이 학 박 사 학 위 논 문

Bayesian regression using non-parametric modeling of Fourier coefficients and its applications

푸리에 계수의 비모수적 추정을 이용한 베이지안 회귀분석과 그 응용

2021년 2월

서울대학교 대학원

통계학과

전 윤 배

Bayesian regression using non-parametric modeling
of Fourier coefficients and its applications
푸리에 계수의 비모수적 추정을 이용한
베이지안 회귀분석과 그 응용

지도교수 임 채 영

이 논문을 이학박사 학위논문으로 제출함
2020년 7월

서울대학교 대학원
통계학과
전 윤 배

전윤배의 이학박사 학위논문을 인준함
2020년 8월

위 원 장	이 재 용	(인)
부위원장	임 채 영	(인)
위 원	Myunghee Cho Paik	(인)
위 원	김 선 영	(인)
위 원	최 정 순	(인)

Bayesian regression using non-parametric modeling of Fourier coefficients and its applications

By

Yoon-Bae Jun

A Thesis
Submitted in fulfillment of the requirement
for the degree of
Doctor of Philosophy
in Statistics

Department of Statistics
College of Natural Sciences
Seoul National University
February, 2021

ABSTRACT

Bayesian regression using non-parametric modeling of Fourier coefficients and its applications

Yoon-Bae Jun
The Department of Statistics
Seoul National University

We illustrate a nonparmetric modeling of Fourier coefficients, known as a spectral density, under a Bayesian framework to forecast a stationary one-dimensional random process and to predict a stationary two-dimensional random field on a regular grid. We switch from the time/space domain to the frequency domain, and introduce a Gaussian process prior to the log-spectral density.

First, we propose Bayesian modeling of spectral density for spatial regression on a regular lattice grid. An interpolation technique to convert an estimated spectral density to a covariance matrix is also proposed to avoid matrix inversion for the spatial prediction. Simulation study shows that our approach is robust in that it does not require a parametric form and/or isotropic assumption of a covariance

function. Also, our approach gives better prediction results over conventional spatial prediction under most parametric covariance models that we considered. We also compare our approach with other existing spatial prediction approaches using two datasets of Korean ozone concentration. Our approach performs reasonably good in terms of mean absolute error and root mean squared error.

Second, we propose Bayesian modeling of spectral density for time series regression with heteroscedastic autocovariance. Heteroskedastic autocovariance is modeled as time varying marginal variance multiplied by stationary autocorrelation. Bayesian Markov-Chain-Monte-Carlo(MCMC) is used to estimate coefficients of the B-spline basis representation of the log marginal variance function as well as a log spectral density at Fourier frequencies so that we can estimate time varying autocovariance function. Simulation results show that the proposed approach successfully detected the temporal pattern of the autocovariance structure. Even though we need to estimate spectral density at all Fourier frequencies during Bayesian procedure, our approach does not lose much efficiency on computation compared to estimating only a few parameters in a parametric model such as $ARMA$ or $GARCH$. We applied the proposed method to forecast foreign exchange rate data and it shows good prediction accuracy in a sense of overall low root mean squared errors.

Keywords: Gaussian process prior, Periodogram, Whittle likelihood, Spectral density

Student Number: $2014 - 21215$

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Bayesian modeling of a spectral density in a non-parametric way such as Carter and Kohn *et al.*(1997)[7] and Dey *et al.*(2018)[12] can naturally be generalized into a heteroscedastic stochastic process over $d(\geq 1)$-dimensional space domain. In this Ph.D. dissertation, we illustrate a nonparmetric modeling of Fourier coefficients, known as a spectral density, under a Bayesian framework to forecast a stationary one-dimensional random process and to predict a stationary two-dimensional random field on a regular grid. The goal is to switch from an auto-correlation on the time/space domain to a spectral density on the frequency domain, and introduce a modeling technique imposing Gaussian process prior to the log-spectral density.

Bayesian spatial regression using non-parametric modeling of Fourier coefficients

Classic spatial prediction methods, also known as kriging in geostatistics, have enabled us not only to make statistical inference about

spatial distribution, but also to predict or interpolate values of a variable of interest at unmonitored locations (Cressie1993; Gelfand, Diggle, Fuentes, & Guttorp 2010; Stein 1999).[10, 19, 54]. Kriging has a well-developed theory providing best linear unbiased predictors and quantified uncertainty at desired locations. Kriging requires covariance information in either parametric or non-parametric way. An empirical covariance function at different lags by non-parametric estimates can be used to construct the covariance matrix, but positive definiteness could fail depending on an estimation method (Cressie 1993)[10].

Instead, we typically assume a parametric structure of an autocorrelation function, $c(\cdot)$, such as spherical, Matérn, or powered exponential covariance functions and use the least square methods or maximum likelihood estimation for estimating their unknown parameters. However, this approach still has a possibility of model misspecification of $c(\cdot)$, which may cause poor prediction of a variable of interest. Also, it requires computation of a matrix inversion which causes great computational burden for high-dimensional spatial data as the spatial covariance matrix is a dense matrix ix (Aune, Simpson, & Eidsvik 2014; Gelfand *et al.* 2010)[19, 2]. Furthermore, typical isotropy assumption that $c(\cdot)$ is characterized only through the distance is rather limited as it is frequently violated in many applications.

Recently developed methodologies have made great progress in fitting a complicated shape of spatial distribution and handling large spatial data. Heaton, Datta, Finley, *et al.*(2019)[29] provides a case study to compare several modern methods for analyzing large spatial data. Fixed Rank Kriging approximates the spatial process as a linear combination using a set of a priori designed basis functions that

is fixed in number (Cressie & Johannesson 2008; Stein 2008)[11, 53]. Lattice Kriging considers multiresolution radial basis functions, which rely on simpler computation than the traditional kriging when sample size is large (Nychka, Bandyopadhyay, Hammerling, Lindgren, & Sain 2015; Nychka, Wikle, & Royle 2002)[47, 46]. Predictive process approach (Banerjee, Gelfand, Finley, & Sang 2008; Finley, Sang, Banerjee, & Gelfand 2009)[3, 17] uses a set of knot locations on which the process is approximated in the form of basis functions representation under a Bayesian framework. Multiresolution approximations (Katzfuss 2017; Katzfuss & Gong 2017)[31, 32] also uses basis function representation with compactly supported basis functions at different resolutions, which can be adapted to any given covariance function. Stochastic partial differential equation approaches (Lindgren, Rue, & Lindström 2011)[42] approximates a Gaussian process with a Matérn covariance function with a Markov random field, which bring an efficient calculation of a likelihood.

Different from the approaches mentioned above which make use of a basis function representation of a spatial process in some way, covariance tapering creates a sparse covariance matrix by multiplying compactly supported covariance function to increase computational efficiency and such approximation is theoretically investigated in in Furrer and Nychka (2006); Kaufman,Schervish, and Nychka (2008) and Du, Zhang, and Mandrekar (2009)[18, 33, 13]. Spatial partitioning also creates a sparse covariance matrix by assuming independence between observations across partitioned subregions and various options for partitioning are suggested in n Heaton, Christensen, and Terres (2017)[52, 36, 28] and [40]. Nearest Neighbor Gaussian process uses conditional specification of spatial processes to model a sparse

structure of a covariance matrix which enables efficient computation.

On the other hand, there are algorithmic approaches in handling large spatial data. Metakriging (Guhaniyogi & Banerjee2018)[23] is an approximate Bayesian method that introduces a combined posterior by subset posteriors from partitioned locations. The gapfill method (Gerber & Schaepman 2018)[20] is purely algorithmic and distribution-free. The approach chooses a subset in a neighborhood and prediction is based on sorting algorithms and quantile regression. Local approximation Gaussian process approach (Gramacy & Apley 2015)[22] focuses on prediction by training a Gaussian process predictor using the values nearby the prediction location based on a criterion related to mean squared prediction error. The algorithm allows adaptive selection of the number of the neighbors for training.

An alternative way to study spatially structured processes is the spectral representation approach. Note that any stationary process can be represented as a superposition of random harmonic oscillations, i.e. it can be represented by some modification of the conventional Fourier integral (Whittle 1954; Yaglom 1987) [57, 59]. Spectral analysis is a study of the spectral measure or spectral density function, which is a Fourier coefficient for the sinusoidal component of an autocorrelation (Gelfand *et al.* 2010)[19]. Once we define a spectral density function $f(\cdot)$ for an autocorrelation function $c(\cdot)$ in a d-dimensional space domain, spatial dependency can be modeled by spectral density because of one-to-one correspondence between $c(\cdot)$ and $f(\cdot)$.

In time series analysis, spectral methods are widely studied and the related theories are well-established (Brillinger 2001;Priestley 1981) [6, 49]. Especially, several studies for spectral density estimation and its application in time series regression have been already presented

(Carter & Kohn 1997; Choudhuri, Ghosal, & Roy 2004; Dey *et al.* 2018)[7, 8, 12]. Once we regard a temporal structure as a one-dimensional spatial structure, several aspects of spectral methods in time series analysis can be generalized into the process with more than one-dimension. Also, there are wide applications whose data are available on a grid so that a spectral method can be applied naturally.

Royle and Wikle(2015)[51] and Paciorek(2007)[48] consider representation of a spatial process using a spectral process so that the corresponding covariance matrix is decomposed into the orthogonal matrix with Fourier basis functions and the diagonal matrix with the values of the spectral density. This construction helps more efficient computation but it is under parametric modeling of a spectral density. Reich and Fuentes (2012)[50] used a Dirichlet process prior for spectral density so that the resulting covariance function is flexible. Guinness and Fuentes (2017)[26] considers discrete spectral approximation of a covariance function so that the approximated covariance matrix has a nested block circulant structure which is computationally efficient and circulant embedding can be done in a smaller size compared to o Stroud, Stein, and Lysen (2017)[55]. However, the approach is under parametric modeling of the spectral density. Guinness (2019)[25] proposes an iterative imputation approach to estimate spectral density non-parametrically from incomplete lattice data. Note that this approach is named as periodic embedding approach in Heaton *et al.*(2019)[29]. This work has been extended to multivariate and spatial-temporal data (Guinness 2018)[24].

We introduce non-parametric modeling of a spectral density under a Bayesian framework by considering a Gaussian process prior on log spectral density which leads estimation of a spatial autocorre-

lation matrix more efficient. Prediction is made concurrently during Bayesian inference. Our work is an extension of Carter and Kohn(1997)[7] and Dey *et al.*(2018)[12] in that we consider a spatial process and allow incomplete lattice data. Given the Gaussian process prior, we expect that our approach produces robust prediction results regardless of a covariance structure as we do not assume any parametric nor isotropic model on the covariance function. The works by Guinness(2019;2018)[25, 24] are comparable to our approach as the spectral density is estimated non-parametrically on incomplete lattice data but it has a different flavor as we handle estimation under the Bayesian framework. The work by Reich and Fuentes (2012)[50] is a Bayesian approach and proposes a flexible modeling for spectral density but it can be computationally demanding due to the nature of posterior sampling with a Dirichlet process prior.

Bayesian time series regression using non-parametric modeling of Fourier coefficients

Heteroscedasticity has played a key role in economic time series analysis. It is well-known that a number of important macroeconomic or financial return processes exhibit the pattern of heteroscedasticity (Engle, 1982) [14]. It has been one of the central issues in the literature how to model such processes to capture their underlying dynamics successfully. Classical time-varying volatility models, such as $ARCH$ and $GARCH$, that accommodate heteroscedasticity are typically stationary under some conditions on the parameter values. The $ARCH/GARCH$ processes are unconditionally uncorrelated while they are still dependent. For the uncorrelated error, the corresponding

spectral density is constant and estimating unconditional variability of classical $ARCH$- or $GARCH$-type models may not be of interest. To that end, one can incorporate heteroscedastic time-varying volatility within the model by assuming a time-varying parameter; for example, see Kim and Kim (2016) [38]. We undertake the task and illustrate how such a framework can be useful.

Specifically, we consider the following framework:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1.1}$$

where $\boldsymbol{y} = (y_1, \cdots, y_T)'$ be a variable of interest, $\boldsymbol{X} = (X_1; \cdots; X_p)$ be a covariate matrix where X_i is a $(T \times 1)$ vector of the i-th predictor, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of unknown parameters and $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_T)'$ is an error vector. Here we further assume $\epsilon_t = \sigma_{\epsilon,t}\, e_t$, where e_t is a weakly stationary continuous Gaussian process with $Var(e_t) \equiv 1$ and $\sigma_{\epsilon,t}$ is a time-varying volatility of ϵ_t. The corresponding autocorrelation function for e_t is $\gamma(\cdot)$. In such a setting, we assume that the log transformation of time-varying volatility $\sigma_{\epsilon,t}$ is represented by B-spline basis functions. We take hierarchical Bayesian framework so that appropriate prior distributions can lead posterior distributions of the spectral density and the time-varying volatility.

As shown in Dey *et al.* (2018) [12], a Gaussian process prior for the log transformation of a spectral density not only helps reduce the bias in regression function estimates, but also enhances the precision of forecasting through a better forecast of the model error. This study employs the prior choice of the spectral density suggested by Dey *et al.* (2018) [12], but also extends their work by allowing the volatility in (1.1) to change over time. It is well-known that a model with heteroscedasticity captures the underlying dynamics of an

economic/financial time series variable more efficiently than a naive constant-volatility model. By adopting the flexible modeling for the volatility of such a process, one can expect more precise estimation and forecasting for the variable of interest.

Estimation of the autocorrelation function for the error process via the spectral density and the related model parameters is conducted under a Bayesian framework here. In particular, the hierarchical Bayesian framework is known to be useful for macro-finance forecasting. Moving into the frequency domain from the time domain, we use an approximate likelihood for the parameters including the spectral density.

Unlike the traditional time-domain-based one, the frequency-domain-based approach adopted by this study ensures the Bayesian estimate of the covariance matrix to be invertible and positive semi-definite, which is a non-trivial advantage in a data-rich environment. In addition, the proposed Bayesian estimation allows one to incorporate relevant information into the covariance matrix estimation through the prior distributions of the model parameters, which also could help generate more precise parameter estimates and forecasts of the variable of interest.

The proposed methodology for (1.1) is applied to the issue of exchange rate forecasting in this study. One of the crying needs in modern macro-finance is to develop a methodology or a model that can forecast foreign exchange rates with decent accuracy. Over the years, many approaches with various degrees of success have been proposed in the literature. Meese and Rogoff (1983) [44] compare out-of-sample forecasting accuracy of various exchange rate models and find out that a random walk model outperforms the rest at one- to twelve-month

horizons for the dollar/pound, dollar/mark and dollar/yen. Engel and Hamilton (1990) [15] develop a long-swing model that explains the uni-directional move in the dollar value for long periods of time. They claim that their model generates better forecasts than a random walk.

Mark (1995) [43] shows that fundamentals such as money stock and output are useful in predicting exchange rates and that the regression with bias-adjusted coefficients can outperform a random walk model in forecasting the spot exchange rate. Kilian (1999) [34] reports no evidence of higher predictability for exchange rates at longer forecast horizons and concludes that the linear model framework underlying many empirical studies are likely to be mis-specified, suggestive of a non-linear data-generating process. To that end, Kilian and Taylor (2003) [35] propose a non-linear, exponential smooth transition autoregressive model to approximate the time series behavior of the real exchange rate. Molodtsova and Papell (2009) [45] find strong evidence of exchange rate predictability with Taylor rule fundamentals at the one-month horizon for 12 OECD countries over the post-Bretton Woods period. Ince, Molodtsova and Papell (2016) [30] further show that the Taylor rule fundamentals model outperform the Taylor rule differentials model (Engel, Mark and West, 2008) [16] and other traditional models in out-of-sample exchange rate forecasting.

On the Bayesian side, Wright (2008) [58] suggests pooling forecasts from several different exchange-rate models, called the Bayesian model averaging, to the problem of pseudo-out-of-sample exchange rate predictions. He shows that the forecasts generated by this model averaging methodology are very close to those based on the random walk forecast. Through our flexible non-parametric modeling approach, the proposed frequency-domain-based Bayesian estimation of the volatil-

ity and the error covariance could lead to further progress in forecasting exchange rate variables.

We will propose Markov Chain Monte Carlo (MCMC) modeling of spectral density for time series regression with heteroscedastic autocovariance. Log-transformed marginal variance function is modeled using B-spline basis functions as well as log spectral density using the Periodogram estimator at Fourier frequencies.

Outline

Chapter2 deals with a Bayesian spatial regression using non-parametric modeling of Fourier coefficients. Section 2.1 introduces brief background relevant to this proposal. Section 2.2 and section 2.3 introduce spectral methods and model description in detail. Section 2.4 provides simulation results for estimation and prediction for various scenarios, and Section 2.5 provides real data analysis with two Korean ozone exposure studies.

Chapter3 illustrates a Bayesian time series regression using non-parametric modeling of Fourier coefficients. Section 3.1 introduces brief background relevant to this proposal. In Section 3.2, we present particulars of the proposed non-parametric modeling of the error covariance and of the time-varying-volatility approach. The estimation procedure is described in Section 3.2. A simulation study to support our approach is in Section 3.3. A real data example involving foreign exchange rate data is provided in Section 3.4.

This dissertation ends with its conclusion and discussion on potential extensions in Chapter 4. Details of the proposed methodologies are relegated to an appendixA.

Chapter 2

Bayesian spatial regression using non-parametric modeling of Fourier coefficients

2.1 Preliminaries

A spatially distributed variable is typically modeled as a continuously indexed stochastic process, $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset R^d\}$, where D is a study region of interest, $\boldsymbol{s}$ represents a point of coordinate in D. Under a common regression structure, we consider $Y(\boldsymbol{s}) = \mu(\boldsymbol{s}; X) + \epsilon(\boldsymbol{s})$, where $\mu(\boldsymbol{s}; X)$ is a deterministic mean function including explanatory variables X with its popular choice $\mu(\boldsymbol{s}; X) = X(s)\boldsymbol{\beta}$ and $\epsilon(\boldsymbol{s})$ is a zero mean spatial process with a spatial dependence structure. We can further decompose $\epsilon(\boldsymbol{s}) = \sigma_\epsilon e(\boldsymbol{s})$, where $\sigma_\epsilon^2 = Var(\epsilon(\boldsymbol{s}))$ is a

marginal variance and $e(\boldsymbol{s})$ is a normalized process characterized by an autocorrelation function $\gamma(\cdot)$. In other words, we consider the following model.

$$Y(\boldsymbol{s}) = X(\boldsymbol{s})\boldsymbol{\beta} + \sigma_\epsilon e(\boldsymbol{s};\gamma),\ \boldsymbol{s} \in D \subset R^d. \tag{2.1}$$

We assume weak stationarity of $e(\boldsymbol{s})$ so that $Cov(e(\boldsymbol{s}), e(\boldsymbol{t})) = \gamma(\boldsymbol{s}-\boldsymbol{t})$, which is a common assumption for spatial data. In addition, we assume that Y is a Gaussian process which is well accepted for tractable modeling.

2.1.1 Spectral representation theorem

To begin with, we review the formal definition of spectral measure and spectral density followed by Bochner's theorem, and the spectral representation of stationary process in d-dimensional continuous space domain. Note that we defined e as a zero mean stationary process in R^d with an autocorrelation function γ in equation 2.1. Our goal is to verify that every e in space domain has a spectral representation in its corresponding frequency domain. From now on, we will consider complex-valued $e(s)$ because harmonic oscillations in complex form is much more convenient and simpler to be manipulated.[59] In order to guarantee identifiable spectral representation of e, we should further assume that e is L^2-continuous, or mean square continuous, i.e. for $\forall s \in D \subset R^d$,

$$\lim_{y \to s} E\{e(y) - e(s)\}^2 = 0$$

Given the assumption of weakly stationarity, e is mean square continuous if and only if γ is continuous at the origin. Namely, if γ is not continuous at the origin, e will not be mean square continuous at

any point. Such discontinuous random process must behave extremely irregularly and can hardly serve as a suitable model for any reasonable applied problem.[59] Thus, we will suppose that this condition can be naturally accepted throughout this paper. Note that the mean square continuity of e does not imply that the realization of e must be continuous. Thus, it is notable that this approach could include some discontinuous classes of realizations of the error process.

Let c be a complex-valued autocorrelation function on $\mathcal{F}^d$ induced by γ. Since γ is assumed to be continuous and nonnegative definite on R^d, it is obvious that c is continuous and nonnegative definite on $\mathcal{F}^d$. Bochner's theorem suggests that the continuous function c is nonnegative definite if and only if it can be represented in the following Fourier integral form

$$c(s) = \int_{W^d} exp(iw^t s)\Lambda(dw) \tag{2.2}$$

where $W = (-\infty, +\infty)$. We might think of (2.2) as the spectral representation of an autocorrelation function γ made by the positive finite measure Λ. Λ is what we call a spectral measure, or a spectral distribution function. If Λ is absolutely continuous, then it has a Radon-Nikodum derivative with respect to Lebesgue measure, i.e. $\lambda = \frac{d\Lambda}{dw}$, which is called a spectral density. Then, (2.2) becomes $c(s) = \int_{W^d} exp(is^t w)\lambda(w)dw$. It means that γ can be decomposed by elementary harmonic oscillations $\{exp(is^t w)\lambda(w), w \in W^d\}$. The spectral density λ can be recovered by inverse Fourier Transformation from $c(\cdot)$:

$$\lambda(w) = \frac{1}{(2\pi)^d} \int_{R^d} exp(-iw^t s)c(s)ds \tag{2.3}$$

If $\{e(s), s \in D \subset R^d\}$ is a zero mean L^2-continuous stationary process with spectral distribution Λ, there exists a complex-valued spectral process $\{Z(w), w \in W^d\}$ with orthogonal increments, such that

$$E\Big(|dZ(w)|^2\Big) = \lambda(w) \tag{2.4}$$

for $w \in W^d$, and

$$e(s) = \int_{W^d} exp(iw^t s) dZ(w) \tag{2.5}$$

This implies that a modelling of spectral density λ in d-dimensional frequency domain can be used to understand fully about the spectral process $\{Z(w)\}$, which corresponds to the desired stochastic process e in d-dimensional space domain.

2.1.2 Whittle Likelihood Approximation

Assuming $n_1 \times n_2$ regular grid over a rectangular study region $D \subset R^2$, let D_1 be the length of D in x-axis, D_2 be the length of D in y-axis, and $N = n_1 n_2$ be the sample size. We denote a complete set of regularly spaced locations $S^{\Delta}_{com} = \{\boldsymbol{s}_{jk} = (s_j, s_k) = (j\delta_1, k\delta_2); j = 0, 1, \cdots, (n_1 - 1), k = 0, 1, \cdots, (n_2 - 1)\}$, where $\delta_1 = \frac{D_1}{n_1}, \delta_2 = \frac{D_2}{n_2}$. We first consider completely observed samples $(\boldsymbol{Y}, \boldsymbol{X}) = \{(Y_{jk}, \boldsymbol{X}_{jk}) = (Y(\boldsymbol{s}_{jk}), X_1(\boldsymbol{s}_{jk}), \cdots, X_p(\boldsymbol{s}_{jk})); \forall \boldsymbol{s}_{jk} \in S^{\Delta}_{com}\}$, where p is the number of covariates. Then, the model (2.1) using the data becomes $Y_{jk} = \sum_{r=1}^{p} X_{rjk}\beta_r + \sigma_\epsilon e_{jk}$, for $j = 0, 1, \cdots, (n_1 - 1), k = 0, 1, \cdots, (n_2 - 1)$ and its matrix form is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sigma_\epsilon \boldsymbol{e} \tag{2.6}$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^t$ and $\boldsymbol{e} = (e_1, \cdots, e_N)^t$. For our Bayesian inference, we wish to model conditional posterior density given by

$$\pi(\boldsymbol{\beta}, \sigma_\epsilon^2, \gamma | \boldsymbol{Y}) \propto \pi(\boldsymbol{Y} | \boldsymbol{\beta}, \sigma_\epsilon^2, \gamma) \pi(\boldsymbol{\beta}, \sigma_\epsilon^2, \gamma)$$

By using $\boldsymbol{e} = (\boldsymbol{Y} - \boldsymbol{X}\beta)/\sigma_\epsilon$, the likelihood $\pi(\boldsymbol{Y} | \boldsymbol{\beta}, \sigma_\epsilon^2, \gamma)$ becomes

$$\pi(\boldsymbol{Y} | \boldsymbol{\beta}, \sigma_\epsilon^2, \gamma) \propto (\sigma_\epsilon^2)^{-n/2} |\boldsymbol{\Gamma}|^{-1/2} \exp(-\boldsymbol{e}' \boldsymbol{\Gamma}^{-1} \boldsymbol{e}/2)$$

If we model a spectral density λ instead of γ using the spectral representation theorem in Section 2.1.1, Whittle(1954) shows that, for N large,

$$\begin{aligned} &\pi(\boldsymbol{e} | \boldsymbol{\beta}, \sigma_\epsilon^2, \gamma) \propto (\sigma_\epsilon^2)^{-n/2} |\Gamma|^{-1/2} \exp(-\boldsymbol{e}' \Gamma^{-1} \boldsymbol{e}/2) \\ \approx\ &\pi(\mathcal{I} | \boldsymbol{\beta}, \sigma_\epsilon^2, \lambda) \propto (\sigma_\epsilon^2)^{-n/2} \prod_{\boldsymbol{w} \in W_\Delta^2} \lambda(\boldsymbol{w})^{-1/2} \exp\left(- \sum_{j=1}^{n} \mathcal{I}(\boldsymbol{w})/2\lambda(\boldsymbol{w}) \right) \\ &= (\sigma_\epsilon^2)^{-n/2} \prod_{\boldsymbol{w} \in W_\Delta^2} 2\lambda(\boldsymbol{w})^{1/2} \left[C_0 \exp\left(- \sum_{j=1}^{n} C_0 \mathcal{I}(\boldsymbol{w}) \right) \right] \end{aligned} \tag{2.7}$$

where $\boldsymbol{w} \in W_\Delta^2 = [-\pi/\delta_1, \pi/\delta_1) \times [-\pi/\delta_2, \pi/\delta_2)$, known as Fourier frequencies, and $\mathcal{I}$ is called a periodogram, which will be described later in Section 2.2.

Whittle's idea was to approximate the quadratic form in such a way as to avoid computing the explicit inverse and determinant of Γ. The approximation is based on the observation that Γ can be approximated by a circulant matrix, $\tilde{\Gamma}$, which in turn can be diagonalized by the FFT-matrix of $(F_n)_{jk} = w^{jk}/\sqrt{n}$ and its conjugate matrix $(F_n)^*$ (Anitescu 2012). In particular,

$$\boldsymbol{e}^t \Gamma^{-1} \boldsymbol{e} \approxeq \boldsymbol{e}^t \tilde{\Gamma}^{-1} \boldsymbol{e} = \boldsymbol{e}^t \tilde{\boldsymbol{\Gamma}}^{-1} \boldsymbol{e} = \boldsymbol{e}^t (F_{n_2} \otimes F_{n_1})^* \Lambda_{n_1 n_2}^{-1} (F_{n_2} \otimes F_{n_1}) \boldsymbol{e}$$

$$= \sum_{j=0}^{n_1-1} \sum_{k=0}^{n_2-1} \frac{|\boldsymbol{e}_{jk}|^2}{\lambda_\Delta(w_j, w_k)}$$

$$log|\sigma^2\Gamma| \approxeq log|\sigma^2(F_{n_2} \otimes F_{n_1})^* \Lambda_{n_1 n_2}(F_{n_2} \otimes F_{n_1})|$$

$$= \sum_{j=0}^{n_1-1} \sum_{k=0}^{n_2-1} log(\sigma^2 \lambda_\Delta(w_j, w_k))$$

These approximation suggest replacing the exact log-likelihood

$$L_W(\Theta) = \sum_{j=0}^{n_1-1} \sum_{k=0}^{n_2-1} \left(log(\sigma^2 \lambda_\Delta(w_j, w_k)) + \frac{|\boldsymbol{e}_{jk}|^2}{\lambda_\Delta(w_j, w_k)} \right)$$

Those approximate equality are to be interpreted as implying that

$$\lim_{N \to \infty} \frac{\boldsymbol{e}^t \Gamma^{-1} \boldsymbol{e}}{\boldsymbol{e}^t \tilde{\Gamma}^{-1} \boldsymbol{e}} = 1$$

almost surely.

Assume A1-A3 provided in Guyon (1982)[27]. All classical models comply with A1-A3, because for them λ is rational with respect to e^{iw}, without zeros or poles, and such that the second derivative of λ in Θ is continuous in Θ. We write $L_N^{(k)}(k = 0, 1, 2)$ for L_N and its first k derivatives in Θ; P_Θ gives the distribution of $\boldsymbol{e}$ for the parameter value Θ.

Proposition 2.1.1. *Let $\boldsymbol{e}$ observed on P_N satisfy the assumption A1-A3, see Guyon (1982)[27], then the whittle log-likelihood L_W and the exact log-likelihood L satisfies*

$$E_{P_\Theta} \sup_{k=0,1,2} \sup_{\Theta} \left| L_W^{(k)} - L^{(k)} \right| = \mathcal{O}(N^{1/2})$$

2.1.3 Fast Fourier Transform algorithm

During our MCMC procedures, we use Discrete Fourier Transform (DFT) operations in two different parts. One is a computation of periodogram estimator $\mathcal{I}_n$, which is a function of an inverse DFT of the normalized error e. The other is a computation of an approximated sample covariance matrix $\tilde{\Gamma}_n$, which is a function of an inverse DFT of the estimated spectral density λ. One of the main benefits of the spectral approach is that these DFT operations can be conducted efficiently by the use of popular numerical algorithm called Fast Fourier Transform (FFT). (Cooley and Tukey 1965, Bracewell 1986)[9, 5]. For example, in two-dimensional case, when the sample size of the data is $N = n_1 n_2$, FFT factorizes the DFT matrix into a product of sparse factors. Then the computational cost of DFT decreases from $\mathcal{O}(n_1^2 n_2^2)$ to $\mathcal{O}(n_1 n_2 log(n_1 n_2))$. FFT is fastest when N is a power of 2, but if not so, FFT is still an effective algorithm by packing the data with zero. Both DFT and inverse-DFT in one or two-dimension domain are readily evaluated using software such as MATLAB or R. More about FFT are introduced in several literatures (Bergland 1969, Good 1997)[4, 21].

Exact calculation of the quantities such as matrix inverse or matrix-vector multiplication is usually not feasible for large data due to computational burden. To solve this issue, various approximation methods have been introduced, and spectral approach is one of the powerful methods to reduce the computational cost of Gaussian process models when we can obtain regularly spaced data. At this study, we approximate Γ_n into $\tilde{\Gamma}_n$ by using FFT with modeling a spectral density λ. As we will see later, this approximation does not lose much on estima-

tion accuracy by the proper prior regularization in Bayesian framework. The use of $\tilde{\Gamma}_n$ instead of Γ_n can greatly improve the ability to compute such quantities in practice.

2.2 Proposed model

2.2.1 Periodogram

The periodogram is a well-known non-parametric estimate of the spectral density using the data observed on regularly spaced lattice. For two dimensional space domain ($d = 2$), assume that observed data are located at $n_1 \times n_2$ regular grid over a rectangular study region $D \subset R^2$. Let $\Delta = (\delta_1, \delta_2)$ be the spacing between neighboring observations in each direction. Then, periodogram is defined as follows:

$$\mathcal{I}_{n_1 n_2}(w_1, w_2) = \frac{1}{4\pi^2 n_1 n_2} \left| \mathcal{D}_{n_1 n_2}(w_1, w_2) \right|^2, \tag{2.8}$$

where

$$\mathcal{D}_{n_1 n_2}(w_1, w_2) = \sum_{j=0}^{n_1-1} \sum_{k=0}^{n_2-1} e(j\delta_1, k\delta_2) \exp[-\iota(w_1 j\delta_1 + w_2 k\delta_2)] \tag{2.9}$$

for $\boldsymbol{w} = (w_1, w_2) \in W_\Delta^2$.

It can be easily shown that $\mathcal{I}_{n_1 n_2}(\boldsymbol{w})$ are asymptotically independent and exponentially distributed with mean
$f_{\delta_1 \delta_2}(\boldsymbol{w}) = \sum_{Q_1 \in \mathcal{Z}} \sum_{Q_2 \in \mathcal{Z}} f\left(w_1 + \frac{2\pi Q_1}{\delta_1}, w_2 + \frac{2\pi Q_2}{\delta_2}\right)$ at the Fourier frequencies, $\boldsymbol{w} \in W_\Delta^2$ as $n_1, n_2 \to \infty$. Also, $\mathcal{I}_{n_1 n_2}$ is symmetric around the half of the Fourier frequencies, i.e.
$\mathcal{I}_{n_1 n_2}(w_1, w_2) = \mathcal{I}_{n_1 n_2}\left(\frac{2\pi}{\delta_1} - w_1, \frac{2\pi}{\delta_2} - w_2\right)$ for $\boldsymbol{w} \in W_\Delta^2$.
(See Appendix B)

2.2.2 Gaussian Mixture Approximation

Given $\boldsymbol{e}$, we can obtain the periodogram $\mathcal{I}_{n_1 n_2}$ at the Fourier frequencies, $\boldsymbol{w} \in W_{\Delta}^2$. Denote $n^h = \lceil \frac{N}{2} \rceil$ as the smallest integer greater than or equal to $\frac{N}{2}$. Then, $\mathcal{I}_{n_1 n_2}$ are asymptotically independent and exponentially distributed for the first half of n^h distinct Fourier frequencies. The other half of frequencies are not needed to evaluate in our approach due to the symmetricity of $\mathcal{I}_{n_1 n_2}$, As [7] suggested that logarithm of exponential distribution can be well-approximated by the five-component mixture of normal distributions with priori known means and variances, we consider the following approximation:

$$\log \mathcal{I}_{n_1 n_2}(\boldsymbol{w}) = \log f_{\delta_1 \delta_2}(\boldsymbol{w}) + \xi(\boldsymbol{w}) \tag{2.10}$$

with ξ having distribution $\pi(\xi)$ such that

$$\pi(\xi) \propto \sum_{l=1}^{5} p_l \phi_{v_l}(\xi - k_l), \tag{2.11}$$

where $\phi_v(\cdot - k)$ is normal density with mean k and variance v^2. (p_l, k_l, v_l) for $l = 1, \cdots, 5$ are provided in [7] as follows:
If a Fourier frequency $\boldsymbol{w} = (w_1, w_2) \in W_{\Delta}^2$ is on the boundary satisfying either $w_1 \in [\{-\pi/\delta_1\}, \{0\}, \{\pi/\delta_1\}]$ or $w_2 \in [\{-\pi/\delta_2\}, \{0\}, \{\pi/\delta_2\}]$,

$$(p_1, p_2, p_3, p_4, p_5) = (0.13, 0.16, 0.23, 0.22, 0.25)$$

$$(k_1, k_2, k_3, k_4, k_5) = (-4.63, -2.87, -1.44, -0.33, 0.76)$$

$$(v_1, v_2, v_3, v_4, v_5) = (8.75, 1.95, 0.88, 0.45, 0.41)$$

Otherwise,

$$(p_1, p_2, p_3, p_4, p_5) = (0.19, 0.11, 0.27, 0.25, 0.18)$$

$$(k_1, k_2, k_3, k_4, k_5) = (-2.20, -0.80, -0.55, -0.035, 0.48)$$

$$(v_1, v_2, v_3, v_4, v_5) = (1.93, 1.01, 0.69, 0.60, 0.29)$$

Let ψ be a latent variable that indicates a component in (2.11), $\boldsymbol{\varphi}$ be a vector of $\log \mathcal{I}_{n_1 n_2}(\boldsymbol{w})$ and $\boldsymbol{\theta}$ be a vector of $\log f_{\delta_1 \delta_2}(\boldsymbol{w})$. We pursue a hierarchical model and Bayesian inference by considering a Gaussian process prior (GP) for $\log f_{\delta_1 \delta_2}(\boldsymbol{w})$ with mean function $\nu(\cdot)$ and covariance function $\tau(\cdot, \cdot)$, and appropriate priors for hyper-parameters.

2.2.3 Proposed Gibbs sampler

The model and prior specifications are summarized as follows:

1. Data model:
   (a) $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sigma_\epsilon \boldsymbol{e}$ (space domain)
   (b) $\boldsymbol{\varphi} = \boldsymbol{\theta} + \boldsymbol{\xi}$ (frequency domain)

2. Process model:
   $\boldsymbol{\theta} \sim GP(\nu(\cdot), \tau(\cdot, \cdot))$ with $\nu(\boldsymbol{w}) \equiv 0$ and
   $\tau(\boldsymbol{w}_1, \boldsymbol{w}_2) = \tau_\theta^{-1} \exp(-\rho_{\theta_1}|w_{11} - w_{21}| - \rho_{\theta_2}|w_{12} - w_{22}|)$, $\boldsymbol{w}_i = (w_{i1}, w_{i2})$ for $i = 1, 2$.

3. Parameter models:
   $\boldsymbol{\beta} \sim N(\mu_\beta \mathbf{1}, \sigma_\beta^2 \boldsymbol{I})$,
   $P(\psi = l) = p_l$, for $l = 1, \cdots, 5$; $\rho_{\theta_1}, \rho_{\theta_2} \sim Unif(0, \rho_0)$ for some $\rho_0 > 0$, $\tau_\epsilon = 1/\sigma_\epsilon^2 \sim G(a, b)$, $\tau_\theta \sim G(c, d)$,
   where $G(a, b)$ is the Gamma distribution with mean ab.

We consider a Gibbs sampler from the above hierarchical structure by obtaining conditional posterior distribution of each parameter given

the data and other parameters. The conditional posterior densities are summarized in Table2.1. The detail construction is provided in the AppendixA.

Prior specification	Posterior distribution
$\boldsymbol{\beta} \sim \boldsymbol{N}(\mu_\beta \mathbf{1}, \sigma_\beta^2 \boldsymbol{I})$	$\boldsymbol{\beta} \mid \cdots \sim \boldsymbol{N}(\boldsymbol{\mu}_*, \boldsymbol{T}_*)$ $\boldsymbol{T}_* = \boldsymbol{X}^t \tilde{\boldsymbol{\Gamma}}_{n_1 n_2}^{-1} \boldsymbol{X} + \sigma_\beta^{-2} \boldsymbol{I}_p$ $\boldsymbol{\mu}_* = \boldsymbol{T}_*^{-1} \left(\boldsymbol{X}^t \tilde{\boldsymbol{\Gamma}}_{n_1 n_2}^{-1} \boldsymbol{Y} + \sigma_\beta^{-2} \mu_\beta \boldsymbol{I}_{n_1 n_2} \right)$
$\tau_\epsilon \sim G(a, b)$	$\tau_\epsilon \mid \cdots \sim G(a_*, b_*)$ $a_* = a + \frac{n_1 n_2}{2}$, $b_* = b + \frac{1}{2} \boldsymbol{r}^t \tilde{\boldsymbol{\Gamma}}_{n_1 n_2}^{-1} \boldsymbol{r}$
$\boldsymbol{\theta} \sim GP(\boldsymbol{\nu}, \boldsymbol{\tau})$	$\boldsymbol{\theta} \mid \cdots \sim \boldsymbol{N}(\boldsymbol{\nu}_*, \boldsymbol{\Upsilon}_*)$ $\boldsymbol{\Upsilon}_* = (\boldsymbol{\Upsilon}^{-1} + \boldsymbol{V}_\psi^{-1})^{-1}$ $\boldsymbol{\nu}_* = \boldsymbol{\Upsilon}_* \boldsymbol{V}_\psi^{-1} (\boldsymbol{\varphi} - \boldsymbol{\kappa}_\psi - \boldsymbol{\nu}) + \boldsymbol{\nu}$
$\tau_\theta \sim G(c, d)$	$\tau_\theta \mid \cdots \sim G(c_*, d_*)$ $c_* = c + \frac{n^h}{2}$, $d_* = d + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\nu})^t \boldsymbol{\Upsilon}^{-1} (\boldsymbol{\theta} - \boldsymbol{\nu})$
$\rho_{\theta_1}, \rho_{\theta_2} \propto 1$	$\pi(\rho_{\theta_j}^{(l)}) \propto \exp\left(-\frac{1}{2} log \lvert \boldsymbol{\Upsilon}^{-1}_{(\rho_{\theta_j}^{(l)}, \rho_{\theta_{-j}})} \rvert - \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\nu})^t \boldsymbol{\Upsilon}^{-1}_{(\rho_{\theta_j}^{(l)}, \rho_{\theta_{-j}})} (\boldsymbol{\theta} - \boldsymbol{\nu}) \right)$
$P(\psi_w = l) = p_l$	$p_l \Phi_{v_l}(\boldsymbol{\varphi}(w) - \boldsymbol{\theta}(w) - k_l)$

Table 2.1: Conditional posterior densities for the proposed Gibbs sampler

2.2.4 Prediction procedures

Once we obtain R Gibbs samples, we predict Y over a given study region D at unmonitored locations. In Bayesian framework, prediction of Y is based on conditional expectation $E(Y|Y_{obs})$ given the observed data, Y_{obs}. Given Gibbs samples, prediction of $Y(\boldsymbol{s}_0)$ at an unmonitored location $\boldsymbol{s}_0 \in D$ is given as $\hat{Y}(\boldsymbol{s}_0) = \frac{1}{R} \sum_{r=1}^{R} E(Y(\boldsymbol{s}_0)|Y_{obs}; \hat{\boldsymbol{\beta}}^{(r)}, \hat{\sigma_\epsilon}^{(r)}, \hat{\boldsymbol{\theta}}^{(r)})$ with $E(Y(\boldsymbol{s}_0)|Y_{obs}; \hat{\boldsymbol{\beta}}, \hat{\sigma_\epsilon}, \hat{\boldsymbol{\theta}}) = X(\boldsymbol{s}_0)\hat{\boldsymbol{\beta}} + \hat{\sigma_\epsilon} \hat{\boldsymbol{h}}^t \tilde{\boldsymbol{\Gamma}}^{-1} (Y_{obs} - X\hat{\boldsymbol{\beta}})$, where

$\hat{\boldsymbol{h}} = \widehat{Cov}(\boldsymbol{e}, e(\boldsymbol{s}_0))$ and $\tilde{\boldsymbol{\Gamma}} = \widehat{Cov}(\boldsymbol{e})$ [10]. The prediction error variance of Y is similarly obtained.

To obtain Gibbs samples and prediction results, we need to compute a matrix-vector multiplication involving $\tilde{\boldsymbol{\Gamma}}^{-1}$. If the sites at S^{Δ}_{com} are ordered from top to bottom and from left to right, the equation (2.2) leads to the covariance matrix $\tilde{\boldsymbol{\Gamma}}$ being $n_2 \times n_2$ block circulant matrix, where each block is also circulant of the size $n_1 \times n_1$. We adopt the approach by Anitescu *et al.*(2012)[1] which makes use of this block-circulant of circulant blocks (BCCB) structure for efficient computation of $\tilde{\boldsymbol{\Gamma}}^{-1}\hat{\boldsymbol{h}}$. The detail explanation is given in the AppendixA.

The computation of $\hat{\boldsymbol{h}}$ requires an additional technique. The covariance estimates retrieved by the estimated spectral density are only at lags in the form of $(j\delta_1, k\delta_2)$ and this is not enough to reconstruct $\hat{\boldsymbol{h}}$ since it requires covariance estimates at a finer resolution. However, a straightforward interpolation of the estimated spectral density, analogous to the approach by Dey *et al.*(2018)[12] is no longer applicable in $d(>1)$ dimensional space due to an aliasing effect of the spectrum of sample observations (Gelfand et al. 2010)[19]. To resolve this issue, we consider an interpolation of covariance estimates at a coarser resolution to get the covariance estimates at a finer resolution so that we can construct $\hat{\boldsymbol{h}}$. For example, $\hat{c}(0.5\delta_1, 0.5\delta_2)$ is obtained by bi-linear interpolation with four neighboring values, $\hat{c}(0,0)$, $\hat{c}(\delta_1, 0)$, $\hat{c}(0, \delta_2)$, $\hat{c}(\delta_1, \delta_2)$.

Our approach requires to sample θ, logarithm of the spectral density at the Fourier frequencies per Gibbs iteration, which can be time consuming. However, we argue that it does not lose much efficiency on computation compared to a conventional Bayesian spatial regres-

sion method under a parametric set-up. This is partly due to the fact that we used Discrete Fourier Transform (DFT) by taking advantage of Fast Fourier Transform (FFT) algorithm [9, 5], whose computation cost is $\mathcal{O}(n_1 n_2 \log(n_1 n_2))$. Also, due to symmetry of the spectral density and the periodogram about the origin, we only need to consider the first half of Fourier frequencies. If we permit to impose more restriction on the true spectral density such as isotropy, we can further improve computation speed and save memories by considering about one fourth of the frequencies.

2.3 Proposed model for observations on an incomplete grid

Let ζ_{jk} be a variable for indicating if Y is observed at $\boldsymbol{s}_{jk} \in S^{\Delta}_{com}$. That is, $\zeta_{jk} = 1$ if $Y(\boldsymbol{s}_{jk})$ is observed and $\zeta_{jk} = 0$ if it is missing. Now we consider a complete set which includes observations as well as missing values with indicators.

$$(\boldsymbol{Y}, \boldsymbol{\zeta}, \boldsymbol{X}) = \{(Y_{jk}, \zeta_{jk}, \boldsymbol{X}_{jk}) = (Y(\boldsymbol{s}_{jk}), \zeta(\boldsymbol{s}_{jk}), X_1(\boldsymbol{s}_{jk}), \cdots, X_p(\boldsymbol{s}_{jk}))\}$$

Denote $\boldsymbol{Y} = (\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis})^t$ where $\boldsymbol{Y}_{obs}$ is an observed part and $\boldsymbol{Y}_{mis}$ is a missing part of $\boldsymbol{Y}$. Recall that the matrix form of the model using the data is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sigma_\epsilon \boldsymbol{e}.$$

Suppose that both $\boldsymbol{X}$ and $\boldsymbol{Y}_{obs}$ are observed but $\boldsymbol{Y}_{mis}$ is missing at random given observations. Let $\boldsymbol{\Theta} = (\boldsymbol{\beta}^t, \sigma_\epsilon, \boldsymbol{\theta}^t, \psi, \tau_\theta, \rho_{\theta_1}, \rho_{\theta_2})^t$ be a vector of the entire model parameters. In Bayesian inference, it is common to treat $\boldsymbol{Y}_{mis}$ as a vector of latent variables. We can augment

missing observations by sampling from the conditional probabilities of missing observations, $h_{mis} = P(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{\zeta}, \boldsymbol{\Theta})$ within the MCMC procedure described in Section 2.2.3.

With Gaussian assumption of $\boldsymbol{Y}$, we can easily show that h_{mis} follows a multivariate normal distribution. Note that

$$\begin{pmatrix} \boldsymbol{Y}_{obs} \\ \boldsymbol{Y}_{mis} \end{pmatrix} \sim \boldsymbol{N}\left(\begin{pmatrix} \boldsymbol{X}_{obs}\boldsymbol{\beta} \\ \boldsymbol{X}_{mis}\boldsymbol{\beta} \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right)$$

so that the conditional distribution of $\boldsymbol{Y}_{mis}$ given $\boldsymbol{Y}_{obs}$ and $\boldsymbol{\Theta}$ is

$$\boldsymbol{Y}_{mis}|(\boldsymbol{Y}_{obs}, \boldsymbol{\Theta}) \sim \boldsymbol{N}\left(\boldsymbol{\mu}_{mis|obs}, \boldsymbol{\Sigma}_{mis|obs}\right),$$

where $\boldsymbol{\mu}_{mis|obs} = \boldsymbol{X}_{mis}\boldsymbol{\beta} + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{Y}_{obs} - \boldsymbol{X}_{obs}\boldsymbol{\beta}\right)$ and $\boldsymbol{\Sigma}_{mis|obs} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{21}$ and $\boldsymbol{\Sigma}_{22}$ are to be recovered from our model parameters. Bayes formula gives

$$h_{mis} = P(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{\zeta}, \boldsymbol{\Theta}) = \frac{\pi(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{\Theta})P(\boldsymbol{\zeta}|\boldsymbol{Y}, \boldsymbol{\Theta})}{\int \pi(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{\Theta})P(\boldsymbol{\zeta}|\boldsymbol{Y}, \boldsymbol{\Theta})d\boldsymbol{Y}_{mis}}$$

When we assume the missingness of $\boldsymbol{Y}$ occurs at random conditioning on both observed data and model parameters, i.e. $P(\boldsymbol{\zeta}|\boldsymbol{Y}, \boldsymbol{\Theta}) = P(\boldsymbol{\zeta}|\boldsymbol{Y_{obs}}, \boldsymbol{\Theta})$, then the component is nothing but a constant with respect to the unknown quantities $\boldsymbol{Y}_{mis}$ given $\boldsymbol{Y}_{obs}$, and $\boldsymbol{\Theta}$ [37]. Therefore, we can get the samples from h_{mis} by sampling from $\pi(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}, \boldsymbol{\Theta})$ within the proposed Gibbs sampler.

2.4 Simulation Study

In this section, we show performance of the proposed approach in terms of both estimation and prediction scheme. Then, we compare

with a parametric Bayesian approach under various simulation settings. Consider a regular grid over a rectangular study region D, denoted by $S_{\delta,n}$, in which the distance between neighboring observations in each direction is δ and the length of each direction is $n\delta$. In other words, $S_{\delta,n} = \{\boldsymbol{s}_{jk} = (s_j, s_k) = (j\delta, k\delta),\ j = 0, \cdots, n;\ k = 0, \cdots, n\}$. We construct two covariates X_1 and X_2 with a constant term such that X_1 is generated from a mixture of two normal distributions, i.e. $X_1 = p\xi_1 + (1-p)\sqrt{5}\xi_2$; $p \sim Ber(0.5), \xi_1, \xi_2 \sim N(0,1)$, and X_2 is generated from a standard exponential distribution. The regression coefficients corresponding to $(1, X_1, X_2)$ are set to $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^t = (0.01, 0.02, 0.03)^t$.

For Bayesian inference, we choose hyper-parameters $\mu_\beta = 0$, $\sigma_\beta^2 = 100$, $a = 100$, $b = 10$, $c = 100$, $d = 100$. Three chains with 10,000 iterations each with 9,000 burn-in are obtained. We call our proposed approach, the non-parametric spectral density Bayesian spatial regression as NSBSR and usual parametric Bayesian spatial regression as PBSR in short.

First, we consider simulated datasets with two different grid sizes ($S_{1,16}$ and $S_{1,32}$) and assuming an exponential covariance model, $\sigma^2 e^{-\|Ah\|/\phi}$ with $\sigma^2 = 1$ and $\phi = 10$ to investigate the estimation result of a spectral density by comparing with the true spectral density. Two choices of A are considered: $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ (isotropy) and $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1/2 \end{smallmatrix}\right)$ (anisotropy). The anisotropic choice of A implies that the x- direction is stretched twice compared to the y- direction. Figure 2.1 shows the estimated log-scale spectral densities in three-dimensional visualization. The first row is the true spectral density and the second row is the NSBSR-estimated log-scale spectral density. Compared to the true spectral densities, estimated spectral densities tend to over-estimate

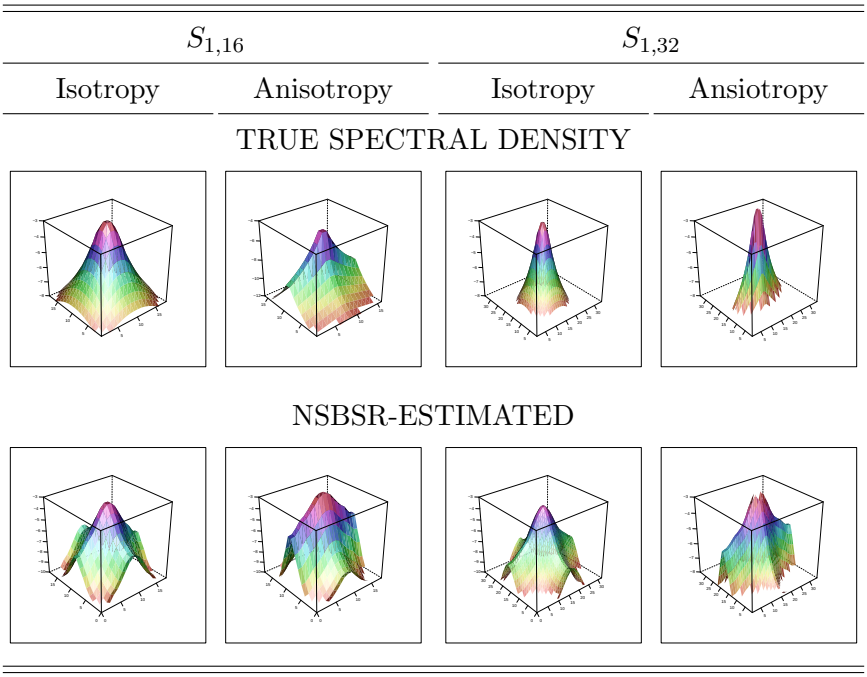

Figure 2.1: Estimated log-scale spectral densities assuming an exponential covariance model with $\sigma^2 e^{-\|Ah\|/\phi}$, $\sigma^2 = 1$, $\phi = 10$, $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ (isotropy), and $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1/2 \end{smallmatrix}\right)$ (anisotropy). First row corresponds to the true log-scale spectral densities while the second row corresponds to the estimated log-scale spectral densities using the proposed method.

at boundaries but they try to capture anisotropic patterns.

Next, we consider simulated datasets on $S_{1,32}$ and assuming a Matérn covariance model, $c(\boldsymbol{h}; \sigma^2, \phi, \alpha) = \sigma^2 \frac{2^{1-\alpha}}{\Gamma(\alpha)} \left(\frac{\|\boldsymbol{h}\|}{\phi}\right)^{\alpha} \mathcal{K}_{\alpha}\left(\frac{\|\boldsymbol{h}\|}{\phi}\right)$ at various smoothing levels α to investigate prediction performance. We set $\sigma^2 = 1$ and $\phi = 10$ as before. For this simulation, we consider $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ (isotropy) and $A = \begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 1/(2\sqrt{2}) \end{pmatrix}$ (anisotropy). The anisotropy choice of A implies that the x- direction is stretched twice compared to the y- direction and the coordinate is rotated by 45 degree. We then fit the model using the data only on $S_{2,16}$, which is a sub-sample of the data on $S_{1,32}$ with neighboring distance in each direction being twice large. The prediction is made on $S_{1,32}$ and compared with the generated data on $S_{1,32}$. Figure 2.2 shows prediction results from our NSBSR approach with observed values (simulated values). We can see that our approach tries to capture observed patterns for both isotropic and anisotropic cases.

Let us investigate several covariance models: (1) Pure Nugget (i.i.d.), (2) Bumpy Matérn ($\alpha = 0.1$), (3) Matérn ($\alpha = 0.5$), (4) Smooth Matérn ($\alpha = 2.0$), (5) Bumpy Powered Exponential ($\alpha = 0.5$), (6) Smooth Powered Exponential ($\alpha = 1.5$) and (7) Gaussian. Note that the form of Matérn covariance function with $\alpha = 0.5$ corresponds to the exponential covariance model. The powered exponential covariance function is $c(\boldsymbol{h}; \sigma^2, \alpha) = \sigma^2 \exp\left(-\|\boldsymbol{A}h\|^{\alpha}\right)$. We set $\sigma^2 = 1$ and $\phi = 10$ as well. In addition to isotropic cases, we investigate anisotropic models with $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1/4 \end{smallmatrix}\right)$, which implies that the x-direction is stretched four times compared to the y-direction. We simulate 100 datasets for each covariance setting with isotropy and

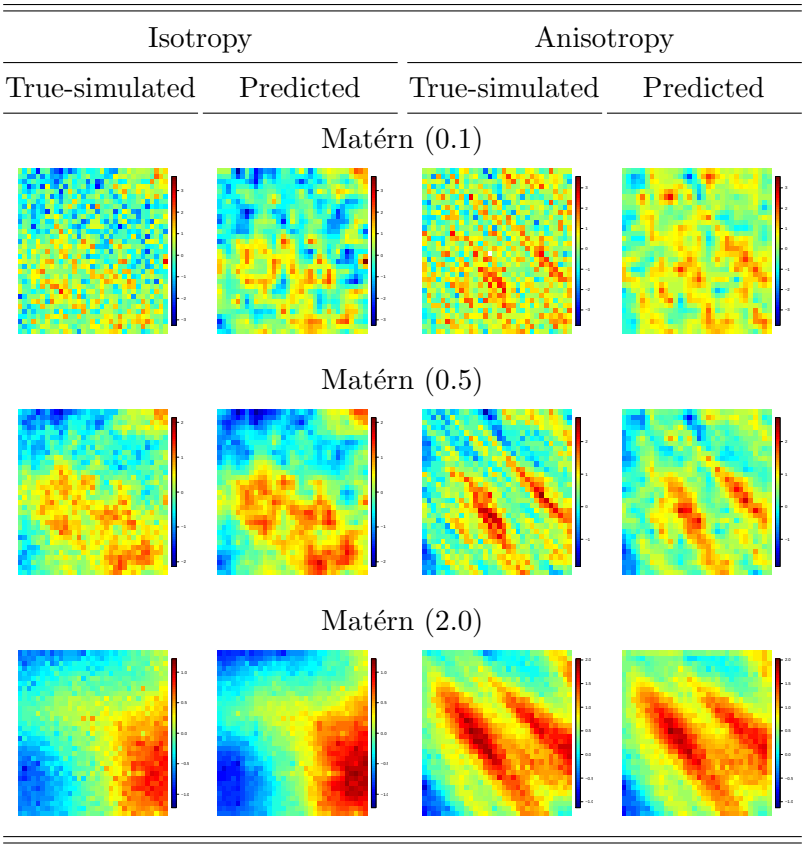

Figure 2.2: True simulated Y (left) and NSBSR predicted Y (right) in isotropic and anisotropic Matérn covariance models. Each row corresponds to a different smoothness parameter, $\alpha = 0.1, 0.5, 2.0$, respectively.

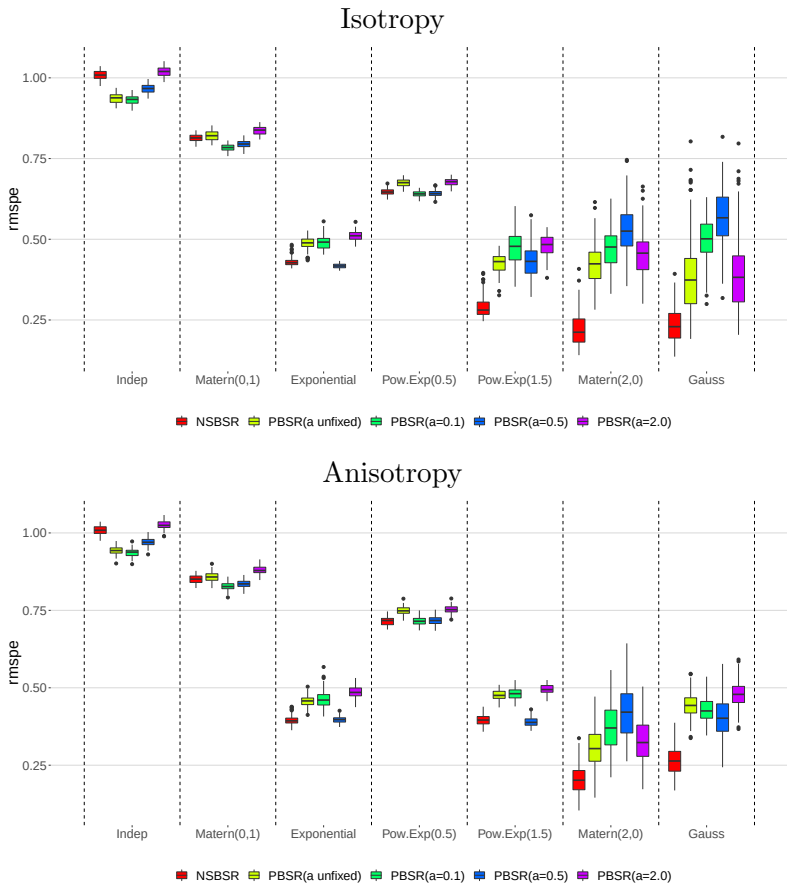

Figure 2.3: Box plots of RMSPE: Isotropic case is the first row and anisotropic case is the second row. Boxplots over 100 datasets using averages of root mean squared prediction errors over locations. Each block in the figure represents a different covariance model for data generation: Pure nugget, Bumpy Matérn ($\alpha = 0.1$), Exponential, Bumpy Powered exponential ($\alpha = 0.5$), Smooth Powered exponential ($\alpha = 1.5$), Smooth Matérn ($\alpha = 2.0$), and Gaussian. In each block, boxplots are ordered by estimation models: NSBSR (red), Bumpy PBSR (α fixed to 0.1; green), Exponential PBSR (α fixed to 0.5; blue), and Smooth PBSR (α fixed to 2.0; purple), and general PBSR with α unfixed (yellow).

ansiotropy. Similar to the second simulation case that produces Figure 2.2, we first generate the data on $S_{1,32}$ and use the data on $S_{2,16}$ to fit the model. The prediction is made on $S_{1,32}$ and compared with the generated data on $S_{1,32}$.

Figure 2.3 shows prediction performance results of NSBSR (red) and PBSR with Matérn model with different degrees of fixed smoothness parameter $\alpha = 0.1$ (green), $\alpha = 0.5$ (blue), $\alpha = 2.0$ (purple), and unfixed smoothness parameter α (yellow) for each simulation setting. A fixed smoothness parameter means that we estimate other parameters except for the smoothness parameter. An unfixed smoothness parameter means we estimate it as well. These are boxplots of root mean squared prediction errors (RMSPE) between observations and predicted values over 100 datasets. RMSPE is averaged over locations for each data set. From the left block (divided by the dotted vertical lines), the covariance models for data generation are, in turn, pure nugget (i.i.d.), Bumpy Matérn ($\alpha = 0.1$), Exponential, Bumpy powered exponential ($\alpha = 0.5$), Smooth powered exponential ($\alpha = 1.5$), Smooth Matérn and Gaussian. RMSPEs for NSBSR are overall comparable to those for PBSR in the case of lower degree of smoothness (first four blocks) and lower in the case of higher degree of smoothness (last three blocks). Sample visualization results for NSBSR in Figure 2.2 also imply that the predicted values for a smoother covariance model is relatively less biased than those for a bumpy covariance model. RMSPEs for NSBSR are quite robust compared to those for the PBSR with various covariance models. Note that unfixed Matérn model results are not impressive. Results for anisotropic cases are also similar, although the difference in prediction performance at smooth covariance models is reduced.

		MSPE					R^2				
Matérn	MR(%)	NSBSR	P00	P01	P05	P20	NSBSR	P00	P01	P05	P20
$\alpha = 0.1$	0	0.691	0.691	0.693	0.690	0.694	0.56	0.56	0.56	0.56	0.56
	10	0.814	0.815	0.820	0.813	0.816	0.54	0.54	0.54	0.54	0.53
	25	0.810	0.808	0.823	0.805	0.808	0.50	0.50	0.51	0.51	0.50
	50	0.778	0.775	0.775	0.777	0.773	0.44	0.45	0.45	0.45	0.45
$\alpha = 0.5$	0	0.187	0.187	0.187	0.187	0.186	0.96	0.96	0.96	0.96	0.96
	10	0.226	0.227	0228	0228	0.227	0.86	0.86	0.86	0.86	0.86
	25	0.230	0.229	0.233	0.232	0.229	0.84	0.84	0.84	0.84	0.84
	50	0.279	0.276	0.280	0.280	0.276	0.74	0.74	0.74	0.74	0.74
$\alpha = 2.0$	0	0.042	0.116	0.127	0.125	0.119	1.00	1.00	1.00	1.00	1.00
	10	0.133	0.172	0.166	0.210	0.170	1.00	1.00	1.00	1.00	1.00
	25	0.196	0.196	0.203	0.247	0.196	0.99	0.99	0.99	0.99	0.99
	50	0.252	0.209	0.220	0.265	0.209	0.80	0.82	0.81	0.82	0.82

Table 2.2: Mean squared prediction error (MSPE) and R squares (R^2) between observations and predicted values under various missing ratios (MR). The first column shows a smoothness parameter setting of Matérn covariance for data generation. NSBSR is the proposed approach. P00, P01, P05 and P20 are results from PBSR by considering Matérn covariance for model fitting. A smoothness parameter for PBSR is unfixed (P00), fixed at 0.1 (P01), fixed at 0.5 (P05), and fixed at 2.0 (P20), respectively.

We also compare NSBSR with PBSR when the data are on an incomplete grid. We generate three exemplary datasets from Matérn covariance models with $\alpha = 0.1, 0.5, 2.0$, respectively on the complete grid $S_{1,32}$. Then, we consider the data only on $S_{2,16}$ for fitting but we randomly remove grid points according to the missing ratio (MR, %). Then, our approach, NSBSR is compared to PBSR with Matérn models with $\alpha = 0.1$ (P01), $\alpha = 0.5$ (P05), $\alpha = 2.0$ (P20) and unfixed α (P00). Table 2.2 shows Mean squared prediction error (MSPE) and R^2

between observations and predicted values. Note that MSPE in this simulation study is the average over locations only as we consider one dataset. The results in Table 2.2 show that MSPEs of NSBSR tend to get increased as MR increases but the increase is comparable to those of PBSR. Likewise, the R^2 of NSBSR get decreased as MR increases but the decrease is comparable to those of PBSR. For example, with the data from the Matérn covariance model with $\alpha = 0.1$, MSPE of NSBSR approach is 0.814 while PBSR with Matérn covariance model fixed at $\alpha = 0.1$ (P01) is 0.820 when MR is 10%. The missing ratio does not affect much on the estimation of regression coefficients for both NSBSR and PBSR as well (results are not shown for brevity). However, we need longer MCMC chains when the missing ratio gets higher for convergence. Although the impact of missing rates is not apparent for this particular simulation study, convergence can be an issue in conditional simulation for imputing the missing data as discussed in [26].

2.5 Real Data Analysis

In this section, we use two ozone datasets. One is from Moderate Resolution Imaging Spectroradiometer (MODIS) atmosphere dataset [1], and the other is from AURA (EOS CH-1) which is a multi-national NASA scientific research satellite studying the Earth's ozone layer, air quality, and climate [2]. We expect that our methodology can be used for exposure assessment to acquire valuable scientific meanings such as health effect estimation of ambient air pollution on mortal-

[1] https://ladsweb.modaps.eosdis.nasa.gov/search/

[2] https://disc.gsfc.nasa.gov/datasets?keywords=aura&page=1

ity/morbidity in general epidemiological studies [41, 39].

Abbreviation	Method
NNGP	Nearest Neighbor Gaussian Process
PP	Predictive Process
Tapering	Covariance Tapering
Gapfill	Gapfill
Partition	Spatial partitioning
FRK	Fixed Rank Kriging
SPDE	Stochastic Partial Differential Equations
Periodic	Periodic Embedding
LK	Lattice Kriging
LAGP	Local Approximate Gaussian Processes
NSBSR	The proposed Bayesian Spectral method

Table 2.3: List of spatial prediction methods in [29] we compared with for prediction performance.

We compare the prediction results with several other approaches reviewed in [29]. Table 2.3 is the list of methods that we compare with and their abbreviations. A brief summary of each method is given in the Introduction section. For implementation, we basically used the codes available from *Heaton et al.*(2018)[29] and some R packages if they are available. Given the use of available codes, we omit two approaches: Metakriging and Multiresolution approximation in comparison. The computation time was rather long for Metakriging and we were not successful to implement the code for Multiresolution approximation.

MODIS Application

The first dataset is from Moderate Resolution Imaging Spectroradiometer (MODIS) Terra Level-3 Aerosol Cloud Water Vapor Ozone Daily Global product (MOD08D3). MOD08D3 contains daily-averaged values of atmospheric parameters related to aerosol particle properties, cloud optical and physical properties, atmospheric water vapor, atmospheric profile and stability indices, and total ozone burden on a $1^\circ \times 1^\circ$ grid. Among them, we obtained quality controlled ozone exposure measurements and the missing values were left untreated. To properly impose a Gaussian assumption, log-transformed ozone exposure is used as a response variable Y. MOD08D3 originally has global coverage in 360×180 pixels, but we restrict our study region to a neighborhood of the Korea peninsula, i.e. longitude ranged from 112° to 141°, latitude from 24° to 53°. The daily average on February 5, 2019 was used for analysis as an example to deal with a forecast of a short-term ambient exposure. For covariates, we used the world geodetic system (WGS 84) information so that X_1 refers to longitude, and X_2 refers to latitude.

MOD08D3 is a rectangular image pixels with 1° resolution as we mentioned above. We take a subset of the pixels with 2° resolution for model fitting. We then predict the values with 1° resolution. The missing rate is 13.0% for the original dataset. Both sample size and missing rate are moderate. As our approach assumes stationarity, we checked the data with a stationarity test introduced by [56], which is designed for testing stationarity of random fields on a regular lattice. The test is available as a `R` function (`TOS2D`) in `LS2Wstat` package. As it requires a complete set of data on a grid, we imputed the data by

ordinary kriging before applying the test. The p-value is 0.851, which indicates that we can not reject stationarity assumption of the data.

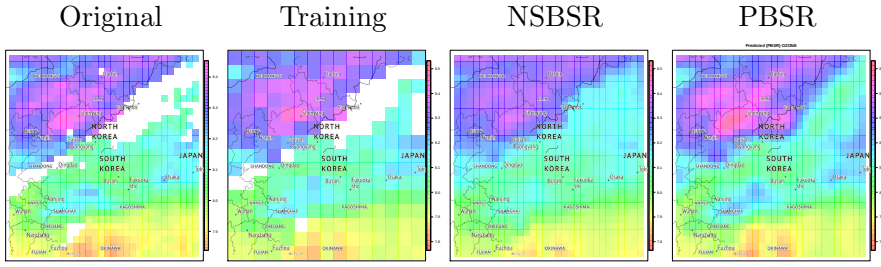

Figure 2.4: MODIS Result: The first plot is the MODIS dataset (1° resolution) (Original). The second plot is a training set (2° resolution) (Training). The third plot is prediction result from NSBSR and the fourth plot is prediction result from PBSR with a Matérn covariance function. The smoothness parameter in PBSR was estimated as well. (Longitude: $112° \sim 141°$; Latitude: $24° \sim 53°$; 30×30 pixels; Time: February 5, 2019)

First, we compare prediction result with a parametric approach by prediction map and computing MSPE and R^2 between observations and predicted values. Figure 2.4 shows prediction results for the ozone data. The first plot shows the original dataset, which we can see some missing values. The second plot shows the data we used to fit the model. The third and fourth plots are prediction maps from our approach (NSBSR) and parametric approach (PBSR). For PBSR, we consider a Matérn covariance model with unfixed smoothness parameter for model fitting. We can see that the prediction map from our approach shows similarity to the original data. Compared with the result by PBSR, R^2 of NSBSR ($R^2 = 0.54$) is higher than R^2 for

PBSR ($R^2 = 0.46$). MSPE of NSBSR (0.0793) is lower than MSPE of PBSR (0.0903).

Name	MAE	RMSE	INT
NNGP	0.2105	0.2392	0.8515
NSBSR	0.2105	0.2395	1.3775
LAGP	0.2135	0.2430	0.1614
SPDE	0.2139	0.2439	0.0705
Partition	0.2144	0.2436	0.2284
LK	0.2146	0.2425	0.1256
Gapfill	0.2147	0.2447	0.1347
FRK	0.2150	0.2409	0.4133
Periodic	0.2168	0.2457	0.4321
PP	0.2505	0.3033	1.2204
Tapering	0.2519	0.2926	2.8318

Table 2.4: Prediction results for the MODIS data from various methods based on mean absolute error (MAE), root mean squared error (RMSE) and confidence/credible interval length (INT). The results are sorted in ascending order by MAE.

Second, we compare prediction results with the methods in [29]. Table 2.4 shows performance results for the MODIS data. We provide mean absolute error (MAE), root mean squared error (RMSE) and the average length of confidence intervals (for Non-Bayesian approaches)/credible interval (for Bayesian approaches) (INT). The results are sorted in ascending order by MAE. Although most methods are comparable as the values of MAE and RMSE are similar, NSBSR provides the best result in terms of MAE (tied with NNGP) and

the second best result in terms of RMSE. NSBSR has relatively large INT, though. It is interesting that values of INT are widely different among the methods compared to MAE and RMSE, which indicates that interval estimation is more challenging.

AURA Application

The second dataset is total column ozone data from TOMS-Like Ozone and Radiative Cloud Fraction L3 1 day $0.25^\circ \times 0.25^\circ$ V3[3] by the Ozone Monitoring Instrument (OMI) onboard the AURA satellite. We consider the following covariates which can affect the level of ozone concentration as the ozone is a secondary pollutant: (1) Radiative Cloud Fraction [3]; (2) Solar Zenith Angle [3]; (3) Total column of Nitrogen Dioxide [4]; (4) Total column of Formaldehyde [5]; (5) Ultra Violet Aerosol Index [6]; (6) Total column of Sulfar Dioxide [7]. (3) and (4) are log-transformed for better interpretability. The achieved OMI/AURA dataset has global coverage in $1440 \times 720 = 14,641$ pixels with $0.25^\circ \times 0.25^\circ$ resolution. We restrict our study region to a neighborhood of the Korea peninsula, i.e. longitude ranged from 112° to 141°, latitude from 24° to 53°. For this time, we consider the seasonal-averaged data between June 1 to August 31, 2019. Note that there were missing values in hourly data due to satellite's orbits and other random sources but we averaged the data wherever available. As a result, there are no missing values for ozone concentration while some covariates still have some missing values. In this case, we imputed those

[3]DOI: *10.5067/Aura/OMI/DATA3002*
[4]DOI: *10.5067/Aura/OMI/DATA3007*
[5]DOI: *10.5067/Aura/OMI/DATA2016*
[6]DOI: *10.5067/Aura/OMI/DATA2025*
[7]DOI: *10.5067/Aura/OMI/DATA2025*

missing values by the ordinary kriging. Stationarity for the AURA data is also tested by [56]. The corresponding $p-$value is 0.672. So, we can not reject the non-stationarity for the AURA data based on this test as well.

For the AURA data, we further randomly removed 20% of the data to see the impact of missing values. Then, we took a subset at 0.5° resolution for fitting each method and prediction is made at 0.25° resolution. Table 2.5 shows the prediction performance for both daily and seasonal studies. Similar to Table 2.4, the results are sorted in ascending order by MAE. RMSE and INT are provided as well. NS-BSR provides the fifth best result in terms of MAE and RMSE. The ranks are not the same compared to the results for the MODIS data. For example, NNGP which show the best result for the MODIS data places the eighth for the AURA data. Variation in values of the prediction measures among the methods is more apparent compared to the MODIS data. The result by our approach is not as good as the one by the MODIS data. However, it shows reasonable output given that our approach is under stationary assumption while several methods allow non-stationarity. The values of INT vary more among the methods compared to MAE and RMSE, which is same as the MODIS data but the variation is more apparent. Given the variation in ranks based on performance measures for two datasets among the methods, our approach performs reasonably good.

Name	MAE	RMSE	INT
LK	0.937	1.303	34.868
SPDE	1.059	1.720	6.784
FRK	1.117	1.620	7.365
LAGP	1.305	1.841	5.315
NSBSR	1.502	1.901	13.344
Partition	1.642	2.898	3.352
Gapfill	2.454	4.192	4.419
NNGP	5.330	6.743	26.258
PP	5.883	7.130	7.431
Periodic	5.894	8.100	26.393
Tapering	7.713	9.560	6.348

Table 2.5: Prediction results for the AURA data from various methods based on mean absolute error (MAE), root mean squared error (RMSE) and confidence/credible interval length (INT). The results are sorted in ascending order by MAE.

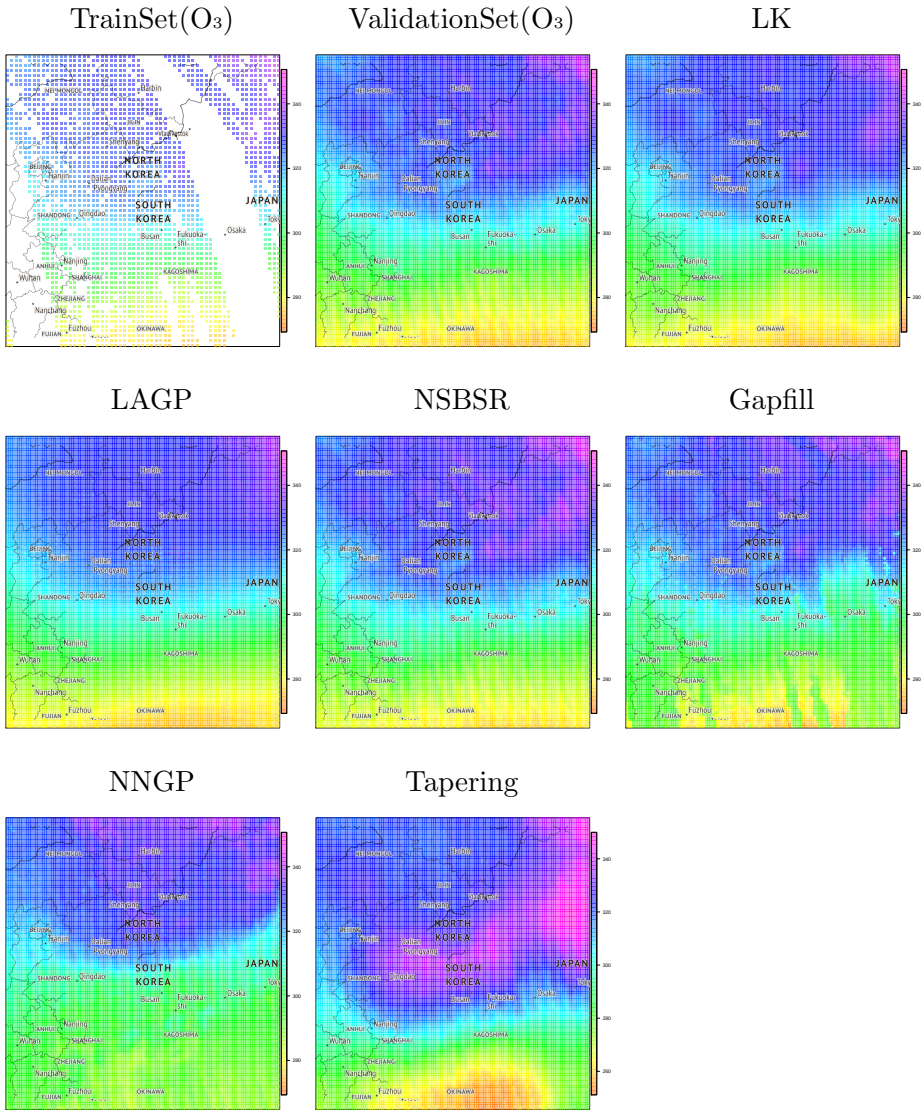

Figure 2.5: AURA Result (SEASONAL) : Cross-validation of the obtained AURA seasonal dataset (unit: 0.25°,0.25°) (Original). (Longitude: 112° $\sim$ 141°; Latitude: 24° $\sim$ 53°; 121 $\times$ 121 pixels; Time window: 2019/06/01/00 : 00 : 00 $\sim$ 2019/08/31/23 : 59 : 59)

Chapter 3

Bayesian time series regression using non-parametric modeling of Fourier coefficients

3.1 Preliminaries

Assuming 1.1, suppose that a dataset $(y_t, \boldsymbol{X}_t)$ is obtained, and consider the standardized error process $e_t = \sigma_\epsilon^{-1}(y_t - \boldsymbol{X}_t\beta)$, for $t = 1, \cdots, n$, of which autocorrelation function $\gamma(t)$ is defined on the discrete time domain. From eq.2.3, we consider a spectral density

$$\lambda(w) = \sum_{t=-\infty}^{+\infty} \gamma(t)\exp(-itw) \tag{3.1}$$

for $w \in [0, 2\pi)$, which is the discrete Fourier transform of γ. Assuming that e_t is invertible and γ is absolutely summable, according to eq.2.2,

we have

$$\gamma(t) = \frac{1}{2\pi}\int_0^{2\pi} \lambda(w)\exp(+itw)dw \tag{3.2}$$

for $t \in \mathcal{Z}$, which is the inverse Fourier transform of λ. Now, let's consider a discrete version of the inverse Fourier transform λ, i.e.

$$\gamma_n(t) = \frac{1}{n}\sum_{j=0}^{n-1} \lambda(w_j)\exp(+itw_j) \tag{3.3}$$

for $w_j \in W = \{w_j = \frac{2\pi j}{n} | j = 0, 1, \cdots, (n-1)\}$, which is so called Fourier frequencies. It is known that $\gamma_n(t) \to \gamma(t)$ as $n \to \infty$, for $t = 0, \cdots, (n-1)$. Since eq.3.3 is computationally easier than eq.3.2, we will approximate $\gamma(t)$ using $\gamma_n(t)$ for large n. Therefore, it suffices to model $\lambda(w)$ for the Fourier frequencies $w \in W$ and use Discrete Fourier Transform (DFT) operations in order to estimate an autocorrelation function γ of our interest.

We define a periodogram, a non-parametric estimate of the spectral density λ, which is defined as

$$\mathcal{I}_n(w) = \frac{1}{2n\pi}\left|\sum_{t=1}^{n} e_t\exp(-itw)\right|^2 \tag{3.4}$$

where $w \in [0, 2\pi)$. For the Fourier frequencies $w_j \in W$, $\boldsymbol{y} = (y_1, \cdots, y_n)'$, $\boldsymbol{e} = (e_1, \cdots, e_n)'$, and $\Gamma = var(\boldsymbol{e})$. Whittle(1954) showed that for n large,

$$\begin{aligned} p(\boldsymbol{y}|\lambda) &\propto |\Gamma|^{1/2}\exp(-\boldsymbol{y}'\Gamma^{-1}\boldsymbol{y}/2) \\ &\approx \prod_{j=1}^{n} \lambda(w_j)^{-1/2}\exp\left(-\sum_{j=1}^{n} \mathcal{I}(w_j)/2\lambda(w_j)\right) \end{aligned} \tag{3.5}$$

3.2 Proposed Method

From the Whittle likelihood(3.5) and Gaussian assumption on $\boldsymbol{y}$, $\mathcal{I}_n(w)$ is asymptotically exponentially distributed with mean $\lambda(w)$ and $\mathcal{I}_n$ are asymptotically independent at Fourier frequencies, $w_j = 2\pi j/n$, for $j = -\lfloor \frac{n}{2} \rfloor, \cdots, \lfloor \frac{n-1}{2} \rfloor$, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x. Carter and Kohn(1997) [7] utilized that the logarithm of a standard exponential random variable can be approximated by a mixture of five Gaussian random variables with known means and variances to bring the following approximated relationship:

$$\log \mathcal{I}_n(w) = \log \lambda(w) + \xi(w) \tag{3.6}$$

with ξ having distribution $\pi(\xi)$ such that

$$\pi(\xi) \propto \sum_{l=1}^{5} p_l \phi_{v_l}(\xi - k_l), \tag{3.7}$$

where $\phi_v(\cdot - k)$ is normal density with mean k and variance v^2. (k_l, v_l) for $l = 1, \cdots, 5$ are provided in Carter and Kohn (1997) [7].

We use the relation (3.6) at Fourier frequencies to bring an approximate likelihood for the spectral density in Bayesian inference. The original data model is $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\sigma}_\epsilon \odot \boldsymbol{e}$ in the time domain, where we express $\boldsymbol{\epsilon}$ as $\boldsymbol{\sigma}_\epsilon \odot \boldsymbol{e}$ with $\boldsymbol{\sigma}_\epsilon = (\sigma_{\epsilon,1}, \cdots, \sigma_{\epsilon,T})'$ and $\boldsymbol{e} = (e_1, \cdots, e_T)'$ to explicitly decompose the error into a time varying volatility and a normalized error process. The autocorrelation function $\gamma(h)$ for $\boldsymbol{e}$ is then implicitly considered in $\boldsymbol{\varphi} = \boldsymbol{\theta} + \boldsymbol{\xi}$ in the frequency domain with the spectral density, where $\boldsymbol{\varphi}$ is a vector of the logarithm of the periodogram at Fourier frequencies and $\boldsymbol{\theta}$ is a vector of the logarithm of spectral density at Fourier frequencies.

$\log(\sigma^2_{\epsilon,t}) = \eta_{\epsilon,t}$ is represented by cubic B-spline basis functions such that $\log(\sigma^2_{\epsilon,t}) = \eta_{\epsilon,t} = \sum_{b=1}^{d} \delta_b \phi_b(t)$. In this model setting, parameters are $\boldsymbol{\beta}$, $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_d)'$, $\boldsymbol{\theta}$ as well as latent label information for $\boldsymbol{\xi}$. Prior distributions for parameters are $\boldsymbol{\beta} \sim N(\mu_\beta \mathbf{1}, \sigma^2_\beta \boldsymbol{I})$, $\boldsymbol{\delta} \sim N(\mu_\alpha \mathbf{1}, \sigma^2_\alpha \boldsymbol{I})$, $\boldsymbol{\theta} \sim GP(\nu(\cdot), \tau(\cdot,\cdot))$ with mean function $\nu(\boldsymbol{w})$ and covariance kernel $\tau(\boldsymbol{w}_i, \boldsymbol{w}_j) = (1/\tau_0^2)\exp(-\rho\|w_i - w_j\|)$. For hyperparameters, $\rho \propto 1$ and $\tau_0 \sim G(a,b)$, where $G(a,b)$ is the Gamma distribution with mean ab.

Prediction at the kth-step ahead with the above model, for $1 \leq k \leq M$ is done in a following way. Let $\sigma^2_{\epsilon,T+k}$ be the (T+k)-th element of the volatility. Prediction of y_{T+k} is based on the estimated conditional expectation given the data, i.e. $\hat{y}_{T+k} = \hat{E}(y_{T+k}|\boldsymbol{y})$ With R samples from the posterior density, we approximate

$$\hat{y}_{T+k} = \frac{1}{R}\sum_{r=1}^{R}\left[X_{T+k}\hat{\boldsymbol{\beta}}^{(r)} + \left(e^{\hat{\boldsymbol{\eta}}^{(r)}_{\epsilon,T+k}}\right)^{\frac{1}{2}} \boldsymbol{h}^{(r)\prime}\tilde{\Gamma}^{(r)(-1)}_n\left(\boldsymbol{y} - X\hat{\boldsymbol{\beta}}^{(r)}\right)\right],$$

where subscript (r) refers to the r-th MCMC sample, $\boldsymbol{h} = \widehat{Cov}(\boldsymbol{e}, e_{T+k})$ and $\tilde{\Gamma}_n = \widehat{Cov}(\boldsymbol{e})$.

A matrix inverse $\tilde{\Gamma}_n^{-1}$ can be obtained efficiently as follows. Eq.3.3 implies that $\tilde{\Gamma}_n = Q_n \Lambda_n Q_n^*$ where $\Lambda_n = \mathrm{diag}[\lambda(w_0), \cdots, \lambda(w_{n-1})]$, Q_n is the $n \times n$ matrix with (u,v)-th entry being $q_{u,v} = \frac{1}{\sqrt{n}} e^{i(u-1)(v-1)2\pi/n}$, and Q_n^* is a complex conjugate matrix of Q_n'. Since Q_n is a unitary matrix, i.e. $Q_n Q_n^* = I$, we have $\tilde{\Gamma}_n^{-1} = Q_n \Lambda_n^{-1} Q_n^*$ where $\Lambda_n^{-1} = \mathrm{diag}[\lambda^{-1}(w_0), \cdots, \lambda^{-1}(w_{n-1})]$ [12].

Next, we investigate the details of an inverse matrix-vector multiplication $\tilde{\Gamma}_n^{-1}\boldsymbol{q}$, for arbitrary vector $\boldsymbol{q} \in R^n$. It is clear that $\tilde{\Gamma}_n$ is not only symmetric positive definite but also circulant matrix of order n in one-dimensional case. In addition, we have shown that the circulant

matrix $\tilde{\Gamma}_n$ can be diagonalized by FFT such that $Q_n^* \tilde{\Gamma}_n Q_n = \Lambda_n$, or equivalently $\tilde{\Gamma}_n = Q_n \Lambda_n Q_n^*$. Let $\boldsymbol{e}_1$ be the 1st column of the $n \times n$ identity matrix. Then we get $\Lambda_n \mathbf{1}_n = \sqrt{n} Q_n^* \tilde{\Gamma}_n \boldsymbol{e}_1$. This implies that the eigenvalues of Q_n can be obtained by performing one FFT on the first column of Q_n. Thus, the time cost of finding the eigenvalues of a circulant matrix is $\mathcal{O}(nlogn)$. We can write

$$\tilde{\Gamma}_n^{-1} \boldsymbol{q} = Q_n \Lambda_n^{-1} Q_n^* \boldsymbol{q} = Q_n \left(\Lambda_n^{-1} (Q_n^* \boldsymbol{q}) \right)$$

This means that $\tilde{\Gamma}_n^{-1} \boldsymbol{q}$ can be obtained by first computing an FFT on $\boldsymbol{q}$, then dividing the resulting vector by the eigenvalues of Q_n elementwise, and performing an inverse FFT on the resulting vector. (Anitescu *et al.*, 2012)[1]

3.3 Simulation Study

A simulated data set consists of three components: covariate ($\boldsymbol{X}$), error($\boldsymbol{\epsilon}$), and the outcome ($\boldsymbol{y}$). We consider two covariates, X_1 and X_2, so that $\boldsymbol{X} = (\mathbf{1}, X_1, X_2)$ has three attributes including intercept $\mathbf{1}$. We assume that X_1 follows a bivariate mixture Gaussian distribution such that $X_1 \sim 0.5N(0,1) + 0.5N(0,2)$ and X_2 follows the standard exponential distribution such that $X_2 \sim Exp(1)$. For $\boldsymbol{\epsilon} = \boldsymbol{\sigma}_\epsilon \odot \boldsymbol{e}$, we construct a total of 6 different models based on two cases of volatility $\sigma_{\epsilon,t}$ and three cases of the normalized error e_t. Specifically, we let $\sigma_{\epsilon,t}$ be (1) constant over time ($\sigma_{\epsilon,t}^2 \equiv 1$); (2) time-varying with a sinusoidal function, $\sigma_{\epsilon,t}^2 = 0.5 \sin\left(\frac{\pi t}{T}\right)$. For e_t, we consider (a) autoregressive model with order 2 [AR(2)]; (b) autoregressive-moving average model with order (1,1) [ARMA(1,1)]; and (c) autoregressive conditional heteroscedasticity model with order 1 [i.e. ARCH(1)]. When

$e_t \sim \text{AR}(2)$, we let the coefficients be $(a_1, a_2) = (0.5, -0.3)$, so that it satisfies $e_t - 0.5e_{t-1} + 0.3e_{t-2} = \xi z_t$, where $z_t \sim N(0,1)$, and ξ is a known scaling factor $\sqrt{\frac{1-a_2}{1+a_2}\frac{var(z_t)}{(1-a_2)^2-a_1^2}} \approx 1.14$ to match its marginal variance one, i.e. $var(e_t) = 1$. The corresponding spectral density is $\lambda(w) = 1/|1-0.5e^{i2\pi w}+0.3e^{i4\pi w}|^2$. When $e_t \sim \text{ARMA}(1,1)$, we set the coefficients to $(a_1, b_1) = (0.5, -0.6)$, so that $e_t - 0.5e_{t-1} = -0.6z_{t-1} + z_t$ with $\lambda(w) = 1|1 - 0.6e^{i2\pi w}|^2/|1 - 0.e^{i2\pi w}|^2$. When $\epsilon_t \sim \text{ARCH}(1)$, we let $(\alpha_0, \alpha_1) = (0.1, 0.9)$, so that it satisfies $\sigma^2_{\epsilon,t} = 0.1 + 0.9\epsilon^2_{t-1}$ and $e_t \sim N(0,1)$ with $\lambda(w) \equiv 1$. Additionally, we let $\boldsymbol{\beta}' = (\beta_0, \beta_1, \beta_2) = (1, 2, 3)$. Finally, $\boldsymbol{y}$ is generated from the model in (1.1).

We generate each data set with the length of 205, so that the sample size for *in-sample* estimation is set to $T = 200$. That is, the last five sample points are used for *out-of-sample* forecasting. We generate a total of 100 data sets for each case. For each data, the three MCMC chains with different initial values are run to generate 10,000 different sets of estimates. The results from the last 1,000 iterations are used for posterior analysis. We report the estimation and prediction results for each set of simulated data. Based on the parameter estimates, we compare our proposed model, the Bayesian time-varying Volatility model (BTV), to some alternative approaches: (1) Bayesian Autoregressive with order 1 [BAR(1)] (2) Bayesian AutoRegressive Conditional Heteroscedasticity with order 1 (BARCH(1)) and (3) Bayesian Fixed Volatility model using spectral density (BFV) proposed by Dey et al. (2018) [12]. Given its well-known usefulness in exchange rate forecasting, our forecasting experiment includes a random walk (RW) process in the pool of competing models.

DGP		Fitted Model	$\hat{\beta}_0$		$\hat{\beta}_1$		$\hat{\beta}_2$	
$\sigma^2_{\epsilon,t}$	e_t		Bias	RMSE	Bias	RMSE	Bias	RMSE
Fixed	AR(2)	BAR(1)	-.0656	.1219	-.0882	.2235	-.1864	.2168
		BARCH(1)	.0306	.1170	.0078	.0385	-.0141	.0826
		BFV	-.0109	.0925	.0022	.0277	-.0066	.0550
		BTV	-.0248	.0962	.0193	.0340	-.0047	.0582
	ARMA (1,1)	BAR(1)	-.0225	.0998	-.0693	.0768	-.1555	.1697
		BARCH(1)	.0152	.1243	.0076	.0373	-.0055	.0851
		BFV	-.0138	.0831	.0019	.0267	-.0037	.0631
		BTV	-.0284	.0886	.0192	.0333	-.0031	.0656
	ARCH(1)	BAR(1)	-.0122	.1168	-.0558	.0656	-.1380	.1506
		BARCH(1)	.0152	.1243	.0076	.0373	-.0055	.0851
		BFV	.0009	.0868	.0052	.0261	-.0130	.0531
		BTV	-.0190	.0614	.0217	.0293	-.0095	.0415
Time-Varying	AR(2)	BAR(1)	-.0684	.1256	-.0661	.0796	-.1832	.1976
		BARCH(1)	.0032	.1283	.0032	.0394	.0042	.0905
		BFV	.0080	.1042	.0003	.0291	-.0092	.0607
		BTV	-.0277	.0942	.0172	.0312	-.0059	.0581
	ARMA (1,1)	BAR(1)	-.0146	.1004	-.0724	.0794	-.1616	.1752
		BARCH(1)	-.0019	.1261	-.0001	.0403	.0035	.0878
		BFV	-.0112	.0918	-.0009	.0281	-.0080	.0677
		BTV	-.0312	.0899	.0174	.0308	-.0029	.0641
	ARCH(1)	BAR(1)	-.0117	.1155	-.0628	.0735	-.1582	.1752
		BARCH(1)	-.0019	.1261	-.0001	.0403	.0035	.0878
		BFV	.0041	.0954	.0057	.0303	-.0151	.0615
		BTV	-.0161	.0644	.0208	.0303	-.0135	.0461

Table 3.1: Bias and Root Mean Squared Error (RMSE) based on the estimates from 100 repeated sets of data for each simulation setting.

Table 3.1 shows bias and mean squared error (MSE) for the estimates of β_0, β_1 and β_2 under the six simulation settings. The first two columns specify the structures on the volatility $\sigma^2_{\epsilon,t}$ and on the normalized stationary error e_t. As discussed before, we let the volatility be either fixed ($\sigma^2_{\epsilon,t} \equiv 1$) or time-varying ($\sigma^2_{\epsilon,t} = 0.5 \sin\left(\frac{\pi t}{T}\right)$). Secondly, the underlying structure for the stationary error e_t is either AR(2), ARMA(1,1) or ARCH(1). For all possible combinations, we

compare the performance of our proposed approach (BTV) to those of the other three specifications (i.e. BAR(1), BARCH(1) and BFV).

Under the fixed volatility, BAR(1) produces a relatively large bias and a large RMSE, compared to the other specifications. In contrast, BARCH(1) generates a small bias but a relatively large RMSE than the others. In general, BFV provides both a small bias and a small RMSE, while BTV provides the smallest RMSE for the majority of time-varying volatility cases. Although it generally provides a larger bias for β_0, β_1 than both BFV and BARCH(1), the performance of BTV is consistently better in the majority of cases than that of BAR(1). In addition, it is interesting to see that the nonparametric modeling of spectral density (i.e. BFV, BTV) provides better outcomes than the parametric counterparts, even when the parametric specification matches the true data generating process.

For those with time-varying volatility, homoscedastic BAR(1) and BFV provide relatively large RMSEs compared to the time-invariant volatility ones. However, the BFV still provides a small bias compared to the other candidates. On the other hand, the BTV shows a relative advantage in terms of RMSE in the majority of simulation settings.

Figure 3.1 shows estimated log-scaled volatility $\sigma^2_{\epsilon,t}$ and log-scaled spectral density $\theta(w)$ from the BTV when the data are simulated from the time-varying volatility cases. The estimated log-scaled $\sigma^2_{\epsilon,t}$ is somewhat biased, especially near the boundary points. However, it does capture the global pattern in the time-varying characteristics of $\sigma^2_{\epsilon,t}$ in all three cases. Regarding the spectral density estimation in the bottom panel of Figure 3.1, the density curve estimates are reasonably close to the true ones in each case. In addition, the credible interval of the log-scaled spectral density covers the true curve of the log-scaled

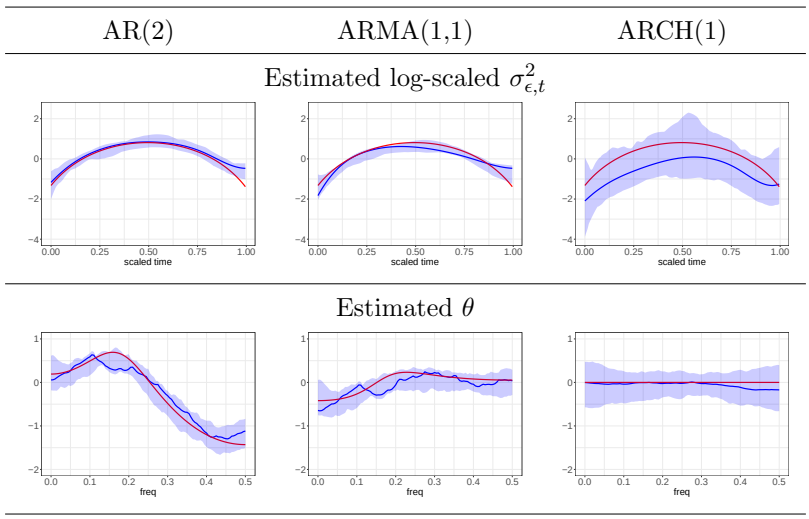

Figure 3.1: Estimated log-scaled volatility $\sigma^2_{\epsilon,t}$ (first row) and log-spectral density θ (second row) from BTV with 95% credible intervals from 100 repeated data. The first row shows the data generating processes. The x-axis in the first row is the time scaled to $(0, 1)$ and the x-axis in the second row is the frequency scaled to $(0, 0.5)$. The red line indicates the true curves, blue line indicates the estimated curves. The shaded regions indicate 95% credible intervals.

spectral density in all three cases.

In Figure 3.2, we compare the mean squared prediction errors (MSPE) of BTV to those from the other specifications at time $T+k$ for $k = 1, \cdots, 5$. In each figure, the black line refers to the MSPE of BTV, the red line refers to BAR(2), the purple line refers to BARCH(1,1) and the blue line refers to BFV. Although the true models have a time-invariant σ_ϵ (first column), MSPE of BTV is similar to those of the BFV and smaller than the other specifications. Quite intuitively, when true models have a time-varying σ_ϵ, then BTV generally shows

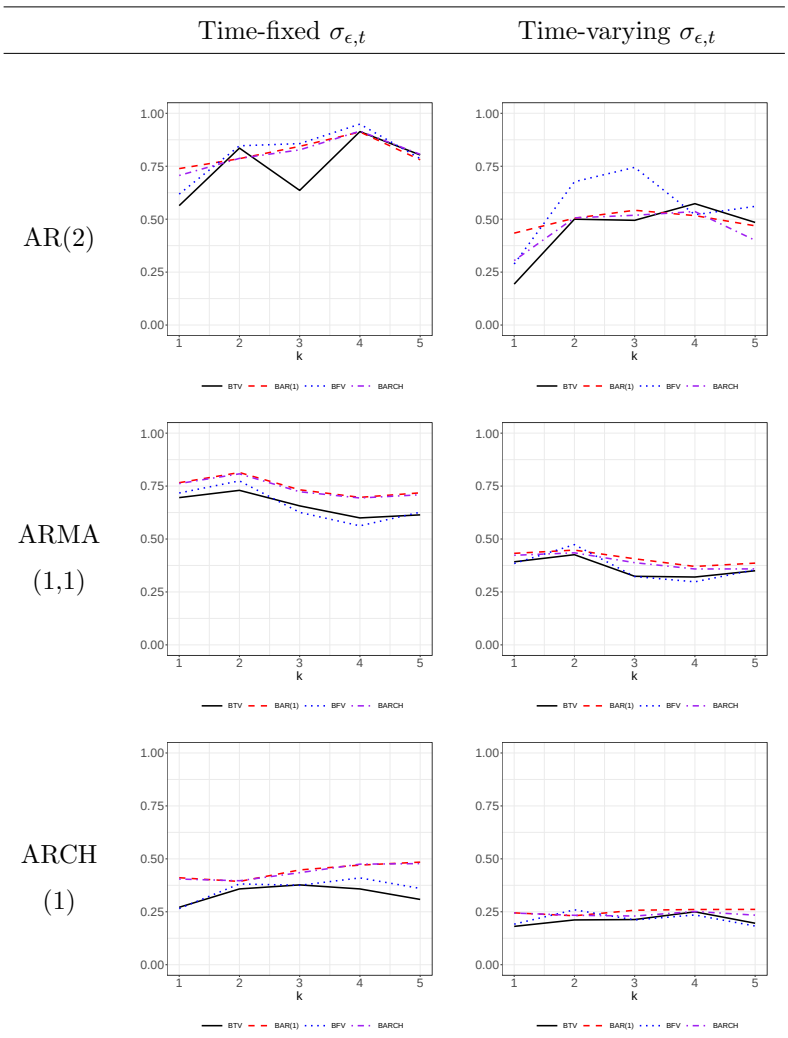

Figure 3.2: The estimated mean squared prediction errors for y_{n+k}, $k = 1, 2, 3, 4, 5$. The model is fitted up to time T=200 and then forecast y_{T+k}, $k = 1, 2, 3, 4, 5$. We compare the proposed model, BTV(black), to the other models: BAR(1)(red); BARCH(1,1)(purple); and BFV(blue)

a lower MSPE than the rest of specifications. However, this relative advantage is rather minor and irregular. Interestingly, the BFV and the BTV behave in a rather similar fashion, although the BFV has a higher MSE than the BTV in the majority of the cases.

3.4 Real Data Analysis

As an application of the proposed methodology, we consider the forward premium regression model introduced by Mark (1995) [43]:

$$s_{t+k} - s_t = \alpha + \beta(f_t - s_t) + u_{t+k}, \tag{3.8}$$

where s_t is log of spot exchange rate at time t, s_{t+k} is log of spot exchange rate at time $t+k$. One can obtain the kth-step-ahead forecast of s_t from the model. Here f_t is the fundamental variable defined by:

$$f_t := m_t - m_t^* - (y_t - y_t^*)$$

where m_t is log of *U.S.* money supply and y_t is log of *U.S.* output, while m_t^* and y_t^* are their foreign equivalents.

This study employs spot exchange rate data and their corresponding macroeconomic data, involving eight currencies of Australian Dollar (AUD), Canadian Dollar (CAD), Swiss Franc (CHF), Danish Krone (DKK), British Pound (GBP), Japanese Yen (JPY), New Zealand Dollar (NZD) vis a vis US Dollar (USD), which is the numeraire currency. We use end-of-month observations from December 1988 through April 2017, including the 2008-2009 financial crisis period.

We have a total of $n = 290$ observations (December 1998 - January 2013) at a monthly frequency for each currency. Note that DKK has

missing values in m_t from January to March 2001, and we thus impute these values by the sample average of m_t in the neighborhood. We start with the first $T = 270$ observations (December 1998 - May 2011) to fit the model. That is, the estimation window length is 270. The remaining $n-T = 290-270 = 20$ observations are used to evaluate the model's *out-of-sample* forecasting performance. We try four different forecast horizons, i.e. $k = 1$ (monthly), $k = 3$ (quarterly), $k = 6$ (half yearly), and $k = 12$ (yearly). Given the starting point l, we conduct k-step-ahead forecasting as follows. That is, we forecast y_{l+T+k} based on the model fitted using the observations from $(l+1)$-th to $(l+T)$-th time points. We move l from 0 to 5 for the four forecast horizons in each currency.

We compare the forecasting performance of our proposed method to those of a pack of competing models, such as a random walk (RW), Bayesian Autoregressive model (BAR(1)), Bayesian Autoregressive Conditional Heteroscedasticity model (BARCH(1)) and the Bayesian fixed volatility model using the spectral density (BFV). Throughout the Bayesian $MCMC$ implementation, we run 3 chains with 10,000 iterations per chain. The last 1,000 iterations of the 3 chains are saved for estimation and inference.

Table 3.2 illustrates the results based on two forecast error measures for the spot return. The numbers represent the mean square of prediction errors (MSPE) for the forecasts, while those inside brackets are the proportions that the corresponding model beats the random walk process in terms of the forecast error. We denote the proportion by RWr. That is, a high RWr implies that the corresponding model performs significantly better than the random walk.

Empirical results indicate that RMSPE tends to increase as the

forecast horizon k grows. BTV shows relatively low RMSPE values among the competing statistical methods for the majority of currencies; CAD, CHF, DKK, JPY, and NZD. Especially, in the short-term prediction, such as one-month-ahead forecasting ($k = 1$), the BTV outperforms both the simple RW and the other conventional time-series regression methods for the majority of currencies. The results show that the RMSPE for the BTV is mostly either the lowest or the second lowest among all the competing ones. They also show that, except for GBP, RWr of the BTV is more than or equal to 50.0%. Rwr of the BTV remains relatively high in most cases. However, we also note that there are some cases where the BVT's performance in terms of RWr is not as good as that in RMSPE. For instance, in seasonal forecasting ($k = 3$) for CAD, the BTV has a lower RMSPE than a RW on average during the study period, but its RWr is lower than 50.0%. Another example is that, in one-year-ahead forecasting ($k = 12$) for DKK, the BTV has a lower RMSPE on average during the study period, but its RWr is lower than that of BAR(1).

Our result indicates that the nonparamteric spectral method can possibly improve the forecast accuracy since both BFV and BTV provide lower RMSPEs and higher RWrs than the other alternatives in many cases. The BFV and BTV do not rely on any parametric covariance structure, while the BAR(1) and BARCH(1) depend on such an assumption, which could lead to model misspecification. However, it is hard to conclude which method is truly superior because the prediction gain by a BARCH(1) could be potentially due to the model's over-fitting. Our results show that the effect is often non-negligible.

Table 3.3 illustrates the forecast outcomes for the spot exchange rate itself, not the log of it. As we can see from the table, the BTV

still shows a relatively low RMSPE and a high RWr for CAD, CHF, DKK, and JPY, among the competing specifications. Note that in short-term predictions, such as one-month-ahead forecasting ($k = 1$), the BTV generally outperforms both a RW and other conventional time-series methods.

Currency	k	RW	BAR(1)	BARCH(1)	BFV	BTV
AUD	1	0.0249 (.)	0.0586 (16.7)	0.0254 (50.0)	**0.0238** (50.0)	**0.0229** (50.0)
	3	0.0448 (.)	**0.0406** (66.7)	**0.0263** (83.3)	0.0463 (33.3)	0.0428 (50.0)
	6	0.0646 (.)	0.0406 (83.3)	**0.0326** (66.7)	0.0423 (66.7)	**0.0313** (66.7)
	12	0.0529 (.)	0.0481 (66.7)	**0.0406** (50.0)	**0.0448** (83.3)	0.0470 (100.0)
CAD	1	**0.0151** (.)	0.0447 (0.0)	0.0523 (0.0)	0.0156 (33.3)	**0.0085** (50.0)
	3	**0.0428** (.)	0.0763 (33.3)	0.0496 (33.3)	0.0445 (50.0)	**0.0393** (33.3)
	6	0.0544 (.)	0.0983 (16.7)	**0.0385** (66.7)	0.0679 (16.7)	**0.0498** (83.3)
	12	0.0917 (.)	0.0645 (50.0)	**0.0254** (50.0)	**0.0247** (66.7)	0.0272 (50.0)
CHF	1	**0.0275** (.)	0.0516 (16.7)	0.0398 (50.0)	0.0542 (16.7)	**0.0145** (66.7)
	3	0.1156 (.)	0.0856 (66.7)	0.2364 (16.7)	**0.0427** (83.3)	**0.0578** (100.0)
	6	0.1917 (.)	0.0639 (66.7)	0.0627 (83.3)	**0.0626** (83.3)	**0.0433** (100.0)
	12	0.2081 (.)	**0.0927** (83.3)	0.1866 (50.0)	0.1357 (83.3)	**0.0855** (83.3)
DKK	1	0.0413 (.)	0.0409 (66.7)	0.0479 (66.7)	**0.0251** (50.0)	**0.0236** (66.7)
	3	0.0530 (.)	0.0598 (50.0)	0.0755 (16.7)	**0.0455** (66.7)	**0.0414** (66.7)
	6	0.1138 (.)	**0.0642** (83.3)	0.1276 (33.3)	0.1548 (50.0)	**0.0520** (100.0)
	12	0.1317 (.)	**0.0598** (83.3)	0.1219 (50.0)	0.0843 (50.0)	**0.0516** (50.0)
GBP	1	0.0260 (.)	0.0447 (16.7)	0.0282 (16.7)	**0.0119** (33.3)	**0.0145** (33.3)
	3	0.0574 (.)	0.0613 (16.7)	**0.0440** (50.0)	**0.0366** (50.0)	0.0493 (33.3)
	6	**0.0261** (.)	0.0425 (33.3)	0.0555 (0.0)	0.0666 (50.0)	**0.0387** (50.0)
	12	0.0538 (.)	**0.0530** (33.3)	**0.0525** (33.3)	0.0874 (16.7)	0.0840 (33.3)
JPY	1	0.0300 (.)	0.0542 (33.3)	**0.0220** (50.0)	0.0634 (0.0)	**0.0144** (83.3)
	3	**0.0289** (.)	0.0838 (16.7)	0.0794 (0.0)	0.0430 (16.7)	**0.0369** (33.3)
	6	**0.0404** (.)	0.0838 (50.0)	0.0580 (50.0)	0.0555 (0.0)	**0.0323** (66.7)
	12	0.0888 (.)	0.0568 (66.7)	**0.0312** (100.0)	0.1782 (0.0)	**0.0492** (100.0)
NZD	1	0.0312 (.)	0.0358 (33.3)	**0.0263** (66.7)	0.0298 (50.0)	**0.0220** (66.7)
	3	0.0989 (.)	0.0655 (66.7)	**0.0550** (66.7)	0.0571 (66.7)	**0.0535** (83.3)
	6	0.1036 (.)	0.0655 (83.3)	**0.0437** (83.3)	0.0527 (83.3)	**0.0392** (83.3)
	12	0.1440 (.)	0.0620 (60.0)	0.0717 (80.0)	**0.0518** (80.0)	**0.0554** (100.0)

Table 3.2: Root mean square prediction error (RMSPE) of log difference in spot exchange rate ($s_{t+k} - s_t$) for $k = 1, 3, 6, 12$-step ahead forecasting from the proposed model (BTV) and other models (RW, BAR(1), BARCH(1), BFV) assuming exchange rate model by Mark (1995). The value in the parenthesis (RWr) is the proportion of the number of time points where the prediction error of the target method is less than that of the random walk over the total number of time points investigated.

Currency	k	RW	BAR(1)	BARCH(1)	BFV	BTV
AUD	1	**0.1495** (.)	0.3753 (0.0)	0.1931 (33.3)	**0.1771** (33.3)	0.2510 (0.0)
	3	0.2604 (.)	0.3706 (33.3)	**0.1621** (83.3)	0.2434 (66.7)	**0.2233** (66.7)
	6	0.4347 (.)	0.3706 (83.3)	**0.1963** (100.0)	0.3194 (83.3)	**0.2573** (100.0)
	12	0.3171 (.)	**0.1997** (100.0)	0.3086 (50.0)	**0.2183** (100.0)	0.2973 (66.7)
CAD	1	**0.1050** (.)	0.2105 (16.7)	0.4127 (0.0)	0.2213 (0.0)	**0.0797** (50.0)
	3	**0.1932** (.)	0.3695 (0.0)	0.4268 (0.0)	0.3151 (0.0)	**0.1888** (50.0)
	6	**0.2711** (.)	0.3418 (33.3)	**0.2543** (33.3)	0.3289 (33.3)	**0.2410** (66.7)
	12	0.5545 (.)	0.3331 (100.0)	**0.2279** (100.0)	0.2764 (100.0)	**0.2597** (100.0)
CHF	1	**0.2268** (.)	0.3799 (0.0)	0.3673 (0.0)	0.2703 (33.3)	**0.1885** (66.7)
	3	0.5727 (.)	0.5108 (83.3)	0.9018 (0.0)	**0.3478** (100.0)	**0.3568** (100.0)
	6	0.6774 (.)	0.4630 (66.7)	0.4814 (66.7)	**0.3770** (66.7)	**0.2519** (100.0)
	12	0.6764 (.)	0.5296 (66.7)	0.7213 (16.7)	**0.4980** (100.0)	**0.3282** (66.7)
DKK	1	0.3441 (.)	0.3132 (50.0)	0.4063 (16.7)	**0.1463** (100.0)	**0.2598** (100.0)
	3	0.3838 (.)	0.3949 (33.3)	0.4102 (33.3)	**0.2975** (50.0)	**0.3451** (50.0)
	6	0.6269 (.)	**0.3865** (100.0)	0.5935 (83.3)	0.6494 (33.3)	**0.3335** (100.0)
	12	**0.3832** (.)	0.3949 (66.7)	0.5840 (16.7)	0.5341 (0.0)	**0.2019** (83.3)
GBP	1	**0.1329** (.)	0.2376 (0.0)	0.2893 (0.0)	0.1629 (33.3)	**0.0952** (83.3)
	3	0.3247 (.)	**0.2433** (83.3)	0.3351 (33.3)	0.3226 (50.0)	**0.2139** (100.0)
	6	0.2990 (.)	**0.2013** (66.7)	0.4235 (0.0)	0.4608 (16.7)	**0.2888** (50.0)
	12	**0.3590** (.)	**0.2828** (83.3)	0.4410 (33.3)	0.4379 (0.0)	0.4019 (33.3)
JPY	1	0.2347 (.)	0.4169 (0.0)	**0.2368** (16.7)	0.4875 (0.0)	**0.0923** (100.0)
	3	**0.2983** (.)	0.5328 (0.0)	0.5452 (0.0)	0.3345 (0.0)	**0.1822** (100.0)
	6	**0.2391** (.)	0.5328 (16.7)	0.4689 (0.0)	0.4085 (0.0)	**0.2903** (0.0)
	12	0.4874 (.)	0.4204 (50.0)	**0.2212** (100.0)	0.8025 (0.0)	**0.3309** (100.0)
NZD	1	0.1760 (.)	0.2275 (33.3)	**0.1395** (66.7)	0.2820 (16.7)	**0.1310** (66.7)
	3	0.5406 (.)	0.3408 (83.3)	0.3878 (83.3)	**0.2756** (83.3)	**0.2730** (83.3)
	6	0.4711 (.)	0.3408 (0.0)	**0.2901** (66.7)	0.3736 (66.7)	**0.1730** (66.7)
	12	0.6613 (.)	**0.3912** (100.0)	0.4830 (100.0)	**0.3387** (100.0)	0.4058 (100.0)

Table 3.3: Root mean absolute prediction error (RMAPE) of spot exchange rate (e^{s_t}) for $k = 1, 3, 6, 12$-step ahead forecasting from the proposed model (BTV) and other models (RW, BAR(1), BARCH(1), BFV) assuming exchange rate model by Mark (1995). The value in the parenthesis (RWr) is the proportion of the number of time points where the prediction error of the target method is less than that of the random walk over the total number of time points investigated.

Figure 3.3: Radar charts of root mean square prediction error of spot exchange rate(e^{s_t}) for the five competing models and the four selected case studies.: Canadian Dollar (CAD, top left); Switzerland Franc (CHF, top right); Danish Krone (DKK, bottom left); and New Zealand Dollar (NZD, bottom right). The line colors represent 1-month forecasting (blue), 3-month forecasting (orange), 6-month forecasting (gray), and 12-month forecasting (yellow)

Chapter 4

Concluding remarks

Bayesian spatial regression using non-parametric modeling of Fourier coefficients

We have proposed Bayesian spatial regression with non-parametric modeling of spectral density derived from Fourier Transform. Our approach, NSBSR, has achieved reasonable computational efficiency in terms of storage and speed by using the Whittle likelihood approximation and the Fast Fourier Transform algorithm, even though there are more parameters to estimate compared to parametric covariance models. Simulation studies show that NSBSR is relatively robust compared to parametric covariance models and/or isotropic assumption in spatial prediction. Also, NSBSR shows better prediction results in a sense that RMSPE is lower than those of parametric covariance models for smoother processes. Our approach requires stationary assumption, which is rather limited given that several methods to handle non-stationary spatial data are available. However, the results in comparison analysis (Tables 2.4 and 2.5) using two different ozone con-

centration datasets show that our approach can provide reasonable prediction given the variation of prediction accuracy among methods for different datasets. These results suggest that NSBSR is a good alternative to the existing prediction approaches in spatial data analysis.

Our approach could be used as a baseline for capturing a more complicated spatial dependence structure than that of stationary Gaussian fields. We can consider the marginal variance σ_ϵ in the equation (2.1) to be $\sigma_\epsilon(s)$ so that it is spatially varying. The resulting process becomes non-stationary. We could apply our approach to a stationary error process component, $e(\boldsymbol{s})$, so that we can handle a class of non-stationary processes.

Estimated spectral densities are not as good as the prediction results. It could be due to the DFT approximation or Whittle likelihood approximation with further approximation using a five-components mixture Gaussian. [59] pointed out that DFT approximation rather than the exact Fourier transform could cause accuracy issue on the autocorrelation estimation. In addition, insufficient sample size or truncated study region could possibly have a negative effect on the spectral density estimation. A possible remedy would be a different likelihood approximation than the Whittle likelihood so that we can avoid using periodogram itself but it requires theoretical justification. An empirical choice of hyperparameters such as ρ_0 for ρ_{θ_1}, ρ_{θ_2} might be beneficial to enhance prediction accuracy, as well. However, these are rather subjective and we tried usual practice of vague priors in our analysis.

Empirical results for multiple AURA datasets by different types of cross-validation procedure have been investigated and they show potential possibilities of better prediction accuracy of our proposed

model than what is reported in section 2.5. Both grid sampling of ρ and conditional sampling of unobserved variable of interest are to be improved to achieve better computing efficiency.

Bayesian time series regression using non-parametric modeling of Fourier coefficients

We have proposed Bayesian time series regression with non-parametric modeling of spectral density to detect heteroscedastic one dimensional stochastic process such as exchange rate. The estimates are applied for prediction of the model dependent variable. The main motivation is to avoid any arbitrary parametric structure on the error in estimating the corresponding model and in forecasting its dependent variable. In addition, one can avoid the notorious issue of nonparametric bandwidth selection by conducting the estimation in a frequency domain. The relevant Bayesian posterior distributions are derived and the estimation and prediction strategy is explained in detail. Our approach, NSBTR, is computationally efficient with the help of Whittle likelihood approximation and Fast Fourier Transform algorithm, even though there are much more parameters to estimate compared to parametric covariance models. Also, the use of B-spline basis function representation enable us to achieve comparable prediction accuracy to the conventional heteroscedastic models such as $ARCH$ or $GARCH$ models. Moreover, NSBTR often shows better prediction accuracy than the conventional ones in that it usually shows lower mean square prediction errors at least in overall time periods. These finding implies that NSBTR is a good alternative to the existing forecasting approaches in financial time-series data analysis.

The current project leaves a couple of interesting issues that deserve future research on: Firstly, the current framework can be extended to accommodate *time-varying coefficients*. To focus on the role of time-varying volatility, our study assumes that the model coefficients are invariant over time. One can generalize this restriction to make our findings more robust. Secondly, we can apply the proposed methodology to *density forecasting*. By forecasting the entire distribution of the dependent variable, one can forecast not only the mean but also any percentile of the underlying distribution. This would result in more comprehensive forecasts, such as interval forecasts, for the corresponding variable. Further insight can be gained by extending the current work in these and other directions.

Acknowledgements

This work is supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIT)
(No. 2016R1A2B4008237, No. 2019R1A2C1002213).

Bibliography

[1] Mihai Anitescu, Jie Chen, and Lei Wang. A matrix-free approach for solving the parametric gaussian process maximum likelihood problem. *SIAM Journal on Scientific Computing*, 34(1):A240–A262, 2012.

[2] Erlend Aune, Daniel P Simpson, and Jo Eidsvik. Parameter estimation in high dimensional gaussian distributions. *Statistics and Computing*, 24(2):247–263, 2014.

[3] Sudipto Banerjee, Alan E Gelfand, Andrew O Finley, and Huiyan Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848, 2008.

[4] GD Bergland. A guided tour of the fast fourier transform. *IEEE spectrum*, 6(7):41–52, 1969.

[5] Ronald Newbold Bracewell and Ronald N Bracewell. *The Fourier transform and its applications*, volume 31999. McGraw-Hill New York, 1986.

[6] David R Brillinger. *Time series: data analysis and theory*, volume 36. Siam, 1981.

[7] Christopher K Carter and Robert Kohn. Semiparametric bayesian inference for time series with mixed spectra. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(1):255–268, 1997.

[8] Nidhan Choudhuri, Subhashis Ghosal, and Anindya Roy. Bayesian estimation of the spectral density of a time series. *Journal of the American Statistical Association*, 99(468):1050–1059, 2004.

[9] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

[10] Noel Cressie. Statistics for spatial data. *Terra Nova*, 4(5):613–617, 1992.

[11] 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.

[12] Tanujit Dey, Kun Ho Kim, and Chae Young Lim. Bayesian time series regression with nonparametric modeling of autocorrelation. *Computational Statistics*, pages 1–17, 2018.

[13] J. Du, H. Zhang, and V. S. Mandrekar. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *The Annals of Statistics*, 37:3330–3361, 2009.

[14] C Engel. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987–1007, 1982.

[15] C Engel and J D Hamilton. Long swings in the dollar: Are they in the data and do markets know it? *American Economic Review*, 80:689–713, 1990.

[16] C Engel, N C Mark, and K D West. Exchange rate models are not as bad as you think. *NBER Macroeconomics Annual 2007*, pages 381–441, 2008.

[17] Andrew O. Finley, Huiyan Sang, Sudipto Banerjee, and Alan E. Gelfand. Improving the performance of predictive process modeling for large datasets. *Computational statistics & data analysis*, 53:2873–2884, 2009.

[18] Genton Marc G. Furrer, Reinhard and Douglas Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15:502–523, 2006.

[19] Alan E Gelfand, Peter Diggle, Peter Guttorp, and Montserrat Fuentes. *Handbook of spatial statistics*. CRC press, 2010.

[20] Furrer R. Schaepman-Strub G. de Jong R. Gerber, F. and M. E. Schaepman. Predicting missing values in spatio-temporal satellite data. *IEEE Transactions on Geoscience and Remote Sensing*, 56:2841–2853, 2018.

[21] Irving John Good. Introduction to cooley and tukey (1965) an algorithm for the machine calculation of complex fourier series. In *Breakthroughs in Statistics*, pages 201–216. Springer, 1997.

[22] R. Gramacy and D. Apley. Local gaussian process approximation for large computer experiments. *Journal of Computational and Graphical Statistics*, 24:561–578, 2015.

[23] R. Guhaniyogi and S. Banerjee. Meta-kriging: Scalable bayesian modeling and inference for massive spatial datasets. *Technometrics*, 60:430–444, 2018.

[24] Joseph Guinness. Nonparametric spectral methods for multivariate spatial and spatial-temporal data. *arXiv preprint arXiv:1811.01280*, 2018.

[25] Joseph Guinness. Spectral density estimation for random fields via periodic embeddings. *Biometrika*, 106(2):267–286, 2019.

[26] Joseph Guinness and Montserrat Fuentes. Circulant embedding of approximate covariances for inference from gaussian data on large lattices. *Journal of computational and Graphical Statistics*, 26(1):88–97, 2017.

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

[28] M. J. Heaton, W. F. Christensen, and M. A. Terres. Nonstationary gaussian process models using spatial hierarchical clustering from finite differences. *Technometrics*, 59:93–101, 2017.

[29] Matthew J Heaton, Abhirup Datta, Andrew O Finley, Reinhard Furrer, Joseph Guinness, Rajarshi Guhaniyogi, Florian Gerber, Robert B Gramacy, Dorit Hammerling, Matthias Katzfuss, et al. A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, pages 1–28, 2018.

[30] O Ince, T Molodtsova, and D H Papell. Taylor rule deviations

and out-of-sample exchange rate predictability. *Journal of International Money and Finance*, 69:22–44, 2016.

[31] Matthias Katzfuss. A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association*, 112:201–214, 2017.

[32] Matthias Katzfuss and Wenlong Gong. Multi-resolution approximations of gaussian processes for large spatial datasets. *arXiv preprint arXiv:1710.08976*, 2017.

[33] C. G. Kaufman, M. J. Schervish, and D.W. Nychka. Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103:1545–1555, 2008.

[34] L Kilian. Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions? *Journal of Applied Econometrics*, 14:491–510, 1999.

[35] L Kilian and M P Taylor. Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, 60:85–107, 2003.

[36] H.-M. Kim, B. K. Mallick, and C. Holmes. Analyzing nonstationary spatial data using piecewise gaussian processes. *Journal of the American Statistical Association*, 100:653–668, 2005.

[37] Jae Kwang Kim and Jun Shao. *Statistical methods for handling incomplete data*. CRC press, 2013.

[38] K H Kim and T Kim. Capital asset pricing model: A time-varying volatility approach. *Journal of Empirical Finance*, 37:268–281, 2016.

[39] S. Y. Kim and I. Song. National-scale exposure prediction for long-term concentrations of particulate matter and nitrogen dioxide in south korea. *Environmental Pollution*, 226:21–29, 2017.

[40] B. A. Konomi, H. Sang, and B. K. Mallick. Adaptive bayesian nonstationary modeling for large spatial datasets using covariance approximations. *Journal of Computational and Graphical Statistics*, 23:802–829, 2014.

[41] F. Laden, J. Schwartz, F. E. Speizer, and D. W. Dockery. Reduction in fine particulate air pollution and mortality: extended follow-up of the harvard six cities study. *American journal of respiratory and critical care medicine*, 173(6):667–672, 2006.

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

[43] N C Mark. Exchange rates and fundamentals: Evidence on long-horizon predictability. *American Economic Review*, 85:201–218, 1995.

[44] R A Meese and K Rogoff. Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of International Economics*, 14:3–24, 1983.

[45] T Molodtsova and D H Papell. Out-of-sample exchange rate predictability with taylor rule fundamentals. *Journal of International Economics*, 77:167–180, 2009.

[46] Douglas Nychka, Soutir Bandyopadhyay, Dorit Hammerling, Finn Lindgren, and Stephan Sain. A multiresolution gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, 24:579–599, 2015.

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

[48] Christopher J Paciorek. Computational techniques for spatial logistic regression with large data sets. *Computational Statistics & Data Analysis*, 51:3631–3653, 2007.

[49] Maurice Bertram Priestley. *Spectral analysis and time series*. Academic press, 1981.

[50] Brian J Reich and Montserrat Fuentes. Nonparametric bayesian models for a spatial covariance. *Statistical methodology*, 9(1-2):265–274, 2012.

[51] Andrew J Royle and Cristopher K Wikle. Efficient statistical mapping of avian count data. *Environmental and Ecological Statistics*, 12:225–243, 2005.

[52] H. Sang, M. Jun, and J. Z. Huang. Covariance approximation for large multivariate spatial data sets with an application to multiple climate model errors. *The Annals of Applied Statistics*, 5:2519–2548, 2011.

[53] Michael L Stein. A modeling approach for large spatial datasets. *Journal of the Korean Statistical Society*, 37(1):3–10, 2008.

[54] Michael L Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.

[55] Jonathan R Stroud, Michael L Stein, and Shaun Lysen. Bayesian and maximum likelihood estimation for gaussian processes on an incomplete lattice. *Journal of computational and Graphical Statistics*, 26(1):108–120, 2017.

[56] Sarah L Taylor, Idris A Eckley, and Matthew A Nunes. A test of stationarity for textured images. *Technometrics*, 56:291–301, 2014.

[57] Peter Whittle. On stationary processes in the plane. *Biometrika*, pages 434–449, 1954.

[58] J H Wright. Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics*, 146:329–341, 2008.

[59] Akira Moiseevich Yaglom. *Correlation theory of stationary and related random functions: Supplementary notes and references*. Springer Science & Business Media, 2012.

Appendix A

Conditional posterior distributions

Conditional posterior distribution of parameters for Gibbs samplers are described below.

Posterior of $\boldsymbol{\beta}$

With a Gaussian prior $\boldsymbol{\beta} \sim \boldsymbol{N}(\mu_\beta \mathbf{1}, \sigma_\beta^2 \boldsymbol{I})$, we have

$$\begin{aligned} \boldsymbol{\beta} \mid \cdots &\sim \boldsymbol{N}(\boldsymbol{\mu}_*, \boldsymbol{T}_*), \\ \boldsymbol{T}_* &= \boldsymbol{X}^t \tilde{\boldsymbol{\Gamma}}^{-1} \boldsymbol{X} + \sigma_\beta^{-2} \boldsymbol{I}_p, \\ \boldsymbol{\mu}_* &= \boldsymbol{T}_*^{-1} \left(\boldsymbol{X}^t \tilde{\boldsymbol{\Gamma}}^{-1} \boldsymbol{Y} + \sigma_\beta^{-2} \mu_\beta \boldsymbol{I} \right), \end{aligned}$$

where $\tilde{\boldsymbol{\Gamma}}$ is the covariance matrix of the data and it is constructed by following. Let

$$\tilde{P}_{n_1 n_2}(u_1, u_2) = \sum_{j=0}^{n_1 - 1} \sum_{k=0}^{n_2 - 1} f_\Delta(w_j, w_k) \exp(\iota(w_j u_1 + w_k u_2))$$

for $w_j = -\frac{\pi}{\delta_1} + \frac{2\pi}{\delta_1} \frac{j}{n_1}$, $w_k = -\frac{\pi}{\delta_2} + \frac{2\pi}{\delta_2} \frac{k}{n_2}$. We construct the circulant matrix $\tilde{B}_{n_1}^{(k)} = Circulant(\tilde{P}_{n_1 n_2}^{(k)})$, where $\tilde{P}_{n_1 n_2}^{(k)}$ is the k-th column

vector of $\tilde{P}_{n_1 n_2}$. Then,

$$\tilde{\boldsymbol{\Gamma}} = \begin{bmatrix} \tilde{B}_{n_1}^{(1)} & \tilde{B}_{n_1}^{(2)} & \dots & \tilde{B}_{n_1}^{(n_2)} \\ \tilde{B}_{n_1}^{(n_2)} & \tilde{B}_{n_1}^{(1)} & \dots & \tilde{B}_{n_1}^{(n_2-1)} \\ \vdots & \ddots & \ddots & \vdots \\ \tilde{B}_{n_1}^{(2)} & \dots & \tilde{B}_{n_1}^{(n_2)} & \tilde{B}_{n_1}^{(1)} \end{bmatrix}.$$

Instead of computing $\tilde{\boldsymbol{\Gamma}}^{-1}$, we actually need to compute $\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{X}$ and $\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{Y}$. We adopt the method by [1] so that we can reduce the computation of both $\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{X}$ and $\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{Y}$ to $\mathcal{O}(n_1 n_2 log(n_1 n_2))$. The approach by [1] is as follows. Since $\tilde{\boldsymbol{\Gamma}}$ is a $n_1 n_2 \times n_1 n_2$ block circulant of circulant block (BCCB) matrix, it can be diagonalized by the Kronecker product of 2D FFT matrices of appropriate orders. Let each block $\tilde{B}_{n_1}^{(k)}$ be diagonalized as

$$F_{n_1} \tilde{B}_{n_1}^{(k)} F_{n_1}^{*} = \Lambda_{n_1}^{(k)}, \quad k = 0, \cdots, n_2 - 1 \tag{A.1}$$

where F_n is a FFT-matrix of $(F_n)_{jk} = w^{jk}/\sqrt{n}$, F_n^* is its conjugate, and $\Lambda_{n_1}^{(k)}$ is its eigenvalues. Then, the matrix inverse-vector multiplication $\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{q}$ with generic vector $\boldsymbol{q}$ is computed by

$$\tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{q} = (F_{n_2} \otimes F_{n_1})^{*} \Big[\Lambda_{n_1 n_2}^{-1} \Big((F_{n_2} \otimes F_{n_1})\boldsymbol{q} \Big) \Big], \tag{A.2}$$

where

$$\Lambda_{n_1 n_2} = diag\Big(\sum_{i=0}^{n_2-1} \Lambda_{n_1}^{(i)}, \sum_{i=0}^{n_2-1} w^i \Lambda_{n_1}^{(i)}, \cdots, \sum_{i=0}^{n_2-1} w^{(n_2-1)i} \Lambda_{n_1}^{(i)} \Big).$$

Since the elements of $\Lambda_{n_1 n_2}$ are modelled from the MCMC procedure, we can evaluate (A.2) during Gibbs iterations by first computing an

2D FFT on $\boldsymbol{q}$, then dividing the resulting vector by the eigenvalues, and performing an inverse 2D FFT on the resulting vector.

Posterior of τ_ϵ

Let $\boldsymbol{r} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}$. With a Gamma prior for τ_ϵ we have

$$\tau_\epsilon \mid \cdots \sim G(a_*, b_*),$$

where $a_* = a + \frac{n_1 n_2}{2}$, $b_* = b + \frac{1}{2}\boldsymbol{r}^t \tilde{\boldsymbol{\Gamma}}^{-1}\boldsymbol{r}$.

Posterior of $\boldsymbol{\theta}$

Let the normalized residual $\boldsymbol{r}_* = \tau_\epsilon^{1/2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$. Given $\boldsymbol{r}_*$, we compute $\boldsymbol{\varphi}$ by (2.2). Let $\boldsymbol{\Upsilon}$ be a covariance kernel matrix of the process $\boldsymbol{\theta}$. With a Gaussian process prior of $\boldsymbol{\theta}$, we obtain

$$\begin{aligned}
\boldsymbol{\theta} \mid \cdots &\sim \boldsymbol{N}(\boldsymbol{\nu}_*, \boldsymbol{\Upsilon}_*), \\
\boldsymbol{\Upsilon}_* &= (\boldsymbol{\Upsilon}^{-1} + \boldsymbol{V}_\psi^{-1})^{-1}, \\
\boldsymbol{\nu}_* &= \boldsymbol{\Upsilon}_* \boldsymbol{V}_\psi^{-1}(\boldsymbol{\varphi} - \boldsymbol{\kappa}_\psi - \boldsymbol{\nu}) + \boldsymbol{\nu}
\end{aligned}$$

where $\boldsymbol{V}_\psi = diag\{v^2_{\psi_0}, \cdots, v^2_{\psi_{n^h}}\}$, and $\boldsymbol{\kappa}_\psi = (\kappa_{\psi_0}, \cdots, \kappa_{\psi_{n^h}})'$ for the assigned n^h Fourier frequencies.

Posterior of ψ

Given the prior $P(\psi = l) = p_l$, for $l = 1, \cdots, 5$,

$$P(\psi_s = l \mid \cdots) = p_l \phi_{v_l}(\varphi_s - \theta_s - \kappa_l),$$

for $s = 1, 2, \cdots, n^h$, the assigned n^h locations of Fourier frequencies.

Posterior of τ_θ

$$\tau_\theta \mid \cdots \sim G(c_*, d_*),$$

where $c_* = c + \frac{n^h}{2}$ and $d_* = d + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\nu})^t \boldsymbol{\Upsilon}^{-1}(\boldsymbol{\theta} - \boldsymbol{\nu})$.

Posterior of $\rho_{\theta_1}, \rho_{\theta_2}$

We use a "grid-search" method to sample ρ_{θ_1}, and ρ_{θ_2}, respectively. That is, for $j = 1, 2$, we sample ρ_{θ_j} among its candidates $\{\rho_{\theta_j}^{(1)}, \rho_{\theta_j}^{(2)}, \cdots, \rho_{\theta_j}^{(M)}\}$ with probability weights $\pi(\rho_{\theta_j}^{(l)}) \Big/ \sum_{m=1}^{M} \pi(\rho_{\theta_j}^{(m)})$, satisfying

$$\pi(\rho_{\theta_j}^{(l)}) \propto \exp\left(-\frac{1}{2} log \Big| \boldsymbol{\Upsilon}^{-1}_{(\rho_{\theta_j}^{(l)}, \rho_{\theta_{-j}})} \Big| \; -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\nu})^t \boldsymbol{\Upsilon}^{-1}_{(\rho_{\theta_j}^{(l)}, \rho_{\theta_{-j}})} (\boldsymbol{\theta} - \boldsymbol{\nu}) \right),$$

where $\Big| \cdot \Big|$ determinant, and $\boldsymbol{\Upsilon}^{-1}_{(\rho_{\theta_j}^{(l)}, \rho_{\theta_{-j}})}$ a covariance kernel matrix with range parameters $(\rho_{\theta_j}^{(l)}, \rho_{\theta_{-j}})$ in replace of $(\rho_{\theta_1}, \rho_{\theta_2})$, for $j = 1, 2$.

Appendix B

Proofs of the main results

Theorem B.0.1. *[Asymptotic distribution]* $\mathcal{I}_{n_1 n_2}$ *are asymptotically independent and exponentially distributed with mean*

$$\lambda_{\delta_1\delta_2}(w_1, w_2) = \sum_{Q_1 \in \mathcal{Z}} \sum_{Q_2 \in \mathcal{Z}} \lambda\big(w_1 + \frac{2\pi Q_1}{\delta_1}, w_2 + \frac{2\pi Q_2}{\delta_2}\big)$$

for the Fourier frequencies $(w_1, w_2) \in W_\Delta^2 = [-\frac{\pi}{\delta_1}, \frac{\pi}{\delta_1}) \times [-\frac{\pi}{\delta_2}, \frac{\pi}{\delta_2})$

Proof. Without loss of generality, we will prove this statement for $(w_1, w_2) = [0, \frac{2\pi}{\delta_1}) \times [0, \frac{2\pi}{\delta_2})$. Suppose $e(j\delta_1, k\delta_2)$, for $j = 0, 1, \cdots, (n_1 - 1)$, $k = 0, 1, \cdots, (n_2 - 1)$, be a zero mean stationary process. Denote $s_j(n_1)$ be an integer with $w_1(n_1) = \frac{2\pi s_j(n_1)}{n_1\delta_1}$ which converges to w_1 as $n_1 \to \infty$, $s_k(n_2)$ be an integer with $w_2(n_2) = \frac{2\pi s_k(n_2)}{n_2\delta_2}$ which converges to w_2 as $n_2 \to \infty$. Let $\mathcal{D}_{n_1 n_2}(w_1(n_1), w_2(n_2)) = \sum_{j=0}^{n_1-1} \sum_{k=0}^{n_2-1} e(j\delta_1, k\delta_2) exp[-i(w_1(n_1)j + w_2(n_2)k)]$. Then it suffices to show that a sequence of random variables $\mathcal{D}_{(n_1 n_2)}(w_1(n_1), w_2(n_2))$ are asymptotically independent with distribution $N\big(0, 4\pi^2 n_1 n_2 \lambda_{\delta_1\delta_2}(w_1, w_2)\big)$.

First, we have

$$
\begin{aligned}
E\big[\mathcal{D}_{(n_1n_2)}(w_1(n_1),w_2(n_2))\big] &= \sum_{j=0}^{n_1-1}\sum_{k=0}^{n_2-1} E\big[e(j\delta_1,k\delta_2)\big]exp[-i(w_1(n_1)j\delta_1+w_2(n_2)k\delta_2)] \\
&= \sum_{j=0}^{n_1-1}\Big[\sum_{k=0}^{n_2-1} E(e)exp[w_2(n_2)k\delta_2]\Big]exp[-i(w_1(n_1)j\delta_1)] \\
&= \begin{cases} 0, & \text{if } w_2(n_2)\not\equiv 0(mod\pi) \\ \sum_{j=0}^{n_1-1} n_2E(e)exp[-i(w_1(n_1)j\delta_1)], & \text{if } w_2(n_2)\equiv 0(mod2\pi) \\ \sum_{j=0}^{n_1-1} E(e)exp[-i(w_1(n_1)j\delta_1)], & \text{otherwise for odd } n_2 \\ 0, & \text{otherwise for even } n_2 \end{cases} \\
&= 0(\because E(e)=0)
\end{aligned}
\tag{B.1}
$$

Next, given $(w_1,w_2)\in W_\Delta^2$, there exist integer sequences $s_{11},s_{12},s_{21},s_{22}$ s.t.

$$
\begin{aligned}
&Cov\Big[\mathcal{D}_{(n_1n_2)}(w_{11}(n_1),w_{12}(n_2)),\mathcal{D}_{(n_1n_2)}(w_{21}(n_1),w_{22}(n_2))\Big] \\
&= Cov\Big[\sum_{j=0}^{n_1-1}\sum_{k=0}^{n_2-1} e(w_{11},w_{12})exp[-i(w_{11}j\delta_1+w_{12}k\delta_2)], \\
&\sum_{l=0}^{n_1-1}\sum_{m=0}^{n_2-1} e(w_{21},w_{22})exp[-i(w_{21}l\delta_1+w_{22}m\delta_2)]\Big] \\
&= \sum_{j=0}^{n_1-1}\sum_{k=0}^{n_2-1}\sum_{l=0}^{n_1-1}\sum_{m=0}^{n_2-1} Cov\Big[e(w_{11},w_{12})exp[-i(w_{11}j\delta_1+w_{12}k\delta_2)], \\
&e(w_{21},w_{22})exp[-i(w_{21}l\delta_1+w_{22}m\delta_2)]\Big] \\
&= \sum_{j=0}^{n_1-1}\sum_{k=0}^{n_2-1}\sum_{l=0}^{n_1-1}\sum_{m=0}^{n_2-1} C_{n_1n_2}exp[-i2\pi(\frac{s_{11}(n_1)j+s_{21}(n_1)l}{n_1}+\frac{s_{12}(n_2)k+s_{22}(n_2)m}{n_2})] \\
&\to (2\pi)^2(n_1n_2)\lambda_{(\delta_1,\delta_2)}(w_1,w_2), a.e. \text{ as } n_1,n_2\to\infty
\end{aligned}
\tag{B.2}
$$

where $s_{11}(n_1)+s_{21}(n_1)\not\equiv 0(modn_1), s_{12}(n_2)+s_{22}(n_2)\not\equiv 0(modn_2)$ at least all but finite n_1,n_2, by the definition of the target spectral

density. Thus,

$$(2\pi)^{-2}(n_1n_2)^{-1}Cov\Big[\mathcal{D}_{(n_1n_2)}(w_{11}(n_1),w_{12}(n_2)),\mathcal{D}_{(n_1n_2)}(w_{21}(n_1),w_{22}(n_2))\Big] \\ \to \lambda_{(\delta_1,\delta_2)}(w_1,w_2), a.e. \text{ as } n_1,n_2\to\infty \quad \text{(B.3)}$$

Finally, we observe joint density functions for $\forall r, \boldsymbol{e}=(e_1,\cdots,e_r)^t$,

$$\mathcal{L}(\mathcal{D}_{e_1},\cdots,\mathcal{D}_{e_r}) \equiv \mathcal{L}\Big[\mathcal{H}^{n_1n_2}_{e_1,\cdots,e_r}(w_1,w_2)\Big(e_1(w_{11},w_{12}),\cdots,e_r(w_{r1},w_{r2})\Big)\Big]$$

$$\mathcal{L}(e_1(w_{11},w_{12}),\cdots,e_r(w_{r1},w_{r2})) = exp\Big(\sigma_0^{-2}\boldsymbol{e}^t\mathcal{G}^{-1}\boldsymbol{e} + log\big|\sigma_0^2\mathcal{G}\big|\Big)$$

where $\mathcal{H}^{n_1n_2}_{e_1,\cdots,e_r}(w_1,w_2)(*) = \sum_{j=0}^{n_1-1}\sum_{k=0}^{n_2-1}(*)exp\{-(w_1j+w_2k)\}$ is an inverse discrete Fourier operator, and $\mathcal{G}=Cov[\boldsymbol{e}]$. Then we find

$$\begin{aligned}
&(n_1n_2)^{-1/2}\mathcal{L}(\mathcal{D}_{e_1}(w_{11},w_{12}),\cdots,\mathcal{D}_{e_r}(w_{r1},w_{r2})) \\
&= (n_1n_2)^{-1/2}\Big[\mathcal{H}^{n_1n_2}_{e_1,\cdots,e_r}(\sum_{j=1}^{r}w_{j1},\sum_{j=1}^{r}w_{j2})\mathcal{L}\Big(e_1(w_{11},w_{12}),\cdots,e_r(w_{r1},w_{r2})\Big)\Big] \\
&\approx (n_1n_2)^{-1/2}\Big[(2\pi)^{2r}\mathcal{H}^{n_1n_2}_{e_1,\cdots,e_r}(\sum_{j=1}^{r}w_{j1},\sum_{j=1}^{r}w_{j2})\prod_{j=1}^{r}\lambda^{(e_j)}_{(\delta_1,\delta_2)}(w_1,w_2)+o(1)\Big] \\
&= (n_1n_2)^{-1/2}\prod_{j=1}^{r}\Big[(2\pi)^2\mathcal{H}^{n_1n_2}_{e_1,\cdots,e_r}(w_{j1},w_{j2})\Big(\lambda_{(\delta_1,\delta_2)}(w_1,w_2)\Big)+o(1)\Big]
\end{aligned} \quad \text{(B.4)}$$

tends to the product of the likelihoods of a normal distribution as $N=n_1n_2\to\infty$. The approximately equal sign "$\approx$" is due to the Whittle's approximation (Whilttle 1953[57]), and $\lambda^{(e_j)}_{(\delta_1,\delta_2)}(w_1,w_2) = \lambda_{(\delta_1,\delta_2)}(w_1,w_2),\forall j$ by stationarity. This verifies both asymptotic independence and asymptotic normality.

□

Theorem B.0.2. *[Symmetricity] $\mathcal{I}_{n_1n_2}$ is symmetric around the half of the Fourier frequencies $(w_1,w_2)\in W^2_\Delta = [-\frac{\pi}{\delta_1},\frac{\pi}{\delta_1})\times[-\frac{\pi}{\delta_2},\frac{\pi}{\delta_2})$, i.e.*

$$\mathcal{I}_{n_1n_2}(w_1,w_2) = \mathcal{I}_{n_1n_2}(\frac{2\pi}{\delta_1}-w_1,\frac{2\pi}{\delta_2}-w_2)$$

Proof. We claim $\mathcal{D}_{n_1n_2}(w_1, w_2) = \overline{\mathcal{D}_{n_1n_2}(\frac{2\pi}{\delta_1} - w_1, \frac{2\pi}{\delta_2} - w_2)}$

$$\begin{aligned}
\mathcal{D}_{n_1n_2}(w_1, w_2) &= \sum_{j=0}^{n_1-1}\sum_{k=0}^{n_2-1} e(j\delta_1, k\delta_2) exp\Big[- i(w_1 j\delta_1 + w_2 k\delta_2)\Big] \\
&= \sum_{j=0}^{n_1-1}\sum_{k=0}^{n_2-1} e(j\delta_1, k\delta_2) exp(+i2\pi(j+k)) exp\Big[- i(w_1\delta_1 j + w_2\delta_2 k)\Big] \\
&= \sum_{j=0}^{n_1-1}\sum_{k=0}^{n_2-1} e(j\delta_1, k\delta_2) exp\Big[+ i\Big\{(2\pi - w_1\delta_1)j + (2\pi - w_2\delta_2)k\Big\}\Big] \\
&= \sum_{j=0}^{n_1-1}\sum_{k=0}^{n_2-1} e(j\delta_1, k\delta_2) exp\Big[+ i\Big\{(\frac{2\pi}{\delta_1} - w_1)j\delta_1 + (\frac{2\pi}{\delta_2} - w_2)k\delta_2\Big\}\Big] \\
&= \overline{\mathcal{D}_{n_1n_2}(\frac{2\pi}{\delta_1} - w_1, \frac{2\pi}{\delta_2} - w_2)}
\end{aligned} \tag{B.5}$$

Thus,

$$\begin{aligned}
\mathcal{I}_{n_1n_2}(w_1, w_2) &= \frac{1}{4\pi^2 n_1 n_2}\Big|\mathcal{D}_{n_1n_2}(w_1, w_2)\Big|^2 \\
&= \frac{1}{4\pi^2 n_1 n_2}\Big|\overline{\mathcal{D}_{n_1n_2}(\frac{2\pi}{\delta_1} - w_1, \frac{2\pi}{\delta_2} - w_2)}\Big|^2 \\
&= \mathcal{I}_{n_1n_2}(\frac{2\pi}{\delta_1} - w_1, \frac{2\pi}{\delta_2} - w_2)
\end{aligned} \tag{B.6}$$

□

국문초록

본 박사학위논문에서는 스펙트럴 밀도라 불리우는 일종의 푸리에(Fourier) 계수를 베이지안(Bayesian) 마코프-체인-몬테-카를로(MCMC) 관점에서 비모수적으로 모형화하는 통계적방법론을 제안하는데, 이는 등간격의 격자점에서 정의된, 정상성(stationarity)을 가진 1 차원 또는 2차원 확률과정을 예측하는 역할을 수행한다. 핵심 원리는 시간 또는 공간 영역에서 정의된 자기공분산함수를 푸리에변환을 통해 주파수 영역에서의 스펙트럴 밀도함수로 전환하는 것, 그리고 사후분석(posterior analysis)을 위해서 그 스펙트럼 밀도의 로그변환된 함수에 가우시안(Gaussian) 과정 사전분포를 부여하는 것이다.

먼저 공간 자료 예측 문제에 적용할 수 있는 모형을 제안한다. 스펙트럴 밀도함수를 공분산 함수로 변환할 때 본 논문에서 제안하는 보간기법은 전통적인 공간예측 모형에서 필요로 했던 역행렬 계산을 생략함으로서 계산 부담을 줄여준다. 본 모형은 어떠한 알려진 형태의 함수나 등방성 등의 가정을 필요로하지 않으면서도 기존에 대표적인 공간예측 모형들과 비교했을 때 비슷하거나 혹은 더 나은 예측력을 가져다 준다는 것이 시뮬레이션 연구를 통해 입증되었다. 또한 이 모형을 MODIS, AURA와 같은 공신력을 가진 위성자료를 이용하여 한국 지역의 오존농도를 예측하는 문제에 적용했을 때에도 비교적 좋은 예측력을 갖는다는 것이 입증되었다.

다음으로 시계열 자료 예측 문제에 적용할 수 있는 모형을 제안한다.

여기서는 특히 정상성(stationarity) 가정이 일부 완화되어 자기공분산의 한계치(marginal auto-covariance)가 시간에 따라 변하는 이분산성(heteroscedasticity) 확률과정을 생각한다. 이 때 자기공분산은 시간에 따라 변하는 한계분산함수와 정상성을 가진 자기상관함수 사이의 곱으로 표현된다. 자기상관함수의 추정은 기존의 아이디어를 따르고, 한계분산함수의 추정에 있어서는 B-spline 기저함수를 이용한 비모수 추정법을 도입한다. 새로 도입된 과정 역시 하나의 베이지안 마코프-체인-몬테-카를로(MCMC) 안에서 구현된다. 시뮬레이션 연구를 통해 제안한 방법이 등분산성 혹은 이분산성을 지닌 시계열자료의 시간에 따른 패턴을 잘 잡아내는 것이 밝혀졌다. 기존에 잘 알려진 방법인 $ARMA$나 $GARCH$와 같은 모수적 방법론보다 훨씬 많은 수의 모수를 추정해야 함에도 계산 효율은 크게 떨어지지 않는 모습을 보여주고 있다. 본 모형을 대표적인 외국환율 자료 분석에 응용했을 때, 많은 경우 평균제곱오차의 관점에서 전반적으로 예측지점별 오차가 비교적 적게 나오는 것으로 확인되었다.

주요어 : 가우시안 과정 사전분포, 피리오도그램, 위틀우도함수, 스펙트럴 밀도함수

학　번 : 2014 – 21215