ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2020)
82, *Part* 3, *pp.* 817–840

On bandwidth choice for spatial data density estimation

Zhenyu Jiang,

University of Southampton, UK

Nengxiang Ling,

Hefei University of Technology, People's Republic of China

Zudi Lu,

University of Southampton, UK

Dag Tjøstheim

University of Bergen, Norway

and Qiang Zhang

Beijing University of Chemical Technology, People's Republic of China

[Received March 2017. Final revision February 2020]

Summary. Bandwidth choice is crucial in spatial kernel estimation in exploring non-Gaussian complex spatial data. The paper investigates the choice of adaptive and non-adaptive bandwidths for density estimation given data on a spatial lattice. An adaptive bandwidth depends on local data and hence adaptively conforms with local features of the spatial data. We propose a spatial cross-validation (SCV) choice of a global bandwidth. This is done first with a pilot density involved in the expression for the adaptive bandwidth. The optimality of the procedure is established, and it is shown that a non-adaptive bandwidth choice comes out as a special case. Although the cross-validation idea has been popular for choosing a non-adaptive bandwidth in data-driven smoothing of independent and time series data, its theory and application have not been much investigated for spatial data. For the adaptive case, there is little theory even for independent data. Conditions that ensure asymptotic optimality of the SCV-selected bandwidth are derived, actually, also extending time series and independent data optimality results. Further, for the adaptive bandwidth with an estimated pilot density, oracle properties of the resultant density estimator are obtained asymptotically as if the true pilot were known. Numerical simulations show that finite sample performance of the SCV adaptive bandwidth choice works quite well. It outperforms the existing R routines such as the 'rule of thumb' and the so-called 'second-generation' Sheather–Jones bandwidths for moderate and big data sets. An empirical application to a set of spatial soil data is further implemented with non-Gaussian features significantly identified.

Keywords: Cross-validation; Kernel density estimation; Optimal bandwidth; Spatial lattice data; Spatially adaptive bandwidth choice

Address for correspondence: Zudi Lu, Statistical Sciences Research Institute and School of Mathematical Sciences, University of Southampton, University Road, Southampton, SO17 1BJ, UK.
E-mail: Z.Lu@soton.ac.uk

 1369–7412/20/82817

1. Introduction

Non-parametric methods have become increasingly popular in exploring spatially dependent complex data. In particular, voluminous geographic data have been, and continue to be, collected with modern data acquisition techniques such as Global Positioning Systems, high resolution remote sensing, location aware services and surveys, and Internet-based volunteered geographic information (Guo and Mennis, 2009). Therefore various data collected over the Earth's surface arise in different disciplines, including environmental science, econometrics, epidemiology, image analysis and oceanography, to list a few. Important applications and developments in the general area of spatial statistics, under linear and/or Gaussian assumptions, can be found widely; see, for example, Cressie (1993), Basawa (1996a, b), Guyon (1995) and Gelfand *et al.* (2010) for comprehensive reviews. However, it has been recognized in the literature that linear and/or Gaussian structures may quite often be violated, or serve only as a crude approximation (see Section 5.2). Non-parametric smoothing methods have provided a powerful methodology for circumventing these shortcomings and gaining insights into spatially dependent data; see, for example, Tran (1990), Hallin *et al.* (2001, 2004a, b, 2009), Zhu and Morgan (2004), Gao *et al.* (2006), Robinson (2008, 2011), Jenish (2012) and Lu and Tjøstheim (2014), among others.

Effective use of non-parametric smoothing methods requires the choice of a smoothing parameter (bandwidth) (see Jones *et al.* (1996)). Arguably, it is the most important aspect of non-parametric density estimation. In this paper, our objective is to develop an adaptive (i.e. local data dependent) as well as a fixed non-adaptive bandwidth choice for spatial data density estimation. The advantage of an adaptive bandwidth is that it is attempting to enhance local, or observationwise, smoothing, rather than a one bandwidth fits all type of smoothing. Here our attention is on spatial data observed on a lattice. Many agricultural experimental data, such as the soil data to be analysed in Section 5.2, and, in particular, with modern data acquisition techniques, remotely sensed data are on a regular grid (see Zhu *et al.* (2010)). For generality, consider the data to be the observations from $\{X_{\mathbf{i}} = (X_{\mathbf{i}}^{(1)}, X_{\mathbf{i}}^{(2)}, \ldots, X_{\mathbf{i}}^{(d)})\}$: a d-dimensional stationary random field with index $\mathbf{i} = (i_1, i_2, \ldots, i_N) \in \mathbb{Z}^N (N \geqslant 1)$, which is defined on a common probability space $(\Omega, \mathscr{F}, \mathscr{P})$. We denote by f the common density function of $X_{\mathbf{i}}$ to be estimated, where a point $\mathbf{i} = (i_1, i_2, \ldots, i_N)$ will be referred to as a site in $\mathbb{Z}^N$ with $\mathbb{Z}$ the set of all integers. When $N = 1$, one may think of the time series data case or data on a line, including independent data, whereas $N = 2$ or $N = 3$ corresponds to data observed on a plane or three-dimensional space.

We shall propose generalizing the popular cross-validation (CV) idea (see Stone (1974)) to the choice of adaptive bandwidth, which includes the fixed non-adaptive bandwidth as a special case, for spatial lattice data. There are bandwidth selection methods that are theoretically justified based on resampling methods for spatial data. Such methods have targeted spectral density or variance estimation (see Nordman and Lahiri (2004)). But these developments are concerned with second-order moments or linear dependence properties, whereas we are seeking to give a theoretical justification with accompanying finite sample numerical experiments for the adaptive and non-adaptive bandwidth choice in general density-based non-parametric analysis of spatial data.

In the special case of density estimation for independent or time series data, bandwidth selection is well known to be a key question when applying any non-parametric smoother, and this question has been addressed in many papers. According to Jones *et al.* (1996), these methods can be categorized into two generations.

CV methods (see Silverman (1986), Fan and Gijbels (1996), Fan and Yao (2003) and Gao (2007)) have been classified as 'first-generation' methods that also include Silverman's (1986) rule-of-thumb bandwidth choice. Non-adaptive CV bandwidth selection has been extensively studied with independent and time series data; see, for example, Hall (1983), Stone (1984), Marron and Härdle (1986), Marron (1985, 1987), Hart and Vieu (1990), Scott (1992) and Kim and Cox (1997), and the references therein. In the present paper, we propose to extend this idea to adaptive density estimation for spatial lattice data.

From the perspective of optimal selection of a non-adaptive bandwidth, many so-called 'second-generation' bandwidth selection methods (e.g. Sheather and Jones (1991) and Marron (1992)) have been proposed. These methods are basically plug-in or bootstrap based, which rely on selection of pilot bandwidths (often by rule of thumb) and have been shown to perform better than the first-generation methods for independent data; see Sheather and Jones (1991), Cao *et al.* (1994), Chiu (1996) and Jones *et al.* (1996) for excellent reviews. However, we have not seen any second-generation methods for adaptive density estimation. This may be because the asymptotic mean integrated squared error (MISE) for the adaptive density estimator is more involved than that for the standard density estimator. It therefore may appear that CV is more natural than the alternative second-generation procedures for choosing the adaptive bandwidth. In view of our use of a pilot density, the spatial CV-based adaptive bandwidth choice in this paper may be seen as a second-generation method.

The main contributions of this paper are summarized: first, we shall propose a spatial bandwidth choice by generalizing CV to dependent lattice data density estimation. Our proposed method works for an adaptive bandwidth choice, but also functions for a non-adaptive bandwidth choice as a special case. Second, conditions that ensure asymptotic optimality of the spatial CV-selected bandwidth in terms of various error measures are derived, actually, also extending time series optimality results. Further, taking a non-adaptive kernel density estimator (KDE) as a pilot for an adaptive estimate of spatial data, oracle properties for the resultant density estimate are obtained as if the pilot density were known. Third, numerical simulations carried out will demonstrate that finite sample performance of the proposed spatial adaptive CV bandwidth choice works quite well. It outperforms the existing non-adaptive R routines such as the 'rule of thumb' and the so-called 'second-generation' Sheather and Jones (1991) bandwidths both for moderate sizes of spatial samples and in particular for big spatial data sets. Our empirical application to a set of spatial soil data will further illustrate that non-Gaussian features of the data are more significantly identified by spatial adaptive density estimation.

These results require an essentially different theoretical approach with a modified, albeit quite mild, set of assumptions. Fundamentally, unlike time series data in which time is unidirectional from the past to the future, space is multidirectional. With spatial data, the fact that the spatial sites do not have a natural ordering makes the problem of CV bandwidth selection more challenging. Several conditions are needed for an asymptotic theory to be established. These are described in detail in appendix A in the on-line supplementary material, but for the convenience of the reader we summarize them here. First a spatial mixing condition is required. Because of the multidirectionality the spatial mixing conditions are more elaborate than the time series strong mixing condition, but by now there are some rather standard spatial mixing conditions as stated in the on-line appendix A. Further, there is a set of conditions on the spatial kernel function to be introduced in equation (2.1). This pertains to the symmetry, continuity, boundedness, differentiability and support (a compact support is assumed) of the kernel function, and an integrability condition on the convolution kernel. There is also a condition on the characteristic function of the kernel linking it to the characteristic function of the d-dimensional standard normal distribution. The order r of the kernel is related to the differentiability of the density function. The density function to be estimated is bounded, with a deviation from the density in the independent case, which is also bounded. Moreover, there is a boundedness condition on conditional densities. In addition there is a weight function to be introduced in equation (3.1), this function being bounded and integrable, and having a compact support with the density function being bounded away from zero on this support. Finally, the bandwidth is restricted to an interval $[a\tilde{\mathbf{n}}^{-1/(2r+d)}, b\tilde{\mathbf{n}}^{-1/(2r+d)}]$ for some constants a and b, $0<a<b<\infty$, and with $\tilde{\mathbf{n}}$ being the total number of observations. All these conditions are quite standard. More details and a discussion of the feasibility of the conditions are given in appendix A.

The organization of the paper is as follows. The basic idea on spatial adaptive density estimation for lattice data, using an adaptive bandwidth that involves a global bandwidth h_0 and an estimated pilot density $\hat{f}_{\mathbf{n}}$, will be introduced in Section 2. In Section 3, spatial CV (SCV) selection of the global bandwidth h_0 will be suggested, where it is noted that CV selection of the non-adaptive bandwidth $h=h_0$ can be seen as a special case. The conditions to ensure optimality results of the bandwidth that is selected by the SCV method, in terms of the integrated squared error (ISE), MISE and average squared error (ASE), will be established in Section 3.1 with a known pilot f. SCV selection for the global bandwidth h_0 with an estimated pilot $\hat{f}_{\mathbf{n}}$ and the oracle properties that are associated with spatial adaptive kernel density estimation will be presented in Section 3.2. In Section 4, some further discussions on data involving spatial trends and a potential extension to bandwidth choice for spatial regression will be provided. The numerical finite sample performances of these estimates by using both simulated and real spatial data sets are examined in Section 5. The technical conditions on spatial mixing and additional regularity assumptions for the optimality theorems will be collected in the on-line appendix A. The technical proofs of the main results together with other supplementary materials are relegated to appendix B. There some useful lemmas are given in appendix B1, where a moment inequality with lattice random fields established by extending that in Gao *et al.* (2008) will play an important role in the technical proof of the main theorems in appendix B2. Some supplementary figures and a table to Section 5.1 are collected in appendix B3. The code that is used in the paper is copied in appendix C in a separate text form and is available from `https://sites.google.com/site/zudiluwebsite/` whereas the data set of soil250 that is used is available from the R package `geoR` (see Ribeiro and Diggle (2016)).

The code is also available from

`https://rss.onlinelibrary.wiley.com/hub/journal/14679868/series-b-datasets.`

2. Spatial adaptive density estimation: basic principles

Throughout the paper, let the random field $\{X_{\mathbf{i}}, \mathbf{i}\in\mathbb{Z}^N\}$ be observed over a rectangular region defined by $\mathcal{I}_{\mathbf{n}}=\{\mathbf{i}=(i_1,i_2,\ldots,i_N)\in\mathbb{Z}^N | 1\leqslant i_k\leqslant n_k, k=1,2,\ldots,N\}$. Thus, the total sample size in $\mathcal{I}_{\mathbf{n}}$ is denoted as $\tilde{\mathbf{n}}=\Pi_{k=1}^N n_k$ for $\mathbf{n}=(n_1,n_2,\ldots,n_N)\in\mathbb{Z}^N$ with strictly positive integer coordinates $n_1,n_2,\ldots,n_N$. As in Hallin *et al.* (2004b), we write that $\mathbf{n}\to\infty$ if $\min_{1\leqslant k\leqslant N}\{n_k\}\to\infty$, without requiring $\max_{1\leqslant j,k\leqslant N}\{n_j/n_k\}\leqslant C$ for some $0<C<\infty$ given in Tran (1990), allowing for multidirectional convergence in the sample size.

The idea of adaptive density estimation is popular in application (see Davies and Hazelton (2010) and Lemke *et al.* (2015)), where the bandwidth (see Abramson (1982a, b) in the independent data case) is defined adaptively depending on the sample $X_{\mathbf{i}}=X_{(i_1,i_2,\ldots,i_N)}$ spatially in estimating $f(x)$:

$$\check{f}_{\mathbf{n}}(x)=\frac{1}{\tilde{\mathbf{n}}}\sum_{\substack{i_k=1\\ \forall k=1,2,\ldots,N}}^{n_k}\frac{1}{h^d_{(i_1,i_2,\ldots,i_N)}}K\left(\frac{x-X_{(i_1,i_2,\ldots,i_N)}}{h_{(i_1,i_2,\ldots,i_N)}}\right),\qquad x\in\mathbb{R}^d, \tag{2.1}$$

where the summation symbol stands for the N-fold summations $\Sigma_{i_1=1}^{n_1}\Sigma_{i_2=1}^{n_2}\dots\Sigma_{i_N=1}^{n_N}$, and $K: \mathbb{R}^d \to \mathbb{R}^+$ is a kernel function. Here the bandwidth, being location dependent and inversely related to the population density $f(\cdot)$, is taken as (see Abramson (1982a, b))

$$h_{\mathbf{i}} \equiv h_{(i_1,i_2,\dots,i_N)} \equiv h_{\mathbf{i}}(h_0; f, \delta) = \frac{h_0}{f(X_{(i_1,i_2,\dots,i_N)})^{\delta}\gamma_f}, \tag{2.2}$$

where $h_0 \equiv h_{0\mathbf{n}}$ is a smoothing multiplier referred to as the global bandwidth, and

$$\gamma_f = \left\{ \prod_{\substack{i_k=1\\ \forall k=1,2,\dots,N}}^{n_k} f(X_{(i_1,i_2,\dots,i_N)})^{-\delta} \right\}^{1/\tilde{\mathbf{n}}}. \tag{2.3}$$

In adaptive estimation for independent data, it has been suggested by Abramson (1982a, b) that $\delta = \frac{1}{2}$ is optimal when f is second-order differentiable. For an introductory exposition on adaptive methods, the reader is referred to, for example, Silverman (1986), Bowman and Azzalini (1997), Pagan and Ullah (1999) and Davies and Hazelton (2010). Inclusion of the geometric mean term γ_f in equation (2.2), as noted in Silverman (1986), is to free the bandwidth factors from dependence on the scale of the data, allowing the global bandwidth h_0 to be considered on the same scale as in the corresponding fixed non-adaptive bandwidth estimate of f that is defined below.

In application of the adaptive density estimator (2.1), we need to have a pilot density estimator of f required in equation (2.2) and resulting in a kernel density estimator $\hat{f}_{\mathbf{n}}$ that is used in equation (3.13) in Section 3.2. This pilot estimator will be chosen as the traditional KDE $\hat{f}_{\mathbf{n}}(x)$ of f, with a fixed bandwidth, which can be seen as a special case of estimator (2.1) with $\delta = 0$ in equations (2.2) and (2.3), defined (see Tran (1990)) as

$$\hat{f}_{\mathbf{n}}(x) = \frac{1}{\tilde{\mathbf{n}}h^d} \sum_{\substack{i_k=1\\ \forall k=1,2,\dots,N}}^{n_k} K\left(\frac{x - X_{(i_1,i_2,\dots,i_N)}}{h}\right), \qquad x \in \mathbb{R}^d, \tag{2.4}$$

where $h \equiv h_{\mathbf{n}}$ (depending on $\mathbf{n}$) is used to distinguish it from h_0 above, for a sequence of fixed bandwidths tending to 0 as $\mathbf{n} \to \infty$. The study of the asymptotic properties for the spatial nonparametric KDE $\hat{f}_{\mathbf{n}}(x)$ is an interesting problem in statistical inference. With a given bandwidth series, Tran (1990) may be the first to establish the asymptotic normality of multivariate KDEs and Carbon *et al.* (1996) for the kernel-type estimator with the convergence in L_1, under stationary spatial random fields satisfying mixing conditions. See also the similar issues that were investigated by Hallin *et al.* (2001, 2004a) under alternative conditions on the spatial processes, and Lu and Tjøstheim (2014) and Harel *et al.* (2016) for some more recent developments. In this paper, we shall focus on the spatial mixing processes, which will be introduced with assumptions in appendix A.

To make the spatial adaptive bandwidth $h_{\mathbf{i}}$ defined in equation (2.2) applicable to adaptive density estimation in equation (2.1), we need to select the global bandwidth h_0 and to have $\hat{f}_{\mathbf{n}}$ in equation (2.4) with a selected bandwidth h as a pilot estimate for f in equations (2.2) and (2.3). Therefore the selection of h_0 as well as h is what we are concerned with in the next section, and selection of h in $\hat{f}_{\mathbf{n}}$ (see equation (2.4)) can be seen as the selection of $h = h_0$ corresponding to $\delta = 0$ in equation (2.2). As commented in Section 1, CV appears to be more natural than the plug-in procedures for selection of h_0, which will be developed in the next section. For the pilot bandwidth h, it seems that any second-generation method such as Sheather and Jones (1991) can also be alternatively applied, which itself, however, needs a pilot bandwidth again and will be examined in the numerical section (Section 5).

3. Spatial cross-validation bandwidth choice

In what follows, we investigate selection of the bandwidths that are needed in the adaptive kernel density estimation $\check{f}_{\mathbf{n}}$ in equation (2.1) by proposing an SCV method with choice of the adaptive bandwidths given in equations (2.2) and (2.3), for spatial lattice processes ($N>1$). Although some CV methods for non-adaptive bandwidths that leave more observations out were suggested for the time series case of $N=1$ in the literature (see Hart and Vieu (1990)), leave-one-out CV is often preferred for its simplicity in applications. Moreover, by our experience with the real spatial data numerical example in Lu *et al.* (2014) with the non-adaptive case, performance of leave-five-out CV seems very similar to that of leave-one-out CV. Further, leave-one-out CV is more popular with spatial data (see Le Rest *et al.* (2014)), and we mainly extend the leave-one-out CV for bandwidth selection below.

3.1. Cross-validation choice for adaptive bandwidth with a given pilot density

We first consider the choice of the adaptive bandwidth $h_{\mathbf{i}}$ by CV with a given pilot density f in equation (2.2), where the choice reduces to the selection of the global bandwidth h_0. We can propose CV bandwidth selection for the global bandwidth h_0 as defined in Section 2, by extending the leave-one-out CV criterion from time series to spatial lattices as follows:

$$\mathrm{SCV}_\delta(h_0)\equiv\mathrm{CV}_\delta(h_0)=\int_{\mathbb{R}^d}\check{f}^2_{\mathbf{n}}(x)w(x)\mathrm{d}x-\frac{2}{\tilde{\mathbf{n}}}\sum_{\substack{i_k=1\\ \forall k=1,2,\ldots,N}}^{n_k}\check{f}^{(\mathbf{i})}_{\mathbf{n}}(X_{\mathbf{i}})w(X_{\mathbf{i}}),\tag{3.1}$$

where SCV_δ stands for SCV for the global bandwidth h_0 with a given δ in equation (2.2), and $\check{f}^{(\mathbf{i})}_{\mathbf{n}}(x)$ is the adaptive kernel estimator of f based on $X_{\mathbf{j}}'$s, $\mathbf{j}=(j_1,\ldots,j_N)\neq\mathbf{i}=(i_1,\ldots,i_N)$, i.e.

$$\check{f}^{(\mathbf{i})}_{\mathbf{n}}(x)=\frac{1}{\tilde{\mathbf{n}}-1}\sum_{\substack{j_k=1\\ \forall k=1,2,\ldots,N\\ \exists k:j_k\neq i_k}}^{n_k}\frac{1}{h^d_{(j_1,j_2,\ldots,j_N)}}K\left(\frac{x-X_{(j_1,j_2,\ldots,j_N)}}{h_{(j_1,j_2,\ldots,j_N)}}\right),$$

with $h_{\mathbf{j}}=h_{(j_1,j_2,\ldots,j_N)}\equiv h_{\mathbf{j}}(h_0;f,\delta)$, as defined in equations (2.2) and (2.3) with $\mathbf{j}$ replacing $\mathbf{i}$, depending on h_0 (and where f is the assumed pilot density and δ is given, e.g. $\delta=\frac{1}{2}$ as explained following equation (2.3) for an optimal adaptive bandwidth), and $w(\cdot)$ is some non-negative weight function. Then, the extended CV optimal smoothing parameter is defined by

$$\check{h}_0(f,\delta)\equiv\check{h}_0=\arg\min_{h_0\in\mathcal{H}_{\mathbf{n}}}\mathrm{SCV}_\delta(h_0),\tag{3.2}$$

where $\mathcal{H}_{\mathbf{n}}$ is an interval that is defined in assumption (*H*) in the on-line appendix A.

As noted in Section 2, the adaptive bandwidth that was defined above includes the fixed bandwidth as a special case. When $\delta=0$, the adaptive KDE (AKDE) (2.1) reduces to the non-adaptive (fixed bandwidth) KDE (2.4) with $h=h_0$, which is independent of the pilot f. Therefore, the bandwidth h of $\hat{f}_{\mathbf{n}}(x)$ in estimator (2.4) can be selected following from equations (3.1) and (3.2) with $\delta=0$, i.e. to minimize an estimated ISE defined by

$$\mathrm{SCV}_0(h) \equiv \mathrm{CV}_0(h) = \int_{\mathbb{R}^d} \hat{f}_{\mathbf{n}}^2(x) w(x) \mathrm{d}x - \frac{2}{\tilde{\mathbf{n}}} \sum_{\substack{i_k=1 \\ \forall k=1,2,\ldots,N}}^{n_k} \hat{f}_{\mathbf{n}}^{(i_1,\ldots,i_N)}(X_{(i_1,\ldots,i_N)}) w(X_{(i_1,\ldots,i_N)}), \quad (3.3)$$

where $\hat{f}_{\mathbf{n}}^{(i_1,\ldots,i_N)}(x)$ is the kernel estimator of f based on $X'_{\mathbf{j}}$s, $\mathbf{j} = (j_1, \ldots, j_N) \neq \mathbf{i} = (i_1, \ldots, i_N)$, i.e.

$$\hat{f}_{\mathbf{n}}^{(i_1,\ldots,i_N)}(x) = \frac{1}{\tilde{\mathbf{n}} - 1} \sum_{\substack{j_k=1 \\ \forall k=1,2,\ldots,N \\ \exists k: j_k \neq i_k}}^{n_k} \frac{1}{h^d} K\left(\frac{x - X_{(j_1,j_2,\ldots,j_N)}}{h}\right).$$

Then, the SCV optimal smoothing parameter for equation (2.4) is defined by

$$\hat{h} = \arg \min_{h \in \mathcal{H}_{\mathbf{n}}} \mathrm{SCV}_0(h), \quad (3.4)$$

where $\mathcal{H}_{\mathbf{n}}$ is as defined in assumption (*H*) in the on-line appendix A. Note that $\hat{h} = \check{h}_0$ (see equation (3.2)) with $\delta = 0$.

In the non-adaptive case, it has been argued (see Xia and Li (2002)) that although, in the sense of MISE (that is data independent), the relative error of a CV bandwidth may be higher than that of, for example, the plug-in selector and the method of Ruppert *et al.* (1995), there is a growing body of opinion for using performance criteria, which are not just the MISE (see Mammen (1990), Jones (1991), Härdle and Vieu (1992) and Loader (1999)), targeting at estimating the unknown probability density function (in the context of this paper). From this point of view, the CV bandwidth performs reasonably well (Hall and Johnstone (1992), page 479). In the time series context, the argument for why CV is an appropriate bandwidth selection method can be found in Hart and Vieu (1990), Kim and Cox (1997), Quintela-del-Rio (1996) and Xia and Li (2002), among others. However, for adaptive kernel density estimation, no plug-in method has been seen even for independent data in the literature. The CV bandwidth looks a more natural and implementable option for the AKDE. There has been little investigation into the issue even with the non-adaptive KDE for spatial data, except for some spatial spectral density or variance estimation concerning second-order moments properties (see Nordman and Lahiri (2004)).

We shall concretely measure the optimality of the selected bandwidth by considering the widely used ISE, MISE and ASE, defined respectively by

$$d_{\mathrm{I}}(\check{f}_{\mathbf{n}}, f)(h) = \mathrm{ISE}(h) = \int_{\mathbb{R}^d} \{\check{f}_{\mathbf{n}}(x) - f(x)\}^2 w(x) \mathrm{d}x, \quad (3.5)$$

$$d_{\mathrm{M}}(\check{f}_{\mathbf{n}}, f)(h) = \mathrm{MISE}(h) = E\left[\int_{\mathbb{R}^d} \{\check{f}_{\mathbf{n}}(x) - f(x)\}^2 w(x) \mathrm{d}x\right] \quad (3.6)$$

and

$$\begin{aligned} d_{\mathrm{A}}(\check{f}_{\mathbf{n}}, f)(h) &= \mathrm{ASE}(h) \\ &= \frac{1}{\tilde{n}} \sum_{\substack{j_k=1 \\ \forall k=1,2,\ldots,N}}^{n_k} \{\check{f}_{\mathbf{n}}(X_{(j_1,\ldots,j_N)}) - f(X_{(j_1,\ldots,j_N)})\}^2 w(X_{(j_1,\ldots,j_N)}), \end{aligned} \quad (3.7)$$

where $\check{f}_{\mathbf{n}}(x)$ is the AKDE defined in equation (2.1), with $h_0 = h$ for ease of notation below, under a given pilot density f and δ.

We shall show that the SCV-selected bandwidth $\check{h}_0 = \check{h}_0(f, \delta)$ enjoys optimality for the AKDE (2.1). Before presenting the main results, we denote

$$R = \frac{1}{\tilde{\mathbf{n}}} \sum_{\substack{j_k=1 \\ \forall k=1,2,\ldots,N}}^{n_k} f(X_{(j_1,\ldots,j_N)} w(X_{(j_1,\ldots,j_N)}) - E\{f(X_{(j_1,\ldots,j_N)}) w(X_{(j_1,\ldots,j_N)})\}$$

and

$$T = -\int_{\mathbb{R}^d} f(x)^2 w(x) \mathrm{d}x - 2R.$$

All the technical assumptions are listed in the on-line appendix A.

First, theorem 1 establishes the rate of convergence of the so-called vertical error for the non-parametric density estimation of lattice data. It demonstrates that the SCV criterion proposed is asymptotically equivalent to the ISE criterion in the selection of bandwidth with spatial kernel density estimation, when compared with the MISE. On the basis of this result, we shall establish the appropriate conditions under which $\check{h}_0$ and $\hat{h}$ are asymptotically optimal not only in terms of the ISE as in Hart and Vieu (1990), but also in terms of the MISE and ASE respectively defined in equations (3.5)–(3.7), thus also extending the time series results of Hart and Vieu (1990).

Theorem 1.

(a) When $\delta = 0$, under assumptions (K1), (K2), (D1), (D2), (M), (H) and (W) listed in the on-line appendix A, we have

$$\frac{|\mathrm{SCV}_\delta(h) - \mathrm{ISE}(h) - T|}{\mathrm{MISE}(h)} = \mathcal{O}_P(\tilde{\mathbf{n}}^{-d/\{2(2r+d)\}}), \tag{3.8}$$

where r is the order of the continuous differentiation of the probability density function f, defined in assumption (D1) in on-line appendix A. Further, if assumption (K3) in appendix A is satisfied, we have

$$\sup_{h \in \mathcal{H}_\mathbf{n}} \frac{|\mathrm{SCV}_\delta(h) - \mathrm{ISE}(h) - T|}{\mathrm{MISE}(h)} = o_P(1) \tag{3.9}$$

as $\mathbf{n} \to \infty$.

(b) When $\delta > 0$, in addition to the conditions in (a), if assumptions (K4) and (D3) in appendix A are satisfied, then conclusions (3.8) and (3.9) hold true.

Remark 1. In theorem 1, conditions have been derived with spatial lattice data for the vertical error convergence that are needed below.

(a) Note that, even in the special case for time series with $N = 1$ and $\delta = 0$, our derived convergence rate is much faster than that in the literature. For example, the convergence rate in Kim and Cox (1997), theorem 1, corresponding to our result (3.8) is $O_P(n^{d(-1/2+\tilde{r}/\mu)/(d+2r)})$ (in the notation of this paper), where $\tilde{r}$ is a positive integer less than $\mu/2$ with μ given in the mixing assumption (M) in the on-line appendix A. This rate of $O_P(n^{d(-1/2+\tilde{r}/\mu)/(d+2r)})$ is much slower than our $O_P(\tilde{\mathbf{n}}^{-d/\{2(2r+d)\}})$, the latter being the same rate as obtained by Marron (1987) for independent and identically distributed (IID) data.

(b) Moreover, our derived conditions in the case of $\delta = 0$ are much weaker than those in the literature. For instance, Kim and Cox (1997), page 195, commented that one may note that the IID rates of convergence that were found in Marron (1987) and Marron and Härdle (1986) correspond to the case $\mu=\infty$ in their theorem 1. Interestingly, our theorem 1 above requires only

$$\mu > \frac{2Nr(2-2/q)}{1-2/q}$$

with $q>2$, roughly corresponding to $\mu>8$ under $N=1$, $r=2$ and $q=\infty$.

Second, the following theorem establishes that both the ISE and the ASE are asymptotically equivalent to the MISE criterion in bandwidth selection with spatial kernel density estimation.

Theorem 2.

(a) When $\delta=0$, under assumptions (K1)–(K3), (D1) and (D2), (M), (H) and (W) listed in the on-line appendix A, we have

$$\sup_{h\in\mathcal{H}_{\mathbf{n}}} \left|\frac{\mathrm{ISE}(h)-\mathrm{MISE}(h)}{\mathrm{MISE}(h)}\right| = o_p(1), \tag{3.10}$$

and

$$\sup_{h\in\mathcal{H}_{\mathbf{n}}} \left|\frac{\mathrm{ASE}(h)-\mathrm{MISE}(h)}{\mathrm{MISE}(h)}\right| = o_p(1), \tag{3.11}$$

as $\mathbf{n}\to\infty$.

(b) When $\delta>0$, in addition to the conditions in (a), if assumptions (K4) and (D3) in appendix A are satisfied, then conclusions (3.10) and (3.11) hold true.

Finally, theorem 3 establishes that the bandwidth $\hat{h}$ that is selected by the suggested CV (see equation (3.4)) is asymptotically optimal in terms of the criteria involving the ISE and the ASE as well as the MISE for spatial kernel density estimation.

Theorem 3.

(a) When $\delta=0$, under the conditions for part (a) of theorem 2, we have

$$\frac{d(\check{f}_{\mathbf{n}}, f)(\check{h}_0)}{\inf_{h\in\mathcal{H}_{\mathbf{n}}} d(\check{f}_{\mathbf{n}}, f)(h)} \to 1, \qquad \text{in probability,} \tag{3.12}$$

as $\mathbf{n}\to\infty$, where d is any of d_{I}, d_{A} and d_{M}, and $\check{f}_{\mathbf{n}}(x)=\hat{f}_{\mathbf{n}}(x)$ and $\check{h}_0=\hat{h}$ as defined in equations (3.2) and (3.4) respectively.

(b) When $\delta>0$, in addition to the conditions in (a), if assumptions (K4) and (D3) in the on-line appendix A are satisfied, then conclusion (3.12) holds true.

Remark 2. In the case $N=1$ for time series data with $\delta=0$, Hart and Vieu (1990) showed that the CV-selected bandwidth is asymptotically optimal in terms of the ISE. Theorem 3 thus extends the time series asymptotic optimality to the ASE and the MISE.

3.2. Cross-validation choice for adaptive bandwidth with an estimated pilot

In practice, for the AKDE with $\delta>0$, we cannot apply equation (2.1) directly because the true unknown density f is involved in the adaptive bandwidth via equation (2.2) in the above procedure. It needs to be replaced by a pilot estimator, say $\hat{f}_{\mathbf{n}}$ given in equation (2.4), where a pilot fixed bandwidth can be selected by any reasonable method, e.g. by CV in equation (3.4) or others such as the Sheather and Jones (1991) plug-in method (Sheather and Jones considered only the non-adaptive case). Thus the practical AKDE can be defined as

$$\check{\check{f}}_{\mathbf{n}}(x)=\frac{1}{\tilde{\mathbf{n}}}\sum_{\substack{i_k=1\\ \forall k=1,2,\ldots,N}}^{n_k}\frac{1}{\check{h}^d_{(i_1,i_2,\ldots,i_N)}}K\left(\frac{x-X_{(i_1,i_2,\ldots,i_N)}}{\check{h}_{(i_1,i_2,\ldots,i_N)}}\right),\qquad x\in\mathbb{R}^d,\tag{3.13}$$

where $\check{h}_{(i_1,i_2,\ldots,i_N)}=h_{(i_1,i_2,\ldots,i_N)}(h_0;\hat{f}_{\mathbf{n}},\delta)$ with $\hat{f}_{\mathbf{n}}$ replacing f in equation (2.2). Then the SCV optimal smoothing parameter is defined by

$$\check{\check{h}}_0(\hat{f}_{\mathbf{n}},\delta)\equiv\check{\check{h}}_0=\arg\min_{h_0\in\mathcal{H}_{\mathbf{n}}}\mathrm{S\check{C}V}_\delta(h_0),\tag{3.14}$$

where $\mathrm{S\check{C}V}_\delta(h_0)\equiv\mathrm{\check{C}V}_\delta(h_0)$ is as defined in equation (3.1) with $h_{(i_1,i_2,\ldots,i_N)}$ replaced by $\check{h}_{(i_1,i_2,\ldots,i_N)}$ in equation (2.2), and $\mathcal{H}_{\mathbf{n}}$ is the same interval as defined in assumption (*H*) in the on-line appendix A. On the basis of theorem 2, we shall show that theorem 4 below is true with $\hat{f}_{\mathbf{n}}$ replacing f in equation (2.2).

We now state the theorem on the selected $\check{\check{h}}_0$ for h_0 in terms of ISE, ASE and MISE.

Theorem 4. Under the conditions for part (b) of theorem 3 with $\sup_{x\in S_w}|\hat{f}_{\mathbf{n}}(x)-f(x)|\to 0$ in probability, we have

$$\frac{d(\check{\check{f}}_{\mathbf{n}},f)(\check{\check{h}}_0)}{\inf_{h_0\in\mathcal{H}_{\mathbf{n}}}d(\check{\check{f}}_{\mathbf{n}},f)(h_0)}\to 1,\qquad\text{in probability}\tag{3.15}$$

as $\mathbf{n}\to\infty$, where d is any of d_{I}, d_{A} and d_{M} defined in equations (3.5)–(3.7).

Theorem 4 confirms that the global bandwidth $\check{\check{h}}_0$ that is selected by the suggested CV (see equation (3.14)) is also asymptotically optimal in terms of the criteria involving the ISE and the ASE as well as the MISE for adaptive spatial kernel density estimation. Note from theorem 4 that the CV bandwidth for the AKDE by using $\hat{f}_{\mathbf{n}}$ instead of f in equation (2.2) is asymptotically optimal like that in theorem 3. We call this property an oracle property in the sense that the asymptotic optimality is achieved as if f were known. In the simulation section (Section 5.1), we shall call the CV-selected $\check{h}_0$ in equation (3.2) with the true f used in equation (2.2) the oracle CV bandwidth, whereas the CV-selected h_0 in equation (3.14) with f replaced by $\hat{f}_{\mathbf{n}}$, the estimated adaptive CV bandwidth for h_0, and the corresponding spatial AKDEs are called the oracle and the (estimated) AKDEs respectively.

4. Some further discussion

We here provide some more discussion regarding the application of the bandwidth selection proposed.

(a) For real spatial data, there may be spatial trends, which make the data non-stationary over space (see Harel *et al.* (2016)). For instance, suppose that on the plane of $N=2$ the observed $\tilde{X}_{\mathbf{i}}$ has a spatial trend, i.e. $\tilde{X}_{\mathbf{i}}=g(\mathbf{s}_{\mathbf{i}})+X_{\mathbf{i}}$, where $\mathbf{i}=(i_1,i_2)$ with $1\leqslant i_1\leqslant n_1$ and $1\leqslant i_2\leqslant n_2$, $\mathbf{s}_{\mathbf{i}}=(i_1/n_1,i_2/n_2)$, $g(\cdot)$ is a smooth trend function and $X_{\mathbf{i}}$ is an unobserved stationary field. In this case, we need to estimate and remove the spatial trends before applying the methodology that was given in Sections 2 and 3. For example, as done in Hallin *et al.* (2009) and Lu *et al.* (2014), the R package `sm` can be used to remove the spatial trends, and it can be proved that the theoretical results for the (estimated) detrended data, say $\hat{X}_{\mathbf{i}}$, replacing the unobservable stationary $X_{\mathbf{i}}$ in the above sections can still hold under some appropriate conditions (see section 3 of Hallin *et al.* (2009)). This is the situation for the real soil data set in Section 5.2.

(b) In this paper, we have developed CV-based adaptive bandwidth selection for density estimation of lattice data. Similarly, with a more involved regression setting, ideas and techniques developed from this paper will be useful in adaptive bandwidth selection for conditional regression of lattice data (see Hallin *et al.* (2004b)). CV-based adaptive bandwidth selection can be extended more naturally than other procedures for spatial regression, which is beyond the scope of this paper and will be left for future research.

5. Numerical finite sample performances

In the previous sections, asymptotic optimality for the CV-bandwidth-based kernel and CV pilot AKDEs was presented for spatial lattice data. In this section, we turn to the numerical finite sample performances of these estimates by using both simulated and real data sets. To simplify our discussion, we consider lattice data on the plane with $N=2$ and $d=1$, and denote X_{i_1,i_2} by $X_{i,j}$ for notational simplicity in this section. We take $\delta=0$ for the usual (non-adaptive) KDE and $\delta=\frac{1}{2}$ for the AKDE as in Abramson (1982a, b) in the following numerical examples.

5.1. Monte Carlo simulation

To evaluate the performance of our CV bandwidth selection procedure with non-Gaussian spatial data, we need to use a spatial lattice process $\{X_{i,j}\}$ on the plane, where the theoretical probability density function of $X_{i,j}$ can be computed analytically. Therefore we consider a special spatial lattice process, $X_{i,j}$, that is generated through a mixture of Gaussian spatial moving average processes, in a way similar to step 2 of section 8.1 of Lu and Tjøstheim (2014), as follows.

Step 1: generate three intermediate processes $Y_{ij,1}$, $Y_{ij,2}$ and $Y_{ij,3}$ from three independent Gaussian spatial moving averages, with $1\leqslant i\leqslant m_1, 1\leqslant j\leqslant m_2, k=1,2,3$:

$$Y_{ij,k}=\mu_k+a_{1,k}Z_{i-1,j;k}+a_{2,k}Z_{i,j-1;k}+a_{3,k}Z_{i,j;k}+a_{4,k}Z_{i+1,j;k}+a_{5,k}Z_{i,j+1;k}, \tag{5.1}$$

where the $Z_{i,j;k}$s are three independent IID samples from $\mathcal{N}(0,\sigma^2_{Z,k})$ for $k=1,2,3$ respectively. We take $\mu_1=-1$, $\mu_2=0.4$ and $\mu_3=1.5$, let $a_{m,1}$, $a_{m,2}$ and $a_{m,3}$ be the mth elements of $a_1=(\frac{1}{5},\frac{2}{5},\frac{3}{5},\frac{4}{5},-\frac{4}{5})$, $a_2=(-1,-\frac{4}{5},-\frac{3}{5},-\frac{2}{5},-\frac{1}{5})$ and $a_3=(\frac{1}{5},\frac{2}{5},\frac{3}{5},\frac{4}{5},1)$ respectively, and $\sigma_{Z,1}=0.3$, $\sigma_{Z,2}=0.2$ and $\sigma_{Z,3}=0.4$. Here the marginal distribution of $Y_{ij,1}$ is Gaussian $\mathcal{N}(\mu_1=-1,\sigma_1^2=0.1656)$, $Y_{ij,2}$ is $\mathcal{N}(\mu_2=0.4,\sigma_2^2=0.088)$ and $Y_{ij,3}$ is $\mathcal{N}(\mu_3=1.5,\sigma_3^2=0.352)$, where $\mathcal{N}(\mu,\sigma^2)$ stands for the univariate Gaussian distribution of mean μ and variance σ^2.
Step 2: we then generate a spatial process by first generating independent $R_{ij}=(R_{ij,1},R_{ij,2},R_{ij,3})\sim\text{multinomial}\{1,(p_1,p_2,p_3)=(0.4,0.3,0.3)\}$, $1\leqslant i\leqslant m_1, 1\leqslant j\leqslant m_2$, and then defining

$$X_{ij}=Y_{ij,1}R_{ij,1}+Y_{ij,2}R_{ij,2}+Y_{ij,3}R_{ij,3}, \qquad 1\leqslant i\leqslant m_1, \quad 1\leqslant j\leqslant m_2, \tag{5.2}$$

where the $R_{ij,k}$s are independent of the $Z_{i,j;k}$s and hence of the $Y_{ij,k}$s, with $k=1,2,3$. Note that the distribution of X_{ij} is a mixture of normal distributions, in the form

$$f(x)=0.4\phi_{(\mu_1=-1,\sigma_1^2=0.1656)}(x)+0.3\phi_{(\mu_2=0.4,\sigma_2^2=0.088)}(x)+0.3\phi_{(\mu_3=1.5,\sigma_3^2=0.352)}(x), \tag{5.3}$$

where $\phi_{(\mu,\sigma^2)}(x)$ stands for the probability density function of a normal distribution $\mathcal{N}(\mu,\sigma^2)$.

We generate the simulated spatial data by using the values of the parameters in the above models. Note, as commented by a referee, that the simulated spatial process is m dependent and

Table 1. Summary of selected bandwidths for 100 simulations with $m_1 = 25$ and $m_2 = 10$

Method	*Minimum*	*1st quartile*	*Median*	*Mean*	*3rd quartile*	*Maximum*
H.CV	0.0669	0.1471	0.1779	0.1732	0.1980	0.2680
H.SJ	0.1470	0.1824	0.1958	0.1989	0.2126	0.3008
H.R	0.3105	0.3302	0.3383	0.3381	0.3434	0.3747
H0.CV	0.0679	0.2104	0.2429	0.2338	0.2685	0.3242
H0.SJ	0.1055	0.2145	0.2438	0.2399	0.2689	0.3264
H0.orac	0.1024	0.2246	0.2704	0.2638	0.3052	0.4679

therefore satisfies the α-mixing assumption in the on-line appendix A. This follows because the Y_{ijk}s (and R_{ijk} also) are m dependent within k and independent across k (with m appropriately defined), following from the simulating models (5.1) and (5.2) and using the fact that $Z_{i,j;1}$, $Z_{i,j;2}$, $Z_{i,j;3}$ and R_{ij} are independent IID processes. It is also easy to see similarly to section 8.1 of Lu and Tjøstheim (2014) that the resultant marginal and the joint density functions are mixtures of Gaussian distributions, which satisfy the assumptions (D1), (D2) and (M) of the on-line appendix A. We repeat the simulation 100 times, with different sizes of samples of $(m_1, m_2) = (25, 10)$, $(m_1, m_2) = (20, 20)$, $(m_1, m_2) = (50, 50)$ and $(m_1, m_2) = (100, 100)$, from smaller to larger sample sizes.

We examine the performance of the density estimates with the spatial CV selection for the global bandwidth h_0 in equation (3.2) with pilot densities of different bandwidths, including the bandwidth h given in equation (3.4). We first consider the case $(m_1, m_2) = (25, 10)$ with total sample size $n = m_1 m_2 = 250$. For CV-based bandwidth selection, after some experiments on the bandwidth interval so that the bandwidths selected are within the interval, we set the bandwidth interval $\mathcal{H}_\mathbf{n} = [0.02, 0.5]$ in the computations below, which roughly corresponds to taking $a = 0.06$ and $b = 1.51$ with $r = 2$, $d = 1$ and $\tilde{\mathbf{n}} = n = 250$ in assumption (H) for the case $(m_1, m_2) = (25, 10)$. (We also tried $\mathcal{H} = [0.001, 1]$, having the same outcomes.) The spatial CV-based global bandwidths for adaptive density estimates of $X_{i,j}$ with different pilot densities, as well as the selected bandwidths with non-adaptive KDE, for the kernel K taken as a standard normal density function, are summarized in Table 1 with the non-adaptive case in the upper and adaptive case in the lower part of the table. Note that the R function `density` with the default (rule-of-thumb) and the Sheather and Jones (1991) bandwidths from package `stats` (R Development Core Team, 2015) are also used for comparison. It is interesting to note from Table 1 that, for the non-adaptive KDE method, the selected bandwidths by CV, H.CV, tend to be smaller than those by the Sheather–Jones method, H.SJ, whereas the rule-of-thumb bandwidth, H.R, is much larger than both H.CV and H.SJ, which may explain that the rule-of-thumb bandwidths tend to lead to a largely biased KDE (as displayed in Fig. 1).

For the AKDEs, the spatial CV-based global bandwidths for h_0 either with CV or Sheather–Jones pilot KDEs (simply denoted by H0.CV and H0.SJ, which may be more clear as H0.CVCV and H0.CVSJ respectively) look quite similar overall in Table 1, which shows that the AKDEs corresponding to those bandwidths either with CV or Sheather–Jones pilot KDEs perform similarly (as again indicated in Fig. 1). In Table 1, we also provided the unpractical oracle bandwidths H0.orac (denoted by H0.or in Fig. 2), i.e. the spatial CV-based global bandwidths with true pilot density, for the AKDE. It follows from Table 1 that both H0.CV and H0.SJ are relatively smaller than H0.orac.

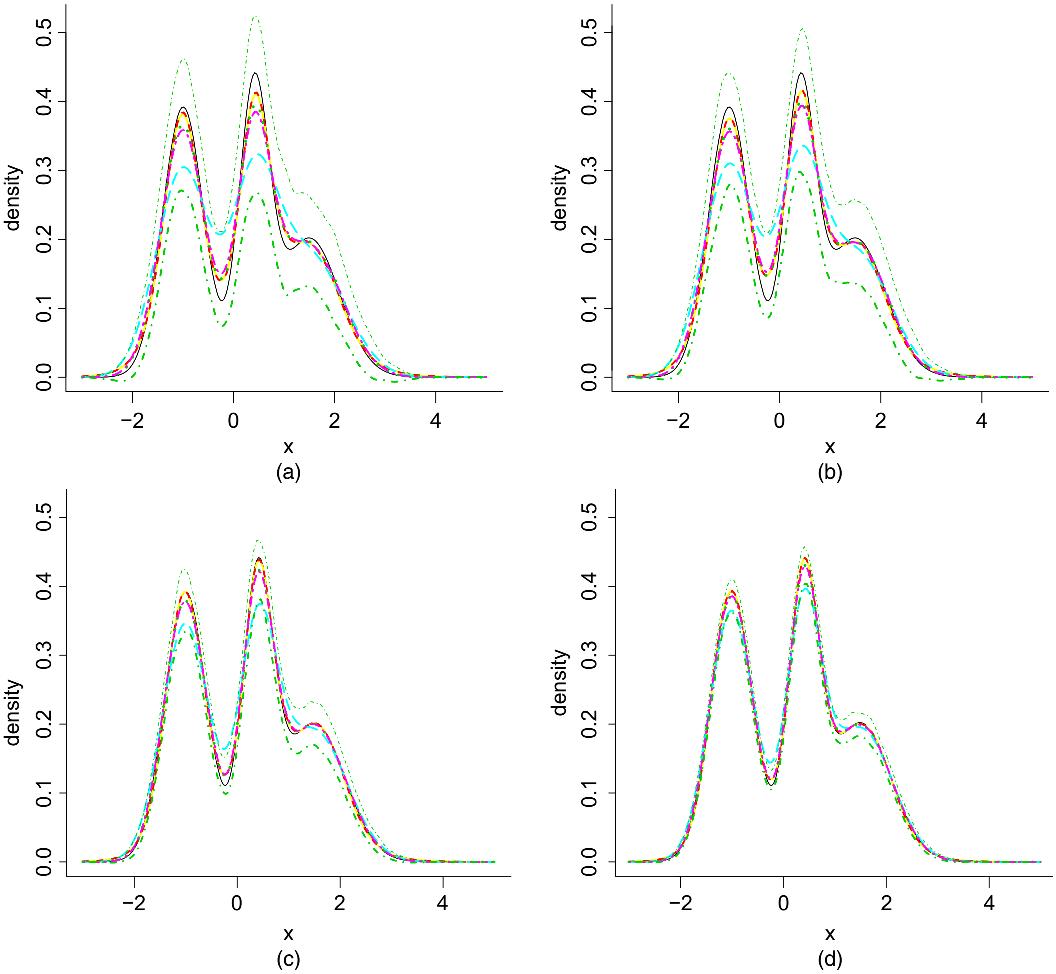

Fig. 1. Mean of 100 simulated estimates of density for various selected bandwidths of 100 simulations of sample sizes of $n = m_1 m_2$ (, ftrue; , akde.cvcv; , kde.cv; , density.r; , kde.SJ; , akde.cvSJ; , kde.cv.upper; , kde.cv.lower): (a) $(m_1, m_2) = (25, 10)$; (b) $(m_1, m_2) = (20, 20)$; (c) $(m_1, m_2) = (50, 50)$; (d) $(m_1, m_2) = (100, 100)$

As a benchmark comparison, we can also calculate the optimal bandwidths that minimize the MISEs for the non-adaptive KDEs and the AKDEs, where the optimality is in terms of the MISE as defined in equation (3.6). For example, given the sample size of $(m_1, m_2) = (25, 10)$, the optimal bandwidth obtained, H.op $= 0.1788$, is the bandwidth that minimizes the MISE of the non-adaptive KDE (2.4) with respect to h. Similarly, the optimal bandwidth, H0.op $=$ 0.2659, is the bandwidth that minimizes the MISE of the AKDE (2.1) with oracle (i.e. true) pilot density f in equations (2.2) and (2.3), with respect to the global bandwidth h_0. This leads to the smallest MISE value among all the AKDEs (including those with other pilot densities, say, the CV-based or Sheather–Jones-based KDE as pilot; see Fig. B.1 in the on-line appendix). Clearly the MISE optimal bandwidths are theoretical quantities independent of data (given the sample size of (m_1, m_2)). These theoretically MISE optimal bandwidths, H.op (for the non-adaptive KDE) and H0.op (for the AKDE), used for a benchmark only, together with boxplots of other

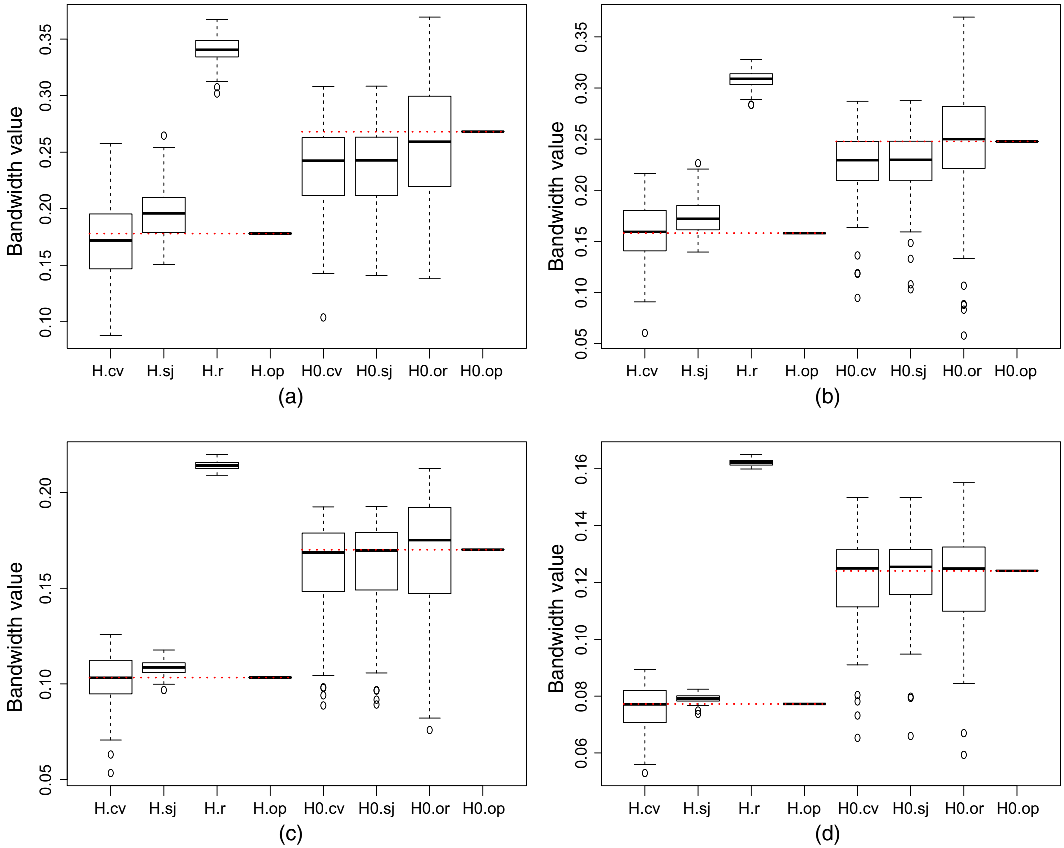

Fig. 2. Boxplots of various selected bandwidths of 100 simulations of sample sizes of $n = m_1 m_2$ (H.CV, H.SJ and H.R are the CV, Sheather–Jones and R default (rule-of-thumb) bandwidths for h respectively, and H.op is the optimal bandwidth of h minimizing the MISE, in KDEs; H0.CV, H0.SJ and H0.or are the CV bandwidths for h_0 in AKDEs with pilot estimates by the CV, the Sheather–Jones and the oracle true density respectively) (• • • • •, MISE optimal H.op and H0.op of the KDE and the AKDE with the pilot of oracle (true) density respectively, provided for ease of comparison with others): (a) $(m_1, m_2) = (25, 10)$; (b) $(m_1, m_2) = (20, 20)$; (c) $(m_1, m_2) = (50, 50)$; (d) $(m_1, m_2) = (100, 100)$

bandwidths summarized in Table 1, can be found in Fig. 2 with dotted horizontal lines indicated for ease of comparison of H.op and H0.op with others.

The means of the AKDEs and non-adaptive KDEs corresponding to those bandwidths in Table 1 for 100 simulations are depicted in Fig. 1; see also boxplots of those bandwidths in Fig. 2. Here the true density function (5.3) together with the upper and lower limits, say, for the mean of the CV-based density estimates as an example (by adding and subtracting respectively the double standard deviation of the CV-based density estimates of 100 simulations) is also provided for comparison. For brevity and to make Fig. 2 legible, the simulated pointwise confidence interval for the mean density estimate over 100 simulations is provided for one KDE only. The sampling variability and simulation error are basically similar for other estimates. Further it can be observed from Fig. 1(a) that the AKDEs by using spatial CV-based global bandwidths either with the CV or the Sheather–Jones pilot KDE (denoted by akde.cvcv and akde.cvSJ) are very similar in the means of the 100 simulations, which are much closer to the true density ftrue than other (non-adaptive) KDEs such as the KDEs with the CV, the Sheather–Jones and the rule-of-thumb bandwidths (denoted by kde.cv, kde.SJ and density.r respectively). Also, we can clearly see from Figs 1(a)–1(d) that these facts remain true for sample sizes sufficiently large, with the AKDEs outperforming the non-adaptive KDEs. In particular, for the case $(m_1, m_2) = (100, 100)$, we cannot distinguish the means of the akde.cvcv and the akde.cvSJ from ftrue in Fig. 1(d). Note in Fig. 2 that boxplots of various bandwidths are provided, where the dotted horizontal lines, standing for the corresponding values of the benchmark MISE optimal bandwidths, are plotted for ease of comparison with other bandwidths. Interestingly, the mean and median of the CV and the Sheather–Jones selected bandwidths given in Table 1 and Fig. 2 are much closer to the optimal bandwidth (say H.op = 0.1788 in Fig. 2(a)) than the rule-of-thumb bandwidths (the CV bandwidths are a little closer than the Sheather–Jones bandwidths) for the non-adaptive KDE. As expected, the optimal bandwidth (say H0.op = 0.2659 in Fig. 2(a)) for the AKDE appears to be quite close to the mean and median of the oracle bandwidths H0.or, which are similar to (only slightly larger than) those for H0.CV and H0.SJ given in Table 1. We can also see from Fig. 2 that, with the sample sizes increasing, the medians for H0.CV and H0.SJ as well as the oracle bandwidths H0.or are becoming similar and close to H0.op for the AKDEs (see Fig. 2(d)).

We now further examine the finite sample optimality of the spatial CV-based adaptive bandwidths compared with other bandwidths in estimates of the density in terms of ISE, ASE and MISE, which are also examined in theorem 3. Here, for a given density estimator $\hat{f}_{\mathbf{n}}$ with the true density function f specified in equation (5.3), $d_{\mathrm{I}}(\hat{f}_{\mathbf{n}}, f)(h) = \mathrm{ISE}(h)$ and $d_{\mathrm{A}}(\hat{f}_{\mathbf{n}}, f)(h) = \mathrm{ASE}(h)$ are easily calculated as defined in equations (3.5) and (3.7) with $d = 1$ and $w(x) \equiv 1$ taken in this section. However, for $d_{\mathrm{M}}(\hat{f}_{\mathbf{n}}, f)(h) = \mathrm{MISE}(h)$ defined in equation (3.6), because of spatial dependence with the simulated random field (2), it becomes very complex and cannot be calculated as simply as with independent or time series observations. But fortunately, as a referee suggested, if the ISE, as a function of h, is easily found, then the MISE can be approximated through simulation. We do this by approximation through the average of the resultant ISE values based on 1000 simulations and we can therefore deal with the MISE for the KDE and the AKDE defined in equations (2.4) and (2.1) respectively. To compare the performance of the estimates, AKDE (2.1) and KDE (2.4) with different bandwidths, defined in Sections 2 and 3, we shall consider the AKDEs that are defined in equation (2.1) with the adaptive bandwidths given in equation (3.2) by using different pilot estimates of f, including the KDE (2.4) with the CV (see equation (3.4)) and the second-generation Sheather and Jones (1991) bandwidth and the oracle (true f) taken as a benchmark. As a comparison, we have also looked at KDE (2.4) with bandwidths of the CV (see equation (3.4)) and the Sheather and Jones (1991) bandwidths (those bandwidths outperforming the R default (rule-of-thumb) bandwidth as shown in Fig. 1). We did not consider the second-generation bandwidth by the bootstrap as bootstrapping the dependent non-Gaussian spatial data is still quite problematic, and not as simple as for independent data (see, for example, Faraway and Jhun (1990)).

Boxplots of the ISEs, ASEs and MISEs (scaled up by 100, 300 and 150 respectively for easy presentation) with different bandwidth-based density estimates are displayed in Fig. 3 for the 100 simulations of various sample sizes of (m_1, m_2) as specified above (for brevity the rule of thumb is not reported here, which performs worse than those reported in Fig. 3, as indicated in Fig. 1). We have also provided boxplots for the minimal values of the ISEs, ASEs and MISEs in Fig. B.1 in the on-line appendix and their corresponding optimal bandwidths (i.e. the bandwidths minimizing the ISE, ASE and MISE in Fig. B.2 for the AKDEs and the KDEs of the simulated data. Here in Fig. 3, k.cv and k.SJ are for the KDE with the CV-based and Sheather–Jones-based bandwidth respectively whereas a.cv, a.SJ or a.o are a simple notation for the AKDE using the spatial CV global bandwidth with the pilot density being the CV-based KDE, the Sheather–

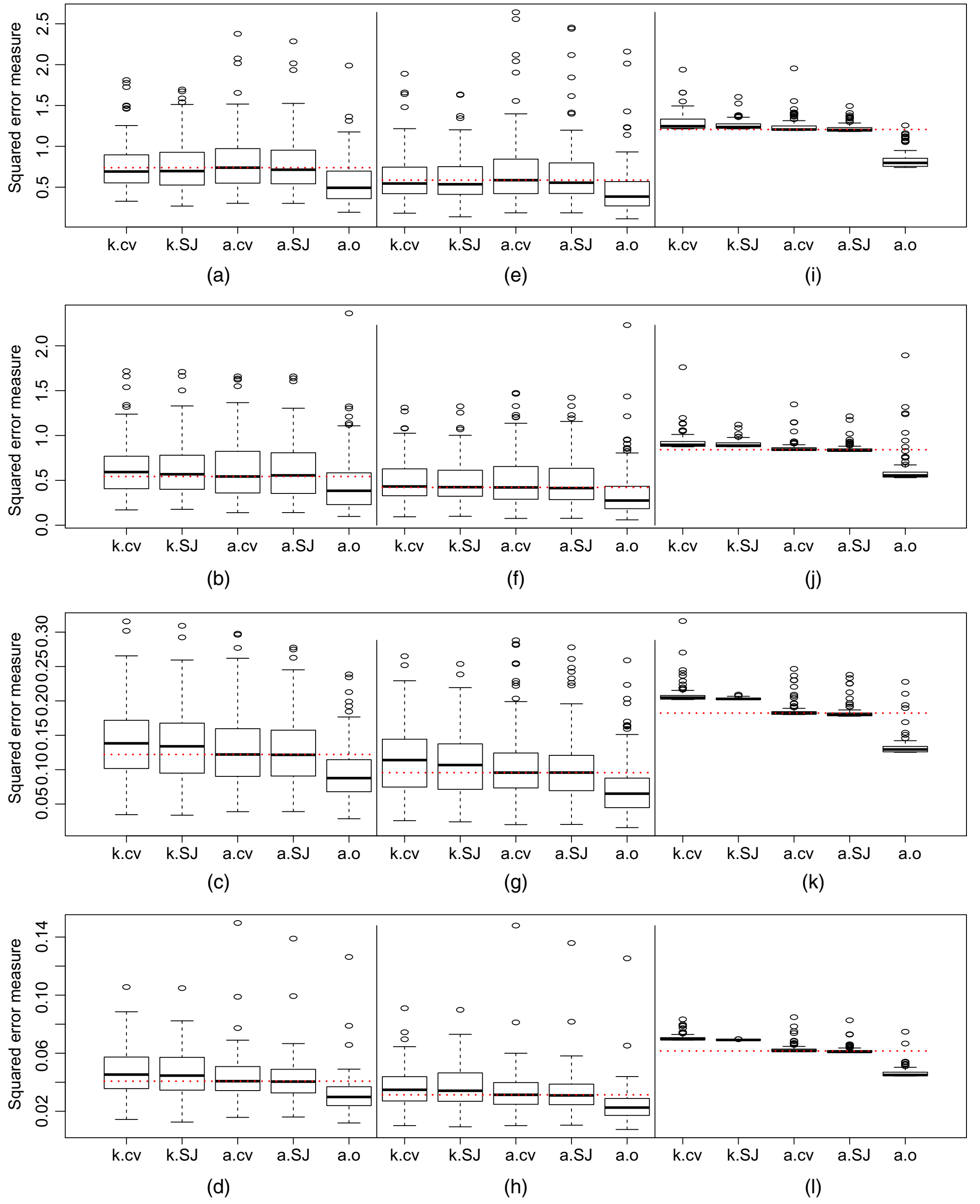

Fig. 3. Boxplots of (a)–(d) the ISE (×100), (e)–(h) ASE (×300) and (i)–(l) MISE (×150) for density estimates with 100 simulations of sample sizes of (m_1, m_2) (k.cv and k.SJ are for the KDEs with CV and Sheather–Jones-based bandwidths, and a.cv, a.SJ and a.o are for the AKDEs by using the SCV global bandwidths with the CV, the Sheather–Jones-based KDE and the oracle true, pilot densities) (• • • • •, medians of the 100 simulated ISEs, ASEs and MISEs of the AKDE with a CV KDE pilot respectively for ease of comparison): (a), (e), (i) $(m_1, m_2) = (25, 10)$; (b), (f), (j) $(m_1, m_2) = (20, 20)$; (c), (g), (k) $(m_1, m_2) = (50, 50)$; (d), (h), (l) $(m_1, m_2) = (100, 100)$

Jones-based KDE or the oracle true density respectively in Section 3. The conclusion of all of these experiments is that, in the KDE case, the CV and Sheather–Jones bandwidths perform similarly (with the Sheather–Jones bandwidth a little better as is known). This similarity holds true in the AKDE case as well, when the CV and Sheather–Jones bandwidths are used in the pilot estimation stage with the global bandwidth by the spatial CV, but the AKDE pair does significantly better than the KDE pair.

Also, as a referee suggested, we have examined the minimal ISE (ASE or MISE) values of the KDE and the AKDE together with their corresponding optimal bandwidths, which are provided, for brevity, in Figs B.1 and B.2 in section B3 of the on-line appendix B. Note that, in Fig. B.1, v.k stands for the minimal ISE (ASE or MISE) value of the KDE, whereas the corresponding optimal bandwidth, denoted by hik (hak or hmk), minimizes with respect to h the ISE (ASE or MISE) of the KDE, and v.acv for the minimal ISE (ASE or MISE) value of the AKDE using the corresponding optimal global bandwidth, denoted by h0icv (h0acv or h0mcv), minimizing with respect to h_0 the ISE (ASE or MISE) of the AKDE with the CV-based KDE as the pilot density in Fig. B.2. Here v.asj (or v.a0) can be defined similarly to v.acv, and h0isj, h0asj or h0msj (or h0io, h0ao or h0mo) to h0icv, h0acv or h0mcv, with the Sheather–Jones-based KDE (the oracle true density) replacing the CV-based KDE as the pilot.

We again have the following observations from Fig. 3 and Fig. B.1.

(a) Overall the AKDEs by using the proposed spatial CV adaptive bandwidths either with the CV or the Sheather–Jones piloted KDE outperform the KDEs with the CV-based and the Sheather–Jones-based bandwidths. This is particularly observed in terms of MISE even with the sample sizes as small as $(m_1, m_2) = (25, 10)$, which is the sample size in the real data below, in Figs 3(a), 3(e) and 3(i) and Fig. B1(a).

(b) With the sample sizes increasing, the AKDEs are clearly preferred to the KDEs, even in terms of the ISE and ASE, as obviously seen in Figs 3(d), 3(h) and 3(l) and Fig. B1(d).

(c) For the AKDEs in Fig. 3 and Figs B1(a)–B1(d), their performances by using the proposed spatial CV adaptive bandwidths either with the CV or the Sheather–Jones pilot KDE appear quite similar although using the Sheather–Jones KDE pilot is sometimes slightly preferred. But, for the non-adaptive KDEs, as indicated in Sheather and Jones (1991) and Jones *et al.* (1996), the Sheather–Jones-based bandwidth is usually preferred.

(d) As expected, the impractical oracle AKDE performs best, but with the sample size becoming large as in Figs 3(d), 3(h) and 3(l) and Fig. B1(d) all the estimates have very small ISE, ASE and MISE (much smaller than those in Figs 3(a)–3(c), 3(e)–3(g) and 3(i)–3(k) and Figs B1(a)–B1(c), so the performance of all the practical estimates including the AKDE and the KDE approaches looks acceptable for the large sample sizes. Also, it is interesting to observe from Fig. B.2 that, as the sample sizes increase, the optimal global bandwidths for AKDE either with the CV KDE or the Sheather–Jones KDE for a pilot approach that of the oracle AKDE with true density as a pilot, all of which are larger than the optimal KDE bandwidths. This is seen particularly clearly under the MISE in Fig. B.2 (note that the asymptotic MISE is the base of the Sheather–Jones bandwidth method; see Jones *et al.* (1996), in the non-adaptive case).

Now we can define the ratios of the left-hand side of equation (3.12) for d_I, d_A and d_M respectively by

$$R_{d\mathrm{I}} = \frac{d_\mathrm{I}(\hat{f}_\mathbf{n}, f)(\hat{h})}{\inf_{h\in\mathcal{H}_\mathbf{n}} d_\mathrm{I}(\hat{f}_\mathbf{n}, f)(h)},$$

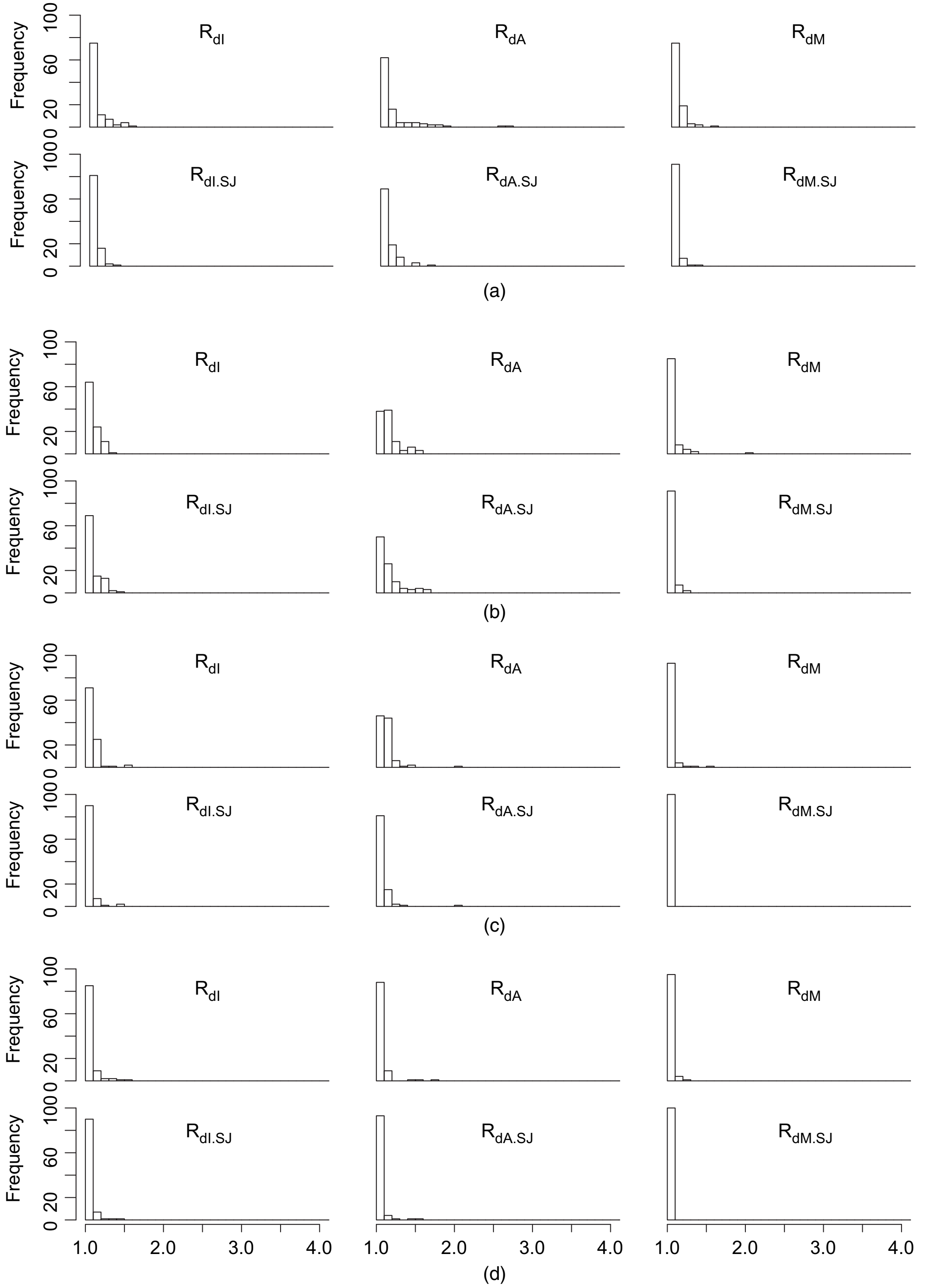

Fig. 4. Histograms of $R_{d\mathrm{I}}$s, $R_{d\mathrm{A}}$s and $R_{d\mathrm{M}}$s, and $R_{d\mathrm{I.SJ}}$s, $R_{d\mathrm{A.SJ}}$s and $R_{d\mathrm{M.SJ}}$s respectively for the CV- and the Sheather–Jones-selected bandwidths of 100 simulations with different sample sizes of $n = m_1 m_2$: (a) $(m_1, m_2) = (25, 10)$; (b) $(m_1, m_2) = (20, 20)$; (c) $(m_1, m_2) = (50, 50)$; (d) $(m_1, m_2) = (100, 100)$

$$R_{d\mathrm{A}} = \frac{d_{\mathrm{A}}(\hat{f}_{\mathbf{n}}, f)(\hat{h})}{\inf_{h \in \mathcal{H}_{\mathbf{n}}} d_{\mathrm{A}}(\hat{f}_{\mathbf{n}}, f)(h)},$$
$$R_{d\mathrm{M}} = \frac{d_{\mathrm{M}}(\hat{f}_{\mathbf{n}}, f)(\hat{h})}{\inf_{h \in \mathcal{H}_{\mathbf{n}}} d_{\mathrm{M}}(\hat{f}_{\mathbf{n}}, f)(h)}.$$

The histograms of these ratios for 100 simulations with different cases of sample sizes for the non-adaptive KDE are depicted in Fig. 4, whereas the histograms for the AKDE, for brevity, in Fig. B.3 in section B.3 of the on-line appendix B, both of which look very similar. It follows from these figures that these ratios tend to be closer to 1 as the sample size increases. In particular $R_{d\mathrm{M}}$ tends to converge faster than $R_{d\mathrm{I}}$, and $R_{d\mathrm{I}}$ faster than $R_{d\mathrm{A}}$, which seem understandable as $R_{d\mathrm{A}}$ is more sample dependent than $R_{d\mathrm{I}}$, and $R_{d\mathrm{I}}$ than $R_{d\mathrm{M}}$. Clearly, the proposed spatial CV-based adaptive bandwidth selection (the non-adaptive bandwidth is its special case) works reasonably well in terms of all d_{I}, d_{A} and d_{M}, and it improves as the sample size becomes larger as the asymptotic optimality is indicated in theorems 3 and 4 with the oracle property. In particular, in terms of the MISE distance d_{M} in equation (3.6), the spatial CV-selected bandwidth (either with the CV or Sheather–Jones bandwidth as a pilot) can approximate the optimal bandwidth well even for the sample size as small as $m_1 = 25$ and $m_2 = 10$, which is the sample size of the real soil data set that is examined in Section 5.2.

Before ending this subsection, we here have a simple look at the computational cost of the approach. As can be seen from equation (3.1), given a pilot density, the leading order of the cost of calculating SCV_{δ} in equation (3.1) for the AKDE is $O(n^2)$, where $n = m_1 m_2$ is the total sample size of an $m_1 \times m_2$ lattice data set here. Likewise, so is the leading order for calculating SCV_0 in equation (3.3) with a non-adaptive KDE. Therefore, in terms of the leading order in computational costs, the adaptive bandwidth selection by equation (3.14) in Section 3.2 either with a CV or a Sheather–Jones pilot density is of the same order $O(n^2)$. In general, the procedure that was used above appears to run well in practice. The real elapsed time for a bandwidth selection run in R by a Dell Precision 7520 laptop with COREi7 for the simulated data sets of different sample sizes is reported in Table B.1 in the on-line appendix B. As is seen from Table B.1, for a sample size of $m_1 \times m_2 = 20 \times 20 = 400$, the real elapsed time that is needed for our proposed spatial CV procedure is about 2.5 s, whereas, for a sample size as large as $m_1 \times m_2 = 100 \times 100 = 10000$, the real elapsed time is about 110 s, which appears quite acceptable. This indicates that our spatial CV approach can work fast for the analysis of the real soil data of sample size $(m_1, m_2) = (25, 10)$ given in Section 5.2.

5.2. An application to soil data analysis

We analyse a spatial soil data set, soil250, in the R package `geoR` (see Ribeiro and Diggle (2016)), which consists of uniformity trials with 250 undisturbed soil samples collected at 25 cm soil depth of a spacing of 5 m, resulting in a regular grid of 25×10 points. The data consist of 250 observations on 22 variables concerning soil chemistry properties measured on the grid. In this analysis, we consider only eight columns of the data table, involving the columns named Linha (the column for x-co-ordinate), Coluna (the column for y-co-ordinate), pHKCl (soil acidity pH by potassium chloride (KCl) solution), Ca (calcium content), K (potassium content), H (hydrogen content), C (carbon content) and CTC (cation exchange capability). Here the acidity by KCl measures the acidity in the soil solution, plus the reserve acidity in the colloids, and is therefore more acidic than water, although both are neutral at pH 7.0 (see Directorate Agricultural Information Services (2007)). Zheng *et al.* (2010) recently studied the

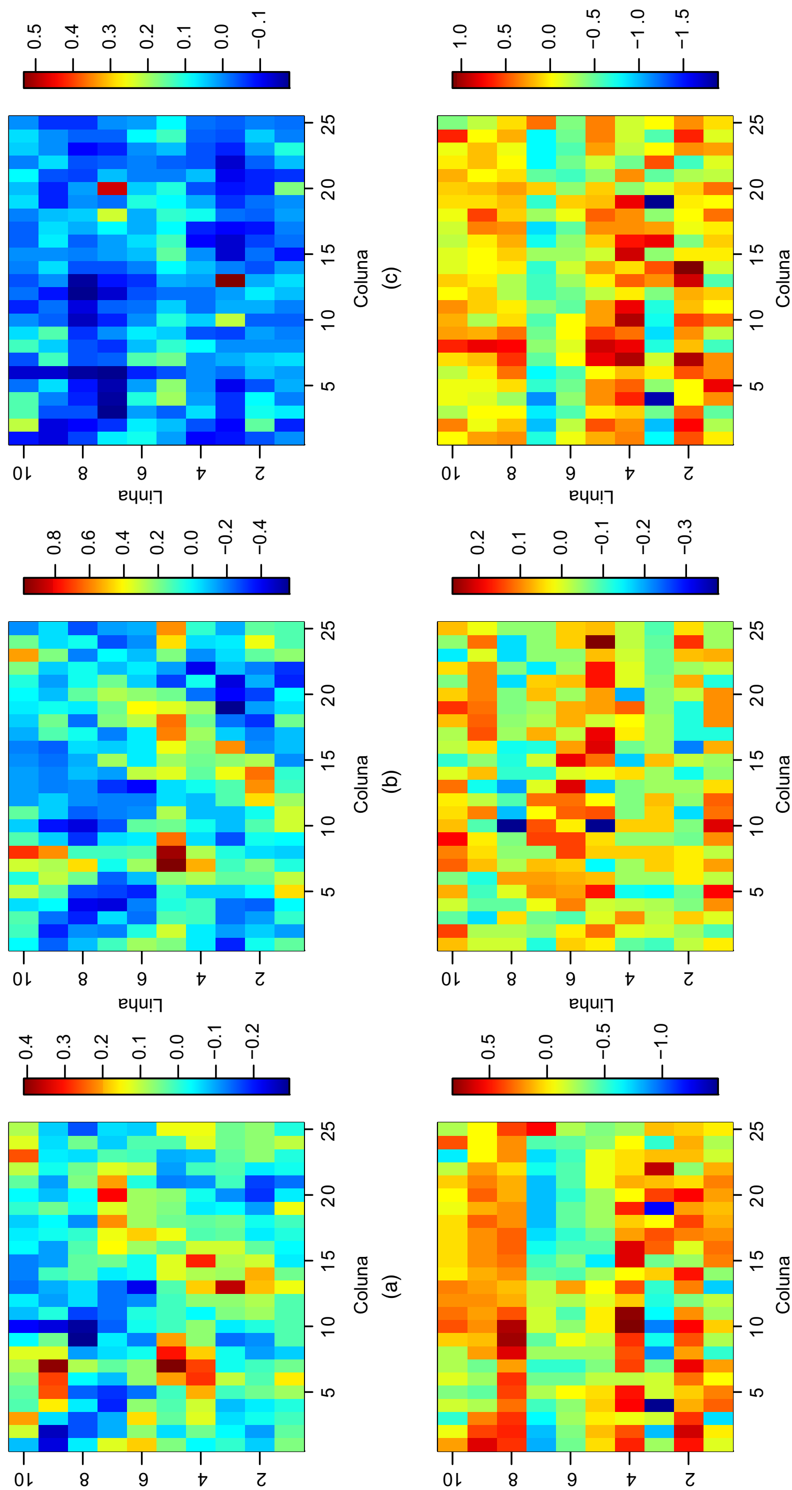

Fig. 5. Soil data—images of six soil properties variables after spatial trend removal by sm.regression, plotted over space (Linha, Coluna): (a) res.pHKCl; (b) res.Ca; (c) res.K; (d) res.H; (e) res.C; (f) res.CTC

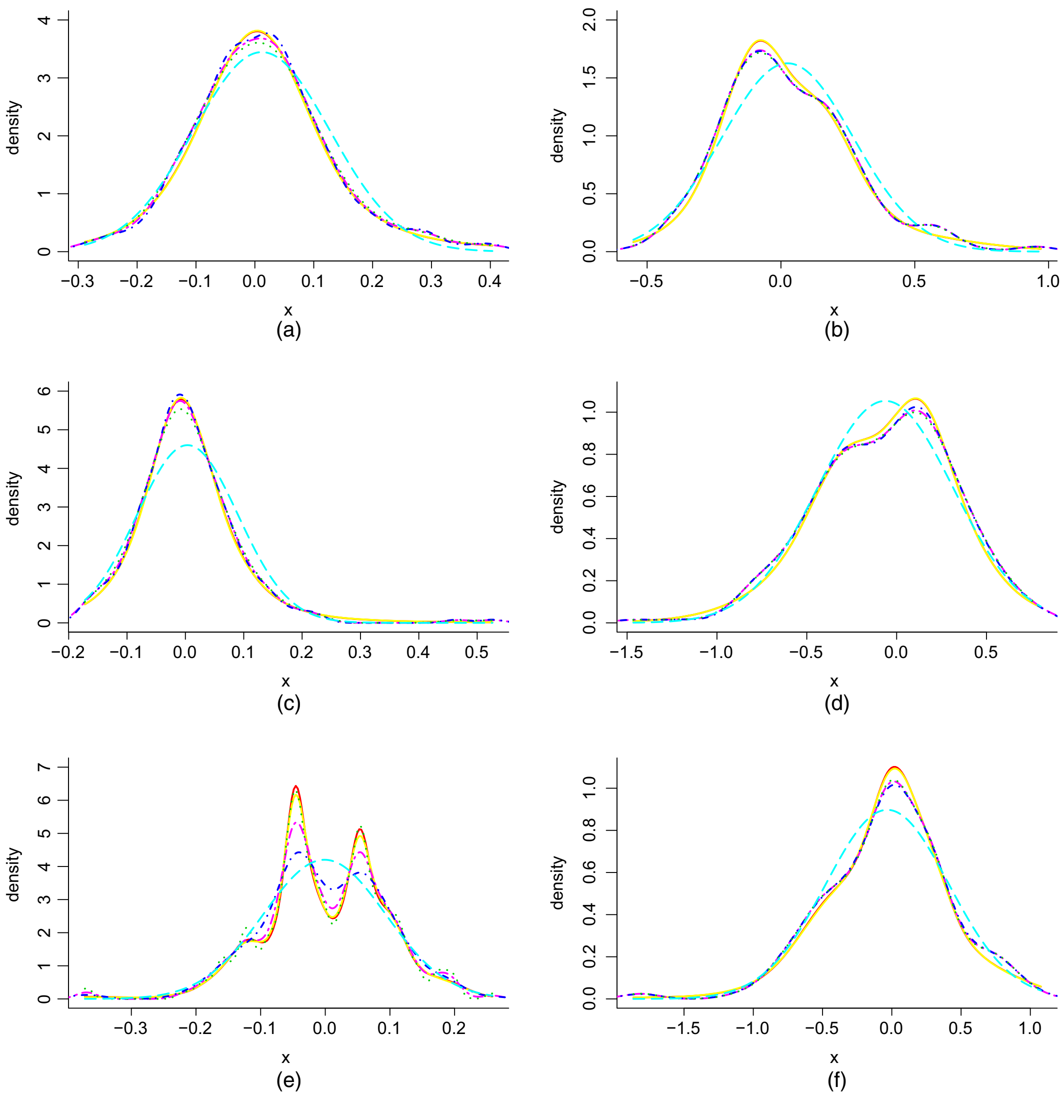

Fig. 6. Density estimation by different methods for six soil chemistry properties of (a) pHKCl, (b) Ca, (c) K, (d) H, (e) C and (f) CTC after spatial trends have been removed: ——, cv.akde; • • • •, cv.kde; -·-·-, r.density; — —, normal; -·-·-, Sheather–Jones; ——, cv.akde.SJ

spatial spectral density for the CTC-variable, and Lu *et al.* (2014) analysed the effects of soil chemistry properties of pHKCl, Ca, K and C as well as other variables on CTC: an important soil property for soil conservation, of concern in agriculture science.

In the original data, as shown in Lu *et al.* (2014), there seem to be some spatial trends for all variables, so we apply `sm.regression` in the R package `sm` (see Bowman and Azzalini (2014)) to remove the spatial trends. The resulting spatial data of these soil chemical variables, denoted by the prefix 'res.', standing for residual, are plotted in Fig. 5, and appear to be stationary. We hence analyse the distribution of these variables on the basis of the residual data; the different density estimates are plotted in Fig. 6: Fig. 6(a) res.pHKCl, Fig. 6(b) res.Ca, Fig. 6(c) res.K, Fig. 6(d) res.H, Fig. 6(e) res.C and Fig. 6(f) res.CTC, where 'cv.akde' and 'cv.akde.SJ' are for

Table 2. CV-selected bandwidths for KDEs and AKDEs with six soil variables

Bandwidth	*Results for the following variables:*					
	res.pHKCl	*res.Ca*	*res.K*	*res.H*	*res.C*	*res.CTC*
h in KDE	0.04256665	0.07467	0.02776	0.12733	0.00969	0.11465
h_0 in AKDE	0.06730755	0.09817	0.04451	0.15311	0.01712	0.16290

adaptive kernel density estimate (2.1) with adaptive bandwidths combined by the CV-piloted and the Sheather–Jones-piloted densities respectively, and 'cv.kde' for the CV-based kernel density estimate (2.4) in Section 3. Here the CV-selected bandwidths in Sections 3.1 and 3.2 are given in Table 2. In addition, as benchmark comparisons, the R function `density` estimates, denoted by 'r.density' and 'r.SJ', with the default (rule-of-thumb) and the Sheather–Jones bandwidths respectively, and the normal density, denoted by 'normal', with the same mean and variance of the sample, are also depicted in Fig. 6. It is clear that the CV-based estimated densities of all these variables indicate that distributions are basically non-Gaussian, where the densities of res.Ca, res.K, res.H, res.C and res.CTC are skewed and one of them is even multimodal (e.g. res.C). Also, from Fig. 6 it is interesting that the estimated densities by the proposed AKDE given in equation (2.1) either with a CV pilot or a Sheather–Jones pilot adaptive bandwidth are similar, which are more significantly illustrating the non-Gaussianity of these residual data than are the KDEs in equation (2.4), including the rule of thumb, the CV and the Sheather–Jones bandwidths, for the six soil variables. Furthermore, in view of the simulation above with the sample size of $m_1 = 25$ and $m_2 = 10$, the (estimated) adaptive CV-piloted AKDE strengthens this impression, demonstrating that non-linear quantile analysis of res.CTC in relation to other covariates is needed, as indicated in Lu *et al.* (2014), for a better understanding of the data. Furthermore, from the estimated probability densities that are plotted in Fig. 6, we can learn more about the soil properties. With the possible exception of Fig. 6(a), the densities in Fig. 6 are all non-Gaussian. From Fig. 6(a), the soil pH by KCl solution, pHKCl, is basically symmetric around its spatial trend; from Fig. 6(b), relative to the spatial trend, the calcium content density Ca is positively skewed with a thicker right-hand tail; from Fig. 6(c), the potassium content K may be seen to be basically symmetric around its spatial trend or only slightly positively skewed; from Fig. 6(d), relative to the spatial trend, the hydrogen content density H is obviously negatively skewed with a thicker left-hand tail; from Fig. 6(e), the property of the carbon content C is more complex having at least two peaks with the negatively located peak higher than the positive peak; and finally, from Fig. 6(f), the cation exchange capability CTC is basically negatively skewed around its spatial trend. These properties would be helpful for soil management and conservation.

Acknowledgements

The authors are grateful to the Editor, Associate Editor and two referees for their insightful comments and suggestions, which have led to a great improvement of this present version. This research was partially supported by European Research Agency's Marie Curie career integration grant PCIG14-GA-2013-631692, and Ling's research by the National Social Science Fund of China (grant 14ATJ005) and the National Natural Science Foundation of China (grant 11171001), which are acknowledged.

References

Abramson, I. (1982a) On bandwidth variation in kernel estimates—a square root law. *Ann. Statist.*, **10**, 1217–1223.
Abramson, I. (1982b) Arbitrariness of the pilot estimator in adaptive kernel methods. *J. Multiv. Anal.*, **12**, 562–567.
Basawa, I. V. (ed.) (1996a) Special issue on spatial statistics: Part 1. *J. Statist. Plann Inf.*, **50**, 311–411.
Basawa, I. V. (ed.) (1996b) Special issue on spatial statistics: Part 2. *J. Statist. Plann Inf.*, **51**, 1–97.
Bowman, A. W. and Azzalini, A. (1997) *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations*. Oxford: Oxford University Press.
Bowman, A. W. and Azzalini, A. (2014) R package 'sm': nonparametric smoothing methods (version 2.2-5.4). University of Glasgow, Glasgow. (Available from `http://www.stats.gla.ac.uk/ adrian/sm` or `http://azzalini.stat.unipd.it/Book_sm`.)
Cao, R., Cuevas, A. and Gonzalez-Manteiga, W. (1994) A comparative study of several smoothing methods in density estimation. *Computnl Statist. Data Anal.*, **17**, 153–176.
Carbon, M., Hallin, M. and Tran, L. T. (1996) Kernel density estimation for random fields: L_1 theory. *J. Nonparam. Statist.*, **6**, 157–170.
Chiu, S.-T. (1996) A comparative review of bandwidth selection for kernel density estimation. *Statist. Sin.*, **6**, 129–145.
Cressie, N. (1993) *Statistics for Spatial Data*, revised edn. New York: Wiley.
Davies, T. M. and Hazelton, M. L. (2010) Adaptive kernel estimation of spatial relative risk. *Statist. Med.*, **29**, 2423–2437.
Directorate Agricultural Information Services (2007) Acid soil and lime. Department of Agriculture, Pretoria. (Available from `http://www.nda.agric.za/docs/Infopaks/lime.pdf`.)
Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.
Fan, J. and Yao, Q. (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer.
Faraway, J. I. and Jhun, M. (1990) Bootstrap choice of bandwidth for density estimation. *J. Am. Statist. Ass.*, **85**, 1119–1122.
Gao, J. (2007) *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. New York: Chapman and Hall.
Gao, J., Lu, Z. and Tjøstheim, D. (2006) Estimation in semi-parametric spatial regression. *Ann. Statist.*, **34**, 1395–1435.
Gao, J., Lu, Z. and Tjøstheim, D. (2008) Moment inequalities for spatial processes. *Statist. Probab. Lett.*, **78**, 687–697.
Gelfand, A. E., Diggle, P. J., Fuentes, M. and Guttorp, P. (2010) *Handbook of Spatial Statistics*. Boca Raton: CRC Press.
Guo, D. and Mennis, J. (2009) Spatial data mining and geographic knowledge discovery—an introduction. *Comput. Environ. Urb. Syst.*, **33**, 403–408.
Guyon, X. (1995) *Random Fields on a Network: Modeling, Statistics, and Application*. New York: Springer.
Hall, P. (1983) Large sample optimality of least squares cross-validation in density estimation. *Ann. Statist.*, **11**, 1156–1174.
Hall, P. and Johnstone, I. (1992) Empirical functionals and efficient smoothing parameter selection (with discussion). *J. R. Statist. Soc.* B, **54**, 475–530.
Hallin, M., Lu, Z. and Tran, L. T. (2001) Density estimation for spatial linear processes. *Bernoulli*, **7**, 657–668.
Hallin, M., Lu, Z. and Tran, L. T. (2004a) Kernel density estimation for spatial processes: the L_1 theory. *J. Multiv. Anal.*, **88**, 61–75.
Hallin, M., Lu, Z. and Tran, L. T. (2004b) Local linear spatial regression. *Ann. Statist.*, **32**, 2469–2500.
Hallin, M., Lu, Z. and Yu, K. (2009) Local linear spatial quantile regression. *Bernoulli*, **15**, 659–686.
Härdle, W. and Vieu, P. (1992) Kernel regression smoothing of time series. *J. Time Ser. Anal.*, **13**, 209–232.
Harel, M., Lenain, J.-F. and Ngatchou-Wandji, J. (2016) Asymptotic behaviour of binned kernel density estimators for locally non-stationary random fields. *J. Nonparam. Statist.*, **28**, 296–321.
Hart, J. D. and Vieu, P. (1990) Data-driven bandwidth choice for density estimation based on dependent data. *Ann. Statist.*, **18**, 873–890.
Jenish, N. (2012) Nonparametric spatial regression under near-epoch dependence. *J. Econmetr.*, **167**, 224–239.
Jones, M. C. (1991) The roles of ISE and MISE in density estimation. *Statist. Probab. Lett.*, **12**, 51–56.
Jones, M. C., Marron, J. S. and Sheather, S. J. (1996) A brief survey of bandwidth selection for density estimation. *J. Am. Statist. Ass.*, **91**, 401–407.
Kim, T. Y. and Cox, D. D. (1997) A study on bandwidth selection in density estimation under dependence. *J. Multiv. Anal.*, **62**, 190–203.
Lemke, D., Mattauch, V., Heidinger, O., Pebesma, E. and Hense, H.-W. (2015) Comparing adaptive and fixed bandwidth-based kernel density estimates in spatial cancer epidemiology. *Int. J. Hlth Geog.*, **14**, article 15.
Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J. and Bretagnolle, V. (2014) Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. *Globl Ecol. Biogeog.*, **23**, 811–820.
Loader, C. R. (1999) Bandwidth selection: classical or plug-in? *Ann. Statist.*, **27**, 415–438.

Lu, Z., Tang, Q. and Cheng, L. (2014) Estimating spatial quantile regression with functional coefficients: a robust semiparametric framework. *Bernoulli*, **20**, 164–189.
Lu, Z. and Tjøstheim, D. (2014) Nonparametric estimation of probability density functions for irregularly observed spatial data. *J. Am. Statist. Ass.*, **109**, 1546–1564.
Mammen, E. (1990) A short note on optimal bandwidth selection for kernel estimators. *Statist. Probab. Lett.*, **9**, 23–25.
Marron, J. S. (1985) An asymptotically efficient solution to the bandwidth problem of kernel density estimation. *Ann. Statist.*, **13**, 1011–1023.
Marron, J. S. (1987) A comparison of cross-validation techniques in density estimation. *Ann. Statist.*, **15**, 152–162.
Marron, J. S. (1992) Bootstrap bandwidth selection. In *Exploring the Limits of Bootstrap* (eds R. LePage and L. Billard), pp. 249–262. New York: Wiley.
Marron, J. S. and Härdle, W. (1986) Random approximations to some measures of accuracy in nonparametric curve estimation. *J. Multiv. Anal.*, **20**, 91–113.
Neaderhouser, C. C. (1980) Convergence of blocks spins defined on random fields. *J. Statist. Phys.*, **22**, 673–684.
Nordman, D. J. and Lahiri, S. N. (2004) On optimal spatial subsample size for variance estimation. *Ann. Statist.*, **32**, 1981–2027.
Pagan, A. and Ullah, A. (1999) *Nonparametric Econometrics*. New York: Cambridge University Press.
Quintela-del-Rio, A. (1996) Comparison of bandwidth selectors in nonparametric regression under dependence. *Computnl Statist. Data Anal.*, **21**, 563–580.
R Development Core Team (2015) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Ribeiro, Jr, P. J. and Diggle, P. J. (2016) geoR: analysis of geostatistical data. *R Package Version 1.7-5.2*. (Available from `http://CRAN.R-project.org/package=geoR`.)
Robinson, P. (2008) Developments in the analysis of spatial data. *J. Jpn Statist. Soc.*, **38**, 87–96.
Robinson, P. (2011) Asymptotic theory for nonparametric regression with spatial data. *J. Econmetr.*, **165**, 5–19.
Ruppert, D., Sheather, S. J. and Wand, M. P. (1995) An effective bandwidth selector for local least squares regression. *J. Am. Statist. Ass.*, **90**, 1257–1270.
Scott, D. W. (1992) *Multivariate Density Estimation: Theory, Practice and Visualization*. New York: Wiley.
Sheather, S. J. and Jones M. C. (1991) A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Statist. Soc.* B, **53**, 683–690.
Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall.
Stone, C. J. (1984) An asymptotically optimal window selection rule for kernel density estimation. *Ann. Statist.*, **12**, 1285–1297.
Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions (with discussion). *J. R. Statist. Soc.* B, **36**, 111–147.
Tran, L. T. (1990) Kernel density estimation on random fields. *J. Multiv. Anal.*, **34**, 37–53.
Xia, Y. and Li, W. K. (2002) Asymptotic behavior of bandwidth selected by cross-validation method under dependence. *J. Multiv. Anal.*, **83**, 265–287.
Zheng, Y., Zhu, J. and Roy, A. (2010) Nonparametric Bayesian inference for the spectral density function of a random field. *Biometrika*, **97**, 238–245.
Zhu, J., Huang, H.-C. and Reyes, P. E. (2010) On selection of spatial linear models for lattice data. *J. R. Statist. Soc.* B, **72**, 389–402.
Zhu, J. and Morgan, G. D. (2004) A nonparametric procedure for analyzing repeated-measures of spatially correlated data. *Environ. Ecol. Statist.*, **11**, 431–443.

Supporting information
Additional 'supporting information' may be found in the on-line version of this article:

'Web-based supplementary materials for "On bandwidth choice for spatial data density estimation"'.