DSSR

Discussion Paper No. 9

GENERALIZED WHITTLE ESTIMATE FOR NONSTATIONARY SPATIAL DATA

YASUMASA MATSUDA

May 2013

Data Science and Service Research
Discussion Paper

Center for Data Science and Service Research
Graduate School of Economic and Management
Tohoku University
27-1 Kawauchi, Aobaku
Sendai 980-8576, JAPAN

GENERALIZED WHITTLE ESTIMATE FOR NONSTATIONARY SPATIAL DATA

YASUMASA MATSUDA

ABSTRACT. This paper considers analysis of nonstationary irregularly spaced data that may have multivariate observations. The nonstationarity we focus on here means a local dependency of parameters that describe covariance structures. Nonparametric and parametric ways to estimate the local dependency of the parameters are proposed by an extension of traditional periodogram for stationary time series to that for nonstationary spatial data We introduce locally stationary processes for which consistency of the estimators are proved as well as demonstrate empirical efficiency of the methods by simulated and real examples.

1. INTRODUCTION

Analysis of large spatial data set has been attracting considerable interests recently. The progress of technology makes it possible to collect point reference data in several tens of thousands of points in such fields as geostatistics, forestry and so on (Diggle, 2010) and a method for analyzing large irregularly spaced data set has been in growing needs. Principal difficulties to analyze large spatial data set lie in the huge dimensionality of covariance matrices that makes the likelihood-based inference infeasible. The calculation of the inverse and determinant for the covariance matrices takes too much time to evaluate the likelihoods in reasonable time. There have been mainly two kinds of approaches to cope with the difficulties associated with the large sample sizes.

One is a method called covariance tapering proposed by Kaufman *et al.* (2008) that focuses more on short lag structures of covariances than on long ones. This method approximates the covariance matrix by the one whose elements with lags longer than a prespecified length are replaced by 0. Efficient sparse matrix algorithm makes it feasible to evaluate the likelihood for the approximated covariance matrix. The method by Stein *et al.* (2004) is regarded as a method in this category.

The other one is a method that focuses more on long lag structures of covariances than on short ones. A typical example is the predictive process approach presented by Banerjee *et al.* (2008), which considered a predictive process on mesh points as latent variables and built a hierarchical model for spatial data and proposed to estimate it by an Bayesian approach. A frequency domain approach proposed by Matsuda and Yajima (2009) is regarded as another example in this category. Bai *et al.* (2012) recently proposed a dual approach that aims at taking merits of both of the two approaches of the covariance tapering and predictive process approaches.

Key words and phrases. cubic B-spline. Gram-Schmidt orthogonalization. local periodogram. locally stationary process. Matérn class covariance. nonstationary. spectral density function. Whittle likelihood function.

Although several methods have been proposed to cope with the difficulties associated with the huge dimensionality of covariance matrices, all the existing methods basically assume that spatial data are stationary, which means that covariance functions depend only on spatial lags. In practice, however, it is rare to find stationarity in spatial data. For example, land price data that we analyze in Section 5 has variances that variates locally. Specifically, the variance of land price becomes larger as the observation points approach the city center, which demonstrates clearly nonstationary covariance structures.

The purpose of this paper is to propose an method for analysis of nonstationarity spatial data which may have multivariate observations by an extension of a frequency domain approach by Matsuda and Yajima (2009). The key features unique in this paper are as follows. The first one is that sinusoidal basis functions employed in Matsuda and Yajima (2009) to detect amplitude of Fourier frequencies are orthogonalized under the inner product characterized by a weight function to cope with nonstationarity. As a result, the weighted Fourier transform by the orthogonalized basis leads to introduction of the local periodogram and proposition of the generalized Whittle likelihood function with it. The next one is the introduction of locally stationary processes for which consistency of the Whittle likelihood estimators are proved. Locally stationary processes were originally proposed by Dahlhaus (1997) for analysis of nonstationary time series in order to conduct rigorous theoretical treatment of Whittle likelihood estimators. This paper borrows the concept of them but define the ones in considerably different manners. Specifically, the main difference between them lies in the definition in terms of Riemannian approximation for the stochastic integral in the original definition, which lets our proposed method more practical in the sense that the generalized Whittle likelihood is not defined by an integral but by a feasible summation on Fourier Frequencies for which a proof of consistency is conducted, although asymptotic normality has not been proved yet unfortunately.

2. Setting of our problems

This paper focuses on a spatial regression model that generates nonstationary observations on irregularly spaced points. Let $X(s) = (X_1(s), \ldots, X_b(s))$ be b independent vectors on s. Then multivariate observations $Y(s_k) = (Y_1(s_k), \ldots, Y_a(s_k))'$ for $s_k, k = 1, \ldots, n$, are described by

$$Y(s_k) \quad = \quad m(s_k) + Z(s_k) + \varepsilon_k, \tag{1}$$

where $m(s) = (m_1(s), \ldots, m_a(s))'$ is the mean function given by the following regression form

$$m_i(s) \quad = \quad X(s)\beta_i(s),$$

for $i = 1, \ldots, a$, $Z(s)$ is a zero mean nonstationary error process and ε_k is a nugget term given by a sequence of independent random vectors with mean 0 and variance matrix $T(s)$. It should be noted that $\beta_i(s)$ and $T(s)$ are allowed to be locally dependent.

The nonstationary process $Z(s)$ has the covariance matrix

$$R(s, h) \quad = \quad EZ(s + h/2)Z'(s - h/2), \tag{2}$$

which does depend not only on h but also on s. The dependency of the covariance on s is attained through a dependency of a parameter θ on s that describes the covariances, namely

$$R(s, h) \quad = \quad R(\theta(s), h).$$

The local dependency of $\theta(s)$ as well as that of $T(s)$ and $\beta_i(s)$ is an unique feature that has not been considered well yet to characterize nonstationarity in spatial data, for which we will work in this paper.

The Fourier transform of the covariance $R(\theta(s), h)$, which is called the Wigner-Ville spectrum (Martin and Flandrin, 1985), is given by

$$F(\theta(s), \lambda) \quad = \quad (2\pi)^{-2} \int_{\mathbb{R}^2} R(\theta(s), h) \exp(-i\lambda' h) dh, \tag{3}$$

which defines the nonstationary structures by the dependency on s of the spectral density F. Both the definitions of nonstationarity on spatial and frequency domains exactly correspond mathematically.

The problem in this paper is how to estimate $\theta(s)$, $T(s)$ and $\beta_i(s)$ for $i = 1, \ldots, a$ as a function of s. After estimation of $\theta(s)$ and $T(s)$ is considered in a nonparametric way, that of $\theta(s)$ and $T(s)$ will be considered in a parametric way. And that of $\beta_i(s)$ will be also considered in a parametric way under the error covariances with the estimated parameters. It should be emphasized that the nonparametric way does not mean assuming no parametric models in covariances but means assuming no parametric models in local dependencies of the parameters that describe covariances. The reason why the nonparametric estimation as well as the parametric one is considered in this paper is that the nonparametric one is helpful in identifying a parametric form of the local dependency and in providing good starting values for the optimization process in the parametric estimation.

For example, let us consider a Matérn class covariance that is described by smoothness, sill and range parameters denoted as $\theta = (\nu, \tau, \rho)$. Then the nonstationary structure is introduced by local dependency of parameters θ on s in the covariance

$$R(\theta(s), h) \quad = \quad \frac{\tau(s) \left\{ 2\sqrt{\nu(s)}|h|/\rho(s) \right\}^{\nu(s)}}{2^{\nu(s)-1}\Gamma(\nu(s))} K_{\nu(s)} \left(\frac{2\sqrt{\nu(s)}|h|}{\rho(s)} \right), \tag{4}$$

or equivalently, by that in the spectral density function

$$F(\theta(s), \lambda) \quad = \quad \frac{\phi(s)}{\{|\lambda|^2 + \alpha^2(s)\}^{1+\nu(s)}}, \tag{5}$$

for $\alpha(s) = 2\sqrt{\nu(s)}/\rho(s)$ and $\phi(s) = \nu(s)\alpha(s)^{2\nu(s)}\tau(s)/\pi$. We will discuss nonparametric estimation of $(\nu(s), \tau(s), \rho(s))$, and will also consider parametric estimation of them when, for example, cubic B-spline functions are fitted to describe their local dependencies.

We keep in mind as a nonstationary model for practical applications a Matérn class with local dependency identified by cubic B-spline functions, although no specific models are identified for $R(\theta(s), h)$ in Sections 3 and 4 to let discussions be as general as possible. In Section 5 for empirical studies, however, a Matérn class covariance with local dependency identified by cubic B splines is applied to simulated and real examples.

3. Estimation

Suppose we have observed $Y(s_k)$ in (1) on irregularly spaced points $s_k \in S, k = 1, \ldots, n$. We discuss estimation of $\theta(s)$ in the spectrum $F(\theta(s), \lambda)$, $T(s)$ in the nugget term and $\beta_i(s)$ for $i = 1, \ldots, a$ in the mean function. First, after introducing nonparametric estimates for $\theta(s)$ and $T(s)$, we will consider parametric ones, when $\theta(s)$ and $T(s)$ are described as $\theta_{\phi_1}(s)$ and $T_{\phi_2}(s)$ for some parameters ϕ_1 and ϕ_2, respectively. Next, under the covariances estimated with $\hat{\phi}_1$ and $\hat{\phi}_2$, the generalized least squares estimator is proposed for the regression coefficients $\beta_i(s)$ when $\beta_i(s)$s are identified as $\beta_i(s)' = U(s)\psi$ for a basis function $U(s)$ and parameters ψ.

The features distinguished from existing frequency domain approaches are the employment of the Gram-Schmidt procedure for empirical orthogonalization of Fourier basis functions, by which the local periodogram is defined.

3.1. **Estimation of the covariance parameters.** Let us start from nonparametric estimation of $\theta(u)$ and $T(u)$ over $u \in [0, B_1] \times [0, B_2]$ which should be included in S. Fix a positive integer r_n such that $b + 2r_n < n$ and a weight function $w_h(x) = w(x/h)$ for a bandwidth $h > 0$ and a positive and continuous function $w(x)$ on $\mathbb{R}^2$ that has a maximum on the origin and converges to 0 as $|x| \to \infty$. Introduce a set of mesh points on $\mathbb{R}^2$ as

$$\Omega^+ = \left\{ \left(\frac{2\pi j}{B_1}, \frac{2\pi k}{B_2} \right), (j, k) \in \mathbb{Z}^2, k > 0 \cup \{j > 0, k = 0\} \right\}.$$

The elements are sorted with ascending order by the distances from the origin and put the first r_n elements and their symmetric points with respect to the origin as Ω_n^+ and Ω_n^-, respectively. Define

$$\Omega_n = \Omega_n^+ \cup \Omega_n^-,$$

which is an extension of Fourier frequencies used in time series (Brockwell and Davis, 1980, page 332) to those in spatial data.

It is well known that the set of Fourier series on Fourier frequencies constitutes an orthogonal basis in time series case (Brockwell and Davis, 1991,page 332). Since it is not in general an orthonormal basis when the data points are irregularly spaced, the set of the basis functions $\exp(i\omega_k' s_j)$ for $\omega_k \in \Omega_n^+$ is forced to be orthonormalized under the inner product weighted on u defined by

$$< f, g >_u = \sum_{p=1}^{n} f(s_p)\overline{g(s_p)} w\left(\frac{s_p - u}{h} \right)$$

for $f, g \in \mathbb{C}^n$ on the observed points $s_p, p = 1, \ldots, n$.

The sequence of $b + r_n$ vectors in $\mathbb{C}^n$, which are the independent vectors and the Fourier series on Ω_n^+, *i.e.*, the union of $X(s) = (X_1(s), \ldots, X_b(s))$ and $\exp(i\omega_j' s), \omega_j \in \Omega_n^+$, is orthogonalized under the inner product weighted on u by the Gram-Schmidt procedure. And define the orthonormalized ones for $\exp(i\omega_j' s)$ as $\xi_u(\omega_j, s_p)$, which satisfy $< \xi_u(\omega_i), \xi_u(\omega_j) >_u = \delta_{ij}$ for Kronecker's delta.

Then the local periodogram on $u \in [0, B_1] \times [0, B_2]$ is, for $\omega_k \in \Omega_n^+$, defined by

$$
\begin{aligned}
d(u,\omega_k) &:= \; < Y, \xi_u(\omega_k) >_u = \sum_{p=1}^{n} Y(s_p)\overline{\xi_u(\omega_k, s_p)} w\left(\frac{s_p - u}{h}\right), \\
H_u &:= \sum_{p=1}^{n} w\left(\frac{s_p - u}{h}\right), \\
I(u,\omega_k) &:= \frac{|B|}{4\pi^2 H_u} d(u,\omega_k)\overline{d(u,\omega_k)}', \qquad (6)
\end{aligned}
$$

where $|B| = B_1 \cdot B_2$, and $I(u, \omega_k)$ is called as the local periodogram matrix on $u \in [0, B_1] \times [0, B_2]$, which is regarded as an extension of traditional one for time series.

Based on the local periodogram on u, the nonparametric estimators are $\hat{\theta}(u)$ and $\hat{T}(u)$ that minimize

$$
\begin{aligned}
Q_u(\theta, T) &= 4\pi^2 |B|^{-1} \sum_{\omega_k \in D} \mathrm{tr}\left[\{F(\theta,\omega_k) + b(u,\omega_k,\omega_k)T\}^{-1} I(u,\omega_k)\right] \\
&\quad + \log|F(\theta,\omega_k) + b(u,\omega_k,\omega_k)T|, \qquad (7)
\end{aligned}
$$

where D is a prefixed region on $\mathbb{R}^2$ and

$$
b(u,\omega,\lambda) = \frac{|B|}{4\pi^2 H_u} \sum_{p=1}^{n} \xi_u(\omega, s_p)\overline{\xi_u(\lambda, s_p)} w^2\left(\frac{s_p - u}{h}\right),
$$

which is the term caused by the nugget effect which does not appear usually in traditional Whittle likelihoods for time series. The proof of Lemma 1 in Section 7 reveals how $b(u, \omega, \omega)$ is inserted in the likelihood to treat the effect caused by the nugget term. We call the objective function in (7) the local Whittle likelihood function on u.

Now let us turn to parametric estimation for $\theta(u)$ and $T(u)$ in which case the parameters ϕ_1 and ϕ_2 describe $\theta(u)$ and $T(u)$ as $\theta_{\phi_1}(u)$ and $T_{\phi_2}(u)$, respectively, over $u \in [0, B_1] \times [0, B_2]$. Let K be a set of mesh points over $[0, B_1] \times [0, B_2]$ given by $(iB_1^{1-\rho}h^\rho, jB_2^{1-\rho}h^\rho)$ for $0 < \rho < 1$ and $i, j = 1, 2, \ldots$.

Based on the local periodograms on $u_j \in K$, the parameters ϕ_1 and ϕ_2 are estimated by the ones, which we denote as $\hat{\phi}_1, \hat{\phi}_2$, that minimize

$$
\begin{aligned}
&L(\phi_1, \phi_2) = 4\pi^2 |B|^{-\rho} h^{2\rho} \\
&\times \sum_{u_j \in K} \sum_{\omega_k \in D} \mathrm{tr}\left[\{F(\theta_{\phi_1}(u_j),\omega_k) + b(u_j,\omega_k,\omega_k)T_{\phi_2}(u_j)\}^{-1} I(u_j,\omega_k)\right] \\
&+ \log|F(\theta_{\phi_1}(u_j),\omega_k) + b(u_j,\omega_k,\omega_k)T_{\phi_2}(u_j)|. \qquad (8)
\end{aligned}
$$

The objective function to be minimized in (8) is called the generalized Whittle likelihood function. Then $\theta(u)$ and $T(u)$ are estimated by $\theta_{\hat{\phi}_1}(s)$ and $T_{\hat{\phi}_2}(s)$, respectively.

Remark 1. It is required to store the values of $I(u_j, \omega_k)$ and $b(u_j, \omega_k, \omega_k)$ for $u_k \in K, \omega_k \in \Omega_N^+$ in the minimization in (8), which requires the orthonormlizations of $\exp(i\omega_j' s_k), \omega_j \in \Omega_n^+$ under each one of the inner products weighted on $u_j \in K$. The orthohonalizations for each point constitute the most time consuming part of the estimation procedure.

Remark 2. The orthonormalizations of $\exp(i\omega_j' s_k), \omega_j \in \Omega_n^+$ are equivalent to those of $\cos(\omega_j' s_k), \sin(\omega_j' s_k), \omega_j \in \Omega_n^+$, which we denote as $a_u(\omega_j), b_u(\omega_j)$, from which $\xi_u(\omega_j, s_k)$ is constructed as $a_u(\omega_j) + ib_u(\omega_j)$.

Remark 3. The order of the basis functions on Fourier frequencies affects seriously the result of the Gram-Schmidt orthogonalization procedure, and hence estimation performances. It is recommended from our experiences that the Gram-Schmidt procedure should be applied to the independent vectors before to $\exp(i\omega_j' s_k), \omega_j \in \Omega_n^+$ in the ascending order by the distances of ω_j from the origin. This is because the ascending order lets the periodograms on high frequencies, which play a crucial role in estimation performance, be more efficient approximations for the spectrum than any other order does. The consistency proof in Section 4, however, requires no special rules for ordering in Ω_n^+.

Remark 4. To conduct the estimation, r_n that determine D and the bandwidth h must be prefixed. The choices of them should be made jointly depending on performances of the Gram-Schmidt procedure applied to Fourier basis functions. Larger r_n and smaller h provide more unbiased but unstable estimates by making Fourier series lose more easily the linear independence as a basis function, and hence make the Gram-Schmidt procedure infeasible, while smaller r_n and larger h provide more biased but stable estimates by making them linearly independent.

3.2. **Mean function estimation.** Let us consider an estimation for the mean function $m_i(s) = X(s)\beta_i(s), i = 1, \ldots, a$ in model (1), when $\beta_i(s)$ is described as

$$\beta_i(s)' \quad = \quad U(s)(\psi_{i,1}, \ldots, \psi_{i,b}), \tag{9}$$

for a basis function $U(s) = (U_1(s), \ldots, U_c(s))$ on $[0, B_1] \times [0, B_2]$ and parameters $\psi_{i,j} = (\psi_{i,j,1}, \ldots \psi_{i,j,c})'$. The reason why we restrict a parametric form for $\beta_i(s)$ to a linear one is that a linear one makes it possible to re-express (1) as a linear regression form and to apply generalized least squares estimate. Namely, substitution of (9) into (1) provides the following linear regression form

$$Y_i(s_k) \quad = \quad X(s_k) \bigotimes U(s_k)\psi_i + Z_{i,s_k} + \varepsilon_{i,k},$$

for $k = 1, \ldots, n$, where $\psi_i = vec(\psi_{i,1}, \ldots, \psi_{i,b})$.

The parametric estimates $\hat{\phi}_1$ and $\hat{\phi}_2$ minimizing the Whittle likelihood in (8) let us estimate the error covariance structures by the i th diagonal element of

$$\hat{V}_{pq} \quad = \quad R\left(\theta_{\hat{\phi}_1}\left(\frac{s_p + s_q}{2}\right), s_p - s_q\right) + T_{\hat{\phi}_2}\left(\frac{s_p + s_q}{2}\right),$$

for $p, q = 1, \ldots, n$, which we denote as $\hat{V}_{i,pq}$. The error covariance matrix is evaluated as the n by n matrix $\hat{V}_i$ whose (p, q)th element is $\hat{V}_{i,pq}$. Let Y_i be $(Y_i(s_1), \ldots, Y_i(s_n))'$ and let L be the $n \times bc$ matrix whose kth row is $X(s_k) \bigotimes U(s_k)$. Then the GLS is

$$\hat{\psi}_i \quad = \quad \left(L'\hat{V}_i^{-1}L\right)^{-1} L'\hat{V}_i^{-1}Y_i, \tag{10}$$

and $\beta_{i,j}(s)$ is estimated by

$$\hat{\beta}_{i,j}(s) \quad = \quad U(s)\hat{\psi}_{i,j},$$

for $j = 1, \ldots, b$.

4. Consistency

This section clarifies in what conditions our proposed estimators are consistent. In irregularly spaced data analysis, special care is necessary for the asymptotics. As Dahlhaus (1997) pointed out, it is impossible to estimate consistently the parameter $\theta = \theta(s), s \in [0, B_1] \times [0, B_2]$ in the asymptotics that the domain B_1, B_2 tends to be large, since finer and finer observations are not available. Even in the fixed domain asymptotics, which is the one based on growing observations under fixed domains (Stein, 1999, page 62), consistency is not available because of a lack of data with longer spatial lags than the fixed domain. The only way to avoid the inconsistency, we need to assume that

$$\theta(s) = \theta\left(\frac{s_1}{B_1}, \frac{s_2}{B_2}\right) \tag{11}$$

under the asymptotics that B_1, B_2 tend to be large.

This section provides the models that satisfy (11) with the asymptotics under which our estimators proposed in the previous section are consistent. The notable features lie in the following two points. First, the models to satisfy (11) is constructed by an extension of locally stationary models by Dahlhaus (1997). Secondly, the asymptotics are built not by introducing randomness but in a deterministic manner to account for the irregularity of the data points.

4.1. **Locally stationary processes.**

Dahlhaus (1997) proposed a locally stationary process for rigorous asymptotic treatment of regular nonstationary time series. We extend the locally stationary process for time series to that for spatial data that may be observed on irregular locations. The main differences are the changes of domain of the spectrum from $[-\pi, \pi]$ to $\mathbb{R}^2$ and the adoption of Riemannian sum for stochastic integral.

In some general conditions, a stationary process on $\mathbb{R}^2$ with the spectral density matrix $F(\lambda)$ has the spectral representation:

$$\int_{\mathbb{R}^2} \exp(i\lambda' s) A(\lambda) d\xi(\lambda), \tag{12}$$

where

$$A(\lambda)\overline{A(\lambda)}' = F(\lambda),$$

and $\xi(\lambda)$ is an orthogonal increment process such that

$$\mathrm{Cov}\left(d\xi(\lambda_1), d\xi(\lambda_2)\right) = \delta_{12} I_a$$

for the Dirac delta function δ.

We generalize the stationary expression (12) to nonstationary one by allowing $A(\lambda)$ to be dependent locally inside $[0, B_1] \times [0, B_2]$. Let Ω be the set of Fourier frequencies given by

$$\Omega = \left\{ \left(\frac{2\pi j}{B_1}, \frac{2\pi k}{B_2}\right), (j, k) \in \mathbb{Z}^2 \right\}.$$

Put $|B| = B_1 \times B_2$ and let us express s/B for $(s_1/B_1, s_2/B_2)$, and ω_j for an element of Ω with an integer j, for notational simplicity.

Definition 1. $Z_B(s)$ is a locally stationary process over $[0,B_1]\times[0,B_2]$ with transfer matrix function A, if there exists a representation

$$Z_B(s) \quad = \quad \sqrt{\frac{4\pi^2}{|B|}}\sum_{\omega_j\in\Omega}\exp(i\omega_j' s)A\left(\frac{s}{B},\omega_j\right)z_j, \tag{13}$$

where the following holds.

(i) $\{z_j, j\in\mathbb{Z}\}$ is a sequence of independent and identically distributed random vectors with $z_k=\overline{z_j}$ for j,k such that $\omega_k=-\omega_j$, and with mean 0, $Ez_j\overline{z_j}'=I_a$ and finite fourth order moments.

(ii) There exists a function $A:[0,1]^2\times\mathbb{R}^2\to\mathbb{R}^{a\times a}$ with $A(u,-\lambda)=\overline{A(u,\lambda)}$.

Let us approximate $Z(s)$ in (1) whose spectrum is identified by the parameter θ that satisfies (11), namely

$$F(\theta(s),\lambda)=F(\theta(s/B),\lambda),$$

by the locally stationary process $Z_B(s)$ in (13) with the transfer function that satisfies

$$A(u,\lambda)\overline{A(u,\lambda)}' \quad = \quad F(\theta(u),\lambda)$$

for $u\in[0,1]^2$. This approximation is justified asymptotically in the sense that the covariances of the latter converges to the ones of the former. Hence it is reasonable to assume later in proving consistency that the error term $Z(s)$ is given by the locally stationary process.

Specifically, let us define the spectral density matrix of $Z_B(s)$ by

$$F_B(u,\lambda):=(2\pi)^{-2}\int_{\mathbb{R}^2}\mathrm{Cov}\left(Z_B(uB+h/2),Z_B(uB-h/2)\right)\exp(-i\lambda' h)dh,$$

where the covariance is defined to be 0 when $u+h/(2B)$ or $u-h/(2B)$ is outside $[0,1]^2$. Then we have the convergence of F_B to F in the following theorem.

Theorem 1. *If $Z_B(s)$ is the locally stationary process in (13) with $A(u,\lambda)\overline{A(u,\lambda)}'=F(\theta(u),\lambda)$, $F(\theta(u),\lambda)\in L^2(\mathbb{R}^2)$ uniformly in $u\in[0,1]^2$ and $A(u,\lambda)$ is continuous with respect to u uniformly in $\lambda\in\mathbb{R}^2$, then we have for all $u\in[0,1]^2$,*

$$\int_{\mathbb{R}^2}\{F_B(u,\lambda)-F(\theta(u),\lambda)\}\overline{\{F_B(u,\lambda)-F(\theta(u),\lambda)\}}'d\lambda=o(1),$$

as B_1,B_2 tends to ∞.

Proof. We have

$$F_B(u,\lambda) \quad = \quad \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\exp(i\lambda' h)dh\frac{4\pi^2}{|B|}\sum_{\omega_j\in\Omega}A(u+\frac{h}{2B},\omega_j)\overline{A(u-\frac{h}{2B},\omega_j)}'\exp(-ih'\omega_j),$$

and

$$F(u,\lambda) \quad = \quad \frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\exp(i\lambda' h)dh\int_{\mathbb{R}^2}F(u,\omega)\exp(-ih'\omega)d\omega.$$

By Perseval equality, we have

$$\int_{\mathbb{R}^2}\{F_B(u,\lambda)-F(u,\lambda)\}\overline{\{F_B(u,\lambda)-F(u,\lambda)\}}'d\lambda \quad = \quad \int_{\mathbb{R}^2}c(h)\overline{c(h)}'dh,$$

where

$$c(h) = \int_{\mathbb{R}^2} F(u,\omega)\exp(-ih'\omega)d\omega - \frac{4\pi^2}{|B|}\sum_{\omega_j\in\Omega} A(u+\frac{h}{2B},\omega_j)\overline{A(u-\frac{h}{2B},\omega_j)}'\exp(-ih'\omega_j),$$

which are both $L^2(\mathbb{R})$ regardless of B, and bounded by

$$\left|\int_{\mathbb{R}^2} F(u,\omega)\exp(-ih'\omega)d\omega - \frac{4\pi^2}{|B|}\sum_{\omega_j\in\Omega} F(u,\omega_j)\exp(-ih'\omega_j)\right| + \frac{4\pi^2}{|B|}\sum_{\omega_j\in\Omega}\left|F(u,\omega_j) - A(u+\frac{h}{2B},\omega_j)\overline{A(u-\frac{h}{2B},\omega_j)}'\right|,$$

both terms of which converge to 0 for each h. It follows that the result follows by the dominated convergence theorem.

4.2. **Assumptions.** Let us give the detailed conditions required for consistency of the nonparametric and parametric estimators. Assumption 2 given below defines the asymptotic regime under which consistency is considered, where we follow a deterministic manner to describe the irregularity of the data points. Assumption 4 (ii) states that the parametric models for $\theta(u)$ and $T(u)$ are not necessarily correct, *i.e.*, $\theta_{\phi_1}(u)$ and $T_{\phi_2}(u)$ do not necessarily include $\theta(u)$ and $T(u)$ for any ϕ_1 and ϕ_2, respectively.

Assumption 1. *In (1), the mean function is given by $m(s) = \beta(s/B)$ with $X(s) \equiv 1$ for a continuous function β on $[0,1]^2$, the error process $Z(s)$ is given by the locally stationary process $Z_B(s)$ with $A(u,\lambda)\overline{A(u,\lambda)}' = F(\theta(u),\lambda)$ for $u \in [0,1]^2$ and the nugget $T(s) = T(s/B)$ for a continuous function T on $[0,1]^2$. The transfer function matrix $A(u,\lambda)$ is uniformly bounded and continuous in u and λ.*

Assumption 2. *Asymptotics is defined with a positive integer k in the following sense. As k tends to be large, the region $[0,B_1]\times[0,B_2] = [0,B_1^k]\times[0,B_2^k]$ tends to $[0,\infty]\times[0,\infty]$, inside which the sample size $n = n_k$ and the observation points $s_p = s_p^{(k)}, p = 1,\ldots,n_k$ must satisfy the followings uniformly in $u \in [0,1]^2$.*

(i) *there exists a continuous function $C(u,\omega)$ on $[0,1]^2\times\mathbb{R}^2$ such that, for $\omega_k,\ \omega_l \in \Omega_n^+$ such that $|\omega_k - \omega_l| > h^{-1}$, $b(uB,\omega_k,\omega_k) \to C(u,\omega_k)$, $b(uB,\omega_k,\omega_l) \to 0$ and $b_2(uB,uB,\omega_k,\omega_l) \to 0$, where*

$$b_2(u_1B,u_2B,\omega_k,\omega_l) = \frac{|B|^2}{16\pi^4 H_{u_1B}H_{u_2B}} \times \sum_{p=1}^{n}\left|\xi_{u_1B}(\omega_k,s_p)\overline{\xi_{u_2B}(\omega_l,s_p)}w\left(\frac{s_p-u_1B}{h}\right)w\left(\frac{s_p-u_2B}{h}\right)\right|^2,$$

(ii) *the empirical correlation between* $\exp(i\omega_j's_p), j = k,l,$ *i.e.*,

$$\rho(u,\omega_k-\omega_l) = H_{uB}^{-1}\sum_{p=1}^{n}\exp\{i(\omega_k-\omega_l)'s_p\}\,w\left(\frac{s_p-uB}{h}\right), \tag{14}$$

satisfies

$$\sum_{\omega_c\in\Omega}|\rho(u,\omega_c)| < C,$$

for a constant C.

Assumption 3. *The bandwidth $h = h_k$ satisfies*

$$h_k \to \infty, B_i^{-1} h_k \to 0,$$

for $i = 1, 2$, as k tends to be large, under the asymptotics in Assumption 2.

Assumption 4. *In the nonparametric and parametric estimation given by minimizing the likelihood (7) and (8), respectively,*

(i) *When we fit $\tilde{\theta}$ and $\tilde{T}$ for $\theta(u)$ and $T(u)$, respectively, in the nonparametric estimation, for each $u \in [0,1]^2$, $(\tilde{\theta}, \tilde{T}) \in \Theta_u \subset \mathbb{R}^p$, where Θ_u is compact and the function $F(\tilde{\theta}, \lambda) + C(u, \lambda)\tilde{T}$ is uniformly bounded from above and below and continuous on $\Theta_u \times D$. $(\tilde{\theta}_1, \tilde{T}_1) \neq (\tilde{\theta}_2, \tilde{T}_2)$ implies $F(\tilde{\theta}_1, \lambda) + C(u, \lambda)\tilde{T}_1 \neq F(\tilde{\theta}_2, \lambda) + C(u, \lambda)\tilde{T}_2$ on a set with positive Lebesgue measure and $(\theta(u), T(u))$ exists in the interior of Θ_u.*

(ii) *When we fit a parametric model $\theta_{\phi_1}(u)$ and $T_{\phi_2}(u)$ for $\theta(u)$ and $T(u)$, respectively, in the parametric estimation, $(\phi_1, \phi_2) \in \Phi \subset \mathbb{R}^q$, where Φ is compact and the function $F(\theta_\phi(u), \lambda) + C(u, \lambda)T_{\phi_2}(u)$ is uniformly bounded from above and below and continuous on $\Phi \times [0,1]^2 \times D$. $(\phi_1, \phi_2) \neq (\tilde{\phi}_1, \tilde{\phi}_2)$ implies $F(\theta_{\phi_1}(u), \lambda) + C(u, \lambda)T_{\phi_2}(u) \neq F(\theta_{\tilde{\phi}_1}(u), \lambda) + C(u, \lambda)\tilde{T}_{\tilde{\phi}_2}(u)$ on a set with positive Lebesgue measure and (ϕ_1^0, ϕ_2^0) that minimize*

$$L_0(\phi_1, \phi_2) = \int_{[0,1]^2} \int_D tr\left[\{F(\theta_{\phi_1}(u), \lambda) + C(u, \lambda)T_{\phi_2}(u)\}^{-1} \{F(\theta(u), \lambda) + C(u, \lambda)T(u)\}\right] + \log|F(\theta_{\phi_1}(u), \lambda) + C(u, \lambda)T_{\phi_2}(u)| \, dud\lambda.$$

exists uniquely and lies in the interior of Φ.

4.3. **Consistency.** Let us prove consistency of the nonparametric and parametric estimates under Assumptions 1-3, 4(i) and Assumptions 1-3, 4(ii), respectively.

Theorem 2. *If Assumptions 1, 2, 3 and 4(i) hold, the nonparametric estimators $\hat{\theta}(uB)$ and $\hat{T}(uB)$ minimizing (7) converge in probability to $\theta(u)$ and $T(u)$ for $u \in [0,1]^2$, respectively, under the asymptotics in Assumption 2.*

Proof. By Lemma 3, we have for a fixed $u \in [0,1]^2$,

$$Q_{uB}(\tilde{\theta}, \tilde{T}) \to Q_u^0(\tilde{\theta}, \tilde{T})$$

in probability, where

$$\begin{aligned} Q_u^0(\tilde{\theta}, \tilde{T}) &= \int_D \mathrm{tr}\left[\left\{F(\tilde{\theta}, \lambda) + C(u, \lambda)\tilde{T}\right\}^{-1} \{F(\theta(u), \lambda) + C(u, \lambda)T(u)\}\right] \\ &\quad + \log\left|F(\tilde{\theta}, \lambda) + C(u, \lambda)\tilde{T}\right| d\lambda. \end{aligned}$$

By the identifiability condition in Assumption 4, for $(\tilde{\theta}, \tilde{T}) \neq (\theta(u), T(u))$,

$$Q_u^0(\tilde{\theta}, \tilde{T}) - Q_u^0(\theta(u), T(u)) = \int_D \left[\log \frac{\left|F(\tilde{\theta}, \lambda) + C(u, \lambda)\tilde{T}\right|}{|F(\theta(u), \lambda) + C(u, \lambda)T(u)|} + \mathrm{tr}\left[\left\{F(\tilde{\theta}, \lambda) + C(u, \lambda)\tilde{T}\right\}^{-1} \{F(\theta(u), \lambda) + C(u, \lambda)T(u)\}\right] - 1\right] d\lambda > 0.$$

For any $\varepsilon > 0$, there is a $\delta > 0$ such that, for any (θ_1, T_1) and (θ_2, T_2) that satisfy $|(\theta_1, T_1) - (\theta_2, T_2)| < \delta$,

$$P\left(|Q_{uB}(\theta_1, T_1) - Q_{uB}(\theta_2, T_2)| < \varepsilon\right) \to 1.$$

It follows that the consistency follows by Walker (1964, Lemma 2).

Theorem 3. *If Assumptions 1, 2, 3 and 4(ii) hold, the parametric estimators $\hat{\phi}_1$ and $\hat{\phi}_2$ minimizing (8) converge in probability to ϕ_1^0 and ϕ_2^0, respectively, under the asymptotics in Assumption 2.*

Proof. By Lemma 5, we have

$$L(\phi_1, \phi_2) \to L_0(\phi_1, \phi_2)$$

in probability. The rest of the proof follows in exactly the same way as that in Theorem 2.

5. Empirical examples

This section considers empirical performances of the estimators proposed in Section 3. In the first experiment, empirical performances of the nonparametric and parametric estimators that minimize (7) and (8), respectively, are examined for simulated data when the mean function is designed to be 0. In the second experiment, the mean function estimation by (10) in addition to the parametric estimation by minimizing (8) are conducted for land price data in Tokyo.

We restrict our attention to uni-dimensional observations for which the empirical studies are conducted. For the two experiments, a Matérn class covariance is fitted in every estimation, where the parametric estimation employs cubic B-splines to describe local dependency of the parameters, and where the weight function is designed as $w(x/h, y/h) = \exp(-x^2/h)\exp(-y^2/h)$.

5.1. **Simulation studies.** We conduct the nonparametric and parametric estimates for simulated data a hundred times to show the empirical properties of them. We simulate nonstationary data by (1) on uniformly scattered points over the region $[0, 30] \times [0, 30]$, where the mean function is designed to be 0 and the covariance function is specified with an isotropic Matérn class in (4), where the local dependencies of the parameters are are designed by

$$\begin{aligned}
\nu(s) &= 2.0 - 1.5\exp\left\{-\frac{(s_1-15)^2+(s_2-15)^2}{100}\right\}, \\
\tau(s) &= 10 + 20\exp\left\{-\frac{(s_1-15)^2+(s_2-15)^2}{100}\right\}, \\
\rho(s) &= 2 + 3\exp\left\{-\frac{(s_1-15)^2+(s_2-15)^2}{100}\right\}, \\
T(s) &= 1 + 3\exp\left\{-\frac{(s_1-15)^2+(s_2-15)^2}{100}\right\}.
\end{aligned} \tag{15}$$

For 100 sets of the simulated data with $n = 8000$, we fit a Matérn class to construct the nonparametric estimators for (ν, τ, ρ, T) over the region $B = [5, 25] \times [5, 25]$. They were conducted by minimizing (7), when we set $h = 4^2$ and took first 1000 elements in Ω_n^+ as D. In Figure 1, the median and the 5 and 95 percentiles for 100 nonparametric estimators are shown in comparison with the true values as

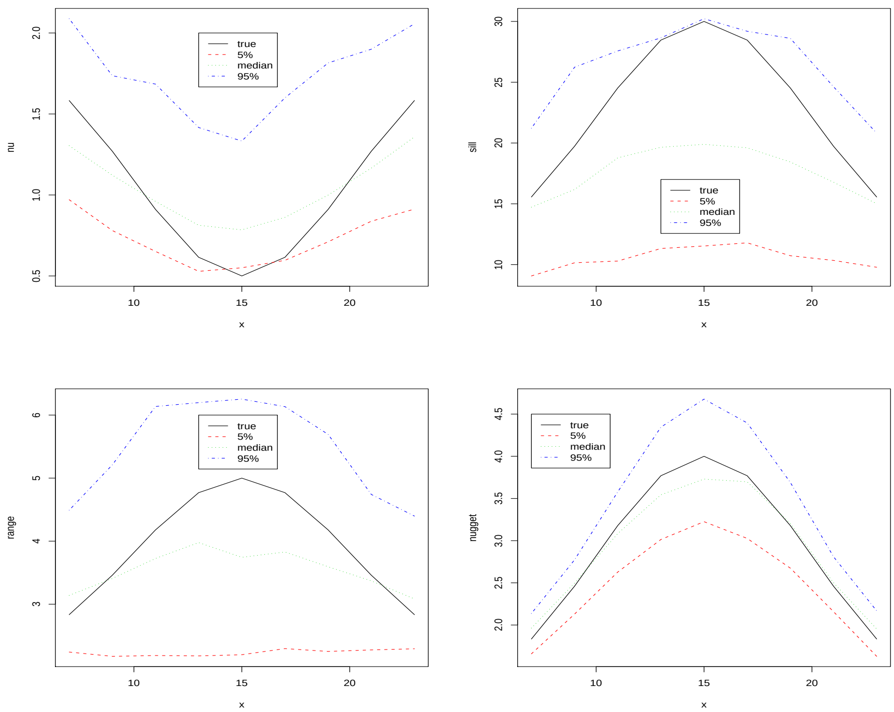

FIGURE 1. The nonparametric estimators for the parameters ν, τ, ρ, T over the region $\{(x,x)|5 \leq x \leq 25\}$.

a function of the location over the line $\{(x,x)|10 \leq x \leq 20\}$. We find that the 90 % confidence intervals of nonparametric estimators for all the parameters include the true values. The nonparametric estimation accounts for the local dependency of the parameters considerably well. The estimation performance for the nugget parameter is better than that of the other parameters.

Also for 100 sets of the simulated data, we fit cubic B-spline functions to $(\nu(s), \tau(s), \rho(s), T(s))$ over the region $[5,25] \times [5,25]$ to construct the parametric estimators for them. A cubic B-spline function, which has minimal support with respect to a given degree, smoothness and domain partition, is numerically evaluated by the de Boor algorithm (de Boor, 2001). Choose the knots (a_i, b_j) for $i, j = 0, \ldots, 8$ over the region by

$$a_i = 5 + 5(i-2), b_j = 5 + 5(j-2),$$

and construct the basis functions $c_{p,m}$ and $d_{q,m}$ for $m = 1,2,3$ and $p, q = 0, \ldots, 8 - m$ by the recursion formula,

$$\begin{aligned} c_{p,m}(s_1) &= \frac{s_1 - a_p}{a_{p+m} - a_p} c_{p,m-1}(s_1) + \frac{a_{p+m+1} - s_1}{a_{p+m+1} - a_{p+1}} c_{p+1,m-1}(s_1), \\ d_{q,m}(s_2) &= \frac{s_2 - b_q}{b_{q+m} - b_q} d_{q,m-1}(s_2) + \frac{b_{q+m+1} - s_2}{b_{q+m+1} - b_{q+1}} d_{q+1,m-1}(s_2), \end{aligned}$$

ν				τ			
true	med.	5%	95%	true	med.	5%	95%
2.85	2.10	1.50	3.58	17.53	17.60	10.48	32.61
1.77	1.62	0.93	2.91	16.49	12.22	5.33	33.81
0.60	0.68	0.19	1.56	27.39	19.26	9.05	36.13
ρ				T			
3.39	4.72	3.35	7.17	1.87	2.11	1.75	2.48
3.02	2.68	1.64	4.90	1.93	2.12	1.76	2.62
4.59	4.07	1.69	8.23	3.63	3.50	2.66	4.10

TABLE 1. The 5, 95 percentiles and medain of the estimated parameters of the cubic B-splines fitted to $\nu(s), \tau(s), \rho(s)$ and $T(s)$ over $[5, 25]^2$, where only the three elements in each case are selected to be shown.

with the initial values given by, for $p, q = 0, \dots, 7$,

$$
\begin{aligned}
c_{p,0}(s_1) &= \begin{cases} 1 & \text{if } a_p \le s_1 < a_{p+1}, \\ 0 & \text{otherwise} \end{cases} \\
d_{q,0}(s_2) &= \begin{cases} 1 & \text{if } b_q \le s_2 < b_{q+1}. \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

By using the basis cubic B-splines $c_{p,3}(s_1) \times d_{q,3}(s_2)$ for $p, q = 0, \dots, 4$ as a basis for our parametric function, the parametric function is given by the linear combination with a parameter θ, namely by

$$
\sum_{p,q=0}^{4} \theta_{p,q} c_{p,3}(s_1) \times d_{q,3}(s_2), \tag{16}
$$

which is fitted to $\nu(s), \tau(s), \rho(s)$ and $T(s)$ over the region $s \in [5, 25]^2$. The dimension of the parameters results in $25 \times 4 = 100$.

We estimated the parameters $\theta_{p,q}$ that identifies each one of $\nu(s), \tau(s), \rho(s)$ and $T(s)$ for 100 sets of simulated data by minimizing the Whittle likelihood functions in (8), when we design the mesh points by $(5 + 2i, 5 + 2j)$ for $i, j = 0, \dots, 10$ over $[5, 25] \times [5, 25]$. Here the optimization procedure was conducted numerically by the BFGS method with the initial values evaluated by the nonparametric estimators in the first experiment. The median, the 5 and 95 percentiles of the estimated parameters were evaluated in comparison with the theoretical values calculated by fitting (16) directly to (15). In Table 1, typical 3 estimators are selected to be shown among 25 ones for each of $\nu(s), \tau(s), \rho(s)$ and $T(s)$ to save space. We find that all the parameters are estimated well enough for the 90 % confidence regions to include the true values. The local dependency of the nugget is estimated best among those of the four parameters as in the first experiment.

5.2. **Applications to real data.** This section applies the nonstationary regression model in (1) to land price data, which were collected by the Japanese Ministry of Land, Infrastructure and Transport in 2001. This is the record of land prices (yen per square meter) with longitudes and latitudes of 5573 irregularly spaced sampling points in the residential areas around Tokyo, which is shown in Figure 2, where the co-ordinates are modified with units of kilometers. Land price in each point attaches

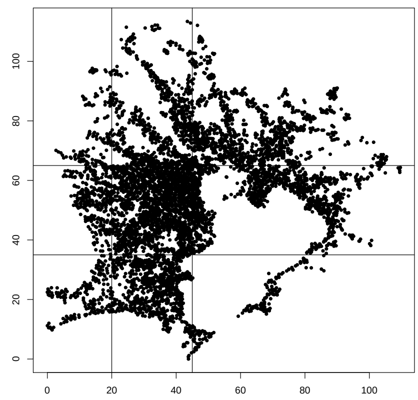

FIGURE 2. The points on which land prices are observed with one unit being adjusted to 1 km. The nonstationary structure is analyzed by local dependency of the spectrum over $\{(x, y)|20 \leq x \leq 45, 35 \leq y \leq 65\}$.

the time in minutes it takes by train from the central Tokyo and the distance in meter from the nearest train station, which we use as the independent variables denoted as d_1 and d_2, respectively.

Cubic B-splines are fitted to describe the local dependencies over the region $B = [20, 45] \times [35, 65]$ of all the parameters that we concern, *i.e.*, the regression coefficients in the mean function and the parameters in the Matérn class in (4) that we fit to the covariance and the nugget variance. Here the knots for the spline functions are selected as

$$a_i = 20 + 25(i - 2)/3, b_j = 35 + 30(j - 2)/4,$$

for $i = 0, \ldots, 7$ and $j = 0, \ldots, 8$, and the cubic B-spline functions are designed exactly as in (16).

The fitted curve of the cubic B-splines that minimize (8) for ν, τ, ρ and T are shown in Figure 3, while the fitted curve of the splines estimated by (10) for the regression coefficients of d_1 and d_2, denoted as β_{d_1} and β_{d_2}, are shown in Figure 4. Here D was designed as first 600 elements in Ω_n^+ with the bandwidth $h = 6^2$ and the mesh points in $(20 + 2i, 30 + 2j)$ for $i = 0, \ldots, 12$ and $j = 0, \ldots, 17$.

Figure 3 detects attracting features of the Matérn class parameters. The area where the smoothness parameter is estimated as high around $(25, 50)$ is the residential region developed in recent periods called Tama new towmn. The highly estimated smoothness corresponds with the reasonable expectation that land price varies more smoothly in recently developed areas than in old areas because the developments conducted together at one time work to ease individual effects caused by various environmental factors. The sill and nugget parameters are estimated as higher in the regions near the center of Tokyo than in the suburbs. This may be because land price in the city center becomes more sensitive to environmental factors such as access to sunlight, distances from roads, train lines and so on.

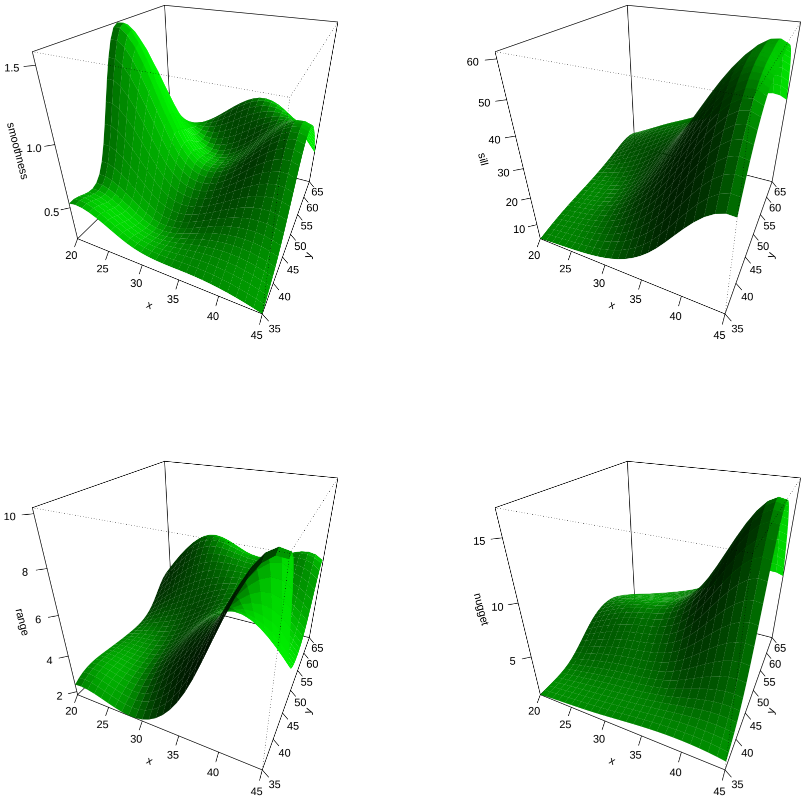

FIGURE 3. The fitted curve for the parameters ν, τ and ρ for the Matérn class and nugget T over $[20, 45] \times [35, 65]$ by the cubic B-spline functions.

Figure 4 detects well the tendency that the negative impact of the distance from the city center and the nearest station becomes higher in the regions near the city center than in the suburbs. This is a reasonable nature of land price.

6. Concluding remarks

Let us state two final comments on the estimation procedures. The first one is the use of the frequency domain approach that works efficiently for large spatial data that may be nonstationary with multivariate observations. Our method requires no modeling of covariances but that of spectral densities in every estimation except for the mean function estimation. As in Im *et al.* (2007), difficulty to check the positive definiteness of covariances often makes us propose covariance models through spectral densities, in which case our frequency domain approach works.

The final one is an easy extension of our method to spatial temporal data, which may be nonstationary both in space and time. The exactly same procedure can be applied to estimate spectral densities on spatial and temporal domains with local

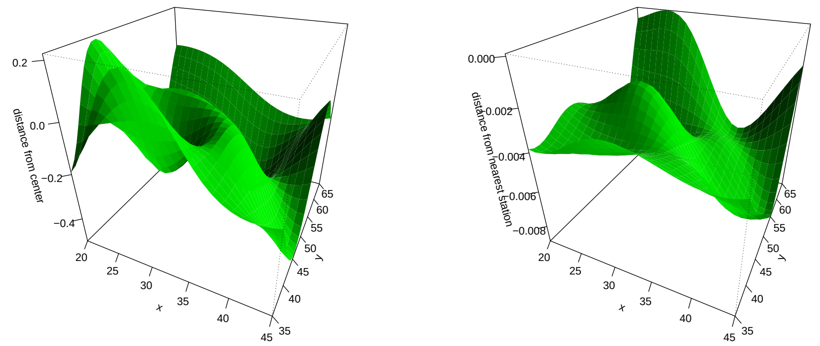

FIGURE 4. The fitted curve for β_{d_1} and β_{d_2} over $[20, 45]\times[35, 65]$ by the cubic B-spline functions, where β_{d_1} and β_{d_2} are the regression coefficients for time distance by train in minutes from the central Tokyo and for distances from the nearest station in meter, respectively.

dependent parameters. The crucial point to be noted is that the separability of Fourier basis functions in space and time given by

$$\exp\{i(\omega' s + \lambda' t)\} \quad = \quad \exp(i\omega' s)\exp(i\lambda' t)$$

makes it possible to orthogonalize them separately, which means that the orthogonalization procedure is conducted just by the union of the time consumed by it on spatial and time domains. Hence our procedure can provide nonstationary spatial temporal data with an efficient way of analysis.

It follows that our method is of greater use than existing spatial domain methods are in the progress of technologies that can collect huge data sets in the forms of spatial or spatial temporal data located in several tens of thousand of data points, which may often show nonstationarity.

7. LEMMAS

Lemma 1. *If Assumptions 1, 2 and 3 hold, then for $\omega_k \in \Omega_n^+$ and $u \in [0,1]^2$,*

$$EI(uB, \omega_k) \quad = \quad F(\theta(u), \omega_k) + C(u, \omega_k)T(u) + o(1),$$

under the asymptotics in Assumption 2.

Proof. From the definition in (13), we have

$$\sqrt{\frac{|B|}{4\pi^2 H_{uB}}} d(uB, \omega_k) \quad = \quad H_{uB}^{-1/2} \sum_{\omega_j \in \Omega} \sum_{p=1}^{n} A\left(\frac{s_p}{B}, \omega_j\right) z_j \exp(i\omega_j' s_p) \overline{\xi_{uB}(\omega_k, s_p)} w\left(\frac{s_p - uB}{h}\right)$$
$$+ \sqrt{\frac{|B|}{4\pi^2 H_{uB}}} \sum_{p=1}^{n} \varepsilon_p \overline{\xi_{uB}(\omega_k, s_p)} w\left(\frac{s_p - uB}{h}\right).$$

It follows that

$$EI(uB, \omega_k) \quad = \quad \sum_{\omega_j \in \Omega} F(\theta(u), \omega_j) K(\omega_j, \omega_k) + b(uB, \omega_k, \omega_k) T(u) + o(1),$$

where

$$K(\omega_j,\omega_k) \quad = \quad H_{uB}^{-1}\left|\sum_{p=1}^{n}\exp(i\omega_j' s_p)\overline{\xi_{uB}(\omega_k,s_p)}w\left(\frac{s_p-uB}{h}\right)\right|^2,$$

and where

$$\begin{aligned}\sum_{\omega_j\in\Omega}K(\omega_j,\omega_k) &= \sum_{\omega_j\in\Omega_n}K(\omega_j,\omega_k)+o(1)\\ &= \sum_{\omega_j\in\Omega_n}K(\omega_k,\omega_j)+o(1)=1+o(1),\end{aligned}$$

by Parseval equality. It follows that, for $p,q=1,\ldots,a$,

$$\begin{aligned}&\left|\sum_{\omega_j\in\Omega}F_{pq}(\theta(u),\omega_j)K(\omega_j,\omega_k)-F_{pq}(\theta(u),\omega_k)\right|\\ =\;&\left|\sum_{\omega_j\in\Omega}\left\{F_{pq}(\theta(u),\omega_j)-F_{pq}(\theta(u),\omega_k)\right\}K(\omega_j,\omega_k)\right|+o(1)\\ \le\;&\sum_{|\omega_j-\omega_k|\le h^{-1}}|F_{pq}(\omega_j)-F_{pq}(\omega_k)|\,K(\omega_j,\omega_k)+C\sum_{|\omega_j-\omega_k|>h^{-1}}K(\omega_j,\omega_k)+o(1)\\ =\;&O\left(h^{-1}\right)+o(1)=o(1),\end{aligned}$$

by the mean value theorem, which completes the proof.

Lemma 2. *If Assumptions 1, 2 and 3 hold, for $u\in[0,1]^2$ and $\omega_k,\omega_l\in\Omega_n^+$ such that $|\omega_k-\omega_l|>h^{-1}$,*

$$V_{pq}(u,\omega_k,\omega_l) \quad := \quad \frac{|B|}{4\pi^2 H_{uB}}\mathrm{Cov}\left(d_p(uB,\omega_k),\overline{d_q(uB,\omega_l)}\right)=o(1),$$

under the asymptotics in Assumption 2.

Proof. By the definition of $d(uB,\omega)$, we have

$$V_{pq}(u,\omega_k,\omega_l) \quad = \quad \sum_{\omega_j\in\Omega}F_{pq}(\theta(u),\omega_j)L(\omega_j,\omega_k,\omega_j,\omega_l)+b(uB,\omega_k,\omega_l)T_{pq}(u)+o(1),$$

where

$$L(\omega_j,\omega_k,\omega_j,\omega_l)=H_{uB}^{-1}\times$$
$$\left\{\sum_{p=1}^{n}\exp(i\omega_j' s_p)\overline{\xi_{uB}(\omega_k,s_p)}w\left(\frac{s_p-uB}{h}\right)\right\}\overline{\left\{\sum_{p=1}^{n}\exp(i\omega_j' s_p)\overline{\xi_{uB}(\omega_l,s_p)}w\left(\frac{s_p-uB}{h}\right)\right\}}.$$

Since

$$\begin{aligned}\sum_{\omega_j\in\Omega}L(\omega_j,\omega_k,\omega_j,\omega_l) &= \sum_{\omega_j\in\Omega_n}L(\omega_k,\omega_j,\omega_l,\omega_j)+o(1)\\ &= H_{uB}^{-1}\sum_{p=1}^{n}\exp\left\{i(\omega_k-\omega_l)' s_p\right\}w\left(\frac{s_p-uB}{h}\right)+o(1),\end{aligned}$$

by Parseval equality, and $\sum_{\omega_j\in\Omega}|L(\omega_j,\omega_k,\omega_j,\omega_l)|\le 1$ by Schwartz inequality, $V_{pq}(u,\omega_k,\omega_l)$ is bounded by

$$\begin{aligned}&\sum_{\omega_j\in\Omega}|F_{pq}(\theta(u),\omega_j)-F_{pq}(\theta(u),\omega_k)|\,|L(\omega_j,\omega_k,\omega_j,\omega_l)|\\ +\;&F_{pq}(\theta(u),\omega_k)H_{uB}^{-1}\left|\sum_{p=1}^{n}\exp\left\{i(\omega_k-\omega_l)s_p\right\}w\left(\frac{s_p-uB}{h}\right)\right|+|b(uB,\omega_k,\omega_l)|\,|T_{pq}(u)|\,.\end{aligned}$$

The first term converges to 0 by the same argument in Lemma 1, while the second and third terms converge to 0 under Assumption 2.

Lemma 3. *If Assumptions 1, 2 and 3 hold, for $u \in [0,1]^2$, a finite region $D \subset \mathbb{R}^2$ and a continuous function $G(\lambda)$ on D,*

$$4\pi^2|B|^{-1}\sum_{\omega_k\in D} tr[G(\omega_k)I(uB,\omega_k)] \to \int_D tr[G(\lambda)\{F(\theta(u),\lambda)+C(u,\lambda)T(u)\}]\,d\lambda,$$

in probability under the asymptotics in Assumption 2.

Proof. The expectation of the left hand side is

$$4\pi^2|B|^{-1}\sum_{\omega_k\in D}\mathrm{tr}\left[G(\omega_k)\{F(\theta(u),\omega_k)+C(u,\omega_j)T(u)\}\right]+o(1),$$

and the first term converges to the right hand side by Lemma 1.

The variance is evaluated by $\sum_{p,q,p',q'=1}^{a} Q_{pqp'q'}$, where $Q_{pqp'q'}$ is

$$\begin{aligned}
& 16\pi^4|B|^{-2}\sum_{\omega_k,\omega_l\in D} G_{pq}(\omega_k)G_{p'q'}(\omega_l)\left[E\{I_{qp}(uB,\omega_k)I_{q'p'}(uB,\omega_l)\}-EI_{qp}(uB,\omega_k)EI_{q'p'}(uB,\omega_l)\right]\\
= \ & 16\pi^4|B|^{-2}\sum_{\omega_k,\omega_l\in D} G_{pq}(\omega_k)G_{p'q'}(\omega_l)\frac{|B|^2}{16\pi^4H_{uB}^2}\mathrm{cum}\left(d_q(uB,\omega_k),\overline{d_p(uB,\omega_k)},d_{q'}(uB,\omega_l),\overline{d_{p'}(uB,\omega_l)}\right)\\
& +16\pi^4|B|^{-2}\sum_{\omega_k,\omega_l\in D} V_{pq}(u,\omega_k,\omega_l)\overline{V_{p'q'}(u,\omega_k,\omega_l)}G_{pq}(\omega_k)G_{p'q'}(\omega_l)\\
& +16\pi^4|B|^{-2}\sum_{\omega_k,\omega_l\in D} V_{pq}(u,\omega_k,-\omega_l)\overline{V_{p'q'}(u,\omega_k,-\omega_l)}G_{pq}(\omega_k)G_{p'q'}(\omega_l).
\end{aligned}$$

The first term is bounded as

$$\begin{aligned}
& C_1|B|^{-2}\sum_{\omega_j\in\Omega} F_{qp}(\theta(u),\omega_j)\overline{F_{q'p'}(\theta(u),\omega_j)}\left[\sum_{\omega_k\in D} G_{pq}(\omega_k)K(\omega_j,\omega_k)\sum_{\omega_l\in D}\overline{G_{p'q'}(\omega_l)K(\omega_j,\omega_l)}\right]\\
& +C_2|B|^{-2}\sum_{\omega_k,\omega_l\in D} G_{pq}(\omega_k)G_{p'q'}(\omega_l)b(u,\omega_k,\omega_l),
\end{aligned}$$

which converge to 0 by the dominated convergence theorem under Assumption 2, while the rest terms converge to 0 by Lemma 2 and the dominated convergence theorem.

Lemma 4. *If Assumptions 1, 2 and 3 hold, for $\omega_k,\omega_l \in \Omega_n^+$ and for $u_1,u_2 \in [0,1]^2$ such that $|u_1-u_2| > (B_i^{-1}h)^\rho$ for $i=1,2$ and $0<\rho<1$,*

$$W_{pq}(u_1,u_2,\omega_k,\omega_l) \ := \ \frac{|B|}{4\pi^2H_{u_1B}^{1/2}H_{u_2B}^{1/2}}\mathrm{Cov}\left(d_p(u_1B,\omega_k),\overline{d_q(u_2B,\omega_l)}\right)=o(1),$$

under the asymptotics in Assumption 2.

Proof.

$$W_{pq}(u_1,u_2,\omega_k,\omega_l) \ = \ \sum_{\omega_j\in\Omega} A_p(u_1,\omega_j)\overline{A_q(u_2,\omega_j)}'M(\omega_j,\omega_k,\omega_l)+N(\omega_k,\omega_l)+o(1),$$

where

$$\begin{aligned}
M(\omega_j,\omega_k,\omega_l) &= (H_{u_1B}H_{u_2B})^{-1/2}\left\{\sum_{p=1}^n \exp(i\omega_j s_p)\overline{\xi_{u_1B}(\omega_k,s_p)}w\left(\frac{s_p-u_1B}{h}\right)\right\}\\
&\quad\overline{\times\left\{\sum_{p=1}^n \exp(i\omega_j s_p)\overline{\xi_{u_2B}(\omega_l,s_p)}w\left(\frac{s_p-u_2B}{h}\right)\right\}} = o(1),\\
N(\omega_k,\omega_l) &= (4\pi)^{-2}|B|\,(H_{u_1B}H_{u_2B})^{-1/2}\\
&\quad\times\sum_{p=1}^n \xi_{u_1B}(\omega_k,s_p)\overline{\xi_{u_2B}(\omega_l,s_p)}w\left(\frac{s_p-u_1B}{h}\right)w\left(\frac{s_p-u_2B}{h}\right) = o(1),
\end{aligned}$$

under the assumption $|u_1-u_2| > (B_i^{-1}h)^\rho$ by the dominated convergence theorem. Since $\sum_{\omega_j\in\Omega}|M(\omega_j,\omega_k,\omega_l)| \le 1$ by Schwartz inequality, we have the conclusion by the dominated convergence theorem.

Lemma 5. *If Assumptions 1, 2 and 3 hold, for a finite region $D\subset\mathbb{R}^2$, a continuous function $G(u,\lambda)$ on $[0,1]^2\times D$ and u_j be a mesh point on $[0,1]^2$ given by $\left(m_1(B_1^{-1}h)^\rho, m_2(B_2^{-1}h)^\rho\right)$ for $m_1,m_2=1,2,\dots$ and $0<\rho<1$,*

$$\begin{aligned}
&4\pi^2|B|^{-1-\rho}h^{2\rho}\sum_{u_j\in[0,1]^2}\sum_{\omega_k\in D} tr[G(u_j,\omega_k)I(u_jB,\omega_k)]\\
&\to \int_{[0,1]^2}\int_D tr[G(u,\lambda)\{F(\theta(u),\lambda)+C(u,\lambda)T(u)\}]\,dud\lambda,
\end{aligned}$$

in probability under the asymptotics in Assumption 2.

Proof. The expectation of the left hand side is

$$4\pi^2|B|^{-1-\rho}h^{2\rho}\sum_{u_j\in[0,1]^2}\sum_{\omega_k\in D}\mathrm{tr}\,[G(u_j,\omega_k)\{F(\theta(u_j),\omega_k)+C(u,\lambda)T(u)\}]+o(1),$$

and the first term converges to the right hand side.

The variance is evaluated as $\sum_{p,q,p',q'=1}^a R_{pqp'q'}$ and $R_{pqp'q'}$ is

$$\begin{aligned}
&16\pi^4|B|^{-2-2\rho}h^{4\rho}\sum_{u_1,u_2\in[0,1]^2}\sum_{\omega_1,\omega_2\in D} G_{pq}(u_1,\omega_1)G_{p'q'}(u_2,\omega_2)\\
&\times[E\{I_{qp}(u_1B,\omega_1)I_{q'p'}(u_2B,\omega_2)\}-EI_{qp}(u_1B,\omega_1)EI_{q'p'}(u_2B,\omega_2)]\\
=\;&16\pi^4|B|^{-2-2\rho}h^{4\rho}\sum_{u_1,u_2\in[0,1]^2}\sum_{\omega_1,\omega_2\in D}\frac{|B|^2}{16\pi^4H_{u_1B}H_{u_2B}}G_{pq}(u_1,\omega_1)G_{p'q'}(u_2,\omega_2)\\
&\times\mathrm{cum}\left(d_q(u_1B,\omega_1),\overline{d_p(u_1B,\omega_1)},d_{q'}(u_2B,\omega_2),\overline{d_{p'}(u_2B,\omega_2)}\right)\\
&+16\pi^4|B|^{-2-2\rho}h^{4\rho}\sum_{u_1,u_2\in[0,1]^2}\sum_{\omega_1,\omega_2\in D}W_{qp}(u_1,u_2,\omega_1,\omega_2)\overline{W_{q'p'}(u_1,u_2,\omega_1,\omega_2)}G_{pq}(u_1,\omega_1)G_{p'q'}(u_2,\omega_2)\\
&+16\pi^4|B|^{-2-2rho}h^{4\rho}\sum_{u_1,u_2\in[0,1]^2}\sum_{\omega_1,\omega_2\in D}W_{qp}(u_1,u_2,\omega_1,-\omega_2)\overline{W_{q'p'}(u_1,u_2,\omega_1,-\omega_2)}G_{pq}(u_1,\omega_1)G_{p'q'}(u_2,\omega_2).
\end{aligned}$$

The first term is bounded by

$$\begin{aligned}
&C_1|B|^{-2-2\rho}h^{4\rho}\sum_{\omega_j\in\Omega}\sum_{u_1\in[0,1]^2}F_{qp}(\theta(u_1),\omega_j)\sum_{\omega_1\in D}G_{pq}(u_1,\omega_1)K(u_1,\omega_j,\omega_1)\\
&\times\sum_{u_2\in[0,1]^2}\overline{F_{q'p'}(\theta(u_2),\omega_j)\sum_{\omega_2\in D}G_{p'q'}(u_2,\omega_2)K(u_2,\omega_j,\omega_2)}\\
&+C_2|B|^{-2-2\rho}h^{4\rho}\sum_{\omega_1,\omega_2\in D}\sum_{u_1,u_2\in[0,1]^2}G_{pq}(\omega_1)G_{p'q'}(\omega_2)b_2(u_1,u_2,\omega_1,\omega_2),
\end{aligned}$$

which converge to 0 under Assumption 2, while the rest terms converge to 0 by Lemma 4.

References

[1] Bai, Y., Song, P. and Raghunathan, T. E. (2012) Joint composite estimating functions in spatio temporal models. *J. R. Statist. Soc. B.*, 74, 799-824.

[2] Banerjee, S., Gelfand A., Finley, A. O. and Sang, H. (2008) "Gaussian predictive process models for large spatial data sets," *J. R. Statist. Soc. B.*, 70, 825-848.

[3] Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods. Second Edition*, Springer-Verlag, New York.

[4] Dahlhaus, R. (1997) Fitting time series models to nonstationary processes, *Ann. Stat.*, 25, 1-37.

[5] de Boor, C. (2001) *A Practical Guide to spines*, Springer-Verlag, New York.

[6] Diggle, P. J. (2010) Histrical introduction. Handbook of Spatial Statistics, CRC Press.

[7] Im, H. K., Stein, M. L. and Zhu, Z. (2007) Semiparametric estimation of spectral density with irregular observations. *J. Am. Statist. Assoc.*, 102, 726-735.

[8] Kaufman, C. G., Schervish, M. J. and Nychka, W. (2008) Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Am. Statist. Assoc.*, 103, 1545-1555.

[9] Martin, W. and Flandlin, P. (1985) Wigner-Ville spectral analysis of nonstationary processes, *IEEE Trans. Acoust. Speech Signal Process.*, 33, 1461-1470.

[10] Matsuda, Y. and Yajima, Y. (2009) Fourier analysis of irregularly spaced data on $\mathbb{R}^d$, *J. R. Statist. Soc. B.*, 71, 191-217.

[11] Stein, M. L. (1999) *Interpolation of Spatial Data*, Springer-Verlag, New York.

[12] Stein, M. L., Chi, Z. and Welty, L. J. (2004) Approximating likelihoods for large spatial data sets. *J. R. Statist. Soc. B.*, 66, 275-296.

[13] Walker, A. M. (1964) Asymptotic properties of least squares estimates of parameter of the spectrum of a stationary non-deterministic time seres. *J. Aust. Math. Soc.*, 4, 363-384.

Graduate School of Economics and Management, Tohoku University, Sendai 980-8576, Japan

E-mail address: `matsuda@econ.tohoku.ac.jp`