Spatial Smoothing, Nugget Effect and Infill Asymptotics*

Zudi Lu[1] Dag Tjøstheim[2] Qiwei Yao[3]

[1]Department of Mathematics & Statistics, Curtin University of Technology, Perth, Australia

[2]Department of Mathematics, University of Bergen, 5007 Bergen, Norway

[3]Department of Statistics, London School of Economics, London, WC2A 2AE, UK

Abstract

For spatio-temporal regression models with observations taken regularly in time but irregularly over space, we investigate the effect of spatial smoothing on the reduction of variance in estimating both parametric and nonparametric regression functions. The processes concerned are stationary in time but may be nonstationary over space. Our study indicates that under the infill asymptotic framework, the existence of the so-called nugget effect in either regressor process or noise process is necessary for spatial smoothing to reduce the estimation variance. In particular the nugget effect in the regressor process may lead to a faster convergence rate in estimating nonparametric regression functions.

Keywords: Fixed-domain asymptotics, Non-lattice data; Nonparametric kernel smoothing; Nonstationarity; Spatio-temporal processes; Time series; Variance reduction

*Partially supported by a Leverhulme Trust research grant and an EPSRC research grant.

1 Introduction

The goal of the paper is to reveal the impact of a nugget effect on spatial smoothing under the infill asymptotic framework which is also called the fixed-domain asymptotics (c.f. Cressie 1993, section 3.3, and Stein 1999, Chapter 3). We consider a class of spatio-temporal models designed for fitting data observed regularly in time and irregularly over space. Since we assume that the process at a given location depends on the time-lagged values of that location and of the spatial neighbouring locations, the process is unilateral in time. This is in marked contrast to purely spatial autoregressive models for which an artificial unilateral order is used in model specification and this artifact limits severely the applicability in modelling real spatial data (Yao and Brockwell 2006).

Another important feature of our setting is that stationarity is assumed in time only, not over space. Indeed the model is capable of catching nonstationary spatial variation, although we require that the functions to be estimated are continuous. Accordingly we adopt a strategy of two step estimation: an initial estimator at each location is based on the observations taken at this location only, and then spatial smoothing to pull information together from neighbouring locations. The initial estimation is the same as that for a time series, which makes effective use of the stationarity in time. The second estimator, i.e. the spatial smoothing, relies on the continuity of the function to be estimated. Naturally one would expect that the spatial smoothing could improve the estimator obtained in the first step. Unfortunately this is only true in the presence of the so-call nugget effect. We illustrate this phenomenon in both parametric and nonparametric settings. To highlight the essence of the problem, we impose some generic assumptions on the initial estimators to avoid some standard and tedious regularity conditions. A special case of the parametric setting (i.e. linear model) was investigated in Zhang *et al.* (2003). A part of the result for nonparametric setting was established in Lu *et al.* (2007) for varying coefficient spatio-temporal models. Both Zhang *et al.* (2003) and Lu *et al.* (2007) contain some simulation studies illustrating the finite sample performance of the two step estimation procedure.

The results reported in this paper are somehow in a similar spirit to those of Zhang (2004) which showed that parameters in some parametric variograms of spatial processes may not be estimated consistently under the infill asymptotic framework.

2 Models and estimation methods

2.1 Spatio-temporal regression model

For each fixed location $\mathbf{s} = (u, v)^\tau \in \mathcal{S}$, the process $\{(Y_t(\mathbf{s}), \mathbf{X}_t(\mathbf{s})),\ t = 1, 2, \cdots\}$ is strictly stationary, where $Y_t(\mathbf{s})$ is a scalar, $\mathbf{X}_t(\mathbf{s})$ is a $d \times 1$ vector, and $\mathcal{S}$ is a bounded subset of $\mathcal{R}^2$. Furthermore, we assume that

$$Y_t(\mathbf{s}) = m_{\mathbf{s}}\{\mathbf{X}_t(\mathbf{s})\} + \varepsilon_t(\mathbf{s}) \equiv m\{\mathbf{s}, \mathbf{X}_t(\mathbf{s})\} + \varepsilon_t(\mathbf{s}), \qquad t = 1, 2, \cdots, \tag{2.1}$$

where the regression function $m_{\mathbf{s}}(\cdot) = m(\mathbf{s}, \cdot)$ may be known upto some unknown parameters, or completely unknown. Nevertheless we always assume $m(\mathbf{s}, \mathbf{x})$ is continuous in $\mathbf{s}$. Furthermore we assume that the noise processes $\varepsilon_t(\mathbf{s})$ satisfy the conditions below.

A1 $\{\varepsilon_1(\mathbf{s}), \mathbf{s} \in \mathcal{S}\}$, $\{\varepsilon_2(\mathbf{s}), \mathbf{s} \in \mathcal{S}\}, \cdots$ is a sequence of independent and identically distributed spatial processes, and $E\{\varepsilon_t(\mathbf{s})\} = 0$. Further, for each $t > 1$, $\{\varepsilon_t(\mathbf{s}), \mathbf{s} \in \mathcal{S}\}$ is independent of $\{(Y_{t-j}(\mathbf{s}), \mathbf{X}_{t+1-j}(\mathbf{s})), \mathbf{s} \in \mathcal{S} \text{ and } j \geq 1\}$. The spatial covariance function

$$\Gamma(\mathbf{s}_1, \mathbf{s}_2) \equiv \mathrm{Cov}\{\varepsilon_t(\mathbf{s}_1), \varepsilon_t(\mathbf{s}_2)\} \tag{2.2}$$

is bounded over $\mathcal{S}^2$.

In addition we assume that the noise $\varepsilon_t(\mathbf{s})$ admits the decomposition below.

A2 For any $t \geq 1$ and $\mathbf{s} \in \mathcal{S}$,

$$\varepsilon_t(\mathbf{s}) = \varepsilon_{1,t}(\mathbf{s}) + \varepsilon_{2,t}(\mathbf{s}), \tag{2.3}$$

where $\{\varepsilon_{1,t}(\mathbf{s}),\ t \geq 1, \mathbf{s} \in \mathcal{S}\}$ and $\{\varepsilon_{2,t}(\mathbf{s}),\ t \geq 1, \mathbf{s} \in \mathcal{S}\}$ are two independent processes, and both fulfill the conditions imposed on $\{\varepsilon_t(\mathbf{s})\}$ in A1 above. Further, $\Gamma_1(\mathbf{s}_1, \mathbf{s}_2) \equiv \mathrm{Cov}\{\varepsilon_{1,t}(\mathbf{s}_1)\,,\, \varepsilon_{1,t}(\mathbf{s}_2)\}$ is continuous in $(\mathbf{s}_1, \mathbf{s}_2)$, and $\mathrm{Cov}\{\varepsilon_{2,t}(\mathbf{s}_1)\,,\, \varepsilon_{2,t}(\mathbf{s}_2)\} = \sigma_2^2(\mathbf{s}_1) \geq 0$ if $\mathbf{s}_1 = \mathbf{s}_2$, and 0 otherwise, where $\sigma_2^2(\mathbf{s})$ is continuous.

When $\sigma_2^2(\mathbf{s}) > 0$, Condition A2 implies that the *nugget effect* exists in the spatial noise process $\{\varepsilon_t(\mathbf{s}),\ \mathbf{s} \in \mathcal{S}\}$. The nugget effect was introduced by G. Matheron in early 1960's. It reflects the fact that the variogram $E\{\varepsilon_t(\mathbf{s}_1) - \varepsilon_t(\mathbf{s}_2)\}^2$ does not converges to 0 as $||\mathbf{s}_1 - \mathbf{s}_2|| \to 0$, where $||\cdot||$ denotes the Euclidean distance. In our notation, it is equivalent to the fact that the function $\gamma(\mathbf{s}) \equiv \Gamma(\mathbf{s}_1 + \mathbf{s},\ \mathbf{s}_1)$ is not continuous at $\mathbf{s} = 0$ for any given $\mathbf{s}_1 \in \mathcal{S}$. For example, A2 implies that $\Gamma(\mathbf{s}_1 + \mathbf{s}, \mathbf{s}_1) = \Gamma_1(\mathbf{s}_1, \mathbf{s}_1) + \sigma_2^2(\mathbf{s}_1)$ if $\mathbf{s} = 0$, and $\Gamma_1(\mathbf{s}_1 + \mathbf{s}, \mathbf{s}_1)$ otherwise. Note that decomposition (2.3) is a convenient way, but not the only way, to model a nugget effect. In this decomposition, $\varepsilon_{1,t}(\mathbf{s})$ represents system noise which typically has continuous sample realizations (in $\mathbf{s}$), while $\varepsilon_{2,t}(\mathbf{s})$ stands for *microscale* variation and/or measurement noise; see, e.g. Cressie (1993, section 2.3.1).

Zhang *et al.* (2003) considered a parametric form $m_{\mathbf{s}}(\cdot) = g\{\cdot,\ \boldsymbol{\theta}(\mathbf{s})\}$ while $\mathbf{X}_t(\mathbf{s})$ consists of some time-lagged value of $Y_t(\mathbf{s})$. Spatio-temporal linear or nonlinear autoregressive model is a special case of (2.1) for which $\mathbf{X}_t(\mathbf{s})$ may consist of time-lagged values of not only $Y_t(\mathbf{s})$ but also $Y_t(\mathbf{i})$ for some $\mathbf{i}$ in the neighbourhood of $\mathbf{s}$.

2.2 Two estimators for $m_{\mathbf{s}}(\cdot)$

The setting presented above assumes stationarity in time t. But the process may be nonstationary over space. Under such a setting, a natural initial estimator for $m_{\mathbf{s}}(\cdot)$ will be based on the observations at location $\mathbf{s}$ only. This is essentially a time series estimation problem. The continuity of $m_{\mathbf{s}}(\cdot) = m(\mathbf{s},\ \cdot)$ invites the possibility that the quality of estimation may be improved by smoothing the time series estimators over a small neighbourhood in space. Below we outline this two-step estimation strategy for both parametric and nonparametric $m_{\mathbf{s}}(\cdot)$.

2.2.1 Parametric estimation

We first consider the case that $m_{\mathbf{s}}(\cdot) = g\{\cdot, \boldsymbol{\theta}(\mathbf{s})\}$, where the form of the smooth function g is known, $\boldsymbol{\theta}(\mathbf{s})$ is a unknown $r \times 1$ parameter vector. For a given location $\mathbf{s} \in \mathcal{S}$, we have the observations in time $\{Y_t(\mathbf{s}), \mathbf{X}_t(\mathbf{s}),\ t = 1, \cdots, T\}$. Without assuming anything about the distribution of $\varepsilon_t(\mathbf{s})$, a natural estimator for $\boldsymbol{\theta}(\mathbf{s})$ is the least squares estimator (LSE) defined as

$$\widehat{\boldsymbol{\theta}}(\mathbf{s}) = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{T} [Y_t(\mathbf{s}) - g\{\mathbf{X}_t(\mathbf{s}), \boldsymbol{\theta}\}]^2. \tag{2.4}$$

Then $\widehat{\boldsymbol{\theta}}(\mathbf{s})$ is the solution of the equation

$$\sum_{t=1}^{T} [Y_t(\mathbf{s}) - g\{\mathbf{X}_t(\mathbf{s}), \boldsymbol{\theta}\}] \dot{g}\{\mathbf{X}_t(\mathbf{s}), \boldsymbol{\theta}\} = 0, \tag{2.5}$$

where $\dot{g}(\cdot,\ \boldsymbol{\theta}) = \partial g(\cdot,\ \boldsymbol{\theta})/\partial \boldsymbol{\theta}$. We assume that

$$\widehat{\boldsymbol{\theta}}(\mathbf{s}) - \boldsymbol{\theta}(\mathbf{s}) = \{\mathcal{X}(\mathbf{s})^{\tau} \mathcal{X}(\mathbf{s})\}^{-1} \mathcal{X}(\mathbf{s})^{\tau} \boldsymbol{\varepsilon}(\mathbf{s}) + o_P(T^{-1/2}), \tag{2.6}$$

where $\mathcal{X}(\mathbf{s})$ is a $T \times r$ matrix with $\dot{g}\{\mathbf{X}_t(\mathbf{s}), \boldsymbol{\theta}(\mathbf{s})\}^\tau$ as its t-th row, $\boldsymbol{\varepsilon}(\mathbf{s}) = \{\varepsilon_1(\mathbf{s}), \cdots, \varepsilon_T(\mathbf{s})\}^\tau$. Note that (2.6) may be derived from (2.5) with additional conditions on g and the underlying distribution. Since the goal of this paper is to highlight the essence of spatial smoothing in relation to the nugget effect, we proceed with our investigation by treating (2.6) as an assumption instead of going through more detailed technical arguments. Obviously for linear g (*i.e.* $g(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{x}^\tau \boldsymbol{\theta}$), (2.6) holds with $o_P(n^{-1/2}) \equiv 0$. In fact it may also be verified for a general class of M-estimators; see, e.g. Chapter 7 of Serfling (1980).

When $\boldsymbol{\theta}(\mathbf{s})$ is continuous in $\mathbf{s}$, intuitively the estimator $\widehat{\boldsymbol{\theta}}(\mathbf{s})$ might be improved by utilizing the information from the neighbourhood locations of $\mathbf{s}$. A simple approach to achieve this is to use Nadaraya-Watson kernel smoothing over space. To this end, we assume availability of the data $\{(Y_t(\mathbf{s}_n), \mathbf{X}_t(\mathbf{s}_n)),\ t = 1, \cdots, T,\ n = 1, \cdots, N\}$. The spatial smoothing estimator at the location $\mathbf{s}_0$ is defined as

$$\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0) = \sum_{n=1}^{N} \widehat{\boldsymbol{\theta}}(\mathbf{s}_n) W_b(\mathbf{s}_n - \mathbf{s}_0) \Big/ \sum_{n=1}^{N} W_b(\mathbf{s}_n - \mathbf{s}_0), \tag{2.7}$$

where $W(\cdot) \geq 0$ is a density function on $\mathcal{R}^2$, $b > 0$ is a bandwidth, and $W_b(\cdot) = b^{-2} W(\cdot/b)$. Note that the estimator $\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0)$ applies regardless whether or not there are observations available at the location $\mathbf{s}_0$.

2.2.2 Nonparametric estimation

Suppose now $m_{\mathbf{s}}(\cdot) = m(\mathbf{s}, \cdot)$ is an unknown smooth function. To simplify our discussion, we use the local linear regression method for the initial estimation based on time series data only. This leads to the estimator $\widehat{m}(\mathbf{s}, \mathbf{x}) = \widehat{a}$, where

$$(\widehat{a}, \widehat{\mathbf{a}}_1) = \arg\min_{a, \mathbf{a}_1} \sum_{t=1}^{T} [Y_t(\mathbf{s}) - a - \mathbf{a}_1^\tau \{\mathbf{X}_t(\mathbf{s}) - \mathbf{x}\}]^2 K_h\{\mathbf{X}_t(\mathbf{s}) - \mathbf{x}\}, \tag{2.8}$$

where $K(\cdot) \geq 0$ is a density function on $\mathcal{R}^d$, $h > 0$ is a bandwidth, and $K_h(\cdot) = h^{-d} K(\cdot/h)$. It may be shown that

$$\begin{aligned} \widehat{m}(\mathbf{s}, \mathbf{x}) - m(\mathbf{s}, \mathbf{x}) &= \frac{1}{Tp(\mathbf{s}, \mathbf{x})} \sum_{t=1}^{T} \varepsilon_t(\mathbf{s}) K_h\{\mathbf{X}_t(\mathbf{s}) - \mathbf{x}\}\{1 + o_P(1)\} \\ &+ \frac{h^2}{2} \int \mathbf{y}^\tau \ddot{m}(\mathbf{s}, \mathbf{x}) \mathbf{y} K(\mathbf{y}) d\mathbf{y} \{1 + o_P(1)\}, \end{aligned} \tag{2.9}$$

where $p(\mathbf{s}, \cdot)$ denotes the probability density of $\mathbf{X}_t(\mathbf{s})$, and $\ddot{m}(\mathbf{s}, \mathbf{x}) = \frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}^\tau} m(\mathbf{s}, \mathbf{x})$.

Similar to (2.7), the spatial smoothing estimator at the location $\mathbf{s}_0$ is defined as

$$\widetilde{m}(\mathbf{s}_0, \mathbf{x}) = \sum_{n=1}^{N} \widehat{m}(\mathbf{s}_n, \mathbf{x}) W_b(\mathbf{s}_n - \mathbf{s}_0) \Big/ \sum_{n=1}^{N} W_b(\mathbf{s}_n - \mathbf{s}_0). \tag{2.10}$$

3 Variance reduction

We study the asymptotic properties of the estimators defined in section 2 above. By comparing the asymptotic variances, we will see that the spatial smoothing leads to an improvement in the estimation for $m_{\mathbf{s}}(\cdot)$ when there exists a nugget effect.

We consider the asymptotic approximations when both T and N tend to ∞. First we introduce some regularity conditions which apply to both the parametric and the nonparametric case.

A3 As $N \to \infty$, $N^{-1} \sum_{i=1}^{N} I(\mathbf{s}_i \in A) \to \int_A f(\mathbf{s}) d\mathbf{s}$ for any measurable set $A \subset \mathcal{S}$, where f is a *sampling intensity* (*i.e.* density) function on $\mathcal{S}$. Further, $f > 0$ has continuous first order partial derivatives in a neighborhood of $\mathbf{s}_0 \in \mathcal{S}$.

A4 The kernel $W(\cdot)$ is a measurable density function on $\mathcal{R}^2$ with a bounded support, and $\int \mathbf{z} W(\mathbf{z}) d\mathbf{z} = 0$. As $N \to \infty$, $b \to 0$ and $Nb^2 \to \infty$.

A5 For any $\mathbf{s} \in \mathcal{S}$, there exists a constant C_0 such that $E\|\mathbf{X}_t(\mathbf{s})\|^{2\delta} < C_0 < \infty$ for some $\delta > 2$. Further, the process $\{(Y_t(\mathbf{s}), \mathbf{X}_t(\mathbf{s})), t \geq 1\}$ is strictly stationary and α-mixing with the mixing coefficient $\alpha(k)$, independent of $\mathbf{s}$, satisfying the condition $\sum_{k=1}^{\infty} \{\alpha(k)\}^{1-2/\delta} < \infty$.

Condition A3 assumes that all the locations are within a fixed area determined by the intensity function f when $N \to \infty$. Note that N is the number of locations where the observations are taken. Our approach belongs to the category of the *fixed-domain asymptotics*. Fixed-domain asymptotics is one of two frequently used asymptotic frameworks in the analysis of spatial statistics; see, e.g. Cressie (1993, §3.3). Condition A4 is standard for kernel smoothing. Note that the number of observations used in a local estimation is of the order Nb^2 which has to tend to ∞ in any asymptotic argument. Condition A5 requires that at each location, the multiple time series $\{(Y_t, \mathbf{X}_t(\mathbf{s})), t \geq 1\}$ is strictly stationary and α-mixing. For example, linear and causal ARMA time series with continuously-distributed innovations are α-mixing with exponentially decaying mixing coefficients. For further discussion on mixing properties of time series, see section 2.6 of Fan and Yao (2003).

3.1 Parametric case

Theorem 1. Let conditions A1 – A5 hold.

(i) Suppose that (2.6) holds for $\mathbf{s} = \mathbf{s}_0$. Then as $T \to \infty$, it holds that

$$\widehat{\boldsymbol{\theta}}(\mathbf{s}_0) - \boldsymbol{\theta}(\mathbf{s}_0) = T^{-1/2}\sigma(\mathbf{s}_0)\,\mathbf{M}(\mathbf{s}_0)^{-1/2}\,\boldsymbol{\xi}_1\,\{1 + o_P(1)\},$$

where $\boldsymbol{\xi}_1 \sim N(0, \mathbf{I}_r)$, $\sigma(\mathbf{s})^2 = \mathrm{Var}\{\varepsilon_t(\mathbf{s})\}$, and $\mathbf{M}(\mathbf{s}) = E[\dot{g}\{\mathbf{X}_t(\mathbf{s}), \boldsymbol{\theta}(\mathbf{s})\}\dot{g}\{\mathbf{X}_t(\mathbf{s}), \boldsymbol{\theta}(\mathbf{s})\}^\tau]$.

(ii) Suppose on a neighbourhood of $\mathbf{s}_0$ (2.6) holds uniformly in $\mathbf{s}$ and $\boldsymbol{\theta}(\mathbf{s})$ is twice continuously differentiable. Then as $T \to \infty$ and $N \to \infty$, it holds that

$$\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0) - \boldsymbol{\theta}(\mathbf{s}_0) = b^2\mathbf{a}(\mathbf{s}_0) + T^{-1/2}\mathbf{Q}(\mathbf{s}_0)^{1/2}\,\boldsymbol{\xi}_2\,\{1 + o_P(1)\},$$

where $\boldsymbol{\xi}_2 \equiv \boldsymbol{\xi}_{2,T,N}$ is a sequence of $r \times 1$ random vectors with zero mean and identity covariance matrix, $\mathbf{a}(\mathbf{s})$ is a $r \times 1$ vector with the j-th element

$$\mathrm{tr}\Big[\Big\{\frac{1}{2}\frac{\partial^2\theta_j(\mathbf{s})}{\partial\mathbf{s}\partial\mathbf{s}^\tau} + \frac{\dot{f}(\mathbf{s})}{f(\mathbf{s})}\frac{\partial\theta_j(\mathbf{s})}{\partial\mathbf{s}^\tau}\Big\}\int \mathbf{z}\mathbf{z}^\tau W(\mathbf{z})d\mathbf{z}\Big] + o(1),$$

and

$$\mathbf{Q}(\mathbf{s}) = \sigma_1(\mathbf{s})^2\mathbf{M}(\mathbf{s})^{-1}\mathbf{M}_1(\mathbf{s})\mathbf{M}(\mathbf{s})^{-1} + \frac{\sigma_2(\mathbf{s})^2}{Nb^2 f(\mathbf{s})}\mathbf{M}(\mathbf{s})^{-1}\int W(z_1, z_2)^2 dz_1 dz_2.$$

In the above expressions, $\theta_j(\mathbf{s})$ denotes the j-th element of $\boldsymbol{\theta}(\mathbf{s})$, $\dot{f}(\mathbf{s}) = \partial f(\mathbf{s})/\partial\mathbf{s}$, $\sigma_1(\mathbf{s})^2 = \Gamma_1(\mathbf{s}, \mathbf{s})$, $\Gamma_1(\cdot, \cdot)$ and $\sigma_2(\cdot)^2$ are defined as in A2, and

$$\mathbf{M}_1(\mathbf{s}) = \lim_{\mathbf{i}\to 0} E[\dot{g}\{\mathbf{X}_t(\mathbf{s}), \boldsymbol{\theta}(\mathbf{s})\}\dot{g}\{\mathbf{X}_t(\mathbf{s} + \mathbf{i}), \boldsymbol{\theta}(\mathbf{s})\}^\tau]$$

Theorem 1(i) follows directly from the LLN and CLT for α-mixing sequences, see, *e.g.* section 2.6.3 of Fan and Yao (2003). With the assumption imposed on (2.6), the proof for (ii) is similar to the proof of Theorem 1 of Zhang *et al.* (2003).

Remark 1. (i) Since $\sigma(\mathbf{s}_0)^2 = \sigma_1(\mathbf{s}_0)^2 + \sigma_2(\mathbf{s}_0)^2$ (see condition A2), Theorem 1 implies that the asymptotic approximation for the variance matrix of $\widehat{\boldsymbol{\theta}}(\mathbf{s}_0)$ may be written as

$$\frac{1}{T}\sigma_1(\mathbf{s}_0)^2\mathbf{M}(\mathbf{s}_0)^{-1} + \frac{1}{T}\sigma_2(\mathbf{s}_0)^2\mathbf{M}(\mathbf{s}_0)^{-1},$$

while the approximation for the variance of $\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0)$ is of the form

$$\frac{1}{T}\sigma_1(\mathbf{s}_0)^2\mathbf{M}(\mathbf{s}_0)^{-1}\mathbf{M}_1(\mathbf{s}_0)\mathbf{M}(\mathbf{s}_0)^{-1} + \frac{1}{T}\sigma_2(\mathbf{s}_0)^2\mathbf{M}(\mathbf{s}_0)^{-1}\frac{1}{Nb^2 f(\mathbf{s}_0)}\int W(z_1, z_2)^2 dz_1 dz_2.$$

Due to possible nugget effect in $\mathbf{X}_t(\mathbf{s})$, $\mathbf{M}_1(\mathbf{s}_0)$ may not be equal to $\mathbf{M}(\mathbf{s}_0)$. If this is the case, typically $\mathbf{M}(\mathbf{s}_0) - \mathbf{M}_1(\mathbf{s}_0)$ is a positive semi-definite matrix. One such example is the linear autoregressive model considered in Zhang *et al.* (2003). Hence the asymptotic variance of $\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0)$ is smaller than that of $\widehat{\boldsymbol{\theta}}(\mathbf{s}_0)$ as long as $\sigma_2^2(\mathbf{s}_0) \neq 0$ and $Nb^2 \to \infty$, *i.e.* the spatial smoothing reduces the variance of the estimation. However in the case of no nugget effect (*i.e.* $\sigma_2^2(\mathbf{s}_0) = 0$ and $\mathbf{M}_1(\mathbf{s}_0) = \mathbf{M}(\mathbf{s}_0)$), no variance reduction may be obtained via spatial smoothing as the asymptotic variances of the two estimators are identical then. This is due to the fact that the spatial smoothing uses effectively the data at locations within distance b from $\mathbf{s}_0$. Due to the continuity of the function $\Gamma_1(\cdot,\ \cdot)$ stated in A2, the variogram $E[\{\varepsilon_t(\mathbf{s}) - \varepsilon_t(\mathbf{s}_0)\}^2] \to 0$ as $\mathbf{s} \to \mathbf{s}_0$. Hence all the $\varepsilon_t(\mathbf{s})'$s from those locations are asymptotically identical as b shrinks to 0. We argue that asymptotic theory under this setting presents an excessively gloomy picture. Adding a nugget effect in the model brings the theory closer to reality since in practice the data used in local spatial smoothing usually contain some noise components which are not identical even within a very small neighborhood. Note that the nugget effect is hardly detectable in practice since we can never estimate $\Gamma(\mathbf{s}+\Delta,\ \mathbf{s})$ defined in (2.2) for $||\Delta||$ less than the minimum pairwise-distance among observed locations.

(ii) Theorem 1(ii) indicates that to minimize the asymptotic mean squared error of the estimator $\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0)$, we should use bandwidth b of the order $(NT)^{-1/6}$. Note that $\text{Var}\{\varepsilon_t(\mathbf{s}_0)\} = \sigma_1(\mathbf{s}_0)^2 + \sigma_2(\mathbf{s}_0)^2$. With $b = O\{(NT)^{-1/6}\}$, the asymptotic mean squared error of the smoothed estimator $\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0)$ is smaller than that of $\widehat{\boldsymbol{\theta}}(\mathbf{s}_0)$ under the condition $T = o(N^2)$. Furthermore, the smaller is $\sigma_1^2(\mathbf{s}_0)/\sigma_2^2(\mathbf{s}_0)$ (the system-noise-to-measurement-noise ratio), the larger is the improvement due to spatial smoothing. In particular, if $\sigma_1^2(\mathbf{s}_0) = 0$, the mean squared error of $\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0)$ is an order of magnitude smaller than that of the method using the data at location $\mathbf{s}_0$ only.

(iii) Without stationarity over the space, we cannot establish the asymptotic normality for the smoothing estimator $\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0)$.

3.2 Nonparametric case

Suppose that $m_{\mathbf{s}}(\mathbf{x}) = m(\mathbf{s}, \mathbf{x})$ is unknown but $\frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}^\tau} m(\mathbf{s}, \mathbf{x})$ is continuous in $\mathbf{x}$. Similar to the parametric case considered above, the property of the regressor process $\{\mathbf{X}_t(\mathbf{s}),\ \mathbf{s} \in \mathcal{S}\}$ also plays an important role in relation to the asymptotic properties of the estimators for $m(\mathbf{s}, \cdot)$. If there is a nugget effect in $\{\mathbf{X}_t(\mathbf{s}),\ \mathbf{s} \in \mathcal{S}\}$, the joint density function of $\mathbf{X}_t(\mathbf{s})$ and $\mathbf{X}_t(\mathbf{i})$, denoted by

$p(\mathbf{s}, \mathbf{i}; \mathbf{x}, \mathbf{y})$, may admit a well-defined limit:

$$\lim_{\mathbf{s}\to\mathbf{s}_0,\, \mathbf{i}\to\mathbf{s}_0} p(\mathbf{s}, \mathbf{i}; \mathbf{x}, \mathbf{y}) = q(\mathbf{s}_0; \mathbf{x}, \mathbf{y}), \tag{3.1}$$

where $q(\mathbf{s}_0; \mathbf{x}, \mathbf{y})$ is continuous in both $\mathbf{x}$ and $\mathbf{y}$. Examples for which such a limit exists include a simple AR(1) model $Y_t(\mathbf{s}) = m\{X_t(\mathbf{s})\} + \varepsilon_t(\mathbf{s})$ with $X_t(\mathbf{s}) = Y_{t-1}(\mathbf{s})$ and $\varepsilon_t(\mathbf{s})$ satisfying conditions A1 and A2, or simply $\mathbf{X}_t(\mathbf{s}) = \mathbf{X}_{t1}(\mathbf{s}) + \mathbf{X}_{t2}(\mathbf{s})$, where $\mathbf{X}_{t1}(\mathbf{s})$ has continuous sample paths over space while $\mathbf{X}_{t2}(\mathbf{s})$, for different $\mathbf{s}$, are i.i.d.. For the latter, $q(\mathbf{s}_0; \mathbf{x}, \mathbf{y})$ is the joint probability density function of $\mathbf{X}_t(\mathbf{s}_0)$ and $\mathbf{X}_{t1}(\mathbf{s}_0) + \mathbf{X}_{t2}(\mathbf{s})$ for any $\mathbf{s} \neq \mathbf{s}_0$.

On the other hand, if there exists no nugget effect in $\{\mathbf{X}_t(\mathbf{s}),\ \mathbf{s} \in \mathcal{S}\}$ in the sense

$$E[||\mathbf{X}_t(\mathbf{s}) - \mathbf{X}_t(\mathbf{i})||^2] \to 0, \quad \text{as } ||\mathbf{s} - \mathbf{i}|| \to 0,$$

the joint density function $p(\mathbf{s}, \mathbf{i}; \mathbf{x}, \mathbf{y})$ is subject to the irregular behavior dictated by the limits

$$\int p(\mathbf{s}, \mathbf{i}; \mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \to 1 \quad \text{and} \quad \int ||\mathbf{x} - \mathbf{y}||^2 p(\mathbf{s}, \mathbf{i}; \mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \to 0 \quad \text{as } ||\mathbf{s} - \mathbf{i}|| \to 0.$$

Therefore $p(\mathbf{s}, \mathbf{i}; \mathbf{x}, \mathbf{y})$ exhibits a ridge along the line $\mathbf{x} = \mathbf{y}$ when $||\mathbf{s} - \mathbf{i}||$ is small. Furthermore, $p(\mathbf{s}, \mathbf{i}; \mathbf{x}, \mathbf{y})$ may diverge to ∞ when $\mathbf{x} = \mathbf{y}$, and 0 otherwise as $||\mathbf{s} - \mathbf{i}|| \to 0$. In this case, the two-dimensional distribution of $\{\mathbf{X}_t(\mathbf{s}), \mathbf{X}_t(\mathbf{i})\}$ degenerates to a one-dimensional distribution when $||\mathbf{s} - \mathbf{i}|| \to 0$. Hence it is reasonable to assume that for any measurable $\psi(\cdot)$, $E[\psi\{\mathbf{X}_t(\mathbf{s})\}|\mathbf{X}_t(\mathbf{i}) = \mathbf{x}] \to \psi(\mathbf{x})$, as $||\mathbf{s} - \mathbf{i}|| \to 0$. To facilitate an asymptotic approximation, we assume that for any $||\mathbf{i}_k - \mathbf{s}_0|| \leq Cb$ $(k = 1, 2)$, where $C > 0$ is a constant, it holds almost surely that

$$\left| E[K_h\{\mathbf{X}_t(\mathbf{i}_1) - \mathbf{x}\} \big| \mathbf{X}_t(\mathbf{i}_2)] \Big/ K_h\{\mathbf{X}_t(\mathbf{i}_2) - \mathbf{x}\} - 1 \right| \to 0, \tag{3.2}$$

as $N, T \to \infty$ and $b, h \to 0$ (see conditions A4 and A6). This condition may be justified by imposing appropriate assumptions on the manner in which the conditional distribution of $\mathbf{X}_t(\mathbf{i}_1)$ given $\mathbf{X}_t(\mathbf{i}_2)$ degenerates and on the relative speeds at which both b and h converge to 0.

Now we introduce an additional regularity condition.

A6 The kernel $K(\cdot)$ is a density function on $\mathcal{R}^d$ with bounded support, and $\int \mathbf{y} K(\mathbf{y}) d\mathbf{y} = 0$. As $T \to \infty$, $h \to 0$ and $Th^d \to \infty$.

Theorem 2. Let conditions A1 – A6 holds, and $\mathbf{x} \in \mathcal{R}^d$ be fixed with $p(\mathbf{s}_0, \mathbf{x}) > 0$, where $p(\mathbf{s}, \cdot) > 0$ denotes the marginal density function of $\mathbf{X}_t(\mathbf{s})$.

(i) As $T \to \infty$, it holds that

$$\widehat{m}(\mathbf{s}_0, \mathbf{x}) - m(\mathbf{s}_0, \mathbf{x}) = h^2 \mu_1(\mathbf{s}_0, \mathbf{x}) + (Th^d)^{-1/2} \nu_1(\mathbf{s}_0, \mathbf{x})\, \xi_3 \{1 + o_P(1)\},$$

where $\xi_3 \equiv \xi_{3,T}$ is a sequence of random variables with mean 0 and variance 1, and

$$\mu_1(\mathbf{s}, \mathbf{x}) = \frac{1}{2} \mathrm{tr}\{\ddot{m}(\mathbf{s}, \mathbf{x}) \int \mathbf{y}\mathbf{y}^\tau K(\mathbf{y}) d\mathbf{y}\} + o_P(1), \quad \nu_1(\mathbf{s}, \mathbf{x})^2 = \sigma(\mathbf{s})^2 \int K(\mathbf{y})^2 d\mathbf{y} / p(\mathbf{s}, \mathbf{x}).$$

(ii) Suppose on a neighbourhood of s_0 (2.9) holds uniformly in $\mathbf{s}$, and $\frac{\partial^2}{\partial \mathbf{s} \partial \mathbf{s}^\tau} m(\mathbf{s}, \mathbf{x})$ and $p(\mathbf{s}, \mathbf{x})$ are uniformly continuous in both $\mathbf{s}$ and $\mathbf{x}$.

(a) Under condition (3.1), it holds that as $T \to \infty$ and $N \to \infty$,

$$\begin{aligned} \widetilde{m}(\mathbf{s}_0, \mathbf{x}) - m(\mathbf{s}_0, \mathbf{x}) &= h^2 \mu_1(\mathbf{s}_0, \mathbf{x}) + b^2 \mu_2(\mathbf{s}_0, \mathbf{x}) \\ &+ (Th^d)^{-1/2} \nu_2(\mathbf{s}_0, \mathbf{x})\, \xi_4 \{1 + o_P(1)\}, \end{aligned} \quad (3.3)$$

where $\xi_4 \equiv \xi_{4,T,N}$ is a sequence of random variables with mean 0 and variance 1, and

$$\mu_2(\mathbf{s}, \mathbf{x}) = \mathrm{tr}\Big[\Big\{\frac{1}{2} \frac{\partial^2 m(\mathbf{s}, \mathbf{x})}{\partial \mathbf{s} \partial \mathbf{s}^\tau} + \frac{\dot{f}(\mathbf{s})}{f(\mathbf{s})} \frac{\partial m(\mathbf{s}, \mathbf{x})}{\partial \mathbf{s}^\tau}\Big\} \int \mathbf{z}\mathbf{z}^\tau w(\mathbf{z}) d\mathbf{z}\Big] + o(1),$$

$$\nu_2(\mathbf{s}, \mathbf{x})^2 = h^d \sigma_1(\mathbf{s})^2 \frac{q(\mathbf{s}; \mathbf{x}, \mathbf{x})}{p(\mathbf{s}, \mathbf{x})^2} + \frac{1}{Nb^2} \frac{\sigma(\mathbf{s})^2}{p(\mathbf{s}, \mathbf{x}) f(\mathbf{s})} \int w(\mathbf{z})^2 d\mathbf{z} \int K(\mathbf{y})^2 d\mathbf{y}.$$

(b) Under condition (3.2), (3.3) still holds but with $\nu_2(\mathbf{s}_0, \mathbf{x})$ replaced by $\nu_3(\mathbf{s}_0, \mathbf{x})$, where

$$\nu_3(\mathbf{s}, \mathbf{x})^2 = \sigma_1(\mathbf{s})^2 \int K(\mathbf{y})^2 d\mathbf{y} / p(\mathbf{s}, \mathbf{x}).$$

Remark 2. (i) The condition (3.3) indicates that the variance of the estimation for $m(\mathbf{s}_0, \mathbf{x})$ may be reduced by the spatial smoothing provided there exists a nugget effect in the process of $\mathbf{X}_t(\mathbf{s})$ in the sense of (3.1) and $Nb^2 = O(h^{-d})$. In fact the convergence rate for $\widetilde{m}$ is $\sqrt{T}$ while that for $\widehat{m}$ is merely $\sqrt{Th^d}$. Since the only difference between $\nu_3(\mathbf{s}, \mathbf{x})^2$ and $\nu_1(\mathbf{s}, \mathbf{x})^2$ is that $\sigma_1(\mathbf{s})^2$ in the definition of $\nu_3(\mathbf{s}, \mathbf{x})^2$ is replaced by $\sigma(\mathbf{s})^2 = \sigma_1(\mathbf{s})^2 + \sigma_2(\mathbf{s})^2$, the variance of the estimator is reduced by the spatial smoothing if there is a nugget effect in the noise process $\varepsilon_t(\mathbf{s})$ (*i.e.* $\sigma_2(\mathbf{s}_0) > 0$) but not in the process of $\mathbf{X}_t(\mathbf{s})$, although the convergence rates for the two estimators are the same. On the other hand, there is no variance reduction in case that there is no nugget effect in both the noise and regressor processes.

(ii) With some additional regularity conditions, it may be proved that $\xi_{3,T}$ converges in distribution to an $N(0,1)$ random variable. See, for example, section 2.6.4 of Fan and Yao (2003). However without the stationarity over the space, ξ_4 is not necessarily normal.

Proof of Theorem 2. We only sketch the proof for (ii), as the proof for (i) is similar and less involved.

Condition A3 implies $\sum_{1\le n\le N} W_b(s_n - s_0) \sim Nf(\mathbf{s}_0)$. Note that $W(\cdot)$ has a bounded support. It follows from the uniform convergence assumption on (2.9) that

$$\begin{aligned}
\widetilde{m}(\mathbf{s}_0,\mathbf{x}) - m(\mathbf{s}_0,\mathbf{x}) &= \frac{1}{NTf(\mathbf{s}_0)}\sum_{n=1}^N \frac{1}{p(\mathbf{s}_n,\mathbf{x})}\sum_{t=1}^T \varepsilon_t(\mathbf{s}_n)K_h\{\mathbf{x}_t(\mathbf{s}_n)-\mathbf{x}\}W_b(\mathbf{s}_n-\mathbf{s}_0)\{1+o_P(1)\}\\
&+ \frac{h^2}{2Nf(\mathbf{s}_0)}\sum_{n=1}^N \int \mathbf{y}^\tau \ddot{m}(\mathbf{s}_n,\mathbf{x})\mathbf{y}K(\mathbf{y})d\mathbf{y}W_b(\mathbf{s}_n-\mathbf{s}_0)\{1+o_P(1)\},\\
&+ \frac{1}{Nf(\mathbf{s}_0)}\sum_{n=1}^N\{m(\mathbf{s}_n,\mathbf{x})-m(\mathbf{s}_0,\mathbf{x})\}W_b(\mathbf{s}_n-\mathbf{s}_0)\{1+o(1)\}\\
&\equiv I_1\{1+o_P(1)\} \;+\; I_2\{1+o_P(1)\} \;+\; I_3.
\end{aligned}$$

Again it follows from condition A3 that

$$I_2 \sim \frac{h^2}{2}\int \mathbf{y}^\tau \ddot{m}(\mathbf{s}_0,\mathbf{x})\mathbf{y}K(\mathbf{y})d\mathbf{y}, \tag{3.4}$$

$$I_3 \sim b^2 \int \Big\{\frac{1}{2}\mathbf{z}^\tau \frac{\partial^2 m(\mathbf{s},\mathbf{x})}{\partial\mathbf{s}\partial\mathbf{s}^\tau}\Big|_{\mathbf{s}=\mathbf{s}_0}\mathbf{z} + \mathbf{z}^\tau\frac{\dot{f}(\mathbf{s}_0)}{f(\mathbf{s}_0)}\frac{\partial m(\mathbf{s},\mathbf{x})}{\partial\mathbf{s}^\tau}\Big|_{\mathbf{s}=\mathbf{s}_0}\mathbf{z}\Big\}W(\mathbf{z})d\mathbf{z}. \tag{3.5}$$

Note that $EI_1 = 0$. Let $\xi_3 = I_1/\sqrt{E(I_1^2)}$. In view of (3.4) and (3.5), Theorem 2(ii) holds if we may show $E(I_1^2) \sim \nu_2(\mathbf{s}_0,\mathbf{x})^2/(Th^d)$ under (3.1), and $E(I_1^2) \sim \nu_3(\mathbf{s}_0,\mathbf{x})^2/(Th^d)$ under condition (3.2). We prove these two asymptotic results below.

It follows from condition A1 that

$$\begin{aligned}
EI_1^2 &= \frac{1}{N^2T^2f(\mathbf{s}_0)^2}\sum_{n=1}^N \frac{1}{p(\mathbf{s}_n,\mathbf{x})^2}\sum_{t=1}^T \sigma(\mathbf{s}_n)^2 W_b(\mathbf{s}_n-\mathbf{s}_0)^2 E[K_h\{\mathbf{X}_t(\mathbf{s}_n)-\mathbf{x}\}^2]\\
&+ \frac{1}{N^2T^2f(\mathbf{s}_0)^2}\sum_{1\le n\ne j\le N}\frac{1}{p(\mathbf{s}_n,\mathbf{x})p(\mathbf{s}_j,\mathbf{x})}\sum_{t=1}^T \Gamma(\mathbf{s}_n,\mathbf{s}_j)W_b(\mathbf{s}_n-\mathbf{s}_0)W_b(\mathbf{s}_j-\mathbf{s}_0)\mathcal{C}(\mathbf{s}_n,\mathbf{s}_j)\\
&= \frac{1}{N^2Th^df(\mathbf{s}_0)^2}\sum_{n=1}^N\frac{1}{p(\mathbf{s}_n,\mathbf{x})}\sigma(\mathbf{s}_n)^2W_b(\mathbf{s}_n-\mathbf{s}_0)^2\int K(\mathbf{y})^2d\mathbf{y}\{1+o(1)\}\\
&+ \frac{1}{N^2Tf(\mathbf{s}_0)^2}\sum_{1\le n\ne j\le N}\frac{1}{p(\mathbf{s}_n,\mathbf{x})p(\mathbf{s}_j,\mathbf{x})}\Gamma(\mathbf{s}_n,\mathbf{s}_j)W_b(\mathbf{s}_n-\mathbf{s}_0)W_b(\mathbf{s}_j-\mathbf{s}_0)\mathcal{C}(\mathbf{s}_n,\mathbf{s}_j)\\
&\equiv I_{11} + I_{12},
\end{aligned} \tag{3.6}$$

where $\mathcal{C}(\mathbf{s}_n, \mathbf{s}_j) = E[\{K_h\{\mathbf{X}_t(\mathbf{s}_n) - \mathbf{x}\}\{K_h\{\mathbf{X}_t(\mathbf{s}_j) - \mathbf{x}\}]$. By condition A3,

$$I_{11} \sim \frac{1}{NTh^d b^2} \frac{\sigma(\mathbf{s}_0)^2}{f(\mathbf{s}_0)p(\mathbf{s}_0, \mathbf{x})} \int W(\mathbf{z})^2 d\mathbf{z} \int K(\mathbf{y})^2 d\mathbf{y}. \tag{3.7}$$

Under condition (3.1),

$$\begin{aligned} I_{12} &\sim \frac{1}{N^2 T f(\mathbf{s}_0)^2} \sum_{1 \le n \ne j \le N} \frac{1}{p(\mathbf{s}_n, \mathbf{x}) p(\mathbf{s}_j, \mathbf{x})} \Gamma(\mathbf{s}_n, \mathbf{s}_j) W_b(\mathbf{s}_n - \mathbf{s}_0) W_b(\mathbf{s}_j - \mathbf{s}_0) p(\mathbf{s}_n, \mathbf{s}_j; \mathbf{x}, \mathbf{x}) \\ &\sim \frac{1}{T} \sigma_1(\mathbf{s}_0)^2 q(\mathbf{s}_0; \mathbf{x}, \mathbf{x}) / p(\mathbf{s}_0, \mathbf{x})^2. \end{aligned}$$

This, together with (3.7), implies $Th^d E(I_1^2) \sim \nu_2(\mathbf{s}_0, \mathbf{x})^2$.

On the other hand, condition (3.2) implies

$$\begin{aligned} \mathcal{C}(\mathbf{s}_n, \mathbf{s}_j) &= E\big\{K_h\{\mathbf{X}_t(\mathbf{s}_n) - \mathbf{x}\} E[K_h\{\mathbf{X}_t(\mathbf{s}_j) - \mathbf{x}\} | \mathbf{X}_t(\mathbf{s}_n)]\big\} \sim E[K_h\{\mathbf{X}_t(\mathbf{s}_n) - \mathbf{x}\}^2] \\ &\sim h^{-d} p(\mathbf{s}_n, \mathbf{x}) \int K(\mathbf{y})^2 d\mathbf{y} \end{aligned}$$

for all $\mathbf{s}_n, \mathbf{s}_j$ with the distances of order b from $\mathbf{s}_0$. Hence

$$\begin{aligned} I_{12} &\sim \frac{1}{N^2 T h^d f(\mathbf{s}_0)^2} \sum_{1 \le n \ne j \le N} \frac{1}{p(\mathbf{s}_n, \mathbf{x})} \Gamma(\mathbf{s}_n, \mathbf{s}_j) W_b(\mathbf{s}_n - \mathbf{s}_0) W_b(\mathbf{s}_j - \mathbf{s}_0) \int K(\mathbf{y})^2 d\mathbf{y} \\ &\sim \frac{1}{Th^d} \frac{\sigma_1(\mathbf{s}_0)^2}{p(\mathbf{s}_0, \mathbf{x})} \int K(\mathbf{y})^2 d\mathbf{y}. \end{aligned}$$

This implies $Th^d E(I_1^2) \sim Th^d I_{12} \sim \nu_3(\mathbf{s}_0, \mathbf{x})^2$, as now $I_{11} = o(I_{12})$. This completes the proof.

References

Cressie, N.A.C. (1993). *Statistics for Spatial Data*. Wiley, New York.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: nonparametric and parametric methods*. Springer, New York.

Lu, Z., Steinskog, D.J., Tjøstheim, D. and Yao, Q. (2007). Adaptive varying coefficient spatial-temporal models. *Available at* http://stats.lse.ac.uk/q.yao/qyao.links/technicalreports.html.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Stein, M.L. (1999). *Interpolation of Spatial Data*. Springer, New York.

Yao, Q. and Brockwell, P.J. (2006). Gaussian maximum likelihood estimation for ARMA models II: spatial processes. *Bernoulli*, **12**, 403-429.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, **99**, 250-261.

Zhang, W., Yao, Q., Tong, H. and Stenseth, N.C. (2003). Smoothing for spatio-temporal models and its application to modeling muskrats-mink interaction. *Biometrics*, **59**, 813-821.