TERG

Yasumasa Matsuda

TOHOKU ECONOMICS RESEARCH GROUP

WAVELET ANALYSIS OF SPATIO-TEMPORAL DATA

YASUMASA MATSUDA

ABSTRACT. This paper aims to provide a wavelet analysis for spatio-temporal data which are observed on irregularly spaced stations at discrete time points, where the spatial covariances show serious non-stationarity caused by local dependency. A specific example that is used for the demonstration is US precipitation data observed on about ten thousand stations in every month. By a reinterpretation of Whittle likelihood function for stationary time series, we propose a kind of Bayesian regression model for spatial data whose regressors are given by modified Haar wavelets and try a spatio-temporal extension by a state space approach. We also propose an empirical Bayes estimation for the parameters, which is regarded as a spatio-temporal extension of Whittle likelihood estimation originally defined for stationary time series. We conduct the extended Whittle estimate and compare mean square errors of the forecasts with those of some benchmarks to evaluate its goodness for the US precipitation data in August from 1987-1997.

1. INTRODUCTION

This paper focuses on analysis of spatio-temporal data, which is observations of regular time series at huge number of stations scattered irregularly over space. We aim at proposing a method of modelling, estimation and kriging for the kind of spatio-temporal data whose spatial covariances are not necessarily stationary.

Several kinds of models for space time covariances have been proposed to analyze spatio-temporal data. The simplest one is a separable covariance that is given by the product of spatial and temporal covariances, which makes it possible to give a covariance model separately in space and time. It provides an easy way to identify models for spatio-temporal data by fitting the product of the temporal and spatial covariances given by time series models such as autoregressive and moving average (ARMA) models (Brockwell and Davis, 1991) and spatial models such as Matérn class (Banerjee et al., 2004), respectively. Separable covariances, however, restrict covariance structures in the very narrow range in which temporal correlations on each spatial point must coincide. Nonseparable covariances can provide a practical class of space time covariances, but require a careful treatment to guarantee the positive definiteness of the covariance functions. Gneiting (2002) provides a sufficient condition for the positive definiteness and propose a useful class of space time covariances that satisfy it.

After a choice of space time covariance models, estimation of the parameters in the space time covariance model is a next step. Since the Gaussian maximum likelihood requires calculations of determinant and inverse of the covariance matrix whose dimension is the number of space time points of the data, MLE is in practice difficult to apply especially for huge data set. In order to avoid this computational

difficulty, several methods have been proposed in the literatures. Typical ones that represents them are the covariance tapering by Kaufman et. al. (2008), predictive process approach by Banerjee et. al. (2008) and the composite function approach by Bai et. al. (2012). The composite approach, which is regarded as the combined one of the other two methods, gave the most satisfactory performances in their empirical studies, while the number of space time points that the predictive approach can cope with is the largest of the three methods. In fact, the numbers of space time points that the three papers analyzed in their empirical studies were about 15,000, 7,000 and 3,000, respectively, and we see that as many as several thousand space time points are supposed in the composite approach, which is regarded as the one to represent the existing methods.

In this paper, we have an interest in analysis of huge data set observed at stations scattered irregularly over so broad space such as all over US continent that stationary spatial covariances are not necessarily satisfactory models. Specifically, our main object is to consider a method of modelling spatio-temporal data over as many as several hundred thousand space time points that shows crucial nonstationarity in spatial covariances, for which growing attentions have been paid recently in areas of both natural and social sciences by rapid progress of data collection technologies. The existing methods, which basically assume stationary covariances over as many as several thousand space time points, do not suppose the kind of spatio-temporal data as their objects of modelling.

This paper proposes a modelling by wavelets, which is an approach completely different from the existing methodology of covariance model fitting by maximizing a likelihood approximated in the clever ways to avoid large dimensional matrix operations. We will employ a kind of Bayesian regression model with regressors given by wavelets whose regression coefficients have prior distributions, which is considered as an extension of Fourier analysis of time series (Brockwell and Davis, 1991). The use of wavelets was originally proposed for analysis of nonstationary time series by Nason et. al. (2000). This paper extends the wavelet model from time series to spatio-temporal data by using the Haar wavelets that are modified to let them empirically orthogonalized under irregular sampling.

The striking features of the wavelet model for spatio-temporal data are the following three points. The first one is that it can express nonseparable space time covariances given with nonstationary spatial covariances and spatially dependent temporal covariances. The second one is that the use of empirically orthogonalized Haar wavelet makes it possible to calculate the likelihood function, kriging and forecasting efficiently by just three dimensional matrix operations in Kalman recursions, which means it opens a way of analysis for huge data set larger than several hundred thousand. The final one is that it allows any numbers of NAs by regarding the data as samples of continuous stochastic process expanded by the modified Haar wavelets.

US precipitation data observed at 11918 stations scattered over US continent in August from 1987 to 1997 is a typical example that has the features we focus in this paper. The space time points are about 70,000 with clear appearance of crucial nonstationarity caused by local dependency of rain falls and the existence of frequent NAs. We fit the wavelet models to the data and examine the goodness of fit by performances of kriging and forecasting in order to demonstrate our methodology by the wavelet modelling.

2. FOURIER ANALYSIS OF STATIONARY TIME SERIES

Let $\{X_t\}, t \in \mathbb{Z}$ be stationary time series with the covariance function

$$\gamma(\theta, h) = \text{Cov}(X_t, X_{t-h}),$$

for a parameter $\theta \in \Theta$. We assume that the spectral density function $f(\theta, \lambda)$ that satisfies

$$(1) \qquad \gamma(\theta, h) = \int_{-\pi}^{\pi} \exp(ih\lambda) f(\theta, \lambda) d\lambda,$$

exists.

To estimate the parameter $\theta \in \Theta$ by the observations $X = (X_1, \ldots, X_n)'$, we usually use the maximum likelihood estimator (MLE) $\hat{\theta}$ that maximizes the log likelihood function

$$\log L(\theta) = -\frac{1}{2} |\Gamma_n(\theta)| - \frac{1}{2} X' \Gamma_n(\theta)^{-1} X,$$

where $\Gamma_n(\theta)$ is the n by n covariance matrix whose (i,j)th element is given by $\gamma(\theta, i - j)$. Since the calculation of the log likelihood function is sometimes time consuming especially for large sample sizes because of large dimensionality of the covariance matrix, we often use an approximated likelihood called Whittle likelihood function given by

$$(2) \qquad \log L_w(\theta) = -\frac{1}{2} \sum_{k=1}^{[n/2]} \left\{ \log f(\theta, \omega_k) + \frac{I(\omega_k)}{f(\theta, \omega_k)} \right\},$$

where $\omega_k = 2\pi k/n$, which is called Fourier frequency, and $I(\lambda)$ is the periodogram given by the squared discrete Fourier transform (DFT) on λ, namely by

$$(3) \qquad I(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t \exp(-i\lambda t) \right|^2.$$

We will give an interpretation for the Whittle likelihood function to show intuitively how it approximates the exact likelihood function. Though many time series researchers have examined the asymptotic equivalence between the estimators that maximize $\log L(\theta)$ and $\log L_w(\theta)$ (see e.g. Proposition 10.8.3 of Brockwell and Davis, 1991), the theories do not directly provide how the exact likelihood is related with the Whittle likelihood.

Let us consider what model the Whittle likelihood function is the exact likelihood function for. For independent random variables $\alpha_k, k = -n/2, \ldots, n/2$ with mean 0 and variance $f(\theta, \omega_k)$, let us consider the following model

$$(4) \qquad \tilde{X}_t = \sqrt{\frac{2\pi}{n}} \sum_{k=-[n/2]}^{[n/2]} \alpha_k \exp(i\omega_k t),$$

which is regarded as a Bayesian regression model whose regression coefficients have the prior distribution with no error terms. Since the regressors $\exp(i\omega_k t)$ for $k = -[n/2], \ldots, [n/2]$ constitute an orthonormal basis that satisfies

$$\frac{1}{n} \sum_{t=1}^{n} \exp(i\omega_p t) \overline{\exp(i\omega_q t)} = \delta_{pq},$$

for Kronecker delta δ_{pq}, the DFT of $\tilde{X}_t$ given by

$$(5) \qquad \sqrt{\frac{1}{2\pi n}} \sum_{t=1}^{n} \tilde{X}_t \exp(-i\omega_k t)$$

reduces to α_k. It follows naturally that the likelihood function for $\tilde{X}_t$ is equal to the Whittle likelihood function (2) by considering the likelihood for DFT instead of $\tilde{X}_t$. In other words, the Whittle likelihood in (2) for X_t is regarded as the likelihood for $\tilde{X}_t$ in (4), and the use of the Whittle likelihood function for MLE is equivalent to approximating X_t by $\tilde{X}_t$.

The approximation of X_t by $\tilde{X}_t$ is justified in the sense that they have asymptotically equivalent covariance functions. In fact, by simple calculation, we have

$$(6) \qquad \mathrm{Cov}(\tilde{X}_t, \tilde{X}_{t-h}) = \frac{2\pi}{n} \sum_{k=-[n/2]}^{[n/2]} \exp(ih\omega_k) f(\theta, \omega_k),$$

which is a Riemannian approximation for $\gamma(\theta, h)$ in (1).

The frequency domain expression in (4) for time series, which we derive from an interpretation for the Whittle likelihood, plays a significant role for extension from time series models to those for spatio-temporal data.

3. Extension to spatio-temporal data

This section proposes a model for spatio-temporal data by an extension of the frequency domain expression for time series in (4). The crucial point for the extension is to find an orthogonal basis for irregularly spaced data points, as the orthogonality of the sinusoidal functions for time series plays a significant role in deriving the Whittle likelihood.

3.1. **Modified Haar wavelets.** For the extension of the model in (4) from time series to spatio-temporal data, it is necessary to find a new orthogonal basis under the inner product defined by observation stations for spatio-temporal data. Let $s_j, j = 1, \ldots, n$ be the locations of observation stations of spatio-temporal data, which we suppose are scattered irregularly over $[0, 1]^2$. For two functions f and g defined on s_j, let us define the inner product by

$$(7) \qquad < f, g >:= \sum_{j=1}^{n} f(s_j) g(s_j).$$

The sinusoidal function $\exp(i\omega s)$ no more constitutes an orthogonal basis under the inner product except for the cases when the stations are regularly spaced to provide a mesh data.

We construct an orthogonal basis under the inner product by modifying the two dimensional Haar wavelets. Let us start from the introduction of the one and two dimensional Haar wavelets that constitute an orthogonal basis over $[0, 1]$ and $[0, 1]^2$, respectively, under the Lebesgue measure. The mother wavelet over $[0, 1]$ denoted as $\psi(x)$, which is shown in the upper of Figure 1, produces the orthogonal basis on $L^2([0, 1])$ given by

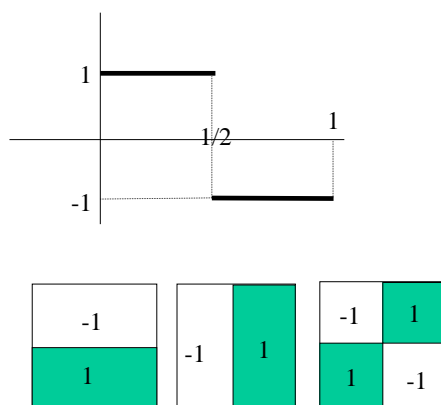$$\phi_{k,j}(x) := \psi(2^k x - j),$$

FIGURE 1. The Haar wavelets on $[0,1]$ and on $[0,1]^2$.

for $k = 0, 1, \ldots, j = 0, \ldots, 2^k - 1$, under the Lebesgue measure (see Theorem 1.1 of Ogden (1997)). In other words, the one dimensional Haar wavelets satisfy

$$\int_{[0,1]} \phi_{k,j}(x)\phi_{k',j'}(x)dx = c_{k,j}\delta_{kk'}\delta_{jj'}$$

for the positive constant $c_{k,j} = 4^{-k}$.

Similarly the mother wavelets on $[0,1]^2$, which are given by the three functions $\psi_1(x), \psi_2(x)$ and $\psi_3(x)$ defined over $[0,1]^2$ in lower part of Figure 1, produce the orthogonal basis on $L^2([0,1]^2)$ given by

$\phi_{k,i,j}(x_1,x_2) :=$

(8) $\quad \left(\psi_1(2^k x_1 - i, 2^k x_2 - j), \psi_2(2^k x_1 - i, 2^k x_2 - j), \psi_3(2^k x_1 - i, 2^k x_2 - j)\right),$

for $k = 0, 1, \ldots, i, j = 0, \ldots, 2^k - 1$, under the Lebesgue measure (see Ogden(1997, page 167)). In other words, the two dimensional Haar wavelets satisfy

$$\int_{[0,1]^2} \phi'_{k,ij}(x_1, x_2)\phi_{k',i'j'}(x_1, x_2)dx_1 dx_2 = C_{k,ij}\delta_{kk'}\delta_{ii'}\delta_{jj'}$$

for the positive definite diagonal matrix $C_{k,ij} = 4^{-k}I_3$.

Let us modify the two dimensional Haar wavelets to let them constitute the orthogonal basis under the empirical measure in (7) for irregularly spaced stations $s_j, j = 1, \ldots, n$ scattered over $[0,1]^2$. Let $N(D)$ be the number of the stations included in $D \subset [0,1]^2$, and $D_{k,ij} \subset [0,1]^2$ be the support for the Haar wavelet $\phi_{k,ij}$, which is give by

$$D_{k,ij} = \{(x_1,x_2)|i/2^k \leq x_1 \leq (i+1)/2^k, j/2^k \leq x_2 \leq (j+1)/2^k\} \subset [0,1]^2.$$

Definition 1. *In the subregions $F^l_{k,ij}, G^l_{k,ij}$ in $D_{k,ij}$ over which the lth component of the Haar wavelet $\phi_{k,ij}$ takes values 1 and -1, we modify the values 1 and -1 to*

those given by

$$\sqrt{\frac{N\left(G_{k,ij}^l\right)}{N\left(F_{k,ij}^l\right)}} \ \ and \ -\sqrt{\frac{N\left(F_{k,ij}^l\right)}{N\left(G_{k,ij}^l\right)}},$$

respectively, which we denote as $\tilde{\psi}_{k,ij}^l(x_1, x_2)$ for $l = 1, 2, 3$. Then the modified Haar wavelet is defined by

$$\tilde{\phi}_{k,ij}(x_1, x_2) = \left(\tilde{\psi}_{k,ij}^1(x_1, x_2), \tilde{\psi}_{k,ij}^2(x_1, x_2), \tilde{\psi}_{k,ij}^3(x_1, x_2)\right). \tag{9}$$

By simple calculation, it is confirmed that the modified Haar wavelets constitute a block orthogonal basis under the empirical measure in (7), namely, satisfy

$$\sum_{p=1}^{n} \tilde{\phi}_{k,ij}'(s_p)\tilde{\phi}_{k',i'j'}(s_p) = L_{k,ij}\delta_{kk'}\delta_{ii'}\delta_{jj'}, \tag{10}$$

where $L_{k,ij}$ is not a diagonal matrix except for mesh data cases, which means that the three components are not necessarily mutually orthogonal, and is not always positive definite. Since the positive definiteness of $L_{k,ij}$ is crucial for modelling of spatio-temporal data, we introduce a sufficient condition for it.

Here we show a sufficient condition to guarantee the positive definiteness of 3×3 matrix $L_{k,ij}$ in (10), or equivalently the linear independence of the three n dimensional vectors given by $\tilde{\phi}_{k,ij} = \left(\tilde{\phi}_{k,ij}'(s_1), \ldots, \tilde{\phi}_{k,ij}'(s_n)\right)'$. Let us denote the four disjoint equal-area sub-regions in $D_{k,ij}$ generated from the division of $D_{k,ij}$ by the lines $x_1 = (i + 0.5)/2^k$ and $x_2 = (j + 0.5)/2^k$ as $E_{k,ij}^l, l = 1, 2, 3, 4$.

Proposition 1. *For observation stations of $s_1, \ldots, s_n$ scattered irregularly over $[0, 1]^2$, let $\tilde{\phi}_{k,ij}(s), k = 0, 1, \ldots, i, j = 0. \ldots, 2^k - 1$ be the modified Haar wavelets. A sufficient condition for the 3×3 matrix $L_{k,ij}$ to be positive definite is that each of the four disjoint sub-regions $E_{k,ij}^l, l = 1, 2, 3, 4$ in $D_{k,ij}$ contains at least one observation station.*

The proof is given in Section 6. Let P be the set of (k, i, j)s that satisfy the sufficient condition for the positive definiteness of $L_{k,ij}$. In the following subsections, the block orthogonal basis $\tilde{\phi}_{k,ij}$ for $(k, i, j) \in P$ plays a role in spatial and spatio-temporal extension of the frequency domain expression in (4) for time series.

3.2. **Wavelet model for spatio-temporal data.** We shall extend the frequency domain expression in (4) for time series to that for spatial and moreover spatio-temporal models by employing the modified Haar wavelets in (8) that constitute the block orthogonal basis under the empirical measure in (7). Let us start from the modelling of spatial data where observation stations are located in $s_p, p = 1, \ldots, n$ scattered irregularly over $[0, 1]^2$, for which we define P as the set of (k, i, j)s that guarantee the positive definiteness of $L_{k,ij}$ in (10) by Proposition 1. In analogy with the frequency domain expression in (4) for time series, we extend it to spatial data model by

$$Z(s_p) = \sum_{(k,i,j) \in P} \tilde{\phi}_{k,ij}(s_p)\beta_{k,ij} + \varepsilon_p, \tag{11}$$

where ε_p is independent and identically distributed random variables with mean 0 and variance σ^2, which we call nugget effect, and $\beta_{k,ij}$ is the three dimensional

independent random vector with mean 0 and variance matrix $A_{k,ij}I_3$ for the 3×3 identity matrix I_3, which is a wavelet version of spectral density in analogy with the time series expression in (4).

Unlike the stationary covariances given by (6) for (4), our proposed model in (11) has nonstationary covariances. In fact, by simple calculations, the covariance function is evaluated as

$$\text{Cov}\left(Z(s_p), Z(s_q)\right) = \sum_{(k,i,j) \in P} \tilde{\phi}_{k,ij}(s_p)\tilde{\phi}'_{k,ij}(s_q)A_{k,ij} + \sigma^2 \delta_{pq},$$

which shows nonstationarity by the nature of the Haar wavelets.

Next we shall conduct a spatio-temporal extension from the spatial model in (11) by state space models. Suppose that , for time points $t = 1, \ldots, T$, stations are located in $s_{t,p}, p = 1, \ldots, n_t$ that are scattered irregularly over $[0,1]^2$, on which we observe $Z(t, s_{t,p})$. We allow $s_{t,p}$ to depend on time to account for an existence of NA. Let $\tilde{\phi}_{k,ij}(t,s)$ be the modified Haar wavelets defined for the stations at time t, and P_t for the set of (k, i, j) that satisfy the sufficient condition for the positive definiteness of $L_{k,ij}$ in (10). Let $P_0 = \cap_{t=1}^T P_t$. For each t, we propose the spatial model in (11) to $Z(t, s_{t,p})$ and conduct temporal extension by regarding the regression coefficients $\beta_{k,ij}$ as the state vector that follows autoregressive models. Let $\beta_{k,ij}(1)$ be the initial state vectors that are three dimensional independent random vectors with mean 0 and variance matrix given by $A_{k,ij}I_3$, we propose the state space models for spatio-temporal data by

$$(12) \qquad Z(t, s_{t,p}) = \sum_{(k,i,j) \in P_0} \tilde{\phi}_{k,ij}(t, s_{t,p})\beta_{k,ij}(t) + \varepsilon_p(t),$$

$$\beta_{k,ij}(t+1) = \rho_{k,ij}\beta_{k,ij}(t) + u_{k,ij}(t),$$

for $t = 1, \ldots, T$, where $u_{k,ij}(t)$ is the three dimensional independent random vectors with mean 0 and variance matrix given by

$$(1 - \rho_{k,ij}^2)A_{k,ij}I_3,$$

and we assume that $|\rho_{k,ij}| < 1$.

The model in (12) can express non-separable space time covariance functions. In fact, by simple calculations, we have

$$(13) \qquad \text{Cov}\left(Z(t, u), Z(t-h, v)\right)$$
$$= \sum_{(k,i,j) \in P_0} \tilde{\phi}_{k,ij}(t, u)\tilde{\phi}'_{k,ij}(t-h, v)A_{k,ij}\rho_{k,ij}^h + \sigma^2 \delta_{uv}\delta_{t,t-h},$$

from which the spatial and temporal covariances are not separated unless $\rho_{k,ij}$ is a constant.

3.3. Parametric modelling of sapce time covariance functions.

We shall propose a specific parametric form for the functions $A_{k,ij}$ and $\rho_{k,ij}$ in the wavelet model (12). Let $D_{k,ij}$ be the support for the modified Haar wavelet $\tilde{\phi}_{k,ij}$. Since $A_{k,ij}$ and $\rho_{k,ij}$ are the quantities that describe the strength for the wavelet $\tilde{\phi}_{k,ij}$ defined over $D_{k,ij}$, they can be regarded as mappings from $D_{k,ij}$ to $\mathbf{R}$. The point for the modelling is how we find the parametric mappings to describe the space time covariance function in (13).

A sufficient condition for (13) to be well defined by the convergent series is that

$$\sum_{k=0}^{\infty} \max_{i,j} |A_{k,ij}| < \infty,$$

and $|\rho_{k,ij}| < 1$. As a model that satisfies the sufficient condition, we propose, for $0 < \tau_{k,ij} < 1$ and $B_{k,ij}$ that is positive and uniformly bounded,

$$A_{k,ij} = B_{k,ij}(1 - \tau_{k,ij})\tau_{k,ij}^k.$$

From the functional form in (13) for $h = 0$, it can be seen that $B_{k,ij}$ and $\tau_{k,ij}$ specify a magnitude and a smoothness for spatial covariance structures, respectively, while $\rho_{k,ij}$ describes temporal correlation structures.

We shall identify $B_{k,ij}, \tau_{k,ij}$ and $\rho_{k,ij}$ with the functions f_1, f_2 and f_3 defined over $[0,1]^2$ through the integral averages

$$(14) \qquad\qquad |D_{k,ij}|^{-1} \int_{D_{k,ij}} f_m(x)dx,$$

for $m = 1, 2$ and 3, respectively, where $|D_{k,ij}|$ is the area of the domain $D_{k,ij}$. The underlying function $f_m(x)$ is given by a Fourier expansion

$$(15) \qquad\qquad f_m(x) = a_0 + \sum_{p \in Q} \left\{ a_{m,p} \cos(\omega_p'x) + b_{m,p}\sin(\omega_p'x) \right\},$$

where ω_p is a frequency $(2\pi i_p/C_1, 2\pi j_p/C_2)'$ for the set Q of mesh points (i_p, j_p) except for the origin over the upper half plane of $\mathbf{R}^2$ and positive constants C_1 and C_2. It follows that the Fourier expansion for each of the underlying functions provides each of $B_{k,ij}, \tau_{k,ij}$ and $\rho_{k,ij}$ with the $2|Q|+1$ dimensional parametric form, which will give a specific parametric form for the space time covariance function as a result.

3.4. **Whittle likelihood estimate.** This section considers estimation for a parameter that describes the scalar functions $A_{k,ij}$ and $\rho_{k,ij}$ in the spatio-temporal model in (12), when we observe in the stations $s_{t,p}, t = 1, : \ldots, T, p = 1, \ldots, n_t$. To show explicitly the dependency on the parameter, we express the functions as $A_{k,ij}(\theta)$ and $\rho_{k,ij}(\theta)$ for the parameter $\theta \in \Theta$.

In analogy with the discrete Fourier transform for time series, let us define the wavelet transform for the spatial data $Z(t) = Z(t, s_{t,1}), \ldots, Z(t, s_{t,n_t}))'$ at time t. Let $\tilde{\Phi}_{k,ij}(t)$ be the $n_t \times 3$ matrix given by $\left(\tilde{\phi}'_{k,ij}(t, s_{t,1}), \ldots, \phi'_{k,ij}(t, s_{t,n_t}) \right)'$. Then the wavelet transform (WT) is defined by the ordinary least square (OLS) estimator for the coefficient on the regressor $\tilde{\Phi}_{k,ij}(t)$ for $(k,i,j) \in P_0$, namely by

$$(16) \qquad\qquad w_{k,ij}(t) = \left(\tilde{\Phi}'_{k,ij}(t)\tilde{\Phi}_{k,ij}(t) \right)^{-1} \tilde{\Phi}'_{k,ij}(t)Z(t),$$

which is the 3 dimensional vector that corresponds with the OLS estimator for $\beta_{k,ij}(t)$ in (12).

By the block orthogonality of $\tilde{\Phi}_{k,ij}(t)$ shown in (10), the state space model for $Z(t, s_{t,p})$ in (12) reduces to, for $(k,i,j) \in P_0$,

$$(17) \qquad\qquad w_{k,ij}(t) = \beta_{k,ij}(t) + f_{k,ij}(t),$$
$$\beta_{k,ij}(t+1) = \rho_{k,ij}(\theta)\beta_{k,ij}(t) + u_{k,ij}(t),$$

where $f_{k,ij}(t)$ is the independent three dimensional random vector with mean 0 and variance matrix $\sigma^2(\tilde{\Phi}'_{k,ij}(t)\tilde{\Phi}_{k,ij}(t))^{-1}$.

Let us evaluate the likelihood function for the WT that follows the state space expression in (17) by Kalman recursion. Put

$$\mu_{k,ij}(t) = E(\beta_{k,ij}(t)|w_{k,ij}(t),\ldots,w_{k,ij}(1)),$$
$$Q_{k,ij}(t) = Var(\beta_{k,ij}(t)|w_{k,ij}(t),\ldots,w_{k,ij}(1)),$$
$$m_{k,ij}(t+1) = E(\beta_{k,ij}(t+1)|w_{k,ij}(t),\ldots,w_{k,ij}(1)),$$
$$P_{k,ij}(t+1) = Var(\beta_{k,ij}(t+1)|w_{k,ij}(t),\ldots,w_{k,ij}(1)),$$

which are evaluated by the Kalman recursion when we initialize by $m_{k,ij}(1) = 0$ and $P_{k,ij}(1) = A_{k,ij}I_3$, namely by

$$v_{k,ij}(t) = w_{k,ij}(t) - m_{k,ij}(t),$$
$$F_{k,ij}(t) = P_{k,ij}(t) + \sigma^2\left(\tilde{\Phi}'_{k,ij}(t)\tilde{\Phi}_{k,ij}(t)\right)^{-1},$$
(18) $$\mu_{k,ij}(t) = m_{k,ij}(t) + P_{k,ij}(t)F_{k,ij}^{-1}(t)v_{k,ij}(t),$$
$$Q_{k,ij}(t) = P_{k,ij}(t) - P_{k,ij}(t)F_{k,ij}^{-1}(t)P_{k,ij}(t),$$
$$m_{k,ij}(t+1) = \rho_{k,ij}\mu_{k,ij}(t),$$
$$P_{k,ij}(t+1) = \rho_{k,ij}^2 Q_{k,ij}(t) + (1 - \rho_{k,ij}^2)A_{k,ij}I_3,$$

for $t = 1,\ldots,T$ and $(k,i,j) \in P_0$.

Define the 3×3 periodogram matrix in analogy with (3) for time series by

$$I_{k,ij}(t) := (w_{k,ij}(t) - m_{k,ij}(t,\theta))(w_{k,ij}(t) - m_{k,ij}(t,\theta))'.$$

Then the likelihood function for the WT is evaluated as

(19) $$\log L_{ww}(\theta) := -\frac{1}{2}\sum_{t=1}^{T}\sum_{(k,i,j)\in P_0}\left\{tr\left(I_{k,ij}(t)F_{k,ij}^{-1}(t,\theta)\right) + \log|F_{k,ij}(t,\theta)|\right\},$$

which we call the wavelet version of Whittle likelihood function in analogy with (2) for time series. The parameter θ that describes the functions $A_{k,ij}$ and $\rho_{k,ij}$ in (12) is estimated by maximizing the Whittle likelihood function, which provides an efficient algorithm conducted just by the three dimensional matrix operations.

3.5. **Kriging and forecasting.** Following the spatio-temporal data analysis literatures, we define estimation for the values on unknown points at $1 \le t \le T$ as kriging and that for the values on any points at $t > T$ as forecasting. Both kriging and forecasting can be conducted efficiently as a result of the Kalman filtering in (18). The kriging for a spatial point u at time $1 \le t \le T$ is given by

$$\sum_{(k,i,j)\in P_0}\tilde{\phi}_{k,ij}(t,u)\mu_{k,ij}(t)$$

for which the mean squared error is evaluated as

$$\sum_{(k,i,j)\in P_0}\tilde{\phi}_{k,ij}(t,u)Q_{k,ij}(t)\tilde{\phi}'_{k,ij}(t,u) + \sigma^2,$$

while the forecasting for a spatial point u at time $T + 1$ is given by

$$\sum_{(k,i,j)\in P_0}\tilde{\phi}_{k,ij}(T,u)m_{k,ij}(T+1)$$

for which the mean squared error is evaluated as

$$\sum_{(k,i,j)\in P_0} \tilde{\phi}_{k,ij}(T,u)P_{k,ij}(T+1)\tilde{\phi}'_{k,ij}(T,u) + \sigma^2.$$

3.6. **Mean function estimation.** We have been considering the case when the mean function is assumed to be 0 for spatio-temporal data. Here we will consider the case when the mean function is not ignored. Let $g(t,s)$ be a mean function at time t and spatial point s. We will consider the case when the mean function given by a regression form is inserted in the spatio-temporal model in (12), which is given by

$$(20) \qquad Z(t,s_{t,p}) = g(t,s_{t,p}) + \sum_{(k,i,j)\in P_0} \tilde{\phi}_{k,ij}(t,s_{t,p})\beta(t) + \varepsilon_{t,p},$$

$$g(t,s_{t,p}) = \sum_{i=1}^{q} X_i(t,s_{t,p})\alpha_i(t),$$

for independent variables $X_i(t,s)$ and temporally dependent regression coefficients $\alpha_i(t), i = 1, \ldots, q$.

First, we assume that the case when $A_{k,ij}$ and σ^2 are known in (20). Suppose we aim to estimate the coefficients $\alpha(t_0) = (\alpha_1(t_0), \ldots, \alpha_q(t_0))'$ at a temporal point $t_0, 1 \le t_0 \le T$. Let $\hat{\alpha}(t_0)$ be the initial estimate for $\alpha(t_0)$, which is typically the OLS

$$\hat{\alpha}(t_0) = \{X'(t_0)X(t_0)\}^{-1} X'(t_0)Z(t_0).$$

Follow the recursion:

(1) Calculate the wavelet transform in (16) at t_0 for the residual data given by

$$(21) \qquad \tilde{Z}(t_0,s_{t_0,p}) = Z(t_0,s_{t_0,p}) - \sum_{i=1}^{q} X_i(t_0,s_{t_0,p})\hat{\alpha}_i(t_0),$$

for $p = 1, \ldots, n_{t_0}$.

(2) Evaluate $\mu_{k,ij}(t_0)$ in the recursion (18) when we initialize the Kalman filter as $m_{k,ij}(t_0 - 1) = 0$ and $P_{k,ij}(t_0 - 1) = A_{k,ij}$.

(3) Calculate the residual process by

$$\hat{Z}(t_0,s_{t_0,p}) = Z(t_0,s_{t_0,p}) - \sum_{(k,i,j)\in P_0} \tilde{\phi}_{k,ij}(t_0,s_{t_0,p})\mu_{k,ij}(t_0)$$

for $p = 1, \ldots, n_{t_0}$.

(4) Estimate $\alpha(t_0)$ by the OLS for the residual

$$\hat{\alpha}(t_0) = \{X'(t_0)X(t_0)\}^{-1} X'(t_0)\hat{Z}(t_0).$$

(5) Return to (1).

Conducting the recursion until it converges, we have the best linear unbiased estimator for $\alpha(t_0)$ given by the converged one under the covariance structures identified by $A_{k,ij}$ and σ^2, which we call the wavelet version of generalized least square (GLS) estimator.

In the practical situation when $A_{k,ij}$ and σ^2 are unknown, we need to estimate them in prior to estimate the mean function. For the residual with the OLS for $\alpha(t_0)$, apply the Whittle likelihood estimation to the OLS residual to obtain the estimators for θ and σ^2. And replacing the parameters with the estimators in the

recursive procedures, we estimate the mean function by following the recursion with the estimated parameters until it converges.

4. EMPIRICAL EXAMPLES

This section examines the empirical performances of the wavelet modelling proposed in Section 3 by applying to real and simulated data. First, we apply it to US precipitation data, which are observations over about 70,000 space time points that shows crucial nonstationarity in the spatial covariances, to catch the empirical properties. Next, we conduct simulation studies to confirm whether they still hold generally for simulated samples generated by the exact model that we identify in the real analysis.
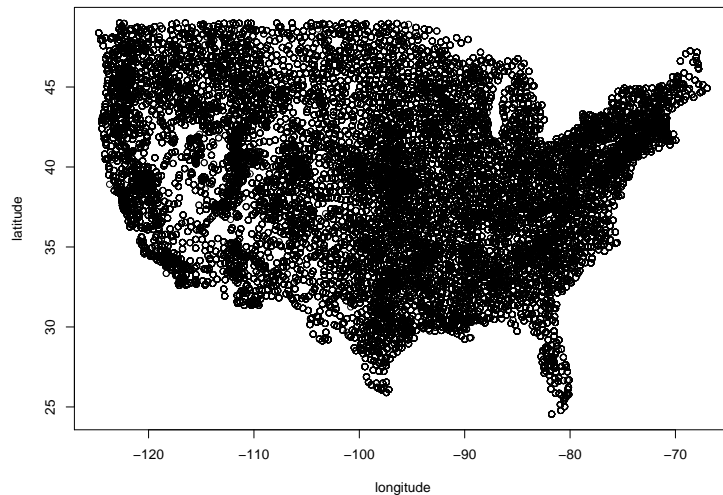

FIGURE 2. The locations of 11918 stations for US precipitation data.

4.1. **Applications to real data.** Let us start from real data analysis by the wavelet methods. We focus on US precipitation data, which are observations of the total monthly precipitation collected at 11918 stations scattered irregularly all over US, and will examine the empirical performances of the wavelet modelling by using specifically the yearly observations in August from 1987 to 1997.

The data are available in the web site:

http://www.image.ucar.edu/Data/US.monthly.met/USmonthlyMet.shtml.

The locations of the observation stations are shown in Figure 2, in which it is found that they are irregularly scattered all over US continent. All of the stations do not necessarily conduct observation every month. Table 1 shows the numbers of the stations that have the observations in August from 1987 to 1997, from which we find that NAs happened frequently.

We fit the wavelet model in (20) to the yearly precipitation in August for the ten years from 1987 to 1996, where we take constant and height of the stations as

year	87	88	89	90	91	92	93	94	95	96	97
no.	7042	7118	7176	7079	7009	7055	7043	6927	6823	6716	6747

TABLE 1. The numbers of stations that have the observations for
US monthly precipitation in August from 1987 to 1997.

regressors for the mean function. Based on the fitted model, we conduct kriging
and forecasting for some spatial points in 1997 to examine the goodness of fit.

Parametric model in (14) for $A_{k,ij}$ and $\rho_{k,ij}$ in the wavelet model (12) must be
specified for application to the yearly data in August. We will employ the specific
case when we identify $A_{k,ij}$ by

$$A_{k,ij} = B_{k,ij}(1 - \tau)\tau^k$$

for a constant τ that does not depend on k, i, j. It means that the smoothness for
the spatial covariances is assumed to be constant all over the US continent, while
the magnitudes of them and temporal correlations may be spatially dependent by
spatial dependencies of B and ρ, respectively And for the specific modelling of the
spatial dependent $B_{k,ij}$ and $\rho_{k,ij}$, we set $C_1 = C_2 = 1.2$ with $Q = \{(i,j)||i| \leq 2, 0 \leq
j \leq 2\} \cap \{(i,j)|i > 0 \text{ or } j > 0\}$ in (15), which means that we employ the model
with the 50 parameters for the data size of 69,988 as a result.

We conduct Whittle estimation in (19) to estimate the parameters in the model
when the mean function estimated by OLS is deleted from the original data in each
year from 1987 to 1996. And with the estimated parameters, we conduct kriging
and forecasting for some points in 1997 when we estimate the mean function by the
recursive procedure in section 3.6. For the locations of the points for kriging and
forecasting, we chose locations with high temporal correlations as well as randomly
scattered points to see the forecasting performances in relations with the temporal
correlations.

As a result of Whittle estimation, τ and σ^2 are estimated as 0.83 and 6.34,
respectively. Since the 24 parameters to describe each of $B_{k,ij}$ and $\rho_{k,ij}$ are too
many to show the estimated values here, we summarise them in Figure 3, which
shows the figures of the underlying functions in (14) for $B_{k,ij}$ and $\rho_{k,ij}$ identified
with the estimated parameters. Table 2 describes the averages of the squared errors
of kriging and forecasting for 100 points randomly chosen from 6747 stations in 1997
and those of forecasting for 125 points in 1997 that have the temporal correlations
identified as being larger than 0.95 by the underlying function for $\rho_{k,ij}$. Note that
kriging and forecasting were calculated by the model estimated with the data until
1996, as if the selected points in 1997 for kriging and forecasting were unknown.
As the benchmarks for comparisons, the weighted averages of the data in 1997 and
1996 for kriging and forecasting, respectively, are calculated by the normal kernel
with the bandwidths given by $0.1, 1, 2$ and ∞, where the one with the bandwidth
∞ means just the sample average of the data.

Figure 3, which are the identified figures of the underlying functions for $B_{k,ij}$
and $\rho_{k,ij}$, detects well the spatial dependencies of the spatial covariances and the
temporal correlations. And Table 2 shows that the wavelet model has the best
performances in kriging and forecasting except for the one case, which means the
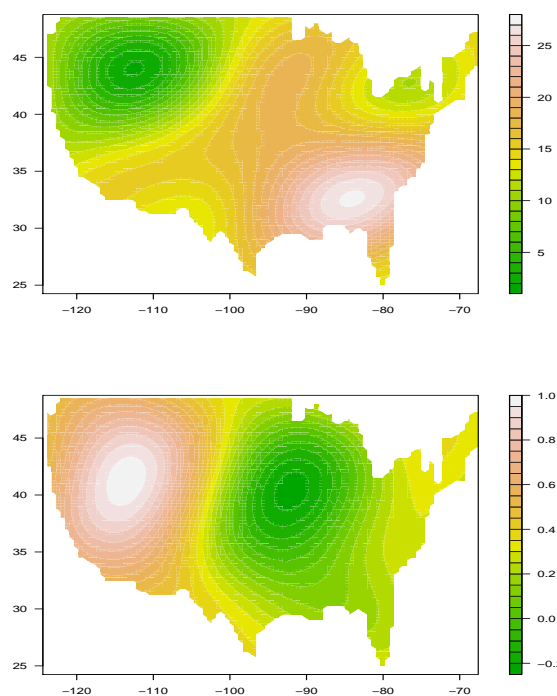
FIGURE 3. The identified figures of the underlying functions in (14) for the magnitudes $B_{k,ij}$ and temporal correlations $\rho_{k,ij}$ in the upper and lower, respectively.

	wavelet	weighted average			
bandwidth		0.1	1	2	∞
kriging	6.61	6.75	7.07	8.27	18.08
forecast 1	20.23	67.91	42.53	36.41	18.22
forecast 2	6.65	14.38	13.73	12.52	9.20

TABLE 2. MSEs of the kriging and forecasting 1 for 100 randomly chosen points in 1997 and those of the forecasting 2 for 125 points that have the temporal correlations identified as being larger than 0.95.

reasonable goodness of fit for the identified wavelet model. The temporal correlation at the location for forecasting has crucial effects for the forecasting performance. Precisely, the forecasting performances of the weighted averages depend on the choice of the bandwidth in relations with the temporal correlations, while the wavelet method adjusts automatically to the temporal correlation to provide a better forecast.

	wavelet	weighted average			
bandwidth		0.1	1	2	∞
kriging	8.15	14.59	16.23	27.27	32.51
	(1.10)	(2.04)	(2.49)	(7.16)	(9.05)
forecast 1	29.58	47.60	36.03	33.94	32.50
	(7.68)	(10.56)	(9.41)	(9.52)	(9.05)
forecast 2	15.58	21.15	19.21	21.83	22.90
	(9.91)	(12.84)	(11.93)	(16.50)	(19.29)

TABLE 3. Averages and standard deviations in parenthesis of MSEs of kriging and forecasting for the data points in the 11th year, where the locations of the points for forecasting and kriging were selected to correspond exactly with those used in the real analysis in Table 2.

4.2. **Simulation studies.** In order to confirm the observations obtained through the applications to US precipitation data, we will conduct simulation studies by the exact model identified in the real analysis.

We simulate 100 sets of spatio-temporal data for 11 years by the estimated model in the real analysis when the locations of the stations are designed to be exactly the same as those of original data. We did not include the mean function in the simulation model for simplicity. By the first 10 years data set, we conduct Whittle likelihood estimation in (19) for the parameters as if they were unknown and compare the estimators with the true values to evaluate the estimation performances. And for the last 11th year spatial data set, we conduct kriging and forecasting based on the estimated model to see their performances in comparisons with the benchmarks by the weighted averages.

We will summarize the results of the estimation, kriging and forecasting for the 100 sets of simulated data to show the performances as simple as possible. Figure 4 shows the figures of median, 5% and 95% points of the estimated underlying functions for B and ρ with the true values as functions of latitudes when longitude is fixed as 36.5. Table 3 shows the averages and standard deviations of the mean squared errors of kriging and forecasting for the 100 points and those of forecasting for the 125 points, where the locations of the points for forecasting and kriging were selected to correspond exactly with those used in the real analysis.

Figure 4 demonstrates the reasonable performances of Whittle estimators. The estimators have almost no bias and the true values are included within the 90% intervals empirically evaluated by the 100 estimators. Table 3 confirms the observations that we stated through the results in Table 2. The averages of MSEs for all the cases in Table 3 are similar to those of Table 2, which are all included in the intervals within one standard deviations. The wavelet method automatically adjusts to the temporal correlation in the point for forecasting to provide reasonable kriging and forecasting, while the weighted averages require careful choice of the bandwidth in relations with the temporal correlation that is usually unknown.

5. DISCUSSION

This paper propose a model for spatio-temporal data by an interpretation of Whittle likelihood function in stationary time series. The striking feature of the
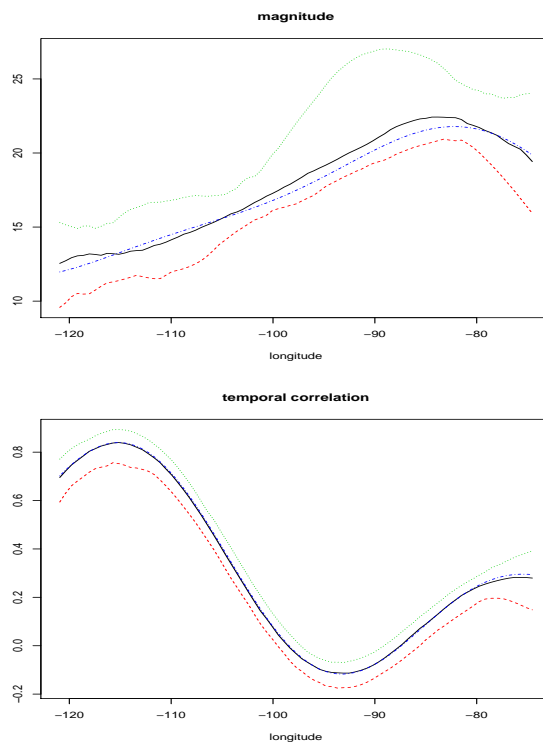
FIGURE 4. The true values and 5%, 95% and 50% points of the
estimators as functions of longitudes with the latitude being 36.5.
The upper and lower figures show those of the underlying functions
for the magnitude and temporal correlation, respectively.

proposed model is that it opens a way for analysis of nonstationary huge spatio-
temporal data set larger than several hundred thousand by the Kalman recursion
for Whittle likelihood estimation, kriging and forecasting.

We state some points that are to be studied in the futures. First one is the
nonstationary extension of the stationary temporal covariances that depend just on
the time lag h in the nonseparable space time covariances in (13). In order to allow
nonstationary also in the temporal dimension, there are two possible ways. One is to
apply random walk model instead of the autoregressive model to the state vectors in
(12). This approach is expected to be effective for nonstationary data in both space
and time when the sampling is conducted regularly in time but irregularly in space
such as yearly collection of land price data at many locations, which are popular in
social science areas. The other one is the use of three dimensional Haar wavelets
to construct an orthogonal basis under irregular sampling over three dimensional
space. Then the model by the modified three dimensional Haar wavelets directly
describe nonstationary covariances not only in space but also in time not via the
state space temporal extension. This approach is expected to be effective for spatio-
temporal data collected irregularly in both space and time, which are frequently
observed in natural science areas.

Second one is the use of continuous wavelets other than the Haar wavelets. Since the Haar wavelets are used as a basis for our modelling, kriged and forecast values model are necessarily piecewise constants. An extension of Haar wavelets to continuous ones that can keep the orthogonality under irregular sampling must be conducted to have continuous estimates. Presently the extension is possible only in mesh data cases.

Third one is a multivariate extension of our wavelet model, which we expect is possible by the method of coregionalization (Banerjee et. al. (2004)). It requires intensive considerations in the space time points where a part of components of observations are missing,

Finally, it is a challenging problem to construct asymptotic theories for the wavelet models in the fixed domain asymptotics (see Stein (1999)), when the length of time is fixed. The problem reduces to the very simple one to estimate the parameters c and δ in independent samples $X_k, k = 1, \ldots, n$ with mean 0 and variance $c\delta^k$ for $0 < \delta < 1$. The difficulty lies in the asymptotic decay of the variance as n tends to infinity.

6. PROOF

In this section, we will prove Proposition 1. Put $N(E_{k,ij}^l) = n_l$ for $l = 1, 2, 3, 4$. Let us prove by contradiction. Assume that $L_{k,ij}$ is singular. Since $\tilde{\phi}_{k,ij}$ is linearly dependent, there exist the non-zero constants k_1, k_2 and k_3 that satisfy

$$k_1 \frac{1}{n_1 + n_2} + k_2 \frac{1}{n_1 + n_3} + k_3 \frac{1}{n_1 + n_4} = 0$$

$$k_1 \frac{1}{n_1 + n_2} - k_2 \frac{1}{n_2 + n_4} - k_3 \frac{1}{n_2 + n_3} = 0$$

$$-k_1 \frac{1}{n_3 + n_4} + k_2 \frac{1}{n_1 + n_3} - k_3 \frac{1}{n_2 + n_3} = 0$$

$$-k_1 \frac{1}{n_3 + n_4} - k_2 \frac{1}{n_2 + n_4} + k_3 \frac{1}{n_1 + n_4} = 0.$$

It follows that the matrix given by

$$B = \begin{bmatrix} \frac{1}{n_1 + n_2} & \frac{1}{n_1 + n_3} & \frac{1}{n_1 + n_4} \\ \frac{1}{n_1 + n_2} & -\frac{1}{n_2 + n_4} & -\frac{1}{n_2 + n_3} \\ -\frac{1}{n_3 + n_4} & \frac{1}{n_1 + n_3} & -\frac{1}{n_2 + n_3} \end{bmatrix}$$

has the rank that must be less than 3. By elementary row operations, the matrix B reduces to the upper triangular matrix given by

$$\begin{bmatrix} \frac{1}{n_1 + n_2} & \frac{1}{n_1 + n_3} & \frac{1}{n_1 + n_4} \\ 0 & -\frac{n_1 + n_2 + n_3 + n_4}{(n_2 + n_4)(n_1 + n_3)} & -\frac{n_1 + n_2 + n_3 + n_4}{(n_1 + n_4)(n_2 + n_3)} \\ 0 & 0 & -\frac{2n_4(n_1 + n_2 + n_3 + n_4)}{(n_1 + n_4)(n_2 + n_3)} \end{bmatrix}.$$

n_4 must be 0 to let the matrix B be singular, which contradicts the assumption that n_1, n_2, n_3, n_4 must be all larger than 1.

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., Carlin, B. and Gelfand A. (2004) *Hierarchical Modeling and Analysis for Spatial Data.* Chapman & Hall/CRC.

[3] 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.

[4] Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*, 2nd edn. Springer-Verlag.

[5] Gneiting, T. (2002) Nonseparable, stationary covariance functions for space-time data. *J. Am. Statist. Assoc.*, 97, 590-600.

[6] 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.

[7] Nason, G., von Sachs, R. and Kroisandt, G. (2000) Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc. B.*, 62, 281-292.

[8] Ogden, R. T. (1997) *Essential Wavelets for Statistical Applications and Data Analysis*. BirkHäuser, Boston.

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

Graduate School of Economics and Management, Tohoku University, 27-1 Kawauchi, Aoba ward, Sendai 980-8576, Japan

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