DSSR

Discussion Paper No. 113

MULTIVARIATE CARMA RANDOM FIELDS

Yasumasa Matsuda
and Xin Yuan

April, 2020

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

MULTIVARIATE CARMA RANDOM FIELDS

YASUMASA MATSUDA AND XIN YUAN

Abstract. This paper conducts a multivariate extension of isotropic Lévy-driven CARMA random fileds on $\mathbb{R}^d$ proposed by Brockwell and Matsuda (2017). Univariate CARMA models are defined as moving averages of a Lévy sheet with CARMA kernels defined by AR and MA polynomials. We define multivariate CARMA models by a multivariate extension of CARMA kernels with matrix valued AR and MA polynomials. For the multivariate CARMA models, we derive the spectral density functions as explicit parametric functions. Given multivariate irregularly spaced data on $\mathbb{R}^2$, we propose Whittle estimation of CARMA parameters to minimize Whittle likelihood given with periodogram matrices and clarify conditions under which consistency and asymptotic normality hold under the so called mixed asymptotics. We finally introduce a method to conduct kriging for irregularly spaced data on $\mathbb{R}^2$ by multivariate CARMA random fields with the estimated parameters in a Bayesian way and demonstrate the empirical properties by tri-variate spatial dataset of simulation and of US precipitation data.

1. Introduction

Continuous-time Autoregressive and Moving Average (CARMA) processes have been applied as a useful tool to analyze continuous time behaviors in physics and engineers for many years. Ornstein and Uhlenbeck process by Uhlenbeck and Ornstein [22] is such a typical example that is regarded as a CARMA $(1, 0)$ process. Doob [10] is one of several papers that examined basic properties and statistical analysis of CARMA processes. Recently CARMA models have been a resurgence of interest by growing needs to analyze high frequency observations in financial time series. CARMA modeling can be a tool to connect partial differential equations with high frequency data in finance. Statistical properties of CARMA time series models, including stationary conditions, parametric forms of covariance and spectral density functions, estimation and prediction, have been investigated by many authors. Brockwell [6] is a good review to see the recent progress in statistical analysis by CARMA processes.

Brockwell and Matsuda [7] extended CARMA models for continuous time series to those for random fields on $\mathbb{R}^d, d \geq 1$, which we call CARMA random fields. The main difficulty for the random field extension from time series is in no trivial order in space like that in time series from past to future. Recalling that CARMA processes are defined by moving averages of Lévy processes specified by CARMA kernels constructed from AR and MA polynomials with negative real roots, Brockwell and Matsuda [7] employed bilateral moving averages by isotropic CARMA kernels on

Key words and phrases. Bayesian kriging, CARMA random fields, Irregularly spaced data, Lévy sheet. periodogram, Spectral density function, Whittle likelihood.

$\mathbb{R}^d$, given by

$$g(s) = g_0(||s||), s \in \mathbb{R}^d,$$

where g_0 is a time series CARMA kernel and $||\cdot||$ is Euclidean norm. to avoid the difficulty. Several choices of AR and MA polynomials in the CARMA kernels provide a flexible class of isotropic random fields.

This paper tries a multivariate extension of the univariate CARMA random fields on $\mathbb{R}^d$. The extension is conducted in a straightforward way of extending AR and MA polynomials in the univariate models to AR and MA polynomials with matrix coefficients, which is similar to the extension from discrete ARMA to vector ARMA models. The extension here is restricted to the cases when AR polynomial matrices are diagonal in order to obtain explicit parametric expressions for CARMA kernel matrices. The spectral density matrices are expressed in explicit parametric forms as a result, which works for parameter estimation on the frequency domain by the Whittle likelihood, while auto-covariance matrices are not explicitly obtained except for $d = 1$, and 3 but are given by Hankel transform of the spectral densities. It should be notified here that multivariate CARMA random fields target multivariate irregularly spaced data where observation points for each component are not supposed to be identical with those for the other components. We develop a series of procedures to conduct statistical analysis for irregularly spaced data by multivariate CARMA random fields, following the standard way of time series analysis in Box, Jenkins and Reinsel [4]. Namely, estimation of model parameters and kriging, which is prediction in spatial data, by the estimated model as if the parameters are known, are developed. We employ frequency domain approach for the parameter estimation by the Whittle likelihood, while we choose spatial domain approach for the kriging.

The Whittle estimation is a classical technique in time series analysis which has been employed by many time series researchers such as Dunsmuir [11], Robinson [21], Hosoya [15] and Dahlhaus [9]. Brockwell and Davis (Chapter 10, 1991) provides an excellent introduction to the Whittle estimation. This paper employs the periodogram extended for irregularly spaced data for the Whittle estimation, which was originally proposed by Matsuda and Yajima [18] and Bandyopadhyay and Lahiri [1], and has been applied to irregularly spaced data analysis by Bandyopadhyay and Subba Rao [3], Bandyopadhyay, Lahiri, and Nordman [2], Matsuda and Yajima [19] and Subba Rao [20]. We define the Whittle likelihood function for multivariate spatial data in a modified form from Matsuda and Yajima [18] to let it be free from the extra nuisance estimation for distributions of sampling points. We have established asymptotic normality of the Whittle estimator for multivariate CARMA random fields which can be non-Gaussian with finite moments of all orders under the so called mixed asymptotics, which is the asymptotic scheme where sample size and sampling region jointly diverge. The asymptotic results are regarded as a non-trivial extension of the classical result for discrete stationary time series by Dunsmuir and Hannan [12] and Dunsmuir [11] to those for continuous random fields. The nontrivial difference between them is in the form of asymptotic variance matrix for random field cases which is not separated as those of CARMA kernel and noise variance unlike time series cases (Remark 4, Dunsmuir [11]). The difference comes from a feature of continuous processes for which Kolmogorov formula (see i.e. sec. 5.8 in Brockwell and Davis [5]) does not hold any more.

Kriging, which is usually referred to as a minimum mean squared error method of spatial prediction that depends on the second order properties of spatial processes (Cressie[8]), is one of main purposes in spatial data analysis. Kriging for multivariate spatial data, which is often called as cokriging, is a challenging topic and lots of methods have been proposed in the literatures. Gelfand and Banerjee [13] is a good review for kriging with multivariate spatial process models. Multivariate CARMA random fields can be regarded as a multivariate extension of the kernel convolution approach by Higdon [14] in spatial statistics literatures. Our approach re-expresses multivariate spatial observations following CARMA random fields as a form of spatial regression model. Assuming Gaussian for Lévy noise terms driving CARMA, we follow Bayesian approach to conduct kriging by regarding it as a Bayesian hierarchical model. With the multivariate model involving a large number of locations, the Bayesian regression requires too heavy computational burden to work for kriging in practice. We apply the technique of Zhang, Sang and Huang [24] to the spatial regression context, which can bee seen also as a kind of covariance tapering by Kaufman, Schervish and Nychka [16]. Specifically, we divide a whole region of spatial observations into several sub-regions to partition the spatial regression into several sub-models, which lets the posterior computation feasible for large multivariate spatial dataset.

Let us start from the definition of multivariate CARMA random fields.

2. Multivariate extension of CARMA random fields

We define multivariate CARMA random fields in a formal way from univariate ones without introductory arguments. For motivating implications on CARMA models, see Brockwell [6] or Brockwell and Matsuda [7]. We start from the definition of multivariate L'evy sheets necessary for the multivariate extension.

2.1. **Multivariate Lévy sheet.**

Define an m-variate Lévy sheet $L(x)$ for $m \geq 1$, which is necessary to multivariate extension of CARMA random fields, by

$$L(x) = \dot{L}((0, x]), x = (x_1, \dots, x_d)' \in \mathbb{R}^d_+,$$

for a random measure $\dot{L}$, which satisfies

(a) if A and B are disjoint Borel sets on $\mathbb{R}^d$, $\dot{L}(A)$ and $\dot{L}(B)$ are independent, and
(b) for every Borel set A on $\mathbb{R}^d$ with finite Lebesgue measure $|A|$,

$$E[\exp\{i\theta'\dot{L}(A)\}] = \exp\{|A|\psi(\theta)\},\ \theta \in \mathbb{R}^m,$$

where ψ is the logarithm of the characteristic function of an infinity divisible distribution.

We follow the tradition in writing the integral of a deterministic function g on $\mathbb{R}^d$ with respect to $\dot{L}$ as $\int_{\mathbb{R}^d} g(x)L(dx)$.

Let us introduce typical examples.

(a) If $\psi(\theta) = -\theta' C\theta/2$ with a positive definite $m \times m$ matrix C, L is a m-variate Brownian sheet.
(b) If, for a Borel set A on $\mathbb{R}^d$,

$$\dot{L}(A) = \sum_{i=1}^{\infty} Y_i 1_{x_i}(A),$$

where x_i denotes the location of the ith unit point mass of a Poisson random measure on $\mathbb{R}^d$ with intensity λ and $\{Y_i\}$ is a sequence of IID random vectors with distribution function F and independent of $\{x_i\}$, L is a m-variate compound Poisson sheet.

We shall restrict attention in this paper to second order Lévy sheets, i.e. those for which $E[L_i(t)^2] < \infty, i = 1, \ldots, m$ at $t = (1, 1, \ldots, 1)$, and then the first and second order moments for the sheets are determined by

$$E\{\dot{L}(A)\} = \mu|A| \text{ and } var\{\dot{L}(A)\} = \Sigma|A|, \tag{1}$$

for a $m \times 1$ vector μ and $m \times m$ positive definite matrix Σ.

If h is a $m \times m$ matrix valued function on R^d of the form $h(x) = \sum_{i=1}^k C_i 1_{A_i}$, where $A_i, i = 1, \ldots, k$ are disjoint Borel subsets on $\mathbb{R}^d$ with finite Lebesgue measure with $m \times m$ matrix $C_i, i = 1, \ldots, k$,

$$\int_{\mathbb{R}^d} h(x) dL(x) := \sum_{i=1}^k C_i \dot{L}(A_i).$$

This definition can be extended, by a standard construction, to include all matrix-valued functions h whose components are in $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$.

2.2. Multivariate CARMA random fields.

Let us recall the definition of univariate CARMA random fields on $\mathbb{R}^d$ by Brockwell and Matsuda [7]. Let $L(x)$ be an univariate Lévy sheet on $\mathbb{R}^d$.

Definition 1. *Let $a_*(z) = z^p + a_1 z^{p-1} + \cdots a_p = \prod_{i=1}^p (z - \lambda_i)$ be a polynomial of degree p with real coefficients and distinct zeros $\lambda_1, \ldots, \lambda_p$ having strictly negative real parts and let $b_*(z) = b_0 + b_1 z + \cdots b_q z^q = \prod_{i=1}^q (z - \xi_i)$ with real coefficient b_j and $0 \le q < p$. Suppose also that $\lambda_i \neq \mu_j$ for all i and j. Then defining*

$$a(z) = \prod_{i=1}^p (z^2 - \lambda_i^2) \text{ and } b(z) = \prod_{i=1}^q (z^2 - \xi_i^2),$$

the univariate CARMA(p, q) random field driven by a Lévy sheet L is

$$S_d(x) = \int_{\mathbb{R}^d} g(||x - u||) dL(u), \ x \in \mathbb{R}^d,$$

where $||x - u||$ denotes the Euclidean norm of the vector $x - u$ and $g(s)$ is the CARMA kernel defined by

$$g(s) = \sum_{j=1}^p \frac{b(\lambda_j)}{a'(\lambda_j)} e^{\lambda_j s}, \ s \in \mathbb{R},$$

where a' denotes the derivative of the polynomial a.

We shall extend univariate models to m-variate CARMA random fields by extending the scaler functions $a(z)$, $b(z)$ in Definition 1 to the matrix ones with an m-variate Lévy sheet on $\mathbb{R}^d$ described by (1). Here we restrict the cases when $a(z)$ is extended trivially to matrix one by $a(z)I_m$ in order to obtain explicit expressions for CARMA kernel matrix.

Definition 2. *Let* $a(z) = \prod_{j=1}^{p}(z^2 - \lambda_j^2)$ *with* $Re(\lambda_j) < 0$ *be the polynomial in Definition 1 and define the matrix polynomial with real* $m \times m$ *matrices* $B_1, \ldots, B_q$ *by*

$$B(z) = z^{2q} I_m + B_1 z^{2q-2} + \cdots + B_{q-1} z^2 + B_q,$$

where $B(\lambda_i) \neq 0$ *for* $i = 1, \ldots, p$. *The* m*-variate CARMA*(p, q) *random field driven by a* m*-variate Lévy sheet* L *is*

$$S_d(x) = \int_{\mathbb{R}^d} G(||x - u||) dL(u), \;\; x \in \mathbb{R}^d,$$

where $G(x)$ *is the* $m \times m$ *CARMA kernel matrix defined by*

$$G(s) = \sum_{j=1}^{p} \frac{1}{a'(\lambda_j)} B(\lambda_j) e^{\lambda_j s}, \;\; s \in \mathbb{R}.$$

Here we introduce two typical kernels for multivariate CARMA(1,0), CARMA(2,1).

Example. *Defining* $a(z)$ *and* $B(z)$ *in Definition 2 as* $(z^2 - \lambda^2)$ *and* I_m, *respectively, we obtain CAR(1) kernel matrix given by*

$$G(s) = \frac{1}{2\lambda} I_m e^{\lambda s}, Re(\lambda) < 0.$$

Defining $a(z)$ *and* $B(z)$ *in Definition 2 as* $(z^2 - \lambda_1^2)(z^2 - \lambda_2^2)$ *and* $I_m z^2 + B_1$, *respectively, we have CARMA(2,1) kernel matrix given by*

$$G(s) = \frac{\lambda_1^2 I_m + B_1}{2\lambda_1(\lambda_1^2 - \lambda_2^2)} e^{\lambda_1 s} + \frac{\lambda_2^2 I_m + B_1}{2\lambda_2(\lambda_2^2 - \lambda_1^2)} e^{\lambda_2 s}, Re(\lambda_1) < Re(\lambda_2) < 0. \quad (2)$$

2.3. **Second order properties.** We show the second order properties of multivariate CARMA(p, q) random fields on $\mathbb{R}^d$, when the driving Lévy sheet has finite variance. We show in Theorem 1 the spectral density matrix and autocovariance matrix without proof, since it is obtained in a straight forward way from the proof in Theorem 2 in Brockwell and Matsuda [7]. Let us notice that the spectral density matrix is explicitly expressed as a function of a, B and Σ, while the auto-covariance matrix is not obtained explicitly except for $d = 1$ and 3, but given by the Hankel transform of the spectral density matrix. The explicit form of the spectral density matrix leads us to propose the Whittle likelihood estimation to estimate the CARMA parameters in the next section.

Theorem 1. *If* L *is a* m*-variate Lévy sheet with parameters* μ *and* Σ *in (1) and if the* m*-variate CARMA random field is defined as in definition 2, then the mean vector is*

$$ES_d(x) = \sum_{i=1}^{p} \frac{1}{a'(\lambda_i)} \frac{\pi^{d/2}\Gamma(d+1)}{|\lambda|^d \Gamma(d/2+1)} B(\lambda_i)\mu,$$

and the spectral density matrix is

$$f_d(\omega) = \tilde{G}_d(\omega) \Sigma \tilde{G}'_d(\omega), \;\; \omega \in \mathbb{R}^d, \quad (3)$$

where

$$\tilde{G}_d(\omega) = c_d \sum_{i=1}^{p} \frac{2\lambda_i}{a'(\lambda_i)\,(||\omega||^2 + \lambda_i^2)^{\frac{d+1}{2}}} B(\lambda_i),$$

with

$$c_d = \begin{cases} -2^{d/2-1}\Gamma\left(\frac{d+1}{2}\right)/\sqrt{\pi}, & \text{if } d \text{ is odd,} \\ -2^{-d/2}\Gamma(d)/\Gamma\left(\frac{d}{2}\right), & \text{if } d \text{ is even.} \end{cases}$$

The autocovariance matrix is

$$\Gamma_d(h) = (2\pi)^{d/2} H_{d/2-1} F_d(||h||), \ h \in \mathbb{R}^d,$$

where $F_d(||\omega||) := f_d(\omega)$ *and* H_r *denotes the modified Hankel transform of order* r *defined by, for* J_r*, the Bessel function of the first kind of order* r*,*

$$H_r F_d(x) = \int_0^\infty F_d(y) \frac{J_r(xy)}{(xy)^r} y^{2r+1} dy, x > 0, r \geq -\frac{1}{2}.$$

The autocovariance matrix, which is given by the Hankel transform of the spectral density matrix, is explicitly evaluated for $d = 1$ and $d = 3$ as

$$\Gamma_1(h) = \sum_{i=1}^p \mathbf{Res}_{z=\lambda_i} \left\{ \frac{\exp(z||h||)}{a(z)^2} B(z)\Sigma B^t(z) \right\}, \ h \in \mathbb{R}$$

and

$$\Gamma_3(h) = -\frac{2\pi}{||h||} \sum_{i=1}^p \mathbf{Res}_{z=\lambda_i} \left\{ \frac{\exp(z||h||)}{za(z)^4} D(z)\Sigma D^t(z) \right\}, \ h \in \mathbb{R}^3,$$

respectively, where

$$D(z) = a'(z)B(z) - a(z)B'(z).$$

3. Parameter estimation

3.1. **Setup.** This section focuses on parameter estimation for multivariate CARMA random fields on $\mathbb{R}^2$. The results here on $\mathbb{R}^2$, however, can be extended to those on $\mathbb{R}^3$ easily. The point to prevent higher dimensional extension is the rate of convergence in Lemma 1, from which the approximation in (11) is not validated for $d \geq 4$.

Let $X(s) = (X_1(s), \ldots, X_m(s))', s \in \mathbb{R}^2$ be a m-variate random field on $\mathbb{R}^2$, given by

$$X(s) = \int_{\mathbb{R}^2} G(\psi; s-u) dL(u), \tag{4}$$

where G is a $m \times m$ CARMA(p, q) kernel matrix defined in Definition 2 with the parameter $\psi = (\lambda_1, \ldots, \lambda_p, B_1, \ldots, B_q)$, and L is a m-variate Lévy sheet on $\mathbb{R}^2$ that satisfies (1) with $\mu = 0$. The spectral density function is given in Theorem 1 by

$$f_2(\theta; \omega) = \tilde{G}_2(\psi; \omega)\Sigma\tilde{G}_2'(\psi; \omega), \omega \in \mathbb{R}^2, \tag{5}$$

for $\theta = (\psi, \Sigma) \in \mathbb{R}^{pdim}$ with

$$\tilde{G}_2(\psi; \omega) = -\frac{1}{2} \sum_{i=1}^p \frac{2\lambda_i}{a'(\lambda_i)\left(||\omega||^2 + \lambda_i^2\right)^{\frac{3}{2}}} B(\lambda_i).$$

Let us propose the Whittle estimation for the parameter θ, which is extended from the one for classical time series analysis, when observation points for each component of $X(s) = (X_1(s), \ldots, X_m(s))'$ are not supposed to be identical. with those for the other components.

3.2. **Whittle Estimation.** Suppose we have observed irregularly spaced data that follow multivariate CARMA random fields (4). Here we allow locations and sample size of the observations for each component not necessarily to be identical. Namely, we suppose that pth component has observations $X_p(s_{pj}), j = 1, \ldots, n_p$ on locations s_{pj} that can depend on p, which we assume distributes irregularly over a region in a rectangular $A = [0, A_1] \times [0, A_2]$ on $\mathbb{R}^2$. Let $|A|$ be the area of A, ie., $|A| = A_1 \times A_2$.
.

First let us define the periodogram matrix whose (p, q)th element by

$$I_{pq}(\omega) = d_p(\omega)\overline{d_q(\omega)}, \omega \in \mathbb{R}^2,$$

where

$$d_p(\omega) = \frac{\sqrt{|A|}}{n_p} \sum_{j=1}^{n_p} X_p(s_{pj}) e^{-i\omega' s_{pj}}.$$

For a grid point $j = (j_1, j_2)$ on $\mathbb{Z}^2$, define a frequency ω_j by

$$\omega_j = \left(\frac{2\pi j_1}{A_1}, \frac{2\pi j_2}{A_2} \right)'.$$

For a symmetric compact set D on $\mathbb{R}^2$ such that $-s \in D$ for $s \in D$, define

$$J_D = \left\{ j = (j_1, j_2) \in \mathbb{Z}^2 | \omega_j \in D \right\},$$

and $|J_D|$ by the number of elements in J_D. The Whittle estimator $\hat{\theta}$ is defined by the one that minimizes the Whittle likelihood:

$$l_w(\theta) = \log \left[\frac{1}{m|J_D|} \sum_{j \in J_D} tr \left\{ \left(f(\theta; \omega_j) + \eta \hat{K} \right)^{-1} I(\omega_j) \right\} \right] + \frac{1}{m|J_D|} \sum_{j \in J_D} \log \left\| f(\theta; \omega_j) + \eta \hat{K} \right\|, \tag{6}$$

where η is the scaler nuisance parameter that is to be estimated jointly with θ, and $\hat{K}$ is the matrix to compensate for the bias of the periodogram, which is defined by, for the set of observation points $S_p = \{s_{pj}, j = 1, \ldots, n_p\}$, $p = 1, \ldots, m$,

$$\hat{K}_{pq} = \frac{|A|}{n_p n_q} \sum_{s_j \in S_p \cap S_q} X_p(s_j) X_q(s_j), p, q = 1, \ldots, m,$$

and defined by 0 if $S_p \cap S_q$ is null.

Notify the following two points for the proposed likelihood function. First, it is modified to be scale invariant in the sense that the spectral density matrix multiplied with any constant provides exactly same parameter estimate. In other words, the variance matrix Σ requires a restriction such as $\Sigma_{11} = 1$ for the identifiability. Second, the reason why we make the likelihood be scale free is because no-additional estimation is necessary to let the Whittle estimation be consistent. In other words, it would be necessary to estimate the quantity related with a density function of sampling points, if we define the Whittle likelihood with a scale parameter τ by

$$\sum_{j \in J_D} \left[tr \left\{ \left(\tau f(\theta; \omega_j) + \tilde{\eta} \hat{K} \right)^{-1} I(\omega_j) \right\} + \log \left\| \tau f(\theta; \omega_j) + \tilde{\eta} \hat{K} \right\| \right].$$

See Remark 1 in Matsuda and Yajima [18]. The scale-free likelihood in (6), which is given by concentrating out the scale parameter, makes it possible to estimate the parameter θ without any additional estimation at the sacrifice of the scale parameter in Σ. For a practical discussion, see the beginning of Section 5.

3.3. **Asymptotic Results.** This section will show the asymptotic results of the Whittle estimator that minimizes (6). First we clarify the scheme under which the asymptotic results shall be derived, which is not trivial unlike time series cases. We state it as assumption given as C1 below. Under the scheme in C1, we consider the asymptotic results for the Whittle estimator.

C1. The sample size n_p and the sampling region $A = [0, A_1] \times [0, A_2]$ diverge jointly such that $A_1 \to \infty, A_2 \to \infty, A_1/A_2 = O(1)$ and $|A|/n_p \to 0, p = 1, \ldots, m$ for the area $|A| = A_1 \times A_2$. We shall employ a suffix k such as $n_p = n_{pk}, A = A_k$ to indicate explicitly that they diverge as k tends to infinity.

C2. Let $S_p, p = 1, \ldots, m$, be the set of sampling points of X_p in $A = [0, A_1] \times [0, A_2]$. We assume that elements in S_p are written as, for $p = 1, \ldots, m$,

$$s_{pj} = (A_1 u_{1,pj}, A_2 u_{2,pj}), j = 1, \ldots, n_p,$$

where $u_{pj} = (u_{1,pj}, u_{2,pj})$ is a sequence of independent and identically distributed random vectors with a probability density function $g(x)$ supported on $[0, 1]^2$ which has continuous first derivatives.

C3. $X(s), s \in \mathbb{R}^2$ follows a random field in (4) driven by a zero-mean Lévy sheet on $\mathbb{R}^2$ with finite moments of all orders. Every component of the CARMA kernel is bounded and integrable and the spectral density matrix has continuous second derivatives.

C4. Let Θ be a compact subset in $\mathbb{R}^{pdim}$ and D be a symmetric compact region on $\mathbb{R}^2$. The parametric spectral density matrix $f(\theta; \omega)$ defined in (5) is positive definite and has continuous second derivatives with respect to θ on $\Theta \times D$. $\theta_1 \neq \theta_2$ implies that $f(\theta_1; \omega) \neq f(\theta_2; \omega)$ on a subset of D with positive Lebesgue measure.

Let us introduce the asymptotic results for the Whittle estimator as k tends to be ∞ under the asymptotic scheme of C1.

Theorem 2. *Under Assumptions C1-C4, the Whittle estimator $\hat{\theta}_k$ minimizing (6) converges to θ_0 in probability as k tends to be infinity.*

Theorem 3. *If $|A_k|^{3/2}/n_p \to 0$ for $p = 1, \ldots, m$. and $g(x), x \in [0, 1]^2$ has continuous second derivatives, in addition with Assumptions C1-C4, $\sqrt{|A_k|}\left(\hat{\theta}_k - \theta_0\right)$ converges in distribution to*

$$N\left(0, b_g(\Omega_1 - \Omega_2)^{-1}(2\Omega_1 + \Pi)(\Omega_1 - \Omega_2)^{-1}\right),$$

as k tends to be infinity, where, for $p, q = 1, \ldots, pdim$,

$$b_g = (2\pi)^2 \left\{ \int_{[0,1]^2} |g(u)|^4 du \right\} \left\{ \int_{[0,1]^2} |g(u)|^2 du \right\}^{-2},$$

$$\Omega_{1,pq} = \int_D tr \left(\frac{\partial f(\theta_0;\omega)}{\partial \theta_p} f^{-1}(\theta_0;\omega) \frac{\partial f(\theta_0;\omega)}{\partial \theta_q} f^{-1}(\theta_0;\omega) \right) d\omega,$$

$$\Omega_{2,pq} = \frac{1}{m|D|} \left(\int_D \frac{\partial \log ||f(\theta_0;\omega)||}{\partial \theta_p} d\omega \right) \left(\int_D \frac{\partial \log ||f(\theta_0;\omega)||}{\partial \theta_q} d\omega \right),$$

$$\Pi_{pq} = \sum_{a,b,c,d=1}^{m} \kappa_{abcd} \times$$

$$\left[\int_D \tilde{G}_2'(\theta_0;\omega) \frac{\partial f^{-1}(\theta_0,\omega)}{\partial \theta_p} \tilde{G}_2(\theta_0;\omega) d\omega \right]_{ab} \left[\int_D \tilde{G}_2'(\theta_0;\omega) \frac{\partial f^{-1}(\theta_0,\omega)}{\partial \theta_q} \tilde{G}_2(\theta_0;\omega) d\omega \right]_{cd},$$

where κ_{abcd} *is the fourth order cumulant of the Lévy sheet given by*

$$cum(L_a(du), L_b(du), L_c(du), L_d(du)) = \kappa_{abcd} du, a, b, c, d = 1, \ldots, m.$$

Remark 1. *There are several interesting differences in the asymptotic variance matrix from the classical one in discrete time series. First,* Ω_2 *vanishes with respect to* ψ *in* $\theta = (\psi, \Sigma)$ *in discrete time series cases, because it holds for a spectral density matrix for discrete time series that*

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log ||f(\omega)|| d\omega = \log \left\| \frac{1}{2\pi} \Sigma \right\|$$

by Kolmogorov formulra (Theorem 5.8.1, Brockwell and Davis, [5]*), while it does not hold any more in continuous random fields. Second,* Π *vanishes in discrete time series except for the components with respect to* Σ *in* $\theta = (\psi, \Sigma)$ *(Remark 3, Dunsmuir,* [11]*), while it does not in continuous random fields. In other words, the asymptotic variance matrix with respect to* ψ *and* Σ *are not separated for continuous processes.*

The twice differentiable assumption for $g(x)$ in Theorem 3 as well as the differentiability in Theorem 2 are strict assumptions around edges of sampling points in $A = [0, A_1] \times [0, A_2]$, which is rare to be satisfied. To avoid the difficulty, let us propose the tapered periodogram. For a taper $h(x)$, a continuous positive function on $[0, 1]^2$, the tapered periodogram is defined by

$$I_{h,pq}(\omega) = \tilde{d}_p(\omega) \overline{\tilde{d}_q(\omega)},$$

for

$$\tilde{d}_p(\omega) = \frac{\sqrt{|A|}}{n_p} \sum_{j=1}^{n_p} X_p(s_{pj}) h(s_{pj,1}/A_1, s_{pj,2}/A_2) e^{-i\omega' s_{pj}}, p = 1, \ldots, m.$$

The Whittle likelihood in (6) replaced with the tapered periodogram provides Theorems 1 and 2 under relaxed conditions on $g(x)$, namely, the first and second differentiability for $g(x)h(x)$, not for $g(x)$.

Let $\tilde{\theta}$ be the estimator minimizing the modified Whittle likelihood with a taper $h(x), x \in [0, 1]^2$.

Theorem 4. *Under C1-C4, where $w(x) = g(x)h(x)$ has continuous first derivatives instead of $g(x)$ in C2, the Whittle estimator $\tilde{\theta}$ constructed with a taper $h(x)$ is consistent. If we assume more that $w(x) = g(x)h(x)$ has continuous second derivatives and that $|A|^{3/2}/n_p \to 0$ for $p = 1, \ldots, m$, the asymptotic normality in Theorem 3 still holds, where b_g in the asymptotic covariance is replaced with*

$$b_h = \left\{ \int_{[0,1]^2} |g(x)h(x)|^4 dx \right\} \left\{ \int_{[0,1]^2} |g(x)h(x)|^2 dx \right\}^{-2}.$$

4. CARMA Kriging

This section shall propose a way to conduct kriging by multivariate CARMA random fields when the parameters are known, although they are estimated in practice by the Whittle estimation we stated in the last section. The kriging is conducted by following the way of time series forecasting by moving average models, where the forecast is constructed with the error process recovered by the moving average recursions. We regard the recovered error term in a Bayesian way as the one obtained from the posterior distribution when iid assumption for the error terms is seen as the prior. We extend the time series forecasting procedure to kriging for multivariate spatial data, where it should be noticed that Gaussian assumption that is not necessary for the Whittle estimation is imposed in this section.

4.1. **Setup as a spatial regression.** We assume that the Lévy sheet driving a m-variate CARMA random field is a compound Poisson sheet, which is as a result given by

$$Y(s) = \sum_{j=1}^{\infty} G(s - u_j) Z_j, \; s \in \mathbb{R}^2,$$

where $\{u_j\}, j = 1, \ldots,$ which are called as knots, are randomly distributed over $\mathbb{R}^2$ and Z_j follows iid with mean 0 and variance matrix Σ. The restriction to compound Poisson sheet does not lose generality in terms of covariances. In other words, the class of covariances by CARMA models driven by Lévy sheets and of those by compound Poisson sheets are identical. We suppose the following special but practical situations under which we shall introduce a method for kriging.

a. We truncate the range of the knots within a compact region C with the number of knots M, which follows a Poisson distribution. In addition, inserting a constant term and an iid measurement error, we employ the following empirically modified CARMA model for kriging by

$$Y(s_i) = \mu + \sum_{\{u_j, j=1,\ldots,M\} \subset C} G(s_i - u_j) Z_j + \varepsilon_i, \; s_i \in \mathbb{R}^2, \tag{7}$$

where we denote the set of knot points as $K_M = \{u_j, j = 1, \ldots, M\} \subset C$.

b. The parameter Σ is a diagonal matrix without loss of generality, because, for the Cholesky decomposition of Σ given by $L\text{diag}(\sigma_1^2, \ldots, \sigma_m^2)L'$ with the lower triangular matrix L with $L_{ii} = 1$, $i = 1, \ldots, m$, and the re-defined G by GL, we obtain the CARMA model driven by the compound Poisson sheet with the diagonal variance matrix $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_m^2)$.

c. The CARMA kernel G is assumed to be known, although it is in practice estimated by the Whittle estimation.

d. The diagonal variance matrices $\Sigma = diag(\sigma_1^2, \ldots, \sigma_m^2)$ and $\Delta = diag(\delta_1^2, \ldots, \delta_m^2)$ for Z_j and ε_i, respectively, are both unknown parameters to be estimated in the kriging procedure.

e. We observe samples of $Y_p(s)$ at $S_p = \{s_{p_j}, j = 1, \ldots, n_p\}$ for $p = 1, \ldots, m$, and aim to krig unknown values of $Y_p(s)$ at $\tilde{S}_p = \{v_{p_j}, j = 1, \ldots, \tilde{n}_p\}$ for $p = 1, \ldots, m$. Denote $V_p = S_p \cup \tilde{S}_p$ with $N_p = n_p + \tilde{n}_p$.

Under the setup, let us express (7) in a componentwise way as

$$Y_p(s_i) = \mu_p + \sum_{q=1}^{m} \sum_{u_j \in K_M} G_{pq}(s_i - u_j) Z_{jq} + \varepsilon_{ip}, \ \ s_i \in V_p,$$

which we stack together for $p = 1, \ldots, m$, rewriting in a vector form as

$$Y = G_u Z + E, \tag{8}$$

where

$$Y = (Y_1', \ldots, Y_m')' \text{ for } Y_p = Y_p(s_i), s_i \in V_p,$$

$$G_u = \begin{pmatrix} 1_{N_1} & 0_{N_1} & \cdots & 0_{N_1} & G_{u,11} & G_{u,12} & \cdots & G_{u,1m} \\ 0_{N_2} & 1_{N_2} & \cdots & 0_{N_2} & G_{u,21} & G_{u,22} & \cdots & G_{u,2m} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0_{N_m} & 0_{N_m} & \cdots & 1_{N_m} & G_{u,m1} & G_{u,m2} & \cdots & G_{u,mm} \end{pmatrix}$$

$$\text{for } G_{u,pq} = G_{pq}(s_i - u_j), s_i \in V_p, u_j \in K_M,$$

$$Z = (\mu', Z_1', \ldots, Z_m')' \text{ for } \mu = (\mu_1, \ldots \mu_m)' \text{ and } Z_q = (Z_{1q}, \ldots, Z_{Mq})',$$

$$E = (\varepsilon_1', \ldots, \varepsilon_m')' \text{ for } \varepsilon_p = (\varepsilon_{1p}, \ldots, \varepsilon_{N_p,p})'.$$

4.2. **Gibbs sampling.** We shall recover Z in (8) in order to construct the krigged value. We evaluate the posterior of Z by

$$p(Z, \Sigma, \Delta | Y, u) \propto p(Y|Z, u, \Delta) p(Z|\Sigma) p(\Sigma) p(\Delta), \tag{9}$$

where we assume Gaussian distributions for the first and second terms in the right hand side with the precision matrices given by $diag(\delta_1^{-2} 1_{N_1}', \cdots, \delta_m^{-2} 1_{N_m}')$, and $diag(0_m', \sigma_1^{-2} 1_M', \cdots, \sigma_m^{-2} 1_M')$ respectively. Assuming the independent non-informative inverse Gamma distributions as the priors of σ_i^2 and δ_i^2 for $i = 1, \ldots, m$, we obtain the posterior samples for Z by the Gibbs sampler. Iteratively sampling from the posteriors in (9) given knots u simulated independently at each iteration, we aim to obtain the marginal posterior samples given Y.

0. Initialize Y at $\tilde{S}_p$ and $\sigma_p^2, \delta_p^2, p = 1, \ldots, m$.
1. Simulate knots $u = (u_1, \ldots, u_M)$ uniformly distributed over C for M that follows a Poisson distribution.
2. Update Z for given Y, u, Σ and Δ by the posterior $N(\Omega a, \Omega)$ with
$$a = G_u' D_E^{-1} Y,$$
$$\Omega = \left(G_u' D_E^{-1} G_u + D_Z^{-1}\right)^{-1}.$$
3. Simulate $\tilde{Y}$ for given u, Z by the posterior $N(G_u Z, D_E)$ and update Y with $\tilde{Y}$ only at $\tilde{S}_p, p = 1, \ldots, m$.
4. Update σ_p^{-2} and δ_p^{-2} for given $u, Z, E = Y - G_u Z$ by the posteriors $Ga(M/2, \sum_{j=1}^{M} Z_{jp}^2/2)$ and $Ga(N_p/2, \sum_{j=1}^{N_p} E_{jp}^2/2)$, respectively, for $p = 1, \ldots, m$.

5. return to 1.

As the result of the iterations, we obtain the posterior samples in Step 3 for Y at $\tilde{S}_p, p = 1, \ldots, m$, by which we can construct interval as well as point estimation as kriging.

Remark 2. *We just simulate the prior samples for u at Step 1 in the Gibbs sampler unlike Zhang, Sang and Huang* [24]*, since knots are not parameters to generate CARMA processes but are randomly distributed ones independent of observations. It means that we construct the krigged values by the marginal posteriors at Step 3 when knots are distributed uniformly over C in (7).*

Step 2 requires the inverse of $m(M+1) \times m(M+1)$ dense matrix, which is infeasible when M is large. In the empirical example shown later, we will consider US precipitation data in which CARMA models with $M = 6,000$ and $m = 3$ are fitted, where the step 2 requires $18,000 \times 18,000$ matrix inversion. In order to avoid the difficulty of huge dimensional matrix inversion, we propose a sub-chain to approximate the inversion in the step 2 with a lower dimensional one.

2-a. We divide the region C randomly into several sub-regions as $C1, \ldots, C_k$. Following the division, we partition Y, u, Z, G_u, D_Z and D_E into the ones corresponding with the devision, denoted as, for $i = 1, \ldots, k$, $Y^{(i)}$, $u^{(i)}$, $Z^{(i)}$, $G_u^{(i)}$, and $D_{Z(i)}$ and $D_{E(i)}$. We include the constant term $\mu^{(i)}$ for all the divisions $C_i, i = 1, \ldots, k$. In other words, we allow μ to be dependent on the sub-regions.

2-b. Initialize Z.

2-c. For $i = 1, \ldots, k$, update $Z^{(i)}$ with $N(\Omega^{(i)} a^{(i)}, \Omega^{(i)})$ for

$$a^{(i)} = G_u'^{(i)} D_{E(i)}^{-1} \left\{ Y^{(i)} - G_u^{(-i)} Z^{(-i)} \right\},$$

$$\Omega^{(i)} = \left(G_u'^{(i)} D_{E(i)}^{-1} G_u^{(i)} + D_{Z(i)}^{-1} \right)^{-1},$$

where $G_u^{(-i)}$ and $Z^{(-i)}$ are the sub-components of G_u and Z excluding the ones corresponding with C_i.

Iterating 2-c in the sub-chain for step 2, we obtain the posterior samples of Z by $(m+1)M_i \times (m+1)M_i$ matrix inversion for M_i given roughly by M/k.

5. Empirical studies

5.1. **tri-variate CARMA(2,1) kernel matrix.**

This section focuses on tri-variate CARMA (2,1) random fields on $\mathbb{R}^2$ and demonstrates the empirical properties in terms of estimation and kriging for simulated and real data. Although it is possible to apply general class of CARMA(p, q) models, CARMA (2,1) shall be employed with the two reasons. First, CARMA (2,1) provides wide enough class of covariance structures to express flexible behaviors both at short and long lags by the CARMA kernel in (10). The first and second terms in (10) control the behaviors at short and long lags, respectively. Second, model selection issues from general CARMA are in a challenging topic currently out of scope in this paper.

We shall employ the CARMA(2,1) kernel in (2) in a modified form. Normalizing the kernel to satisfy $G(0) = I_m$ to guarantee the identifiability of Σ with a new parameter $m \times m$ matrix Φ, we obtain

$$G(s) = \Phi e^{\lambda_1 s} + (I_m - \Phi) e^{\lambda_2 s}, Re(\lambda_1) < Re(\lambda_2) < 0.$$

	λ_1	λ_2	ϕ_{11}	ϕ_{22}	ϕ_{33}	ϕ_{21}	ψ_{21}	ϕ_{31}	ψ_{31}	ϕ_{32}	ψ_{32}	$\log\sigma_2^2$	$\log\sigma_3^2$
true	−3.951	−0.619	0.822	0.864	0.825	1.595	0.160	1.017	0.032	0.608	0.079	0	0
mean	−3.938	−0.648	0.825	0.858	0.805	1.547	0.156	0.963	0.030	0.694	0.079	−0.002	−0.191
RMSE	0.325	0.083	0.027	0.145	0.069	0.204	0.033	0.163	0.033	0.293	0.096	0.500	0.372

TABLE 1. The means and root mean squared errors of the Whittle estimators for tri-variate CARMA (2,1) random field in (10), which were evaluated by 100 simulations.

We impose a restriction on Φ and Σ, the variance matrix of Lévy sheet, to increase the identifiability of CARMA kernel by the Whittle likelihood in (6).

For the Cholesky decomposition for Σ, which is given by

$$\Sigma = L\mathrm{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)L',$$

for the lower triangular matrix L with $L_{ii} = 1, i = 1, 2, 3$, modify the CARMA(2,1) kernel as

$$G(s)L = \Phi L e^{\lambda_1 s} + (I_3 - \Phi)L e^{\lambda_2 s}.$$

We restrict the parameter matrix Φ to be within the class of lower triangular, which improves significantly low identifiability of Φ by the Whittle likelihood. As a result, the tri-variate CARMA(2,1) kernel we shall employ in this section is expressed as

$$(10) \quad G(s) = \begin{pmatrix} \phi_{11} & 0 & 0 \\ \phi_{21} & \phi_{22} & 0 \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} e^{\lambda_1 s} + \begin{pmatrix} 1-\phi_{11} & 0 & 0 \\ \psi_{21} & 1-\phi_{22} & 0 \\ \psi_{31} & \psi_{32} & 1-\phi_{33} \end{pmatrix} e^{\lambda_2 s},$$

with the diagonal variance matrix Σ for Lévy sheet, which is re-parametrized as

$$\Sigma = \sigma_1^2 \mathrm{diag}(1, \sigma_2^2, \sigma_3^2).$$

Recalling that the Whittle likelihood in (6) is scale invariant, we notice that the estimable parameters in the CARMA(2,1) model are $\lambda_1, \lambda_2, \phi_{ij}, \psi_{ij}$ and σ_2^2, σ_3^2.

5.2. **Simulation studies.** W examined the Whittle estimation for simulated tri-variate data by the CARMA (2,1) kernel in (10) on irregularly spaced points over the compact set $A = [0, 50] \times [0, 30]$. More specifically, we employed the empirically modified expression in (7) driven by a Poisson Lévy sheet, where 4,000 knots uniformly distributed over $[0, 60] \times [0, 60]$ including A. We designed three independent sets of 5,000 uniformly distributed points over A as observation points for each component of tri-variate data. In the notation in the last section, S_1, S_2 and S_3 were designed to have no intersections. We simulated 100 sets of the tri-variate data under the setting, where the parameter values for Φ in Table 1 were taken from the empirical analysis for US precipitation data shown in the next section.

We estimated the CARMA (2,1) parameters Φ in (10) and σ_2^2, σ_3^2 in Σ to minimize the Whitttle likelihood for the 100 sets of simulated data, where the compact region D in (6) was designed by $\{\omega \in \mathbb{R}^2, ||\omega|| < 2\pi\}$. In Table 1, we showed the means and root mean squared errors evaluated by 100 simulations.

We find from Table 1 that the Whittle estimation works overall for all the parameters in terms of bias and RMSE. The variance parameters has larger RMSE than the other parameters. The Whittle likelihood evaluates the likelihood over the compact region within D, namely it ignores the periodograms on higher frequency regions than D. The ignorance on higher frequencies leads to poor estimation for

Figure 1. Weather stations in United States

the variance parameters in Lévy sheets in comparisons with those for the other parameters in CARMA kernels.

5.3. **Real example.** This section demonstrates the applications of the tri-variate CARMA (2,1) random field in (10) to real dataset of US precipitation. Monthly total precipitation observed at weather stations all over US from 1895 through 1997 are available in the web page of Institute for Mathematics Applied to Geosciences (IMAG):

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

Around 6,000 weather stations, which are scattered over United Stats, are shown in Figure 1.

Regarding the monthly precipitation in November, December, 1996 and January, 1997 as tri-variate spatial observations, we fitted the tri-variate CARMA(2,1) model to them to examine the estimation and kriging performances. Dividing the whole dataset into the two sets of in-samples and out-of-samples, we estimated the CARMA parameters to minimize the Whittle likelihood in (6) by the in-samples and evaluated the mean squared errors of the kriging constructed with the estimated CARMA model by the out-of-samples. More specifically, the numbers of the data points in Nov., Dec. and Jan. were 6,841, 6,838 and 6,463, respectively and we divided each of them into randomly chosen 500 out-of-samples and the rest as in-samples. The in-sample datasets in Nov. and Dec., in Nov. and Jan. and in Dec. and in Jan. respectively have intersections of 5,772, 5,400 and 5,415 stations, respectively. For the Whittle estimation, we took $A = [0, 50] \times [0, 30]$ to construct the periodogram and chose D in (6) as the compact region of $\{\omega \in \mathbb{R}^2, ||\omega|| < 2\pi\}$. At Step 1 in the Bayesian kriging procedure, we designed as knots the points chosen randomly from among the 6,838 stations in Dec. with M following the Poisson distribution with mean 6,000. To avoid the difficulty of huge matrix inversion in Step

	λ_1	λ_2	ϕ_{11}	ϕ_{22}	ϕ_{33}	ϕ_{21}	ψ_{21}	ϕ_{31}	ψ_{31}	ϕ_{32}	ψ_{32}	$\log \sigma_2^2$	$\log \sigma_3^2$
est.	−3.951	−0.619	0.822	0.864	0.825	1.595	0.160	1.017	0.032	0.608	0.079	0.879	−0.903
s.e.	0.318	0.068	0.023	0.020	0.029	0.083	0.044	0.067	0.032	0.027	0.016	0.088	0.114

TABLE 2. The Whittle estimation with the standard error for the tri-variate CARMA (2,1) random field in (10) fitted to US precipitation data in Nov., Dec., 1996 and Jan., 1997.

MSE		kernel smoother	uni-variate CARMA(2,1)	tri-variate CARMA(2,1)
in-sample	Nov.	4.79	6.36	4.32
	Dec.	15.84	17.67	10.05
	Jan.	7.41	9.22	5.37
out-of-sample	Nov.	8.83	7.94	4.88
	Dec.	26.43	27.94	13.84
	Jan.	18.42	15.82	10.12

TABLE 3. Kriging MSE by the estimated tri-variate CARMA(2,1) random field in comparisons with those of the benchmarks of kernel smoother and univariate CARMA(2,1) random field.

2, we employed the sub-chain of 2-a, b and c with randomly chosen 50 sub-regions in the whole US continent and iterated Step 2-c four times. We iterated 200 times Steps 1-3 in the kriging procedure and constructed krigigged values by the averages of the last 100 posterior samples.

The Whittle estimators for the CARMA(2,1) parameters are in Table 2 and the identified autocorrelation matrix by the formula in Theorem 1 is in Figure 2, while the mean squared errors of the tri-variate CARMA kriging are shown in Table 3 with those of the Gaussian kernel smoother and univariate CARMA (2,1) model as benchmarks, where the bandwidth for the kernel smoothing was optimized to give the best kriging performances. The standard errors in Table 2 were evaluated by $2b_g H^{-1}/|A|$, where H is the Hessian matrix of the Whittle likelihood in (6) and b_g was evaluated by the kernel density estimation, $4\pi^2 \times 3.09$. Notice that it ignored the asymptotic variance terms of Ω_2 and Π in Theorem 3, which would cause negatively biased approximation.

We summarize the results in the three points. First, the comparison between Tables 1 and 2 demonstrates the negatively biased standard errors in Table 2. Namely, the standard errors obtained via the inverse Hessian are smaller than the ones via the simulations, following the expectation from the asymptotic variance matrix in Theorem 3. Second, the univariate CARMA models and kernel smoothing provide the close kriging MSEs. This is because kriging by uni-variate CARMA models is regarded as a kind of smoothing with the kernel and bandwidth specified by CARMA kernels. Finally, it is found from Figure 2 that correlation of Dec. and Jan. is more durable than those of the other two especially in short lag distances, which accounts for the significantly better performances of the tri-variate CARMA kriging over the univariate benchmarks. A local shock at a point that the univariate benchmarks cannot account can be caught by the tri-variate CARMA models via the correlations between Dec. and Jan. or Dec. and Nov., which resulted in the significant kriging improvements.

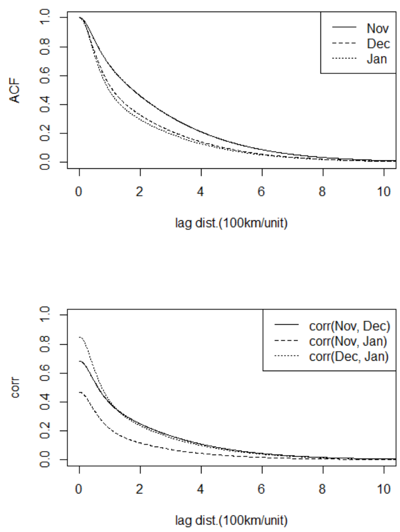

FIGURE 2. Autocorrelations as a function of lag distance identified by the tri-variate CARMA(2,1) random field for US precipitation in Nov., Dec., 1996 and Jan., 1997.

6. CONCLUSION

This paper conducts a multivariate extension of univariate CARMA random fields proposed by Brockwell and Matsuda [7], which is obtained as a continuous analogue of discrete time moving average models. The features of the extension are summarized in the following points. First, the extension is simply designed to provide with explicit parametric expressions of multivariate CARMA kernel matrix and as a result with the spectral density matrix. Second, we propose the Whittle likelihood to estimate the CARMA parameters efficiently. Unlike usual normal likelihood in spatial domains that requires huge dimensional matrix inversion of covariances, the Whittle likelihood is efficiently evaluated by the matrix inversion of spectral density matrices. Third, we successfully derived the asymptotic normality of the Whittle estimation without imposing Gaussian assumption but with the existence of all finite order moments. The asymptotic variance matrix is similar but different from the one in traditional discrete time series analysis. The difference comes from the feature of continuous random fields on which the celebrated Kolmogorov formula does not hold any more. Fourth, we propose a Bayesian way of kriging by multivariate CARMA random fields under Gaussian conditions. Although it requires huge matrix inversion, we give a way to avoid the difficulty to provide kriging efficiently. Finally, our proposed methodology of multivariate CARMA random fields works well in practice. Multivariate CARMA model captures well durable spatial correlations among components of multivariate observations to provide better kriging than those by univariate modeling.

The summarized features are all related with second order properties of CARMA random fields, which are resulted from the assumption that driving Lévy sheet has finite variance matrix. Our future study is to examine CARMA models driven by infinite variance Lévy sheets, which would open fruitful applications in theories and practices for random fields.

7. Proofs

Proof of Theorem 2. Let $\theta_1 \in \Theta$ be a parameter that is not equal to θ_0. By Lemmas 1 and 2,

$$l_w(\theta_1) \to \log\left[\frac{1}{m|D|}\int_D tr\left\{f^{-1}(\theta_1;\omega)f(\theta_0,\omega)\right\}d\omega\right] + \frac{1}{m|D|}\int_D \log\|f(\theta_1;\omega)\|\,d\omega + \log\tau_0,$$
$$:= l_\infty(\theta_1),$$

say, in probability as k tends to be infinity, where τ_0 is defined in Lemma 2. Hence, by Jensen's inequality,

$$l_\infty(\theta_1) - l_\infty(\theta_0) =$$
$$\log\left[\frac{1}{m|D|}\int_D tr\left\{f^{-1}(\theta_1;\omega)f(\theta_0,\omega)\right\}d\omega\right] - \frac{1}{m|D|}\int_D \log\left\|f^{-1}(\theta_1;\omega)f(\theta_0;\omega)\right\|d\omega$$
$$> 0.$$

It follows that, for any positive constant $L(\theta_0,\theta_1)$ that is smaller than $l_\infty(\theta_1) - l_\infty(\theta_0)$,

$$\lim_{k\to\infty} P\left(l_w(\theta_0) - l_w(\theta_1) < -L(\theta_0,\theta_1)\right) = 1.$$

For any $\delta > 0$, there exists an $H_{k,\delta}$ of the form,

$$\delta\left\{\frac{C_1}{|J_D|}\sum_{j\in J_D} tr\left(I(\omega_j)\right)\right\}^{-1}\left\{\frac{C2}{|J_D|}\sum_{j\in J_D} tr\left(I(\omega_j)\right) + C_3\right\}$$

such that, for any θ_1 and θ_2 that satisfy $|\theta_2 - \theta_1| < \delta$,

$$|l_w(\theta_2) - l_w(\theta_1)| < H_{k,\delta},$$

because of the non-negative definiteness of $\hat{K}$ and of the mean value theorem. It is seen from the form of $H_{k,\delta}$ that there exists a $\delta > 0$ such that

$$\lim_{k\to\infty} P(H_{k,\delta} < K(\theta_0,\theta_1)) = 1.$$

Applying Lemma 2 of Walker [23], we have the consistency. □

Proof of Theorem 3. By Taylor series expansion,

$$0 = \frac{\partial l_w(\hat\theta)}{\partial\theta} = \frac{\partial l_w(\theta_0)}{\partial\theta} + \frac{\partial^2 l_w(\theta^*)}{\partial\theta\partial\theta'}(\hat\theta - \theta_0),$$

where θ^* is the mean value between θ_0 and $\hat\theta$. Hence

$$\sqrt{|A|}\left(\hat\theta - \theta_0\right) = \left\{m|D|\frac{\partial^2 l_w(\theta^*)}{\partial\theta\partial\theta'}\right\}^{-1}\sqrt{|A|}\left\{-m|D|\frac{\partial l_w(\theta_0)}{\partial\theta}\right\}.$$

The (p,q)th element of the first factor, which is the Hessian matrix, is evaluated as, for $p,q=1,\ldots,pdim$,

$$
\begin{aligned}
&-\frac{|D|}{|J_D|}\sum_{j\in J_D} tr\left[\frac{\partial f^{-1}(\theta^*;\omega_j)}{\partial\theta_p}\frac{\partial f(\theta^*;\omega_j)}{\partial\theta_q}\right]\\
&-\frac{1}{m|D|}\frac{|D|}{|J_D|}\sum_{j\in J_D} tr\left[\frac{\partial f^{-1}(\theta^*;\omega_j)}{\partial\theta_p}\frac{I(\omega_j)}{\hat\tau(\theta^*)}\right]\times\frac{|D|}{|J_D|}\sum_{j\in J_D} tr\left[\frac{\partial f^{-1}(\theta^*;\omega_j)}{\partial\theta_q}\frac{I(\omega_j)}{\hat\tau(\theta^*)}\right]\\
&+o_p(1),
\end{aligned}
$$

where

$$
\hat\tau(\theta)=\frac{1}{m|J_D|}\sum_{j\in J_D} tr\left[f^{-1}(\theta;\omega_j)I(\omega_j)\right].
$$

Noting that θ^* converges to θ_0, we find that the Hessian converges in probability to $\Omega_1-\Omega_2$ by Lemmas 1 and 2.

The pth element of the second factor, which is the score vector, is evaluated as, for $p=1,\ldots,pdim$,

$$
\frac{\sqrt{|A|}|D|}{|J_D|}\sum_{j\in J_D} tr\left[\left\{\frac{I(\omega_j)}{\hat\tau(\theta_0)}-f(\theta_0;\omega_j)\right\}\frac{\partial f^{-1}(\theta_0;\omega_j)}{\partial\theta_p}\right]+o_p(1),
$$

which is, by Lemma 1, equal to

$$
\frac{\sqrt{|A|}|D|}{|J_D|}\sum_{j\in J_D} tr\left[\left\{\frac{I(\omega_j)}{\tau_0}-\frac{EI(\omega_j)}{\tau_0}\right\}\frac{\partial f^{-1}(\theta_0;\omega_j)}{\partial\theta_p}\right]+o_p(1). \tag{11}
$$

By applying Lemma 3 to the first term, it is equivalent to show the asymptotic distribution of

$$
J_p=\sqrt{|A|}\int_D tr\left[\left\{\frac{I(\omega)}{\tau_0}-\frac{EI(\omega)}{\tau_0}\right\}\frac{\partial f^{-1}(\theta_0;\omega)}{\partial\theta_p}\right]d\omega, p=1,\ldots,pdim.
$$

For simplicity, we define

$$
\begin{aligned}
\phi_p(\omega)&=\frac{\partial f^{-1}(\theta_0;\omega)}{\partial\theta_p},\\
\tilde\phi_p(s)&=\int_D\phi_p(\omega)e^{-i\omega' s}d\omega,
\end{aligned}
$$

and re-express the random term for J_p as

$$
T_p=\sum_{a=1}^m\sum_{b=1}^m\frac{|A|^{3/2}}{n_an_b}\sum_{c=1}^{n_a}\sum_{d=1}^{n_b}\tau_0^{-1}X_a(s_{ac})X_b(s_{bd})\tilde\phi_{ba,p}(s_{ac}-s_{bd}). \tag{12}
$$

We shall show in Lemmas 4 and 5 that

$$
\begin{aligned}
Cov\,(T_p,T_q)&\to b_g(2\Omega_{1,pq}+\Pi_{pq}), p,q=1,\ldots,pdim,\\
cum\,(T_{p_1},\ldots,T_{p_r})&\to 0, p_1,\ldots,p_r=1,\ldots,pdim,\ \text{for } r\geq 3,
\end{aligned}
$$

as k tends to ∞, which proves the asymptotic normality in Theorem 3. □

8. Lemmas

Lemma 1.

$$EI_{ab}(\omega) = \tau_0 f_{ab}(\omega) + O\left(\frac{n_{ab}|A|}{n_a n_b} + A_1^{-2} + A_2^{-2}\right), a, b = 1, \ldots, m,$$

where n_a, n_b and n_{ab} are the number of elements in S_a, S_b and $S_a \cap S_b$, respectively, and

$$\tau_0 = (2\pi)^2 \int_{[0,1]^2} |g(x)|^2 dx.$$

Proof. We find from Lemma 3 in Matsuda and Yajima [18] that the expectation is evaluated as

$$\tau_0 f_{ab}(\omega) + \frac{|A|}{n_a n_b} \sum_{s_j \in S_a \cap S_b} EX_a(s_j) X_b(s_j) + O\left(A_1^{-2} + A_2^{-2}\right),$$

which completes the proof. □

Lemma 2. *For a square integrable function $\psi(\omega), \omega \in \mathbb{R}^2$,*

$$var\left\{\sqrt{|A|}\frac{|D|}{|J_D|}\sum_{j \in J_D} I_{ab}(\omega_j)\psi(\omega_j)\right\} = O(1), a, b = 1, \ldots, m.$$

Proof. Let

$$\hat{\psi}(s) = \frac{|D|}{|J_D|}\sum_{j \in J_D} \psi(\omega_j) e^{-i\omega_j' s}, s \in A = [0, A_1] \times [0, A_2],$$

which is extended periodically to $[-A_1, A_1] \times [-A_2, A_2]$. Then the object for the variance is evaluated as

$$\frac{|A|^{3/2}}{n_a n_b}\sum_{c=1}^{n_a}\sum_{d=1}^{n_b} X_a(s_{ac}) X_b(s_{bd}) \hat{\psi}(s_{ac} - s_{bd}).$$

The variance is given by

$$\begin{aligned}
&E\frac{|A|^3}{n_a^2 n_b^2}\sum_{c_1}\sum_{d_1}\sum_{c_2}\sum_{d_2}\Big(cum\left(X_a(s_{ac_1}), X_b(s_{bd_1}), X_a(s_{ac_2}), X_b(s_{bd_2})\right) \\
&+ \gamma_{aa}(s_{ac_1} - s_{ac_2})\gamma_{bb}(s_{bd_1} - s_{bd_2}) + \gamma_{ab}(s_{ac_1} - s_{bd_2})\gamma_{ba}(s_{bd_1} - s_{ac_2})\Big) \\
&\times \hat{\psi}(s_{ac_1} - s_{bd_1})\hat{\psi}(s_{ac_2} - s_{bd_2}) \\
=&\int_A\int_A\int_A\int_A\Big(cum(X_a(u_1), X_b(v_1), X_a(u_2), X_b(v_2)) \\
&+ \gamma_{aa}(u_1 - u_2)\gamma_{bb}(v_1 - v_2) + \gamma_{ab}(u_1 - v_2)\gamma_{ba}(v_1 - u_2)\Big) \\
&\times \delta(u_1 - v_1)\delta(u_2 - v_2)|A|^{-1} g(u_1/A) g(v_1/A) g(u_2/A) g(v_2/A) du_1 dv_1 du_2 dv_2 + o(1).
\end{aligned}$$

The first term is, by expressing the cumulant term with the cumulant spectrum:

$$f_{abab}(\omega_1, \omega_2, \omega_3) = \sum_{e,f,g,h=1}^{m} \kappa_{efgh}\tilde{G}_{ae}(\omega_1)\tilde{G}_{bf}(\omega_2)\tilde{G}_{ag}(\omega_3)\overline{\tilde{G}_{bh}(\omega_1 + \omega_2 + \omega_3)}, \tag{13}$$

given by

$$\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} f_{abab}(\omega_1,\omega_2,\omega_3)e^{i\omega_1'(u_1-v_2)}e^{i\omega_2'(v_1-v_2)}e^{i\omega_3'(u_2-v_2)}d\omega_1 d\omega_2 d\omega_3$$
$$|A|^{-1}\int_A\int_A\int_A\int_A \delta(u_1-v_1)\delta(u_2-v_2)g(u_1/A)g(v_1/A)g(u_2)/Ag(v_2/A)du_1dv_1du_2dv_2,$$

which is, by Schwarz inequality, bounded by

$$\prod_{j=1}^{2}\sqrt{\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}|f_{abab}(\omega_1,\omega_2,\omega_3)|P_j d\omega_1 d\omega_2 d\omega_3}, \tag{14}$$

for

$$P_1 = |A|^{-1}\left|\int_A\int_A \hat{\psi}(u_1-v_1)g(u_1/A)g(v_1/A)e^{i\omega_1' u_1}e^{i\omega_2' v_1}du_1dv_1\right|^2,$$
$$P_2 = |A|^{-1}\left|\int_A\int_A \hat{\psi}(u_2-v_2)g(u_2/A)g(v_2/A)e^{-i(\omega_1+\omega_2)' v_2}e^{i\omega_3'(u_2-v_2)}du_2dv_2\right|^2.$$

By applying Perseval's equality to both terms, (14) is bounded by

$$C\int_{-A_1}^{A_1}\int_{-A_2}^{A_2}\left|\hat{\psi}(s)\right|^2 ds, \tag{15}$$

which is evaluated as

$$4C|A|\frac{|D|^2}{|J_D|^2}\sum_{j\in J_D}|\psi(\omega_j)|^2 < C'\int_D |\psi(\omega)|^2 d\omega = O(1).$$

Also the second and third terms in the variance are bounded by a constant with the same argument, which completes the proof. □

Lemma 3. *For a square integrable function* $\psi(\omega), \omega \in \mathbb{R}^2$,

$$\frac{\sqrt{|A|}|D|}{|J_D|}\sum_{j\in J_D} I_{ab}(\omega_j)\psi(\omega_j) - \sqrt{|A|}\int_D I_{ab}(\omega)\psi(\omega)d\omega = o_p(1).$$

Proof. Let

$$\hat{\psi}(s) = \frac{|D|}{|J_D|}\sum_{j\in J_D}\psi(\omega_j)e^{-i\omega_j' s}, s \in A = [0,A_1]\times[0,A_2],$$
$$\tilde{\psi}(s) = \int_D \psi(\omega)e^{-i\omega' s}d\omega, s\in\mathbb{R}^2,$$

the first one of which is extended periodically to $[-A_1,A_1]\times[-A_2,A_2]$. For $\delta(s) = \hat{\psi}(s)-\tilde{\psi}(s)$, the difference is evaluated as

$$\frac{|A|^{3/2}}{n_a n_b}\sum_{c=1}^{n_a}\sum_{d=1}^{n_b} X_a(s_{ac})X_b(s_{bd})\delta(s_{ac}-s_{bd}).$$

The variance, which is similarly evaluated till (15) in Lemma 2, is bounded by

$$C\int_{-A_1}^{A_1}\int_{-A_2}^{A_2}|\delta(s)|^2 ds. \tag{16}$$

Notice that $\hat{\psi}(s), \tilde{\psi}(s)$ are square integrable, since

$$\int_0^{A_1}\int_0^{A_2} |\hat{\psi}(s)|^2 ds = |A||D|^2|J_D|^{-2}\sum_{j\in J_D}|\psi(\omega_j)|^2 < C\int_D |\psi(\omega)|^2 d\omega < \infty,$$

$$\int_{\mathbb{R}^2}|\tilde{\psi}(s)|^2 ds = (2\pi)^2\int_D |\psi(\omega)|^2 d\omega < \infty.$$

It follows that, for any $\varepsilon > 0$, there exists a compact set $B_M = [-M_1, M_1] \times [-M_2, M_2] \subset [-A_1, A_1] \times [-A_2, A_2]$ such that

$$\int_{-A_1}^{A_1}\int_{-A_2}^{A_2}\left|\hat{\psi}(s) - \hat{\psi}(s)I_{B_M}(s)\right|^2 ds < \varepsilon,$$

$$\int_{\mathbb{R}^2}\left|\tilde{\psi}(s) - \tilde{\psi}(s)I_{B_M}(s)\right|^2 ds < \varepsilon.$$

Then (16) is bounded by

$$C\left\{\int_{-A_1}^{A_1}\int_{-A_2}^{A_2}\left|\hat{\psi}(s) - \hat{\psi}(s)I_{B_M}(s)\right|^2 ds + \int_{B_M}\left|\hat{\psi}(s) - \tilde{\psi}(s)\right|^2 ds + \int_{\mathbb{R}^2}\left|\tilde{\psi}(s)I_{B_M}(s) - \tilde{\psi}(s)\right|^2 ds\right\}$$
$$< C'\varepsilon,$$

which completes the proof. □

Lemma 4.

$$Cov(T_p, T_q) \to b_g(2\Omega_{1,pq} + \Pi_{pq}), p, q = 1, \ldots, pdim.$$

Proof. First, $\tilde{\phi}_p(s)$ in T_p defined in (12) may be replaced with

$$\tilde{\phi}_p^M(s) = \tilde{\phi}_p(s)I_{B_M}(s), p = 1, \ldots, pdim,$$

for a sufficiently large compact set $B_M = [-M_1, M_1] \times [-M_2, M_2]$, since the variance of the difference between them may be made arbitrary small by following the argument till (15) in Lemma 2. The covariance for the one replaced with $\tilde{\phi}_p^M(s)$ is evaluated as

$$\sum_{a_1,a_2=1}^{m}\sum_{b_1,b_2=1}^{m}|A|^{-1}\int_A\int_A\int_A\int_A \tau_0^{-2}\{cum(X_{a_1}(u_1), X_{b_1}(v_1), X_{a_2}(u_2), X_{b_2}(v_2))$$
$$+ \gamma_{a_1a_2}(u_1 - u_2)\gamma_{b_1b_2}(v_1 - v_2) + \gamma_{a_1b_2}(u_1 - v_2)\gamma_{b_1a_2}(v_1 - u_2)\}$$
$$\times \tilde{\phi}_{p,b_1a_1}^M(u_1 - v_1)\tilde{\phi}_{q,b_2a_2}^M(u_2 - v_2)g(u_1/A)g(v_1/A)g(u_2/A)g(v2/A)du_1dv_1du_2dv_2$$
$$+ o(1).$$

The first term is, by using the cumulant spectrum defined in (13), evaluated as

$$\sum_{a_1,a_2}\sum_{b_1,b_2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} f_{a_1b_1a_2b_2}(\omega_1, \omega_2, \omega_3)|A|^{-1}\tau_0^{-2}P_1P_2 d\omega_1 d\omega_2 d\omega_3$$

for

$$P_1 = \int_A\int_A \tilde{\phi}_{p,b_1a_1}^M(u_1 - v_1)e^{i\omega_1' u_1}e^{i\omega_2' v_1}g(u_1/A)g(v_1/A)du_1dv_1,$$

$$P_2 = \int_A\int_A \tilde{\phi}_{q,b_2a_2}^M(u_2 - v_2)e^{i\omega_3' u_2}e^{-i(\omega_1+\omega_2+\omega_3)' v_2}g(u_2/A)g(v_2/A)du_2dv_2.$$

By change of variables by $u_1 - v_1 = l_1, u_2 - v_2 = l_2$ and the compactness of the supports of $\tilde{\phi}^M$, $|A|^{-1}P_1P_2$ is evaluated as

$$\phi^M_{p,b_1a_1}(\omega_1)\phi^M_{q,b_2a_2}(\omega_3)F(\omega_1+\omega_2)+o(1),$$

where

$$F(\omega) = |A|^{-1}\left|\int_A g^2(u/A)e^{i\omega' u}du\right|^2,$$
$$\phi^M_p(\omega) = (2\pi)^{-2}\int_{B_M}\tilde{\phi}^M_p(s)e^{i\omega' s}ds.$$

It follows by Lemma 1(c) in Matsuda and Yajima [18] that the first term converges to

$$\sum_{a_1,a_2=1}^{m}\sum_{b_1,b_2=1}^{m} b_g\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} f_{a_1b_1a_2b_2}(\omega_1,-\omega_1,\omega_3)\phi^M_{p,b_1a_1}(\omega_1)\phi^M_{q,b_2a_2}(\omega_3)d\omega_1 d\omega_3.$$

Replacing the cumulant spectrum with (13), we have the result arbitrary close to $b_g\Pi_{pq}$ by taking B_M large. By the same argument, the second and third terms converge to $b_g\Omega_{1,pq}$, which completes the proof. □

Lemma 5. *For $r \geq 3$,*

$$cum\left(T_{p_1},\dots,T_{p_r}\right) = O\left(|A|^{-r/2+1}\right), p_1,\dots,p_r = 1,\dots,pdim.$$

Proof. First, $\tilde{\phi}_p(s)$ in T_p defined in (12) may be replaced with

$$\tilde{\phi}^M_p(s) = \tilde{\phi}_p(s)I_{B_M}(s), p = 1,\dots,pdim,$$

for a sufficiently large compact set $B_M = [-M_1, M_1]\times[-M_2, M_2]$, since the variance of the difference between them may be made arbitrary small by following the argument till (15) in Lemma 2.

The cumulant for the one replaced with $\tilde{\phi}^M_p(s)$ is evaluated as

$$\begin{aligned} cum(T_{p_1},\dots,T_{p_r}) &= \sum_{a_1,b_1=1}^{m}\cdots\sum_{a_r,b_r=1}^{m}|A|^{-r/2} \\ &\times\int_A\cdots\int_A cum(X_{a_1}(u_1)X_{b_1}(v_1),\dots,X_{a_r}(u_r)X_{b_r}(v_r)) \\ &\times\prod_{j=1}^{r}\tilde{\phi}^M_{p_j,b_ja_j}(u_j-v_j)g(u_j/A)g(v_j/A)du_jdv_j + o(1). \end{aligned}$$

Let $Y_i = (Y_{i1}, Y_{i2})$ be the two dimensional random vector defined by $(X_{a_i}(u_i), X_{b_i}(v_i))$ for $i = 1,\dots,r$. The cumulant term in the equation is expressed by the formula in Leonov and Shiryaev [17] as

$$\sum_{\cup_{p=1}^{q}D_p}\prod_{p=1}^{q} cum(Y(D_p)),$$

where the summation is taken over all the indecomposable partition $\cup_{p=1}^{q}D_p$ of the two way table of indices:

$$\begin{matrix} (1,1) & (1,2) \\ \vdots & \vdots \\ (r,1) & (r,2) \end{matrix}.$$

Let us prove only for $cum(Y_{11}, Y_{12}, \dots, Y_{r1}, Y_{r2})$, the highest order term, since the other cases are similarly evaluated. We express the highest cumulant term by the $2r$th order cumulant spectrum as

$$cum(X_{a_1}(u_1), X_{b_1}(v_1), \dots, X_{a_r}(u_r), X_{b_r}(v_r))$$
$$= \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} f_{a_1 b_1 \cdots a_r b_r}(\omega_1, \dots, \omega_{2r-1}) \prod_{j=1}^{r} e^{i\omega_j'(u_j - v_r)} \prod_{j=1}^{r-1} e^{i\omega_j'(v_j - v_r)} d\omega_1 \cdots d\omega_{2r-1},$$

where

$$(17) \quad f_{a_1 b_1 \cdots a_r b_r}(\omega_1, \dots, \omega_{2r-1}) = \sum_{e_1, \dots, e_{2r}=1}^{m} \kappa_{e_1, \dots, e_{2r}}$$
$$\times \overline{\tilde{G}_{b_r, e_{2r}}(\omega_1 + \ldots + \omega_{2r-1})} \prod_{j=1}^{r} \tilde{G}_{a_j, e_{2j-1}}(\omega_{2j-1}) \prod_{j=1}^{r-1} \tilde{G}_{b_j, e_{2j}}(\omega_{2j}).$$

By replacing the highest cumulant term with the spectrum expression, the summand of the corresponding term is given by

$$|A|^{-r/2} \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} f_{a_1 b_1 \cdots a_r b_r}(\omega_1, \dots, \omega_{2r-1}) d\omega_1 \cdots \omega_{2r-1}$$
$$\times \prod_{j=1}^{r-1} \int_A \int_A \tilde{\phi}^M_{p_j, b_j a_j}(u_j - v_j) e^{i\omega'_{2j-1} u_j} e^{i\omega'_{2j} v_j} g(u_j/A) g(v_j/A) du_j dv_j$$
$$\int_A \int_A \tilde{\phi}^M_{p_r, b_r a_r}(u_r - v_r) e^{i\omega'_{2r-1} u_r} e^{-i(\omega_1 + \cdots + \omega_{2r-1})' v_r} g(u_r/A) g(v_r/A) du_r dv_r.$$

which is, by changes of variables by $u_j - v_j = l_j, j = 1, \dots, r$ and the compactness of the supports of $\tilde{\phi}^M$, evaluated as

$$|A|^{-r/2} \int_{\mathbb{R}^2} \cdots \int_{\mathbb{R}^2} f_{a_1 b_1 \cdots a_r b_r}(\omega_1, \dots, \omega_{2r-1}) d\omega_1 \cdots \omega_{2r-1}$$
$$\times \phi^M_{p_r, b_r a_r}(\omega_{2r-1}) D(-\omega_1 - \cdots - \omega_{2r-2}) \prod_{j=1}^{r-1} \phi^M_{p_j, b_j a_j}(\omega_{2j-1}) D(\omega_{2j-1} + \omega_{2j}) + o(1),$$

where

$$D(\omega) = \int_A g^2(u/A) e^{i\omega' u} du.$$

Again by change of variables of $\omega_{2j-1}+\omega_{2j}=\lambda_j, j=1,\ldots,r-1$, and by replacing the spectrum with (17), it is evaluated as

$$|A|^{-r/2}\sum_{e_1,\ldots,e_{2r}=1}^{m}\kappa_{e_1,\ldots,e_{2r}}\int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2}\overline{\tilde{G}_{b_r,e_{2r}}(\lambda_1+\ldots+\lambda_{r-1}+\omega_{2r-1})}$$
$$\times\prod_{j=1}^{r}\tilde{G}_{a_j,e_{2j-1}}(\omega_{2j-1})\prod_{j=1}^{r-1}\tilde{G}_{b_j,e_{2j}}(\lambda_j-\omega_{2j-1})d\omega_1 d\lambda_1\cdots d\omega_{2r-3}d\lambda_{r-1}d\omega_{2r-1}$$
$$\times\,\phi^M_{p_r,b_r a_r}(\omega_{2r-1})D(-\lambda_1-\cdots-\lambda_{r-1})\prod_{j=1}^{r-1}\phi^M_{p_j,b_j a_j}(\omega_{2j-1})D(\lambda_j),$$

whose summand is bounded by

$$C|A|^{-r/2+1}\prod_{j=1}^{r}\int_{\mathbb{R}^2}\tilde{G}_{a_j,e_{2j-1}}(\omega_{2j-1})d\omega_{2j-1}\prod_{j=2}^{r-1}\int_{\mathbb{R}^2}\tilde{G}_{b_j,e_{2j}}(\lambda_j-\omega_{2j-1})D(\lambda_j)d\lambda_j$$
$$\times\,|A|^{-1}\int_{\mathbb{R}^2}\overline{\tilde{G}_{b_r,e_{2r}}(\lambda_1+\ldots+\lambda_{r-1}+\omega_{2r-1})}\tilde{G}_{b_1,e_2}(\lambda_1-\omega_1)D(\lambda_1)D(-\lambda_1-\cdots-\lambda_{r-1})d\lambda_1,$$

which we find to be $O(|A|^{-r/2+1})$ by Lemma 2 in Matsuda and Yajima [18]. □

Acknowledgments We sincerely thank Professor Peter Brockwell of Colorado State University for his kind advice to the derivation of the auto-covariance matrices in Theorem 1.

References

[1] Bandyopadhyay, S. and Lahiri, S. (2009). Asymptotic properties of discrete fourier transforms for spatial data. *Sankhya*, Series A, 71, 221259.

[2] Bandyopadhyay, S., Lahiri, S. N. and Nordman, D. (2015). A frequency domain empirical likelihood method for irregular spaced spatial data. *Annals of Statistics*, 43, 2. 519545.

[3] Bandyopadhyay, S., and Subba Rao, S. (2015). A test for stationarity of irregular spaced spatial data. *Journal of the Royal Statistical Society:Series B*, 79, 95-123.

[4] Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Englewood Cliffs: Prentice Hall.

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

[6] Brockwell, P. J. (2014). Recent results in the theory and applications of CARMA processes, *Ann. Inst. Stat. Math.*, **66**, 647-685.

[7] Brockwell, P. J. and Matsuda, Y. (2017). Continuous auto-regressive moving average random fields on $\mathbb{R}^n$. *Journal of the Royal Statistical Society: Series B*, 79(3), 833-857.

[8] Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.

[9] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 16. 137.

[10] Doob, J. L. (1944). The elementary Gaussian processes. *Annals of Mathematical Statistics*, 25, 229-282.

[11] Dunsmuir, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise. *Annals of Statistics*, 7, 490-506.

[12] Dunsmuir, W. and Hannan, E. J. (1976). Vector linear time series models. *Adv. Appl. Probability*, 8, 339-364.

[13] Gelfand, A. E. and Banerjee, S. (2010). Multivariate spatial process models. In: Gelfand, A. E., Diggle, P. J., Fuentes, M. and Guttorp, P. (eds) *Handbook of Spatial Statistics*, CRC Press. 493-515.

[14] Higdon D. (2002). Space and Space-Time Modeling using Process Convolutions. In: Anderson C.W., Barnett V., Chatwin P.C., El-Shaarawi A.H. (eds) Quantitative Methods for Current Environmental Issues. Springer, London.

[15] Hosoya, Y. (1997). A limit theory for long-range dependence and statistical inference on related models *Annals of Statistics*, .25, 1, 105-37.

[16] Kaufman, C., Schervish, M., Nychka, D. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *J. Amer. Statist. Assoc.* 103, 484, 1545-1555.

[17] Leonov, V. P. and Shiryaev, A. N. (1959). On a method of calculation of semi-invariants. *Theory of Probability and its Applications*, 4(3), 319-329.

[18] Matsuda, Y. and Yajima, Y. (2009). Fourier analysis of irregularly spaced data on Rd. *Journal of the Royal Statistical Society: Series B*, 71(1), 191-217.

[19] Matsuda, Y. and Yajima, Y. (2018). Locally stationary spatio-temporal processes. *Jpn. J. Stat. Data Sci.*, 1, 41-57.

[20] Subba Rao, S. (2018). Statistical inference for spatial statistics defined in the Fourier domain. *Annals of Statistics*, 46, 469-499.

[21] Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 23, 1630-1661.

[22] Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of Brownian Motion, *Physical Review*, **36**, 823-841.

[23] Walker, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *Journal of the Australian Mathematical Society*, 4, 363-384.

[24] Zhang, B., Sang, H. and Huang, J. Z. (2013). Full-scale approximations of spatio-temporal covariance models for large datasets. *Statist. Sin.*, 25, 99-114.

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

E-mail address: yasumasa.matsuda.a4@tohoku.ac.jp