Modelling columnarity of pyramidal cells in the human cerebral cortex

Andreas Dyreborg Christoffersen¹, Jesper Møller¹,* and Heidi Søgaard Christensen²

Aalborg University

Summary

For modelling the location of pyramidal cells in the human cerebral cortex, we suggest a hierarchical point process in \( \mathbb{R}^3 \) that exhibits anisotropy in the form of cylinders extending along the \( z \)-axis. The model consists first of a generalised shot noise Cox process for the \( xy \)-coordinates, providing cylindrical clusters, and next of a Markov random field model for the \( z \)-coordinates conditioned on the \( xy \)-coordinates, providing either repulsion, aggregation or both within specified areas of interaction. Several cases of these hierarchical point processes are fitted to two pyramidal cell data sets, and of these a final model allowing for both repulsion and attraction between the points seem adequate. We discuss how the final model relates to the so-called minicolumn hypothesis in neuroscience.

Key words: anisotropy; cylindrical \( K \)-function; determinantal point process; hierarchical point process model; line cluster point process; Markov random field; minicolumn hypothesis; pseudo-likelihood.

1. Introduction

The structuring of neurons in the human brain is a subject of great interest since abnormal structures may be linked to certain neurological diseases (see Buxhoeveden & Casanova 2002; Casanova et al. 2006; Esiri & Chance 2006; Casanova 2007; Chance et al. 2011). A specific structure that has been extensively studied in the biological literature is the so called ‘minicolumn’ structure of the cells in the cerebral cortex (see Buxhoeveden & Casanova 2002; Rafati et al. 2016 and references therein). Rafati et al. (2016) characterised these minicolumns as ‘linear aggregates of neurons organised vertically in units that traverse the cortical layer II–VI, and have in humans a diameter of 35–60 \( \mu \text{m} \) and consist typically of 80–100 neurons’.

1.1. Data

In this paper, we analyse the structuring of pyramidal cells, which make up approximately 75–80% of all neurons (Buxhoeveden & Casanova 2002) and are pyramid-shaped cells, where the so-called apical dendrite extends from the top/apex of a pyramidal cell. Specifically, the

*Author to whom correspondence should be addressed.
¹Department of Mathematical Sciences, Aalborg University, Skjernvej 4A, 9220, Aalborg Ø, Denmark.
²Department of Clinical Medicin, Aalborg University, Søndre Skovvej 15, 9000, Aalborg, Denmark. e-mail: jm@math.aau.dk

Acknowledgements. This work was supported by The Danish Council for Independent Research – Natural Sciences, grant DFF – 7014-00074 ‘Statistics for point processes in space and beyond’, and by the ‘Centre for Stochastic Geometry and Advanced Bioimaging’, funded by grant 8721 from the Villum Foundation. We are thankful to Ali H. Rafati for collecting the data analysed in this paper and to Jens R. Nyengaard and Ninna Vihrs for helpful comments.
paper is concerned with pyramidal cells from the so-called Brodmann area four of the human cerebral cortex. The neocortex constitutes most of the cerebral cortex and can be divided into six layers. We consider two data sets consisting of the locations and orientations of pyramidal cells in a section of the third and fifth layer, respectively. Here, each location is a three-dimensional (3D) coordinate representing the centre of a pyramidal cell’s nucleolus and each orientation is a unit vector representing the apical dendrite’s position relative to the corresponding nucleolus.

The left part of Figure 1 shows two point pattern data sets of 634 and 548 nucleolus locations which will be referred to as $L_3$ (top) and $L_5$ (bottom) respectively (for plot of the orientations for $L_3$, see Møller et al. 2019). Note that the observation window $W$ for the cell locations is a rectangular region with side lengths 492.70 $\mu$m, 132.03 $\mu$m and 407.70 $\mu$m for $L_3$, and 488.40 $\mu$m, 138.33 $\mu$m and 495.40 $\mu$m for $L_5$. Notice also that the nucleolus locations are recorded such that the $z$-axis is perpendicular to the so-called pial surface of the brain. In accordance to the minicolumn hypothesis, this implies that the minicolumns extend parallel to the $z$-axis. In the right part of Figure 1, we have therefore shown the two-dimensional $xy$-locations of the two 3D point pattern data sets.

1.2. Background and objective

Møller et al. (2019) found independence between locations and orientations for $L_3$ meaning that the two components may be modelled separately; the same conclusion has afterwards been drawn for $L_5$. As they also found a suitable inhomogeneous Poisson process model for the orientations, and since it is difficult to visually detect any structure in the point patterns shown in Figure 1, the focus of this paper is on modelling the nucleolus locations. In particular, we aim at modelling the nucleolus locations for $L_3$ respective $L_5$ by a spatial point process with a columnar structure and discuss to what extent this relates to the minicolumn hypothesis. Note that for the two data sets we use the same notation $X$ for the spatial point process, and we view $X$ as a random finite subset of $W$.

To the best of our knowledge, the so-called Poisson line cluster point process (see Møller, Safavimanesh & Rasmussen 2016) is the only existing point process model for modelling columnar structures. This model was considered by Rafati et al. (2016) in connection to another data set of pyramidal nucleolus locations, but it was not fitted to that data set. For each point pattern considered in the present paper, we notice later that a more advanced model than the Poisson line cluster point process is needed; below we describe such a model for $X$.

1.3. Hierarchical point process models

We consider a hierarchical model for $X$ as follows. Note that the observation window is a product space, $W = W_{xy} \times W_z$, where $W_{xy}$ is a rectangular region in the $xy$-plane and $W_z$ is an interval on the $z$-axis. Let $X_{xy}$ be the point process of $xy$-coordinates in $X$ (i.e. given by the projection of $X$ onto the $xy$-plane). To each point $(x_i, y_i) \in X_{xy}$ associate the corresponding, $z$-coordinate $z_i$ so that $(x_i, y_i, z_i) \in X$. Conditioned on the $xy$-points $X_{xy} = \{(x_i, y_i)\}_{i=1}^n$, let $X_z = (z_i)_{i=1}^n$ be the random vector of $z$-coordinates in $X$ (with an arbitrary ordering of these $n$ points and where $(z_i)_{i=1}^n$ is shorthand notation for $(z_1, \ldots, z_n)$). Clearly, $X$ is in a one-to-one correspondence with $X_{xy}$ and $X_z$, and we model first $X_{xy}$ and second $X_z$ conditioned on $X_{xy}$. Further details are given below and in Sections 3–5.
1.3.1. The model for $X_{xy}$

For $X_{xy}$, we consider the restriction of a cluster point process in $\mathbb{R}^2$ to $W_{xy}$ defined as follows. Let $\Phi$ be a stationary point process on $\mathbb{R}^2$ (i.e. its distribution is invariant under planar translations) with intensity $\kappa > 0$, and associate to each point $(\xi, \eta) \in \Phi$ a point process $X_{(\xi, \eta)}$ that is concentrated around the line in $\mathbb{R}^3$, which is perpendicular to the $xy$-plane, with intersection point $(\xi, \eta, 0)$. We refer to $X_{(\xi, \eta)}$ as the cylindrical cluster associated with $(\xi, \eta)$. Let $P_{xy}(X_{(\xi, \eta)} \cap W)$ denote the projection onto the $xy$-plane of the observed part of the cylindrical cluster. For short, we refer to the non-empty $P_{xy}(X_{(\xi, \eta)} \cap W)$ as the projected cluster with centre point $(\xi, \eta)$. Then we let

$$X_{xy} = \bigcup_{(\xi, \eta) \in \Phi} P_{xy}(X_{(\xi, \eta)} \cap W).$$

Furthermore, conditioned on $\Phi$, we assume that the projected clusters are independent and each non-empty $P_{xy}(X_{(\xi, \eta)} \cap W)$ is distributed as the intersection of $W_{xy}$ with a finite planar Poisson process translated by the centre point $(\xi, \eta)$; this Poisson process has intensity function $af$, where $a$ is the length of the interval $W_z$, $a > 0$ is a parameter, and $f$ is the probability density function of a bivariate zero-mean isotropic normal distribution with standard deviation $\sigma > 0$. Thus, ignoring boundary effects, $a\sigma$ is the expected size (or number of points) of a projected cluster and $\sigma$ controls the spread of points in a projected cluster. Specifically, we let first $\Phi$ be a planar stationary Poisson process and later a stationary
determinantal point process (Lavancier, Møller & Rubak 2015), since later we observe for a fitted Poisson process a very low expected number of points in a projected cluster and that there is a need for a repulsive model in order to obtain less overlap between the projected clusters.

1.3.2. The model for $X_z$ conditioned on $X_{xy}$

Consider the special case where $\Phi$ is a planar stationary Poisson process and $X_z$ conditioned on $X_{xy}$ is a homogeneous binomial point process, that is the $n$ points in $X_z$ are independent and uniformly distributed on $W_z$. Note that $X_z$ depends only on $X_{xy}$ through the number of points in $X_{xy}$. It becomes clear in Section 4 that this special case corresponds to a degenerate case of a Poisson line cluster point process (PLCPP) as considered in Møller et al. (2016).

In Section 5, we consider several other cases than a homogeneous binomial point process for $X_z$ conditioned on $X_{xy}$. In the end, in comparison with the PLCPP model suggested in Møller et al. (2016) and Rafati et al. (2016) for the description of another data set of pyramidal nucleolus locations, we suggest and fit a rather complex hierarchical point process model. Our final model describes columnar structures of the nucleolus locations in each data set $L3$ and $L5$, with repulsion between nucleolus locations given by a hard core condition on a small scale and a stunted cylindrical interaction region on a larger scale, as well as clustering between nucleolus locations given by an elongated cylindrical interaction region. In particular, our final fitted model describes a columnar structure but with much smaller columns than expected under the minicolumn hypothesis.

1.4. Outline

The remainder of this paper is organised as follows. In Section 2, we introduce some basic concepts and definitions needed when introducing and fitting the models in the subsequent sections. In Section 3, we investigate how the nucleolus locations deviate from complete spatial randomness. In Section 4, we also notice a deviation from a fitted degenerate PLCPP model. In Section 5, we introduce and fit various generalisations of the degenerate PLCPP model, including the final model briefly described in Sections 1.3.1 and 1.3.2. In particular, for estimation of parameters used in models for $X_z$ conditioned on $X_{xy}$, we propose in Section 5 a maximum pseudo-likelihood procedure, the performance of which is studied in a simulation study reported in the Appendix. Finally, Section 6 summarises our findings and discuss directions for future research.

2. Preliminaries

The point processes $X$, $X_{xy}$ and $X_z$ introduced above can be viewed as the restriction to the bounded sets $W$, $W_{xy}$ and $W_z$ of a locally finite point process on $\mathbb{R}^d$ with $d = 3, 2, 1$ respectively. Below we recall a few basic statistical tools needed in this paper, using the generic notation $Y$ for a locally finite point process defined on $\mathbb{R}^d$ (apart from the cases above, we have in mind that $Y$ could also be the centre process $\Phi$ from Section 1.3). Briefly speaking, this means that $Y$ is a random subset of $\mathbb{R}^d$ satisfying the condition that $Y_B = Y \cap B$ is finite for any bounded set $B \subset \mathbb{R}^d$; for a more rigorous definition of point processes, see, for example Daley & Vere-Jones (2003) or Møller & Waagepetersen (2004).

© 2021 Australian Statistical Publishing Association Inc.
2.1. Moments

For each integer $k \geq 1$, to describe the $k$th-order moment properties of $Y$, we consider the so-called $k$th-order intensity function $\lambda^{(k)}(\mathbb{R}^d)^k \to [0, \infty)$ given that it exists. This means that for any pairwise distinct and bounded Borel sets $B_1, \ldots, B_k \subset \mathbb{R}^d$,

$$E\left[n(Y_{B_1}) \cdots n(Y_{B_k})\right] = \int_{B_1} \cdots \int_{B_k} \lambda^{(k)}(x_1, \ldots, x_k)dx_1 \cdots dx_k$$

is finite, where $n(Y_B)$ denotes the cardinality of $Y_B$.

The first-order intensity function $\lambda^{(1)} = \lambda$ is of particular interest and is simply referred to as the intensity function. Heuristically, $\lambda(\cdot)du$ can be interpreted as the probability of observing a point from $Y$ in the infinitesimal ball of volume $du$ centred at $u$. If the intensity function $\lambda(\cdot) = \lambda$ is constant, then $\lambda[B] = E[n(Y_B)]$ for any bounded Borel set $B \subset \mathbb{R}^d$, where $|\cdot|$ is the Lebesgue measure. In this case, $Y$ is said to be homogeneous and otherwise inhomogeneous. Clearly, stationarity of $Y$ (meaning that its distribution is invariant under translations in $\mathbb{R}^d$) implies homogeneity.

2.2. Functional summaries

The functional summaries described in this section will be used both for model fitting as described in Section 2.3 and for model checking as described in Section 2.4.

To summarise the second-order moment properties, it is custom to consider the pair correlation function (PCF), $g$, which is defined as the ratio of the second- and first-order intensity function, that is

$$g(x_1, x_2) = \frac{\lambda^{(2)}(x_1, x_2)}{\lambda(x_1)\lambda(x_2)}, \quad x_1, x_2 \in \mathbb{R}^d.$$  

Heuristically, $\lambda(x_1)\lambda(x_2) g(x_1, x_2)$ is the probability of simultaneously observing a point from $X$ in each of the two infinitesimal balls of volume $dx_1$ and $dx_2$ centred at $x_1$ and $x_2$ respectively.

For a Poisson process $Y$, $g = 1$. The PCF is said to be stationary when (with abuse of notation) $g(x_1, x_2) = g(x_1 - x_2)$; this is the case when $Y$ is stationary.

If the PCF is stationary, it is closely related to the so-called second-order reduced moment measure, $K$, given by

$$K(B) = \int_B g(x)dx,$$

where $B \subset \mathbb{R}^d$ is a Borel set (see Möller & Waagepetersen 2004). If $Y$ is stationary, then $\lambda K(B)$ can be interpreted as the expected number of points in $Y \setminus \{o\}$ within $B$ given that $Y$ has a point at the origin $o$ of $\mathbb{R}^d$; when considering scalings of $B$, we refer to $B$ as a structuring element. The simplest example occurs when $B$ is a ball centred at the origin and with radius $r > 0$; then $K(r) = K(B)$ becomes the $K$-function introduced by Ripley (1976); and often we instead consider a transformation of the $K$-function, which is called the $L$-function and defined by $L(r) = (K(r)/\omega_d)^{1/d}$, where $\omega_d$ is the volume of the $d$-dimensional unit ball. In particular, if $Y$ is a stationary Poisson process, then $L(r) = r$.

For detecting cylindrical structures, Möller et al. (2016) introduced the cylindrical $K$-function which corresponds to $K(B)$ when $B$ is a cylinder of height $2r$, base-radius $r$, and centre of mass at the origin. Note that Ripley’s $K$-function depends only on one argument,
r, while the cylindrical \(K\)-function depends both on \(r, t\), and the direction of the cylinder. However, when \(d = 3\) and since the minicolumns are expected to extend along the \(z\)-axis, we only consider cylinders extending in this direction, effectively reducing the number of arguments to two.

We will also consider the commonly used \(F\)-, \(G\)- and \(J\)-function when performing model control; see Van Lieshout & Baddeley (1996) for definitions. Briefly, if \(Y\) is stationary, \(F(r)\) is the probability that \(Y\) has a point within distance \(r > 0\) from an arbitrary fixed location in \(\mathbb{R}^d\); \(G(r)\) is the probability that \(Y\) has another point within distance \(r > 0\) from an arbitrary fixed point in \(Y\); and \(J(r) = (1 - G(r))/(1 - F(r))\) when \(F(r) < 1\).

2.3. Model fitting

In Möller et al. (2016), parameter estimation for the degenerate PLCPP model was simply done by a moment based procedure which included minimisation of a certain contrast between a theoretical second-order moment functional summary and its empirical estimate. Below we describe a similar minimum contrast procedure for estimating the parameters of models for \(X_{xy}\). For the models of \(X_z\) conditioned on \(X_{xy}\), we find it convenient to use a maximum pseudo-likelihood procedure as detailed in Section 5.3.2.

Minimum contrast estimation is a computationally simple fitting procedure introduced by Diggle & Gratton (1984) that is applicable when a closed form expression of a functional summary, \(T\), exists. The idea is to minimise the distance from the theoretical function \(\hat{T}\) to its empirical estimate \(\hat{T}\) for the data. Specifically, if \(T\) depends on the parameter vector \(\theta\) and is a function of ‘distance’ \(r > 0\) (as, e.g. in case of Ripley’s \(K\)-function), the minimum contrast estimate of \(\theta\) is given by

\[
\hat{\theta} = \arg\min_{\theta} \int_{r_{\min}}^{r_{\max}} |T(\theta, r)^q - \hat{T}(r)^q| \, d\theta,
\]

where \(0 \leq r_{\min} < r_{\max}\), \(q > 0\), and \(p > 0\) are tuning parameters. General recommendations on \(q\) are given in Guan (2009) and Diggle (2014), when \(T(r) = g(r)\) or \(T(r) = K(r)\); when fitting a model to \(X_{xy}\), we let \(p = 2\), \(q = 1/4\), \(r_{\min} = 0\), and \(r_{\max}\) be one fourth of the shortest side length of \(W_{xy}\).

When the PCF has a closed form expression, alternative estimation procedures can be used, including the second-order composite likelihood (see Guan 2006; Waagepetersen 2007), adapted second-order composite likelihood (see Lavancier, Poinas & Waagepetersen 2018) and Palm likelihood (see Ogata & Katsura 1991; Prokešová, Dvořák & Jensen 2016; Baddeley, Rubak & Turner 2016).

2.4. GERL envelope procedure

For model checking, we consider informative global extreme rank length (GERL) envelope procedures (Myllymäki et al. 2017; Mrkvička et al. 2018) based on various functional summaries as described below.

In the GERL envelope procedure, the distribution of an empirical functional summary is estimated by simulations under a fitted model of interest. The procedure is a refinement of the global rank envelope procedure (Myllymäki et al. 2017), where it is recommended to use 2499 simulations for a single one-dimensional functional summary and at least 9999 simulations for a single two-dimensional functional summary (Mrkvička et al. 2016). However, we consider
a concatenation of the $L$, $G$, $F$ and $J$-function in which case Mrkvička, Myllymäki & Hahn (2017) recommend using more simulations. Particularly for a concatenation of $k$ one-dimensional summary functions they recommend using $k \times 2499$ simulations. For the GERL envelope procedure, Mrkvička et al. (2018) suggest that a lower number of simulations may be enough. Therefore, we use 9999 simulations. Since we consider a concatenation of one-dimensional functional summaries, we ensure that each of the functional summaries are weighted equally in the GERL envelope test by evaluating them at the same number of arguments (Mrkvička et al. 2017). Specifically, we consider $64 \times r$-values for each of the functions $L$, $G$, $F$ and $J$. The cylindrical $K$-function is not part of the concatenation, and it will be evaluated over a square grid consisting of $64 \times r$-values and $64 \times t$-values.

2.5. Summary

We remark that the summary function used for model fitting will not be used for checking goodness of fit using the GERL envelope procedure. In summary, we use:

- empirical estimates of the cylindrical $K$-function for investigating anisotropy and in particular columnarity in the 3D point patterns;
- when considering isotropic point process models for the two-dimensional projected locations (projections into the two-dimensional $xy$-plane), Ripley’s $K$-function for parameter estimation (both the theoretical $K$-function and its parametric estimate are used as explained in Section 2.3) and empirical estimates of the $G$, $F$- and $J$-function when checking for goodness of fit;
- when checking for goodness of fit for anisotropic point process models fitted to the 3D point patterns, empirical estimates of both the $L$, $G$, $F$- and $J$-function (which are not informative about anisotropy) and the cylindrical $K$-function (which is informative about anisotropy).

3. Complete spatial randomness

The most natural place to begin our point pattern analysis is by testing whether a homogeneous Poisson process $X$ with intensity $\lambda > 0$ (we then view $Y$ as a stationary Poisson process with the same intensity), also called complete spatial randomness (CSR), adequately describe each nucleolus point pattern data set. Recall that this means that $n(X)$ is Poisson distributed with parameter $\lambda |W|$ and conditional on $n(X)$ the points in $X$ are independent and uniformly distributed within $W$. We shall see that CSR is too simple a model for the description of $L3$ and $L5$, and that the noticed deviations from CSR become useful for suggesting new models.

The CSR model is fully specified by its intensity, which naturally is estimated by $n(X)/|W|$, which is equal to $2.37 \times 10^{-5}$ for $L3$ and $1.63 \times 10^{-5}$ for $L5$. For that fitted model, Figure 2 summarises the results of the GERL envelope procedure, based on the empirical cylindrical $K$-function with the cylindrical structuring element extending in the $x$-, $y$- and $z$-directions, along with the areas at which the empirical cylindrical $K$-function falls outside the GERL 95% envelope. It is clearly seen that CSR is rejected for both point patterns in all three directions. This is supported by the associated $p$-values of $10^{-4}$ for all three directions for $L3$, and $4 \times 10^{-4}$ for the $x$ and $y$-directions and $10^{-4}$ for the $z$-direction for $L5$. We notice that the empirical cylindrical $K$-function extending in the $z$-direction falls above the upper GERL envelopes for cylinders that have a height larger than approximately 35 $\mu$m for both
Figure 2. Results of the global extreme rank length (GERL) envelope procedure under complete spatial randomness based on the empirical cylindrical $K$-function (grey scale) with the cylindrical structuring element extending in the $x$-, $y$- and $z$-directions (top, middle and bottom respectively). The shaded horizontal/vertical lines indicate that the function falls above/below the 95% GERL envelope. The white line indicates the values for which the cylinder height is equal to the base diameter. Left: results for the data set $L3$. Right: results for the data set $L5$.

data sets, together with a base radius of approximately 5–15 $\mu$m for $L3$ and 5–20 $\mu$m for $L5$. This trend is not observed for the empirical cylindrical $K$-functions extending in the $x$- and $y$-directions. Furthermore, the observed cylindrical $K$-functions extending in the $z$-direction falls below the lower 95% GERL envelope for cylinders with a height of approximately 10–25 $\mu$m and a base radius larger than 5 $\mu$m. We further observe that the empirical cylindrical $K$-functions extending in the $x$- and $y$-directions falls below the lower 95% GERL envelope for cylinders with a height of approximately 0–60 $\mu$m and a base radius of approximately 5–25 $\mu$m. Hence, for elongated cylinders extending in the $z$-direction, we tend to see more points in the data than we expect under CSR, while for stunted cylinders we tend to see fewer points. Similarly, for cylinders that are neither very stunted nor very elongated in the $x$- and $y$-directions, we see fewer points than we expect to see under CSR. This seems to be in correspondence with columnar structures where the columns extend in the $z$-direction.

4. The degenerate Poisson line cluster point process

Møller et al. (2016) presented the so-called Poisson line cluster point process (PLCPP), which is useful for modelling columnar structures. Specifically, we consider a degenerate PLCPP $Y \subset \mathbb{R}^3$ constructed as follows.

(i) Generate a stationary Poisson process $\Phi = \{(\xi_i, \eta_i)\}_{i=1}^{\infty} \subset \mathbb{R}^2$ with finite intensity $\kappa > 0$.

Each point $(\xi_i, \eta_i) \in \Phi$ corresponds to an infinite line $l_i$ in $\mathbb{R}^3$, which is perpendicular to the $xy$-plane, that is, $l_i = \{(\xi_i, \eta_i, z) | z \in \mathbb{R}\}$. 

© 2021 Australian Statistical Publishing Association Inc.
(ii) Conditional on \( \Phi \), generate independent stationary Poisson processes \( L_1 \subset l_1, L_2 \subset l_2, \ldots \) with identical and finite intensity \( \kappa > 0 \).

(iii) Generate point processes \( X_1, X_2, \ldots \subset \mathbb{R}^3 \) by independently displacing the points of \( L_1, L_2, \ldots \) by the zero-mean isotropic normal distribution with standard deviation \( \sigma > 0 \).

(iv) Finally, set \( Y = \bigcup_{i=1}^{\infty} X_i \) and \( X = Y_W \).

Some comments to the construction in items 1–4 are in order.

In the general definition of the PLCPP in Møller et al. (2016), the lines \( l_1, l_2, \ldots \) are modelled as a stationary Poisson line process. That is, the lines are not required to be perpendicular to the \( xy \)-plane nor does the Poisson line process need to be degenerate (meaning that the lines are not required to be mutually parallel). Furthermore, the dispersion density (used in item 3) can be arbitrary. However, the construction is still such that \( Y \) becomes stationary. Furthermore, the same distribution of \( Y \) is obtained whether we consider a 3D zero-mean isotropic normal distribution for displacements in item 3 or a bivariate zero-mean isotropic normal distribution with displacements of the \( xy \)-coordinates for the points of \( L_1, L_2, \ldots \), provided the variances of the two normal distributions are identical, cf. Møller et al. (2016).

Returning to the degenerate PLCPP of items 1–4, we imagine that each \( X_i \) is a cylindrical cluster of points around the line \( l_i \), where these cylindrical clusters are parallel to the \( z \)-axis. Furthermore, the interpretation of the parameters \( \kappa, \alpha \) and \( \sigma \) in terms of a Poisson cluster point process is similar to that in Section 1.3.1 except that we now also consider lines not intersecting \( W \): if \( Y \) as defined by items 1–4 is restricted to a subset \( S \subset \mathbb{R}^3 \) bounded by two planes parallel to the \( xy \)-plane, for specificity \( S = \{(x, y, z) \in \mathbb{R}^3 | z \in W_z \} \), this restricted point process can be seen as a (modified) Thomas process (see Thomas 1949; Møller & Waagepetersen 2004) on \( \mathbb{R}^2 \) along with independent \( z \)-coordinates following a uniform distribution on \( W_z \).

To see this, first note that conditional on \( \Phi \) a Poisson process \( \{ \xi_i, \eta_i \}_{i=1}^{\infty} \) and for all \( i = 1, 2, \ldots, X_i \) is a Poisson process in \( \mathbb{R}^3 \) with intensity function \( \lambda_i((x, y, z)) = \alpha f(x - \xi_i, y - \eta_i) \), where \( f \) is the probability density function of the bivariate isotropic normal distribution given in item 3. In turn, this implies that \( Y \) conditioned on \( \Phi \) is a Poisson process in \( \mathbb{R}^3 \) with intensity function \( \sum_{i=1}^{\infty} \lambda_i((x, y, z)) \). Furthermore, since \( \lambda_i((x, y, z)) = \lambda_i((x, y)) \) does not depend on \( z \) for all \( i = 1, 2, \ldots, \) the projection of \( Y_S \) onto the \( xy \)-plane, \( P_{xy}(Y_S) \), conditioned on \( \Phi \) is a Poisson process with intensity \( \sum_{i=1}^{\infty} \lambda_i((x, y)) \), where \( a \) is the length of the interval \( W_z \). Since \( \Phi \) is a stationary Poisson process, \( P_{xy}(Y_S) \) is a Thomas process with centre process intensity \( \kappa \) and expected cluster size \( \alpha a \) (i.e. the expected number of points in \( X_i \cap S \)). Finally, from items 2 to 4, it follows that the \( z \)-coordinates of \( X_i \) are independent and uniformly distributed on \( W_z \), and they are independent of \( X_{xy} \).

Consequently, simulating \( X = Y_W \) is straightforwardly done by simulating a Thomas point process (on a larger set than \( W_{xy} \) in order to avoid boundary effects) along with independent uniform \( z \)-coordinates on \( W_z \). For simulating the Thomas point process we apply standard software from the \( R \)-package spatstat (Baddeley et al. 2016). Similarly, fitting a degenerate PLCPP to a realisation of \( X \) is simply a matter of fitting a Thomas process to the point pattern consisting of the \( xy \)-coordinates of the points in that realisation. Since the \( K \)-function of the Thomas process has a closed form expression, the model can easily be fitted using minimum contrast estimation with \( T(r) = K(r) \) in (1). Table 1 summarises the parameter estimates, where most notably the expected cluster size \( \tilde{a} \alpha \) is \( < 1 \) for both \( L3 \) and \( L5 \). Understanding each cylindrical cluster within \( W \) as (a part of) a minicolumn, ‘these
Table 1. Minimum contrast estimates of the degenerate Poisson line cluster point process (PLCPP).

|   | $\hat{\kappa}$ | $\hat{\sigma}$ | $\hat{\alpha}$ |
|---|-----------------|-----------------|-----------------|
| L3 | 0.027           | 2.86            | 0.36            |
| L5 | 0.0085          | 4.58            | 0.95            |

Figure 3. Empirical estimates of the cylindrical $K$-function (grey scale) where shaded horizontal/vertical lines indicate that the function falls above/below the 95% global extreme rank length (GERL) envelope under the fitted degenerate Poisson line cluster point process. The white line indicates the values for which the cylinder height is equal to the base diameter. Top: results for the data set L3. Bottom: results for the data set L5.

parameter estimates result in very unnatural models for the datasets, since each minicolumn within $W$ is expected to consist of less than one point’ (personal communication with Jens R. Nyengaard).

Despite the fact that the fitted degenerate PLCPP models are somewhat unnatural and hardly can be interpreted as a model with (the hypothesised) minicolumnar structures, GERL envelope procedure based on a concatenation of the $F$, $G$- and $J$-function shows that the Thomas process suitably fit the projected locations (projections into the two-dimensional $xy$-plane) with a $p$-value of 0.76 for L3 and 0.87 for L5. However, results from the concatenated GERL envelope procedure described in Section 2.4 indicated that the model did not suitably describe the 3D nucleolus locations with $p$-values of $10^{-4}$ and $10^{-4}$ for L3 and $10^{-4}$ and $6 \times 10^{-4}$ for L5 when using the GERL envelope procedure based on the concatenation of the one-dimensional summary functions ($L - r$, $G$, $F$, $J$) and the cylindrical $K$-function in the $z$-direction respectively. Specifically, Figure 3 shows the empirical cylindrical $K$-function and
indicates where it deviates from the 95% GERL envelope. Clearly, the model does account for some of the hypothesized columnarity of the data as opposed to CSR, but the empirical cylindrical $K$-function for $L_3$ still falls above the 95% GERL envelope. Furthermore, the empirical cylindrical $K$-function for both data sets falls below the 95% GERL envelope similar to what was seen under CSR, indicating a lack of regularity of the model, which in fact is supported by the one-dimensional functional summaries (not shown). This could suggest that the cylindrical clusters should be more distinct; motivating us to generalise the degenerate PLCPP model as in the following section.

5. A generalisation of the degenerate PLCPP

As some but not all features of the data were explained by the degenerate PLCPP fitted in Section 4, we propose in this section two generalisations as follows:

(i) The centre process $\Phi$ is a planar stationary point process.  
(ii) $X_c$ conditioned on $X_{xy}$ is a Markov random field.

The first modification is straightforward and although the Thomas point process suitably fitted the projected point patterns we found the parameter estimates unnatural for describing cylindrical clustering as discussed in Section 4. Therefore, we chose a repulsive centre process in order to obtain more distinguishable cylindrical clusters; this is detailed in Section 5.1. Furthermore, the assumption of stationarity of $\Phi$ is made in order to apply a similar minimum contrast estimation procedure as in Section 4, so implicitly we make the assumption that the PCF or the $K$-function is expressible in a closed form. For the second modification, we suggest a conditional model inspired by the multiscale point process and particularly the Strauss hard core point process (see, e.g. Møller & Waagepetersen 2004), which will allow for further repulsion or even aggregation between the points; this is detailed in Sections 5.2 and 5.3.

5.1. A determinantal point process model for the centre points

Consider a point process $Y \subset \mathbb{R}^3$ specified by items 1–4 in Section 4 with the exception that the centre process $\Phi$ now is an arbitrary stationary planar point process. Then, recalling the notation from Section 4, $P_{xy}(Y_S)$ is a planar Cox process (see Møller & Waagepetersen 2004) and even a planar generalised shot-noise Cox process (see Møller & Torrisi 2005) driven by the random intensity function $\Lambda(x,y) = a \sum_{i=1}^{\infty} \lambda_i(x,y)$, for $(x,y) \in \mathbb{R}^2$. Moreover, $P_{xy}(Y_S)$ corresponds to the Thomas process, but with a different centre point process (unless of course $\Phi$ is a stationary Poisson process).

In this section, we focus on the case where $\Phi$ is a stationary determinantal point process (DPP; see Lavancier et al. 2015), in which case we will refer to $Y$ as the determinantal line cluster point process (DLCPP). Let $C: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{C}$ be a function, then $\Phi$ is a DPP with kernel $C$ if its intensity functions satisfy

$$\lambda^{(k)}(u_1, \ldots, u_k) = \det[C](u_1, \ldots, u_k), \quad \text{for } k = 1, 2, \ldots, u_1, \ldots, u_k \in \mathbb{R}^2,$$

where $\det[C](u_1, \ldots, u_k)$ is the determinant of the $k \times k$ matrix with $(i,j)$th entry $C(u_i, u_j)$. For further details on DPPs, we refer to Lavancier et al. (2015) and the references therein. When $\Phi$ is a DPP, we call $P_{xy}(Y_S)$ a determinantal Thomas point process (DTPP). The DTPP is discussed to some extent in Møller & Christoffersen (2018), where a closed form expression
Table 2. For each data set L3 and L5, minimum contrast estimates for the parameters of our final model for $X_{xy}$ (the determinantal line cluster point process (DLCPP) model in Section 5.1).

|   | $\hat{\kappa}$ | $\hat{\sigma}$ | $\hat{\alpha}$ |
|---|----------------|----------------|----------------|
| L3 | 0.0040         | 5.45           | 2.42           |
| L5 | 0.0021         | 6.53           | 3.87           |

of its PCF is found. Thus, the DLCPP can be fitted by fitting a DTPP to the projected data using a minimum contrast procedure (see Section 2.3).

Specifically, we let $\Phi$ be the jinc-like DPP given by the kernel $C(u_1, u_2) = \sqrt{\kappa/\pi} J_1 \left(2\sqrt{\pi\kappa\|u_1 - u_2\|/\|u_1 - u_2\|}\right)/\|u_1 - u_2\|$, where $\kappa > 0$ is the intensity of $\Phi$, $J_1$ is the first-order Bessel function of the first kind, and $\|\cdot\|$ denotes the usual planar distance. In the sense of Lavancier et al. (2015), this is the most repulsive DPP with a stationary kernel (see also Biscio & Lavancier 2016). Simulation of the DTPP is done by first simulating $\Phi$ on a larger region than $W_{xy}$ in order to avoid boundary effects, for which we use the functionality of spatstat; second we generate for each cluster a Poisson distributed number of points with intensity $\alpha$; and third we displace these points by a bivariate zero-mean isotropic normal distribution.

The parameter estimates of the jinc-like DTPP model were obtained by minimum contrast with $T(r) = g(r)$. The parameter estimates are given in Table 2, where the estimated expected cluster size $\hat{\alpha}$ is ‘much smaller than expected for a minicolumn when restricting it to the observation window – provided the minicolumn hypothesis is true’ (personal communication with Jens R. Nyengaard). So we neither claim that we have a fitted model for minicolumns nor that the minicolumn hypothesis is true. Instead we have fitted a model with cylindrical clusters: from Table 2 we see, if $|W_{xy}|$ denotes the area of $W_{xy}$, the estimated number of projected clusters is $|W_{xy}|\hat{\kappa}$, which is approximately 260 for L3 and 142 for L5; the estimated expected size of a projected cluster is only 2.42 for L3 and 3.87 for L5.

Despite the expectation under the minicolumn hypothesis of having much higher values of $\hat{\alpha}$ than in Table 2, simulations of the fitted jinc-like DPP in the $xy$-plane seem in reasonable correspondence to the projected data; see Figure 4. Furthermore, results from the GERL envelope procedure based on a concatenation of one-dimensional functional summaries as well as the cylindrical $K$-function do not provide any evidence against the jinc-like DPP model for the projected points with $p$-values of 0.67 for L3 and 0.83 for L5.

Since the jinc-like DTPP model fits the projected data well, we proceeded and added independent uniform $z$-coordinates on $W_z$ to the simulations, thereby obtaining simulations of the jinc-like DLCPP. Figure 5 summarises the result of the 95% GERL test based on the concatenation of one-dimensional functional summaries as well as the cylindrical $K$-function as described in Section 2.2. These plots show that the models do not account for the regularity of the data, but accounts for more clustering compared to the degenerate PLCPP discussed in Section 4. This leads us to our next generalisation in Section 5.2 and the specific models in Section 5.3.

© 2021 Australian Statistical Publishing Association Inc.
Figure 4. Projection of observed nucleolus locations onto the $xy$-plane (left) and simulations from the fitted jinc-like determinantal Thomas point process (right) for the data sets $L3$ (top) and $L5$ (bottom).

Figure 5. Results of the global extreme rank length (GERL) envelope procedure under the fitted determinantal line cluster point process (DLCPP). Left: concatenation of the one-dimensional empirical $L$, $G$, $F$- and $J$-function for the data (solid line) together with 95% GERL envelopes (grey region), and theoretical function under CSR (dotted line); for ease of visualisation, the functions have been scaled. Right: empirical cylindrical $K$-function (grey scale) where shaded horizontal/vertical lines indicate that the function falls above/below the 95% GERL envelope. The white line indicates the values for which the cylinder height is equal to the base diameter. Top: results for the data set $L3$. Bottom: results for the data set $L5$.

5.2. General MRF models for the $z$-coordinates given the $xy$-coordinates

Motivated by the observations at the end of the previous section, in this section, we propose to model the vector of $z$-coordinates conditioned on the $xy$-coordinates by a pairwise interaction MRF given in (2) below.
In general, conditioned on $X_{xy} = \{ (x_i, y_i) \}_{i=1}^n$, we propose a model for $X_z$ with a conditional probability density function of the form

$$f((z_i)_{i=1}^n | (x_i, y_i)_{i=1}^n) \propto \gamma_1^{s_{B_{1}}(\{(z_i)_{i=1}^n | (x_i, y_i)_{i=1}^n\})} \cdot \gamma_2^{s_{B_{2}}(\{(z_i)_{i=1}^n | (x_i, y_i)_{i=1}^n\})} \times 1(||(x_i, y_i, z_i) - (x_j, y_j, z_j)|| > h, \text{ for } 1 \leq i < j \leq n),$$

with notation defined as follows. The right hand side in (2) is an unnormalised (conditional) density with respect to $n$-fold Lebesgue measure on $W_z$; $1(\cdot)$ denotes the indicator function; and $\gamma_1 > 0, \gamma_2 > 0$ and $h > 0$ are unknown parameters. When $h > 0$, it is a hard core parameter ensuring a minimum distance $h$ between all pair of points in $X$; for the pyramidal cell data, it seems natural to include a hard core condition since the cells cannot overlap. Clearly, in order that the conditional density is well-defined, $h$ needs to be smaller that the length of $W_z$ (which is much smaller than the diameter of a cell). Furthermore, for $k = 1, 2$,

$$s_{B_{k}}(\{(z_i)_{i=1}^n | (x_i, y_i)_{i=1}^n\}) = \sum_{1 \leq i < j \leq n} 1((x_i, y_i, z_i) \in B_k(x_j, y_j, z_j; \theta_k)),$$

where $B_k(x, y, z; \theta_k) \subset \mathbb{R}^3$ is an interaction region, with centre of mass $(x, y, z)$ and a ‘size and shape parameter’, $\theta_k$, that determines the interaction between points. It is additionally assumed that the hard core ball, given by the 3D closed ball of radius $h$ and centre $(x, y, z)$ contains neither $B_1(x, y, z; \theta_1)$ nor $B_2(x, y, z; \theta_2)$. Finally, it is assumed that the symmetry condition

$$(x_i, y_i, z_i) \in B_k(x_j, y_j, z_j; \theta_k), \quad \text{if and only if} \quad (x_j, y_j, z_j) \in B_k(x_i, y_i, z_i; \theta_k),$$

and the disjointness condition

$$B_1(x, y, z; \theta_1) \cap B_2(x, y, z; \theta_2) = \emptyset$$

are satisfied. The symmetry condition is imposed to ensure that we can interchange the roles of $i$ and $j$.

If $\gamma_1 = \gamma_2 = 1$ and $h = 0$, then $X_z$ becomes the homogeneous binomial point process which depends only on $X_{xy}$ through the number of points in $X_{xy}$. In general, using Markov random field (MRF) terminology, $X_z$ conditioned on $X_{xy} = \{ x_i, y_i \}_{i=1}^n$ is a pairwise interaction MRF with sites given by the lattice $\{ (x_i, y_i) \}_{i=1}^n$, where two sites $(x_i, y_i)$ and $(x_j, y_j)$ are neighbours if and only if there exist $z'_i, z'_j \in W_z$ and a $k \in \{1, 2\}$ such that $((x_i, y_i, z'_i) \in B_k(x_j, y_j, z'_j; \theta_k)$ or $||(x_i, y_i, z'_i) - (x_j, y_j, z'_j)|| \leq h$. In other words, the conditional distribution of $z_i$ (the $i$th coordinate in $X_z$) given both $\{ x_i, y_i \}_{i=1}^n$ and all $z_j$ with $j \neq i$ depends only on those $z_j$ where $(x_i, y_i)$ and $(x_j, y_j)$ are neighbours. We may also say that $(x_i, y_i, z_i)$ and $(x_j, y_j, z_j)$ interact if $(x_i, y_i, z_i)$ is in the union of $B_1(x_j, y_j, z_j; \theta_1)$ and $B_2(x_j, y_j, z_j; \theta_2)$ – we refer to this union of sets as the region of interaction of $(x_j, y_j, z_j)$ (note that we can interchange the roles of $i$ and $j$) or if $||(x_i, y_i, z_i) - (x_j, y_j, z_j)|| \leq h$ (i.e. the hard core condition is not satisfied, which happens with probability 0). The interaction can either cause repulsion/inhibition or attraction/clustering of the points in $X$ depending on whether $\gamma_k < 1$ or $\gamma_k > 1$ for $k = 1, 2$. Thus, apart from the hard core condition, the model allows for both repulsion and attraction but within different interaction regions $B_1$ and $B_2$.

Note that our hierarchical model construction yields a more flexible model for $X$ but we ignore edge effects in the sense that we have only specified a model for first $P_{xy}(Y_S)$ and second $X_z$ conditioned on $X_{xy} = P_{xy}(Y_S) \cap W_{xy}$, thereby ignoring a possible influence
of points in \( Y \backslash W \) when (2) is used in the latter step (unless it specifies a binomial point process). This simplification is just made for mathematical convenience; indeed, it would be interesting to construct a model taking edge effects into account so that \( Y \) becomes stationary, but we leave this challenging issue for future research.

### 5.3. Fitting specific MRF models for the \( z \)-coordinates given the \( xy \)-coordinates

Below we first specify the ingredients of the conditional probability density function given in (2) for various models and discuss the overall conclusions, next describe how to find parameter estimates, and finally, discuss how well the estimated models fit the data. Note that although we have not specified a stationary model for \( Y \), it still make sense to interpret plots of the empirical cylindrical \( K \)-function and of the \( \hat{F} \)-, \( \hat{G} \)-, \( \hat{J} \)- and \( \hat{L} \)-function, since we have stationarity in the \( xy \)-plane and approximately stationarity in the \( z \)-direction (as the density (2) is invariant under ‘translations of \((z_1, \ldots, z_n)\) within \( W_\circ \)).

#### 5.3.1. Selected models

In our search for a suitable model for the nucleolus locations, we considered many special cases of (2). Table 3 summarises five selected models, where \( b((x, y, z); r) \) is the ball with centre \((x, y, z)\) and radius \( r \), and where \( c((x, y, z); r, t) \) and \( d((x, y, z); r, t) \) denote the cylinder and double cone, respectively, with centre of mass at \((x, y, z)\), height \( 2t \), base radius \( r \) and extending in the \( z \)-direction. Note that in Table 3, we do not need to specify \( B_k(\cdot; \theta_k) \) when \( \gamma_k = 1 \). For the final model 5, (2) becomes the conditional density

\[
 f((z_i')_{i=1}^n | (x_i, y_i')_{i=1}^n) \propto \prod_{1 \leq i < j \leq n} I(\| (x_i, y_i, z_i) - (x_j, y_j, z_j) \| > h) \\
 \times \gamma_1 \prod_{1 \leq i < j \leq n} I(\| (x_i, y_i) - (x_j, y_j) \| \leq r_1, |z_i - z_j| \leq t_1) \\
 \times \gamma_2 \prod_{1 \leq i < j \leq n} I(\| (x_i, y_i) - (x_j, y_j) \| \leq r_2, t_1 < |z_i - z_j| \leq t_2),
\]

where the cylindrical interaction regions \( B_1(x_j, y_j, z_j; \theta_1) = c(x_j, y_j, z_j; r_1, t_1) \) and \( B_2(x_j, y_j, z_j; \theta_2) = c(x_j, y_j, z_j; r_2, t_2) \cap c(x_j, y_j, z_j; r_1, t_1) \) are illustrated in Figure 6. \( h \geq 0 \) is still the hard core parameter, \( \gamma_1 > 0 \) and \( \gamma_2 > 0 \) are interaction parameters, and \( 0 < r_2 \leq r_1 \) and \( 0 < t_1 < t_2 \) are parameters which determine the ‘range of interaction’ satisfying \( h < \sqrt{r_k^2 + r_k^2} \) for \( k = 1, 2 \).

| Model | \( \gamma_1 \) | \( \gamma_2 \) | \( B_1(\cdot; \theta_1) \) | \( B_2(\cdot; \theta_2) \) | \( \theta_1 \) |
|-------|-------------|-------------|----------------|----------------|-------------|
| 1     | 1           | 1           | -              | -              | -           |
| 2     | > 0         | 1           | \( b(\cdot; r) \) | -              | \( r > h \) |
| 3     | > 0         | 1           | \( c(\cdot; r, t) \) | -              | \( r, t > 0 \) |
| 4     | > 0         | 1           | \( c(\cdot; r, t) \) \& \( d(\cdot; r, t) \) | -              | \( r, t > 0 \) |
| 5     | > 0         | > 0         | \( c(\cdot; r_1, t_1) \) \& \( c(\cdot; r_2, t_2) \) | -              | \( r_1, t_1 > 0 \) |
Figure 6. Visualisation of the hard core region ball (in dark) and the cylindrical interaction regions $B_1$ (the cylinder) and $B_2$ (the union of the two elongated cylinders) used in our final model 5 for $L^3$. The relative dimensions of the objects correspond to the parameter estimates in Table 4.

These restrictions on the parameters are empirically motivated by use of functional summaries as detailed below.

First, we considered model 1, which is a hard core model if $h > 0$ and one of the simplest ways of modelling regularity; note that model 1 with $h = 0$ is the binomial point process with a uniform density as considered in Section 4. Though accounting for small distance repulsion, when fitted to the data, model 1 turned out not to account for the repulsion at larger scales. Second, we considered model 2 which is a conditional Strauss model with a hard core condition (see Møller & Waagepetersen 2004, and the references therein). For this model, the scale of repulsion for the $z$-coordinates seemed too great for points with similar $xy$-coordinates, and therefore we found it natural to replace the spherical interaction region with a cylinder, yielding model 3. However, model 3 did not correct the problem, and continuing with a single region of interaction we next suggested model 4 with a region given by a cylinder minus a double cone. Model 4 does to a smaller degree penalise the occurrence of points with similar $xy$-coordinates. However, this model was not suitable either. Models 1–4 were discarded by GERL tests with extremely small $p$-values. Finally, we considered model 5 which is a more flexible model that allows for both repulsion and aggregation within cylinder-shaped interaction regions, cf. the discussion in Section 1.3.2. For simplicity, all the models were also considered without a hard core condition, that is $h = 0$, but was in every case found inadequate.

© 2021 Australian Statistical Publishing Association Inc.
5.3.2. Results based on maximum pseudo-likelihood

The likelihood function corresponding to (2) involves a normalising constant which needs to be approximated by Markov chain Monte Carlo methods. We propose an easier alternative based on the pseudo-likelihood function (Besag 1975) defined as follows when the data is given by \{({x_i}, {y_i}, {z_i})\}_{i=1}^n \subset W. For \(i = 1, \ldots, n\), the \(i\)th full conditional density associated with (2) is

\[
f(z_i|({z_1}, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n), ({x_j}, {y_j})_{j=1}^n)
= \mathbf{1}(\|({x_i}, {y_i}, z_i) - ({x_j}, {y_j}, z_j)\| > h \text{ for } j \neq i) s_{s_{1,i}, s_{2,i}}^h / c_i,
\]

where we define

\[
s_{k,i} = \sum_{j\neq i} \mathbf{1}(({x_j}, {y_j}, z_j) \in B_k(({x_i}, {y_i}, z_i); \theta_k)), \quad k = 1, 2,
\]

and where the normalising constant is given by

\[
c_i = \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \gamma_1^k \gamma_2^l \int_W \mathbf{1}(\|({x_i}, {y_i}, z) - ({x_j}, {y_j}, z_j)\| > h \text{ for } j \neq i)
\times \mathbf{1} \left( \sum_{j\neq i} \mathbf{1}(({x_j}, {y_j}, z_j) \in B_1(({x_i}, {y_i}, z); \theta_1)) = k \right)
\times \mathbf{1} \left( \sum_{j\neq i} \mathbf{1}(({x_j}, {y_j}, z_j) \in B_2(({x_i}, {y_i}, z); \theta_2)) = l \right) dz.
\]

To estimate the model parameters, we maximise the log pseudo-likelihood given by

\[
LP(\gamma_1, \gamma_2, h, \theta_1, \theta_2) = \sum_{i=1}^n \log f(z_i|({z_1}, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n), ({x_j}, {y_j})_{j=1}^n).
\]

Clearly, by (3), the maximum pseudo-likelihood estimate (MPLE) \(\hat{h}\) of \(h\) is the minimum distance between any distinct pair of points \(({x_i}, {y_i}, z_i)\) and \(({x_j}, {y_j}, z_j)\) in the data. This in fact also corresponds to the maximum likelihood estimate. For \(h = \hat{h}\) and for fixed \(\theta_1\) and \(\theta_2\), we easily obtain the following. For each of models 2–4, the MPLE of \(\gamma_1\) exists if and only if \(s_{1,i} \neq 0\) for some \(i\), and then the log pseudo-likelihood function is strictly concave, with respect to \(\log \gamma_1\). For model 5, the MPLE of \((\gamma_1, \gamma_2)\) exists if and only if \(s_{1,i} \neq 0\) for some \(i\) and \(s_{2,j} \neq 0\) for some \(j\), and then the log pseudo-likelihood function is strictly concave with respect to \((\log \gamma_1, \log \gamma_2)\). Therefore, the (profile) log pseudo-likelihood can be maximised by a combination of a grid search over \(\theta_1\) and \(\theta_2\) and numerical optimisation with respect to \(\gamma_1\) and \(\gamma_2\).

Each of the five models in Table 3 were fitted to L3 and L5 by finding the maximum pseudo-likelihood estimate, where for the numerical optimisation we used optim (a general-purpose optimisation function from the R-package stats). Table 4 shows the maximum pseudo-likelihood estimates of model 5 for the two data sets. Most notably, around each point in \(X\), there is a repulsive interaction region (since \(\hat{\gamma}_1 < 1\)) within a stunted cylinder \(B_1\) and an attractive interaction region (since \(\hat{\gamma}_2 > 1\)) within the union \(B_2\) of two elongated...
Table 4. Pseudo-likelihood estimates of our final model (model 5 from Table 3 in Section 5.2) for the data sets L3 and L5.

|   | \( \hat{\gamma}_1 \) | \( \hat{\gamma}_2 \) | \( \hat{h} \) | \( \hat{r}_1 \) | \( \hat{r}_2 \) | \( \hat{r}_t \) |
|---|---|---|---|---|---|---|
| L3 | 0.41 | 1.78 | 6.25 | 20 | 11.5 | 11 | 35.5 |
| L5 | 0.51 | 1.68 | 6.77 | 24.25 | 15.5 | 14.75 | 37.25 |

Figure 7. Results of the GERL envelope procedure under the fitted model 5. Left: concatenation of the one-dimensional empirical \( L_-, G_-, F_-, \) and \( J_- \)-function for the data (solid line) together with 95% global extreme rank length (GERL) envelopes (grey region), and theoretical function under complete spatial randomness (dotted line); for ease of visualisation, the functions have been scaled. Right: empirical cylindrical \( K \)-function (grey scale) where shaded horizontal/vertical lines indicate that the function falls above/below the 95% GERL envelope. The white line indicates the values for which the cylinder height is equal to the base diameter. Top: results for the data set L3. Bottom: results for the data set L5.

cylinders, see again Figure 6. In particular, the fitted model is in accordance to the empirical findings as noted later when the cylindrical \( K \)-function extending in the \( z \)-direction of Figure 2 is discussed. Specifically, there is repulsion between two points \((x_i, y_i, z_i)\) and \((x_j, y_j, z_j)\) in \( X \), if \((x_i, y_i)\) and \((x_j, y_j)\) lie within distance \(20 \mu m\) for L3 and \(24.25 \mu m\) for L5 and if \(z_i\) and \(z_j\) lie within distance \(11.5 \mu m\) for L3 and \(15.5 \mu m\) for L5. Analogously, there is attraction between \((x_i, y_i, z_i)\) and \((x_j, y_j, z_j)\) if \((x_i, y_i)\) and \((x_j, y_j)\) lie within distance \(11 \mu m\) for L3 and \(14.75 \mu m\) for L5 and if \(|z_i - z_j|\) is between \(11.5–35.5 \mu m\) for L3 and \(15.5–37.25 \mu m\) for L5. Moreover, the estimated hard core \( \hat{h} \) is between 6 and 7 \( \mu m\), which is in accordance with ‘distance between the nucleolus and the membrane of a pyramidal cell’ (personal communication with Jens R. Nyengaard). Note that \(2\hat{h} \) (the diameter of an estimated hard core ball) is about half.
as small as $2\hat{r}_1$ (the estimated height of $B_1$). Finally, comparing Tables 2–4, we note that the two ‘clustering parameters’ $2\hat{\sigma}$ and $\hat{r}_3$ are of the same order.

Model checking was performed using GERL envelope procedures based on the concatenation of one-dimensional functional summaries as well as the cylindrical $K$-function as discussed in Section 2.4. For the fitted models, model 5 was the most appropriate with $p$-values of 0.15 and 0.32 for $L_3$ and 0.23 and 0.02 for $L_5$ when using the GERL envelope procedure based on the concatenation of the one-dimensional summary functions ($L - r, G, F, J$) and the cylindrical $K$-function in the $z$-direction respectively; the 95% GERL envelopes are visualised in Figure 7. Thus no evidence is seen against the fitted models summarised in Table 4 for $L_3$ while only slight evidence is present for $L_5$.

Finally, note that simulations from each of models 1–5 can straightforwardly be obtained using a Metropolis-Hastings algorithm for a fixed number of points and given a realisation of the $xy$-coordinates. Specifically, we used Algorithm 7.1 in Møller & Waagepetersen (2004) but with a systematic updating scheme cycling over the point indexes 1 to $n$, using a uniform proposal for a new point in $W_z$ and a Hastings ratio calculated from the full conditional (3). We successively updated each point 100 times under the systematic updating scheme, corresponding to 63400 and 54800 point updates for $L_3$ and $L_5$, respectively.

6. Concluding remarks

The structure of minicolumns (clusters) of points has been investigated in relation to illnesses such as Down’s syndrome (Buxhoeveden et al. 2002), schizophrenia (Casanova 2007), autism (Casanova et al. 2006) and Alzheimer’s disease (Esiri & Chance 2006; Chance et al. 2011). However, none of the methods used treat the positions of cells as a 3D point pattern. This paper has contributed to 3D point process methodology by developing new models and inference procedures based on a combination of moment-based estimation and maximum pseudolikelihood estimation, and by illustrating the usefulness of various 3D functional summaries, in particular the cylindrical $K$-function, in combination with global rank envelope test procedures.

We fitted several hierarchical models for the two 3D point pattern data sets of nucleolus locations and related our findings to the minicolumn hypothesis, which claims that minicolumns extend parallel to the $z$-axis. Starting with a homogeneous Poisson process and second with the degenerated Poisson line cluster point process model in Møller et al. (2016), we proceeded by developing more advanced point process models for a columnar structure along the $z$-axis. Our final model can be summarised as follows. For the $xy$-coordinates, we argued a need for a cluster point process given by a generalized shot noise Cox process, where the cluster centres are described by a repulsive point process given by a stationary determinantal point process, and where the offspring distribution is given by a bivariate isotropic normal distribution. For the $z$-coordinates conditioned on the $xy$-coordinates, we specified a pairwise interaction Markov random field model with repulsion between nucleolus locations given by a hard core condition on a small scale and a stunted cylindrical interaction region on a larger scale, as well as clustering between nucleolus locations given by an elongated cylindrical interaction region. The final model specifies much smaller columns than expected under the minicolumn hypothesis. Although the same type of model describes the two data sets, very different parameter estimates were obtained.

All the studies in Buxhoeveden et al. (2002), Casanova et al. (2006), Casanova (2007) and Chance et al. (2011) show deviations in the minicolumn structure in the brains of sick
MODELLING COLUMNARITY OF PYRAMIDAL CELLS

subjects compared to normal ones. For future applications with several 3D point pattern
data sets belonging to different groups (related to normal and sick objects), it remains to
develop statistical methods for comparison of the groups. We hope that our methodology
may serve as an inspiration for this and other future developments.

Appendix

A simulation study was performed to investigate the performance of the maximum
pseudo-likelihood estimation procedure used for fitting our final model (model 5 from Table
3) with a DTPP for the centre points in the xy-plane. The simulations were obtained by
(i) simulating 100 DTPPs with $\kappa=0.0040$, $\sigma=5.45$ and $\alpha=2.42$, corresponding to the
parameter estimates in Table 2 for the data set L3;
(ii) for each of the 100 simulated DTPPs, simulating the associated vector of z-coordinates
from the pairwise interaction MRF specified by the conditional density (2) and with
regions of interaction as specified for model 5 in Table 3. For this, we used a Metropolis–
Hastings algorithm as described in Section 5.2 with $\gamma_1=0.41$, $\gamma_2=1.78$, $h=6.25$,$r_1=20$, $t_1=11.5$, $r_2=11$ and $t_2=35.5$, corresponding to the parameter estimates in
Table 4 for the data set L3.

The results of the simulation study are presented in Table 5 and Figure 8. Despite the high
number of parameters, the estimates are distributed with most mass close to true parameter
values and with a small standard deviation. Not surprisingly, the hard core parameter $h$ is an

Table 5. Parameter values, their estimate and standard deviation from the simulation
study.

|       | $\gamma_1$ | $\gamma_2$ | $h$  | $r_1$ | $t_1$ | $r_2$ | $t_2$ |
|-------|------------|------------|------|-------|-------|-------|-------|
| True value | 0.41       | 1.78       | 6.25 | 20    | 11.5  | 11    | 35.5  |
| Mean     | 0.40       | 1.82       | 6.73 | 19.96 | 11.56 | 11.06 | 35.18 |
| Standard deviation | 0.048 | 0.081 | 0.447 | 0.608 | 0.292 | 0.618 | 1.357 |

Figure 8. Distribution of the parameter estimates obtained under the simulation study. The histograms
are in reading order associated with $\gamma_1, \gamma_2, h, r_1, t_1, r_2, t_2$. 

© 2021 Australian Statistical Publishing Association Inc.
exception that is biased with a right skewed distribution, since the simulated point patterns always have to satisfy the hard core condition, whilst for a given simulated point pattern the estimate of \( h \) is the shortest distance between any two points.

References

BADDELEY, A., RUBAK, E. & TURNER, R. (2016). Spatial Point Patterns: Methodology and Applications with R. New York: Chapman & Hall/CRC.

BESAG, J. (1975). Statistical analysis of non-lattice data. The Statistician 24, 179–195.

BISCIO, C.A.N. & LAVANCIER, F. (2016). Quantifying repulsiveness of determinantal point processes. Bernoulli 22, 2001–2028.

BUXHOEVEDEN, D., FOBBS, A., ROY, E. & CASANOVA, M. (2002). Quantitative comparison of radial cell columns in children with Down’s syndrome and controls. Journal of Intellectual Disability Research 46, 76–81.

BUXHOEVEDEN, D.P. & CASANOVA, M.F. (2002). The minicolumn hypothesis in neuroscience. Brain 125, 935–951.

CASANOVA, M.F. (2007). Schizophrenia seen as a deficit in the modulation of cortical minicolumns by monoaminergic systems. International Review of Psychiatry 19, 361–372.

CASANOVA, M.F., VAN KOOTEN, I.A.J., SWITALA, A.E., VAN ENGELAND, H., HEINSEN, H., STEINBUSCH, H.W.M., HOF, P.R., TRIPPE, J., STONE, J. & SCHMITZ, C. (2006). Minicolumnar abnormalities in autism. Acta Neuropathologica (Berl) 112, 287–303.

CHANCE, S.A., CLOVER, L., COUSIN, H., CURRAH, L., PETTINGILL, R. & ESIRI, M.M. (2011). Microanatomical correlates of cognitive ability and decline: Normal ageing, MCI, and Alzheimer’s disease. Cerebral Cortex 21, 1870–1878.

DALEY, D.J. & VERE-JONES, D. (2003). An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods 2nd edn. New York: Springer-Verlag.

DIGGLE, P.J. (2014). Statistical Analysis of Spatial and Spatio-temporal Point Patterns. Boca Raton, Florida: Chapman & Hall/CRC.

DIGGLE, P.J. & GRATTON, R.J. (1984). Monte Carlo methods of inference for implicit statistical models (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology) 46, 193–227.

ESIRI, M.M. & CHANCE, S.A. (2006). Vulnerability to Alzheimer’s pathology in neocortex: the roles of plasticity and columnar organization. Journal of Alzheimer’s Disease 9, 79–89.

GUAN, Y. (2006). A composite likelihood approach in fitting spatial point process models. Journal of the Americal Statistical Association 101, 1502–1512.

GUAN, Y. (2009). A minimum contrast estimation procedure for estimating the second-order parameters of inhomogeneous spatial point processes. Statistics and Its Interface 2, 91–99.

LAVANCIER, F., MØLLER, J. & RUBAK, E. (2015). Determinantal point process models and statistical inference. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77, 853–877.

LAVANCIER, F., POINAS, A. & WAAGEPETERSEN, R.P. (2018). Adaptive estimating function inference for non-stationary determinantal point processes. Available on arXiv:1806.06231.

MØLLER, J., CHRISTENSEN, H.S., CUEVAS-PACHECO, F. & CHRISTOFFERSEN, A.D. (2019). Structured space-sphere point processes and K-functions. Methodology and Computing in Applied Probability Available at https://doi.org/10.1007/s11009-019-09712-w.

MØLLER, J. & CHRISTOFFERSEN, A.D. (2018). Pair correlation functions and limiting distributions of iterated cluster point processes. Journal of Applied Probability 55, 789–809.

MØLLER, J., SAFAVIMANESH, F. & RASMUSSEN, J.G. (2016). The cylindrical K-function and Poisson line cluster point process. Biometrika 103, 937–954.

MØLLER, J. & TORRISSI, G.L. (2005). Generalised shot noise Cox processes. Advances in Applied Probability 37, 48–74.

MØLLER, J. & WAAGEPETERSEN, R.P. (2004). Statistical Inference and Simulation for Spatial Point Processes. Boca Raton, Florida: Chapman & Hall/CRC.

MRKVIČKA, T., MYLLYMÄKI, M. & HAHN, U. (2017). Multiple Monte Carlo testing, with applications in spatial point processes. Statistics and Computing 27, 1239–1255.

MRKVIČKA, T., MYLLYMÄKI, M., JILEK, M. & HAHN, U. (2018). A one-way ANOVA test for functional data with graphical interpretation. Available on arXiv:1612.03608.
Mrkvička, T., Soubeyrand, S., Myllymäki, M., Grabarnik, P. & Hahn, U. (2016). Monte Carlo testing in spatial statistics, with applications to spatial residuals. *Spatial Statistics* **18**, 40–53.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H. & Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**, 381–404.

Ogata, Y. & Katsura, K. (1991). Maximum likelihood estimates of the fractal dimension for random spatial patterns. *Biometrika* **78**, 463–474.

Prokešová, M., Dvořák, J. & Jensen, E.B.V. (2016). Two-step estimation procedures for inhomogeneous shot-noise Cox processes. *Annals of the Institute of Statistical Mathematics* **69**, 1–30.

Rafati, A.H., Safavimanesh, F., Dorph-Petersen, K.A., Rasmussen, J.G., Møller, J. & Nyengaard, J.R. (2016). Detection and spatial characterization of minicolumnarity in the human cerebral cortex. *Journal of Microscopy* **261**, 115–126.

Ripley, B.D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability* **13**, 255–266.

Thomas, M. (1949). A generalization of Poisson’s binomial limit for use in ecology. *Biometrika* **36**, 18–25.

Van Lieshout, M.N.M. & Baddeley, A.J. (1996). A nonparametric measure of spatial interaction in point patterns. *Statistica Neerlandica* **50**, 344–361.

Waagepetersen, R.P. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics* **63**, 252–258.