{12}+\frac{d}{2})}$.

Now, we can extend this theorem to various α in $\mathbf{M}(h|\mathbf{v},\alpha)$:

Theorem 5. Let $\mathbf{v}=(v_1,v_2,v_{12})$,$\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_{12})$, $\boldsymbol{\alpha'}=(\alpha_1',\alpha_2',\alpha_{12}')$, $\boldsymbol{\beta}=(\beta_1,\beta_2)$ be constant vectors. $v_k\geq 0,\alpha_k\geq 0,\alpha_k'\geq 0,-1\leq\beta_k\leq 1$, $\mathbb{D}\subset\mathbb{R}^d$ or $\mathbb{S}^d$. A sufficient and necessary condition for $p\times p$ matrix function, $p\leq 2$

$$C(h;t)=c\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha},\rho_{12})\beta_1^{|t|}+(1-c)\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha'},\rho_{12}')\beta_2^{|t|},\ t\in\mathbb{Z},h\in\mathbb{D} \tag{1.14}$$

to be a correlation matrix function on $\mathbb{D} \times \mathbb{Z}$ is that the constant c satisfies

$$\underset{h\geq 0, D(h)>0}{inf} \frac{c^2(\beta_1^*)^2 H(h) + (1-c)^2(\beta_2^*)^2 \tilde{H}(h)}{(\beta_1^*)(\beta_2^*)D(h)} \geq c(c-1). \tag{1.15}$$

Where

$$\mathbf{M}(h|\mathbf{v}, \boldsymbol{\alpha}, \rho_{12}) = \begin{bmatrix} m_{11}(h|v_1, \alpha_1) & \rho_{12} m_{12}(h|v_{12}, \alpha_{12}) \\ \rho_{12} m_{12}(h|v_{12}, \alpha_{12}) & m_{22}(h|v_2, \alpha_2) \end{bmatrix},$$

$m_{ij}(h|v_k, \alpha_k) = \frac{2^{1-v_k}}{\Gamma(v_k)}(\alpha_k h)^{v_k} K_{v_k}(\alpha_k h)$, $\beta_i^* = \frac{1-\beta_i^2}{1+\beta_i^2 - 2\beta_i cos(\omega)}, i, j = 1, 2, k = 1, 2, 12$

$$H(h) = \frac{\alpha_1^{2v_1}\alpha_2^{2v_2} c_{v_1} c_{v_2}}{(\alpha_1^2 + h^2)^{v_1 + d/2}(\alpha_2^2 + h^2)^{v_2 + d/2}} - \frac{\rho_{12}^2 \alpha_{12}^{4v_{12}} c_{v_{12}}^2}{(\alpha_{12}^2 + h^2)^{2v_{12} + d}},$$

$\widetilde{H}(h)$ is defined like $H(h)$, with α_i replaced with α_i', $i = 1, 2, 12$.

$$D(h) = \frac{\alpha_1^{2v_1}\alpha_2'^{2v_2} c_{v_1} c_{v_2}}{(\alpha_1^2 + h^2)^{v_1 + d/2}(\alpha_2'^2 + h^2)^{v_2 + d/2}} + \frac{\alpha_1'^{2v_1}\alpha_2^{2v_2} c_{v_1} c_{v_2}}{(\alpha_1'^2 + h^2)^{v_1 + d/2}(\alpha_2^2 + h^2)^{v_2 + d/2}} - \frac{2\rho_{12}\rho_{12}' \alpha_{12}^{2v_{12}} \alpha_{12}'^{2v_1 2} c_{v_{12}}^2}{((\alpha_{12}^2 + h^2)(\alpha_{12}'^2 + h^2))^{v_{12} + d/2}}.$$

Adding different α_i in the model can provide us more specific spatial parameters to capture the spatial trend preciously. Again, the condition of this theorem can be simplified to several special cases:

Corollary 5.1. The sufficient and necessary condition of (1.14) to be a correlation matrix function can be reduced to $0 \leq c \leq 1$ in the following cases:
(a) When $\alpha_{12} \leq min(\alpha_1, \alpha_2)$, $\alpha_{12}' \leq min(\alpha_1', \alpha_2')$, $v_{12} = \frac{v_1 + v_2}{2}$,

$$\rho_{12}^2 = \frac{c_{v_1} c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha_{12}^2}{\alpha_1 \alpha_2})^d, \rho_{12}'^2 = \frac{c_{v_1} c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha_{12}'^2}{\alpha_1' \alpha_2'})^d$$

(b) When $\alpha_{12} \geq max(\alpha_1, \alpha_2)$, $\alpha'_{12} \geq max(\alpha'_1, \alpha'_2)$, $v_{12} = \frac{v_1+v_2}{2}$,

$$\rho_{12}^2 = \frac{c_{v_1}c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha_1}{\alpha_{12}})^{2v_1}(\frac{\alpha_2}{\alpha_{12}})^{2v_2}, \rho_{12}'^2 = \frac{c_{v_1}c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha'_1}{\alpha'_{12}})^{2v_1}(\frac{\alpha'_2}{\alpha'_{12}})^{2v_2}$$

Proof of this Corollary is similar to Corollary 3.1. The temporal margin in both theorem is $C(0,t) = cI_{2,2}\beta_1^{|t|} + (1-c)I_{2,2}\beta_2^{|t|}, t \in \mathbb{Z}$. This linear combination of correlation matrix includes families of valid space-time correlation function with stationary AR(1), AR(2) and ARMA(2,1) type of temporal margin. α_k and ν_k , $k = 1, 2, 12$ can be viewed as the scaling parameter and smoothness parameter for the spatial component. β_1, β_2 are the parameters for time series component and c is balancing parameter in between space and time interaction.

When applying the proposed parametric models, we can use time series techniques to fit time series for individual location developing ARMA order and starting values for β_1, β_2, and c, so that the final parameter estimation can be achieved by maximum likelihood estimation, or Cressie (2015) weighted least square estimation (see Eq. (22) of Gneiting (2002) and least square estimation. For the spatial aspect we can use procedures in spatial statistics to find starting values for α_i and ρ, ρ'. For example, we can set the starting value as the fitted parameters in the marginal spatial correlation and cross correlation at different time lag. We can also take advantage of employing more commonly used time series knowledge, such as ACF, PAC as well as likelihood-based criteria like AIC and BIC to determine the temporal patterns and orders, since the temporal margin here can be first analyzed independently. This step can provides an initial idea of what the marginal time series looks like for model selection. However, the choice of appropriate models will eventually be justified by the final space-time fitting criteria, which are often not quite sensitive to very mild difference of marginal time series choices. The simplicity is also of concern in final model selection. Hence the proposed models along with this stepwise estimation procedure is relatively convenient for it's convenience of splitting a complex problem into two simple steps. Our proposed model presented can serve as an attempt in seeking of more straightforward approach to studying multivariate spatial-temporal data where at each location the temporal process can be modeled with some ARMA-type covariance structure. Also, The benefit of following these

steps is that it can capture the whole multivariate MA(1) process by looking at marginally trend. Since the spatial correlation structure is different for each variable at each time lag, it's better for us to fit the trend separately in different time lag in order to deterring the more precise initial value. Then we can combine them together into the complete model and fit the whole process again. Estimation for the data application in next section was done using Cressie (2015) least squares and techniques introduced by Gneiting (2002). Expanding these techniques to the general ARMA(p, q) would require careful expansion of the theorems presented herein and the computation should still be manageable if Cressie weighted least squares method is employed. We will consider more complex temporal margins in our future work.

1.4 Data Example: Kansas Daily Temperature Data

This data set is sourced from National Oceanic and Atmospheric Administration(NOAA) and 105 weather stations across Kansas. We pick up two highly correlated variables in our real data application: The 8030 days of maximum temperature and minimum temperature for 105 aggregated weather stations of the 105 counties in Kansas from January 1, 1990 to December 31, 2011. We pre-process the data by taking weekly average for 8030 days and results in 1144 weeks of average maximum temperature and minimum temperature then use it as our raw dataset. The next step was model selection: We split the dataset into training and testing data by using first 800 weeks in raw data, first fifteen years, of minimum and maximum temperature as the training data and the 801 to 1144 weeks, last seven years, as the testing data. We make data de-trending by computing mean of minimum and maximum temperature for each week in training data and remove this weekly mean among all 1144 weeks. The ACF and CCF for de-trended minimum and maximum temperature for these 105 counties in first 800 weeks has been computed. Based on the time series plots of ACF and CCF, we can see an approximately moving average one type of trend for both minimum and maximum temperature, showing the reasonable use of space-time model with MA(1) temporal margin. The figure 1.1 and 1.2 shows the ACF of four randomly selected stations

for both maximum and minimum temperature.

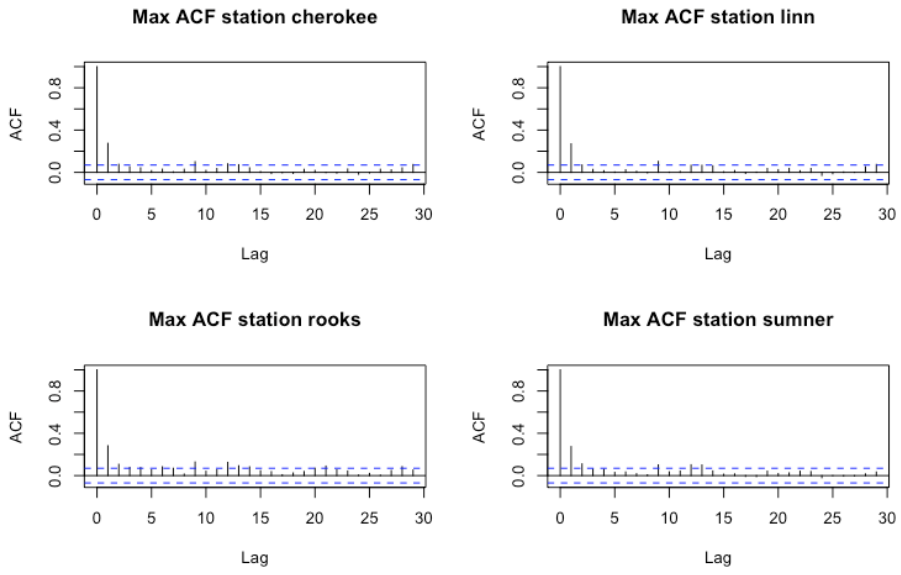

Figure 1.1: *ACFs of Maximum temperature in Kansas counties*

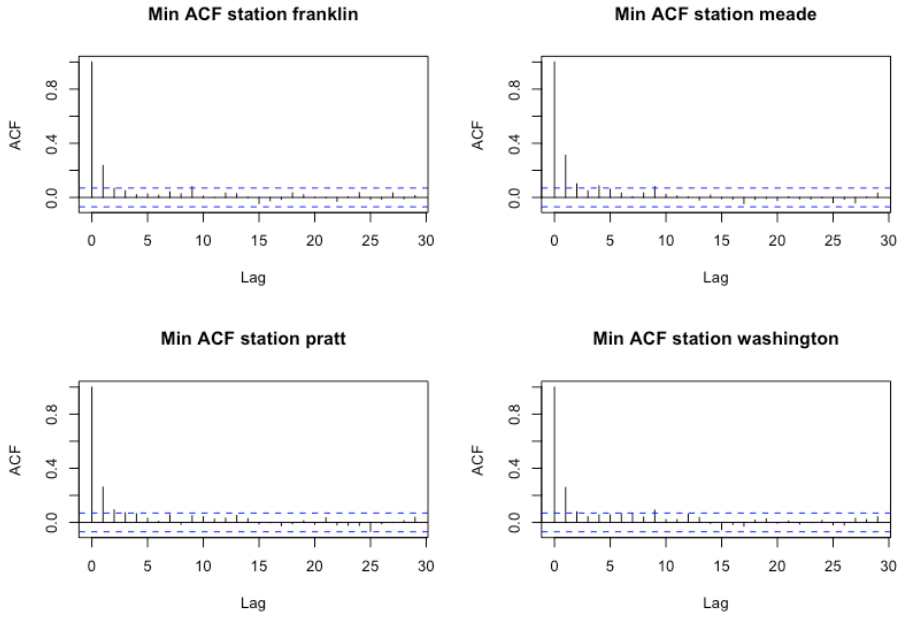

Figure 1.2: *ACFs of Minimum temperature in Kansas counties*

Since the raw data including too many data points is too messy to display the spatial trend in different time lag, we make data binning by using $h = 4$, $\delta = 2$, which means we take

the average of spatial correlation within each 4 kilometres of distance and then we remove all the empty points. Based on the binned data, we use leas square method to fit the empirical spacial correlation for minimum temperature, maximum temperature, cross correlation at time lag 0 to find appropriate initial values for PMM and Separable model below.

Suggested by these exploratory analysis, an MA(1) type of temporal margin is reasonable to choose. After incorporating the nugget effect in theorem 3, the proposed model is below, which is called PMM model:

$$\begin{bmatrix} (1-\eta_1) & 1 \\ 1 & (1-\eta_2) \end{bmatrix} \circ \{c\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha},\rho_{12}) + (1-c)\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha'},\rho'_{12})\} + \begin{bmatrix} \eta_1 \mathbb{1}_{h=0} & 0 \\ 0 & \eta_2 \mathbb{1}_{h=0} \end{bmatrix}, t = 0$$

$$\begin{bmatrix} (1-\eta_1) & 1 \\ 1 & (1-\eta_2) \end{bmatrix} \circ \{c\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha},\rho_{12})\beta_1 + (1-c)\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha'},\rho'_{12})\beta_2\} + \begin{bmatrix} \eta_1 \mathbb{1}_{h=0} & 0 \\ 0 & \eta_2 \mathbb{1}_{h=0} \end{bmatrix} \{c\beta_1 + (1-c)\beta_2\}, t = 1$$

$$0, \text{ otherwise.}$$

Finally, the fitted and estimated parameter values are: $\eta_1 = 0.101395958, \eta_2 = 0.128030019, \alpha_1 = 0.000025000, \alpha'_1 = 0.004088524, \alpha_2 = 0.003852290, \alpha'_2 = 0.000025000, \alpha_{12} = 0.002867771, \alpha'_{12} = 0.000100000,\ c = 0.525378193, \beta_1 = 0.249596718 \beta_2 = 0.259122085, \rho_{12} = 0.696423223, \rho'_{12} = 0.652322061$ and all of the v_{ij} are equal to 2.5.

Next we use the traditional MA(1) space time separable model for comparison,

$$\mathbf{C}(h;t) = \begin{cases} \mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha}), & t = 0, \\ \mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha})\beta, & t = \pm 1, h \in \mathbb{D} \\ \mathbf{0}, & otherwise, \end{cases} \tag{1.16}$$

$$\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha}) = \begin{bmatrix} m_{11}(h|v_1,\alpha_1) & \rho_{12}m_{12}(h|v_{12},\alpha_{12}) \\ \rho_{12}m_{12}(h|v_{12},\alpha_{12}) & m_{22}(h|v_2,\alpha_2) \end{bmatrix},$$

Where $m_{ij}(h|v_{ij},\alpha) = \frac{2^{1-v_{ij}}}{\Gamma(v_{ij})}(\alpha h)^{v_{ij}} K_{v_{ij}}(\alpha h)$, $i = 1,2; j = 1,2$.

After incorporating the nugget effect, the model comes to:

$$\begin{bmatrix} (1-\eta_1) & 1 \\ 1 & (1-\eta_2) \end{bmatrix} \circ \{\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha})\} + \begin{bmatrix} \eta_1 \mathbb{1}_{h=0} & 0 \\ 0 & \eta_2 \mathbb{1}_{h=0} \end{bmatrix}, t = 0$$

$$\begin{bmatrix} (1-\eta_1) & 1 \\ 1 & (1-\eta_2) \end{bmatrix} \circ \{c\mathbf{M}(h|\mathbf{v},\boldsymbol{\alpha})\beta\} + \begin{bmatrix} \eta_1 \mathbb{1}_{h=0} & 0 \\ 0 & \eta_2 \mathbb{1}_{h=0} \end{bmatrix}, t = 1$$

0, otherwise.

The fitted parameters upon least square method are: $\eta_1 = 0.135148323, \eta_2 = 0.145652449, \alpha_1 = 0.002092260, \alpha_2 = 0.002293880, \alpha_{12} = 0.001474033, \beta = 0.258625355, \rho_{12} = 0.649892569$.

Also, we compared the performance of Cauchy's separable model in continuous time from

Gneiting et al. (2010) with nugget effect incorporated.

$$\mathbf{C}(h;t) = \{ \begin{bmatrix} (1-\eta_1) & 1 \\ 1 & (1-\eta_2) \end{bmatrix} \circ \mathbf{M}(h|\mathbf{v}, \boldsymbol{\alpha}) + \begin{bmatrix} \eta_1 \mathbb{1}_{h=0} & 0 \\ 0 & \eta_2 \mathbb{1}_{h=0} \end{bmatrix} \} \cdot \{(1 + a|t|^{2\alpha})^{-1}\}$$

Where $t \in \mathbb{R}, h \in \mathbb{D}$.

The figure 1.3 and 1.4 are the fitted PMM, separatable model and Cauchy's separatable model with time lag 0, 1, 2 for maximum, minimum temperature and cross space-time correlation:

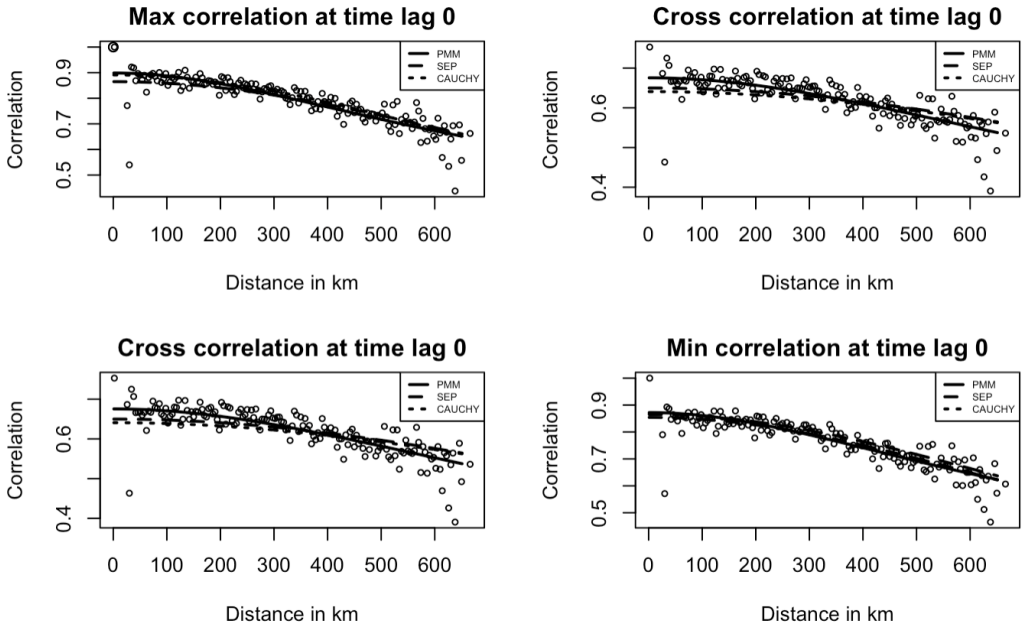

Figure 1.3: *Empirical temperature space-time correlations and fitted models at time lag 0 in Kansas*

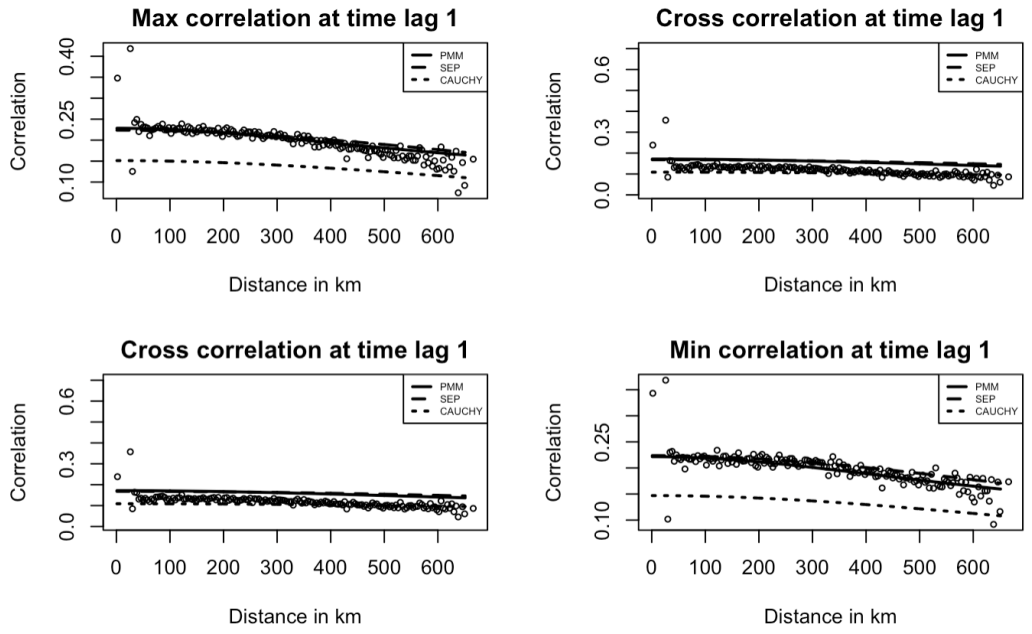

Figure 1.4: *Empirical temperature space-time correlations and fitted models at time lag 1 in Kansas*

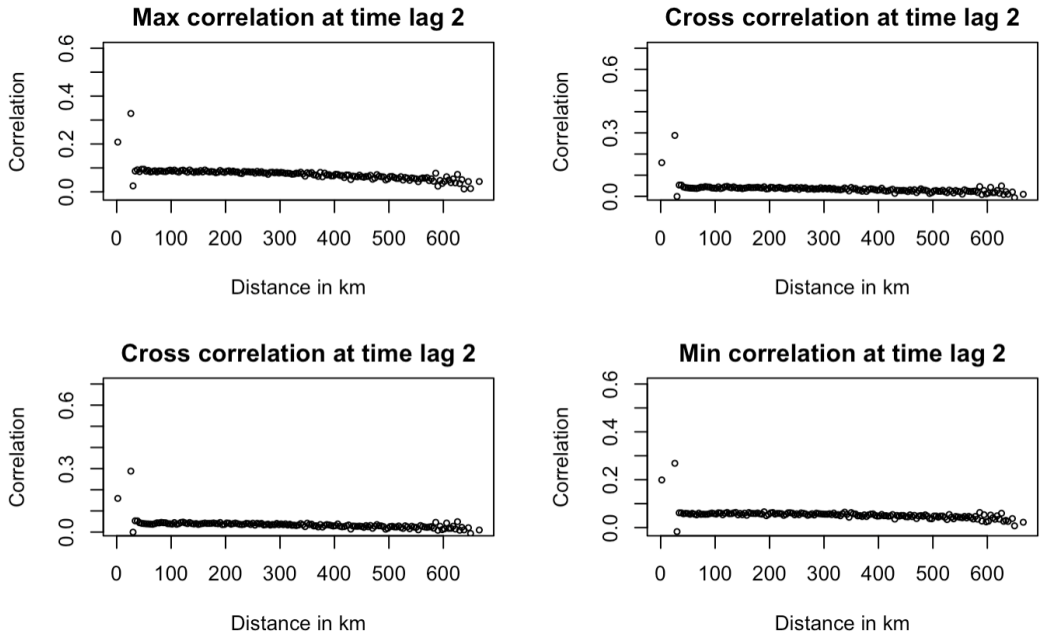

Figure 1.5: *Empirical temperature space-time correlations at time lag 2 in Kansas*

In addition, time series model is another commonly used traditional way to consider. Since the traditional time series prediction in the R package cannot predict data with fixed parameters, we built our own time series code by using the Innovations algorithm in Cressie and Huang (1999) to perform the prediction. We fit the time series model in training data for both maximum and minimum temperature at all 105 stations and made prediction in testing data meaning that for each station, we fit the θ and σ for maximum temperature and minimum temperature, which are the key parameters for moving average one type of time series model.

Finally, the prediction has been performed in testing data and the root mean-square error (RMSE) for each methods have been examined to evaluate the prediction performance of the models. Table 1.1 and 1.2 gives the average RMSE, standard deviation over all counties on maximum temperatures and minimum temperatures.

Measure	PMM	SEPARABLE	CAUCHY	TIME SERIES
AVG. RMSE	3.887092	3.901797	3.938303	3.914282
STD. DEV.	0.3810256	0.3795923	0.4256573	0.3911837
Low Count	100	1	0	4

Table 1.1: *Kansas Maximum Temperature RMSE Statistics*

Measure	PMM	SEPARABLE	CAUCHY	TIME SERIES
AVG. RMSE	2.944538	2.947995	2.982429	2.959063
STD. DEV.	0.5142587	0.507793	0.564688	0.5166872
Low Count	78	19	3	5

Table 1.2: *Kansas Minimum Temperature RMSE Statistics*

The Low Count gives the number of counties have the lowest average RMSE per model. Although all the models have a similar average RMSE around 3.9 on maximum temperature and 2.9 on minimum temperature, the proposed PMM model does have the lowest average

RMSE, showing the best among all the models in prediction. The low count means the number of counties with lowest RMSE per model. The PMM model has the largest number of counties with the lowest RMSE. For maximum temperature, 100 out of 105 counties has the smallest RMSE by using PMM model. About 80% of counties has the smallest RMSE by using PMM model for minimum temperature. Based on this analysis, the proposed model preforms much better than Gneiting's models when the temporal margin of the space-time process can be modeled with a MA(1) structure. It seems that incorporating the highly correlated spatial correlation and discrete temporal margin does improve the predictability of the model.

Appendix A: Proof of Theorem 1

Proof of Theorem 1. Under the assumptions (i) and (ii), we are going to verify that (1.1) is a covariance matrix function on $\mathbb{D} \times \mathbb{Z}$. Clearly, $\{\mathbf{C}(\mathbf{s}_1, \mathbf{s}_2; t)\}' = \mathbf{C}(\mathbf{s}_2, \mathbf{s}_1; -t), \mathbf{s}_1, \mathbf{s}_2 \in \mathbb{D}, t \in \mathbb{Z}$. Thus, it suffices that the inequality

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{a}_i' \mathbf{C}(\mathbf{s}_i, \mathbf{s}_j; i-j) \mathbf{a}_j \geq 0$$

or, equivalently,

$$\sum_{i=1}^{n} \mathbf{a}_i' \mathbf{G}_0(\mathbf{s}_i, \mathbf{s}_i) \mathbf{a}_i + \sum_{i=1}^{n-1} \{\mathbf{a}_i' \mathbf{G}_1'(\mathbf{s}_{i+1}, \mathbf{s}_i) \mathbf{a}_{i+1} + \mathbf{a}_{i+1}' \mathbf{G}_1(\mathbf{s}_{i+1}, \mathbf{s}_i) \mathbf{a}_i\} \geq 0, \tag{1.17}$$

holds for every positive integer n, any $\mathbf{s}_k \in \mathbb{D}$, and any $\mathbf{a}_k \in \mathbb{R}^m$.

Since $\mathbf{G}_0(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{G}_1(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{G}_1'(\mathbf{s}_1, \mathbf{s}_2)$ is a covariance matrix function on $\mathbb{D}$, its transpose $\{\mathbf{G}_0(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{G}_1(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{G}_1'(\mathbf{s}_1, \mathbf{s}_2)\}' = \mathbf{G}_0(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{G}_1'(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{G}_1(\mathbf{s}_1, \mathbf{s}_2)$ equals $\mathbf{G}_0(\mathbf{s}_2, \mathbf{s}_1) + \mathbf{G}_1(\mathbf{s}_2, \mathbf{s}_1) + \mathbf{G}_1'(\mathbf{s}_2, \mathbf{s}_1)$, so that

$$\mathbf{G}_1'(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{G}_1(\mathbf{s}_1, \mathbf{s}_2) = \mathbf{G}_1(\mathbf{s}_2, \mathbf{s}_1) + \mathbf{G}_1'(\mathbf{s}_2, \mathbf{s}_1), \ \mathbf{s}_1, \mathbf{s}_2 \in \mathbb{D}.$$

Notice that the matrix function $\frac{1}{2}(\mathbf{C}(\mathbf{s}_1,\mathbf{s}_2;t)+\mathbf{C}'(\mathbf{s}_1,\mathbf{s}_2;t))$ can be written as

$$
\begin{aligned}
&\frac{1}{2}(\mathbf{C}(\mathbf{s}_1,\mathbf{s}_2;t)+\mathbf{C}'(\mathbf{s}_1,\mathbf{s}_2;t)) \\
= \;& \frac{\mathbf{G}_0(\mathbf{s}_1,\mathbf{s}_2)+\mathbf{G}_1(\mathbf{s}_1,\mathbf{s}_2)+\mathbf{G}_1'(\mathbf{s}_1,\mathbf{s}_2)}{2}\cdot\begin{cases}\mathbf{I}, & t=0,\\ \frac{1}{2}\mathbf{I}, & t=1,\\ \frac{1}{2}\mathbf{I}, & t=-1,\\ \mathbf{0}, & t=\pm 2,\pm 3,\ldots,\ \mathbf{s}_1,\mathbf{s}_2\in\mathbb{D},\end{cases} \\
&+\frac{\mathbf{G}_0(\mathbf{s}_1,\mathbf{s}_2)-\mathbf{G}_1(\mathbf{s}_1,\mathbf{s}_2)-\mathbf{G}_1'(\mathbf{s}_1,\mathbf{s}_2)}{2}\cdot\begin{cases}\mathbf{I}, & t=0,\\ -\frac{1}{2}\mathbf{I}, & t=1,\\ -\frac{1}{2}\mathbf{I}, & t=-1,\\ \mathbf{0}, & t=\pm 2,\pm 3,\ldots,\ \mathbf{s}_1,\mathbf{s}_2\in\mathbb{D}.\end{cases}
\end{aligned}
$$

This is a sum of two separable covariance matrix functions, and is thus a covariance matrix function on $\mathbb{D}\times\mathbb{Z}$. By Theorem 8 of Ma (2011), we obtain

$$
\begin{aligned}
0 \;\le\; & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{a}_i'\{\mathbf{C}(\mathbf{s}_i,\mathbf{s}_j;i-j)+\mathbf{C}'(\mathbf{s}_i,\mathbf{s}_j;i-j)\}\mathbf{a}_j \\
= \;& \sum_{i=1}^{n}\mathbf{a}_i'\mathbf{G}_0(\mathbf{s}_i,\mathbf{s}_i)\mathbf{a}_i+\frac{1}{2}\sum_{i=1}^{n-1}\mathbf{a}_i'\{\mathbf{G}_1'(\mathbf{s}_{i+1},\mathbf{s}_i)+\mathbf{G}_1(\mathbf{s}_{i+1},\mathbf{s}_i)\}\mathbf{a}_{i+1} \\
& +\frac{1}{2}\sum_{i=1}^{n-1}\mathbf{a}_{i+1}'\{\mathbf{G}_1(\mathbf{s}_{i+1},\mathbf{s}_i)+\mathbf{G}_1'(\mathbf{s}_{i+1},\mathbf{s}_i)\}\mathbf{a}_i \\
= \;& \sum_{i=1}^{n}\mathbf{a}_i'\mathbf{G}_0(\mathbf{s}_i,\mathbf{s}_i)\mathbf{a}_i+\sum_{i=1}^{n-1}\{\mathbf{a}_i'\mathbf{G}_1'(\mathbf{s}_{i+1},\mathbf{s}_i)\mathbf{a}_{i+1}+\mathbf{a}_{i+1}'\mathbf{G}_1(\mathbf{s}_{i+1},\mathbf{s}_i)\mathbf{a}_i\},
\end{aligned}
$$

where the last equality follows from $\mathbf{a}_i'\mathbf{G}_1(\mathbf{s}_{i+1},\mathbf{s}_i)\mathbf{a}_{i+1}=\mathbf{a}_{i+1}'\mathbf{G}_1'(\mathbf{s}_{i+1},\mathbf{s}_i)\}\mathbf{a}_i$. Thus, inequality (1.17) is derived. Conversely, suppose that (1.16) is a covariance matrix function on $\mathbb{D}\times\mathbb{Z}$. Then for arbitrary n locations and l integer time points at each location, we formulate nm pairs $\mathbf{s}_i$ and $t_j=j$, choose the corresponding vectors as the products $\mathbf{a}_i b_j$, $i=1,\ldots,n, j=$

$1, \ldots, l$, and obtain

$$\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j=1}^{l}\sum_{j'=1}^{l} b_j b_{j'} \mathbf{a}_i' \mathbf{C}(\mathbf{s}_i, \mathbf{s}_{i'}; j - j') \mathbf{a}_{i'} \geq 0,$$

or

$$\sum_{i=1}^{n}\sum_{i'=1}^{n} \mathbf{a}_i' \left(\sum_{j=1}^{l} b_j^2 \mathbf{G}_0(\mathbf{s}_i, \mathbf{s}_{i'}) + \sum_{j=1}^{l-1} b_j b_{j+1} (\mathbf{G}_1(\mathbf{s}_i, \mathbf{s}_{i'}) + \mathbf{G}_1'(\mathbf{s}_{i'}, \mathbf{s}_i)) \right) \mathbf{a}_{i'} \geq 0. \tag{1.18}$$

In particular, in (1.18) taking $b_j = 1, j = 1, \ldots, l$, and both sides dividing by l yields

$$\sum_{i=1}^{n}\sum_{i'=1}^{n} \mathbf{a}_i' \left(\mathbf{G}_0(\mathbf{s}_i, \mathbf{s}_{i'}) + \frac{l-1}{l} (\mathbf{G}_1(\mathbf{s}_i, \mathbf{s}_{i'}) + \mathbf{G}_1'(\mathbf{s}_{i'}, \mathbf{s}_i)) \right) \mathbf{a}_{i'} \geq 0.$$

Letting $l \to \infty$ gives

$$\sum_{i=1}^{n}\sum_{i'=1}^{n} \mathbf{a}_i' (\mathbf{G}_0(\mathbf{s}_i, \mathbf{s}_{i'}) + \mathbf{G}_1(\mathbf{s}_i, \mathbf{s}_{i'}) + \mathbf{G}_1'(\mathbf{s}_{i'}, \mathbf{s}_i)) \mathbf{a}_{i'} \geq 0.$$

It implies that $\mathbf{G}_0(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{G}_1(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{G}_1'(\mathbf{s}_2, \mathbf{s}_1)$ is a covariance matrix function on $\mathbb{S}$, by Theorem 8 of Ma (2011). Thus, condition (i) is confirmed.

Similarly, in order to confirm condition (ii), in (1.18) we take $b_j = (-1)^j, j = 1, \ldots, l$, divide both sides by l, and letting $l \to \infty$ gives

$$\sum_{i=1}^{n}\sum_{i'=1}^{n} \mathbf{a}_i' \mathbf{G}_0(\mathbf{s}_i, \mathbf{s}_{i'}) - \mathbf{G}_1(\mathbf{s}_i, \mathbf{s}_{i'}) - \mathbf{G}_1'(\mathbf{s}_{i'}, \mathbf{s}_i)) \mathbf{a}_{i'} \geq 0.$$

It implies that $\mathbf{G}_0(\mathbf{s}_1, \mathbf{s}_2) - \mathbf{G}_1(\mathbf{s}_1, \mathbf{s}_2) - \mathbf{G}_1'(\mathbf{s}_2, \mathbf{s}_1)$ is a covariance matrix function on $\mathbb{S}$, by Theorem 8 of Ma (2011).

Appendix B: Proof of Theorem 2

Proof of Theorem 2. Following from Theorem 1, it is equivalent to show that the inequality (1.3) is necessary and sufficient condition for $\mathbf{G}_0(h) \pm \mathbf{G}_1(h) \pm \mathbf{G}_1(h)'$ to be valid covariance

matrix function on $\mathbb{D}$ with $h = ||\mathbf{s}_i - \mathbf{s}_j||$. Under the scenario of this theorem,

$$\mathbf{G}_0(h) \pm \mathbf{G}_1(h) \pm \mathbf{G}_1(h)' = c\mathbf{M}(h|\mathbf{v}, \alpha_1)(1 \pm 2\beta_1) + (1-c)\mathbf{M}(h|\mathbf{v}, \alpha_2)(1 \pm 2\beta_2), h \in \mathbb{D}. \quad (1.19)$$

For sufficiency, from the using the Theorem 2 in Ma (2012) , the sum of two positive defined matrix is also a positive definable matrix. Since $0 \leq c \leq 1$, $-\frac{1}{2} \leq \beta_i \leq \frac{1}{2}, i = 1, 2$, (1.19) is a positive definite matrix, thus, (1.1) is valid covariance matrix.

For necessity, consider the spectral density of Matérn class of function (see Eq.(32) of stein(1999), the Fourier transforms of $\mathbf{G}_0(h) + \mathbf{G}_1(h) + \mathbf{G}_1(h)'$ and $\mathbf{G}_0(h) - \mathbf{G}_1(h) - \mathbf{G}_1(h)'$ are given by

$$\mathbf{F}(h) = \begin{bmatrix} c_{\nu_1} f_{11}(h) & c_{\nu_{12}} \rho_{12} f_{12}(h) \\ c_{\nu_{12}} \rho_{12} f_{21}(h) & c_{\nu_2} f_{22}(h) \end{bmatrix}; \mathbf{G}(h) = \begin{bmatrix} c_{\nu_1} g_{11}(h) & c_{\nu_{12}} \rho_{12} g_{12}(h) \\ c_{\nu_{12}} \rho_{12} g_{21}(h) & c_{\nu_2} g_{22}(h) \end{bmatrix}$$

Where $c_\nu = \pi^{-d/2}\Gamma(\nu + d/2)/\Gamma(\nu)$

$$f_{ij}(h) = c(\alpha_1)^{v_i+v_j}(h^2 + \alpha_1^2)^{-\frac{v_i+v_j}{2}-d/2}(1 + 2\beta_1) + (1-c)\alpha_2^{v_i+v_j}(h^2 + \alpha_2^2)^{-\frac{v_i+v_j}{2}-d/2}(1 + 2\beta_2)$$

and

$$g_{ij}(h) = c(\alpha_1)^{v_i+v_j}(h^2 + \alpha_1^2)^{-\frac{v_i+v_j}{2}-d/2}(1 - 2\beta_1) + (1-c)\alpha_2^{v_i+v_j}(h^2 + \alpha_2^2)^{-\frac{v_i+v_j}{2}-d/2}(1 - 2\beta_2)$$

$$\mathbf{s} \in \mathbb{D}$$

respectively. Hence it is reduced to show that inequality (1.3) is necessary and sufficient to $\mathbf{F}(h)$ and $\mathbf{G}(h)$ to be positive definite, which means $f_{11}(h) \geq 0$, $f_{22}(h) \geq 0$, $g_{11}(h) \geq 0$, $g_{22}(h) \geq 0$ and

$$c_{v_1} c_{v_2} f_{11}(h) f_{22}(h) - c_{v_{12}}^2 \rho_{12}^2 f_{12}(h) f_{21}(h) \geq 0 \quad (1.20)$$

$$c_{v_1} c_{v_2} g_{11}(h) g_{22}(h) - c_{v_{12}}^2 \rho_{12}^2 g_{12}(h) g_{21}(h) \geq 0 \quad (1.21)$$

by Cramér's Theorem. From Theorem 2 in Demel and Juan (2012), we already know that $f_{11}(h) \geq 0$ and, $g_{11}(h) \geq 0$ if and only if

$$\{1 - \frac{\alpha_2^d(1-2\beta_1)}{\alpha_1^d(1-2\beta_2)}\}^{-1} \leq c \leq \{1 - \frac{\alpha_2^{2v_1}(1+2\beta_1)}{\alpha_1^{2v_1}(1+2\beta_2)}\}^{-1}. \tag{1.22}$$

Also, $f_{22}(h) \geq 0$ and, $g_{22}(h) \geq 0$ if and only if

$$\{1 - \frac{\alpha_2^d(1-2\beta_1)}{\alpha_1^d(1-2\beta_2)}\}^{-1} \leq c \leq \{1 - \frac{\alpha_2^{2v_2}(1+2\beta_1)}{\alpha_1^{2v_2}(1+2\beta_2)}\}^{-1}. \tag{1.23}$$

Since $0 \leq v_1 \leq v_2, 0 \leq \alpha_1 \leq \alpha_2, -1/2 \leq \beta_1 \leq \beta_2 \leq 1/2$, $f_{ii} \geq 0$ and $g_{ii} \geq 0$, $i = 1, 2$, entail

$$\{1 - \frac{\alpha_2^d(1-2\beta_1)}{\alpha_1^d(1-2\beta_2)}\}^{-1} \leq c \leq \{1 - \frac{\alpha_2^{2v_2}(1+2\beta_1)}{\alpha_1^{2v_2}(1+2\beta_2)}\}^{-1}. \tag{1.24}$$

To evaluate (1.20) and (1.21), noting that $c_{v_1}c_{v_2} = c_{v_{12}}^2\rho_{12}^2$ we expand the LHS of (1.20) and

(1.21) with this positive factor removed.

$$
\begin{aligned}
&c_{v_1}c_{v_2}(c\alpha_1^{2v_1}(h^2+\alpha_1^2)^{-v_1-d/2}(1\pm 2\beta_1)+(1-c)\alpha_2^{2v_1}(h^2+\alpha_2^2)^{-v_1-d/2}(1\pm 2\beta_2))\\
&\cdot(c\alpha_1^{2v_2}(h^2+\alpha_1^2)^{-v_2-d/2}(1\pm 2\beta_1)+(1-c)\alpha_2^{2v_2}(h^2+\alpha_2^2)^{-v_2-d/2}(1\pm 2\beta_2))\\
&-c_{v_{12}}^2\rho_{12}^2(c\alpha_1^{v_1+v_2}(h^2+\alpha_1^2)^{-\frac{v_1+v_2}{2}-d/2}(1\pm 2\beta_1)+(1-c)\alpha_2^{v_1+v_2}(h^2+\alpha_2^2)^{-\frac{v_1+v_2}{2}-d/2}(1\pm 2\beta_2))^2
\end{aligned}
$$

$$
\begin{aligned}
=\ & c^2\alpha_1^{2v_1+2v_2}(h^2+\alpha_1^2)^{-v_1-d/2}(h^2+\alpha_1^2)^{-v_2-d/2}(1\pm 2\beta_1)^2\\
&+c(1-c)\alpha_1^{2v_2}\alpha_2^{2v_1}(h^2+\alpha_1^2)^{-v_2-d/2}(h^2+\alpha_2^2)^{-v_1-d/2}(1\pm 2\beta_1)(1\pm 2\beta_2)\\
&+c(1-c)\alpha_1^{2v_1}\alpha_2^{2v_2}(h^2+\alpha_1^2)^{-v_1-d/2}(h^2+\alpha_2^2)^{-v_2-d/2}(1\pm 2\beta_1)(1\pm 2\beta_2)\\
&+(1-c)^2\alpha_2^{2v_1+2v_2}(h^2+\alpha_2^2)^{-v_1-d/2}(h^2+\alpha_2^2)^{-v_2-d/2}(1\pm 2\beta_2)^2\\
&-c^2\alpha_1^{2v_1+2v_2}(h^2+\alpha_1^2)^{-(v_1+v_2)-d}(1\pm 2\beta_1)^2\\
&-2c(1-c)\alpha_1^{v_1+v_2}\alpha_2^{v_1+v_2}(h^2+\alpha_1^2)^{-(v_1+v_2)/2-d/2}(h^2+\alpha_2^2)^{-(v_1+v_2)/2-d/2}(1\pm 2\beta_1)(1\pm 2\beta_2)\\
&-(1-c)^2\alpha_2^{2v_1+2v_2}(h^2+\alpha_2^2)^{-(v_1+v_2)-d}(1\pm 2\beta_2)^2
\end{aligned}
$$

$$
\begin{aligned}
=\ & c(1-c)\{\alpha_1^{2v_2}\alpha_2^{2v_1}(h^2+\alpha_1^2)^{-v_2-2/d}(h^2+\alpha_2^2)^{-v_1-2/d}\\
&+\alpha_1^{2v_1}\alpha_2^{2v_2}(h^2+\alpha_1^2)^{-v_1-2/d}(h^2+\alpha_2^2)^{-v_2-2/d}\\
&-2\alpha_1^{v_1+v_2}\alpha_2^{v_1+v_2}(h^2+\alpha_2^2)^{-(v_1+v_2)/2-2/d}(h^2+\alpha_2^2)^{-(v_1+v_2)/2-2/d}\}(1\pm 2\beta_1)(1\pm 2\beta_2)
\end{aligned}
$$

$$
\begin{aligned}
=\ & c(1-c)\{\alpha_1^{v_2-v_1}\alpha_2^{v_1-v_2}(h^2+\alpha_1^2)^{(v_1-v_2)/2}(h^2+\alpha_2^2)^{(v_2-v_1)/2}\\
&+\alpha_1^{v_1-v_2}\alpha_2^{v_2-v_1}(h^2+\alpha_1^2)^{(v_2-v_1)/2}(h^2+\alpha_2^2)^{(v_1-v_2)/2}-2\}(1\pm 2\beta_1)(1\pm 2\beta_2)\\
=\ & c(1-c)\{(\frac{\alpha_1^2(h^2+\alpha_2^2)}{\alpha_2^2(h^2+\alpha_1^2)})^{(v_2-v_1)/2}+(\frac{\alpha_2^2(h^2+\alpha_1^2)}{\alpha_1^2(h^2+\alpha_2^2)})^{(v_2-v_1)/2}-2\}(1\pm 2\beta_1)(1\pm 2\beta_2)
\end{aligned}
$$

For the necessary part, with $(1\pm 2\beta_1)(1\pm 2\beta_2)\geq 0$, letting $h\to+\infty$ in $(\frac{\alpha_1^2(h^2+\alpha_2^2)}{\alpha_2^2(h^2+\alpha_1^2)})^{(v_2-v_1)/2}+(\frac{\alpha_2^2(h^2+\alpha_1^2)}{\alpha_1^2(h^2+\alpha_2^2)})^{(v_2-v_1)/2}-2$ yields $(\frac{\alpha_1^2}{\alpha_2^2})^{(v_2-v_1)/2}+(\frac{\alpha_2^2}{\alpha_1^2})^{(v_2-v_1)/2}-2$, which is greater than zero, so $c\in(0,1)$. Or if can be proved by summation of two positive definite matrix is also a positive

definite matrix.

For the sufficient part, other than using the Theorem 2 in Ma (2012), we can work on $(\frac{\alpha_1^2(h^2+\alpha_2^2)}{\alpha_2^2(h^2+\alpha_1^2)})^{(v_2-v_1)/2} + (\frac{\alpha_2^2(h^2+\alpha_1^2)}{\alpha_1^2(h^2+\alpha_2^2)})^{(v_2-v_1)/2} \geq 2$, by using inequality $a+b \geq \sqrt{2ab}$, so $c \in (0,1)$. Because $c \in (0,1)$ is with in (3), finally, (1) is positive definite if and only if $c \in (0,1)$.

Appendix C: Proof of Theorem 3

Proof of Theorem 3.

If we let

$$\mathbf{G}_0(h) = c\mathbf{M}(h|\mathbf{v}, \boldsymbol{\alpha}, \rho_{12}) + (1-c)\mathbf{M}(h|\mathbf{v}, \boldsymbol{\alpha'}, \rho'_{12})$$

$$\mathbf{G}_1(h) = c\mathbf{M}(h|\mathbf{v}, \boldsymbol{\alpha}, \rho_{12})\beta_1 + (1-c)\mathbf{M}(h|\mathbf{v}, \boldsymbol{\alpha'}, \rho'_{12})\beta_2$$

From Theorem 1, we only need to show $\mathbf{G}_0(h) \pm 2\mathbf{G}_1(h)$ is a valid covariance matrix function. The Fourier transform of $\mathbf{G}_0(h) \pm 2\mathbf{G}_1(h)$ is

$$\begin{pmatrix} f_{11}(h) & f_{12}(h) \\ f_{21}(h) & f_{22}(h) \end{pmatrix}$$

Where

$$f_{11}(h) = c\sigma^2(h^2+\alpha_1^2)^{-v_1-d/2} \cdot \frac{\Gamma(v_1+d/2)\alpha_1^{2v_1}}{\Gamma(v_1)\pi^{d/2}}(1 \pm 2\beta_1) +$$
$$(1-c)\sigma^2(h^2+\alpha_1'^2)^{-v_1-d/2} \cdot \frac{\Gamma(v_1+d/2)\alpha_1'^{2v_1}}{\Gamma(v_1)\pi^{d/2}}(1 \pm 2\beta_2)$$

$$f_{12}(\mathbf{s}) = c\rho_{12}\sigma^2(h^2+\alpha_{12}^2)^{-v_{12}-d/2}\cdot\frac{\Gamma(v_{12}+d/2)\alpha_{12}^{2v_{12}}}{\Gamma((v_1+v_2)/2)\pi^{d/2}}(1\pm 2\beta_1)+$$
$$(1-c)\rho'_{12}\sigma^2(h^2+\alpha_{12}'^2)^{-v_{12}-d/2}\cdot\frac{\Gamma((v_1+v_2)/2+d/2)\alpha_{12}'^{2v_{12}}}{\Gamma((v_1+v_2)/2)\pi^{d/2}}(1\pm 2\beta_2)$$

$$f_{21}(h) = f_{12}(h)$$

$$f_{22}(h) = c\sigma^2(h^2+\alpha_2^2)^{-v_2-d/2}\cdot\frac{\Gamma(v_2+d/2)\alpha_2^{2v_1}}{\Gamma(v_2)\pi^{d/2}}(1\pm 2\beta_1)+$$
$$(1-c)\sigma^2(h^2+\alpha_2'^2)^{-v_2-d/2}\cdot\frac{\Gamma(v_2+d/2)\alpha_2'^{2v_1}}{\Gamma(v_2)\pi^{d/2}}(1\pm 2\beta_2)$$

We need to show $f_{11}(h)\geq 0$, $f_{22}(h)\geq 0$,$f_{11}(h)f_{22}(h)-f_{12}(h)f_{21}(h)\geq 0$, by Cramér's theorem.

Let $h_1 = h^2+\alpha_1^2, h_1' = h^2+\alpha_1'^2, h_2 = h^2+\alpha_2'^2, h_2' = h^2+\alpha_2'^2, c(v) = \pi^{-d/2}\Gamma(v+d/2)/\Gamma(v)$.

Then

$$
\begin{aligned}
&f_{11}(h)f_{22}(h) - f_{12}(h)f_{21}(h) \\
&= c^2 h_1^{-v_1-\frac{d}{2}} c_{v_1} \alpha_1^{2v_1}(1 \pm 2\beta_1) h_2^{-v_2-\frac{d}{2}} c_{v_2} \alpha_2^{2v_2}(1 \pm 2\beta_1) \\
&+ c(1-c) h_1^{-v_1-\frac{d}{2}} c_{v_1} \alpha_1^{2v_1}(1 \pm 2\beta_1) h_2'^{-v_2-\frac{d}{2}} c_{v_2} \alpha_2'^{2v_2}(1 \pm 2\beta_2) \\
&+ c(1-c) h_1'^{-v_1-\frac{d}{2}} c_{v_1} \alpha_1'^{2v_1}(1 \pm 2\beta_2) h_2^{-2v_2-\frac{d}{2}} c_{v_2} \alpha_2^{2v_2}(1 \pm 2\beta_2) \\
&+ (1-c)^2 h_1'^{-v_1-d/2} c_{v_1} \alpha_1'^{2v_1}(1 \pm 2\beta_2) h_2'^{-v_2-d/2} c_{v_2} \alpha_2'^{2v_2}(1 \pm 2\beta_2) \\
&- c^2 \rho_{12}^2 h_{12}^{-2v_{12}-d} c_{v_{12}}^2 \alpha_{12}^{4v_{12}}(1 \pm 2\beta_1)^2 \\
&- 2c(1-c)\rho_{12}\rho_{12}' h_{12}^{-v_{12}-\frac{d}{2}} c_{v_{12}} \alpha_{12}^{2v_{12}}(1 \pm 2\beta_1) h_{12}'^{-v_{12}-\frac{d}{2}} c_{v_{12}} \alpha_{12}'^{2v_{12}}(1 \pm 2\beta_2) \\
&- (1-c)^2 \rho_{12}'^2 h_{12}'^{-2v_{12}-d} c_{v_{12}}^2 \alpha_{12}'^{4v_{12}}(1 \pm 2\beta_2)^2
\end{aligned}
$$

$$
\begin{aligned}
&= c^2(1 \pm 2\beta_1)^2 (h_1^{-v_1-\frac{d}{2}} h_2^{-v_2-\frac{d}{2}} \alpha_1^{2v_1} \alpha_2^{2v_2} c_{v_1} c_{v_2} - \rho_{12}^2 h_{12}^{-2v_{12}-d} \alpha_{12}^{4v_{12}} c_{v_{12}}^2) \\
&+ c(1-c)(1 \pm 2\beta_1)(1 \pm 2\beta_2)(h_1^{-v_1-\frac{d}{2}} h_2'^{-v_2-\frac{d}{2}} \alpha_1^{2v_1} \alpha_2'^{2v_2} c_{v_1} c_{v_2} + h_1'^{-v_1-d/2} h_2^{-v_2-d/2} \alpha_1'^{2v_1} \alpha_2^{2v_2} c_{v_1} c_{v_2} \\
&- 2\rho_{12}\rho_{12}' h_{12}^{-v_{12}-\frac{d}{2}} h_{12}'^{-v_{12}-\frac{d}{2}} \alpha_{12}^{2v_{12}} \alpha_{12}'^{2v_{12}} c_{v_{12}}^2) \\
&+ (1-c)^2(1 \pm 2\beta_2)^2 (h_1'^{-v_1-d/2} h_2'^{-v_2-d/2} \alpha_1'^{2v_1} \alpha_2'^{2v_2} c_{v_1} c_{v_2} - \rho_{12}'^2 h_{12}'^{-2v_{12}-d} \alpha_{12}'^{4v_{12}} c_{v_{12}}^2) \geq 0
\end{aligned}
$$

Since

$$
\begin{aligned}
H(h) &= \frac{\alpha_1^{2v_1} \alpha_2^{2v_2} c_{v_1} c_{v_2}}{(\alpha_1^2 + h^2)^{v_1+d/2} (\alpha_2^2 + h^2)^{v_2+d/2}} - \frac{\rho_{12}^2 \alpha_{12}^{4v_{12}} c_{v_{12}}^2}{(\alpha_{12}^2 + h^2)^{2v_{12}+d}} \\
&= h_1^{-v_1-\frac{d}{2}} h_2^{-v_2-\frac{d}{2}} \alpha_1^{2v_1} \alpha_2^{2v_2} c_{v_1} c_{v_2} - \rho_{12}^2 h_{12}^{-2v_{12}-d} \alpha_{12}^{4v_{12}} c_{v_{12}}^2.
\end{aligned}
$$

$\widetilde{H}(h)$ is with order α replaced by α', which is equal to

$$
h_1'^{-v_1-d/2} h_2'^{-v_2-d/2} \alpha_1'^{2v_1} \alpha_2'^{2v_2} c_{v_1} c_{v_2} - \rho_{12}'^2 h_{12}'^{-2v_{12}-d} \alpha_{12}'^{4v_{12}} c_{v_{12}}^2,
$$

$$D(h) = \frac{\alpha_1^{2v_1}\alpha_2'^{2v_2}c_{v_1}c_{v_2}}{(\alpha_1^2+h^2)^{v_1+d/2}(\alpha_2'^2+h^2)^{v_2+d/2}} + \frac{\alpha_1'^{2v_1}\alpha_2^{2v_2}c_{v_1}c_{v_2}}{(\alpha_1'^2+h^2)^{v_1+d/2}(\alpha_2^2+h^2)^{v_2+d/2}} - \frac{2\rho_{12}\rho_{12}'\alpha_{12}^{2v_{12}}\alpha_{12}'^{2v_1 2}c_{v_{12}}^2}{((\alpha_{12}^2+h^2)(\alpha_{12}'^2+h^2))^{v_{12}+d/2}}.$$
$$= h_1^{-v_1-\frac{d}{2}}h_2'^{-v_2-\frac{d}{2}}\alpha_1^{2v_1}\alpha_2'^{2v_2}c_{v_1}c_{v_2} + h_1'^{-v_1-d/2}h_2^{-v_2-d/2}\alpha_1'^{2v_1}\alpha_2^{2v_2}c_{v_1}c_{v_2}$$
$$- 2\rho_{12}\rho_{12}'h_{12}^{-v_{12}-\frac{d}{2}}h_{12}'^{-v_{12}-\frac{d}{2}}\alpha_{12}^{2v_{12}}\alpha_{12}'^{2v_{12}}c_{v_{12}}^2.$$

This inequality can be written as:

$$c^2(1\pm 2\beta_1)^2H(h) + c(1-c)(1\pm 2\beta_1)(1\pm 2\beta_2)D(h) + (1-c)^2(1\pm 2\beta_2)^2\widetilde{H}(h) \geq 0.$$

Where c satisfies inequality (14) to ensure $f_{11}(h) \geq 0$ and $f_{22}(h) \geq 0$, ρ_{12}, ρ_{12}' satisfies inequality (4) and (5) to ensure $H(h) \geq 0, \widetilde{H}(h) \geq 0$.
To show $D(h) \geq 0$: Since

$$H(h) = \frac{\alpha_1^{2v_1}\alpha_2^{2v_2}c_{v_1}c_{v_2}}{(\alpha_1^2+h^2)^{v_1+d/2}(\alpha_2^2+h^2)^{v_2+d/2}} - \frac{\rho_{12}^2\alpha_{12}^{4v_{12}}c_{v_{12}}^2}{(\alpha_{12}^2+h^2)^{2v_{12}+d}} \geq 0$$

$$\widetilde{H}(h) = \frac{\alpha_1'^{2v_1}\alpha_2'^{2v_2}c_{v_1}c_{v_2}}{(\alpha_1'^2+h^2)^{v_1+d/2}(\alpha_2'^2+h^2)^{v_2+d/2}} - \frac{\rho_{12}^2\alpha_{12}'^{4v_{12}}c_{v_{12}}^2}{(\alpha_{12}'^2+h^2)^{2v_{12}+d}} \geq 0$$

By using $a^2+b^2 \geq 2ab$,

$$D(h) = \frac{\alpha_1^{2v_1}\alpha_2'^{2v_2}c_{v_1}c_{v_2}}{(\alpha_1^2+h^2)^{v_1+d/2}(\alpha_2'^2+h^2)^{v_2+d/2}} + \frac{\alpha_1'^{2v_1}\alpha_2^{2v_2}c_{v_1}c_{v_2}}{(\alpha_1'^2+h^2)^{v_1+d/2}(\alpha_2^2+h^2)^{v_2+d/2}} - \frac{2\rho_{12}\rho_{12}'\alpha_{12}^{2v_{12}}\alpha_{12}'^{2v_{12}}c_{v_{12}}^2}{((\alpha_{12}^2+h^2)(\alpha_{12}'^2+h^2))^{v_{12}+d/2}}$$
$$\geq \frac{2c_{v_1}c_{v_2}\alpha_1^{v_1}\alpha_1'^{v_1}\alpha_2^{v_2}\alpha_2'^{v_2}}{(\alpha_1^2+h^2)^{v_1/2+d/4}(\alpha_1'^2+h^2)^{v_1/2+d/4}(\alpha_2^2+h^2)^{v_2/2+d/4}(\alpha_2'^2+h^2)^{v_2/2+d/4}}$$
$$- \frac{2\rho_{12}\rho_{12}'\alpha_{12}^{2v_{12}}\alpha_{12}'^{2v_1 2}c_{v_{12}}^2}{((\alpha_{12}^2+h^2)(\alpha_{12}'^2+h^2))^{v_{12}+d/2}}$$

Since if $a^2 \geq b^2$, $a'^2 \geq b'^2$, then $aa' \geq bb'$, if a, a', b, b' are all greater than zero. Thus, $D(h) \geq 0$.

$D(h) = 0$ when

$$(\frac{1}{\alpha_1^2+h^2})^{v_1+d/2}(\frac{1}{\alpha_2'^2+h^2})^{v_2+d/2}\alpha_1^{2v_1}\alpha_2^{2v_2} = (\frac{1}{\alpha_1'^2+h^2})^{v_1+d/2}(\frac{1}{\alpha_2^2+h^2})^{v_2+d/2}\alpha_1'^{2v_1}\alpha_2^{2v_2}$$

When $D(h) = 0$, this inequality holds automatically, so the remaining part is to discuss the case when $D(h) > 0$. We can conclude that

$$\inf_{h\geq 0, D(h)>0} \frac{c^2(1\pm 2\beta_1)^2 H(h) + (1-c)^2(1\pm 2\beta_2)^2\tilde{H}(h)}{(1\pm 2\beta_1)(1\pm 2\beta_2)D(h)} \geq c(c-1) \tag{1.25}$$

Appendix D: Proof of Corollary 3.1

Proof of Corollary 3.1.

For the necessary condition: The inequality (1.6) is:

$$\inf_{h\geq 0, D(h)>0} \frac{c^2(1\pm 2\beta_1)^2 H(h) + (1-c)^2(1\pm 2\beta_2)^2\tilde{H}(h)}{(1\pm 2\beta_1)(1\pm 2\beta_2)D(h)} \geq c(c-1)$$

(a) When $\alpha_{12} \leq min(\alpha_1, \alpha_2)$, $\alpha_{12}' \leq min(\alpha_1', \alpha_2')$, $v_{12} = \frac{v_1+v_2}{2}$, equality $\rho_{12}^2 = \frac{c_{v_1}c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha_{12}^2}{\alpha_1\alpha_2})^d$, $\rho_{12}'^2 = \frac{c_{v_1}c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha_{12}'^2}{\alpha_1'\alpha_2'})^d$holds. The minimum zero of numerator will be attached when $h = 0$,

$$\begin{aligned}
H(0) &= \frac{\alpha_1^{2v_1}\alpha_2^{2v_2}c_{v_1}c_{v_2}}{(\alpha_1^2+0^2)^{v_1+d/2}(\alpha_2^2+0^2)^{v_2+d/2}} - \frac{\rho_{12}^2\alpha_{12}^{4v_{12}}c_{v_{12}}^2}{(\alpha_{12}^2+0^2)^{2v_{12}+d}} \\
&= \frac{\alpha_1^{2v_1}\alpha_2^{2v_2}c_{v_1}c_{v_2}}{(\alpha_1^2+0^2)^{v_1+d/2}(\alpha_2^2+0^2)^{v_2+d/2}} - \frac{c_{v_1}c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha_{12}'^2}{\alpha_1'\alpha_2'})^d\frac{\alpha_{12}^{4v_{12}}c_{v_{12}}^2}{(\alpha_{12}^2+0^2)^{2v_{12}+d}} \\
&= \alpha_1^{-d}\alpha_2^{-d}c_{v_1}c_{v_2} - \alpha_1^{-d}\alpha_2^{-d}c_{v_1}c_{v_2} \\
&= 0
\end{aligned}$$

Similarly, $\tilde{H}(0) = 0$. We can conclude LFS of inequality (1.6) is equal to zero, so $0 \leq c \leq 1$.

(b) When $\alpha_{12} \geq max(\alpha_1, \alpha_2)$, $\alpha_{12}' \geq max(\alpha_1', \alpha_2')$,$v_{12} = \frac{v_1+v_2}{2}$, equality $\rho_{12}^2 = \frac{c_{v_1}c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha_1}{\alpha_{12}})^{2v_1}(\frac{\alpha_2}{\alpha_{12}})^{2v_2}$, $\rho_{12}'^2 = \frac{c_{v_1}c_{v_2}}{c_{v_{12}}^2}(\frac{\alpha_1'}{\alpha_{12}'})^{2v_1}(\frac{\alpha_2'}{\alpha_{12}'})^{2v_2}$ holds, the minimum zero of numerator will be attached when

$h \to \infty$,

$$H(\infty) \to \frac{\alpha_1^{2v_1}\alpha_2^{2v_2}c_{v_1}c_{v_2}}{(\alpha_1^2+h^2)^{v_1+d/2}(\alpha_2^2+h^2)^{v_2+d/2}} - \frac{\rho_{12}^2\alpha_{12}^{4v_{12}}c_{v_{12}}^2}{(\alpha_{12}^2+h^2)^{2v_{12}+d}}$$
$$\to 0$$

Similarly, $\tilde{H}(\infty) = 0$. We can conclude LFS of inequality (1.6) is equal to zero, so $0 \leq c \leq 1$.

Appendix E: Proof of Theorem 4

The proof of sufficient is the same as Theorem 2. The necessary part is as follows. The Fourier transform of (1.11) is equal to

$$\mathbf{F}(h) = \begin{bmatrix} c_{\nu_1} f_{11}(h) & c_{\nu_{12}}\rho_{12} f_{12}(h) \\ c_{\nu_{12}}\rho_{12} f_{21}(h) & c_{\nu_2} f_{22}(h) \end{bmatrix}$$

Where $c_\nu = \pi^{-d/2}\Gamma(\nu + d/2)/\Gamma(\nu)$

$$f_{ij}(h) = c(\alpha_1)^{v_i+v_j}(h^2+\alpha_1^2)^{-\frac{v_i+v_j}{2}-d/2}\beta_1^* + (1-c)\alpha_2^{v_i+v_j}(h^2+\alpha_2^2)^{-\frac{v_i+v_j}{2}-d/2}\beta_2^*$$

Hence it is reduced to show that $0 \leq c \leq 1$ is necessary and sufficient to $\mathbf{F}(h)$ to be positive definite, which means $f_{11}(h) \geq 0$, $f_{22}(h) \geq 0$, and

$$c_{v_1}c_{v_2}f_{11}(h)f_{22}(h) - c_{v_{12}}^2\rho_{12}^2 f_{12}(h)f_{21}(h) \geq 0 \tag{1.26}$$

by Cramér's Theorem. From Demel and Juan (2012), we already know that $f_{11}(h) \geq 0$ andif and only if

$$\{1 - \frac{\alpha_2^d(1-\beta_1)(1+\beta_2)}{\alpha_1^d(1+\beta_1)(1-\beta_2)}\}^{-1} \leq c \leq \{1 - \frac{\alpha_2^{2v_1}(1+\beta_1)(1-\beta_2)}{\alpha_1^{2v_1}(1-\beta_1)(1+\beta_2)}\}^{-1}. \tag{1.27}$$

Also, $f_{22}(h) \geq 0$ if and only if

$$\{1 - \frac{\alpha_2^d(1-\beta_1)(1+\beta_2)}{\alpha_1^d(1+\beta_1)(1-\beta_2)}\}^{-1} \leq c \leq \{1 - \frac{\alpha_2^{2v_2}(1+\beta_1)(1-\beta_2)}{\alpha_1^{2v_2}(1-\beta_1)(1+\beta_2)}\}^{-1}. \tag{1.28}$$

To evaluate (1.26), noting that $c_{v_1}c_{v_2} = c_{v_{12}}^2\rho_{12}^2$ we expand the LHS of (1.26) with this positive factor removed.

$$\begin{aligned}
&c_{v_1}c_{v_2}(c\alpha_1^{2v_1}(h^2+\alpha_1^2)^{-v_1-d/2}\beta_1^* + (1-c)\alpha_2^{2v_1}(h^2+\alpha_2^2)^{-v_1-d/2}\beta_2^*) \\
&\cdot (c\alpha_1^{2v_2}(h^2+\alpha_1^2)^{-v_2-d/2}\beta_1^* + (1-c)\alpha_2^{2v_2}(h^2+\alpha_2^2)^{-v_2-d/2}\beta_2^*) \\
&- c_{v_{12}}^2\rho_{12}^2(c\alpha_1^{v_1+v_2}(h^2+\alpha_1^2)^{-\frac{v_1+v_2}{2}-d/2}\beta_1^* + (1-c)\alpha_2^{v_1+v_2}(h^2+\alpha_2^2)^{-\frac{v_1+v_2}{2}-d/2}\beta_2^*)^2
\end{aligned}$$

Let $f_{\alpha v} = (h^2+\alpha^2)^{-v-d/2}\alpha^{2v}$

$$\begin{aligned}
&= c^2 f_{\alpha_1 v_1} f_{\alpha_1 v_2}(\beta_1^*)^2 + c(1-c) f_{\alpha_1 v_2} f_{\alpha_2 v_1}\beta_1^*\beta_2^* + c(1-c) f_{\alpha_1 v_1} f_{\alpha_2 v_2}\beta_1^*\beta_2^* + (1-c)^2 f_{\alpha_2 v_1} f_{\alpha_2 v_2}(\beta_2^*)^2 \\
&\quad - c^2 f_{\alpha_1 v_{12}} f_{\alpha_1 v_{12}}(\beta_1^*)^2 - 2c(1-c) f_{\alpha_1 v_{12}} f_{\alpha_2 v_{12}}\beta_1^*\beta_2^* - (1-c)^2 f_{\alpha_2 v_{12}} f_{\alpha_2 v_{12}}(\beta_2^*)^2
\end{aligned}$$

$$= c(1-c)\{f_{\alpha_1 v_2} f_{\alpha_2 v_1} + f_{\alpha_1 v_1} f_{\alpha_2 v_2} - 2 f_{\alpha_1 v_{12}} f_{\alpha_2 v_{12}}\}\beta_1^*\beta_2^*$$

$$\begin{aligned}
&= c(1-c)\{\alpha_1^{v_2-v_1}\alpha_2^{v_1-v_2}(h^2+\alpha_1^2)^{(v_1-v_2)/2}(h^2+\alpha_2^2)^{(v_2-v_1)/2} \\
&\quad + \alpha_1^{v_1-v_2}\alpha_2^{v_2-v_1}(h^2+\alpha_1^2)^{(v_2-v_1)/2}(h^2+\alpha_2^2)^{(v_1-v_2)/2} - 2\}\beta_1^*\beta_2^* \\
&= c(1-c)\{(\frac{\alpha_1^2(h^2+\alpha_2^2)}{\alpha_2^2(h^2+\alpha_1^2)})^{(v_2-v_1)/2} + (\frac{\alpha_2^2(h^2+\alpha_1^2)}{\alpha_1^2(h^2+\alpha_2^2)})^{(v_2-v_1)/2} - 2\}\beta_1^*\beta_2^*
\end{aligned}$$

For the necessary part, with $\beta_1^*\beta_2^* \geq 0$, $f_{\alpha_1 v_{12}} f_{\alpha_2 v_{12}} \geq 0$, letting $h \to +\infty$ in $(\frac{\alpha_1^2(h^2+\alpha_2^2)}{\alpha_2^2(h^2+\alpha_1^2)})^{(v_2-v_1)/2} + (\frac{\alpha_2^2(h^2+\alpha_1^2)}{\alpha_1^2(h^2+\alpha_2^2)})^{(v_2-v_1)/2} - 2$ yields $(\frac{\alpha_1^2}{\alpha_2^2})^{(v_2-v_1)/2} + (\frac{\alpha_2^2}{\alpha_1^2})^{(v_2-v_1)/2} - 2$, which is greater than zero, so $c \in (0,1)$. Or if can be proved by summation of two positive definite matrix is also a positive definite matrix.

For the sufficient part, other than using the Theorem 2 in Ma(2012), we can work on $(\frac{\alpha_1^2(h^2+\alpha_2^2)}{\alpha_2^2(h^2+\alpha_1^2)})^{(v_2-v_1)/2} + (\frac{\alpha_2^2(h^2+\alpha_1^2)}{\alpha_1^2(h^2+\alpha_2^2)})^{(v_2-v_1)/2} \geq 2$, by using inequality $a+b \geq \sqrt{2ab}$, so $c \in (0,1)$. Because $c \in (0,1)$ is with in (3), finally, (1) is positive definite if and only if $c \in (0,1)$.

Appendix F: Proof of Theorem 5

The Fourier transform of (1.14) is

$$\begin{pmatrix} f_{11}(h) & f_{12}(h) \\ f_{21}(h) & f_{22}(h) \end{pmatrix} = \begin{pmatrix} cf_{\alpha_1 v_1}\beta_1^* + (1-c)f_{\alpha_1' v_1}\beta_2^* & cf_{\alpha_{12} v_{12}}\beta_1^*\rho_{12} + (1-c)f_{\alpha_{12}' v_{12}}\beta_2^*\rho_{12}' \\ cf_{\alpha_{12} v_{12}}\beta_1^*\rho_{12} + (1-c)f_{\alpha_{12}' v_{12}}\beta_2^*\rho_{12}' & cf_{\alpha_2 v_2}\beta_1^* + (1-c)f_{\alpha_2' v_2}\beta_2^* \end{pmatrix}$$

Where

$$f_{\alpha v} = (h^2+\alpha^2)^{-v-d/2} \cdot \frac{\Gamma(v+d/2)\alpha^{2v}}{\Gamma(v)\pi^{d/2}}$$

We need to show $f_{11}(h) \geq 0$, $f_{22}(h) \geq 0$,$f_{11}(h)f_{22}(h) - f_{12}(h)f_{21}(h) \geq 0$, by Cramér's theorem.

Then

$$f_{11}(h)f_{22}(h) - f_{12}(h)f_{21}(h)$$
$$= c^2 f_{\alpha_1 v_1} f_{\alpha_2 v_2} (\beta_1^*)^2 + c(1-c) f_{\alpha_1 v_1} f_{\alpha_2' v_2} \beta_1^* \beta_2^* + c(1-c) f_{\alpha_1' v_1} f_{\alpha_2 v_2} \beta_1^* \beta_2^* + (1-c)^2 f_{\alpha_1' v_1} f_{\alpha_2' v_2} (\beta_2^*)^2$$
$$- c^2 \rho_{12}^2 f_{\alpha_{12} v_{12}}^2 (\beta_1^*)^2 - 2c(1-c)\rho_{12}\rho_{12}' f_{\alpha_{12} v_{12}} f_{\alpha_{12}' v_{12}} \beta_1^* \beta_2^* - (1-c)^2 \rho_{12}'^2 f_{\alpha_{12}' v_{12}}^2 (\beta_2^*)^2$$

$$= c^2 (\beta_1^*)^2 (f_{\alpha_1 v_1} f_{\alpha_2 v_2} - f_{\alpha_{12} v_{12}}^2 \rho_{12}^2)$$
$$+ c(1-c)\beta_1^* \beta_2^* (f_{\alpha_1 v_1} f_{\alpha_2' v_2} + f_{\alpha_1' v_1} f_{\alpha_2 v_2} - 2\rho_{12}\rho_{12}' f_{\alpha_{12} v_{12}} f_{\alpha_{12}' v_{12}})$$
$$+ (1-c)^2 (\beta_2^*)^2 (f_{\alpha_1' v_1} f_{\alpha_2' v_2} - f_{\alpha_{12}' v_{12}}^2 \rho_{12}^2) \geq 0$$

This inequality can be written as:

$$c^2 (\beta_1^*)^2 H(h) + c(1-c)\beta_1^* \beta_2^* D(h) + (1-c)^2 (\beta_2^*)^2 \widetilde{H}(h) \geq 0.$$

Where c satisfies inequality (14) to ensure $f_{11}(h) \geq 0$ and $f_{22}(h) \geq 0$, ρ_{12}, ρ_{12}' satisfies inequality (4) and (5) to ensure $H(h) \geq 0, \widetilde{H}(h) \geq 0$.
Since we already proved $D(h) \geq 0$ and $D(h) = 0$ when

$$(\frac{1}{\alpha_1^2 + h^2})^{v_1 + d/2} (\frac{1}{\alpha_2'^2 + h^2})^{v_2 + d/2} \alpha_1^{2v_1} \alpha_2^{2v_2} = (\frac{1}{\alpha_1'^2 + h^2})^{v_1 + d/2} (\frac{1}{\alpha_2^2 + h^2})^{v_2 + d/2} \alpha_1'^{2v_1} \alpha_2^{2v_2}$$

When $D(h) = 0$, this inequality holds automatically, so the remaining part is to discuss the case when $D(h) > 0$. We can conclude that

$$\inf_{h \geq 0, D(h) > 0} \frac{c^2 (\beta_1^*)^2 H(h) + (1-c)^2 (\beta_2^*)^2 \tilde{H}(h)}{\beta_1^* \beta_2^* D(h)} \geq c(c-1). \tag{1.29}$$

Chapter 2

Simulation of multivariate space-time processes on a sphere

2.1 Introduction

With ever- growing collection of data in global scale in many fields, like geosciences, climatology and oceanology, the statistical modeling and simulations of the underlying stochastic processes on the surface of a sphere has drawn more and more attention. As noted in Jeong and Jun (2015), the distortion using Euclidean distance in place of spherical distance is not negligible when the spatial range of the data is large and the models originally defined on Euclidean space are often not physically justifiable on spheres. To assess the performance of certain inference or model on these types of data, simulation plays an important role in statistical analysis. This chapter will investigate an efficient method for simulation of isotropic multivariate random fields in space and/or time on spheres.

There has been an abundant simulation algorithms for Gaussian random fields in $\mathbb{R}^d$, such as sequential Gaussian, Cholesky decomposition of the covariance matrix, Gibbs sampling, autoregressive and moving average, circulant embedding, and discrete spectral methods (Chiles and Delfiner (2009)). However, to the best of our knowledge, there has limited simulation method available for univariate or multivariate random fields on spheres or sphere

and time. The method proposed in a recent paper by Lantuéjoul et al. (2019) only works for the scalar case.

Covariance models on spheres has gain more and more interest to identify a suitable spatial dependence structure for large scale data. Ma (2012) gives the characterization of the covariance matrix function of a Gaussian or second-order elliptically contoured vector random field on the sphere. Du et al. (2013a) derived characterization of the continuous and isotropic variogram matrix function on a sphere, which is based on the infinite sum of the products of positive definite matrices and ultra-spherical polynomials. Porcu et al. (2016) proposed stationary covariance functions for processes that evolve temporally over a sphere, as well as cross-covariance functions for multivariate random fields defined over a sphere.

This chapter we will mainly focus on proposing simulation methods of multivariate random fields on a d-dimensional compact two-point homogeneous space $\mathbb{M}^d$, and $\mathbb{M}^d \times \mathbb{T}$. The spatial domain $\mathbb{M}^d$ is a compact Riemannian symmetric space of rank one and includes sphere $\mathbb{S}^d$, as a special case equipped with the geodesic distance $\theta(x1, x2)$. Then application will be conducted on $\mathbb{S}^d, \mathbb{S}^d \times \mathbb{T}$ and model comparison will be conducted among different models on sphere via simulated data.

There is few investigation in the literature on isotropic vector random fields on $\mathbb{M}^d$ or time-varying isotropic vector random fields on $\mathbb{M}^d \times \mathbb{T}$, except for the particular case $\mathbb{M}^d = \mathbb{S}^d$. In the scalar case where $m = 1$, Gangolli (1967) extended Paul Lévy's Brownian motion on $\mathbb{S}^d$ to the Schoenberg-Lévy kernel on $\mathbb{M}^d$, and Askey and Bingham (1976) further studied some isotropic Gaussian random fields on $\mathbb{M}^d$. See also Malyarenko (2004), Malyarenko (2012). Studies on scalar and vector isotropic random fields on $\mathbb{S}^d$ may be found in Yadrenko and Balakrishnan (1983), Yaglom (1987), Ma (2012, 2015b, 2016a-c, 2017a), Du et al. (2013b), Du and Ma (2012), Cheng and Xiao (2016), Bingham and Symons (2019), Leonenko et al. (2018), Lang and Schwab (2015), Creasey and Lang (2018) among others. Recently, Malyarenko and Ma (2018) derive a general form of the covariance matrix function and present a series representation for a vector random field that is isotropic and mean square continuous on $\mathbb{M}^d$ and stationary on $\mathbb{T}$, and call for constructing parametric and semiparametric covariance matrix models for theoretical developments and practical applications.

In this chapter, we will use the following notations and definitions. Let's consider the m-variate spatio-temporal random field $\{\mathbf{Z}(\mathbf{x},t), \mathbf{x} \in \mathbb{S}^d, t \in \mathbb{T}\}$, where $\mathbb{S}^d$ is the spherical shell of radius 1 and center at $0_{(d+1)\times 1}$ in $\mathbb{R}^{d+1}$, i.e., $\mathbb{S}^d = \{||\mathbf{x}|| = 1, \mathbf{x} \in \mathbb{R}^{d+1}\}$, where $||\mathbf{x}||$ is the Euclidean norm of vector $\mathbf{x} \in \mathbb{R}^{d+1}$, $\mathbb{T}$ is either in $\mathbb{R}$ or $\mathbb{Z}$ is called a time varying or time dependent random field on sphere. For any two points x_1, x_2 on $\mathbb{S}^d$, their spherical (angular, or geodesic) distance can be denoted by $\theta(x_1, x_2)$, while the Euclidean distance is equal to $||x_1 - x_2||$. Let $x_1'x_2$ be the inner product of x_1, x_2 then the spherical distance can be represented by

$$\theta(\mathbf{x}_1, \mathbf{x}_2) = arccos(\mathbf{x}_1'\mathbf{x}_2), \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^d.$$

With this distance, any isometry between two pairs of points can be extended to an isometry of $\mathbb{S}^d$. A metric space with such a property is called two-point homogeneous. Besides spheres, many projective spaces over different algebras can be concluded into such space called manifold. We denote it by $\mathbb{M}^d$, where d is the topological dimension of the manifold.

When $\{\mathbf{Z}(\mathbf{x},t), \mathbf{x} \in \mathbb{M}^d, t \in \mathbb{T}\}$ has finite second-order moments, its mean function and covariance matrix function are given by $E\mathbf{Z}(\mathbf{x},t)$ and

$$cov(\mathbf{Z}(\mathbf{x}_1;t_1), \mathbf{Z}(\mathbf{x}_2;t_2)) = E\{(\mathbf{Z}(\mathbf{x}_1;t_1) - E\mathbf{Z}(\mathbf{x}_2;t_2))(\mathbf{Z}(\mathbf{x}_2;t_2) - E\mathbf{Z}(\mathbf{x}_2;t_2))'\}, \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{M}^d, t_1, t_2 \in \mathbb{T}.$$

In the following sections, we will introduce an efficient method for simulation of isotropic vector random fields on $\mathbb{M}^d$ or $\mathbb{M}^d \times \mathbb{T}$ to which the key theoretical background is supported by the series representations of such random fields established by Malyarenko and Ma (2018), Ma (2016a, 2016b, 2017), Li et al (2019).

2.2 Simulation of isotropic vector random fields on $\mathbb{M}^d$

Increasing number of models on $\mathbb{M}^d$ requires efficient simulation methods on $\mathbb{M}^d$, as well as $\mathbb{M}^d \times \mathbb{T}$. Just $\mathbb{M}^d$, As its remarkable feature, the series representation (1) in Malyarenko and

Ma (2018) is an imitator of the covariance matrix function (2), and is useful for modelling and simulation. Motivated by this feature, we propose a truncation simulation method, which starts by choosing a relatively large n, and then approximates the random field by the mean square error of this $(n+1)$-term truncation of the series representation (1) is closely related to the decay of the series expansion (2) of the corresponding covariance matrix function $C(\theta(\mathbf{x}_1, \mathbf{x}_2))$. The value of n can be chosen by using cross validation in model prediction. We will derive an upper bound of the mean square error.

As a comparison in the scalar case over $\mathbb{S}^d$, our simulation method is simpler and more efficient and straightforward than that proposed in Lang and Schwab (2015), which is based on the series representation of an isotropic random field $\mathbb{Z}(\mathbf{x})$, $\mathbf{x} \in \mathbb{S}^d$ in terms of spherical harmonic. In this paper, simulation isotropic Gaussian random fields on the sphere are based on the Karhunen–Loéve expansions with respect to the spherical harmonic functions and the angular power spectrum.

This method is motivated by Malyarenko and Ma (2018), a method based on a series representation of random fields on $\mathbb{M}^d$ with d is assumed to be greater or equal than 1. In other words, it can not only works for univariate random fields, but also multivariate one. Suppose that $\{\mathbf{V}_n : n \in \mathbb{N}_0\}$ is a sequence of independent m-variate random vectors with expectation equal to zero and covariance equal to $a_n^2\mathbf{I}_m$. $\mathbf{U}$ is a random vector on $\mathbb{M}^d$ which is independent of $\mathbf{V}_n$. $\{\mathbf{B}_n : n \in \mathbb{N}_0\}$ is a series of $m \times m$ symmetric nonnegative definite matrices and $(\mathbf{B}_n^{\frac{1}{2}})^2 = \mathbf{B}_n$. $\mathbf{I}_m$ is a $m \times m$ identity matrix. $P_n^{(\alpha,\beta)}$ are the Jacobi polynomials of degree n with a pair of parameters (α, β) and

$$a_n = \left(\frac{\Gamma(\beta+1)(2n+\alpha+\beta+1)\Gamma(n+\alpha+\beta+1)}{\Gamma(\alpha+\beta+2)\Gamma(n+\beta+1)}\right)^{\frac{1}{2}}, \ n \in \mathbb{N}_0.$$

If the series $\sum_{n=0}^{\infty} \mathbf{B}_n P_n^{(\alpha,\beta)}(1)$ converges, a centred m-variate isotropic spatial random field can be simulated under several conditions by using

$$\mathbf{Z}(\mathbf{x}) = \sum_{n=0}^{L} \mathbf{B}_n^{\frac{1}{2}} \mathbf{V}_n P_n^{(\alpha,\beta)}(cos\rho(\mathbf{x}, \mathbf{U})), \ \mathbf{x} \in \mathbb{M}^d, L \in \mathbb{L}^+. \tag{2.1}$$

with the covariance matrix function

$$cov(\mathbf{Z}(\mathbf{x}_1), \mathbf{Z}(\mathbf{x}_2)) = \sum_{n=0}^{\infty} \mathbf{B}_n P_n^{(\alpha,\beta)}(cos\rho(\mathbf{x}_1, \mathbf{x}_2)), \ \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{M}^d. \tag{2.2}$$

A key ingredient or building block is a random vector uniformly distributed on $\mathbb{M}^d$. Simulation of uniform distributions on other $\mathbb{M}^d$ may be found in Kent et al. (2018). Another important ingredient in (10) is a closed form of $\mathbf{B}_n$. This can be derived from the series expansion (2) of $C(\theta(\mathbf{x}_1, \mathbf{x}_2))$, or, recursively from the cosine expansion of $C(\theta)$, with the help of the recursive relationships among Jacobi polynomials Szeg (1939), Olver et al. (2010). $\mathbf{B}_n$ as a sequence of $m \times m$ symmetric nonnegative definite matrices, will be the key to the spatial correlation pattern of simulated data. Later, we will expand the ways of constructing $\mathbf{B}_n$ to achieve more complicated data structures. Since $\mathbf{B}_n$, $\mathbf{V}_n$, P_n are now well defined and feasible to realize, this theorem gives a key idea and general form of vector random field on $\mathbb{M}^d$ and $\mathbb{S}^d$ as well.

Unlike other simulation methods that can determine the spatial correlation of data structure before conducted, our series expansion method can easily avoid this defect. The significance of our model lies in the fact that we can only get a slight hint about the spatial correlation and cross correlation pattern between each simulated variables based on the choice of parameters. This will be extremely meaningful in model comparison since if the data structure has already been determined, the model with closet form will preform best so that fair comparison will never achieve. Extensions will be made to more complicated cases in the following section.

2.3 Simulation of Isotropic vector random fields on $\mathbb{M}^d \times \mathbb{T}$ with discrete ARMA margin

As spatial data always comes with temporal variation properties, especially in related studies like astrophysics, geophysics, optics, etc. These studies shows the increasing demand of

simulation focusing on time-varying isotropic vector random fields in $\mathbb{M}^d \times \mathbb{T}$ especially on $\mathbb{S}^d \times \mathbb{T}$. Results on $\mathbb{M}^d \times \mathbb{T}$. Similarly, our simulation algorithm for an m-variate random field on $\mathbb{S}^d \times \mathbb{T}$ is based on an $(l+1)$-term truncation of its series representation, and an upper bound of its mean square error will be derived based on Theorem 5 of Malyarenko and Ma (2018), which expands the previous theorem onto $\mathbb{M}^d \times \mathbb{T}$.

Theorem 5 indicated that a m-variate random field

$$\mathbf{Z}(\mathbf{x},t) = \sum_{n=0}^{\infty} \mathbf{V}_n(t) P_n^{(\alpha,\beta)}(cos\rho(\mathbf{x},\mathbf{U})), \mathbf{x} \in \mathbb{M}^d, t \in \mathbb{T},$$

is isotropic and mean square continuous on $\mathbb{M}^d$, stationary on $\mathbb{T}$ with mean $\mathbf{0}$ and covariance matrix function equal to

$$cov(\rho(\mathbf{x}_1,\mathbf{x}_2);t) = \sum_{n=0}^{\infty} \mathbf{B}_n(t) P_n^{(\alpha,\beta)}(cos\rho(\mathbf{x}_1,\mathbf{x}_2)), \mathbf{x}_1,\mathbf{x}_2 \in \mathbb{M}^d, t \in \mathbb{T}. \tag{2.3}$$

$\{\mathbf{V}_n(t) : n \in \mathbb{N}_0\}$ is a sequence of independent m-variate stationary stochastic process on $\mathbb{T}$ with expectation equal to zero and covariance equal to $a_n^2\mathbf{B}(t_1 - t_2), n \in \mathbb{N}_0$. Also, a random vector $\mathbf{U}$ is uniformly distributed on $\mathbb{M}^d$ and need to be independent with $\{\mathbf{V}_n(t) : n \in \mathbb{N}_0\}$ and $\sum_{n=0}^{\infty} \mathbf{B}_n(0) P_n^{(\alpha,\beta)}(1)$ converges. The key difference here is $\mathbf{B}_n$. Instead of simply defining it as a $m \times m$ covariance matrix, temporal correlations can also be incorporated so that will became a correlation matrix function. Multivariate spacial temporal correlated random field could then be simulated. Variety types of structures can be flexibly taken by $\mathbf{B}_n$, thereby, this method can achieve the realisation of plenty types of spatial temporal correlated data. Another difference is the covariance of $\mathbf{V}_n$, instead of taking $a_n^2\mathbf{I}_m$ as the previous method, it will also be controlled by $\mathbf{B}_n$.

2.4 Proposition

Based on the theorem represented by Malyarenko and Ma (2018), we elaborate the series representation of space-time process on spheres so in the following section, we will provide

theorem at the special case with ARMA margins.

Theorem 6. Assume that $\mathbf{V}_n(t), t \in \mathbb{Z}$ is a m-variate stationary stochastic process with $E\mathbf{V}_n = \mathbf{0}$ and

$$cov(\mathbf{V}_n(t_1), \mathbf{V}_n(t_2)) = a_n^2 \mathbf{B}_n(t_1 - t_2).$$

Where

$$\mathbf{B}_n(t_1 - t_2) = \frac{1}{n} \begin{cases} (\mathbf{\Sigma} + \mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}')^{\circ n}, & t_1 - t_2 = 0, \\ (\mathbf{\Phi}\mathbf{\Sigma})^{\circ n}, & t_1 - t_2 = -1, \\ (\mathbf{\Sigma}\mathbf{\Phi}')^{\circ n}, & t_1 - t_2 = 1, \\ \mathbf{0}, & t_1 - t_2 \pm 2, \pm 3, \ldots, \end{cases} , n \in \mathbb{Z}^+. \tag{2.4}$$

In particular, $\mathbf{B}_0(t_1 - t_2) = \mathbf{B}_1(t_1 - t_2)$ for each fixed $n = 1, 2...$. The n-th Hadamard power of a matrix B is denoted by $B^{\circ n}$. $a_n = (\frac{2n+d-1}{d-1})^{\frac{1}{2}}, n \in \mathbb{N}_0$. Now suppose $\mathbf{U}$ is a $(d+1)$ dimensional random vector uniformly distributed on $\mathbb{S}^d$ $(d \geq 2)$, and $\mathbf{U}$ and $\mathbf{V}_n(t),\ t \in \mathbb{Z}, n \in \mathbb{N}_0$, are independent. $P_n^{\lambda}(x), n \in \mathbb{N}_0$ are the Gegenbauer's polynomials can be defined through the recurrence formula:

$$\begin{cases} P_0^{(\lambda)}(x) \equiv 1, \\ P_1^{(\lambda)}(x) = 2\lambda x \\ P_n^{(\lambda)}(x) = \frac{2(\lambda+n-1)xP_{n-1}^{(\lambda)}(x) - (2\lambda+n-1)P_{n-1}^{(\lambda)}(x)}{n}, \ x \in \mathbb{R}, \ n \geq 2. \end{cases}$$

If $\sum_{n=0}^{\infty} \mathbf{B}_n(0) P_n^{(\frac{d-1}{2})}(1)$ converges. then an m-variate random field

$$\mathbf{Z}(\mathbf{x}; t) = \sum_{n=0}^{L} \mathbf{V}_n(t) P_n^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U}), \ \mathbf{x} \in \mathbb{S}^d, t \in \mathbb{Z}, L \in \mathbb{L}^+ \tag{2.5}$$

is isotropic and mean square continuous on $\mathbb{S}^d$, stationary on $\mathbb{Z}$, and possesses mean and covariance matrix function:

$$\mathbf{C}(\theta;t)=\mathbf{B}_0(t)+\sum_{n=1}^{\infty}\mathbf{B}_n(t)P_n^{(\frac{d-1}{2})}(cos\theta)\ \theta\in[0,\pi],t\in\mathbb{Z},n\in\mathbb{Z}^+. \tag{2.6}$$

The reason of n will goes from 0 to L, instead of infinity is that it is impossible for us to reach infinity in the real world simulation. In this case, we propose a truncated simulation method to utilize a finite, relatively large L such that $\sum_{n=0}^{L}\mathbf{B}_n(0)P_n^{(\frac{d-1}{2})}(1)$ almost converges.

In order to generate $\mathbf{Z}(\mathbf{x},t)$, the key part is to find appropriate $\mathbf{V}_n$ that meets all conditions above. $\mathbf{B}_n$ then will be important to control the covariance structure of $\mathbf{V}_n(t)$. When $\mathbf{B}_n$ equal to (2.9), it is just the covariance matrix function of an m-variate first order moving average time series data with the structure: $\mathbf{Z}(t)=\epsilon(t)+\mathbf{\Phi}\epsilon(t-1)$, where $\{\epsilon(t),t\in\mathbb{Z}\}$ is m-variate white noise with $E\epsilon(t)=\mathbf{0}$ and $var(\epsilon(t))=\mathbf{\Sigma}$. $\frac{1}{n}$ in $\mathbf{B}_n(t_1-t_2)$ structure is to accelerate decay of $\mathbf{B}_n$, so that converge speed will be increased.

This theorem gives a more concrete form to be easily utilize and interpret.

2.5 Simulation

Based on the previous theorem, we simulate m-variate random field on the unit sphere, which allows a fair comparison on the performance of exiting space-time model and model proposed by chapter 1. A bi-variate isotropic random field has been simulated for such comparison. Thus, in this case, $d=2$. Then

$$\mathbf{B}_n(t_1-t_2)=\frac{1}{n}\begin{cases}([\begin{smallmatrix}\sigma_1 & \sigma_{12}\\ \sigma_{21} & \sigma_2\end{smallmatrix}]+[\begin{smallmatrix}\phi_1 & \phi_{12}\\ \phi_{21} & \phi_2\end{smallmatrix}][\begin{smallmatrix}\sigma_1 & \sigma_{12}\\ \sigma_{21} & \sigma_2\end{smallmatrix}][\begin{smallmatrix}\phi_1 & \phi_{21}\\ \phi_{12} & \phi_2\end{smallmatrix}])^{\circ n}, & t_1-t_2=0,\\ ([\begin{smallmatrix}\phi_1 & \phi_{12}\\ \phi_{21} & \phi_2\end{smallmatrix}][\begin{smallmatrix}\sigma_1 & \sigma_{12}\\ \sigma_{21} & \sigma_2\end{smallmatrix}])^{\circ n}, & t_1-t_2=-1,\\ ([\begin{smallmatrix}\sigma_1 & \sigma_{12}\\ \sigma_{21} & \sigma_2\end{smallmatrix}][\begin{smallmatrix}\phi_1 & \phi_{21}\\ \phi_{12} & \phi_2\end{smallmatrix}])^{\circ n}, & t_1-t_2=1,\\ \mathbf{0}, & t_1-t_2=\pm2,\pm3,\dots,\end{cases} \tag{2.7}$$

and the covariance matrix function then can be written as:

$$\mathbf{C}(\theta;t)=\mathbf{B}_0(t)+\sum_{n=1}^{\infty}P_n^{\frac{d-1}{2}}(cos\theta)\frac{1}{n}\begin{cases}([\begin{smallmatrix}\sigma_1 & \sigma_{12}\\ \sigma_{21} & \sigma_2\end{smallmatrix}]+[\begin{smallmatrix}\phi_1 & \phi_{12}\\ \phi_{21} & \phi_2\end{smallmatrix}][\begin{smallmatrix}\sigma_1 & \sigma_{12}\\ \sigma_{21} & \sigma_2\end{smallmatrix}][\begin{smallmatrix}\phi_1 & \phi_{21}\\ \phi_{12} & \phi_2\end{smallmatrix}])^{\circ n}, & t=0,\\ ([\begin{smallmatrix}\phi_1 & \phi_{12}\\ \phi_{21} & \phi_2\end{smallmatrix}][\begin{smallmatrix}\sigma_1 & \sigma_{12}\\ \sigma_{21} & \sigma_2\end{smallmatrix}])^{\circ n}, & t=-1,\\ ([\begin{smallmatrix}\sigma_1 & \sigma_{12}\\ \sigma_{21} & \sigma_2\end{smallmatrix}][\begin{smallmatrix}\phi_1 & \phi_{21}\\ \phi_{12} & \phi_2\end{smallmatrix}])^{\circ n}, & t=1,\\ \mathbf{0}, & t=\pm2,\pm3,\ldots,\end{cases} \tag{2.8}$$

with $\mathbf{B}_0(t)=\mathbf{B}_1(t)$. Σ is more related to the spatial marginal variability while Σ and ϕ works together to affect the cross covariance between each variable. We randomly sample 200 locations on a unit sphere and compute corresponding two correlated variables in each locations, denoted as y_1, y_2 through 200 time lags by using theorem 6. The training data set is the two variables in first 160 time lags among all 200 locations and thus the testing data set is the remaining last 40 time lags of data.

The same as data example in chapter 1, we first calculate and correlation and cross correlation correlation of y_1, y_2 in the training data set with time lag 0 and 1 separately. Then we bin the correlation and cross correlation by using $h=0.01, \delta=0.005$. Since simulation can just consider the ideal cases and avoid several disturbance, the nugget effect doesn't need to be take in to considered. We directly use theorem (3) and Separable model in equation (15) to fit the correlation and cross correlation and perform Kriging. Moreover, time series model also has been used for comparison. The average RMSE has been computed in testing data for each locations through all methods.

As for the choice of L, as we know, the most suitable value range of n is depended on the value of $\mathbf{B}_n(0)$ matrix and $\mathbf{B}_n(0)$ is depended on the choice of σ and ϕ. To achieve converge, we need to make sure all the elements of $\mathbf{B}_0(0)$ less than 1. The choice of L is depended on the value of elements. When at least one elements is very close to 1, L need to be pretty large, say, around 2000 is a good choice. When the elements are close to zero, L can be under 100. In most of cases, L could be between 80 and 200, varying from the value of elements.

From chapter 2 of Reinsel (2003), all eigenvalue of $\mathbf{\Phi}$, that is, all roots of λ from

$det\{\lambda I - \mathbf{\Phi}\} = 0$ need to be less than 1 in absolute value to meet the convertibility condition. $\mathbf{\Sigma}$ must be a positively defined matrix. $\mathbf{\Phi}$ mainly controls the spatial and temporal correlation between two variables. Therefore, the shape of correlation plot through distance with different time lags will be mainly depended on the choose of $\mathbf{\Phi}$

Jeong and Jun (2015) indicated that the Matérn covariance model using the great circle distance, which may not be positive definite on the surface of a sphere, unless the smoothness parameter, ν, is $\nu \in (0, 0.5]$. Thus, when ν in our fitted model is less than 0.5, we can directly apply it on the sphere. But when $\nu > 0.5$, a simple transaction between Euclidean distance and spherical distance can be made to use the model on Euclidean space:

$$||\mathbf{x}_1 - \mathbf{x}_2|| = (2 - 2\mathbf{x}_1'\mathbf{x}_2)^{\frac{1}{2}} = (2 - 2cos\theta(\mathbf{x}_1, \mathbf{x}_2))^{\frac{1}{2}} = 2sin(\frac{\theta(\mathbf{x}_1, \mathbf{x}_2)}{2}), \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}^d.$$

In our simulation, most of the cases separable model and PMM model choose the smooth parameter $\nu = 2.5$, which is more suitable to fit the correlation patten of the simulated data, so we reasonably made the distance transformation.

The following tables gives fitted results in different choose of $n, \mathbf{\Sigma}$ and $\mathbf{\Phi}$.

2.5.1 Fix L ,Phi, change Sigma

Sigma	Variable	Criteria	PMM	SEP	Time Series
$\sigma_1 = 0.6$	y1	RMSE	1.6304	1.7206	1.7210
$\sigma_{12} = 0.1$		LOW COUNT	180	13	7
$\sigma_{21} = 0.1$	y2	RMSE	1.7985	1.8960	1.8873
$\sigma_2 = 0.6$		LOW COUNT	193	7	0
$\sigma_1 = 0.6$	y1	RMSE	1.4374	1.5535	1.5525
$\sigma_{12} = 0.1$		LOW COUNT	194	3	3
$\sigma_{21} = 0.1$	y2	RMSE	0.8568	0.8664	0.8647
$\sigma_2 = 0.2$		LOW COUNT	81	56	63
$\sigma_1 = 0.2$	y1	RMSE	0.8301	0.8365	0.8352
$\sigma_{12} = 0.2$		LOW COUNT	125	18	57
$\sigma_{21} = 0.2$	y2	RMSE	1.8374	1.9291	1.8405
$\sigma_2 = 0.6$		LOW COUNT	87	19	94
$\sigma_1 = 0.5$	y1	RMSE	1.4614	1.5470	1.5400
$\sigma_{12} = 0.3$		LOW COUNT	189	8	3
$\sigma_{21} = 0.3$	y2	RMSE	1.5797	1.6823	1.6657
$\sigma_2 = 0.5$		LOW COUNT	200	0	0
$\sigma_1 = 0.2$	y1	RMSE	0.6908	0.6908	0.6858
$\sigma_{12} = 0.1$		LOW COUNT	55	46	99
$\sigma_{21} = 0.1$	y2	RMSE	0.7718	0.7718	0.7630
$\sigma_2 = 0.2$		LOW COUNT	49	25	126

Table 2.1: *Simulation results for fixed L ,Phi, and changed Sigma*

This table 2.1 clearly shows a whole picture of simulation results for fixed $L, \mathbf{\Phi}$ and different value of $\mathbf{\Sigma}$ matrix with $\phi_1 = \phi_2 = 0.6, \phi_{12} = \phi_{21} = 0.35$, $L = 120$. When we change $\sigma_1, \sigma_2, \sigma_{12}$ or σ_{21} into a relevantly small number, PMM and SEP model will perform worse than time series model. This is easy to understand since when we decrease these elements, their summation and multiplication $\mathbf{B}_n(t)$ will also decrease, so that the spatial-temporal correlation of y_1, y_2 and their cross correlation will be not so obvious. In this case, it's reasonable to understand why PMM and SEP model loose their advantage. The last line of the table clearly confirmed our explanation. As is showed, all the elements in $\mathbf{\Sigma}$ matrix are very small so that the result of PMM model becomes the worst one in the table.

2.5.2 Fix Sigma, Phi, change L

L	Variable	Criteria	PMM	SEP	Time Series
50	y1	RMSE	1.6285	1.6216	1.6116
		LOW COUNT	2	74	124
	y2	RMSE	1.4061	1.3929	1.3770
		LOW COUNT	0	62	138
100	y1	RMSE	1.5744	1.5034	1.4906
		LOW COUNT	43	53	104
	y2	RMSE	1.6250	1.5839	1.5629
		LOW COUNT	16	63	121
150	y1	RMSE	1.3706	1.3932	1.4006
		LOW COUNT	157	15	28
	y2	RMSE	1.4576	1.4823	1.4942
		LOW COUNT	158	32	10
200	y1	RMSE	1.2650	1.2957	1.3135
		LOW COUNT	151	32	17
	y2	RMSE	1.2708	1.3167	1.3306
		LOW COUNT	188	0	12

Table 2.2: *Simulation results for fixed Sigma, Phi, and changed L*

In this table we fix $\phi_1 = \phi_2 = 0.4, \phi_{12} = \phi_{21} = 0.3$, $\sigma_1 = \sigma_2 = 0.4, \sigma_{12} = \sigma_{21} = 0.3$ and change the L. The result is pretty clear based on this table 2.2. The preform of models goes better and better as L increases. When L is relevantly small, spatial-temporal models, especially PMM doesn't works better than other time series model. That's because $\sum_{n=0}^{L} \mathbf{B}_n(0) P_n^{(\frac{d-1}{2})}(1)$ couldn't be converged based on small L so that the spatial temporal

pattern won't be obvious too, which means PMM cannot take advantage of these two information and get ideal results. Also, the performance of other two models can be explained under this reason. Since time series model only based on temporal information, as spatial-temporal structure become more and more clear, it will loose the advantage to beat PMM model.

2.5.3 Fix Sigma, L, change Phi

Phi	Variable	Criteria	PMM	SEP	Time Series
$\phi_1 = 0.7$	y1	RMSE	1.2216	1.2531	1.2495
$\phi_{12} = 0.1$		LOW COUNT	167	0	33
$\phi_{21} = 0.1$	y2	RMSE	1.4739	1.4939	1.5022
$\phi_2 = 0.7$		LOW COUNT	133	0	67
$\phi_1 = 0.6$	y1	RMSE	1.2653	1.3177	1.3182
$\phi_{12} = 0.35$		LOW COUNT	200	0	0
$\phi_{21} = 0.35$	y2	RMSE	1.5150	1.5456	1.5636
$\phi_2 = 0.6$		LOW COUNT	172	14	14
$\phi_1 = 0.4$	y1	RMSE	1.2624	1.3003	1.3061
$\phi_{12} = 0.5$		LOW COUNT	200	0	0
$\phi_{21} = 0.5$	y2	RMSE	1.5006	1.5218	1.5464
$\phi_2 = 0.4$		LOW COUNT	177	23	0
$\phi_1 = 0.2$	y1	RMSE	1.0240	1.0281	1.0361
$\phi_{12} = 0.1$		LOW COUNT	145	42	13
$\phi_{21} = 0.1$	y2	RMSE	1.2280	1.2272	1.2529
$\phi_2 = 0.2$		LOW COUNT	85	0	115

Table 2.3: *Simulation results for fixed Sigma, L, and changed Phi*

In table 2.3, we fix $\sigma_1 = \sigma_2 = 0.5, \sigma_{12} = \sigma_{21} = 0.1$ and $L = 160$, change the value of Φ matrix. This table shows that in most of cases, PMM model performs much better than Separable and time series model. With increasing value of $\mathbf{\Phi}$, the PMM performance will became better and better. Figure 2.1 and 2.2 are the fitted correlation plots for one typical simulation for PMM:

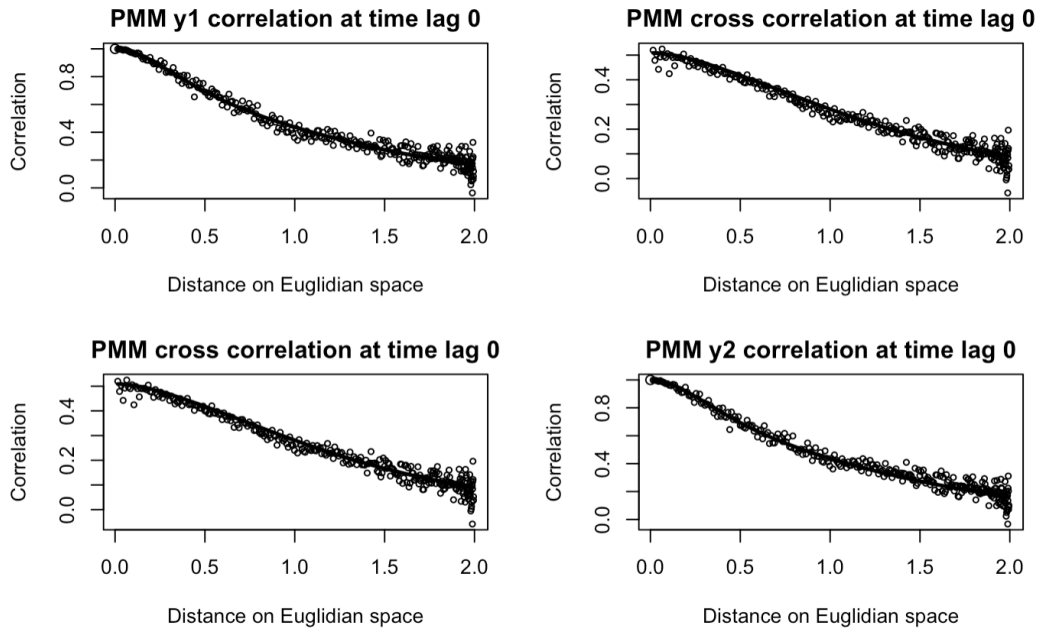

Figure 2.1: *Empirical space-time correlations and fitted models at time lag 0 for simulated data*

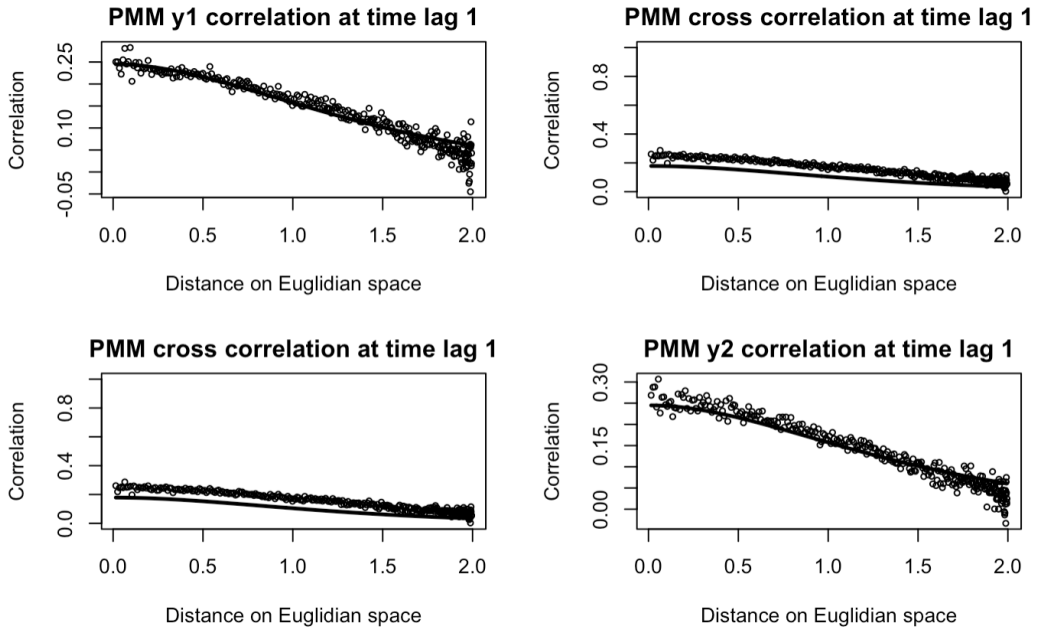

Figure 2.2: *Empirical space-time correlations and fitted models at time lag 1 for simulated data*

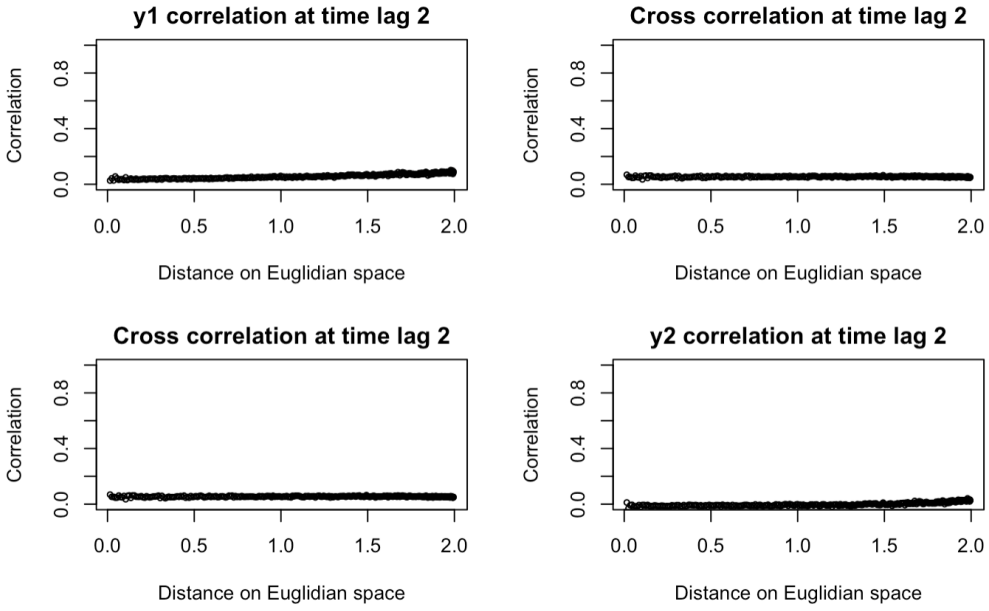

Figure 2.3: *Empirical space-time correlations at time lag 2 for simulated data*

The maximum distance equals to 2 is because we transformed the great arch distance

into the Euclidean distance, so the distance range changed from $(0,\pi)$ to (0,2). These figures show that the model fits the correlation well in all types of correlation and different time lag.

2.6 Appendix: Proof of theorem 1

We use Cauchy's convergence test to proof theorem 1. The key will be using the convergent assumption $\sum_{n=0}^{\infty}\mathbf{B}_n(0)P_n^{(\frac{d-1}{2})}(1)$ to proof the mean square converge of the series at the right hand of (2.5). For $n_1, n_2 \in \mathbb{N}$, we have

$$
\begin{aligned}
& E(\sum_{i=n_1}^{n_1+n_2}\mathbf{V}_i(t)P_i^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U}) - E(\sum_{i=n_1}^{n_1+n_2}\mathbf{V}_i(t)P_i^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U}))(\sum_{j=n_1}^{n_1+n_2}\mathbf{V}_j(t)P_j^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U})- \\
& E(\sum_{j=n_1}^{n_1+n_2}\mathbf{V}_j(t)P_j^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U})))' \\
=& E(\sum_{i=n_1}^{n_1+n_2}\mathbf{V}_i(t)P_i^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U}) - 0)(\sum_{j=n_1}^{n_1+n_2}\mathbf{V}_j(t)P_j^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U}) - 0)' \\
=& E(\sum_{i=n_1}^{n_1+n_2}\sum_{j=n_1}^{n_1+n_2}\mathbf{V}_i(t)\mathbf{V}_j'(t)P_i^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U})P_j^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U}))' \\
=& \sum_{i=n_1}^{n_1+n_2}\sum_{j=n_1}^{n_1+n_2}E(\mathbf{V}_i(t)\mathbf{V}_j'(t))E(P_i^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U})P_j^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U})) \\
=& \sum_{i=n_1}^{n_1+n_2}\sum_{j=n_1}^{n_1+n_2}E(\mathbf{V}_i(t) - 0)(\mathbf{V}_j(t)' - 0)E(P_i^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U})P_j^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U})) \\
=& \sum_{i=n_1}^{n_1+n_2}a_i^2\mathbf{B}_i(0)E(P_i^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U})P_i^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U})) \\
=& \sum_{i=n_1}^{n_1+n_2}\mathbf{B}_i(0)cov(a_iP_i^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U}), a_iP_i^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U})) \\
=& w_d\sum_{i=n_1}^{n_1+n_2}\mathbf{B}_i(0)P_i^{(\frac{d-1}{2})}(1) \\
\to & 0, \ as \ n_1, n_2 \to \infty
\end{aligned}
$$

where the third equality follows from the independent assumption between $\mathbf{U}$ and $\{\mathbf{V}_n(t), t \in \mathbb{Z}\}$, and the fourth one follows from lemma 3 of Malyarenko and Ma (2018). Applying Lemma

3 we obtain the mean and covariance matrix functions of $\{\mathbf{Z}(\mathbf{x};t), \mathbf{x} \in \mathbb{S}^d, \mathbf{x} \in \mathbb{Z}\}$ with

$$E\mathbf{Z}(\mathbf{x};t) = \sum_{n=0}^{\infty} E\mathbf{V}_n(t) E P_n^{(\frac{d-1}{2})}(\mathbf{x}'\mathbf{U}) = \mathbf{0},\ \mathbf{x} \in \mathbb{S}^d, t \in \mathbb{Z},$$

and

$$\begin{aligned}
&cov(\mathbf{Z}(\mathbf{x}_1;t_1), \mathbf{Z}(\mathbf{x}_2;t_2)) \\
=&cov(\sum_{i=0}^{\infty} \mathbf{V}_i(t_1) P_i^{(\frac{d-1}{2})}(\mathbf{x}_1'\mathbf{U}), \sum_{j=0}^{\infty} \mathbf{V}_j(t_2) P_j^{(\frac{d-1}{2})}(\mathbf{x}_2'\mathbf{U})) \\
=&\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} E(\mathbf{V}_i(t_1) P_i^{(\frac{d-1}{2})}(\mathbf{x}_1'\mathbf{U}) \mathbf{V}_j'(t_2) P_j^{(\frac{d-1}{2})}(\mathbf{x}_2'\mathbf{U})) \\
=&\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} E(\mathbf{V}_i(t_1)\mathbf{V}_j'(t_2)) E(P_i^{(\frac{d-1}{2})}(\mathbf{x}_1'\mathbf{U}) P_j^{(\frac{d-1}{2})}(\mathbf{x}_2'\mathbf{U})) \\
=&\sum_{i=0}^{\infty} \mathbf{B}_n(t_1-t_2) \times cov(a_n P_n^{(\frac{d-1}{2})}(\mathbf{x}_1'\mathbf{U}), a_n P_n^{(\frac{d-1}{2})}(\mathbf{x}_2'\mathbf{U})) \\
=&\sum_{i=0}^{\infty} \mathbf{B}_n(t_1-t_2) P_n^{(\frac{d-1}{2})}(cos\theta(\mathbf{x}_1,\mathbf{x}_2)),\ \mathbf{x}_1,\mathbf{x}_2 \in \mathbb{S}^d, t_1,t_2 \in \mathbb{Z}.
\end{aligned}$$

The latter is obviously isotropic and continuous on $\mathbb{S}^d$ and stationary on $\mathbb{Z}$. In particular,

$$\mathbf{B}_n(t_1-t_2) = \frac{1}{n}\begin{cases} (\mathbf{\Sigma}+\mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}')^{\circ n}, & t_1-t_2=0, \\ (\mathbf{\Phi}\mathbf{\Sigma})^{\circ n}, & t_1-t_2=-1, \\ (\mathbf{\Sigma}\mathbf{\Phi}')^{\circ n}, & t_1-t_2=1, \\ \mathbf{0}, & t_1-t_2 \pm 2, \pm 3, \ldots, \end{cases} , n \in \mathbb{Z}^+ \tag{2.9}$$

Again, $\mathbf{B}_0(t_1-t_2) = \mathbf{B}_1(t_1-t_2)$

Chapter 3

Deep Probabilistic Space and Time Imputation

3.1 Introduction

In electronics engineering, distribution system state estimation (DSSE) techniques infer the states of the system based on the network model and measurements taken from it. Nevertheless, the distribution system often has a restricted number of measurement equipment, especially for the medium and low voltage feeders which hides the development and use of DSSE. In the recent years, there has been an increase in installing different measurement sensors to improve monitoring the distribution system. For example, smart meters have been developed to measure the distribution systems on the secondary side and usually, measurements are taken every 15 minutes for customer's invoicing purposes. Also, the use of PMU (Phasor Measurement units) and SCADA (supervisory control and data acquisition) has increased significantly nowadays. Aggregating measurements from those smart meters as well as PMU and SCADA can significantly increase the information for the distribution system.

Despite the collecting multiple sources of data is critical for developing and use of DSSE, the main challenge faced by many researchers is how to aggregate this information. Un-

fortunately, this poses several significant challenges (Gómez-Expósito et al. (2014)): First, measurements from different sources have different sampling rates and are rarely in sync with each other. For example, typically the PMUs and SCADA systems are collecting data with range from each milli-seconds to minutes, while slow rate measures, such as smart meters from primary feeders, collect data over 15 minutes or an hour a time. Second, due to communication network impairments, some of this data could be missing or corrupted. While some of this data might be available to the utility directly, some may have to be accessed via the cloud leading to added latencies. Those issues indicate that the key challenges to maintain a reliable DSSE is to properly aggregate and resolve noisy, corrupted, heterogeneous, and insufficient time series data.

Similar problems can be also found in other types of multivariate spatio-temporal data measured from different time scales, for instance, medical data like patients' heart rate, blood pressure, etc. at different hospital are monitored at different unit of time. In environmental science and agricultural research, variables such as yield, precipitation, and temperature are frequently monitored at separate times and places and are rarely synchronized. Again, information can be intermittent with missing values due to malfunctioning measurement devices, partially observed states, pricey measurement products, or communication network impairments.

Multivariate spatial temporal correlated data with multiple time scales gives rise to three different ways of imputing missing information: Firstly, by using temporal correlations between each measurement. Secondly, by using exploiting the cross correlation between variables. In environment research, for instance, if a station's temperature for a specific time period is unrecorded, it might be informative if the precipitation at the time is higher, indicating that the temperature decreased. Finally, the spatial correlation between each location is extremely beneficial. Stations close to each other will have higher correlation than those with greater distances. Stations that are close to each other will have a higher correlation than those that are further apart.

3.2 Motivations and related work

There are many studies focused on data imputation, however, approaches fall short with respect to at least one of these desiderata. For example, classical methods like reconciling two time-scale measurements using linear interpolation/extrapolation based on weighted least squares approach (Gómez-Expósito et al. (2015)), classical time series imputation methods (Little and Rubin (2002), Pedersen et al. (2017)), but they do not take the potentially spatial-temporal interactions into account. While many statistical methods for multivariate time series analysis like Gaussian processes (Roberts et al. (2013)) works well with complete data in forecasting, these methods are generally inapplicable when there are missing values. Madbhavi et al. (2021) used tensor completion to fill the missing data in tensors by exploring the spatio-temporal correlation of the measurements, however, it ignored multi-time-scale measurements.

While several recent non-linear works like variational autoencoders (Ainsworth et al. (2018), Ma et al. (2018), Nazabal et al. (2020)) have investigated in missing value imputations, they have not utilized the essential time series correlations. There are several deep generative model for missing value imputation that does account for the time series dynamics like the GRUI-GAN (Luo et al. (2018)) and BRITS (Cao et al. (2018)) that uses recurrent neural networks (RNNs). Fortuin et al. (2020) used deep variational autoencoders(VAE) to map the missing time series data into a low dimensional space without messiness for data imputation. However, all of these methods fail to consider spatial correlations among locations.

Traditional statistical multivariate spatial-temporal Co-kriging for missing values methods are time consuming and have costly computational complexity, when the number of variables increase. It has been challenging to model multivariate spatio-temporal processes directly using a global covariance structure with lattice layout, which needs to be non-negative definite matrix function across different variables at different time and space. Also, the dimension is too large to be tractable for a reasonable output using analytical method. Models will get even harder to parameterize and justify if nonstationary is more preferred.

The multi-task Gaussian process frame-work methods proposed by Balasbramaniam Natarajan and Shweta Dahale (2022) provides one way to incorporate the spatio-temporal correlation as well as multiple time scales, however, it relies on choice of stationary covariance kernel ignoring the correlation among different variables (tasks). Other existing statistical or machine learning approaches either ignore temporal or spatial dynamics, and fall short in giving reliable uncertainty estimates with the imputations.

Along this line, the specific objective of this study is to utilize the information of spatial, temporal and correlations between each variables for data imputation. We will combine the non-linear dimensionality reduction with an expressive spatio-temporal model to achieve a more differentiable, higher computational efficiency method than traditional statistical as well as machine-learning models. We combined the idea from Fortuin et al. (2020) and propose a method of using deep probabilistic variational auto-encoders (VAE) to impute multivariate spatio-temporal data with multiple time scales. This can be done by learning the missing data from projecting the data space into a low dimensional space, which called latent space, without missingness, where we model the low dimensional dynamics with a Gaussian Process(GP). We propose a prior model that efficiently capture the space-time dependence structure at multiple time scales. Finally, our variational inference approach makes use of efficient structured variational approximations, where we fit another multivariate Gaussian process in order to approximate the intractable true posterior.

3.3 Methodology

3.3.1 Background

The goal is to study the data like network situational awareness, the energy data from smart meters, solar inverts, grid automation/SCADA sensors and micro PMU along with weather data that are collected in a geographical lattice over time. The complex correlation between and within different variables in space and time will be modeled using a multivariate spatio-temporal dynamic random filed $\mathbf{x}(s,t), t = 1, 2, ...T$ where t refers the the time

lag. For missing data imputation and resolution alignment, a multivariate spatio-temporal variational autoencoders (GP-VAE) will be used based on Gaussian processes and Bayesian deep learning, which allows for uncertainty estimation. The main idea is to project high dimensional random field $\mathbf{x}$ on a latent low-dimensional space with full determined spatial temporal structure and then impute missing values using the posterior distribution of $\mathbf{x}(s,t)$ given learned latent processes. The use of GP-VAE brings many benefits: Fist, the latent space is able to utilize all the correlations among variables to reconstruct the missing values. Second, through the encoding process, the model is able to smooth the original time series data to remove the noise and make the pattern more explainable. Third, since the constructed variational distribution include the temporal information, the imputing process can not only use other correlated variables' information but also get benefit of the time series association.

To summarize, this method comes from a combination of ideas from VAEs (Kingma and Welling (2013)), GPs (Rasmussen (2003)), Cauchy kernels (Jähnichen et al. (2018)), structured variational distributionswith efficient inference and a special ELBO for missing data (Nazabal et al. (2020)).

Let's suppose a multivariate spatial temporal data set $\mathbf{X} \in \mathbb{R}^{S\times T\times d}$ with $S \times T \times d$ data points: $\mathbf{x}_t = [\mathbf{x}_{1,t}^\intercal, \mathbf{x}_{2,t}^\intercal, ..., \mathbf{x}_{s,t}^\intercal]^\intercal \in \mathbb{R}^d$, where $S \in R, T \in R, d \in R$ represents locations, time lags, dimensions of $\mathbf{X}$ respectively. We assume that value from any place of $\mathbf{X}$ can be missing. In this case, we split the data into two partitions: Observed and unobserved data. The observed data at time t and location s can be denoted as $\mathbf{x}_{s,t}^o = [x_{s,t,k}|x_{s,t,k}$ is observed] and the missing data are $\mathbf{x}_{s,t}^m = [x_{s,t,k}|x_{s,t,k}$ is missing], $k = 1, 2, ...d$. $\mathbf{x}_{s,t}^o \cup \mathbf{x}_{s,t}^m = \mathbf{x}_{s,t}$. Now we're able to define our problem using the notations: Missing data imputation is to estimate the true values of missing variables: $\mathbf{X}^m = [x_{s,t,k}^m]$ given the observed variables: $\mathbf{X}^o = [x_{s,t,k}^o]$, where data are dependent on time, location and other variables.

3.3.2 Generative model

It is common to model the original partial data directly with the Gaussian Process (Roberts et al. (2013)), however it is not practical for some reasons: The cost of inverting the kernel matrix is high with time complexity of $O(n^3)$. Another problem is that to assure the kernel computable, the missing values are usually assigned to zeros, resulting in estimation bias. Also, it's hard to design a kernel function that can measure the complicated multivariate spatial temporal correlations. Although it is possible to treat the entire dataset as a collection of many single time series and use the GP to infer missing values separately, it will ignore the well-structured relationship between variables and across space.

This GP-VAE method can overcome those problems by using the GP in the latent space of a variational autoencoder where the encoded feature representations are complete. It assign a latent variable $\mathbf{z}_t \in \mathbb{R}^l$ for every $\mathbf{x}_t$ then use GP to model the temporal correlation in that latent space, with the prior $\mathbf{z}(t) \sim GP(m_z(.), k_z(.,.)), z \in 1, 2, ..., l$. This step can not only avoid the missing values in the latent space but also capture the spatial correlation and correlations between each variables (Fortuin et al. (2020)). As of the GP kernels, the Cauchy kernel for each latent variable

$$k_z(t, t') = \sigma^2(1 + \frac{(t - t')^2}{l^2})^{-1} \tag{3.1}$$

is well behaved for multi-scale time dynamics (Jähnichen et al. (2018)).

The likelihood given the latent variable $\mathbf{z}_t$ can be written as

$$p_\theta(\mathbf{x}_t|\mathbf{z}_t) = N(g_\theta(\mathbf{z}_t), \sigma^2\mathbf{I}) \tag{3.2}$$

where $g_\theta(\mathbf{z}_t)$ is a nonlinear function with the parameter vector θ. As one example, if we consider $\mathbf{z}_t$ as latent weather situations, then g_θ would be the process to generate observable measurements $\mathbf{x}_t$ like the temperature, wind speed, rainfall, etc. The g_θ function is estimated by a deep learning neural network.

3.3.3 Inference model

Now let's discuss the posterior distribution $p(\mathbf{z}_{1:T}|\mathbf{x}_{1:T})$. Since the exact posterior is intractable, variational inference is utilized (Blei et al. (2017), Jordan et al. (1999)). This method utilized a structured GP variational distribution with efficient inference to approximate true posterial distribution. To make the variational distribution more expressive and capture the temporal correlations of the data, we employ a structured variational distribution (Wainwright et al. (2008)) on Gaussian Process. The true posterior $p(\mathbf{z}_{1:T,j}|\mathbf{x}_{1:T})$ can be approximated by a multivariate Gaussian variational distribution:

$$q(\mathbf{z}_{1:T,j}|\mathbf{x}^o_{1:T}) = N(\mathbf{m}_j, \mathbf{\Lambda}_j^{-1}). \tag{3.3}$$

Where $\mathbf{\Lambda}_j = \mathbf{B}_j^\intercal \mathbf{B}_j$ and

$$\mathbf{B}_j = \begin{pmatrix} b^j_{11} & b^j_{12} & 0 & \cdots & 0 \\ 0 & b^j_{22} & b^j_{23} & \cdots & 0 \\ 0 & 0 & b^j_{33} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & b^j_{T-1T} \\ 0 & \cdots & \cdots & \cdots & b^j_{TT} \end{pmatrix} \tag{3.4}$$

j refers to the dimensions in the latent space. This approximation is to measure the temporal correlations in for each latent variable independently, which is commonly used in VAE (Kingma and Welling (2013), Rezende et al. (2014)). $b_{tt'}, t \in 1, 2, ...T$ are the local variational parameters. The benefit of this matrix is that it comes to positive definite, symmetric automatically. Samples from posterior distribution can be generated in linear time (Bamler and Mandt (2017), Huang and McColl (1997), Mallik (2001)). While the precision matrix is sparse, this covariance matrix can still be dense, allowing to reflect long-range dependencies in time.

For the VAE approach, we uses an inference network to optimize the $\mathbf{m}$ and $\mathbf{B}$ at the

same time. Let's suppose $q_\psi(\cdot)$ as the the variational distribution, where $\psi = \{\mathbf{m}, \mathbf{B}\}$. As mentioned before, the parameters for the generative model can be denoted as θ. θ and ψ can be trained together by minimizing the KL distance (Kullback and Leibler (1951)) to approximate $p(\mathbf{z}_{1:T}|\mathbf{x}_{1:T})$ using $q_\psi(\mathbf{z}_{1:T}|\mathbf{x}^o_{1:T})$. Where

$$
\begin{aligned}
(\theta^*, \psi^*) &= argmin\ (D_{KL}(q_\psi(\mathbf{z}_{1:T}|\mathbf{x}^o_{1:T})||p(\mathbf{z}_{1:T}|\mathbf{x}^o_{1:T}))),\ \theta \in \Theta, \psi \in \Psi \\
&= argmin\ (\mathbb{E}_{q_\psi(\mathbf{z}_{1:T}|\mathbf{x}_{1:T})}[\log\ q_\psi(\mathbf{z}_{1:T}|\mathbf{x}^o_{1:T})] - \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{x}_{1:T})}[\log \frac{p_\theta(\mathbf{x}^o_{1:T}|\mathbf{z}_{1:T})p(\mathbf{z}_{1:T})}{p(\mathbf{x}^o_{1:T})}]) \\
&= argmin\ (\mathbb{E}_{q_\psi(\mathbf{z}_{1:T}|\mathbf{x}_{1:T})}[\log\ q_\psi(\mathbf{z}_{1:T}|\mathbf{x}^o_{1:T})] - \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{x}_{1:T})}[\log\ p_\theta(\mathbf{x}^o_{1:T}|\mathbf{z}_{1:T})] - \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{x}_{1:T})}[\log\ p(\mathbf{z}_{1:T})] \\
&+ \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{x}_{1:T})}[\log\ p(\mathbf{x}^o_{1:T})]) \\
&= argmin\ (-\mathbb{E}_{q_\psi(\mathbf{z}_{1:T}|\mathbf{x}_{1:T})}[\log\ p_\theta(\mathbf{x}^o_{1:T}|\mathbf{z}_{1:T})] + D_{KL}(q(\mathbf{z}_{1:T}|\mathbf{x}^o_{1:T})||p(\mathbf{z}_{1:T})) + \log\ p(\mathbf{x}^o_{1:T})) \\
&= argmin\ (-\sum_{t=1}^{T} \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{x}_{1:T})}[\log\ p_\theta(\mathbf{x}^o_t|\mathbf{z}_t)] + D_{KL}(q(\mathbf{z}_{1:T}|\mathbf{x}^o_{1:T})||p(\mathbf{z}_{1:T})) + \log\ p(\mathbf{x}^o_{1:T})).
\end{aligned}
$$

Since $\log\ p(\mathbf{x}^o_{1:T}) \geq 0$, minimizing KL distance is equivalent to optimize the evidence lower bound (ELBO):

$$
\log\ p(\mathbf{X^o}) \geq \sum_{t=1}^{T} \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{x_{1:T}})}[\log\ p_\theta(\mathbf{x}^o_t|\mathbf{z}_t)] - \beta D_{KL}[q_\psi(\mathbf{z}_{1:T}|\mathbf{x}^o_{1:T})||p(\mathbf{z}_{1:T}))] \tag{3.5}
$$

Now this optimizing problem is simplified to maximize the RHS of (3.5) that is only based on the observed data with setting the missing values as zeros which is followed by Nazabal et al. (2020). Given that the likelihood term is the sum of all observed data, its highly related to the missing rate, so a trade-off parameter β is added to balance the influence of likelihood and KL distance term (Higgins et al. (2016)).

3.4 Simulation study

We performed simulation study based on simulated data of three-phase unbalanced IEEE 37 bus test system from Balasbramaniam Natarajan and Shweta Dahale. The loads connected

in this system are assumed to comprise of residential homes. In this paper, we simulate time series AMI measurements with active power (P) and reactive power (Q) in two phases at 37 different nodes, which are used to perform state estimate task and measure customer's power consumption. There are totally 6 measurements in 24 hours at each nodes over 1-min interval.

As is well known, missing AMI data from the real world can occur for a variety of reasons, including the sensors in a practical network may not always transmit data to the utility. Also, the measurements may be lost due to the communication network impairments. Thus, data can be missed at random, or at particular time periods. In many occurrences, AMI measurements are taken at 15-minute intervals, so we also include thinned data by removing gaps between every 1 and 15-minute intervals. On the basis of these details, we simulated different types of missing data: Missing at random, missing on the right, missing in the middle for 1 and 15 min interval data. GP-VAE data imputation can be accomplished in any of the time-scale.

There are a number of crucial tuning parameters, such as latent dimension, which denotes the dimension of latent space. The selection of the latent dimension is primarily based on the correlation between the various variables, the dimension of original data and the missing rate. The higher the missing rate, the correlations, the lower latent dimension can be considered. Another crucial tuning parameter that may affect the model's performance is window size. It's the smoothness time series parameter when transmit original data into latent space. GP-VAE favors a larger window size if the original data is smoother and more stable. In contrast, if the data is highly volatile, a smaller window size is more effective. Window size is between 1 and total dimension of time points. β is a trade-off parameter that to balance the influence of likelihood and KL distance in the optimization process. β can be approximately equal to 1 minus the missing rate. Another tuning parameter is the length scale that affects the time series correlation between two distinct time lags for a latent variable. When the length scale is decreased, the correlation approaches to zero more rapidly as the time gap comes raises.

We compared the performance of spatial GP-VAE with Universal-Kriging(UK) (Wack-

ernagel et al. (1997)), which is one of the most commonly used traditional data imputation methods. The RMSE has been used for model evaluation. The following tables shows the performance under different types of missing data settings.

Missing Rate	Universial Kriging	Spatial GP-VAE	GP-VAE
10%	30.4549	17.2429	216.6744
20%	271.1314	21.9390	233.9716
30%	273.1933	25.2602	228.4552
40%	104.5111	25.5899	258.8100
50%	86.68882	28.4501	267.3308

Table 3.1: *Missing on the right RMSE Statistics*

Missing Rate	Universial Kriging	Spatial GP-VAE	GP-VAE
10%	11.9339	19.6938	251.9476
20%	26.3303	25.3891	257.1676
30%	37.5289	27.7698	265.0235
40%	101.8372	28.1762	266.2166
50%	116.8307	30.2798	284.0116

Table 3.2: *Missing in the middle RMSE Statistics*

The missing rate represents the percent of missing data from the complete data set of a chosen variable. In this example, we use reactive power(Q) at phase C as testing variable. Missing on the right means we remove all the data after a specific time lag for imputation. For example, as this 1-minutes interval data contains 1440 minutes of records, imputation at 50% of missing rate means we use the 0 to 720 minutes of data to predict 720 to 1440 minutes data by different models. Missing in the middle means we removed the middle part of data with certain missing rate. For instance, missing rate 50% means we use 0 to 360, 1080 to 1440 minutes of data to impute the resting part. The table (3.1), (3.2) gives the

RMSE results for UK, spatial GP-VAE and VAE from this 1-minute interval data. From table (3.1), the Spatial GP-VAE outperforms all of other models at all the missing rates. Especially at missing rate 20% and 30%, spatial GP-VAE successfully decreased more than 93% of RMSE for both GP-VAE and UK. In table (3.2), although UK performs better on missing rate 10%, it's is getting worse when missing rate increase, while the performance of spatial GP-VAE is smooth and significantly better than other models from missing rate 20%.

Missing Rate	Universial Kriging	Spatial GP-VAE	GP-VAE
85%	42.7607	23.9191	160.0096
90%	68.6659	23.3511	163.1013
93%	66.7290	22.4084	161.0080
95%	68.4164	25.4699	164.4334
98%	269.511	22.8372	166.7708

Table 3.3: *15 minutes thinned data nested missing by random RMSE Statistics*

Missing Rate	Universial Kriging	Spatial GP-VAE	GP-VAE
10%	24.7876	20.9388	229.1943
20%	29.7614	18.4679	236.5785
30%	101.441	20.0650	234.8750
40%	215.1648	24.6416	265.6006
50%	73.8632	23.8794	270.7572

Table 3.4: *15 minutes thinned data missing on the right RMSE Statistics*

Missing Rate	Universial Kriging	Spatial GP-VAE	GP-VAE
10%	9.1739	16.8369	242.2727
20%	15.5523	18.8003	255.2829
30%	22.1169	17.0495	263.8181
40%	96.5531	26.7188	253.9683
50%	101.8183	31.7313	277.5901

Table 3.5: *15 minutes thinned data missing in the middle RMSE Statistics*

The table (3.3), (3.5) and (3.5) are the results from different missing rate for 15 minutes thinned data. In this case, there are total 96 time lags. Nested missing by random data are created starting from randomly removing 85% of data from 15 minutes thinned data, then for the remaining part, we continually randomly remove data until it reaches 90% of missing rate from whole 15 minutes thinned data. Then we repeat those steps until 98% of missing rate. Missing on the right and missing in the middle data are created by the similar technique as table (3.1) and (3.2). Again, for the nested missing by random and missing on the right data, the spatial GP-VAE reduced RMSE at all missing rate with best performance of reducing 90% and 92% of RMSE comparing to UK and GP-VAE. AS for missing in the middle, the performance of spatial GP-VAE is very stable and reduced more than 60% of RMSE at 40% and 50% of missing rate comparing to UK and GP-VAE. In conclusion, the benefit of spatial GP-VAE lies in that it's imputation result is smooth at all the missing rates. Especially for the data includes high percent of missing values, the traditional methods couldn't work properly however the spatial GP-VAE is able to obtain a relevantly high accuracy. Another advantage of spatial GP-VAE is its ability to impute multiple time series data at the same time, which can save a significant amount of time when compared to imputing each variable one at a time using the traditional UK method. Also, it's able to capture the spatial correlation between different locations to improve the prediction accuracy.

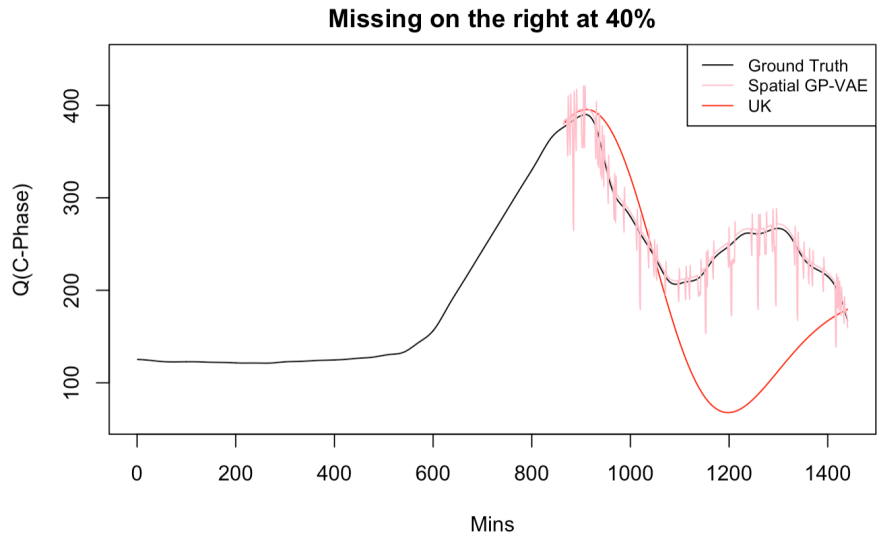

Figure 3.1: *Data imputation for AMI time series missing on the right*

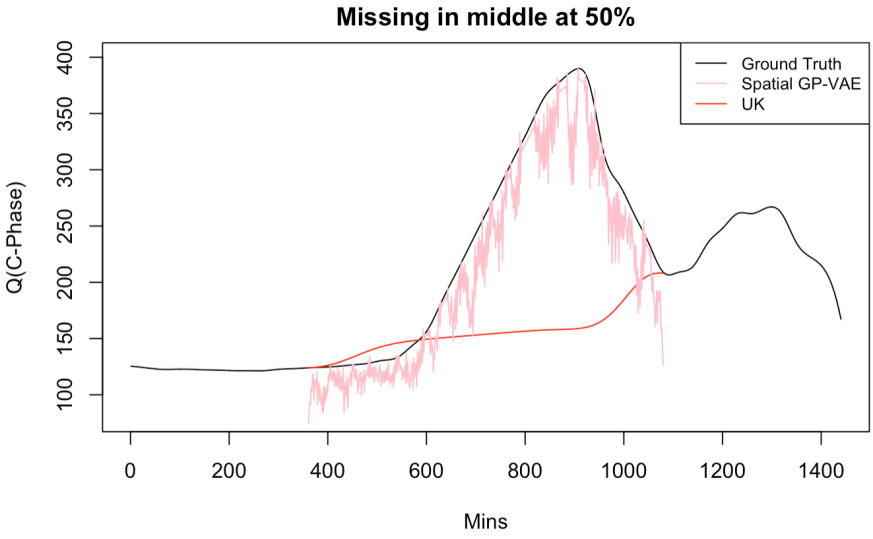

Figure 3.2: *Data imputation for AMI time series missing in the middle*

Fig.3.1 and Fig.3.2 shows the results obtained from imputation performed for reactive power (C-Phase) profiles with 40% missing on the right and 50% missing in the middle measurements. Comparing to the results from UK method that is a far way from the ground truth value, the trend could be accurately depicted by GP-VAE. This is due to the fact that spatial-temporal and cross correlation between variables in this model are exploited.

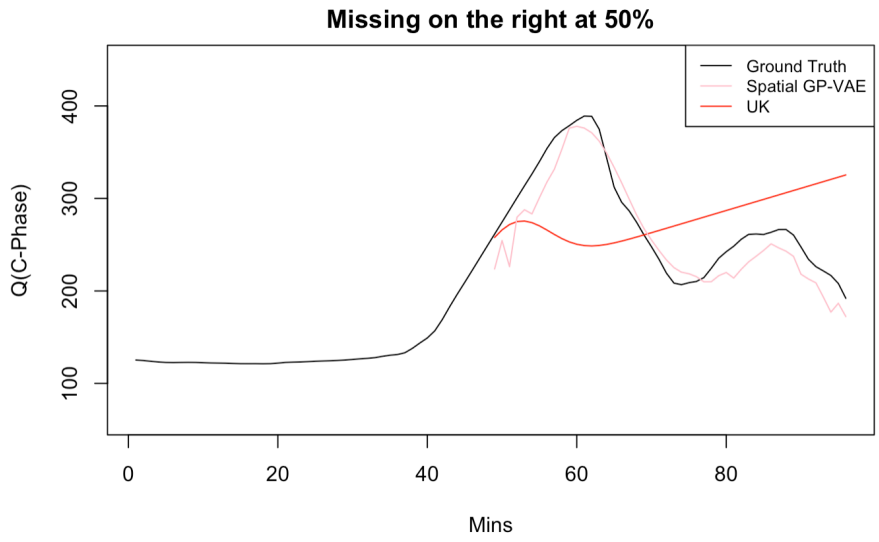

Figure 3.3: *Data imputation for AMI time series 15 minutes thinned data missing on the right*

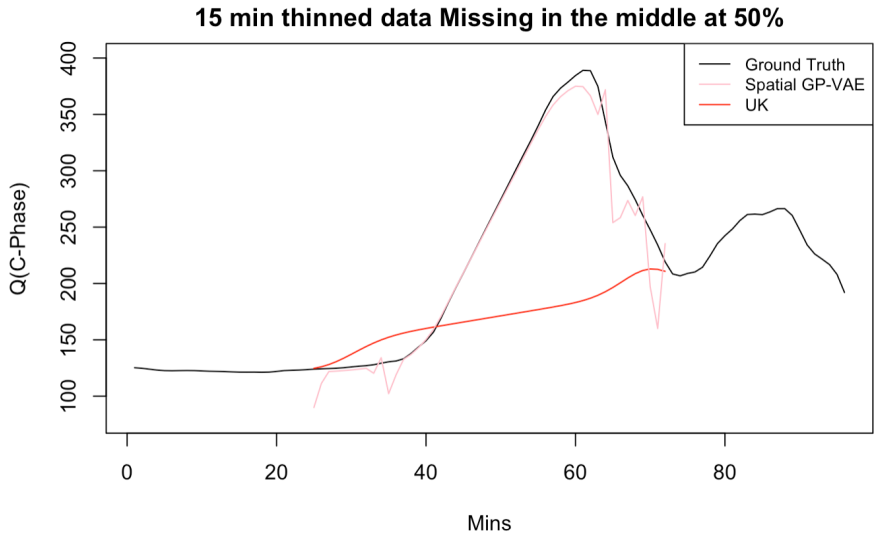

Figure 3.4: *Data imputation for AMI time series 15 minutes thinned data missing in the middle*

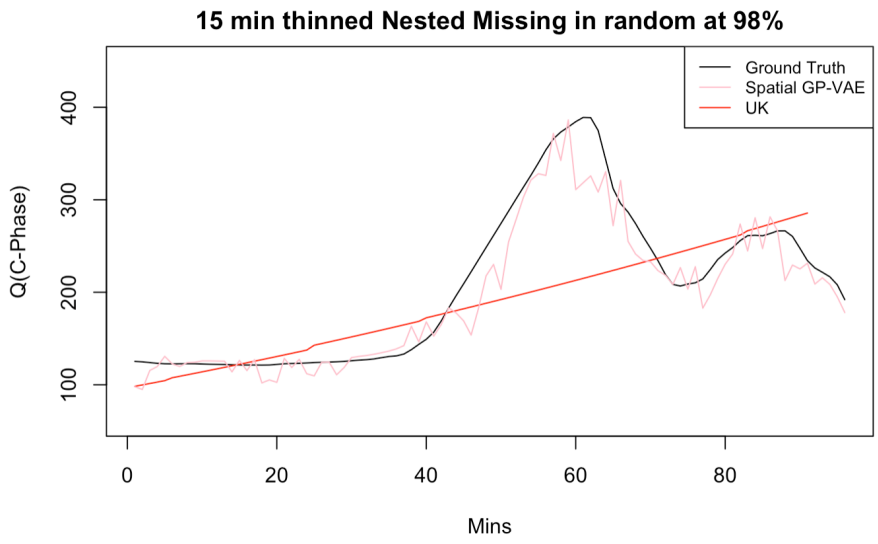

Figure 3.5: *Data imputation for AMI time series 15 minutes thinned data missing at random*

As seen from the figures(3.3),(3.4),(3.5), the results for 15 minutes thinned data, our approach provides accurate and smoother imputation than UK even only few available measurements are applicable. Another advantage of spatial GP-VAE is that when multivariate UK can only impute a small number of variables simultaneously, spatial GP-VAE can handle a large number of variables and the performance could be even better since more information is included.

3.5 Future work

In the first chapter, we will consider the more complex type of temporal pattern in the model for theorems 2,3,4. Also, instead of using least square method for parameter estimation, we can try to fit the correlation by using maximum likelihood method. Other types of spatial covariance structure, instead of Matérn can also be considered in the model.

In the second chapter, referring to the question about the representation in Malyarenko and Ma (2018), we conjecture that (2.8) might be a general form of $C(\theta(\mathbf{x}_1, \mathbf{x}_2); t)$. With the assumptions on $\{\mathbf{V}_n(t), n \in N_0, t \in \mathbb{T}\}$ or U changed, the process $\mathbf{Z}$ therein could represent

a large class of vector random fields on $\mathbb{M}^d \times \mathbb{T}$. For example, it would be of interest to see what $\mathbf{Z}$ produces, if U follows a distribution on $\mathbb{M}^d$ rather than the uniform one, such as angular central Gaussian, projected normal, Von Mises-Fisher, Bingham, Fisher-Bingham, or Kent distribution (Mardia and Jupp (2009), Kent et al. (2018)). Also, temporal structure can not be limited to first order of moving average one. Future simulation work may be capable of generating more general type of temporal margin like autoregressive and moving average(ARMA). This can be realized by changing the structure of $\mathbf{B}_n(t)$. In this chapter, (4) is more like the covariance structure of the MA(1) model so if we change $\mathbf{B}_n(t)$ to a MA(2), AR(1), or even ARMA type, the simulation may result in a spatial-temporal correlated data with corresponding temporal margin.

In the third chapter, our efforts will be focus on better measure the spatial correlation in between each channels. We're planing to assign a reduced dimensional latent dynamic random vector $\mathbf{z}_t$ to $\mathbf{x}(.,t)$ via Moran's I (MI) basis functions (Bradley et al. (2016)) for each component variable. Consequently, the generative network can be modeled using the vector autoregressive Gaussian process and the spectral representation of the MI operator. Therefore, the spatial-temporal correlation can be captured.

To be more specific, the Gaussian random field $\mathbf{x}(s,t)$ representing smart meter data can be characterized by the following spatio-temporal model:

$$\mathbf{x}(s,t) = \mu(s,t) + \mathbf{S}(s,t)\mathbf{z}_t + \xi(s,t) \tag{3.6}$$

The first term $\mu_x(s,t)$ is a fixed effect, which is unknown, and requires estimation. We set $\mu(s,t) = \mathbf{y}(s,t)\beta(t)$, where $\mathbf{y}(s,t)$ is a known p-dimensional vector of covariates and $\beta(t) \in \mathbb{R}^p$ is an associated unknown parameter vector; $t = 1, ...T$. The second term on the right-hand side of (1), i.e., $\mathbf{S}(s,t)\mathbf{z}_t$ represents multivariate spatio-temporal dependencies. The r-dimensional vectors of multivariate spatio-temporal basis functions $\mathbf{S}(s,t) \equiv (S(s,t)_1, S(s,t)_2, ..., S(s,t)_r)'$ are respecified for each locations and $t = 1, ..., T$. By a class of basis functions, we mean a collection of functions that form a basis for a function space (Franklin (2005), Chapter 5). A simple example of a class of basis functions would be

the Fourier basis functions, which are comprised of sine and cosine curves (used, e.g., by Royle and Wikle (2005)). Another popular choice are the bisquare spatial basis functions, which have the ability to be multiresolu tional (Cressie and Johannesson (2008); Katzfuss and Cressie (2011)). This multiresolutional property is a motivating factor for the use of wavelets as well. W-wavelets (Kwong and Tang (1994)) were introduced into the spatio-temporal literature by Nychka et al. (2002), and they have also been used to define S_r in spatial and spatio-temporal mixed-effects models. The r-dimensional random vector $\mathbf{z}_t$, $r \leq p$ is assumed to follow a spatio-temporal VAR(1) model (Cressie (2015), Chapter 7).

$$\mathbf{z}_t = \mathbf{M}_t \mathbf{z}_{t-1} + \mathbf{u}_t; \ t = 2, 3, 4.. \tag{3.7}$$

Where for all t the r-dimensional random vector $\mathbf{z}_t$ is Gaussian with mean $\mu_z(s,t)$ and has an unknown $r \times r$ covariance matrix $\mathbf{K}_t$; $\mathbf{M}_t$ is a $r \times r$ known propagator matrix and $\mathbf{u}_t$, is an r-dimensional Gaussian random vector with mean zero and unknown $r \times r$ covariance matrix $\mathbf{W}_t$ and is independent of $\mathbf{z}_{t-1}$. Finally, the third term on the right-hand side of (1), i.e., $\xi(s,t)$ represents fine-scale variability and is assumed to be Gaussian white noise with mean zero and unknown variance $\{\sigma^2_{\xi,t}\}$. On the other hand, following the similar technique in Fortuin et al. (2020), we employ inference amortization for the inference model that leads to an approximated posterior of $\mathbf{z}$ given $\mathbf{x}$. Again, the covariance can be specified based on reduced rank structure of the MI basis function and conditional autoregressive model (CAR). In addition, computational efficiency can be also achieved by applying Sherman–Morrison–Woodbury formula. Third, VAE training gives the parameters in both the generative model and inference network by optimizing the evidence lower bound.

3.6 Appendix

Attached is an example of the hyperparameters utilized in the spatial GP-VAE model for the AMI measurements data set.

Hyperparameter	Value
Number of CNN layers in inference network	1
Number of filters per CNN layer	128
Filter size (i.e., time window size)	10
Number of feedforward layers in inference network	1
Width of feedforward layers	126
Dimensionality of latent space	3
Length scale of Cauchy kernel	7
Width of feedforward layers	256
Activation function of all layers	RELU
Learning rate during training	0.001
Optimizer	Adam
Number of training epochs	30
Train/val/test split of data set	3
Dimensionality of time points	720
Length of time series	720
Tradeoff parameter β	0.5

Table 3.6: *Hyperparameters used in the spatial GP-VAE model.*

Bibliography

Samuel K Ainsworth, Nicholas J Foti, and Emily B Fox. Disentangled vae representations for multi-aspect and missing data. *arXiv preprint arXiv:1806.09060*, 2018.

Tatiyana V. Apanasovich, Marc G. Genton, and Ying Sun. A valid matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association*, 107(497):180–193, 2012. doi: 10.1080/01621459.2011.643197.

Richard Askey and NH Bingham. Gaussian processes on compact symmetric spaces. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 37(2):127–143, 1976.

Robert Bamler and Stephan Mandt. Structured black box variational inference for latent time series models. *arXiv preprint arXiv:1707.01069*, 2017.

Nicholas H Bingham and Tasmin L Symons. Gaussian random fields on the sphere and sphere cross line. *Stochastic Processes and their Applications*, 2019.

David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017.

Wei Cao, Dong Wang, Jian Li, Hao Zhou, Lei Li, and Yitan Li. Brits: Bidirectional recurrent imputation for time series. *Advances in neural information processing systems*, 31, 2018.

Stefano Castruccio and Michael L Stein. Global space-time models for climate ensembles. *The Annals of Applied Statistics*, pages 1593–1611, 2013.

Dan Cheng and Yimin Xiao. Excursion probability of gaussian random fields on sphere. *Bernoulli*, 22(2):1113–1130, 2016.

Jean-Paul Chiles and Pierre Delfiner. *Geostatistics: modeling spatial uncertainty*, volume 497. John Wiley & Sons, 2009.

Harald Cramér. On the theory of stationary random processes. *Annals of Mathematics*, pages 215–230, 1940.

Harald Cramér and M Ross Leadbetter. *Stationary and related stochastic processes: Sample function properties and their applications*. Courier Corporation, 2013.

Peter E Creasey and Annika Lang. Fast generation of isotropic gaussian random fields on the sphere. *Monte Carlo Methods and Applications*, 24(1):1–11, 2018.

N. Cressie. *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. Wiley, 2015. ISBN 9781119115175. URL https://books.google.com/books?id=gyLPBwAAQBAJ.

Noel Cressie and Hsin-Cheng Huang. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448):1330–1340, 1999. ISSN 01621459. URL http://www.jstor.org/stable/2669946.

Noel Cressie and Gardar Johannesson. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226, 2008.

Samuel Seth Demel and Juan Du. Spatio-temporal models for some data sets in continuous space and discrete time. *Statistica Sinica*, pages 81–98, 2015.

J. Du, C. Ma, and Y. Li. Isotropic variogram matrix functions on spheres. *Mathematical Geosciences*, 45:341–357, 2013a.

Juan Du and Chunsheng Ma. Spherically invariant vector random fields in space and time. *IEEE transactions on signal processing*, 59(12):5921–5929, 2011.

Juan Du and Chunsheng Ma. Variogram matrix functions for vector random fields with second-order increments. *Mathematical Geosciences*, 44(4):411–425, 2012.

Juan Du, Chunsheng Ma, and Yang Li. Isotropic variogram matrix functions on spheres. *Mathematical Geosciences*, 45(3):341–357, 2013b.

Vincent Fortuin, Dmitry Baranchuk, Gunnar Rätsch, and Stephan Mandt. Gp-vae: Deep probabilistic time series imputation. In *International conference on artificial intelligence and statistics*, pages 1651–1661. PMLR, 2020.

James Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005.

Ramesh Gangolli. Positive definite kernels on homogeneous spaces and certain stochastic processes related to levy's brownian motion of several parameters. In *Annales de l'IHP Probabilités et statistiques*, volume 3, pages 121–226, 1967.

Gregory Gaspari and Stephen E Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757, 1999.

Tilmann Gneiting. Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97(458):590–600, 2002. ISSN 01621459. URL http://www.jstor.org/stable/3085674.

TILMANN Gneiting. Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, 19(4):1327–1349, 2013. ISSN 13507265. URL http://www.jstor.org/stable/23525753.

Tilmann Gneiting, William Kleiber, and Martin Schlather. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491): 1167–1177, 2010. doi: 10.1198/jasa.2010.tm09420. URL https://doi.org/10.1198/jasa.2010.tm09420.

Antonio Gómez-Expósito, Catalina Gómez-Quiles, and Izudin Džafić. State estimation in two time scales for smart distribution systems. *IEEE Transactions on Smart Grid*, 6(1): 421–430, 2014.

Jean Carlo Guella, Valdir Antonio Menegatto, and Emilio Porcu. Strictly positive definite multivariate covariance functions on spheres. *Journal of Multivariate Analysis*, 166:150–159, 2018. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2018.03.001. URL https://www.sciencedirect.com/science/article/pii/S0047259X1730489X.

Antonio Gómez-Expósito, Catalina Gómez-Quiles, and Izudin Džafić. State estimation in two time scales for smart distribution systems. *IEEE Transactions on Smart Grid*, 6(1): 421–430, 2015. doi: 10.1109/TSG.2014.2335611.

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. 2016.

Mehdi Hosseinpour, Sina Sahebi, Zamira Hasanah Zamzuri, Ahmad Shukri Yahaya, and Noriszura Ismail. Predicting crash frequency for multi-vehicle collision types using multivariate poisson-lognormal spatial model: A comparative analysis. *Accident; analysis and prevention*, 118:277—288, September 2018. ISSN 0001-4575. doi: 10.1016/j.aap.2018.05.003. URL https://doi.org/10.1016/j.aap.2018.05.003.

Y Huang and WF McColl. Analytical inversion of general tridiagonal matrices. *Journal of Physics A: Mathematical and General*, 30(22):7919, 1997.

Patrick Jähnichen, Florian Wenzel, Marius Kloft, and Stephan Mandt. Scalable generalized dynamic topic models. In *International Conference on Artificial Intelligence and Statistics*, pages 1427–1435. PMLR, 2018.

Jaehong Jeong and Mikyoung Jun. Covariance models on the surface of a sphere: when does it matter? *Stat*, 4(1):167–182, 2015.

Jaehong Jeong, Mikyoung Jun, and M. Genton. Spherical process models for global spatial statistics. *Statistical science : a review journal of the Institute of Mathematical Statistics*, 32 4:501–513, 2017.

Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233, 1999.

Matthias Katzfuss and Noel Cressie. Spatio-temporal smoothing and em estimation for massive remote-sensing data sets. *Journal of Time Series Analysis*, 32(4):430–446, 2011.

John T Kent, Asaad M Ganeiber, and Kanti V Mardia. A new unified approach for the simulation of a wide class of directional distributions. *Journal of Computational and Graphical Statistics*, 27(2):291–301, 2018.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Pavel Krupskii and Marc G. Genton. A copula model for non-gaussian multivariate spatial data. *Journal of Multivariate Analysis*, 169:264–277, 2019. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2018.09.007. URL https://www.sciencedirect.com/science/article/pii/S0047259X18301696.

Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

Man Kam Kwong and P TP Tang. W-matrices, nonorthogonal multiresolution analysis, and finite signals of arbitrary length. Technical report, Argonne National Lab., IL (United States), 1994.

Annika Lang and Christoph Schwab. Isotropic gaussian random fields on the sphere: regularity, fast simulation and stochastic partial differential equations. *The Annals of Applied Probability*, 25(6):3047–3094, 2015.

Nikolai N Leonenko, Murad S Taqqu, and Gyorgy H Terdik. Estimation of the covariance function of gaussian isotropic random fields on spheres, related rosenblatt-type distributions and the cosmic variance problem. *Electronic Journal of Statistics*, 12(2):3114–3146, 2018.

Roderick JA Little and Donald B Rubin. Single imputation methods. *Statistical analysis with missing data*, pages 59–74, 2002.

Yonghong Luo, Xiangrui Cai, Ying Zhang, Jun Xu, et al. Multivariate time series imputation with generative adversarial networks. *Advances in neural information processing systems*, 31, 2018.

C. Ma. Stationary and isotropic vector random fields on spheres. *Mathematical Geosciences*, 44:765–778, 2012.

Chao Ma, Sebastian Tschiatschek, Konstantina Palla, José Miguel Hernández-Lobato, Sebastian Nowozin, and Cheng Zhang. Eddi: Efficient dynamic discovery of high-value information with partial vae. *arXiv preprint arXiv:1809.11142*, 2018.

Chunsheng Ma. Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference*, 116(2):489–501, 2003. ISSN 0378-3758. doi: https://doi.org/10.1016/S0378-3758(02)00353-1. URL https://www.sciencedirect.com/science/article/pii/S0378375802003531.

Chunsheng Ma. Spatio-temporal variograms and covariance models. *Advances in Applied Probability*, 37(3):706–725, 2005.

Chunsheng Ma. Vector random fields with second-order moments or second-order increments. *Stochastic analysis and applications*, 29(2):197–215, 2011.

Chunsheng Ma. Time varying isotropic vector random fields on spheres. *Journal of Theoretical Probability*, 30(4):1763–1785, 2017.

Rahul Madbhavi, Balasubramaniam Natarajan, and Babji Srinivasan. Enhanced tensor completion based approaches for state estimation in distribution systems. *IEEE Transactions on Industrial Informatics*, 17(9):5938–5947, 2021. doi: 10.1109/TII.2020.3035449.

Ranjan K Mallik. The inverse of a tridiagonal matrix. *Linear Algebra and its Applications*, 325(1-3):109–139, 2001.

Anatoliy Malyarenko. Abelian and tauberian theorems for random fields on two-point homogeneous spaces. *Theory of Probability and Mathematical Statistics*, 69:115–127, 2004.

Anatoliy Malyarenko. *Invariant random fields on spaces with a group action*. Springer Science & Business Media, 2012.

Anatoliy Malyarenko and Chunsheng Ma. Time-varying isotropic vector random fields on compact two-point homogeneous spaces. 2018.

Kanti V Mardia and Peter E Jupp. *Directional statistics*, volume 494. John Wiley & Sons, 2009.

Elias Silva de Medeiros, Renato Ribeiro de Lima, Ricardo Alves de Olinda, Leydson G. Dantas, and Carlos Antonio Costa dos Santos. Space–time kriging of precipitation: Modeling the large-scale variation with model gamlss. *Water*, 11(11), 2019. ISSN 2073-4441. doi: 10.3390/w11112368. URL https://www.mdpi.com/2073-4441/11/11/2368.

Alfredo Nazabal, Pablo M Olmos, Zoubin Ghahramani, and Isabel Valera. Handling incomplete heterogeneous data using vaes. *Pattern Recognition*, 107:107501, 2020.

Douglas Nychka, Christopher Wikle, and J Andrew Royle. Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling*, 2(4):315–331, 2002.

Frank WJ Olver, Daniel W Lozier, Ronald F Boisvert, and Charles W Clark. *NIST handbook of mathematical functions hardback and CD-ROM*. Cambridge university press, 2010.

Alma B Pedersen, Ellen M Mikkelsen, Deirdre Cronin-Fenton, Nickolaj R Kristensen, Tra My Pham, Lars Pedersen, and Irene Petersen. Missing data and multiple imputation in clinical epidemiological research. *Clinical epidemiology*, 9:157, 2017.

Emilio Porcu, Moreno Bevilacqua, and Marc G. Genton. Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *Journal of the American Statistical Association*, 111(514):888–898, 2016. doi: 10.1080/01621459.2015.1072541. URL https://doi.org/10.1080/01621459.2015.1072541.

Carl Edward Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.

Gregory C Reinsel. *Elements of multivariate time series analysis*. Springer Science & Business Media, 2003.

Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pages 1278–1286. PMLR, 2014.

Stephen Roberts, Michael Osborne, Mark Ebden, Steven Reece, Neale Gibson, and Suzanne Aigrain. Gaussian processes for time-series modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):20110550, 2013.

J Andrew Royle and Christopher K Wikle. Efficient statistical mapping of avian count data. *Environmental and Ecological Statistics*, 12(2):225–243, 2005.

Stephan R. Sain and Noel Cressie. A spatial model for multivariate lattice data. *Journal of Econometrics*, 140(1):226–259, 2007. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2006.09.010. URL https://www.sciencedirect.com/science/article/pii/S0304407606002302. Analysis of spatially dependent data.

Stephan R Sain, Reinhard Furrer, and Noel Cressie. A spatial analysis of multivariate output from regional climate models. *The Annals of Applied Statistics*, pages 150–175, 2011.

Wayan Somayasa, Makulau, Yulius Bara Pasolon, and Desak Ketut Sutiari. Universal kriging of multivariate spatial data under multivariate isotropic power type variogram model. *AIP Conference Proceedings*, 2326(1):020035, 2021. doi: 10.1063/5.0039429. URL https://aip.scitation.org/doi/abs/10.1063/5.0039429.

J.J. Song, M. Ghosh, S. Miaou, and B. Mallick. Bayesian multivariate spatial models for roadway traffic crash mapping. *Journal of Multivariate Analysis*, 97(1):246–273, 2006.

ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2005.03.007. URL https://www.sciencedirect.com/science/article/pii/S0047259X05000308.

Gabor Szeg. *Orthogonal polynomials*, volume 23. American Mathematical Soc., 1939.

C Tebaldi and DB Lobell. Towards probabilistic projections of climate change impacts on global crop yields. *Geophysical Research Letters*, 35(8), 2008.

Hans Wackernagel, Victor De Oliveira, and Benjamin Kedem. Multivariate geostatistics. *SIAM Review*, 39(2):340–340, 1997.

Martin J Wainwright, Michael I Jordan, et al. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

Yating Wan, Minya Xu, Hui Huang, and Song Xi Chen. A spatio-temporal model for the analysis and prediction of fine particulate matter concentration in beijing. *Environmetrics*, 32(1):e2648, 2021. doi: https://doi.org/10.1002/env.2648. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2648.

Jia Xu, Wen Yang, Bin Han, Meng Wang, Zhanshan Wang, Zhiping Zhao, Zhipeng Bai, and Sverre Vedal. An advanced spatio-temporal model for particulate matter and gaseous pollutants in beijing, china. *Atmospheric Environment*, 211:120–127, 2019. ISSN 1352-2310. doi: https://doi.org/10.1016/j.atmosenv.2019.04.011. URL https://www.sciencedirect.com/science/article/pii/S1352231019302262.

Mihail Iosifovic Yadrenko and Alampallam Venkatachalaiyer Balakrishnan. *Spectral Theory of Random Fields (Spektral'naja Teorija Sluchajnykh Polej)*. Optimization Software, Publications Division, 1983.

Akira Moiseevich Yaglom. Correlation theory of stationary and related random functions. *Volume I: Basic Results.*, 526, 1987.

Xuening Zhu, Danyang Huang, Rui Pan, and Hansheng Wang. Multivariate spatial autoregressive model for large scale social networks. *Journal of Econometrics*, 215(2):591–

606, 2020. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2018.11.018. URL https://www.sciencedirect.com/science/article/pii/S030440761930212X.