Statistical analysis of a spatio-temporal model with location dependent parameters and a test for spatial stationarity

Suhasini Subba Rao
Department of Statistics,
Texas A&M University,
College Station, TX 77843-3143, USA
Email: `suhasini.subbarao@stat.tamu.edu`

23rd September, 2007

Abstract

In this paper we define a spatio-temporal model with location dependent parameters to describe temporal variation and spatial nonstationarity. We consider the prediction of observations at unknown locations using known neighbouring observations. Further we propose a local least squares based method to estimate the parameters at unobserved locations. The sampling properties of these estimators are investigated. We also develop a statistical test for spatial stationarity. In order to derive the asymptotic results we show that the spatially nonstationary process can be locally approximated by a spatially stationary process. We illustrate the methods of estimation with some simulations.

key words: Autoregressive process, ground ozone data, kriging, local least squares, local stationarity, polynomial interpolation, spatio-temporal models, testing for spatial stationarity.

1 Introduction

The modelling of spatial data has been an active area of research, because of its vast potential in applications to ecology, the environmental sciences and finance amongst others. If at a given time, we have observations over various locations (equally or unequally spaced) we can find a suitable spatial model or covariance function to describe the dependence over

space (cf. Whittle (1954), Mardia and Marshall (1984) Cressie (1993), Johns et al. (2003), Hallin et al. (2004), Guan et al. (2004) and Lu et al. (2004)). In the situation where time is not fixed, then we have observations over both space and time, and there are various ways to model this type of data. For instance, often it is assumed that the observations are Gaussian, therefore to model the dependence a covariance is often fitted. Generally it is supposed that the process is spatially stationary and the covariance has a particular structure (usually isotropic or anisotropic). In this case likelihood methods are often used to estimate the parameters (c.f. Cressie and Huang (1999), Shitan and Brockwell (1995), Matsuda and Yajima (2004), Zhang (2004) and Jun and Stein (2007)). However many factors could cause the process to be spatially nonstationary, therefore it would be of interest to develop methods of estimation and theory for such processes.

We observe if we fix a location, the observations at that location can be considered as a time series. This inspires us to define a spatio-temporal process in terms of its time dynamics. Let us suppose that for every fixed location, the resulting time series has an AR representation, where the innovations are samples from a spatial process. By assuming the innovations are observations on a spatial process, dependence between two observations in space can be modelled. More precisely, we define the location dependent autoregressive (AR) process $\{X_t(u) : u \in [0,1]^2\}_t$, where $X_t(u)$ satisfies the representation

$$X_t(u) = \sum_{j=1}^{p} a_j(u) X_{t-j}(u) + \sigma(u) \xi_t(u) \qquad t = 1, \ldots, T, \tag{1}$$

with $u = (x, y) \in [0,1]^2$, and $\{a_j(\cdot); j = 1, \ldots, p\}$ and $\sigma(\cdot)$ are nonparametric functions. We suppose the innovations $\{\xi_t(u) : u \in [0,1]^2\}$ are independent over time and are spatially stationary processes, with $\mathbb{E}[\xi_t(u)] = 0$ and $\text{var}[\xi_t(u)] = 1$. We observe if the $\{a_j(\cdot)\}$ are not constant over space, then $\{X_t(u)\}$ is a spatially nonstationary process. We mention that the location dependent AR process is used to fit ozone and house price data in Gilleland and Nychka (2005) and Gelfand et al. (2003) respectively. An integrated spatially stationary AR process is considered Storvik et al. (2002). We note that the results in this paper do not rely on any distributional assumptions on $\xi_t(u)$.

In Section 2 we consider the prediction of observations at unknown locations, using known neighbouring observations. The predictor requires an estimate of $a_j(\cdot)$ at the unobserved location. In Section 2 we propose two methods for estimating the AR functions $\{a_j(\cdot)\}$. Both methods are based on a localised least squares criterion. The first estimator is a localised least squares estimator with constant regressors, whereas the second estimator is a local linear least squares estimator. In Section 3 we consider the sampling properties of both estimators. We consider the two cases where (i) the number of locations are kept fixed and time $T \to \infty$ and (ii) both the number of locations and $T \to \infty$. In the case that the number of locations is fixed, we show that both estimators are asymptotically normal but biased (in probability). However, if the parameters are sufficiently smooth, the linear interpolating

least squares estimator yields a bias which is smaller than the constant interpolating least squares estimator. In the case that the number of locations (m) also grow, the estimators are asymptotically consistent.

In Section 4 we develop a test for spatial stationarity, which is based on testing for homogeneity. We evaluate the limiting distribution of the test statistic under the null and alternative hypotheses of spatial stationarity and nonstationarity. We note that the 'roughness' of the parameters $\{a_j(\cdot)\}$ determine the power of the test.

To illustrate the methods and test for spatial stationarity, in Section 5 we consider some simulations. If $a_j(\cdot)$ is smooth we show that the local linear estimator is better than the local least squares estimators. However when the parameter $a_j(\cdot)$ is relatively rough (its first derivative does not exists everywhere), then the two estimation methods are comparable.

An outline of the proofs can be found in the Appendix. The full details and some additional results can be found in the accompanying technical report Subba Rao (2007).

2 Estimation of location parameters at an unobserved location

2.1 The model and assumptions

Throughout this paper we let $u = (x, y)$ and for $s = 0, \ldots, m$, we suppose $u_s = (x_s, y_s)$. We note that (x, y) can denote the spatial coordinates, but it is straightforward to generalise the results in this paper to higher dimensions. Let $\{\xi_t(u) : u \in [0, 1]^2\}$ be a spatially stationary process and $c_\xi(u) = \text{cov}\{\xi_t(0), \xi_t(u)\}$. Let $\|\cdot\|_\infty$ denote the sup-norm of a vector, $\|\cdot\|_2$ denote the Euclidean norm and $\|\cdot\|_1$ the ℓ_1-norm. Suppose A is a $p \times p$ matrix, then $\|A\|_{spec}$ denotes the spectral norm of A, A_{ij} denotes the (i, j)th element of A, $A_{\cdot, j}$, the jth column of A, and b_j the jth element of the vector $\underline{b}$. Let A' denote the transpose of matrix A.

We make the following assumptions

Assumption 2.1 *Suppose the process $\{X_t(u)\}$ satisfies (1). Furthermore,*

(i) *The innovations $\xi_t(u)$ are independent over time, and spatially, strictly stationary.*

(ii) *Let $\{\lambda_j(u) : j = 1, \ldots, p\}$ be the roots of the characteristic polynomial $x^p - \sum_{j=1}^p a_j(u) x^{p-j}$. For some $\delta > 0$, we assume that $\sup_{u, j} |\lambda_j(u)| \leq (1 - \delta)$.*

(iii) *The second order partial derivatives of $a_j(u)$ and $\sigma(\cdot)$ exist and all the partial derivatives up to the second order are uniformly bounded over $[0, 1]^2$.*

(iv) *For some $C > 0$, $\inf_u \sigma(u) > C$.*

(v) *For some $\eta > 0$, $\mathbb{E}(|\xi_t(u)|^{4+\eta}) < \infty$.*

Assumption 2.1(i,ii) means that the process $\{X_t(u)\}$ is stationary over time. We use Assumption 2.1(ii,iii) to show that in a local neighbourhood of u, $X_t(u)$ can be approximated by a spatially stationary process. We use Assumption 2.1(iv,v) to prove asymptotic normality of the estimators defined below. It is worth mentioning that most of the results in this paper do not require any additional assumptions on the distribution of $\{X_t(u)\}$.

In order to study the least squares estimators, below, we use the notation $\dot{f}(u) = (\frac{\partial f(u)}{\partial x}, \frac{\partial f(u)}{\partial y})'$.

We will make frequent use of the well known result that since for fixed u, $\{X_t(u)\}_t$ is an AR(p) process it has the unique solution

$$X_t(u) = \sum_{k=0}^{\infty} \zeta_k(u)\xi_{t-k}(u), \tag{2}$$

where $\zeta_k(u) = \sigma(u)[A(u)^k]_{1,1}$, with

$$A(u) = \begin{pmatrix} a_1(u) & a_2(u) & \dots & a_{p-1}(u) & a_p(u) \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \dots & 1 & 0 \end{pmatrix} \tag{3}$$

and satisfies $\sup_u |\zeta_k(u)| \leq K\rho^k$ for some finite K and $1 - \delta < \rho < 1$.

Remark 2.1 (Spatio-temporal Covariance) *In spatial statistics often the dependence of a process is characterised through its covariance. In the case of the location dependent AR process, the covariance across both space and time is*

$$cov[X_{t_1}(u_1), X_{t_2}(u_2)] = c_\xi(u_1 - u_2) \sum_{k=0}^{\infty} \zeta_k(u_1)\zeta_{k+t_2-t_1}(u_2). \tag{4}$$

A spatio-temporal process $\{Y_t(u)\}$ is said to be space-time separable, if there exists functions $d : \mathbb{R}^2 \to \mathbb{R}$ and $e : \mathbb{R} \to \mathbb{R}$ such that for all u_1, u_2, t_1 and t_2 we have $cov(Y_{t_1}(u_1), Y_{t_2}(u_2)) = d(u_1, u_2)e(t_1 - t_2)$. It is easy to show that the location dependent AR process $\{X_t(u)\}$ is only space-time separable if $a_j(\cdot)$ and $\sigma(\cdot)$ are constant over $[0, 1]^2$, for all $1 \leq j \leq p$. In this case $\{X_t(u)\}$ is a spatially stationary process.

2.2 Kriging for nonstationary processes

In this section we consider kriging for spatially nonstationary models which satisfy (1). More precisely, we consider prediction of the spatio-temporal process $X_t(u_0)$ at any arbitrary unobserved location u_0 given the observations $\{X_1(u_s), \dots, X_T(u_s) : s = 1, \dots, m\}$. The predictor

depends on the parameters $\{a_j(u_0)\}$. However, the parameters $\{a_j(u_0)\}$ are unknown, and cannot be estimated using standard methods. Therefore in Section 2.3 we consider methods for estimating $\{a_j(\cdot)\}$. To simplify notation we assume Gaussianity of the innovations. If the innovations are not Gaussian the discussion below applies to the best linear predictor.

Let $\mathbf{X} = [\{X_j(u_s) : s = 1, \ldots, m\}_{j=1}^T]$. To evaluate $\mathbb{E}[X_t(u_0)|\mathbf{X}]$, for $p+1 \le t \le T$, we use (2) and note that $\underline{X}_t(u_0) = A(u_0)\underline{X}_{t-1}(u_0) + \underline{\xi}_t(u_0)$, where $\underline{X}_t(u_0)' = (X_t(u_0), \ldots, X_{t-p+1}(u_0))$, $\underline{\xi}_t(u_0)' = (\xi_t(u_0), \ldots, 0)$ are p-dimensional random vectors and $A(\cdot)$ is defined in (3), to obtain

$$\mathbb{E}[X_t(u_0)|\mathbf{X}] = \sum_{k=0}^{t-p-1} \zeta_k(u_0)\mathbb{E}[\xi_{t-k}(u_0)|\mathbf{X}] + [A(u_0)^{t-p}\mathbb{E}\{\underline{X}_p(u_0)|\mathbf{X}\}]_1.$$

For $p + 1 \le t - k \le T$ we have $\mathbb{E}[\xi_{t-k}(u_0)|\mathbf{X}] = \mathbb{E}[\xi_{t-k}(u_0)|\{\xi_{t-k}(u_s) : s = 1, \ldots, m\}]$, substituting this into the above gives

$$\mathbb{E}[X_t(u_0)|\mathbf{X}] = \sum_{k=0}^{t-p-1} \zeta_k(u_0)\mathbb{E}[\xi_{t-k}(u_0)|\{\xi_{t-k}(u_s) : s = 1, \ldots, m\}] + [A(u_0)^{t-p}\mathbb{E}\{\underline{X}_p(u_0)|\mathbf{X}\}]_1.$$

Under Assumption 2.1(ii), we have the bound $\|A(u_0)^k\|_{spec} \le K\rho^k$, where $1 - \delta \le \rho < 1$, which means

$$\mathbb{E}[X_t(u_0)|\mathbf{X}] = \sum_{k=0}^{t-p-1} \zeta_k(u_0)\mathbb{E}[\xi_{t-k}(u_0)|\{\xi_{t-k}(u_s) : s = 1, \ldots, m\}] + O_p(\rho^{t-p}). \qquad (5)$$

We now show that the first term on the right hand side of (5) can be evaluated recursively. Let us suppose the random process $\{X_t(u_0|m)\}$ satisfies

$$X_t(u_0|m) = \sum_{j=1}^p a_j(u_0)X_{t-j}(u_0|m) + \sigma(u_0)\mathbb{E}[\xi_t(u_0)|\xi_t(u_1), \ldots, \xi_t(u_m)], \qquad (6)$$

and set $X_t(u_0|m) = 0$ for $t \le p$. Define the p-dimensional vectors

$$\begin{aligned} \underline{X}_t(u_0|m)' &= (X_t(u_0|m), \ldots, X_{t-p+1}(u_0|m)) \\ \underline{\xi}_t(u_0|m)' &= (\sigma(u_0)\mathbb{E}[\xi_t(u_0)|\xi_t(u_1), \ldots, \xi_t(u_m)], 0, \ldots, 0). \end{aligned}$$

We observe that (6) can be written as $\underline{X}_t(u_0|m) = A(u_0)\underline{X}_{t-1}(u_0|m) + \underline{\xi}_t(u_0|m)$. Now, this together with the initial condition $\underline{X}_p(u_0|m)' = (0, \ldots, 0)$, imply that $X_t(u_0|m)$ has the solution $\sum_{k=0}^{t-p-1} \zeta_k(u_0)\mathbb{E}[\xi_{t-k}(u_0)|\{\xi_{t-k}(u_s) : s = 1, \ldots, m\}]$. Comparing $X_t(u_0|m)$ with (5), we see that for large t, $X_t(u_0|m)$ and $\mathbb{E}[X_t(u_0)|\mathbf{X}]$ are asymptotically equivalent. Therefore one can use $X_t(u_0|m)$ as an estimator of $\mathbb{E}[X_t(u_0)|\mathbf{X}]$.

5

In order to evaluate $X_t(u_0|m)$ we need to know $\mathbb{E}\big[\xi_t(u_0)\big|\xi_t(u_1),\ldots,\xi_t(u_m)\big]$, $\{a_j(\cdot)\}$ and $\sigma(\cdot)$. If the joint distribution of $\{\xi_t(u_s) : s = 0,\ldots,m\}$ were known, then we can evaluate the conditional expectation $\mathbb{E}\big[\xi_t(u_0)\big|\{\xi_t(u_s)\}_{s=1}^m\big]$. Furthermore, if $\{\xi_t(u); u \in [0,1]^2\}$ were a spatial Gaussian process, then it is straightforward to show that

$$\mathbb{E}\big[\xi_t(u_0)\big|\{\xi_t(u_s)\}_{s=1}^m\big] = \sum_{s=1}^m \gamma_j(\underline{u})\xi_t(u_s), \tag{7}$$

where $\underline{u} = (u_0,\ldots,u_m)$, $\gamma(\underline{u}) = (\gamma_1(\underline{u}),\ldots,\gamma_m(\underline{u}))' = R(\underline{u})^{-1}r(\underline{u})$, with $R(\underline{u})_{ij} = \text{cov}(\xi_t(u_i),\xi_t(u_j)) = c_\xi(u_i - u_j)$ and $r(\underline{u})_i = \text{cov}(\xi_t(u_0),\xi_t(u_i)) = c_\xi(u_i - u_0)$. This means by using estimates of the innovations $\{\xi_t(u_s)\}_t$ at the observed locations, we can estimate the covariance function $c_\xi(\cdot)$. For example, if $\{\xi_t(\cdot)\}$ were Gaussian and the spatial covariance belonged to the Matérn family, then likelihood methods could be used to estimate the parameters of the Matérn covariance (cf. Cressie (1993) and the references therein).

Nevertheless, it is still not possible to evaluate $X_t(u_0|m)$, since the parameters $\{a_j(u_0)\}$ are unknown. In the following section we use a least squares approach for estimating the parameters $\{a_j(\cdot)\}$ from the observations $\{X_1(u_s),\ldots,X_T(u_s) : s = 1,\ldots,m\}$, and study their asymptotic properties.

2.3   Estimation

Ordinary least squares is often used to estimate the parameters of a stationary autoregressive process. However, if $\{X_t(u_0)\}_t$ is unobserved such methods cannot be used to estimate $\{a_j(u_0)\}$. Nevertheless, since the autoregressive parameters are continuous over space (see Assumption 2.1(iii)) nonparametric methods can be developed, to estimate the parameters at unobserved locations. In nonparametric regression, often the local likelihood is used to obtain an estimator of a nonparametric function (cf. Tibshirani and Hastie (1987)). Motivated by these methods, to estimate the autoregressive parameters at u_0, we use a localised least squares approach. More precisely, we consider the localised least squares criterion

$$\mathcal{L}_{1,T}(u_0, \underline{\alpha}) = \frac{1}{mT'} \sum_{t=p+1}^T \sum_{s=1}^m W_b(u_0 - u_s)\Big\{X_t(u_s) - \sum_{j=1}^p \alpha_j X_{t-j}(u_s)\Big\}^2, \tag{8}$$

where $T' = T - p - 1$ and $K : \mathbb{R}^2 \to \mathbb{R}$ is a kernel function with $\int K(x)dx = 1$, $W(u) = K(x)K(y)$ and $W_b(\cdot) = \frac{1}{b}W(\frac{\cdot}{b})$. Let $\underline{a}(u_0)' = (a_1(u_0),\ldots,a_p(u_0))$. We use $\underline{\hat{a}}_{1,T}(u_0)$ as an estimator of $\underline{a}(u_0)$, where

$$\underline{\hat{a}}_{1,T}(u_0) = \arg\min_\alpha \mathcal{L}_{1,T}(u_0, \alpha). \tag{9}$$

We call $\underline{\hat{a}}_{1,T}(u_0) = (\hat{a}_{1,1}(u_0),\ldots,\hat{a}_{1,p}(u_0))$ the *local least squares estimator*.

The local least squares criterion (8) motivates a generalisation, which is similar to the local polynomials methods often used in nonparametric regression. In the context of nonparametric regression, Fan (1993) showed that by replacing the constant term in the local least squares criterion with a linear polynomial, thereby estimating both the parameter and its derivative, a reduction in bias can be achieved. Under Assumption 2.1(iii) we can make a Taylor expansion of u_s about u_0 $(a_j(u_s) = a_j(u_0) + (u_0 - u_s)\dot{a}_j(u_0) + O(\|u_s - u_0\|^2_\infty))$, to yield the least squares criterion

$$\mathcal{L}_{2,T}(u_0, \mathcal{A}) = \frac{1}{mT'} \sum_{t=p+1}^{T} \sum_{s=1}^{m} W_b(u_0 - u_s) \left\{ X_t(u_s) - \sum_{j=1}^{p} [\alpha_j + (u_0 - u_s)\beta_j] X_{t-j}(u_s) \right\}^2, \quad (10)$$

where $\beta'_j = (\beta_{1j}, \beta_{2j})$ and $\mathcal{A}$ is a $3 \times p$-dimensional vector. Let $\mathcal{A}(u_0)' = \big(\underline{a}(u_0), \dot{a}_1(u_0), \ldots, \dot{a}_p(u_0)\big)$ and use $\hat{\mathcal{A}}_{2,T}(u_0)$ as an estimator of $\mathcal{A}(u_0)$, where

$$\hat{\mathcal{A}}_{2,T}(u_0) = \arg\min_{\mathcal{A}} \mathcal{L}_{2,T}(u_0, \mathcal{A}) \quad (11)$$

and $\hat{\mathcal{A}}_{2,T}(u_0) = (\underline{\hat{a}}_{2,T}(u_0), \hat{\beta}_{2,1}(u_0), \ldots \hat{\beta}_{2,p}(u_0))$. In particular $\underline{\hat{a}}_{2,T}(u_0)$ is an estimator of $\underline{a}(u_0)$. We call $\hat{\mathcal{A}}_{2,T}(u_0)$ the *local linear least squares estimator*. It is clear that $\hat{\mathcal{A}}_{2,T}(u_0)$ can easily be evaluated using $\hat{\mathcal{A}}_{2,T}(u_0) = \hat{R}_T(\underline{u})^{-1}\hat{\underline{r}}_T(\underline{u})$, where $\underline{u} = (u_0, \ldots, u_s)$,

$$\hat{R}_T(\underline{u}) = \frac{1}{mT'} \sum_{s=1}^{m} W_b(u_0 - u_s)\, \mathbf{u}(s)\mathbf{u}(s)' \otimes \sum_{t=p+1}^{T} \left\{ \underline{X}_{t-1}(u_s)\underline{X}_{t-1}(u_s)' \right\} \quad (12)$$

$$\hat{\underline{r}}_T(\underline{u}) = \frac{1}{mT'} \sum_{s=1}^{m} W_b(u_0 - u_s)\, \mathbf{u}(s) \otimes \sum_{t=p+1}^{T}{}' \underline{X}_{t-1}(u_s),$$

$\underline{X}_{t-1}(u_s)' = (X_{t-1}(u_s), \ldots, X_{t-p}(u)), \otimes$ denotes the Kronecker product and $\mathbf{u}(s) = (1, (u_s - u_0))'$.

Studying (12), we see if the observed locations are clustered together, the matrix $\hat{R}_T(\underline{u})$ can be close to singular. This problem can be overcome by using a commonly used technique in linear regression, where a small diagonal matrix is added to the Hessian matrix to make the estimator more stable. Adding a diagonal matrix to the Hessian, ensures invertibility and means that the smallest eigenvalues of the 'regularised' Hessian is bounded below by the smallest eigenvalue of the diagonal matrix. We use the same trick here, and add a 'small' diagonal matrix to $\hat{R}_T(\underline{u})$. This makes calculating the estimator simpler and guarantees that the resulting estimator is consistent. We define the estimator $\hat{\mathcal{A}}_{3,T}(u_0)$

$$\hat{\mathcal{A}}_{3,T}(u_0) = [\hat{R}_T(\underline{u}) + \eta^\gamma I]^{-1}\, \hat{\underline{r}}_T(\underline{u}), \quad (13)$$

where $\eta = \frac{1}{m} \sum_{s=1}^{m} W_b(u_0 - u_s)\|u_0 - u_s\|^2_\infty$, I is the identity matrix and $\nu > 0$. Let $\hat{\mathcal{A}}_{3,T}(u_0) = (\underline{\hat{a}}_{3,T}(u_0), \hat{\beta}_{3,1}(u_0), \ldots, \hat{\beta}_{3,p}(u_0))$, then we use $\underline{\hat{a}}_{3,T}(u_0)$ as an estimator of $\underline{a}(u_0)$.

To estimate the variance $\sigma(\cdot)^2$, we assume $c_\xi(0) = 1$ and use $\hat{\sigma}(u_0)^2$ as an estimator of $\sigma(u_0)^2$, where

$$\sigma(u_0)^2 = \min_{\underline{\alpha}} \frac{1}{mT'} \sum_{t=p+1}^{T} \sum_{s=1}^{m} W_b(u_0 - u_s)\big\{X_t(u_s) - \sum_{j=1}^{p} \alpha_j X_{t-j}(u_s)\big\}^2.$$

3 Sampling properties of the estimators

In this section we consider the sampling properties of the estimators $\hat{\underline{a}}_{1,T}(\cdot)$, $\hat{\underline{a}}_{2,T}(\cdot)$ and $\hat{\underline{a}}_{3,T}(\cdot)$, defined in the section above. We consider the two situations where (i) m is kept fixed and $T \to \infty$ and (ii) both $m \to \infty$ and $T \to \infty$.

We show in the theorem below, in the case that m is kept fixed, that $\hat{\underline{a}}_{1,T}(u_0)$ is biased, even when $T \to \infty$. The bias can be quantified by defining the following process $\{X_{t,u_0}(u)\}_t$ which satisfies the representation

$$X_{t,u_0}(u) = \sum_{j=1}^{p} a_j(u_0)X_{t-j,u_0}(u) + \sigma(u_0)\xi_t(u), \tag{14}$$

and has the unique solution

$$X_{t,u_0}(u) \;=\; \sum_{k=0}^{\infty} \zeta_k(u_0)\xi_{t-k}(u), \tag{15}$$

where $\{\zeta_k(u_0)\}_k$ is defined in (2). We note that for fixed u_0, the parameters $\{a_j(u_0)\}$ are constant over space, hence $\{X_{t,u_0}(u) : u \in [0,1]^2\}$ is a spatially stationary process, unlike $\{X_t(u)\}$. Now in the special case that $X_t(u) \equiv X_{t,u_0}(u)$ for all $u \in [0,1]^2$, even when m is fixed, the estimator $\hat{\underline{a}}_{1,T}(u_0)$ is asymptotically unbiased. However, in the general case where $X_t(u) \neq X_{t,u_0}(u)$, we show in the theorem below that the bias of the estimator is due to the difference $\{X_{t-j}(u) - X_{t-j,u_0}(u)\}$.

Theorem 3.1 *Suppose $X_t(u)$ satisfies model (1) and Assumption 2.1. Let $X_{t,u_0}(u)$, $\hat{\underline{a}}_{1,T}(u_0)$ and $\{\zeta_k(u)\}$ be defined as in (14), (9) and (2) respectively. Define*

$$\mathcal{B}(u_0) \;=\; \frac{-2c_\xi(0)}{m}\{\sum_{s=1}^{m} W_b(u_0 - u_s) \cdot (u_s - u_0)\} \sum_{j=1}^{p} \dot{a}_j(u_0)\{\Gamma(u_0)\}_{\cdot,j}$$

$$\Delta_1(u_0) \;=\; \frac{1}{m} \sum_{s=1}^{m} W_b(u_0 - u_s)\mathbb{E}\{\underline{X}_0(u_s)\underline{X}_0(u_s)'\}, \quad \Sigma(u_0) = \frac{\sigma(u_0)^2}{m^2}\|\Sigma_\xi\|_2^2\Gamma(u_0),$$

$\Gamma(u_0)$ is a $p \times p$ matrix with $\{\Gamma(u_0)\}_{a,b} = \sum_{k=0}^{\infty} \zeta_k(u_0)\zeta_{k+|a-b|}(u_0)$, $\Sigma_\xi = var\{\underline{\xi}_t\}$ and $\underline{\xi}'_t = (W_b(u_0 - u_1)\xi_t(u_1), \ldots, W_b(u_0 - u_m)\xi_t(u_m))$. Then we have

(i)

$$\{\hat{a}_{1,T}(u_0) - a(u_0)\} = \Delta_1(u_0)^{-1}\mathcal{B}(u_0) + O_p\{\frac{1}{m}\sum_{s=1}^{m} W_b(u_0 - u_s)\|u_s - u_0\|_\infty^2\}, \quad (16)$$

(ii)

$$\sqrt{T}\left\{[\hat{a}_{1,T}(u_0) - a(u_0)] + \Delta_1(u_0)^{-1}\mathcal{B}_{1,T}(u_0, a(u_0))\right\} \xrightarrow{\mathcal{D}} \mathcal{N}\{0, \Delta_1(u_0)^{-1}\Sigma(u_0)\Delta_1(u_0)^{-1}\}, (17)$$

where

$$\mathcal{B}_{1,T}\{u_0, a(u_0)\} = \frac{-2}{mT'}\sum_{t=p+1}^{T}\sum_{s=1}^{m} W_b(u_0 - u_s) \times$$

$$\left\{\sum_{j=0}^{p} a_j(u_0)[X_{t-j}(u_s)\underline{X}_{t-1}(u_s) - X_{t-j,u_0}(u_s)\underline{X}_{t-1,u_0}(u_s)]\right\}, (18)$$

$\underline{X}_t(u_s)' = (X_t(u_s), \ldots, X_{t-p+1}(u_s)),\ \underline{X}_{t,u_0}(u_s)' = (X_{t,u_0}(u_s), \ldots, X_{t-p+1,u_0}(u_s))$ and $a_0(u_0) = -1$.

Studying the theorem above we see that the bias, $\mathcal{B}_{1,T}\{u_0, a(u_0)\}$, depends, primarily, on the differences $\{X_{t-j}(u_s) - X_{t-j,u_0}(u_s)\}$ and the asymptotic bias depends on the magnitude of the derivatives $\dot{a}_j(u_0)$. Since $\{X_{t-j}(u_s) - X_{t-j,u_0}(u_s)\}$ and $\dot{a}_j(u_0)$ can both be considered as measures of spatial nonstationarity, the bias of $\underline{\hat{a}}_{1,T}(u_0)$ really depends on how close the location dependent process $\{X_t(u)\}$ is to spatially stationary.

We now consider the sampling properties of the local linear least squares estimator.

Theorem 3.2 *Suppose $X_t(u)$ satisfies model (1) and Assumption 2.1 and in addition let $\Theta \subset [0,1]^2$ be a region which includes $\{u_s : s = 0, \ldots, m\}$, where for some $\tilde{\delta} > 0$ and for all $u \in \Theta$, the absolute values of the roots of the characteristic polynomial $x^p - \sum_{j=1}^{p}\{a_j(u_0) + (u - u_0)\dot{a}_j(u_0)\}x^{p-j}$ are less than $1 - \tilde{\delta}$. Let $\underline{\hat{a}}_{2,T}(u_0)$ and $\underline{\hat{a}}_{3,T}(u_0)$ be defined as in (11) and (13) respectively. Then we have*

$$\{\underline{\hat{a}}_{2,T}(u_0) - \underline{a}(u_0)\} = O_p\left\{\Delta_2(u_0)^{-1}\left(\frac{1}{m}\sum_{s=1}^{m} W_b(u_0 - u_s)\|u_s - u_0\|_\infty\right)^2\right\} \quad (19)$$

and

$$\{\underline{\hat{a}}_{3,T}(u_0) - \underline{a}(u_0)\} = O_p\left\{\left(\frac{1}{m}\sum_{s=1}^{m} W_b(u_0 - u_s)\|u_s - u_0\|_\infty\right)^{2-\nu}\right\}. \quad (20)$$

9

where

$$\Delta_2(\underline{u}) = \frac{1}{m} \sum_{s=1}^m W_b(u_0 - u_s) U(s) \otimes \mathbb{E}\{\underline{X}_0(u_s)\underline{X}_0(u_s)'\}.$$

Initially it would appear that the estimator $\hat{\underline{a}}_{3,T}(u_0)$ has a slower rate of convergence than $\hat{\underline{a}}_{2,T}(u_0)$. However, on closer inspection, we notice that the rate in $\hat{\underline{a}}_{2,T}(u_0)$ includes $\Delta_2(u_0)^{-1}$, in other words if $\Delta_2(u_0)$ were close to singular, this can influence the rate of convergence of $\hat{\underline{a}}_{2,T}(u_0)$. On the other hand ν can be arbitrarily small and $\hat{\underline{a}}_{3,T}(u_0)$ is uneffected by $\Delta_2(u_0)^{-1}$.

It is worth noting that, asymptotic normality of $\hat{\underline{a}}_{2,T}(u_0)$ and $\hat{\underline{a}}_{3,T}(u_0)$ can also be shown, and the proof is similar to the proof of Theorem 3.1

In the results above, we have shown that in the case where $T \to \infty$ but m is kept fixed, the estimator can have a significant bias. Often in spatial statistics, 'infill' asymptotics are used (where the number of locations grow) to prove consistency of an estimator (c.f. Mukherjee and Lahiri (2004)). To show that $\hat{\underline{a}}_{1,T}(\cdot)$ is a consistent estimator of $\underline{a}(\cdot)$ we will also use an infilling argument. In the theorem below we consider the case where $m \to \infty$ (the number of sites grow) as $T \to \infty$. In this situation we will show that $\hat{\underline{a}}_{1,T}(u_0)$ is a consistent estimator of $a(u_0)$. Similar results hold true for $\hat{\underline{a}}_{2,T}(u_0)$ and $\hat{\underline{a}}_{3,T}(u_0)$.

Theorem 3.3 *Suppose $X_t(u)$ satisfies model (1) and Assumption 2.1 and let $\hat{\underline{a}}_{1,T}(u_0)$ be defined as in (9). Suppose $u_1, \ldots, u_m$ are independent, identically distributed random variables, whose density, f_U, exists. Further f_U is Hölder continuous and bounded away from zero on $[0,1]^2$. Then for all $u_0 \in [0,1]^2$ we have*

$$\left|\hat{\underline{a}}_{1,T}(u_0) - \underline{a}(u_0)\right|^2 = O_p\left(b^2 + \frac{1}{b^2 m} + \frac{1}{mTb^2} + \frac{1}{T}\right),$$

Furthermore for $b = O(m^{-1/4})$ and $T = O(m^{1/2})$ we have

$$\left|\hat{\underline{a}}_{1,T}(u_0) - \underline{a}(u_0)\right|^2 = O_p(m^{-1/2}).$$

We see from Theorem 3.3 that the estimator is consistent, with $\hat{\underline{a}}_{1,T}(u_0) \xrightarrow{\mathcal{P}} \underline{a}(u_0)$, as $b \to 0$, $m^2 b \to \infty$ and $T \to \infty$.

4 Testing for spatial stationarity

In this section we develop a test for spatial stationarity in the region $\Omega \subset [0,1]^2$. More precisely, our object is to test if $\{X_t(u_s)\}_t$ satisfies the representation

$$X_t(u_s) = \sum_{j=1}^p a_j X_{t-j}(u_s) + \sigma \xi_t(u_s) \text{ for } s = 0, \ldots, m, \tag{21}$$

where $\{\xi_t(u)\}$ is a spatially stationary process which is independent over time and $\mathrm{var}\{\xi_t(u)\} = 1$.

Let us suppose we observe $\{X_t(u_s) : s = 0, \ldots, m, t = 1, \ldots, T\}$. If the estimates of the parameters of the AR models at different locations are significantly different from each other, then this would indicate deviation from spatial stationarity. In view of this, we will compare the estimates of the AR(p) parameters at different locations.

The test is based on the test for homogeneity for time series proposed by Basawa et al. (1984). Using Lütkepohl (1993), Section 4.2.2, to define the likelihood, let

$$\mathcal{S}_T(\hat{a}_{H_0}, \Sigma) = \min_{\alpha, \Sigma} \left[\frac{T}{2} \log |\Sigma| + \frac{1}{2} \sum_{t=p+1}^{T} \left\{ \mathcal{X}_t(\underline{u}) - \sum_{j=1}^{p} \alpha_j \mathcal{X}_{t-j}(\underline{u}) \right\}' \Sigma^{-1} \left\{ \mathcal{X}_t(\underline{u}) - \sum_{j=1}^{p} \alpha_j \mathcal{X}_{t-j}(\underline{u}) \right\} \right] \quad (22)$$

and

$$\begin{aligned}
\mathcal{S}_T(\hat{a}_{H_1}) &= \min_{\alpha_0, \ldots, \alpha_m, \Sigma} \left[\frac{T}{2} \log |\Sigma| + \frac{1}{2} \sum_{t=p+1}^{T} \left\{ \mathcal{X}_t(\underline{u}) - \sum_{j=1}^{p} \mathrm{diag}\{\alpha_{0,j}, \ldots, \alpha_{m,j}\} \mathcal{X}_{t-j}(\underline{u}) \right\}' \Sigma^{-1} \times \right. \\
&\quad \left. \times \left\{ \mathcal{X}_t(\underline{u}) - \sum_{j=1}^{p} \mathrm{diag}\{\alpha_{0,j}, \ldots, \alpha_{m,j}\} \mathcal{X}_{t-j}(\underline{u}) \right\} \right],
\end{aligned} \quad (23)$$

where $\mathcal{X}_t(\underline{u}) = (X_t(u_0), \ldots, X_t(u_m))'$, Σ is a $(p+1) \times (p+1)$ dimensional matrix where $\Sigma_{ij} = \Sigma_{ji}$ and $\underline{u} = (u_0, \ldots, u_m)$. We use as the test statistic

$$Z(\underline{u}) = 2 \left\{ \mathcal{S}_T(\hat{a}_{H_0}) - \mathcal{S}_T(\hat{a}_{H_1}) \right\}. \quad (24)$$

Before deriving the asymptotic distribution of $Z(\underline{u})$ under the null and alternative hypothesis, we first define some notation. Let $\underline{\eta}$ and $\delta(\underline{u})$ be $(m+1)p$-dimensional and mp-dimensional vectors respectively, where

$$\begin{aligned}
\underline{\eta}' &= (a_1(u_0), \ldots, a_1(u_m), \ldots, a_p(u_0), \ldots, a_p(u_m)) \quad (25) \\
\delta(\underline{u})' &= ([a_1(u_1) - a_1(u_0)], \ldots, [a_1(u_m) - a_1(u_0)], \ldots, [a_p(u_1) - a_p(u_0)], \ldots, [a_p(u_m) - a_p(u_0)]).
\end{aligned}$$

We note that under the alternative hypothesis, the Fisher information of $\mathcal{X}_t(\underline{u})$ given the past (evaluated using (23)) is

$$J(\eta) = \begin{pmatrix} \mathbb{E}(\Upsilon_{t-1} \tilde{\Sigma}_\xi^{-1}(\underline{u}) \Upsilon_{t-1}) & \ldots & \mathbb{E}(\Upsilon_{t-1} \tilde{\Sigma}_\xi^{-1}(\underline{u}) \Upsilon_{t-p}) \\ \vdots & \ddots & \vdots \\ \mathbb{E}(\Upsilon_{t-p} \tilde{\Sigma}_\xi^{-1}(\underline{u}) \Upsilon_{t-1}) & \ldots & \mathbb{E}(\Upsilon_{t-p} \tilde{\Sigma}_\xi^{-1}(\underline{u}) \Upsilon_{t-p}) \end{pmatrix}, \quad (26)$$

where $\tilde{\Sigma}_\xi(\underline{u}) = \mathrm{var}((\xi_t(u_0), \ldots, \xi_t(u_m)))$ and

$$\Upsilon_t = \mathrm{diag}(X_t(u_0), \ldots, X_t(u_m)).$$

11

Theorem 4.1 *Supppose the process $\{X_t(u)\}$ satisfies model (1) and Assumption 2.1. Let $Z(\underline{u})$, $\delta(\underline{u})$, $J(\eta)$ and D be defined as in (24), (25) and (26). Then*

(i) under the null hypothesis, as $T \to \infty$, the limiting distribution of $Z(\underline{u})$ is $\chi^2(mp)$.

(ii) under the altervative hypothesis the limiting distribution of $Z(\underline{u})$ is a non-central $\chi^2(mp, \delta^2)$, where for large T

$$\delta^2 = \frac{T}{\sigma^2}\delta(\underline{u})'\big\{D'J(\eta)D\big\}^{-1}\delta(\underline{u}) \tag{27}$$

and D is a $(m+1)p \times mp$ matrix where $D_{ij} = \frac{\partial \delta(\underline{u})'_j}{\partial \eta_i}$.

PROOF. To prove the result we use Basawa et al. (1984), Theorem 2.1, see Subba Rao (2007) for the details. $\square$

Studying the non-centrality parameter δ^2 (defined in (27)) we see that it depends on the differences $\{\underline{a}(u_i) - \underline{a}(u_j)\}$ $(i, j = 0, \ldots, m)$. In other words, the magnitude of the non-centrality parameter is determined by how close $\{X_t(u)\}$ is to spatially stationary in the region Ω.

5 Simulations

To assess the performance of the estimation procedures proposed, we simulate two test processes. That is for $r = 1, 2$ we suppose that $\{X_t^{(r)}(u)\}$ satisfies the representation

$$X_t^{(r)}(u) = a^{(r)}(u)X_{t-j}^{(r)}(u) + \xi_t(u) \qquad t = 1, \ldots, T, \tag{28}$$

where $a^{(r)} : [0, 10]^2 \to [-1, 1]$, with

$$a^{(1)}(x, y) = 0.99 \cdot \sin(0.08x) \cdot \cos(0.2y)$$
$$a^{(2)}(x, y) = \begin{cases} 0.19 \cdot \cos(0.5y) & \text{for } 0 \le x \le 5 \\ 0.19 \cdot (5-x) \cdot \cos(0.5y) & \text{for } 5 < x \le 10 \end{cases}$$

and the innovations $\{\xi_t(u)\}$ is a spatial Gaussian process with mean zero which are independent over time and the covariance of the spatial is defined by the Matérn covariance

$$c(x, y) = \frac{\phi}{2^{\nu-1}\Gamma(\nu)}(\alpha\|x-y\|)^\nu \mathcal{K}_\nu(\alpha\|x-y\|),$$

where $\mathcal{K}$ is the modified Bessel function of the second kind, $\phi = \exp(1)$, $\alpha = \exp(2)$ and $\nu = 2.5 \times \frac{\exp(0.9)}{1+\exp(0.9)}$.

	average mean sq. error	
	local least squares	local linear least squares
parameter $a^{(1)}$	$AISE_{1,1} = 0.01007$	$AISE_{2,1} = 0.0008$
parameter $a^{(2)}$	$AISE_{1,2} = 0.0721$	$AISE_{2,2} = 0.0517$

Table 1: Average integrated squared error of the local least squares and local linear estimators of $a^{(1)}(\cdot)$ and $a^{(2)}(\cdot)$ respectively.

Using the uniform distribution, we randomly sample 50 points on the square $[0, 10]^2$ (note that using $[0, 10]^2$ rather than $[0, 1]^2$ makes no difference to the estimation), we label these points $\{u_i\}_{i=1}^{50}$. The selected points are plotted in Figure 1. The location dependent AR process is simulated for both parameters $a^{(1)}(\cdot)$ and $a^{(2)}(\cdot)$ at the locations $\{u_i\}_{i=1}^{50}$, each realisation is of length 200. We simulate each process 100 times, and for each simulation the local least squares and local linear estimates is evaluated. We denote the local least squares and local linear estimators at location u, for the kth realisation and parameter $a^{(r)}(\cdot)$, $r = 1, 2$, as $\hat{a}_{k,1}^{(r)}(u)$ and $\hat{a}_{k,2}^{(r)}(u)$ respectively. We mention that for these examples there is very little difference between the local linear and regularised local linear estimators defined in (11) and (13) respectively (because the matrix $R_T(\underline{u})$ is non-singular). Therefore we only report the results from the local linear estimator (in practice one uses a regularisation when one believe $R_T(\underline{u})$ is close to singular). The average integrated squared error for both parameters and estimators over the uniform grid is evaluated using

$$AISE_{\ell,r} = \frac{1}{100} \sum_{k=1}^{100} \frac{1}{100} \sum_{i=1}^{10} \sum_{j=1}^{10} [\hat{a}_{k,\ell}^{(r)}(i, j) - a^{(r)}(i, j)]^2$$

and is given in Table 1.

We randomly sample 5 additional points on $[0, 10]^2$ we use these points as locations where there are no observations, and denote these points $w_1, \ldots, w_5$. These points are plotted in Figure 1, where the point w_n is indicated by 'n' (for $n = 1, \ldots, 5$) in the plot. Using the prediction method described in Section 2.2, and the local least square and local linear estimators, we predict the observations at the locations $\{w_n\}_{n=1}^5$ given the observations at the locations $\{u_i; i = 1, \ldots, 50\}$. We estimate the parameters of the Matérn using the empirical residuals of the observed locations and the Gaussian likelihood.

To illustrate how different two observations, which are geographically close, can be, a sample realisation of $X_t^{(r)}(w_1)$ together with its nearest observed neighbour $X_t^{(r)}(u_1)$ and the difference $X_t^{(r)}(w_1) - X_t^{(r)}(u_1)$ is given in Figure 2.

We now study the performance of the estimators of $a^{(1)}(\cdot)$ and $a^{(2)}(\cdot)$, which are given in Table 1, noting that function $a^{(1)}(\cdot)$ is smoother than $a^{(2)}(\cdot)$. From Table 1 we see that the local linear estimator of $a^{(1)}(\cdot)$ is far better than the local least squares estimator. However the difference between the methods is less notable in the estimation of $a^{(2)}(\cdot)$. We recall, the

	unobserved locations				
	w_1	w_2	w_3	w_4	w_5
$\mathrm{var}(X_t^{(1)}(w_n))$	3.236	2.988	2.757	2.742	2.716
prediction with $\hat{\underline{a}}_1^{(1)}$	0.802	2.990	2.137	2.491	2.714
prediction with $\hat{\underline{a}}_2^{(1)}$	0.798	2.990	2.126	2.486	2.714
prediction with `loess`	2.990	3.121	2.606	2.922	3.003
$\mathrm{var}(X_t^{(2)}(w_n))$	2.703	2.784	2.999	2.860	2.779
prediction with $\hat{\underline{a}}_1^{(2)}$	0.790	2.759	2.411	2.585	2.768
prediction with $\hat{\underline{a}}_2^{(2)}$	0.778	2.758	2.506	2.609	2.768
prediction with `loess`	2.512	2.898	2.879	3.059	3.063

Table 2: Prediction when the location dependent parameter is $a^{(1)}(\cdot)$ and $a^{(2)}(\cdot)$.

local linear estimator is constructed under the assumption that the first derivative of $a(\cdot)$ exists everywhere, but the partial derivative of $a^{(2)}(\cdot)$ with respect to x does not exist at $x = 5$. This is the reason why we don't see a significant difference between the two estimators in this case.

Now we consider the average squared error of the predictions given in Table 2. The predictions are compared with the `loess` function predictions which can be found in the statistical software `R`. We first observe that the prediction using the local linear estimator is better than the prediction using the local least squares estimator. And both estimators have average squared errors which are less than the variance of the observations and are consistently better than the `loess` estimator. The improvement in our method over the `loess` method appears to depend on the proximity of the unobserved location with relation to the observed locations. We see that when the unobserved location is close to the observed locations (see location w_1) our estimator performs significantly better than the `loess` estimator. The difference between the estimation performance is not so apparent when the unobserved locations are far from the observed locations.
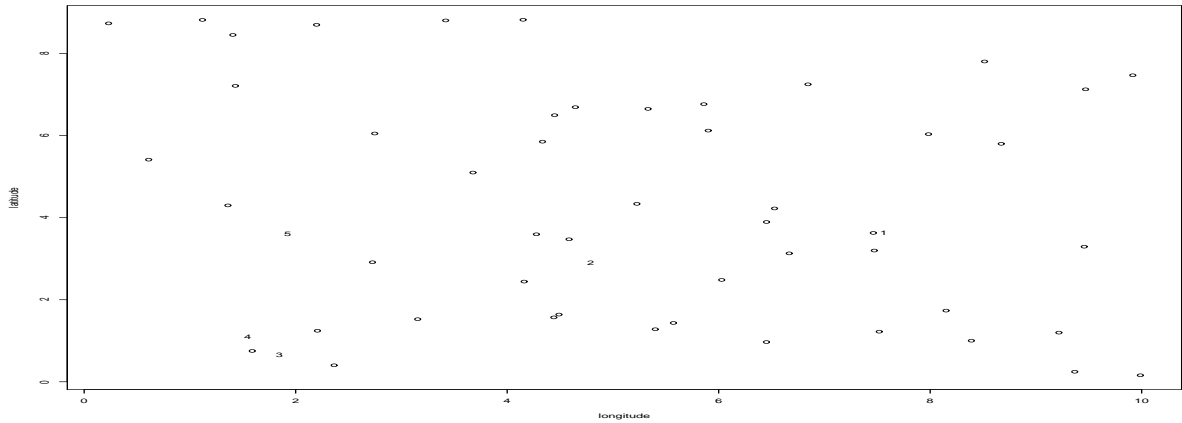
Acknowledgments

14

Figure 1: The circles are the observed locations on the square $[0, 10]^2$. The numbered points are the unobserved locations.
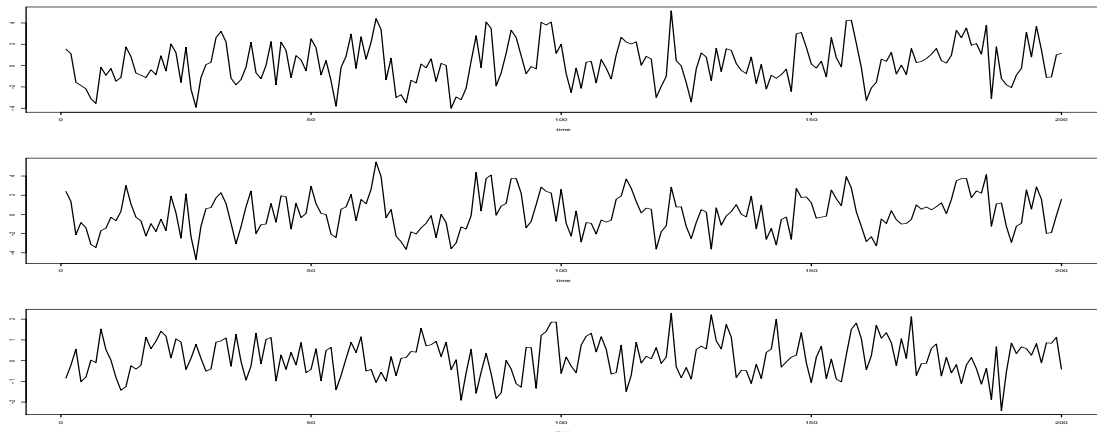

Figure 2: The upper plot is time series at location $(7.4622, 3.6285)$, the middle plot is the time series at $(7.5616, 3.6296)$ and the lower plot is the difference between the two time series.

A   Appendix

The purpose of this appendix is to give a rough outline for the proofs in the sections above. The full details can be found in the accompanying technical report (Subba Rao, 2007).

A.1   The location dependent AR process and spatial local stationarity

In this section we consider some of the probabilistic properties of the location dependent AR process. In particular, we will show that the spatially nonstationary process $X_t(u)$ can be approximated by the spatially stationary process $X_{t,u_0}(u)$. The results in this section are used to prove the results in Section 3. The proofs of the results in this section can be found in Subba Rao (2007). The results may also be of independent interest.

In the theorem below we will show that the spatially stationary process $X_{t,u_0}(u)$ locally approximates the spatially nonstationary process $X_t(u)$ in some neighbourhood of u_0.

Theorem A.1 *Suppose $X_t(u)$ satisfies model (1) and Assumption 2.1. Let $X_{t,u_0}(u)$ be defined as in (14). Then we have*

$$|X_t(u) - X_{t,u_0}(u)| \leq \|u - u_0\|_\infty V_t(u), \tag{29}$$

where for some $1 - \delta \leq \rho < 1$,

$$V_t(u) = C \sum_{k=0}^{\infty} (k+1)^2 \rho^k |\xi_{t-k}(u)|, \tag{30}$$

C is a finite constant independent of u, and $V_t(u)$ is a stationary process in both space and time with a finite variance.

The theorem above gives an approximation for $X_t(u)$ in terms of the spatially stationary process $X_{t,u_0}(u)$, we now refine this approximation. To do this we define the 'derivative' spatial process , which we show in Subba Rao (2007), is the derivative of the spatially stationary process $X_{t,u_0}(u)$ with respect to u_0. Let us suppose the 2-dimensional vector process $\{\dot{X}_{t,u_0}(u) = (\frac{\partial X_{t,u_0}(u)}{\partial x_0}, \frac{\partial X_{t,u_0}(u)}{\partial y_0})'\}_t$ satisfies the representation

$$\dot{X}_{t,u_0}(u) = \sum_{j=1}^{p} \left\{ \dot{a}_j(u_0) X_{t-j,u_0}(u) + a_j(u_0) \dot{X}_{t-j,u_0}(u) \right\} + \dot{\sigma}(u_0) \xi_t(u). \tag{31}$$

We use this process in the results below. We use this result to derive the bias of the estimator $\hat{\underline{a}}_{1,T}$.

Theorem A.2 *Suppose $X_t(u)$ satisfies model (1) and Assumption 2.1. Let $X_{t,u_0}(u)$, $\dot{X}_{t,u_0}(u)$ and $V_t(u)$ be defined as in (14), (31) and (30) respectively. Then we have*

$$|X_t(u) - X_{t,u_0}(u) - (u - u_0)\dot{X}_{t,u_0}(u)| \leq \|u - u_0\|_\infty^2 V_t(u). \tag{32}$$

We observe that (32) is a representation of a spatially nonstationary process in terms of spatially stationary processes. And if u is close to u_0, then $X_t(u) \approx X_{t,u_0}(u) + (u-u_0)\dot{X}_{t,u_0}(u)$.

In order to study the properties of $\hat{a}_{2,T}(u_0)$ we define the process $\{\tilde{X}_{t,u_0}(u)\}_t$. Let u_0 be a fixed location and suppose the process $\{\tilde{X}_{t,u_0}(u)\}_t$ satisfies the representation

$$\tilde{X}_{t,u_0}(u) = \sum_{j=1}^{p} \{a_j(u_0) + (u - u_0)\dot{a}_j(u_0)\}\tilde{X}_{t-j,u_0}(u) + \{\sigma(u_0) + (u - u_0)\dot{\sigma}(u_0)\}\xi_t(u). \tag{33}$$

We show in the theorem below that $X_t(u)$ can also be approximated by $\tilde{X}_{t,u_0}(u)$. This theorem is an important component in the proof of Theorem 3.2.

Theorem A.3 *Suppose $X_t(u)$ satisfies model (1) and Assumption 2.1. Let $\tilde{X}_{t,u_0}(u)$ be defined as in (33). Let Θ be a region which includes $\{u_s; s = 0, \ldots, m\}$, where for some $\tilde{\delta} > 0$ and for all $u \in \Theta$ the roots of the characteristic polynomial $x^p + \sum_{j=1}^{p}\{a_j(u_0) + (u - u_0)\dot{a}_j(u_0)\}x^{p-j}$ are less than $1 - \tilde{\delta}$. Then for some $1 - \tilde{\delta} < \tilde{\rho} < 1$, we have*

$$|X_t(u) - \tilde{X}_{t,u_0}(u)| \leq \|u - u_0\|_\infty^2 \tilde{V}_t(u), \tag{34}$$

where

$$\tilde{V}_t(u) = C \sum_{k=0}^{\infty} (k + 1)^2 \tilde{\rho}^k |\xi_{t-k}(u)|, \tag{35}$$

C is a finite constant independent of u, and $\tilde{V}_t(u)$ is a spatially stationary process with finite variance.

A.2 Proof of Theorem 3.1

We use the following definitions. Suppose $f : \mathbb{R}^{p+1} \to \mathbb{R}$, let $\nabla f(u, x_1, \ldots, x_p) = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_p})'$ and let $\nabla^2 f(u, x_1, \ldots, x_p)$ be a $p \times p$ matrix, where $(\nabla^2 f)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$.

Since $\nabla \mathcal{L}_{1,T}\{u_0, \underline{\hat{a}}_{1,T}(u_0)\} = 0$, by expanding $\nabla \mathcal{L}_{1,T}(u_0, \underline{\hat{a}}_{1,T}(u_0))$ about $a(u_0)$, we have

$$\nabla \mathcal{L}_{1,T}\{u_0, \underline{a}(u_0)\} = -\nabla^2 \mathcal{L}_{1,T}\{u_0, \underline{a}(u_0)\}\{\underline{\hat{a}}_{1,T}(u_0) - \underline{a}(u_0)\}. \tag{36}$$

To obtain the sampling properties of $\underline{\hat{a}}_{1,T}(u_0)$ we consider $\nabla\mathcal{L}_{1,T}\{u_0,\underline{a}(u_0)\}$. In order to obtain the bias and show asymptotic normality of $\underline{\hat{a}}_{1,T}(u_0)$ we replace $X_t(u_s)$ with $X_{t,u_0}(u_s)$ (where $X_{t,u_0}(u_s)$ is defined as in (14)) and obtain

$$\nabla\mathcal{L}_{1,T}\{u_0,\underline{a}(u_0)\} = \nabla\tilde{\mathcal{L}}_{1,T}\{u_0,\underline{a}(u_0)\} + \mathcal{B}_{1,T}\{u_0,\underline{a}(u_0)\}, \tag{37}$$

where $\mathcal{B}_{1,T}\{u_0,\underline{a}(u_0)\}$ is defined in (18) and

$$\nabla\tilde{\mathcal{L}}_{1,T}\{u_0,a(u_0)\} = \frac{-2}{mT}\sum_{t=p+1}^{'}\sum_{s=1}^{m} w_s\sigma(u_0)\xi_t(u_s)\underline{X}_{t-1,u_0}(u_s) \tag{38}$$

with $w_s = W_b(u_0 - u_s)$.

We use $\nabla\tilde{\mathcal{L}}_{1,T}\{u_0,a(u_0)\}$ to show asymptotic normality of the estimator and $\mathcal{B}_{1,T}\{u_0,a(u_0)\}$ to obtain an expression for the bias of the estimator.

In order to evaluate the limit of the bias $\mathcal{B}_{1,T}\{u_0,a(u_0)\}$, we use Theorem A.2 to obtain

$$X_{t-i}(u)X_{t-j}(u) - X_{t-i}(u_0,u)X_{t-j}(u_0,u)$$
$$= (u-u_0)\left\{\dot{X}_{t-i,u_0}(u)X_{t-j,u_0}(u) + X_{t-i,u_0}(u)\dot{X}_{t-j,u_0}(u)\right\} + R_t^{ij}(u_0,u) \tag{39}$$

where $|R_t^{ij}(u_0,u)| \le 4\|u-u_0\|_\infty^2 V_{t-i}(u)V_{t-j}(u)$.

Lemma A.1 *Suppose the assumptions in Theorem 3.1 hold, then*

$$\mathcal{B}_{1,T}\{u_0,\underline{a}(u_0)\} \xrightarrow{\mathcal{P}} \mathcal{B}(u_0) + O\{\frac{1}{m}\sum_{s=1}^{m} W_b(u_0-u_s)\|u_s-u_0\|_\infty^2\}, \tag{40}$$

as $T \to \infty$.

PROOF. By substituting (39) into $\mathcal{B}_{1,T}\{u_0,a(u_0)\}$ and using the ergodic theorem it is straightforward to show (40). $\qquad\square$

In the lemma below we show asymptotic normality of the stochastic term $\nabla\tilde{\mathcal{L}}_{1,T}\{u_0,a(u_0)\}$.

Lemma A.2 *Suppose $X_t(u)$ satisfies Assumption 2.1. Let $\nabla\tilde{\mathcal{L}}_{1,T}\{u_0,\underline{a}(u_0)\}$ and $\Sigma_1(u_0)$ be defined as in (38) and (16) respectively. We have that*

$$var\left[\sqrt{T'}\nabla\tilde{\mathcal{L}}_{1,T}\{u_0,\underline{a}(u_0)\}\right] \to 4\Sigma_1(u_0) \tag{41}$$

$$\nabla\tilde{\mathcal{L}}_{1,T}\{u_0,\underline{a}(u_0)\} \xrightarrow{\mathcal{P}} 0, \tag{42}$$

$$\sqrt{T'}\nabla\tilde{\mathcal{L}}_{1,T}\{u_0,\underline{a}(u_0)\} \xrightarrow{\mathcal{D}} \mathcal{N}\{0,4\Sigma_1(u_0)\}, \tag{43}$$

as $T \to \infty$.

PROOF. We prove (41) and (42) by using the ergodic theorem and (43) by verifying Lindeberg's condition and the conditions given in (Hall & Heyde, 1980), Theorem 3.2. □

We now prove Theorem 3.1. Substituting (37) into (36) we have

$$\{\hat{\underline{a}}_T(u_0) - \underline{a}(u_0)\} = -[\nabla^2 \mathcal{L}_{1,T}\{u_0, \underline{a}(u_0)\}]^{-1}\big[\nabla \tilde{\mathcal{L}}_{1,T}\{u_0, a(u_0)\} + \mathcal{B}_{1,T}\{u_0, a(u_0)\}\big]. \quad (44)$$

By using (40) and (42) we have

$$\nabla \tilde{\mathcal{L}}_{1,T}\{u_0, \underline{a}(u_0)\} + \mathcal{B}_{1,T}(u_0, \underline{a}(u_0)) \xrightarrow{\mathcal{P}} \mathcal{B}(u_0) + O\big\{\frac{1}{m}\sum_{s=1}^{m} w_s\|u_s - u_0\|_\infty^2\big\}, \quad (45)$$

Furthermore, by using that $\nabla^2 \mathcal{L}_{1,T}\{u_0, \alpha\} \xrightarrow{\mathcal{P}} 2\Delta_1(u_0)$, and Slutsky's theorem we have (16), as $T \to \infty$. Finally, (17) follows from (44) and the asymptotic normality of $\nabla \tilde{\mathcal{L}}_{1,T}\{u_0, \underline{a}(u_0)\}$, given in (43).

A.3   Proof of Theorem 3.2

We now study the estimators $\hat{\underline{a}}_{2,T}(u_0)$ and $\hat{\underline{a}}_{3,T}(u_0)$, and evaluate their asymptotic bias.

Define $\hat{\underline{r}}_{T,u_0}(\underline{u})$ and $\hat{R}_{T,u_0}(\underline{u})$ in the same way as $\hat{r}_T(\underline{u})$ and $\hat{R}_T(\underline{u})$ but with $\tilde{X}_{t,u_0}(u_s)$ (defined in (33)) and $\underline{\tilde{X}}_{t-1,u_0} = (\tilde{X}_{t-1,u_0}(u_s), \ldots, \tilde{X}_{t-p,u_0}(u_s))$ replacing $X_t(u_s)$ and $\underline{X}_{t-1}(u_s)$ respectively. We now state the following result, which follows immediately from the ergodicity of the process, and will be used later in the proof,

$$\hat{\underline{r}}_{T,u_0}(\underline{u}) - \hat{R}_{T,u_0}(\underline{u})\mathcal{A}(\underline{u})$$

$$= \frac{1}{m(T-p-1)}\sum_{s=1}^{m}\sum_{t=p+1}^{\prime} w_s\{\sigma(u_0) + (u_s - u_0)\dot{\sigma}(u_0)\}\xi_t(u_s)\begin{pmatrix} 1 \\ u_s - u_0 \end{pmatrix} \otimes \underline{X}_{t-1,u_0}(u_s)$$

$$\xrightarrow{\mathcal{P}} 0, \quad \text{as } T \to \infty. \quad (46)$$

Furthermore, it is straightforward to show

$$\hat{R}_T(\underline{u})^{-1}\hat{r}_T(\underline{u}) - \hat{R}_{T,u_0}(\underline{u})^{-1}\hat{r}_{T,u_0}(\underline{u}) = \hat{R}_T(\underline{u})^{-1}\Big\{[\hat{R}_T(\underline{u}) - \hat{R}_{T,u_0}(\underline{u})]\mathcal{A}(u_0) + [\hat{\underline{r}}_T(\underline{u}) - \hat{\underline{r}}_{T,u_0}(\underline{u})]\Big\}$$

$$+\hat{R}_T(\underline{u})^{-1}\Big\{\hat{R}_T(\underline{u}) - \hat{R}_{T,u_0}(\underline{u})\Big\}\Big\{\hat{R}_{T,u_0}(\underline{u})^{-1}\hat{\underline{r}}_{T,u_0}(\underline{u}) - \mathcal{A}(u_0)\Big\}.$$

Now let us consider the estimator $\hat{\mathcal{A}}_{2,T}$. By using the above it is straightforward to show

19

that

$$
\begin{aligned}
\hat{\mathcal{A}}_{2,T} - \mathcal{A}(u_0) &= \Delta_2(\underline{u})^{-1}\mathcal{B}_{2,T}(u_0) + \left\{\hat{R}_T(\underline{u})^{-1} - \Delta_2(\underline{u})^{-1}\right\}\mathcal{B}_{2,T}(u_0), \\
&\quad + \hat{R}_T(\underline{u})^{-1}\left\{\hat{R}_T(\underline{u}) - \hat{R}_{T,u_0}(\underline{u})\right\}\hat{R}_{T,u_0}(\underline{u})^{-1}\left\{\hat{\underline{r}}_{T,u_0}(\underline{u}) - \hat{R}_{T,u_0}(\underline{u})\mathcal{A}(u_0)\right\} \\
&\quad + \hat{R}_{T,u_0}(\underline{u})^{-1}\left\{\hat{r}_{T,u_0}(\underline{u}) - \hat{R}_{T,u_0}(\underline{u})\mathcal{A}(\underline{u})\right\} \\
&= I_1 + I_2 + I_3 + I_4, \tag{47}
\end{aligned}
$$

where $\mathcal{A}(u_0) = \left(\underline{a}(u_0)', \dot{a}_1(u_0), \ldots, \dot{a}_p(u_0)\right)$ and

$$
\mathcal{B}_{2,T}(u_0) = \left\{\hat{\underline{r}}_T(\underline{u}) - \hat{\underline{r}}_{T,u_0}(\underline{u})\right\} + \left\{\hat{R}_T(\underline{u}) - \hat{R}_{T,u_0}(\underline{u})\right\}A(u_0).
$$

We now show that $I_2, I_3, I_4 \xrightarrow{\mathcal{P}} 0$ and that $I_1 = O(\frac{1}{m}\sum_{s=1}^m w_s\|u_s - u_0\|_\infty^2)$, which leads to the desired result. By using (46) and $\hat{R}_T(\underline{u}) \xrightarrow{\mathcal{P}} \Delta_2(\underline{u})$ as $T \to \infty$, it follows that $I_2 \xrightarrow{\mathcal{P}} 0$, $I_3 \xrightarrow{\mathcal{P}} 0$ and $I_4 \xrightarrow{\mathcal{P}} 0$ as $T \to \infty$.

To bound I_1 we use that

$$
\begin{aligned}
&|X_{t-j}(u)X_{t-i}(u) - \tilde{X}_{t-j}(u_0, u)\tilde{X}_{t-i}(u_0, u)| \\
&\leq \|u - u_0\|_\infty^2\{\tilde{V}_{t-j}(u)|X_{t-i}(u)| + \tilde{V}_{t-i}(u)|\tilde{X}_{t-j}(u_0, u)|\}, \tag{48}
\end{aligned}
$$

which follows immediately from Theorem A.3. Substituting the above bound into $\mathcal{B}_{2,T}(u_0)$, and appealing to the ergodic theorm gives

$$
\begin{aligned}
\|\mathcal{B}_{2,T}(u_0)\|_1 &\leq \frac{1}{mT'}\sideset{}{'}\sum_{t=p+1}\sum_{s=1}^m w_s\|u_s - u_0\|_\infty^2\left(\sum_{j=0}^p\{a_j(u_0) + (u_s - u_0)\dot{a}_j(u_0)\} \times\right. \\
&\qquad \left. \times\{\sum_{i=1}^p\{\tilde{V}_{t-j}(u_s)|X_{t-i}(u_s)| + \tilde{V}_{t-i}(u_s)|\tilde{X}_{t-j,u_0}(u_s)|\}\right) \\
&\xrightarrow{\mathcal{P}} \frac{1}{m}\sum_{s=1}^m w_s\|u_s - u_0\|_\infty^2\sum_{j=0}^p\{a_j(u_0) + (u_s - u_0)\}\dot{a}_j(u_0) \times \\
&\qquad \mathbb{E}\left(\sum_{i=1}^p\tilde{V}_{t-j}(u_s)|X_{t-i}(u_s)| + \tilde{V}_{t-i}(u_s)|\tilde{X}_{t-j,u_0}(u_s)|\right) \leq K\frac{1}{m}\sum_{s=1}^m w_s\|u_s - u_0\|_\infty^2
\end{aligned}
$$

as $T \to \infty$ and K is a finite constant. By substituting the above into I_1, recalling that $I_2, I_3, I_4 \xrightarrow{\mathcal{P}} 0$ and using (47), we obtain (19).

By using the same arguments as those above and that the smallest eigenvalue of $(\hat{R}_T(\underline{u}) + \eta^\gamma I)$ is bounded from below by η^γ, we obtain (20).

A.4 Proof of Theorem 3.3

To reduce notation we prove the result for location dependent AR(1) processes, the proof for the general location dependent AR(p) process is the same. We recall that the estimator $\hat{a}_{1,T}(u_0)$ can be written as

$$\hat{a}_{1,T}(u_0) \ = \ \frac{\sum_{t=2}^{T}\sum_{s=1}^{m} W_b(u_0 - U_s)X_t(U_s)X_{t-1}(U_s)}{\sum_{t=1}^{T-1}\sum_{s=1}^{m} W_b(u_0 - U_s)X_t(U_s)^2} := \frac{\hat{c}_1(u_0)}{\hat{c}_0(u_0)},$$

where we have replaced u_s by U_s to emphasis that the locations are random variables. Let f_U denote the density of U_s. It is straightforward to show that $\hat{a}_{1,T}(\cdot)$ can be written as

$$\hat{a}_{1,T}(u_0) - a(u_0) = \frac{\{\hat{c}_1(u_0) - c_1(u_0)f_U(u_0)\}}{\hat{c}_0(u_0)} + \frac{a(u_0)\{\hat{c}_0(u_0) - c_0(u_0)f_U(u_0)\}}{\hat{c}_0(u_0)}. \tag{49}$$

We start by evaluating the mean squared error of the numerator of the above, that is $\mathbb{E}\{\hat{c}_0(u_0) - c_0(u_0)f_U(u_0)\}^2$ and $\mathbb{E}\{\hat{c}_1(u_0) - c_1(u_0)f_U(u_0)\}^2$. To obtain a bound for the mean squared errors we use the classical variance/bias decomposition. To obtain the bias we note that under Assumption 2.1(iii) it is straightforward to show that

$$c_{t_1-t_2}(u) = c_{t_1-t_2}(u_0) + (u - u_0)\dot{c}_{t_1-t_2}(u_0) + O(\|u - u_0\|_\infty^2),$$

where $\text{cov}(X_{t_1}(u), X_{t_2}(u)) = c_{t_1-t_2}(u_1)$. Using the expansion above, the Hölder continuity of f_U and iterated expectation arguments it is straightforward to show that $\mathbb{E}[\hat{c}_1(u_0)] = c_1(u_0)f(u_0)+O(b)$ and $\mathbb{E}[\hat{c}_0(u_0)] = c_0(u_0)f(u_0)+O(b)$. Appealing to the results in Subba Rao (2007) we have $\text{var}\{\hat{c}_1(u_0)\}, \text{var}\{\hat{c}_0(u_0)\} = O(\frac{1}{b^2Tm} + \frac{1}{b^2m} + \frac{1}{T})$. Altogether this gives

$$\mathbb{E}\{\hat{c}_0(u_0) - c_0(u_0)f_U(u_0)\}^2 \ = \ O\left(b^2 + \frac{1}{b^2Tm} + \frac{1}{b^2m} + \frac{1}{T}\right)$$

$$\text{and similarly} \quad \mathbb{E}\{\hat{c}_1(u_0) - c_1(u_0)f_U(u_0)\}^2 \ = \ O\left(b^2 + \frac{1}{b^2Tm} + \frac{1}{b^2m} + \frac{1}{T}\right). \tag{50}$$

Studying $c_0(\cdot)f_U(\cdot)$ we observe that since $\inf_u \sigma(u) > 0$, $c_0(u_0)$ is bounded away from zero and by assumption $f_U(u_0)$ is bounded away from zero, therefore $c_0(\cdot)f_U(\cdot)$ is bounded away from zero. Since (50) implies $\hat{c}_0(u_0) \xrightarrow{\mathcal{P}} c_0(\cdot)f_U(\cdot)$ and $c_0(\cdot)f_U(\cdot)$ is bounded away from zero, by Slutsky's lemma we have $\hat{c}_0(u_0)^{-1} \xrightarrow{\mathcal{P}} c_0(\cdot)^{-1}f_U(\cdot)^{-1}$. Therefore by applying the delta method, the bounds in (50) determine the bounds for $|\hat{a}_1(u_0) - a_1(u_0)|$ and we have

$$|\hat{a}_{1,T}(u_0) - a(u_0)|^2 = O\left(b^2 + \frac{1}{b^2Tm} + \frac{1}{b^2m} + \frac{1}{T}\right).$$

It follows from the above that $\hat{a}_{1,T}(u_0)$ is a consistent estimator of $a(u_0)$ if $b \to 0$ and $b^2m \to \infty$ as $m \to \infty$ and $T \to \infty$. And the optimal rate of convergence is obtained when $b = O(m^{-1/4})$. $\qquad\square$

Reference

Basawa, I., Billard, L., & Srinivas, R. (1984). Large-sample test of homogeneity for time series models. *Biometrika*, *71*, 203-6.

Cressie, N., & Huang, H. (1999). Classes of nonseparable spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, *94*, 1330-40.

Cressie, N. A. (1993). *Statistics for Spatial Data.* New York: John Wiley and sons.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.*, *21*, 196-216.

Gelfand, A. E., Kim, H.-J., Sirmans, C. F., & Banerjee, S. (2003). Spatial modelling with spatially varying coefficient processes. *Journal of the American Statistical Association*, *98*.

Gilleland, E., & Nychka, D. (2005). Statistical models for the monitoring and regulating ground-level ozone. *Environmetrics*, *16*, 535-546.

Guan, Y., Sherman, M., & Calvin, J. (2004). A nonparametric test for spatial isoptropy using subsampling. *Journal of the American Statistical Association*, *99*, 810-821.

Hall, P., & Heyde, C. (1980). *Martingale Limit Theory and its Application.* New York: Academic Press.

Hallin, M., Lu, Z., & Tran, L. (2004). Local linear spatial regression. *Ann. Statist.*, *32*, 2469-2500.

Johns, C., Nychka, D., Kittel, T., & Daly, C. (2003). Infilling sparce records of spatial fields. *Journal of the American Statistical Association*, *98*, 796-806.

Jun, M., & Stein, M. L. (2007). An approach to producing space-time covariance functions on spheres. *Technometrics*, To appear.

Lu, Z., Lundervold, A., Tjøstheim, D., & Yao, Q. (2004). Exploring spatial nonlinearity using additive approximation. *Preprint.*

Lütkepohl, H. (1993). *Introduction to multiple time series analysis.* Berlin: Springer.

Mardia, K., & Marshall, R. J. (1984). Maximum likelihood estimation of models of residual covariance in spatial statistics. *Biometrika*, *71*, 135-146.

Matsuda, Y., & Yajima, Y. (2004). On testing for separable correlations in multivariate time series. *Journal of Time Series Analysis*, *25*, 501-528.

Mukherjee, K., & Lahiri, S. (2004). Asymptotic distribution of m-estimators in spatial regression under fixed and some stochastic spatial sampling designs. *Annals of the Institute of Statistical Mathematics*, *56*, 225-250.

Shitan, M., & Brockwell, P. (1995). An asymptotic test for separability of a spatial autoregressive model. *Commun. Statist-Theory Meth.*, *24*, 2027-2040.

Storvik, G., Frigessi, A., & Hirst, D. (2002). Space Time Gaussian fields and their autoregressive representation. *Statistical Modelling*, *2*, 139-61.

Subba Rao, S. (2007). Statistical analysis of a spatio-temporal model with location dependent parameters and a test for spatial stationarity - Technical Report. `http://www.stat.tamu.edu/∼suhasini/papers/spatial_temporal_technical_report.pdf`.

Tibshirani, R., & Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, *82*, 559-567.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, **41**, 434-49.

Zhang, H. (2004). In consistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Ass.*, *99*, 250-252.