Title
Local Linear Estimation for Spatial Random Processes with Stochastic Trend and Stationary Noise

Journal
Sankhya B, 80(2)

ISSN
0581-5738

Authors
Hyun, Jung Won
Burman, Prabir
Paul, Debashis

Publication Date
2018-11-01

DOI
10.1007/s13571-018-0155-4

Peer reviewed

HHS Public Access
Author manuscript
Sankhya Ser B. Author manuscript; available in PMC 2019 November 04.

Published in final edited form as:
Sankhya Ser B. 2018 November ; 80(2): 369–394. doi:10.1007/s13571-018-0155-4.

Local linear estimation for spatial random processes with stochastic trend and stationary noise

Jung Won Hyun[1], **Prabir Burman**[2], **Debashis Paul**[2,*]

[1]Department of Biostatistics, St. Jude Children's Research Hospital, 262 Danny Thomas Place Memphis, TN 38105

[2]Department of Statistics, University of California, Davis, CA 95616

Abstract

We consider the problem of estimating the trend for a spatial random process model expressed as $Z(x) = \mu(x) + \varepsilon(x) + \delta(x)$, where the trend μ is a smooth random function, $\varepsilon(x)$ is a mean zero, stationary random process, and $\{\delta(x)\}$ are assumed to be i.i.d. noise with zero mean. We propose a new model for stochastic trend in $\mathbb{R}^d$ by generalizing the notion of a structural model for trend in time series. We estimate the stochastic trend nonparametrically using a local linear regression method and derive the asymptotic mean squared error of the trend estimate under the proposed model for trend. Our results show that the asymptotic mean squared error for the stochastic trend is of the same order of magnitude as that of a deterministic trend of comparable complexity. This result suggests from the point of view of estimation under stationary noise, it is immaterial whether the trend is treated as deterministic or stochastic. Moreover, we show that the rate of convergence of the estimator is determined by the degree of decay of the correlation function of the stationary process $\varepsilon(x)$ and this rate can be different from the usual rate of convergence found in the literature on nonparametric function estimation. We also propose a data dependent selection procedure for the bandwidth parameter which is based on a generalization of Mallow's C_p criterion. We illustrate the methodology by simulation studies and by analyzing a data on surface temperature anomalies.

Keywords

spatial process; stochastic trend; local polynomial smoothing; bandwidth selection; Mallows' C_p

1 Introduction

We consider a random process $\{Z(x), x \in D\}$, where D is a subset of $\mathbb{R}^d$ for $d \geq 1$ observed at locations $S_1,\ldots,S_n$. We suppose that the observed random process is generated from a trend plus a short term stationary error and a measurement error. Thus, the model for the observations is of the form

*Corresponding author; debpaul@ucdavis.edu.

$$Z(S_i) = \mu(S_i) + \varepsilon(S_i) + \delta(S_i), \quad i = 1, \ldots, n, \tag{1}$$

where S_i's are the observation locations, $\{Z(S_i)\}$ is the observed random process, $\varepsilon(x)$ is a mean zero, stationary process with $\mathrm{Cov}(\varepsilon(x+h), \varepsilon(x)) = \gamma(Nh)$, where $\gamma(\cdot)$ is a covariance function on $\mathbb{R}^d$ and N is an unknown scalar parameter, and $\delta(S_i)$'s are i.i.d. observational noise with zero mean and variance τ^2. Moreover, $\mu(x)$, $\varepsilon(x)$ and $\delta(x)$ are assumed to be independent. Our goal is to estimate the trend $\mu(x)$, which is throughout assumed to be stochastic. We estimate the trend using a local linear regression method. The proposed methodology generalizes the formulation by Burman (1991) for the univariate setting.

When the trend is assumed to be a nonrandom function of the spatial location, one common approach is to model the trend as a known parametric function. Such a trend surface modeling approach assumes that the trend is represented in a given basis, such as a polynomial or a fixed-knot spline basis, in the spatial coordinates. However, often it is unrealistic to assume that the trend has a known parametric form, and it is reasonable to model the trend as an arbitrary smooth function of the spatial location. Under such settings, one can estimate the trend using a nonparametric smoothing method such as local polynomial regression (Schabenberger & Gotway 2004) or spline smoothing (Wood et al. 2002). When a deterministic trend is assumed and a local polynomial method is employed to estimate the trend, the asymptotic properties of the estimator have been thoroughly investigated in the literature under the assumption of spatially uncorrelated noise. As a representative text, Ruppert & Wand (1994) study the asymptotic bias and variance of multivariate local regression estimators.

Stochastic trend models have been considered primarily in the time series literature (Box et al. 1994, Durbin & Koopman 2001, Harvey 1991, Shumway & Stoffer 2000), where it is often referred to as a *structural model* Burman & Shumway (2009) consider a time series model with a random trend and a stationary error and derive an expression for the asymptotic mean squared error of the trend estimate. A discussion on deterministic versus stochastic trend can be found in Chapter 4.1 of Box et al. (1994).

When modeling the random trend in the structural model, it is typically assumed that, for a given $k > 0$, the k-th order differences of the trend are i.i.d. mean zero random variables. We extend this idea to model the trend in (1). Specifically, we propose a model for μ by defining it locally through an integral with respect to a Gaussian process. However, the assumption of Gaussianity is not essential for the asymptotic properties of the proposed estimator of the trend.

In this paper we estimate the stochastic trend nonparametrically using a local linear regression method and derive an expression for its asymptotic mean squared error. Throughout we assume that the observation locations are randomly distributed over a fixed finite domain. The results of this paper show that the asymptotic mean squared error for the stochastic trend model is of the same order of magnitude as that for the deterministic trend model, which suggests that it does not matter whether μ is considered to be a nonrandom or a random function. We also show that the rate of convergence of the estimator of μ is determined by N, the parameter controlling the degree of correlation in the stationary noise ε. Note that in our asymptotic analysis, we allow $N \to \infty$ as $n \to \infty$, and show that the rate of convergence for the local linear estimates is of order $(\min(n, N))^{-4/(4+d)}$ Indeed, if $N \le n$, then the variance of the estimator is dominated by N, which plays the role of the effective number of measurements (see Theorem 1). Only if $N > n$, the rate of convergence is similar to the case of i.i.d. noise with n measurements. In practice, we need to select the smoothing parameter for estimating μ. Moreover, we propose a data-driven selection procedure for the bandwidth which is based on a generalization of Mallow's C_p criterion and takes into account the spatial correlation of the residual process. The analysis techniques can be generalized to obtain qualitatively similar results for a local polynomial regression estimator of stochastic trend using a higher order polynomial.

The rest of the paper is organized as follows. We present the stochastic trend model in Section 2. In Section 3, we discuss the local linear estimation of the trend and derive an expression for its asymptotic bias and variance. In Section 4, we present the method for data dependent selection of the bandwidth for the smoother. In Section 5, we conduct a simulation study to demonstrate the performance of the bandwidth selector for the local linear estimator. In Section 6, we analyze the data on surface temperature anomalies in the northern America using the proposed estimator. Proofs of the asymptotic results are given in Section 7.

2 A model for stochastic trend

In this section, we describe the statistical model for the random trend μ. This is a spatial generalization of a stochastic trend model commonly used in time series through a state-space formulation (Shumway & Stoffer 2000, Burman & Shumway 2009). Later, we use this random trend model for the true trend and establish expressions for asymptotic MSE of a local linear estimator of the trend.

The idea of the proposed stochastic trend is to generalize the notion from the state space model in time series that the k-th divided difference of the series is a white noise process, for a given integer k. In the setting of the continuum, this notion is implemented by defining the process through stochastic integration of a spline-type kernel with respect to the standard Brownian motion. This kernel is defined through iterated convolution of a boxcar function, and therefore the definition generalizes the notion that the divided difference of a certain order is a white noise process.

Specifically, we define a univariate kernel $L_1^{(k)}(\cdot)$ by

$$L_1^{(k)}(t) = \frac{1}{F_k(R)} \frac{1}{k!} \sum_{j=0}^{[k/2]} (-1)^j \binom{k+1}{j} \{(k+1-2j)R - |t|\}_+^k, \quad t \in \mathbb{R}, \tag{2}$$

where k is a positive integer which determines the smoothness of the kernel, $[x]$ denotes the largest integer less than or equal to $x \in \mathbb{R}$, $R > 0$ is a parameter determining the support of the kernel, and $F_k(R)$ is the normalizing constant given by

$$
\begin{aligned}
F_k(R) &= \int \frac{1}{k!} \sum_{j=0}^{[k/2]} (-1)^j \binom{k+1}{j} \{(k+1-2j)R - |t|\}_+^k dt \\
&= \frac{2\{(k+1)R\}^{k+1}}{(k+1)!} \sum_{j=0}^{[k/2]} (-1)^j \binom{k+1}{j} \left(1 - \frac{2j}{k+1}\right)^{k+1} = (2R)^{k+1} .
\end{aligned}
$$

Notice that $L_1^{(k)}$ is actually a B-spline of degree k with knots at the points $\{0\} \cup \{\pm(k+1-2j)R\}_{j=0}^{[k/2]}$. We plot the kernels $L_1^{(k)}(t)$ for $k = 1,2,3$, respectively, when $R = 1$ in Figure 1.

Next, we define the d-dimensional product kernel based on (2) as

$$
L_d^{(k)}(x) = \prod_{i=1}^{d} L_1^{(k)}(x_i), \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d . \tag{3}
$$

Since the random trend is supposed to be a repeated local average of a spatial white noise, we define $\mu(x)$ on $\mathbb{R}_+^d$ as

$$
\begin{aligned}
\mu(x) &= \int L_d^{(k)}(x-u) dB(u) \\
&= \int \prod_{i=1}^{d} \frac{1}{(2R)^{k+1}} \frac{1}{k!} \sum_{j=0}^{[k/2]} (-1)^j \binom{k+1}{j} \{(k+1-2j)R - |x_i - u_i|\}_+^k dB(u),
\end{aligned} \tag{4}
$$

where $B(u)$ is the standard Brownian sheet process on $\mathbb{R}_+^d$. Notice that μ is $k-1$ times continuously differentiable on $\mathbb{R}_+^d$. The complexity of μ is determined by the parameters R and k.

Figure 2 shows the realizations of $\mu(x)$ for $d = 1$ and $k = 2$ when R is fixed at 0.5 and 1, respectively. As expected, larger R results in smoother trend μ.

Remark 1. *It should be pointed out that estimation of the underlying trend* μ *does not require the knowledge of the parameters R and k. Local linear or polynomial estimation of* μ *depends on a bandwidth (a tuning parameter) which can be estimated using a Mallows'* C_P (Mallows 1973) *type criterion discussed in* Section 4.

3 Local linear regression

Before we begin this Section, we first briefly describe the main results presented below. The optimal rate of convergence for the local linear estimation of μ in the one-dimensional case (i.e., $d = 1$) is of order $(\min(n, N))^{-4/5}$, and is of order $(\min(n, N))^{-2/3}$ when $d = 2$. In the general case, this rate is of order $(\min(n, N))^{-4/(4+d)}$.

In the nonparametric literature the rate of convergence is $n^{-4/(4+d)}$. However, in our setup the rate is of order $N^{-4/(4+d)}$ if $N \leq n$. As argued in Section 1, N is the effective sample size and this rate given here can be substantially lower that the usual rate $n^{-4/(4+d)}$ Even though we do not explicitly write down the local polynomial estimate of μ, it can be shown that, under appropriate technical assumptions, a local polynomial estimate achieves the rate of $(\min(n, N))^{-2k/(2k+d)}$ where k is the degree of smoothness of the trend μ.

We start with a brief description of the local linear estimation procedure. We use the same notation as in Ruppert & Wand (1994). Let H be a $d \times d$ symmetric positive definite matrix depending on n. Then $H^{1/2}$ is called the bandwidth matrix. For simplicity, we take a diagonal bandwidth matrix so that $H = \mathrm{diag}\left(h_1^2, \ldots, h_d^2\right)$, where $h_i > 0$ for all i. Let K be a d-variate kernel such that $\int K(u)du = 1$. We also write $K_H(u) = |H|^{-1/2}K(H^{-1/2}u)$. We consider the optimization problem:

$$\text{Minimize} \sum_{i=1}^{n} \left\{Z\left(S_i\right) - \beta_0 - \beta_1^T\left(S_i - x\right)\right\}^2 K_H\left(S_i - x\right)$$

with respect to β_0 and β_1. The local linear estimator of the trend at location x is

$$\hat{\mu}(x; H) = \hat{\beta}_0 . \tag{5}$$

Equation (5) can be expressed in a matrix form. Let

$$X_x = \begin{bmatrix} 1 & \left(S_1 - x\right)^T \\ \vdots & \vdots \\ 1 & \left(S_n - x\right)^T \end{bmatrix}.$$

$Z = [Z_1,\ldots,Z_n]^T$, and $W_x = \mathrm{diag}\{K_H(S_1 - x),\ldots, K_H(S_n - x)\}$. Assuming that $X_x^T W_x X_x$ is nonsingular, the local linear estimator (5) is

$$\hat{\mu}(x; H) = e_1^T\left(X_x^T W_x X_x\right)^{-1} X_x^T W_x Z,$$

where e_1 is the $(d + 1) \times 1$ vector having 1 in the first entry and all other entries 0.

We now use the stochastic trend model introduced in Section 2 to study the asymptotic behavior of the MSE of the local linear estimator of the trend. Recall that, if $k \geq 2$ in (4), all first-order derivatives of $\mu(x)$ are continuous. Hence, we choose $k = 2$ in (4). Then the statistical model, which we assume to be correct throughout, is

$$Z(x) = \mu(x) + \varepsilon(x) + \delta(x), \quad \text{where}$$

$$\mu(x) = \int L_d^{(2)}(x-u)dB(u) = \int \prod_{i=1}^{d} \frac{1}{16R^3}\Big\{\big(3R - |x_i - u_i|\big)_+^2 - 3\big(R - |x_i - u_i|\big)_+^2 \Big\} dB(u), \tag{6}$$

$\varepsilon(x)$ is a mean zero, stationary process with $\mathrm{Cov}(\varepsilon(x+h), \varepsilon(x)) = \gamma(Nh)$ and $\gamma(0) = \sigma^2$,

$\delta(x)$ are i.i.d. with mean zero and variance τ^2 and independent of $\varepsilon(x)$.

In order to present the theoretical results, we assume that the locations S_i's are randomly distributed, even though the results would hold if the S_i's follow a reasonably regular design.

We additionally make the following assumptions.

(A1) The kernel K is a compactly supported, bounded kernel such that, $\int uu^T K(u)du = \mu_2(K)I$, where $\mu_2(K)$ is a nonzero scalar and I is the $d \times d$ identity matrix. Also all odd-order moments of K vanish, that is, $\int u_1^{l_1} \dots u_d^{l_d} K(u)du = 0$ for all nonnegative integers $l_1, \dots, l_d$ adding up to an odd number. For simplicity, the multivariate kernel K is taken to be a product kernel based on symmetric univariate kernels.

(A2) S_i are i.i.d. with a common density f having a bounded $\mathrm{supp}(f) \subseteq \mathbb{R}^d$. Without loss of generality, the support of f is taken to be a unit cube in $\mathbb{R}^d$. The point x is in supp(f) and f is continuously differentiable at x with $f(x) > 0$.

(A3) The sequence of bandwidth matrices $H^{1/2}$ where $H = \mathrm{diag}\left(h_1^2, \dots, h_d^2\right)$ is such that each entry of H tends to zero, and $n^{-1}|H|^{-1/2}$ and $N^{-1}|H|^{-1/2} \to 0$ as $n, N \to \infty$. In addition, there is a fixed constant C_H such that the condition number of H is at most C_H for all n, N.

The main results are concerned with the mean squared error characteristics of $\hat{\mu}(x; H)$ when x is an interior point. All the asymptotic results are conditional on the location $\{S_i\}$, and to emphasize this, the mean and variances are denoted by $\mathbb{E}_S$ and Var_S, respectively.

Let $\bar{\mu}(x) = \mathbb{E}_S\{\hat{\mu}(x; H) | \mu\}$. Then $\bar{\mu}(x) - \mu(x)$ is the model bias, and we can get the usual variance plus bias-square decomposition of the mean squared error of $\hat{\mu}(x; H)$ as

$$\mathbb{E}_S\left[\{\hat{\mu}(x; H) - \mu(x)\}^2\right] = \mathbb{E}_S[\mathrm{Var}\{\hat{\mu}(x; H) | \mu\}] + \mathbb{E}_S\left[\{\bar{\mu}(x) - \mu(x)\}^2\right]. \tag{7}$$

In order to describe our asymptotic results, we define the kernel K^* as

$$K^*(u;x) = e_1^T\left(X_x^T W_x X_x\right)^{-1}\left[1, \quad (u-x)^T\right]^T K_H(u-x). \tag{8}$$

Then we can express the local linear estimator given in (5) as

$$\hat{\mu}(x;H) = \sum_{i=1}^{n} K^*\left(S_i;x\right)Z_i. \tag{9}$$

The kernel K^* given in (8) has the well-known properties

$$\sum_{i=1}^{n} K^*\left(S_i;x\right) = 1 \quad \text{and} \quad \sum_{i=1}^{n} K^*\left(S_i;x\right)\left(S_i - x\right) = 0. \tag{10}$$

We use (10) to prove Theorems 1 and 2. Also let $P(K) = \int (K(u))^2 du$.

Theorem 1. *Let x be a fixed point in the interior of supp*(f). *Assume that the model given in* (6) *holds for* $d = 1$. *Also assume that (A1)–(A3) hold. Then*

$$\mathbb{E}_S\left[\{\bar{\mu}(x) - \mu(x)\}^2\right] = \frac{3}{16R^5}\left(\int_0^1 K(t)t^2 dt\right)^2 h^4\left(1 + o_p(1)\right),$$

and

$$\mathbb{E}_S\left[\mathrm{Var}_S\{\hat{\mu}(x;H) | \mu\}\right] = n^{-1}h^{-1}P(K)\left(\sigma^2 + \tau^2\right)/f(x)\left(1 + o_p(1)\right) + N^{-1}h^{-1}P(K)\int \gamma(z)dz\left(1 + o_p(1)\right),$$

where the o_p *terms are with respect to the distribution of the* S_i*'s.*

Remark 2. Theorem 1 shows that the asymptotic mean squared error of the estimated μ is determined by N and n. In this result, as well as Theorem 2 below, the complexity parameter R for the stochastic trend μ is treated as fixed, even though its role in the expression for the leading term in the bias is explicit. Note that the mean squared error

$$D(h) = E\left[\{\hat{\mu}(x;H) - \mu(x)\}^2\right] = \left(q_1 h^4 + q_2 n^{-1}h^{-1} + q_3 N^{-1}h^{-1}\right)\left(1 + o_p(1)\right), \quad \text{where}$$
$$q_1 = \frac{3}{16R^5}\left(\int_0^1 K(t)t^2 dt\right)^2, q_2 = P(K)\left(\sigma^2 + \tau^2\right)/f(x) \quad \text{and} \quad q_3 = P(K)\int \gamma(z)dz,$$

is minimized at

$$h^* = \left(\frac{q_2}{4q_1 n} + \frac{q_3}{4q_1 N}\right)^{1/5}\left(1 + o_p(1)\right),$$

and the smallest asymptotic mean squared error is

$$D(h^*) = q_1^{1/5}\left\{\left(\frac{1}{4}\right)^{4/5} + 4^{1/5}\right\}\left(\frac{q_2}{n} + \frac{q_3}{N}\right)^{4/5}\left\{1 + o_p(1)\right\} \asymp (\min(n, N))^{-4/5}.$$

Thus the optimal rate of convergence of the estimator of μ, as well as the optimal choice of bandwidth, depend on both sample size n and N, the parameter controlling the degree of correlation in the stationary noise. Moreover, these choices, at the level of rates as a function of N and n, are the same as that when μ is a deterministic function with continuous second derivative.

Theorem 2. *Let x be a fixed point in the interior of supp*(f). *Assume that the model given in* (6) *holds for* $d = 2$. *Also assume that (A1)-(A3) hold. Then*

$$\mathbb{E}_S\left[\{\bar{\mu}(x) - \mu(x)\}^2\right] = \left[\frac{33}{160R^6}h_1^4\left(\int_0^1\int_0^1 K(t)t_1^2 dt_1 dt_2\right)^2 + \frac{33}{160R^6}h_2^4\left\{\int_0^1\int_0^1 K(t)t_2^2 dt_1 dt_2\right\}^2 + \frac{1}{8R^6}h_1^2h_2^2\left(\int_0^1\int_0^1 K(t)t_1^2 dt_1 dt_2\right)\left(\int_0^1\int_0^1 K(t)t_2^2 dt_1 dt_2\right)\right]\left(1 + o_p(1)\right),$$

$$\mathbb{E}_S\left[\mathrm{Var}_S\{\hat{\mu}(x; H)|\mu\}\right] = n^{-1}\left(h_1h_2\right)^{-1}P(K)\left(\sigma^2 + \tau^2\right)/f(x)\left(1 + o_p(1)\right) + N^{-1}\left(h_1h_2\right)^{-1}P(K)\int\gamma(z)dz\left(1 + o_p(1)\right).$$

Remark 3. Theorem 2 shows that the asymptotic mean squared error of the estimated μ is determined by both N and n. In particular, the optimal bandwidth is of the order of $(1/n + 1/N)^{1/6}$ and the optimal mean squared error is of the order of $(\min(n, N))^{-2/3}$ as shown below. The mean squared error is

$$\begin{aligned}
D\left(h_1, h_2\right) &= \mathbb{E}_S\left[\left(\hat{\mu}(x; H) - \mu(x)\right)^2\right] \\
&= \left(q_{11}h_1^4 + q_{12}h_2^4 + q_{13}h_1^2h_2^2 + q_{21}n^{-1}\left(h_1h_2\right)^{-1} + q_{22}N^{-1}\left(h_1h_2\right)^{-1}\right)\left\{1 + o_p(1)\right\}, \text{ where} \\
q_{11} &= \frac{33}{160R^6}\left(\int_0^1\int_0^1 K(t)t_1^2 dt_1 dt_2\right)^2, \quad q_{12} = \frac{33}{160R^6}\left(\int_0^1\int_0^1 K(t)t_2^2 dt_1 dt_2\right)^2, \\
q_{13} &= \frac{1}{8R^6}\left(\int_0^1\int_0^1 K(t)t_1^2 dt_1 dt_2\right)\left(\int_0^1\int_0^1 K(t)t_2^2 dt_1 dt_2\right), \\
q_{21} &= P(K)\left(\sigma^2 + \tau^2\right)/f(x), \quad \text{and} \quad q_{22} = P(K)\int\gamma(z)dz.
\end{aligned}$$

It is minimized at the bandwidth

$$h_1^* = \frac{q_{12}^{1/8}}{q_{11}}\left[4\left(q_{11}q_{12}\right)^{1/2} + 2q_{13}\right]^{1/6}\left(\frac{q_{21}}{n} + \frac{q_{22}}{N}\right)^{1/6}\left\{1 + o_p(1)\right\} \text{ and } h_2^* = \frac{q_{11}{}^{1/4}}{q_{12}}{}^{1/4} h_1^*,$$

and the smallest mean squared error is

$$D\left(h_1^*, h_2^*\right) = \left(2^{-2/3} + 2^{1/3}\right)\left[2\left(q_{11}q_{12}\right)^{1/2} + q_{13}\right]^{1/3}\left(\frac{q_{21}}{n} + \frac{q_{22}}{N}\right)^{2/3}\left\{1 + o_p(1)\right\} \asymp (\min(n, N))^{-2/3}.$$

4 Data-driven selection of bandwidth

In this section, we propose a method for selecting the bandwidth matrix $H^{1/2}$. Ideally we would like to select the bandwidth matrix which minimizes

$$L(H) = \sum_{i=1}^{n} E\left\{\hat{\mu}\left(S_i; H\right) - \mu\left(S_i\right)\right\}^2. \quad (11)$$

Since $L(H)$ involves unknown parameters, we cannot use (11) directly. Hence we first obtain a good (meaning, nearly unbiased) estimator of (11) and then minimize the estimate with respect to the bandwidth matrix $H^{1/2}$.

We begin by examining the residual sum of squares, $E\{SSE(H)\} := E\left[\sum_{i=1}^{n}\left\{Z\left(S_i\right) - \hat{\mu}\left(S_i; H\right)\right\}^2\right]$. It is straightforward to show that $E\{SSE(H)\}$ equals $E\left[\sum_{i=1}^{n}\left\{\varepsilon^2\left(S_i\right) + \delta^2\left(S_i\right) + 2\varepsilon\left(S_i\right)\delta\left(S_i\right)\right\}\right] + L(H) - 2E\left[\sum_{i=1}^{n}\left\{\varepsilon\left(S_i\right) + \delta\left(S_i\right)\right\}\left\{\hat{\mu}\left(S_i; H\right) - \mu\left(S_i\right)\right\}\right]$.

Lemma 1. *Assume that the second and third conditions given in* (6) *hold. Then* $\mathbb{E}_S\left[\sum_{i=1}^{n}\left\{\varepsilon\left(S_i\right) + \delta\left(S_i\right)\right\}\left\{\hat{\mu}\left(S_i; H\right) - \mu\left(S_i\right)\right\}\right] = \sum_{i=1}^{n}\sum_{j=1}^{n}\gamma\left(N\left(S_i - S_j\right)\right)K^*\left(S_j; S_i\right)$ *where the* $+ \tau^2\sum_{i=1}^{n}K^*\left(S_i; S_i\right)$
kernel K^* is given by (8) *in case of local linear regression.*

Assuming that we have preliminary estimates $\hat{\gamma}$ and $\hat{\tau}^2$, and using Lemma 1, $L(H)$ can be estimated by

$$SSE(H) + 2\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{\gamma}_{ij}K^*\left(S_j; S_i\right) + 2\hat{\tau}^2\sum_{i=1}^{n}K^*\left(S_i; S_i\right) - E\left[\sum_{i=1}^{n}\left\{\varepsilon^2\left(S_i\right) + \delta^2\left(S_i\right) + 2\varepsilon\left(S_i\right)\delta\left(S_i\right)\right\}\right].$$

Since the last term in the above expression does not involve H, we can ignore it and then obtain the following criterion function

$$\phi(H) = SSE(H) + 2\sum_{i=1}^{n}\sum_{j=1}^{n}\hat{\gamma}_{ij}K^*\left(S_j; S_i\right) + 2\hat{\tau}^2\sum_{i=1}^{n}K^*\left(S_i; S_i\right). \quad (12)$$

We minimize the criterion function $\phi(H)$ over H. If the minimum of $\phi(H)$ is attained at $\widehat{H}$, then we take $\hat{\mu}(x; \widehat{H})$ to be our estimate of the trend.

We briefly discuss how to estimate $\gamma(\cdot)$ and τ^2. Assume that the error component can be modeled using a known parametric form. We first obtain a preliminary estimate of $\mu(x)$ by using a k-nearest neighbor regression where the integer k is not too large, so that the estimate has low bias. In our examples, we used $k \approx n^{2/3}$. If $\tilde{\mu}$ is the preliminary estimate, then we use $Z(S_i) - \tilde{\mu}(S_i)$ as an estimate of the spatially correlated error $\varepsilon(S_i) + \delta(S_i)$. Then we can fit a covariogram model to these residuals, and estimate the corresponding parameters for the covariance function by one of the well-known methods, such as a likelihood-based method or a least squares method.

5 Simulation

We perform a brief simulation study primarily to assess the effectiveness of the bandwidth selection criterion proposed in Section 4. Additionally, we demonstrate the effect of stochastic variation on the accuracy of the trend estimation by local linear regression.

Fifty realizations, each consisting of 200 observations, were generated from model (1) with a deterministic trend $\mu(x) = A\sin(\pi x_1)\sin(\pi x_2)$, and $\varepsilon(x)$ is a mean zero Gaussian process with $\mathrm{Cov}(\varepsilon(x+h),\varepsilon(x)) = \exp(-\theta h)$ and $\delta(x)$ are i.i.d. $N(0, \tau^2)$ independent of $\varepsilon(x)$. For each realization, we obtained an estimator $\hat{\mu}(x; \hat{H})$ and calculated the mean squared error given by

$$\mathrm{MSE}(\hat{\mu}) = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{\mu}(S_i; \hat{H}) - \mu(S_i)\right\}^2.$$

The bandwidth was selected by the method described in Section 4, where we assumed that the stationary error follows the exponential covairogram model (same as that used in the simulation), and estimated the corresponding parameters by the estimation procedure developed in Hyun et al. (2012). The mean and standard deviation of MSE for various combinations of the model parameters are listed in Table 1. The results show that as the value of θ increases, implying smaller degree of correlation among the observations, the accuracy of the estimated trend increases. This is an indication that the bandwidth selection scheme is effective in choosing a suitable estimator of the trend.

We also report the result from a stochastic trend model with μ modeled through $\mu(x) = A\sin(\pi\gamma_1 x_1)\sin(\pi\gamma_2 x_2)$ where the amplitude A and the phase γ_1, γ_2 of the trend follow Chi-Square distributions. We choose $A \sim \chi^2_{40}/20$ and $\gamma_1, \gamma_2 \overset{i.i.d.}{\sim} \chi^2_{20}/20$ so that the expectations of A, γ_1 and γ_2 are 2, 1 and 1, respectively. The mean and standard deviation of MSE for the stochastic trend model are shown in Table 2. We observe that both the mean and standard deviation pf MSE for a stochastic trend model tend to be greater than those for the corresponding deterministic trend model (compare with the first column of Table 1), reflecting the additional variability in the estimates accrued due to randomness. The behavior of the MSE, with respect to the change in the parameter θ (determining spatial correlation among residuals), is qualitatively similar to that in the deterministic trend setting.

6 Analysis of surface temperature anomalies data

As an application of our method to a real problem, we analyzed the data on monthly surface temperature anomalies collected in the region corresponding to the latitude range of 52.5 degrees south to 22.5 degrees north. The data is from the National Oceanic and Atmospheric

Administration (Smith et al. 2008) and can be downloaded from the website http://www.esrl.noaa.gov/psd/data/gridded/data.mlost.html

The data set consists of merged land air and sea surface temperature anomalies on a 5 × 5 grid-box basis spanning 1880 to the present at monthly resolution. We considered the monthly data in March, June, September, and December in 1983, 1993, and 2003. For each monthly data set, roughly 1,150 observations were available. We applied a local linear method to each spatial data set and compared the results. We used the criterion given by (12) to select the bandwidths. We first obtained an estimate of the error using the method described in Section 4. Then we used the residuals after subtracting the trend to fit a Matérn covariance model (Diggle & Ribeiro 2007, Schabenberger & Gotway 2004), with a nugget effect, to account for the observational noise. We set the smoothness parameter of the Matérn covariance to be at $\nu = 1.5$, so that the covariance function is given by $\gamma(u) = \sigma^2(2/\pi)^{0.5}(\|u\|/\phi)^{1.5}K_{1.5}(\|u\|/\phi)$, where $\phi > 0$ is the range parameter, and $K_\nu(\cdot)$ denotes a modified Bessel function of the second kind of order ν. We estimated the parameters σ^2 and ϕ of the covariance model, together with the nugget effect τ^2, by maximizing the likelihood (assuming Gaussianity of the residuals) through a grid search.

Table 3 shows the estimation results along with the mean squared error (MSE) calculated from the estimated trend. The key observations are summarized here, (a) Estimates of ϕ are stable across years and months, indicating the overall pattern of spatial correlation in the small scale variability of the temperature anomalies remains fairly static across time, (b) Estimates of σ^2 are more variable, with higher values in the year 1983, reflecting significant changes in the variability of the temperature anomalies across years, (c) Small values of estimated τ^2 indicate the near absence of observational errors, which is understandable given the precise measuring devices typically used in collecting the observations.

We also display the observed monthly surface temperature anomalies along with the difference between the observed temperature anomalies and the estimated trend in Figure 3. These plots show that the proposed estimation and bandwidth selection method is able to capture the large scale component of the temperature anomalies quite effectively. The only portion showing significant residual effect corresponds to a patch around longitude 300° and latitude between 0° and 40° south, which corresponds to the west coast of South America. Here, the changes in the temperature profile are rather sharp and the selected bandwidth appears to oversmooth the estimated trend in this region.

7 Proofs

Proof of Theorem 1: Let $D_g(x)$ denotes the $d \times 1$ vector of first-order partial derivatives of a sufficiently smooth d-variate function g at x. Also **1** denotes a generic matrix with each entry equal to 1, the dimensions of which will be determined in the context. For a random matrix U_n, $O_p(U_n)$ and $o_p(U_n)$ are to be taken componentwise.

Using (9) and (10), we can get

$$\begin{aligned}\mathbb{E}_S[\hat{\mu}(x;H)\big|\mu] &= \sum_{i=1}^{n} K^*(S_i;x)\mu(S_i)\\ &= \sum_{i=1}^{n} K^*(S_i;x)\Big[\mu(x)+\dot{\mu}(x)^T(S_i-x)+\Big\{\mu(S_i)-\mu(x)-\dot{\mu}(x)^T(S_i-x)\Big\}\Big]\\ &= \mu(x)+\sum_{i=1}^{n} K^*(S_i;x)\Big\{\mu(S_i)-\mu(x)-\dot{\mu}(x)^T(S_i-x)\Big\}.\end{aligned}$$

Recall from (6) that $\mu(x) = \int L_d^{(2)}(x-u)dB(u)$. For simplicity we write $L^{(2)}(x-u)$ to denote $L_d^{(2)}(x-u)$. Hence

$$\begin{aligned}&\mathbb{E}_S[\hat{\mu}(x;H)|\mu]-\mu(x) = \sum_{i=1}^{n} K^*(S_i;x)\Big\{\int L^{(2)}(S_i-u)dB(u)-\int L^{(2)}(x-u)dB(u)-\\ &\int \dot{L}^{(2)}(x-u)^T(S_i-x)dB(u)\Big\}\\ &= \sum_{i=1}^{n} K^*(S_i;x)\int\Big\{L^{(2)}(S_i-u)-L^{(2)}(x-u)-\dot{L}^{(2)}(x-u)^T(S_i-x)\Big\}dB(u).\end{aligned}$$

Let $R(x)$ be the $n \times 1$ vector given by

$$\begin{aligned}&R(x)\\ &=\Big[\int\Big\{L^{(2)}(S_1-u)-L^{(2)}(x-u)-\dot{L}^{(2)}(x-u)^T(S_1-x)\Big\}dB(u),\ldots,\int\Big\{L^{(2)}(S_n-u)-L^{(2)}(x-u)-\dot{L}^{(2)}(x-u)^T(S_n-x)\Big\}dB(u)\Big]^T\\ &.\end{aligned}$$

Then

$$\mathbb{E}_S[\hat{\mu}(x;H)|\mu]-\mu(x) = e_1^T\left(X_x^T W_x X_x\right)^{-1} X_x^T W_x R(x),$$

where X_x and W_x are given in Section 3.

Ruppert and Wand (1994) showed that

$$\left(n^{-1}X_x^T W_x X_x\right)^{-1} = \begin{bmatrix} f(x)^{-1}+o_p(1) & -D_f(x)^T f(x)^{-2}+o_p(1)\\ -D_f(x)f(x)^{-2}+o_p(1) & \{\mu_2(K)f(x)H\}^{-1}+o_p\left(H^{-1}\right)\end{bmatrix}. \qquad (13)$$

It is also easily seen that

$$n^{-1}X_x^T W_x R(x) = \begin{bmatrix} n^{-1}\sum_{i=1}^{n} K_H(S_i - x)\int\left\{L^{(2)}(S_i - u) - L^{(2)}(x - u) - \dot{L}^{(2)}(x - u)^T(S_i - x)\right\}dB(u) \\ n^{-1}\sum_{i=1}^{n} K_H(S_i - x)\int\left\{L^{(2)}(S_i - u) - L^{(2)}(x - u) - \dot{L}^{(2)}(x - u)^T(S_i - x)\right\}dB(u)(S_i - x) \end{bmatrix} \tag{14}$$

and

$$n^{-1}\sum_{i=1}^{n} K_H(S_i - x)\int\left\{L^{(2)}(S_i - u) - L^{(2)}(x - u) - \dot{L}^{(2)}(x - u)^T(S_i - x)\right\}dB(u)(S_i - x) = \int K(t)$$
$$\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x - u) - \dot{L}^{(2)}(x - u)^T H^{1/2}t\right\}dB(u) \times \left(H^{1/2}t\right)f\left(x + H^{1/2}t\right)dt\left(1 + o_p(\mathbf{1})\right).$$

It follows from (13) and (14) that

$$\begin{aligned} &\mathbb{E}_S[\hat{\mu}(x;H)|\mu] - \mu(x) = f(x)^{-1}n^{-1}\sum_{i=1}^{n} K_H(S_i - x) \\ &\int\left\{L^{(2)}(S_i - u) - L^{(2)}(x - u) - \dot{L}^{(2)}(x - u)^T(S_i - x)\right\}dB(u) \times \left(1 + o_p(\mathbf{1})\right) \\ &= f(x)^{-1}\int K(t)\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x - u) - \dot{L}^{(2)}(x - u)^T H^{1/2}t\right\}dB(u \\ &) \times f\left(x + H^{1/2}t\right)dt\left(1 + o_p(1)\right) \\ &= \int K(t)\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x - u) - \dot{L}^{(2)}(x - u)^T H^{1/2}t\right\}dB(u)dt \\ &\left(1 + o_p(1)\right). \end{aligned} \tag{15}$$

Recall from (6) that, for $d = 1$,

$$L^{(2)}(x - u) = \frac{1}{16R^3}\left\{(3R - |x - u|)_+^2 - 3(R - |x - u|)_+^2\right\}. \tag{16}$$

It follows from (15) and (16) that

$$\int K(t)\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x - u) - \dot{L}^{(2)}(x - u)^T H^{1/2}t\right\}dB(u)dt \tag{17}$$

is a sum of integrals of $K(t)$ and component of the kernel $L^{(2)}(x-u)$ over various rectangular regions corresponding to the points of non-differentiability of $\dot{L}^{(2)}$. The dominant term in (17) is given by

$$\frac{1}{16R^3}\int_{x-3R+h}^{x-R}\int_0^1 K(t)h^2t^2dtdB(u) + \frac{1}{16R^3}\int_{x-R+h}^{x} K(t)\left(-2h^2t^2\right)dtdB(u)$$
$$\frac{1}{16R^3}\int_{x+h}^{x+R}\int_0^1 K(t)\left(-2h^2t^2\right)dtdB(u) + \frac{1}{16R^3}\int_{x+R+h}^{x+3R}\int_0^1 K(t)h^2t^2dtdB(u)$$
$$\frac{1}{16R^3}\int_{x-3R}^{x-R-h}\int_{-1}^0 K(t)h^2t^2dtdB(u) + \frac{1}{16R^3}\int_{-1}^0 K(t)\left(-2h^2t^2\right)dtdB(u) \tag{18}$$
$$\frac{1}{16R^3}\int_{x}^{x+R-h}\int_{-1}^0 K(t)\left(-2h^2t^2\right)dtdB(u) + \frac{1}{16R^3}\int_{x+R}^{x+3R-h}\int_{-1}^0 K(t)h^2t^2dtdB$$
$$(u)\,.$$

One important component in obtaining the leading order terms in the expression for the squared bias $\mathbb{E}_S\left[\left(\mathbb{E}_S[\hat{\mu}(x;H)|\mu] - \mu(x)\right)^2\right]$ is to use the following “identities”

$$\mathbb{E}\left[(dB(u))^2\right] = du \quad \text{and} \quad \mathbb{E}[dB(u)dB(v)] = 0 \quad \text{for} \quad u \neq v\,. \tag{19}$$

Then we can show that the squared bias is given by

$$\mathbb{E}_S\left[\left(\mathbb{E}_S[\hat{\mu}(x;H)|\mu] - \mu(x)\right)^2\right] = \frac{3}{16R^5}\left(\int_0^1 K(t)t^2dt\right)^2 h^4\left(1 + o_p(1)\right).$$

For the variance, we first calculate $\mathbb{E}_S[\mathrm{Var}[\hat{\mu}(x;H)|\mu]]$ for $x \in \mathbb{R}$. Let V be the $n \times n$ variance-covariance matrix given by

$$V = \begin{bmatrix} \sigma^2+\tau^2 & \gamma\left(N\left(S_1-S_2\right)\right) & \dots & \gamma\left(N\left(S_1-S_n\right)\right) \\ \gamma\left(N\left(S_2-S_1\right)\right) & \sigma^2+\tau^2 & \dots & \gamma\left(N\left(S_2-S_n\right)\right) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma\left(N\left(S_n-S_1\right)\right) & \gamma\left(N\left(S_n-S_2\right)\right) & \dots & \sigma^2+\tau^2 \end{bmatrix}.$$

Then

$$\mathbb{E}_S\left[\mathrm{Var}_S[\hat{\mu}(x;H)|\mu]\right] = e_1^T\left(X_x^TW_xX_x\right)^{-1}X_x^TW_xVW_xX_x\left(X_x^TW_xX_x\right)^{-1}e_1\,.$$

The upper-left entry of $n^{-2}X_x^TW_xVW_xX_x$ is

$$\frac{\sigma^2+\tau^2}{n^2}\sum_{i=1}^{n} K_H(S_i-x)^2+\frac{1}{n^2}\sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n} K_H(S_i-x)K_H(S_j-x)\gamma\left(N\left(S_i-S_j\right)\right)$$
$$=\frac{\sigma^2+\tau^2}{n}|H|^{-1/2}\int K^2(u)f\left(x+H^{1/2}u\right)du\left(1+o_p(1)\right)+\iint K(u)K(v)\gamma\left(NH^{1/2}(u-v)\right)f\left(x+H^{1/2}u\right)f\left(x+H^{1/2}v\right)dudv\left(1+o_p(1)\right)$$
$$=n^{-1}|H|^{-1/2}P(K)\left(\sigma^2+\tau^2\right)f(x)\left(1+o_p(1)\right)+N^{-1}|H|^{-1/2}P(K)\int\gamma(z)dzf^2(x)\left(1+o_p(1)\right), \tag{20}$$

the upper-right block is

$$\frac{\sigma^2+\tau^2}{n^2}\sum_{i=1}^{n}(S_i-x)^TK_H(S_i-x)^2+\frac{1}{n^2}\sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n}(S_i-x)^TK_H(S_i-x)K_H\left(S_j-x\right)\gamma\left(N\left(S_i-S_j\right)\right)=\frac{\sigma^2+\tau^2}{n}|H|^{-1/2}\int u^TH^{1/2}K^2(u)f\left(x+H^{1/2}u\right)du\left(1+o_p(1)\right)$$
$$+\iint u^TH^{1/2}K(u)K(v)\gamma\left(NH^{1/2}(u-v)\right)f\left(x+H^{1/2}u\right)f\left(x+H^{1/2}v\right)dudv\times\left(1+o_p(1)\right)$$
$$=\frac{\sigma^2+\tau^2}{n}|H|^{-1/2}\int u^TH^{1/2}K^2(u)D_f(x)^TH^{1/2}udu\left(1+o_p(1)\right)+N^{-2}|H|^{-1/2}\int K(v)v^TH^{1/2}D_k(v)^TH^{-1/2}dv\int z\gamma(z)dzf^2(x)\left(1+o_p(1)\right)+N^{-2}|H|^{-1/2}P(K)\int z^T\gamma(z)dzf^2(x)\left(1+o_p(1)\right)$$
$$=O_p\left(n^{-1}|H|^{-1/2}\mathbf{1}H\right)+O_p\left(N^{-2}|H|^{-1/2}\mathbf{1}\right), \tag{21}$$

and the lower-right block is

$$\frac{\sigma^2+\tau^2}{n^2}\sum_{i=1}^{n}\left(S_i-x\right)\left(S_i-x\right)^T K_H\left(S_i-x\right)^2$$

$$+\frac{1}{n^2}\sum_{j=1}^{n}\sum_{\substack{i=1\\ i\neq j}}^{n}\left(S_j-x\right)K_H\left(S_j-x\right)\left(S_i-x\right)^T K_H\left(S_i-x\right)\gamma\left(N\left(S_i-S_j\right)\right)$$

$$=\frac{\sigma^2+\tau^2}{n}|H|^{-1/2}\int H^{1/2}uu^T H^{1/2}K^2(u)f\left(x+H^{1/2}u\right)du\left(1+o_p(1)\right)$$

$$+\iint H^{1/2}vK(v)K(u)u^T H^{1/2}\gamma\left(NH^{1/2}(u-v)\right)f\left(x+H^{1/2}u\right)f\left(x+H^{1/2}v\right)dudv \tag{22}$$

$$\times\left(1+o_p(1)\right)$$

$$=\frac{\sigma^2+\tau^2}{n}|H|^{-1/2}H^{1/2}\int K^2(u)uu^T du H^{1/2}f(x)\left(1+o_p(1)\right)+N^{-1}|H|^{-1/2}\int\gamma(z$$

$$)dzH^{1/2}\int K^2(v)vv^T dv H^{1/2}f^2(x)\left(1+o_p(1)\right).$$

So, using (13) and (20)–(22), we obtain

$$\mathbb{E}_S[\mathrm{Var}[\hat{\mu}(x;H)|\mu]]=n^{-1}|H|^{-1/2}P(K)\left(\sigma^2+\tau^2\right)/f(x)\left(1+o_p(1)\right)+N^{-1}|H|^{-1/2}P(K)\int\gamma(z)dz\left(1+o_p(1)\right). \tag{23}$$

Now, let $d = 1$. Then $|H| = h^2$. Hence,

$$\mathbb{E}_S[\mathrm{Var}[\hat{\mu}(x;H)|\mu]]=n^{-1}h^{-1}R(K)\left(\sigma^2+\tau^2\right)/f(x)\left\{1+o_p(1)\right\}+N^{-1}h^{-1}R(K)\int\gamma(z)dz\left(1+o_p(1)\right).$$

□

Proof of Theorem 2: It follows from (15) that, for $d = 2$,

$$\mathbb{E}_S[\hat{\mu}(x;H)|\mu] - \mu(x) = \int K(t)\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x-u) - \dot{L}^{(2)}(x-u)^T H^{1/2}t\right\}dB(u)dt\left(1 + o_p(1)\right)$$
$$= \left[\int_0^1\int_0^1 K(t)\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x-u) - \dot{L}^{(2)}(x-u)^T H^{1/2}t\right\}dB(u)dt_1dt_2 + \int_0^1\int_{-1}^0 K(t)\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x-u) - \dot{L}^{(2)}(x-u)^T H^{1/2}t\right\}dB(u)dt_1dt_2 + \int_{-1}^0\int_0^1 K(t)\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x-u) - \dot{L}^{(2)}(x-u)^T H^{1/2}t\right\}dB(u)dt_1dt_2 + \int_{-1}^0\int_{-1}^0 K(t)\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x-u) - \dot{L}^{(2)}(x-u)^T H^{1/2}t\right\}dB(u)dt_1dt_2\right] \times \left(1 + o_p(1)\right).$$

Hence,

$$\mathbb{E}_S\left[\left(\mathbb{E}_S[\hat{\mu}(x;H)|\mu] - \mu(x)\right)^2\right] = \mathbb{E}_S\left[\left\{\int_0^1\int_0^1 K(t)\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x-u) - \dot{L}^{(2)}(x-u)^T H^{1/2}t\right\}dB(u)dt_1dt_2\right\}^2\right] \quad (24)$$

$$+\mathbb{E}_S\left[\left\{\int_0^1\int_{-1}^0 K(t)\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x-u) - \dot{L}^{(2)}(x-u)^T H^{1/2}t\right\}dB(u)dt_1dt_2\right\}^2\right] \quad (25)$$

$$+\mathbb{E}_S\left[\left\{\int_{-1}^0\int_0^1 K(t)\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x-u) - \dot{L}^{(2)}(x-u)^T H^{1/2}t\right\}dB(u)dt_1dt_2\right\}^2\right] \quad (26)$$

$$+\mathbb{E}_S\left[\left\{\int_{-1}^0\int_{-1}^0 K(t)\int\left\{L^{(2)}\left(x + H^{1/2}t - u\right) - L^{(2)}(x-u) - \dot{L}^{(2)}(x-u)^T H^{1/2}t\right\}dB(u)dt_1dt_2\right\}^2\right] \quad (27)$$

$$+2\mathbb{E}_S\Bigg[\Bigg\{\int_0^1\int_0^1 K(t)\int\Big\{L^{(2)}\Big(x+H^{1/2}t-u\Big)-L^{(2)}(x-u)-\dot{L}^{(2)}(x-u)^TH^{1/2}t \\ \Big\}dB(u)dt_1dt_2\Bigg\}\times\Bigg\{\int_0^1\int_{-1}^0 K(t)\int\Big\{L^{(2)}\Big(x+H^{1/2}t-u\Big)-L^{(2)}(x-u) \\ -\dot{L}^{(2)}(x-u)^TH^{1/2}t\Big\}dB(u)dt_1dt_2\Bigg\}\Bigg] \tag{28}$$

$$+2\mathbb{E}_S\Bigg[\Bigg\{\int_0^1\int_0^1 K(t)\int\Big\{L^{(2)}\Big(x+H^{1/2}t-u\Big)-L^{(2)}(x-u)-\dot{L}^{(2)}(x-u)^TH^{1/2}t \\ \Big\}dB(u)dt_1dt_2\Bigg\}\times\Bigg\{\int_{-1}^0\int_0^1 K(t)\int\Big\{L^{(2)}\Big(x+H^{1/2}t-u\Big)-L^{(2)}(x-u) \\ -\dot{L}^{(2)}(x-u)^TH^{1/2}t\Big\}dB(u)dt_1dt_2\Bigg\}\Bigg] \tag{29}$$

$$+2\mathbb{E}_S\Bigg[\Bigg\{\int_0^1\int_0^1 K(t)\int\Big\{L^{(2)}\Big(x+H^{1/2}t-u\Big)-L^{(2)}(x-u)-\dot{L}^{(2)}(x-u)^TH^{1/2}t \\ \Big\}dB(u)dt_1dt_2\Bigg\}\times\Bigg\{\int_{-1}^0\int_{-1}^0 K(t)\int\Big\{L^{(2)}\Big(x+H^{1/2}t-u\Big)-L^{(2)}(x-u) \\ -\dot{L}^{(2)}(x-u)^TH^{1/2}t\Big\}dB(u)dt_1dt_2\Bigg\}\Bigg] \tag{30}$$

$$+2\mathbb{E}_S\Bigg[\Bigg\{\int_0^1\int_{-1}^0 K(t)\int\Big\{L^{(2)}\Big(x+H^{1/2}t-u\Big)-L^{(2)}(x-u)-\dot{L}^{(2)}(x-u)^TH^{1/2}t \\ \Big\}dB(u)dt_1dt_2\Bigg\}\times\Bigg\{\int_{-1}^0\int_0^1 K(t)\int\Big\{L^{(2)}\Big(x+H^{1/2}t-u\Big)-L^{(2)}(x-u) \\ -\dot{L}^{(2)}(x-u)^TH^{1/2}t\Big\}dB(u)dt_1dt_2\Bigg\}\Bigg] \tag{31}$$

$$+2\mathbb{E}_S\Bigg[\Bigg\{\int_0^1\int_{-1}^0 K(t)\int\Big\{L^{(2)}\Big(x+H^{1/2}t-u\Big)-L^{(2)}(x-u)-\dot{L}^{(2)}(x-u)^TH^{1/2}t \\ \Big\}dB(u)dt_1dt_2\Bigg\}\times\Bigg\{\int_{-1}^0\int_{-1}^0 K(t)\int\Big\{L^{(2)}\Big(x+H^{1/2}t-u\Big)-L^{(2)}(x-u) \\ -\dot{L}^{(2)}(x-u)^TH^{1/2}t\Big\}dB(u)dt_1dt_2\Bigg\}\Bigg] \tag{32}$$

$$+2\mathbb{E}_S\Bigg[\Bigg\{\int_{-1}^0\int_0^1 K(t)\int\Big\{L^{(2)}\Big(x+H^{1/2}t-u\Big)-L^{(2)}(x-u)-\dot{L}^{(2)}(x-u)^TH^{1/2}t \\ \Big\}dB(u)dt_1dt_2\Bigg\}\times\Bigg\{\int_{-1}^0\int_{-1}^0 K(t)\int\Big\{L^{(2)}\Big(x+H^{1/2}t-u\Big)-L^{(2)}(x-u) \\ -\dot{L}^{(2)}(x-u)^TH^{1/2}t\Big\}dB(u)dt_1dt_2\Bigg\}\Bigg] \tag{33}$$

By expanding (24)–(33), repeatedly using the identities (19) and then collecting only the terms that involve the leading order, we can show that the squared bias is given by

$$\mathbb{E}_S\Big[\big(\mathbb{E}_S[\hat{\mu}(x;H)\big|\mu] - \mu(x)\big)^2\Big] = \Bigg[\frac{33}{160R^6}h_1^4\Big\{\int_0^1\int_0^1 K(t)t_1^2dt_1dt_2\Big\}^2 + \frac{33}{160R^6}h_2^4\Big(\int_0^1\int_0^1 K(t)t_2^2dt_1dt_2\Big)^2 + \frac{1}{8R^6}h_1^2h_2^2\Big(\int_0^1\int_0^1 K(t)t_1^2dt_1dt_2\Big)\Big(\int_0^1\int_0^1 K(t)t_2^2dt_1dt_2\Big)\Bigg]\big(1 + o_p(1)\big).$$

For the variance, (23) shows that, for $x \in \mathbb{R}$,

$$\mathbb{E}_S\big[\mathrm{Var}_S[\hat{\mu}(x;H)|\mu]\big] = n^{-1}|H|^{-1/2}R(K)\big(\sigma^2 + \tau^2\big)/f(x)\big(1 + o_p(1)\big) + N^{-1}|H|^{-1/2}R(K)\int\gamma(z)dz\big(1 + o_p(1)\big).$$

Now take $d = 2$. Then $|H| = h_1^2h_2^2$. Thus it follows that

$$\mathbb{E}_S\big[\mathrm{Var}_S[\hat{\mu}(x;H)|\mu]\big] = n^{-1}\big(h_1h_2\big)^{-1}R(K)\big(\sigma^2 + \tau^2\big)/f(x)\big(1 + o_p(1)\big) + N^{-1}\big(h_1h_2\big)^{-1}R(K)\int\gamma(z)dz\big(1 + o_p(1)\big).$$

□

Proof of Lemma 1: Using (9), we get

$$\bar{\mu}(x) = \mathbb{E}_S[\hat{\mu}(x;H)|\mu] = \sum_{i=1}^{n} K^*\big(S_i;x\big)\mu\big(S_i\big).$$

Thus

$$\begin{aligned}
\mathbb{E}_S\Bigg[\sum_{i=1}^{n}\big\{\varepsilon\big(S_i\big) + \delta\big(S_i\big)\big\}\big\{\hat{\mu}\big(S_i;H\big) - \mu\big(S_i\big)\big\}\Bigg] &= \mathbb{E}_S\Bigg[\sum_{i=1}^{n}\big\{\varepsilon\big(S_i\big) + \delta\big(S_i\big)\big\}\big\{\hat{\mu}\big(S_i;H\big) - \bar{\mu}\big(S_i\big)\big\}\Bigg] \\
&= \mathbb{E}_S\Bigg[\sum_{i=1}^{n}\big\{\varepsilon\big(S_i\big) + \delta\big(S_i\big)\big\}\sum_{j=1}^{n}K^*\big(S_j;S_i\big)\big\{Z\big(S_j\big) - \mu\big(S_j\big)\big\}\Bigg] \\
&= \mathbb{E}_S\Bigg[\sum_{i=1}^{n}\sum_{j=1}^{n}\big\{\varepsilon\big(S_i\big) + \delta\big(S_i\big)\big\}\big\{\varepsilon\big(S_j\big) + \delta\big(S_j\big)\big\}K^*\big(S_j;S_i\big)\Bigg] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}_S\big[\varepsilon\big(S_i\big),\varepsilon\big(S_j\big)\big]K^*\big(S_j;S_i\big) + \tau^2\sum_{i=1}^{n}K^*\big(S_i;S_i\big) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\gamma\big(N\big(S_i - S_j\big)\big)K^*\big(S_j;S_i\big) + \tau^2\sum_{i=1}^{n}K^*\big(S_i;S_i\big).
\end{aligned}$$

□

Acknowledgments

Research of the Prabir Burman is partially supported by the NSF grants DMS-9108295, DMS-0907622, DMS-1148643 and NSA grant H98230-04-1-0109. Research of Debashis Paul is partially supported by the NSF grant DMS-1407530 and the NIH grant 1R01EB021707.

References

Box GE, Jenkins GM & Reinsel GC (1994), Time Series Analysis: Forecasting and Control, Prentice Hall, Englewood Cliffs, NJ.

Burman P (1991), 'Regression function estimation from dependent observations', Journal of Multivariate Analysis 36, 263–279.

Burman P & Shumway RH (2009), 'Estimation of trend in state-space models: Asymptotic mean square error and rate of convergence', The Annals of Statistics 37, 3715–3742.

Diggle P & Ribeiro PJ (2007), Model-based Geo statistics, Springer Science & Business Media, New York.

Durbin J & Koopman SJ (2001), Time series analysis by state space methods, Oxford University Press, Oxford.

Harvey AC (1991), Forecasting, structural time series models and the Kalman filter, Cambridge university press, Cambridge.

Hyun JW, Burman P & Paul D (2012), 'A regression approach for estimating the parameters of the covariance function of a stationary spatial random process', Journal of Statistical Planning and Inference 142, 2330–2344.

Mallows CL (1973), 'Some comments on C_p' Technometrics 15, 661–675.

Ruppert D & Wand MP (1994), 'Multivariate locally weighted least squares regression', The Annals of Statistics 22, 1346–1370.

Schabenberger O & Gotway CA (2004), Statistical Methods for Spatial Data Analysis, Chapman and Hall/CRC, Boca Raton, FL.

Shumway RH & Stoffer DS (2000), Time series analysis and its applications, Springer, New York.

Smith TM, Reynolds RW, Peterson TC & Lawrimore J (2008), 'Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006)', Journal of Climate 21, 2283–2296.

Wood SA, Jiang W & Tanner M (2002), 'Bayesian mixture of splines for spatially adaptive nonparametric regression', Biometrika 89, 513–528.

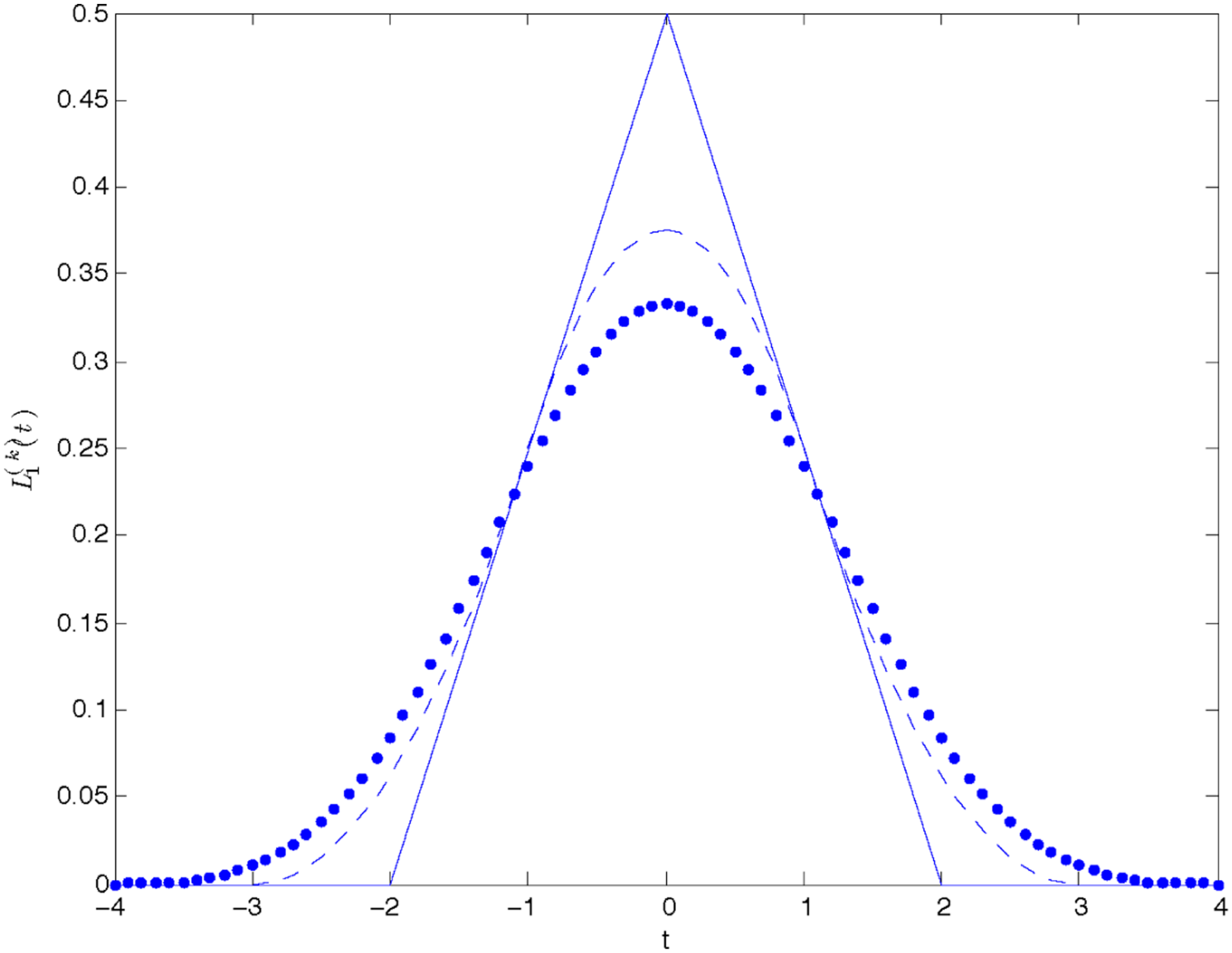

Figure 1:
Univariate kernels $L_1^{(k)}(t)$ for $R = 1$. Solid line corresponds to $L_1^{(1)}(t) = 0.25(2 - |t|)_+$, dashed line $L_1^{(2)}(t) = \left\{(3 - |t|)_+^2 - 3(1 - |t|)_+^2\right\}/16$, and dotted line $L_1^{(3)}(t) = \left\{(4 - |t|)_+^3 - 4(2 - |t|)_+^3\right\}/96$.

Author Manuscript
Author Manuscript
Author Manuscript

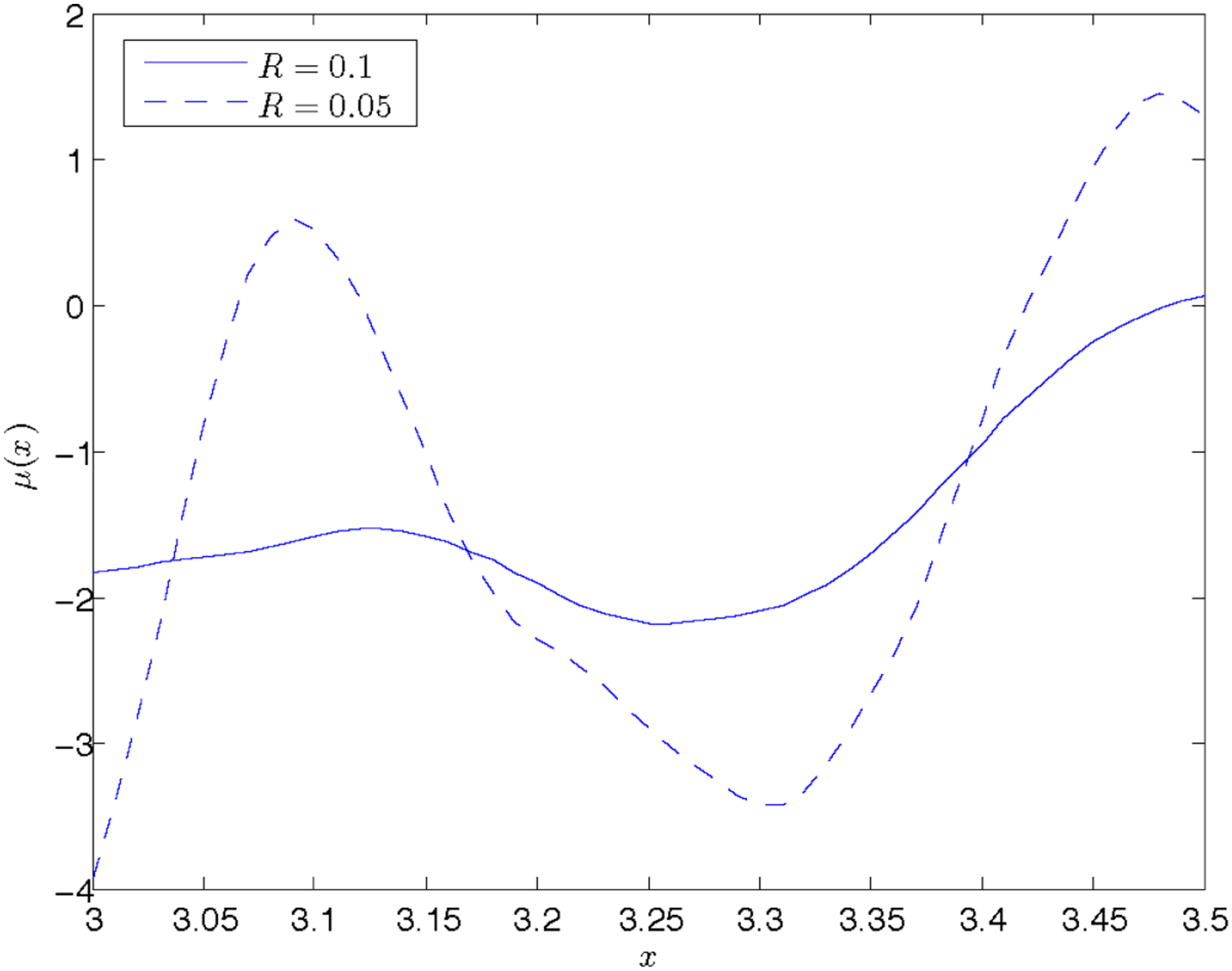

Figure 2:
Realizations of the random trend μ for $d = 1$ and $k = 2$ when $R = 0.1$ (solid line) and $R = 0.05$ (dashed line).

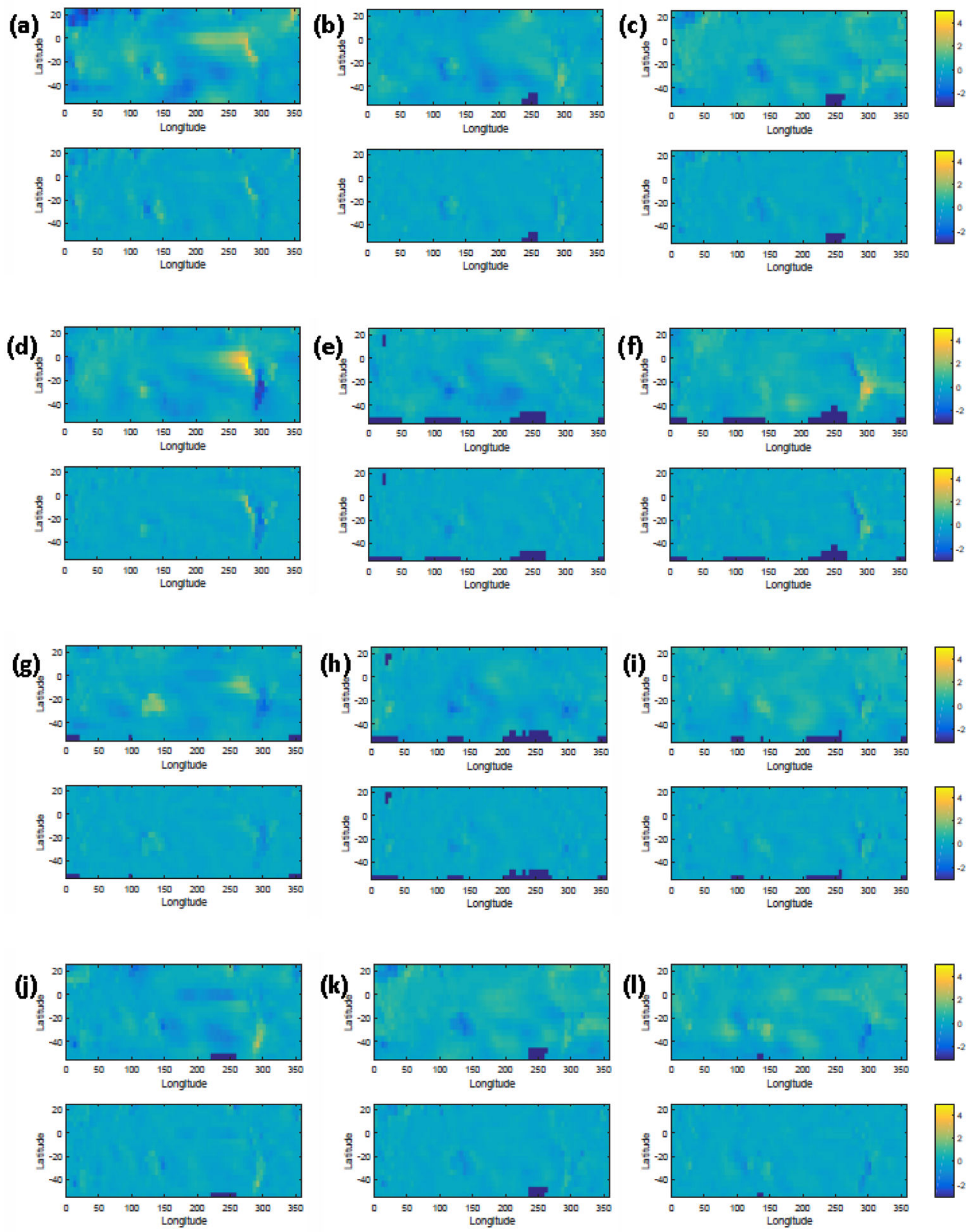

Figure 3:

The observed temperature anomalies (upper panel) and the difference between the observed temperature anomalies and the estimated trend (bottom panel) for (a) March in 1983, (b) March in 1993, (c) March in 2003, (d) June in 1983, (e) June in 1993, (f) June in 2003, (g) September in 1983, (h) September in 1993, (i) September in 2003, (j) December in 1983, (k) December in 1993, and (l) December in 2003.

Table 1:

MSE for the deterministic trend model

		A=2	A=8
		Mean (SD)	**Mean (SD)**
$\theta = 8$	$\tau^2 = 0.1$	0.4886 (0.1796)	0.5840 (0.2033)
	$\tau^2 = 0.4$	0.5008(0.1891)	0.6208 (0.2156)
$\theta = 32$	$\tau^2 = 0.1$	0.1705 (0.0596)	0.2760 (0.0718)
	$\tau^2 = 0.4$	0.1991 (0.0692)	0.3173 (0.0800)

Table 2:

MSE for the stochastic trend model

		Mean (SD)
$\theta = 8$	$\tau^2 = 0.1$	0.8933 (0.2398)
$\theta = 32$	$\tau^2 = 0.1$	0.6293 (0.1014)

Table 3:

Estimation results for temperature anomalies data. The selected bandwidths h_1 and h_2 correspond to longitude and latitude, respectively.

Year	Month	$\hat{\phi}$	$\hat{\sigma}^2$	$\hat{\tau}^2$	h_1	h_2	MSE
1983	March	3.95	0.15	0.01	14.30	3.08	0.0884
	June	4.48	0.14	0.02	8.71	4.62	0.0796
	September	4.17	0.07	0.01	12.14	2.77	0.0360
	December	3.61	0.09	0	14.23	3.01	0.0569
1993	March	4.17	0.06	0.01	11.12	3.70	0.0333
	June	4.04	0.05	0	12.48	3.04	0.0276
	September	3.85	0.05	0	12.34	3.50	0.0274
	December	3.86	0.07	0	13.35	3.46	0.0385
2003	March	3.88	0.08	0.01	14.02	3.18	0.0433
	June	4.04	0.09	0.01	10.72	5.32	0.0569
	September	4.18	0.07	0.01	13.94	5.25	0.0449
	December	4.15	0.07	0.01	11.44	3.26	0.0365