MDPI

Article

~~Nonparametric~~ Edge-preserving Denoising of Image Sequences

Fan Yi [1,*] and Peihua Qiu [1]

[1] Department of Biostatistics, University of Florida, Gainesville, Florida, USA
* Correspondence: yifan@ufl.edu; Tel.: +1-352-745-4977

Abstract: To monitor the Earth's surface, the satellite of the NASA Landsat program provides us image sequences of any region on the Earth constantly over time. These image sequences give us a unique resource to study the Earth surface, changes of the Earth resource over time, and their implications in agriculture, geology, forestry, and more. Besides natural sciences, image sequences are also commonly used in functional magnetic resonance imaging (fMRI) of medical studies for understanding the functioning of brains and other organs. In practice, observed images almost always contain noise and other contaminations. For a reliable subsequent image analysis, it is important to remove such contaminations in advance. This paper focuses on image sequence denoising, which has not been well discussed in the literature yet. To this end, an edge-preserving image denoising procedure is suggested. The suggested method is based on a jump-preserving local smoothing procedure, in which the bandwidths are chosen such that the possible spatio-temporal correlation in the observed image intensities is accommodated properly. Both theoretical arguments and numerical studies show that it works well in various cases considered.

Keywords: Bandwidth selection; Correlation; Edge-preserving image denoising; Image sequence; Jump regression analysis; Local smoothing; Nonparametric regression; Spatio-temporal data

1. Introduction

The Landsat project of the US Geological Survey (USGS) and NASA has launched 8 satellites since 1972 to continuously provide scientifically valuable images of the Earth's surface. These images can be freely accessed by researchers around the world (cf., Zanter 2016). This rich archive of the Landsat images has become a major resource for scientific research about the Earth's surface and resources in different scientific disciplines, including forest science, climate science, agriculture, ecology, fire science, and many more. As a demonstration, Figure 1 shows two images of the Las Vegas area in Nevada taken in 1984 and 2007, respectively. These two images clearly show the increasing urban sprawl of Las Vegas during the 23-year period, and consequently the environment in that region has changed dramatically. The current satellite (i.e., the Landsat 8) can deliver an image of a given region roughly every 16 days. So, we have an image sequence of that region collected sequentially over time stored in the Landsat database, and that sequence is increasing all the time. Image sequences are commonly used in many other applications, including functional magnetic resonance imaging (fMRI) in neuroscience and quality control in manufacturing industries (Qiu 2018). In practice, observed images usually contain noise and other contaminations (Gonzalez and Woods 2018). For reliable subsequent image analyses, such contaminations should be removed in advance. In the image processing literature, to remove noise from an observed image is referred to as *image denoising*. This paper focuses on image denoising for analyzing observed image sequences.

In the literature, there has been extensive discussion on image denoising (Qiu 2007). Many early methods in the computer science literature are based on the Markov random field (MRF) framework, in which observed image intensities of an image are assumed to have the Markov property that the observed intensity at a given pixel depends only on

Citation: Yi, F.; Qiu, P. Nonparametric Denoising of Image Sequences. *Entropy* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

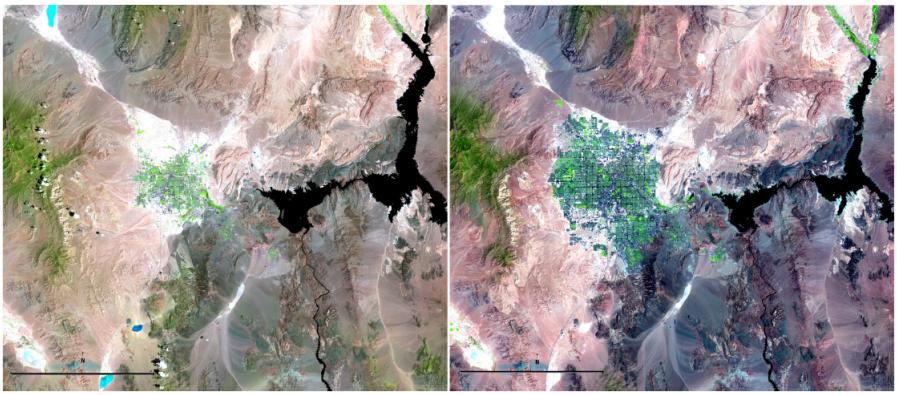


Figure 1. Two Landsat images of Las Vegas area taken in 1984 (left panel) and 2007 (right panel).

the observed intensities in a neighborhood of the given pixel (Geman and Geman 1984). Then, if the true image is assumed to have a prior distribution which is also an MRF, its posterior distribution would be an MRF too and consequently the true image can be estimated by the *maximum a posteriori (MAP)* estimator (e.g., Besag 1986, Fessler et al. 2000, Geman and Geman 1984). Other popular image denoising methods include those based on diffusion equations (e.g., Perona and Malik 1990, Weickert 1998), total variation (Beck and Teboulle 2009, Rudin et al. 1992, Yuan et al. 2012), wavelet transformations (e.g., Chang et al. 2000, Mrázek et al. 2003), jump regression analysis (e.g., Gijbels et al. 2006, Qiu 1998, 2009, Qiu and Mukherjee 2012), adaptive weights smoothing (e.g., Polzehl and Spokoiny 2000), spatial adaption (e.g., Kervrann and Boulanger 2006) and more.

Although there are already many existing methods for image denoising, almost all of them are for handling individual observed images. So far, we have not found much discussion on denoising of image sequences which is the focus of the current paper. [strength of our method] For a given image sequence, it often describes a gradual change over time of the underlying process. For instance, the sequence of images of the Las Vegas area acquired by the Landsat satellite (cf., Figure 1) describes the gradual change of the Earth surface in that area over time. As mentioned above, two consecutive images in the sequence acquired by the current Landsat satellite are only about 16-day apart. So, their difference should be very small. But, the images could be substantially different after a long period of time, as shown in Figure 1. In such applications, it should be reasonable to assume that edge locations in different images either do not change or change gradually over time. To handle such image sequences, the neighboring images should be useful when denoising the image at a given time point, or *information in neighboring images should be shared during image denoising*. By noticing such features of image sequences, we propose an edge-preserving image denoising procedure for analyzing image sequences in this paper. [strength of our method] Our proposed method is based on the jump regression analysis (JRA) that is for regression modeling when the underlying regression function has jumps or other singularities (Qiu 2005). It is a local smoothing procedure, and the possible spatio-temporal correlation in the observed image data has been accommodated properly in its construction. [some novelties of the proposed method, besides its information sharing feature mentioned above] Both theoretical arguments and numerical studies show that this method works well in various different cases.

The remaining parts of the article are organized as follows. The proposed method is described in detail in Section 2. Its statistical properties and numerical studies about its performance in different finite-sample cases are presented in Section 3. Several concluding remarks are provided in Section 4. Some technical details are given in Appendices.

2. Materials and Methods

This section describes our proposed method in two parts. A JRA model for describing an image sequence and the model estimation are discussed in Subsection 2.1. Selection of several parameters used in model estimation is discussed in Subsection 2.2.

2.1. JRA model and its estimation

To describe an image sequence, let us consider the following JRA model:

$$Z_{ijk} = f(x_i, y_j; t_k) + \varepsilon_{ijk}, \quad i = 1, 2, \ldots, n_x, j = 1, 2, \ldots, n_y, k = 1, 2, \ldots, n_t, \tag{1}$$

where Z_{ijk} is the observed image intensity level at the (i,j)th pixel (x_i, y_j) and at the kth time point t_k, $f(x_i, y_j; t_k)$ is the true image intensity level, and ε_{ijk} is the pointwise random noise with mean 0 and variance σ^2. In model (1), spatio-temporal data correlation is allowed, namely, $\{\varepsilon_{ijk}\}$ could be correlated over i, j and k. For image data, the pixel locations are usually regularly spaced. Without loss of generality, it is assumed that they are equally spaced in the design space $\Omega = [0,1] \times [0,1]$, namely, $(x_i, y_j) = (i/n_x, j/n_y)$, for all i and j, where n_x and n_y are the numbers of rows and columns, respectively. The observation times $\{t_k, k = 1, 2, \ldots, n_t\}$ are also assumed to be equally spaced in the time interval $[0,1]$. The true image intensity function $f(x,y;t)$, for $(x,y) \in \Omega$, is continuous in the design space Ω at each $t \in [0,1]$, except on the edges where it has jumps.

To estimate the unknown image intensity function $f(x,y;t)$ in model (1), we consider using a local smoothing method, instead of a global smoothing method (e.g., smoothing spline method), because of a large amount of data involved in the current problem. Also, it has been well discussed in the JRA literature that conventional smoothing methods (e.g., conventional local kernel smoothing methods) would not work well for estimating models like (1) where the true image intensity function $f(x,y;t)$ has jumps at the edges, because the jumps would be blurred by such conventional methods (cf., Qiu 2005). In this paper, we suggest a jump-preserving local smoothing method for estimating (1), described in detail below. For a given point $(x,y;t) \in \Omega \times [0,1]$, define a local neighborhood

$$\begin{aligned} O(x,y;t) = &\{\left(x^{'}, y^{'}; t^{'}\right) : \left(x^{'}, y^{'}; t^{'}\right) \in \Omega \times [0,1], \\ &\sqrt{\frac{(x^{'}-x)^2}{h_x^2} + \frac{(y^{'}-y)^2}{h_y^2}} \le 1, |t^{'}-t|/h_t \le 1\}, \end{aligned}$$

where h_x, h_y and h_t are the bandwidths in the $x-$, $y-$, and $t-$axis, respectively. In $O(x,y;t)$, we first consider the following local linear kernel (LLK) smoothing procedure (Fan and Gijbels 1996):

$$\begin{aligned} \min_{a,b,c,d} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \sum_{k=1}^{n_t} &\Big\{ Z_{ijk} - \left[a + b(x_i - x) + c(y_j - y) + d(t_k - t)\right] \Big\}^2 \\ &K\left(\frac{x_i - x}{h_x}, \frac{y_j - y}{h_y}\right) K\left(\frac{t_k - t}{h_t}\right), \end{aligned} \tag{2}$$

where $K(v)$ is a density kernel function with the support $\{v : |v| \le 1\}$. The solutions to (a,b,c,d) of the minimization problem (2) are denoted as $\widehat{a}(x,y;t)$, $\widehat{b}(x,y;t)$, $\widehat{c}(x,y;t)$, and $\widehat{d}(x,y;t)$, respectively. It can be checked that they have the following expressions:

$$\begin{bmatrix} \widehat{a}(x,y;t) \\ \widehat{b}(x,y;t) \\ \widehat{c}(x,y;t) \\ \widehat{d}(x,y;t) \end{bmatrix} = \begin{bmatrix} m_{000} & m_{100} & m_{010} & m_{001} \\ m_{100} & m_{200} & m_{110} & m_{101} \\ m_{010} & m_{110} & m_{020} & m_{011} \\ m_{001} & m_{101} & m_{011} & m_{002} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{ijk} Z_{ijk} K_{ijk} \\ \sum_{ijk} (x_i - x) Z_{ijk} K_{ijk} \\ \sum_{ijk} (y_j - y) Z_{ijk} K_{ijk} \\ \sum_{ijk} (t_k - t) Z_{ijk} K_{ijk} \end{bmatrix}, \tag{3}$$

where $\sum_{ijk}$ denotes $\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\sum_{k=1}^{n_t}$, K_{ijk} denotes $K(\frac{x_i-x}{h_x}, \frac{y_j-y}{h_y})K(\frac{t_k-t}{h_t})$, and $m_{rsl} = \sum_{ijk}(x_i - x)^r(y_j-y)^s(t_k-t)^l K_{ijk}$, for $r,s,l = 0,1,2$. The LLK estimator of $f(x,y;t)$ is defined to be $\widehat{a}(x,y;t)$. The estimated gradient direction of $f(x,y;t)$ at $(x,y;t)$ is $\widehat{G}(x,y;t) = (\widehat{b}(x,y;t), \widehat{c}(x,y;t), \widehat{d}(x,y;t))'$ which indicates the direction in which the estimated plane in $O(x,y;t)$ by the LLK procedure (2) increases the fastest. If there is an edge surface in $O(x,y;t)$, then $\widehat{G}(x,y;t)$ would be (approximately) orthogonal to that surface.

In cases when there are no edges in the neighborhood $O(x,y;t)$, $\widehat{a}(x,y;t)$ would be a good estimate of $f(x,y;t)$. Otherwise, it cannot be a good estimate because $\widehat{a}(x,y;t)$ is a weighted average of all observed image intensities in $O(x,y;t)$, the jumps in the image intensity surface would be smoothed out in the weighted average, and the estimate $\widehat{a}(x,y;t)$ would be biased for estimating $f(x,y;t)$. To overcome that limitation, we consider the following one-sided smoothing idea. Let $O(x,y;t)$ be divided into two parts $O^{(1)}(x,y;t)$ and $O^{(2)}(x,y;t)$ by a plane that passes $(x,y;t)$ and is perpendicular to $\widehat{G}(x,y;t)$. See Firgure 2 for a demonstration. Then, in cases when there is an edge

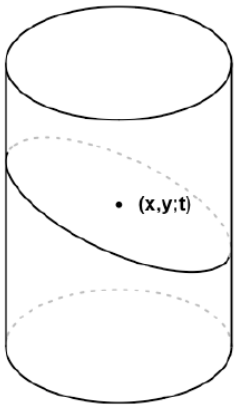


Figure 2. The neighborhood $O(x,y;t)$ is divided into two parts by a plane that passes $(x,y;t)$ and is perpendicular to the estimated gradient direction $\widehat{G}(x,y;t)$.

surface in $O(x,y;t)$, that plane would be (approximately) parallel to the edge surface. Consequently, at least one of $O^{(1)}(x,y;t)$ and $O^{(2)}(x,y;t)$ would be (mostly) located on a single side of the edge surface in such cases. Now, let us consider the following one-sided LLK smoothing procedure: for $l = 1,2$,

$$\min_{a,b,c,d} \sum_{(x_i,y_j;t_k)\in O^{(l)}(x,y;t)} \Big\{ Z_{ijk} - \big[a + b(x_i - x) + c(y_j - y) + d(t_k - t)\big] \Big\}^2 K\Big(\frac{x_i - x}{h_x}, \frac{y_j - y}{h_y}\Big) K\Big(\frac{t_k - t}{h_t}\Big). \tag{4}$$

The solutions of (4) to (a,b,c,d) are denoted as $(\widehat{a}^{(l)}(x,y;t), \widehat{b}^{(l)}(x,y;t), \widehat{c}^{(l)}(x,y;t), \widehat{d}^{(l)}(x,y;t))$, for $l = 1,2$. Intuitively, when there are no edges in $O(x,y;t)$, $\widehat{a}(x,y;t)$, $\widehat{a}^{(1)}(x,y;t)$ and $\widehat{a}^{(2)}(x,y;t)$ are all consistent estimates of $f(x,y;t)$ under some regular conditions. In such cases, $\widehat{a}(x,y;t)$ would be preferred since it averages more observations and consequently it would have a smaller variance. When there are edges in $O(x,y;t)$, $\widehat{a}(x,y;t)$ would not be a good estimate of $f(x,y;t)$ as explained above, but one of $\widehat{a}^{(1)}(x,y;t)$ and $\widehat{a}^{(2)}(x,y;t)$ should estimate $f(x,y;t)$ well. Therefore, in all cases, at least one of the three estimators $\widehat{a}(x,y;t)$, $\widehat{a}^{(1)}(x,y;t)$ and $\widehat{a}^{(2)}(x,y;t)$ should estimate $f(x,y;t)$ well.

Next, we need to choose a good estimator from $\widehat{a}(x,y;t)$, $\widehat{a}^{(1)}(x,y;t)$ and $\widehat{a}^{(2)}(x,y;t)$ based on the observed data, which is not straightforward, partly because we don't know in advance whether there are edges in the neighborhood $O(x,y;t)$ and whether the edges are mostly contained in $O^{(1)}(x,y;t)$ or $O^{(2)}(x,y;t)$ if the answer to the first question is

positive. To overcome this difficulty, let us consider the following weighted residual mean squares (WRMS) of the fitted local plane by the LLK procedure (2):

$$e(x,y;t) = \Big\{ \sum_{ijk} [Z_{ijk} - \widehat{a}(x,y;t) - \widehat{b}(x,y;t)(x_i - x) - \widehat{c}(x,y;t)(y_j - y) - \widehat{d}(x,y;t)(t_k - t)]^2 K_{ijk} \Big\} / \sum_{ijk} K_{ijk}. \tag{5}$$

The above WRMS measures how well the fitted local plane describes the observed data in $O(x,y;t)$. If there are edges in $O(x,y;t)$, this quantity would be relatively large, due mainly to the jumps in the image intensity surface. Otherwise, it would be relatively small. So, the quantity $e(x,y;t)$ contains useful information about the existence of edges in $O(x,y;t)$. Similarly, we can define WRMS values for the two one-sided local planes fitted in $O^{(1)}(x,y;t)$ and $O^{(2)}(x,y;t)$. They are denoted as $e^{(1)}(x,y;t)$ and $e^{(2)}(x,y;t)$. Based on these WRMS values, we define our edge-preserving estimator of $f(x,y;t)$ to be

$$\begin{aligned} \widehat{f}(x,y;t) = {} & \widehat{a}(x,y;t) I(D(x,y;t) \le u) \\ & + \widehat{a}^{(1)}(x,y;t) I(D(x,y;t) > u) I(e^{(1)}(x,y;t) < e^{(2)}(x,y;t)) \\ & + \widehat{a}^{(2)}(x,y;t) I(D(x,y;t) > u) I(e^{(1)}(x,y;t) > e^{(2)}(x,y;t)) \\ & + \frac{\widehat{a}^{(1)}(x,y;t) + \widehat{a}^{(2)}(x,y;t)}{2} I(D(x,y;t) > u) I(e^{(1)}(x,y;t) = e^{(2)}(x,y;t)), \end{aligned} \tag{6}$$

where $D(x,y;t) = \max(e(x,y;t) - e^{(1)}(x,y;t), e(x,y;t) - e^{(2)}(x,y;t))$, $I(\cdot)$ is the indicator function, and $u > 0$ is a threshold parameter. By (6), it is obvious that $\widehat{f}(x,y;t)$ is defined to be one of $\widehat{a}(x,y;t)$, $\widehat{a}^{(1)}(x,y;t)$ and $\widehat{a}^{(2)}(x,y;t)$. The quantity $\widehat{a}(x,y;t)$, which is obtained from the entire neighborhood $O(x,y;t)$, is chosen if the observed data indicate no edges in $O(x,y;t)$, supported by the event $D(x,y;t) \le u$. Otherwise, one of the two one-sided quantities $\widehat{a}^{(1)}(x,y;t)$ and $\widehat{a}^{(2)}(x,y;t)$ with a smaller WRMS value is chosen. Although theoretically the event $(e^{(1)}(x,y;t) = e^{(2)}(x,y;t))$ would have probability zero to happen, the last term on the right-hand-side of (6) is still included for completeness of the definition of $\widehat{f}(x,y;t)$ and for the consideration that $e^{(1)}(x,y;t)$ and $e^{(2)}(x,y;t)$ could be considered the same in certain algorithms when their values are close.

2.2. Parameter selection

In our proposed method described in Subsection 2.1, there are four parameters h_x, h_y, h_t and u that need to be chosen properly in advance. For that purpose, it is natural to consider the cross validation (CV) procedure, especially in the current research problem where the observed data are quite large in size. However, it has been well demonstrated in the literature that the conventional CV procedure would not work well in cases when the observed data are autocorrelated, because it cannot effectively distinguish the data correlation structure from the mean structure (cf., Altman 1990, Opsomer et al. 2001). In the current problem, spatio-temperal data correlation is possible in almost all applications. Thus, the conventional CV procedure is not feasible in such cases. In the univariate regression setup, Brabanter et al. (2011) suggested a modified CV procedure for choosing smoothing parameters in cases with correlated data. This procedure is generalized here for choosing the parameters h_x, h_y, h_t and u used in the proposed method, which is described below. Let the modified CV score for choosing h_x, h_y, h_t and u be defined as

$$CV(h_x,h_y,h_t,u) = \frac{1}{n_x n_y n_t} \sum_{ijk} \Big[\widehat{f}_{-(ijk)}(x_i,y_j;t_k) - Z(x_i,y_j;t_k) \Big]^2, \tag{7}$$

where $\widehat{f}_{-(ijk)}(x_i, y_j; t_k)$ is the leave-one-out estimate of $f(x_i, y_j; t_k)$ by (2)-(6) after the observation Z_{ijk} is removed from the estimation process and after the kernel function is replaced by the so-called ϵ-optimal bimodal kernel function $K_\epsilon(v)$ defined to be

$$K_\epsilon(v) = \frac{4}{4-3\epsilon-\epsilon^3} \times \begin{cases} \frac{3}{4}(1-v^2)I(|v| \le 1), & \text{if } |v| \ge \epsilon, \\ \frac{3(1-\epsilon^2)}{4\epsilon}|v|, & \text{if } |v| < \epsilon, \end{cases} \tag{8}$$

where $0 < \epsilon < 1$ is a parameter. Based on a large simulation study, Brabanter et al. (2011) suggested choosing ϵ to be 0.1, which is adopted in this paper. Then, by the above modified CV procedure (7)-(8), the parameters h_x, h_y, h_t and u can be chosen by minimizing the modified CV score $CV(h_x, h_y, h_t, u)$.

3. Results

3.1. Statistical Properties

In this part, we discuss some statistical properties of the proposed edge-preserving image sequence denoising method (2)-(6). First, we have the following proposition.

Proposition 1. *Assume that i) the kernel function $K(v)$ used in (2) is a Lipschitz-1 continuous density function, and ii) the noise terms $\{\varepsilon_{ijk}, i = 1, 2, \ldots, n_x, j = 1, 2, \ldots, n_y, k = 1, 2, \ldots, n_t\}$ in model (1) form a strong mixing stochastic process with the following strong mixing coefficients:*

$$\begin{aligned} \alpha(d) \quad = \quad & \sup_{(ijk),(i'j'k')} \sup_{A,B} \Big\{ |P(A \cap B) - P(A)P(B)|, A \in \sigma(\varepsilon_{ijk}), B \in \sigma(\varepsilon_{i'j'k'}), \\ & \max\{|i - i'|, |j - j'|, |k - k'|\} > d \Big\}, \end{aligned}$$

which have the property that $\alpha(d) \le c_1\sigma^2\rho^{c_2 d}$, where $c_1, c_2 > 0$ and $0 < \rho < 1$ are constants, and iii) $E(\varepsilon_{111}^6) < \infty$. Let $N = n_x n_y n_t$, $H = h_x h_y h_t$, $n_{min} = \min(n_x, n_y, n_t)$, and $h_{min} = \min(h_x, h_y, h_t)$. Then, for any $(x, y; t) \in \Omega_h = [h_x, 1 - h_x] \times [h_y, 1 - h_y] \times [h_t, 1 - h_t]$, we have

$$\left| \frac{1}{NH} \sum_{ijk} K\left(\frac{x_i - x}{h_x}, \frac{y_i - y}{h_y}\right) K\left(\frac{t_i - t}{h_t}\right) - 1 \right| = O\left(\frac{1}{n_{min} h_{min}}\right),$$

$$E\left[\left| \frac{1}{NH} \sum_{ijk} \varepsilon_{ijk} K\left(\frac{x_i - x}{h_x}, \frac{y_i - y}{h_y}\right) K\left(\frac{t_i - t}{h_t}\right) \right|^2 \right] = O\left(\frac{1}{NH}\right),$$

$$E\left[\left| \frac{1}{NH} \sum_{ijk} (\varepsilon_{ijk}^2 - \sigma^2) K\left(\frac{x_i - x}{h_x}, \frac{y_i - y}{h_y}\right) K\left(\frac{t_i - t}{h_t}\right) \right|^2 \right] = O\left(\frac{1}{NH}\right).$$

Based on the results in Proposition 1, we can derive the following properties of the LLK estimates defined in (3).

Theorem 1. *Besides the conditions in Proposition 1, we further assume that the true image intensity function $f(x, y; t)$ has continuous first order partial derivatives with respect to x, y and t in the design space Ω except at the edge curves. Then, for any $(x, y; t) \in \Omega_h \setminus J_h$, we have*

$$\begin{bmatrix} \widehat{a}(x,y;t) \\ \widehat{b}(x,y;t) \\ \widehat{c}(x,y;t) \\ \widehat{d}(x,y;t) \end{bmatrix} = \begin{bmatrix} f(x,y;t) \\ f'_x(x,y;t) \\ f'_y(x,y;t) \\ f'_t(x,y;t) \end{bmatrix} + \begin{bmatrix} O(h_x^2 + h_y^2 + h_t^2) \\ O(\frac{h_x^2+h_y^2+h_t^2}{h_x}) \\ O(\frac{h_x^2+h_y^2+h_t^2}{h_y}) \\ O(\frac{h_x^2+h_y^2+h_t^2}{h_t}) \end{bmatrix} + \begin{bmatrix} O_p(\frac{1}{\sqrt{NH}}) \\ O_p(\frac{1}{h_x\sqrt{NH}}) \\ O_p(\frac{1}{h_y\sqrt{NH}}) \\ O_p(\frac{1}{h_t\sqrt{NH}}) \end{bmatrix}.$$

for any $(x,y,t) \in J_h \setminus S_h$, *we have*

$$\begin{bmatrix} \widehat{a}(x,y;t) \\ \widehat{b}(x,y;t) \\ \widehat{c}(x,y;t) \\ \widehat{d}(x,y;t) \end{bmatrix} = \begin{bmatrix} f_-(x_\tau,y_\tau;t_\tau) + d_\tau \xi_{000}^{(2)} \\ \frac{d_\tau}{\xi_{200}h_x}\xi_{100}^{(2)} \\ \frac{d_\tau}{\xi_{020}h_y}\xi_{010}^{(2)} \\ \frac{d_\tau}{\xi_{002}h_t}\xi_{001}^{(2)} \end{bmatrix} + \begin{bmatrix} O(\sqrt{h_x^2+h_y^2+h_t^2}) \\ O(\frac{\sqrt{h_x^2+h_y^2+h_t^2}}{h_x}) \\ O(\frac{\sqrt{h_x^2+h_y^2+h_t^2}}{h_y}) \\ O(\frac{\sqrt{h_x^2+h_y^2+h_t^2}}{h_t}) \end{bmatrix} + \begin{bmatrix} O_p(\frac{1}{\sqrt{NH}}) \\ O_p(\frac{1}{h_x\sqrt{NH}}) \\ O_p(\frac{1}{h_y\sqrt{NH}}) \\ O_p(\frac{1}{h_t\sqrt{NH}}) \end{bmatrix} \quad (9)$$

where $\xi_{rsl} = \int_{\Omega\times[0,1]} u^r v^s w^l K(u,v)K(w)\,dudvdw$, $\xi_{rsl}^{(2)} = \int_{Q^{(2)}} u^r v^s w^l K(u,v)K(w)\,dudvdw$, *for* $r,s,l = 0,1,2$, J *is the closure of the set of all jump points of* $f(x,y;t)$, $J_h = \{(x,y;t) : (x,y;t) \in \Omega_h, \sqrt{(x-x^*)^2/h_x^2 + (y-y^*)^2/h_y^2} \le 1, |t-t^*|/h_t \le 1$, *for any* $(x^*,y^*,t^*) \in J\}$, S *is the set of singular points in* J, *including the crossing points of two or more edges, points on an edge surface at which the edge surface does not have a unique tangent surface, and points in* J *at which the jump sizes in* $f(x,y;t)$ *are zero,* $S_h = \{(x,y;t) : (x,y;t) \in \Omega_h, \sqrt{(x-x^*)^2/h_x^2 + (y-y^*)^2/h_y^2} \le 1, |t-t^*|/h_t \le 1$, *for any* $(x^*,y^*,t^*) \in S\}$, $(x_\tau,y_\tau;t_\tau) \in J \setminus S$ *is the projection of* $(x,y;t)$ *to* J *with the Euclidean distance between the two points being* $c\sqrt{h_x^2+h_y^2+h_t^2}$, *for a constant* $0 < c < 1$, *and* $f_-(x_\tau,y_\tau;t_\tau)$ *is the smaller one of the two one-sided limits of* $f(x,y;t)$ *at* $(x_\tau,y_\tau;t_\tau)$. *In cases when* $O(x,y;t)$ *contains jumps, without loss of generality, it is assumed that* $O(x,y;t)$ *is divided by the edge surface into two parts* I_1 *and* I_2 *with a positive jump size* d_τ *from* I_1 *to* I_2 *at* $(x_\tau,y_\tau;t_\tau)$, *and* $Q^{(1)}$ *and* $Q^{(2)}$ *are the two corresponding parts in the support of* $K(u,v)K(w)$.

The next two theorems establish the consistency of the proposed edge-preserving image denoising procedure (2)-(6). First, we have the following theorem about the WRMS values defined in (5).

Theorem 2. *Assume that the conditions in Theorem 1 are satisfied,* $h_x^2+h_y^2+h_t^2 = o(1)$, $(h_x^2+h_y^2+h_t^2)/h_{min} = o(1)$, $1/(NH) = o(1)$ *and* $1/(NHh_{min}^2) = o(1)$. *Then, we have the following results: for any* $(x,y;t) \in \Omega_h \setminus J_h$,

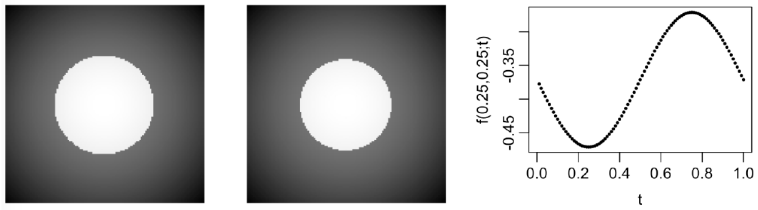
$$\begin{aligned} e(x,y;t) &= \sigma^2 + o_p(1), \\ e^{(l)}(x,y;t) &= \sigma^2 + o_p(1), \quad \text{for } l = 1,2; \end{aligned} \quad (10)$$

for any $(x,y;t) \in J_h \setminus S_h$,

$$\begin{aligned} e(x,y;t) &= \sigma^2 + d_\tau C_\tau^2 + o_p(1), \\ e^{(l)}(x,y;t) &= \sigma^2 + d_\tau \left[C_\tau^{(l)}\right]^2 + o_p(1), \quad \text{for } l = 1,2, \end{aligned} \quad (11)$$

where

$$\begin{aligned} C_\tau \;=\; & \Bigg(\int\int\int_{Q^{(1)}} \left[\xi_{000}^{(2)} + \frac{\xi_{100}^{(2)}}{\xi_{200}}u + \frac{\xi_{010}^{(2)}}{\xi_{020}}v + \frac{\xi_{001}^{(2)}}{\xi_{002}}w\right]^2 K(u,v)K(w)dudvdw + \\ & \int\int\int_{Q^{(2)}} \left[1 - \xi_{000}^{(2)} - \frac{\xi_{100}^{(2)}}{\xi_{200}}u - \frac{\xi_{010}^{(2)}}{\xi_{020}}v - \frac{\xi_{001}^{(2)}}{\xi_{002}}w\right]^2 K(u,v)K(w)dudvdw \Bigg)^{1/2}. \end{aligned}$$



(a) (b)

Figure 3. (a) The true image intensity function $f(x,y;t)$ at $t = 0.01$ (left) and $t = 0.25$ (right). (b) The temporal profile $f(0.25, 0.25; t)$ when t changes in $[0,1]$.

and

$$C_\tau^{(l)} = \left(2\int\int\int_{Q^{(1l)}}\left[B_{0l} + \frac{B_{1l}}{\xi_{200}}u + \frac{B_{2l}}{\xi_{020}}v + \frac{B_{3l}}{\xi_{002}}w\right]^2 K(u,v)K(w)dudvdw + 2\int\int\int_{Q^{(2l)}}\left[1 - B_{0l} - \frac{B_{1l}}{\xi_{200}}u - \frac{B_{2l}}{\xi_{020}}v - \frac{B_{3l}}{\xi_{002}}w\right]^2 K(u,v)K(w)dudvdw\right)^{1/2}.$$

with the quantities $Q^{(1l)}$, $Q^{(2l)}$, B_{0l}, B_{1l}, B_{2l} and B_{3l} defined as follows. Let $\overrightarrow{g} = (\frac{d_\tau}{\xi_{200}h_x}\xi^{(2)}_{100}, \frac{d_\tau}{\xi_{020}h_y}\xi^{(2)}_{010}, \frac{d_\tau}{\xi_{002}h_t}\xi^{(2)}_{001})$. Then, from (9), $\overrightarrow{g}$ is actually the asymptotic direction of the gradient vector $\widehat{G}(x,y;t)$. Let $\widetilde{O}^{(l)}(x,y;t)$, for $l = 1,2$, be two halves of the neighborhood $O(x,y;t)$ separated by a plane passing the point $(x,y;t)$ in the direction perpendicular to $\overrightarrow{g}$ and $\widetilde{Q}^{(l)}$ be the two corresponding parts in the support of $K(u,v)K(w)$. Then, $Q^{(1l)} = Q^{(1)} \cap \widetilde{Q}^{(l)}$, $Q^{(2l)} = Q^{(2)} \cap \widetilde{Q}^{(l)}$, $B_{0l} = \int\int\int_{Q^{(2l)}} K(u,v)K(w)dudvdw$, $B_{1l} = \int\int\int_{Q^{(2l)}} uK(u,v)K(w)dudvdw$, $B_{2l} = \int\int\int_{Q^{(2l)}} vK(u,v)K(w)dudvdw$, and $B_{3l} = \int\int\int_{Q^{(2l)}} wK(u,v)K(w)dudvdw$, for $l = 1,2$.

Theorem 3. *Under the conditions in Theorem 2 and the extra assumption that threshold parameter $u = u_N \to 0$ as $N \to \infty$, we have, for any $(x,y;t) \in \Omega_h$,*

$$\widehat{f}(x,y;t) = f(x,y;t) + o_p(1).$$

The proofs of these theoretical results are given in Appendices.

3.2. Numerical Studies

In this part, we study the numerical performance of our proposed method for denoising an image sequence. First, we consider a simulation example in which the true image intensity function in model (1) has the following expression:

$$f(x,y;t) = \begin{cases} -2(x-0.5)^2 - 2(y-0.5)^2 - 0.1\sin(2\pi t) + 1, & \text{if } r(x,y;t) \le 0.25^2, \\ -2(x-0.5)^2 - 2(y-0.5)^2 - 0.1\sin(2\pi t), & \text{otherwise,} \end{cases}$$

where $r(x,y;t) = (x-0.5)^2 + (y-0.5)^2 + 0.01\sin(2\pi t)$, $(x,y) \in \Omega = [0,1] \times [0,1]$, and $t \in [0,1]$. At a given value of t, $f(x,y;t)$ has a circular edge curve $r(x,y;t) = 0.25^2$ with a constant jump size 1 in $f(x,y;t)$ at the edges. The radius of the circular edge curve, $\sqrt{0.25^2 - 0.01\sin(2\pi t)}$, changes periodically over $t \in [0,1]$. The image intensity function $f(x,y;t)$ at $t = 0.01$ and 0.25 and its temporal profile $f(0.25, 0.25; t)$ are shown in Figure 3. It can be seen that both the image intensity level at a given pixel and the edge curve change gradually when t changes in $[0,1]$. In model (1), the random errors $\{\varepsilon_{ijk}, i = 1,2,\ldots,n_x, j = 1,2,\ldots,n_y, k = 1,2,\ldots,n_t\}$ are generated by the function spatialnoise() in the R-package neuRosim (cf., Welvaert et al. 2011). In that R function,

there are two parameters ρ and σ to specify in advance, where ρ controls the data autocorrelation in all three dimensions and σ is the common standard deviation of the random errors. In all our examples, σ is fixed at 0.1, 0.2 or 0.3, and ρ is fixed at 0.1, 0.3 or 0.5, to study the possible impact of data noise level and data correlation on the performance of the proposed method. Without loss of generality, we set $n_x = n_y$ in all examples. In the model estimation procedure (2)-(6), we set $h_x = h_y$, and the kernel function $K(v)$ is chosen to be the following truncated Gaussian density function:

$$K(v) = \begin{cases} \frac{\exp(-v^2/2)-\exp(-0.5)}{2\pi-3\pi\exp(-0.5)}, & \text{if } |v| \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

In cases when $\sigma = 0.1$, 0.2 or 0.3, $n_x = 64$ or 128, $n_t = 50$ or 100, $\rho = 0.1$, 0.3 or 0.5, the MSE values of the estimator $\widehat{f}(x, y; t)$ defined in (6) are presented in Table 1, along with the corresponding parameters h_x, h_t and u selected by the modified CV procedure (7)-(8). In each case considered, the MSE value is computed based on 10 replicated simulations. For comparison purposes, the optimal MSE value of the estimator $\widehat{f}(x, y; t)$ when its parameters h_x, h_t and u are chosen such that the MSE value reaches the minimum in each case considered is also presented in the table, along with the corresponding parameter values. From the table, we can have the following conclusions. i) The MSE values are smaller when either n_x or n_t is larger, which confirms the consistency results discussed in Section 3. ii) When ρ is larger (i.e., the spatio-temporal data correlation is stronger), the MSE values are larger. So, data correlation does have an impact on the performance of the proposed method, which is intuitively reasonable. iii) By comparing the MSE and the optimal MSE values, we can see that the MSE values are usually larger than their optimal values, but their differences are not that big in almost all cases considered. This conclusion indicates that the modified CV procedure (7)-(8) for determining the values of the parameters (h_x, h_t, u) is quite effective. iv) The parameter values chosen by the modified CV procedure (7)-(8) are quite close to the optimal parameter values in most cases considered.

Next, we compare our proposed method, denoted as NEW, with some alternative methods described below. The first alternative method is the conventional LLK procedure (2), by which $f(x, y; t)$ is estimated by $\widehat{a}(x, y; t)$ defined in (3). Its bandwidths are chosen by the conventional CV procedure, without considering any possible spatio-temporal data correlation. As explained in Subsection 2.1, this estimator would blur edges while removing noise. The second alternative method is to use $\widehat{a}(x, y; t)$ for estimating $f(x, y; t)$, but its bandwidths are chosen by the modified CV procedure (7)-(8). The above two alternative methods are denoted as LLK-C and LLK, respectively, where LLK-C denotes the first conventional LLK procedure that does not accommodate data correlation. The third alternative method is the one by Gijbel et al. (2006) which is for edge-preserving image denoising of a single image. To apply this method to the current problem, individual images collected at different time points can be denoised by it separately. This method assumes that the observed image intensities at different pixels are independent of each other, and thus its bandwidths can be chosen by the conventional CV procedure. This method is denoted as GLQ. The fourth alternative method is to use $\widehat{f}(x, y; t)$ in (6) to estimate $f(x, y; t)$, but the parameters (h_x, h_t, u) are chosen by the conventional CV procedure. This method is denoted as NEW-C. [Explanations of the alternative methods in numerical comparison] By considering all these four alternative methods (i.e., LLK-C, LLK, GLQ and NEW-C), we can check whether the current problem to denoise an image sequence can be handled properly by the conventional LLK procedure with or without using the modified CV procedure, by an existing edge-preserving image denoising method designed for denoising a single image, or by the proposed method without considering the possible spatio-temporal data correlation. To evaluate their

Table 1: In each entry, MSE of $\widehat{f}(x, y; t)$ in (6) is presented in the first line with its standard error (in parenthesis); the corresponding values of (h_x, h_t, u) chosen by the modified CV procedure (7)-(8) is presented in the second line; the optimal MSE is presented in the third line with its standard error (in parenthesis); the optimal values of (h_{xy}, h_t, u) are presented in the fourth line. MSE in the table has been multiplied by 10^3 and standard error has been multiplied by 10^5.

		n_t=50		n_t=100	
σ	ρ	n_x=64	n_x=128	n_x=64	n_x=128
0.1	0.1	0.65(0.80) (0.03,0.10,0.05) 0.32(0.46) (0.04,0.07,0.025)	0.30(0.25) (0.03,0.08,0.025) 0.20(0.14) (0.03,0.05,0.025)	0.48(0.43) (0.03,0.10,0.05) 0.37(0.36) (0.03,0.08,0.025)	0.26(0.10) (0.02,0.07,0.05) 0.19(0.08) (0.02,0.05,0.025)
	0.3	0.60(0.45) (0.04,0.10,0.05) 0.49(0.35) (0.04,0.08,0.025)	0.33(0.16) (0.03,0.07,0.025) 0.30(0.16) (0.03,0.06,0.025)	0.59(0.39) (0.03,0.10,0.05) 0.50(0.37) (0.03,0.08,0.025)	0.33(0.15) (0.02,0.07,0.025) 0.29(0.22) (0.03,0.04,0.025)
	0.5	1.25(1.24) (0.03,0.10,0.05) 0.77(0.65) (0.04,0.09,0.025)	0.80(0.22) (0.02,0.07,0.025) 0.49(0.24) (0.03,0.06,0.025)	0.81(0.55) (0.03,0.10,0.05) 0.74(0.46) (0.03,0.09,0.025)	0.64(0.21) (0.02,0.04,0.025) 0.45(0.25) (0.03,0.04,0.025)
0.2	0.1	1.14(1.13) (0.04,0.10,0.025) 1.11(0.86) (0.04,0.09,0.025)	0.68(0.38) (0.03,0.08,0.025) 0.66(0.33) (0.03,0.07,0.025)	1.02(0.74) (0.04,0.10,0.025) 0.93(0.71) (0.04,0.08,0.025)	0.56(0.26) (0.03,0.07,0.025) 0.54(0.31) (0.03,0.05,0.025)
	0.3	1.69(0.91) (0.04,0.10,0.025) 1.69(1.24) (0.04,0.11,0.025)	1.03(0.54) (0.03,0.08,0.025) 1.03(0.54) (0.03,0.08,0.025)	1.32(1.08) (0.04,0.10,0.025) 1.29(1.12) (0.04,0.09,0.025)	0.78(0.41) (0.03,0.07,0.025) 0.78(0.41) (0.03,0.07,0.025)
	0.5	3.25(1.74) (0.04,0.07,0.025) 2.59(2.23) (0.05,0.10,0.025)	2.88(0.78) (0.02,0.07,0.025) 1.54(1.32) (0.04,0.09,0.025)	1.95(1.85) (0.04,0.09,0.025) 1.91(1.78) (0.04,0.11,0.025)	2.61(0.58) (0.02,0.04,0.025) 1.21(0.43) (0.03,0.08,0.025)
0.3	0.1	2.32(1.91) (0.05,0.13,0.025) 2.28(2.58) (0.05,0.11,0.025)	1.26(1.03) (0.04,0.09,0.025) 1.26(1.03) (0.04,0.09,0.025)	1.59(0.81) (0.04,0.11,0.025) 1.59(0.65) (0.04,0.10,0.025)	0.92(0.34) (0.03,0.08,0.025) 0.92(0.34) (0.03,0.08,0.025)
	0.3	3.15(2.28) (0.05,0.13,0.025) 3.14(2.45) (0.05,0.14,0.025)	1.72(1.37) (0.04,0.09,0.025) 1.71(1.52) (0.04,0.10,0.025)	2.26(1.53) (0.04,0.11,0.025) 2.21(1.31) (0.04,0.13,0.025)	1.36(0.50) (0.03,0.08,0.025) 1.33(0.41) (0.04,0.09,0.025)
	0.5	6.78(3.46) (0.04,0.09,0.05) 4.46(4.94) (0.06,0.16,0.025)	6.81(2.00) (0.02,0.07,0.05) 2.48(2.38) (0.05,0.11,0.025)	4.18(2.72) (0.04,0.10,0.025) 3.18(3.42) (0.05,0.14,0.025)	6.33(1.43) (0.02,0.04,0.05) 1.88(0.56) (0.04,0.10,0.025)

performance, in addition to the regular MSE criterion, we also consider the following edge-preservation (EP) criterion originally discussed in Hall and Qiu (2007):

$$EP(\widehat{f}) = |JS(\widehat{f}) - JS(f)|/JS(f),$$

where

$$JS(f) = \frac{1}{(n_x-2)(n_y-2)(n_t-2)} \sum_{i=2}^{n_x-1} \sum_{j=2}^{n_y-1} \sum_{k=2}^{n_t-1} \Big([f(x_{i+1}, y_j; t_k) - f(x_{i-1}, y_j; t_k)]^2 + [f(x_i, y_{j+1}; t_k) - f(x_i, y_{j-1}; t_k)]^2 + [f(x_i, y_j; t_{k+1}) - f(x_i, y_j; t_{k-1})]^2 \Big)^{1/2},$$

and JS($\widehat{f}$) is defined similarly. According to Hall and Qiu (2007), JS(f) is a reasonable measure of the cumulative jump magnitude of f at the edge locations. So, $EP(\widehat{f})$ provides a measure of the percentage of the cumulative jump magnitude of f that has been lost during data smoothing by using the estimator $\widehat{f}$. By this explanation, the smaller its value, the better. In cases when $\sigma = 0.1$, 0.2 or 0.3, $n_x = 128$, $n_t = 100$, and $\rho = 0.1$, 0.3 or 0.5, the MSE and EP values of the related methods are presented in Table 2. From the table, it can be seen that the proposed method NEW has the smallest MSE values with quite large margins among all five methods in all cases considered, except the case when $\sigma = 0.1$ and $\rho = 0.1$ where NEW-C has a lightly smaller MSE value than that of NEW due to the weak data correlation in that case. Also, NEW has much smaller EP values in all cases considered, compared to the four competing methods. This example confirms that it is nessary to consider edge-preserving procedures when denoising image sequences and the possible spatio-temporal data correlation should be taken into account during the denoising process. It also confirms the benefit to share useful information among neighboring images when denoising an image sequence.

Table 2: In each entry, the first line is the MSE value with its standard error (in parenthesis), and the second line is the EP value. MSE values in the table are in the unit of 10^3 and the standard error values are in the unit of 10^5.

σ	ρ	LLK-C	LLK	GLQ	NEW-C	NEW
0.1	0.1	2.06(0.08) 73.68%	2.10(0.06) 18.43%	0.60(0.18) 28.24%	0.24(0.11) 12.32%	0.26(0.10) 7.48%
	0.3	3.04(0.14) 124.48%	2.28(0.09) 34.40%	0.95(0.18) 43.69%	2.93(0.40) 131.28%	0.33(0.15) 10.58%
	0.5	3.89(0.24) 141.47%	3.23(0.21) 95.86%	1.42(0.42) 57.40%	3.77(0.48) 148.17%	0.64(0.21) 28.86%
0.2	0.1	4.16(0.25) 142.65%	2.93(0.15) 51.78%	1.51(0.38) 54.40%	0.86(0.25) 39.01%	0.56(0.26) 9.14%
	0.3	9.39(0.52) 291.31%	3.67(0.25) 82.84%	2.87(0.51) 94.59%	9.60(0.78) 295.72%	0.78(0.41) 15.08%
	0.5	12.80(0.94) 326.38%	11.21(0.86) 289.71%	7.75(1.32) 203.86%	13.12(1.16) 334.62%	2.61(0.58) 84.24%
0.3	0.1	7.88(0.57) 235.43%	3.94(0.26) 82.24%	3.17(0.86) 73.18%	1.01(0.37) 23.36%	0.92(0.34) 15.41%
	0.3	19.97(1.15) 461.12%	5.56(0.50) 133.33%	12.36(0.63) 261.31%	19.97(1.16) 461.13%	1.36(0.50) 25.78%
	0.5	27.64(2.09) 514.22%	23.75(1.92) 458.82%	15.75(1.71) 292.50%	28.04(2.29) 518.16%	6.33(1.43) 144.58%

In the cases when $\sigma = 0.2$ and $\rho = 0.1$, 0.3 or 0.5, Figure 4 shows the observed images at $t = 0.5$ in the first column, and the denoised images by the methods LLK-C, LLK, GLQ, NEW-C and NEW in columns 2-6. From the figure, it can be seen that the denoised images by NEW are the best in removing noise and preserving edges. As a comparison, the denoised images by LLK-C, and NEW-C are quite noisy because their

selected bandwidths by the conventional CV procedure are relatively small due to the fact the conventional CV procedure cannot distinguish the data correlation from the mean structure, as discussed in Subsection 2.2. The denoised images by LLK are quite blurry because it does not take the edges into account when denoising the images. The denoised images by GLQ is quite blurry as well since GLQ denoises individual images at different time points separately and the serial data correlation is ignored in that method.

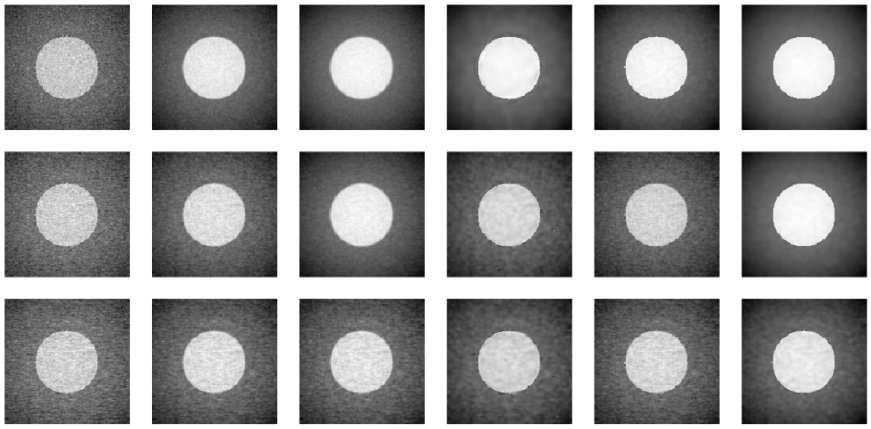


Figure 4. First column shows the observed images at $t = 0.5$ when $\sigma = 0.2$ and $\rho = 0.1$ (1st row), 0.3 (2nd row), and 0.5 (3rd row). Second to sixth columns show the denoised images by LLK-C, LLK, GLQ, NEW-C and NEW, respectively.

Next, we apply the proposed method NEW and the four alternative methods LLK-C, LLK, GLQ and NEW-C to a sequence of cell images that records the vasculogenesis process. The sequence has 100 images, and each image has 128×128 pixels. A detailed description of the data can be found in Svoboda et al. (2016). The 1st, 50th and 100th images of the sequence are shown in Figure 5. In the image denoising literature, to test

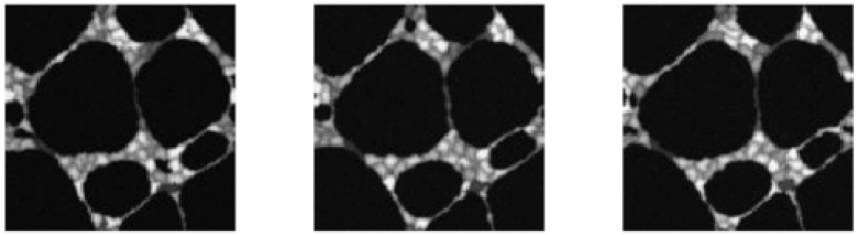


Figure 5. The 1st, 50th and 100th cell images of the image sequence for describing a vasculogenesis process.

the noise removal ability of a image denoising method, it is a common practice to add random noise at a certain level to the test images and then apply the image denoising method to the noisy test images (cf., Gijbels et al. 2006). To follow this convention, spatio-temporally correlated noise is first generated using the R-package `neuRosim` and then added to the sequence of 100 cell images described above. When generating the noise, σ is chosen to be 0.1, 0.2 or 0.3 and ρ is chosen to be 0.1, 0.3 or 0.5, as in the simulation examples presented above. The MSE and EP values of the five image denoising methods based on 10 replicated simulations are presented in Table 3. From the table, it can be seen that NEW still has smaller MSE and EP values in this example, compared to the four competing methods, except in a small number of cases when σ and ρ are relatively small.

The 50th observed test image after the spatio-temporally correlated noise with $\rho = 0.1$, 0.3 or 0.5 being added is shown in the first column of Figure 6. The denoised images by the five methods LLK-C, LLK, GLQ, NEW-C and NEW are shown in columns

Table 3: Results for denoising a sequence of 100 cell images. In each entry, the first line is the MSE value and its standard error (in parenthesis), and the second line is the EP value. MSE values in the table are in the unit of 10^3 and the standard errors are in the unit of 10^5.

σ	ρ	LLK-C	LLK	GLQ	NEW-C	NEW
0.1	0.1	1.69(0.11) 63.30%	0.97(0.08) 5.53%	1.67(0.12) 18.88%	1.69(0.12) 63.31%	1.35(0.12) 18.52%
	0.3	2.36(0.16) 77.54%	1.43(0.14) 31.64%	1.94(0.18) 25.72%	2.36(0.16) 77.55%	1.51(0.19) 7.28%
	0.5	3.21(0.25) 88.68%	2.82(0.24) 75.95%	2.28(0.29) 30.68%	3.21(0.25) 88.68%	1.92(0.31) 10.11%
0.2	0.1	3.22(17.00) 85.64%	1.47(5.54) 13.57%	3.93(0.29) 76.53%	3.22(17.00) 85.64%	1.67(0.25) 16.28%
	0.3	8.71(0.56) 189.74%	2.34(0.35) 42.07%	5.00(0.43) 91.44%	8.71(0.56) 189.75%	2.17(0.45) 4.88%
	0.5	12.12(0.94) 213.90%	10.35(0.88) 187.93%	6.41(0.86) 102.68%	12.14(0.96) 214.07%	4.48(0.90) 59.86%
0.3	0.1	3.16(0.50) 47.15%	2.01(0.28) 22.46%	5.47(0.53) 54.20%	3.16(0.50) 47.15%	1.93(0.40) 10.91%
	0.3	19.30(1.23) 308.32%	4.29(0.71) 79.75%	10.11(0.85) 161.91%	19.30(1.23) 308.32%	2.82(0.77) 14.37%
	0.5	26.96(2.09) 345.91%	22.88(1.95) 306.28%	13.36(1.82) 180.35%	27.00(2.13) 346.14%	8.75(1.85) 113.48%

2-6 of the figure. It can be seen that similar conclusions to those from Figure 4 can be made here, and the denoised images by NEW look reasonably well in removing noise and preserving edges.

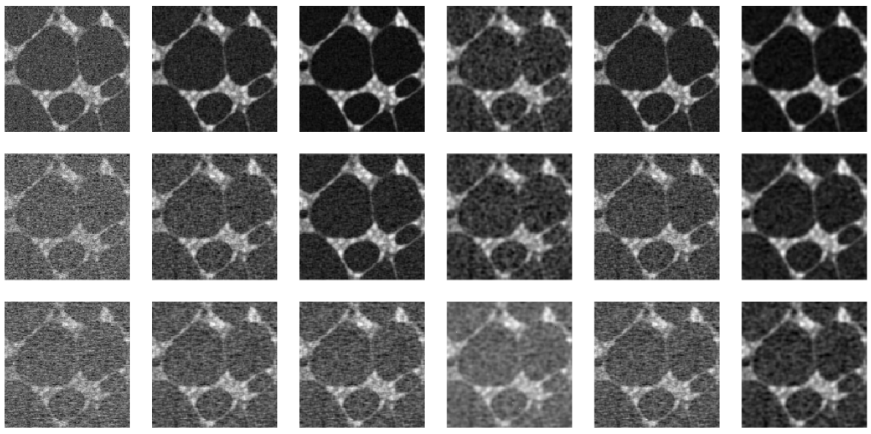


Figure 6. First column shows the 50th observed cell image after the spatio-temporally correlated noise with $\rho = 0.1$ (1st row), 0.3 (2nd row) or 0.5 (3rd row) being added. Second to sixth columns show the denoised images by LLK-C, LLK, GLQ, NEW-C and NEW, respectively.

4. ~~Discussion~~Conclusion

In the previous sections of this paper, we have described our proposed edge-preserving image denoising method for handling image sequences. Some major features of the proposed method include i) helpful information in neighboring images is shared during image denoising, ii) edge structures in the observed images can be preserved when removing noise, and iii) possible sptio-temporal data correlation can be accommodated in the related local smoothing procedure. [Novelties of the proposed method] Theoretical arguments given in Section 3 and numerical studies presented in Section 4 show that the proposed method works well in various cases considered. There are still some issues about the proposed method for future research. For instance, the current method assumes

that the true image intensity function $f(x, y; t)$ is a continuous function of x, y and t, except on the edge curves/surfaces (cf., the conditions of Theorems 1-3 in Section 3). This assumption implies that the two true images $f(x, y; t_1)$ and $f(x, y; t_2)$ at the two different time points t_1 and t_2 should be similar when t_1 and t_2 are close. In many applications, this condition should be satisfied. But, in some applications (e.g., the Landsat images), there could be a systematic shift from $f(x, y; t_1)$ to $f(x, y; t_2)$, due to a move in relative position between the camera and the image object, even when t_1 and t_2 are close. For such applications, geometric alignment among images acquired at different time points might be needed before applying the proposed method for image denoising. This is related to image registration that has been discussed extensively in the image processing literature (e.g., Zitova and Flusser 2003). [A main reason why Landsat images have not been used for testing the proposed method in this paper] In the proposed local smoothing procedure (2)-(6), each of the bandwidths (h_x, h_y, h_t) is chosen by the modified CV procedure (7)-(8) to be the same in the entire design space $\Omega \times [0, 1]$. Intuitively, relatively small bandwidths are preferred at places where the image intensity surface $f(x, y; t)$ has large curvature and relatively large bandwidths are preferred at places where the curvature of $f(x, y; t)$ is small. Thus, in some applications where the curvature of $f(x, y; t)$ could change quite dramatically in the design space, variable bandwidths might be helpful. All such issues will be studied carefully in our future research.

Author Contributions: methodology, Qiu, P.; formal analysis, Yi, F.; writing—original draft preparation, Yi, F.; writing—review and editing, Qiu, P; funding acquisition, Qiu, P; supervision, Qiu, P. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: We thank the four referees for many constructive comments and suggestions about the paper which greatly improved its quality. This research is supported in part by the NSF grant DMS-1914639.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Altman, N.S. Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association* **1990**, *85*, 749–759.
2. Beck, A.; Teboulle, M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing* **2009**, *18*, 2419–2434.
3. Besag, J. Spatial interaction and the statistical analysis of lattice systems (with discussions). *Journal of the Royal Statistical Society (Series B)* **1974**, *36*, 192–236.
4. Brabanter, K.D.; Brabanter, J.D; Suykens, J.A.K.; Moor, B.D. Kernel regression in the presence of correlated errors. *Journal of Machine Learning Research* **2011**, *12*, 1955–1976.
5. Chang, G.S.; Yu, B.; Vetterli, M. Spatially adaptive wavelet thresholding with context modeling for image denoising. *IEEE Transactions on Image Processing* **2000**, *9*, 1522–1531.
6. Fan, J.; Gijbels, I. *Local Polynomial Modelling and Its Applications*; New York: Chapman and Hall, 1996.
7. Fessler, J.A.; Erdogan, H.; Wu, W.B. Exact distribution of edgepreserving MAP estimators for linear signal models with Gaussian measurement noise. *IEEE Transactions on Image Processing3* **2000**, *9*, 1049–1055.
8. Geman, S.; Geman, D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1984**, *6*, 721–741.
9. Gijbels, I.; Lambert, A.; Qiu, P. Edge-preserving image denoising and estimation of discontinuous surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2006**, *28*, 1075–1087.
10. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 4th ed.; Pearson: New York, USA, 2018.
11. Hall, P.; Qiu, P. Blind deconvolution and deblurring in image analysis. *Statistica Sinica* **2007**, *17*, 1483–1509.
12. Kervrann, C.; Boulanger, J. Optimal Spatial Adaptation for Patch-Based Image Denoising. *IEEE Transactions on Image Processing* **2006**, *15*, 2866–2878.

13. Mrázek, P.; Weickert, J.; Steidl, G. Correspondences between wavelet shrinkage and nonlinear diffusion. In Proceedings of Scale Space Methods in Computer Vision, Isle of Skye, UK, 10-12 June 2003; Griffin, L.D., Lillholm, M., Eds..
14. Opsomer, J.; Wang, Y.; Yang, Y. Nonparametric regression with correlated errors. *Statistical Science* **2001**, *16*, 134–153.
15. Perona, P.; Malik, J. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1990**, *12*, 629–639.
16. Polzehl, J.;Spokoiny, V.G. Adaptive weights smoothing with applications to image restoration. *Journal of the Royal Statistical Society (Series B)* **2000**, *62*, 335–354.
17. Qiu, P. Discontinuous regression surfaces fitting. *The Annals of Statistics* **1998**, *26*, 2218–2245.
18. Qiu, P. *Image Processing and Jump Regression Analysis*; John Wiley & Sons: New York, USA, 2005.
19. Qiu, P. Jump surface estimation, edge detection, and image restoration. *Journal of the American Statistical Association* **2007**, *102*, 745–756.
20. Qiu, P. Jump-preserving surface reconstruction from noisy data. *Annals of the Institute of Statistical Mathematics* **2009**, *61*, 715–751.
21. Qiu, P. Jump regression, image processing and quality control (with discussions). *Quality Engineering* **2018**, *30*, 137–153.
22. Qiu, P.; Mukherjee, P.S. Edge structure preserving 3-D image denoising by local surface approximation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2012**, *34*, 1457–1468.
23. Rudin, L.; Osher, S.; Fatemi, E. Jump regression, Nonlinear total variation based noise removal algorithms. *Physica D, vol* **1992**, *60*, 259–268.
24. Svoboda, D.; Ulman, V.; Kováč, P.; Šalingová, B.; Tesařová, L.; Koutná, IK.; Matula, P. Vascular network formation in silico using the extended cellular potts model. *IEEE International Conference of Image Processing* **2016**, 3180–3183.
25. Weickert, J. *Anisotropic Diffusion in Imaging Processing*; Teubner: Stuttgart, Germany, 1998.
26. Rudin, L.; Osher, S.; Fatemi, E. **neuRosim:** An R package for generating fMRI data. *Journal of Statistical Software* **2011**, *44*, 1–18.
27. Yuan, Q.; Zhang, L.; Shen, H. Hyperspectral Image Denoising Employing a Spectral–Spatial Adaptive Total Variation Model. *IEEE Transactions on Geoscience and Remote Sensing* **2012**, *50*, 3660–3677.
28. Zanter, K. *Landsat 8 (L8) Data Users Handbook (version 2)*; Department of the Interior, U.S. Geological Survey, 2016.

Appendix A

Appendix A.1 Proof of Proposition 1

Define $B_h(x,y,t) = \{(x^{'},y^{'};t^{'}) : \sqrt{(|x^{'}-x|/h_x)^2+(|y^{'}-y|/h_y)^2} \leq 1, |t-t^{'}| \leq h_t, (x^{'},y^{'};t^{'}) \in [0,1]\times[0,1]\times[0,1]\}$, $\Delta_{ijk} = [x_{i-1},x_i]\times[y_{j-1},y_j]\times[t_{k-1},t_k]$, $x_0 = y_0 = t_0 = 0$. Then it can be seen that

$$
\begin{aligned}
& \left|\frac{1}{NH}\sum_{ijk}K\Big(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\Big)K\Big(\frac{t_k-t}{h_t}\Big)-1\right| \\
= & \left|\frac{1}{H}\sum_{ijk}\int\int\int_{\Delta_{ijk}}K\Big(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\Big)K\Big(\frac{t_k-t}{h_t}\Big)dudvdw-1\right| \\
= & \left|\frac{1}{H}\sum_{ijk}\int\int\int_{\Delta_{ijk}}K\Big(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\Big)K\Big(\frac{t_k-t}{h_t}\Big)dudvdw-\right. \\
& \left.\frac{1}{H}\int\int\int_{B_h(x,y,t)}K\Big(\frac{u-x}{h_x},\frac{v-y}{h_y}\Big)K\Big(\frac{w-t}{h_t}\Big)dudvdw\right| \\
= & \left|\frac{1}{H}\sum_{ijk}\int\int\int_{\Delta_{ijk}}K\Big(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\Big)K\Big(\frac{t_k-t}{h_t}\Big)dudvdw-\right. \\
& \left.\frac{1}{H}\sum_{ijk}\int\int\int_{B_h(x,y,t)\cap\Delta_{ijk}}K\Big(\frac{u-x}{h_x},\frac{v-y}{h_y}\Big)K\Big(\frac{w-t}{h_t}\Big)dudvdw\right| \\
= & \left|\frac{1}{H}\sum_{ijk}\int\int\int_{B_h(x,y,t)^c\cap\Delta_{ijk}}K\Big(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\Big)K\Big(\frac{t_k-t}{h_t}\Big)dudvdw+\right. \\
& \frac{1}{H}\sum_{ijk}\int\int\int_{B_h(x,y,t)\cap\Delta_{ijk}}K\Big(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\Big)K\Big(\frac{t_k-t}{h_t}\Big)dudvdw- \\
& \left.\frac{1}{H}\sum_{ijk}\int\int\int_{B_h(x,y,t)\cap\Delta_{ijk}}K\Big(\frac{u-x}{h_x},\frac{v-y}{h_y}\Big)K\Big(\frac{w-t}{h_t}\Big)dudvdw\right| \\
\leq & O(\frac{1}{n_{min}h_{min}})+\frac{1}{H}\sum_{ijk}\int\int\int_{B_h(x,y,t)\cap\Delta_{ijk}}\left|K\Big(\frac{x_i-x}{h_x},\frac{y_j-y}{h_y}\Big)K\Big(\frac{t_k-t}{h_t}\Big)-\right. \\
& \left.K\Big(\frac{u-x}{h_x},\frac{v-y}{h_y}\Big)K\Big(\frac{w-t}{h_t}\Big)\right|dudvdw \\
\leq & O(\frac{1}{n_{min}h_{min}})+\frac{1}{H}\sum_{ijk}\int\int\int_{B_h(x,y,t)\cap\Delta_{ijk}}\frac{(1+\sqrt{2})C}{n_{min}h_{min}}dudvdw \\
= & O(\frac{1}{n_{min}h_{min}})+\frac{1}{H}\frac{(1+\sqrt{2})C}{n_{min}h_{min}}\int\int\int_{B_h(x,y,t)}1dudvdw \\
= & O(\frac{1}{n_{min}h_{min}}),
\end{aligned}
$$

where $C \geq 0$ is the Lipschitz constant that satisfies the condition $|K(u)-K(u^{'})| \leq C|u-u^{'}|$. So, the first result in Proposition 1 is valid.

To prove the second result, it can be checked that

$$\begin{aligned}
& E\left|\frac{1}{NH}\sum_{ijk}\varepsilon_{ijk}K\left(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\right)K\left(\frac{t_i-x}{h_t}\right)\right|^2 \\
= \;& Var(\frac{1}{NH}\sum_{ijk}\varepsilon_{ijk}K\left(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\right)K\left(\frac{t_k-x}{h_t}\right)) \\
= \;& \frac{1}{N^2H^2}\sum_{ijk}\sum_{i'j'k'}K\left(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\right)K\left(\frac{t_k-x}{h_t}\right) \\
& K\left(\frac{x_{i'}-x}{h_x},\frac{y_{i'}-y}{h_y}\right)K\left(\frac{t_{k'}-x}{h_t}\right)Cov(\varepsilon_{ijk},\varepsilon_{i'j'k'}) \\
\leq \;& \frac{1}{N^2H^2}\sum_{ijk}\sum_{i'j'k'}K\left(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\right)K\left(\frac{t_k-x}{h_t}\right) \\
& K\left(\frac{x_{i'}-x}{h_x},\frac{y_{i'}-y}{h_y}\right)K\left(\frac{t_{k'}-x}{h_t}\right)c_1\sigma^2\rho^{c_2\max\{|i-i'|,|j-j'|,|k-k'|\}} \\
\leq \;& \frac{1}{N^2H^2}\sum_{ijk}K\left(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\right)K\left(\frac{t_k-x}{h_t}\right)c_1\sigma^2 24\int_0^\infty\tau^2\rho^\tau d\tau \\
= \;& O(\frac{1}{NH}).
\end{aligned}$$

Similarly, it can be checked that

$$\begin{aligned}
& E\left|\frac{1}{NH}\sum_{ijk}(\varepsilon_{ijk}^2-\sigma^2)K\left(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\right)K\left(\frac{t_i-x}{h_t}\right)\right|^2 \\
= \;& Var(\frac{1}{NH}\sum_{ijk}\varepsilon_{ijk}^2K\left(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\right)K\left(\frac{t_k-x}{h_t}\right)) \\
= \;& \frac{1}{N^2H^2}\sum_{ijk}\sum_{i'j'k'}K\left(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\right)K\left(\frac{t_k-x}{h_t}\right) \\
& K\left(\frac{x_{i'}-x}{h_x},\frac{y_{i'}-y}{h_y}\right)K\left(\frac{t_{k'}-x}{h_t}\right)Cov(\varepsilon_{ijk}^2,\varepsilon_{i'j'k'}^2) \\
\leq \;& \frac{1}{N^2H^2}\sum_{ijk}\sum_{i'j'k'}K\left(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\right)K\left(\frac{t_k-x}{h_t}\right) \\
& K\left(\frac{x_{i'}-x}{h_x},\frac{y_{i'}-y}{h_y}\right)K\left(\frac{t_{k'}-x}{h_t}\right)12(c_1\sigma^2\rho^{c_2\max\{|i-i'|,|j-j'|,|k-k'|\}})^{1/4}E(\varepsilon_{111}^4) \\
\leq \;& \frac{1}{N^2H^2}\sum_{ijk}K\left(\frac{x_i-x}{h_x},\frac{y_i-y}{h_y}\right)K\left(\frac{t_k-x}{h_t}\right)12(c_1\sigma^2 24\int_0^\infty\tau^2\rho^\tau d\tau)^{1/3}(E(\varepsilon_{111}^6))^{2/3} \\
= \;& O(\frac{1}{NH}).
\end{aligned}$$

The first inequality in the above expression is based on the result in Davydov (1968). So, the third result is valid.

Appendix A.2 Proof of Theorem 1

We first consider the case when $(x,y;t) \in \Omega_h \setminus J_h$. By the Taylor's expansion, we have

$$\begin{aligned} Z_{ijk} =& f(x_i,y_j;t_k) + \epsilon_{ijk} \\ =& f(x,y;t) + (x_i - x)f_x^{'}(x,y;t) + (y_j - y)f_y^{'}(x,y;t) + (t_k - t)f_t^{'}(x,y;t) + \\ & O(h_x^2 + h_y^2 + h_t^2) + \epsilon_{ijk}. \end{aligned}$$

So, it can be checked that

$$\begin{aligned} \begin{bmatrix} \sum_{ijk} Z_{ijk}K_{ijk} \\ \sum_{ijk}(x_i - x)Z_{ijk}K_{ijk} \\ \sum_{ijk}(y_j - y)Z_{ijk}K_{ijk} \\ \sum_{ijk}(t_k - t)Z_{ijk}K_{ijk} \end{bmatrix} =& M \begin{bmatrix} f(x,y;t) \\ f_x^{'}(x,y;t) \\ f_y^{'}(x,y;t) \\ f_t^{'}(x,y;t) \end{bmatrix} + \begin{bmatrix} \sum_{ijk} O(h_x^2 + h_y^2 + h_t^2)K_{ijk} \\ \sum_{ijk}(x_i - x)O(h_x^2 + h_y^2 + h_t^2)K_{ijk} \\ \sum_{ijk}(y_j - y)O(h_x^2 + h_y^2 + h_t^2)K_{ijk} \\ \sum_{ijk}(t_k - t)O(h_x^2 + h_y^2 + h_t^2)K_{ijk} \end{bmatrix} + \\ & \begin{bmatrix} \sum_{ijk} \epsilon_{ijk}K_{ijk} \\ \sum_{ijk}(x_i - x)\epsilon_{ijk}K_{ijk} \\ \sum_{ijk}(y_j - y)\epsilon_{ijk}K_{ijk} \\ \sum_{ijk}(t_k - t)\epsilon_{ijk}K_{ijk} \end{bmatrix}, \end{aligned}$$

where

$$M = \begin{bmatrix} m_{000} & m_{100} & m_{010} & m_{001} \\ m_{100} & m_{200} & m_{110} & m_{101} \\ m_{010} & m_{110} & m_{020} & m_{011} \\ m_{001} & m_{101} & m_{011} & m_{002} \end{bmatrix}.$$

From Expression (3), we have

$$\begin{aligned} \begin{bmatrix} \widehat{a}(x,y;t) \\ \widehat{b}(x,y;t) \\ \widehat{c}(x,y;t) \\ \widehat{d}(x,y;t) \end{bmatrix} =& \begin{bmatrix} f(x,y;t) \\ f_x^{'}(x,y;t) \\ f_y^{'}(x,y;t) \\ f_t^{'}(x,y;t) \end{bmatrix} + M^{-1} \begin{bmatrix} \sum_{ijk} O(h_x^2 + h_y^2 + h_t^2)K_{ijk} \\ \sum_{ijk}(x_i - x)O(h_x^2 + h_y^2 + h_t^2)K_{ijk} \\ \sum_{ijk}(y_j - y)O(h_x^2 + h_y^2 + h_t^2)K_{ijk} \\ \sum_{ijk}(t_k - t)O(h_x^2 + h_y^2 + h_t^2)K_{ijk} \end{bmatrix} + \\ & M^{-1} \begin{bmatrix} \sum_{ijk} \epsilon_{ijk}K_{ijk} \\ \sum_{ijk}(x_i - x)\epsilon_{ijk}K_{ijk} \\ \sum_{ijk}(y_j - y)\epsilon_{ijk}K_{ijk} \\ \sum_{ijk}(t_k - t)\epsilon_{ijk}K_{ijk} \end{bmatrix}. \end{aligned}$$

By some simple algebraic manipulations, we have

$$M^{-1} = \begin{bmatrix} O(\frac{1}{NH}) & O(\frac{1}{NH \cdot h_x}) & O(\frac{1}{NH \cdot h_y}) & O(\frac{1}{NH \cdot h_t}) \\ O(\frac{1}{NH \cdot h_x}) & O(\frac{1}{NH \cdot h_x^2}) & O(\frac{1}{NH \cdot h_x h_y}) & O(\frac{1}{NH \cdot h_x h_t}) \\ O(\frac{1}{NH \cdot h_y}) & O(\frac{1}{NH \cdot h_x h_y}) & O(\frac{1}{NH \cdot h_y^2}) & O(\frac{1}{NH \cdot h_y h_t}) \\ O(\frac{1}{NH \cdot h_t}) & O(\frac{1}{NH \cdot h_x h_t}) & O(\frac{1}{NH \cdot h_y h_t}) & O(\frac{1}{NH \cdot h_t^2}) \end{bmatrix}.$$

Then,

$$\begin{bmatrix} \widehat{a}(x,y;t) \\ \widehat{b}(x,y;t) \\ \widehat{c}(x,y;t) \\ \widehat{d}(x,y;t) \end{bmatrix} = \begin{bmatrix} f(x,y;t) \\ f_x^{'}(x,y;t) \\ f_y^{'}(x,y;t) \\ f_t^{'}(x,y;t) \end{bmatrix} + \begin{bmatrix} O(h_x^2 + h_y^2 + h_t^2) \\ O(\frac{h_x^2 + h_y^2 + h_t^2}{h_x}) \\ O(\frac{h_x^2 + h_y^2 + h_t^2}{h_y}) \\ O(\frac{h_x^2 + h_y^2 + h_t^2}{h_t}) \end{bmatrix} + \begin{bmatrix} O_p(\frac{1}{\sqrt{NH}}) \\ O_p(\frac{1}{h_x\sqrt{NH}}) \\ O_p(\frac{1}{h_y\sqrt{NH}}) \\ O_p(\frac{1}{h_t\sqrt{NH}}) \end{bmatrix}.$$

Now, we consider the case when $(x,y;t) \in J_h \setminus S_h$. If $(x_i,y_j;t_k) \in I_1$, then we have

$$\begin{aligned} Z_{ijk} &= f(x_i,y_j;t_k) + \varepsilon_{ijk} \\ &= f_-(x_\tau,y_\tau;t_\tau) + O(\sqrt{h_x^2+h_y^2+h_t^2}) + \varepsilon_{ijk}, \end{aligned}$$

and if $(x_i,y_j;t_k) \in I_2$, we have

$$\begin{aligned} Z_{ijk} &= f(x_i,y_j;t_k) + \varepsilon_{ijk} \\ &= f_-(x_\tau,y_\tau;t_\tau) + d_\tau + O(\sqrt{h_x^2+h_y^2+h_t^2}) + \varepsilon_{ijk}. \end{aligned}$$

By some similar arguments to those in the case considered above, we have

$$\begin{aligned} \begin{bmatrix} \widehat{a}(x,y;t) \\ \widehat{b}(x,y;t) \\ \widehat{c}(x,y;t) \\ \widehat{d}(x,y;t) \end{bmatrix} &= \begin{bmatrix} f_-(x_\tau,y_\tau;t_\tau) + d_\tau \frac{\sum_{(x_i,y_j;t_k)\in I_2} K_{ijk}}{\sum_{ijk} K_{ijk}} \\ \frac{d_\tau}{h_x} \frac{\sum_{(x_i,y_j;t_k)\in I_2} [(x_i-x)/h_x] K_{ijk}}{\sum_{ijk} [(x_i-x)/h_x]^2 K_{ijk}} \\ \frac{d_\tau}{h_y} \frac{\sum_{(x_i,y_j;t_k)\in I_2} [(y_j-y)/h_y] K_{ijk}}{\sum_{ijk} [(y_j-y)/h_y]^2 K_{ijk}} \\ \frac{d_\tau}{h_t} \frac{\sum_{(x_i,y_j;t_k)\in I_2} [(t_k-t)/h_t] K_{ijk}}{\sum_{ijk} [(t_k-t)/h_t]^2 K_{ijk}} \end{bmatrix} + \\ & \begin{bmatrix} O(\sqrt{h_x^2+h_y^2+h_t^2}) \\ O(\frac{\sqrt{h_x^2+h_y^2+h_t^2}}{h_x}) \\ O(\frac{\sqrt{h_x^2+h_y^2+h_t^2}}{h_y}) \\ O(\frac{\sqrt{h_x^2+h_y^2+h_t^2}}{h_t}) \end{bmatrix} + \begin{bmatrix} O_p(\frac{1}{\sqrt{NH}}) \\ O_p(\frac{1}{h_x\sqrt{NH}}) \\ O_p(\frac{1}{h_y\sqrt{NH}}) \\ O_p(\frac{1}{h_t\sqrt{NH}}) \end{bmatrix} \\ &= \begin{bmatrix} f_-(x_\tau,y_\tau;t_\tau) + d_\tau \xi_{000}^{(2)} \\ \frac{d_\tau}{\xi_{200} h_x} \xi_{100}^{(2)} \\ \frac{d_\tau}{\xi_{020} h_y} \xi_{010}^{(2)} \\ \frac{d_\tau}{\xi_{002} h_t} \xi_{001}^{(2)} \end{bmatrix} + \begin{bmatrix} O(\sqrt{h_x^2+h_y^2+h_t^2}) \\ O(\frac{\sqrt{h_x^2+h_y^2+h_t^2}}{h_x}) \\ O(\frac{\sqrt{h_x^2+h_y^2+h_t^2}}{h_y}) \\ O(\frac{\sqrt{h_x^2+h_y^2+h_t^2}}{h_t}) \end{bmatrix} + \begin{bmatrix} O_p(\frac{1}{\sqrt{NH}}) \\ O_p(\frac{1}{h_x\sqrt{NH}}) \\ O_p(\frac{1}{h_y\sqrt{NH}}) \\ O_p(\frac{1}{h_t\sqrt{NH}}) \end{bmatrix} \end{aligned}$$

Appendix A.3 Proof of Theorem 2

We prove the second equations in (10) and (11) here. The first equations can be proved similarly. For simplicity, we write $\widehat{a}^{(l)}(x,y;t)$, $\widehat{b}^{(l)}(x,y;t)$, $\widehat{c}^{(l)}(x,y;t)$, $\widehat{d}^{(l)}(x,y;t)$, $O^{(l)}(x,y;t)$ and $\widetilde{O}^{(l)}(x,y;t)$ as $\widehat{a}^{(l)}$, $\widehat{b}^{(l)}$, $\widehat{c}^{(l)}$, $\widehat{d}^{(l)}$, $O^{(l)}$ and $\widetilde{O}^{(l)}$respectively from now. First, by Proposition 1, it is easy to show that

$$\frac{\sum_{ijk} \varepsilon_{ijk} K\left(\frac{x_i-x}{h_x}, \frac{y_i-y}{h_y}\right) K\left(\frac{t_i-x}{h_t}\right)}{\sum_{ijk} K\left(\frac{x_i-x}{h_x}, \frac{y_i-y}{h_y}\right) K\left(\frac{t_i-x}{h_t}\right)} = O_p(\frac{1}{\sqrt{NH}}), \tag{A1}$$

$$\frac{\sum_{ijk} (\varepsilon_{ijk}^2 - \sigma^2) K\left(\frac{x_i-x}{h_x}, \frac{y_i-y}{h_y}\right) K\left(\frac{t_i-x}{h_t}\right)}{\sum_{ijk} K\left(\frac{x_i-x}{h_x}, \frac{y_i-y}{h_y}\right) K\left(\frac{t_i-x}{h_t}\right)} = o_p(1). \tag{A2}$$

Let us first consider the case when $(x,y;t) \in \Omega_h \setminus J_h$. In such a case, it can be checked that

$$\begin{aligned}
e^{(l)}(x,y;t) &= \Big\{ \sum_{(x_i,y_j;t_k)\in O^{(l)}} [\varepsilon_{ijk} + f(x_i,y_j;t_k) - \widehat{a}^{(l)} - \widehat{b}^{(l)}(x_i - x) - \\
&\quad \widehat{c}^{(l)}(y_j - y) - \widehat{d}^{(l)}(t_k - t)]^2 K_{ijk} \Big\} / \sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk} \\
&= \Big\{ \sum_{(x_i,y_j;t_k)\in O^{(l)}} \varepsilon_{ijk}^2 K_{ijk} \Big\} / \sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk} + \\
&\quad \Big\{ 2 \sum_{(x_i,y_j;t_k)\in O^{(l)}} \varepsilon_{ijk} [f(x_i,y_j;t_k) - \widehat{a}^{(l)} - \widehat{b}^{(l)}(x_i - x) - \\
&\quad \widehat{c}^{(l)}(y_j - y) - \widehat{d}^{(l)}(t_k - t)] K_{ijk} \Big\} / \sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk} + \\
&\quad \Big\{ \sum_{(x_i,y_j;t_k)\in O^{(l)}} [f(x_i,y_j;t_k) - \widehat{a}^{(l)} - \widehat{b}^{(l)}(x_i - x) - \\
&\quad \widehat{c}^{(l)}(y_j - y) - \widehat{d}^{(l)}(t_k - t)]^2 K_{ijk} \Big\} / \sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk} \\
&=: A_1^{(l)}(x,y;t) + A_2^{(l)}(x,y;t) + A_3^{(l)}(x,y;t).
\end{aligned}$$

Similar to (A2), we have

$$A_1^{(l)}(x,y;t) = \sigma^2 + o_p(1). \tag{A3}$$

By the Taylor's expansion of $f(x_i,y_j;t_k)$ at point $(x,y;t)$, results in Theorem 1, and similar arguments for (A1), we have

$$\begin{aligned}
A_2^{(l)}(x,y;t) &\le 2|f(x,y;t) - \widehat{a}^{(l)}| \left| \frac{\sum_{(x_i,y_j;t_k)\in O^{(l)}} \varepsilon_{ijk} K_{ijk}}{\sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk}} \right| + \\
&\quad 2h_x |f_x'(x,y;t) - \widehat{b}^{(l)}| \left| \frac{\sum_{(x_i,y_j;t_k)\in O^{(l)}} \varepsilon_{ijk} \frac{x_i - x}{h_x} K_{ijk}}{\sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk}} \right| + \\
&\quad 2h_y |f_y'(x,y;t) - \widehat{c}^{(l)}| \left| \frac{\sum_{(x_i,y_j;t_k)\in O^{(l)}(x,y;t)} \varepsilon_{ijk} \frac{y_j - y}{h_y} K_{ijk}}{\sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk}} \right| + \\
&\quad 2h_t |f_t'(x,y;t) - \widehat{d}^{(l)}| \left| \frac{\sum_{(x_i,y_j;t_k)\in O^{(l)}} \varepsilon_{ijk} \frac{t_k - t}{h_t} K_{ijk}}{\sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk}} \right| \\
&= o_p(1).
\end{aligned} \tag{A4}$$

Similarly, we have

$$A_3^{(l)}(x,y;t) = o_p(1). \tag{A5}$$

By combining (A3)-(A5), we have

$$e^{(l)}(x,y;t) = \sigma^2 + o_p(1).$$

Now, let us consider the case when $(x,y;t) \in J_h \setminus S_h$. Similar to the above case, let us write

$$e^{(l)}(x,y;t) = A_1^{(l)}(x,y;t) + A_2^{(l)}(x,y;t) + A_3^{(l)}(x,y;t).$$

Here, we still have

$$A_1^{(l)}(x,y;t) = \sigma^2 + o_p(1). \tag{A6}$$

For $A_2^{(l)}(x,y;t)$, we have

$$\begin{aligned} A_2^{(l)}(x,y;t) &= \Big\{2 \sum_{(x_i,y_j;t_k)\in I^1\cap O^{(l)}} \varepsilon_{ijk}[f(x_i,y_j;t_k) - \widehat{a}^{(l)} - \widehat{b}^{(l)}(x_i-x) - \\ &\quad \widehat{c}^{(l)}(y_j-y) - \widehat{d}^{(l)}(t_k-t)]K_{ijk}\Big\} / \sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk} + \\ &\quad \Big\{2 \sum_{(x_i,y_j;t_k)\in I^2\cap O^{(l)}} \varepsilon_{ijk}[f(x_i,y_j;t_k) - \widehat{a}^{(l)} - \widehat{b}^{(l)}(x_i-x) - \\ &\quad \widehat{c}^{(l)}(y_j-y) - \widehat{d}^{(l)}(t_k-t)]K_{ijk}\Big\} / \sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk} \\ &=: A_{21}^{(l)}(x,y;t) + A_{22}^{(l)}(x,y;t). \end{aligned}$$

By the results in Theorem 1, we have

$$\begin{aligned} A_{21}^{(l)}(x,y;t) &= \frac{2\sum_{(x_i,y_j;t_k)\in I^1\cap O^{(l)}} \varepsilon_{ijk}\big[f(x_i,y_j;t_k) - f_-(x_\tau,y_\tau;t_\tau)\big]K_{ijk}}{\sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk}} - \\ &\quad \frac{(D_1+o_p(1))\sum_{(x_i,y_j;t_k)\in I^1\cap O^{(l)}} \varepsilon_{ijk}K_{ijk}}{\sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk}} - \\ &\quad \frac{(D_2+o_p(1))\sum_{(x_i,y_j;t_k)\in I^1\cap O^{(l)}} \varepsilon_{ijk}\frac{x_i-x}{h_x}K_{ijk}}{\sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk}} - \\ &\quad \frac{(D_3+o_p(1))\sum_{(x_i,y_j;t_k)\in I^1\cap O^{(l)}} \varepsilon_{ijk}\frac{y_j-y}{h_y}K_{ijk}}{\sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk}} - \\ &\quad \frac{(D_4+o_p(1))\sum_{(x_i,y_j;t_k)\in I^1\cap O^{(l)}} \varepsilon_{ijk}\frac{t_k-t}{h_t}K_{ijk}}{\sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk}}, \end{aligned}$$

where D_1, D_2, D_3 and D_4 are constants. By similar arguments for (A1), we can conclude that

$$A_{21}^{(l)} = o_p(1).$$

Similarly, we have

$$A_{22}^{(l)} = o_p(1).$$

So,

$$A_2^{(l)} = o_p(1). \tag{A7}$$

By similar arguments to those about Proposition 1, we have

$$\left|\frac{1}{NH}\sum_{(x_i,y_j;t_k)\in O^{(l)}} K_{ijk} - \frac{1}{2}\right| = o(1).$$

For a function $\phi(x,y;t)$ satisfying the condition that $\sup_{x^2+y^2+t^2\leq 1}|\phi(x,y;t)| \leq b_\phi < \infty$, we can have

$$\begin{aligned}
& \Big| \frac{1}{NH} \sum_{(x_i,y_j;t_k)\in I^1 \bigcap O^{(l)}} \phi(\frac{x_i-x}{h_x}, \frac{y_j-y}{h_y}; \frac{t_k-t}{h_t}) K_{ijk} - \\
& \frac{1}{NH} \sum_{(x_i,y_j;t_k)\in I^1 \bigcap \widetilde{O}^{(l)}} \phi(\frac{x_i-x}{h_x}, \frac{y_j-y}{h_y}; \frac{t_k-t}{h_t}) K_{ijk} \Big| \\
\leq \; & b_\phi ||K|| \frac{1}{NH} \sum_{(x_i,y_j;t_k)\in O^{(l)} \Delta \widetilde{O}^{(l)}} 1 \\
= \; & o(1),
\end{aligned}$$

where $O^{(l)}\Delta\widetilde{O}^{(l)} = (O^{(l)} \bigcup \widetilde{O}^{(l)}) \setminus (O^{(l)} \bigcap \widetilde{O}^{(l)})$. The last equation above is a direct conclusion of (9). By the above results, we have

$$
\begin{aligned}
A_3^{(l)}(x,y;t) &= \frac{2}{NH}\sum_{(x_i,y_j;t_k)\in O^{(l)}}\Big[f(x_i,y_j;t_k) - \widehat{a}^{(l)} - \widehat{b}^{(l)}(x_i-x) - \\
&\quad \widehat{c}^{(l)}(y_j-y) - \widehat{d}^{(l)}(t_k-t)\Big]^2 K_{ijk} \\
&= \frac{2}{NH}\sum_{(x_i,y_j;t_k)\in O^{(l)}}\Big[f(x_i,y_j;t_k) - f_-(x_\tau,y_\tau;t_\tau) - d_\tau B_{0l} - \frac{d_\tau B_{1l}}{\xi_{200}}\frac{x_i-x}{h_x} - \\
&\quad \frac{d_\tau B_{2l}}{\xi_{020}}\frac{y_j-y}{h_y} - \frac{d_\tau B_{3l}}{\xi_{002}}\frac{t_k-t}{h_t}\Big]^2 K_{ijk} + o_p(1) \\
&= \frac{2}{NH}\Big(\sum_{(x_i,y_j;t_k)\in I^1\cap O^{(l)}} + \sum_{(x_i,y_j;t_k)\in I^2\cap O^{(l)}}\Big) \\
&\quad \Big[f(x_i,y_j;t_k) - f_-(x_\tau,y_\tau;t_\tau) - d_\tau B_{0l} - \frac{d_\tau B_{1l}}{\xi_{200}}\frac{x_i-x}{h_x} - \\
&\quad \frac{d_\tau B_{2l}}{\xi_{020}}\frac{y_j-y}{h_y} - \frac{d_\tau B_{3l}}{\xi_{002}}\frac{t_k-t}{h_t}\Big]^2 K_{ijk} + o_p(1) \\
&= \frac{2}{NH}\Big(\sum_{(x_i,y_j;t_k)\in I^1\cap \widetilde{O}^{(l)}} + \sum_{(x_i,y_j;t_k)\in I^2\cap \widetilde{O}^{(l)}}\Big) \\
&\quad \Big[f(x_i,y_j;t_k) - f_-(x_\tau,y_\tau;t_\tau) - d_\tau B_{0l} - \frac{d_\tau B_{1l}}{\xi_{200}}\frac{x_i-x}{h_x} - \\
&\quad \frac{d_\tau B_{2l}}{\xi_{020}}\frac{y_j-y}{h_y} - \frac{d_\tau B_{3l}}{\xi_{002}}\frac{t_k-t}{h_t}\Big]^2 K_{ijk} + o_p(1) \\
&= \frac{2}{NH}\sum_{(x_i,y_j;t_k)\in I^1\cap \widetilde{O}^{(l)}}\Big[- d_\tau B_{0l} - \frac{d_\tau B_{1l}}{\xi_{200}}\frac{x_i-x}{h_x} - \\
&\quad \frac{d_\tau B_{2l}}{\xi_{020}}\frac{y_j-y}{h_y} - \frac{d_\tau B_{3l}}{\xi_{002}}\frac{t_k-t}{h_t}\Big]^2 K_{ijk} + \\
&\quad \frac{2}{NH}\sum_{(x_i,y_j;t_k)\in I^2\cap \widetilde{O}^{(l)}}\Big[d_\tau - d_\tau B_{0l} - \frac{d_\tau B_{1l}}{\xi_{200}}\frac{x_i-x}{h_x} - \\
&\quad \frac{d_\tau B_{2l}}{\xi_{020}}\frac{y_j-y}{h_y} - \frac{d_\tau B_{3l}}{\xi_{002}}\frac{t_k-t}{h_t}\Big]^2 K_{ijk} + o_p(1) \\
&= 2d_\tau^2\int\int\int_{Q^{(1l)}}\Big[B_{0l} + \frac{B_{1l}}{\xi_{200}}u + \frac{B_{2l}}{\xi_{020}}v + \frac{B_{3l}}{\xi_{002}}w\Big]^2 K(u,v)K(w)dudvdw + \\
&\quad 2d_\tau^2\int\int\int_{Q^{(2l)}}\Big[1 - B_{0l} - \frac{B_{1l}}{\xi_{200}}u - \frac{B_{2l}}{\xi_{020}}v - \frac{B_{3l}}{\xi_{002}}w\Big]^2 K(u,v)K(w)dudvdw \\
&\quad + o_p(1) \\
&= d_\tau^2(C_\tau^{(l)})^2 + o_p(1),
\end{aligned}
\tag{A8}
$$

where

$$C_\tau^{(l)} = \left(2\int\int\int_{Q^{(1l)}}\left[B_{0l}+\frac{B_{1l}}{\xi_{200}}u+\frac{B_{2l}}{\xi_{020}}v+\frac{B_{3l}}{\xi_{002}}w\right]^2 K(u,v)K(w)dudvdw+ 2\int\int\int_{Q^{(2l)}}\left[1-B_{0l}-\frac{B_{1l}}{\xi_{200}}u-\frac{B_{2l}}{\xi_{020}}v-\frac{B_{3l}}{\xi_{002}}w\right]^2 K(u,v)K(w)dudvdw\right)^{1/2}.$$

Then by equation (A6)-(A8), we have

$$e^{(l)}(x,y;t) = \sigma^2 + d_\tau^2 (C_\tau^{(l)})^2 + o_p(1).$$

Similarly, we can prove that

$$e(x,y;t) = \sigma^2 + d_\tau^2 (C_\tau)^2 + o_p(1),$$

where

$$C_\tau = \left(\int\int\int_{Q^{(1)}}\left[\xi_{000}^{(2)}+\frac{\xi_{100}^{(2)}}{\xi_{200}}u+\frac{\xi_{010}^{(2)}}{\xi_{020}}v+\frac{\xi_{001}^{(2)}}{\xi_{002}}w\right]^2 K(u,v)K(w)dudvdw+ \int\int\int_{Q^{(2)}}\left[1-\xi_{000}^{(2)}-\frac{\xi_{100}^{(2)}}{\xi_{200}}u-\frac{\xi_{010}^{(2)}}{\xi_{020}}v-\frac{\xi_{001}^{(2)}}{\xi_{002}}w\right]^2 K(u,v)K(w)dudvdw\right)^{1/2}.$$

The main difference between this case and the previous case in the proof is in the derivation of the result of (19). For $e(x,y;t)$, the corresponding result is

$$\begin{aligned}
A_3(x,y;t) &= \frac{1}{NH}\sum_{(x_i,y_j;t_k)}\Big[f(x_i,y_j;t_k)-\widehat{a}(x,y;t)-\widehat{b}(x,y;t)(x_i-x)- \\
&\quad \widehat{c}(x,y;t)(y_j-y)-\widehat{d}(x,y;t)(t_k-t)\Big]^2K_{ijk} \\
&= \frac{1}{NH}\sum_{(x_i,y_j;t_k)}\Big[f(x_i,y_j;t_k)-f_-(x_\tau,y_\tau;t_\tau)-d_\tau\xi^{(2)}_{000}-\frac{d_\tau\xi^{(2)}_{100}}{\xi_{200}}\frac{x_i-x}{h_x}- \\
&\quad \frac{d_\tau\xi^{(2)}_{010}}{\xi_{020}}\frac{y_j-y}{h_y}-\frac{d_\tau\xi^{(2)}_{001}}{\xi_{002}}\frac{t_k-t}{h_t}\Big]^2K_{ijk}+o_p(1) \\
&= \frac{1}{NH}\Big(\sum_{(x_i,y_j;t_k)\in I^1}+\sum_{(x_i,y_j;t_k)\in I^2}\Big) \\
&\quad \Big[f(x_i,y_j;t_k)-f_-(x_\tau,y_\tau;t_\tau)-d_\tau\xi^{(2)}_{000}-\frac{d_\tau\xi^{(2)}_{100}}{\xi_{200}}\frac{x_i-x}{h_x}- \\
&\quad \frac{d_\tau\xi^{(2)}_{010}}{\xi_{020}}\frac{y_j-y}{h_y}-\frac{d_\tau\xi^{(2)}_{001}}{\xi_{002}}\frac{t_k-t}{h_t}\Big]^2K_{ijk}+o_p(1) \\
&= \frac{1}{NH}\sum_{(x_i,y_j;t_k)\in I^1}\Big[-d_\tau\xi^{(2)}_{000}-\frac{d_\tau\xi^{(2)}_{100}}{\xi_{200}}\frac{x_i-x}{h_x}- \\
&\quad \frac{d_\tau\xi^{(2)}_{010}}{\xi_{020}}\frac{y_j-y}{h_y}-\frac{d_\tau\xi^{(2)}_{001}}{\xi_{002}}\frac{t_k-t}{h_t}\Big]^2K_{ijk}+ \\
&\quad \frac{1}{NH}\sum_{(x_i,y_j;t_k)\in I^2}\Big[d_\tau-d_\tau\xi^{(2)}_{000}-\frac{d_\tau\xi^{(2)}_{100}}{\xi_{200}}\frac{x_i-x}{h_x}- \\
&\quad \frac{d_\tau\xi^{(2)}_{010}}{\xi_{020}}\frac{y_j-y}{h_y}-\frac{d_\tau\xi^{(2)}_{001}}{\xi_{002}}\frac{t_k-t}{h_t}\Big]^2K_{ijk}+o_p(1) \\
&= d_\tau^2\int\int\int_{Q^{(1)}}\Big[\xi^{(2)}_{000}+\frac{\xi^{(2)}_{100}}{\xi_{200}}u+\frac{\xi^{(2)}_{010}}{\xi_{020}}v+\frac{\xi^{(2)}_{001}}{\xi_{002}}w\Big]^2K(u,v)K(w)dudvdw+ \\
&\quad d_\tau^2\int\int\int_{Q^{(2)}}\Big[1-\xi^{(2)}_{000}-\frac{\xi^{(2)}_{100}}{\xi_{200}}u-\frac{\xi^{(2)}_{010}}{\xi_{020}}v-\frac{\xi^{(2)}_{001}}{\xi_{002}}w\Big]^2K(u,v)K(w)dudvdw \\
&\quad +o_p(1) \\
&= d_\tau^2(C_\tau)^2+o_p(1).
\end{aligned}$$

Appendix A.4 Proof of Theorem 3

For the case when $(x,y;t)\in\Omega_h\setminus J_h$, the estimator $\widehat{f}(x,y;t)$ is one of $\widehat{a}(x,y;t)$, $\widehat{a}^{(1)}(x,y;t)$, $\widehat{a}^{(2)}(x,y;t)$ and $(\widehat{a}^{(1)}(x,y;t)+\widehat{a}^{(2)}(x,y;t))/2$, all of which are consistent estimators of $f(x,y;t)$. So, we have the result in the theorem.

For the case when $(x,y;t)\in J_h\setminus S_h$, it is easy to see that we have either i) $e(x,y;t)=\sigma^2+d_\tau^2(C_\tau)^2+o_p(1)$, $e^{(1)}(x,y;t)=\sigma^2+o_p(1)$, and $e^{(2)}(x,y;t)=\sigma^2+d_\tau^2(C_\tau^{(2)})^2+o_p(1)$, or ii) $e(x,y;t)=\sigma^2+d_\tau^2(C_\tau)^2+o_p(1)$, $e^{(1)}(x,y;t)=\sigma^2+d_\tau^2(C_\tau^{(1)})^2+o_p(1)$, and $e^{(2)}(x,y;t)=\sigma^2+o_p(1)$. In both cases, we have $D(x,y;t)=d_\tau^2(C_\tau)^2+o_p(1)$. Therefore, asymptotically $D(x,y;t)>u$. Since $e^{(1)}(x,y;t)<e^{(2)}(x,y;t)$ in i), the estimator $\widehat{f}(x,y;t)$ is $\widehat{a}^{(1)}(x,y;t)$ in this case, which is a consistent estimator of $f(x,y;t)$. A similar result follows in the case ii).