

## TESTING ADDITIVE ASSUMPTIONS ON MEANS OF REGULAR MONITORING DATA: A MULTIVARIATE NONSTATIONARY TIME SERIES APPROACH

Ting Zhang

*Boston University*

*Abstract:* In the analysis of surface meteorological data, observations are usually recorded regularly and frequently in time at multiple but fixed locations in space. The data can thus be viewed as multivariate time series in which a small number of lengthy time series is observed. Motivated by a temperature data, the current paper considers the problem of testing the additive assumption of location and time effects via a multivariate time series approach. Test statistics based on both maximum absolute and integrated squared deviations are proposed and their asymptotic properties are studied for a general class of multivariate nonstationary processes. The results are illustrated in a simulation study and applied to temperature data.

*Key words and phrases:* Local linear estimation, multivariate nonlinear time series, nonparametric hypothesis testing, nonstationary processes, regular monitoring data.

### 1. Introduction

For meteorological and environmental data, it is common that observations are recorded at regular and frequent intervals in time on a fixed monitoring network in space; see for example the daily Irish wind data that is collected at 12 locations, considered by Haslett and Raftery (1989) and Stein (2005), and the minute-by-minute atmospheric pressure data that is collected at 13 monitoring sites in north-central Oklahoma, studied by Stein (2009). Let  $X(\mathbf{u}, t)$  denote the underlying spatial-temporal process which is observed at  $(\mathbf{u}_i, t_j)$  for locations  $\mathbf{u}_1, \dots, \mathbf{u}_p$  and times  $t_1, \dots, t_n$ . Based on the available data points, an important task is to provide predictions of the underlying process at locations and times for which no observations are available, and to generate meteorological maps at a broad range of spatial and temporal scales. In general, this goal is achieved by modeling the underlying data generating mechanism by statistical models that impose certain structural assumptions on the underlying process. The current paper aims to provide rigorous statistical tests for an additive assumption that has been employed in the literature.

The problem is motivated by temperature data provided by the Atmospheric Radiation Measurement (ARM) Program established by the U.S. Department of Energy. It contains minute-by-minute surface measurements of air temperature at 15 monitoring sites in the Southern Great Plains region of the United States; see Guinness and Stein (2013) for more details. They analyzed the data from the first 30 days of October, 2005, and developed methods for interpolating available data to locations and times for which no observations were available. Despite the slow-moving cold front on day 5, they proposed to model the underlying spatial-temporal process as

$$X(\mathbf{u}, t) = s(\mathbf{u}) + m(t) + Y(\mathbf{u}, t), \quad (1.1)$$

where  $s(\cdot)$  and  $m(\cdot)$  are the spatial and temporal mean functions respectively, and  $Y(\mathbf{u}, t)$  is the residual term. Model (1.1) assumes that the spatial and temporal effects are additive, and was used by Guinness and Stein (2013) to model the temperature data during the period from 10/06/2005 to 10/30/2005. Hence, for the  $i$ -th monitoring site, by (1.1), the mean function is parallel to  $m(t)$  with a constant shift  $s(\mathbf{u}_i)$  caused by the site effect. As commented by Stein (2005) and Guinness and Stein (2013), regular monitoring data on a fixed monitoring network can be viewed as multivariate time series, in which a small number of lengthy time series is observed. This motivates us to consider the problem of testing for parallelism among mean trends in multiple time series.

There is a huge literature on comparing mean functions in the regression setting. Härdle and Marron (1990) compared two regression curves by testing whether they are equal up to a parametric transformation of the axes. Hall and Hart (1990) proposed a bootstrap method for testing the equality of two or several mean functions in the nonparametric regression setting. The latter problem has also been studied by King, Hart, and Wehrly (1991), Delgado (1993), Kulasekera (1995), Young and Bowman (1995), Munk and Dette (1998), Dette and Neumeyer (2001), Neumeyer and Dette (2003) and Park and Kang (2008), among others. Much of this work considers models with independent errors, and thus is not suitable for analyzing time series data where dependence is the rule rather than the exception. Fan and Lin (1998) considered the problem of testing the equality of mean functions for stationary linear error processes; see also Li (2006). However, the assumption of stationarity can also be quite strong in practice and, as argued by Guinness and Stein (2013), a flexible nonstationary time series model is needed for the aforementioned temperature data to accurately capture its statistical properties. Zhou and Wu (2010) demonstrated that the existence of nonstationarity can have a large effect on the distribution of the test statistic, and should be taken into account to avoid erroneous conclusions. They considered the problem of constructing simultaneous confidence tubes for

coefficient function estimators of linear models. We focus here on the problem of testing for mean trend parallelism among multiple nonstationary time series.

Recently, Degras et al. (2012) considered phone download data and proposed a test for determining whether the mean download trends after a logarithmic transformation in different area codes are identical up to vertical shifts. A crucial assumption underlying their test is that the error processes for different area codes are independent and identically distributed (i.i.d.): the time series  $\{X(\mathbf{u}_1, t)\}_t$  and  $\{X(\mathbf{u}_2, t)\}_t$  observed at different locations  $\mathbf{u}_1 \neq \mathbf{u}_2$  are completely independent and have the same joint distribution. Hence, their method is not able to capture the spatial dependence of the observed data that is a key feature of the aforementioned temperature data; see for example Guinness and Stein (2013). As seen in the simulation results of Section 4.1, ignoring the spatial dependence among observations collected at different monitoring sites and applying the test of Degras et al. (2012) can yield misleading  $p$ -values leading to erroneous conclusions. This motivates us to consider statistically valid tests that can survive under the temporal and spatial dependence contemplated here. Another limitation caused by the i.i.d. assumption of Degras et al. (2012) is that their error processes are spatially homogeneous:  $[X(\mathbf{u}_1, t) - E\{X(\mathbf{u}_1, t)\}]_t$  and  $[X(\mathbf{u}_2, t) - E\{X(\mathbf{u}_2, t)\}]_t$  have the same (joint) distribution for any  $\mathbf{u}_1 \neq \mathbf{u}_2$ . This can also limit the applicability of their test, as different measurement devices can be used at different monitoring sites leading to different types of measurement errors. We seek statistically valid tests that allow inhomogeneous error processes across different monitoring sites with an unknown and possibly time-varying spatial dependence structure, so that our results can be widely applicable.

The rest of the paper is organized as follows. Section 2 introduces the framework and basic assumptions. Section 3 contains our proposed test statistics and their asymptotic properties. Section 4 provides applications of the proposed methods to simulated and data examples. Proofs are deferred to the Supplementary Material.

## 2. Multivariate Nonstationary Processes

Given observations

$$X_{i,j} = X(\mathbf{u}_i, t_j), \quad i = 1, \dots, p, \quad j = 1, \dots, n,$$

we seek a rigorous statistical test for the additive assumption in (1.1), under which the mean functions for different monitoring sites are parallel to each other with the vertical shifts caused by the site effect. In order to capture the nonstationary feature, in addition to the spatial and temporal dependence, of the temperature data as described in Guinness and Stein (2013), we consider modeling the joint process  $\mathbf{X}_j = (X_{1,j}, \dots, X_{p,j})^\top$ ,  $j = 1, \dots, n$ , as a multivariate

nonstationary process. The problem of modeling nonstationary processes has been studied in the literature by means of spectral representations (Priestley (1965), Dahlhaus (1996, 1997)), pseudo-differential operators (Mallat, Papanicolaou, and Zhang (1998)), discrete non-decimated wavelets (Nason, von Sachs, and Kroisandt (2000)) and smooth localized complex exponentials (Ombao, von Sachs, and Guo (2005)). Other contributions can be found in Cheng and Tong (1998), Giurcanu and Spokoiny (2004), and Draghicescu, Guillas, and Wu (2009), among others. We follow the framework of Zhou and Wu (2010) and assume that

$$\mathbf{X}_j = \mathbf{G}\left(\frac{j}{n}; \mathcal{F}_j\right), \quad \mathcal{F}_j = (\dots, \epsilon_{j-1}, \epsilon_j), \quad (2.1)$$

where  $\epsilon_k$ ,  $k \in \mathbb{Z}$ , are i.i.d. random vectors, and  $\mathbf{G}$  is a measurable function. If the function  $\mathbf{G}$  does not depend on time, then (2.1) reduces to a multivariate stationary process with causal representation

$$\mathbf{X}_j = \mathbf{G}(\mathcal{F}_j), \quad \mathcal{F}_j = (\dots, \epsilon_{j-1}, \epsilon_j);$$

this includes linear processes, bilinear processes, Volterra processes and many other time series models as special cases; see for example Wiener (1958), Tong (1990) and Wu (2005). Hence, (2.1) provides a natural generalization of a wide range of existing time series models to their nonstationary counterparts. In addition, by allowing the joint process as a time-varying nonlinear transformation of past innovations as in (2.1), components of the vector  $\mathbf{X}_j$  can be correlated in a time-varying and complicated fashion.

We introduce the functional dependence measure that enables us to develop an asymptotic theory for complicated statistics of time series. For a random vector  $\mathbf{X}$ , write  $\mathbf{X} \in \mathcal{L}^q$ ,  $q > 0$ , if  $\|\mathbf{X}\|_q = \{E(|\mathbf{X}|^q)\}^{1/q} < \infty$  where  $|\cdot|$  denotes the Euclidean norm, and let  $\|\cdot\| = \|\cdot\|_2$ . Let  $(\epsilon_k^*)_{k \in \mathbb{Z}}$  be an i.i.d. copy of  $(\epsilon_k)_{k \in \mathbb{Z}}$ , and  $\mathcal{F}_k^* = (\mathcal{F}_{-1}, \epsilon_0^*, \epsilon_1, \dots, \epsilon_k)$  be the coupled shift process, the functional dependence measure is defined as

$$\theta_{k,q} = \sup_{t \in [0,1]} \|\mathbf{G}(t; \mathcal{F}_k) - \mathbf{G}(t; \mathcal{F}_k^*)\|_q. \quad (2.2)$$

According to the idea of coupling, the quantity  $\theta_{k,q}$  measures the  $k$ -th step ahead impact of the current innovation. Let  $\Theta_{m,q} = \sum_{k=m}^{\infty} \theta_{k,q}$ , then  $\Theta_{0,q}$  measures the cumulative impact of the current innovation. We need some assumptions in establishing the main results.

(A1)  $\Theta_{0,2} < \infty$  and  $\sup_{t \in [0,1]} \|\mathbf{G}(t; \mathcal{F}_k)\|_q < \infty$  for some  $q > 2$ .

(A2) There exists a constant  $c_0 < \infty$  such that  $\|\mathbf{G}(t_1; \mathcal{F}_k) - \mathbf{G}(t_2; \mathcal{F}_k)\| \leq c_0|t_1 - t_2|$  uniformly for all  $t_1, t_2 \in [0, 1]$ .

(A3) The smallest eigenvalue of  $\mathbf{\Lambda}(t) = \sum_{l \in \mathbb{Z}} \text{cov}\{\mathbf{G}(t; \mathcal{F}_0), \mathbf{G}(t; \mathcal{F}_l)\}$  is bounded away from zero on  $[0, 1]$ .

Condition (A1) is the short range dependence condition, under which the long-run covariance  $\mathbf{\Lambda}(t)$  is uniformly bounded over  $t \in [0, 1]$ . Condition (A2) can be interpreted as the stochastically Lipschitz continuous condition, under which the underlying data generating mechanism evolve smoothly over time; see Zhang and Wu (2011) for more discussion.

### 3. Main Results

#### 3.1. Nonparametric estimation

For the observed multivariate time series  $\mathbf{X}_j = \mathbf{G}(j/n; \mathcal{F}_j)$ ,  $j = 1, \dots, n$ , we write

$$\mathbf{X}_j = \boldsymbol{\mu}\left(\frac{j}{n}\right) + \mathbf{e}_j,$$

where  $\boldsymbol{\mu}(t) = \{\mu_1(t), \dots, \mu_p(t)\}^\top$ ,  $t \in [0, 1]$ , is the unknown mean function, and  $(\mathbf{e}_j)_{j=1}^n$  is the corresponding zero-mean error process. If the parallelism assumption holds, then  $\mu_i(\cdot) - \mu_1(\cdot) \equiv c_i$  for some constant  $c_i \in \mathbb{R}$  for all  $i = 2, \dots, p$ , and vice versa. Let  $\mathbf{A} \in \mathbb{R}^{r \times p}$  be a prespecified matrix with rank  $r \leq p$  whose nonzero entries are bounded away from zero. We consider testing the more general null hypothesis

$$H_0 : \mathbf{A}\boldsymbol{\mu}(t) = \mathbf{a}, \quad t \in [0, 1], \tag{3.1}$$

for some vector  $\mathbf{a} \in \mathbb{R}^r$ . For example, if

$$\mathbf{A} = \begin{pmatrix} -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{(p-1) \times p}, \tag{3.2}$$

then (3.1) is equivalent to testing the parallelism assumption. In this case,  $\mathbf{a} \in \mathbb{R}^{p-1}$  is unknown and needs to be estimated, while the case with  $\mathbf{a}$  a vector of zeros corresponds to testing for the equality of mean functions.

A natural strategy for testing the null hypothesis (3.1) is to compare the estimator obtained under the null with a nonparametric estimator which is consistent under both the null and the alternative, and reject the null hypothesis if their difference is statistically significant. Li, Genton, and Sherman (2007) considered a similar form to (3.1) on spatial covariances by comparing parametric covariance estimators with a vector of zeros; see Bliznyuk et al. (2012) for recent developments on parameterizing spatial covariance functions. We focus on the mean trend function and use nonparametric methods without specifying

any parametric covariance structure. There are a number of ways of performing nonparametric estimation of  $\boldsymbol{\mu}(\cdot)$  including kernel, spline, and wavelet methods. We use the local linear estimator (Fan and Gijbels (1996))

$$\hat{\boldsymbol{\mu}}_n(t) = \sum_{j=1}^n w_{j,n}(t) \mathbf{X}_j, \quad (3.3)$$

where  $w_{j,n}(t) = K\{(j/n-t)/b_n\}\{S_2(t) - (t-j/n)S_1(t)\}/\{S_2(t)S_0(t) - S_1^2(t)\}$  are the local linear weights,  $b_n$  is the bandwidth,  $K(\cdot)$  is a kernel function and  $S_l(t) = \sum_{j=1}^n (t-j/n)^l K\{(j/n-t)/b_n\}$ . Throughout, we assume that the kernel function  $K(\cdot)$  is a symmetric and bounded function in  $\mathcal{C}^1[-1, 1]$  with  $\int_{-1}^1 K(v)dv = 1$ . For example, it can be the Epanechnikov kernel  $K(v) = 3 \max(0, 1 - v^2)/4$ , or the Bartlett kernel  $K(v) = \max(0, 1 - |v|)$ . If the vector  $\mathbf{a}$  in (3.1) is unknown, one can estimate it by

$$\hat{\mathbf{a}}_n = \int_0^1 \mathbf{A} \hat{\boldsymbol{\mu}}_n(t) dt. \quad (3.4)$$

A similar averaging estimator as in (3.4) was used by Zhang, Lee, and Song (2002) for estimating semivarying coefficient models with i.i.d. observations. Take  $\kappa_2 = \int_{-1}^1 v^2 K(v)dv$ ,  $K_2 = \int_{-1}^1 K(v)^2 dv$  and  $\boldsymbol{\Sigma}(t) = \mathbf{A} \boldsymbol{\Lambda}(t) \mathbf{A}^\top$ ,  $t \in [0, 1]$ . Theorem 1 provides central limit theorems for estimators (3.3) and (3.4). It is remarkable that the estimator (3.4) can achieve the parametric convergence rate.

**Theorem 1.** *Assume (A1), (A2) and  $\boldsymbol{\mu} \in \mathcal{C}^3[0, 1]$ . If  $nb_n \rightarrow \infty$  and  $nb_n^7 \rightarrow 0$ , then for any  $t \in (0, 1)$  as  $n \rightarrow \infty$ ,*

$$(nb_n)^{1/2} \{ \hat{\boldsymbol{\mu}}_n(t) - \boldsymbol{\mu}(t) - b_n^2 \kappa_2 \frac{\boldsymbol{\mu}''(t)}{2} \} \Rightarrow N\{0, K_2 \boldsymbol{\Lambda}(t)\},$$

and

$$n^{1/2} \{ \hat{\mathbf{a}}_n - E(\hat{\mathbf{a}}_n) \} \Rightarrow N \left\{ 0, \int_0^1 \boldsymbol{\Sigma}(t) dt \right\},$$

where  $E(\hat{\mathbf{a}}_n) = \mathbf{a}$  under (3.1).

Under the null hypothesis (3.1),  $\mathbf{A} \boldsymbol{\mu}(\cdot) \equiv \mathbf{a}$  and  $\mathbf{A} \boldsymbol{\mu}''(t) = 0$ ,  $t \in [0, 1]$ . By Theorem 1,

$$(nb_n)^{1/2} \{ \mathbf{A} \hat{\boldsymbol{\mu}}_n(t) - \hat{\mathbf{a}}_n \} \Rightarrow N\{0, K_2 \boldsymbol{\Sigma}(t)\}.$$

Let  $\mathbf{D}_n(t) = \boldsymbol{\Sigma}(t)^{-1/2} \{ \mathbf{A} \hat{\boldsymbol{\mu}}_n(t) - \hat{\mathbf{a}}_n \}$  be the normalized difference between  $\mathbf{A} \hat{\boldsymbol{\mu}}_n(t)$  and  $\hat{\mathbf{a}}_n$ , we reject the null hypothesis (3.1) at level  $\alpha \in (0, 1)$  if

$$|\mathbf{D}_n(t)| > (nb_n)^{-1/2} K_2^{1/2} z_{1-\alpha, r}, \quad (3.5)$$

where  $z_{1-\alpha, r}$  is the  $(1 - \alpha)$ -th quantile of the Euclidean norm of a standard  $r$ -dimensional multivariate normal distribution. The test (3.5) depends on the

choice of  $t \in (0, 1)$ , while the time point at which the null hypothesis (3.1) is the least likely to hold is often unknown in practice. We consider the  $\mathcal{L}^\infty$ - and  $\mathcal{L}^2$ -norms of the discrepancy function (3.5), and study their asymptotic properties in Sections 3.2 and 3.3, respectively.

### 3.2. The maximum absolute deviation

With  $\mathbf{A} \in \mathbb{R}^{r \times p}$  a prespecified matrix with rank  $r \leq p$ , and  $\mathbf{D}_n(t)$  the normalized difference between  $\mathbf{A}\hat{\boldsymbol{\mu}}_n(t)$  and  $\hat{\mathbf{a}}_n$ . Theorem 2 states that, after proper centering and scaling, the maximum absolute deviation has the asymptotic extreme value distribution. The maximum absolute deviation was used by Fan and Zhang (2000) to construct simultaneous confidence bands for state-domain varying coefficient models with i.i.d. observations, and we here consider nonstationary and dependent observations.

**Theorem 2.** *Assume (A1)–(A3) and  $\theta_{n,4} = O(n^{-\iota})$  for some  $\iota \geq 2$ . If  $b_n(\log n)^2 \rightarrow 0$  and  $nb_n^{5/2}/(\log n)^{15/2} \rightarrow \infty$ , then under (3.1) as  $n \rightarrow \infty$ ,*

$$\text{pr} \left\{ \left( \frac{2nb_n \log m_n}{K_2} \right)^{1/2} \sup_{t \in [b_n, 1-b_n]} |\mathbf{D}_n(t)| - B_K(m_n) \leq u \right\} \rightarrow \exp\{-2 \exp(-u)\},$$

where  $m_n = 1/b_n$  and

$$B_K(m_n) = 2 \log m_n + \log C_K + \left\{ \frac{(r-1)}{2} \right\} \log \log m_n - \log 2$$

with  $C_K = \Gamma(r/2)^{-1} \{ (K_2\pi)^{-1} \int_{-1}^1 |K'(v)|^2 dv \}^{1/2}$ .

By Theorem 2, we reject the null hypothesis (3.1) at level  $\alpha \in (0, 1)$  if the maximum absolute deviation from the null satisfies

$$\sup_{t \in [b_n, 1-b_n]} |\mathbf{D}_n(t)| > \frac{K_2^{1/2}}{(2nb_n \log m_n)^{1/2}} \left[ B_K(m_n) - \log \log \{ (1-\alpha)^{-1/2} \} \right]. \quad (3.6)$$

To investigate the power of this testing procedure, we consider the local alternative

$$H_1 : \mathbf{A}\boldsymbol{\mu}(t) = \mathbf{a} + d_n \mathbf{f}(t), \quad t \in [0, 1], \quad (3.7)$$

where  $\mathbf{f} \in \mathcal{C}^2[0, 1]$  is a known nonconstant function, and  $d_n \rightarrow 0$  is a sequence of positive real numbers indicating the magnitude of the deviation from (3.1). For two positive real sequences  $s_n$  and  $t_n$ , we say that  $s_n \gg t_n$  if  $s_n/t_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Proposition 1 states that the test (3.6) can detect alternatives with  $d_n \gg (nb_n)^{-1/2}(\log m_n)^{1/2}$ .

**Proposition 1.** *Assume conditions of Theorem 2. If  $(nb_n/\log m_n)^{1/2}d_n \rightarrow \infty$ , then under the alternative hypothesis (3.7) as  $n \rightarrow \infty$ ,*

$$\text{pr} \left( \sup_{t \in [b_n, 1-b_n]} |\mathbf{D}_n(t)| > \frac{K_2^{1/2}}{(2nb_n \log m_n)^{1/2}} \left[ B_K(m_n) - \log \log \{(1-\alpha)^{-1/2}\} \right] \right) \rightarrow 1.$$

### 3.3. The integrated squared deviation

Recall that the kernel function  $K(\cdot)$  is a symmetric and bounded function in  $\mathcal{C}^1[-1, 1]$  with  $\int_{-1}^1 K(v)dv = 1$ . Let

$$K^*(x) = \int_{-1}^{1-2|x|} K(v)K(v+2|x|)dv,$$

then  $K^* \in \mathcal{C}^1[-1, 1]$  and is also symmetric. Let  $K_2^* = \int_{-1}^1 K^*(v)^2 dv$ .

**Theorem 3.** *Assume (A1)–(A3) and  $\theta_{n,4} = O(n^{-\iota})$  for some  $\iota > 2$ . If  $b_n \rightarrow 0$  and  $nb_n^{3/2} \rightarrow \infty$ , then under the null hypothesis (3.1) as  $n \rightarrow \infty$ ,*

$$\text{pr} \left\{ \frac{nb_n^{1/2}}{(4rK_2^*)^{1/2}} \int_0^1 |\mathbf{D}_n(t)|^2 dt - \frac{r^{1/2}K^*(0)}{(4b_nK_2^*)^{1/2}} \leq u \right\} \rightarrow \Phi(u),$$

where  $\Phi(\cdot)$  is the distribution function of the standard normal.

Accordingly, we reject the null hypothesis (3.1) at level  $\alpha \in (0, 1)$  if

$$\int_0^1 |\mathbf{D}_n(t)|^2 dt > (nb_n)^{-1}rK^*(0) + n^{-1}b_n^{-1/2}(4rK_2^*)^{1/2}q_{1-\alpha}, \quad (3.8)$$

where  $q_{1-\alpha}$  is the  $(1-\alpha)$ -th quantile of the standard normal. This is an alternative to the test at (3.6). We here investigate the asymptotic power of the test (3.8) under the local alternative (3.7).

**Proposition 2.** *Assume the conditions of Theorem 3. If  $nb_n^{1/2}d_n^2 \rightarrow \infty$ , then under the alternative hypothesis (3.7) as  $n \rightarrow \infty$ ,*

$$\text{pr} \left\{ \int_0^1 |\mathbf{D}_n(t)|^2 dt > (nb_n)^{-1}rK^*(0) + n^{-1}b_n^{-1/2}(4rK_2^*)^{1/2}q_{1-\alpha} \right\} \rightarrow 1.$$

With this, the test at (3.8) has unit asymptotic power if  $d_n \gg (nb_n^{1/2})^{-1/2}$ . As a comparison, the test at (3.6) based on the maximum absolute deviation can detect alternatives with departure rate satisfying  $d_n \gg (nb_n)^{-1/2}(\log m_n)^{1/2}$ , where  $m_n = 1/b_n$ . As  $n \rightarrow \infty$ ,

$$\frac{(nb_n^{1/2})^{-1/2}}{(nb_n)^{-1/2}(\log m_n)^{1/2}} = b_n^{1/4}(-\log b_n)^{-1/2} \rightarrow 0, \quad (3.9)$$

which suggests that the  $\mathcal{L}^2$ -based test (3.8) is asymptotically more powerful than its  $\mathcal{L}^\infty$ -based counterpart. Nevertheless, the convergence speed in (3.9) can be quite slow. For example, if one uses the popular bandwidth  $b_n = cn^{-1/5}$  for some  $0 < c < \infty$  that has the asymptotic mean integrated squared error (AMISE) optimal rate, then the rate in (3.9) is  $n^{-1/20}(\log n)^{-1/2}$  which approaches zero quite slowly as  $n \rightarrow \infty$ . Therefore, the superiority of the  $\mathcal{L}^2$ -based test is ambiguous for finite sample problems, and we carried out simulation studies (see Section 4.1) to examine the finite sample power performance of the two tests.

### 3.4. Bandwidth selection

The problem of selecting an optimal bandwidth for nonparametric estimation and testing problems has been extensively studied in the literature; see for example Härdle and Marron (1985), Härdle, Hall, and Marron (1988), Ruppert, Sheather, and Wand (1995), Xia (1998), and Gao and Gijbels (2008), among many others. Most of the existing results are developed for independent observations, while Opsomer, Wang, and Yang (2001) demonstrated that bandwidth selectors developed under independence can often break down for dependent data. The problem is even more complicated in the current setting because of the underlying nonstationarity and spatial dependence. As commented by Wang (2008), there exists no uniform guidance for an optimal choice of bandwidth. On the positive side, the simulation results in Section 4.1 suggest that the empirical acceptance probabilities are not so sensitive to the choice of the bandwidth. Hence, one can simply choose  $b_n = n^{-1/5}$  which has the AMISE optimal rate. Unlike data-driven bandwidth selectors, this does not introduce an extra amount of randomness into the testing procedure; the asymptotic results developed in Sections 3.1–3.3 remain unharmed and theoretically rigorous tests are feasible. In addition, it produces practically reasonable bandwidths for nonparametric hypothesis testing problems, as suggested by Zhang and Wu (2011).

As an alternative, we consider the generalized cross-validation (GCV) methods of Craven and Wahba (1979), and correct the dependence by incorporating estimates of the covariance matrices into the criterion function, as suggested by Wang (1998). In particular, if  $\boldsymbol{\zeta}_i^\top = (\zeta_{i,1}, \dots, \zeta_{i,n})$  is the  $i$ -th row of the data matrix  $(\mathbf{A}\mathbf{X}_1, \dots, \mathbf{A}\mathbf{X}_n)$ ,  $i = 1, \dots, r$ , then the local linear estimate with bandwidth  $b$  can be written as  $\hat{\boldsymbol{\zeta}}_i(b) = \mathbf{H}(b)\boldsymbol{\zeta}_i$ , where  $\mathbf{H}(b)$  is the corresponding hat matrix. We choose the bandwidth  $\hat{b}_n$  by minimizing

$$\text{GCV}(b) = \frac{\sum_{i=1}^r \{\hat{\boldsymbol{\zeta}}_i(b) - \boldsymbol{\zeta}_i\}^\top \boldsymbol{\Upsilon}_i^{-1} \{\hat{\boldsymbol{\zeta}}_i(b) - \boldsymbol{\zeta}_i\}}{[1 - \text{tr}\{\mathbf{H}(b)\}/n]^2}, \quad (3.10)$$

where  $\boldsymbol{\Upsilon}_i = \{\text{cov}(\zeta_{i,k}, \zeta_{i,l})\}_{1 \leq k, l \leq n}$  is the covariance matrix of  $\boldsymbol{\zeta}_i$ ,  $i = 1, \dots, r$ ; this can be estimated by using the banding technique in Bickel and Levina (2008a)

and Wu and Pourahmadi (2009). Simulation results in Section 4.1 suggest that the GCV selector (3.10) works reasonably well as it is able to capture the actual degree of smoothness of the underlying function. In practice we suggest choosing the bandwidth by searching the minimizer of the GCV criterion (3.10) over the range  $[c^{-1}n^{-1/5}, cn^{-1/5}]$  with  $c = 3/2$ .

### 3.5. Covariance matrix estimation

The implementation of the tests (3.6) and (3.8) involves unknown quantities depending on the covariance matrix  $\mathbf{\Lambda}(t)$ ,  $t \in [0, 1]$ . The problem of estimating covariance matrices has been extensively studied in the literature; see for example Newey and West (1987), Andrews (1993), Lumley and Heagerty (1999), and Zhou and Wu (2010), among others. We use the estimate proposed by Zhang and Wu (2012), obtained by applying techniques of banding and local linear smoothing. Asymptotic properties including the consistency and convergence rate can be found in Zhang and Wu (2012). For completeness, we present the estimate. Let  $\tau_n$  and  $\varrho_n$  be tuning parameter sequences satisfying  $\tau_n \rightarrow 0$ ,  $\varrho_n \rightarrow 0$ , and  $n\tau_n\varrho_n \rightarrow \infty$ . Let  $\mathcal{I}_{n,1} = [0, \tau_n\varrho_n]$ ,  $\mathcal{I}_{n,2} = (\tau_n\varrho_n, 1 - \tau_n\varrho_n)$ ,  $\mathcal{I}_{n,3} = [1 - \tau_n\varrho_n, 1]$ , and

$$\boldsymbol{\lambda}_j = \begin{cases} \mathbf{e}_j \mathbf{e}_j^\top + 2\mathbf{e}_j \sum_{k=1}^n \mathbf{e}_k^\top \mathbf{I}(0 < \frac{k}{n} - \frac{j}{n} \leq \tau_n\varrho_n), & \text{if } \frac{j}{n} \in \mathcal{I}_{n,1}; \\ \mathbf{e}_j \sum_{k=1}^n \mathbf{e}_k^\top \mathbf{I}(|\frac{k}{n} - \frac{j}{n}| \leq \tau_n\varrho_n), & \text{if } \frac{j}{n} \in \mathcal{I}_{n,2}; \\ \mathbf{e}_j \mathbf{e}_j^\top + 2\mathbf{e}_j \sum_{k=1}^n \mathbf{e}_k^\top \mathbf{I}(0 < \frac{j}{n} - \frac{k}{n} \leq \tau_n\varrho_n), & \text{if } \frac{j}{n} \in \mathcal{I}_{n,3}, \end{cases}$$

where  $\mathbf{I}(\cdot)$  denotes the indicator function. We estimate the long-run covariance matrices by

$$\hat{\mathbf{\Lambda}}_n(t) = \sum_{j=1}^n w_{j,n,\tau_n}(t) \frac{(\boldsymbol{\lambda}_j + \boldsymbol{\lambda}_j^\top)}{2}, \quad t \in [0, 1], \quad (3.11)$$

where  $w_{j,n,\tau_n}(t)$  is the local linear weight with bandwidth  $\tau_n$ . In computing (3.11), since the error sequence  $(\mathbf{e}_j)_{j=1}^n$  is usually not observable, we replace it by the local linear residuals  $(\hat{\mathbf{e}}_j)_{j=1}^n$ , and refer to Zhang and Wu (2012) for a discussion on selecting the tuning parameters  $\tau_n$  and  $\varrho_n$ .

### 3.6. A simulation-assisted testing procedure

An important observation of Theorems 2 and 3 is that both the maximum absolute deviation and the integrated squared deviation are asymptotically pivotal, namely their asymptotic distributions do not depend on quantities related to the underlying process. This motivates us to consider a simulation-assisted testing procedure that can substantially improve the finite sample performance.

- (i) Select the bandwidth  $\hat{b}_n$  by minimizing the GCV criterion (3.10).
- (ii) Obtain the local linear estimate  $\hat{\boldsymbol{\mu}}_n(t)$ ,  $t \in [0, 1]$ , compute the residuals  $\hat{\boldsymbol{e}}_j = \boldsymbol{X}_j - \hat{\boldsymbol{\mu}}_n(j/n)$ ,  $j = 1, \dots, n$ , and  $\hat{\boldsymbol{\Lambda}}_n(t)$ ,  $t \in [0, 1]$ , by (3.11).
- (iii) Compute  $\boldsymbol{D}_n(t)$ ,  $t \in [0, 1]$ , and form the test statistics  $T_{n,\infty} = \sup_{t \in [0,1]} |\boldsymbol{D}_n(t)|$  and  $T_{n,2} = \int_0^1 |\boldsymbol{D}_n(t)|^2 dt$ .
- (iv) Generate i.i.d. standard multivariate normal random vectors  $\boldsymbol{X}_j^\circ$ , and compute the corresponding  $T_{n,\infty}^\circ$  and  $T_{n,2}^\circ$  in the same manner as in steps (ii) and (iii).
- (v) Let  $\alpha \in (0, 1)$  be the chosen significance level. Repeat step (iv) and obtain the estimated quantiles  $\hat{q}_{1-\alpha,\infty}^\circ$  of  $T_{n,\infty}^\circ$  and  $\hat{q}_{1-\alpha,2}^\circ$  of  $T_{n,2}^\circ$ .
- (vi) Reject the null hypothesis (3.1) at level  $\alpha$  if  $T_{n,\infty} > \hat{q}_{1-\alpha,\infty}^\circ$  based on the maximum absolute deviation, or  $T_{n,2} > \hat{q}_{1-\alpha,2}^\circ$  based on the integrated squared deviation.

## 4. Numerical Experiments

### 4.1. Simulation studies

Degras et al. (2012) considered the problem of testing for mean trend parallelism among multiple time series  $(X_{i,j})_{j=1}^n$ ,  $i = 1, \dots, p$ , where the asymptotic distribution of their test statistic  $T_{n,i.i.d.}$  is approximated by its counterpart  $T_{n,i.i.d.}^\circ$  for i.i.d. Gaussian processes (Wu and Zhou (2011)). Let  $e_{i,j} = X_{i,j} - E(X_{i,j})$  denote the measurement error, a key assumption in Degras et al. (2012) is that the error processes  $(e_{i,j})_{j=1}^n$ ,  $i = 1, \dots, p$ , are i.i.d.. As commented by Guinness and Stein (2013), this i.i.d. assumption can be quite strong and is easily violated for spatial-temporal processes. We provide a comparison of the proposed methods with the test of Degras et al. (2012).

Let  $\boldsymbol{\epsilon}_k = (\epsilon_{1,k}, \dots, \epsilon_{p,k})$ ,  $k \in \mathbb{Z}$ , be i.i.d. random vectors with independent Rademacher components satisfying  $\text{pr}(\epsilon_{i,k} = -1) = \text{pr}(\epsilon_{i,k} = 1) = 1/2$ . Let  $\boldsymbol{M} = (0.25^{|i-j|})_{1 \leq i,j \leq p}$ , then  $\boldsymbol{\xi}_k = \boldsymbol{M}\boldsymbol{\epsilon}_k$ ,  $k \in \mathbb{Z}$ , forms a sequence of i.i.d. random vectors but with correlated components. Let  $L_i(t)$  be the  $i$ -th order Legendre polynomial, and  $\boldsymbol{L}(t) \in \mathbb{R}^{p \times p}$  be the diagonal matrix with  $i$ -th diagonal component  $L_i(2t - 1)/4$ . Then

$$\boldsymbol{e}_j = \sum_{k=0}^{\infty} \boldsymbol{L}\left(\frac{j}{n}\right)^k \boldsymbol{\xi}_{j-k}, \quad j = 1, \dots, n, \tag{4.1}$$

is a sequence of dependent and nonstationary noises with correlated components. With  $n = 1,000$ ,  $p = 10$ , and the bandwidth  $b_n = n^{-1/5} = 0.25$ , we consider testing whether the mean trend functions of  $(e_{i,j})_{j=1}^n$ ,  $i = 1, \dots, p$ , are parallel to each other. The Epanechnikov kernel was used throughout our numerical

experiments. The cut-off values were obtained by 5,000 simulated  $T_{n,\infty}^\circ$ ,  $T_{n,2}^\circ$ , and  $T_{n,\text{i.i.d.}}^\circ$ , and the results are summarized in Figure 1 based on 5,000 realizations of  $T_{n,\infty}$ ,  $T_{n,2}$ , and  $T_{n,\text{i.i.d.}}$ . From Figure 1 (c), we can see that the test of Degras et al. (2012) fails to provide valid  $p$ -values when the error processes are not i.i.d.. The proposed tests based on the maximum absolute deviation as in Figure 1 (a) and the integrated squared deviation as in Figure 1 (b) perform reasonably well. At nominal levels (90%, 95%, 99%), the empirical acceptance probabilities are (89.8%, 94.2%, 98.7%) for Figure 1 (a) and (91.4%, 95.2%, 98.9%) for Figure 1 (b).

Although both proposed tests survive in the presence of spatial dependence and inhomogeneity, which is more powerful for finite sample problems is ambiguous due to the slow convergence in (3.9). We examine the finite sample power of the proposed two tests via simulation studies. Let  $\mathbf{X}_j = \boldsymbol{\mu}(j/n) + \mathbf{e}_j$ ,  $j = 1, \dots, n$ , with the error process  $(\mathbf{e}_j)_{j=1}^n$  specified as in (4.1) and the mean function  $\boldsymbol{\mu}(t) = \{\mu_1(t), \dots, \mu_p(t)\}$  satisfying  $\mu_i(t) = 0$  for  $i \leq p/2$  and  $\mu_i(t) = df(t)$  for  $i > p/2$ , where  $d$  represents the magnitude of deviation from the null and  $f(t)$  is a nonconstant function. We consider two types of alternatives.

- (SA) Systematic alternative:  $f(t) = t$ ,  $t \in [0, 1]$ , which deviates from the constancy in an even and systematic manner; see Figure 2 (a).
- (BA) Bumpy alternative:  $f(t) = 3 \exp\{-2^{11}(t - 1/2)^2\}$ ,  $t \in [0, 1]$ , which deviates from the constancy in an abrupt and bumpy manner; see Figure 2 (b).

With the nominal level  $\alpha = 0.05$ , the cut-off values were obtained by the simulation-assisted testing procedure described in Section 3.6, with 5,000 simulated  $T_{n,\infty}^\circ$  and  $T_{n,2}^\circ$ . The results, based on 5,000 realizations of the underlying process, are summarized in Figure 2 (c) and (d) for alternatives (SA) and (BA), respectively, where the case  $d = 0$  corresponds to the null.

In Figure 2 the power of both tests increases to one as  $d$ , the deviation from the null, increases. The  $\mathcal{L}^2$ -based test seems to be more powerful than its  $\mathcal{L}^\infty$ -based counterpart, in line with the discussion in Section 3.3. As commented by Zhou (2010),  $\mathcal{L}^2$ -based methods should be preferred over their  $\mathcal{L}^\infty$ -based counterparts when one is interested in verifying a specific null hypothesis, while the  $\mathcal{L}^\infty$ -based methods can be useful in assessing the overall pattern of the curve as it relates to the construction of simultaneous confidence tubes. That paper considered the problem of testing parametric forms of quantile curves for univariate time series, where the underlying process has a functional dependence measure  $\theta_{n,q} = O(\rho^n)$  for some  $\rho \in (0, 1)$ , while our multivariate nonstationary time series has a functional dependence measure  $\theta_{n,q} = O(n^{-\iota})$  for some  $\iota > 2$ .

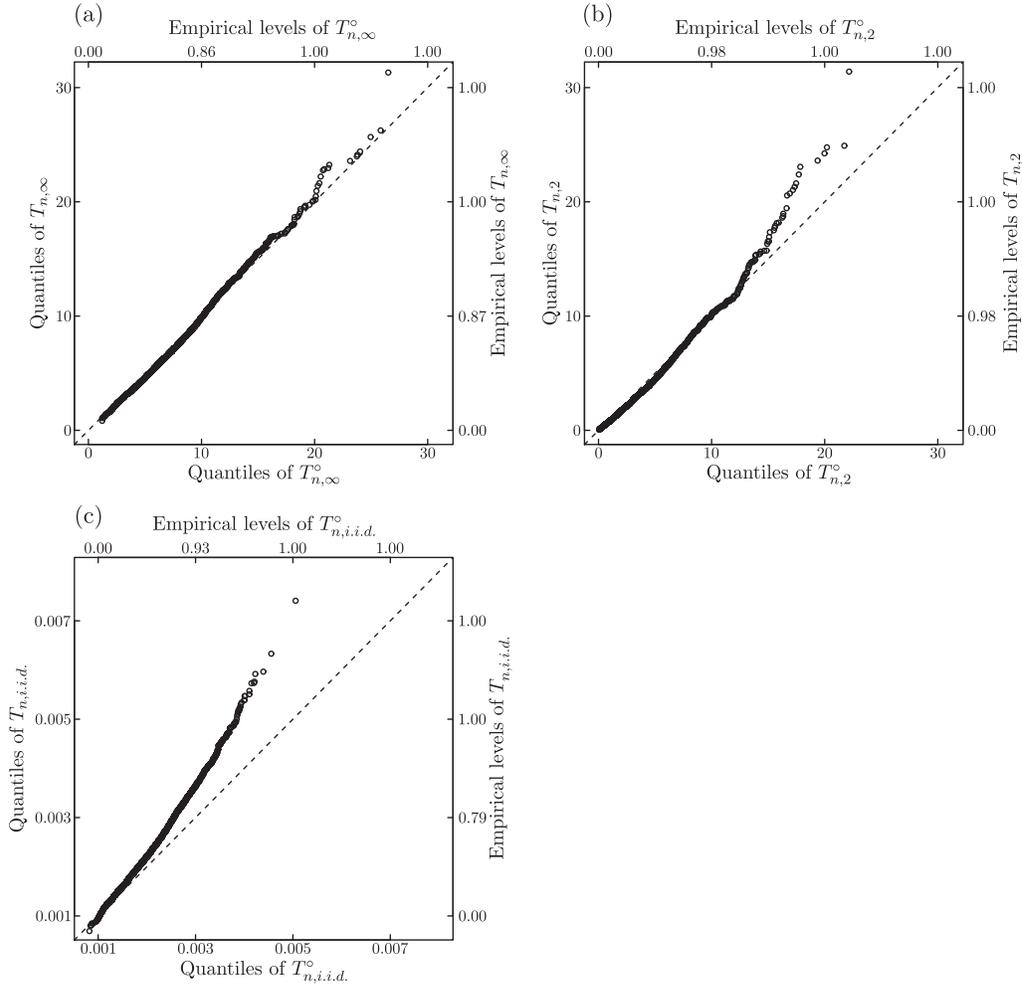


Figure 1. A comparison of (a) the proposed test based on the maximum absolute deviation, (b) the proposed test based on the integrated squared deviation, and (c) the test of Degras et al. (2012). Q-Q plots of (a)  $T_{n,\infty}$  against  $T_{n,\infty}^\circ$ ; (b)  $T_{n,2}$  against  $T_{n,2}^\circ$ ; and (c)  $T_{n,i.i.d.}$  against  $T_{n,i.i.d.}^\circ$ . For all the plots, the dashed lines have unit slope and zero intercept.

We carried out simulation studies to examine the sensitivity of the proposed tests to the choice of the bandwidth. Let

$$\mathbf{X}_j = \boldsymbol{\mu}\left(\frac{j}{n}\right) + \sigma \mathbf{e}_j, \quad j = 1, \dots, n,$$

where the error process  $(\mathbf{e}_j)_{j=1}^n$  is specified as in (4.1) and the mean function  $\boldsymbol{\mu}(t) = \{\mu_1(t), \dots, \mu_p(t)\}$  with  $\mu_i(t) = \sin(2\pi t) + i/p$ ,  $i = 1, \dots, p$ . For  $n = 1,000$  and  $p = 10$ , we considered testing the null hypothesis (3.1) for  $(\mathbf{X}_j)_{j=1}^n$  with  $\mathbf{A}$

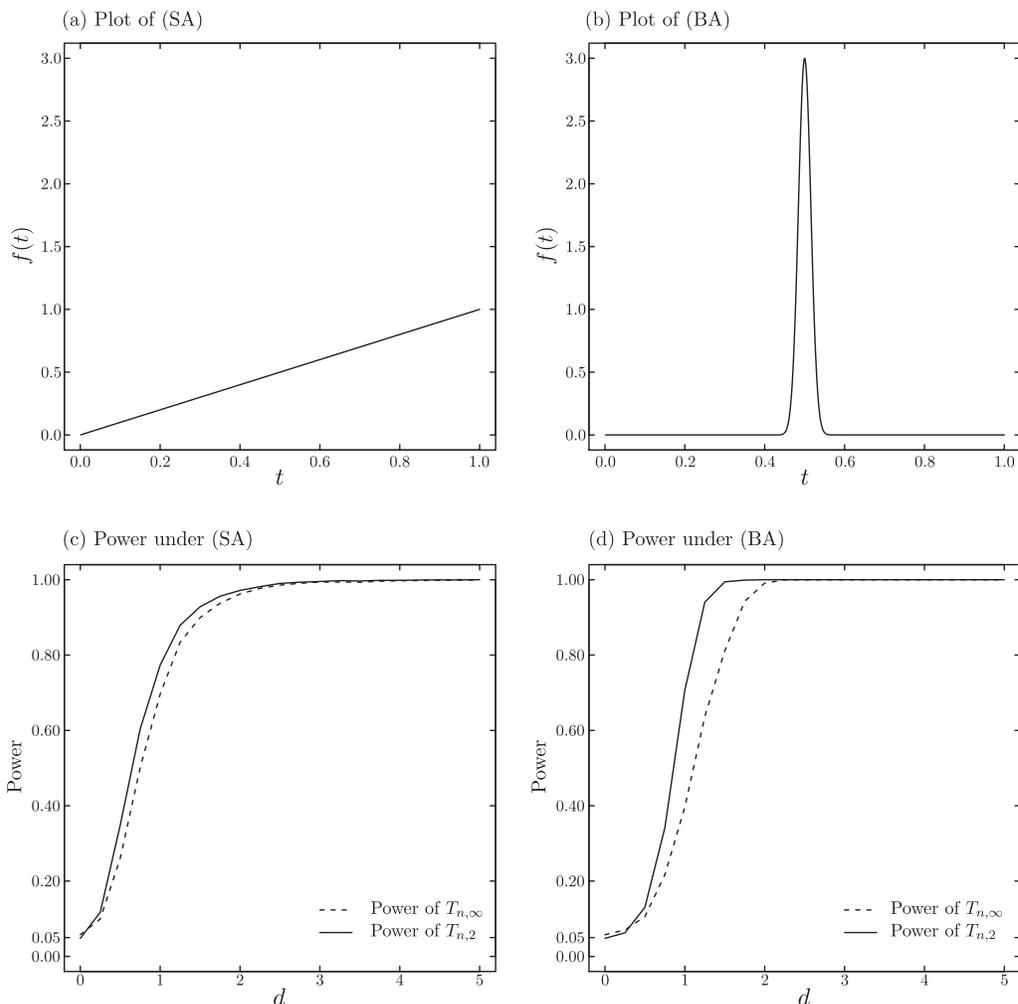


Figure 2. In (a) and (b) are plots of the alternative function  $f(t)$ ,  $t \in [0, 1]$ , under (SA) and (BA), respectively. In (c) and (d) are power curves of the  $\mathcal{L}^\infty$ -based test in Section 3.2 (dashed curves) and the  $\mathcal{L}^2$ -based test in Section 3.3 (solid curves) under alternatives (SA) and (BA), respectively.

specified as in (3.2). The cut-off values were obtained by using the simulation-assisted method in Section 3.6 with 5,000 simulated  $T_{n,\infty}^\circ$  and  $T_{n,2}^\circ$ . The results are summarized in Table 1 based on 5,000 realizations for each different combination of bandwidths  $b \in \{0.1, \dots, 0.9\}$  and noise-to-signal levels  $\sigma^2 \in \{1, 2\}$ . It can be seen that the empirical acceptance probabilities are fairly close to their nominal levels (90%, 95%, 99%) as long as the bandwidth is not chosen to be too small, in which case only a small number of data points are used to compute the local mean and covariance estimates. For the simulated processes,  $\mathbf{A}\boldsymbol{\mu}(\cdot)$  is

Table 1. Empirical acceptance probabilities (in percentage) of the simulation-assisted hypothesis testing procedure described in Section 3.6 for  $n = 1,000$  and  $p = 10$ , where the noise-to-signal level  $\sigma^2 = 1$ .

$\sigma^2$	$b$	$\mathcal{L}^\infty$ -based			$\mathcal{L}^2$ -based		
		90%	95%	99%	90%	95%	99%
1	0.1	84.0	92.4	98.7	85.4	91.9	98.5
	0.2	88.0	94.2	99.1	86.7	93.2	99.0
	0.3	87.4	93.5	98.3	87.0	92.9	98.4
	0.4	89.2	94.4	98.8	88.7	94.2	98.5
	0.5	88.5	93.9	98.3	88.5	93.4	98.2
	0.6	89.4	94.9	98.9	89.9	95.0	98.7
	0.7	89.7	94.7	98.6	88.5	94.1	98.7
	0.8	89.5	94.3	98.6	89.0	94.1	98.5
	0.9	89.4	94.7	98.5	89.5	94.6	98.8
2	0.1	84.4	92.4	98.5	86.3	92.7	98.5
	0.2	88.3	94.5	98.9	87.1	93.5	98.6
	0.3	88.0	94.2	98.4	87.2	93.5	98.4
	0.4	88.3	93.9	98.6	87.8	93.5	98.6
	0.5	88.7	94.3	98.5	88.4	93.3	98.3
	0.6	89.5	94.9	98.9	89.3	94.6	98.4
	0.7	89.3	94.2	98.6	88.1	93.9	98.5
	0.8	88.7	94.2	98.9	87.8	93.5	98.9
	0.9	90.2	94.8	98.6	89.6	94.2	98.7

a constant and thus a large bandwidth is preferred to reduce the mean squared error. This is reflected by the data-driven bandwidth selector discussed in Section 3.4 as the medians of the selected bandwidths under a full search are above 0.9 for both  $\sigma^2 = 1$  and  $\sigma^2 = 2$ .

#### 4.2. Application to temperature data

We applied the proposed tests to measurements of air temperature at 15 measurement facilities, a central facility (E-13) in northern Oklahoma along with 14 extended facilities, in the Southern Great Plains region of the United States. The data has been studied by Guinness and Stein (2013) under the additive model (1.1) for the purpose of spatial and temporal interpolation. It can be accessed from the ARM website at <http://www.archive.arm.gov>. We use the 30-minute averaged data during the period from 10/06/2005 to 10/30/2005, and provide a statistical test for the additive assumption in model (1.1) used by Guinness and Stein (2013); this with  $p = 15$  and  $n = 1,200$ . This is equivalent to testing the mean trend parallelism assumption with vertical shifts caused by the site effect; this seems plausible from the time series plot given in Figure 3 (a). In particular, if we take the differences with respect to the records from the central facility

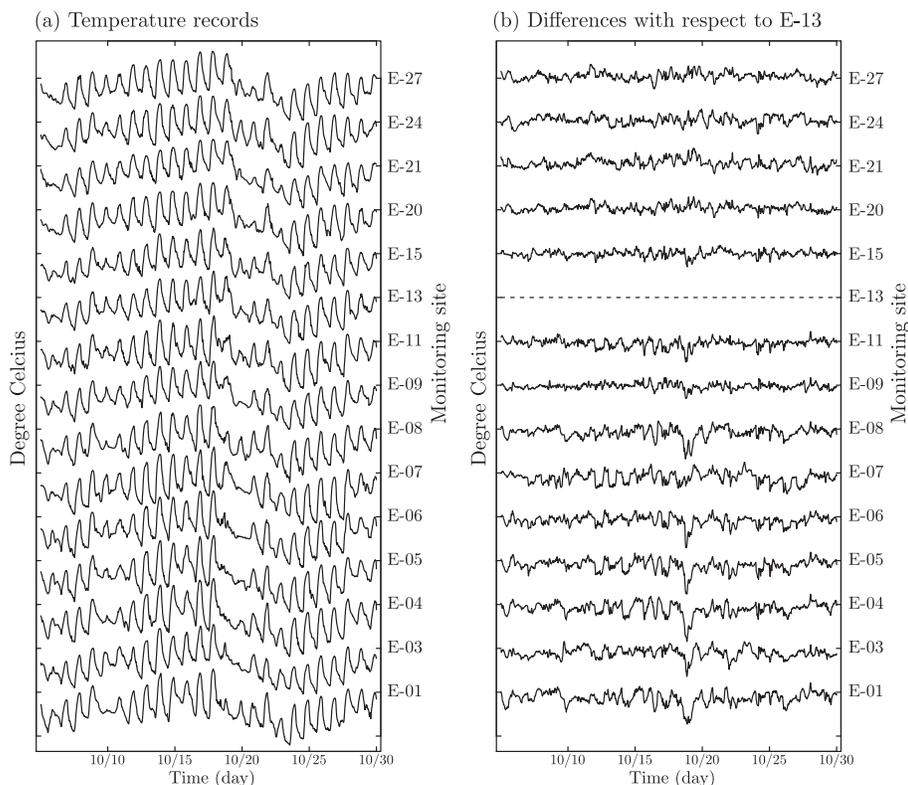


Figure 3. Time series plot of (a) air temperature records; and (b) differences with respect to the central facility E-13 at 15 monitoring sites in the Southern Great Plains region of the United States. The tick mark on the vertical axis are separated by  $20^{\circ}\text{C}$ , and records from different monitoring sites are offset from each other by  $20^{\circ}\text{C}$ .

E-13, then the mean trend parallelism assumption of the original temperature series is equivalent to the mean trend constancy assumption of the differences; see Figure 3 (b) for a time series plot of the corresponding differences.

We modeled the joint process by (2.1) and applied our tests. The selected bandwidth  $\hat{b}_n = 0.36$ , and the corresponding test statistics were  $T_{n,\infty} = 78.18$  and  $T_{n,2} = 208.93$ . With 5,000 repetitions of step (iv) in Section 3.6, the  $p$ -values are 0.000 for both  $T_{n,\infty}$  and  $T_{n,2}$ . In particular, at 1% significance level the cut-off values are 16.86 and 21.69 for  $T_{n,\infty}$  and  $T_{n,2}$ , respectively. We reject the null hypothesis at the 1% significance level, so the additive model (1.1) is not statistically sufficient for modeling this data. Although the work of Guinness and Stein (2013) has provided a decent interpolation scheme for the temperature data collected in the Southern Great Plains region of the United States, the current analysis suggests that there is still room for improvement of the models they used

to generate their results. We leave this as a future research topic for climate scientists, and a possible direction is to allow spatial-temporal interactions to produce a more sophisticated interpolation of the temperature data considered.

## 5. Discussions

The paper models regular monitoring data as a multivariate time series in which a fixed number of lengthy time series is associated with fixed measurement facilities, and proposes statistical tests that can survive for a general class of non-stationary and dependent processes so that our results can be widely applicable.

Although the paper is motivated by regular monitoring data where the number of measurement facilities  $p$  is usually fixed and much smaller than the length of each time series  $n$ , the proposed tests possess the potential of being generalized to situations where  $p = p_n \rightarrow \infty$  as  $n \rightarrow \infty$ . For this, we consider a variant of the functional dependence measure (2.2) and use

$$\theta_{k,q}^\# = \max_{1 \leq i \leq p} \left\{ \sup_{t \in [0,1]} \|G_i(t; \mathcal{F}_k) - G_i(t; \mathcal{F}_k^*)\|_q \right\}, \quad (5.1)$$

where  $G_i(t; \mathcal{F}_k)$  is the  $i$ -th component of the vector  $\mathbf{G}(t; \mathcal{F}_k)$ ; the process is said to be stochastically Lipschitz continuous or  $\mathbf{G} \in \mathcal{C}_{SL}$  if there exists a constant  $c_0^\# < \infty$  such that  $\max_{1 \leq i \leq p} \|G_i(t_1; \mathcal{F}_k) - G_i(t_2; \mathcal{F}_k)\| \leq c_0^\# |t_1 - t_2|$  holds uniformly for all  $t_1, t_2 \in [0, 1]$ . The dependence measure (5.1) provides a useful counterpart of (2.2) for studying high-dimensional time series, and has been used by Chen, Xu, and Wu (2013) for estimating large covariance and precision matrices. In order to make statistical inference for high-dimensional objects, additional regularity conditions are usually needed among which sparsity is one of the most popular; see for example Bickel and Levina (2008b), Cai and Zhou (2012) and references therein. Corollary 1 states that the maximum absolute deviation and the integrated squared deviation considered in Sections 3.2 and 3.3 respectively will continue to have the desired asymptotic distributions provided that the matrix  $\mathbf{A}$  is sparse in the sense of the following condition.

(H1) There exists a constant  $c_F$  such that the Frobenius norm  $\|\mathbf{A}\|_F \leq c_F < \infty$ .

**Corollary 1.** *Suppose  $p = p_n \rightarrow \infty$  as  $n \rightarrow \infty$ . If  $\mathbf{G} \in \mathcal{C}_{SL}$  and (H1) holds, then the results of Theorems 2 and 3 continue to hold under corresponding conditions with  $\theta_{k,q}$  replaced by  $\theta_{k,q}^\#$ .*

Although the sparsity condition (H1) serves as a reasonable requirement for making statistical inference about high-dimensional objects, we conjecture that it is not the weakest possible and can be relaxed to allow a growing Frobenius norm if we have further knowledge about the underlying data generating mechanism (for example measurement errors at different monitoring sites and different

time points are i.i.d.) and the growth rate is expected to depend on the interplay between the inherited dependence and nonstationarity across both time and space. We leave this as a possible future research topic.

### Supplementary Materials

The Supplementary Material contains technical proofs of Theorems 1–3, Propositions 1 and 2, and Corollary 1.

### Acknowledgement

I am grateful to the co-Editor, an associate editor, and two anonymous referees for their helpful comments and suggestions. I thank Joseph Guinness for providing the temperature data analyzed in Section 4.2, and Dale Zimmerman for mentioning a few references on additive space-time models. The research was supported in part by NSF Grant DMS-1461796.

### References

- Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61**, 821-856.
- Bickel, P. J. and Levina, E. (2008a). Regularized estimation of large covariance matrices. *Ann. Statist.* **36**, 199-227.
- Bickel, P. J. and Levina, E. (2008b). Covariance regularization by thresholding. *Ann. Statist.* **36**, 2577-2604.
- Bliznyuk, N., Carroll, R. J., Genton, M. G. and Wang, Y. (2012). Variogram estimation in the presence of trend. *Stat. Interface* **5**, 159-168.
- Cai, T. T. and Zhou, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Statist.* **40**, 2389-2420.
- Chen, X., Xu, M. and Wu, W. B. (2013). Covariance and precision matrix estimation for high-dimensional time series. *Ann. Statist.* **41**, 2994-3021.
- Cheng, B. and Tong, H. (1998). K-stationarity and wavelets. *J. Statist. Plann. Inference* **68**, 129-144.
- Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.* **31**, 377-403.
- Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.* **62**, 139-168.
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1-37.
- Degras, D., Xu, Z., Zhang, T. and Wu, W. B. (2012). Testing for parallelism among trends in multiple time series. *IEEE Trans. Signal Process.* **60**, 1087-1097.
- Delgado, M. A. (1993). Testing the equality of nonparametric regression curves. *Statist. Probab. Lett.* **17**, 199-204.
- Dette, H. and Neumeier, N. (2001). Nonparametric analysis of covariance. *Ann. Statist.* **29**, 1361-1400.

- Draghicescu, D., Guillas, S. and Wu, W. B. (2009). Quantile curve estimation and visualization for non-stationary time series. *J. Comput. Graph. Statist.* **18**, 1-20.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications*. Chapman & Hall, London, U.K.
- Fan, J. and Lin, S.-K. (1998). Test of significance when data are curves. *J. Amer. Stat. Assoc.* **93**, 1007-1021.
- Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Statist.* **27**, 715-731.
- Gao, J. and Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *J. Amer. Statist. Assoc.* **484**, 1584-1594.
- Giurcanu, M. and Spokoiny V. G. (2004). Confidence estimation of the covariance function of stationary and locally stationary processes. *Statist. Decisions* **22**, 283-300.
- Guinness, J. and Stein, M. L. (2013). Interpolation of nonstationary high frequency spatial-temporal temperature data. *Ann. Appl. Statist.* **7**, 1684-1708.
- Hall, P. and Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Amer. Statist. Assoc.* **85**, 1039-1049.
- Härdle, W., Hall, P. and Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *J. Amer. Statist. Assoc.* **83**, 86-95.
- Härdle, W. and Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* **13**, 1465-1481.
- Härdle, W. and Marron, J. S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.* **18**, 63-89.
- Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: assessing Ireland's wind power resource. *Appl. Statist.* **38**, 1-50.
- King, E., Hart, J. D. and Wehrly, T. E. (1991). Testing the equality of two regression curves using linear smoothers. *Statist. Probab. Lett.* **12**, 239-247.
- Kulasekera, K. B. (1995). Comparison of regression curves using quasi-residuals. *J. Amer. Stat. Assoc.* **90**, 1085-1093.
- Li, B., Genton, M. G. and Sherman, M. (2007). A nonparametric assessment of properties of space-time covariance functions. *J. Amer. Statist. Assoc.* **102**, 736-744.
- Li, F. (2006). Testing for the equality of two nonparametric regression curves with long memory errors. *Comm. Statist. Simulation Comput.* **35**, 621-643.
- Lumley, T. and Heagerty, P. (1999). Weighted empirical adaptive variance estimators for correlated data regression. *J. Roy. Statist. Soc. Ser. B* **61**, 459-477.
- Mallat, S., Papanicolaou, G. and Zhang, Z. (1998). Adaptive covariance estimation of locally stationary processes. *Ann. Statist.* **26**, 1-47.
- Munk, A. and Dette, H. (1998). Nonparametric comparison of several regression functions: Exact and asymptotic theory. *Ann. Statist.* **26**, 2339-2368.
- Nason, G. P., von Sachs, R. and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. Roy. Statist. Soc. Ser. B* **62**, 271-292.
- Neumeier, N. and Dette, H. (2003). Nonparametric comparison of regression curves: an Empirical process approach. *Ann. Statist.* **31**, 880-920.
- Newey, W. K. and West, K. D. (1987). A simple, positive-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703-708.

- Ombao, H., von Sachs, R. and Guo, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Amer. Statist. Assoc.* **100**, 519-531.
- Opsomer, J., Wang, Y. D. and Yang, Y. H. (2001). Nonparametric regression with correlated errors. *Statist. Sci.* **16**, 134-153.
- Park, C. and Kang, K.-H. (2008). SiZer analysis for the comparison of regression curves. *Comput. Statist. Data Anal.* **52**, 3954-3970.
- Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *J. Roy. Statist. Soc. Ser. B* **27**, 204-237.
- Ruppert, D., Sheather, S. J. and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* **90**, 1257-1270.
- Stein, M. L. (2005). Statistical methods for regular monitoring data. *J. Roy. Statist. Soc. Ser. B* **67**, 667-687.
- Stein, M. L. (2009). Spatial interpolation of high-frequency monitoring data. *Ann. Appl. Statist.* **3**, 272-291.
- Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press, U.K.
- Wang, L. (2008). Nonparametric test for checking lack of fit of the quantile regression model under random censoring. *Canad. J. Statist.* **36**, 321-336.
- Wang, Y. (1998). Smoothing spline models with correlated random errors. *J. Amer. Statist. Assoc.* **93**, 341-348.
- Wiener, N. (1958). *Nonlinear Problems In Random Theory*. The MIT Press.
- Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci.* **102**, 14150-14154.
- Wu, W. B. and Pourahmadi, M. (2009). Banding sample covariance matrices of stationary processes. *Statist. Sinica* **19**, 1755-1768.
- Wu, W. B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica* **21**, 1397-1413.
- Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. Roy. Statist. Soc. Ser. B* **60**, 797-811.
- Young, S. G. and Bowman, A. W. (1995). Non-parametric analysis of covariance. *Biometrics* **51**, 920-931.
- Zhang, W., Lee, S.-Y. and Song, X. (2002). Local polynomial fitting in semivarying coefficient model. *J. Multivariate Anal.* **82**, 166-188.
- Zhang, T. and Wu, W. B. (2011). Testing parametric assumptions of trends of a nonstationary time series. *Biometrika* **98**, 599-614.
- Zhang, T. and Wu, W. B. (2012). Inference of time-varying regression models. *Ann. Statist.* **40**, 1376-1402.
- Zhou, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *Ann. Statist.* **38**, 2187-2217.
- Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *J. Roy. Statist. Soc. Ser. B* **72**, 513-531.

Department of Mathematics and Statistics, Boston University, Boston, MA 02215, U.S.A.

E-mail: tingz@bu.edu

(Received May 2014; accepted December 2014)