An evaluation of likelihood-based bandwidth selectors for spatial and spatiotemporal kernel estimates

Tilman M. Davies a and Andrew B. Lawson b

aDepartment of Mathematics & Statistics, University of Otago, Dunedin, New Zealand; bDepartment of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina, USA

ABSTRACT
Spatial point pattern data sets are commonplace in a variety of different research disciplines. The use of kernel methods to smooth such data is a flexible way to explore spatial trends and make inference about underlying processes without, or perhaps prior to, the design and fitting of more intricate semiparametric or parametric models to quantify specific effects. The long-standing issue of ‘optimal’ data-driven bandwidth selection is complicated in these settings by issues such as high heterogeneity in observed patterns and the need to consider edge correction factors. We scrutinize bandwidth selectors built on leave-one-out cross-validation approximation to likelihood functions. A key outcome relates to previously unconsidered adaptive smoothing regimens for spatiotemporal density and multitype conditional probability surface estimation, whereby we propose a novel simultaneous pilot-global selection strategy. Motivated by applications in epidemiology, the results of both simulated and real-world analyses suggest this strategy to be largely preferable to classical fixed-bandwidth estimation for such data.

ARTICLE HISTORY
Received 9 September 2018
Accepted 22 January 2019

KEYWORDS
Cross-validation; density estimation; geographical epidemiology; optimal bandwidth; point pattern

2010 MATHEMATICS SUBJECT CLASSIFICATIONS
62H11; 62G07

1. Introduction
The analysis of spatial point process data usually involves estimation of the continuous density or intensity function deemed to have given rise to the observed pattern. In many research disciplines it is common to encounter high heterogeneity in observed patterns, with points clustered densely in some areas and dispersed sparsely elsewhere. Studies in geographical epidemiology, for example, might note tight clusters of disease occurrences in populated urban centres, with comparatively far fewer observations made in rural regions.

Appealing flexibility is therefore found in nonparametric techniques – multivariate kernel smoothing methods in particular [1,2] – for spatial density/intensity estimation. While in isolation nonparametric methods are not equipped to quantify, say, specific covariate effects on the process of interest, they are a powerful tool in initial data exploration and
visualization; in making comparative inferences between different types of points [3–6]; and as a building block for more complicated semiparametric models [7–9].

The issue of appropriate bandwidth selection is of paramount importance in applications of kernel methods, impacting greatly on the resulting spatial and spatiotemporal surfaces. Data driven selection methods are complicated in this setting by the need to take into account corrective factors to combat boundary bias. In recent years, spatially adaptive kernel estimation [10] has been shown to be particularly useful in spatial applications, with the potential to substantially outperform the classical fixed bandwidth estimator in certain settings [7,11,12]. Bandwidth selection is even more difficult for such estimators, however, with the need to specify both a pilot and a global bandwidth to compute the variable bandwidth factors.

In this work we address the issue of bandwidth selection for kernel estimators of spatial and spatiotemporal point patterns, motivated by smoothing problems in epidemiology. Specifically, we focus on data-driven selection methods built on approximations to likelihood functions, the fundamental ideas being readily interpretable and applicable to several variants of kernel estimators. In so doing, we introduce adaptive kernel estimators for constructing conditional probability surfaces of multitype patterns, and for spatiotemporal densities. We also propose using a common pilot-global bandwidth selection strategy for all adaptive estimators considered. The methodological exposé is interspersed with a comprehensive collection of comparative simulations.

The article is structured as follows. Standalone spatial density estimation is examined in Section 2, including fixed and adaptive estimators, and the concept of likelihood-based bandwidth selection is introduced. Simulations for these estimators are designed and executed in Section 3. In Section 4 we consider spatial relative risk via conditional probability surface estimation; the accompanying simulations following in Section 5. Spatiotemporal density estimators are discussed in Section 6 and numerical performance is compared in Section 7. We apply the common pilot-global selection strategies to a pair of real-data examples in Section 8. Concluding remarks and suggestions for future research appear in Section 9.

2. Spatial density estimation

In what follows, suppose we observe a point pattern \( X = \{x_1, \ldots, x_n\} \) falling within some bounded region \( W \subset \mathbb{R}^2 \). The goal is estimation of \( f \), the density function on \( W \) assumed to have given rise to \( X \). A common assumption is that \( X \) is a realization of a Poisson point process on \( W \), in which case the kernel density estimator is directly proportional to the same estimator of the process intensity, which yields the expected number of observations per unit area.

2.1. Fixed bandwidth

The classical 2D kernel estimator [1,2] of the unknown density is given as

\[
\hat{f}_X(x | h) = n^{-1}h^{-2} \sum_{u \in X}K \left( \frac{x-u}{h} \right) q_W(x; h)^{-1}; \quad x \in W.
\]
In (1), $h > 0$ is the bandwidth or smoothing parameter; and $K(\cdot)$ is the kernel function taken as a zero-mean probability density function (pdf). We shall assume a Gaussian $K$ throughout as is often used in spatial applications. Furthermore,

$$q_W(x; h) = h^{-2} \int_W K\left(\frac{y-x}{h}\right) \, dy$$

(2)

is an edge-correction factor in place to mitigate the potentially serious negative bias that can occur in $\tilde{f}$ at the boundaries of $W$. Edge-correction is a non-trivial issue, particularly so for spatial applications where it is common for $W$ to form one (or even several) irregular polygonal shapes. While there are different forms of such correction factors [13,14], we shall for simplicity focus on the form of $q_W$ as given in (1) and (2), where the factor is found as the proportion of mass of the kernel $K$ centred on the query location $x$ (and with bandwidth $h$) that falls within $W$. In any case, it is important to note that any such solutions are approximative in nature and merely mitigate, as opposed to completely annihilate, the adverse effects of estimation at region boundaries.

In papers that deal more generally with multivariate density estimation, $h$ above is often replaced by an appropriately dimensioned, symmetric positive definite matrix. For data in $\mathbb{R}^2$, this would entail specification of a $2 \times 2$ matrix with up to three unique components allowing differing degrees of smoothing along each axis, as well as altering the orientation of the kernel. In this work, motivated by applications in spatial epidemiology, we restrict attention to isotropic smoothing via a scalar bandwidth. Imparting equal amounts of smoothing in all directions in such analyses is generally preferable in the absence of external information defining the disease under scrutiny or significant prevailing geographic features. Computational complexity is also a factor that cannot be ignored, especially for adaptive smoothing (Section 2.2), and use of a scalar bandwidth helps minimize the cost of relevant operations.

It is universally agreed that a ‘sensible’ choice of bandwidth, i.e. one that is neither excessively large nor impractically small, is crucial in obtaining a reliable estimate of the unknown density. As such, the development of data-driven bandwidth selectors which aim to optimize this balance with respect to some appropriate criterion has been a key research focus for many years. While the bulk of early efforts focused on selectors for univariate data, for which there exist several suitable methods [2,15,16], not all are straightforwardly adaptable to the multivariate setting [17]. This is particularly true when we consider the aforementioned issue of boundary bias – approximative edge-correction techniques such as (2) are frequently absent from papers that deal with multivariate kernel density estimation in a more general sense. As a result, most bandwidth selection methods are not developed with edge-correction in mind, and the corresponding practical and empirical comparisons do not concern themselves with the bounded region upon which the data are generated.

A popular approach to bandwidth selection that can be applied in the spatial and spatiotemporal settings is likelihood cross-validation. This technique is based on a leave-one-out approximation derived from $D(\cdot, \cdot)$, the Kullback-Leibler (KL) information, a measure of disparity between two densities. With respect to the density estimate $\tilde{f}$ as per (1) and the true target density $f$, we have
\[
D(f, \tilde{f}_X; h) = \int_W \log \left\{ \frac{f(x)}{\tilde{f}_X(x | h)} \right\} f(x) \, dx
\]
\[
= \int_W \log \{ f(x) \} f(x) \, dx - \int_W \log \{ \tilde{f}_X(x | h) \} f(x) \, dx. \tag{3}
\]

Minimisation of (3) with respect to \( h \) is therefore equivalent to maximization of

\[
\mathbb{E} \log \{ \tilde{f}_X \} = \int_W \log \{ \tilde{f}_X(x | h) \} f(x) \, dx,
\]

which can be approximated via leave-one-out averaging as

\[
\text{FIXLIK}(h) = n^{-1} \sum_{i=1}^{n} \log \{ \tilde{f}_{X[-i]}(x_i | h) \} \approx \mathbb{E} \log \{ \tilde{f}_X \}, \tag{4}
\]

where \( X^{[-i]} \equiv X \setminus x_i \), such that

\[
\tilde{f}_{X[-i]}(x_i | h) = (n-1)^{-1} h^{-2} \sum_{u \in X \setminus x_i} K \left( \frac{x_i - u}{h} \right) q_W(x_i; h)^{-1} \tag{5}
\]
is the edge-corrected density estimate evaluated at the location of the \( i \)th observation assuming its omission from \( X \). Thus, finding \( \arg \max_h \text{FIXLIK}(h) \) is a direct data-driven method by which the smoothing parameter can be optimized for a fixed bandwidth kernel density estimate.

The result in (5), barring the presence of the corrective factor \( q_W \), is well established in the literature [1]. However, the role such likelihood-based bandwidth selection can play in more intricate smoothing applications, such as in variable-bandwidth and spatial relative risk estimation, has yet to receive any attention in the literature.

### 2.2. Adaptive bandwidth

Allowing the amount of smoothing to vary leads to an adaptive or variable-bandwidth estimator, certain forms of which have been shown to be superior to fixed bandwidth estimators in various scenarios. For highly heterogeneous spatial point patterns, such as those often encountered in geographical epidemiology, ‘Abramson’ observation-specific adaptation [10,18] performs particularly well. Under this smoothing regimen, (1) becomes

\[
\hat{f}_X(x | h_0, h_p) = n^{-1} \sum_{u \in X} h_X(u; h_0, h_p)^{-2} K \left( \frac{x - u}{h_X(u; h_0, h_p)} \right) \times q_W(x; h_X(x; h_0, h_p))^{-1}; \quad x \in W, \tag{6}
\]

where the isotropic smoothing parameter varies according to the following expression:

\[
h_X(u; h_0, h_p) = h_0 \gamma_{\hat{f}}^{-1} \tilde{f}_X(u | h_p)^{-0.5}. \tag{7}
\]

Here, note that a bandwidth at any location \( x \in W \) is inversely proportional to the square root of a fixed-bandwidth density estimate of the observed data \( X \). This estimate is found
using a pilot bandwidth $h_p$; the pilot density must also be clipped at some nominal value $> 0$ to prevent excessively large variable bandwidths adversely affecting the final smooth [18]. The required global bandwidth $h_0$ sets an overall level of smoothing for the variable bandwidths. Finally, $\gamma_f = \exp\{n^{-1} \sum_{u \in X} \log \hat{f}_X(u | h_p)^{-0.5}\}$ is the geometric mean of the inverse square root pilot density values at each observation. Its inclusion in (7) is optional, but is useful in practice as it allows the global bandwidth $h_0$ to be considered on the same scale as the pilot bandwidth $h_p$ [1].

Additionally noteworthy in (6) is the modification to the edge-correction factor, which now depends on the variable bandwidth function (7) [19]. While the kernel estimator itself requires a bandwidth at each observation $u \in X$, the edge-correction factors require the 'hypothetical' adaptive bandwidths at all desired evaluation coordinates $x \in W$, conditional upon the observed data $X$.

Abramson adaptation decreases the relative amount of smoothing in areas of high pilot density, where points are more densely grouped, and increases smoothing for observations that lie in relative isolation where the pilot density is small. In the final density estimate $\hat{f}$, this has the effect of preserving more detail in areas where we have more data, while mitigating the potential for undersmoothing in sparsely populated regions. Compelling theoretical and practical benefits have been noted in comparison to standalone fixed-bandwidth estimation [12].

Potential advantages of adaptive smoothing over fixed smoothing notwithstanding, we are still faced with a bandwidth selection problem in (6). Indeed, this problem is now more complex than earlier – we require some optimal selection of $(h_p, h_0)$, preferably jointly so. Furthermore, recent work has shown certain approximations based on asymptotic theory are not practically useful for this task [20]. As such, applications to date have largely seen the pilot and global bandwidths set subjectively or in an overtly ad-hoc manner [7,11,21]. Finding a data-driven method to choose these values simultaneously and objectively is desirable.

With this in mind, we may consider an appropriate alteration to (4), and identify a naive, joint pilot-global bandwidth selector for the adaptive kernel estimator based on

$$\text{ADALIK}_2(h_0, h_p) = n^{-1} \sum_{i=1}^n \log \left\{ \hat{f}_{\hat{f}^{-1}}(x_i | h_0, h_p) \right\} \approx \mathbb{E} \log \left\{ \hat{f}_X \right\},$$  \hspace{1cm} (8)$$

where we choose the values of $(h_p, h_0)$ that satisfy $\arg\max_{h_p, h_0} [\text{ADALIK}_2(h_0, h_p)]$; the subscript ‘2’ utilized to emphasize bivariate optimization over the two scalar smoothing parameters. Note that, similar to (5), we define

$$\hat{f}_{\hat{f}^{-1}}(x_i | h_0, h_p) = (n - 1)^{-1} \sum_{u \in X \setminus x_i} h_{\hat{f}^{-1}}(u; h_0, h_p)^{-2} K\left( \frac{x_i - u}{h_{\hat{f}^{-1}}(u; h_0, h_p)} \right)$$

$$\times q_W(x_i; h_{\hat{f}^{-1}}(x_i; h_0, h_p))^{-1},$$  \hspace{1cm} (9)$$

where

$$h_{\hat{f}^{-1}}(x_i; h_0, h_p) = h_0 \gamma_f^{-1} \hat{f}_{\hat{f}^{-1}}(x_i | h_p)^{-0.5}.$$  \hspace{1cm} (10)$$
Crucially, (9) and (10) reveal that the leave-one-out operation affects both the main kernel sum and the bandwidth function, due to the dependence of the latter on the pilot density estimate. This presents a computational challenge for even modest sample sizes.

We refer to ADALIK2 as ‘naive’ because of the complex relationship between the pilot and global bandwidths in the Abramson estimator. It is reasonable to anticipate identifiability issues in such unconstrained optimization: The effect of a small pilot bandwidth could be tempered by a large global bandwidth, and vice-versa. Just how convoluted the two bandwidths are with respect to the optimization criterion is difficult to formally ascertain, and we turn to the simulations in the latter sections of the paper to shed light on this phenomenon.

In anticipation of this potential problem, we also propose a constrained version of (8), where we force $h_c \equiv h_0 = h_p$, and define

$$
\text{ADALIK}_1 (h_c) = n^{-1} \sum_{i=1}^{n} \log \left\{ \hat{f}_{X \mid \cdot} (x_i \mid h_c, h_c) \right\},
$$

which uses (9) and (10) with the shared $h_c$. While setting the pilot and global bandwidths to be equal reduces the flexibility of the adaptive smoother in general, it is of interest to determine if the optimal choice – $\arg\max_{h_c} [\text{ADALIK}_1 (h_c)]$ – still provides a satisfactory final estimate, particularly in comparison to a fixed-bandwidth estimate.

3. Simulation study 1: spatial densities

Our first round of simulations will examine standalone density estimation using the likelihood bandwidth selectors for the estimators given in the previous section. We define four scenarios, S1 through S4. They are plotted in the top row of Figure 1; formal details and corresponding plots of example data sets appear in the supplement. From S1 to S3, which are defined as Gaussian mixtures, we move from a relatively smooth setting with mild spatial fluctuations to a more heterogeneous scenario. S4 is based on a single realization of a log-Gaussian Cox process [22], which is able to provide a ‘less smooth’ true scenario than Gaussian mixtures naturally permit. It is in the latter scenarios that we expect adaptive estimation to perform better than fixed, owing to the tendency of a constant bandwidth to simultaneously undersmooth and oversmooth different areas of the spatial region depending on local point concentrations.

Repeated over 100 iterations we simulate a Poisson process realization of 500 points for each scenario. The FIXLIK, ADALIK2 and ADALIK1 selectors are engaged to find optimal bandwidths for the relevant fixed or adaptive density estimate. In addition to these bandwidths, and to provide a ‘baseline’ performance upon which to improve, we also compute Terrell’s oversmoothing bandwidth [23], a rule-of-thumb measure that provides an estimate of the upper bound on the amount of fixed-bandwidth smoothing compatible with the estimated scalar standard deviation of the data $\hat{\sigma}$; $h_{OS} = \hat{\sigma} \left\{ 625 / (384n) \right\}^{1/6}$. We reference these results as ‘FIXOS’.

Performance itself is measured using the integrated squared error between a given density estimate $g$ and the true density according to the corresponding scenario $f$; $\text{ISE} = \int_w (g - f)^2$, as well as the KL distance (3). The bottom row of Figure 1 shows boxplots
Figure 1. The four synthetic scenarios (top row, logarithmic color scales) and ISE results (bottom row, corresponding column, logarithmic y-axis scales) for Simulation Study 1.

Figure 2. Distributions of optimal bandwidths for selectors utilized in Simulation Study 1. Top row: Results for selectors FIXOS, FIXLIK, and ADALIK1. Bottom row: Selected pilot and global bandwidths for ADALIK2.

of the ISE measures for each of S1 through S4; the KL results tell the same story and are deferred to the supplement.

Adaptive estimation with a constrained pilot-global bandwidth $h_c$ chosen via ADALIK1 is generally the better performer for all scenarios. Its improvement over either FIXOS or FIXLIK is more obvious in the latter three scenarios where point spreads and overall heterogeneity becomes more variable; one could argue that fixed bandwidth smoothing is comparatively acceptable in situations like S1. Additionally we note poor performance of ADALIK2; due at least in part to the anticipated inability of the unconstrained adaptive selector to tease apart the effects of $h_0$ and $h_p$. This can be seen in Figure 2, which shows the distributions of selected bandwidths. While the curves for FIXOS, FIXLIK, and ADALIK1 provide guides as to good ranges of bandwidth values, there is high variability in the ADALIK2 selections. This is extreme for S1, for example, where pilot bandwidths erring toward zero are attempted to be compensated for by massive global bandwidths. Even for
S4, which shows the best performance for ADALIK2 out of the four (Figure 1), the high variability of the chosen pilot and global bandwidths render its overall performance far less stable than in the constrained setting.

4. Spatial relative risk and conditional probability surfaces

A common extension to standalone density estimation is to compare the estimated spatial densities of different types of observations collected in the same region. For example, researchers are often interested in estimating spatial disease risk after adjusting for the natural background heterogeneity of the population under scrutiny. The density estimates are based on an observed pattern of confirmed ‘case’ points, and another of at-risk ‘control’ points [3,4,24,25]. This concept can be further extended to situations in which there are more than two different categories of points, such as the comparison of spatial patterning of several different subtypes of a certain disease [26,27].

4.1. Notation and assumptions

Fully nonparametric assessment of any such spatial segregation (or lack thereof) exhibited by the distributions of different types of observations is typically performed under certain assumptions. The conditions that follow mirror those laid out Diggle et al. [27].

Suppose we are interested in comparing the spatial distributions of \( m \) different types of disease, all affecting an at-risk population taken to follow a Poisson point process with intensity function \( \eta_0(x); x \in W \subseteq \mathbb{R}^2 \). Furthermore, let \( \eta_k(x); k = 1, \ldots, m \) represent the intensity functions of \( m \) independent Poisson point processes on \( W \), corresponding to each of the \( m \) different types of observations, whereby \( \eta_k(x) = \rho_k(x) \eta_0(x) \). We interpret the functions \( \rho_k(x) \) as the probability a point at \( x \) will yield an observation of type \( k \).

Of interest is estimation of the \( m \) surfaces \( p_k(x) \), which are interpreted as the conditional probability that an observation is of type \( k \ given \) a point has occurred at location \( x \):

\[
p_k(x) = \frac{\eta_k(x)}{\sum_{j=1}^{m} \eta_j(x)}; \quad x \in W, \ k = 1, \ldots, m. \tag{12}
\]

Note that the assumption of a common spatial intensity \( \eta_0(x) \) in the expression for \( \eta_k(x) \) above is important because it is revealed to cancel out in the ratio (12). This subsequently permits us to estimate the type-specific probabilities \( p_k(x) \) using kernel estimation of the type-specific data sets only, in other words, without specifying any particular spatial form of the at-risk population in \( W \).

In case-control settings we have \( m = 2 \) and may use (12) to obtain the conditional probability of a case, given an observation at \( x \). A reformulation of this approach in such a ‘two-type’ scenario sees the relative risk of disease directly estimated as the ratio of the case density to the control density [4,24,28]. Such density-ratio estimators have been used in a variety of applications [21,29–32] though are less capable of handling the multitype scenario when \( m > 2 \), and do not naturally lend themselves to likelihood-based bandwidth selection. We therefore focus on estimators for the conditional probability surfaces of \( p_k \) in the remainder of this section.
4.2. Fixed estimator

Let $X_k; k = 1, \ldots, m$ represent realisations of the $m$ Poisson point processes on $W \subset \mathbb{R}^2$, with corresponding sample sizes $n_1, \ldots, n_m$. We collect all $n = \sum_k n_k$ points into a single data set $X = \{x_1, \ldots, x_n\}$, and define a corresponding vector $Y = \{y_1, \ldots, y_n\}$ that identifies the type of each point i.e. $y_i = k$ if the $i$th observation is of type $k$.

The fixed-bandwidth estimator of (12) can be found as

$$
\tilde{p}_k(x \mid h, X) = \frac{n_k \tilde{f}_{X_k}(x \mid h)}{n f_X(x \mid h)} = \frac{\sum_{u \in X_k} K\left(\frac{x-u}{h}\right)}{\sum_{u \in X} K\left(\frac{x-u}{h}\right)},
$$

where \(\tilde{f}_X(\cdot)\) is the fixed-bandwidth density estimator (1). This estimator for $p_k$ can be seen as proportional to the quotient of the kernel density estimate of the points of type $k$ only and the kernel density estimate of the pooled dataset of all spatial observations.

While choosing optimal bandwidths for each separate density estimate in (13) is possible use of a common bandwidth $h$ throughout is preferable [27]: A common bandwidth ensures all type-specific probabilities sum to unity at any given location; i.e. $\sum_k \tilde{p}_k(x \mid h, X) = 1$. As can be seen in the rightmost term in (13), a common bandwidth also obviates the need to calculate edge-correction factors $q_W$ per (2) as seen in (1) – their equivalence in the ratio leads to cancellation [33,34].

A bandwidth selector can be identified by recognizing the vector $Y$ as a realization of $n$ multinomial outcomes. These are associated with their observation locations $X$ which in turn reference the probabilities of success over the $m$ categories through the estimated $\tilde{p}_k(x_i \mid h, X)$. The leave-one-out criterion for choosing an optimal $h$ based on the log-likelihood of the multinomial distribution is therefore

$$
\text{FIXMULT}(h) = \sum_{i=1}^n \sum_{k=1}^m I(y_i = k) \log \left\{ \tilde{p}_k(x_i \mid h, X^{[-i]}) \right\},
$$

which sees (13) evaluated at $x_i$ given its deletion from $X$.

4.3. Adaptive estimator

Spatially adaptive estimation for nonparametric comparisons of spatial distributions is appealing for many of the same reasons noted in Section 2.2. Abramson adaptation has been suggested for the $m=2$ density-ratio method [11], but has not previously been considered for the kernel regression estimator of conditional probability surfaces.

Choosing a universally common global bandwidth $h_0$ and pilot bandwidth $h_p$, direct substitution of adaptive kernel density estimates (6) in place of their fixed counterparts in (13) leads to the following:

$$
\hat{p}_k^*(x \mid h_0, h_p, X) = \frac{n_k \hat{f}_{X_k}(x \mid h_0, h_p)}{n \hat{f}_X(x \mid h_0, h_p)}
= \frac{\sum_{u \in X_k} h_{X_k}(u; h_0, h_p)^{-2} K\left(\frac{x-u}{h_{X_k}(u; h_0, h_p)}\right) q_W(x; h_{X}(x; h_0, h_p))}{\sum_{u \in X} h_{X}(u; h_0, h_p)^{-2} K\left(\frac{x-u}{h_{X}(u; h_0, h_p)}\right) q_W(x; h_{X_k}(x; h_0, h_p))}.
$$
The above shows that adaptive smoothing of such surfaces requires additional care with respect to bandwidth assignments. Even if common global and pilot bandwidths are chosen for all individual density estimates, this is not enough to yield the favoured behaviour of total summation of the \( m \) probabilities to 1 at a given spatial site, nor does it circumvent the need to edge-correct. This is owed to the reliance of the adaptive density estimate on the pilot density, which, being based on different sets of spatial data in numerator and denominator positions (\( X_k \) and \( X \) respectively) of \( \tilde{p}_k^* \), produces different bandwidth factors on either side of the fraction; cf. (7).

We can impose the desired behaviour by basing the bandwidth function on a common dataset, meaning the bandwidth factors at the pilot estimation stage are identical in numerator and denominator at any given location \( x \). This is akin to symmetric adaptive smoothing regimens discussed for density-ratio estimators of relative risk [35]. First, extend slightly the notation of the adaptive estimator (6) to highlight this modification,

\[
\hat{f}_B(x | h_0, h_p, A) = n^{-1} \sum_{u \in B} h_A(u; h_0, h_p)^{-2} K \left( \frac{x - u}{h_A(u; h_0, h_p)} \right) \times q_W(x; h_A(x; h_0, h_p))^{-1}; \quad x \in W, (15)
\]

where \( A \in \mathbb{R}^2 \) and \( B \in W \) are data sets of \( n_A \) and \( n_B \) spatial observations respectively. The above shows that the density estimate is constructed by variable-bandwidth kernels situated at the observations of \( B \), but that the variable bandwidths at each \( u \in B \) are found using pilot estimation referencing \( A \). Note that either \( A \subset B \) or \( A \supset B \) is permitted, and if \( A = B \) then (15) simplifies to (6).

While (15) itself is no longer optimal in the sense of standalone density estimation via Abramson adaptation, it has been shown to possess excellent practical behaviour in ratio-based density estimators, particularly when \( A \) is chosen ‘sensibly’, for example, as a superset of \( B \) [35]. Similar benefits might be reasonably anticipated for the conditional probability estimator, though it remains to be seen if a high degree of spatial segregation adversely affects performance.

We further take into account the highly favourable simulation results of bandwidth selection based on ADALIK1 shown in Section 3, and hold pilot and global bandwidths equal at a single value \( h_c \). Our proposed symmetric constrained adaptive conditional probability estimator is therefore given as

\[
\hat{p}_k(x | h_c, X) = \frac{n_k f_X(x | h_c, h_c, X)}{n f_X(x | h_c, h_c, X)} = \frac{\sum_{u \in X_k} h_X(u; h_c)^{-2} K \left( \frac{x - u}{h_X(u; h_c)} \right)}{\sum_{u \in X} h_X(u; h_c)^{-2} K \left( \frac{x - u}{h_X(u; h_c)} \right)}, (16)
\]

where equivalent edge-correction factors have cancelled out, and the imposed symmetry ensures \( \sum_{k=1}^m \hat{p}_k(x | h_c, X) = 1 \). Note in our implementation the shared bandwidth factors are taken to be based on the pooled observations \( X \) regardless of corresponding type in \( Y \).

Optimization of \( h_c \) via the multinomial likelihood may thus proceed in a manner identical to (14). We seek \( \text{argmax}_{h_c} [\text{ADAMULT}(h_c)] \), where

\[
\text{ADAMULT}(h_c) = \sum_{i=1}^n \sum_{k=1}^m I(y_i = k) \log \left\{ \hat{p}_k(x_i | h_c, X^{[-i]}) \right\}. (17)
\]
5. Simulation study 2: spatial conditional probabilities

The numeric evaluation of the fixed and symmetric constrained adaptive conditional probability estimators of the previous section proceeds as follows. We base four scenarios on two themes, mild segregation and strong segregation. For each theme, we specify two multitype problems; one with \( m = 2 \), the other with \( m = 3 \). Figure 3 provides plots of these designs; where the \( m = 2 \) scenarios are made up of densities A and B, and the \( m = 3 \) scenarios are made up of A, B and C. We label the four multitype scenarios MS1 through MS4, respectively referencing \( m = 2 \) and \( m = 3 \) mild; and \( m = 2 \) and \( m = 3 \) strong themes.

The densities A, B and C are used to define the corresponding type-specific probability surfaces \( p_A \) and \( p_B \) for MS1 and MS3, and \( p_A \), \( p_B \) and \( p_C \) for MS2 and MS4 as per (12). Figure 4 displays the surfaces of \( p_A \) for the \( m = 2 \) scenarios (the \( p_B \) surface is simply its pixel-wise complement) as well as all three type-specific probability surfaces \( p_A \), \( p_B \), and \( p_C \) for the \( m = 3 \) scenarios. These are computed assuming sample sizes of 500, 1000, and 650 for point types A, B and C respectively.

Results are reported for 100 iterations in Figure 5. The top row shows the ISEs between the estimated type-specific probability surfaces and the true surfaces; note that for the \( m = 2 \) scenarios, MS1 and MS3, identical errors arise for both \( p_A \) and \( p_B \), so only one set for each bandwidth selector/estimator is reported. Across all scenarios, the symmetric constrained adaptive estimator is clearly shown to hold superior performance. This is particularly pleasing given the \( h_0 = h_p \) restriction in place for all contributing densities in the ADAMULT selector and the imposed symmetry on the corresponding estimator (16). The improvement over the FIXOS or FIXMULT versions of the estimator is especially obvious.
in the high-segregation scenarios MS3 and MS4; the adaptive estimator appears far better equipped to handle locally abrupt changes in probability in the type-specific surfaces.

In the bottom row of Figure 5, we see the distribution of the common bandwidths $h_c$ chosen for ADAMULT are somewhat similar to those of FIXMULT for the mild segregation scenarios MS1 and MS2. This is visibly different to those of MS3 and MS4, where the chosen ADAMULT bandwidths are notably smaller than FIXMULT. In general, reducing the pilot bandwidth will increase the range of variable bandwidths by decreasing the smoothness of the pilot density in (7). This tends to be desirable for surfaces of high fluctuation such as those making up MS3 and MS4 – a wider gamut of point-specific bandwidths allows for more accurate capturing of local peaks and troughs.
6. Spatiotemporal density estimation

Our final methodological investigation concerns kernel density estimation of spatiotemporal data where, in addition to \( n \) spatial coordinates \( X = \{x_1, \ldots, x_n\}; X \in W \), we are provided a timestamp \( T = \{t_1, \ldots, t_n\} \) where \( T \in D \subset \mathbb{R} \). It is assumed \( D \) is a single closed interval on the real line.

We follow Fernando and Hazelton [36] in expressing a fixed-bandwidth trivariate kernel density estimate as the sum of \( n \) products of a 2D and 1D kernel function \( K \) and \( L \) respectively; each is assumed to be a zero-mean symmetric pdf with finite variance on the respective domain. The estimator is

\[
\tilde{f}_{X,T}(x, t \mid h, \lambda) = n^{-1}h^{-2} \lambda^{-1} \sum_{i=1}^{n} K \left( \frac{x - x_i}{h} \right) L \left( \frac{t - t_i}{\lambda} \right) q_W(x; h)^{-1} q_D(t; \lambda)^{-1},
\]

for \( x \in W \) and \( t \in D \), where \( h \) and \( \lambda \) are the spatial and temporal bandwidths respectively. Edge-correction factors are in general necessary on both spatial and temporal domains; \( q_W \) is given in (2), and a similar definition serves \( q_D(t; \lambda) = \lambda^{-1} \int_D L((t - u)/\lambda) \, du \). The same principle behind derivation of (4) applies in this setting, and the likelihood-optimal bandwidths are found via bivariate optimization of \( \arg \max_{h_\lambda} \{\text{FIXST}(h, \lambda)\} \), where

\[
\text{FIXST}(h, \lambda) = n^{-1} \sum_{i=1}^{n} \log \left\{ \tilde{f}_{X,T^{[-i]}, T^{[-i]}}(x_i, t_i \mid h, \lambda) \right\}.
\]

The leave-one-out operation is based on deletion of (and kernel density evaluation at) the \( i^{th} \) observation in space-time; denoted \( X^{[-i]}, T^{[-i]} \).

Let \( H = \{h_0, h_p, \lambda_0, \lambda_p\} \) be a set of four scalar bandwidth values; a global and pilot bandwidth for spatial and temporal margins respectively. A spatiotemporally adaptive version of (18) per Abramson’s regimen can be written as

\[
\tilde{f}_{X,T}(x, t \mid H) = n^{-1} \sum_{i=1}^{n} h_{X,T}(x_i, t_i; H \backslash \lambda_0)^{-2} \lambda_{X,T}(x_i, t_i; H \backslash h_0)^{-1}
\]

\[
\times K \left( \frac{x - x_i}{h_{X,T}(x_i, t_i; H \backslash \lambda_0)} \right) L \left( \frac{t - t_i}{\lambda_{X,T}(x_i, t_i; H \backslash h_0)} \right)
\]

\[
\times q_W(x; h_{X,T}(x, t; H \backslash \lambda_0))^{-1} q_D(t; \lambda_{X,T}(x, t; H \backslash h_0))^{-1}
\]

for \( x \in W; t \in D \), where \( A \backslash a \) refers to the set \( A \) without member \( a \). The bandwidth functions are

\[
h_{X,T}(x, t; H \backslash \lambda_0) = h_0 \gamma_f^{-1} \tilde{f}_{X,T}(x, t \mid h_p, \lambda_p)^{-0.5}
\]

and

\[
\lambda_{X,T}(x, t; H \backslash h_0) = \lambda_0 \gamma_f^{-1} \tilde{f}_{X,T}(x, t \mid h_p, \lambda_p)^{-0.5},
\]

where \( \tilde{f}_{X,T}(x, t \mid h_p, \lambda_p) \) is the fixed bandwidth pilot density estimate obtained via (19), and \( \gamma_f = \exp\{n^{-1} \sum_i \log \tilde{f}_{X,T}(x_i, t_i \mid h_p, \lambda_p)^{-0.5}\} \). As with purely spatial estimates, we clip the
pilot density at some small nominal value to prevent excessive variable bandwidths arising from either (21) or (22).

Completely unconstrained bandwidth selection for the new estimator (20) would involve quadivariate optimization over $H$. As noted in Section 2.2 and revealed by the simulations in Section 3, this is inadvisable. We therefore constrain the pilot and global bandwidths in a manner similar to the purely spatial estimator; this time, we are required to retain separate scales in the spatial and temporal margins. Setting $h_c \equiv h_0 = h_p$ and $\lambda_c \equiv \lambda_0 = \lambda_p$, the likelihood selector is given by

$$ADAST(h_c, \lambda_c) = n^{-1} \sum_{i=1}^{n} \log \left\{ \hat{f}_{X[-i], T[-i]} \left( x_i, t_i \mid h_c, \lambda_c \right) \right\}. \tag{23}$$

7. Simulation study 3: spatiotemporal densities

The computational cost of brute-force leave-one-out operations for adaptive spatiotemporal kernel estimates is high. There is a substantial bottleneck induced by the variable bandwidth and edge-correction factors, necessitating repeated recalculation of the adaptive density over $W \times \mathcal{D}$ for each deleted point. We mitigate this expense to a certain extent by implementing (20) using a data-partitioning approach whereby the adaptive estimate is approximated by a sum of fixed bandwidth estimates operating on subsets of the observed data. Appropriate subsets are identified based on a preselected sequence of unique quantiles of the variable bandwidth factors [12]. Such an approximation yields final estimates of lower precision when compared to direct evaluation of (20), though is arbitrarily improved by using a finer sequence of bandwidth quantiles. Further reductions in total running time can be achieved by parallelizing the leave-one-out steps of (23) over multiple CPU cores.

Even with such measures in place and constraining optimization to the two values $(h_c, \lambda_c)$, complete calculation of the ADAST selector for a single data set is costly (roughly 150 minutes for direct evaluation on one of our modern desktop machines, based on a sample size of $n = 500$; an evaluation lattice of size $64^3$; and setting a modest relative tolerance for the stopping rule defined using successive steps of the optimization routine). We therefore limit ourselves to two scenarios for the final simulation experiment and label these STS1 and STS2.

The scenarios are designed by first defining several pixel images over $W$ to pose as ‘keyframes’. We let $\mathcal{D} = [1, 10]$, and associate the endpoints of this interval with two chosen keyframes; remaining keyframes are assigned unique times within $\mathcal{D}$. This sequence of spatial images is linearly interpolated in a pixel-wise fashion so that the spatial appearance of the density is available for any $t \in \mathcal{D}$. We define a smooth univariate function over $\mathcal{D}$ to describe relative point intensity at each $t$. The resulting spatiotemporal surface, when evaluated over a discretized 3D array spanning $W \times \mathcal{D}$, is rescaled to ensure integration to 1. A 3D rejection sampling algorithm subsequently permits generation of independent and identically distributed observations.

Figures 6 and 7 plot aspects of the design of STS1 and STS2 respectively. The first scenario is composed of two main spatial centres of observations, the northernmost manifesting itself in the middle of the time period; both dissipate somewhat as time progresses. There are ongoing sporadic observations throughout, with relative point counts peaking around $t = 5$. STS2 is more complex, with single tight groupings of points both starting
Figure 6. Design of STS1, comprised of three unique spatial keyframes, alongside a plot showing the relative point intensity over time and an example data set.

Figure 7. Design of STS2, comprised of six unique spatial keyframes, alongside a plot showing the relative point intensity over time and an example data set.

and disappearing at staggered stages during $D$. Multiple observation clusters of different sizes develop between $t = 6$ and $t = 7$, with a single prominent grouping remaining at the end of the study period. The marginal point distribution over time oscillates between low and high density, with increased frequency for $t > 8$. There are greater space-time pockets devoid of observations than in STS1.

For 100 iterations, datasets of size 500 points are generated for each scenario. Densities are estimated via the constrained adaptive estimator (20) using ADAST (23), and via the fixed bandwidth estimator using FIXST (19). Again, a fixed bandwidth estimate is also computed based on Terrell’s oversmoothing rule-of-thumb as FIXOS; the two bandwidth values are computed separately using the marginal spatial and temporal data values. For
this we have $h_{OS}$ as defined in Section 3, and $\lambda_{OS} = \hat{\sigma}_T(2187/(96n\Gamma[4.5]))^{1/5}$ where $\hat{\sigma}_T$ is the standard deviation of the observation times and $\Gamma[\cdot]$ is the gamma function.

Results of the simulations are given in Figure 8, and clearly favor the adaptive approach in both scenarios on average. Noticeable is the increased variability in the ISE measures for the ADAST densities, owed largely to a flatter distribution of selected spatial-margin bandwidths. As noted above, this variability can be reduced by increasing the resolution of the partitioning approximation to the spatiotemporal density estimator for a computational price. Importantly, however, the variability shown by our current implementation does not hinder the potential of the adaptive estimator to significantly outperform its fixed bandwidth counterpart.

8. Applications

We conclude our investigation with a pair of applications drawn from human and veterinary epidemiology. While, unlike in the simulations, a lack of knowledge about the true function of interest prevents an exact error comparison, it is nevertheless of interest to contrast the fixed and the proposed constrained adaptive estimates.

8.1. Bovine tuberculosis

The first example deals with the spatial locations of 873 farms affected by bovine tuberculosis in Cornwall, UK, recorded between 1989 and 2002. Of interest in the study that originally presented and analysed these data was estimation and visualization of the degree...
of spatial segregation in the four most common spoligotypes via (13) – 9 \((n_9 = 494)\), 12 \((n_{12} = 109)\), 15 \((n_{15} = 166)\), and 20 \((n_{20} = 104)\) – of the disease [27]. The data are displayed in the left-hand panel of Figure 9: Visual evidence of segregation is apparent with spoligotypes 9 and 15 situated primarily in the northern portion of the region and the other types aggregated mainly in the southern area.

We estimate the \(m=4\) type-specific surfaces \(p_9\), \(p_{12}\), \(p_{15}\) and \(p_{20}\) using both fixed and constrained adaptive estimators, as well as a second fixed bandwidth estimate based on the oversmoothing factor. Bandwidth selection proceeds exactly as outlined for FIXMULT (14) and ADAMULT (17); and we find optimal bandwidths of \(h=4.8264\) km and \(h_c=3.7722\) km respectively. The objective curves of FIXMULT and ADAMULT are shown on the right of Figure 9; the common FIXOS bandwidth is found as \(h_{OS}=12.2312\) km.

Figure 10 shows the three estimated type-specific probability surfaces for spoligotype 20, with the full collection of images for all four types available in the supplementary materials. Dashed tolerance contours mark off statistically significant departures from uniformity that would occur for completely unsegregated patterns (i.e. assuming a null hypothesis of constant probability of occurrence over \(W\) for all observation types). These are based on \(p\)-values surfaces computed via Monte-Carlo permutation as \(\omega(x) = (M + 1)^{-1}(1 + \sum_{i=1}^{M}I[\tilde{p}^{(i)}(x) > \bar{p}(x)])\); \(x \in W\), where \(\tilde{p}\) is an estimate of a given type-specific probability surface from observed data, and \(\tilde{p}^{(i)}\) is the \(i\)th estimate found following random relabelling of the point types; this is repeated for \(M=999\) iterates [4,6]. The contours are displayed at a significance level of 1%.

The FIXOS surface is suggestive of oversmoothing: Almost no detail is present in the raised area of probability to the south. We capture additional detail through a reduced bandwidth in FIXMULT, though a closer inspection of the contours in the FIXMULT and ADAMULT surfaces reveal the latter to preserve slightly more detail than the former. Crucially, the additional detail in the ADAMULT surface does not appear to have come at the cost of undersmoothing elsewhere in the region – a clear benefit of Abramson adaptation.
Figure 10. Estimated type-specific probability surfaces for spoligotype 20 of the bovine tuberculosis data. Dashed green contours delineate statistically significant increases in type-specific occurrence probability at the 1% level. Common color scale.

Figure 11. Spatial and temporal margins of the Burkitt’s lymphoma data.

8.2. Burkitt’s lymphoma

Our second example concerns the 188 cases of Burkitt’s lymphoma recorded mostly in children in the Western Nile district of Uganda over an approximate 15-year period from 1960 to 1975 [37,38]. In addition to spatial coordinate data, we have the time of onset recorded as the number of days from January 1, 1960. The spatial and temporal margins of the data are plotted in Figure 11. A relatively diffuse yet large aggregation of observations lie at the western border, and over time, the number of reported cases appears to steadily increase within the first decade.

To assess structure in the data, we seek kernel density estimates of these spatiotemporal observations as outlined in Section 6, using (18) and (20). The required spatial- and temporal-axis bandwidths are selected using FIXST (19) and ADAST (23) for the fixed and constrained adaptive estimates respectively. We obtain $h = 6.641$ km; $\lambda = 2.556$ years for the former, and $h_c = 7.594$ km; $\lambda_c = 2.737$ years for the latter.
Figure 12. FIXST (left) and ADAST (right) spatiotemporal density estimates of the Burkitt’s lymphoma data; isosurfaces are drawn at identical numeric levels on both plots (darkening color depicts increasing density), and the observations are shown as black spheres.

The two density estimates are displayed as 3D plots in Figure 12. For details on an interactive version of the graphic that can be rotated with the mouse in a web browser to gain different viewing perspectives, see the supplementary material. Superimposed upon the observations themselves are 3D contours (‘isosurfaces’) drawn at $\alpha \times \max[\tilde{f}_{X,T}]$ for $\alpha \in \{0.5, 0.75, 0.95\}$ in darkening colors on both plots, where $\tilde{f}_{X,T}$ is the fixed bandwidth estimate.

Comparing the two estimates, we see both depict a bimodal structure with the drawn isosurfaces. The northernmost mode near the village of Omugo commences roughly at the fifth year, and persists for the remainder of the period with minimal spatial variation – shown with the ‘tubular’ appearance of the 3D contour. A higher-density mode situated further south around the Arua township is manifested after the first decade. In contrast to the fixed estimate, the adaptive estimate yields larger pockets of higher density within these modes – again illustrating the ability of Abramson adaptation to capture more detail in observation-rich subregions of $W \times D$ without compromising by undersmoothing elsewhere.

9. Concluding remarks

Kernel smoothing is a highly flexible tool for the nonparametric analysis of spatial and spatiotemporal point patterns. Its role can range from basic data exploration; to estimation and testing of relative risk and conditional probability surfaces; to capturing deterministic heterogeneity in semiparametric models.

Classical fixed bandwidth estimation, requiring only a single level of smoothing be set and therefore being relatively easy to implement, suffers somewhat when smoothing highly heterogeneous patterns. Adaptive smoothing is intuitively appealing in many applications, though the need to initialize pilot and global smoothing levels renders the bandwidth selection problem more complex.
Focusing on likelihood cross-validation, we have demonstrated the promising performance of Abramson adaptation for various nonparametric smoothing problems. By constraining pilot and global bandwidths to be equal, we circumvent identifiability problems in automated selection, providing a reliable means by which the two smoothing parameters could be chosen simultaneously in both spatial and spatiotemporal settings. These ideas extend to our proposed adaptive estimation of conditional probability surfaces for multitype patterns, which additionally benefit from symmetrization of bandwidth factors. Results of both synthetic and real-data analyses show the proposed approach has the capacity to perform rather well when compared to fixed bandwidth estimation.

Likelihood-based bandwidth selectors were subject to scrutiny here due to their ability to deal with both standalone density and conditional probability surface estimation. That said, there is no reason to suggest the strategy of constraining pilot and global bandwidths could not also be applied to different kinds of bandwidth selectors for adaptive density estimation. Future research could involve an investigation of comparative performance in this context. We also anticipate symmetric constrained adaptive spatiotemporal estimators of conditional probability will yield similar benefits as demonstrated here in the purely spatial case, though computational expense is a significant hindrance, and this could warrant additional attention. Finally, we consider the further development of Bayesian selection methods, where bandwidths are treated as parameters and have priors placed upon them [39,40], is a worthwhile avenue of pursuit: This could have benefits for construction and estimation of inhomogeneous Cox processes for data assumed to reflect both deterministic and stochastic influences.

Acknowledgments

All computations and analyses were performed in the R programming environment [41] using the contributed packages spatstat [42] and sparr [43]. Simulation study design was aided by the spagmix package [44], and the 3D plots were created using functionality of the rgl [45] and misc3d [46] packages.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The first author acknowledges partial financial support from the Royal Society of New Zealand, Marsden Fast-start grant 15-U00-092.

ORCID

Tilman M. Davies http://orcid.org/0000-0003-0565-1825

References

[1] Silverman BW. Density estimation for statistics and data analysis. New York (NY): Chapman and Hall; 1986.
[2] Wand MP, Jones MC. Kernel smoothing. Boca Raton: Chapman and Hall; 1995.
[3] Lawson AB, Williams FLR. Applications of extraction mapping in environmental epidemiology. Stat Med. 1993;12:1249–1258.
[4] Kelsall JE, Diggle PJ. Kernel estimation of relative risk. Bernoulli. 1995;1:3–16.
[5] Clark AB, Lawson AB. An evaluation of non-parametric relative risk estimators for disease maps. Comput Stat Data Anal. 2004;47:63–78.
[6] Hazelton ML, Davies TM. Inference based on kernel estimates of the relative risk function in geographical epidemiology. Biom J. 2009;51:98–109.
[7] Diggle PJ, Rowlingson B, Su TL. Point process methodology for on-line spatio-temporal disease surveillance. Environmetrics. 2005;16(5):423–434.
[8] Liang S, Carlin BP, Gelfand AE. Analysis of minnesota colon and rectum cancer point patterns with spatial and nonspatial covariate information. Ann Appl Stat. 2009;3(3):943–962.
[9] Davies TM, Hazelton ML. Assessing minimum contrast parameter estimation for spatial and spatiotemporal log-Gaussian Cox processes. Stat Neerl. 2013;67(4):355–389.
[10] Abramson IS. On bandwidth variation in kernel estimates – a square root law. Ann Stat. 1982;10(4):1217–1223.
[11] Davies TM, Hazelton ML. Adaptive kernel estimation of spatial relative risk. Stat Med. 2010;29(23):2423–2437.
[12] Davies TM, Baddeley A. Fast computation of spatially adaptive kernel estimates. Stat Comp. 2018;28:937–956.
[13] Diggle PJ. A kernel method for smoothing point process data. J R Stat Soc Ser C Appl Stat. 1985;34(2):138–147.
[14] Jones MC. Simple boundary correction for kernel density estimation. Stat Comp. 1993;3:135–146.
[15] Cao R, Cuevas A, Gonzáles-Manteiga W. A comparative study of several smoothing methods in density estimation. Comput Stat Data Anal. 1994;17:153–176.
[16] Jones MC, Marron JS, Sheather SJ. A brief survey of bandwidth selection for density estimation. J Am Stat Assoc. 1996;91:401–407.
[17] Sain SR, Baggerly KA, Scott DW. Cross-validation of multivariate densities. J Am Stat Assoc. 1994;89:807–817.
[18] Hall P, Marron JS. Variable window width kernel density estimates of probability densities. Probab Theory Rel. 1988;80:37–49.
[19] Marshall JC, Hazelton ML. Boundary kernels for adaptive density estimators on regions with irregular boundaries. J Multivariate Anal. 2010;101:949–963.
[20] Davies TM, Flynn CR, Hazelton ML. On the utility of asymptotic bandwidth selectors for spatially adaptive kernel density estimation. Stat Probab Lett. 2018;138:75–81.
[21] Zhang ZJ, Davies TM, Gao J, et al. Identification of high-risk regions for schistosomiasis in the Guichi region of China: an adaptive kernel density estimation-based approach. Parasitology. 2013;140(7):868–875.
[22] Møller J, Syversveen AR, Waagepetersen RP. Log Gaussian Cox processes. Scand J Stat. 1998;25:451–482.
[23] Terrell GR. The maximal smoothing principle in density estimation. J Am Stat Assoc. 1990;85:470–477.
[24] Bithell JF. An application of density estimation to geographical epidemiology. Stat Med. 1990;9:691–701.
[25] Bithell JF. Estimation of relative risk functions. Stat Med. 1991;10:1745–1751.
[26] Lawson AB, Williams FLR. Spatial competing risk modelling with applications in spatial epidemiology. Stat Med. 2000;19:2451–2468.
[27] Diggle PJ, Zheng P, Durr P. Nonparametric estimation of spatial segregation in a multivariate point process: bovine tuberculosis in Cornwall, UK. J R Stat Soc Ser C Appl Stat. 2005;54(3):645–658.
[28] Kelsall JE, Diggle PJ. Spatial variation in risk of disease: a nonparametric binary regression approach. J R Stat Soc Ser C Appl Stat. 1998;47(4):559–573.
[29] Kelsall JE, Diggle PJ. Non-parametric estimation of spatial variation in relative risk. Stat Med. 1995;14:2335–2342.
[30] Lawson AB, Williams FLR. Armadale: a case study in environmental epidemiology. J R Stat Soc Ser A. 1994;157(2):285–298.

[31] Wheeler DC. A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996–2003. Int J Health Geogr. 2007;6(13):1–16.

[32] Davies TM, Cornwall J, Sheard PW. Modelling dichotomously marked muscle fibre configurations. Stat Med. 2013;32:4240–4258.

[33] Zheng P, Durr P, Diggle PJ. Edge-correction for spatial kernel smoothing methods – when is it necessary? In: 2nd International Conference on the Applications of GIS and Spatial Analysis to Veterinary Science; University of Guelph, Ontario; 2004. p. 36–38; GISVET 04.

[34] Hazelton ML. Letter to the editor: kernel estimation of risk surfaces without the need for edge correction. Stat Med. 2008;27:2269–2272.

[35] Davies TM, Jones K, Hazelton ML. Symmetric adaptive smoothing regimens for estimation of the spatial relative risk function. Comput Stat Data Anal. 2016;101:12–28.

[36] Fernando WTPS, Hazelton ML. Generalizing the spatial relative risk function. Spatial Spatio-temporal Epidemiol. 2014;8:1–10.

[37] Bailey TC, Gatrell AC. Interactive spatial data analysis. Harlow: Longman; 1995.

[38] Rowlingson B, Diggle P. splancs: spatial and space-time point pattern analysis; 2017. R package version 2.01-40. Available from: https://CRAN.R-project.org/package = splancs.

[39] Zhang X, King ML, Hyndman RJ. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. Comput Stat Data Anal. 2006;50:3009–3031.

[40] Hu S, Poskitt DS, Zhang X. Bayesian adaptive bandwidth kernel density estimation of irregular multivariate distributions. Comput Stat Data Anal. 2012;56:732–740.

[41] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available from: https://www.R-project.org/.

[42] Baddeley A, Rubak E, Turner R. Spatial point patterns: methodology and applications with R. London: Chapman and Hall/CRC; 2015.

[43] Davies TM, Marshall JC, Hazelton ML. Tutorial on kernel estimation of continuous spatial and spatiotemporal relative risk. Stat Med. 2018;37(7):1191–1221.

[44] Redmond AK, Davies TM. spagmix: artificial spatial and spatiotemporal densities on bounded windows; 2018. R package version 0.3-0. Available from: https://CRAN.R-project.org/package = spagmix.

[45] Adler D, Murdoch D. rgl: 3d visualization using opengl; 2018. R package version 0.99.16. Available from: https://CRAN.R-project.org/package = rgl.

[46] Feng D, Tierney L. Computing and displaying isosurfaces in R. J Stat Soft. 2008;28(1). Available from: http://www.jstatsoft.org/v28/i01/.