



# Multitype point process analysis of spines on the dendrite network of a neuron

Adrian Baddeley,

University of Western Australia and Commonwealth Scientific and Industrial  
Research Organisation, Perth, Australia

Aruna Jammalamadaka

University of California at Santa Barbara, USA

and Gopalan Nair

University of Western Australia, Perth, Australia

[Received May 2013. Revised October 2013]

**Summary.** We develop methods for analysing the spatial pattern of events, classified into several types, that occur on a network of lines. The motivation is the study of small protrusions called 'spines' which occur on the dendrite network of a neuron. The spatially varying density of spines is modelled by using relative distributions and regression trees. Spatial correlations are investigated by using counterparts of the  $K$ -function and pair correlation function, where the main problem is to compensate for the network geometry. This application illustrates the need for careful analysis of spatial variation in the intensity of points, before assessing any evidence of clustering.

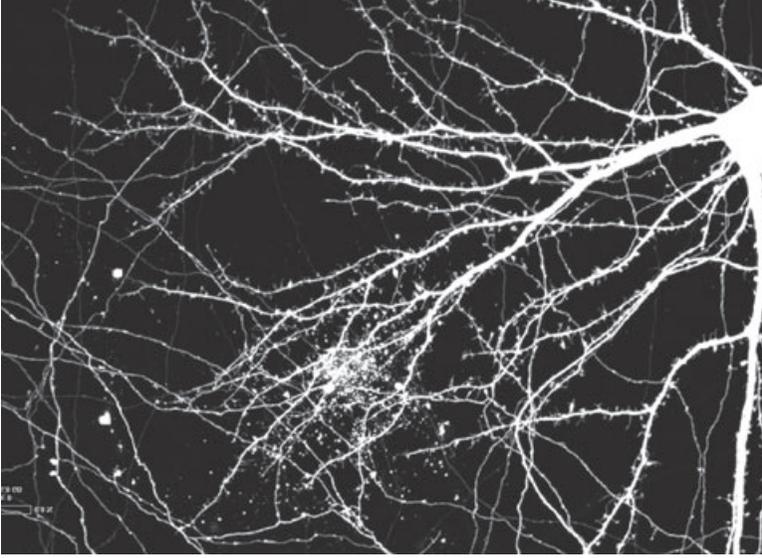
**Keywords:** Inhomogeneous  $K$ -function;  $K$ -function; Linear network; Pair correlation function; Point process; Regression trees; Spatial point patterns

## 1. Introduction

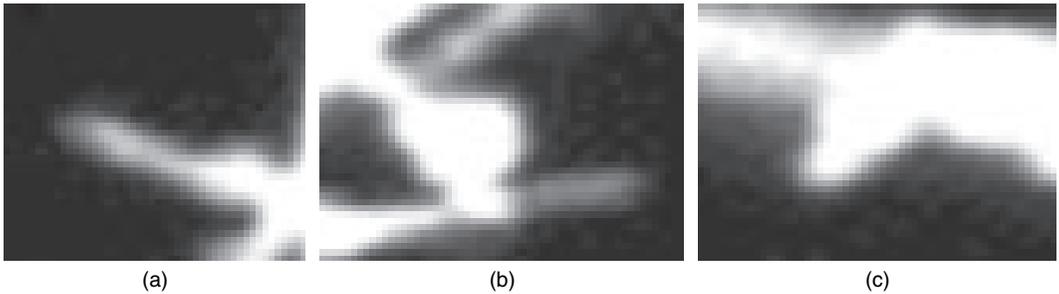
Fig. 1 shows a microscope image of part of the dendrite network of a rat neuron in a cell culture. Small protrusions called *spines* are more clearly visible at higher digital magnification, as shown in Fig. 2. Three types of spines are distinguished by their shapes, exemplified in Fig. 2. For a better understanding of normal function and disease processes, it is important to characterize the joint spatial distribution of spines of different types. For example, changes in spine shape and distribution have been linked to neurological disorders (Irwin *et al.*, 2001).

Spatial point process methods (Diggle, 2003; Illian *et al.*, 2008) have been used since the 1970s to analyse the spatial distribution of cells (Bell and Grunwald, 2004; Bjaalie and Diggle, 1990; Diggle, 1986; Diggle *et al.*, 1991, 2005; Fleischer *et al.*, 2005; Millet *et al.*, 2011; Mamaghani *et al.*, 2010; Ripley, 1977) and subcellular objects (Pedro *et al.*, 1984; Webster *et al.*, 2005; Chen *et al.*, 2008) observed in microscope imagery. The unusual feature here is that the spines are not free to lie anywhere on the two-dimensional image plane but are constrained to lie on the one-dimensional dendrite network. Since the dendrites propagate electrical signals, and convey most

*Address for correspondence:* Gopalan Nair, School of Mathematics and Statistics, University of Western Australia, 35 Stirling Highway, Crawley, Perth, WA 2009, Australia.  
E-mail: gopalan.nair@uwa.edu.au



**Fig. 1.** Microscope image of a dendrite network (white lines) and part of a cell body (white area) of a rat neuron in cell culture (width,  $232 \mu\text{m}$ ; height  $168 \mu\text{m}$ ; depth  $2.6 \mu\text{m}$ ; projected image; laser scanning confocal microscope)



**Fig. 2.** Examples of spines of three types in Fig. 1: (a) thin; (b) mushroom; (c) stubby

of the nutrients and molecular genetic signals, the network structure is highly relevant. To our knowledge, very few previous studies have attempted to analyse the arrangement of dendritic spines along the dendrites (Yadav *et al.*, 2012; Jammalamadaka *et al.*, 2013). The application of methods from spatial statistics may provide a deeper understanding of the spatial organization of dendritic spines.

Methods for spatial analysis on linear networks have been developed over the past decade, principally by Professor A. Okabe and collaborators (Okabe and Satoh, 2009; Okabe and Sugihara, 2012), and include analogues of standard point process techniques such as Ripley's  $K$ -function (Okabe and Yamada, 2001). Recently it has been shown that these methods can be improved by adjusting for the geometry of the network (Ang *et al.*, 2012).

Fig. 3 is a representation of the dendrite network extracted from Fig. 1, together with the locations and types of the spines. These data can be described as a multitype point pattern on a linear network. In a 'multitype' point pattern, the points are classified into several different categories or 'types'. Equivalently, to each point  $x_i$  of the point pattern we associate a categorical



**Fig. 3.** Extracted representation of a branch of the dendrite network (—) and multitype point pattern of spines (○, mushroom; △, stubby; +, thin)

variable  $t_i$  indicating its type. In other examples the types could be different kinds of road accident, criminal offences, and so on. Although it is usually straightforward to generalize point process tools to apply to multitype point processes, experience shows that their *interpretation* may be subtly different in the multitype case, and that statistical inference requires careful attention (Baddeley, 2010; Grabarnik and Särkkä, 2009; Harkness and Isham, 1983; Illian *et al.*, 2008; van Lieshout and Baddeley, 1999).

This paper develops non-parametric methods for analysing multitype point patterns on a linear network and applies them to the dendritic spines data. The methodology is based on first and second moments of a point process, extends the results of Ang *et al.* (2012) to multitype patterns and includes some additional techniques for estimating first-order intensity functions on a tree-like network by using relative distributions and regression trees.

The plan of the paper is as follows. Section 2 sketches the scientific background to the dendrite example. Section 3 gives some formal definitions and background. Methods for second-order analysis using the  $K$ -function and pair correlation function are described in Section 4 under the assumption of homogeneity, and in Section 5 for the case where inhomogeneity is present. The dendrite data are analysed in Section 6, requiring the development of some additional methodology. We end with a discussion.

## 2. Scientific background

### 2.1. Importance of spatial pattern of spines

Spines are protrusions that occur on the neuronal dendrites of most mammalian neurons. They contain the postsynaptic apparatus and have a role in learning and memory storage. The shapes of dendritic spines contribute to synaptic plasticity. The distribution of spine shapes is biologically important because the electrical properties of spines, such as the spine neck resistance, promote non-linear dendritic processing and associated forms of plasticity and storage (Harnett *et al.*, 2012) which enhance the computational capabilities of neurons. For example,

mushroom-type spines are generally thought to be the most electrochemically mature and are therefore more likely to create synapses with neighbouring neurons than are the stubby-type or thin-type spines (Nimchinsky *et al.*, 2002). Changes in spine shape and spatial distribution have been linked to neurological disorders such as fragile X syndrome (Irwin *et al.*, 2001).

The spatial distribution of spines is critically important. It determines the extent to which electrical connections (synapses) will be established with the surrounding neural circuits (Yuste, 2011). It may also reflect activity patterns in these circuits, because the synaptic pruning that occurs during neural development is dependent on this activity. Given that neighbouring spines on the same short segment of dendrite can express a full range of structural dimensions, individual spines might act as separate computational units (Harris and Kater, 1994). Nevertheless, the dendrite acts in a co-ordinated manner and thus the joint spatial distribution of spines of different types is likely to be important.

## 2.2. Data

The data originate from a study of neuronal development in cell cultures (Jammalamadaka *et al.*, 2013). In a series of replicate experiments, neurons were grown in glial culture and visualized once. The analysis of this controlled, replicated experiment is beyond the scope of the present paper. We confine attention to the single example pattern shown in Fig. 1, which was taken from the fifth neuron in the second biological replicate experiment, observed on the 14th day *in vitro*. The network shown in Fig. 1 is one of the 10 dendritic trees of this neuron. A dendritic tree consists of all dendrites issuing from a single root branch off the cell body; each neuron typically has 4–10 dendritic trees. This example was chosen because it is sufficiently large to demonstrate our techniques clearly, without being too large for graphical purposes.

For more detail on how the image data were obtained, we refer the reader to Jammalamadaka *et al.* (2013). To avoid errors that are inherent in automated reconstruction algorithms, the dendrite network was traced manually, and spine locations and types were verified manually, by trained observers. The linear network trace of each dendritic branch as well as the spine locations and types as shown in Fig. 1 were obtained from the images by using the software package *NeuronStudio* (Rodriguez *et al.*, 2008; Wearne *et al.*, 2005). The full image stack as well as the annotation files are publicly available from the web site of *Bisque* (Kvilekval *et al.*, 2010) as detailed in the acknowledgements.

Although the material is three dimensional, and was originally visualized in three dimensions, it is very shallow in the third dimension, so a two-dimensional projection is adequate for representing the spatial layout of the dendrites. Three-dimensional information was nevertheless essential to determine which parts of the network are physically joined. The resulting linear network is shown in two-dimensional projection in Fig. 3.

## 2.3. Previous studies

In Yadav *et al.* (2012) clusters of three or more spines were identified by using hierarchical clustering of distances along the dendrite network. In Jammalamadaka *et al.* (2013), the spatial pattern of spine locations was studied by using the linear network *K*-function of Ang *et al.* (2012) and was found to be completely spatially random. To assess dependence between the types of neighbouring spines, Jammalamadaka *et al.* (2013) fitted a multinomial logistic regression of spine type against the types of the three nearest neighbours. This suggested positive association between the type of a spine and those of its neighbours.

## 3. Multitype point processes on a linear network

Mathematical definitions of a linear network in two-dimensional space, and of a point process

on a network, were given in Ang *et al.* (2012). These require only minor adaptation to deal with the fact that the dendrite network lies in three-dimensional space. Following is a brief summary.

A *linear network* in  $\mathbb{R}^3$  is defined as the union  $L = \cup_{i=1}^n l_i$  of a finite collection of line segments  $l_1, \dots, l_n$  in  $\mathbb{R}^3$ . The total length of all line segments in  $L$  is denoted by  $|L|$ . The *shortest path distance*  $d_L(u, v)$  between two points  $u$  and  $v$  in  $L$  is the minimum of the lengths of all paths along the network from  $u$  to  $v$ . If  $v$  cannot be reached from  $u$  along the network, then we set  $d_L(u, v) = \infty$ .

Our notation for multitype point patterns and point processes is borrowed from van Lieshout and Baddeley (1999). A multitype point pattern data set  $\mathbf{y}$  on a linear network  $L$ , with each point belonging to one of  $c$  possible categories or types, is a finite set  $\mathbf{y} = \{(x_1, t_1), \dots, (x_n, t_n)\}$  where  $x_k \in L$  is the location of the  $k$ th point and  $t_k \in C = \{1, 2, \dots, c\}$  is the type classification of the  $k$ th point,  $k = 1, 2, \dots, n$ , and the number  $n$  of random points is not fixed in advance.

A multitype point process  $\mathbf{Y}$  on a linear network  $L$  is a stochastic process whose realizations  $\mathbf{y}$  are multitype point patterns. It can be regarded as a point process on  $L \times C$ . We write  $\mathbf{X} = \{x_k : (x_k, t_k) \in \mathbf{Y}\}$  for the ‘projected’ or ‘unmarked’ process consisting of the locations of points of  $\mathbf{Y}$  ignoring their types. For each possible type  $i$ , we write  $\mathbf{X}_i = \{x_k : (x_k, t_k) \in \mathbf{Y}, t_k = i\}$  for the point process of locations of random points of type  $i$ . Then we may regard  $\mathbf{Y}$  as equivalent to the multivariate process  $(\mathbf{X}_1, \dots, \mathbf{X}_c)$ . It is also useful to define, for any non-empty set  $I \subseteq C$ , the point process  $\mathbf{X}_I = \cup_{i \in I} \mathbf{X}_i$  of points whose types belong to the set  $I$ .

More generally it would be possible to replace the set of categories  $C$  by any space  $\mathcal{C}$  of possible ‘marks’. Then we would define a *marked point process* on  $L$  with marks in  $\mathcal{C}$  as a point process in  $L \times \mathcal{C}$  such that the projected process  $\mathbf{X}$  is a point process (meaning in particular that the number of random points, regardless of their mark value, is almost surely finite). The mark  $t_k$  that is attached to the random point  $x_k$  may specify any quantitative or qualitative characteristics of the random point that are relevant to the study, such as size, arrival time and colour. The methods of this paper extend to the general case of a marked point process. However, we shall confine attention to multitype point processes for simplicity.

### 3.1. Intensity

The most important property of a point process is its ‘intensity’ or ‘rate’ which describes the expected frequency of occurrence of random points of the process.

For an unmarked point process  $\mathbf{X}$  on the linear network  $L$  we say that  $\mathbf{X}$  has constant *intensity* or *rate*  $\lambda > 0$  if, for any  $B \subseteq L$ ,

$$\mathbb{E}[n(\mathbf{X} \cap B)] = \lambda|B|,$$

where  $n(\mathbf{X} \cap B)$  denotes the number of points of  $\mathbf{X}$  falling in  $B$ . Thus  $\lambda$  is the average density of random points per unit length of the network. An unbiased estimator of  $\lambda$ , given a point pattern data set  $\mathbf{x}$ , is  $\hat{\lambda} = n(\mathbf{x})/|L|$  where  $n(\mathbf{x})$  is the number of points in  $\mathbf{x}$ .

More generally the intensity may be spatially varying, and  $\mathbf{X}$  has *intensity function*  $\lambda(u)$  if, for any  $B \subseteq L$ ,

$$\mathbb{E}[n(\mathbf{X} \cap B)] = \int_B \lambda(u) d_1u, \quad (1)$$

where  $d_1u$  denotes integration with respect to one-dimensional arc length along the linear network. Heuristically if  $[u, u + d_1u]$  denotes an infinitesimal segment in  $L$ , then the probability that a point of  $\mathbf{X}$  falls in the segment is  $\mathbb{P}\{n(\mathbf{X} \cap [u, u + d_1u]) > 0\} = \lambda(u) d_1u$ . In some applications it may be inappropriate to assume that an intensity function exists, and we may have to rely on the intensity measure  $\Lambda$  defined by  $\Lambda(B) = \mathbb{E}[n(\mathbf{X} \cap B)]$ .

The intensity function of a point process  $\mathbf{X}$  on a linear network can be estimated by using various kernel smoothing estimators (Xie and Yan, 2008; Okabe *et al.*, 2009; Shiode and Shiode, 2009) although the statistical properties of these estimators are not very well understood.

If  $\mathbf{Y}$  is a *multitype* point process on  $L$ , we write  $\lambda_i(u)$  for the intensity function of  $\mathbf{X}_i$ , and  $\lambda_I(u)$  for the intensity function of  $\mathbf{X}_I$  where  $I \subseteq C$ . It follows that  $\lambda_I(u) = \sum_{i \in I} \lambda_i(u)$  and in particular the intensity of the unmarked process  $\mathbf{X}$ . is  $\lambda \cdot(u) = \sum_{i=1}^c \lambda_i(u)$ .

### 3.2. Multitype Poisson processes on linear networks

A multitype *Poisson process* on  $L$  can be defined in three equivalent ways, following Kingman (1993): firstly as a Poisson point process  $\mathbf{Y}$  on  $L \times C$ ; or secondly as a multivariate point process  $\mathbf{Y} = (\mathbf{X}_1, \dots, \mathbf{X}_c)$  such that the processes  $\mathbf{X}_i$  of random points of each type are Poisson processes, and  $\mathbf{X}_1, \dots, \mathbf{X}_c$  are *independent* processes; or thirdly as a multitype point process  $\mathbf{Y}$  with the property that the process of locations  $\mathbf{X}$ . is a Poisson process on  $L$ , and the marks are conditionally independent given  $\mathbf{X}$ .

A *homogeneous* multitype Poisson process is a process in which each component process  $\mathbf{X}_i$  has a constant intensity  $\lambda_i > 0$  for  $i \in C$ . The unmarked process  $\mathbf{X}$ . then has constant intensity  $\lambda \cdot = \sum_{i=1}^c \lambda_i$ . The marks are independent and identically distributed, with probability  $p_i = \lambda_i/\lambda \cdot$  for mark  $i \in C$ . For further explanation, see Baddeley (2010) and Illian *et al.* (2008).

## 4. Second-order statistics

Here we develop the analogue, for point patterns on a linear network, of the second-order analysis of processes of several types of points. Ripley's  $K$ -function (Ripley, 1976, 1977) was generalized to multitype point patterns in two dimensions by Lotwick and Silverman (1982) and Harkness and Isham (1983). The  $K$ -function was adapted to linear networks by Okabe and Yamada (2001) and a geometrically corrected version of the  $K$ -function was proposed by Ang *et al.* (2012). Here we extend the geometrically corrected  $K$ -function to the multitype case.

### 4.1. Key quantities

An important quantity that was introduced in Ang *et al.* (2012) is

$$m(u, r) = \#\{v \in L : d_L(u, v) = r\}, \quad (2)$$

the number of locations  $v$  on the network which lie exactly  $r$  units away from the location  $u$  by the shortest path. This quantity can be regarded as the perimeter of a 'disc' of radius  $r$  in the linear network, centred at  $u$ . Let

$$R = \sup\{r : m(u, r) > 0 \text{ for all } u \in L\}. \quad (3)$$

This can be interpreted as the '*circumradius*' of the network as explained in Ang *et al.* (2012).

### 4.2. Multitype pair correlation function

For simplicity we make the regularity assumption that the multitype point process  $\mathbf{Y}$  has intensity functions of first and second order, i.e. for any  $i \in C$  the subprocess  $\mathbf{X}_i$  has an intensity function  $\lambda_i(u)$  as defined in equation (1); and for any  $i, j \in C$  the subprocesses  $\mathbf{X}_i$  and  $\mathbf{X}_j$  have a *second-moment intensity function*  $\lambda_{ij}(u, v)$  defined to satisfy

$$\mathbb{E}[n(\mathbf{X}_i \cap A)n(\mathbf{X}_j \cap B)] = \int_A \int_B \lambda_{ij}(u, v) d_1 u d_1 v \quad (4)$$

for *disjoint* line segments  $A, B \subset L$ . Heuristically  $\lambda_{ij}(u, v) d_1 u d_1 v$  is the joint probability that two given infinitesimal intervals, of lengths  $d_1 u$  and  $d_1 v$ , around the locations  $u$  and  $v$  will each contain a random point, of types  $i$  and  $j$  respectively. For a multitype Poisson process  $\mathbf{Y}$  we have  $\lambda_{ij}(u, v) = \lambda_i(u) \lambda_j(v)$  for  $u \neq v$ .

Equation (4) implies the ‘second-order Campbell formula’

$$\mathbb{E} \left[ \sum_{x_k \in \mathbf{X}_i} \sum_{x_l \in \mathbf{X}_j} h(x_k, x_l) \right] = \int_L \int_L h(u, v) \lambda_{ij}(u, v) d_1 u d_1 v, \quad (5)$$

which holds for any measurable real function  $h$  on  $L \times L$  for which the right-hand side is finite.

We define the *multitype pair correlation function* between  $\mathbf{X}_i$  and  $\mathbf{X}_j$  by

$$\rho_{ij}^L(u, v) = \frac{\lambda_{ij}(u, v)}{\lambda_i(u) \lambda_j(v)}, \quad u, v \in L. \quad (6)$$

This is a non-centred correlation function with the heuristic interpretation

$$\rho_{ij}^L(u, v) = \mathbb{E}[I_i(u) I_j(v)] / \{\mathbb{E}[I_i(u)] \mathbb{E}[I_j(v)]\},$$

where, for any possible type  $k$ ,  $I_k(u)$  is the indicator that equals 1 if the interval of length  $d_1 u$  around  $u$  contains a random point of type  $k$ .

The following useful result can easily be proved.

*Lemma 1.* Suppose that  $\mathbf{Y}$  is a multitype point process on a linear network  $L$  whose first- and second-moment intensities exist. Then:

- (a) if  $\mathbf{Y}$  is a multitype Poisson process,  $\rho_{ij}^L \equiv 1$  for all  $i$  and  $j$ ;
- (b) if the component processes  $\mathbf{X}_i$ ,  $i \in C$ , are *independent*,  $\rho_{ij}^L \equiv 1$  for all  $i$  and  $j$ ,  $i \neq j$ ;
- (c) if  $\mathbf{Y}$  has the *random-labelling property* that the marks are conditionally independent and identically distributed given the locations  $\mathbf{X}_.$ ,  $\rho_{ij}^L \equiv \rho^L$  for all  $i$  and  $j$ , where  $\rho^L$  is the univariate pair correlation function of  $\mathbf{X}_.$

For an unmarked point process, a pair correlation function identically equal to 1 would usually be taken as indicating that the point process is consistent with a Poisson process, despite some *caveats* (Baddeley and Silverman, 1984). However, for a multitype point process, the finding that  $\rho_{ij}^L \equiv 1$  for  $i \neq j$  suggests merely that the component processes  $\mathbf{X}_i$  and  $\mathbf{X}_j$  are uncorrelated.

In the application to dendritic spines, there is a possibility of misclassification of spine types. It is useful to note that cases (a) and (c) of lemma 1 still apply if spine types are independently randomly misclassified. Suppose that the observed types  $t_k^{\text{obs}}$  are conditionally independent given the true types  $t_k$ , and that  $\mathbb{P}(t_k^{\text{obs}} = j | t_k = i)$  does not depend on  $k$ . Then case (c) remains true in that, if the true process  $\mathbf{Y}$  has the random-labelling property, then the observed process  $\mathbf{Y}^{\text{obs}}$  also has the random-labelling property, and the conclusion of case (c) holds for  $\mathbf{Y}^{\text{obs}}$ . The Poisson process case (a) is a special case of (c).

### 4.3. Estimation assuming homogeneity

For the rest of this section we assume that the point process is homogeneous in the following sense.

*Definition 1.* A multitype point process  $\mathbf{Y}$  on a linear network  $L$  is called *second order pseudo-stationary* if the first-order intensities are constant,  $\lambda_i(u) \equiv \lambda_i$ , and the second-order intensities depend only on shortest path distance,

$$\lambda_{ij}(u, v) = \lambda_{ij} \{d_L(u, v)\}, \quad d_L(u, v) < \infty. \quad (7)$$

It follows that the multitype pair correlations also depend only on the shortest path distance,

$$\rho_{ij}^L(u, v) = \rho_{ij}^L\{d_L(u, v)\}, \quad d_L(u, v) < \infty. \quad (8)$$

The condition that  $d_L(u, v) < \infty$  means that we consider only pairs of points  $u$  and  $v$  that are connected by a path in  $L$ . See Ang *et al.* (2012) for further discussion. In the dendrite network, all points are connected.

Given a multitype point pattern data set  $\mathbf{y} = \{(x_1, t_1), \dots, (x_n, t_n)\}$  assumed to come from a second-order pseudostationary process, we estimate the intensity  $\lambda_i$  for  $i \in C$  by  $\hat{\lambda}_i = n_i/|L|$ , where  $n_i = n(\mathbf{x}_i)$  is the number of points of type  $i$ . The multitype pair correlation function  $\rho_{ij}^L(r)$  can be estimated by kernel smoothing of the interpoint distances, with appropriate weighting for the geometry of the network (Ang *et al.*, 2012):

$$\hat{\rho}_{ij}^L(r) = \frac{|L|}{n_{ij}} \sum_{x_k \in \mathbf{X}_i} \sum_{x_l \in \mathbf{X}_j} \frac{\kappa\{d_L(x_k, x_l) - r\}}{m\{x_k, d_L(x_k, x_l)\}}, \quad (9)$$

where  $\kappa(\cdot)$  is any smoothing kernel function on  $\mathbb{R}$ , with  $n_{ij} = n_i n_j$  if  $i \neq j$  and  $n_{ii} = n_i(n_i - 1)$ . The weighting factor  $m\{x_k, d_L(x_k, x_l)\}$  in equation (9), defined in equation (2), counts the number of locations  $v$  on the network such that  $d_L(x_k, v) = d_L(x_k, x_l)$ . This compensates for the variable geometry of the network and ensures the following ‘unbiasedness’ property.

*Lemma 2.* If  $\mathbf{Y}$  is second order pseudostationary, the smoothing estimator (9) satisfies

$$\mathbb{E}[N_{ij} \hat{\rho}_{ij}^L(r)] = \lambda_i \lambda_j |L|^2 \bar{\rho}_{ij}(r), \quad (10)$$

where  $\bar{\rho}_{ij}(r) = \int \kappa(t - r) \rho_{ij}^L(t) dt$  is a kernel-smoothed version of  $\rho_{ij}^L(r)$ . Here  $N_{ij}$  is the random variable  $N_{ij} = N_i N_j$  for  $i \neq j$  and  $N_{ii} = N_i(N_i - 1)$  with  $N_i = n(\mathbf{X}_i)$  and  $N_j = n(\mathbf{X}_j)$ .

A proof is given in Appendix A. In the special case of a multitype Poisson process, it is easy to show that  $\mathbb{E}[N_{ij}] = \lambda_i \lambda_j |L|^2$ , so equation (10) implies that  $\hat{\rho}_{ij}^L(r)$  is the ratio of unbiased estimators. It will be a consistent and asymptotically normal estimator of  $\bar{\rho}_{ij}(r)$  under a large sample limit regime and will be consistent for  $\rho_{ij}(r)$  with an appropriate bandwidth selection rule. For non-Poisson processes, the quantity  $n_{ij}$  could be a biased and even inconsistent estimator of  $\lambda_i \lambda_j |L|^2$ , leading to possible bias in the estimation of the pair correlation. This problem is familiar from two-dimensional spatial statistics (Ripley, 1976; Diggle, 2003).

#### 4.4. Multitype $K$ -function

Following is the multitype version of the geometrically corrected  $K$ -function that was introduced in Ang *et al.* (2012).

*Definition 2.* Let  $\mathbf{Y}$  be a multitype point process on the linear network  $L$ . For any types  $i, j \in C$  define

$$K_{ij}^L(u, r) = \frac{1}{\lambda_j} \mathbb{E} \left[ \sum_{x_k \in \mathbf{X}_j} \frac{\mathbf{1}\{0 < d_L(u, x_k) \leq r\}}{m\{u, d_L(u, x_k)\}} \middle| u \in \mathbf{X}_i \right] \quad (11)$$

for any location  $u \in L$  and any  $r \in [0, R)$  where  $R$  is the circumradius defined in expression (3).

In heuristic terms, the  $K$ -function of the process gives the expected number of random points of type  $j$  that lie within a given distance  $r$  of a typical random point of type  $i$ , normalized by the intensity of points of type  $j$ . The conditional expectation on the right-hand side of equation (11) is formally defined as an expectation with respect to the *Palm distribution* of  $\mathbf{Y}$  given a point of  $\mathbf{X}_i$  at the location  $u$ . See van Lieshout and Baddeley (1999) for explanation.

Again the denominator  $m\{u, d_L(u, x_k)\}$  in equation (11) compensates for the variable geometry of the network and ensures the following result, which effectively states that the  $K$ -function is well defined.

*Lemma 3.* If  $\mathbf{Y}$  is second order pseudostationary, then, for any possible types  $i$  and  $j$ ,  $K_{ij}^L(u, r) = K_{ij}^L(r)$  does not depend on the choice of  $u$ .

We call  $K_{ij}^L(r)$  the (geometrically corrected) *multitype  $K$ -function*. When  $i = j$ ,  $K_{ii}^L(r)$  reduces to the geometrically corrected  $K$ -function of  $\mathbf{X}_i$  as defined in Ang *et al.* (2012).

*Lemma 4.* If  $\mathbf{Y}$  is second order pseudostationary, then, for any possible types  $i$  and  $j$ ,

$$K_{ij}^L(r) = \int_0^r \rho_{ij}^L(t) dt. \quad (12)$$

Lemma 4 is analogous to the connection between the  $K$ -function and pair correlation function in the two-dimensional case. It leads to the following, practically important, result.

*Lemma 5.* Suppose that  $\mathbf{Y}$  is second order pseudostationary. Then:

- (a) if  $\mathbf{Y}$  is a multitype Poisson process,  $K_{ij}^L(r) = r$  for all  $i$  and  $j$  and  $0 \leq r < R$ ;
- (b) if the component processes  $\mathbf{X}_i$ ,  $i \in C$ , are *independent*,  $K_{ij}^L(r) \equiv r$  for all  $i \neq j$  and  $0 \leq r < R$ ;
- (c) if  $\mathbf{Y}$  has the *random-labelling property* that the marks are conditionally independent and identically distributed given the locations  $\mathbf{X}$ , then  $K_{ij}^L(r) \equiv K^L(r)$ , where  $K^L$  is the geometrically corrected  $K$ -function of  $\mathbf{X}$ .

The proof is straightforward.

Result (a) of lemma 5 is important because it means that it is valid to compare  $K$ -functions on different linear networks. It is further evidence that the geometrical correction factor  $m(u, r)$  in equation (11) is appropriate.

Given a multitype point pattern data set  $\mathbf{y} = \{(x_1, t_1), \dots, (x_n, t_n)\}$  assumed to come from a second-order pseudostationary process, we estimate the multitype  $K$ -function by

$$\hat{K}_{ij}^L(r) = \frac{|L|}{n_i n_j} \sum_{x_k \in \mathbf{X}_i} \sum_{x_l \in \mathbf{X}_j} \frac{\mathbf{1}\{d_L(x_k, x_l) \leq r\}}{m\{x_k, d_L(x_k, x_l)\}}. \quad (13)$$

The double sum in equation (13) has an unbiasedness property

$$\mathbb{E}[N_{ij} \hat{K}_{ij}^L(r)] = \lambda_i \lambda_j |L|^2 K_{ij}^L(r) \quad (14)$$

for  $r < R$ . If  $\mathbf{Y}$  is a multitype Poisson process, then  $\hat{K}_{ij}^L(r)$  is the ratio of unbiased estimators and is consistent and asymptotically normal under an appropriate large sample limit. For a non-Poisson process, the denominator  $N_{ij}$  may contribute bias.

Note that result (14) is valid only for  $r < R$ . This is similar to the constraint that was imposed by Ohser and Stoyan (1981) for validity of Ripley's estimator of the two-dimensional  $K$ -function.

The variance of estimator (13) can be calculated by a straightforward adaptation of the results of Ang *et al.* (2012). For a homogeneous multitype Poisson process, the variance of  $\hat{K}_{ij}^L(r)$  is approximately constant as a function of  $r$ , over a large range of  $r$ -values.

#### 4.5. Mark connection function

Experience with the analysis of two-dimensional point patterns (Diggle, 2003; Gelfand *et al.*, 2010; Illian *et al.*, 2008) suggests that it may be useful, especially when investigating the random-labelling property, to estimate the *mark connection function* (Illian *et al.*, 2008) between marks

$i$  and  $j$ ,

$$p_{ij}(r) = \frac{\lambda_i \lambda_j \rho_{ij}^L(r)}{\lambda^2 \rho^L(r)}, \quad (15)$$

and the *mark equality function*

$$p(r) = \sum_i p_{ii}(r). \quad (16)$$

Loosely speaking  $p_{ij}(r)$  is the conditional probability, given that there is a pair of points separated by a distance equal to  $r$ , that the points have types  $i$  and  $j$  respectively. Similarly,  $p(r)$  is the conditional probability that the two points have the same type. Under any of the scenarios listed in lemma 1, the functions  $p_{ij}(r)$  and  $p(r)$  are constant.

A practical strategy for analysis (assuming second-order pseudostationarity) is to start by plotting the mark equality function  $p(r)$ . If this appears *not* to be a constant function, then the data are apparently inconsistent with each of the scenarios in lemma 1; the form of  $p(r)$  may suggest the type of dependence. Alternatively if  $p(r)$  appears to be a constant function, and if the individual functions  $p_{ij}(r)$  also appear to be constant, then the pair correlation functions  $\rho_{ij}^L$  should be inspected to discriminate between the three scenarios in lemma 1.

The plug-in estimator of the mark connection function  $p_{ij}(r)$ , which is obtained by substituting equation (9) into equation (15), collapses to

$$\hat{p}_{ij}(r) = \sum_{x_k \in \mathbf{X}_i} \sum_{x_l \in \mathbf{X}_j} s_{k,l}(r) / \sum_k \sum_{l \neq k} s_{k,l}(r) \quad (17)$$

where  $s_{k,l}(r) = \kappa \{d_L(x_k, x_l) - r\} / m \{x_k, d_L(x_k, x_l)\}$ . Up to a constant factor, the denominator and numerator are unbiased estimators of the second-moment density of all points, and of pairs of types  $i$  and  $j$  respectively.

## 5. Inhomogeneous second-order statistics

For a spatial point process with *non-constant* intensity, inhomogeneous analogues of the  $K$ -function and pair correlation function were proposed in Baddeley *et al.* (2000) for two-dimensional point processes, and in Ang *et al.* (2012) for point processes on a linear network. Here we extend this idea to multitype point processes on a linear network.

*Definition 3.* Let  $\mathbf{Y}$  be a multitype point process on the linear network  $L$  for which the first- and second-moment intensity functions exist. The process will be called (multitype) *correlation stationary* if, for all  $i, j \in C$ , the multitype pair correlation  $\rho_{ij}^L$  is a function of distance only, i.e.  $\rho_{ij}^L(u, v) = \rho_{ij}^L\{d_L(u, v)\}$ .

*Theorem 1.* Let  $\mathbf{Y}$  be a correlation stationary multitype point process on a linear network  $L$ . For fixed  $u \in L$  and for subsets  $i, j \in C$  define

$$K_{ij}^{L,\text{ih}}(u, r) = \mathbb{E} \left[ \sum_{x_k \in \mathbf{X}_j} \frac{\mathbf{1}\{d_L(u, x_k) \leq r\}}{\lambda_j(x_k) m\{u, d_L(u, x_k)\}} \middle| u \in X_i \right]. \quad (18)$$

Then

$$K_{ij}^{L,\text{ih}}(u, r) = K_{ij}^{L,\text{ih}}(r) = \int_0^r \rho_{ij}^L(t) dt \quad (19)$$

does not depend on  $u$  and will be called the *multitype inhomogeneous K-function* of  $\mathbf{Y}$ . Furthermore

$$K_{ij}^{L,\text{ih}}(r) = \frac{1}{|L|} \mathbb{E} \left[ \sum_{x_l \in \mathbf{X}_i} \sum_{x_k \in \mathbf{X}_j} \frac{\mathbf{1}\{0 < d_L(x_l, x_k) \leq r\}}{\lambda_i(x_l) \lambda_j(x_k) m\{x_l, d_L(x_l, x_k)\}} \right]. \quad (20)$$

The intensity function  $\lambda_i(u)$ ,  $u \in L$ , can be estimated by parametric or non-parametric methods as described in Ang *et al.* (2012). Then a plug-in geometrically corrected estimator of  $K_{ij}^{L,\text{ih}}(r)$  is given by

$$\hat{K}_{ij}^{L,\text{ih}}(r) = \frac{1}{|L|} \sum_{x_l \in \mathbf{X}_i} \sum_{x_k \in \mathbf{X}_j} \frac{\mathbf{1}\{0 < d_L(x_l, x_k) \leq r\}}{\hat{\lambda}_i(x_l) \hat{\lambda}_j(x_k) m\{x_l, d_L(x_l, x_k)\}}, \quad (21)$$

where  $\hat{\lambda}_i(u)$  is an estimator of  $\lambda_i(u)$ .

Similarly, the multitype pair correlation function of a correlation stationary point process can be estimated by

$$\hat{\rho}_{ij}^{L,\text{ih}}(r) = \frac{1}{|L|} \sum_{x_l \in \mathbf{X}_i} \sum_{x_k \in \mathbf{X}_j} \frac{\kappa\{d_L(x_l, x_k) - r\}}{\hat{\lambda}_i(x_l) \hat{\lambda}_j(x_k) m\{x_l, d_L(x_l, x_k)\}}, \quad (22)$$

with  $\kappa$  again being a kernel function on  $\mathbb{R}$ .

An analogue of the mark connection function (15) is available for inhomogeneous processes, under the additional assumption that  $p_i(u) = \lambda_i(u)/\lambda(u)$  is constant for each type  $i$ . Since  $p_i(u)$  is the probability that a point at location  $u$  has type  $i$ , this amounts to assuming that the distribution of types does not depend on location. Then for points  $u, v \in L$ , with  $d_L(u, v) = r$  we have

$$\frac{\lambda_i(u) \lambda_j(v) \rho_{ij}^L(u, v)}{\lambda(u) \lambda(v) \rho^L(u, v)} = p_i p_j \frac{\rho_{ij}^L(u, v)}{\rho^L(u, v)} \quad (23)$$

and we define this quantity to be the generalization of  $p_{ij}(r)$  to the inhomogeneous case. A little algebra shows that equation (23) can be estimated by using the *same* ratio-of-sums estimator (17) as in the homogeneous case. The explanation given below equation (17) continues to hold.

## 6. Analysis of dendritic spines data

The techniques that were developed in the previous sections will now be applied to the dendritic spines data. Intensities are studied in Section 6.1. The results of this analysis lead us to split the data into two subsets, which are analysed respectively in Section 6.2 (assuming pseudostationarity) and Section 6.3 (using inhomogeneous summary functions).

### 6.1. Intensity of spines

The linear network that was depicted in Fig. 3 has a total edge length of 1934  $\mu\text{m}$ . There are  $n = 566$  spines in total, broken down into  $n_1 = 228$  mushroom,  $n_2 = 223$  stubby and  $n_3 = 115$  thin spines, where we henceforth use the numerals 1, 2 and 3 to refer to mushroom, stubby and thin spines respectively. Assuming constant intensity for each spine type, unbiased estimates  $\hat{\lambda}_j = n_j/|L|$  of the intensities are 0.118, 0.115, 0.059 spines  $\mu\text{m}^{-1}$  for the mushroom, stubby and thin spines respectively, and  $\hat{\lambda} = n./|L| = 0.293$  total spines  $\mu\text{m}^{-1}$ .

Kernel estimation of intensity is discussed in Xie and Yan (2008), Okabe *et al.* (2009), Shiode and Shiode (2009) and McSwiggan *et al.* (2013). Fig. 4 shows a kernel smoothing estimate of the spatially varying intensity of the spines regardless of type, using the ‘equal split continuous’

method of Okabe *et al.* (2009), section 5, with a Gaussian kernel with standard deviation  $10 \mu\text{m}$ . The ribbon width in Fig. 4 is proportional to the intensity estimate, which ranges between 0.01 and  $0.78 \text{ spines } \mu\text{m}^{-1}$ .

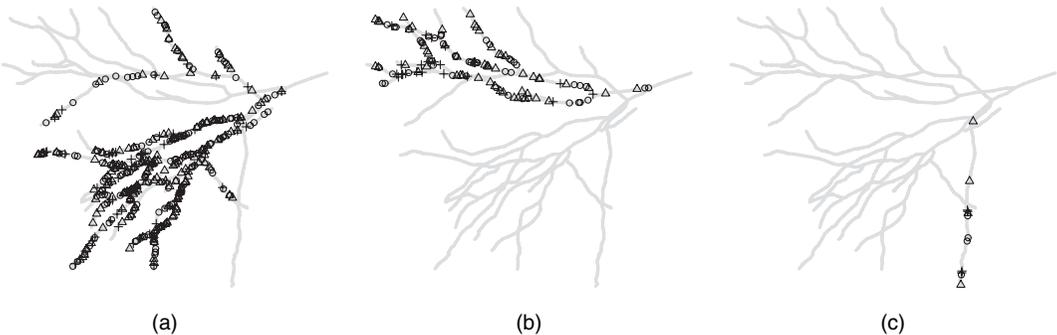
Despite the risk that spatial inhomogeneity may be mistaken for clustering (Bartlett, 1964, 1975), Fig. 4 strongly suggests that different branches of the dendrite network have different intensities of spines. The single unbroken filament in the lower right-hand side appears to have relatively low intensity of spines. The remainder of the network could be divided into upper and lower halves, with the lower half having greater intensity than the upper half. A similar conclusion is suggested when the same technique is applied separately to the spines of each type.

In biological terms it is conceivable that a dendrite network may exhibit different structural characteristics in different branches. A neuron has a single cell body which exerts centralized control over the transcription of genes into molecular messages which are then distributed throughout the entire dendritic tree. Uneven distribution of the messages may result in uneven structural development.

For any proposed split of the network into several subsets  $S_1, \dots, S_K$ , the  $\chi^2$ -test of the null hypothesis of constant intensity, against the alternative of different constant intensities in each subset, is based on  $X^2 = \sum_k \{n(\mathbf{X} \cap S_k) - e_k\}^2 / e_k$  with  $e_k = nl_k / l$ , where  $n(\mathbf{X} \cap S_k)$  is the number of spines in the  $k$ th subset and  $l_k$  is the length of dendrite in the  $k$ th subset, and  $l = \sum_k l_k$  and



**Fig. 4.** Kernel smoothing estimates of the intensity of spines (smoothing bandwidth  $10 \mu\text{m}$ ; the intensity value is proportional to the width of the ribbon)

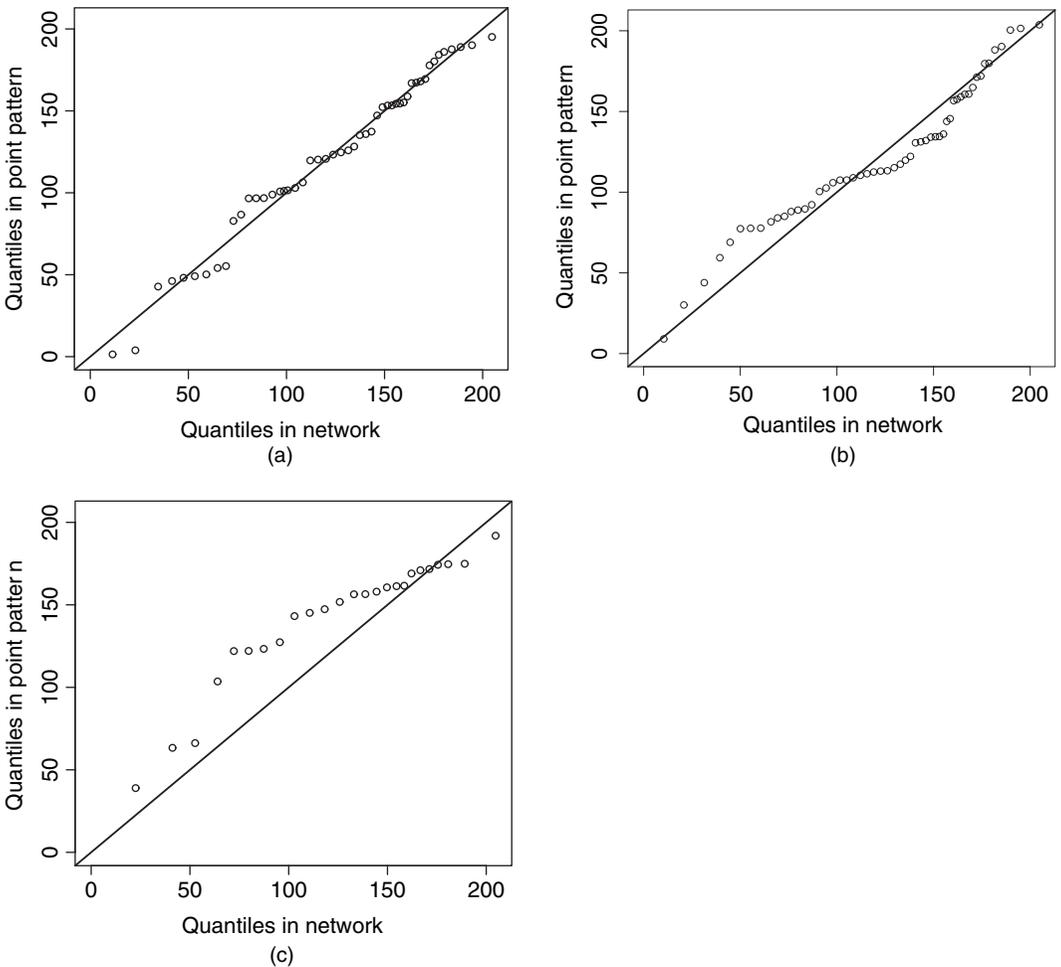


**Fig. 5.** Division of the dendrite spine data into three branches labelled (a) A, (b) B and (c) Z

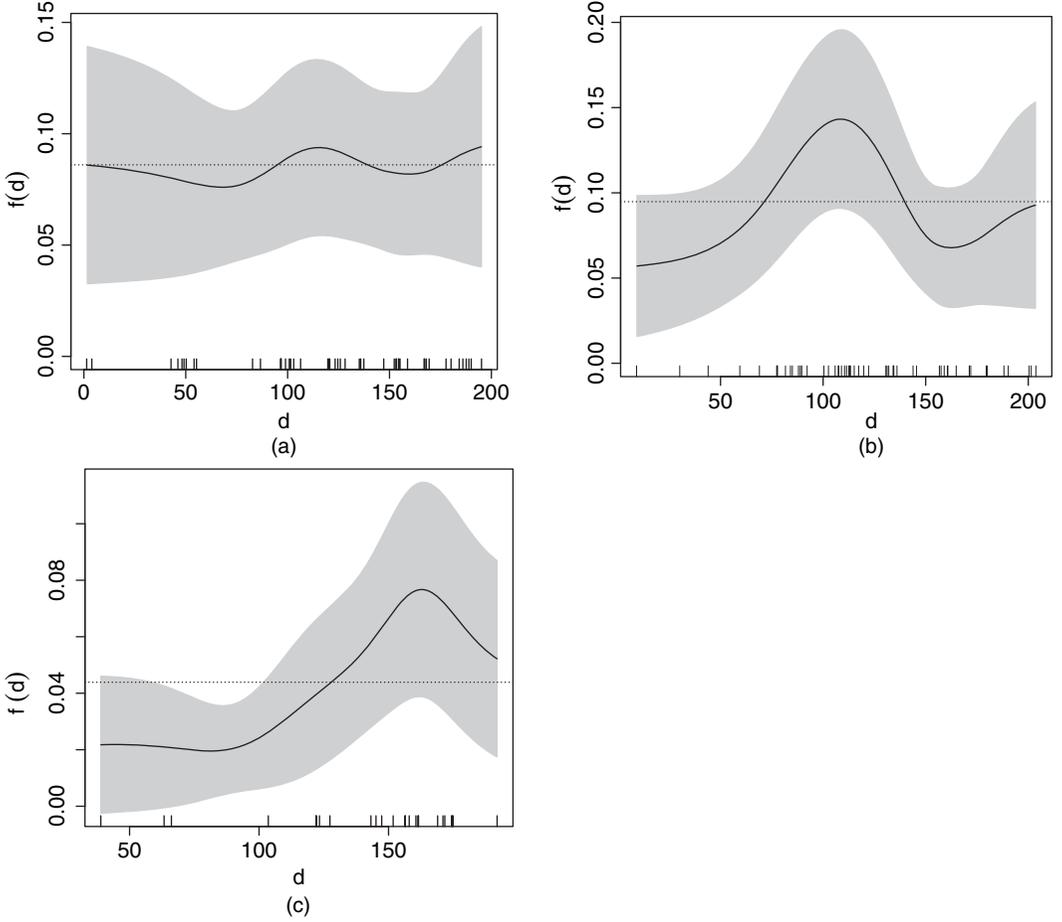
$n = \sum_k n(\mathbf{X} \cap S_k)$  are totals, and similarly for spines of a given type  $i$ . This test ignores the spine types, but the analogous test could also be performed on the subprocess  $\mathbf{X}_i$  of points of type  $i$ .

Selection of an appropriate split of the network into branches is a problem of model selection; in our case, because the network is a tree, we adopt a recursive splitting approach similar to that used for classification and regression trees (Breiman *et al.*, 1984). Starting from the cell body as the root of the tree, we visit each successive branching point of the network and apply the  $\chi^2$ -test statistic of uniformity to the subset *beyond* this branch point. The data will be split if the null hypothesis is rejected at, say, the 5% level. Of course the usual significance interpretation of the tests is not applicable in this context where multiple tests are performed on the same data.

The result is a split into the three branches that are shown in Fig. 5. Branch A has  $n = 419$  spines along  $|L| = 1203.2 \mu\text{m}$  of dendrites, with average intensity  $\hat{\lambda} = n/|L| = 0.348 \text{ spines } \mu\text{m}^{-1}$ , and the breakdown by spine type is  $n_1 = 173$ ,  $n_2 = 161$  and  $n_3 = 85$ . Branch B has  $n = 128$  spines along  $|L| = 569.6 \mu\text{m}$  of dendrites, with average intensity  $\hat{\lambda} = 0.225 \text{ spines } \mu\text{m}^{-1}$ , broken down



**Fig. 6.**  $Q-Q$ -plots of distance to soma for (a) mushroom, (b) stubby and (c) thin spine types in branch B of the dendrite network: order statistics of the observed distances from spines to the cell body (the vertical axis) are plotted against quantiles of distance from a uniformly random point to the cell body (the horizontal axis)



**Fig. 7.** Smoothing estimates of the function  $f_j$  for (a) mushroom, (b) stubby and (c) thin spine types in branch B of the dendrite network

into  $n_1 = 49$ ,  $n_2 = 54$  and  $n_3 = 25$ . Branch Z contains only four spines of each type in  $132 \mu\text{m}$  of dendrite and has several idiosyncrasies; for simplicity we delete this subset in the analysis that is reported here.

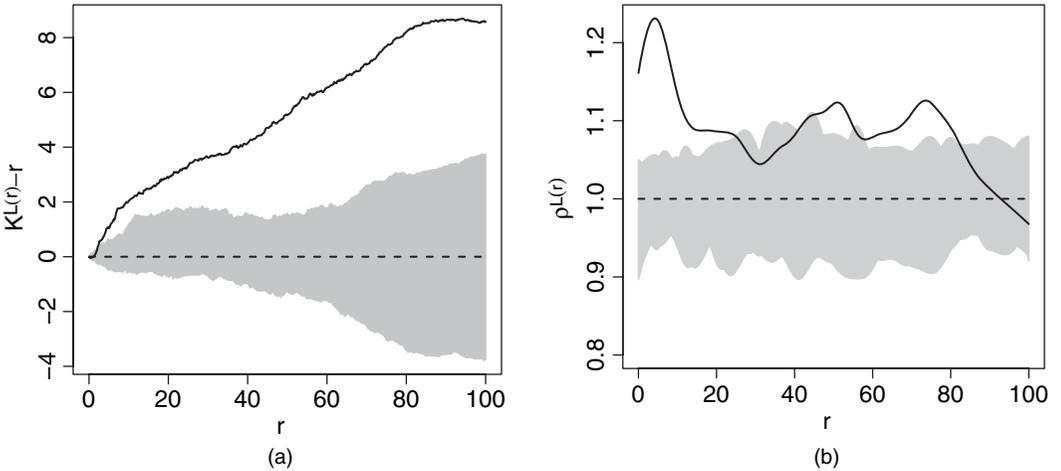
It is also of interest whether the intensity depends on distance from the cell body. Assume that, within a particular branch of the network, the intensity function  $\lambda_j(u)$ ,  $u \in L$ , of spines of type  $j$  depends only on distance to the soma (cell body)

$$\lambda_j(u) = f_j\{d(u)\}, \quad u \in L, \quad (24)$$

where  $f_j$  is a function to be estimated, and  $d(u)$  is the distance from the location  $u \in L$  to the cell body, measured by the shortest path in the dendrite network. Inference about  $f_j$  can be performed by comparing the empirical and theoretical distributions of distance, i.e. comparing the observed distribution of distance values  $d(x_i)$  at the data points  $x_i$  with the theoretical distribution of  $d(U)$  for a random point  $U$  uniformly distributed over the network  $L$ . This approach is practical because it does not involve the geometry of the linear network, once the distance values have been computed.

**Table 1.**  $p$ -values for tests of constant intensity for each spine type, assuming that equation (24) holds, within each branch

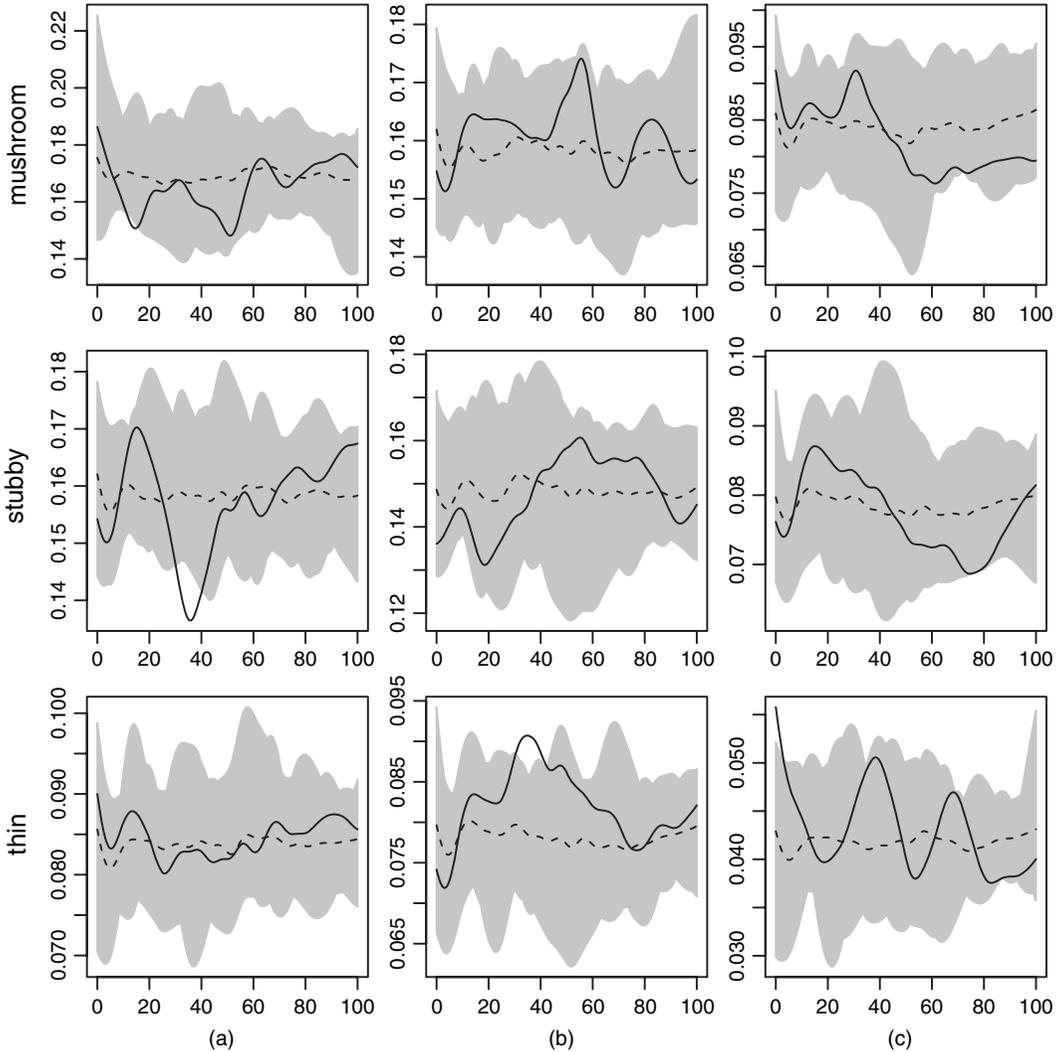
Branch	Spine type	Test statistic		
		Kolmogorov–Smirnov	$Z_1$	$Z_2$
A	Mushroom	0.299	0.917	0.590
	Stubby	0.373	0.902	0.953
	Thin	0.228	0.588	0.188
B	Mushroom	0.662	0.912	0.751
	Stubby	0.187	0.911	0.941
	Thin	0.018	0.411	0.034



**Fig. 8.** Second-order summaries of spine locations (regardless of type) in branch A, assuming constant intensity (—, empirical estimate; ■, pointwise envelope of the summary functions obtained from 39 simulations of a uniform Poisson process with the same estimated intensity): (a) centred  $K$ -function  $K^L(r) - r$  plotted against  $r$ ; (b) pair correlation function  $\rho^L(r)$

Our analysis suggests that, within each branch of the dendrite network, spines of a given type have constant intensity, except for the thin spines in branch B. Fig. 6 shows  $Q-Q$ -plots of distance to the soma for spines of each type, in branch B. Order statistics of the observed distances  $d(x_i)$  for spines of a given type are plotted against theoretical quantiles of distance  $d(U)$  at a uniformly random point  $U$  on the network. We computed the quantiles of  $d(U)$  by creating a fine grid of equally spaced locations  $u_k \in L$ , evaluating  $d(u_k)$  at each grid point, and sorting the values. Another approach would be to generate a large number of independent uniform random points  $U$  in  $L$  and to evaluate the distance  $d(U)$  at these points. The plot suggests that  $f_j$  may be constant for the mushroom and stubby types  $j=1$  and  $j=2$ , but not for the thin-type spines.

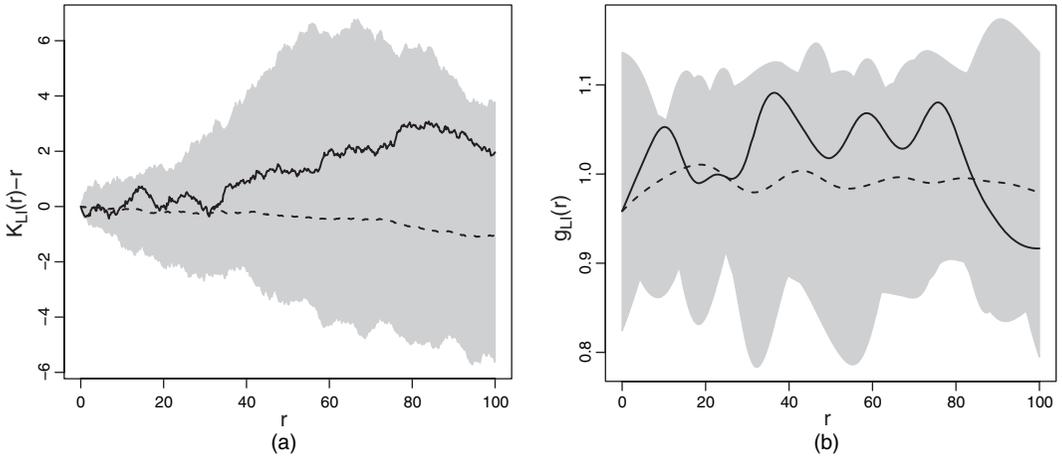
Assuming that equation (24) holds, the function  $f_j$  is related to the slope of the  $Q-Q$ -plot. One can use the non-parametric kernel smoothing estimators of  $f_j$  that were developed for spatial point processes (Baddeley *et al.*, 2012; Guan, 2008) which apply without modification to the case of a linear network, since again these do not depend on the geometry of the spatial



**Fig. 9.** Estimates of the mark connection function (15) between each pair of types, for branch A, assuming constant intensity of each type (■), pointwise envelope of the summary functions obtained from 39 simulations of random labelling): (a) mushroom type; (b) stubby type; (c) thin type

domain. The resulting estimates of  $f_j$  are shown in Fig. 7. Grey shading shows pointwise 95% confidence intervals based on the asymptotic normal approximation. The horizontal broken line shows the estimate by assuming that  $f_j$  is constant. Rug plots (Tufté, 1983) show the observed values  $d(x_j)$ .

Formal tests of the hypothesis that  $f_j$  is constant (assuming that equation (24) holds and that the process is Poisson) are also available for spatial point processes (Berman, 1986; Lawson, 1988; Waller *et al.*, 1992) and again these can be adapted immediately to linear networks. The tests compare the observed distribution of the covariate  $d$  at the data points with the null distribution of the covariate at a random point on the network. Table 1 shows the  $p$ -values that were obtained for the Kolmogorov–Smirnov test and for Berman’s  $Z_1$ - and  $Z_2$ -tests. The  $Z_1$ -test statistic of Berman (1986) is a standardized version of  $\sum_i d(x_i)$  whereas the  $Z_2$ -statistic



**Fig. 10.** Second-order summaries of spine locations (regardless of type) in branch B by using inhomogeneous intensity estimate (25) (■), pointwise envelope of the summary functions obtained from 39 simulations of an inhomogeneous Poisson process with the same estimated intensity): (a) centred inhomogeneous  $K$ -function  $K^L,ih(r) - r$  plotted against  $r$ ; (b) inhomogeneous estimate of pair correlation function  $\hat{\rho}^{L,ih}(r)$

is a standardized version of  $\sum_i F_0\{d(x_i)\}$  where  $F_0$  is the cumulative distribution function of  $d(U)$  under the null hypothesis. The provisional conclusion is that the mushroom and stubby spine types have constant intensity, but the intensity of the thin spines is increasing with greater distance from the soma, in branch B.

## 6.2. Second-order analysis of dendrite branch A

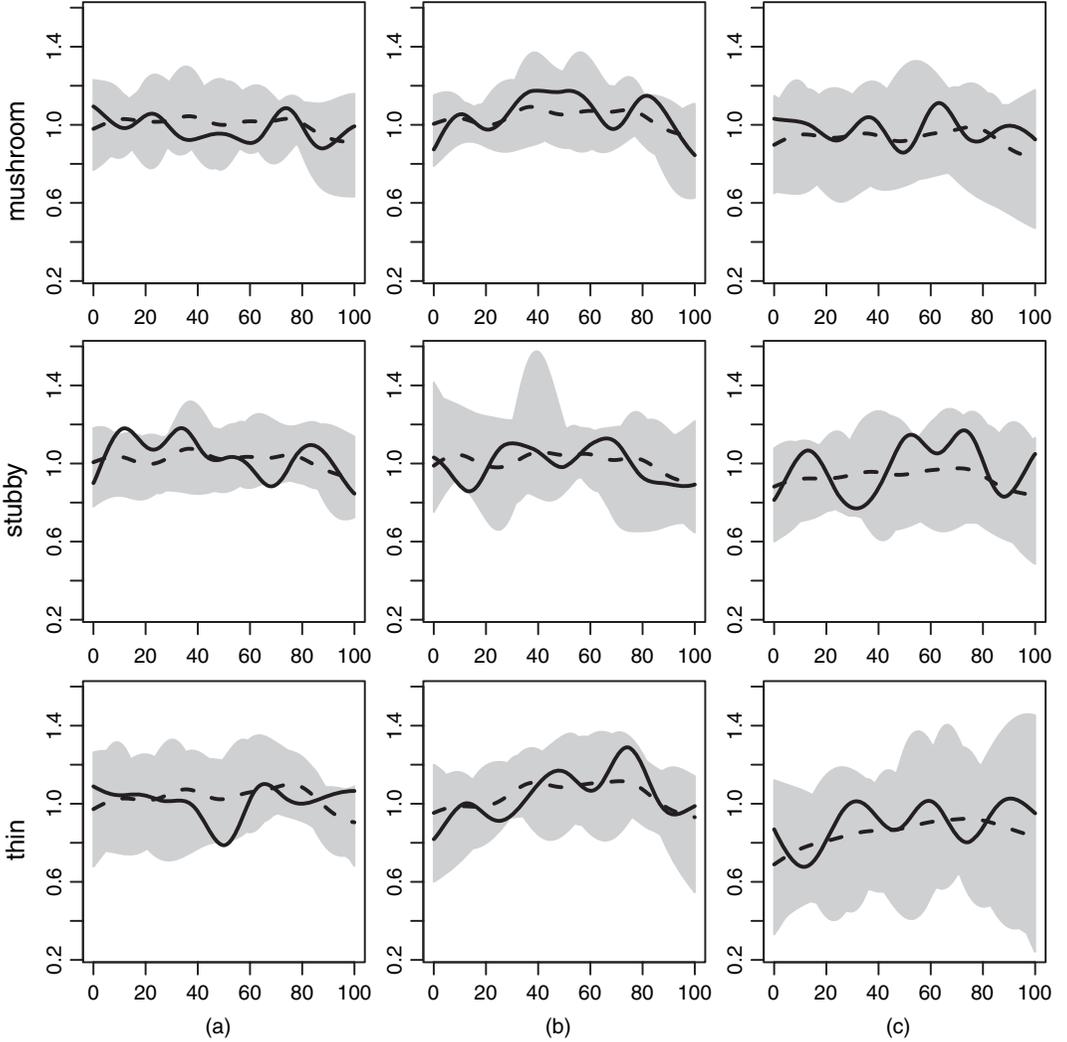
Here we apply the methods of Section 4 to branch A of the dendritic spines data set that was identified in Fig. 5. Branch A is tentatively believed to have uniform intensity of each spine type. We assume that the underlying process is second order pseudostationary.

For estimating pair correlation functions, the smoothing kernel  $\kappa$  in equations (9), (17) or (22) will be the Gaussian density, with standard deviation selected by Silverman's rule of thumb (Silverman (1986), page 48, equation (3.31)) although this seems to produce slight undersmoothing.

Fig. 8 shows estimates of the geometrically corrected  $K$ -function  $K^L(r)$  and pair correlation function  $\rho^L(r)$  for the unmarked point pattern  $\mathbf{X}$  of spines regardless of type. Grey shading represents the pointwise envelope of the summary functions that were obtained from 39 simulations of a uniform Poisson process with the same estimated intensity. There is strong evidence of spatial clustering (assuming uniform intensity), confirmed by formal Monte Carlo tests based on the maximum absolute deviation or integrated squared deviation of the summary statistic (Cressie (1991), page 667, equation (8.5.42), and Diggle (2003), page 12, equation (2.7)).

Fig. 9 shows the estimates (17) of the mark connection function  $p_{ij}(r)$  of equation (15), computed for the data in branch A. Grey shading shows the pointwise envelope of the estimates that were obtained from 39 random patterns obtained by randomly permuting the spine-type labels while holding the spine locations fixed. These plots and the associated Monte Carlo tests suggest no evidence against the hypothesis of random labelling.

Thus a tentative conclusion for branch A is that spine locations have uniform intensity but are spatially clustered; the distribution of spine types does not depend on location; and the types of neighbouring spines are independent. The evidence for spatial clustering seems strong but is



**Fig. 11.** Inhomogeneous multitype pair correlation functions (—) for each pair of spine types in branch B, together with envelopes of simulations from inhomogeneous random labelling, as explained in the text: (a) mushroom type; (b) stubby type; (c) thin type

quite sensitive to misspecification of the intensity; for example, the strength of evidence depends on how the network is divided into subbranches.

### 6.3. Inhomogeneous second-order analysis of branch B

Here we apply the methods of Section 5 to branch B of the dendritic spines data set that was identified in Fig. 5. For branch B, the intensity function of  $\mathbf{X}$  is estimated by

$$\hat{\lambda} \cdot (u) = \hat{\lambda}_1 + \hat{\lambda}_2 + \hat{f}_3\{d(u)\}, \quad (25)$$

where  $\hat{\lambda}_1$  and  $\hat{\lambda}_2$  are the estimated intensities of mushroom and stubby spines respectively, assuming that these are constant, whereas  $\hat{f}_3\{d(u)\}$  is the estimated intensity of thin spines at a distance  $d(u)$  from the cell body, assuming that equation (24) holds for the thin spines.

Fig. 10 shows the estimated inhomogeneous  $K$ -function and inhomogeneous pair correlation function for the unmarked pattern of spines of all types in branch B, using the estimated intensity function  $\hat{\lambda} \cdot (u)$ . It does not suggest any evidence of spatial clustering.

Fig. 11 shows the multitype pair correlation functions for each pair of spine types in branch B, together with pointwise significance bands computed in the following way. Simulated point patterns were generated by assigning new random types to the spines, holding their locations fixed. For a spine at location  $u$ , the probabilities of assigning the labels *mushroom*, *stubby* and *thin* were  $p_1(u) = \hat{\lambda}_1 / \hat{\lambda} \cdot (u)$ ,  $p_2(u) = \hat{\lambda}_2 / \hat{\lambda} \cdot (u)$  and  $p_3(u) = \hat{f}_3\{d(u)\} / \hat{\lambda} \cdot (u)$  respectively. Figs 10 and 11 suggest no evidence against the null hypothesis of an inhomogeneous Poisson process.

In conclusion, there is strong evidence that different branches of the dendrite network may have different patterns of spines. There are large branches within which the mushroom and stubby spines appear to have uniform intensity. In some branches of the network (B and Z) there is evidence that the intensity of thin spines is increasing with distance from the cell body. In one branch (A) there is evidence of spatial clustering of the locations of spines, assuming uniform intensity; this requires further investigation to validate the assumption of uniformity. Conditionally on the spine locations, there is no evidence of dependence between the spine types.

## 7. Discussion

This paper has developed and demonstrated generic tools, based mainly on first and second moments, for analysing a multitype point pattern on a linear network. In any field of statistics, estimates of correlations or interactions are highly sensitive to misspecification of the first moment or main effect. This *caveat* also applies to spatial point processes (Bartlett, 1964) and in particular to the data analysis in this paper.

For the application to dendritic networks, careful attention to the intensity was crucial. Visual inspection of the raw and kernel-smoothed data suggested several new models for inhomogeneous intensity functions that are scientifically meaningful. The second-order analysis was highly sensitive to the fitted intensity. This was not so with the other applications that we have studied (Ang *et al.*, 2012). It is conceivable that strong inhomogeneity is more likely to occur in tree-like branching networks, such as the dendritic network, than in networks with loops, such as road networks. Alternatively the dendrite example could be anomalous, perhaps by virtue of the highly structured molecular and genetic messaging in the network.

A different analysis (Jammalamadaka *et al.*, 2013), of a suite of data that included Fig. 1, found a positive association between the types of immediately neighbouring spines, conditionally on the spine locations. That analysis needs to be revisited to determine whether the positive association still persists when the network is divided into homogeneous branches as we did above. If it does persist, then the most likely explanation of the different conclusions from the two analyses is that there is very short-range nearest neighbour dependence between spines.

Dendritic networks are three dimensional, but existing computational techniques for spatial data on linear networks (Okabe and Sugihara, 2012) mostly assume that the network lies in two-dimensional space. Fortunately, neurons in cell culture *in vitro* are almost flat, so we may ignore the third dimension, except when resolving the connectivity of the network. Thus, existing computational techniques are applicable to neurons in cell culture but would require algorithmic modification to deal with neurons *in vivo*.

The data came from a designed experiment in which the ‘response’ for each experimental unit is a point pattern. It is possible to pool first- and second-order summary statistics across replicate point patterns (Baddeley *et al.*, 1993; Bell and Grunwald, 2004; Diggle *et al.*, 1991) but additional methodology needs to be developed.

Numerous *caveats* apply to the biological interpretation of our results. The findings of a cell culture experiment do not necessarily extrapolate to cells *in vivo*. However, subject to those *caveats*, we have demonstrated evidence that different branches of a dendrite network appear to have different, homogeneous, concentrations of spines. Evidence for clustering is at best equivocal.

A more searching analysis of the dendrite data would require the ability to fit explicit statistical models to the data. Point process modelling on a linear network is under development and will be demonstrated in a sequel paper.

## Acknowledgements

We thank Dr Kenneth Kosik, Dr Sourav Banerjee and the Kosik Laboratory at the University of California at Santa Barbara for introducing us to the problem and for generously granting access to the neuronal data. The data are available from [http://bisque.ece.ucsb.edu/client\\_service/view?resource=http://bisque.ece.ucsb.edu/data\\_service/data\\_set/2653471](http://bisque.ece.ucsb.edu/client_service/view?resource=http://bisque.ece.ucsb.edu/data_service/data_set/2653471).

Aruna Jammalamadaka was supported by National Science Foundation award III-0808772. Adrian Baddeley and Gopalan Nair were supported by the Australian Research Council (discovery grants DP 130102322 and DP 130104470 and an Australian Research Council Discovery Outstanding Researcher Award). The majority of the research was conducted during Gopalan Nair's sabbatical with the Commonwealth Scientific and Industrial Research Organisation's Division of Mathematics, Informatics and Statistics.

All the calculations were performed in the R language using the package `spatstat` (Baddeley and Turner, 2005).

We thank the referees and Eddy Campbell, Robin Milne, Andrew Hardegen and Tom Lawrence for helpful advice on the paper.

## Appendix A: Proof of lemma 2

Let  $Z(r) = n_{ij} \bar{\rho}_{ij}^L(r) / |L|$  denote the double sum in equation (9). By the second-order Campbell formula (5)

$$\mathbb{E}[Z(r)] = \lambda_i \lambda_j \int_L \int_L \frac{\kappa\{d_L(u, v) - r\}}{m\{u, d_L(u, v)\}} \rho_{ij}^L(u, v) \, d_1 u \, d_1 v. \quad (26)$$

For fixed  $u$ , the mapping  $v \mapsto d_L(u, v)$  is a piecewise linear function with unit Jacobian. Invoking the change of variables

$$\int_L h\{d_L(u, v)\} \, d_1 v = \int_0^\infty h(t) m(u, t) \, dt \quad (27)$$

for any measurable function  $h: [0, \infty) \rightarrow \mathbb{R}$  (shown in Ang *et al.* (2012)) equation (26) becomes

$$\begin{aligned} \int_L \frac{\kappa\{d_L(u, v) - r\}}{m\{u, d_L(u, v)\}} \rho_{ij}^L(u, v) \, d_1 v &= \int_0^\infty \sum_{v: d_L(u, v)=t} \frac{\kappa\{d_L(u, v) - r\}}{m\{u, d_L(u, v)\}} \rho_{ij}^L(t) \, dt \\ &= \int_0^\infty \sum_{v: d_L(u, v)=t} \frac{\kappa(t - r)}{m(u, t)} \rho_{ij}^L(t) \, dt \\ &= \int_0^\infty \frac{\kappa(t - r)}{m(u, t)} m(u, t) \rho_{ij}^L(t) \, dt \\ &= \int_0^\infty \kappa(t - r) \rho_{ij}^L(t) \, dt = \bar{\rho}_{ij}^L(r) \end{aligned}$$

(provided that  $m(u, t) > 0$  for all  $t < r$ ) since the sum contains  $m(u, t)$  terms. Hence

$$\mathbb{E}[Z(r)] = \lambda_i \lambda_j |L| \bar{\rho}_{ij}^L(r)$$

and the result follows.

## References

- Ang, Q., Baddeley, A. and Nair, G. (2012) Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scand. J. Statist.*, **39**, 591–617.
- Baddeley, A. (2010) Multivariate and marked point processes. In *Handbook of Spatial Statistics* (eds A. Gelfand, P. Diggle, M. Fuentes and P. Guttorp), ch. 21, pp. 371–402. Boca Raton: CRC Press.
- Baddeley, A., Chang, Y., Song, Y. and Turner, R. (2012) Nonparametric estimation of the dependence of a spatial point process on a spatial covariate. *Statist. Interf.*, **5**, 221–236.
- Baddeley, A., Møller, J. and Waagepetersen, R. (2000) Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statist. Neerland.*, **54**, 329–350.
- Baddeley, A. J., Moyeed, R. A., Howard, C. V. and Boyde, A. (1993) Analysis of a three-dimensional point pattern with replication. *Appl. Statist.*, **42**, 641–668.
- Baddeley, A. and Silverman, B. (1984) A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics*, **40**, 1089–1094.
- Baddeley, A. and Turner, R. (2005) Spatstat: an R package for analyzing spatial point patterns. *J. Statist. Softw.*, **12**, 1–42.
- Bartlett, M. (1964) A note on spatial pattern. *Biometrics*, **20**, 891–892.
- Bartlett, M. (1975) *The Statistical Analysis of Spatial Pattern*. London: Chapman and Hall.
- Bell, M. and Grunwald, G. (2004) Mixed models for the analysis of replicated spatial point patterns. *Biostatistics*, **5**, 633–648.
- Berman, M. (1986) Testing for spatial association between a point process and another stochastic process. *Appl. Statist.*, **35**, 54–62.
- Bjaalie, J. and Diggle, P. (1990) Statistical analysis of corticopontine neuron distribution in visual areas 17, 18 and 19 of the cat. *J. Compar. Neurol.*, **295**, 15–32.
- Breiman, L., Friedman, J., Stone, C. and Olshen, R. (1984) *Classification and Regression Trees*. New York: Chapman and Hall.
- Chen, B., Leser, G., Jackson, D. and Lamb, R. (2008) The influenza virus M2 protein cytoplasmic tail interacts with the M1 protein and influences virus assembly at the site of virus budding. *J. Virol.*, **82**, 10059–10070.
- Cressie, N. (1991) *Statistics for Spatial Data*. New York: Wiley.
- Diggle, P. (1986) Displaced amacrine cells in the retina of a rabbit: analysis of a bivariate spatial point pattern. *J. Neurosci. Meth.*, **18**, 115–125.
- Diggle, P. (2003) *Statistical Analysis of Spatial Point Patterns*, 2nd edn. London: Hodder Arnold.
- Diggle, P., Eglén, S. and Troy, J. (2005) Modelling the bivariate spatial distribution of amacrine cells. In *Case Studies in Spatial Point Process Modeling* (eds A. Baddeley, P. Gregori, J. Mateu, R. Stoica and D. Stoyan), pp. 215–233. New York: Springer.
- Diggle, P., Lange, N. and Benes, F. M. (1991) Analysis of variance for replicated spatial point patterns in clinical neuroanatomy. *J. Am. Statist. Ass.*, **86**, 618–625.
- Fleischer, F., Beil, M., Kazda, M. and Schmidt, V. (2005) Analysis of spatial point patterns in microscopic and macroscopic biological image data. In *Case Studies in Spatial Point Process Modeling* (eds A. Baddeley, P. Gregori, J. Mateu, R. Stoica and D. Stoyan), pp. 235–260. New York: Springer.
- Gelfand, A., Diggle, P., Fuentes, M. and Guttorp, P. (eds) (2010) *Handbook of Spatial Statistics*. Boca Raton: CRC Press.
- Grabarnik, P. and Särkkä, A. (2009) Modelling the spatial structure of forest stands by multivariate point processes with hierarchical interactions. *Ecol. Modelling*, **220**, 1232–1240.
- Guan, Y. (2008) On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. *J. Am. Statist. Ass.*, **103**, 1238–1247.
- Harkness, R. D. and Isham, V. (1983) A bivariate spatial point pattern of ants' nests. *Appl. Statist.*, **32**, 293–303.
- Harnett, M., Makara, J., Spruston, N., Kath, W. and Magee, J. (2012) Synaptic amplification by dendritic spines enhances input cooperativity. *Nature*, **491**, 599–602.
- Harris, K. and Kater, S. (1994) Dendritic spines: cellular specializations imparting both stability and flexibility to synaptic function. *A. Rev. Neurosci.*, **17**, 341–371.
- Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008) *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester: Wiley.
- Irwin, S., Patel, B., Idupulapati, M., Harris, J., Crisostomo, R., Larsen, B., Kooy, F., Willems, P., Cras, P., Kozłowski, P., Swain, R., Weiler, I. and Greenough, W. (2001) Abnormal dendritic spine characteristics in the temporal and visual cortices of patients with Fragile-X syndrome: a quantitative examination. *Am. J. Med. Genet.*, **98**, 161–167.
- Jammalamadaka, A., Banerjee, S., Manjunath, B. and Kosik, K. (2013) Statistical analysis of dendritic spine distributions in rat hippocampal cultures. *BMC Bioinform.*, **14**, article 287.
- Kingman, J. (1993) *Poisson Processes*. New York: Oxford University Press.
- Kivlekväl, K., Fedorov, D., Obara, B., Singh, A. and Manjunath, B. (2010) Bisque: a platform for bioimage analysis and management. *Bioinformatics*, **26**, 544–552.
- Lawson, A. (1988) On tests for spatial trend in a nonhomogeneous Poisson process. *J. Appl. Statist.*, **15**, 225–234.

- van Lieshout, M. and Baddeley, A. (1999) Indices of dependence between types in multivariate point patterns. *Scand. J. Statist.*, **26**, 511–532.
- Lotwick, H. W. and Silverman, B. W. (1982) Methods for analysing spatial processes of several types of points. *J. R. Statist. Soc. B*, **44**, 406–413.
- Mamaghani, M., Andersson, M. and Krieger, P. (2010) Spatial point pattern analysis of neurons using Ripley's K-function in 3D. *Front. Neuroinform.*, **4**, no. 9, 1–10.
- McSwiggan, G., Baddeley, A. and Nair, G. (2013) Kernel smoothing on a linear network. To be published.
- Millet, L., Collens, M., Perry, G. and Bashir, R. (2011) Pattern analysis and spatial distribution of neurons in culture. *Integr. Biol.*, **3**, 1167–1178.
- Nimchinsky, E., Sabatini, B. and Svoboda, K. (2002) Structure and function of dendritic spines. *A. Rev. Physiol.*, **64**, 313–353.
- Ohser, J. and Stoyan, D. (1981) On the second-order and orientation analysis of planar stationary point processes. *Biometr. J.*, **23**, 523–533.
- Okabe, A. and Satoh, T. (2009) Spatial analysis on a network. In *The SAGE Handbook on Spatial Analysis* (eds A. Fotheringham and P. Rogers), ch. 23, pp. 443–464. London: Sage.
- Okabe, A., Satoh, T. and Sugihara, K. (2009) A kernel density estimation method for networks, its computational method and a GIS-based tool. *Int. J. Geog. Inform. Sci.*, **23**, 7–32.
- Okabe, A. and Sugihara, K. (2012) *Spatial Analysis along Networks*. Hoboken: Wiley.
- Okabe, A. and Yamada, I. (2001) The K-function method on a network and its computational implementation. *Geog. Anal.*, **33**, 271–290.
- Pedro, N., Carmo-Fonseca, M. and Fernandes, P. (1984) Quantitative analysis of pore patterns on rat prostate nuclei using spatial statistics methods. *J. Microsc.*, **134**, 271–280.
- Ripley, B. (1976) The second-order analysis of stationary point processes. *J. Appl. Probab.*, **13**, 255–266.
- Ripley, B. D. (1977) Modelling spatial patterns (with discussion). *J. R. Statist. Soc. B*, **39**, 172–212.
- Rodriguez, A., Ehlenberger, D., Dickstein, D., Hof, P. and Wearne, S. (2008) Automated three-dimensional detection and shape classification of dendritic spines from fluorescence microscopy images. *PLOS ONE*, **3**, article e1997.
- Shiode, S. and Shiode, N. (2009) Detection of hierarchical point agglomerations by the network-based variable clumping method. *Int. J. Geog. Inform. Sci.*, **23**, 75–92.
- Silverman, B. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- Tufte, E. (1983) *The Visual Display of Quantitative Information*, 1st edn. Cheshire: Graphics Press.
- Waller, L., Turnbull, B., Clark, L. and Nasca, P. (1992) Chronic disease surveillance and testing of clustering of disease and exposure: application to leukaemia incidence and TCE-contaminated dumpsites in upstate New York. *Environmetrics*, **3**, 281–300.
- Wearne, S., Rodriguez, A., Ehlenberger, D., Rocher, A., Henderson, S. and Hof, P. (2005) New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales. *Neuroscience*, **136**, 661–680.
- Webster, S., Diggle, P., Clough, H., Green, R. and French, N. (2005) Strain-typing transmissible spongiform encephalopathies using replicated spatial data. In *Case Studies in Spatial Point Process Modeling* (eds A. Baddeley, P. Gregori, J. Mateu, R. Stoica and D. Stoyan), pp. 197–214. New York: Springer.
- Xie, Z. and Yan, J. (2008) Kernel density estimation of traffic accidents in a network space. *Comput. Environ. Urb. Syst.*, **32**, 396–406.
- Yadav, A., Gao, Y., Rodriguez, A., Dickstein, D., Wearne, S., Luebke, J., Hof, P. and Weaver, C. (2012) Morphologic evidence for spatially clustered spines in apical dendrites of monkey neocortical pyramidal cells. *J. Comp. Neurol.*, **520**, 2888–2902.
- Yuste, R. (2011) Dendritic spines and distributed circuits. *Neuron*, **71**, 772–781.