Smoothing for Spatio-Temporal Models and its Application in Modelling Muskrat-Mink Interaction *

Wenyang Zhang

Institute of Mathematics and Statistics

University of Kent

Canterbury, Kent CT2 7NF, UK

Qiwei Yao

Department of Statistics

London School of Economics

London WC2A 2AE, UK

Howell Tong

Department of Statistics

London School of Economics, and

Department of Statistics & Actuarial Science

Hong Kong University, Hong Kong

Nils Chr. Stenseth

Department of Biology

University of Oslo

N-0316 Oslo, Norway

Abstract

For a set of spatially dependent dynamical models, we propose to estimate parameters which control temporal dynamics by spatial smoothing. The new approach is particularly relevant for analysing spatially distributed panels of short time series over space. The asymptotic results show that spatial smoothing will improve the estimation

*This research was partially supported by BBSRC/EPSRC Grant 96-MMI09785.

in the presence of nugget effect even when the sample size in each location is large. The proposed methodology is used to analyse the annual mink and muskrat data collected in a period of 25 years over 81 locations in Canada. Based on the proposed method, we are able to model the temporal dynamics which reflects the food-chain-interaction of the two species.

KEY WORDS: α-mixing; Canadian muskrat and mink data; Fixed-domain asymptotics; Food chain interaction; Local linear regression; Nugget effect; Spatial smoothing; Spatio-temporal model; Threshold model; Time series.

1 Introduction

In the context of purely spatial data analysis, *kriging* is one of the most frequently used methods; it is typically used for the prediction of a random variable from neighbouring observations by exploiting the statistical dependence between these observations and the unknown random variable (see e.g. Chapter 3, Cressie, 1993). Unfortunately this approach is no longer pertinent when we deal with spatio-temporal data and when we are particularly interested in estimating the temporal dynamics. In fact, we argue that it will be difficult, if ever possible, to estimate the temporal dynamics by relying on an approach which focuses on prediction. In this paper, we assume that (i) the temporal dynamics follows a certain known structure governed by an unknown parameter vector which varies smoothly over space, and (ii) the spatial dynamics is driven by a series of noise processes which are dependent over space but independent over time. This setting can be viewed as an extension of Hjellvik and Tjøstheim (1999), in which they dealt with a panel of dependent linear time series and accommodated the (spatial) dependence in terms of a common noise term across the panel.

We shall propose a nonlinear spatio-temporal model approach, in which the the functional form is assumed known up to some unknown parameters. We estimate the unknown parameters by spatial smoothing technique through local linear kernel regression. A spin-off of the proposed spatial smoothing is that, in the presence of the *nugget effect* (see e.g. p.59, Cressie, 1993), even the estimation at locations at which observations are available can be further improved .

Spatial smoothing is one of the very frequently used ideas in analysing spatial data especially for spatial interpolation, although the form of smoothing is diverse. Kernel smoothing was used to estimate the intensity of the spatial pattern by Diggle (1985) and Diggle and Marron (1988), marginal density functions in Hallin, Lu and Tran (2002) and Yao (2003). The approach proposed in this paper is motivated by modelling the food chain interaction between mink and muskrat in Canada. The available data consist of the annual numbers of both mink and muskrat fur sales at 81 posts over a period of 25 years, namely 1925 – 1949. In short, we have 81×2 time series, but each consisting of 25 points only. Most of these series exhibit cycles with a period of around 10 years. It is well-known that for many parts of Canada (such as those covered by boreal forests) there exists a close food-chain-interaction between the mink (as predator) and the muskrat (as prey). Biological studies suggest that the food-chain-interaction should be nonlinear, reflecting the changing behaviour of the animals at different stages of the population cycle. The statistical analysis of this particular data set aims at a deeper understanding of the food-chain-interaction from a quantitative point of view. Thus, we seek a statistical model which can capture much of the temporal dynamical fluctuation and interaction of the two species as well as the pattern change in the fluctuation and the interaction over space. Briefly, our approach is as follows. We adopt the

food-chain-interaction idea proposed by May (1981) and Stenseth *et al.* (1997) and couple it with the idea of regime change dictated by a threshold variable (Tong 1990) to arrive at a model for the temporal dynamics at each of the 81 posts. We clothe the resulting temporal model for each post with spatially dependent noise to reflect spatial dependence in the data. By pooling the information from neighbouring posts, we overcome the difficulties in estimation due to the shortage of data at each post. Further, we are able to identify the main factors which dominate the pattern change of the food-chain-interaction over space using the estimated models.

The paper is organised as follows. In section 2, we present a local linear approach for the estimation of spatially dependent parameters in a fairly general spatio-temporal model setup. In section 3, we explain why it is practically relevant to include the nugget effect in our setting, and present the asymptotic properties of the proposed estimators in the form of the *fixed-domain asymptotics*. We conduct some simulation in Section 4, which illustrates the improvement due to spatial smoothing. The analysis on mink and muskrat interaction over time and space based on the proposed method is reported in Section 5. The technical proof of our main result is given in an appendix.

2 Methodology

2.1 Model

Suppose $\{(Y_t(\mathbf{s}), \mathbf{X}_t(\mathbf{s})), t = 1, 2, \cdots\}$ is a strictly stationary process indexed by a *location* variable $\mathbf{s} \equiv (u, v)^\tau \in \mathcal{S}$, where $Y_t(\mathbf{s})$ is a scalar, $\mathbf{X}_t(\mathbf{s})$ is a $p \times 1$ vector, and $\mathcal{S}$ is a subset of R^2. In a time series context, $\mathbf{X}_t(\mathbf{s})$ may include some lagged values of $Y_t(\mathbf{s})$. At each fixed

location, we assume that

$$Y_t(\mathbf{s}) = g(\mathbf{X}_t(\mathbf{s}), \boldsymbol{\theta}(\mathbf{s})) + \varepsilon_t(\mathbf{s}), \quad t = 1, 2, \cdots, \tag{2.1}$$

where the form of function g is given and $\boldsymbol{\theta}(\mathbf{s})$ is an unknown parameter vector and is continuous in $\mathbf{s}$. We assume that the noise processes $\varepsilon_t(\mathbf{s})$ satisfy the condition below.

(C1) $\{\varepsilon_1(\mathbf{s}), \mathbf{s} \in \mathcal{S}\}$, $\{\varepsilon_2(\mathbf{s}), \mathbf{s} \in \mathcal{S}\}, \cdots$ is a sequence of independent spatial processes with identical distribution. 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 correlation 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 $\mathcal{S}^2$.

We further assume that the noise $\varepsilon_t(\mathbf{s})$ admits the decomposition below.

(C2) 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 $\{\varepsilon_{1,t}(\mathbf{s})\}$ and $\{\varepsilon_{2,t}(\mathbf{s})\}$ are two independent processes, and both fulfil conditions imposed on $\{\varepsilon_t(\mathbf{s})\}$ in (C1) above. Further, $\Gamma_1(\mathbf{s} + \Delta\mathbf{s}, \mathbf{s}) \equiv \mathrm{Cov}(\varepsilon_{1,t}(\mathbf{s} + \Delta\mathbf{s}), \varepsilon_{1,t}(\mathbf{s}))$ is continuous, and $\mathrm{Cov}(\varepsilon_{2,t}(\mathbf{s} + \Delta\mathbf{s}), \varepsilon_{2,t}(\mathbf{s})) = \sigma_0^2$ if $\Delta\mathbf{s} = 0$, and 0 otherwise.

Condition (C2) 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 B. Matheron in early 1960's. It reflects the fact that the variogram of a spatial process does not converg to 0 at the origin. In our notation,

it is equivalent to the fact that the function $\Gamma(.)$ defined as in (2.2) is not continuous at 0. In fact, it follows from (C2) that $\Gamma(\mathbf{s}_0+\Delta\mathbf{s},\mathbf{s}_0)=\Gamma_1(\mathbf{s}_0,\mathbf{s}_0)+\sigma_0^2$ for $\Delta\mathbf{s}=0$, and $\Gamma_1(\mathbf{s}_0+\Delta\mathbf{s},\mathbf{s}_0)$ otherwise. In the decomposition (2.3), $\varepsilon_{1,t}(\mathbf{s})$ represents system noise which typically has continuous sample path (in $\mathbf{s}$), while $\varepsilon_{2,t}(\mathbf{s})$ stands for *microscale* variation and/or measurement noise; see, e.g. Cressie (1993, esp. §2.3.1) and also Remark (b) in Section 3 below. Hjellvik and Tjøstheim (1999) adopted a similar, but different, noise decomposition to model the dependence in panels of time series data.

2.2 Estimation

Suppose that we have observations at N locations denoted by $\mathcal{S}_N \equiv \{\mathbf{s}_1,\cdots,\mathbf{s}_N\} \subset \mathcal{S}$. For each location, we have observed data $\{(Y_t(\mathbf{s}),\mathbf{X}_t(\mathbf{s})), 1 \le t \le T\}$. The goal is to estimate $\boldsymbol{\theta}(\mathbf{s}_0)$ for a given location $\mathbf{s}_0 \in \mathcal{S}$. If we have observations at $\mathbf{s}_0$ (*i.e.* $\mathbf{s}_0 \in \mathcal{S}_N$), the least squares estimator $\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0)$ for $\boldsymbol{\theta}(\mathbf{s}_0)$ based on the data at the location $\mathbf{s}_0$ is the solution of minimising the sum $\sum_{t=1}^{T}\{Y_t(\mathbf{s}_0) - g(\mathbf{X}_t(\mathbf{s}_0),\mathbf{a})\}^2$ over vector $\mathbf{a}$. When $\mathbf{s}_0 \notin \mathcal{S}_N$, we may estimate $\boldsymbol{\theta}(\mathbf{s}_0)$ by pooling the information from observations at nearby locations. This leads us to consider the local linear estimator $\widehat{\boldsymbol{\theta}}(\mathbf{s}_0) \equiv \hat{\mathbf{a}}$, where $(\hat{\mathbf{a}},\widehat{\mathbf{B}})$ is the minimiser of

$$\sum_{\mathbf{S}\in\mathcal{S}_N}\sum_{t=1}^{T}\{Y_t(\mathbf{s}) - g(\mathbf{X}_t(\mathbf{s}),\mathbf{a}) - \dot{g}(\mathbf{X}_t(\mathbf{s}),\mathbf{a})^\tau\mathbf{B}(\mathbf{s}-\mathbf{s}_0)\}^2 K_h(\mathbf{s}-\mathbf{s}_0). \tag{2.4}$$

Here $\dot{g}(\mathbf{x},\mathbf{a}) = (\frac{\partial}{\partial\mathbf{a}})g(\mathbf{x},\mathbf{a})$, $\mathbf{B}$ is a matrix, $K(\cdot)$ is a kernel function, and $K_h(\mathbf{s}) = h^{-2}K(\mathbf{s}/h)$. Obviously, $\widehat{\mathbf{B}}$ is an estimator for $\dot{\boldsymbol{\theta}}(\mathbf{s}_0)$, where $\dot{\boldsymbol{\theta}}(\mathbf{s}) = (\frac{\partial}{\partial\mathbf{S}})\boldsymbol{\theta}(\mathbf{s})$. For linear model $g(\mathbf{x},\boldsymbol{\theta}) = \mathbf{x}^\tau\boldsymbol{\theta}$, the above sum of squares reduces to

$$\sum_{\mathbf{S}\in\mathcal{S}_N}\sum_{t=1}^{T}\{Y_t(\mathbf{s}) - \mathbf{X}_t^\tau(\mathbf{s})(\mathbf{a}+\mathbf{B}(\mathbf{s}-\mathbf{s}_0))\}^2 K_h(\mathbf{s}-\mathbf{s}_0).$$

When $\mathbf{s}_0 \in \mathcal{S}_N$, the estimator $\widehat{\boldsymbol{\theta}}(\mathbf{s}_0)$ based on the combined data from neighbouring locations has a smaller variance than the ordinary least squares estimator $\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0)$ based on

the data at the location $\mathbf{s}_0$ only; see Remark (a) below. From now on, we call $\widehat{\boldsymbol{\theta}}(\mathbf{s}_0)$ the *smoothed estimator* and $\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0)$ the *unsmoothed estimator*.

2.3 Bandwidth selection

The bandwidth h plays a crucial role in kernel smoothing. For a comprehensive discussion on bandwidth selection, we refer to Fan and Gijbels (1996, Ch. 4) and Simonoff (1996, Ch. 5). In this paper, we adopt the generalised cross-validation method proposed by Wahba (1977) and Craven and Wahba (1979), to select h.

Let $\widehat{Y}_{t_0}(\mathbf{s}_0) = g(\mathbf{X}_{t_0}(\mathbf{s}_0), \widehat{\boldsymbol{\theta}}(\mathbf{s}_0))$, where $\widehat{\boldsymbol{\theta}}(\mathbf{s}_0)$ is the local linear estimator derived from (2.4). Then $\widehat{Y}_{t_0}(\mathbf{s}_0)$ can be written as a linear combination of $\{Y_t(\mathbf{s}), 1 \le t \le T, \mathbf{s} \in \mathcal{S}_N\}$ with coefficients depending on $\{\mathbf{X}_t(\mathbf{s})\}$ and h only. Let $\mathbf{Y}$ be the $(NT) \times 1$ vector with $\{Y_t(\mathbf{s}), 1 \le t \le T, \mathbf{s} \in \mathcal{S}_N\}$ as its NT components and $\widehat{\mathbf{Y}}$ be the corresponding vector with $\{Y_t(\mathbf{s})\}$ replaced by $\{\widehat{Y}_t(\mathbf{s})\}$. Then we may write $\widehat{\mathbf{Y}} = H(h)\mathbf{Y}$, where $H(h)$ is a $(NT) \times (NT)$ coefficient matrix independent of $\{Y_t(\mathbf{s})\}$. The GCV selects h which minimises

$$\mathrm{GCV}(h) = NT \left\{ \mathrm{tr}\Big(I - H(h)\Big) \right\}^{-2} (\mathbf{Y} - \widehat{\mathbf{Y}})^T (\mathbf{Y} - \widehat{\mathbf{Y}}).$$

3 Asymptotic Properties

We study the asymptotic properties of our estimator when both T and N tend to infinity. Note that N is the number of locations where the observations are taken. Our approach belongs to the category of the *fixed-domain asymptotics* which assumes that all the locations are within a fixed area when $N \to \infty$; see also condition (C5) below. Fixed-domain asymptotics is one of two frequently used asymptotic frameworks in the analysis of spatial

statistics; see, e.g. Cressie (1993, §3.3).

For the sake of simplicity, we only present the asymptotic results for linear models. Specifically, we always assume in this section that model (2.1) holds with $g(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{x}^\tau \boldsymbol{\theta}$, and the smoothed estimator $\widehat{\boldsymbol{\theta}}(\mathbf{s}_0)$ is derived from (2.2). Similar results hold for a general nonlinear g but with more complicated notation. First we state some regularity conditions.

(C3) 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 α-mixing with the mixing coefficient $\alpha(k)$ satisfying the condition $\sum_{k=1}^{\infty}\{\alpha(k)\}^{1-2/\delta} < \infty$.

(C4) The kernel $K(\cdot)$ is a symmetric density function on R^2 with a bounded support.

(C5) As $N \to \infty$, $N^{-1}\sum_{\mathbf{s}\in\mathcal{S}_N} I(\mathbf{s} \in A) \to \int_A f(\mathbf{s})d\mathbf{s}$ for any measurable set $A \subset \mathcal{S}$, where f is a *sampling intensity* (*i.e.* density) function on $\mathcal{S}$. Further, $f > 0$ in a neighbourhood of $\mathbf{s}_0 \in \mathcal{S}$.

(C6) The matrix function $\mathbf{A}_1(\mathbf{s}_1, \mathbf{s}_2) \equiv E\{\mathbf{X}_t(\mathbf{s}_1)\,\mathbf{X}_t(\mathbf{s}_2)^\tau\}$ is continuous, and $\boldsymbol{\theta}(\mathbf{s})$ is twice continuously differentiable.

We introduce some notation. Let $\mathbf{A}(\mathbf{s}) = \mathbf{A}_1(\mathbf{s}, \mathbf{s})$, $\sigma_1^2 = \Gamma_1(\mathbf{s}_0, \mathbf{s}_0)$, and

$$\mu_{i,1} = \int\int u^i K(u,v)dudv, \quad \mu_{i,2} = \int\int v^i K(u,v)dudv,$$
$$\nu_{i,1} = \int\int u^i K^2(u,v)dudv, \quad \nu_{i,2} = \int\int v^i K^2(u,v)dudv.$$

Theorem 1. *Let conditions (C1) – (C6) hold. As $T \longrightarrow \infty$, $n \longrightarrow \infty$ and $h \longrightarrow 0$, it holds for $\mathbf{s} \in \mathcal{S}$ that,*

$$\widehat{\boldsymbol{\theta}}(\mathbf{s}_0) - \boldsymbol{\theta}(\mathbf{s}_0) = \frac{1}{2}h^2\mathbf{b}(\mathbf{s}_0)(1 + o_P(1)) + \gamma T^{-1/2}\mathbf{A}^{-1/2}(\mathbf{s}_0)\boldsymbol{\xi}(1 + o_P(1)).$$

where $\boldsymbol{\xi}$ *is a* $p \times 1$ *random vector with zero mean and identity covariance matrix, and*

$$\gamma^2 = \sigma_1^2 + \frac{\nu_{0,1}\sigma_0^2}{Nh^2 f(\mathbf{s}_0)}, \quad \mathbf{b}(\mathbf{s}) = \mu_{2,1}\frac{\partial^2 \boldsymbol{\theta}(\mathbf{s})}{\partial u^2} + \mu_{2,2}\frac{\partial^2 \boldsymbol{\theta}(\mathbf{s})}{\partial v^2}.$$

Remark (a). It may be shown that when $\mathbf{s}_0 \in \mathcal{S}_N$,

$$\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0) - \boldsymbol{\theta}(\mathbf{s}_0) = \left(\sigma_1^2 + \sigma_0^2\right)^{1/2} T^{-1/2}\mathbf{A}^{-1/2}(\mathbf{s}_0)\boldsymbol{\xi}(1 + o_P(1)).$$

By choosing bandwidth $h = O\left((nT)^{-1/6}\right)$, we can see from Theorem 1 that the mean squared error of the smoothed estimator $\widehat{\boldsymbol{\theta}}(\mathbf{s}_0)$ is smaller than that of $\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0)$ under the condition $T = o(n^2)$. Furthermore, the smaller is σ_1^2/σ_0^2 (the system-noise-to-measurement-noise ratio), the larger is the improvement due to spatial smoothing. In particular, if $\sigma_1^2 = 0$, the mean squared error of $\widehat{\boldsymbol{\theta}}(\mathbf{s}_0)$ is an order of magnitude smaller than that of the method using the data at location $\mathbf{s}_0$ only.

Remark (b). In the case of no nugget effect (*i.e.* $\sigma_0^2 = 0$), the spatial smoothing cannot reduce the asymptotic variance of the unsmoothed estimator $\widetilde{\boldsymbol{\theta}}(\mathbf{s}_0)$. This is due to the fact that the spatial smoothing uses effectively the data at locations within a distance h from $\mathbf{s}_0$. Due to the continuity of the function $\Gamma_1(\cdot)$ stated in (C2), all the $\varepsilon_t(\mathbf{s})'$s from those locations are asymptotically identical. (The conventional weighted least squares approach using correlation coefficients as weights leads to an estimator with a constant bias, and is therefore not applicable here.) We argue that asymptotic theory under this setting presents an excessively gloomy picture; see Example 1 in §4 below. Adding a nugget effect in the model brings the theory closer to reality since in practice the data used in local spatial smoothing usually, if not always, contain some independent noise components. Note that the nugget effect is not detectable in practice since we can never observe or estimate $\Gamma(x)$

defined in (2.2) for x less than the minimum pairwise-distance among observed locations.

4 Numerical properties

Theorem 1 above shows that a smoothed estimator has a smaller asymptotic variance than its unsmoothed counterpart. In this section, we conduct simulation with two examples to illustrate the improvement due to spatial smoothing in the context of small sample size. Here and also in section 5, we use kernel $K(\mathbf{s}) = (1 - u^2 - v^2)_+$ and select the bandwidth h by the GCV method described in Section 2.3.

We set the sample size $T = 25$. The observations are taken over $N = m^2$ grid points on the square $[0, 6]^2$ with $m = 3, 6$ and 9, and the grid points $\{(u_i, v_j), 1 \le i, j \le m\}$ are defined as

$$u_i = 6(i-1)/(m-1), \quad v_j = 6(j-1)/(m-1).$$

We generate the noise process $\{\varepsilon_t(u_i, v_j), 1 \le i, j \le m\}$, which is independent in t, with the formula

$$\varepsilon_t(u_i, v_j) = \frac{\sigma}{2m_0 + 1} \sum_{k=-m_0}^{m_0} \sum_{l=-m_0}^{m_0} e_{i+m_0+k, j+m_0+l},$$

where $e_{i,j}$ are independent $N(0, 1)$ random variables, and $m_0 = 2$. We select $\sigma > 0$ such that the ratio of noise to signal defined as

$$\mathrm{Var}(\varepsilon_t(u, v)/\mathrm{Var}\{g(\mathbf{X}_t(u, v), \boldsymbol{\theta}(u, v)\}$$

is equal to 0.2. We define a process $\{X_t(u, v)\}$ in the same way as $\{\varepsilon_t(u, v)\}$ except with $\sigma = 1$ and $m_0 = 3$. We replicate the simulation 100 times for each setting.

Example 1. Consider the threshold model

$$Y_{t+1}(u,v) = \begin{cases} a_1(u,v) + a_2(u,v)X_t(u,v) + a_3(u,v)Y_t(u,v) + \varepsilon_{t+1}(u,v) & X_t(u,v) \le 0.6 \\ a_4(u,v) + a_5(u,v)X_t(u,v) + a_6(u,v)Y_t(u,v) + \varepsilon_{t+1}(u,v) & X_t(u,v) > 0.6, \end{cases}$$

where

$$a_1(u,v) = 1 + u + v, \quad a_2(u,v) = 0.5(u^2+v^2), \quad a_3(u,v) = 0.2\sin(u+v),$$
$$a_4(u,v) = u + 3v, \quad a_5(u,v) = 0.5(u^2+2v^2), \quad a_6(u,v) = 0.5\cos(u+v).$$

For each $k = 1, \cdots, 6$, MSE, which is defined by

$$\text{MSE}(\hat{a}_k) = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} E\Big\{\hat{a}_k(u_i,v_j) - a_k(u_i,v_j)\Big\}^2,$$

is employed as the criterion to compare the empirical performance of the (new) smoothed estimator and the (more conventional) unsmoothed estimator for $a_k(u_i,v_j)$, $i,\ j = 1, \cdots, m$.

(Put Table 1 here)

Table 1 reports the results of the comparison between the proposed smoothed estimator and the conventional unsmoothed estimator, in which the RMSE is defined as

$$\text{RMSE}(a_k) = \text{MSE1}(a_k)/\text{MSE2}(a_k),$$

where $\text{MSE1}(a_k)$ is the MSE for the smoothed estimator of $a_k(u_i,v_j)$, $i,\ j = 1, \cdots, m$, and $\text{MSE2}(a_k)$ is the MSE for the conventional least squares estimator of the same. In Table 1, we can see that all entries are less than 1, sometimes by a large margin; thus the gain from spatial smoothing is substantial even when m is moderate.

Example 2: Consider the model

$$Y_{t+1}(u_i,v_j) = a_1(u_i,v_j) + a_2(u_i,v_j)X_t(u_i,v_j) + a_3(u_i,v_j)Y_t(u_i,v_j) + \varepsilon_{t+1}(u_i,v_j)$$

$i,\ j = 1, \cdots, m,\ t = 1, \cdots, T$. Here (u_i, v_j), $X_t(u_i, v_j)$, $\varepsilon_t(u_i, v_j)$ are the same as those in Example 1, and

$$a_1(u, v) = 0.2\cos(u + v), \quad a_2(u, v) = u + v, \quad a_3(u, v) = \frac{1}{12}\sin(u + v).$$

T is still 25. Table 2 gives the comparison between the smoothed estimator and its unsmoothed counterpart in this case.

(Put Table 2 here.)

Table 2 shows a similar pattern as in Table 1, although the improvement due to spatial smoothing is less substantial. This lends support to the intuition that the pooling of information from nearby locations is more crucial when the underlying temporal dynamics is more complex.

5 Application to Mink and Muskrat Data

From the records compiled by the Hudson Bay Company on fur sales at auction in 1925-1949, we can extract annual numbers of mink and muskrats caught over 81 trapping regions in Canada for a period of 25 years. Based on the population dynamic structure, Yao *et al.* (2000) arrived at a grouping which divided the whole observation region into three groups, namely the western group (29 posts), the central group (43 posts), and the eastern group (9 posts). We now apply the proposed method to analyse the data in each of these three groups. Let $Y_t(\mathbf{s}_i)$ and $X_t(\mathbf{s}_i)$ be the mink observation and the muskrat observation respectively, on a natural logarithmic scale, for the ith station and at time t.

5.1 Food-chain-interaction models: initial fitting

Biological considerations suggest nonlinear temporal dynamics. In ecological population modelling, threshold autoregression has been found to be a simple approach which often offers interesting biological insights; see, for example, Framstad *et al.* (1997) and Stenseth *et al.* (1998a,b). We apply our smoothing approach to this class of models in the context of predator-prey interaction. Now, for each given station, i,

$$X_{t+1}(\mathbf{s}_i) = \begin{cases} a_1(\mathbf{s}_i) + a_2(\mathbf{s}_i)Y_t(\mathbf{s}_i) + a_3(\mathbf{s}_i)X_t(\mathbf{s}_i) + \varepsilon_{t+1}(\mathbf{s}_i) & X_t(\mathbf{s}_i) \le r_1(\mathbf{s}_i), \\ a_4(\mathbf{s}_i) + a_5(\mathbf{s}_i)Y_t(\mathbf{s}_i) + a_6(\mathbf{s}_i)X_t(\mathbf{s}_i) + \varepsilon_{t+1}(\mathbf{s}_i) & X_t(\mathbf{s}_i) > r_1(\mathbf{s}_i), \end{cases} \tag{5.1}$$

$$Y_{t+1}(\mathbf{s}_i) = \begin{cases} b_1(\mathbf{s}_i) + b_2(\mathbf{s}_i)X_t(\mathbf{s}_i) + b_3(\mathbf{s}_i)Y_t(\mathbf{s}_i) + \varepsilon_{t+1}(\mathbf{s}_i) & X_t(\mathbf{s}_i) \le r_2(\mathbf{s}_i), \\ b_4(\mathbf{s}_i) + b_5(\mathbf{s}_i)X_t(\mathbf{s}_i) + b_6(\mathbf{s}_i)Y_t(\mathbf{s}_i) + \varepsilon_{t+1}(\mathbf{s}_i) & X_t(\mathbf{s}_i) > r_2(\mathbf{s}_i), \end{cases} \tag{5.2}$$

$t = 1, \cdots, 24$, $i = 1, \cdots, 81$. Here, $\mathbf{s}_i$ =(latitude, longitude) is the location of the ith station.

It is clear that if we estimate the above parameters for each post using observations from that post alone, then large variability will result because only 24 observations are available to estimate the 7 parameters. The estimation proposed in section 2.2, which involves smoothing, provides a natural remedy for this problem by pooling the information from nearby posts within the same group.

We have also conducted an analysis of variance for the estimated coefficients $\hat{a}_i(\mathbf{s}), \hat{b}_i(\mathbf{s})$ and $\hat{r}_i(\mathbf{s})$. As expected, the variation within (the three) groups are much smaller than the variation between groups.

5.2 PCA for estimated coefficients

We now apply principal component analysis for the smoothing estimates $[\{\hat{b}_1(\mathbf{s}_j), \cdots, \hat{b}_6(\mathbf{s}_j), \hat{r}_2(\mathbf{s}_j)\}^T, j = 1, \cdots, 81]$ derived above. The coefficients of the first two principal components,

produced by Splus, are presented in Table 3, which account for 93.6%, 94.5% and 88.2% of total variation in western, central and eastern areas respectively. Note that Splus suppresses small coefficients by default. Although such a censoring lacks firm statistical underpinnings, it nevertheless suggests that, for example, in the western region the major (spatial) variation of model (5.2) occurs in the coefficients $b_1(\mathbf{s}), b_4(\mathbf{s})$ and $r_2(\mathbf{s})$, and we may set $b_j(\mathbf{s}) \equiv b_j$ for $j = 2, 3, 5, 6$. If we ignore a 'small' coefficient in the fourth row of Table 3, the same argument may apply to the central region. In the same vein, we may set $b_j(\mathbf{s}) \equiv b_j$ for $j = 2, 3, 5, 6$ in the eastern region.

(Put Table 3 here.)

The same analysis for $[\{(\hat{a}_1(\mathbf{s}_j), \cdots, \hat{a}_6(\mathbf{s}_j), \hat{r}_1(\mathbf{s}_j)\}^T,\ j = 1, \cdots, 81]$ leads to the suggestion that $a_j(\mathbf{s}) \equiv a_j$ for $j = 2, 3, 5, 6$ in model (5.1) for all the three regions.

5.3 Identifying nonlinear effect based on estimated coefficients

In threshold models (5.1) and (5.2), nonlinearity is reflected by different coefficients in the upper and lower regimes. As a diagnostic tool, we may conduct a simple one-sample t-test for, for example, $[\{\hat{a}_1(s_j) - \hat{a}_4(\mathbf{s}_j)\},\ j = 1, \cdots, 81]$. The p-values are reported in Tables 4 and 5. These results suggest that $a_2(\mathbf{s}) = a_5(\mathbf{s})$ in model (5.1) for all the three regions. Similarly, we may let $b_3(\mathbf{s}) = b_6(\mathbf{s})$ for central region, and $a_1(\mathbf{s}) = a_4(\mathbf{s})$ and $b_1(\mathbf{s}) = b_4(\mathbf{s})$ in the western and the eastern region.

(Put Table 4 here.)

5.4 Food-chain-interaction model: refined fitting

The above analysis of the estimated coefficients suggests that we may impose some constraints on models (5.1) and (5.2) while fitting the mink and muskrat data. We only report the results from the western and central regions in detail since the data in these regions suggest that the food-chain-interaction is strong in these regions. For example, in western region the fitted models, under appropriate constraints, are

$$X_{t+1}(\mathbf{s}_i) = \begin{cases} \hat{a}_1(\mathbf{s}_i) - 0.226Y_t(\mathbf{s}_i) + 0.856X_t(\mathbf{s}_i) + \hat{\varepsilon}_{t+1}(\mathbf{s}_i) & X_t(\mathbf{s}_i) \le \hat{r}_1(\mathbf{s}_i), \\ \hat{a}_4(\mathbf{s}_i) - 0.226Y_t(\mathbf{s}_i) + 1.009X_t(\mathbf{s}_i) + \hat{\varepsilon}_{t+1}(\mathbf{s}_i) & X_t(\mathbf{s}_i) > \hat{r}_1(\mathbf{s}_i), \end{cases}$$

$$Y_{t+1}(\mathbf{s}_i) = \begin{cases} \hat{b}_1(\mathbf{s}_i) + 0.260X_t(\mathbf{s}_i) + 0.478Y_t(\mathbf{s}_i) + \hat{\varepsilon}_{t+1}(\mathbf{s}_i) & X_t(\mathbf{s}_i) \le \hat{r}_2(\mathbf{s}_i), \\ \hat{b}_1(\mathbf{s}_i) + 0.182X_t(\mathbf{s}_i) + 0.656Y_t(\mathbf{s}_i) + \hat{\varepsilon}_{t+1}(\mathbf{s}_i) & X_t(\mathbf{s}_i) > \hat{r}_2(\mathbf{s}_i), \end{cases}$$

The above model is the result of the following fitting procedure. We first estimate model (5.1) with constraint $a_2(\mathbf{s}) \equiv a_5(\mathbf{s})$ and model (5.2) using the proposed smoothing method. We then replace $\hat{a}_j(\mathbf{s}_i)$ and $\hat{b}_j(\mathbf{s}_i)$, respectively, by their mean values over the 29 posts for $j = 2, 3, 5, 6$. This idea of estimating constant coefficients was explored in Fan and Zhang (2000). For the muskrat model in this region, the regulation of mink over muskrat is phase-independent, and the spatial variation is limited to intercepts and thresholds. The corresponding estimated coefficient functions are plotted in Figure 1.

(*Put Figure 1 here.*)

Similarly the simplified models in the central region are

$$X_{t+1}(\mathbf{s}_i) = \begin{cases} \hat{a}_1(\mathbf{s}_i) - 0.208Y_t(\mathbf{s}_i) + 0.848X_t(\mathbf{s}_i) + \hat{\varepsilon}_{t+1}(\mathbf{s}_i) & X_t(\mathbf{s}_i) \le \hat{r}_1(\mathbf{s}_i), \\ \hat{a}_4(\mathbf{s}_i) - 0.208Y_t(\mathbf{s}_i) + 1.024X_t(\mathbf{s}_i) + \hat{\varepsilon}_{t+1}(\mathbf{s}_i) & X_t(\mathbf{s}_i) > \hat{r}_1(\mathbf{s}_i), \end{cases}$$

$$Y_{t+1}(\mathbf{s}_i) = \begin{cases} \hat{b}_1(\mathbf{s}_i) + 0.220X_t(\mathbf{s}_i) + 0.534Y_t(\mathbf{s}_i) + \hat{\varepsilon}_{t+1}(\mathbf{s}_i) & X_t(\mathbf{s}_i) \leq \hat{r}_2(\mathbf{s}_i), \\ \hat{b}_4(\mathbf{s}_i) + 0.351X_t(\mathbf{s}_i) + 0.534Y_t(\mathbf{s}_i) + \hat{\varepsilon}_{t+1}(\mathbf{s}_i) & X_t(\mathbf{s}_i) > \hat{r}_2(\mathbf{s}_i), \end{cases}$$

These models present largely a similar pattern as in the western region, except that now the self-regulation of mink is also phase-independent. The estimated coefficient functions are plotted in Figure 2.

(*Put Figure 2 here.*)

The above fitted models may be interpreted biologically as follows. First in both regions the mink is affected positively by the presence of its prey muskrat. The effect is about the same in both regions. Further the effect of the mink on the muskrat population is the same in both upper and lower regimes, whereas the effect of the muskrat on the mink is not. This seems to indicate that the muskrat is more important for the mink in the upper regime (corresponding to the increase and peak population phase; see Yao *et al.* 2000) in the central region which indeed may be explained by reference to the possibility that the central region may be the mink-muskrat's core area within which they are most closely linked with each other. This interpretation is in fact supported by observing that the muskrat is less affected by the mink in the upper regime; this pattern may be due to the possibility that the two species have been mutually adapted to each other. Furthermore, in both regions the self-regulation within the muskrat population is the strongest in the lower regime (corresponding to the decrease phase of the population cycle), during which competition within the population is likely to be the strongest. A similar pattern is found for the mink in the western region but not in the central region, which indeed is consistent with the central region being a better region for the species.

In conclusion, ecological time series are typically short, but there may be many of them for each single species or system (see, e.g., Stenseth 1999). By pooling information from nearby locations through kernel smoothing, we have derived reliable estimation for the food-chain-interaction models between mink and muskrat thereby enabling relevant biological interpretation.

Appendix – Proof of Theorem 1

Let

$$\mathbf{X}_i = \Omega_i \otimes \Big(1,\ (\mathbf{s}_i - \mathbf{s}_0)^\tau/h\Big), \quad \text{with} \quad \Omega_i = \Big(\mathbf{X}_1(\mathbf{s}_i), \cdots, \mathbf{X}_T(\mathbf{s}_i)\Big)^\tau, \quad i = 1, \cdots, n.$$

$$\mathbf{Y} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_n), \quad \text{with} \quad \mathbf{Y}_i = \Big(Y_1(\mathbf{s}_i), \cdots, Y_T(\mathbf{s}_i)\Big).$$

$$\mathbf{X} = (\mathbf{X}_1^\tau, \cdots, \mathbf{X}_n^\tau)^\tau, \quad W = \text{diag}\Big(K_h(\mathbf{s}_1 - \mathbf{s}_0), \cdots, K_h(\mathbf{s}_n - \mathbf{s}_0)\Big) \otimes I_T.$$

Solving (2.2), we obtain the estimator of $\boldsymbol{\theta}(\mathbf{s}_0)$

$$\hat{\boldsymbol{\theta}}(\mathbf{s}_0) = I_p \otimes e_{1,3}^\tau (\mathbf{X}^\tau W \mathbf{X})^{-1} \mathbf{X}^\tau W \mathbf{Y}. \tag{A.1}$$

Put

$$\boldsymbol{\alpha}_i^\tau = \Big(1, (\mathbf{s}_i - \mathbf{s}_0)^\tau/h\Big), \quad B = \Big(I_p \otimes \text{diag}(1, h, h)\Big),$$

note that $\mathbf{X}_i = (\Omega_i \otimes \boldsymbol{\alpha}_i^\tau)B$ we have

$$B^{-1}\mathbf{X}^\tau W \mathbf{X} B^{-1} = \sum_{i=1}^n (\Omega_i^\tau \Omega_i) \otimes (\boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^\tau) K_h(\mathbf{s}_i - \mathbf{s}_0),$$

and

$$B^{-1}\mathbf{X}^\tau W \mathbf{Y} = \sum_{i=1}^n (\Omega_i^\tau \otimes \boldsymbol{\alpha}_i) \mathbf{Y}_i K_h(\mathbf{s}_i - \mathbf{s}_0).$$

For any $3p$ dimension vector Z with $Z^\tau Z = 1$, we have

$$\frac{1}{nT} Z^\tau B^{-1}(\mathbf{X}^\tau W \mathbf{X}) B^{-1} Z = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T} Z^\tau \Big\{(\Omega_i^\tau \Omega_i) \otimes (\boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^\tau)\Big\} Z\right] K_h(\mathbf{s}_i - \mathbf{s}_0)$$

Let

$$\xi_i = \frac{1}{T} Z^\tau \Big\{(\Omega_i^\tau \Omega_i) \otimes (\boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^\tau)\Big\} Z,$$

we have

$$\mathrm{Var}\Big\{(nT)^{-1} Z^\tau B^{-1}(\mathbf{X}^\tau W \mathbf{X}) B^{-1} Z\Big\} \le \left\{\frac{1}{n}\sum_{i=1}^{n}\Big(\mathrm{Var}(\xi_i)\Big)^{1/2} K_h(\mathbf{s}_i - \mathbf{s}_0)\right\}^2.$$

Let

$$\xi_{i,j} = Z^\tau \Big\{\mathbf{X}_j(\mathbf{s}_i)\mathbf{X}_j^\tau(\mathbf{s}_i) \otimes (\boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^\tau)\Big\} Z,$$

we have

$$\xi_i = \frac{1}{T}\sum_{j=1}^{T} \xi_{i,j}.$$

For $k < 0$, let $\mathrm{Cov}(\xi_{i,1},\ \xi_{i,1+k}) = \mathrm{Cov}(\xi_{i,1},\ \xi_{i,1-k})$, we get

$$\begin{aligned}
\mathrm{Var}(\xi_i) &= \frac{1}{T^2}\sum_{1\le j,k\le T} \mathrm{Cov}(\xi_{i,j},\ \xi_{i,k}) = \frac{1}{T}\sum_{k=-(T-1)}^{T-1}\left(1 - \frac{|k|}{T}\right)\mathrm{Cov}(\xi_{i,1},\ \xi_{i,1+k}) \\
&\le \frac{1}{T}\sum_{k=-(T-1)}^{T-1} \mathrm{Cov}(\xi_{i,1},\ \xi_{i,1+k}).
\end{aligned}$$

By Davydov's inequality, we get

$$\mathrm{Cov}(\xi_{i,1},\ \xi_{i,1+k}) \le \frac{2\delta}{\delta - 2}\Big\{2\alpha(k)\Big\}^{1-2/\delta}\Big\{E|\xi_{i,1}|^\delta\Big\}^{2/\delta},$$

this together with condition (C3) leads to

$$\mathrm{Var}(\xi_i) \le \frac{C}{T}\sum_{k=0}^{T-1}\Big\{\alpha(k)\Big\}^{1-2/\delta} \le C_1 T^{-1}$$

where C and C_1 are constants, and free of i, T. This gives

$$\mathrm{Var}\Big\{(nT)^{-1} Z^\tau B^{-1}(\mathbf{X}^\tau W \mathbf{X}) B^{-1} Z\Big\} = O(T^{-1}).$$

Moreover,

$$\begin{aligned} E\Big\{(nT)^{-1}Z^\tau B^{-1}(\mathbf{X}^\tau W\mathbf{X})B^{-1}Z\Big\} &= \frac{1}{n}\sum_{i=1}^{n} Z^\tau\Big\{\mathbf{A}(\mathbf{s}_i)\otimes(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^\tau)K_h(\mathbf{s}_i-\mathbf{s}_0)\Big\}Z \\ &= Z^\tau\mathbf{A}(\mathbf{s}_0)\otimes \mathrm{diag}(1,\mu_{2,1},\mu_{2,2})f(\mathbf{s}_0)Z(1+o(1)). \end{aligned}$$

So,

$$\frac{1}{nT}B^{-1}\mathbf{X}^\tau W\mathbf{X}B^{-1} = \mathbf{A}(\mathbf{s}_0)\otimes \mathrm{diag}(1,\mu_{2,1},\mu_{2,2})f(\mathbf{s}_0)(1+o_P(1)). \tag{A.2}$$

Let $\Sigma = \Sigma_2 \otimes I_T$, Σ_2 is a $n\times n$ matrix with the (i,j)th element $\Gamma_1(\mathbf{s}_i,\mathbf{s}_j)$.

$$B^{-1}\mathbf{X}^\tau W\Sigma W\mathbf{X}B^{-1} = \sum_{i=1}^{n}\sum_{j=1}^{n}(\Omega_i^\tau\Omega_j)\otimes(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j^\tau)K_h(\mathbf{s}_i-\mathbf{s}_0)K_h(\mathbf{s}_j-\mathbf{s}_0)\Gamma_1(\mathbf{s}_i,\mathbf{s}_j),$$

similar to the derivation for (A.2), and note that

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{A}_1(\mathbf{s}_i,\mathbf{s}_j)\otimes(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j^\tau)K_h(\mathbf{s}_i-\mathbf{s}_0)K_h(\mathbf{s}_j-\mathbf{s}_0)\Gamma_1(\mathbf{s}_i,\mathbf{s}_j) \\ = \;& n^2 f(\mathbf{s}_0)^2\mathbf{A}(\mathbf{s}_0)\otimes\Delta(1+o(1)), \end{aligned}$$

where

$$\Delta = \int\int (1,\mathbf{u}_1^\tau)^\tau(1,\mathbf{u}_2^\tau)K(\mathbf{u}_1)K(\mathbf{u}_2)\Gamma_1(\mathbf{s}_0+h\mathbf{u}_1,\mathbf{s}_0+h\mathbf{u}_2)d\mathbf{u}_1 d\mathbf{u}_2,$$

we obtain

$$B^{-1}\mathbf{X}^\tau W\Sigma W\mathbf{X}B^{-1} = n^2 T f(\mathbf{s}_0)^2\mathbf{A}(\mathbf{s}_0)\otimes\Delta(1+o_P(1)), \tag{A.3}$$

and

$$e_{1,3}^\tau\Delta e_{1,3} = \int\int K(\mathbf{u}_1)K(\mathbf{u}_2)\Gamma_1(\mathbf{s}_0+h\mathbf{u}_1,\mathbf{s}_0+h\mathbf{u}_2)d\mathbf{u}_1 d\mathbf{u}_2 = \sigma_1^2(1+o(1)).$$

Similar to (A.2), we have

$$\frac{1}{nT}B^{-1}\mathbf{X}^\tau W^2\mathbf{X}B^{-1} = h^{-2}\mathbf{A}(\mathbf{s}_0)\otimes\mathrm{diag}(\nu_{0,1},\nu_{2,1},\nu_{2,2})f(\mathbf{s}_0)(1+o_P(1)). \tag{A.4}$$

By Taylor expansion, we have

$$\begin{aligned}\theta_j(u,v) &= \theta_j(u_0,v_0) + \left(\Delta_u\frac{\partial}{\partial u} + \Delta_v\frac{\partial}{\partial v}\right)\theta_j(u_0,v_0) \\ &+ \frac{1}{2}\left(\Delta_u\frac{\partial}{\partial u} + \Delta_v\frac{\partial}{\partial v}\right)^2\theta_j(u_0,v_0) + O\Big\{(\Delta_u^2+\Delta_v^2)^{3/2}\Big\}\end{aligned}$$

where $(u,v)=\mathbf{s}$, $(u_0,v_0)=\mathbf{s}_0$, $\Delta_u = u-u_0$, $\Delta_v = v-v_0$. This gives

$$\hat{\boldsymbol{\theta}}(s_0) - \boldsymbol{\theta}(s_0) = \mathbf{J}_1 h^2 + \mathbf{J}_2 + o_P(h^2), \tag{A.5}$$

where

$$\mathbf{J}_1 = \frac{1}{2}(I_p\otimes e_{1,3}^\tau)(\mathbf{X}^\tau W\mathbf{X})^{-1}\mathbf{X}^\tau W\boldsymbol{\Psi}\boldsymbol{\gamma}, \quad \mathbf{J}_2 = (I_p\otimes e_{1,3}^\tau)(\mathbf{X}^\tau W\mathbf{X})^{-1}\mathbf{X}^\tau W\Big(\boldsymbol{\epsilon}^{(1)}+\boldsymbol{\epsilon}^{(2)}\Big),$$

$$\boldsymbol{\gamma} = \frac{1}{2}(\ddot{\theta}_1^\tau,\cdots,\ddot{\theta}_p^\tau)^\tau, \quad \ddot{\theta}_i = \left(\frac{\partial^2\theta_i(u_0,v_0)}{\partial u^2}, \frac{\partial^2\theta_i(u_0,v_0)}{\partial u\partial v}, \frac{\partial^2\theta_i(u_0,v_0)}{\partial v^2}\right)^\tau$$

$$\boldsymbol{\epsilon}^{(1)} = \Big(\varepsilon_1^{(1)}(\mathbf{s}_1),\cdots,\varepsilon_T^{(1)}(\mathbf{s}_1),\cdots,\varepsilon_1^{(1)}(\mathbf{s}_n),\cdots,\varepsilon_T^{(1)}(\mathbf{s}_n)\Big)^\tau,$$

$$\boldsymbol{\epsilon}^{(2)} = \Big(\varepsilon_1^{(2)}(\mathbf{s}_1),\cdots,\varepsilon_T^{(2)}(\mathbf{s}_1),\cdots,\varepsilon_1^{(2)}(\mathbf{s}_n),\cdots,\varepsilon_T^{(2)}(\mathbf{s}_n)\Big)^\tau, \quad \boldsymbol{\Psi} = (\boldsymbol{\Psi}_1^\tau,\cdots,\boldsymbol{\Psi}_n^\tau)^\tau,$$

with

$$\boldsymbol{\Psi}_i = \Omega_i\otimes\boldsymbol{\beta}_i^\tau, \quad \boldsymbol{\beta}_i^\tau = \Big((u_i-u_0)^2, 2(u_i-u_0)(v_i-v_0), (v_i-v_0)^2\Big)/h^2.$$

Note that

$$B^{-1}\mathbf{X}^\tau W\boldsymbol{\Psi} = \sum_{i=1}^n(\Omega_i^\tau\Omega_i)\otimes(\boldsymbol{\alpha}_i\boldsymbol{\beta}_i^\tau)K_h(\mathbf{s}_i-\mathbf{s}_0),$$

and using the similar argument for getting (A.2), we can obtain

$$\frac{1}{nT}B^{-1}\mathbf{X}^\tau W\boldsymbol{\Psi} = \mathbf{A}(\mathbf{s}_0)\otimes\begin{pmatrix}\mu_{2,1} & 0 & \mu_{2,2}\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix} f(\mathbf{s}_0)(1+o_P(1)).$$

This together with (A.2) leads to

$$\mathbf{J}_1 = \frac{1}{2}\mathbf{b}(1+o_P(1)) \tag{A.6}$$

Now, we discuss $\mathbf{J}_2$. Let

$$\boldsymbol{\eta} = \left\{T^{-1}\sigma_1^2 + (nTh^2 f(s_0))^{-1}\nu_{0,1}\sigma_0^2\right\}^{-1/2} \mathbf{A}^{1/2}(s_0)\mathbf{J}_2 \tag{A.7}$$

From (A.2), we get

$$\begin{aligned}\boldsymbol{\eta} &= \left\{T^{-1}\sigma_1^2 + (nTh^2 f(s_0))^{-1}\nu_{0,1}\sigma_0^2\right\}^{-1/2} \times \\ &\quad (nT)^{-1}\left(\mathbf{A}^{-1/2}(\mathbf{s}_0) \otimes e_{1,3}^{\tau}\right) f(\mathbf{s}_0)^{-1} B^{-1}\mathbf{X}^{\tau} W(\boldsymbol{\epsilon}^{(1)} + \boldsymbol{\epsilon}^{(2)})(1 + o_P(1)).\end{aligned}$$

By (A.3), (A.4)

$$\mathrm{Cov}(\boldsymbol{\eta}) = I_p(1 + o(1)),$$

this together with (A.5), (A.6) and (A.7) leads to the result of Theorem 1.

References

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, **31**, 377-403.

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

Diggle, P.J. (1985). A kernel method for smoothing point process data. *Appl. Statist.* **34**, 138-147.

Diggle, P.J. and Marron, J.S. (1988). Equivalence of smoothing parameter selectors in density and intensity estimation. *J. Amer. Statist. Assoc.* **83**, 793-800.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman & Hall, London.

Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*, **27**, 1-17.

Framstad, E., Stenseth, N.C., Bjørnstad, O.N., and Falck, W. (1997). Limit cycles in Norwegian lemmings: tensions between phase-dependence and density-dependence. *Proc. Roy. Soc.* **B**, **264**, 31-38.

Hallin, M., Lu, Z. and Tran, L.T. (2002). Density estimation for linear processes. *Bernoulli*, to appear.

Hjellvik, V. and Tjøstheim, D. (1999). Modeling panels of intercorrelated autoregressive time series. *Biometrika*, **86**, 573-590.

May, R.M. (1981). Models for two interacting populations. In *Theoretical Ecology* ed by R.M. May, 78-104, Oxford: Blackwell.

Simonoff, J. S. (1996). *Smoothing Methods in Statistics*. Springer, New York.

Stenseth, N.C., Falck, W., Bjørnstad, O.N., and Krebs, C.J. (1997). Population regulation in snowshoe hare and Canadian lynx; Asymmetric food web configurations between hare and lynx. *Proc. Natl. Acad. Sci. USA* **94**, 5147-5152.

Stenseth, N.C., Falck, W., Chan, K.S., Bjørnstad, O.N., Tong, H., O'Donoghue, M. Boonstra, R., Boutin, S., Krebs, C.J., and Yoccoz, N.G. (1998a). From patterns to processes: Phase and density dependencies in the Canadian lynx cycle. *Proc. Natl. Acad. Sci. USA* **95**, 15430-15435.

Stenseth, N.C., Chan, K.S., Framstad, E., and Tong, H. (1998b). Phase- and density-dependent population dynamics in Norwegian lemmings: interaction between deterministic and stochastic processes. *Proc. Roy. Soc.* **B**, **265**, 1957-1968.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.

Wahba, G. (1977). A survey of some smoothing problems and the method of generalized cross-validation for solving them. In *Applications of Statistics* (P. R. Krisnaiah, ed.), 507-523. North Holland, Amsterdam.

Yao, Q. (2003). Exponential inequalities for spatial processes and uniform convergence rates for density estimation. In *Development of Modern Statistics and Related Topics — In Celebration of Professor Yaoting Zhang's 70th Birthday*, J. Huang and H. Zhang (edit.). World Scientific, Singapore, to appear.

Yao, Q., Tong, H., Finkenstädt, B. and Stenseth, N., C. (2000). Common structure in panels of short ecological time series. *Proceeding Royal Soc. Lond.*, **B**, **267**, 2457-2467.

Table 1: The Empirical Performance Comparison

m	RMSE(a_1)	RMSE(a_2)	RMSE(a_3)	RMSE(a_4)	RMSE(a_5)	RMSE(a_6)
3	0.865	0.941	0.907	0.680	0.772	0.883
6	0.403	0.463	0.378	0.502	0.495	0.562
9	0.263	0.342	0.026	0.046	0.088	0.008

Table 2: The Empirical Performance Comparison

m	RMSE(a_1)	RMSE(a_2)	RMSE(a_3)
3	0.960	0.925	0.999
6	0.724	0.647	0.920
9	0.652	0.530	0.653

Table 3: First Two Principal Components of Smoothing Estimates

western (first)	-0.15	0	0	0.50	0	0	0.85
western (second)	-0.76	0	0	-0.60	0	0	0.21
central (first)	-0.27	0	0	-0.85	0	0	0.44
central (second)	-0.93	0	0.14	0.33	0	0	0
eastern (first)	0	0	0	-0.85	0	0	0.51
eastern (second)	-0.98	0	0.13	0	0	0	0

Table 4: The p-values for one-sample t-tests

Null Hypothesis	Western	Central	Eastern
$a_1(\mathbf{s}) = a_4(\mathbf{s})$	0.000	0.000	0.377
$a_2(\mathbf{s}) = a_5(\mathbf{s})$	0.422	0.282	0.456
$a_3(\mathbf{s}) = a_6(\mathbf{s})$	0.000	0.000	0.090
$b_1(\mathbf{s}) = b_4(\mathbf{s})$	0.053	0.000	0.292
$b_2(\mathbf{s}) = b_5(\mathbf{s})$	0.000	0.000	0.118
$b_3(\mathbf{s}) = b_6(\mathbf{s})$	0.000	0.639	0.050

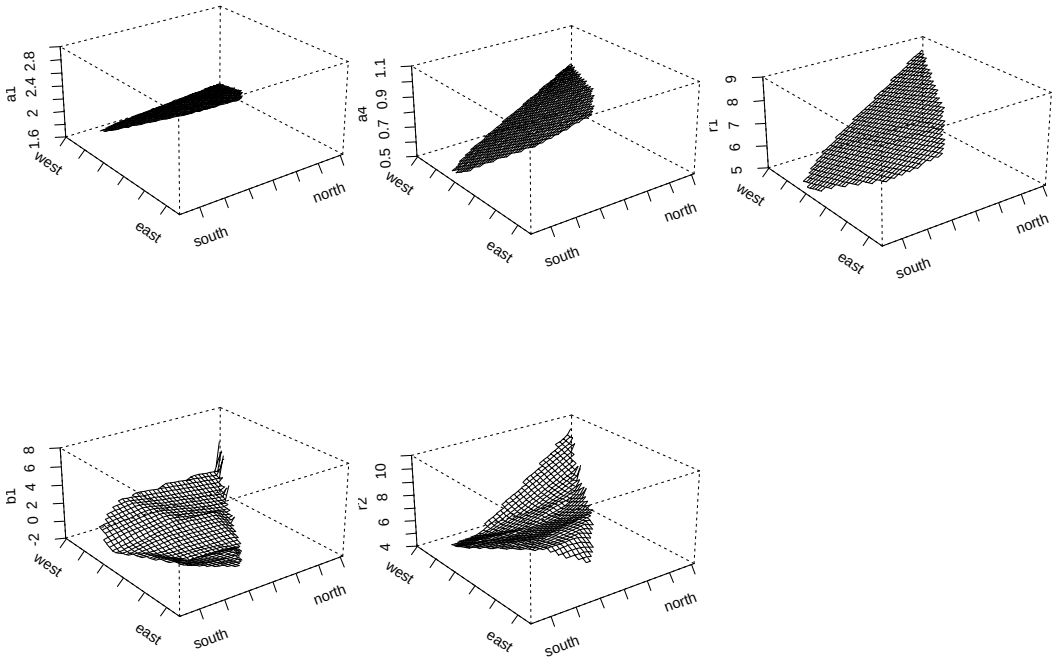

Figure 1: The estimates for the functional coefficients of the model for Western region.

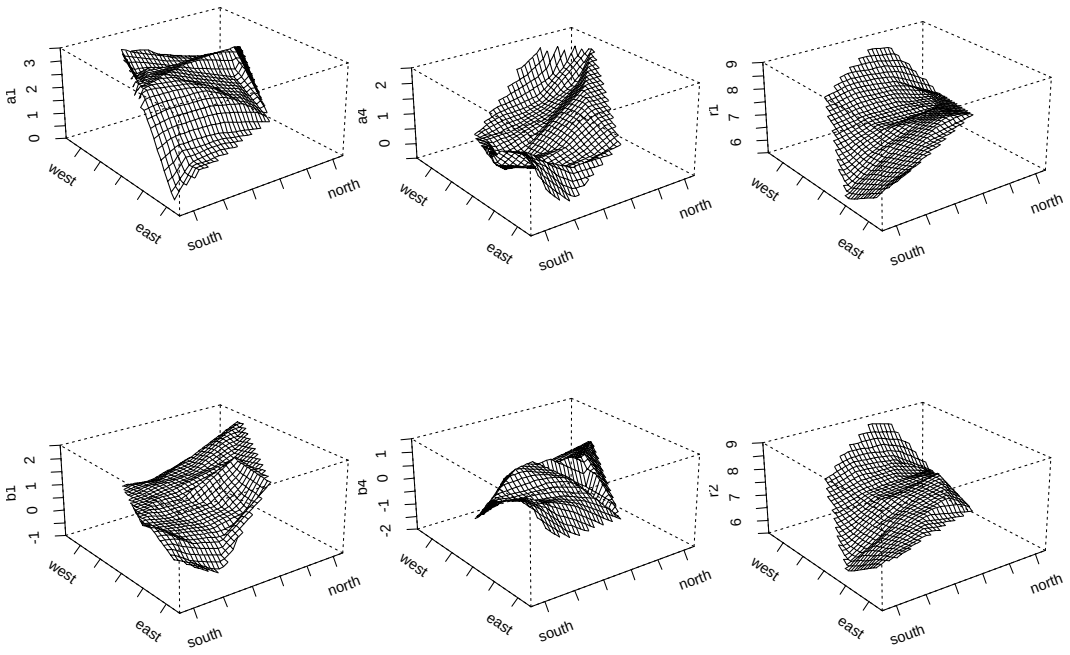

Figure 2: The estimates for the functional coefficients of the model for Central region.