Monitoring of significant changes over time in fluorescence microscopy imaging of living cells

Kathrin Bissantz,
Technische Universität Dortmund
Fakultät für Statistik
44221 Dortmund, Germany

Nicolai Bissantz,
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany

Katharina Proksch
Universität Siegen
Fachbereich Mathematik
57068 Siegen, Germany

Abstract

The question whether structural changes in time-resolved images are of statistical significance, and therefore of scientific interest, or merely emerge from random noise is of great relevance in many practical applications such as live cell fluorescence microscopy, where intracellular diffusion processes are investigated.

In this paper the statistical recovery of such time-resolved images from fluorescence microscopy of living cells is discussed, based on which a bootstrap method is introduced that allows to both monitor and visualize statistically significant structural changes between individual frames over time. The method can be adopted for use in other imaging systems. It yields a criterion to assess time-resolved small scale structural changes e. g. in the nanometer range.

The proposed bootstrap method is based on data reconstruction with a regularization technique as well as new theoretical results on uniform confidence bands for the function of interest in a two-dimensional heteroscedastic nonparametric convolution-type inverse regression model of Poisson-type.

Moreover, a data-driven selection method for the regularization parameter based on statistical multiscale methods is discussed. The method can be used for an automatic, data-driven data analysis.

The theoretical results are demonstrated in a simulation study and are used to analyze data of fluorescently labeled intracellular transport compartments in living cells.

Key words: bootstrap, confidence bands, deconvolution, fluorescence microscopy, live cell microscopic imaging,

1 Introduction

In many applications, data show a dynamic behavior, that is, making observations of the same objects at different times will give different data sets not only due to random noise but also due to systematic changes. The question of interest then is which changes are significant and which ones are not. As a particular example we consider microscopic live cell imaging in biology in which a sequence of images is taken over time. Figure 1 shows such a sequence of images taken over a total period of 44.844 seconds with equidistant time steps.

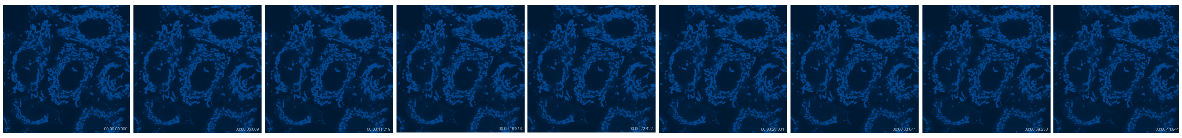

Figure 1: Sequence of images of living HeLa cells in which membrane compartments inside the cell are stained with a fluorescent dye, taken over a total period of 44.844 seconds with equidistant time steps.

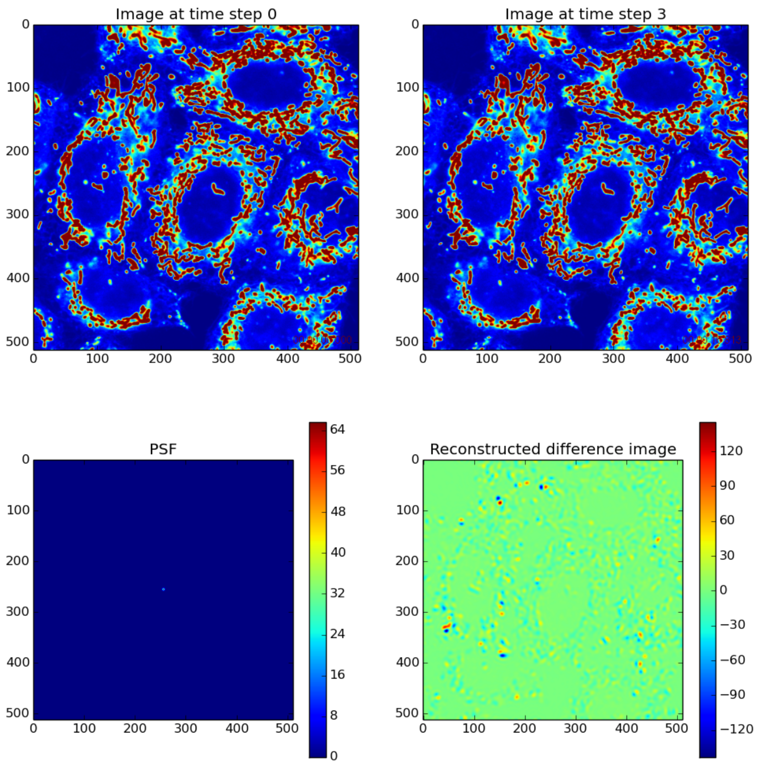

Figure 2: Exemplary demonstration of the determination of regions of significant difference between the images at different time steps. Clockwise from upper left: Image at time steps 0 and 3, estimated PSF and reconstruction of a difference image. The picture shows living HeLa cells in which membrane compartments inside the cell are stained with a fluorescent dye.

A simple visual inspection of the sequence of images in Figure 1 does not reveal much of the evolution over time. By taking differences of images at different time steps, changes can be emphasized (see for instance Figure 2 where the difference image of the first and the fourth image of the above sequence is displayed). This hints at several differences which are in the focus of our investigations. The example of microscopic imaging bears additional complications, though, as it is intrinsically multivariate and it is actually an ill-posed inverse problem. It is a well-known fact (see for example Adorf (1995), Wallace et al. (2001) or Bertero et al. (2009)) that unprocessed images, taken with imaging devices such as optical microscopes or telescopes, are blurry which is due to the physical characteristics of the propagation of light at surfaces of mirrors and lenses. The process of optical distortion can mathematically be modeled as convolution of the "true image" with a so-called point-spread function ψ (PSF). This means that in this case we only have empirical access to

$$T_\psi f = f * \psi,$$

where f is the true image, which the object of interest, ψ denotes the point-spread function, the operation $*$ denotes convolution and T_ψ is the associated (linear) convolution operator. In many other inverse problems the connection between the quantity of interest and the observed one can be expressed in terms of a linear operator equation as well. Well-known examples are Positron Emission Tomography, which involves the Radon transform, (Cavalier (2000)), the heat equation (Mair and Ruymgaart (1996)) or the Laplace transform (Saitoh (1997)). In all these examples the first step in data analysis is the recovery of the quantity of interest which involves the approximate inversion of the operator considered.

In the context of imaging we often face a further characteristic property as it is composed of two parts: The optical device (such as the microscope or the telescope) and a detector (for example a CCD). While the inverse problem character is is due to the properties of the optical components, the detectors usually entail peculiarities as well and provide count rates rather than continuously distributed data, a fact that is often ignored. In this paper we take all properties discussed above into account and consider a nonparametric inverse regression model of Poisson-type:

$$Y \sim \text{Poisson}((f * \psi)(x_1, x_2)). \tag{1}$$

In this application the variable $\mathbf{x} = (x_1, x_2)$ represents a pixel of a CCD and we can only observe a blurred version of the true image modeled by the signal f with Poisson noise. In contrast to other authors (see, e. g., Cavalier and Tsybakov (2002)) we do not assume that the function ψ in model (1) is periodic, because in the reconstruction of astronomical or biological images from telescopes or microscopic imaging devices this assumption is often unrealistic.

The purpose of the present paper is to suggest a procedure for the reconstruction of images observed according to model (1) with a particular focus on determining regions of significant change between images at subsequent time steps as, for instance, in live-cell imaging. To this end we adapt a multiresolution method, presented in Bissantz et al. (2008) in a one-dimensional framework, to the reconstruction of two-dimensional images from fluorescence microscopy. Furthermore, cf. Hotz et al. (2012) for a related approach to image denoising. Furthermore, we extend results from Proksch et al. (2014), who consider the construction of uniform confidence surfaces in multivariate inverse regression models with convolution operator with homoscedastic, additive errors, to the Poisson model (1). Based on these asymptotic results, bootstrap confidence surfaces are derived, which provide both a graphical tool but also a rigorous statistical testing procedure to a given significance level for our problem of determining regions of significant change.

We apply our method to data from live-cell imaging of cultured mammalian cells. The HeLa cells were stained with the fluorescent dye DiI which labels lipid membranes inside the cell that belong to intracellular transport compartments. It is therefore used to monitor intracellular transport processes in living mammalian cells by fluorescence microscopy (see *The Molecular Probes Handbook*).

The organization of this paper is as follows. In Section 2 we introduce the mathematical methods used for data analysis. This includes image reconstruction, asymptotic and bootstrap-based confidence bands and a multiresolution based approach to the selection of the smoothing parameter required for the image reconstruction. All technical details are given in Section 6 and all proofs are deferred to an appendix. All other sections are devoted to data analysis. First we summarize the data to which our methods are applied and give a description of the pipeline for data reduction in Section 3. Finally, in Sections 4 and 5 we discuss the results of our data analysis and its implications for live cell imaging.

2 Statistical modeling and mathematical preliminaries

Suppose that we have available a sequence of images of some object evolving over time, e. g., in live cell imaging as discussed in the introduction. Assume that for a fixed time we observe

$$Y_{i,j} \sim \text{Poisson}((f * \psi)(x_i, x_j)) = \text{Poisson}(g(x_i, x_j)), \quad (i,j) \in \{-n, \dots, n\}^2, \tag{2}$$

where the design points $(x_i, x_j) = (i/na_n, j/na_n)$ correspond to a value of a CCD at pixel (i,j) of the image and the numbering is such that the center of the image is at $(0,0)$. The sequence $(a_n)_{n\in\mathbb{N}}$ is a sequence of design parameters satisfying the condition $a_n \to 0$ as well as $na_n \to \infty$ as $n \to \infty$. Recall that we are interested in the function f itself and not in the convolution of f and ψ. Throughout this paper the point-spread function $\psi(\cdot,\cdot)$ is assumed to be known. For a discussion of this assumption and the choice of the PSF in practical applications we refer to Section 3.

In the following $\boldsymbol{\alpha} \in \mathbb{N}^2$ denotes a double-index, the bold type letters $\mathbf{x}, \mathbf{y}, \mathbf{z}$ and $\boldsymbol{\xi}$ denote elements of $\mathbb{R}^2$ and $\mathbf{x}_{i,j} = (x_i, x_j)^T$. Further, with a slight abuse of notation we shall denote the vector $(\frac{y_1 - x_i}{h}, \frac{y_2 - x_j}{h})^T$ by $\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}$ for a scalar $h \in \mathbb{R}\backslash\{0\}$ for the sake of a clearer display of the results and the proofs.

Throughout this paper, some conventional multi-index notation is used. For a double-index $\boldsymbol{\alpha} \in \mathbb{N}^2$ we denote

$$\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdot x_2^{\alpha_2}, \quad |\boldsymbol{\alpha}| = \alpha_1 + \alpha_2, \quad \boldsymbol{\alpha}! = \alpha_1! \cdot \alpha_2!$$

and

$$\boldsymbol{\alpha} \leq (i,j) \quad \text{if} \quad \alpha_1 \leq i \quad \text{and} \quad \alpha_2 \leq j.$$

2.1 Estimation

Given Poisson model (2), the first goal is to define a suitable estimator for the signal (image) f. To this end, let $\mathcal{F}f$ define the Fourier transform of f. As a consequence of the convolution theorem and the formula for Fourier inversion we obtain the representation

$$f(\mathbf{x}) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \frac{\mathcal{F}g(\boldsymbol{\xi})}{\mathcal{F}\psi(\boldsymbol{\xi})} \exp(i\boldsymbol{\xi}^T \mathbf{x})\, d\boldsymbol{\xi}. \tag{3}$$

An estimator for the regression function f can now easily be obtained from the data by replacing the unknown quantity $\mathcal{F}g = \mathcal{F}(f * \psi)$ by an estimator $\widehat{\mathcal{F}g}$. The random fluctuations in the estimator $\widehat{\mathcal{F}g}$ cause instability of the ratio $\frac{\widehat{\mathcal{F}g}(\boldsymbol{\xi})}{\mathcal{F}\psi(\boldsymbol{\xi})}$ if at least one of the components of $\boldsymbol{\xi}$ is large. As a consequence, the problem at hand is ill-posed and requires regularization. We address this issue by suppressing large values of ξ_j for $j = 1, 2$ from the domain of integration, i.e. we multiply the integrand in (3) with a sequence of smooth Fourier transforms $\mathcal{F}\eta(h\cdot)$ with compact support $[-h^{-1}, h^{-1}]^2$. Here, $h = h_n$ is a regularization parameter which corresponds to a bandwidth in non-parametric curve estimation and satisfies $h \to 0$ if $n \to \infty$. The choice of the regularization h is discussed in Section 2.6 below. For the exact properties and for possible choices of the function η we refer to Assumption 2 below and Section 3.2, respectively.
An estimator $\hat{f}_n$ for the signal f in model (2) is now easily obtained as

$$\hat{f}_n(\mathbf{x}) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \frac{\widehat{\mathcal{F}g}(\boldsymbol{\xi})}{\mathcal{F}\psi(\boldsymbol{\xi})} \exp(i\boldsymbol{\xi}^T\mathbf{x}) \mathcal{F}\eta(h\boldsymbol{\xi})\, d\boldsymbol{\xi}, \tag{4}$$

where

$$\widehat{\mathcal{F}g}(\boldsymbol{\xi}) = \frac{1}{2\pi n^2 a_n^2} \sum_{(i,j)\in\{-n,\ldots,n\}^2} Y_{i,j} \exp(-i(\xi_1 x_i + \xi_2 x_j))$$

is the empirical analogue of the Fourier transform of g. Note that with the definition of the kernel

$$K_n(\mathbf{x}) = \frac{1}{2\pi} \int_{\mathbb{R}^2} \frac{\mathcal{F}\eta(\boldsymbol{\xi})}{\mathcal{F}\psi(\frac{\boldsymbol{\xi}}{h})} \exp(i\boldsymbol{\xi}^T\mathbf{x})\, d\boldsymbol{\xi}, \tag{5}$$

the estimator (4) can be written in the following form:

$$\hat{f}_n(y_1, y_2) = \frac{1}{(2\pi)^2 n^d a_n^2 h^2} \sum_{(i,j)\in\{-n,\ldots,n\}^2} Y_{i,j} K_n\left(\frac{y_1 - x_i}{h}, \frac{y_2 - x_j}{h}\right). \tag{6}$$

2.2 Assumptions

We now state the assumptions that are used for our theoretical considerations.

Assumption 1. Assume that there exist constants $S_f > 0$ and $S_{\mathcal{F}g} > 0$ such that

(i) $\displaystyle\int_{\mathbb{R}^2} |\boldsymbol{\xi}^{\boldsymbol{\alpha}}|\, |\mathcal{F}f(\boldsymbol{\xi})|\, d\boldsymbol{\xi} < \infty$ for all $|\boldsymbol{\alpha}| \leq S_f$.

(ii) $\displaystyle\int_{\mathbb{R}^2} |\mathbf{z}^{\boldsymbol{\alpha}}|\, |g(\mathbf{z})|\, d\mathbf{z} < \infty$ for all $|\boldsymbol{\alpha}| \leq S_{\mathcal{F}g}$, where $g = f * \psi$.

(iii) The function g is bounded away from zero, and the functions $\partial^{\boldsymbol{\alpha}} g$, $|\boldsymbol{\alpha}| \leq 1$ are bounded, that is, there exist positive constants g^* and G^* such that $0 < g^* \leq g(\mathbf{z}) \leq G^*$ and $|\partial^{\boldsymbol{\alpha}} g(\mathbf{z})| \leq G^*$ for all $|\boldsymbol{\alpha}| \leq 1$, $\mathbf{z} \in \mathbb{R}^2$.
Furthermore, $\partial^{\boldsymbol{\alpha}} g$ is Lipschitz-continuous for $|\boldsymbol{\alpha}| = 1$.

Remark 1. The constants S_f and $S_{\mathcal{F}g}$ in Assumption 1 quantify the degree of smoothness of the functions f and $\mathcal{F}g$, respectively, by the connection between the decay of the respective Fourier transforms and the derivatives of the transformed functions. If Assumption 1 holds, this implies that the functions f and $\mathcal{F}g$ are S_f-times, respectively $S_{\mathcal{F}g}$-times, continuously differentiable.

Assumption 2. The Fourier transform $\mathcal{F}\eta$ of η is symmetric, supported on $[-1,1]^2$, $\mathcal{F}\eta(\boldsymbol{\xi}) \equiv 1$ for all $\boldsymbol{\xi} \in D := [-\delta,\delta]^2$ for some $\delta > 0$ and $|\mathcal{F}\eta| \leq 1$. Further, there exists a constant $S_{\mathcal{F}\eta} \geq 3$ such that all partial derivatives of order up to $S_{\mathcal{F}\eta}$ exist and are continuous.

Remark 2. A straightforward choice for a regularization function $\mathcal{F}\eta$ is

$$\mathcal{F}\eta(\xi_1,\xi_2) = \begin{cases} \exp\left(-\frac{1}{1-|\boldsymbol{\xi}|}+1\right) & |\boldsymbol{\xi}| < 1 \\ 0 & |\boldsymbol{\xi}| \geq 1 \end{cases} \tag{7}$$

where $|\cdot|$ denotes the Euclidean distance. Other widely used choices for $\mathcal{F}\eta$ include, for instance, the family of functions $\mathcal{F}\eta_\gamma = (1-|\cdot|^2)^\gamma I_{[0,1]}(|\cdot|)$ where the degree of smoothness increases monotonically with the parameter γ. The choice of $\mathcal{F}\eta = I_{[-1,1]^2}$ corresponds to the popular spectral cut-off spectral regularization method. Our approach is similar as we also cut off high frequencies but the smoothness of the function (7) at the boundary of its support results in theoretical properties of the Kernel K_n which are more suitable for our purposes.

Assumption 3. For a positive integer β' and a constant $\gamma \geq 1$ assume that

$$\frac{1}{\mathcal{F}\psi(\boldsymbol{\xi})} = \big(P(\boldsymbol{\xi})\big)^\gamma, \tag{8}$$

where p is a polynomial of degree β', that is

$$P(\boldsymbol{\xi}) = \sum_{\substack{(i,j)\in\{0,\dots,\beta'\}^2\\ i+j\leq\beta'}} a_{i,j}\boldsymbol{\xi}^{(i,j)} = \sum_{\substack{(i,j)\in\{0,\dots,\beta'\}^2\\ i+j\leq\beta'}} a_{i,j}\xi_1^i\cdot\xi_2^j,$$

with possibly complex coefficients $a_{i,j}$.

Remark 3.

(i) If $\gamma \in \mathbb{N}$, there exists a representation of $1/\mathcal{F}\psi(\boldsymbol{\xi})$ in which $\gamma = 1$. Thus, in this case we choose the representation with $\gamma = 1$. If $\gamma \notin \mathbb{N}$, we choose the largest number γ for which (8) holds, in case this representation is not unique.

(ii) Assumption 3 guarantees that $\mathcal{F}\psi$ decays polynomially which ensures that the convolution function ψ is not too smooth. An example for such a function is the joint density of two independent Laplace distributed random variables:

$$\psi(\mathbf{z}) = \frac{1}{4}\exp\big(-|z_1|-|z_2|\big) \quad \text{with} \quad \frac{1}{\mathcal{F}\psi(\boldsymbol{\xi})} = 2\pi\big(1+\xi_1^2+\xi_2^2+\xi_1^2\xi_2^2\big),$$

where $\beta' = 4$ and $\gamma = 1$.

(iii) Prominent examples of rotationally invariant functions are also included in Assumption 3, such as

$$\psi(\mathbf{z}) = 2^{-\frac{3}{2}} \exp\bigl(-\sqrt{z_1^2 + z_2^2}\bigr) \quad \text{with} \quad \frac{1}{\mathcal{F}\psi(\boldsymbol{\xi})} = 2^{\frac{3}{2}}\bigl(1 + \xi_1^2 + \xi_2^2\bigr)^{\frac{3}{2}}.$$

In this example we have $\beta' = 2$ and $\gamma = 3/2$.

(iv) Some implications of Assumption 3 that will be used throughout this paper are listed in Lemma 3 in Section 6.

2.3 Theoretical results

In this paper we consider graphical analysis of the data. In particular, we discuss image reconstruction, i.e. removal of the blur introduced by the PSF and the detection of regions of statistically significant changes between images at different points in time. The first task is complicated by the necessity of choosing a regularization parameter for image reconstruction methods which controls the trade-off between fit to the (noisy) data and expected smoothness of the true image. Subsequently, we solve the problem of deciding which changes between image frames are statistically significant i.e. pointing towards a real change in the object's appearance and deciding which changes are just due to image noise. We adapt an approach based on uniform confidence bands which have only recently been developed by Proksch et al. (2014) to our model (2). The following sections will introduce some relevant theory and a bootstrap approach for the determination of quantiles for the confidence bands.

2.4 Graphical model choice

Throughout this paper the image to be reconstructed is modeled as a bi-variate function whose graph is a surface in $\mathbb{R}^3$. Asymptotic confidence surfaces are two random surfaces, forming a corridor, which are constructed in such a way that the object of interest, that is, the graph of the true function, is fully contained between the surfaces with high probability converging to the nominal level $1-\alpha$ as the sample size increases. The basic part is a limit theorem for the maximal deviation between the estimator and the function of interest which can be derived by means of extreme value theory. This yields a result of the form

$$\lim_{n\to\infty} \mathbb{P}\Bigl(\bigl(\sup_{\mathbf{x}\in[0,1]^2} v_n^{-1/2}(\mathbf{x})|\hat{f}_n(\mathbf{x}) - f(\mathbf{x})| - d_n\bigr)d_n < \kappa\Bigr) = e^{-2e^{-\kappa}}, \tag{9}$$

where $v_n(\mathbf{x}) = (na_nh^{1+\beta}(2\pi)^2)^{-2}\|K\|_2^2 g(\mathbf{x})$ is the asymptotic variance of the estimator $\hat{f}_n(\mathbf{x})$ (see Lemma 2 in Section 6). Note that (9) implies

$$\lim_{n\to\infty} \mathbb{P}\Bigl(f(\mathbf{x}) \in \bigl[\hat{f}_n(\mathbf{x}) - \Phi_{n,\alpha}(\mathbf{x})\,,\, \hat{f}_n(\mathbf{x}) + \Phi_{n,\alpha}(\mathbf{x})\bigr] \quad \forall\, \mathbf{x} \in [0,1]^2\Bigr) = 1-\alpha, \tag{10}$$

where $\Phi_{n,\alpha}(x) = (q_{1-\alpha}/d_n + d_n)v_n^{1/2}(x)$ with the $1-\alpha$-quantile $q_{1-\alpha} = -\log(-0.5\log(1-\alpha))$ of the Gumbel limit distribution (9). Eq. (10) implies that for all $\mathbf{x} \in [0,1]^2$ at the same time, with probability of approximately $1-\alpha$, the true value $f(\mathbf{x})$ lies between $\hat{f}_n(\mathbf{x}) - \Phi_{n,\alpha}(\mathbf{x})$ and $\hat{f}_n(\mathbf{x}) + \Phi_{n,\alpha}(\mathbf{x})$. Thus, the set $\bigl\{\bigl[\hat{f}_n(\mathbf{x}) - \Phi_{n,\alpha}(\mathbf{x})\,,\, \hat{f}_n(\mathbf{x}) + \Phi_{n,\alpha}(\mathbf{x})\bigr] \,|\, \mathbf{x} \in [0,1]^2\bigr\}$ defines

a lower and an upper surface between which the true surface $\left\{f(\mathbf{x})\,|\,\mathbf{x}\in[0,1]^2\right\}$ is contained with high probability.
Confidence surfaces for the difference between two images at different time steps immediately provide us with a mean for the determination of regions of statistically significant changes in the image. An alternative approach, based on the confidence surfaces for single images, would be to compare those surfaces for the two images which are to be compared. A region would be classified as showing a significant change in the image, if the surfaces do not overlap. However, this yields a procedure which will in general result in a more conservative test for significance of the differences between the two images than the approach based on the confidence surface for the difference of the two images. Hence, we will use the latter approach in the data analysis discussed below.
Next, we state one of our main results which is a limit theorem for the maximal deviation of the estimate $\hat{f}_n$ and the function of interest f of the form (9) for the Poisson-model (2).

Theorem 1. *If Assumptions 1, 2 and 3 are satisfied and that* $\frac{\log(n)^2}{na_n^2h^2}=o(1)$*, and*

$$na_nh\sqrt{\log(n)}\Big(\frac{1}{na_n}+h^{\beta+S_f}+(a_n^2h)^{S_{\mathcal{F}\eta}}I_{\mathbb{N}}(\gamma)+(a_n^2h)^{S_{\mathcal{F}\eta}\wedge\lfloor\gamma\rfloor}(1-I_{\mathbb{N}}(\gamma))\Big)=o(1),$$

as $n\to\infty$*, then, in the Poisson-model* (2)*, limit theorem* (9) *holds with* $d_n=\sqrt{2\ln(C_{n,1})}+\frac{\ln\left(2\ln(C_{n,1})\right)}{2\sqrt{2\ln(C_{n,1})}}$*, where*

$$C_{n,1}=\sqrt{\frac{C_2}{(2\pi)^3}\frac{1}{h^2}},\quad \textit{and}\quad C_2=\mathbf{det}\left(\left(\frac{(2\pi)^4}{\|K\|_2^2}\int_{\mathbb{R}^2}|\Psi(\boldsymbol{\xi})\mathcal{F}\eta(\boldsymbol{\xi})|^2\boldsymbol{\xi}^{(i,j)}\,d\boldsymbol{\xi}\right)_{i,j=1}^2\right).$$

Since for many regularization functions that are frequently used in practice, such as those discussed in Remark 2, not all conditions of Assumption 2 are satisfied, we now give a version of Theorem 1 where the condition $\mathcal{F}\eta(\boldsymbol{\xi})\equiv 1$ for all $\boldsymbol{\xi}\in[-\delta,\delta]^2$, $\delta>0$ is dropped.

Theorem 2. *Let all assumptions of Theorem 1 hold. Assume that* $\mathcal{F}\eta$ *satisfies the conditions listed in Assumption 2 for* $\delta=0$*, that is* $\mathcal{F}\eta(\boldsymbol{\xi})=1$ *is only required for* $\boldsymbol{\xi}=0$*. If the constant* S_f*, defined in Assumption 1, satisfies* $S_f\geq 2$ *and*

$$na_nh^{3+\beta}\log(n)\to 0\quad \textit{for}\quad n\to\infty,$$

limit theorem (9) *holds with the same constants as defined in Theorem 1.*

Note that the limit theorems both contain the unknown parameter (function) g. The following corollary assures that the result still holds if $g=f*\psi$ is estimated from the data. To this end let $\tilde{g}_n$ be the following estimator for $g=T_\psi f$:

$$\tilde{g}_n(\mathbf{y})=\frac{1}{(2\pi)^2}\int_{\mathbb{R}^2}\mathcal{F}\hat{g}(\boldsymbol{\xi})\exp(i\boldsymbol{\xi}^T\mathbf{y})\mathcal{F}\eta(h\boldsymbol{\xi})\,d\boldsymbol{\xi}.$$

Define

$$\hat{g}_n(\mathbf{y})=\begin{cases}g^* & \text{if } \tilde{g}_n(\mathbf{y})<g^*,\\ \tilde{g}_n(\mathbf{y}) & \text{if } \tilde{g}_n(\mathbf{y})\in[g^*,G^*],\\ G^* & \text{if} \tilde{g}_n(\mathbf{y})>G^*.\end{cases}\tag{11}$$

For the following corollary to hold we need to impose an additional assumption on the function η.

Assumption 4. *Let the function $\eta : \mathbb{R}^2 \to \mathbb{R}$ be such that its Fourier transform $\mathcal{F}\eta$ satisfies the condition of Assumption 2. Furthermore, let either*

$$\eta(\mathbf{z}) = \widetilde{\eta}\big(\|A\mathbf{z} + \mathbf{b}\|_{l^2}\big), \quad A \in \mathbb{R}^{2\times 2},\ \mathbf{b} \in \mathbb{R}^2$$

or

$$\eta(\mathbf{z}) = \check{\eta}\big(\mathbf{a}^T\mathbf{z} + b\big), \quad \mathbf{a} \in \mathbb{R}^2,\ b \in \mathbb{R},$$

where $\widetilde{\eta} : \mathbb{R} \to \mathbb{R}$ and $\check{\eta} : \mathbb{R} \to \mathbb{R}$ are functions of bounded variation on $\mathbb{R}$.

This assumption guarantees the strong uniform convergence of $\hat{g}_n$ to g as is shown in Lemma 6 and Corollary 4 in Section 6 below.

Corollary 1. *Assume that the conditions of Theorem 1 and Assumption 4 are satisfied. Then we have for any $\kappa \in \mathbb{R}$*

$$\lim_{n\to\infty} \mathbb{P}\left(\hat{f}_n(\mathbf{x}) - \Phi_{n,\kappa} \le f(\mathbf{x}) \le \hat{f}_n(\mathbf{x}) + \Phi_{n,\kappa}(\mathbf{x}) \ \textit{for all}\ \mathbf{x} \in [0,1]^2\right) = e^{-2e^{-\kappa}},$$

where the sequence $\Phi_{n,\kappa}(\mathbf{x})$ is defined by

$$\Phi_{n,\kappa}(\mathbf{x}) = \frac{\sqrt{\hat{g}_n(\mathbf{x})}(\frac{\kappa}{d_n} + d_n)\|K\|_2}{(2\pi)^2 n a_n h^{\beta+1}}.$$

2.5 Obtaining Quantiles

Since the speed of convergence in limit theorems of the form (9) is known to be only of logarithmic order (cf. Hall (1991)), the error in coverage accuracy of the asymptotic bands decays also only logarithmically. For this reason we propose two alternative ways to obtain quantiles that can be used instead of $q_{1-\alpha}$ for the construction of the confidence surfaces in this section. The first one is based on a strong invariance principle (see Lemma 4) and is referred to as Gaussian method in the following sections and the second way is based on a bootstrap re-sampling method.

2.5.1 Gaussian method

For the Gaussian approach we simulate realizations of *iid* normally distributed random variables. In Section 6 we show that, even though the residuals $\big(Y_{i,j} - g(x_i, x_j)\big)/\sqrt{g(x_i, x_j)}$ are not identically distributed and follow a discrete distribution,

$$Z_{n,0}(\mathbf{y}) = nh^{1+\beta}(\hat{f}_n(\mathbf{y}) - \mathbb{E}\hat{f}_n(\mathbf{y})) = \frac{h^\beta}{n a_n h} \sum_{i,j=-n}^{n} (Y_{i,j} - g(x_i, x_j)) K_n\Big(\frac{y_1 - x_i}{h}, \frac{y_2 - x_j}{h}\Big)$$

can be approximated in sup-norm, with an error of sufficiently small order, by

$$Z_{n,1}(\mathbf{y}) = \frac{h^\beta}{n a_n h} \sum_{i,j=-n}^{n} \eta_{i,j} \sqrt{g(x_i, x_j)} K_n\Big(\frac{y_1 - x_i}{h}, \frac{y_2 - x_j}{h}\Big), \tag{12}$$

where the $\eta_{i,j}$ are *iid* standard normally distributed random variables.

An application of Lemma 9 backwards yields that $Z_{n,1}$ in (12) can be replaced by

$$Z_{n,2}(\mathbf{y}) = \frac{h^\beta \sqrt{g(y_1,y_2)}}{nh} \sum_{i,j=-n}^{n} \eta_{i,j} K_n\Big(\frac{y_1-x_i}{h}, \frac{y_2-x_j}{h}\Big). \tag{13}$$

For the Gaussian approach to simulating quantiles, generate M fields $\{\eta^l_{i,j} \,|\, (i,j) \in \{-n,\dots,n\}^2\}$, $l=1,\dots,M$ of realizations of *iid* standard normally distributed random variables. For each l, calculate

$$Z^l_{n,2}(y_1,y_2) = \frac{h^\beta \sqrt{g(y_1,y_2)}}{na_n h} \sum_{i,j=-n}^{n} \eta^l_{i,j} K_n\Big(\frac{y_1-x_i}{h}, \frac{y_2-x_j}{h}\Big)$$

on a grid of values $(y_1,y_2) \in I^2$ and compute the suprema

$$S^l_n = \sup_{(y_1,y_2)\in I^2} |Z^l_{n,2}(y)|,\ l=1,\dots,M.$$

Finally, estimate the $(1-\alpha)$-quantile $q_{\eta,1-\alpha}$ from the sample $S^1_n,\dots,S^M_n$.

Corollary 2. *Assume that the conditions of Corollary 1 are satisfied. Then we have for any* $\kappa \in \mathbb{R}$, $\alpha \in (0,1)$

$$\lim_{n\to\infty} \mathbb{P}\left(\hat{f}_n(\mathbf{x}) - \Phi^\eta_{n,\alpha} \le f(\mathbf{x}) \le \hat{f}_n(\mathbf{x}) + \Phi^\eta_{n,\alpha}(\mathbf{x}) \ \textit{ for all } \ \mathbf{x} \in [0,1]^2\right) = 1-\alpha,$$

where the sequence $\Phi^\eta_{n,\alpha}(\mathbf{x})$ *is defined by*

$$\Phi^\eta_{n,\alpha}(\mathbf{x}) = \frac{\sqrt{\hat{g}_n(\mathbf{x})}(\frac{q_{\eta,1-\alpha}}{d_n} + d_n)\|K\|_2}{(2\pi)^2 n a_n h^{\beta+1}},$$

where $q_{\eta,1-\alpha}$ *is the estimated* $(1-\alpha)-$*quantile of* $\sup_{\mathbf{y}\in[0,1]^2} |Z_{n,2}(\mathbf{y})|$.

2.5.2 Poisson Bootstrap

For the bootstrap approach we suggest to generate data

$$Y^*_{i,j} \sim \text{Poisson}\big(\hat{g}_n(x_i,x_j)\big).$$

Conditionally on the observations, the field $\mathcal{Y} = \{Y^*_{i,j} \,|\, (i,j) \in \{-n,\dots,n\}^2\}$ consists of independent, Poisson distributed random variables with $\mathbb{E}^*(Y^*_{i,j}) = \text{Var}^*(Y^*_{i,j}) = \hat{g}_n(x_i,x_j)$. Define $\hat{f}^*_n(\mathbf{y})$ and $\hat{g}^*(\mathbf{y})$ as the bootstrap versions of the estimators $\hat{f}_n$ and $\hat{g}_n$, that is

$$\hat{f}^*_n(y_1,y_2) = \frac{1}{(2\pi)^2 n^d a_n^2 h^2} \sum_{(i,j)\in\{-n,\dots,n\}^2} Y^*_{i,j} K_n\left(\frac{y_1-x_i}{h}, \frac{y_2-x_j}{h}\right).$$

and

$$\hat{g}^*_n(\mathbf{y}) = \begin{cases} g^* & \text{if } \tilde{g}^*_n(\mathbf{y}) < g^*, \\ \tilde{g}^*_n(\mathbf{y}) & \text{if } \tilde{g}^*_n(\mathbf{y}) \in [g^*, G^*], \\ G^* & \text{if} \tilde{\ } g^*_n(\mathbf{y}) > G^*, \end{cases}$$

where $\tilde{g}_n^*(\mathbf{y}) = \big(\hat{f}_n^* * \psi\big)(\mathbf{y})$.

This way, generate M bootstrap fields $\{Y_{i,j}^{l*} \,|\, (i,j) \in \{-n,\dots,n\}^2\}$, $l = 1,\dots,M$ and define

$$Z_{n,1}^{l*}(y_1,y_2) = \frac{h^\beta}{nh} \sum_{i,j=-n}^{n} \big(Y_{i,j}^{l*} - \hat{g}(x_i,x_j)\big) K_n\Big(\frac{y_1 - x_i}{h}, \frac{y_2 - x_j}{h}\Big)$$

on a grid of values $(y_1,y_2) \in I^2$ and calculate the suprema

$$S_n^{l*} = \sup_{(y_1,y_2)\in I^2} |Z_{i,j}^{l*}(y_1,y_2)|,\ l = 1,\dots,M.$$

Finally, estimate the $(1-\alpha)$-quantile $q_{1-\alpha}^*$ from the bootstrap sample $S_n^{1*},\dots,S_n^{M*}$.

Lemma 1. *Let the assumptions of Lemma 6 be satisfied. Then, as $n \to \infty$*

$$\sup_{\mathbf{y}\in[0,1]^2} |\hat{g}_n^*(\mathbf{y}) - \hat{g}_n(\mathbf{y})| = o_P^*\big(\log(n)^{-\frac{1}{2}}\big) \quad a.s..$$

The following corollary is a direct consequence of Theorems 1 and 3 and Lemma 1.
The next Theorem justifies the use of a bootstrap procedure in order to approximate the quantiles.

Theorem 3. *If the assumptions of Theorem 1 are satisfied, the following two limit theorems hold:*

$$\lim_{n\to\infty} \mathbb{P}^*\left(\sup_{\mathbf{x}\in[0,1]^2} \Big(v_n^*(\mathbf{x})^{-\frac{1}{2}} |\hat{f}_n^*(\mathbf{x}) - \hat{f}_n(\mathbf{x})| - d_n\Big) \cdot d_n < \kappa\right) = e^{-2e^{-\kappa}} \quad a.s., \tag{14}$$

and

$$\lim_{n\to\infty} \mathbb{P}^*\left(\sup_{\mathbf{x}\in[0,1]^2} \Big(\hat{v}_n^*(\mathbf{x})^{-\frac{1}{2}} |\hat{f}_n^*(\mathbf{x}) - \hat{f}_n(\mathbf{x})| - d_n\Big) \cdot d_n < \kappa\right) = e^{-2e^{-\kappa}} \quad a.s., \tag{15}$$

where

$$v_n^*(\mathbf{x}) = \frac{\hat{g}(\mathbf{x})\,\|K\|_2}{(2\pi)^4 n^2 a_n^2 h^{2+2\beta}} \quad \text{and} \quad \hat{v}_n^*(\mathbf{x}) = \frac{\hat{g}^*(\mathbf{x})\,\|K\|_2}{(2\pi)^4 n^2 a_n^2 h^{2+2\beta}}.$$

The sequence d_n is the same as in Theorem 1.

Corollary 3. *Assume that the conditions of Corollary 1 are satisfied. Then we have for any $\kappa \in \mathbb{R}$*

$$\lim_{n\to\infty} \mathbb{P}^*\left(\hat{f}_n(\mathbf{x}) - \Phi_{n,\alpha}^* \le f(\mathbf{x}) \le \hat{f}_n(\mathbf{x}) + \Phi_{n,\alpha}^*(\mathbf{x}) \ \text{for all}\ \mathbf{x} \in [0,1]^2\right) = 1 - \alpha \quad a.s.,$$

where the sequence $\Phi_{n,\alpha}^*(\mathbf{x})$ is defined by

$$\Phi_{n,\alpha}^*(\mathbf{x}) = \frac{\sqrt{\hat{g}_n(\mathbf{x})}(\frac{q_{1-\alpha}^*}{d_n} + d_n)\|K\|_2}{(2\pi)^2 n a_n h^{\beta+1}},$$

where $q_{1-\alpha}^*$ is the $(1-\alpha)-$quantile of either bootstrap procedure (14) or (15).

2.6 Stochastic multiresolution analysis in nonparametric regression

In this Section we give a short summary of the basic ideas of stochastic multiresolution analysis, which will be used below to select the regularization parameter for the estimator (4) (see Bissantz et al. (2008)). The basic idea amounts to testing the distribution of the residuals on all scales for being distinguishable from pure noise and is based on the following theoretical result.

Assume that we have data at our disposal according to model (2) and let $\hat{f}_n$ be some reconstruction of the (unknown) true image f. We aim to decide whether $\hat{f}_n$ is a reasonable reconstruction of the data, which, in turn, will indicate whether the regularization parameter h involved in the computation of (4) is acceptable. If so, we expect $\tilde{g}_n = \psi * \hat{f}_n$ to be a good approximation to the true observable image $g := \psi * f$ and the residuals

$$R_{i,j} := \frac{Y_{i,j} - \tilde{g}_n(x_i, x_j)}{\sqrt{\hat{g}_n(x_i, x_j)}}, \tag{16}$$

should resemble standardized Poisson noise (model (2)). Large values of $R_{i,j}$ indicate substantial remaining signal in the residuals due to over-smoothing of the estimator (4) whereas too small values of $R_{i,j}$ indicate over-fitting. In both cases we would consider the corresponding regularization parameter improperly chosen.

In stochastic multiresolution analysis we aim at testing the distribution of residuals simultaneously on all scales for significant deviation from randomness and for indications of over- or under-fitting. The basic idea is to check the residual properties on all scales by means of the increments of their partial sums which in the most simple (i.e. one-dimensional) setting implies controlling the value of the statistic

$$\mathcal{D}(n) = \max_{0 \le i < j \le n} \frac{|S_j - S_i|}{k\alpha(k/\log(n))}, \quad \text{where} \quad S_l := \sum_{i=1}^{l} R_i,$$

where $k = j - i$ and α is the *inverse Chernoff* function of the R_i, which depends on the distribution of the residuals (16). Note that this test considers all scales for testing if the residuals are consistent with pure Poisson noise, which is very different from a test based on a global statistic such as the sum of squares of the residuals.

By a result from Steinebach (1998), $\mathcal{D}(n) \to 1$ almost surely as $n \to \infty$, which suggests to consider choices of the regularization parameter to be reasonable if $\mathcal{D}(n) \approx 1$. Note that for fixed n, the difference $S_j - S_i$ is the increment of the (discrete) partial sum process $S_{(\cdot)}$ over the set $[i, j] \cap \{1, \ldots, n\}$.

In image analysis we have to consider two-dimensional signals $f(x_1, x_2)$ and hence two-dimensional arrays of residuals as well as two-dimensional partial sums and their increments. To this end define the double-indexed partial sum S_{i_1,i_2} by

$$S_{i_1,i_2} = \sum_{p=1}^{i_1} \sum_{q=1}^{i_2} R_{p,q}$$

and its increment $\mathcal{I}_{(i_1,i_2),(j_1,j_2),(n_1,n_2)}$ over the discrete grid $\mathcal{X}_{(i_1,i_2),(j_1,j_2),(n_1,n_2)} := [i_1, j_1] \times [i_2, j_2] \cap \{1, \ldots, n_1\} \times \{1, \ldots, n_2\}$ by

$$\mathcal{I}_{(i_1,i_2),(j_1,j_2),(n_1,n_2)} := S_{j_1,j_2} - S_{i_1,j_2} + S_{i_1,i_2} - S_{j_1,i_2} = \sum_{p,q \in \mathcal{X}_{(i_1,i_2),(j_1,j_2),(n_1,n_2)}} R_{p,q}. \tag{17}$$

This means that in the two-dimensional multiscale approach we need to control the normalized increments $\mathcal{I}_{(i_1,i_2),(j_1,j_2),(n_1,n_2)}$ for all rectangles $\mathcal{X}_{(i_1,i_2),(j_1,j_2),(n_1,n_2)}$. Kabluchko and Munk (2009) showed that

$$\lim_{n\to\infty} \max_{\substack{\mathcal{X}_{(i_1,i_2),(j_1,j_2),(n_1,n_2)} \\ 1\le i_k<j_k\le n_k,\, k=1,2}} \frac{\mathcal{I}_{(i_1,i_2),(j_1,j_2),(n_1,n_2)}}{(j_1-i_1)(j_2-i_2)\alpha\Big(\frac{(j_1-i_1)(j_2-i_2)}{\log(n)}\Big)} = 1 \quad \text{a.s.}$$

if $\{R_{p,q} \,|\, (p,q) \in \mathbb{N}^2\}$ is an array of independent and identically distributed random variables with unit variances such that $\log \mathbb{E}\big(\exp(\theta R_{p,q})\big) < \infty$ for some $\theta > 0$. Hence, the same ideas as in the one-dimensional case apply in the two-dimensional setting as well.

3 Material and methods

3.1 Microscopic images

For live-cell imaging HeLa-Cells (ATCC) were cultured on cover-slips with Dulbecco's Modified Eagle Medium (DMEM), 5% Fetal Calf Serum (FCS) at $37°C$. Prior to staining the cells were washed three times with Phosphate Buffered Saline (PBS). Cells were then incubated with the lipophilic carbocyan dye DiI (Molecular Probes) in DMEM for 5 min at $37°C$. Cells were washed with DMEM to remove excess dye and were then subjected to fluorescence microscopy using a confocal laser scanning microscope (Leica TCS) equipped with a HeNe-Laser. The samples were excited at 543 nm and imaged for about 50 seconds. Finally, the number of pixels per image is $512 \times 512 = 262144$ and the number of time steps is 8+1.

3.2 Image reconstruction and parameter selection

For the image reconstruction we use the estimator (4) proposed in section 2.1. Such regularization methods require to fix a regularization parameter for the method. In the case of the damped spectral cut-off estimator (4) this amounts to the selection of the regularization parameter h.

We suggest the following procedure based on the multiresolution introduced in Section 2.6 and which is based on the multiscale statistic

$$MR = \max_{L,k,l} \frac{\Big|\sum\limits_{(i,j)\in A_{ij}^{L,k,l}} R_{i,j}\Big|}{L^2 \rho_\mu^{-1}(L^2\log(n^2))},$$

where $A^{L,k,l}$ is the partition with number (k,l) of a partitioning of the residuals

$$R_{i,j} = \frac{\hat{r}_{i,j}}{\mathrm{sd}(\hat{r}_{i,j})} \quad \text{with} \quad \hat{r}_{i,j} := Y_{i,j} - \hat{f}_n(x_1^{(i)}, x_2^{(j)}).$$

of the image after deconvolution. L is the number of pixels along each coordinate axis of the partitions and ρ_μ^{-1} the inverse Chernoff function of the $R_{i,j}$. Fig. 3 shows a map of the image partitioning in two different scales.

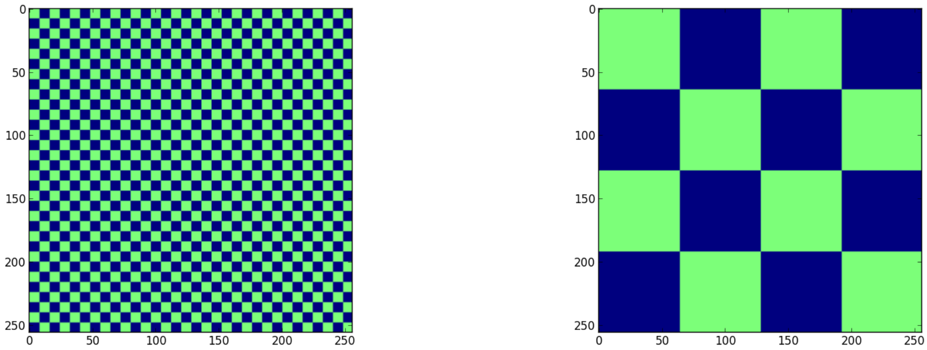

Figure 3: Partitioning of the residual image along two scales with $L = 8$ and $L = 64$, respectively.

The multiresolution statistic 'measures' the image quality on all scales considered. Whereas it can be shown theoretically under suitable assumptions that the multiscale statistic MR converges almost surely to 1 for $n \to \infty$, for finite size data the statistic MR shows a distribution of finite, non-zero variance due to the randomness of the residuals. Hence, we suggest the following approach. First, we simulate MR and determine the 5% and 95%-quantile of the simulated distribution (see Section 3.3 below for more details). In other words, in 90% of the cases the residuals, if due to random fluctuations for a Poisson data model, should be within these two quantiles. Then, we consider all regularization parameters which result in MR-values within this confidence interval to be acceptable and use the mean of these values in our subsequent computations. It turns out that the results from such simulations are only insignificantly dependent on the true test image used for the generation of the artificial data in the simulations. Hence, it is sufficient to determine the distribution of MR only once as long as the the main image characteristics (in particular the mean signal in the pixels) have not changed severely. Most importantly, for our application in live cell imaging, the differences between pictures from different time points taken with similar imaging properties (in particular) exposure time are negligibly small, such that a re-simulation of the statistic MR is clearly not required.

Numerical simulations indicate that the reconstruction can be difficult close to the edges of an image. Moreover, the PSF can be substantially different close to the edges as well (see **?**). Hence, we propose to focus on the most central part of the image for determining the value of the statistic MR. For computational reasons it is preferable to have a power of 2 as the number of pixels along each coordinate direction. Therefore, we use the central quarter of the image, however this can be changed as appears reasonable for the image under consideration. Figure 3 shows the smallest and largest partitioning of the image used in the computation of the multiscale statistic. In general, it is not feasible to use smaller scales to avoid problems which can be introduced by numerical artifacts, imperfect modeling of the PSF or similar problems. All these issues can result in a breakdown of the Poisson assumption for single pixels, where this problem is averaged out to a sufficient extent if a minimum number of pixels is included in the sum of residuals.

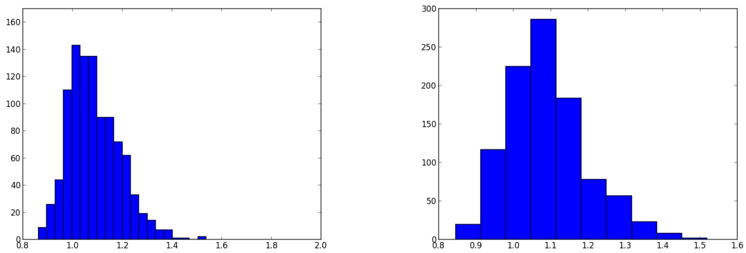

Figure 4: Simulated multiscale statistic for Poisson residuals. Left: Residuals based on artificial images generated with constant Poisson parameter approximately equal to the mean signal in the true images and right: Residuals for the difference of two random images generated according to the procedure on which the left panel is based. See text for details.

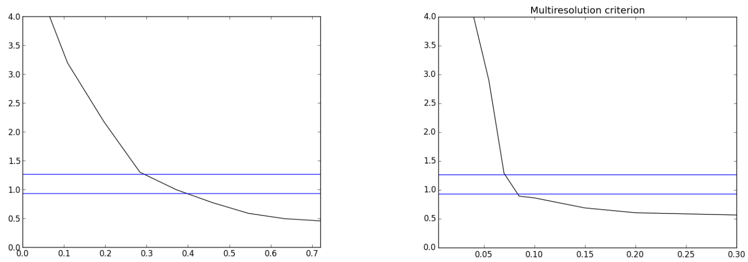

Figure 5: Left: Multiscale statistic for different regularization parameters of the damped Fourier estimator for the reconstruction of the image at time 0 with confidence limits from simulations of the multiscale statistic. Right: Results for the test image.

3.3 Data reduction

Based on the methods discussed in the previous sections we suggest the following pipeline for data analysis.

1. Definition of the PSF for estimator (4). In practical applications, the PSF is often only imperfectly known and has to be estimated from data. Fortunately, it is known that the convergence properties of estimators in so-called blind deconvolution problems are not deteriorated if (additional) data is available for estimation of the PSF (see Hoffman and Reiss (2008); Hall and Qiu (2007)).

   Here, the PSF is estimated from high resolution images of a solid ball of 200nm diameter (bead), which was imaged at $\approx 7\times$ the resolution of the images of interest along each axis, i.e. ≈ 50 pixels per pixel in the production images (see Figure 2). Hence, due to the significantly larger amount of data available, we assume the PSF to be known a priori. The PSF has the shape of a Laplace distribution with parameter λ, i.e. its shape is
$$\psi(x_1, x_2) = \frac{1}{2\lambda} \cdot \exp(-\sqrt{x_1^2 + x_2^2}/\lambda),$$
   where $\lambda = 0.065$ was determined with the method explained above. Whereas the theoretical results on which our confidence bands are based have not been shown for a

PSF with the shape of a normal density, we have used it empirically in the simulations and it turns out to yield on insignificantly different results as compared to the Laplace PSF results.

2. Selection of the regularization parameter with the multiscale method. For this we use simulated quantiles of the test statistic, where the simulations are based on an image with pixel intensities similar to the mean signal in the image at time step 0. For comparison we have also performed simulations based on the difference of two images, similar to the previous approach.

   In Figure 4 we compare the resulting distributions. To this end, we have generated artificial images and determined the value of the multiscale statistics for 1000 such artificial images, where the pixel values are generated as follows. In the left panel, each pixel value is distributed according to a Poisson distribution with constant Poisson parameter approximately equal to the mean value of observed counts in the true images, and in the right panel, each pixel value is the difference between two realizations of the image generated according to the procedure described for the left panel. Using an image or a difference map (with much smaller mean count rate than the image) does not have a significant impact.

   Figure 5 shows the multiscale statistic for estimates with the damped spectral cut-off method in dependence of the regularization parameters from 1000 simulations each. The horizontal lines indicate 90% confidence intervals determined from the empirical quantiles of the simulated distributions of the test statistic. We will use a regularization parameter of 0.23 in both cases in the subsequent reconstructions. Note that all regularization parameters given below are in terms of the Nyquist frequency.

3. Reconstruction of the images both with a damped spectral cut-off approach (4) and determination of the differences between the reconstructed images at different time steps.

4. Bootstrap simulations of the quantiles for the confidence bands for the difference between the reconstructions at two different points in time. Here, we use only image pixels which are at least 5% of the axis length from the edges of the image to avoid problems with reconstruction at the image edges.

5. Determination of regions of significant change in the difference maps between images. To this end, we determine whether the observed difference at some pixel position (i, j) is larger than the 90% or 95% quantile (depending on the significance level α) of the simulated quantiles. If yes, a significant change is indicated.

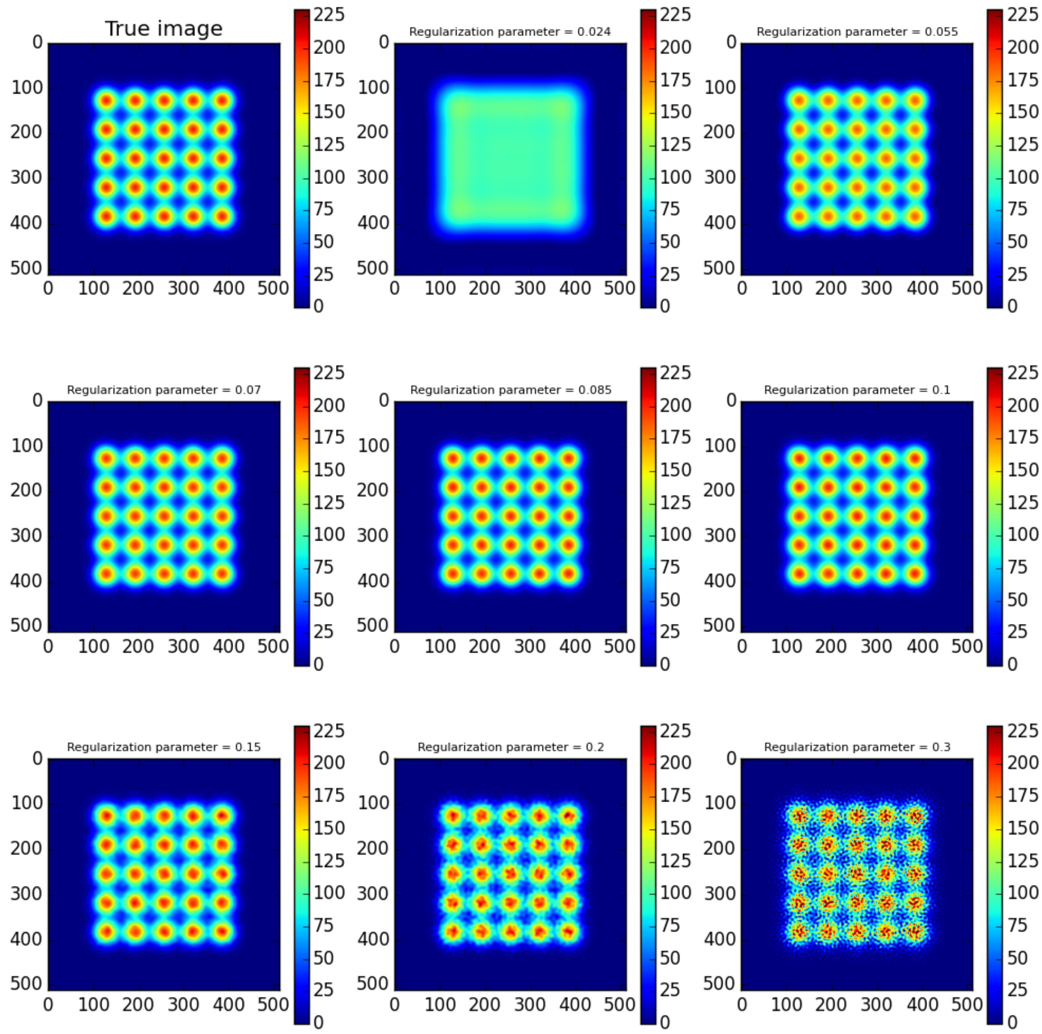

Figure 6: Results for the test image. All regularization parameters given below are in terms of the Nyquist frequency. See text for details.

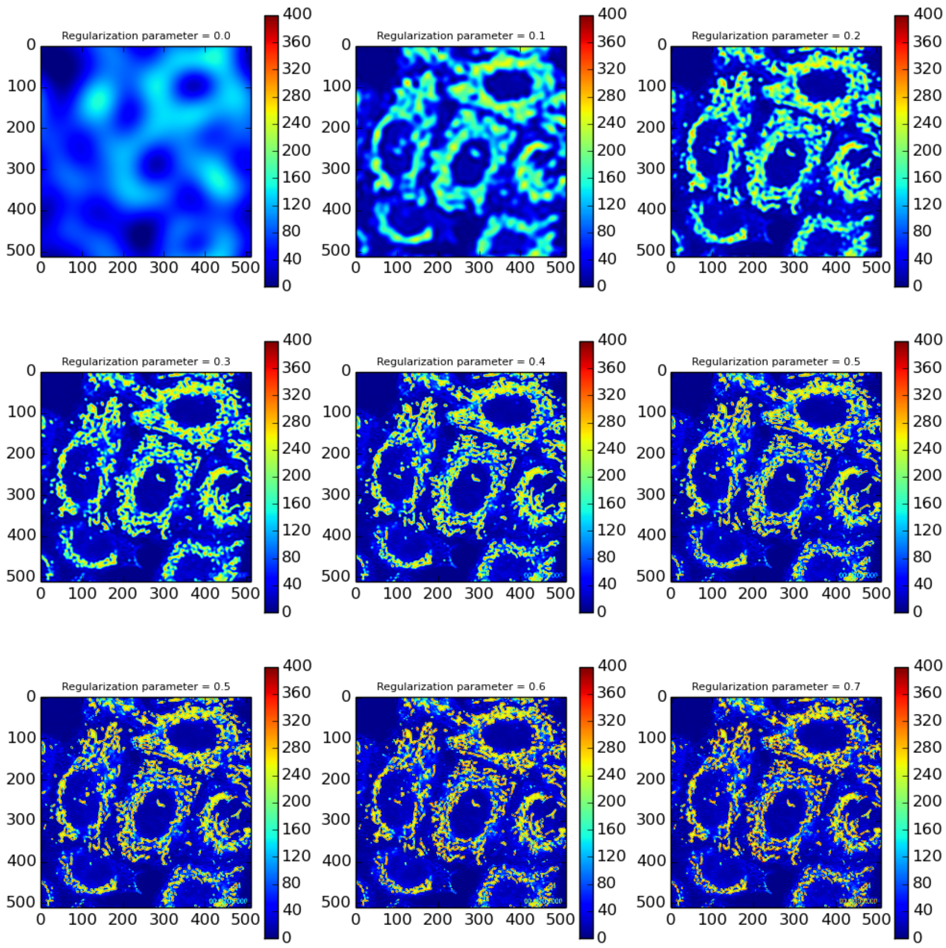

Figure 7: Damped SC method reconstructions for a selection of regularization parameters for time step 0. All regularization parameters given below are in terms of the Nyquist frequency.

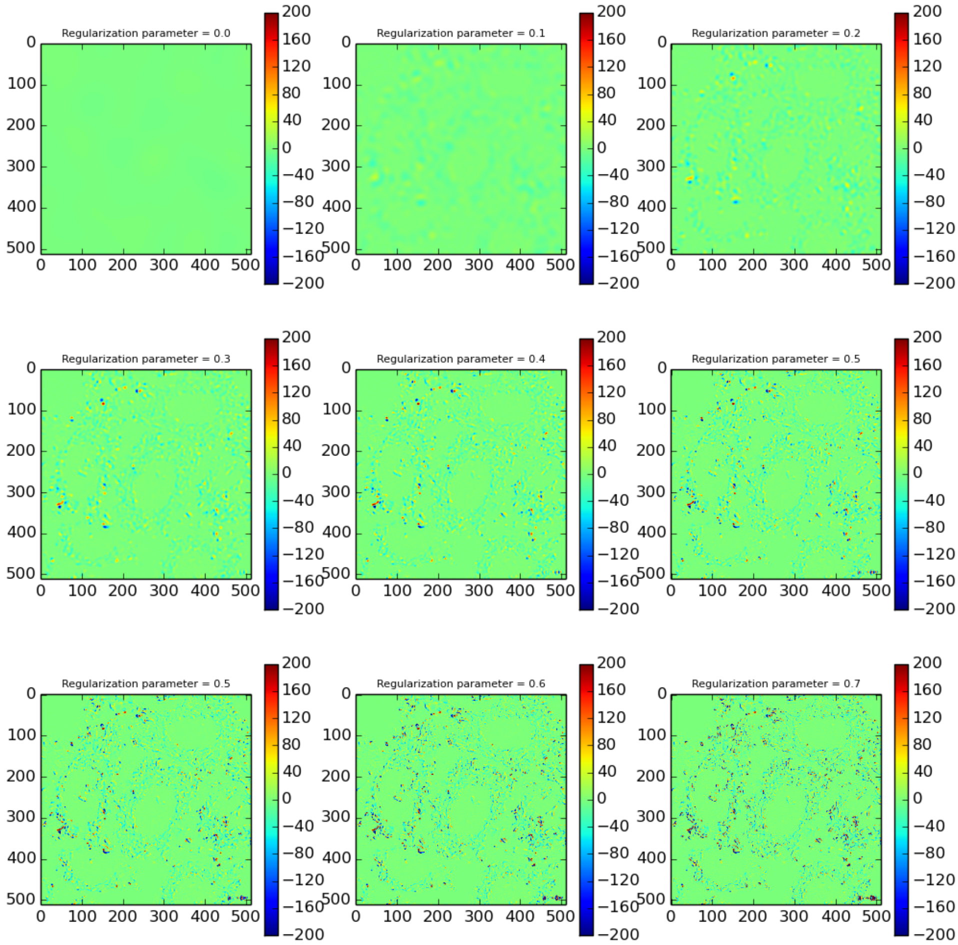

Figure 8: Damped SC method reconstructions for a selection of regularization parameters for the difference between time steps 0 and 3. All regularization parameters given below are in terms of the Nyquist frequency.

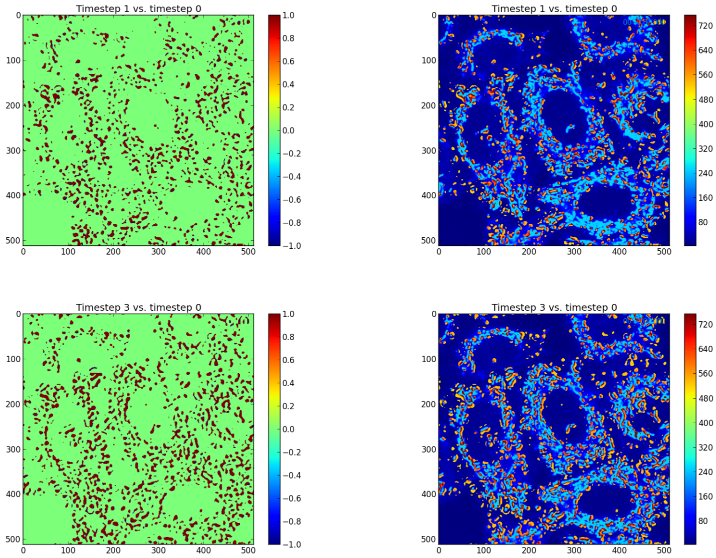

Figure 9: Significant differences between time steps 1 and 0 and 3 and 0 from different methods. Left: Regions with significant change with time of the image. Right: Difference maps overlayed on the image at time 0. In the middle of the right pictures from time-point 1 to 3 the budding of a small vesicle from a larger tubular structure can be observed.

4 Results

In a first step we have used artificial data to test the estimator in combination with the multiresolution method for feasibility. Figure 6 shows the results for the damped spectral cut-off estimator and the multiresolution parameter selection method for a test image with approximately the same mean Poisson parameter in the central part of the image (which is essentially used in the determination of the multiresolution statistic). Both from a visual selection of the optimal regularization parameters and the multiresolution statistic we conclude on optimal regularization parameters which are the same (and the images for regularization parameters 0.07 and 0.085 as most preferable choices).

In the second part of our numerical study we have applied our method to the Hela cell data. Figures 7 and 8 present the estimates for some regularization parameters for the image at time step 0 and for the difference between the images at time steps 0 and 3. Note that the multiresolution method indicates a best choice for the regularization parameter of 0.23.

In more detail, living HeLa cells were stained with the fluorescent dye DiI which labels lipid membranes in the cell. These so called intracellular transport compartments consist of round vesicular and tubular structures (see lower picture on the left side of Figure 9). The vesicles and micro-tubules mediate intracellular transport processes and are rapidly moving within the cell. The presented method is also suited to detect structural changes of biological relevance in the observed object: E. g. in the middle of the right pictures of Figure 9 the budding of a round vesicle from a tubular structure can be observed between the time-points 1 and 3. A second example in the center of the upper right picture of Figure 9 shows the budding of two vesicles from a tubular structure : At time step 3 a tubular structure is located in the middle of the center cell. At time step 7 in the bottom right picture two small vesicular structures (in yellow) have been emerging from the tubular structure which shows the pinching off of two vesicles.

5 Discussion

The results indicate that the method presented above provides biologically relevant information on small scale structural changes in time resolved fluorescence microscopy of living cells. Intracellular transport, which was monitored in the present experiment, plays a vital role in the functioning of an organism. It is e. g. involved in receptor signaling processes in the nervous system and the immune system. Defects in intracellular trafficking in the nervous system play a key role in the onset of neurodegenerative disorders such as Alzheimer's or Huntington's disease. In summary, this method may contribute to elucidate processes of intracellular trafficking in living cells and thereby also provide insight into defects of these transport processes which may lead to neurodegenerative diseases.

Acknowledgements. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt C4) of the German Research Foundation (DFG).

6 Auxiliary results

Lemma 2. *Suppose that Assumptions 1, 2 and 3 are satisfied and that* $na_nh^2 \to 0$ *for* $n \to \infty$. *Then*

(i)

$$\sup_{\mathbf{y}\in[0,1]^2} \left|\hat{f}_n(\mathbf{y}) - \mathbb{E}\big[\hat{f}_n(\mathbf{y})\big]\right| = \begin{cases} O\Big(\frac{1}{na_nh^\beta} + h^{S_f} + a_n^{S_{\mathcal{F}g}} h^{-\beta}(a_nh)^{S_{\mathcal{F}\eta}}\Big) & \gamma \in \mathbb{N} \\ O\Big(\frac{1}{na_nh^\beta} + h^{S_f} + a_n^{S_{\mathcal{F}g}} h^{-\beta}(a_nh)^{\lfloor\gamma\rfloor \wedge S_{\mathcal{F}\eta}}\Big) & \gamma \notin \mathbb{N}, \end{cases}$$

where S_f *and* $S_{\mathcal{F}g}$ *denote the constants defined in Assumption 1.*

(ii) With the definition

$$v_n(\mathbf{y}) := \frac{g(\mathbf{y})\,\|K\|_2}{(2\pi)^4 n^2 a_n^2 h^{2+2\beta}} \tag{18}$$

the following asymptotic expansion holds for the variance of the estimator $\hat{f}_n$*:*

$$Var\big[\hat{f}_n(\mathbf{y})\big] = v_n(\mathbf{y}) + o\left(\frac{1}{n^2a_n^2h^{2+2\beta}}\right).$$

Lemma 3. *Suppose that Assumptions 2 and 3 are satisfied and recall that* $\beta = \beta' \cdot \gamma$, $1/\mathcal{F}\psi(\boldsymbol{\xi}) = (P(\boldsymbol{\xi}))^\gamma$, *where* P *is a polynomial of degree* β'. *Define*

$$\Psi(\boldsymbol{\xi}) := \sum_{\substack{(i,j)\in\{0,\dots,\beta'\}^2 \\ i+j=\beta'}} a_{i,j}\boldsymbol{\xi}^{(i,j)} \quad \text{and} \quad K(\mathbf{z}) := \int_{\mathbb{R}^2} \mathcal{F}\eta(\boldsymbol{\xi})\Psi(\boldsymbol{\xi})\exp(i\boldsymbol{\xi}^T\mathbf{z})\,d\boldsymbol{\xi}. \tag{19}$$

Then the following properties hold:

(i) $\sup_{\boldsymbol{\xi}\in[-1,1]^2}\Big|\dfrac{h^\beta}{\mathcal{F}\psi\big(\frac{\boldsymbol{\xi}}{h}\big)} - \Psi(\boldsymbol{\xi})\Big| = O(h).$

(ii) $\|h^\beta K_n - K\|_p = O(h)$, *for all* $p \in [2,\infty)$.

(iii) Recall that $S_{\mathcal{F}\eta}$ *is the degree of smoothness of* $\mathcal{F}\eta$ *(see Assumption 2). If* $\gamma \in \mathbb{N}$, *all partial derivatives of order* $\boldsymbol{\alpha}$ $\partial^{\boldsymbol{\alpha}} h^\beta \mathcal{F}\eta/\mathcal{F}\psi(\cdot/h)$ *are bounded for all* $|\boldsymbol{\alpha}| \leq S_{\mathcal{F}\eta}$*:*

$$\lim_{n\to\infty}\Big\|\partial^{\boldsymbol{\alpha}} h^\beta \frac{\mathcal{F}\eta}{\mathcal{F}\psi(\cdot/h)}\Big\|_\infty < \infty. \tag{20}$$

(iv) If $\gamma \notin \mathbb{N}$, (20) *holds for all double-indices* $\boldsymbol{\alpha}$ *with* $|\boldsymbol{\alpha}| \leq \min\{S_{\mathcal{F}\eta}, \lfloor\gamma\rfloor\}$.

(v) There exists a constant C *such that*

$$\Big|\partial^{\boldsymbol{\alpha}} h^\beta\Big(\frac{\mathcal{F}\eta(\boldsymbol{\xi})}{\mathcal{F}\psi\big(\frac{\boldsymbol{\xi}}{h}\big)}\Big) - \partial^{\boldsymbol{\alpha}}\Big(\mathcal{F}\eta(\boldsymbol{\xi})\Psi(\boldsymbol{\xi})\Big)\Big| \leq ChI_{[-1,1]^2}(\boldsymbol{\xi}),$$

if $|\boldsymbol{\alpha}| = 1$.

(vi) *a) For all* $\boldsymbol{\alpha} \in \mathbb{N}^2$ *the following estimate is valid:*

$$\int_{[-1/(a_n h),1/(a_n h)]^2}\big|h^\beta\partial^{\boldsymbol{\alpha}}K_n(\mathbf{z}) - \partial^{\boldsymbol{\alpha}}K(\mathbf{z})\big|\,d\mathbf{z} = O\big(h\sqrt{\log(n)}\big).$$

b) For all $\boldsymbol{\alpha} \in \mathbb{N}^2$, $j = 1,2$ *the following order of convergence is obtained:*

$$\int_{\mathbb{R}}\big|h^\beta\partial^{\boldsymbol{\alpha}}K_n((z_1,z_2)) - \partial^{\boldsymbol{\alpha}}K(z_1,z_2)\big|\,dz_j = O\big(h\big).$$

Lemma 4. *Suppose that Assumption 1 holds and recall that* $S_{i,j}$ *is defined as the* (i,j)*-th partial sum of the standardized Poisson variables* $Y_{k,l}$ *from model* (2)*, that is,*

$$S_{i,j} = \sum_{k=1}^{i}\sum_{l=1}^{j}\varepsilon_{k,l} \quad \textit{with} \quad \varepsilon_{k,l} = \frac{Y_{k,l} - g(\mathbf{x}_{k,l})}{\sqrt{g(\mathbf{x}_{k,l})}}.$$

There exists a Wiener field W *on* $[0,\infty)^2$ *such that for arbitrarily small* $\mu > 0$

$$\sup_{(i,j)\in\{1,\ldots,n\}^2}\big|S_{(i,j)} - W(i,j)\big| = o(n^\mu) \ \ a.s.$$

Lemma 5. *Let* $\{W(\mathbf{t})\,|\,\mathbf{t}\in\mathbb{R}^2\}$ *be a Wiener sheet,* K *some kernel with either bounded Fourier transform* $|\mathcal{F}K| \leq K^*$ *or finite* L^1*-norm* $\|K\|_1 < \infty$ *whose partial derivatives* $\partial^{\boldsymbol{\alpha}}K$ *of order up to* $|\boldsymbol{\alpha}| = 3$ *exist and are square-integrable. Then, if* $h = h_n \to 0$, *the field*

$$Z_n(\mathbf{x}) := \frac{1}{h}\int_{\mathbb{R}^2} K\Big(\frac{\mathbf{x}-\mathbf{y}}{h}\Big)\,dW(\mathbf{y})$$

satisfies

$$\sup_{\mathbf{x}\in[0,1]^2}|Z_n(\mathbf{x})| = O_P\big(\sqrt{\log(n)}\big).$$

Lemma 6. *Let $\tilde{g}_n$ be the following estimator for $g = T_\psi f$:*

$$\tilde{g}_n(\mathbf{y}) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \mathcal{F}\hat{g}(\boldsymbol{\xi}) \exp(i\boldsymbol{\xi}^T \mathbf{y}) \mathcal{F}\eta(h\boldsymbol{\xi})\, d\boldsymbol{\xi}.$$

Under the assumptions of Corollary 1

$$\sup_{(y_1,y_2)\in\mathbb{R}^2} |\tilde{g}_n(y_1,y_2) - g(y_1,y_2)| = o\Big(\frac{1}{\sqrt{\log(n)}}\Big) \quad a.s.$$

The following simple observation allows us to transfer the convergence result for $\tilde{g}_n$ from Lemma 6 to the estimator $\hat{g}_n$ defined in (11). If the function g is bounded from below and above, $g^* \le g \le G^*$ we find

$$\|g - \hat{g}_n\|_\infty \le \|g - \tilde{g}_n\|_\infty$$

Corollary 4. *Under the assumptions of Corollary 1*

$$\sup_{(y_1,y_2)\in\mathbb{R}^2} |\hat{g}_n(y_1,y_2) - g(y_1,y_2)| = o\Big(\frac{1}{\sqrt{\log(n)}}\Big)\ a.s.$$

References

Adorf, H. (1995). Hubble space telescope image restauration in its fourth year. *Inverse Problems*, 11:639–653.

Bertero, M., Boccacci, P., Desiderà, G., and Vicidomini, G. (2009). Image deblurring with Poisson data: from cells to galaxies. *Inverse Problems*, 25(12):123006, 26.

Bissantz, N., Mair, B., and Munk, A. (2008). A statistical stopping rule for mlem reconstructions in pet. *IEEE Nucl. Sci. Symp. Conf. Rec.*, 8:4198–4200.

Cavalier, L. (2000). Efficient estimation of a density in a problem of tomography. *Annals of Statistics*, 28:630–647.

Cavalier, L. and Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. *Probab. Theory Relat. Fields*, 123:323–354.

Hall, P. (1991). On convergence rates of suprema. *Probability theory and related fields*, 89:447–455.

Hall, P. and Qiu, P. (2007). Nonparametric estimation of a point-spread function in multivariate problems. *Ann. Statist.*, 35:1512–1534.

Hoffman, M. and Reiss, M. (2008). Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Statist.*, 36:310–336.

Hotz, T., Marnitz, P., Stichtenoth, R., Davies, L., Kabluchko, Z., and Munk, A. (2012). Locally adaptive image denoising by a statistical multiresolution criterion. *Comput. Stat. Data An.*, 56:543 –558.

Johnson, I. and Spence, M., editors (2010). *The Molecular Probes Handbook, A Guide to Fluorescent Probes and Labeling Technologies*, chapter 13.4. "", 11 edition.

Kabluchko, Z. and Munk, A. (2009). Shao's theorem on the maximum of standardized random walk increments for multidimensional arrays. *ESAIM Probab. Stat.*, 13:409–416.

Mair, B. A. and Ruymgaart, F. H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56:1424–1444.

Proksch, K., Bissantz, N., and Dette, H. (2014). Confidence bands for multivariate and time dependent inverse regression models. To appear in *Bernoulli*.

Saitoh, S. (1997). *Integral Transforms, Reproducing Kernels and their Applications*. Longman, Harlow.

Steinebach, J. (1998). On a conjecture of revesz and ist analogue for renewal processes. In *Proceed. Intern. Conf. Asympt. Meth. Probab. Statist. (ICAMPS'97)*, pages 311–322, Amsterdam. Elsevier.

Wallace, W., Schaefer, L., and Swedlow, J. (2001). A workingperson's guide to deconvolution in light microscopy. *BioTechniques*, 31:1076–1097.

7 Proofs

7.1 Proofs of the main results

Proof of Theorem 1

Note that by the assumptions of the theorem Lemma 2 implies

$$\sup_{\mathbf{y}\in[0,1]^2}\big|\hat{f}_n(\mathbf{y})-\mathbb{E}\big[\hat{f}_n(\mathbf{y})\big]\big| = o\Big(\frac{1}{\sqrt{\log(n)}}\Big).$$

Thus, we only need to consider the stochastic part $\hat{f}_n(\mathbf{y})-\mathbb{E}[\hat{f}_n(\mathbf{y})]$. In the subsequent proof we aim at approximating the standardized quantity

$$\begin{aligned}Z_{n,0}(\mathbf{y}) &:= v_n^{-\frac{1}{2}}(\mathbf{y})\Big(\hat{f}_n(\mathbf{y})-\mathbb{E}[\hat{f}_n(\mathbf{y})]\Big)\\ &= \frac{h^\beta}{\|K\|_2\sqrt{g(\mathbf{y})}na_nh}\sum_{i,j=-n}^{n}\big(Y_{(i,j)}-g(\mathbf{x}_{i,j})\big)K_n\Big(\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}\Big)\end{aligned} \tag{21}$$

by a suitable Gaussian process uniformly with respect to the variable $\mathbf{y}\in[0,1]^2$. To this end we split the grid of indices $\{-n,\dots,n\}^2$ into the intersections with the four quadrants in $\mathbb{R}^2$ and the intersection with the axes, that is,

$$\{-n,\dots,n\}^2 = \mathcal{I}_{n,-,-}\,\dot\cup\,\mathcal{I}_{n,-,+}\,\dot\cup\,\mathcal{I}_{n,+,-}\,\dot\cup\mathcal{I}_{n,+}\,\dot\cup\,\mathcal{R}_n,$$

where $\mathcal{R}_n=\{0\}\times\{-n,\dots,n\}\cup\{-n,\dots,n\}\times\{0\}$, $\mathcal{I}_{n,-,+}=\{-n,\dots,-1\}\times\{1,\dots,n\}$ and so on. We first show an approximation result for the sum over the index-set $\mathcal{I}_{n,+,+}$. The result implies the respective results for the sums over the remaining quadrants. Finally we show that the sum over $\mathcal{R}_n$ is negligible.

The approximation is realized in several steps. We start with the initial process $Z^+_{n,0}$:

$$
\begin{aligned}
Z^+_{n,0}(\mathbf{y}) &:= v_n^{-\frac{1}{2}}(\mathbf{y})\Big(\hat{f}_n(\mathbf{y}) - \mathbb{E}[\hat{f}_n(\mathbf{y})]\Big) \\
&= \frac{h^\beta}{\|K\|_2\sqrt{g(\mathbf{y})}na_nh}\sum_{i,j=1}^{n}\big(Y_{(i,j)} - g(\mathbf{x}_{i,j})\big)K_n\Big(\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}\Big)
\end{aligned} \tag{22}
$$

In the first step (22) is approximated by

$$
\begin{aligned}
Z^+_{n,1}(\mathbf{y}) := \frac{h^\beta}{\sqrt{g(\mathbf{y})}na_nh}\Bigg\{ &\sum_{i,j=1}^{n} W(i,j)\int_{[\mathbf{x}_{i,j},\mathbf{x}_{i+1,j+1}]}\partial_{\mathbf{z}}^{(1,1)}\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{z})\,d\mathbf{z} \\
&- \sum_{i=0}^{n-1} W(i,n)\int_{[x_i,x_{i+1}]}\partial_{\mathbf{z}}^{(0,1)}\widetilde{\mathcal{K}}_{n,\mathbf{y}}\Big(\frac{1}{a_n},z_2\Big)\,dz_2 \\
&- \sum_{j=0}^{n-1} W(n,j)\int_{[x_j,x_{j+1}]}\partial_{\mathbf{z}}^{(1,0)}\widetilde{\mathcal{K}}_{n,\mathbf{y}}\Big(z_1,\frac{1}{a_n}\Big)\,dz_1 \\
&+ W(n,n)\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{n,n})\Bigg\},
\end{aligned} \tag{23}
$$

uniformly in $\mathbf{y}$ with an error of sufficiently small magnitude, where

$$
\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i,j}) := \frac{\sqrt{g(\mathbf{x}_{i,j})}}{\|K\|_2}K_n\Big(\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}\Big).
$$

Subsequently (23) will be is approximated by the integral

$$
Z^+_{n,2}(\mathbf{y}) := \frac{h^\beta}{h\sqrt{g(\mathbf{y})}}\int_{[0,1/a_n]^2}\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{z})\,dW(\mathbf{z}). \tag{24}
$$

In order to get to (25)

$$
Z^+_{n,3}(\mathbf{y}) := \frac{h^\beta}{\|K\|_2h}\int_{[0,1/a_n]^2}K_n\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}), \tag{25}
$$

from (24) the localizing property of the scale family of kernels is used.

In two final steps the sequence of kernels K_n is replaced by the limit kernel K, defined in (19), and the domain of integration is extended from $[0,1/a_n]^2$ to $[0,\infty)$. Those approximation steps are performed in Lemma 7 to Lemma 11. The corresponding approximating quantities are

$$
Z^+_{n,4}(\mathbf{y}) := \frac{1}{\|K\|_2h}\int_{[0,1/a_n]^2}K\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}) \tag{26}
$$

and

$$
Z^+_{n,5}(\mathbf{y}) := \frac{1}{\|K\|_2h}\int_{[0,\infty)^2}K\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}). \tag{27}
$$

Step 1 of the proof of Theorem 1: $Z^+_{n,0} \longrightarrow Z^+_{n,1}$

Lemma 7. *Under the assumptions of Theorem 1 there exist a Wiener field on a suitable probability space and a small constant $\nu > 0$ such that*

$$\sup_{\mathbf{y}\in[0,1]^2} \left| Z^+_{n,0}(\mathbf{y}) - Z^+_{n,1}(\mathbf{y}) \right| = O(n^{-\nu}) \quad a.s.$$

Proof. Define

$$\varepsilon_{i,j} := \frac{Y_{i,j} - g(\mathbf{x}_{i,j})}{\sqrt{g(\mathbf{x}_{i,j})}}.$$

Then $\{\varepsilon_{i,j} \,|\, (i,j) \in \{1,\ldots,n\}^2\}$ is a field of independent, centered random variables with unit variances and

$$Z^+_{n,0}(\mathbf{y}) = \frac{h^\beta}{\sqrt{g(\mathbf{y})}na_n h} \sum_{i,j=1}^n \varepsilon_{i,j} \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i,j}). \tag{28}$$

We now define the double-indexed partial sum $S_{k,l}$ of the $\varepsilon_{i,j}$ by

$$S_{k,l} := \sum_{i=1}^k \sum_{j=1}^l \varepsilon_{i,j}, \quad S_{k,l} \equiv 0 \quad \text{if} \quad i \cdot j = 0.$$

Using the identity

$$\varepsilon_{i,j} = S_{i,j} - S_{i-1,j} + S_{i-1,j-1} - S_{i,j-1},$$

we obtain

$$\begin{aligned} Z^+_{n,0}(\mathbf{y}) = \frac{h^\beta}{\sqrt{g(\mathbf{y})}na_n h} \Bigg\{ & \sum_{i,j=1}^n S_{i,j} \big(\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i+1,j+1}) - \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i+1,j}) + \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i,j}) - \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i,j+1}) \big) \\ & - \sum_{i=0}^{n-1} S_{i,n} \big(\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i+1,n}) - \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i,n}) \big) \\ & - \sum_{j=0}^{n-1} S_{n,j} \big(\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{n,j+1}) - \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{n,j}) \big) \\ & + S_{n,n} \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{n,n}) \Bigg\}. \end{aligned}$$

Next, we make use of the following three identities

$$\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i+1,j+1}) - \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i+1,j}) + \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i,j}) - \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i,j+1}) = \int_{[\mathbf{x}_{i,j},\mathbf{x}_{i+1,j+1}]} \partial^{(1,1)}_{\mathbf{z}} \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{z})\, d\mathbf{z},$$

$$\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i+1,n}) - \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{i,n}) = \int_{[x_i,x_{i+1}]} \partial^{(0,1)}_{\mathbf{z}} \widetilde{\mathcal{K}}_{n,\mathbf{y}}\Big(\frac{1}{a_n}, z_2\Big)\, dz_2,$$

and

$$\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{n,j+1}) - \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{n,j}) = \int_{[x_j,x_{j+1}]} \partial_{\mathbf{z}}^{(1,0)} \widetilde{\mathcal{K}}_{n,\mathbf{y}}\Big(z_1, \frac{1}{a_n}\Big)\, dz_1.$$

Let W be the Wiener sheet defined in Lemma 4. We obtain

$$\begin{aligned}\big|Z_{n,0}^{+}(\mathbf{y}) - Z_{n,1}^{+}(\mathbf{y})\big| &\le \frac{h^\beta}{\sqrt{g(\mathbf{y})}na_n h} \sup_{0\le i,j\le n} \big|S_{i,j} - W(i,j)\big| \Big\{ \int_{[0,1/a_n]^2} |\partial_{\mathbf{z}}^{(1,1)} \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{z})|\, d\mathbf{z} \\ &+ \int_{[0,1/a_n]} |\partial_{\mathbf{z}}^{(0,1)} \widetilde{\mathcal{K}}_{n,\mathbf{y}}\Big(\frac{1}{a_n}, z_2\Big)|\, dz_2 + \int_{[0,1/a_n]} |\partial_{\mathbf{z}}^{(1,0)} \widetilde{\mathcal{K}}_{n,\mathbf{y}}\Big(z_1, \frac{1}{a_n}\Big)|\, dz_1 + |\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{n,n})| \Big\}\end{aligned}$$

By Lemma 3 we find the estimate

$$\begin{aligned}\big|Z_{n,0}^{+}(\mathbf{y}) - Z_{n,1}^{+}(\mathbf{y})\big| &\le \frac{1}{na_n h} \sup_{0\le i,j\le n} \big|S_{i,j} - W(i,j)\big| C(g^*, G^*) \big[\|\partial^{(1,1)} K\|_1 \\ &+ \|\partial^{(1,0)} K\|_1 + \|\partial^{(0,1)} K\|_1 + \|K\|_\infty + O\big(\sqrt{\log(n) h}\big)\big],\end{aligned}$$

where all the estimates on the right hand side are independent of the variable $\mathbf{y}$. Hence, by Lemma 4

$$\sup_{\mathbf{y}\in[0,1]^2} \big|Z_{n,0}^{+}(\mathbf{y}) - Z_{n,1}^{+}(\mathbf{y})\big| = O\Big(\frac{n^\mu}{na_n h}\Big) \quad a.s.,$$

for some arbitrarily small $\mu > 0$. Since, by the assumptions of Theorem 1, $na_n h^{1+\beta} \to \infty$ as $n \to \infty$, $\beta \ge 1$, we can find a positive constant μ such that for some small constant $\nu > 0$

$$\frac{n^\mu}{na_n h} = o\big(n^{-\nu}\big),$$

which concludes the proof of this lemma. $\square$

Step 2 of the proof of Theorem 1 : $Z_{n,1}^{+} \longrightarrow Z_{n,2}^{+}$

Lemma 8. *Under the assumptions of Theorem 1*

$$\sup_{\mathbf{y}\in[0,1]^2} \big|Z_{n,1}^{+}(\mathbf{y}) - Z_{n,2}^{+}(\mathbf{y})\big| = O_P\big(\sqrt{\log(n)}\big).$$

Proof. Integration by parts and scaling of the Wiener sheet gives

$$\begin{aligned}Z_{n,2}^{+}(\mathbf{y}) = \frac{h^\beta}{a_n h\sqrt{g(\mathbf{y})}} \Big\{ &\int_{[\mathbf{x}_{i,j}, \mathbf{x}_{i+1,j+1}]} \partial_{\mathbf{z}}^{(1,1)} \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{z}) W(a_n \mathbf{z})\, d\mathbf{z} \\ &- \int_{[x_i,x_{i+1}]} \partial_{\mathbf{z}}^{(0,1)} \widetilde{\mathcal{K}}_{n,\mathbf{y}}\Big(\frac{1}{a_n}, z_2\Big) W(1, a_n z_2)\, dz_2 \\ &- \int_{[x_j,x_{j+1}]} \partial_{\mathbf{z}}^{(1,0)} \widetilde{\mathcal{K}}_{n,\mathbf{y}}\Big(z_1, \frac{1}{a_n}\Big) W(a_n z_1, 1)\, dz_1 \\ &+ W(1,1) \widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{n,n}) \Big\},\end{aligned}$$

since $W(i,j)=0$ if $i\cdot j=0$. Hence

$$Z_{n,2}^{+}(\mathbf{y})-Z_{n,1}^{+}(\mathbf{y})=\frac{h^{\beta}}{a_n h\sqrt{g(\mathbf{y})}}\sum_{i,j=1}^{n}\int_{[\mathbf{x}_{i,j},\mathbf{x}_{i+1,j+1}]}\partial_{\mathbf{z}}^{(1,1)}\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{z})\big(W(i/n,j/n)-W(a_n\mathbf{z})\big)\,d\mathbf{z}$$
$$-\sum_{i=0}^{n-1}\int_{[x_i,x_{i+1}]}\partial_{\mathbf{z}}^{(0,1)}\widetilde{\mathcal{K}}_{n,\mathbf{y}}\Big(\frac{1}{a_n},z_2\Big)\big(W(1,a_nz_2)-W(i/n,1)\big)\,dz_2$$
$$-\sum_{j=0}^{n-1}\int_{[x_j,x_{j+1}]}\partial_{\mathbf{z}}^{(1,0)}\widetilde{\mathcal{K}}_{n,\mathbf{y}}\Big(z_1,\frac{1}{a_n}\Big)\big(W(1,j/n)-W(a_nz_1,1)\big)\,dz_1\Big\}.$$

By the modulus of continuity of the Wiener sheet (see ?, Theorem 3.2.1) we obtain

$$\limsup_{n\to\infty}\sup_{\substack{\mathbf{s},\mathbf{t}\in[0,1]^2\\ |s_i-t_i|<\frac{1}{n},i=1,2}}\frac{|W(\mathbf{s})-W(\mathbf{t})|}{\sqrt{\frac{1}{n}\log(n)}}\le 48.$$

Thus

$$\big|Z_{n,1}^{+}(\mathbf{y})-Z_{n,2}^{+}(\mathbf{y})\big|\le\frac{h^{\beta}\sqrt{\log(n)}}{g(\mathbf{y})\sqrt{n}a_nh}\sup_{\substack{\mathbf{s},\mathbf{t}\in[0,1]^2\\ |s_i-t_i|<\frac{1}{n},i=1,2}}\frac{|W(\mathbf{s})-W(\mathbf{t})|}{\sqrt{\frac{1}{n}\log(n)}}\Big\{\int_{[0,1/a_n]^2}|\partial_{\mathbf{z}}^{(1,1)}\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{z})|\,d\mathbf{z}$$
$$+\int_{[0,1/a_n]}|\partial_{\mathbf{z}}^{(0,1)}\widetilde{\mathcal{K}}_{n,\mathbf{y}}\Big(\frac{1}{a_n},z_2\Big)|\,dz_2+\int_{[0,1/a_n]}|\partial_{\mathbf{z}}^{(1,0)}\widetilde{\mathcal{K}}_{n,\mathbf{y}}\Big(z_1,\frac{1}{a_n}\Big)|\,dz_1+|\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{x}_{n,n})|\Big\}.$$

Similar to the proof of Lemma 7 we conclude

$$\big|Z_{n,1}^{+}(\mathbf{y})-Z_{n,2}^{+}(\mathbf{y})\big|=O_P\Big(\frac{\sqrt{\log(n)}}{ha_n\sqrt{n}}\Big)=o_P\Big(\frac{1}{\sqrt{\log(n)}}\Big).$$

□

Step 3 of the proof of Theorem 1: $Z_{n,2}^{+}\longrightarrow Z_{n,3}^{+}$

Lemma 9. *Under the assumptions of Theorem 1*

$$\sup_{\mathbf{y}\in[0,1]^2}|Z_{n,2}^{+}(\mathbf{y})-Z_{n,3}^{+}(\mathbf{y})|=O_P(\sqrt{\log(n)}h)$$

Proof. Recall that

$$Z_{n,2}^{+}(\mathbf{y}):=\frac{h^{\beta}}{h\sqrt{g(\mathbf{y})}}\int_{[0,1/a_n]^2}\widetilde{\mathcal{K}}_{n,\mathbf{y}}(\mathbf{z})\,dW(\mathbf{z})$$
$$=\frac{h^{\beta}}{h\sqrt{g(\mathbf{y})}\|K\|_2}\int_{[0,1/a_n]^2}\sqrt{g(\mathbf{z})}K_n\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}).$$

With the definition

$$G(\mathbf{y},\mathbf{z}):=\sqrt{g(\mathbf{y})}+\frac{z_1-y_1}{2\sqrt{g(\mathbf{y})}}\partial^{(1,0)}g(\mathbf{y})+\frac{z_2-y_2}{2\sqrt{g(\mathbf{y})}}\partial^{(0,1)}g(\mathbf{y})$$

we obtain the following decomposition of $Z_{n,2}(\mathbf{y})$:

$$\begin{aligned} Z_{n,2}(\mathbf{y}) &= \frac{h^\beta}{h\sqrt{g(\mathbf{y})}\|K\|_2}\int_{[0,1/a_n]^2} G(\mathbf{y},\mathbf{z})K_n\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}) \\ &+ \frac{h^\beta}{h\sqrt{g(\mathbf{y})}\|K\|_2}\int_{[0,1/a_n]^2} \big(\sqrt{g(\mathbf{z})}-G(\mathbf{y},\mathbf{z})\big)K_n\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}) \\ &=: Z^+_{n,2,I}(\mathbf{y}) + Z^+_{n,2,II}(\mathbf{y}), \end{aligned}$$

where $Z^+_{n,2,I}(\mathbf{y})$ and $Z^+_{n,2,II}(\mathbf{y})$ are defined in an obvious manner. Notice that

$$\begin{aligned} Z^+_{n,2,I}(\mathbf{y}) = Z^+_{n,3}(\mathbf{y}) &+ \frac{h^\beta\partial^{(1,0)}g(\mathbf{y})}{2\sqrt{g(\mathbf{y})}\|K\|_2}\int_{[0,1/a_n]^2}\frac{z_1-y_1}{h}K_n\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}) \\ &+ \frac{h^\beta\partial^{(0,1)}g(\mathbf{y})}{2\sqrt{g(\mathbf{y})}\|K\|_2}\int_{[0,1/a_n]^2}\frac{z_2-y_2}{h}K_n\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}). \end{aligned}$$

Further

$$\begin{aligned} Z^+_{n,2,I}(\mathbf{y}) = Z^+_{n,3}(\mathbf{y}) &+ \frac{\partial^{(1,0)}g(\mathbf{y})\|K\|_2}{2\sqrt{g(\mathbf{y})}}\int_{[0,1/a_n]^2}\frac{z_1-y_1}{h}\big(h^\beta K_n - K\big)\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}) \\ &+ \frac{\partial^{(0,1)}g(\mathbf{y})}{2\sqrt{g(\mathbf{y})}\|K\|_2}\int_{[0,1/a_n]^2}\frac{z_2-y_2}{h}\big(h^\beta K_n - K\big)\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}) \\ &+ \frac{\partial^{(1,0)}g(\mathbf{y})}{2\sqrt{g(\mathbf{y})}\|K\|_2}\int_{[0,1/a_n]^2}\frac{z_1-y_1}{h}K\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}) \\ &+ \frac{\partial^{(0,1)}g(\mathbf{y})}{2\sqrt{g(\mathbf{y})}\|K\|_2}\int_{[0,1/a_n]^2}\frac{z_2-y_2}{h}K\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}) \\ =: Z^+_{n,3}(\mathbf{y}) &+ \frac{h^\beta}{2\sqrt{g(\mathbf{y})}}\big(Z^{+,1}_{n,2,I}(\mathbf{y}) + Z^{+,2}_{n,2,I}(\mathbf{y}) + Z^{+,3}_{n,2,I}(\mathbf{y}) + Z^{+,4}_{n,2,I}(\mathbf{y})\big). \end{aligned}$$

Define

$$\widetilde{Z}_{n,2,3}(\mathbf{y}) := \frac{CG^*}{\|K\|h}\int_{\mathbb{R}^2}\mathcal{F}I_{[-1,1]^2}\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}),$$

where C is the constant defined in Lemma 3 (v). Then, since $I_{[-1,1]^2}$ is compactly supported, all derivatives of $\mathcal{F}I_{[-1,1]^2}$ exist and are square-integrable and $|\mathcal{F}(\mathcal{F}I_{[-1,1]^2})| = |I_{[-1,1]^2}| \le 1$. By Corollary 2.2.8 in **?** we find

$$\mathbb{E}\Big[\sup_{\mathbf{y}\in[0,1]^2}|\widetilde{Z}_{n,2,3}(\mathbf{y})|\Big] = O\big(\sqrt{\log(n)}\big).$$

Further, by the Plancherel equality

$$\mathbb{E}\big|\widetilde{Z}_{n,2,3}(\mathbf{s}) - \widetilde{Z}_{n,2,3}(\mathbf{t})\big|^2 = \frac{C^2}{\|K\|_2}\int_{[-1,1]^2}\Big|\exp\Big(i\boldsymbol{\xi}^T\frac{\mathbf{t}}{h}\Big) - \exp\Big(-i\boldsymbol{\xi}^T\frac{\mathbf{s}}{h}\Big)\Big|^2\,d\boldsymbol{\xi}.$$

$$\mathbb{E}\big|Z^{+,1}_{n,2,I}(\mathbf{s}) - Z^{+,1}_{n,2,I}(\mathbf{t})\big|^2 \le \mathbb{E}\big|\widetilde{Z}_{n,2,3}(\mathbf{s}) - \widetilde{Z}_{n,2,3}(\mathbf{t})\big|^2 \quad \forall\, \mathbf{s},\mathbf{t}\in[0,1]^2,$$

by Lemma 3 (v) since

$$\Big|\frac{z_1-y_1}{h}\big(h^\beta K_n - K\big)\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\Big| = \Big|\frac{1}{(2\pi)^3}\int \partial^{(0,1)}\Big(\frac{\mathcal{F}\eta(\boldsymbol{\xi})}{\mathcal{F}\psi\big(\frac{\boldsymbol{\xi}}{h}\big)} - \mathcal{F}\eta(\boldsymbol{\xi})\Psi(\boldsymbol{\xi})\Big)\exp\Big(-i\boldsymbol{\xi}^T\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,d\boldsymbol{\xi}\Big|.$$

Hence, by the Sudakov-Fernique comparison inequality (see, e. g. **?**, Theorem 2.2.3) it follows that

$$\mathbb{E}\big[\sup_{\mathbf{y}\in[0,1]^2}|Z^{+,1}_{n,2,I}(\mathbf{y})|/h\big] \le \mathbb{E}\big[\sup_{\mathbf{y}\in[0,1]^2}|\widetilde{Z}_{n,2,3}(\mathbf{y})|\big] = O\big(\sqrt{\log(n)}\big).$$

Thus

$$\sup_{\mathbf{y}\in[0,1]^2}|Z^{+,1}_{n,2,I}(\mathbf{y})| = O_P\big(\sqrt{\log(n)}h\big).$$

Along the same line of arguments we obtain

$$\sup_{\mathbf{y}\in[0,1]^2}|Z^{+,j}_{n,2,I}(\mathbf{y})| = O_P\big(\sqrt{\log(n)}h\big)\ j=2,3,4,$$

by comparing $Z^{+,3}_{n,2,I}$ and $Z^{+,4}_{n,2,I}$ to the Gaussian fields

$$\frac{G^*}{\|K\|_2}\int\frac{z_j-y_j}{h}K\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z})\ j=1,2.$$

By Assumption 1 the derivatives $\partial^{\boldsymbol{\alpha}}g,\ |\boldsymbol{\alpha}|=1$ and the function $1/\sqrt{g}$ are Lipschitz-continuous. Thus, there exists a constant D such that

$$\big|\sqrt{g(\mathbf{z})} - G(\mathbf{y},\mathbf{z})\big| \le D\big(|z_1-y_1| + |z_2-y_2|\big).$$

The assertion of the lemma now follows by a repetition of the steps for the estimation of $Z^{+}_{n,2,I}(\mathbf{y})$ for the field $Z^{+}_{n,2,II}(\mathbf{y})$.

□

Step 4 of the proof of Theorem 1: $Z^+_{n,3} \longrightarrow Z^+_{n,4}$

Lemma 10.

$$\sup_{\mathbf{y}\in[0,1]^2}|Z^+_{n,3}(\mathbf{y}) - Z^+_{n,4}(\mathbf{y})| = O_P\big(h\sqrt{\log(n)}\big)$$

Proof. Define

$$\widetilde{Z}_{n,3,4}(\mathbf{y}) := \frac{C_E}{\|K\|}\int_{\mathbb{R}^2}\eta\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}),$$

where η is the function defined in Assumption 2. Then $\|\mathcal{F}\eta\|\le 1$ and, since $\mathcal{F}\eta$ is compactly supported, η is smooth and each derivative $\partial^{\boldsymbol{\alpha}}\eta$, $\boldsymbol{\alpha}\in\mathbb{N}_0^2$ is square integrable. Again, by Corollary 2.2.8 in **?**,

$$\mathbb{E}\big[\sup_{\mathbf{y}\in[0,1]^2}|\widetilde{Z}_{n,3,4}(\mathbf{y})|\big] = O\big(\sqrt{\log(n)}h\big).$$

and

$$
\begin{aligned}
\widetilde{E}_{n,\mathbf{s},\mathbf{t}} &:= \mathbb{E}\big|\widetilde{Z}_{n,3,4}(\mathbf{s}) - \widetilde{Z}_{n,3,4}(\mathbf{t})\big|^2 \\
&= \frac{C_E^2}{\|K\|}\int_{\mathbb{R}^2}\Big|\eta\Big(\frac{\mathbf{s}-\mathbf{z}}{h}\Big) - \eta\Big(\frac{\mathbf{t}-\mathbf{z}}{h}\Big)\Big|^2\,d\mathbf{z} \\
&= \frac{h^2C_E^2}{\|K\|}\int_{\mathbb{R}^2}\Big|\mathcal{F}\eta(\mathbf{z})\Big(\exp\Big(i\xi^T\frac{\mathbf{t}}{h}\Big) - \exp\Big(-i\xi^T\frac{\mathbf{s}}{h}\Big)\Big)\Big|^2\,d\mathbf{z},
\end{aligned}
$$

where the last equality follows from the Plancherel identity.

$$
\begin{aligned}
E_{\mathbf{s},\mathbf{t}} &:= \mathbb{E}\big|Z_{n,2}^+(\mathbf{s}) - Z_{n,3}^+(\mathbf{s}) - (Z_{n,2}^+(\mathbf{t}) - Z_{n,3}^+(\mathbf{t}))\big|^2 \\
&= \frac{1}{h^2\|K\|}\int_{\mathbb{R}^2}\Big|(h^\beta K_n - K)\Big(\frac{\mathbf{s}-\mathbf{z}}{h}\Big) - (h^\beta K_n - K)\Big(\frac{\mathbf{t}-\mathbf{z}}{h}\Big)\Big|^2\,d\mathbf{z} \\
&= \frac{1}{\|K\|}\int_{\mathbb{R}^2}\Big|(h^\beta K_n - K)\Big(\frac{\mathbf{s}}{h}-\mathbf{z}\Big) - (h^\beta K_n - K)\Big(\frac{\mathbf{t}}{h}-\mathbf{z}\Big)\Big|^2\,d\mathbf{z}.
\end{aligned}
$$

As before, the Plancherel identity yields

$$
\begin{aligned}
E_{\mathbf{s},\mathbf{t}} &= \frac{1}{\|K\|}\int_{\mathbb{R}^2}\Big|\mathcal{F}\eta(\mathbf{z})\Big(\frac{h^\beta}{\mathcal{F}\Psi(\frac{\xi}{h})} - \Psi(\xi)\Big)\Big(\exp\Big(i\xi^T\frac{\mathbf{t}}{h}\Big) - \exp\Big(-i\xi^T\frac{\mathbf{s}}{h}\Big)\Big)\Big|^2\,d\boldsymbol{\xi} \\
&\le \frac{h^2C_E^2}{\|K\|}\int_{\mathbb{R}^2}\Big|\mathcal{F}\eta(\mathbf{z})\Big(\exp\Big(i\xi^T\frac{\mathbf{t}}{h}\Big) - \exp\Big(-i\xi^T\frac{\mathbf{s}}{h}\Big)\Big)\Big|^2\,d\boldsymbol{\xi} = \widetilde{E}_{n,\mathbf{s},\mathbf{t}},
\end{aligned}
$$

where the last estimate follows from Lemma 3 (i). Since $E_{n,\mathbf{s},\mathbf{t}} \le \widetilde{E}_{n,\mathbf{s},\mathbf{t}}$ for all $\mathbf{s},\mathbf{t} \in [0,1]^2$ and $\mathbb{E}[\widetilde{Z}_{n,3,4}(\mathbf{y})] = \mathbb{E}[Z_{n,3}^+(\mathbf{y}) - Z_{n,4}^+(\mathbf{y})] = 0$ for all $\mathbf{y} \in [0,1]^2$ it follows by the Sudakov-Fernique comparison inequality (see, e. g. **?**, Theorem 2.2.3)

$$
\mathbb{E}\big[\sup_{\mathbf{y}\in[0,1]^2}|Z_{n,3}^+(\mathbf{y}) - Z_{n,4}^+(\mathbf{y})|\big] \le \mathbb{E}\big[\sup_{\mathbf{y}\in[0,1]^2}|\widetilde{Z}_{n,3,4}(\mathbf{y})|\big] = O\big(h\sqrt{\log(n)}\big),
$$

which implies

$$
\sup_{\mathbf{y}\in[0,1]^2}|Z_{n,3}^+(\mathbf{y}) - Z_{n,4}^+(\mathbf{y})| = O_P\big(h\sqrt{\log(n)}\big).
$$

□

Step 5 of the proof of Theorem 1: $Z_{n,4}^+ \longrightarrow Z_{n,5}^+$

Lemma 11. *Under the assumptions of Theorem 1*

$$
\sup_{\mathbf{y}\in[0,1]^2}|Z_{n,4}^+(\mathbf{y}) - Z_{n,5}^+(\mathbf{y})| = O_P\big(ha_n\sqrt{\log(n)}\big).
$$

Proof. Define

$$
\begin{aligned}
\Delta_{n,4,5}(\mathbf{y}) &:= Z_{n,5}^+(\mathbf{y}) - Z_{n,4}^+(\mathbf{y}) \\
&= \frac{1}{h\|K\|_2}\int\big(I_{[0,\infty)}(\mathbf{y}) - I_{[0,1/a_n]^2}(\mathbf{z})\big)K\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,dW(\mathbf{z}).
\end{aligned}
$$

Then

$$\mathbb{E}\big|\Delta_{n,4,5}(\mathbf{s}) - \Delta_{n,4,5}(\mathbf{t})\big|^2 \le \frac{1}{h^2\|K\|_2^2}\int \sum_{j=1}^{2} I_{(1/a_n,\infty)}(y_j)\Big|K\Big(\frac{\mathbf{s}-\mathbf{z}}{h}\Big) - K\Big(\frac{\mathbf{t}-\mathbf{z}}{h}\Big)\Big|^2 d\mathbf{z}.$$

If $y_j > 1/a_n$ and $s_j \in [0,1],\ j = 1,2$ we have

$$\Big|\frac{s_j - y_j}{h}\Big| > \frac{1}{2a_n h}$$

for sufficiently large n. This implies

$$2a_n h\Big|\frac{s_j - y_j}{h}\Big| > 1,$$

which yields the estimate

$$\begin{aligned}\mathbb{E}\big|\Delta_{n,4,5}(\mathbf{s}) - \Delta_{n,4,5}(\mathbf{t})\big|^2 &\le \frac{4a_n^2}{\|K\|_2^2}\int \sum_{j=1}^{2}\Big|\frac{s_j-y_j}{h}\Big|^2\Big|K\Big(\frac{\mathbf{s}-\mathbf{z}}{h}\Big) - K\Big(\frac{\mathbf{t}-\mathbf{z}}{h}\Big)\Big|^2 d\mathbf{z}\\ &\le \int\Big|\partial^{(0,1)}\big(\mathcal{F}\eta(\mathbf{z})\Psi(\mathbf{z})\big) + \partial^{(1,0)}\big(\mathcal{F}\eta(\mathbf{z})\Psi(\mathbf{z})\big)\Big|^2\Big|\exp\Big(i\xi^T\frac{\mathbf{t}}{h}\Big) - \exp\Big(-i\xi^T\frac{\mathbf{s}}{h}\Big)\Big|^2 dz.\end{aligned}$$

Again, by Gaussian comparison (Theorem 2.2.3, **?**) with the Gaussian field

$$\tilde{Z}_{n,4,5}(\mathbf{y}) := \frac{2a_n}{\|K\|}\int_{\mathbb{R}^2}\mathcal{F}\big(\partial^{(0,1)}\big(\mathcal{F}\eta\Psi\big) + \partial^{(1,0)}\big(\mathcal{F}\eta\Psi\big)\big)(\boldsymbol{\xi})\Big(\frac{\mathbf{y}-\boldsymbol{\xi}}{h}\Big)\,d\boldsymbol{\xi}\,dW(\boldsymbol{\xi})$$

and a further application of Corollary 2.2.8 in **?** we conclude the proof of this lemma. □

Step 6 of the proof of Theorem 1: Negligibility of the remainder

Lemma 12. *Under the assumptions of Theorem 1*

$$\sup_{\mathbf{y}\in[0,1]^2}\Big|\frac{h^\beta}{\|K\|_2\sqrt{g(\mathbf{y})}na_n h}\sum_{(i,j)\in\mathcal{R}_n}(Y_{(i,j)} - g(\mathbf{x}_{ij}))K_n\Big(\frac{\mathbf{y}-\mathbf{x}_{ij}}{h}\Big)\Big| = O\Big(\frac{1}{\sqrt{n}a_n h}\Big) \quad a.s.$$

Proof. Define

$$\epsilon_i := Y_{i,0} - g(\mathbf{x}_{i,0}) \quad \text{and} \quad S_i := \sum_{j=1}^{i}\epsilon_i, \quad i = 1,\dots,n$$

and write

$$R_n(\mathbf{y}) = \frac{h^\beta}{\sqrt{g(\mathbf{y})}}\sum_{i=1}^{n}(Y_{i,0} - g(\mathbf{x}_{i,0})K_n\Big(\frac{\mathbf{y}-\mathbf{x}_{i,0}}{h}\Big) = \frac{h^\beta}{\sqrt{g(\mathbf{y})}}\sum_{i=1}^{n}\epsilon_i K_n\Big(\frac{\mathbf{y}-\mathbf{x}_{i,0}}{h}\Big).$$

It follows that

$$\begin{aligned}\sup_{\mathbf{y}\in[0,1]^2}|R_n(\mathbf{y})| &\le \frac{1}{\sqrt{g^*}}\sup_{\mathbf{y}\in[0,1]^2}\Big|\sum_{i=1}^{n}\epsilon_i\big(h^\beta K_n - K\big)\Big(\frac{y_1-x_i}{h},\frac{y_2}{h}\Big)\Big|\\ &\quad + \frac{1}{\sqrt{g^*}}\sup_{\mathbf{y}\in[0,1]^2}\Big|\sum_{i=1}^{n}\epsilon_i K\Big(\frac{y_1-x_i}{h},\frac{y_2}{h}\Big)\Big|.\end{aligned}$$

Furthermore, since K and $1/h(h^\beta K_n - K)$ are uniformly bounded, it follows for some constant $C > 0$

$$\sup_{\mathbf{y}\in[0,1]^2} |R_n(\mathbf{y})| \le C \sup_{0\le c_1\le\ldots\le c_n\le 1} \Big|\sum_{i=1}^n c_i\epsilon_i\Big| = C \sup_{\max_{1\le j\le n}} |S_j|$$
$$\le C(\sup_{\max_{1\le j\le n}} |S_j - \tilde{S}_j|) + C \sup_{\max_{1\le j\le n}} |\tilde{S}_j|,$$

where $\tilde{S}_j$ is a partial sum of independent and identically standard normal random variables as in Lemma 4 below. Since the increments $\tilde{S}_j$ are symmetric, it follows by Lévy's maximal inequality for all $\lambda > 0$

$$\mathbb{P}\big(\max_{1\le j\le n} |\tilde{S}_j| > \lambda\big) \le 2\mathbb{P}\big(|\tilde{S}_n| > \lambda\big).$$

Hence, by Lemma 4, for any small $\mu > 0$

$$\sup_{\mathbf{y}\in[0,1]^2} |R_n(\mathbf{y})| = O\Big(\frac{1}{\sqrt{n}a_n h}\Big) + O\Big(\frac{1}{n^{1-\mu}a_n h}\Big) a.s. = O\Big(\frac{1}{\sqrt{n}a_n h}\Big) \quad a.s.$$

□

Step 7 of the proof of Theorem 1: Combining the results from steps 1-6

Together, Lemmas 7 to 11 yield

$$\sup_{\mathbf{y}\in[0,1]^2} \Big| Z_{n,0}^+(\mathbf{y}) - \frac{1}{h\|K\|_2} \int_{[0,\infty)^2} K\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\, dW(\mathbf{z})\Big| = o_P\Big(\log(n)^{-\frac{1}{2}}\Big). \tag{29}$$

Let $\widetilde{W}$ be a continuous version of the Wiener sheet in the above integral approximation. Set $\widetilde{W}(\mathbf{z}) = 0$ if $z_1 \vee z_2 < 0$. Let

$$\big\{\widetilde{W}_{\boldsymbol{\alpha}} \,|\, \boldsymbol{\alpha} \in \{0,1\}^2\big\}$$

be four mutually independent copies of $\widetilde{W}$. For $\mathbf{z} \in \mathbb{R}^2$, $\boldsymbol{\alpha} \in \{0,1\}^2$ define

$$W_{\boldsymbol{\alpha}}(\mathbf{z}) := \widetilde{W}_{\boldsymbol{\alpha}}((-1)^{\alpha_1} z_1, (-1)^{\alpha_2} z_2).$$

With the same arguments used to prove (29) and an application of Lemma 12 we find for the full process $Z_{n,0}$

$$\sup_{\mathbf{y}\in[0,1]^2} \Big| Z_{n,0}(\mathbf{y}) - \frac{1}{h\|K\|_2} \int_{\mathbb{R}^2} K\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\, dW(\mathbf{z})\Big| = o_P\Big(log(n)^{-\frac{1}{2}}\Big),$$

where $W(\mathbf{z}) = \sum_{\boldsymbol{\alpha}\in\{0,1\}^2} W_{\boldsymbol{\alpha}}(\mathbf{z})$ is a Wiener sheet on $\mathbb{R}^2$. By the scaling-property of the integral process we find

$$\sup_{\mathbf{y}\in[0,1]^2} \Big| \frac{1}{h} \int_{\mathbb{R}^2} K\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\, dW(\mathbf{z})\Big| = \sup_{\mathbf{y}\in\frac{1}{h}[0,1]^2} \Big| \int_{\mathbb{R}^2} K(\mathbf{y}-\mathbf{z})\, dW(\mathbf{z})\Big|.$$

The assertion of Theorem 1 now follows from an application of Theorem 14.2 of **?**.

□

7.2 Proof of Theorem 2

The proof of Theorem 2 essentially follows the line of proof of Theorem 1. The only difference is in the estimation of the bias of the estimator $\hat{f}_n$. In order to estimate the bias we use the representation of the function f in equation (30) below and replace the difference $1 - \mathcal{F}\eta(\boldsymbol{\xi})$ by the expansion

$$1 - \mathcal{F}\eta(\boldsymbol{\xi}) = \mathcal{F}\eta(\mathbf{0}) - \mathcal{F}\eta(\boldsymbol{\xi}) = \xi_1 \frac{\partial}{\partial \xi_1}\mathcal{F}\eta(\mathbf{0}) + \xi_2 \frac{\partial}{\partial \xi_2}\mathcal{F}\eta(\mathbf{0}), + \sum_{\substack{\boldsymbol{\alpha} \in \{0,1,2\}^2 \\ |\boldsymbol{\alpha}|=2}} \frac{\boldsymbol{\xi}^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} \frac{\partial^{\boldsymbol{\alpha}}}{\partial \xi_1^{\alpha_1} \partial \xi_2^{\alpha_2}} \mathcal{F}\eta(\boldsymbol{\xi}) \bigg|_{\boldsymbol{\xi}=\boldsymbol{\xi}^*},$$

where $\boldsymbol{\xi}^*$ is an intermediate point and $\mathbf{0} = (0,0)^T$. Here, we make use of the fact that

$$\frac{\partial^{\boldsymbol{\alpha}}}{\partial \xi_1^{\alpha_1} \partial \xi_2^{\alpha_2}} \mathcal{F}\eta(\boldsymbol{\xi}) = (-i)^{\boldsymbol{\alpha}} \mathcal{F}\big(\cdot^{\boldsymbol{\alpha}} \eta(\cdot)\big)(\boldsymbol{\xi})$$

and that

$$\|\mathcal{F}\big(\cdot^{\boldsymbol{\alpha}} \eta(\cdot)\big)\|_\infty \leq \| \cdot^{\boldsymbol{\alpha}} \eta\|_1 \leq C$$

for some $C < \infty$ and for all $|\boldsymbol{\alpha}| \leq 2$. Thus,

$$\begin{aligned}
|1 - \mathcal{F}\eta(\boldsymbol{\xi})| &\leq |\xi_1| \big|\mathcal{F}\big(t_1 \eta(\cdot)\big)(\mathbf{0})\big| + |\xi_2| \big|\mathcal{F}\big(t_2 \eta(\cdot)\big)(\mathbf{0})\big| + 2C(|\xi_1|^2 + |\xi_2|^2 + |\xi_1 \xi_2|) \\
&\leq |\xi_1| \left| \int t_1 \eta(\mathbf{t})\, d\mathbf{t} \right| + |\xi_2| \left| \int t_2 \eta(\mathbf{t})\, d\mathbf{t} \right| + 2C(|\xi_1|^2 + |\xi_2|^2 + |\xi_1 \xi_2|).
\end{aligned}$$

By symmetry of η we thus find

$$|1 - \mathcal{F}\eta(\boldsymbol{\xi})| \leq 2C(|\xi_1|^2 + |\xi_2|^2 + |\xi_1 \xi_2|).$$

Furthermore,

$$\begin{aligned}
&\frac{1}{(2\pi)^3 h^2} \left| \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \frac{g(\mathbf{z})}{\mathcal{F}\psi\Big(\frac{\boldsymbol{\xi}}{h}\Big)} \Big(1 - \mathcal{F}\eta(\boldsymbol{\xi})\Big) \exp\Big(i \boldsymbol{\xi}^T \frac{\mathbf{y} - \mathbf{z}}{h}\Big)\, d\mathbf{z}\, d\boldsymbol{\xi} \right| \\
&\leq \frac{1}{(2\pi)^3 h^2} \int_{\mathbb{R}^2} \frac{\xi_1^2 + \xi_2^2 + |\xi_1 \xi_2|}{\Big|\mathcal{F}\psi\Big(\frac{\boldsymbol{\xi}}{h}\Big)\Big|} \left| \int_{\mathbb{R}^2} g(\mathbf{z}) \exp\Big(i \boldsymbol{\xi}^T \frac{\mathbf{y} - \mathbf{z}}{h}\Big)\, d\mathbf{z} \right| d\boldsymbol{\xi}.
\end{aligned}$$

By a change of variables

$$\begin{aligned}
&\frac{1}{(2\pi)^3 h^2} \left| \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \frac{g(\mathbf{z})}{\mathcal{F}\psi\Big(\frac{\boldsymbol{\xi}}{h}\Big)} \Big(1 - \mathcal{F}\eta(\boldsymbol{\xi})\Big) \exp\Big(i \boldsymbol{\xi}^T \frac{\mathbf{y} - \mathbf{z}}{h}\Big)\, d\mathbf{z}\, d\boldsymbol{\xi} \right| \\
&\frac{1}{(2\pi)^3} \int_{\mathbb{R}^2} \frac{h^2 \xi_1^2 + h^2 \xi_2^2 + h^2 |\xi_1 \xi_2|}{|\mathcal{F}\psi(\boldsymbol{\xi})|} \left| \int_{\mathbb{R}^2} g(\mathbf{z}) \exp\Big(i \boldsymbol{\xi}^T (\mathbf{y} - \mathbf{z})\Big)\, d\mathbf{z} \right| d\boldsymbol{\xi} \\
&= \frac{h^2}{2\pi} \int_{\mathbb{R}^2} \frac{\xi_1^2 + \xi_2^2 + |\xi_1 \xi_2|}{|\mathcal{F}\psi(\boldsymbol{\xi})|} |\mathcal{F}\psi(\boldsymbol{\xi})| |\mathcal{F}f(\boldsymbol{\xi})|\, d\boldsymbol{\xi} = O(h^2)
\end{aligned}$$

by assumption. The claim of the Theorem now follows by Theorem 1 and Lemma 2.

$\square$

Proof of Lemma 1

Since $0 < g^* \leq \hat{g}_n \leq G^*$ by definition, conditionally on the data $\mathcal{Y}$, the bootstrap residuals $\epsilon^*_{i,j} := Y^*_{i,j} - \hat{g}(\mathbf{x}_{i,j})$ are distributed according to a centered Poisson distribution for which all moments exist, because higher moments of the Poisson distribution are polynomials in the parameter, and are uniformly bounded with respect to the indices (i, j). Hence, the same arguments as in Lemma 6 apply and the assertion of the lemma follows. □

Proof of Theorem 3

Define the bootstrap analogue $Z^{+*}_{n,0}$ of $Z^{+}_{n,0}$ by

$$Z^{+*}_{n,0}(\mathbf{y}) := \frac{h^\beta}{\|K\|_2\sqrt{\hat{g}_n(\mathbf{y})}na_nh}\sum_{i,j=1}^{n}\big(Y^*_{(i,j)} - \hat{g}_n(\mathbf{x}_{i,j})\big)K_n\Big(\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}\Big)$$
$$= \frac{h^\beta}{\|K\|_2\sqrt{\hat{g}_n(\mathbf{y})}na_nh}\sum_{i,j=1}^{n}\sqrt{\hat{g}_n(\mathbf{x}_{i,j})}\varepsilon^*_{i,j}K_n\Big(\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}\Big).$$

Since $0 < g^* \leq \hat{g}_n \leq G^*$ by definition, conditionally on the data $\mathcal{Y}$, the bootstrap residuals $\varepsilon^*_{i,j} := 1/\hat{g}(\mathbf{x}_{i,j})(Y^*_{i,j} - \hat{g}(\mathbf{x}_{i,j}))$ are centered, independent with unit variances, where all moments exist and are uniformly bounded with respect to the indices (i, j). Hence, by Lemma 4, there exists a Wiener field W^*, defined conditionally on the sample $\mathcal{Y}$, such that

$$|S_{(i,j)^*} - W^*(i,j)| = o^*_P(n^\mu) \quad a.s.$$

With the same approximation steps as in the proof of Theorem 1 we obtain that $Z^{+*}_{n,0}$ can be approximated by

$$\frac{1}{h\|K\|_2}\int_{\mathbb{R}^2}K\Big(\frac{\mathbf{y}-\mathbf{x}}{h}\Big)\,dW^*(\mathbf{z}) + o^*_P\big((\log(n))^{-\frac{1}{2}}\big)$$

uniformly with respect to $\mathbf{y} \in [0,1]^2$. We further conclude that

$$\mathbb{P}\Big(\sup_{\mathbf{y}\in[0,1]^2}|Z^*_{n,5}(\mathbf{y})| \leq \kappa \,\big|\, \mathcal{Y}\Big) = \mathbb{P}\Big(\sup_{\mathbf{y}\in[0,1]^2}|Z_{n,5}(\mathbf{y})| \leq \kappa\Big)$$

for all $\kappa \in \mathbb{R}$. An application of Theorem 1 concludes the proof of this theorem.

□

7.3 Proofs of the auxiliary results

Proof of Lemma 2

Recall from Remark 3 (iv) that some useful properties of the reciprocal $1/\mathcal{F}\psi$ of the Fourier transform of the function ψ, which are implied by Assumption 3, are listed in Lemma 3 in Section 6. These results can be used to derive convergence properties of the sequence of kernels $(K_n)_{n\in\mathbb{N}}$ which allow for the definition of a "limit kernel" (see (19)).
(i) Recall that

$$\hat{f}_n(\mathbf{y}) = \frac{1}{(2\pi)^2n^2a_n^2h^2}\sum_{i,j=-n}^{n}Y_{i,j}K_n\Big(\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}\Big),$$

where

$$K_n(\mathbf{z}) = \frac{1}{(2\pi)^3} \int_{\mathbb{R}^2} \frac{\mathcal{F}\eta(\boldsymbol{\xi})}{\mathcal{F}\psi\left(\frac{\boldsymbol{\xi}}{h}\right)} \exp(i\boldsymbol{\xi}^T \mathbf{z})\, d\mathbf{z}$$

and $Y_{i,j} \sim$Poisson$\big(g(\mathbf{x}_{i,j})\big)$ such that

$$\mathbb{E}\left[\hat{f}_n(\mathbf{y})\right] = \frac{1}{(2\pi)^2 n^2 a_n^2 h^2} \sum_{i,j=-n}^{n} g(\mathbf{x}_{i,j}) K_n\Big(\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}\Big).$$

We now express f in terms of its Fourier transform $\mathcal{F}f$ and make use of the relation

$$\mathcal{F}f = \frac{1}{2\pi}\frac{\mathcal{F}g}{\mathcal{F}\psi},$$

which yields

$$\begin{aligned}
f(\mathbf{y}) &= \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \frac{\mathcal{F}g(\boldsymbol{\xi})}{\mathcal{F}\psi(\boldsymbol{\xi})} \exp(i\boldsymbol{\xi}^T \mathbf{y})\, d\boldsymbol{\xi} \\
&= \frac{1}{(2\pi)^3} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \frac{g(\mathbf{z})}{\mathcal{F}\psi(\boldsymbol{\xi})} \exp(i\boldsymbol{\xi}^T(\mathbf{y}-\mathbf{z}))\, d\mathbf{z}\, d\boldsymbol{\xi} \\
&= \frac{1}{(2\pi)^3 h^2} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \frac{g(\mathbf{z})}{\mathcal{F}\psi\left(\frac{\boldsymbol{\xi}}{h}\right)} \Big(\mathcal{F}\eta(\boldsymbol{\xi}) + 1 - \mathcal{F}\eta(\boldsymbol{\xi})\Big) \exp\Big(i\boldsymbol{\xi}^T \frac{\mathbf{y}-\mathbf{z}}{h}\Big)\, d\mathbf{z}\, d\boldsymbol{\xi}.
\end{aligned}$$

Hence

$$\begin{aligned}
f(\mathbf{y}) &= \frac{1}{(2\pi)^3 h^2} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \frac{g(\mathbf{z})}{\mathcal{F}\psi\left(\frac{\boldsymbol{\xi}}{h}\right)} \mathcal{F}\eta(\boldsymbol{\xi}) \exp\Big(i\boldsymbol{\xi}^T \frac{\mathbf{y}-\mathbf{z}}{h}\Big)\, d\mathbf{z}\, d\boldsymbol{\xi} \\
&+ \frac{1}{(2\pi)^3 h^2} \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \frac{g(\mathbf{z})}{\mathcal{F}\psi\left(\frac{\boldsymbol{\xi}}{h}\right)} \Big(1 - \mathcal{F}\eta(\boldsymbol{\xi})\Big) \exp\Big(i\boldsymbol{\xi}^T \frac{\mathbf{y}-\mathbf{z}}{h}\Big)\, d\mathbf{z}\, d\boldsymbol{\xi}. \qquad (30)
\end{aligned}$$

Further, with the definition $A_n := [-1/a_n, 1/a_n]^2$ we write

$$\begin{aligned}
f(\mathbf{y}) &= \frac{1}{(2\pi)^3 h^2} \int_{A_n} g(\mathbf{z}) K_n\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\, d\mathbf{z} \\
&+ \frac{1}{(2\pi)^3 h^2} \int_{A_n^C} g(\mathbf{z}) K_n\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\, d\mathbf{z} \\
&+ \frac{1}{(2\pi)^3 h^2} \int_{D^C} \int_{\mathbb{R}^2} \frac{g(\mathbf{z})}{\mathcal{F}\psi\left(\frac{\boldsymbol{\xi}}{h}\right)} \Big(1 - \mathcal{F}\eta(\boldsymbol{\xi})\Big) \exp\Big(i\boldsymbol{\xi}^T \frac{\mathbf{y}-\mathbf{z}}{h}\Big)\, d\mathbf{z}\, d\boldsymbol{\xi} \\
&=: f_1(\mathbf{y}) + f_2(\mathbf{y}) + f_3(\mathbf{y}),
\end{aligned}$$

where f_1 and f_2 are defined in an obvious manner. To derive the latter decomposition we used that $1 - \mathcal{F}\eta(\boldsymbol{\xi}) = 0$ for all $\boldsymbol{\xi} \in D$ by Assumption 2.

$$|f_3(\mathbf{y})| = \frac{1}{2\pi}\left| \int_{\frac{1}{h}D^C} \mathcal{F}f(\boldsymbol{\xi})(1 - \mathcal{F}\eta(\boldsymbol{\xi})) \exp(i\mathbf{y}^T\boldsymbol{\xi})\, d\boldsymbol{\xi} \right| \le \frac{1}{\pi} \int_{\frac{1}{h}D^C} |\mathcal{F}f(\boldsymbol{\xi})|\, d\boldsymbol{\xi} = O\big(h^{S_f}\big)$$

by Assumption 1. By Lemma 3 (ii) and (vi) we have

$$\|K_n - K\|_p = O(h) \quad \text{if} \quad p \in [2,\infty) \quad \text{and} \quad \|(K_n - K)I_{A_n}\|_1 = O(h\sqrt{\log(n)}).$$

Straightforward calculations for the integral approximation yield

$$\mathbb{E}\big[\hat{f}_n(\mathbf{y})\big] = f_1(\mathbf{y}) + O\Big(\frac{1}{na_nh^{\beta}}\Big).$$

We now estimate f_2. Notice that K_n can be expressed as a Fourier transform as follows:

$$K_n(\mathbf{z}) = \overline{\mathcal{F}\bigg(\frac{\mathcal{F}\eta}{\mathcal{F}\psi(\cdot/h)}(\mathbf{z})\bigg)}.$$

(ii) For each double-index (i,j), for which the derivative $\partial^{(i,j)}\big(\mathcal{F}\eta/\mathcal{F}\psi(\cdot/h)\big)$ exists, we find

$$\big\|h^{\beta}\boldsymbol{\xi}^{(i,j)}K_n(\mathbf{z})\big\|_{\infty} = \big\|h^{\beta}\mathcal{F}\partial^{(i,j)}\big(\mathcal{F}\eta/\mathcal{F}\psi(\cdot/h)\big)\big\|_{\infty} \leq \frac{1}{2\pi}\big\|h^{\beta}\partial^{(i,j)}\big(\mathcal{F}\eta/\mathcal{F}\psi(\cdot/h)\big)\big\|_1 \tag{31}$$

by the Haussdorf-Young inequality. If $\gamma \in \mathbb{N}$ inequality (31) holds for all double-indices (i,j) with $i+j \leq S_{\mathcal{F}\eta}$. If $\gamma \notin \mathbb{N}$, $\gamma \geq 1$, inequality (31) holds if $i+j \leq \min\{S_{\mathcal{F}\eta}, \lfloor\gamma\rfloor\}$. Assertion (i) now follows by Assumption 2, Assumption 3 and claim (v) of Lemma 3 in Section 6.

For the variance of the estimator we obtain

$$\begin{aligned}
\mathrm{Var}\big[\hat{f}_n(\mathbf{y})\big] &= \frac{1}{(2\pi)^4n^4a_n^4h^4}\sum_{i,j=-n}^{n} g(\mathbf{x}_{i,j})K_n^2\Big(\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}\Big)\\
&= \frac{1}{(2\pi)^4n^2a_n^2h^4}\bigg(\int_{A_n} g(\mathbf{z})K_n^2\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,d\mathbf{z} + O\Big(\frac{1}{na_nh^{2\beta}}\Big)\bigg)\\
&= \frac{1}{(2\pi)^4n^2a_n^2h^4}\bigg(\int_{\mathbb{R}^2} g(\mathbf{z})K_n^2\Big(\frac{\mathbf{y}-\mathbf{z}}{h}\Big)\,d\mathbf{z} + O\Big(h^{-2\beta}a_n^{2S_{\mathcal{F}g}}(h^2a_n^2) + \frac{1}{na_nh^{2\beta}}\Big)\bigg)\\
&= \frac{1}{(2\pi)^4n^2a_n^2h^{2+2\beta}}\Big(g(\mathbf{y})\|K\|^2 + O\Big(h + a_n^{2S_{\mathcal{F}g}+2} + \frac{1}{na_nh^2}\Big)\Big).
\end{aligned}$$

□

Proof of Lemma 3

(i) By definition

$$\begin{aligned}
\frac{h^{\beta}}{\mathcal{F}\psi\big(\frac{\boldsymbol{\xi}}{h}\big)} &= h^{\beta'\gamma}\Big(P\Big(\frac{\boldsymbol{\xi}}{h}\Big)\Big)^{\gamma} = h^{\beta}\Bigg(\sum_{\substack{(i,j)\in\{0,\dots,\beta'\}^2\\ i+j\leq\beta'}} a_{i,j}\Big(\frac{\boldsymbol{\xi}}{h}\Big)^{(i,j)}\Bigg)^{\gamma}\\
&= \Bigg(\sum_{\substack{(i,j)\in\{0,\dots,\beta'\}^2\\ i+j=\beta'}} a_{i,j}\boldsymbol{\xi}^{(i,j)} + \sum_{\substack{(i,j)\in\{0,\dots,\beta'\}^2\\ i+j<\beta'}} a_{i,j}h^{\beta'-i-j}\boldsymbol{\xi}^{(i,j)}\Bigg)^{\gamma}.
\end{aligned}$$

Consider the function $F:\mathbb{R}\to\mathbb{R}$ with $F(x)=x^{\gamma}$, $\gamma\geq 1$. By Taylor's theorem

$$F(x) - F(x_0) = F'(x_0) + o(|x-x_0|)$$

and hence, for $x = h^\beta P(\boldsymbol{\xi}/h)$ and $x_0 = \Psi(\boldsymbol{\xi})$

$$\sup_{\boldsymbol{\xi}\in[-1,1]^2}\Big|\frac{h^\beta}{\mathcal{F}\psi(\boldsymbol{\xi})} - \Psi(\boldsymbol{\xi})\Big| = \sup_{\boldsymbol{\xi}\in[-1,1]^2} \gamma\Big|\sum_{\substack{(i,j)\in\{0,\ldots,\beta'\}^2\\ i+j=\beta'}} a_{i,j}\boldsymbol{\xi}^{(i,j)}\Big|^{\gamma-1}\Big|h^\beta P\Big(\frac{\boldsymbol{\xi}}{h}\Big) - \Psi(\boldsymbol{\xi})\Big| + o(h),$$

since there exists a constant $C < \infty$ such that $|x - x_0| \le Ch$ for all $\boldsymbol{\xi} \in [-1,1]^2$.

(ii) For $p \ge 2$ and q such that $1/p + 1/q = 1$ the Hausdorff-Young inequality yields

$$\begin{aligned}\|h^\beta K_n - K\|_p &= \Big\|\mathcal{F}\Big[\frac{h^\beta \mathcal{F}\eta}{\mathcal{F}\psi(\cdot/h)} - \mathcal{F}\eta\Psi\Big]\Big\|_p\\ &\le \frac{1}{(2\pi)^{\frac{2}{q}-1}}\Big\|\frac{h^\beta\mathcal{F}\eta}{\mathcal{F}\psi(\cdot/h)} - \mathcal{F}\eta\Psi\Big\|_q.\end{aligned}$$

The claim now follows from (i) since $\mathcal{F}\eta$ is bounded and compactly supported on $[-1,1]^2$.

(iii) For $\gamma \in \mathbb{N}$ we can assume without loss of generality that $\gamma = 1$ (see Remark 3). In this case all partial derivatives of order $S_{\mathcal{F}\eta}$ of $\mathcal{F}\eta$ and $1/\mathcal{F}\psi(\cdot/h)$ exist. An application of the general Leibniz rule gives

$$\partial^{\boldsymbol{\alpha}}\Big(\frac{\mathcal{F}\eta}{\mathcal{F}\psi(\cdot/h)}\Big) = \sum_{\substack{\{\boldsymbol{\alpha}':\boldsymbol{\alpha}'\le\boldsymbol{\alpha}\}\\ |\boldsymbol{\alpha}|\le S_{\mathcal{F}\eta}}} \binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}\partial^{\boldsymbol{\alpha}-\boldsymbol{\alpha}'}(\mathcal{F}\eta)\partial^{\boldsymbol{\alpha}'}\Big(\frac{1}{\mathcal{F}\psi(\cdot/h)}\Big),$$

where

$$\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'} = \frac{\boldsymbol{\alpha}!}{\boldsymbol{\alpha}'!(\boldsymbol{\alpha}-\boldsymbol{\alpha}')!} = \frac{\alpha_1!\alpha_2!}{\alpha_1'!\alpha_2'!(\alpha_1-\alpha_1')!(\alpha_2-\alpha_2')!}.$$

This yields

$$\partial^{\boldsymbol{\alpha}'}\Big(\frac{h^\beta}{\mathcal{F}\psi(\boldsymbol{\xi}/h)}\Big) = \sum_{\substack{(i,j)\in\{0,\ldots,\beta'\}^2\\ i+j=\beta'}} a_{i,j}\partial^{\boldsymbol{\alpha}'}\boldsymbol{\xi}^{(i,j)} + \sum_{\substack{(i,j)\in\{0,\ldots,\beta'\}^2\\ i+j<\beta'}} a_{i,j}\partial^{\boldsymbol{\alpha}'}h^{\beta'-i-j}\boldsymbol{\xi}^{(i,j)},$$

which is uniformly bounded in both $\boldsymbol{\xi} \in [-1,1]^2$ and $n \in \mathbb{N}$ for all multi-indices $\boldsymbol{\alpha}'$ with $|\boldsymbol{\alpha}'| \le S_{\mathcal{F}\eta}$.

(iv) Assertion (iv) follows from (iii) by multiple applications of the chain rule.

(v) The proof of (v) follows the line of proof of (iii) and (iv).

(vi) a) Consider first the case of $\boldsymbol{\alpha} = (0,0)$, that is

$$\begin{aligned}&\int_{[-1/(a_nh),1/(a_nh)]^2}\big|h^\beta K_n(\mathbf{z}) - K(\mathbf{z})\big|\,d\mathbf{z}\\ &= \int_{[-1/(a_nh),1/(a_nh)]^2}\big|h^\beta K_n(\mathbf{z}) - K(\mathbf{z})\big|\frac{\sqrt{1+z_1^2+z_2^2}}{\sqrt{1+z_1^2+z_2^2}}\,d\mathbf{z}\\ &\le \Big(\int_{\mathbb{R}^2}\big|h^\beta K_n(\mathbf{z}) - K(\mathbf{z})\big|^2(1+z_1^2+z_2^2)\,d\mathbf{z}\Big)^{\frac{1}{2}}\Big(\int_{[-1/(a_nh),1/(a_nh)]^2}\frac{1}{1+z_1^2+z_2^2}\,d\mathbf{z}\Big)^{\frac{1}{2}},\end{aligned}$$

By the Cauchy-Schwarz inequality and using polar coordinates we obtain

$$\int_{[-1/(a_nh),1/(a_nh)]^2} \frac{1}{1+z_1^2+z_2^2}\,d\mathbf{z} = 2\pi \int_{(0,1/(a_nh)]} \frac{r}{1+r^2}\,dr = \pi\log(1+(a_nh)^2) = O(\log(n)).$$

Next, the Plancherel identity and an application of (v) yields

$$\begin{aligned}\int_{\mathbb{R}^2} \big|h^\beta z_1 K_n(\mathbf{z}) - z_1 K(\mathbf{z})\big|^2\,d\mathbf{z} &= \int_{\mathbb{R}^2} \Big|h^\beta \partial^{(1,0)} \frac{\mathcal{F}\eta(\boldsymbol{\xi})}{\mathcal{F}\psi\big(\frac{\boldsymbol{\xi}}{h}\big)} - \partial^{(1,0)}\mathcal{F}\eta(\boldsymbol{\xi})\Psi(\boldsymbol{\xi})\Big|^2\,d\boldsymbol{\xi}\\ &\le C^2h^2 \int_{\mathbb{R}^2} I_{[-1,1]^2}(\boldsymbol{\xi})\,d\boldsymbol{\xi}.\end{aligned}$$

Combining the latter results, an application of (ii) finally yields

$$\int_{[-1/(a_nh),1/(a_nh)]^2} \big|h^\beta K_n(\mathbf{z}) - K(\mathbf{z})\big|\,d\mathbf{z} = O\big(h\sqrt{\log(n)}\big).$$

Since

$$\partial^{\boldsymbol{\alpha}} K_n = C_{\boldsymbol{\alpha}}\mathcal{F}^{-1}\Big(\frac{\boldsymbol{\xi}^{\boldsymbol{\alpha}}\mathcal{F}\eta(\boldsymbol{\xi})}{\mathcal{F}\psi\big(\frac{\boldsymbol{\xi}}{h}\big)}\Big) \quad \text{and} \quad \partial^{\boldsymbol{\alpha}} K = C_{\boldsymbol{\alpha}}\mathcal{F}^{-1}\Big(\boldsymbol{\xi}^{\boldsymbol{\alpha}}\mathcal{F}\eta(\boldsymbol{\xi})\Psi(\boldsymbol{\xi})\Big)$$

for a constant $C_{\boldsymbol{\alpha}} \in \mathbb{C}$, the same arguments as above apply to the integrals

$$\int_{[-1/(a_nh),1/(a_nh)]^2} \big|h^\beta \partial^{\boldsymbol{\alpha}} K_n(\mathbf{z}) - \partial^{\boldsymbol{\alpha}} K(\mathbf{z})\big|\,d\mathbf{z}, \quad |\boldsymbol{\alpha}| > 0,$$

since multiplication with $\boldsymbol{\xi}^{\boldsymbol{\alpha}}$ does neither change the smoothness properties nor the integrability properties of the functions $\frac{\mathcal{F}\eta(\boldsymbol{\xi})}{\mathcal{F}\psi\left(\frac{\boldsymbol{\xi}}{h}\right)}$ and $\mathcal{F}\eta(\boldsymbol{\xi})\Psi(\boldsymbol{\xi})$, by the fact that $\mathcal{F}\eta$ is compactly supported.

(vi) b) Consider first the case of $\boldsymbol{\alpha} = (0,0)$. Then

$$\begin{aligned}&\int_{\mathbb{R}} \big|h^\beta K_n((z_1,z_2)) - K(z_1,z_2)\big|\,dz_j\\ &= \Big(\int_{\mathbb{R}} \frac{1}{1+z_j^2}\,dz_j\Big)^{\frac{1}{2}} \Big(\int_{\mathbb{R}} \big|h^\beta K_n((z_1,z_2)) - K(z_1,z_2)\big|^2(1+z_j^2)\,dz_j\Big)^{\frac{1}{2}}.\end{aligned}$$

The claim now follows with the same arguments as used to prove (vi)a).

□

Proof of Lemma 4

For a fixed constant $\mu > 0$ let M be an even, (large) positive integer ($M \in 2\mathbb{N}$) such that $M \cdot \mu > 3$. By Assumption 1 we obtain

$$\|g^{\widetilde{M}}\|_1 \le \big(G^*\big)^{\widetilde{M}-1}\|g\|_1 < \infty \quad \text{for all} \quad \widetilde{M} \le M. \tag{32}$$

Since (32) holds true for all $M \in 2\mathbb{N}$ we can choose μ arbitrarily small. Further

$$\mathbb{E}\big[\varepsilon_{i,j}^M\big] = \mathbb{E}\Big[\frac{(Y_{k,l} - g(\mathbf{x}_{k,l}))^M}{\sqrt{g(\mathbf{x}_{k,l})^M}}\Big] = \frac{\mathbb{E}\big[(Y_{k,l} - g(\mathbf{x}_{k,l}))^M\big]}{\sqrt{g(\mathbf{x}_{k,l})^M}} = \frac{P_{\widetilde{M}}(g(\mathbf{x}_{i,j}))}{\sqrt{g(\mathbf{x}_{k,l})^M}}, \tag{33}$$

where $P_{\widetilde{M}}$ is some polynomial of degree $\widetilde{M} \leq M$.
Define

$$L_M := \sum_{i,j=1}^{n} \mathbb{E}\big|\varepsilon_{i,j}\big|^M = \sum_{i,j=1}^{n} \mathbb{E}\varepsilon_{i,j}^M,$$

since $M \in 2\mathbb{N}$. From relations (32) and (33) it follows that there exists a constant $C(g^*, G^*)$ such that

$$\frac{1}{n^2 a_n^2} L_M \leq C(g^*, G^*)\|g\|_1. \tag{34}$$

Let $\Phi : \{1, \ldots n^2\} \to \{1, \ldots n\}^2$ be a bijective map that satisfies the additional requirement $\Phi(k^2) = (k,k)$ for all $k \leq n$ and

$$\Phi\big(\{(k-1)^2 + 1, \ldots, k^2\}\big) = \{(i,j)\,|\,1 \leq i \leq k, j = k\} \cup \{(i,j)\,|\,1 \leq j < k, i = k\}.$$

This way the quadratic sums coincide

$$\sum_{j=1}^{k^2} Z_{\Phi(j)} = \sum_{i=1}^{k}\sum_{j=1}^{k} Z_{i,j}$$

just the order of summation might differ. By Corollary 5.4 of **?** there exists a field of independent identically $\mathcal{N}(0,1)$-distributed random variables $\{Z_{i,j}\,|\,(i,j) \in \mathbb{N}^2\}$ ordered in such a way that for

$$W(i,j) := \sum_{k=1}^{i}\sum_{l=1}^{j} Z_{k,l}$$

and $\nu > 0$

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{n} \varepsilon_{\Phi(i)} - \sum_{i=1}^{n} Z_{\Phi(i)}\Big| > CMn^{\nu}\Big) \leq \frac{L_M}{n^{n\cdot M}} + \mathbb{P}\Big(\max_{1\leq i\leq n}\big|\varepsilon_{\Phi(i)} - Z_{\Phi(i)}\big| > n^{\nu}\Big)$$

and consequently

$$\begin{aligned}
&\mathbb{P}\Big(\big|S_{(n,n)} - W(n,n)\big| > CMn^{2\nu}\Big) \\
&\quad = \mathbb{P}\Big(\big|S_{(\Phi(n^2))} - W(\Phi(n^2))\big| > CMn^{2\nu}\Big) \leq \frac{L_M}{n^{2\nu\cdot M}} + \mathbb{P}\Big(\max_{1\leq i\leq n^2} |\varepsilon_{\Phi(i)} - Z_{\Phi(i)}| > n^{2\nu}\Big).
\end{aligned} \tag{35}$$

We now estimate the second term on the right hand side of inequality (35). We set $\mu := 2\nu$ and find

$$\begin{aligned}
\mathbb{P}\Big(\max_{1\leq i\leq n^2} |\varepsilon_{\Phi(i)} - Z_{\Phi(i)}| > n^{\mu}\Big) &\leq \mathbb{P}\Big(\max_{1\leq i\leq n^2} |\varepsilon_{\Phi(i)}| + \max_{1\leq i\leq n^2} |Z_{\Phi(i)}| > n^{\mu}\Big) \\
&\leq \mathbb{P}\Big(\max_{1\leq i\leq n^2} |\varepsilon_{\Phi(i)}| > \frac{n^{\mu}}{2}\Big) + \mathbb{P}\Big(\max_{1\leq i\leq n^2} |Z_{\Phi(i)}| > \frac{n^{\mu}}{2}\Big).
\end{aligned}$$

Since the $Z_{i,j}$ are *iid* standard normally distributed with distribution function ϕ we find

$$\mathbb{P}\Big(\max_{1\le i\le n^2}|Z_{\Phi(i)}|>\frac{n^\mu}{2}\Big)=1-\Big(2\phi(n^\mu/2)-1\Big)^{n^2}\sim 1-\Big(2-\frac{2}{\sqrt{2\pi}}\frac{n^\mu}{2}\exp\Big(-\frac{n^{2\mu}}{8}\Big)-1\Big)^{n^2}$$
$$=\sum_{k=1}^{n^2}\binom{n^2}{k}\frac{1}{n^{2k}}\Big(-\frac{n^{\mu+2}}{\sqrt{2\pi}}\exp\Big(-\frac{n^{2\mu}}{8}\Big)\Big)^k\le n^{\mu+2}\exp\Big(-\frac{n^{2\mu}}{8}\Big)\sum_{k=1}^{n^2}\binom{n^2}{k}\frac{1}{n^{2k}}$$
$$\le e\cdot n^{\mu+2}\exp\Big(-\frac{n^{2\mu}}{8}\Big)\le\frac{1}{2n^2}$$

for sufficiently large n. Furthermore

$$\mathbb{P}\Big(\max_{1\le i\le n^2}|\varepsilon_{\Phi(i)}|\le\frac{n^\mu}{2}\Big)=\mathbb{P}\Big(|\varepsilon_{\Phi(i)}|\le\frac{n^\mu}{2}\ \text{ for all }\ 1\le i\le n^2\Big)$$
$$=\mathbb{P}\bigg(\frac{|Y_{\Phi(i)}-g(\mathbf{x}_{\Phi(i)})|}{\sqrt{g(\mathbf{x}_{\Phi(i)})}}\le\frac{n^\mu}{2}\ \text{ for all }\ 1\le i\le n^2\bigg).$$

Since $Y_{\Phi(i)}\ge 0$ for all $1\le i\le n^2$ we obtain for sufficiently large n

$$\mathbb{P}\Big(\max_{1\le i\le n^2}|\varepsilon_{\Phi(i)}|\le\frac{n^\mu}{2}\Big)=\mathbb{P}\bigg(0\le Y_{\Phi(i)}\le g(\mathbf{x}_{\Phi(i)})+\frac{n^\mu}{2}\sqrt{g(\mathbf{x}_{\Phi(i)})}\ \text{ for all }\ 1\le i\le n^2\bigg).$$

For independent and identically distributed random variables $\widetilde{Y}_{\Phi(i)}\sim$Poisson$(G^*)$ $1\le i\le n^2$ we find

$$\mathbb{P}\Big(\max_{1\le i\le n^2}|\varepsilon_{\Phi(i)}|\le\frac{n^\mu}{2}\Big)\ge\mathbb{P}\bigg(0\le\widetilde{Y}_{\Phi(i)}\le g(\mathbf{x}_{\Phi(i)})+\frac{n^\mu}{2}\sqrt{g(\mathbf{x}_{\Phi(i)})}\ \text{ for all }\ 1\le i\le n^2\bigg)$$
$$\ge\mathbb{P}\bigg(0\le\widetilde{Y}_{\Phi(i)}\le\frac{n^\mu}{2}\sqrt{g^*}\ \text{ for all }\ 1\le i\le n^2\bigg)$$
$$\ge\bigg(1-\mathbb{P}\bigg(\widetilde{Y}_{\Phi(1)}>\frac{n^\mu}{2}\sqrt{g^*}\bigg)\bigg)^{n^2},$$

since $\mathbb{P}(X_{\lambda_1}\le k)\le\mathbb{P}(X_{\lambda_2}\le k)$ for arbitrary numbers k if $X_{\lambda_j}\sim$ Poisson(λ_j), $j=1,2$ with $\lambda_1\ge\lambda_2$. By Theorem 5.4 in **?** we obtain the estimate

$$\mathbb{P}\bigg(\widetilde{Y}_{\Phi(1)}\ge\frac{n^\mu}{2}\sqrt{g^*}\bigg)\le e^{G^*}\bigg(\frac{2eG^*}{n^\mu\sqrt{g^*}}\bigg)^{\frac{n^\mu}{2}\sqrt{g^*}},$$

which yields

$$\mathbb{P}\Big(\max_{1\le i\le n^2}|\varepsilon_{\Phi(i)}|\le\frac{n^\mu}{2}\Big)\ge\bigg(1-e^{G^*}\bigg(\frac{2eG^*}{n^\mu\sqrt{g^*}}\bigg)^{\frac{n^\mu}{2}\sqrt{g^*}}\bigg)^{n^2}.$$

With the same arguments as those used in order to estimate the maximum of the independent, normally distributed random variables we obtain the estimate

$$\mathbb{P}\Big(\max_{1\le i\le n^2}|\varepsilon_{\Phi(i)}|\ge\frac{n^\mu}{2}\Big)\le n^2e^{G^*+1}\bigg(\frac{2eG^*}{n^\mu\sqrt{g^*}}\bigg)^{\frac{n^\mu}{2}\sqrt{g^*}}\le\frac{1}{2n^2}$$

for sufficiently large n. In relation (35) we now replace the term

$$\mathbb{P}\Big(\max_{1\le i\le n^2}|\varepsilon_{\Phi(i)}-Z_{\Phi(i)}|>n^{\mu}\Big)$$

by the estimate $1/n^2$, which gives

$$\mathbb{P}\Big(\big|S_{(n,n)}-W(n,n)\big|>CMn^{\mu}\Big)\le\frac{L_M}{n^{\mu\cdot M}}+\frac{1}{n^2}\le\frac{C(g^*,G^*)\|g\|_1a_n^2}{n^{\mu\cdot M-2}}+\frac{1}{n^2},$$

where we used (34) to obtain the last estimate. This yields

$$\big|S_{(n,n)}-W(n,n)\big|=O(n^{\mu})\ a.s.$$

Recall that our objective is to show that

$$\sup_{(i,j)\in\{1,\dots,n\}^2}\big|S_{(i,j)}-W(i,j)\big|=O(n^{\mu})\ \ a.s.$$

To this end we apply a fluctuation inequality by **?** that allows to compare the orders of magnitude of the quantities $\sup_{(i,j)\in\{1,\dots,n\}^2}\big|S_{(i,j)}-W(i,j)\big|$ and $|S(n,n)-W(n,n)|$ by an inequality of the form

$$\mathbb{P}\Big(\sup_{(i,j)\in\{1,\dots,n\}^2}\big|S_{(i,j)}-W(i,j)\big|>C_1\mu\Big)\le C_2\mathbb{P}\Big(|S(n,n)-W(n,n)|>C_3n^{\mu}\Big),$$

where the constants C_1, C_2 and C_3 do not depend on n. Since

$$\frac{S_{(n,n)}-W(n,n)}{\sqrt{\mathrm{Var}(S_{(n,n)}-W(n,n))}}=\frac{\sum_{i=1}^{n^2}(\varepsilon_{\Phi(i)}-Z_{\Phi(i)})}{\sqrt{\sum_{i=1}^{n^2}\mathrm{Var}(\varepsilon_{\Phi(i)}-Z_{\Phi(i)})}}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)\quad\text{as}\quad n\to\infty,$$

it follows that $\mathrm{Var}(S_{(n,n)}-W(n,n))/n^{2\mu}=O(1)$ for $n\to\infty$ and hence there exists a constant $\tilde{C}>0$ such that for all n

$$\tilde{C}\frac{\sqrt{\mathrm{Var}(S_{(n,n)}-W(n,n))}}{n^{\mu}}\le\frac{1}{2}.$$

Finally, an application of Lemma 3.1 of **?** yields

$$\begin{aligned}\mathbb{P}\Big(\sup_{\substack{1\le i\le n\\1\le j\le n}}\big|S_{i,j}-W(i,j)\big|>4\frac{n^{\mu}}{\tilde{C}}\Big)&\le\frac{16}{9}\mathbb{P}\Big(\big|S_{n,n}-W(n,n)\big|>CM\frac{n^{\mu}}{\tilde{C}}\Big)\\&\le\frac{16}{9}\Big(\frac{\tilde{C}C(g^*,G^*)\|g\|_1a_n^2}{n^{\mu\cdot M-2}}+\frac{1}{n^2}\Big),\end{aligned}$$

for large enough $M>3/\mu$. Since the latter estimation shows that the probability is summable the assertion of the lemma now immediately follows by Borel-Cantelli and the zero-one law.

□

Proof of Lemma 5

First, note that

$$\sup_{\mathbf{x}\in[0,1]^2} |Z_n(\mathbf{x})| \stackrel{\mathcal{D}}{=} \sup_{\mathbf{x}\in\frac{1}{h}[0,1]^2} |\widetilde{Z}(\mathbf{x})|,$$

where

$$\widetilde{Z}(\mathbf{x}) := \int K(\mathbf{x}-\mathbf{y})\, dW(\mathbf{y})$$

and that the covariance function r of the stationary Gaussian field $\widetilde{Z}$ is given by

$$r(\mathbf{t}) = \mathrm{Cov}\big(\widetilde{Z}(\mathbf{x}), \widetilde{Z}(\mathbf{x}+\mathbf{t})\big) = \int K(\mathbf{y})K(\mathbf{y}+\mathbf{t})\, d\mathbf{y}.$$

Hence, the function r is square-integrable if $\mathcal{F}K$ is bounded:

$$\int |r(\mathbf{t})|^2\, d\mathbf{t} = 4\pi^2 \int |\mathcal{F}K(\boldsymbol{\xi})|^4\, d\boldsymbol{\xi} \leq 4\pi^2 (K^*)^2 \int |\mathcal{F}K(\boldsymbol{\xi})|^2\, d\boldsymbol{\xi} < \infty.$$

The function r is integrable if $K \in L^1(\mathbb{R}^2)$

$$\int |r(\mathbf{t})|\, d\mathbf{t} \leq \|K\|_1^2 < \infty.$$

By assumption, the function r is three times differentiable in the mean-squared sense and thus an application of Theorem 14.2 of **?** yields the result of this lemma under either of the assumptions on the kernel K.

$\square$

Proof of Lemma 6

Note that we can express $\tilde{g}_n$ as

$$\tilde{g}_n(y_1, y_2) = \frac{1}{n^2 a_n^2 h^2} \sum_{i,j=-n}^{n} Y_{i,j}\, \eta\Big(\frac{y_1 - x_i}{h}, \frac{y_2 - x_j}{h}\Big),$$

that is, $\tilde{g}_n$ is a kernel-type estimator for the function g with kernel η. Define $M_n := \log(n)$ and $\epsilon_{i,j} := Y_{i,j} - \mathbb{E}[Y_{i,j}]$. Then we can write

$$\begin{aligned}\epsilon_{i,j} &= \big(\epsilon_{i,j} I\{|\epsilon_{i,j}| < M_n\} - \mathbb{E}[\epsilon_{i,j} I\{|\epsilon_{i,j}| < M_n\}]\big) + \big(\epsilon_{i,j} I\{|\epsilon_{i,j}| \geq M_n\} - \mathbb{E}[\epsilon_{i,j} I\{|\epsilon_{i,j}| \geq M_n\}]\big)\\ &:= \tau_{i,j} + \rho_{i,j},\end{aligned}$$

where $\tau_{i,j}$ and $\rho_{i,j}$ are defined in a obvious manner. We now first show that

$$\mathbb{P}\Big(\frac{1}{n^2 a_n^2 h^2} \sup_{\mathbf{y}\in\mathbb{R}^2} \sum_{i,j=1}^{n} |\rho_{i,j}| \Big|\eta\Big(\frac{\mathbf{y} - \mathbf{x}_{i,j}}{h}\Big)\Big| > \frac{1}{n}\Big) = O\Big(\frac{1}{n^2}\Big)$$

and subsequently that

$$\mathbb{P}\Big(\frac{1}{n^2a_n^2h^2}\sup_{\mathbf{y}\in\mathbb{R}^2}\sum_{i,j=1}^{n}|\tau_{i,j}|\Big|\eta\Big(\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}\Big)\Big|>\frac{\log(n)}{na_n^2h^2}\Big)=O\Big(\frac{1}{n^2}\Big).$$

Recall that $\|\mathcal{F}\eta\|_1\le 4$, which implies by the Hausdorff-Young inequality that $\|\eta\|_\infty\le 1/(2\pi)\|\mathcal{F}\eta\|_1\le 4/(2\pi)\le 1$. Hence, by an application of Markov's inequality, we find that

$$\begin{aligned}\mathbb{P}\Big(\frac{1}{n^2a_n^2h^2}\sup_{\mathbf{y}\in\mathbb{R}^2}\sum_{i,j=1}^{n}|\rho_{i,j}|\Big|\eta\Big(\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}\Big)\Big|>\frac{1}{n}\Big)&\le\frac{1}{na_n^2h^2}\sum_{i,j=-n}^{n}\mathbb{E}[|\rho_{i,j}|]\\&\le\frac{2}{na_n^2h^2}\sum_{i,j=-n}^{n}\mathbb{E}\big[|\epsilon_{i,j}|I\{|\epsilon_{i,j}|\ge M_n\}\big].\end{aligned}$$

Next, the Cauchy-Schwarz inequality yields the estimate

$$\mathbb{E}\big[|\epsilon_{i,j}|I\{|\epsilon_{i,j}|\ge M_n\}\big]\le\big(\mathbb{E}\big[|\epsilon_{i,j}|^2\big]\big)^{\frac12}\big(\mathbb{E}\big[I\{|\epsilon_{i,j}|\ge M_n\}\big]\big)^{\frac12}\le\sqrt{G^*}\big(\mathbb{P}\big(|\epsilon_{i,j}|\ge M_n\big)\big)^{\frac12}.$$

Further, since g is bounded and $M_n\to\infty$ as $n\to\infty$, we obtain for sufficiently large n

$$\mathbb{P}\big(|\epsilon_{i,j}|\ge M_n\big)=\mathbb{P}\big(\epsilon_{i,j}\ge M_n\big)=\mathbb{P}\big(Y_{i,j}\ge M_n+g(\mathbf{x}_{i,j})\big)\le\mathbb{P}\big(Y_{i,j}\ge M_n+g^*\big).$$

By Theorem 5.4 of **?** we obtain the estimate

$$\begin{aligned}\mathbb{P}\big(Y_{i,j}\ge M_n+g^*\big)&\le\frac{\exp(-g(\mathbf{x}_{i,j}))\big(eg(\mathbf{x}_{i,j})\big)^{g^*+M_n}}{\big(M_n+g^*\big)^{M_n+g^*}}\le\Big(\frac{G^*}{M_n+g^*}\Big)^{g^*}\Big(\frac{eG^*}{M_n+g^*}\Big)^{M_n}\\&\le\Big(\frac{G^*}{M_n}\Big)^{g^*}\Big(\frac{eG^*}{M_n}\Big)^{M_n}\le\Big(\frac{G^*}{M_n}\Big)^{g^*}\frac{1}{n^{10}}\end{aligned}$$

for sufficiently large n. This yields

$$\mathbb{P}\Big(\frac{1}{n^2a_n^2h^2}\sup_{\mathbf{y}\in\mathbb{R}^2}\sum_{i,j=1}^{n}|\rho_{i,j}|\Big|\eta\Big(\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}\Big)\Big|>\frac{1}{n}\Big)\le\frac{2(2n+1)^2\sqrt{G^*}}{na_n^2h^2}\Big(\frac{G^*}{M_n}\Big)^{g^*/2}\frac{1}{n^5}=o\Big(\frac{1}{n^2}\Big).$$

Next, we consider

$$\check{g}_n(\mathbf{y}):=\frac{\log(n)}{n^2a_n^2h^2}\sum_{i,j=1}^{n}\check{\tau}_{i,j}\eta\Big(\frac{\mathbf{y}-\mathbf{x}_{i,j}}{h}\Big),$$

where $\check{\tau}_{i,j}:=\tau_{i,j}/\log(n)$. The random variables $\check{\tau}_{i,j}$ are independent, centered and bounded but not identically distributed. By Theorem 1 of **?** there exists a universal constant $C\in(0,\infty)$ and independent and identically distributed random variables $(\widetilde{\tau}_{i,j},\widetilde{\mathbf{x}}_{i,j})$ such that

$$\mathbb{P}\Big(\sup_{\mathbf{y}\in\mathbb{R}^2}|\check{g}_n(\mathbf{y})|>\lambda\Big)\le C\,\mathbb{P}\Big(\sup_{\mathbf{y}\in\mathbb{R}^2}|\widetilde{g}_n(\mathbf{y})|>\frac{\lambda}{C}\Big),\tag{36}$$

where

$$\widetilde{g}_n(\mathbf{y}):=\frac{\log(n)}{n^2a_n^2h^2}\sum_{i,j=1}^{n}\widetilde{\tau}_{i,j}\eta\Big(\frac{\mathbf{y}-\widetilde{\mathbf{x}}_{ij}}{h}\Big).$$

Note that now the design-points $\widetilde{\mathbf{x}}_{ij}$ are random. Define the class of functions

$$\mathcal{G} := \Big\{ f : \mathbb{R}^3 \to \mathbb{R} \,|\, f(z_1, z_2, z_3) = z_1 \cdot I\{|z_1| \le 1\} \cdot \eta\Big(\frac{\mathbf{y} - (z_2, z_3)^T}{h}\Big),\, h \in \mathbb{R}_+,\, \mathbf{y} \in \mathbb{R}^2 \Big\}.$$

By Assumption 4 and Lemma 22 of **?** the classes

$$\mathcal{G}_1 := \Big\{ f : \mathbb{R}^3 \to \mathbb{R} \,|\, f(z_1, z_2, z_3) = z_1 \cdot I\{|z_1| \le 1\} \Big\}$$

and

$$\mathcal{G}_2 := \Big\{ f : \mathbb{R}^3 \to \mathbb{R} \,|\, f(z_1, z_2, z_3) = \eta\Big(\frac{\mathbf{y} - (z_2, z_3)^T}{h}\Big),\, h \in \mathbb{R}_+,\, \mathbf{y} \in \mathbb{R}^2 \Big\}$$

are VC-classes of functions. Since $\mathcal{G} = \{f_1 \cdot f_2 \,|\, f_1 \in \mathcal{G}_1 \text{ and } f_2 \in \mathcal{G}_2\}$, by Lemma 2.14 of **?** also $\mathcal{G}$ is a VC-class of functions. Furthermore, $\mathcal{G}$ is measurable and uniformly bounded and we can apply Theorem 2.1 of **?** to find an estimate for the probability

$$\mathbb{P}\Big(\sup_{\mathbf{y} \in \mathbb{R}^2} |\widetilde{g}_n(\mathbf{y})| > \frac{\lambda}{C} \Big)$$

which also bounds the probability

$$\mathbb{P}\Big(\sup_{\mathbf{y} \in \mathbb{R}^2} |\breve{g}_n(\mathbf{y})| > \lambda \Big)$$

by eq. (36). Let $\Phi : \{1, \dots, (2n+1)^2\} \to \{-n, \dots, n\}^2$ be a bijective map. By Theorem 3.1 of **?** there exists a constant L depending only on the VC-characteristics of the set $\mathcal{G}$, such that

$$\begin{aligned}\mathbb{P}\Big(\sup_{\mathbf{y} \in \mathbb{R}^2} |\widetilde{g}_n(\mathbf{y})| > \frac{\log(n)}{Cna_n^2h^2} \Big) &= \mathbb{P}\Big(\sup_{f \in \mathcal{G}} \Big| \sum_{i=1}^{(2n+1)^2} f(\widetilde{\tau}_{\Phi(i)}, \widetilde{\mathbf{x}}_{\Phi(i)}) - \mathbb{E}[f(\widetilde{\tau}_{\Phi(i)}, \widetilde{\mathbf{x}}_{\Phi(i)})] \Big| > \frac{\log(n)}{Cna_n^2h^2} \Big) \\ &\le L \exp\Big(-\frac{n}{LC} \log\Big(1 + \frac{\log(n)^2}{LCn(2G^* + G^*/n + \log(n)/n)^2} \Big) \Big).\end{aligned}$$

Thus, for sufficiently large n we obtain the estimate

$$\mathbb{P}\Big(\sup_{\mathbf{y} \in \mathbb{R}^2} |\widetilde{g}_n(\mathbf{y})| > \frac{\log(n)}{Cna_n^2h^2} \Big) \le L \exp\Big(-\frac{2\log(n)^2}{L^2C^2(2G^* + G^*/n + \log(n)/n)^2} \Big) \le \exp(-2\log(n)).$$

With (36) this implies

$$\mathbb{P}\Big(\sup_{\mathbf{y} \in \mathbb{R}^2} |\breve{g}_n(\mathbf{y})| > \frac{\log(n)}{na_n^2h^2} \Big) = O\Big(\frac{1}{n^2}\Big),$$

which yields

$$\sup_{\mathbf{y} \in \mathbb{R}^2} |\breve{g}_n(\mathbf{y})| = O\Big(\frac{\log(n)}{na_n^2h^2}\Big) = o\Big(\frac{1}{\sqrt{\log(n)}}\Big) \text{ a.s.}$$

$\square$