Nonparametric Estimation of Local Smoothness for Spatially Inhomogeneous Function

Dongik Jang and Hee-Seok Oh†

Department of Statistics

Seoul National University

Seoul 151-747, Korea

Philippe Naveau

Laboratoire des Sciences du Climat et l'Environnement

91191 Gif-sur-Yvette, France

†Correspondence to heeseok.oh@gmail.com

Draft: version of May 14, 2015

Abstract

We investigate the local smoothness of a spatially inhomogeneous function from noisy data under the framework of smoothing splines. Most previous studies related to this problem deal with the estimation induced by a single smoothing parameter or partially local smoothing parameters, which may not be efficient to characterize various degrees of smoothness of the underlying function when it is spatially varying. In this paper we propose a new nonparametric method to estimate the local smoothness of the function based on a moving local risk minimization coupled with spatially adaptive smoothing splines. The proposed method provides full information of the local smoothness of the function at every design point on the entire data domain, so that it is able to understand the degrees of spatial inhomogeneity of the function. A successful estimate of the local smoothness is useful for identifying abrupt changes of smoothness of the data, performing an efficient functional clustering, and improving the uniformity of coverage of the confidence intervals of smoothing splines. Moreover we consider an extension of the local smoothness of inhomogeneous two-dimensional functions or spatial fields. Results from a simulation study and real data analysis demonstrate that the proposed method is effective for estimating the local smoothness of inhomogeneous functions.

Keywords: Clustering; Confidence interval; Local risk; Local smoothness; Smoothing splines.

1 Introduction

1.1 Background

Suppose that we observe a set of noisy data $\{(x_i, y_i) : i = 1, \ldots, n\}$ from the model

$$y_i = f(x_i) + \epsilon_i,$$

where the errors ϵ_i are independently and identically distributed (i.i.d.) random errors and the distribution of the errors is typically assumed to be Gaussian, and the underlying mean function $f \in L^2(\mathbb{R})$ defined on the interval $\Omega \in \mathbb{R}$ may be spatially inhomogeneous, that is, the function possesses various degrees of smoothness over the range of design points. Here we assume, without loss of generality, $\Omega = [0, 1]$.

The primary goal of this study is to identify the local smoothness of the unknown function $f(x)$ at every $x(\in \Omega)$ given the observed data, instead of estimating the function itself. This helps understanding the local inhomogeneity of the underlying function. To achieve this goal, we focus on spatially smoothing splines as studied by Pintore, Speckman and Holmes (2006) and Wahba (1995), which are expressed as the following framework

$$\frac{1}{n}\sum_{i=1}^{n}\big\{y_i - f(x_i)\big\}^2 + \int_0^1 \lambda(x)\big\{f''(x)\big\}^2 dx, \tag{1}$$

where $\lambda(x)$ is the (local) smoothness function as a function of x, which represents varying degrees of the smoothness of the function. A single smoothing parameter $\lambda = \lambda(x)$ used in the conventional smoothing splines is designed for a function with a constant degree of smoothness (Green and Silverman, 1994; Wahba, 1990).

In the literature, there are several attempts to estimate $\lambda(x)$. Lee (2004) proposed localized smoothing splines with a discretization $\lambda(x_i) = \lambda_i$ at each design point x_i, and suggested an estimation method of λ_i by minimizing a local risk. Due to focusing only each point x_i without neighbourhood information, the localized smoothing spline fit typically enjoys a good local property, but it fails to capture the global pattern of the data. Thus, it is not capable of handling unequally spaced data or scattered data. Pintore *et al.* (2006) developed a spatially adaptive smoothing spline method in which the smoothing parameter function $\lambda(x)$ is modelled

as a piecewise-constant function with jumps at the knots τ_k, *i.e.*, $\lambda(x) = \lambda_k$, $x \in [\tau_k, \tau_{k+1}]$ for $k = 1, \ldots, K$. However, this approach suffers from discontinuities caused by the piecewise-constant modeling even though the true function is continuous. More recently, Jang and Oh (2011) suggested a method to estimate $\lambda(x)$ based on a combination of a blockwise risk minimization with a parameterization of $\lambda(x)$ modelled as a smooth step function. However, this blockwise approach is not able to reflect spatial inhomogeneity of the function fully at every point, and hence, the resulting $\hat{\lambda}(x)$ tends to be biased. The choice of blocks is one of the key components of this method, but it remains a practical roadblock.

In this paper, along with the line that takes both advantages from Lee (2004) and Jang and Oh (2011), we consider a new approach, termed *moving local risk minimization* that can adapt the multiscale characteristics of spatially inhomogeneous functions. That is, it is capable of identifying local features as well as capturing global trend of the smoothness simultaneously.

1.2 An Example

For motivation of the proposed method, we consider the following particular example. Figure 1 shows four noisy data generated from underlying functions with different degrees of smoothness. The generic forms of test functions in Figures 1(a)–(d) are listed as $f_a(x) = \sin(2\pi x)$, $f_b(x) = \sin((2\pi x)^2)$, $f_c(x) = \sin((2\pi x)^3)$ and $f_d(x) = \sin((2\pi(x + 0.0026))^3)I(x < .5) + \sin((2\pi(0.9974 - x))^3)I(x \geq .5)$, respectively. The design points x_i's are generated from uniform distribution on $[0, 1]$, the signal-to-noise ratio (SNR) is set to be 5, and the sample size is $n = 800$. As shown, the function in Figure 1(a) holds a single degree of smoothness, whereas the smoothnesses of the all other functions vary over the domain of x. The proposed method to be discussed in Section 2 is applied to the four types of noisy data. The resulting estimates $\hat{\lambda}(x)$ (log-scaled) are shown in Figure 1(e). As one can see, the $\hat{\lambda}(x)$ for the noisy data in Figure 1(a) is almost flat, which represents the constant smoothness of the function, and both $\hat{\lambda}(x)$s for the data in Figures 1(b) and (c) decrease linearly over the domain of x. The smoothness of Figure 1(c) rapidly decreases, compared to that of Figure 1(b). On the other hand, the smoothness estimation $\hat{\lambda}(x)$ of the data in Figure 1(d) decreases linearly in the first half and increases linearly in the second half. It seems

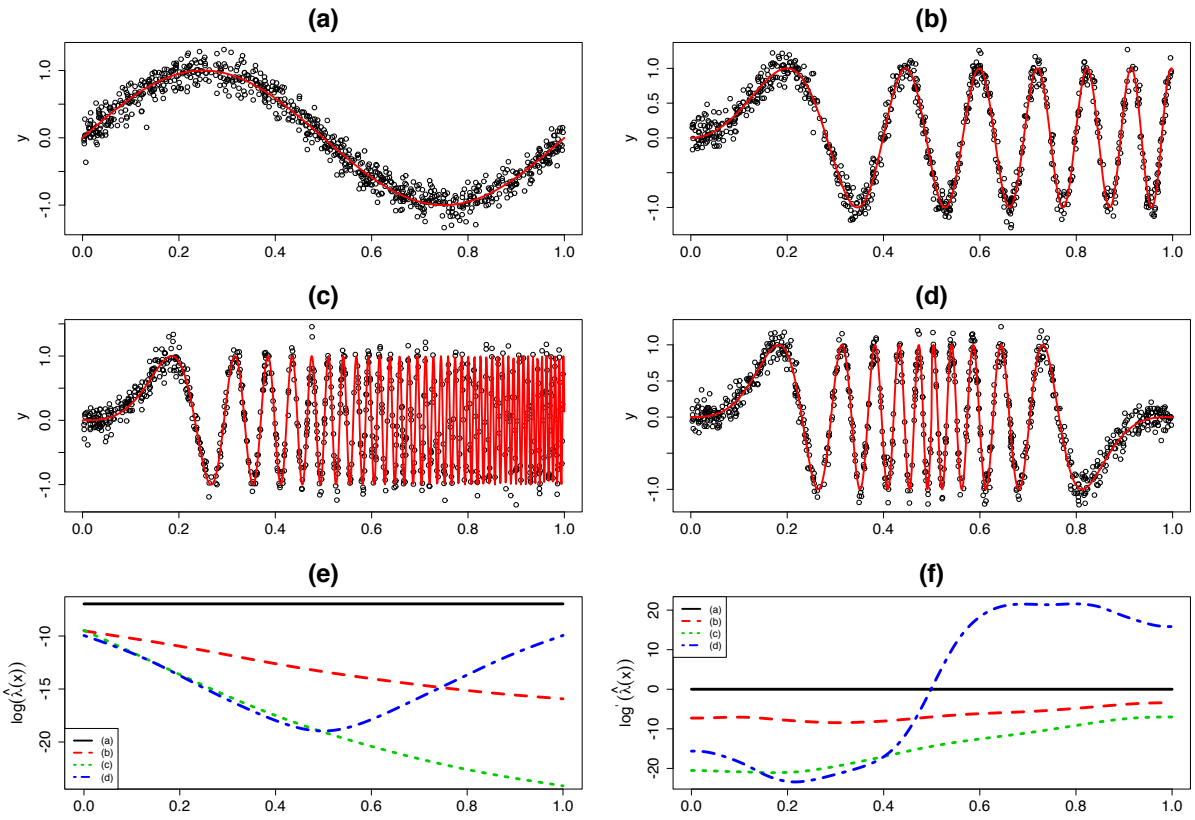

Figure 1: Estimates of smoothness functions $\lambda(x)$ and their derivatives.

that all $\hat{\lambda}(x)$'s represent the local smoothness for the underlying functions well. Furthermore, Figure 1(f) shows the derivatives of the estimated $\hat{\lambda}(x)$'s with respect to x, *i.e.*, the rate of change of the smoothness over x, which may be useful for identifying the inhomogeneity of a function and clustering functional data.

1.3 Main Contribution

The main contribution of this paper can be summarized as follows.

(a) We develop an adaptive method of estimating the smoothness function $\lambda(x)$ on the entire domain of x under the spatially smoothing splines framework in (1). For this purpose, we introduce a moving local risk $R_\lambda(x)$ and propose an iterative procedure to estimate $R_\lambda(x)$ efficiently. Therefore, by minimizing the estimated moving local risk $\hat{R}_\lambda(x)$, it is capable of identifying the local smoothness of a spatially inhomogeneous function.

(b) We extend the proposed method to two-dimensional data or spatial data. In other words, we propose a nonparametric method to estimate the smoothness field of two-dimensional function over the spatial domain (u, v) under the following spatially bivariate smoothing splines framework

$$\frac{1}{n}\sum_{i=1}^{n}\{z_i - f(u_i, v_i)\}^2 + \int\int_{\mathbb{R}^2} \lambda(u, v)\Big(\frac{\partial^2 f}{\partial u^2} + 2\frac{\partial^2 f}{\partial u \partial v} + \frac{\partial^2 f}{\partial v^2}\Big)dudv.$$

It can be expected that an accurate information of $\lambda(u, v)$ may provide a better understanding of the spatial inhomogeneity.

(c) We further consider some possible applications based on the information of the local smoothness and its derivative: (i) clustering functional data that are spatially inhomogeneous, (ii) change point detection of the local smoothness, and (iii) improvement of confidence interval estimation of smoothing splines.

The remainder of this paper is organized as follows. Section 2 presents our method of estimating the smoothing function $\lambda(x)$ based on a moving local risk and its related issues. In addition, an extension to inhomogeneous spatial data is discussed. Section 3 discusses some statistical applications including functional data clustering, change point detection of the smoothness, and confidence interval of smoothing splines. For each application, some numerical studies are presented for evaluation of empirical performance. Finally, concluding remarks are given in Section 4.

2 Estimation of Local Smoothness

2.1 Moving Local Risk and Estimation of Local Smoothness

Let $\mathcal{I}_x = \{i : |x_i - x| \le r\}$ for a local point x be a set that contains data points contributing to the smoothing spline estimate $\hat{f}_\lambda(x)$ with a smoothing parameter λ. We define a local risk at the local point x on the set $\mathcal{I}_x$ as

$$R_\lambda(x) := \sum_{i \in \mathcal{I}_x} E\big\{f(x_i) - \hat{f}_\lambda(x_i)\big\}^2.$$

For the bias-variance decomposition of $E\{f(x_i) - \hat{f}_\lambda(x_i)\}^2$, we use matrix-vector notation as $\boldsymbol{y} = (y_1, \ldots, y_n)^T$, $\boldsymbol{f} = (f(x_1), \ldots, f(x_n))^T$, and $\hat{\boldsymbol{f}}_\lambda = (\hat{f}_\lambda(x_1), \ldots, \hat{f}_\lambda(x_n))^T$. Furthermore, let S_λ be the smoothing matrix that maps $\boldsymbol{y}$ onto $\hat{\boldsymbol{f}}_\lambda : \hat{\boldsymbol{f}}_\lambda = S_\lambda \boldsymbol{y}$. We define the ith element of the vector $S_\lambda \boldsymbol{f}$ as $S_\lambda \boldsymbol{f}(x_i)$ and the ith diagonal element of the square matrix $S_\lambda S_\lambda^T$ as $s_\lambda(x_i)$. Then, by straightforward calculation, the local risk at the local point x is expressed as

$$R_\lambda(x) = \sum_{i \in \mathcal{I}_x} \left[\left\{ S_\lambda \boldsymbol{f}(x_i) - f(x_i) \right\}^2 + \sigma^2 s_\lambda(x_i) \right], \tag{2}$$

where σ^2 denotes the noise level. However, in practice $\boldsymbol{f}$ and σ are unknown, so we consider instead a fixed-point analogy to the local risk of (2) as

$$\hat{R}_\lambda(x) = \sum_{i \in \mathcal{I}_x} \left[\left\{ S_\lambda \hat{\boldsymbol{f}}_{\lambda_p}(x_i) - \hat{f}_{\lambda_p}(x_i) \right\}^2 + \hat{\sigma}^2_{\lambda_p} s_\lambda(x_i) \right],$$

where $\hat{f}_{\lambda_p}$ denotes a smoothing spline fit with a smoothing parameter λ_p, and $\hat{\sigma}^2_{\lambda_p}$ is an estimate of σ^2. Then, by moving the local point x across the data domain Ω, we first obtain the estimated local risk $\hat{R}_\lambda(x)$, and then find an optimal smoothing function $\lambda(x)$ to minimize the moving local risk over Ω.

For implementation of the above moving local risk derivation and the corresponding accurate estimate of $\lambda(x)$, we further consider the following iterative procedure:

1. Given a pilot smoothing parameter $\lambda^{(0)}$, obtain an initial estimated moving local risk at the point x

$$\hat{R}^{(0)}_{\lambda(x)}(x) = \sum_{i \in \mathcal{I}_x} \left[\left\{ S_{\lambda(x)} \hat{\boldsymbol{f}}_{\lambda^{(0)}}(x_i) - \hat{f}_{\lambda^{(0)}}(x_i) \right\}^2 + \hat{\sigma}^2_{(0)} s_{\lambda(x)}(x_i) \right], \tag{3}$$

   which is considered as a function of $\lambda(x)$.

2. For the step $\ell = 1, 2, \ldots$, estimate, until convergence, the moving local risk at the point x

$$\hat{R}^{(\ell)}_{\lambda(x)}(x) = \sum_{i \in \mathcal{I}_x} \left[\left\{ S_{\lambda(x)} \hat{\boldsymbol{f}}_{\lambda_i^{(\ell)}}(x_i) - \hat{f}_{\lambda_i^{(\ell)}}(x_i) \right\}^2 + \hat{\sigma}^2_{(\ell)} s_{\lambda(x)}(x_i) \right], \tag{4}$$

   where $\lambda_i^{(\ell)} \triangleq \lambda^{(\ell)}(x_i)$ denotes the local smoothness at x_i obtained from the previous level $\ell - 1$; hence, define the estimated local smoothness at the point x as

$$\lambda^{(\ell+1)}(x) \triangleq \arg\min_{\lambda(x)} \hat{R}^{(\ell)}_{\lambda(x)}(x).$$

3. Take the converged estimate as the final local smoothness $\lambda(x)$ at the point x.

4. Repeat Steps 1-3 over $x \in \Omega$ (i.e., $x = x_1, \ldots, x_n$), and then finally obtain the smoothing function $\lambda(x)$ in the entire domain of x.

We have some remarks regarding the above approach.

(a) *Initial estimate $\hat{R}_\lambda^{(0)}(x)$ of (3)*: The pilot smoothing parameter $\lambda^{(0)}$ is chosen by GCV, and the corresponding estimate of σ^2 is given by

$$\hat{\sigma}^2_{(0)} = \frac{\sum_{i=1}^n \left\{ y_i - \hat{f}_{\lambda^{(0)}}(x_i) \right\}^2}{\mathrm{tr}\big(I - S_{\lambda^{(0)}}\big)}.$$

Due to the global property of $\lambda^{(0)}$, the estimate $\hat{R}_\lambda^{(0)}(x)$ at the local point x might suffer from bias.

(b) *Update of (4)*: Given a set of $\{\lambda_i^{(\ell)}\}_{i=1}^n$ from the previous $(\ell-1)$th step, we obtain the spatially adaptive smoothing spline fits $\hat{f}_{\lambda_i^{(\ell)}}(x_i)$ $(i = 1, \ldots, n)$ and the corresponding estimate of σ^2

$$\hat{\sigma}^2_{(\ell)} = \frac{\sum_{i=1}^n \left\{ y_i - \hat{f}_{\lambda_i^{(\ell)}}(x_i) \right\}^2}{n - \sum_{i=1}^n S_{\lambda_i^{(\ell)}}(i,i)},$$

where $S_{\lambda_i^{(\ell)}}(i,i)$ denotes the ith diagonal element of $S_{\lambda_i^{(\ell)}}$. We then have the estimate of the moving local risk, $\hat{R}^{(\ell)}_{\lambda(x)}(x)$ at the point x.

(c) *Bias correction*: In general, we decompose the following term as

$$\begin{aligned}
\|S_\lambda \hat{\boldsymbol{f}}_{\lambda'} - \hat{\boldsymbol{f}}_{\lambda'}\|^2 &= \|(S_\lambda - I)\hat{\boldsymbol{f}}_{\lambda'}\|^2 \\
&= \|(S_\lambda - I)(\hat{\boldsymbol{f}}_{\lambda'} - \boldsymbol{f}) + (S_\lambda - I)\boldsymbol{f}\|^2 \\
&= \|(S_\lambda - I)(\hat{\boldsymbol{f}}_{\lambda'} - \boldsymbol{f})\|^2 + 2(\hat{\boldsymbol{f}}_{\lambda'} - \boldsymbol{f})^T (S_\lambda - I)^T (S_\lambda - I)\boldsymbol{f} + \|(S_\lambda - I)\boldsymbol{f}\|^2,
\end{aligned}$$

where $\lambda \neq \lambda'$. Then, under $\hat{\sigma}^2_{\lambda'} \approx \sigma^2$, it follows that the estimated risk is expressed as

$$\begin{aligned}
\hat{R}_\lambda &= \|S_\lambda \hat{\boldsymbol{f}}_{\lambda'} - \hat{\boldsymbol{f}}_{\lambda'}\|^2 + \hat{\sigma}^2_{\lambda'} tr(S_\lambda S_\lambda^T) \\
&\approx \|S_\lambda \hat{\boldsymbol{f}}_{\lambda'} - \hat{\boldsymbol{f}}_{\lambda'}\|^2 + \sigma^2 tr(S_\lambda S_\lambda^T) \\
&= \|(S_\lambda - I)(\hat{\boldsymbol{f}}_{\lambda'} - \boldsymbol{f})\|^2 + 2(\hat{\boldsymbol{f}}_{\lambda'} - \boldsymbol{f})^T (S_\lambda - I)^T (S_\lambda - I)\boldsymbol{f} + \|(S_\lambda - I)\boldsymbol{f}\|^2 + \sigma^2 tr(S_\lambda S_\lambda^T) \\
&= \|(S_\lambda - I)(\hat{\boldsymbol{f}}_{\lambda'} - \boldsymbol{f})\|^2 + 2(\hat{\boldsymbol{f}}_{\lambda'} - \boldsymbol{f})^T (S_\lambda - I)^T (S_\lambda - I)\boldsymbol{f} + R_\lambda \\
&:= B_\lambda + R_\lambda.
\end{aligned}$$

The bias term B_λ tends to be negligible as long as the smoothing spline fit $\hat{\boldsymbol{f}}_{\lambda'}$ is close to $\boldsymbol{f}$; hence $R_\lambda \approx \hat{R}_\lambda$. It seems that the correction of the bias term B_λ is improved by the update procedure in (4).

(d) *Choice of window size* r: When the size of r is very small such that the window contains only a single point, the proposed procedure is exactly identical to the local risk minimization of Lee (2004), and when the size of r is large enough to cover the entire data points, it is equivalent to GCV. We note that the optimal choice of r is not the main focus of this paper. Instead we are interested in investigation of multiscale structure of the local smoothness according to the size r, which will be discussed in the next section. For some possibility of the optimal choice of r, we will leave for a future research.

As an alternative related to the above procedure, we consider an unbiased estimator of the local risk defined in (2). Globally the risk can be expressed as

$$\begin{aligned} R_\lambda &= \|\boldsymbol{f} - S_\lambda \boldsymbol{f}\|^2 + \sigma^2 tr(S_\lambda S_\lambda^T) \\ &= E(RSS_\lambda) + \sigma^2\{tr(2S_\lambda) - n\}, \end{aligned}$$

where $RSS_\lambda = \|\boldsymbol{y} - S_\lambda \boldsymbol{y}\|^2$ denotes the residual sum of squares. Thus, a typical plug-in-type estimator of R_λ is

$$\tilde{R}_\lambda = RSS_\lambda + \hat{\sigma}^2_{\lambda_p}\{tr(2S_\lambda) - n\}.$$

We then consider a localized version of $\tilde{R}_\lambda$ at the local point x on the set $\mathcal{I}_x$, which is an unbiased estimator of the local risk, defined as

$$\tilde{R}_\lambda(x) = \|\big(I^x - S^x_\lambda\big)\boldsymbol{y}\|^2 + \hat{\sigma}^2 tr(2S^x_\lambda - I^x), \tag{5}$$

where $S^x_\lambda = I^x S_\lambda$ and $I^x = \mathrm{diag}(\xi_i)$ with $\xi_i = 1$ if $i \in \mathcal{I}_x$ and $\xi_i = 0$ otherwise for $i = 1, 2, \ldots, n$. Furthermore, since the generalized cross-validation (GCV) of Craven and Wahba (1979) can be expressed as

$$\begin{aligned} \mathrm{GCV}_\lambda &= \frac{RSS_\lambda/n}{(1 - tr(S_\lambda)/n)^2} \\ &\approx \frac{RSS_\lambda}{n} + \frac{\hat{\sigma}^2}{n} tr(2S_\lambda) \\ &= \frac{\tilde{R}_\lambda}{n} - \hat{\sigma}^2, \end{aligned}$$

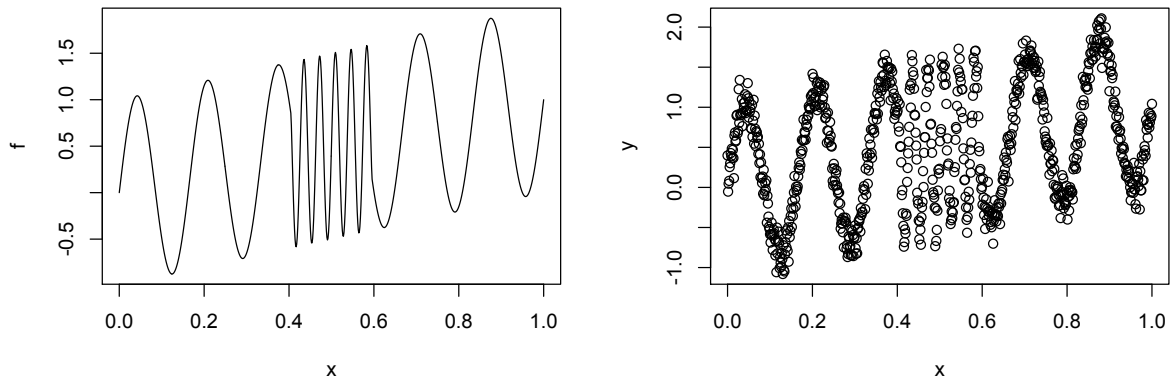

Figure 2: Spatially inhomogeneous function and a set of noisy observations with SNR=5.

we also consider a localized version of GCV, named *local GCV* as

$$\mathrm{LGCV}_\lambda(x) = \frac{\tilde{R}_\lambda(x)}{\kappa} - \hat{\sigma}^2,$$

where κ denotes the number of data points in the set I_x.

Through a simple numerical example, we now compare the proposed iterative procedure based on the moving local risk minimization with the unbiased local risk of (5). To that end, we generate a noisy signal from test function $f(x) = \sin(12\pi x)I(|x - 0.5| \geq 0.095) - \sin(54\pi x)I(|x - 0.5| < 0.095) + x$ and SNR=5, which are displayed in Figure 2. We then estimate the smooth function $\lambda(x)$ by minimizing the true local risk of (2), the unbiased local risk of (5), and the proposed moving local risk, respectively. Figure 3 shows the estimates of the local smoothness by the true local risk, the unbiased local risk, and the proposed local risk (from left to right), respectively. The first row of the figure show the estimation results of the local smoothness with a small r (say, $r = 0.05$), and histograms of the results in the second row. Note that the vertical line of the plots in the bottom separates two different degrees of the smoothness of the true function in Figure 2. As one can see, the smooth function estimated by the proposed approach is almost identical to that by minimizing the true risk, while it seems that the smooth function obtained from the unbiased local risk suffers from large variability which induces unstable results.

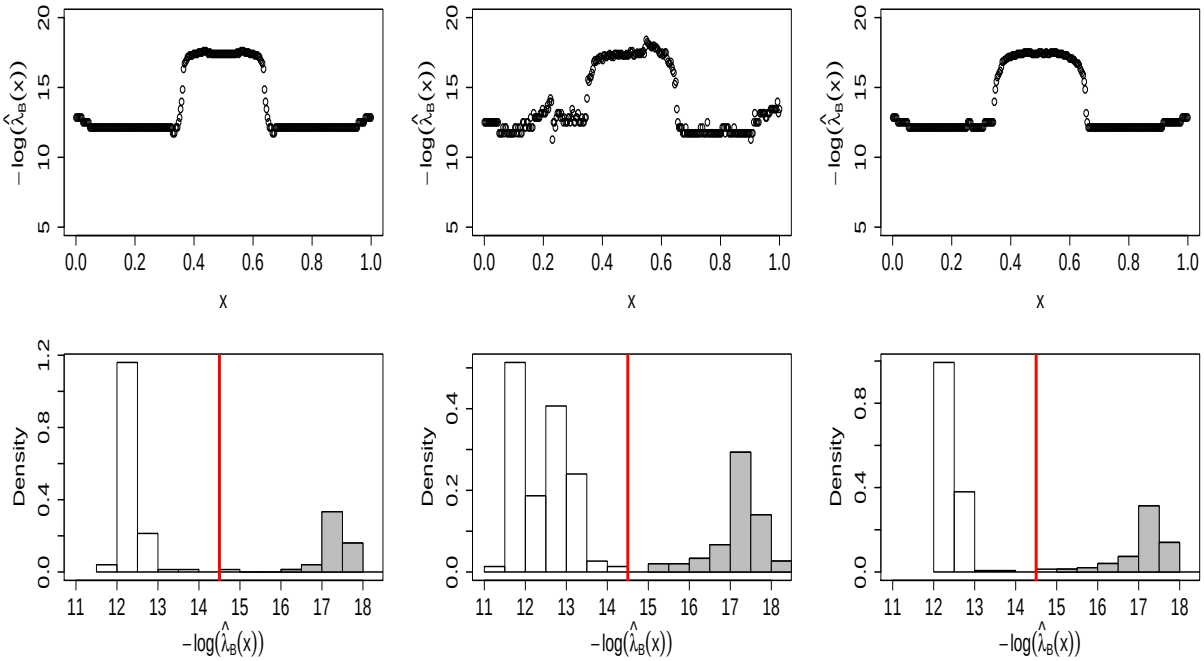

Figure 3: From left to right, the results by minimizing the true local risk of (2), the unbiased local risk of (5), and the proposed moving local risk.

2.2 Smoothness Map

The smoothness derived by the proposed moving local risk minimization is closely related to the choice of window size r. Here we investigate the change of the local smoothness using a range of varying window size r. In other words, we estimate the smooth function $\lambda(x, r)$ on the (x, r) plane, which is termed *smoothness map*. It might provide valuable information of the local smoothness of the underlying function. In fact, this multi-level smoothness map is motivated by the scale-space approach in computer vision. Although the smoothness map may share the motivation with SiZer of Chaudhuri and Marron (1999), it focuses on the behaviour of the local smoothness $\lambda(x, r)$ with various window size r's, whereas SiZer provides fitted values of the underlying function with various smoothing parameter λ's.

For a particular example that provides possible utility of the smoothness map, we consider the following three functions that are spatially inhomogeneous: Doppler, reflecting chirp and the function in Figure 2. For each function, we generate noisy data contaminated by Gaussian noise with SNR=5, and then obtain the corresponding smoothness maps. The first column of Figure 4

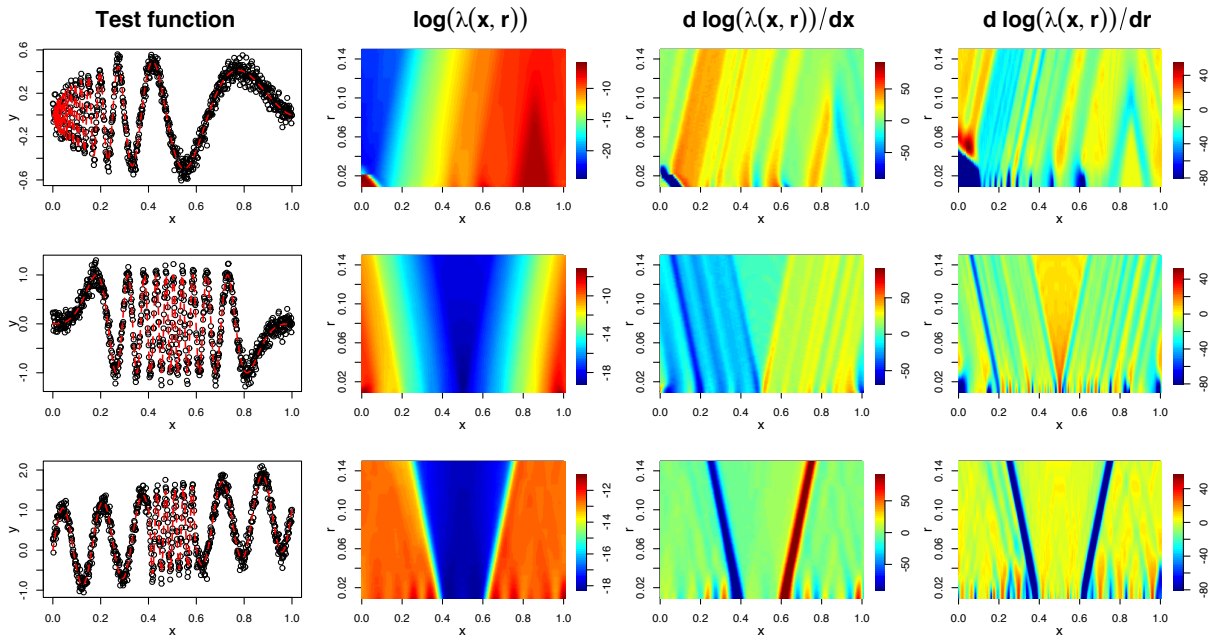

Figure 4: Three test functions with noisy observations with snr=5 (the first column), the estimated smoothness map $\lambda(x, r)$ on the (x, r) plane (the second column), and derivatives of $\lambda(x, r)$ w.r.t x and r, respectively (the third and fourth columns).

shows the three functions and their noisy observations, and the second column displays estimated smoothness maps for each test function. Furthermore, it is possible to obtain the rate of change of smoothness with respect to x as well as r, which might be useful for identifying the smoothness and further statistical analyses. The third and last columns show the derivative of $\lambda(x, r)$ with respective to x and the derivative of $\lambda(x, r)$ with respective to r, respectively. As shown in the figures, we obtain a better understanding of the characteristics of the inhomogeneous functions via smoothness map. Moreover, from the results in the third row, it seems that the smoothness map with a small r is capable of detecting change points of the local smoothness effectively.

2.3 Extension to Spatial Data

Here we consider a nonparametric estimation of the smoothness of a spatial field (or a two-dimensional function) defined on the spatial domain. This extension to spatial data is straightforward. The main difference is to replace univariate smoothing splines with thin-plate smoothing

splines. Suppose that we observe a set of noisy data $\{z\}_{i=1}^n$ on the domain $\{(u_i, v_i)\}_{i=1}^n$ from the model

$$z_i = f(u_i, v_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

where the set of location (u_i, v_i) is randomly selected on $[0,1]\times[0,1]$, and ε_i are independently and identically distributed Gaussian errors. Under the following penalized least squares framework

$$\frac{1}{n}\sum_{i=1}^{n}\{z_i - f(u_i, v_i)\}^2 + \int\int_{\mathbb{R}^2} \lambda(u,v)\Big(\frac{\partial^2 f}{\partial u^2} + 2\frac{\partial^2 f}{\partial u \partial v} + \frac{\partial^2 f}{\partial v^2}\Big) du dv,$$

we would like to estimate the two-dimensional local smoothness $\lambda(u, v)$ of a spatially inhomogeneous two-dimensional function $f(u, v)$. For this purpose, we extend the proposed algorithm explained in (4), that is, the estimated local smoothness at the spatial location (u, v) is defined as

$$\lambda^{(L)}(u,v) \triangleq \arg\min_{\lambda(u,v)} \hat{R}^{(L)}_{\lambda(u,v)}(u,v),$$

where $\hat{R}^{(L)}_{\lambda(u,v)}(u,v)$ denotes the converged two-dimensional the moving local risk at the spatial location (u, v) in the rectangular region $[u-r, u+r]\times[v-r, v+r]$.

To evaluate the empirical performance of the above extension, we consider two test functions which are spatially homogeneous and inhomogeneous, respectively, as shown in the left column of Figure 5. The generic forms of the test function are listed as: $f_1(u,v) = \sin(6\pi(u-0.5))\sin(6\pi(v-0.5))$ and $f_2(u,v) = \sqrt{u(1-u)}\sin\left(\frac{2\pi(1+2^{-2.2})}{u+2^{-2.2}}\right) + \sqrt{v(1-v)}\sin\left(\frac{2\pi(1+2^{-2.2})}{v+2^{-2.2}}\right)$.

For the noisy data shown in the middle column of the figure, the SNR is set to be 5 and the sample size is $n = 1000$. In this example, we use $r = 0.15$. The right column of Figure 5 shows the estimates of the smoothness functions $\log(\lambda(u,v))$ obtained by the proposed method. As one can see, the estimate of $\lambda(u,v)$ for the noisy data in Figure 5(b) is almost flat, which represents the constant smoothness of the first test function in Figure 5(a), whereas the estimate of $\lambda(u,v)$ for the noisy data in Figure 5(e) increases as u and v values are increasing on the spatial domain, and hence, the result reflects the spatial variability of the second test function effectively.

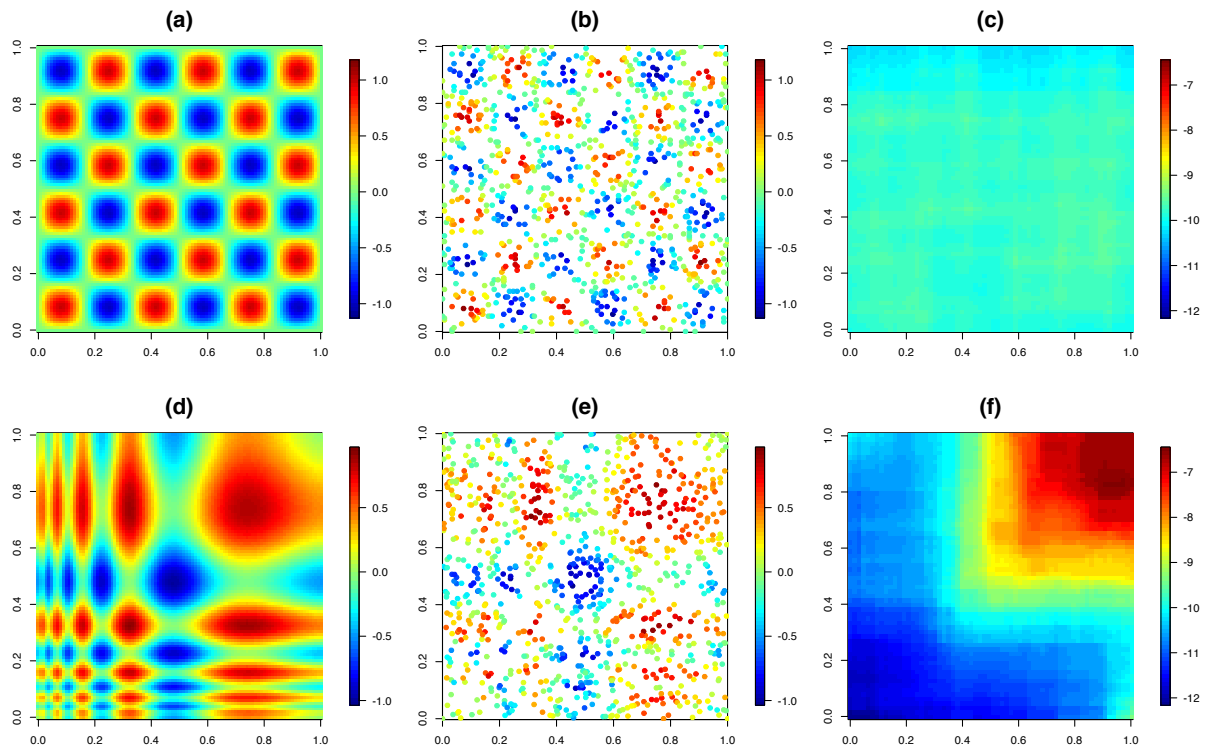

Figure 5: The left panel displays the test function, the middle panel displays noisy observation with snr=5 and the right panel displays $\log(\lambda(u, v))$ with $r = 0.15$.

3 Three Applications

In this section, we discuss some applications based on the estimated local smoothness by the proposed local risk minimization, which include functional clustering and change point detection problems. Furthermore, we investigate an enhancement of constructing confidence intervals of smoothing splines, which provides more uniform coverage rate than the conventional confidence intervals based on a single smoothing parameter.

3.1 Functional Clustering

We would like to perform clustering of functions that hold various types of smoothness. For this purpose, we use the information of the local smoothness obtained from the proposed approach, instead of the data on the physical domain. In some cases, the information of the local smoothness is rather easy to identify, compared to that of function itself; hence, it can be expected that the

functional clustering based on the local smoothness performs well.

To illustrate the motivation, we present a simple example for functional clustering. To that end, we consider three types of chirp signals shown in Figure 6. For each chirp signal, we generate 50 sets of noisy data contaminated by Gaussian noise with SNR=5 and $n = 800$. The design points x_i's are generated from uniform distribution on $[0, 1]$. For each case of $150(= 3 \times 50)$ datasets, we estimate its smooth function $\lambda(x)$.

As a discrepancy measure for clustering, we consider the following distance,

$$d(f, g) = \Big(\int \big|\lambda_f(x) - \lambda_g(x)\big|^2 dx \Big)^{1/2},$$

where $\lambda_f(x)$ and $\lambda_g(x)$ denote the estimated local smoothness functions of f and g, respectively. We then perform a hierarchical cluster analysis with the distance measure, $d(f, g)$. The clustering results are shown in Figure 6. In the second row of the figure, the dendrogram for clustering with total 150 noisy signals is displayed. As one can see, the above approach performs a perfect clustering. See the bottom row of Figure 6 for grouping according to class. The third row shows the estimated smooth functions according to group.

3.2 Change Point Detection of Spatial Homogeneity

The goal of this application is to develop a method of detecting change points of spatial homogeneity. The key idea of the proposed detection method is to split the physical domain that has the maximum derivative of difference between local smoothnesses of adjacent regions. To be specific, we define a possible change point as

$$x^* = \arg\max_x \big|\widehat{Q}'(x)\big|, \tag{6}$$

where $Q(x) = \big|\log(\lambda_\ell(x)) - \log(\lambda_r(x))\big|$, the local smoothness in the left region of the point x is obtained by

$$\lambda_\ell(x) = \arg\min_{\lambda(x)} \sum_{\{i:\ x_i < x\}} \big[\{S_{\lambda(x)}\hat{\boldsymbol{f}}_\lambda(x_i) - \hat{f}_\lambda(x_i)\}^2 + \hat{\sigma}^2 s_{\lambda(x)}(x_i)\big],$$

and similarly the local smoothness in the right region of the point x is estimated by

$$\lambda_r(x) = \arg\min_{\lambda(x)} \sum_{\{i:\ x_i \geq x\}} \big[\{S_{\lambda(x)}\hat{\boldsymbol{f}}_\lambda(x_i) - \hat{f}_\lambda(x_i)\}^2 + \hat{\sigma}^2 s_{\lambda(x)}(x_i)\big].$$

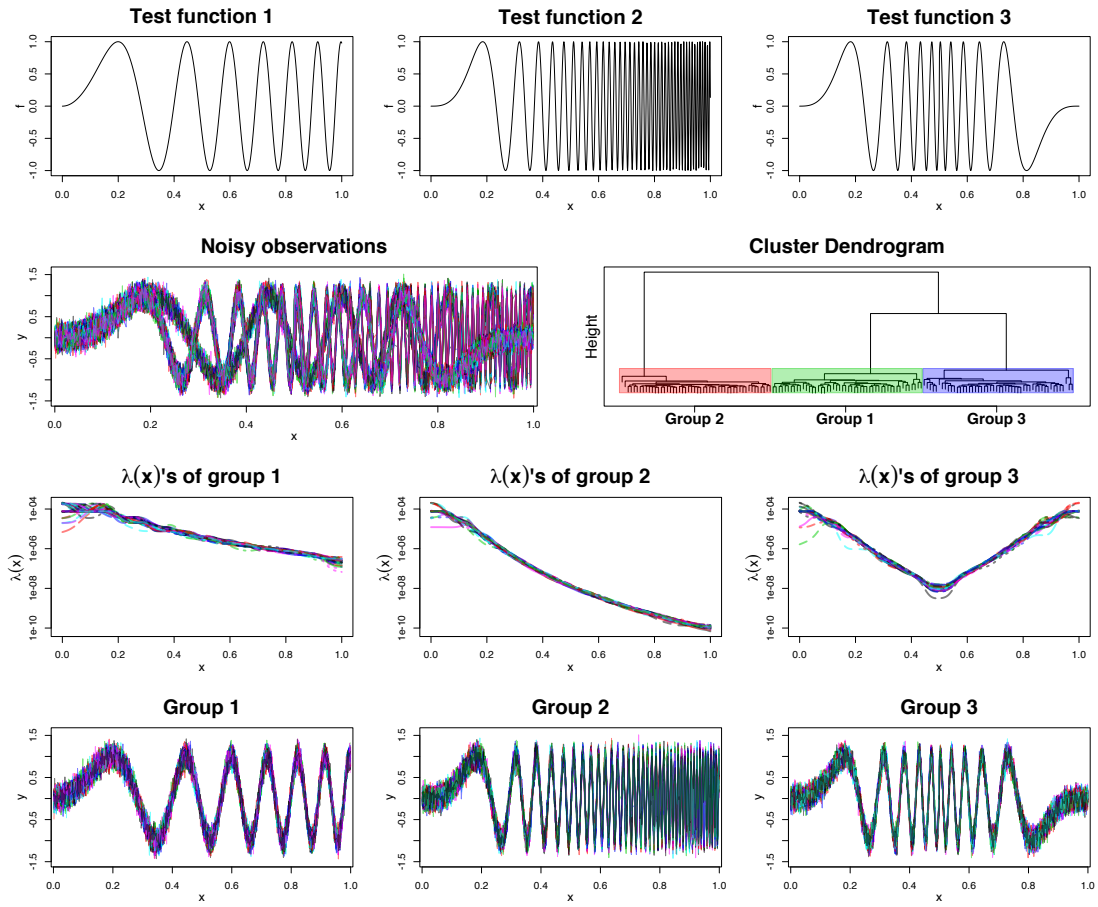

Figure 6: Clustering of three types of chirp signals.

Finally $\widehat{Q}(x)$ is defined as a smooth version of $Q(x)$ in order to obtain its derivative. For estimation of $\lambda_\ell(x)$ and $\lambda_r(x)$, we use the proposed moving local risk minimization.

For a particular example, we apply the above criterion to the noisy data in Figure 2 for detection of smoothness change. We note that the test function in Figure 2 holds two change points of spatial homogeneity. To estimate the local smoothness function, we use the true local risk of (2), the unbiased local risk of (5), and the proposed moving local risk. Figure 7 shows the results obtained by (a) the true local risk, (b) the unbiased local risk and (c) the proposed moving local risk, respectively. The top row of Figure 7 shows the negative $\log(\lambda_\ell(x))$ (solid line) as the smoothness function on the left region and the negative $\log(\lambda_r(x))$ (dashed line) as the smoothness function on the right one. The middle row shows the results of $Q(x)$ (dot) and $\widehat{Q}(x)$ (dashed line). As shown, $\widehat{Q}(x)$ by the proposed local risk is almost identical to that by the true

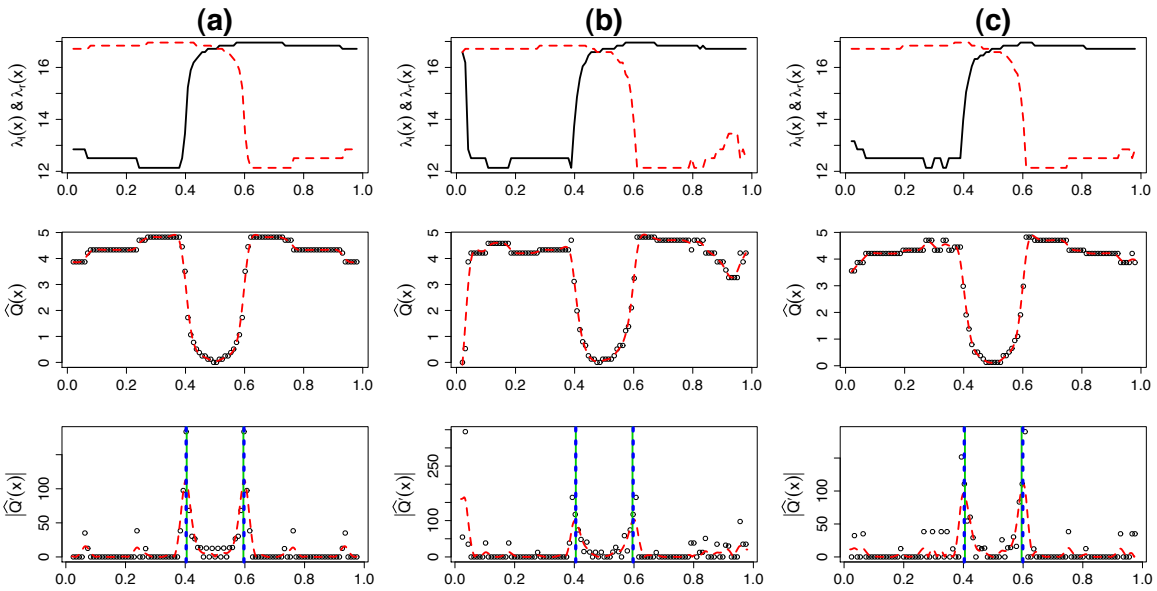

Figure 7: Change points detection by minimization of (a) true local risk, (b) unbiased local risk, and (c) the proposed moving local risk, respectively.

local risk, while $\widehat{Q}(x)$ based on the unbiased local risk suffers from large variability especially at boundary. The bottom row shows the results of $|\hat{Q}'(x)|$ for detection of changing points. As one can see, it seems that the criterion (6) based on the local smoothness is useful for detecting the change points of spatial homogeneity.

3.3 Improved Confidence Intervals of Smoothing Splines

In this section, we consider a problem of constructing confidence intervals for nonparametric quantile regression with an emphasis of smoothing splines. More specifically, we propose a method for improving confidence intervals based on the proposed moving local risk minimization, which provides more uniform pointwise coverage, compared to confidence intervals constructed by a global smoothing parameter. Wahba (1983) suggested $100(1-\alpha)\%$ pointwise confidence intervals for the underlying mean function $f(x)$ at the point x_i as

$$C_0(x_i, \alpha) := \hat{f}_\lambda(x_i) \pm z_{\alpha/2}\sqrt{\hat{\sigma}^2 S_\lambda(i,i)}, \tag{7}$$

where $\hat{\sigma}^2 = \sum_{i=1}^n \{y_i - \hat{f}_\lambda(x_i)\}^2/\mathrm{tr}(I - S_\lambda)$ and the smoothing parameter is chosen by GCV. A function estimate based on a single global smoothing parameter performs well in the average sense,

but it occasionally underfits and/or overfits the data; hence the corresponding confidence intervals may not hold the desired confidence level $1-\alpha$ uniformly at all design points when the bias of the function estimate are not uniform. To overcome this problem, Cummins, Filloon and Nychka (2001) proposed an approach based on local cross-validation criterion (LCV), which improves the coverage probability at design points, compared to confidence interval based on GCV.

Along with the line of Cummins *et al.* (2001), in order to obtain improved confidence intervals that provides more uniform pointwise coverage, we consider the use of local smoothness $\lambda(x_i)$ that minimizes the proposed moving local risk. Hence, the proposed pointwise confidence intervals for the function $f(x_i)$ with local smoothness estimation $\lambda(x_i)$ is given by

$$C_1(x_i,\alpha) := \hat{f}_{\lambda_i}(x_i) \pm z_{\alpha/2}\sqrt{\hat{\sigma}^2 S_{\lambda_i}(i,i)}, \tag{8}$$

where $\hat{f}_{\lambda_i}(x_i)$ denotes smoothing spline estimates with $\lambda_i = \lambda(x_i)$ at the point x_i.

For a numerical example, we consider a noisy Doppler signal contaminated by Gaussian noises with SNR=5 and $n = 400$ in Figure 8(a). We generate 500 sets of noisy data and then compute the pointwise confidence intervals $C_0(x,\alpha)$ and $C_1(x_i,\alpha)$ with $\alpha = 0.05$ for each set, respectively. Figure 8(b) displays the average values of pointwise confidence intervals over all 500 sets. As one can see, it seems that the width of the proposed confidence interval $C_1(x_i,\alpha)$ is narrower that $C_0(x_i,\alpha)$ by GCV at the smooth area of the function. For further evaluation, we consider a pointwise coverage probability at x_i defined as $\sum_{k=1}^{500} I\big(f(x_i) \in C(x_i,\alpha)\big)/500$, where $C(x_i,\alpha)$ is either $C_0(x_i,\alpha)$ of (7) or $C_1(x_i,\alpha)$ of (8). Figures 8(c) and (d) show the pointwise coverage probability by $C_0(x_i,\alpha)$ and $C_1(x_i,\alpha)$, respectively. Furthermore, we compute 'breakouts above' which implies the number of times that the true function at the local point is higher than the upper confidence interval, and 'breakouts below' which means the number of times that the true function at the local point is lower than the lower confidence interval. The results are shown in Figures 8(e) and (f). As shown in the figures, we observe that the proposed $C_1(x_i,\alpha)$ provides much more uniform confidence interval than $C_0(x_i,\alpha)$.

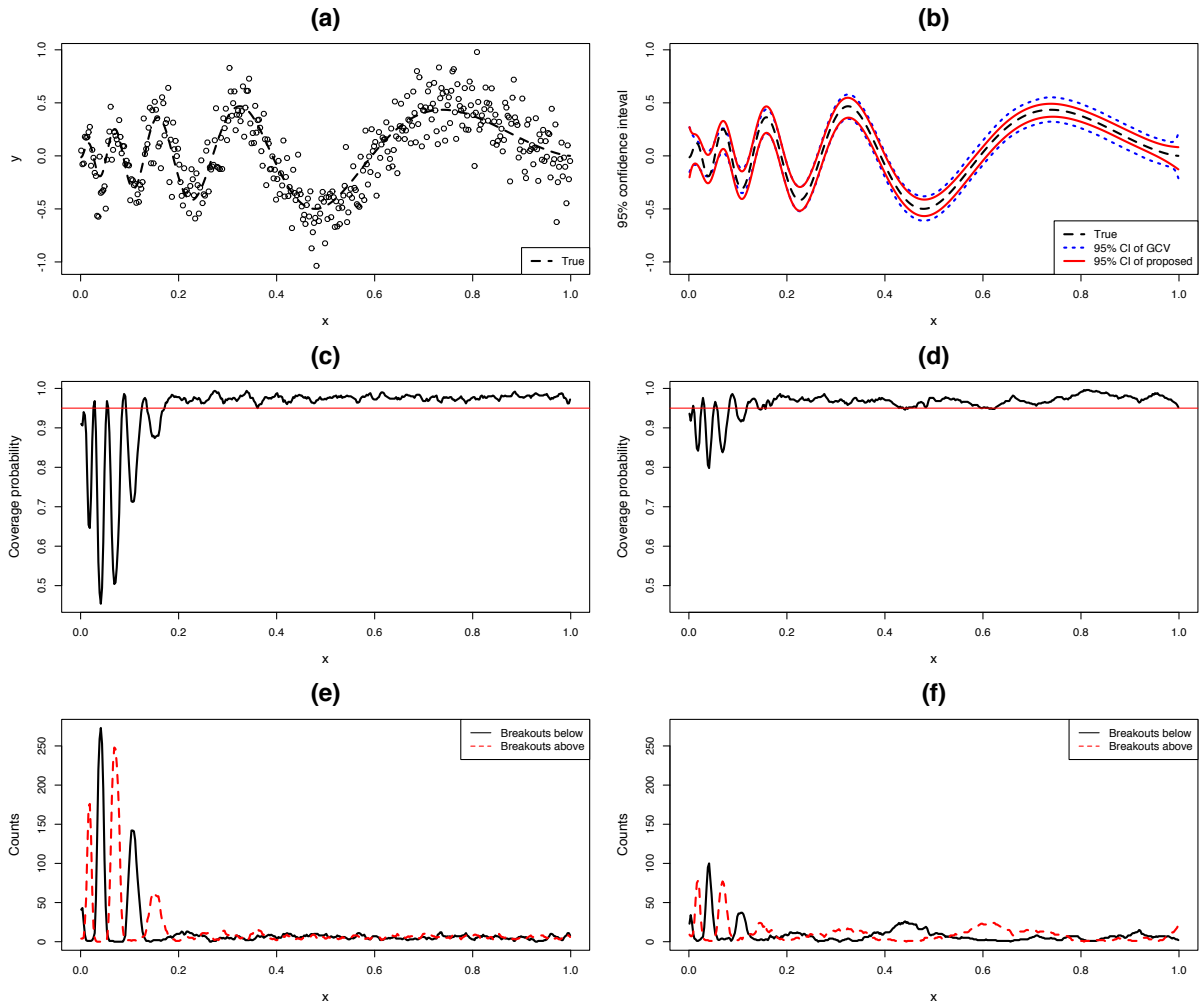

Figure 8: The pointwise coverage probability of 95% confidence intervals of estimates and counts of breakouts based on GCV and proposed methods.

4 Conclusion

We have proposed a nonparametric method to estimate the local smoothness of spatially inhomogeneous functions. To achieve this goal, we have introduced a local risk of a particular point on an interval around the point and computed the local risk at the entire domain by moving the window along with all design points. The proposed iterative procedure has provided an accurate estimate of the true risk, so that it is able to obtain an efficient estimation of the local smoothness of the function. Throughout some applications, the proposed method shows its potential usage for functional clustering, change point detection and enhancement of confidence interval for

smoothing splines. Furthermore, we have generalized the procedure to the bivariate cases.

There are some issues to be further studied. An objective selection of the window size is needed since an automatic choice might help analyzing a very large data. In addition, it is worth to consider the case that noises are skewed or correlated.

References

Chaudhuri, P. and Marron, J. S. (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association*, **94**, 807–823.

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

Cummins, D. J., Filloon, T. G. and Nychka, D. (2001). Confidence intervals for nonparametric curve estimates: Toward more uniform pointwise coverage. *Journal of the American Statistical Association*, **96**, 233–246.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*, Chapman & Hall, London.

Jang, D. and Oh, H.-S. (2011). Enhancement of spatially adaptive smoothing splines via parameterization of smoothing parameters. *Computational Statistics & Data Analysis*, **55**, 1029–1040.

Lee, T. C. M. (2004). Improved smoothing spline regression by combining estimates of different smoothness. *Statistics & Probability Letters*, **67**, 133–140.

Pintore, A., Speckman, P. and Holmes, C. C. (2006). Spatially adaptive smoothing splines. *Biometrika*, **93**, 113–125.

Wahba, G.(1983). Bayesian 'confidence intervals' for the cross-validated smoothing spline. *Journal of the Royal Statistical Society*, Series B, **45**, 133–150.

Wahba, G. (1990). Spline models for observational data. In *CBMS-NSF, Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia.

Wahba, G. (1995). Discussion of a paper by Donoho et al. *Journal of the Royal Statistical Society, Series* B, **57**, 360–361.