DSSR

Discussion Paper No. 72

Locally stationary spatio-temporal processes

Yasumasa Matsuda
Yoshihiro Yajima

January , 2018

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

Locally stationary spatio-temporal processes

Yasumasa Matsuda* Yoshihiro Yajima†

Abstract

This paper proposes a locally stationary spatio-temporal processes to analyze the motivating example of US precipitation data, which is a huge data set composed of monthly observations of precipitation on thousands of monitoring points scattered irregularly all over US continent. Allowing the parameters of continuous autoregressive and moving average (CARMA) random fields by Brockwell and Matsuda [2] to be dependent spatially, we generalize locally stationary time series by Dahlhaus [3] to spatio-temporal processes that are locally stationary in space. We develop Whittle likelihood estimation for the spatially dependent parameters and derive the asymptotic properties rigorously. We demonstrate that the spatio-temporal models actually work to account for nonstationary spatial covariance structures in US precipitation data.

Keywords: CARMA kernel, Compound Poisson, Locally stationary process, Seasonal AR model, Spatially dependent spectral density function, Spatial nonstationarity, Whittle likelihood estimation.

1 Introduction

Figure 1 shows the locations of monitoring stations scattered all over US continent on which monthly precipitation has been observed since 1895. The huge spatio-temporal data set of US precipitation is the motivating example in this paper to let us consider nonstationary spatio-temporal models. US precipitation data has the following features: First, thousands of monitoring points are scattered irregularly over US continent while temporal observations are sampled in usual discrete time points. Secondly the space time covariance is obviously nonstationary. More precisely the covariance depends on space, although it may not critically on time. Thirdly data size of US precipitation is huge, namely, more than one hundred thousands even for three years period. Spatio-temporal models that account for the features are required for the analysis of US precipitation data.

Continuous autoregressive and moving average (CARMA) random fields, which were proposed by [2] as stationary spatial model defined on $\mathbf{R}^d, d \geq 2$, shall be extended for the motivating example. Extensions to spatio-temporal random fields with stationary temporal and nonstationary spatial covariances are to be tried to describe spatially dependent behaviors in US precipitation data. Stationary temporal extension can be done easily by discrete ARMA time series models, while nonstaionary spatial extension requires some careful considerations.

*Graduate School of Economics and Management, Tohoku University, Sendai 980-8576, Japan. `matsuda@econ.tohoku.ac.jp`.

†Graduate School of Economics and Management, Tohoku University, Sendai 980-8576, Japan. `yajima@e.u-tokyo.ac.jp`

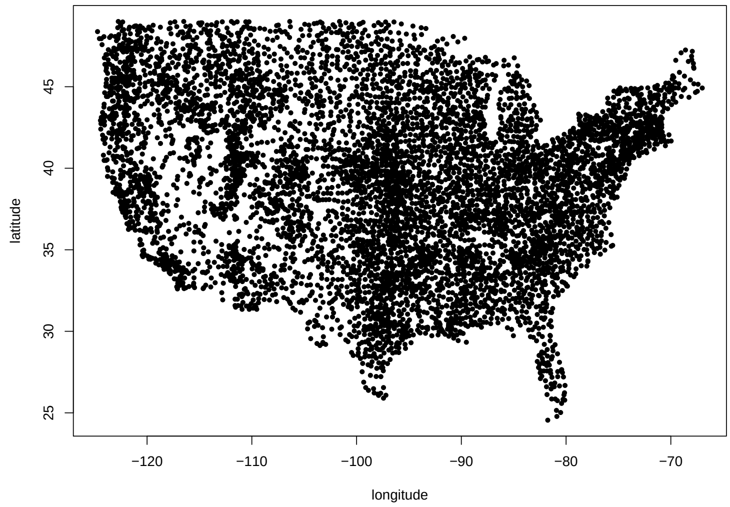

Figure 1: The locations of weather stations in US continent, on which monthly precipitation has been recorded since 1895.

Nonstationary spatial models have been attracting great interests in spatial statistics areas, since it is usual to find nonstationary features in environmental data whose covariances depend not only on lags but also on locations (Sampson [11]). Kernel-based methods by Fuentes [5], basis function approach by Nychka et al. [9], convolution models by Higdon [7], spatial deformation methods by Guttorp and Sampson [6] are the typical researches proposing nonstationary spatial models. Although all of the approaches work well to express nonstationary spatial covariances in theoretically sophisticated ways, they have often difficulties in conducting estimation and kriging for huge spatial data sets, which are often the case recently because of rapid progress of data collecting technology such as remote sensing data by satellites. US precipitation is a typical case of huge spatio-temporal data set that requires nonstationary spatial covariance models.

Locally stationary processes, proposed by Dahlhaus [3], are nonstationary time series by allowing parameters to be dependent on time. Dahlhaus [3] succeeded in estimating the time dependency of parameters by the frequency domain based method and derived the asymptotic properties rigorously. His essential idea that makes it possible to establish the asymptotic theories is in the expression of the time dependency of parameters θ, which is denoted as as $\theta(t/T)$ for sample size T. Similar researches in prior to his paper expressed time dependent parameters as $\theta(t)$, for which asymptotic arguments were difficult to formulate (Priestley [10]).

Extending locally stationary time series by Dahlhaus [3] to random fields, we propose locally stationary spatio-temporal processes. CARMA random fields with spatially dependent parameters are special cases of locally stationary spatio-temporal processes with separable covariances given by the product of stationary temporal and locally stationary spatial covariances. Following Dahlhaus [3] in estimation, we develop Whittle likelihood estimation for spatially dependent parameters in locally stationary spatio-temporal processes. To establish asymptotic theories for the estimation, we need to generalize a asymptotic scheme for time series to that for spatio-temporal data. Extending the so called mixed asymptotics in spatial data (Stein [12]) to that for spatio-temporal data, in which sample size and sampling region jointly diverge, we derive the asymptotic properties rigorously.

The striking features of locally stationary spatio-temporal CARMA random fields are as follows: First, the parameters are efficiently estimated by minimizing Whittle likelihood which requires no matrix operations. Secondly, asymptotic theories for Whittle estimations are established under the asymptotic scheme regarded as an extension of mixed asymptotics in spatial statistics literature. Thirdly, kriging and forecasting, which usually require huge matrix inversions for large spatial data set, are conducted with light computational burdens. Applying an approximation to the kriging procedure in Brockwell and Matsuda ([2]), we conduct efficient kriging that does not require matrix inversions. Finally, locally stationary CARMA models provide an easy way of simulating spatio-temporal data with spatially nonstationary and temporally stationary covariances. Simulating spatial data with nonstationary covariances is also possible as a part of simulating spatio-temporal data.

We use the following notation. For $A = (A_1, A_2), s = (s_1, s_2)$, $[0, A] = [0, A_1] \times [0, A_2]$, $|A| = A_1 \times A_2$, $s/A = (s_1/A_1, s_2/A_2)$.

2 Locally stationary random fields

2.1 Extension of stationary CARMA random fields

CARMA random fields were introduced by Brockwell and Matsuda [2] as stationary models over $\mathbf{R}^d, d \geq 2$. We shall extend them to spatio-temporal models with spatially nonstationary and temporarily stationary covariances. Consider CARMA random fields driven by a compound Poisson sheet on $\mathbf{R}^2$,

$$X(s) = \sum_j g(\theta, s - e_j) Z_j, s \in \mathbf{R}^2, \tag{1}$$

where $g(\theta, s)$ is a CARMA kernel with parameters θ, e_js are knot points distributed randomly over $\mathbf{R}^2$ and Z_js are independent and identical random variables with mean 0 and variances τ^2. Here we normalize the CARMA kernel $g(s)$ to satisfy $g(0) = 1$. Let $n(dx)$ be the number of knot points contained in the region $dx \in \mathbf{R}^2$. Then we normalize them to satisfy $E(n(dx)) = var(n(dx) = dx$. The two normalizations are necessary in order to guarantee the identifiability for τ^2.

We shall begin from the stationary temporal extension of CARMA random fields by a discrete ARMA model. Extending iid variables Z_js to stationary time series Z_{jt} by discrete ARMA models, which is defined by

$$\phi(B) Z_{jt} = \psi(B) \sigma \varepsilon_{jt}, \tag{2}$$

where ε_{jt}s are mutually independent and identically distributed random variables with mean 0 and variance 1, and ϕ and ψ are autoregressive and moving average polynomials given by

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p,$$
$$\psi(B) = 1 + \psi_1 B + \cdots + \psi_q B^q,$$

where B is the backward shift operator defined by $BZ_{jt} = Z_{j,t-1}$, we have temporally extended CARMA random fields expressed by

$$X(s, t) = \sum_j g(\theta, s - e_j) Z_{jt}, s \in \mathbf{R}^2, t = 1, 2, \ldots,$$

which provides separable space time covariances that are stationary both in space and time.

Next let us try a nonstationary extension. Allowing the parameters θ, ϕ, ψ and σ to depend spatially on s, we have the spatially nonstationary model denoted as

$$X(s,t) = \sum_j g\left(\theta(s), s - e_j\right) Z_{jt}(s), s \in \mathbf{R}^2, t = 1, 2, \ldots, \tag{3}$$

where $Z_{jt}(s)$ is the stationary ARMA time series generated by (2) with spatially dependent parameters $\phi(s), \psi(s)$ and $\sigma(s)$.

Inference for the spatially dependent parameters $\theta(s), \phi(s), \psi(s), \sigma(s)$ in (3) are inconsistent, since the domains of the parameter diverges as the observation regions $[0, A]$ for s in $X(s,t)$ tends to be large. Consistent estimation for the spatially dependent parameters requires finer samples over the domain as the sample size tends to be large. Following Dahlhaus (1997), we replace the spatial dependencies for the parameters with the local dependencies defined by $\theta(s/A), \phi(s/A), \psi(s/A), \sigma(s/A)$, which leads to the expression,

$$X_A(s,t) = \sum_j g\left(\theta\left(\frac{s}{A}\right), s - e_j\right) Z_{jt}\left(\frac{s}{A}\right), s \in [0, A], t = 1, 2, \ldots. \tag{4}$$

We call it the locally stationary spatio-temporal CARMA processes in the followings. We shall apply the spatio-temporal model to US precipitation data later in Section 4 in order to check empirically if it can actually catch the spatially nonstationary behaviors.

2.2 Locally stationary spatio-temporal processes

Here we generalize the locally stationary spatio-temporal CARMA processes in (4) to locally stationary spatio-temporal processes. Dahlhaus [3] proposed locally stationary processes to express nonstationarity with valid asymptotic theories. Here we extend the one for nonstationary time series to that for spatio-temporal data. We consider the cases when locally stationary in space but stationary in time that include (4) as a special case.

Definition 2.1. *A spatio-temporal process $X_A(s,t), s \in [0, A] \subset \mathbf{R}^2, t = 1, 2, \ldots$ is called a temporally stationary and spatially locally stationary process with transfer function K, if there exists a representation*

$$X_A(s,t) = \int_{\mathbf{R}^2} \int_{-\pi}^{\pi} K\left(\frac{s}{A}, \omega, \lambda\right) \exp(i\omega' s + i\lambda t) d\xi(\omega) d\zeta(\lambda), \tag{5}$$

where $\xi(\omega)$ and $\zeta(\lambda)$ are mutually independent stochastic processes on $\mathbf{R}^2$ and $[-\pi, \pi]$ with $\overline{\xi(\omega)} = \xi(-\omega)$ and $\overline{\zeta(\lambda)} = \zeta(-\lambda)$, respectively, and satisfy

$$cum(d\xi(\omega_1), \ldots, d\xi(\omega_k)) = \eta\left(\sum_{i=1}^k \omega_{i1}\right) \eta\left(\sum_{i=1}^k \omega_{i2}\right) a_k(\omega_1, \ldots, \omega_{k-1}) d\omega_1 \cdots d\omega_{k-1},$$
$$cum(d\zeta(\lambda_1), \ldots, d\zeta(\lambda_k)) = \eta\left(\sum_{i=1}^k \lambda_i\right) b_k(\lambda_1, \ldots, \lambda_{k-1}) d\lambda_1 \cdots d\lambda_{k-1},$$

where cum is the cumulant function, $a_1 = b_1 = 0, a_2 = b_2 = 1$, $|a_k(\omega_1, \ldots, \omega_{k-1})| \leq const_k$, $|b_k(\lambda_1, \ldots, \lambda_{k-1})| \leq const_k$ for $k \geq 3$ and $\eta(x) = \sum_{j=-\infty}^{\infty} \delta(x + 2\pi j)$ for the Dirac delta function δ.

Let us define spatially dependent spectral density function by, for $u = s/A$,

$$
\begin{aligned}
f(u,&\omega, \lambda) = \\
&\lim_{A \to (\infty,\infty)} (2\pi)^{-3} \int_{\mathbf{R}^2} \sum_{k=-\infty}^{\infty} cov(X_A(s+h/2,t), X_A(s-h/2,t-k)) \exp(-ih'\omega - ik\lambda) dh \\
&= |K(u,\omega,\lambda)|^2.
\end{aligned}
$$

Then the spatio-temporal CARMA model in (4) is a special case in (5) in the sense that the spatially dependent spectral density is expressed by

$$
|\tilde{g}_{sp}(u,\omega)|^2 \times |\tilde{g}_{tmp}(u,\lambda)|^2
$$

where

$$
\begin{aligned}
\tilde{g}_{sp}(u,\omega) &= \frac{1}{2\pi} \int_{\mathbf{R}^2} g(\theta(u), s) \exp(-i\omega' s) ds, \\
\tilde{g}_{tmp}(u,\lambda) &= \frac{\sigma(u)}{\sqrt{2\pi}} \frac{\psi(u, \exp(-i\lambda))}{\phi(u, \exp(-i\lambda))}.
\end{aligned}
$$

In other words, the model in (4) is regarded as the separable case when the transfer function is given by

$$
K(u,\omega,\lambda) = \tilde{g}_{sp}(u,\omega) \times \tilde{g}_{tmp}(u,\lambda)
$$

Example 2.2. *Consider an example for (4) when CARMA(2,1) is temporally extended by a seasonal AR polynomial* $(1-\phi_1(u)B)(1-\phi_2(u)B^{12})$. *CARMA(2,1) kernel (see eq. (31) in Brockwell and Matsuda [2]) is expressed as*

$$
g(u,s) = (1-\theta_3(u)) \exp(\theta_1(u)||s||) + \theta_3(u) \exp(\theta_2(u)||s||), \theta_1(u) < \theta_2(u) < 0.
$$

Hence the transfer function in (5) is expressed as the product of

$$
\tilde{g}_{sp}(u,\omega) = (1-\theta_3(u))\theta_1(u) \left\{||\omega||^2 + \theta_1(u)^2\right\}^{-3/2} + \theta_3(u)\theta_2(u) \left\{||\omega||^2 + \theta_2(u)^2\right\}^{-3/2}
$$

and

$$
\tilde{g}_{tmp}(u,\lambda) = \frac{\sigma(u)}{\sqrt{2\pi}} \frac{1}{(1-\phi_1(u)\exp(-i\lambda))(1-\phi_2(u)\exp(-12i\lambda))}.
$$

Finally we comment on the method to simulate spatio-temporal data that follows (5). Simulate bivariate standard normal numbers ω_j over $\mathbf{R}^2$ and uniform random numbers λ_k over $[-\pi,\pi]$. Let $\phi(\cdot)$ be the bivariate standard normal density function and suppose the transfer function is expressed by $K(\cdot) = K_1(\cdot) + iK_2(\cdot)$ for real-valued functions K_1, K_2. Simulate zero-mean iid variables ε_j and z_k with variances $1/\phi(\omega_j)$ and one, respectively. Then generate spatio-temporal data by

$$
\begin{aligned}
X_A(s,t) =&2 \sum_j \sum_k \left\{K_1(s/A,\omega_j,\lambda_k)\cos(\omega_j' s + \lambda_k t) - K_2(s/A,\omega_j,\lambda_k)\sin(\omega_j' s + \lambda_k t)\right\} \\
&\times \varepsilon_j z_k, s \in \mathbf{R}^2, t = 1,2,\ldots,
\end{aligned}
$$

which are simulated spatio-temporal data with the spatially dependent spectral density $|K(u,\omega,\lambda)|^2$.

3 Estimation of parameters

3.1 Whittle likelihood

Suppose we have observed spatio-temporal data $X_A(s_p, t), p = 1, \ldots, N, t = 1, \ldots, T$ that follow locally stationary models in (5) with the spatially dependent spectral density function $f(u, \omega, \lambda) = |K(u, \omega, \lambda)|^2$, which is expressed as $f(\theta_u, \omega, \lambda)$ with spatially dependent parameters θ_u. Our aim is to estimate θ_u for a fixed $u \in [0, 1]^2$ in a nonparametric way that would not specify any parametric form for the dependency of θ on u. In other words, we assume parametric function for the spectral density with parameter θ that may depend u, but do not give any parametric form for the functional form of $\theta(u)$. We assume that all the observation points $\{s_p\} \subset [0, A]$.

Let $B = (B_1, B_2)$ and let $w_{sp}(x)$ and $w_{tmp}(x)$ be tapers defined on $[-1/2, 1/2]^2$ and $[0, 1]$, respectively. Then the local discrete Fourier transform and periodogram for $u \in [0, 1]^2$ is defined by

$$
\begin{aligned}
d_B(u, \omega, \lambda) &= \frac{(2\pi)^{-3/2}|A|}{N\sqrt{T|B|}} \sum_{p=1}^{N} \sum_{t=1}^{T} X_A(s_p, t) \exp(-i\omega' s_p - i\lambda t) w_{sp}\left(\frac{s_p - Au}{B}\right) w_{tmp}\left(\frac{t}{T}\right), \\
I_B(u, \omega, \lambda) &= |d_B(u, \omega, \lambda)|^2.
\end{aligned}
$$

Let $h(x)$ be a probability density function over $[0, 1]^2$. We assume that s_ps are independently and identically distributed over $[0, A]$ with the density $|A|^{-1}h(s/A)$. Under conditions that will be clarified later, we find that $I_B(u, \omega, \lambda)$ is biased unlike discrete time series case, and that

$$
EI_B(u, \omega, \lambda) \to C_u f(u, \omega, \lambda) + \tilde{C}_u k_u(\lambda),
$$

as $A = (A_1, A_2) \to (\infty, \infty)$, where the second term is the bias term and

$$
\begin{aligned}
C_u &= h(u)^2 \int_{[-1/2,1/2]^2} w_{sp}(x)^2 dx \int_{[0,1]} w_{tmp}(x)^2 dx, \\
\tilde{C}_u &= (2\pi)^{-2} N^{-1} |A| h(u) \int_{[-1/2,1/2]^2} w_{sp}(x)^2 dx \int_{[0,1]} w_{tmp}(x)^2 dx, \\
k(u, \lambda) &= \int_{\mathbf{R}^2} f(u, \alpha, \lambda) d\alpha
\end{aligned}
$$

Noticing that $f(u, \omega, \lambda) = f(\theta_u, \omega, \lambda), k(u, \lambda) = k(\theta_u, \lambda)$, we propose to estimate θ_u by minimizing Whittle likelihood function with respect to θ, which is defined by

$$
\begin{aligned}
&l_w(\theta) = \\
&\int_D \int_{-\pi}^{\pi} \left\{ \frac{I_B(u, \omega, \lambda)}{C_u f(\theta, \omega, \lambda) + \tilde{C}_u k(\theta, \lambda)} + \log(C_u f(\theta, \omega, \lambda) + \tilde{C}_u k(\theta, \lambda)) \right\} d\omega d\lambda, \qquad (6)
\end{aligned}
$$

where D is a compact and symmetric region on $\mathbf{R}^2$ such that $-\omega \in D$ whenever $\omega \in D$. Regarding $C_u, \tilde{C}_u$ as nuisance parameters and concentrating out C_u from the function, we have the concentrated likelihood

$$
\begin{aligned}
l_c(\theta) = &\log\left\{ \frac{1}{2\pi|D|} \int_D \int_{-\pi}^{\pi} \frac{I_B(u, \omega, \lambda)}{f(\theta, \omega, \lambda) + ck(\theta, \lambda)} d\omega d\lambda \right\} \\
&+ \frac{1}{2\pi|D|} \int_D \int_{-\pi}^{\pi} \log\{f(\theta, \omega, \lambda) + ck(\theta, \lambda)\} d\omega d\lambda, \qquad (7)
\end{aligned}
$$

where $c = \tilde{C}_u/C_u > 0$, the nuisance parameter. Minimizing $l_c(\theta)$ with respect to θ and c for a fixed u, we estimate θ_u by $\hat{\theta}$, which means that the dependencies of θ_u on u are estimated in the nonparametric way.

Notice that $l_c(\theta)$ cannot identify the scale parameter σ_u^2 when $f(u,\omega,\lambda)$ is given by $\sigma_u^2 f_0(u,\omega,\lambda)$, as it is seen easily that $l_c(\theta)$ does not depend on σ_u^2. Hence Whittle estimation proposed here just provides the estimators only for the parameters included in $f_0(u,\omega,\lambda)$. In addition, l_c in which the periodogram is replaced with the modified one multiplied with any constant would provide exactly the same values of the estimators by the same reason. In Example 2.2, all the parameters except for $\sigma(u)^2$ are identifiable by the likelihood l_c and can be estimated by minimizing it.

Remark 3.1. *In practice, the integration in (7) should be replaced with the Riemannian summation. Let ω_j, λ_j be jth element in the set of Fourier frequency*

$$\{(2\pi p_1/A_1, 2\pi p_2/A_2), 2\pi q/T | p_1, p_2, q = 0, \pm 1, \pm 2, \ldots\}.$$

Namely, the parameter θ is estimated practically by minimizing

$$\tilde{l}_c(\theta) = \log\left\{ M^{-1} \sum_{(\omega_j,\lambda_j)\in D\times[-\pi,\pi]} \frac{I_B(u,\omega_j,\lambda_j)}{f(\theta,\omega_j,\lambda_j) + ck(\theta,\lambda_j)} \right\}$$
$$+ M^{-1} \sum_{(\omega_j,\lambda_j)\in D\times[-\pi,\pi]} \log\left\{ f(\theta,\omega_j,\lambda_j) + ck(\theta,\lambda_j) \right\},$$

with respect to θ, where M is the cardinality of the Fourier frequencies included in $D \times [-\pi,\pi]$.

3.2 Assumptions

One of the advantage of employing locally stationary models is in making it possible to establish the asymptotic properties of the Whittle estimator. The followings are the assumptions required to derive them. The first and second assumptions specify the asymptotic scheme in which the estimator is consistent and asymptotic normal, which is the mixed asymptotics where the sample size and sampling region diverge jointly.

(A1) Suppose $X_A(s,t)$ follows locally stationary processes in (5) with the spatially dependent spectral density function $f(u,\omega,\lambda) = |K(u,\omega,\lambda)|^2$, and is observed on $(s_p,t), p = 1,\ldots,N, t = 1,\ldots,T, s_p \in [0,A]$. $s_p, p = 1,\ldots,N$ are written as

$$s_p = (A_1\varepsilon_{p1}, A_2\varepsilon_{p2})',$$

where $\varepsilon_p = (\varepsilon_{p1}, \varepsilon_{p2})$ is a sequence of independently and identically distributed random vectors with a probability density function $h(x)$ over the compact region $[0,1]^2$.

(A2) We assume $A_j, B_j, j = 1,2$, N and T are the functions of k such that $A_j = A_j(k), B_j = B_j(k) \to \infty$, $N = N_k \to \infty$ and $T = T_k \to \infty$ as $k \to \infty$. $N_k^{-1}|A_k| \to 0$, $B_j(k)/A_j(k) \to 0$, $\sqrt{T_k|B_k|}B_j(k)^{-2} \to 0$, $\sqrt{T_k^{-3}|B_k|} \to 0$ and $\sqrt{T_k|B_k|}B_j(k)/A_j(k) \to 0$ for $j = 1,2$ as $k \to \infty$.

(A3) The spatially dependent spectral density function $f(u,\omega,\lambda)$ is an integrable, bounded and twice partially differentiable function with respect to $\omega \in \mathbf{R}^2$, $\lambda \in [-\pi,\pi]$, and partially differentiable with respect to $u \in [0,1]^2$.

(A4) The tapers $w_{sp}(x), x \in [-.5, .5]^2$ and $w_{tmp}(x), x \in [0, 1]$ are twice partially differentiable functions when they are regarded as functions over $\mathbf{R}^2$ and $\mathbf{R}$, respectively.

(A5) We fit, for a fixed $u \in [0, 1]^2$, the parametric spectral density $f(\theta_u, \omega, \lambda)$, $\theta_u \in \Theta$, a compact subset in $\mathbf{R}^d$. $f(\theta_u, \omega, \lambda)$ is positive on $\Theta \times D \times [-\pi, \pi]$ and twice differentiable with respect to θ_u for $(\omega, \lambda) \in D \times [-\pi, \pi]$. $\theta_1(u) \neq \theta_2(u)$ implies that $f(\theta_1(u), \omega, \lambda) \neq f(\theta_2(u), \omega, \lambda)$ on a subset of $D \times [-\pi, \pi]$ with positive Lebesgue measure. The true parameter denoted by $\theta_0(u)$ lies in the interior of Θ, namely $f(\theta_0(u), \omega, \lambda) = f(u, \omega, \lambda)$.

3.3 Asymptotic results

Consider the asymptotic results under the scheme in (A1) and (A2). Let $\hat{\theta}_k(u)$ be the estimator minimizing $l_c(\theta)$ in (7) for a fixed $u \in [0, 1]^2$ under the asymptotic scheme in (A1) and (A2) for $k = 1, 2, \ldots$.

Theorem 3.2. *Under Assumptions A1-A5,*

(*a*) *For a fixed* $u \in [0, 1]^2$ *such that* $h(u) > 0$, $\hat{\theta}_k(u)$ *converges to* $\theta_0(u)$ *in probability as* $k \to \infty$.

(*b*) *For a fixed* $u \in [0, 1]^2$ *such that* $h(u) > 0$,

$$\sqrt{T_k|B_k|}\left(\hat{\theta}_k(u) - \theta_0(u)\right) \to N\left(0, b_w\left(\Gamma_{0u} - \Phi_{0u}\right)^{-1}\left(2\Gamma_{0u} + \Delta_{0u}\right)\left(\Gamma_{0u} - \Phi_{0u}\right)^{-1}\right),$$

in distribution as $k \to \infty$, *where* $\Gamma_{0u} = \Gamma(\theta_0(u)), \Phi_{0u} = \Phi(\theta_0(u)), \Delta_{0u} = \Delta(\theta_0(u))$ *with*

$$\begin{aligned}
b_w &= \left\{\int\int |w_{sp}(x)|^4 |w_{tmp}(y)|^4 dxdy\right\}\left\{\int\int |w_{sp}(x)|^2 |w_{tmp}(y)|^2 dxdy\right\}^{-2},\\
\Gamma(\theta) &= (2\pi)^{-3}\int_D\int_{-\pi}^{\pi}\left(\frac{\partial \log f(\theta,\omega,\lambda)}{\partial\theta}\right)\left(\frac{\partial \log f(\theta,\omega,\lambda)}{\partial\theta}\right)' d\omega d\lambda,\\
\Phi(\theta) &= (2\pi)^{-3}(2\pi|D|)^{-1}\int_D\int_{-\pi}^{\pi}\left(\frac{\partial \log f(\theta,\omega,\lambda)}{\partial\theta}\right) d\omega d\lambda \int_D\int_{-\pi}^{\pi}\left(\frac{\partial \log f(\theta,\omega,\lambda)}{\partial\theta}\right)' d\omega d\lambda,\\
\Delta(\theta) &= (2\pi)^{-3}\int_D\int_{-\pi}^{\pi}\int_D\int_{-\pi}^{\pi}\left(\frac{\partial \log f(\theta,\omega_1,\lambda_1)}{\partial\theta}\right)\left(\frac{\partial \log f(\theta,\omega_2,\lambda_2)}{\partial\theta}\right)'\\
&\quad \times a_4(\omega_1, -\omega_1, \omega_2) b_4(\lambda_1, -\lambda_1, \lambda_2) d\omega_1 d\lambda_1 d\omega_2 d\lambda_2.
\end{aligned}$$

The asymptotic variance is different from the popular one in discrete time series models (Dunsmuir[4]). Precisely, $\Phi(\theta)$ in the asymptotic variance disappears in the cases of discrete time series models, since the integration of logged spectral density is the constant, i.e., the logged innovation variance (see Theorem 5.8.1 in Brockwell and Davis[1]). It is different also from the one in Matsuda and Yajima[8], which employs the non-concentrated Whittle likelihood in (6). The non-concentrated likelihood estimator does not include $\Phi(\theta)$ in the asymptotic variance. Hessian matrices between (6) and (7) correspond in the cases of discrete time series, while they do not in our cases, which is the reason for the difference.

4 Empirical studies

We apply locally stationary spatio-temporal CARMA models in (4) to US precipitation data, the motivating example for the temporal and nonstationary extensions of CARMA random fields, in

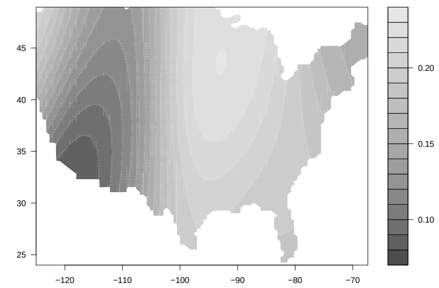

Figure 2: Estimated values for the smoothness parameter $\theta_3(u)$ in the CARMA(2,1) kernel

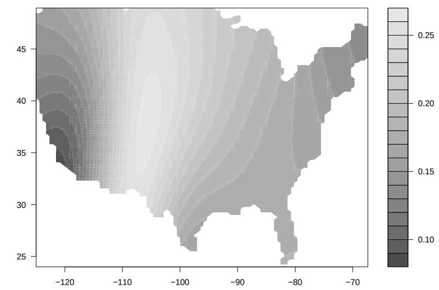

Figure 3: Estimated values for the autoregressive parameter $\phi_1(u)$ in the seasonal AR model

order to check empirical properties of Whittle likelihood estimation and forecasting performances based on the identified model. US precipitation data is the monthly weather station data for the continental US from 1895 through 1997, which is available in the web page of Institute for Mathematics Applied to Geosciences (IMAG):

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

We downloaded monthly total precipitation observed in the weather stations for 48 months period from January, 1994 till December, 1997. They are regarded as a spatio-temporal data, namely monthly observations of spatial data. Precisely, total millimeters of precipitation during the month in the weather station were recorded with the longitude and latitude. See Figure 1 in Section 1 for locations of weather stations, which are seen to be irregularly spaced all over US continent. We transformed the longitudes and latitudes to rectangular coordinates with one unit of 100 kilometers to identify the locations of weather stations.

We fit the locally stationary spatio-temporal CARMA(2,1) model introduced in Example 2.2, where θ_3, the smoothness parameter and ϕ, ψ, the AR parameters, are designed to be dependent spatially. The other two of θ_1 and θ_2 were fixed as 3.80 and 0.56, which were estimated by the usual Whittle estimation, which are obtained by minimizing (7) in which the periodogram was modified with the one for the spatial weight $w_{sp} = 1$. The samples for 36 months from Jan. in 1994 to Dec. in 1996 were used for conducting the estimation minimizing the Whittle likelihood function in (7), where the weight $w_{sp}(x) = \exp(-x^2/8^2), w_{tmp}(x) = 1$ were employed. The estimated spatial dependency of the parameters, which were smoothed figures of the estimators in 27 mesh points over $[0, 1]^2$, are depicted in Figures 2-4.

We find that the estimators caught the spatial dependencies of the three parameters well in

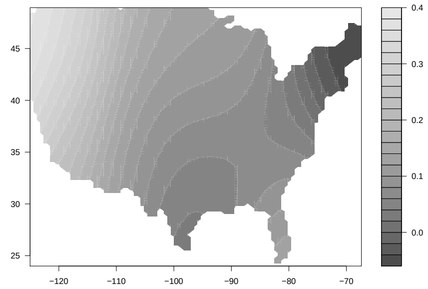

Figure 4: Estimated values for the seasonal autoregressive parameter $\phi_2(u)$ in the seasonal AR model

the nonparametric way. Figure 2 of the smoothness parameter shows that smoothness of covariances decreases over the range in the Rocky mountains in comparison with that in plain fields, which appeals to our intuitive observations of spiky behaviors of precipitations in mountainous areas. Figures 3 and 4 show that seasonal (12 moth lag) correlations are more obvious than one month lag correlations. Figure 4 demonstrates that the seasonal correlations become higher gradually from east to west in US continent. In the east coast area, even negative seasonal correlations are found.

Finally, we conduct one, two and three months ahead forecasts for the samples from Jan. till Dec. in 1997 by the identified spatio-temporal CARMA(2,1) model based on the samples till Dec. 1996. Table 1 shows the MSEs of the forecasts for precipitation in 100 randomly selected stations from the ones in 1997, where the two benchmarks are the three year averages of precipitation on the previous and same months in the past years. For example, the two benchmark forecasts in March, 1997 are three year averages of precipitation in February, 1995-1997 and those of March, 1994-1996.

The forecasts by the identified spatio-temporal CARMA model are constructed as follows. Suppose we construct one step ahead forecast for $X_A(w, t+1)$ at a location w by the samples $X_A(s_j, k), j = 1, \ldots, N, k \leq t$. By (4), we have, for I_t being the information generated by $X_A(s, k), s \in [0, A], k \leq t$,

$$\begin{aligned} E(X_A(w, t+1)|I_t) &= \sum_j g(w/A, w - e_j) E(Z_{j,t+1}(w/A)|I_t) \\ &= \sum_j \phi_1(w/A) g(w/A, w - e_j) Z_{j,t}(w/A) + \sum_j \phi_2(w/A) g(w/A, w - e_j) Z_{j,t-11}(w/A) \\ &= \phi_1(w/A) X_A(w, t) + \phi_2(w/A) X_A(w, t-11). \end{aligned} \tag{8}$$

When $Z_{jt}(w/A)$ is not zero mean, the forecast should be modified with

$$\mu(w/A) + \phi_1(w/A)(X_A(w, t) - \mu(w/A)) + \phi_2(w/A)(X_A(w, t-11) - \mu(w/A)),$$

where $\mu(w/A)$ is the temporal mean of $X_A(w, t)$. $X_A(w, t)$ and $\mu(w/A)$, when there are no observations at w, should be estimated with

$$\hat{X}_A(w, t) = \sum_j c_j(w) X_A(s_j, t) \text{ and } \hat{\mu}(w/A) = \frac{1}{36} \sum_{k=1}^{36} \hat{X}_A(w, t+1-k),$$

	MSE				
month	bmrk1	bmrk2	1step	2step	3step
Jan.	25.2	21.9	24.8	27.2	27.1
Feb.	86.1	46.4	37.4	34.9	33.0
Mar.	22.9	30.6	19.2	20.5	19.7
Apr.	42.8	38.3	36.6	33.9	35.0
May	15.1	16.3	16.9	16.9	17.1
Jun.	26.8	26.3	28.9	30.8	31.3
Jul.	23.8	22.9	21.5	20.4	20.3
Aug.	15.2	17.0	19.3	20.1	20.1
Sep.	28.8	23.9	22.5	22.6	22.9
Oct.	22.5	18.4	15.7	15.7	15.8
Nov.	19.5	21.1	18.7	18.1	17.9
Dec.	21.6	43.7	26.4	28.3	28.4
average	29.2	27.2	24.0	24.1	24.0

Table 1: Comparisons of MSEs among the forecasts for US monthly precipitation in 1997 by CARMA(2,1) and two benchmarks identified by the samples from Jan. 1994 till Dec. 1996, where 100 stations were randomly chosen for the precipitation to be predicted. The three year averages of precipitation on the previous and same months in the past three years (1994-1996) are adopted as the benchmark1 and 2, respectively. By the identified CARMA, we conduct the 1, 2 and 3 step forecasts. For example, the two benchmark forecasts in March, 1997 are three year averages of precipitation in February, 1995-1997 and those of March, 1994-1996.

respectively, where $c_j(w) = \{\sum_k g(w/A, w-s_k)\}^{-1} g(w/A, w-s_j)$, the identified CARMA kernel normalized to let the total summation to be 1. It is found by (4) that $\hat{X}_A(w,t)$ is the estimator obtained by replacing the knots $\{e_j\}$ with $\{s_j\}$ and $Z_{jt}(w/A)$ with $X_A(s_j,t)$, which is an computationally feasible approximation for the exact kriging $E(X_A(w,t)|X_A(s_j,t), j = 1,2,\ldots)$ that can work for huge data set. Multi-step (h-step, say) ahead forecasts for $X_A(w,t+h)$ are constructed recursively for $h = 2,3,\ldots$ by replacing the unobserved values of $h-1$ step ahead in (8) with the predicted values.

First we found from Table 1 that MSEs of multi-step ahead forecasts by the CARMA(2,1) model are not necessarily larger than those of one step ahead forecasts. This may be because of the weaker lag one temporal correlation than seasonal correlation demonstrated by the estimation results shown in Figures 3 and 4. Comparing MSEs of the two benchmarks with those of CARMA (2,1) forecasts, we observe that the spatio-temporal CARMA (2,1) models can catch the temporal correlations jointly with nonstationary spatial correlations, which results in the improvement of forecasts over the benchmarks.

5 discussion

This paper has proposed locally stationary spatio-temporal processes to describe the empirical properties of US precipitation data, the huge set of spatio-tempotal data. Extending stationary CARMA random fields on $\mathbf{R}^2$ to spatio-temporal models with spatially nonstationary and temporarily stationary covariances, we have locally stationary spatio-temporal CARMA processes, which are moreover generalized to locally stationary spatio-temporal processes. Following Dahlhaus[3], we estimate the spatially dependent parameter by minimizing Whittle likelihood and derive the asymptotic properties rigorously. Applications them to US precipitation data demonstrate that the nonstationary features of temporal and spatial covariances are accounted well by the locally stationary CARMA(2,1) model.

The critical restriction of spatio-temporal CARMA processes is that the covariances are confined to separable ones given by the products of spatial and temporal covariances. Nonseparable extensions that can express fruitful class of covariance structures are our next target. One more interesting extension is to allow Lévy sheets that drive CARMA random fields to have infinite variances, which makes it possible to express several varieties of spiky behaviors in spatial data. New parameter estimation method over Whittle estimation, which may not work for the infinite variance cases, is required Its asymptotic properties are important issues that attract empirical as well as mathematical interests.

6 Sketch of the proof

This section shows the outline of the proof for Theorem 1. Complete proof is available on request from the first author. *Proof of Theorem 1 (a)*

Let $\theta_1(u) \neq \theta_0(u)$ for a fixed $u \in [0,1]^2$ such that $h(u) > 0$. By Lemmas 3 and 6 in Matsuda and Yajima[8], we have

$$\begin{aligned} &l_c(\theta_1(u)) \to \\ &\quad \log\left\{\frac{1}{2\pi|D|}\int_D\int_{-\pi}^{\pi}\frac{f(\theta_0(u),\omega,\lambda)}{f(\theta_1(u),\omega,\lambda)}d\omega d\lambda\right\} + \frac{1}{2\pi|D|}\int_D\int_{-\pi}^{\pi}\log\left\{f(\theta_1(u),\omega,\lambda)\right\}d\omega d\lambda \\ &\quad + Const(u) := l_\infty(\theta_1(u)), \end{aligned}$$

in probability as $k \to \infty$. By the identifiability condition in (A5) and Jensen's inequality, we have

$$\begin{aligned} &l_\infty(\theta_1(u)) - l_\infty(\theta_0(u)) = \\ &\quad \log\left\{\frac{1}{2\pi|D|}\int_D\int_{-\pi}^{\pi}\frac{f(\theta_0(u),\omega,\lambda)}{f(\theta_1(u),\omega,\lambda)}d\omega d\lambda\right\} - \frac{1}{2\pi|D|}\int_D\int_{-\pi}^{\pi}\log\left\{\frac{f(\theta_0(u),\omega,\lambda)}{f(\theta_1(u),\omega,\lambda)}\right\}d\omega d\lambda \\ &\quad > 0. \end{aligned}$$

It follows that, for any positive constant $K(\theta_0(u),\theta_1(u))$ that is less than $l_\infty(\theta_1(u)) - l_\infty(\theta_0(u))$,

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

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

$$\delta\left(C_1\int_D\int_{-\pi}^{\pi} I_B(u,\omega,\lambda)d\omega d\lambda\right)^{-1}\left\{C_2\int_D\int_{-\pi}^{\pi} I_B(u,\omega,\lambda)d\omega d\lambda + C_3\right\}$$

such that, for any $\theta_1(u)$ and $\theta_2(u)$ that satisfy $||\theta_1(u) - \theta_2(u)|| < \delta$,

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

because, letting $a(\theta,\lambda) = bk(\theta,\lambda)$ and θ^* be the mean value between $(\theta_1(u)$ and $\theta_2(u)$, we have

$$\begin{aligned} |l_c(\theta_2(u)) - l_c(\theta_1(u))| &\leq \left(\int_D\int_{-\pi}^{\pi}\frac{I_B(u,\omega,\lambda)}{f(\theta^*,\omega,\lambda) + a(\theta^*,\lambda)}d\omega d\lambda\right)^{-1} \\ &\times \int_D\int_{-\pi}^{\pi} I_B(u,\omega,\lambda)\left|\frac{1}{f(\theta_2(u),\omega,\lambda) + a(\theta_2(u),\lambda)} - \frac{1}{f(\theta_1(u),\omega,\lambda) + a(\theta_1(u),\lambda)}\right| d\omega d\lambda \\ &+ \left|\log\{f(\theta_2(u),\omega,\lambda) + a(\theta_2(u),\lambda)\} - \log\{f(\theta_1(u),\omega,\lambda) + a(\theta_1(u),\lambda)\}\right|. \end{aligned}$$

And it is easily seen from the form of $H_{k,\delta}$ that there exisrs a $\delta > 0$ such that

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

Applying lemma 2 of Walker[13], we have the consistency.

Proof of Theorem 1 (b)

By Taylor series expansion,

$$0 = \sqrt{T_k|B_k|}\frac{\partial l_c(\theta_0(u))}{\partial \theta} + \frac{\partial^2 l_c(\theta^*)}{\partial\theta\partial\theta'}\sqrt{T_k|B_k|}\left(\hat{\theta}_u - \theta_0(u)\right),$$

where θ^* is the mean value between $\theta_0(u)$ and $\hat{\theta}_u$. Hence

$$\sqrt{T_k|B_k|}\left(\hat{\theta}_u - \theta_0(u)\right) = \left(-\frac{\partial^2 l_c(\theta^*)}{\partial\theta\partial\theta'}\right)^{-1} \times \left\{\sqrt{T_k|B_k|}\frac{\partial l_c(\theta_0(u))}{\partial\theta}\right\}.$$

Then by Lemma 7 in Matsuda and Yajima[8],

$$\begin{aligned}
&\sqrt{T_k|B_k|}\frac{\partial l_c(\theta_0(u))}{\partial\theta} \\
&= \sqrt{T_k|B_k|}(2\pi|D|)^{-1}\int_D\int_{-\pi}^{\pi} C_u^{-1}\left\{I_B(u,\omega,\lambda) - EI_B(u,\omega,\lambda)\right\}\frac{\partial f^{-1}(\theta_0(u),\omega,\lambda)}{\partial\theta} + o_p(1) \\
&\to N\left\{0, (2\pi)^6(2\pi|D|)^{-2}b_w\left(2\Gamma_{0u} + \Delta_{0u}\right)\right\}.
\end{aligned}$$

By the consistency of $\hat{\theta}_u$,

$$\begin{aligned}
-\frac{\partial^2 l_c(\theta^*)}{\partial\theta\partial\theta'} &= -(2\pi|D|)^{-1}\int_D\int_{-\pi}^{\pi}\left\{C_u^{-1}I_B(u,\omega,\lambda)\frac{\partial^2}{\partial\theta\partial\theta'}f^{-1}(\theta^*,\omega,\lambda)\right\}d\omega d\lambda \\
&+ (2\pi|D|)^{-2}\int_D\int_{-\pi}^{\pi}\left\{C_u^{-1}I_B(u,\omega,\lambda)\frac{\partial}{\partial\theta}f^{-1}(\theta^*,\omega,\lambda)\right\}d\omega d\lambda\int_D\int_{-\pi}^{\pi}\left\{C_u^{-1}I_B(u,\omega,\lambda)\frac{\partial}{\partial\theta}f^{-1}(\theta^*,\omega,\lambda)\right\}'d\omega d\lambda \\
&+ (2\pi|D|)^{-1}\int_D\int_{-\pi}^{\pi}\left\{f(\theta^*,\omega,\lambda)\frac{\partial^2}{\partial\theta\partial\theta'}f^{-1}(\theta^*,\omega,\lambda) + \frac{\partial}{\partial\theta}f(\theta^*,\omega,\lambda)\frac{\partial}{\partial\theta'}f^{-1}(\theta^*,\omega,\lambda)\right\}d\omega d\lambda + o_p(1) \\
&\to -(2\pi)^3(2\pi|D|)^{-1}\left(\Gamma_{0u} - (2\pi|D|)^{-1}\Phi_{0u}\right).
\end{aligned}$$

References

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

[2] Brockwell, P. J. and Matsuda, Y., Continuous auto-regressive moving average random fields on R^n. *J. Roy. Statist. Soc. Ser. B.* 79(3), 833-857 (2017)

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

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

[5] Fuentes, M. A new high frequency kriging approach for nonstationary environmental processes. *Environmetrics* 12, 469-483.

[6] Guttorp, P. and Sampson, P. D., Methods for estimating heterogeneous spatial covariance functions with environmental applications. *In: Handbook of Statistics*, 12, G.P. Patil and C.R.Rao, Eds., Elsevier Science, New York, 661-689 (1994)

[7] Higdon, A process-convolution approach to modelling temperatures in the North Atlantic ocean, *Environ. Ecol. Stat.*, 5, 173-190 (1998)

[8] Matsuda, Y. and Yajima, Y., Fourier analysis of irregularly spaced data on R^d. *J. Roy. Statist. Soc. B.* 71(1), 191-217 (2009).

[9] Nychka, D., Wikle, C. K. and Royle, J. A. Multiresolution models for nonstationary spatial covariance functions, *Stat. Modeling*, 2, 315-332, (2002).

[10] Priestley, M. B. Time-dependent spectral analysis and its application in prediction and control. *J. Sound and Vibration*, 17, 517-534, (1971)

[11] Sampson, P. D., Constructions for Nonstationary Spatial Processes. *In Handbook of Spatial Statistics*, Gelfand, Diggle, Fuentes and Guttorp, Eds., CRC press, Boca Ration, 119-130 (2010)

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

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