Differences in spatial point patterns with application to subcellular biological structures

Thomas R. Honnor, Julia A. Brettschneider, Adam M. Johansen

January 31, 2017

Abstract

In many real-world scenarios point pattern data may be obtained from two populations, with a desire to determine the extent and form of any difference between the point pattern structure. We have developed nonparametric statistics and a testing procedure free of distributional assumptions to quantify the significance of calculated differences between point pattern structures. The validity of the methodology is illustrated and its sensitivity investigated through application to simulated unmarked and marked point patterns. Analysis of a biological data set in which points represent microtubule locations within K-fibers is then carried out for control cells and cells for which TACC3 has been overexpressed. The results of our methodology when applied to this data indicate that TACC3 overexpression has a significant impact on K-fiber structure, shedding some light on the interplay between the structures and chemicals involved in mitosis and informing biologists future directions of study.

1 Introduction

Advances in sensor (Kanoun and Trankler, 2004) and storage technology (Grochowski and Hoyt, 1996) allow parallelisation of data collection across sensor networks and continuous monitoring. Improvements in communication networks have also made collected data more accessible. One result of this is the production of large, specialised spatial point pattern data sets, the analysis of which requires development of new statistical techniques. A particular area in which this is apparent is the imaging of large numbers of biological samples at high magnifications. The resulting images may be analysed computationally to determine the location of subcellular structures of interest. Further investigation can shed light on the inner workings of the cell and the effect of applied external conditions. This report introduces and applies a methodology for comparing the structure of point patterns with a particular biological application in mind, the structure of microtubules within kinetochore fibers, but with further applicability to more general data sets. Analysis of microtubule structure is of particular importance as microtubules perform a vital role during chromosome separation in mitosis, where errors can lead to aneuploidy, a common cause of genetic disorders.

Point pattern data comprising observations from two populations may arise in numerous ways. Plant locations, divided into two populations based upon the species of plant (Mateu et al., 2014). The location of neurons within the brain, divided based upon whether the individual suffers from mental illness (Diggle et al., 1991). The locations of a particular subcellular structure within

Paper No. 17-01, www.warwick.ac.uk/go/crism

multiple cells, divided by control cells and those which have received a treatment (Nixon et al., 2015).

It may be desirable to determine if there is variation in the structure of point patterns to make inference on underlying differences between the two populations. Such variation may occur consistently, but at a small enough scale to make detection by eye impossible. This report describes a statistical methodology for application to point patterns and a class of marked point patterns, to test for the existence of structural differences between two collections of point patterns.

There exists a large literature on spatial point processes. A basis for the statistical modelling of point processes is generally provided by Poisson point processes (Kingman, 1993) for which there is no interaction between point locations. Further generalisations to Cox processes (Cox, 1955) and Neyman-Scott processes (Neyman and Scott, 1958) account for interaction between points through the specification of Poisson point processes with random intensity. Alternatively, pairwise interaction point processes (Ripley, 1977) and more general Markov point processes (van Lieshout, 2000) account for interaction between points through the specification of the likelihood.

One modelling approach is to model each of the populations individually and compare the models. Due to the wide variety of models and difficulties fitting them to data, we instead compare the collections of point patterns directly using a number of nonparametric summary statistics, which are then combined across and compared between collections to produce test statistics. Nonparametric permutation testing is then be used to quantify the significance of reported test statistic values.

We introduce a variety of test statistics, such as the number of points and the distances between points, and a number of comparison methods, for example pointwise and functional comparisons. Decisions on which of the suggested testing procedures are most suitable will depend upon the property of interest, the required sensitivity of the testing procedure and the desired interpretability of the test statistic. The performance of each of the test statistics is quantified through application to a number of simulated data sets. Following convincing application to simulated data we apply the methodology to a real biological data set for which the populations are control cells and cells modified to overexpress TACC3. The results of this investigation could have an impact on the understanding of mitosis, influencing the direction of further biological research.

This report begins with an introduction to the problem and methodology. A concise background on spatial point patterns and permutation testing is then given in Section 2, after which the data is introduced and methodology is described in detail in Section 3. The following section describes the formulation and results of a simulation study of the effectiveness of the proposed methodology. After successful application to simulations a set of biological data is analysed in Section 5, including sensitivity analysis, before conclusions are presented.

2 Statistical background

2.1 Point processes

A summary of the evolution of the study of point processes including state of the art approaches may be found in works by Møller (2003), Gaetan and Guyon (2009) and Diggle (2013). This section defines some of the terms used in reference to point processes and provides a description

of the Poisson point process, the foundational tractable point process model. Point processes are denoted by underlined capital letters $\underline{X}, \underline{Y}, \ldots$, point patterns by underlined lower case letters $\underline{x}, \underline{y}, \ldots$, while letters $x_j, x_k, \ldots$ are used to denote points in patterns and letters $x, y, \ldots$ are used to denote points in the general space $\mathbb{R}^d$.

For a point process $\underline{X}$ on the space $\mathbb{R}^d$ the count function is the random variable given by $N(B) = n(\underline{X}_B)$, equal to the number of points of the pattern $\underline{X}$ in the set B. The set B is a member of the Borel σ-algebra on $\mathbb{R}^d$, $\mathcal{B}(\mathbb{R}^d)$, and the notation $\underline{X}_B = \underline{X} \cap B$ is used. The first order properties of the count function are described by the intensity measure μ, the expected value of the count function.

$$\mu(B) = \mathbb{E}[N(B)], \quad \forall B \in \mathcal{B}(\mathbb{R}^d).$$

If the intensity measure μ may be written as

$$\mu(B) = \int_{x \in B} \rho(x) dx, \quad \forall B \in \mathcal{B}(\mathbb{R}^d),$$

for some function $\rho : \mathbb{R}^d \to [0, \infty)$, then ρ is referred to as the intensity function of the point process $\underline{X}$.

For f a density function on the set $B \subseteq \mathbb{R}^d$ and $n \in \mathbb{N} = \{1, 2, 3, \ldots\}$, the point process consisting of n i.i.d. points with density f is referred to as a binomial point process and denoted by $\underline{X} \sim \text{binomial}(B, n, f)$. A point process $\underline{X}$ on $\mathbb{R}^d$ is defined to be a Poisson point process with intensity function $\rho(x)$, denoted by $\underline{X} \sim \text{Poisson}(\mathbb{R}^d, \rho)$, if the following properties are satisfied (Kingman, 1993):

1. For any $B \subseteq \mathbb{R}^d$ with $\mu(B) = \mathbb{E}(N(B)) < \infty$, $N(B) \sim \text{po}(\mu(B))$, the Poisson distribution with mean $\mu(B)$.

2. For any $n \in \mathbb{N}$ and $B \subseteq \mathbb{R}^d$ with $0 < \mu(B) < \infty$, conditional on $N(B) = n$, $\underline{X}_B \sim \text{binomial}(B, n, f)$ with $f(x) = \rho(x)/\mu(B)$.

Poisson point processes therefore have no interaction between point locations, as $N(B_1)$ and $N(B_2)$ are independent for disjoint sets B_1 and B_2.

A point process $\underline{X}$ on $\mathbb{R}^d$ is defined to be stationary if its distribution is invariant under translations, that is the distribution of $\underline{X} + y = \{x_j + y : x_j \in \underline{X}\}$ must be the same as that of $\underline{X}$ for any $y \in \mathbb{R}^d$. The Poisson point process with intensity function $\rho(x)$ is therefore stationary, or homogeneous, if $\rho(x) \equiv \rho \in [0, \infty)$ a constant for all $x \in \mathbb{R}^d$.

2.2 Permutation testing

A null hypothesis H_0 is typically tested in the presence of data x, in the context of this report a collection of observed point patterns, using a test statistic $t(x)$. In the case where the distribution of the corresponding random statistic $T(X)$ under H_0 is known analytically, the value of $t(x)$ may be compared to this distribution to obtain a p-value, the probability of observing values of the test statistic as or more extreme than that calculated for the observed data. Alternatively, the distribution of $T(X)$ under H_0 may not be known analytically. In such cases, if under H_0 there exists exchangeability of X under a set of operations $\Gamma = \{\gamma_0, \gamma_1, \ldots, \gamma_m\}$, that is the joint

distribution of γX is identical to that of X for all $\gamma \in \Gamma$, then H_0 may be tested using what is known as a permutation test.

Denoting by $\gamma_0 \in \Gamma$ the identity operation, that is $\gamma_0 X = X$, an exact permutation test (Edgington, 1964) requires calculation of the test statistic under all exchangeability operations $\gamma_i \in \Gamma$, producing $\{t(\gamma_0 x), t(\gamma_1 x), \ldots, t(\gamma_m x)\}$. The reported p-value in the case of a two-sided test of H_0 is then given by

$$p = \frac{1}{m+1} \sum_{\gamma \in \Gamma} \mathbb{1}[|t(\gamma x)| \geq |t(\gamma_0 x)|],$$

the proportion of the calculated $t(\gamma_i x)$ under exchangeability determined by H_0 which are as or more extreme than the observed value of the test statistic.

In the case where it is not feasible to calculate every test statistic $t(\gamma x)$ under H_0, potentially because calculation of $t(\gamma x)$ is computationally expensive or the number of possibilities $m+1$ is simply too large, the null hypothesis may be tested using what is known as an approximate permutation test (Edgington, 1969). Of the $m+1$ possible exchangeability operations, a simple random sample Γ', including γ_0, of size $m' < m+1$ is taken from Γ. The reported approximate p-value in the case of a two-sided test of H_0 is then given by

$$p = \frac{1}{m'} \sum_{\gamma \in \Gamma'} \mathbb{1}[|t(\gamma x)| \geq |t(\gamma_0 x)|],$$

the proportion of test statistics under exchangeability determined by H_0 which are as or more extreme than the observed value of the test statistic.

For both exact and approximate permutation tests the null hypothesis is then rejected at significance level α if $p < \alpha$. Both the exact and approximate permutation tests are valid as exchangeability under permutations γ under the null hypothesis implies that the $t(\gamma X)$ are identically distributed. Thus, the probability that the observed value $t(\gamma_0 x)$ lies in the most extreme α proportion of the set of all considered $t(\gamma x)$, i.e. $p < \alpha$, is α. In the case of the approximate permutation test power increases with the size of the random sample, m'. When testing at the $\alpha = 0.05$ significance level with $m' = 999$ the power of the approximate permutation test is at least 94.5 percent of the exact test (Jöckel, 1986), rising to at least 98.3 percent when $m' = 9999$.

3 Quantification of differences in point pattern structure

3.1 Data description

3.1.1 2D point patterns

The methodology in the following section is designed for application to point patterns $\underline{x} \in \chi_2$, where χ_2 is the set of all finite point patterns on $\mathbb{R}^2$. That is $\underline{x} = (x_1, x_2, \ldots, x_{n(\underline{x})})$ with $x_j \in \mathbb{R}^2$ and $n(\underline{x}) \in \mathbb{N}$. The point pattern $\underline{x}$ is represented as an ordered collection of point locations for identifiability purposes. To summarise,

$$\chi_2 := \{(x_1, x_2, \ldots, x_{n(\underline{x})}) : n(\underline{x}) \in \mathbb{N}, x_i \in \mathbb{R}^2 \text{ for } i = 1, 2, \ldots, n\}.$$

We will consider a collection of point patterns $\underline{x}^i \in \chi_2$ indexed by the set of $i \in I$. The notation $\underline{x}^J = (\underline{x}^i : i \in J)$ is used to represent the collection of point patterns indexed by the set J, where

ordering is again simply for identifiability. Two collections of point patterns, indexed by I_0 and I_1 and denoted by $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$, are obtained by partitioning the set I. Under this notation, x_j^i is the location of the j^{th} point in the i^{th} point pattern, $\underline{x}^i$.

3.1.2 Marked point patterns

Additional methodologies are also designed for application to marked point patterns $\underline{y} \in \chi_2^+$. The point space χ_2^+ is the set of all finite point patterns on $\mathbb{R}^2$ augmented with a third coordinate equal to zero, with each point x_j assigned a corresponding mark v_j which is itself a vector in $\mathbb{R}^2$ augmented with a third coordinate equal to h. That is $\underline{y} = ((x_1, v_1), (x_2, v_2), \ldots, (x_{n(\underline{y})}, v_{n(\underline{y})}))$ with $x_j \in \mathbb{R}^2 \times \{0\}$ and $v_j \in \mathbb{R}^2 \times \{h\}$. Again, the points of $\underline{y}$ are ordered for identifiability purposes. The reason for the particular formulation of χ_2^+ will become clear in light of the particular application described in Section 5, but in brief is chosen to represent paired point patterns in parallel planes separated by a distance of h by a marked point pattern. To summarise,

$$\chi_2^+ := \{((x_1, v_1),(x_2, v_2), \ldots, (x_{n(\underline{y})}, v_{n(\underline{y})})) : n(\underline{y}) \in \mathbb{N}, x_i \in \mathbb{R}^2 \times \{0\} \\ \text{and } v_i \in \mathbb{R}^2 \times \{h\} \text{ for } i = 1, 2, \ldots, n(\underline{y})\}.$$

We again consider a collection of marked point patterns $\underline{y}^i \in \chi_2^+$ indexed by the set I, with $\underline{y}^J$ the collection of marked point patterns indexed by the set J. The sets of indices I_0 and I_1 are defined analogously to those for the unmarked point patterns.

3.2 Statistical problem

Point patterns $\underline{x}^{I_0}$ may be considered to be independent realisations of a point process $\underline{X}^0$ and point patterns $\underline{x}^{I_1}$ to be independent realisations of a point process $\underline{X}^1$. The aim of the proposed methodology for point patterns is to make inference on the existence and form of a difference between the processes $\underline{X}^0$ and $\underline{X}^1$ using the available data $\underline{x}^I$.

Similarly, marked point patterns $\underline{y}^{I_0}$ may be considered to be independent realisations of a marked point process $\underline{Y}^0$ and marked point patterns $\underline{y}^{I_1}$ to be independent realisations of a marked point process $\underline{Y}^1$. The aim of the proposed methodology for marked point patterns is to make inference on the existence and form of a difference between the processes $\underline{Y}^0$ and $\underline{Y}^1$ using the available data $\underline{y}^I$.

3.3 Summary statistics

Summary statistics provide information on the distribution of observed points and may be compared between observations. Comparison may also be made to the homogeneous Poisson process for which summary statistics are sometimes tractable. Summary statistics for more advanced point process models are typically intractable. This section introduces a number of summary statistics, each of which summarises a property of a single (marked) point pattern, with the following section combining these values over the index sets I_0 and I_1 to produce test statistics.

In applications, point locations may be reported relative to an origin which is not consistent between observations. Under such circumstances, comparison of point locations between observations is uninformative due to this inconsistency. As such, we focus on point location comparison within samples through summary statistics which for example consider the distances between points.

For $\underline{X}$ a homogeneous point process on the observation window W, the intensity is given by

$$\rho = \frac{\mu(W)}{|W|},$$

where $|W|$ denotes the volume of the observation window W. For $\underline{x}$ a realisation of a homogeneous point process, obtained by first fixing the observation window W and then reporting all point locations within W, the intensity may be estimated by

$$\hat{\rho} = \frac{n(\underline{x})}{|W|}. \tag{1}$$

The nearest neighbour of point x_j in pattern $\underline{x}$ may be denoted by $\mathrm{nn}(x_j)$

$$\mathrm{nn}(x_j) = \left\{ x_k : k = \arg\min_l ||x_l - x_j|| \right\},$$

where the set definition accounts for the possibile existence of more than one nearest neighbour. The nearest neighbour distance of point x_j may then be denoted by $\mathrm{nnd}(x_j)$ with

$$\mathrm{nnd}(x_j) = \inf_{x \in \mathrm{nn}(x_j)} \{||x_j - x||\},$$

the unique infimum over $\mathrm{nn}(x_j)$ which gives the Euclidean distance between x_j and its nearest neighbours in $\underline{x}$. The mean nearest neighbour distance for the point process $\underline{X}$ is then

$$\overline{\mathrm{nnd}}(\underline{X}) = \mathbb{E}[\mathrm{nnd}(x), x \sim \mathrm{Uniform}\{\underline{X}\}],$$

an estimator of which for the point pattern $\underline{x}$ is given by

$$\overline{\mathrm{nnd}}(\underline{x}) = \frac{1}{n(\underline{x})} \sum_{j=1}^{n(\underline{x})} \mathrm{nnd}(x_j). \tag{2}$$

The graph produced by including only those edges which represent a nearest neighbour relationship is unlikely to be connected. As a result, the disjoint connected subgraphs may be translated to produce clearly different point patterns which retain the same nearest neighbour properties. The minimum spanning tree is a connected alternative to the graph of nearest neighbours for which the translation of subgraphs typically results in a difference in a corresponding mean minimum spanning distance.

Considering the weighted graph associated with the point pattern $\underline{x}$ denoted by

$$\mathcal{G}(\underline{x}) = (V(\underline{x}), E(\underline{x}), D(\underline{x})),$$

the vertex set of $\mathcal{G}(\underline{x})$, denoted by $V(\underline{x})$, is the set of point locations $x_j \in \underline{x}$. The graph $\mathcal{G}(\underline{x})$ is the complete graph on $V(\underline{x})$, meaning that the edge set $E(\underline{x})$ is the set of edges joining every vertex to every other vertex. Graph edges are weighted by the distances between points, with the edge joining the vertices at x_j and x_k being attached the weight $||x_j - x_k||$, producing the set of edge weights $D(\underline{x})$. The minimum spanning tree of the graph $\mathcal{G}(\underline{x})$, denoted by $\mathcal{G}^*(\underline{x})$, is the spanning subgraph of $\mathcal{G}(\underline{x})$ with minimum weight. Let $E^*(\underline{x}) \subset E(\underline{x})$ denote the edge set of $\mathcal{G}^*(\underline{x})$ and $D^*(\underline{x}) \subset D(\underline{x})$ denote the corresponding set of edge weights. The expected edge length of the minimum spanning tree of the point process $\underline{X}$, referred to as the mean minimum spanning distance $\overline{\mathrm{msd}}(\underline{X})$, is then

$$\overline{\mathrm{msd}}(\underline{X}) = \mathbb{E}[d, d \sim \mathrm{Uniform}\{\mathcal{D}^*(\underline{X})\}].$$

An estimate of the mean minimum spanning distance for a point pattern $\underline{x}$ is given by

$$\overline{\text{msd}}(\underline{x}) = \frac{1}{n(\underline{x}) - 1} \sum_{d \in \mathcal{D}^*(\underline{x})} d, \tag{3}$$

where the divisor is $n(\underline{x}) - 1$ as the minimum spanning tree of n points is made up of $n - 1$ edges.

The K-function (Ripley, 1977) for a stationary point process $\underline{X}$ on the space $S \subseteq \mathbb{R}^d$ is given by

$$K(r) = \frac{1}{\rho} \mathbb{E} \left[\frac{1}{N(S)} \sum_{x_j \neq x_k \in \underline{X}} \mathbb{1}[||x_j - x_k|| < r] \right]. \tag{4}$$

Under this formulation, $\rho K(r)$ is the expected number of further points within a distance of r from a randomly selected point in $\underline{X}$ (Diggle, 2013). For ease of future reference we therefore refer to the K-function as the scaled neighbourhood count function. An estimate of the scaled neighbourhood count function for point pattern $\underline{x}$ evaluated at distance r, denoted by $\hat{K}(\underline{x}, r)$, is given by

$$\hat{K}(\underline{x}, r) = \frac{|W|}{n(\underline{x})^2} \sum_{j \neq k} e_{j,k} \mathbb{1}(||x_j - x_k|| \leq r), \tag{5}$$

where $e_{j,k}$ is an edge correction term such that $e_{j,k}$ is the proportion of the circumference of the circle with centre x_j and radius $||x_j - x_k||$ which is contained within the observation window W. For $\underline{X}$ a homogeneous Poisson point process on $S \subseteq \mathbb{R}^2$ the scaled neighbourhood count function is given by $K(r) = \pi r^2$. Values of $K(r) > \pi r^2$ are evidence for aggregation of points within $\underline{X}$ at distances of less than r. Values of $K(r) < \pi r^2$ are evidence of regularity of points within $\underline{X}$ at distances of less than r.

The nearest neighbour function (Diggle, 2003), G, for a stationary point process $\underline{X}$ on the space $S \subseteq \mathbb{R}^d$ with intensity ρ is given by

$$G(r) = \frac{1}{\rho |B|} \mathbb{E} \left[\sum_{x \in \underline{X}_B} \mathbb{1}[(\underline{X} \setminus x) \cap b(x, r) \neq 0] \right], \quad r > 0,$$

for an arbitrary set B in the Borel σ-algebra of $\mathbb{R}^d$ with $0 < |B| < \infty$ and $b(x, r)$ the d-dimensional ball centred at x with radius r. As $\underline{X}$ is assumed to be stationary, the nearest neighbour function is independent of the chosen set B. The nearest neighbour function $G(r)$ is the distribution function of the distance from a randomly selected point in $\underline{X}$ to its nearest neighbour in $\underline{X}$. An uncorrected estimate of the nearest neighbour function for point pattern $\underline{x}$ evaluated at distance r, denoted by $\hat{G}(\underline{x}, r)$, is given by

$$\hat{G}(\underline{x}, r) = \frac{1}{n(\underline{x})} \sum_{j=1}^{n(\underline{x})} \mathbb{1}\left[\text{nnd}(x_j) \leq r\right], \quad r \in [0, \infty). \tag{6}$$

For $\underline{X}$ a homogeneous Poisson point process on $\mathbb{R}^2$ with constant intensity ρ the nearest neighbour function is given by $G(r) = 1 - \exp(-\rho \pi r^2)$.

Edge corrections to estimates of $K(r)$ and $G(r)$ are often necessary to account for unobserved points of $\underline{X}$ in the space $\mathbb{R}^d \setminus W$. Possible edge correction procedures include the use of an edge correction factor, minus sampling and the utilisation of Kaplan-Meier estimators (Møller, 2003).

The scaled neighbourhood count function, $K(r)$, and the nearest neighbour function, $G(r)$, are cumulative functions which summarise information across distances of less than or equal to r. Care

must therefore be taken when making inference upon values of $K(r)$ or $G(r)$ at a single separation distance r. In cases where the point pattern is non-stationary, deviation from the theoretical values proposed in this section may be caused by inhomogeneous intensity rather than specific interaction between points. Furthermore, similarity between K-functions for observed point processes is no guarantee that they are identical, as very different point process models may share the same K-function (Baddeley and Silverman, 1984).

The location of points within marked point patterns may be summarised by the previous statistics described in this section. A further summary statistic for the marked point pattern $\underline{y}$ may be formulated using only the marks $v_j \in \mathbb{R}^2 \times \{h\}$, each of which represents the vector connecting related points in paired point patterns lying in parallel planes separated by a distance of h. Let $\hat{v}_j$ denote the unit vector in the direction of the mark v_j and β be given by

$$\beta = \sum_{j=1}^{n(\underline{y})} \hat{v}_j. \tag{7}$$

We then propose the effective force transference summary statistic for the marked point process $\underline{Y}$ as

$$EFT(\underline{Y}) = \mathbb{E}[\cos(\theta(v, \beta)), (x, v) \sim \text{Uniform}\{\underline{Y}\}],$$

where $\theta(v, \beta)$ is the angle between vectors v and β. An estimate of the effective force transference summary statistic for a marked point pattern $\underline{y}$ is given by

$$EFT(\underline{y}) = \frac{1}{n(\underline{y})} \sum_{j=1}^{n(\underline{y})} \cos(\theta(v_j, \beta)).$$

This approximation may alternatively be expressed as

$$\begin{aligned} EFT(\underline{y}) &= \frac{1}{n(\underline{y})} \sum_{j=1}^{n(\underline{y})} \frac{\hat{v}_j \cdot \beta}{||\beta||} \\ &= \frac{||\beta||}{n(\underline{y})}. \end{aligned} \tag{8}$$

A further expression of the effective force transference estimator is given by

$$\begin{aligned} EFT(\underline{y}) &= \frac{1}{n(\underline{y})} \sqrt{\beta \cdot \beta} \\ &= \frac{1}{n(\underline{y})} \sqrt{\sum_{j=1}^{n(\underline{y})} \hat{v}_j \cdot \sum_{k=1}^{n(\underline{y})} \hat{v}_k} \\ &= \frac{1}{n(\underline{y})} \sqrt{\sum_{j=1}^{n(\underline{y})} \sum_{k=1}^{n(\underline{y})} \cos(\theta(v_j, v_k))}. \end{aligned} \tag{9}$$

From this formulation it is clear that the effective force transference is dependent only upon the angles between mark vectors, independently of point locations and the length of mark vectors.

As β is the sum of $n(\underline{y})$ unit length vectors, $0 \leq ||\beta|| \leq n(\underline{y})$ and $EFT(\underline{y}) \in [0, 1]$. The expression of effective force transference given in (9) makes it clear that $EFT(\underline{y}) = 1$ if and only if

all mark vectors are parallel and the greater the angles between lines, the smaller the value of the effective force transference. As a result $EFT(\underline{y})$ is a measure of the degree to which mark vectors v_j are similarly oriented. Further to the general interpretation of the effective force transference test statistic for any $\underline{y} \in \chi_2^+$, Section 5.4 introduces a specific interpretation in the case of the TACC3 data which motivates the name.

3.4 Test statistics

The first order summary statistic, $n(\underline{x})$, may be compared between point pattern collections using the pattern size test statistic

$$\delta_N(I) = \frac{1}{|I_0|}\sum_{i\in I_0} n(\underline{x}^i) - \frac{1}{|I_1|}\sum_{i\in I_1} n(\underline{x}^i).$$

Similarly, when the area of the observation window of point patterns, $|W|$, is random, it may be compared between point pattern collections using the pattern area test statistic

$$\delta_W(I) = \frac{1}{|I_0|}\sum_{i\in I_0} |W^i| - \frac{1}{|I_1|}\sum_{i\in I_1} |W^i|.$$

The estimated homogeneous intensity of point patterns, $\hat{\rho}(\underline{x})$, defined in (1), may be compared between collections using the intensity test statistic.

$$\delta_\rho(I) = \sum_{i\in I_0} \omega_0(\underline{x}^i)\hat{\rho}(\underline{x}^i) - \sum_{i\in I_1} \omega_1(\underline{x}^i)\hat{\rho}(\underline{x}^i),$$

where weights ω_0 and ω_1 may be formulated to produce unweighted test statistics with

$$\omega_0(\underline{x}^i) = \frac{1}{|I_0|}, \qquad \omega_1(\underline{x}^i) = \frac{1}{|I_1|},$$

or weighted test statistics with

$$\omega_0(\underline{x}^i) = \frac{n(\underline{x}^i)}{\sum_{j\in I_0} n(\underline{x}^j)}, \qquad \omega_1(\underline{x}^i) = \frac{n(\underline{x}^i)}{\sum_{j\in I_1} n(\underline{x}^j)}. \tag{10}$$

The notation δ_ρ is used to denote unweighted test statistics, with $\delta_{\rho,\omega}$ used to denote weighted test statistics. Uniform weighting takes into account the evidence of each point pattern equally, while weighting according to the number of points attaches equal weight of evidence to each point.

Pattern size, area and intensity test statistics summarise point pattern properties through single numbers dependent only upon first order properties, independently of the location of observed points x_j within the observation window W. In cases where point patterns are expected to be realisations of homogeneous Poisson point patterns, specification of the intensity or number of points completely describes the distribution of point locations. However, knowledge of the number of points or the estimated density under the assumption of homogeneity is inadequate for specification of the distribution of point locations for inhomogeneous Poisson point processes and alternative classes of point processes. As a result, test statistics based upon the second order property of distances between points may be more suitable when the patterns are not believed to be realisations of homogeneous Poisson processes.

We propose comparison of $\overline{\text{nnd}}(\underline{x})$, defined in (2), between collections using the mean nearest neighbour test statistic

$$\delta_{\text{nnd}}(I) = \sum_{i\in I_0} \omega_0(\underline{x}^i)\overline{\text{nnd}}(\underline{x}^i) - \sum_{i\in I_1} \omega_1(\underline{x}^i)\overline{\text{nnd}}(\underline{x}^i). \tag{11}$$

Similarly, we propose comparison of the mean minimum spanning distance $\overline{\text{msd}}(\underline{x})$, defined in (3), between collections using the mean minimum spanning distance test statistic

$$\delta_{\text{msd}}(I) = \sum_{i \in I_0} \omega_0(\underline{x}^i)\overline{\text{msd}}(\underline{x}^i) - \sum_{i \in I_1} \omega_1(\underline{x}^i)\overline{\text{msd}}(\underline{x}^i).$$

The mean nearest neighbour distance and mean minimum spanning distance test statistics compare average separation distances, ensuring that the statistics summarise information only on the distances between points and not on the number of points in each pattern. While nearest neighbour and minimum spanning distances both summarise point separation distances, they are expected to differ in particular for cluster processes. For a cluster process the mean nearest neighbour distance summarises within cluster point separation distances, while the mean minimum spanning distance tends to summarise both within cluster point separation distances and between cluster distances.

Summarising the information contained within a point pattern by a single number before comparison between collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$ leads to a large loss of information. Alternatively, summarising the information in each observation by a function allows the preservation of a greater amount of information, with comparison between functions revealing differences which are unobservable when comparing single number summary statistics. The additional information can be carried forwards even in cases where the comparison of functions results in a single number.

Diggle et al. (2000) propose a test statistic for the comparison of the scaled neighbourhood count function, $K(r)$, an estimator of which is defined in (5). An estimate of the average scaled neighbourhood count function evaluated at distance r for the collection of point patterns $\underline{x}^J$ indexed by the set J, denoted by $\bar{K}(\underline{x}^J, r)$, is given by

$$\bar{K}(\underline{x}^J, r) = \frac{1}{\sum_{i \in J} n(\underline{x}^i)} \sum_{i \in J} n(\underline{x}^i)\hat{K}(\underline{x}^i, r). \tag{12}$$

The proposed test statistic, which will in future be referred to as the scaled neighbourhood count test statistic, is given by

$$\begin{aligned}\delta_K(I) = &\int_0^{r_0} \frac{1}{r^2} \left(\sum_{i \in I_0} n(\underline{x}^i) \right) \left[\bar{K}(\underline{x}^{I_0}, r) - \bar{K}(\underline{x}^I, r) \right]^2 dr \\ &+ \int_0^{r_0} \frac{1}{r^2} \left(\sum_{i \in I_1} n(\underline{x}^i) \right) \left[\bar{K}(\underline{x}^{I_1}, r) - \bar{K}(\underline{x}^I, r) \right]^2 dr.\end{aligned} \tag{13}$$

The integration in (13) is carried out across the range $r = 0$ to $r = r_0$, where r_0 is chosen to be large enough to summarise the information present in the point patterns, but small enough in comparison to the dimensions of W that the impact of edge corrections is not too great. The dominant term in the sampling variance of $\hat{K}(\underline{x}, r)$ is of order $1/n(\underline{x})$ for a homogeneous Poisson process (Diggle et al., 2000), motivating the weighted average in the expression of $\bar{K}(\underline{x}^J, r)$ in (12). Furthermore, for a homogeneous Poisson point pattern and values of r which are small in relation to the dimensions of the observation window W, the dominant term in the sampling variance of $\hat{K}(\underline{x}, r)$ with r is of order r^2 (Diggle et al., 2000). As a result a multiplier of $1/r^2$ is included in the expression of the scaled neighbourhood count test statistic given by (13).

Estimated nearest neighbour functions, defined in (6), may be averaged over the collection of

point patterns $\underline{x}^J$ indexed by the set J, to produce $\hat{G}(\underline{x}^J, r)$ given by

$$\hat{G}(\underline{x}^J, r) = \sum_{i \in J} \omega_J(\underline{x}^i) \hat{G}(\underline{x}^i, r), \tag{14}$$

where ω_J is defined for the set J as ω_0 is defined for the set I_0 in (10). We propose two test statistics for comparison of $\hat{G}(\underline{x}^{I_0}, r)$ and $\hat{G}(\underline{x}^{I_1}, r)$ across the range of distances r. These will be referred to as the L_1 nearest neighbour distribution test statistic, denoted by $\delta_{G,1}(I)$, and the L_∞ nearest neighbour distribution test statistic, denoted by $\delta_{G,\infty}(I)$, given by

$$\begin{aligned} \delta_{G,1}(I) &= ||\hat{G}(\underline{x}^{I_0}, r) - \hat{G}(\underline{x}^{I_1}, r)||_1 \\ &= \int_0^\infty |\hat{G}(\underline{x}^{I_0}, r) - \hat{G}(\underline{x}^{I_1}, r)| dr. \end{aligned} \tag{15}$$

$$\begin{aligned} \delta_{G,\infty}(I) &= ||\hat{G}(\underline{x}^{I_0}, r) - \hat{G}(\underline{x}^{(I_1}, r)||_\infty \\ &= \sup_r |\hat{G}(\underline{x}^{I_0}, r) - \hat{G}(\underline{x}^{I_1}, r)|. \end{aligned} \tag{16}$$

Both the L_1 and L_∞ nearest neighbour distribution test statistics propose comparison of the nearest neighbour function, $G(r)$, over an infinite range of r values. In practise, there is a finite value of $r = r_1$, equal to the largest observation window diameter, which all observed nearest neighbour distances are less than and, as a result, $\hat{G}(\underline{x}, r) = 1, \forall r \geq r_1$. The upper limit r_1 could then replace the upper limit of integration in the L_1 nearest neighbour distribution test statistic and limit the set over which the supremum is taken in the L_∞ nearest neighbour distribution test statistic to $r \in [0, r_1]$.

The L_∞ and L_1 norms are the two extreme L_p norms which respectively attribute all weight to the maximum difference between functions across the argument r and equal weight to the difference between functions across r. As a result, the L_1 distance may be preferred as it explicitly compares across the entire range of values of r. The scaled neighbourhood count test statistic may be interpreted as the sum of L_2 distances between $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$ collection averages and the overall $\underline{x}^I$ collection average scaled neighbourhood count functions.

We propose comparison of the effective force transference test statistic between collections using the effective force transference test statistic

$$\delta_{EFT}(I) = \sum_{i \in I_0} \omega_0(\underline{y}^i) EFT(\underline{y}^i) - \sum_{i \in I_1} \omega_1(\underline{y}^i) EFT(\underline{y}^i).$$

3.5 Significance quantification

In the formulation of the statistical problem, we state that observations $\underline{x}^{I_0}$ are independent realisations of the point process $\underline{X}^0$ and observations $\underline{x}^{I_1}$ are independent realisations of the point process $\underline{X}^1$. It is not possible that any one-dimensional single test statistic totally summarises all possible differences between $\underline{X}^0$ and $\underline{X}^1$. In the previous section we therefore proposed a number of test statistics, each of which compares particular properties of $\underline{X}^0$ and $\underline{X}^1$. Despite the variety of proposed test statistics our null hypothesis is unchanged for each test statistic, H_0: $\underline{X}^0$ and $\underline{X}^1$ are equal in distribution, with straightforward extension to $\underline{Y}^0$ and $\underline{Y}^1$.

Without making further assumptions on the properties of $\underline{X}^0$ and $\underline{X}^1$, the distribution of proposed test statistics under the null hypothesis cannot be analytically determined. Analysis of the significance of calculated test statistics is therefore carried out using permutation testing.

Under the null hypothesis the labelling of point patterns as members of the collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$ is exchangeable because they are assumed to independent realisations of the same point process, $\underline{X}$, say.

Exchangeability of allocation to collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$ is theoretically described by the application of permutations $\gamma \in S_{|I|} = \{\gamma_0, \gamma_1, \ldots, \gamma_m\}$, the symmetric group of degree $|I|$, to the set of indices I. Application of the permutation γ_j to the set of indices I results in subsets $I_0^{(j)}$ and $I_1^{(j)}$ satisfying

$$I_0^{(j)} \cup I_1^{(j)} = I, \qquad I_0^{(j)} \cap I_0^{(j)} = \emptyset,$$
$$|I_0^{(j)}| = |I_0|, \qquad |I_1^{(j)}| = |I_1|,$$

between which point patterns are compared. The observed test statistic under the identity permutation, $\delta(\gamma_0(I))$ with $\gamma_0 : \gamma_0(I) = I$, is then be compared to the set $\{\delta(\gamma_0(I)), \delta(\gamma_1(I)), \ldots, \delta(\gamma_m(I))\}$.

In practice, the total number of permutations is too large to feasibly calculate for both simulated data and the real data examples considered in the following sections. An approximate permutation test is therefore carried out using a subset of $S_{|I|}$ of size m' typically equal to 10^3 or 10^4 to ensure good power (Jöckel, 1986). In the case where $|I_0| = k$, $I_0^{(j)}$ is obtained by randomly sampling k elements of I without replacement such that $I_0^{(j)} \neq I_0^{(l)}$ $\forall l \in \{0, 1, \ldots, j\}$. The corresponding set $I_1^{(j)}$ is then simply $I \setminus I_0^{(j)}$ and the first permutation is chosen specifically to be the identity permutations such that $I_0^{(0)} = I_0$ and $I_1^{(0)} = I_1$.

Across all proposed test statistics $\delta(I) = 0$ indicates no difference between $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$. Scaled neighbourhood count and nearest neighbour distribution test statistics take values in $[0, \infty)$, as a result one-sided p-values are calculated. All other proposed test statistics take values in $(-\infty, \infty)$, as a result two-sided p-values are calculated.

4 Validation study

To confirm the suitability of the methods described in the previous section, we first apply them to simulated data which models the features that we wish to differentiate between.

4.1 Simulation description

Let $HPPP(\rho, W)$ denote the homogeneous Poisson point process with intensity ρ on the window $W \subset \mathbb{R}^2$. Similarly, let $IPPP(\rho(x), W)$ denote the inhomogeneous Poisson point process with intensity $\rho(x)$ on the window $W \subset \mathbb{R}^2$.

Let $CPP(d, W)$ denote the cluster point pattern containing $n + 1$ points distributed over the observation window $W \subset \mathbb{R}^2$, according to $d = \{d_1, d_2, \ldots, d_n\}$, an ordered set of n fixed strictly positive separation distances arranged in ascending order. Point location x_{j+1} is determined by rejection sampling using as a proposal distribution Uniform$\{x \in \mathbb{R}^2 : \exists k \in \{1, 2, \ldots, j\}$ st $||x - x_k|| = d_j\}$, with acceptance probability $\mathbf{1}\{x \in W, \inf_{k \in 1,2,\ldots,j} ||x - x_j|| \geq d_j\}$. The result of this procedure is point locations distributed according to the conditional distribution

$$x_{i+1}|x_1, x_2, \ldots, x_i \sim \text{Uniform}\left(W \cap \left\{x \in \mathbb{R}^2 : \min_{j \in \{1,2,\ldots,i\}} ||x - x_j|| = d_i\right\}\right), \quad i = 1, 2, \ldots, n.$$

Data: d, W
Result: Point pattern $\underline{x}$ on W with nearest neighbour distances $\{d_1\} \cup d$
$x_1 \leftarrow$ centre of W;
for i *in* $1 : n$ **do**
  repeat
    $j \sim \text{Uniform}\{1, \ldots, i\}$;
    $\theta \sim \text{Uniform}[0, 2\pi]$;
    $x_{i+1} \leftarrow x_j + d_i(\cos\theta, \sin\theta)$;
  until $\min_{k \in \{1,2,\ldots,i\}} ||x_{i+1} - x_k|| \geq d_i$ *and* $x_{i+1} \in W$;
end
$\underline{x} \leftarrow (x_1, x_2, \ldots, x_{n+1})$;

Algorithm 1: $CPP(d, W)$ generation

An algorithmic description for the generation of a realisation of a $CPP(d, W)$ process is given in Algorithm 1. From Algorithm 1, it is clear that $\text{nnd}(x_1) = d_1$ and $\text{nnd}(x_i) = d_{i-1}, i \in \{2, 3, \ldots, n+1\}$. Similarly, the minimum spanning tree of $\underline{X} \sim CPP(d, W)$ has weight $\sum_{i \in \{1,2,\ldots,n\}} d_i$ by construction.

Let $MPP(n, u, \phi)$ denote marked point process on χ_2^+ with a fixed number of points n and mark directions v_j deviating from u by angles of up to ϕ. As the effective force transference summary statistic depends only upon the distribution of marks v_j, point locations y_j are fixed at the origin. Marks are then simulated for $u' = (0, 0, 1)$ with

$$\phi_j \sim \text{Uniform}[0, \phi], \qquad \theta_j \sim \text{Uniform}[0, 2\pi], \qquad v_j' = (\sin\phi_j \cos\theta_j, \sin\phi_j \sin\theta_j, \cos\phi_j).$$

Marks are transformed by the rotation R for which $Ru' = u$ and scaled by the constant η_j such that $v_j \cdot \hat{z} = h$, producing $v_j = \eta_j R v_j'$.

4.2 Study design

For each simulation type, the required collection of point patterns $\underline{x}^I$ or $\underline{y}^I$ is simulated before permutation testing is carried out for each of the appropriate test statistics, with the resulting p-value being recorded. One hundred replicates of each simulation and testing procedure are carried out to provide information on the sensitivity and specificity of proposed tests and the variability of these properties. In the case of point patterns $\underline{x}$ collection sizes are $|I_0| = |I_1| = 30$ and for marked point patterns $\underline{y}$ the collection sizes are $|I_0| = |I_1| = 14$ to approximately match the TACC3 data. A realisation of each simulation type is displayed in Figure 1.

Homogeneous intensity simulations are made up of point patterns $\underline{x}$ simulated according to $HPPP(\rho_0, W)$ for $\underline{x} \in \underline{x}^{I_0}$ and $HPPP(\alpha\rho_0, W)$ for $\underline{x} \in \underline{x}^{I_1}$. The base intensity $\rho_0 = 10^{-4}$ is chosen to approximately match that of the TACC3 data, with each W the square window with area sampled independently from $\text{Uniform}[500^2 - 10^5, 500^2 + 10^5]$ to produce approximately the same number of points per pattern as the TACC3 data. Tested values of α vary across the set $\{1, 1.1, 1.2, 1.5\}$, resulting in various strengths of difference between collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$.

Inhomogeneous intensity simulations are made up of point patterns $\underline{x}$ simulated according to $HPPP(\rho_0, W_0)$ for $\underline{x} \in \underline{x}^{I_0}$ and $IPPP(\rho, W_0)$ for $\underline{x} \in \underline{x}^{I_1}$. The observation window W_0 is fixed as

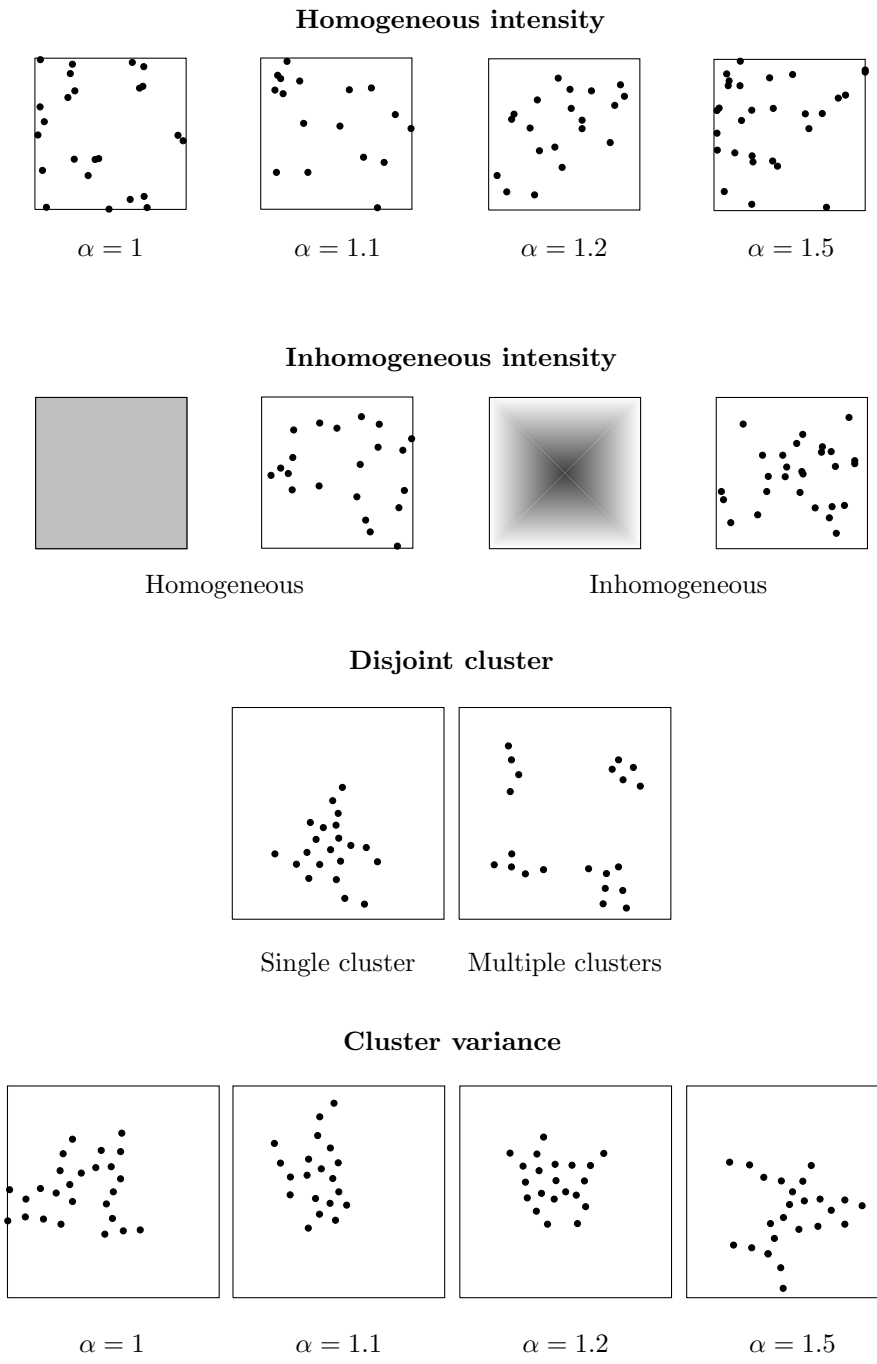

Figure 1: Realisations of each of the simulation types.

the square window with side length 500. The inhomogeneous intensity at the point with coordinate location $(x, y) \in \mathbb{R}^2$ is given by

$$\rho((x,y)) = \begin{cases} \frac{3\rho_0 x}{250} & x+y<500, x>y \\ \frac{3\rho_0 y}{250} & x+y<500, x<y \\ \frac{3\rho_0(500-x)}{250} & x+y>500, x>y \\ \frac{3\rho_0(500-y)}{250} & x+y>500, x<y. \end{cases}$$

The shape of the inhomogeneous intensity is then a square based pyramid with maximum height $3\rho_0$, chosen such that

$$\int_{W_0} \rho(x)dx = \int_{W_0} \rho_0 dx.$$

Under this formulation the first order properties of $\underline{x}$ are identical for both collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$, while the second order properties differ.

Disjoint cluster simulations aim to compare point patterns made up of a single cluster with point patterns made up of four clusters, one in each quadrant of the square observation window. Point patterns $\underline{x}$ are simulated according to $CPP(d, W_1)$ for $\underline{x} \in \underline{x}^{I_0}$. W_1 is the square window with area $|W_1| = 2|W_0| = 2 \times 500^2$ and the number of points $n(\underline{x})$ is simulated from $\text{po}(\rho_0|W_0|)$, truncated to produce values greater than eight. Separation distances d are determined by $\tilde{d}$ a sample of size $n(\underline{x}) - 4$ from Normal$(55, 10^2)$, truncated at zero to produce only non-negative separation distances, with

$$d = \{\tilde{d}_{(1)}, \tilde{d}_{(1)}, \tilde{d}_{(2)}, \tilde{d}_{(2)}, \tilde{d}_{(3)}, \tilde{d}_{(3)}, \tilde{d}_{(4)}, \tilde{d}_{(4)}, \tilde{d}_{(5)}, \tilde{d}_{(6)}, \ldots, \tilde{d}_{(n(\underline{x})-4)}\},$$

where $\tilde{d}_{(j)}$ is the j^{th} order statistic of $\tilde{d}$. For $\underline{x} \in \underline{x}^{I_1}$ point patterns $\underline{x} = \cup_{j=1}^4 \underline{x}^j$ are the union of four simulated realisations $\underline{x}^j$ of $CPP(d^j, W_1^j)$. The W_1^j divide W_1 into four square quadrants each with area $|W_0|/2$, with $n(\underline{x}) \sim \text{po}(\rho_0|W_0|)$, again truncated to be greater than eight, and d an independent sample of size $n(\underline{x}) - 4$ from Normal$(55, 10^2)$, again truncated to be non-negative. Separation distances d are partitioned into d^j with

$$\begin{aligned} d_j \in d^j, \quad j \in \{1,2,3,4\}, \\ \mathbb{P}(d_j \in d^k) = 1/4, \quad j \in \{5,6,\ldots,n(\underline{x})\}, \quad k \in \{1,2,3,4\}. \end{aligned}$$

All point patterns $\underline{x} \in \underline{x}^I$ have identical first order properties and identically distributed nearest neighbour distances, but $\underline{x} \in \underline{x}^{I_0}$ are made up of a single cluster while $\underline{x} \in \underline{x}^{I_1}$ are made up of four clusters.

Cluster variance simulations are made up of point patterns $\underline{x}$ simulated by $CPP(d^{(1)}, W_1)$ for $\underline{x} \in \underline{x}^{I_0}$ and $CPP(d^{(2)}, W_1)$ for $\underline{x} \in \underline{x}^{I_1}$. For all observations $n(\underline{x})$ is simulated from $\text{po}(\rho_0|W_0|)$. Separation distances $d^{(1)}$ are an independent sample of size $n(\underline{x})$ from Normal$(55, 10^2)$, truncated to be non-negative, while $d^{(2)}$ are an independent sample of size $n(\underline{x})$ from Normal$(55, 10^2\alpha)$, again truncated to be non-negative, for α varying across the set $\{1, 1.1, 1.2, 1.5\}$. Point patterns therefore have identical first order properties and identical mean nearest neighbour distances, but different nearest neighbour distance variances between collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$.

Varying angle simulations are made up of marked point patterns $\underline{y}$ simulated according to $MPP(n, u, \phi)$ for $\underline{y} \in \underline{y}^{I_0}$ and $MPP(n, u, \alpha\phi)$ for $\underline{y} \in \underline{y}^{I_1}$. The number of points, n are sampled

	Homogeneous intensity, α				Inhom. intensity	Disjoint cluster	Cluster variance, α			
	1	1.1	1.2	1.5			1	1.1	1.2	1.5
δ_N	0.07	0.23†	0.65†	1.00†	0.07	0.06	0.04	0.03	0.06	0.05
δ_W	0.06	0.10	0.04	0.04	-	-	-	-	-	-
δ_ρ	0.03	0.44†	0.96†	1.00†	0.07	0.06	0.04	0.03	0.06	0.05
$\delta_{\rho,\omega}$	0.03	0.46†	0.96†	1.00†	0.07	0.07	0.06	0.03	0.07	0.05
δ_{nnd}	0.05	0.16†	0.64†	1.00†	0.98†	0.06	0.05	0.05	0.06	0.06
$\delta_{\text{nnd},\omega}$	0.04	0.21†	0.70†	1.00†	0.99†	0.04	0.05	0.05	0.06	0.08
δ_{msd}	0.03	0.29†	0.80†	1.00†	1.00†	1.00†	0.05	0.05	0.07	0.02
$\delta_{\text{msd},\omega}$	0.04	0.30†	0.85†	1.00†	1.00†	1.00†	0.05	0.05	0.08	0.02
δ_K	0.08	0.00	0.06	0.04	1.00†	1.00†	0.04	0.06	0.07	0.06†
$\delta_{G,1}$	0.07	0.16†	0.66†	1.00†	0.97†	0.09	0.05	0.32†	0.91†	1.00†
$\delta_{G,1,\omega}$	0.05	0.19†	0.71†	1.00†	0.97†	0.09	0.03	0.33†	0.91†	1.00†
$\delta_{G,\infty}$	0.06	0.14†	0.51†	0.99†	0.95†	0.09	0.05	0.19†	0.44†	1.00†
$\delta_{G,\infty,\omega}$	0.06	0.15†	0.54†	1.00†	0.97†	0.08	0.04	0.16†	0.50†	1.00†

Table 1: Proportion of p-values in the range $[0, 0.05]$, † indicates non-uniformity of p-values under the Kolmogorov-Smirnov test at the Bonferroni corrected (Dunn, 1961) $5/13 = 0.38$ (Homogeneous intensity) or $5/12 = 0.42$ (Inhomogeneous intensity, Disjoint cluster, Cluster variance) percent significance level.

independently from po(30) to approximately match the TACC3 data. Direction u is fixed at $(0, 0, 1)$ as results are independent of its value. Maximum angle ϕ is fixed at 10 degrees, the approximate average for TACC3 data, with α varying across the set $\{1, 1.05, 1.1, 1.2, 1.5\}$. There is therefore a greater variance in line trajectories for $\underline{y} \in \underline{y}^{I_1}$ than for $\underline{y} \in \underline{y}^{I_0}$ with increasing α.

4.3 Study results

Simulated point patterns are tested using each of the relevant test statistics using a total of 1000 permutations and this procedure is repeated for 100 independent sets of simulations, resulting in 100 p-values for each test statistic for each parameter value. The distribution of p-values is compared to the Uniform distribution on $[0, 1]$ via the Kolmogorov-Smirnov test (Massey, 1951) at the five percent significance level, to determine whether there is a significant difference between collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$. The proportion of p-values in the interval $[0, 0.05]$ is also considered as the strength of evidence for rejection of the null hypothesis. These values are presented for simulated point patterns $\underline{x}$ in Table 1 and simulated marked point patterns $\underline{y}$ in Table 2.

Test statistics based upon first order statistics, δ_N, δ_ρ and $\delta_{\rho,\omega}$, successfully report a difference for homogeneous intensity simulations where the difference in intensity is at least ten percent, but fail to report a difference in any other case. This behaviour is to be expected as the alternative classes of simulations produce differences in second order properties, while keeping first order properties constant. There is negligible difference between unweighted and weighted intensity test

statistics, δ_ρ and $\delta_{\rho,\omega}$.

The pattern area test statistic, δ_W, consistently fails to reject the uniform distribution of p-values. This is the expected behaviour as in the homogeneous intensity simulations the pattern area, $|W|$, is identically distributed for both collections of point patterns $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$. For all other simulations the pattern area is exactly identical for all simulated point patterns and it is unnecessary to test the pattern area test statistic.

Test statistics based upon single number second order summary statistics, δ_{nnd}, $\delta_{\text{nnd},\omega}$, δ_{msd} and $\delta_{\text{msd},\omega}$, successfully report a difference for both homogeneous and inhomogeneous intensity simulations. Mean minimum spanning distance test statistics, δ_{msd} and $\delta_{\text{msd},\omega}$, are additionally able to report the difference between disjoint cluster simulations. This is expected by construction of the mean minimum spanning distance summary statistic which summarises both within and between cluster distances, in comparison to the mean nearest neighbour distance summary statistic which summarises only within cluster distances. As these test statistics compare the means of separation distances they are unable to report a difference when only the variance of separation distances changes, as in the case of the cluster variance simulations. The difference between weighted and unweighted test statistics is small, but in general weighted test statistics are more accurate at detecting differences. Test statistics based upon the mean minimum spanning distance generally outperform those based upon the mean nearest neighbour distance.

Test statistics based upon functional summary statistics of nearest neighbour distances, $\delta_{G,1}$, $\delta_{G,1,\omega}$, $\delta_{G,\infty}$ and $\delta_{G,\infty,\omega}$, successfully report a difference in every case except for disjoint cluster simulations. This behaviour is expected as by design the nearest neighbour distribution depends upon within cluster separation distances that are unchanged for disjoint cluster simulations, while the between cluster separation distance does change. The difference between weighted and unweighted test statistics is again small, with weighted test statistics performing slightly better for homogeneous and inhomogeneous intensity simulations and no clear difference for cluster variance simulations. Test statistics comparing nearest neighbour functions using the L_1 distance outperform those which compare using the L_∞ distance, likely because they directly summarise the difference over the entire range of distances r.

The scaled neighbourhood count test statistic, δ_k, accurately reports a difference in the case of inhomogeneous intensity and disjoint cluster simulations, but is largely ineffective for alternative simulations. As the scaled neighbourhood count summary statistic is normalised by the estimated intensity, $\hat{\rho}$, it is not expected to report a difference for homogeneous intensity simulations. Accurate detection in the case of disjoint cluster simulations may result from the fact that the scaled neighbourhood count function summarises more point separation distances than simply nearest neighbour distances.

The effective force transference test statistics, δ_{EFT} and $\delta_{EFT,\omega}$, accurately report a difference in mark orientation when the difference in maximum deviation is increased by at least five percent, consistently so when the increase is at least 20 percent.

As a result of the simulation study we have confirmed the scenarios in which each of the test statistics are effective, supporting the use of specific test statistics when it is desired to test for particular differences in structure. We have also confirmed the suitability of the permutation testing approach, with an absence of evidence to reject uniformity of p-values when there is no difference between simulated point pattern collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$.

	Varying angle, α				
	1	1.05	1.1	1.2	1.5
δ_{EFT}	0.03	$0.32^{\dagger}$	$0.83^{\dagger}$	$1.00^{\dagger}$	$1.00^{\dagger}$
$\delta_{EFT,\omega}$	0.01	$0.32^{\dagger}$	$0.84^{\dagger}$	$1.00^{\dagger}$	$1.00^{\dagger}$

Table 2: Proportion of p-values in the range $[0, 0.05]$, † indicates non-uniformity of p-values under the Kolmogorov-Smirnov test at the Bonferroni corrected (Dunn, 1961) $5/2 = 2.5$ percent significance level.

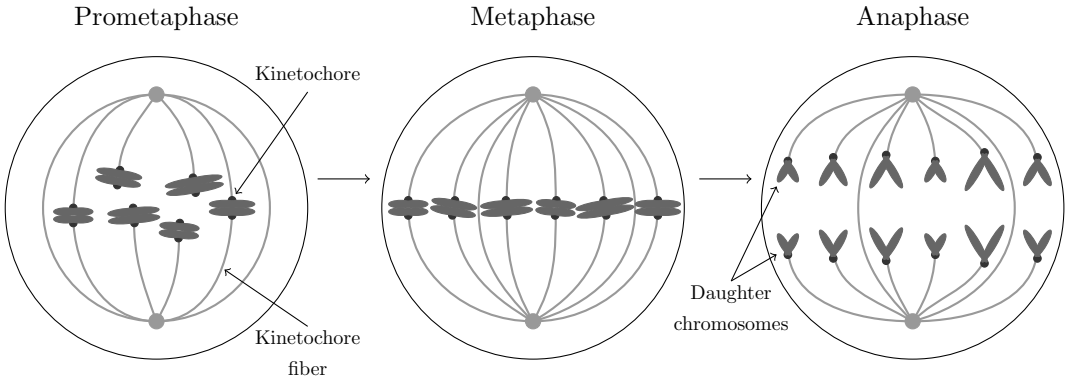

Figure 2: Diagram of the intermediate stages of mitosis.

5 Investigation of changes in K-fiber microtubule organisation following TACC3 overexpresison

5.1 Biological background

During mitosis, subcellular structures known as kinetochore fibers (K-fibers) connect two anchor points, centrosomes, within the cell to each of the chromosome pairs (Booth et al., 2011). The chromosomes contain the genetic information and successful mitosis requires the pairs to be evenly divided, such that each of the two daughter cells contains one of each chromosome (Holland and Cleveland, 2009). K-fibers are believed to apply the force necessary to separate chromosome pairs and as a result they are important for successful cell division. An illustration of the intermediate stages in the process of mitosis including the function of the K-fibers can be seen in Figure 2.

Each K-fiber is made up of a number of microtubules, approximately cylindrical structures which are bound together by a mesh structure to form the rigid K-fiber. Our collaborators, Dr. Steve Royle and his research group within the Centre for Mechanochemical Cell Biology at the University of Warwick, are interested in the effect that overexpression of TACC3, Transforming acidic coiled-coil containing protein 3, may have on the structure of microtubules within K-fibers. The structure may be visualised by microscopy imaging of cells at the correct point in the cell cycle under a control regime and a treatment regime for which there is overexpression of TACC3. Images are collected in planes perpendicular to the K-fiber axis, resulting in microtubules visible through their cross-sections as dark circles. Example plots of microtubule locations may be seen in Figure 3. Initial analysis of the microtubule centres within a single image results in a collection of

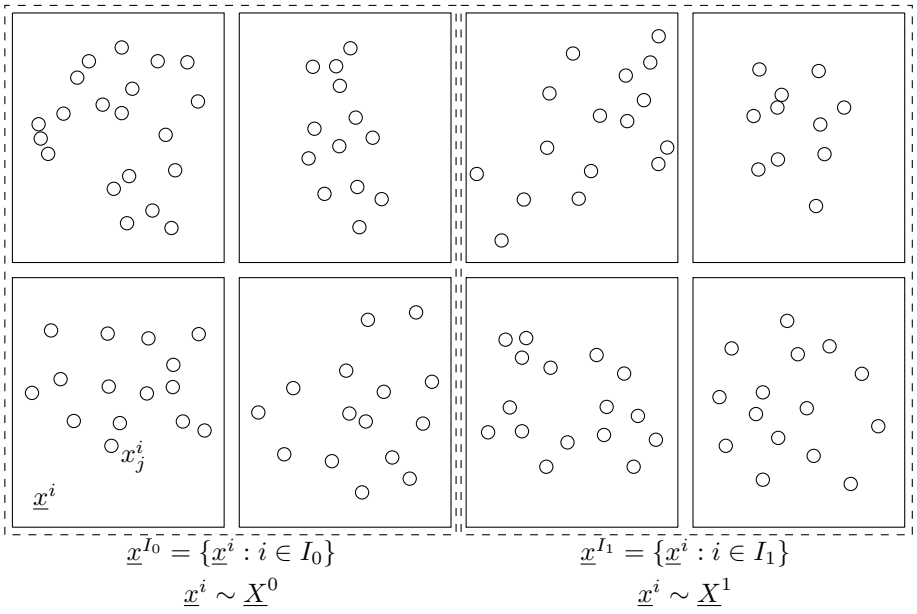

Figure 3: Plots of sample microtubule location data. Each pattern represents a single K-fiber with microtubules represented as circles, the centres of which are analysed as point locations and the radii of which are consistent with the expected microtubule radius.

coordinate locations, with imaging of multiple cells under each regime producing the two collections of coordinate locations, $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$, for analysis.

We choose to investigate 2D coordinate locations obtained from imaging biological samples as point patterns. Choosing to model the TACC3 2D point pattern data within this framework, we have $|I| = 63$, I_0 indexing observations under the control regime with $|I_0| = 26$ and I_1 indexing observations under the treatment regime with $|I_1| = 37$.

Also available is a data set comprising paired 2D microtubule coordinate locations obtained from two parallel image slices through the same sample approximately perpendicular to microtubule directions, including information on which coordinate locations represent the same microtubule in each of the slices. Paired coordinate locations are reported as

$$\underline{x} = \{(x_{1,0}, x_{1,1}), (x_{2,0}, x_{2,1}), \ldots, (x_{n(\underline{x}),0}, x_{n(\underline{x}),1})\}.$$

We choose to investigate paired 2D coordinate locations obtained by imaging biological samples as marked point patterns $\underline{y}$ with $x_j = x_{j,0} \times \{0\}$ and $v_j = (x_{j,1} - x_{j,0}) \times \{h\}$. An example plot of marked point pattern data may be seen in Figure 4.

Paired coordinate observations are similarly collected for multiple samples under each of the two experimental regimes. Choosing to model the TACC3 paired point pattern data within this framework, we have $|I| = 28$, I_0 indexing observations under the control regime with $|I_0| = 13$ and I_1 indexing observations under the treatment regime with $|I_1| = 15$.

5.2 TACC3 data features

Using the notation introduced in Section 3.1.1, point patterns $\underline{x}$ are realisations of point processes, $\underline{X}$, obtained by reporting the locations of all points contained within a fixed observation window W.

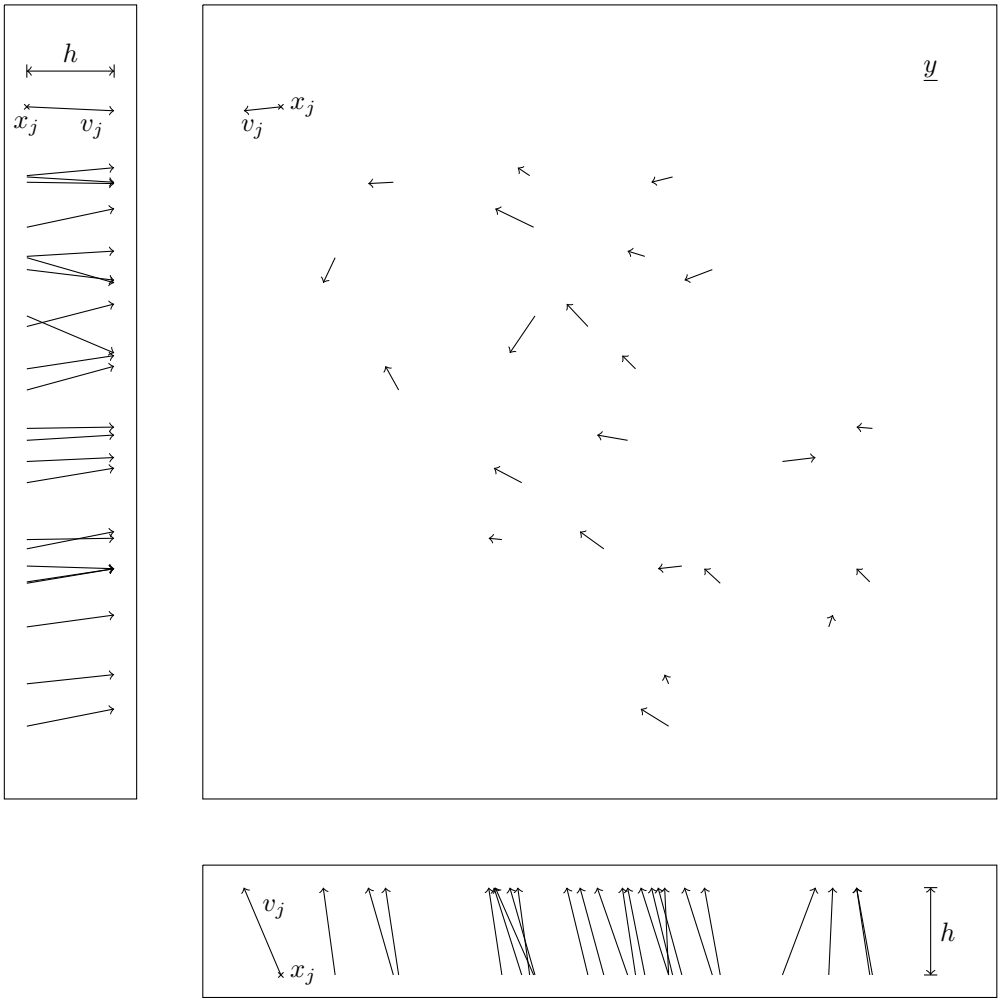

Figure 4: Plots of sample marked point pattern data. Point locations and mark orientations representing the location and direction of microtubules when viewed parallel to the K-fiber axis (left and bottom) and perpendicular to the K-fiber axis (top right).

The TACC3 data set instead adds points to patterns using an iterative process. The experimenter first selects a single dark circle within the microscope image, taken to represent a microtubule within the K-fiber of interest, reporting the coordinates of its centre as the first point location x_1. The coordinates corresponding to the centres of any further dark circles which lie within a distance of 105nm of x_1 are then added to the point pattern. This procedure is iterated, until there exists no further dark circles with centre within 105nm of a point in the pattern. The specific distance of 105nm has been suggested by previous biological investigations to be the maximum centre-to-centre distance at which K-fiber microtubules may be connected by mesh, as microtubules have an average radius of 12.5nm and the mesh forms edge-to-edge connections at distances up to 80nm (Nixon et al., 2015). All locations and distances are on the scale of nanometres and as a result we omit explicit reference to units in future, describing instead points within a distance of 105.

The purpose of the iterative procedure is to isolate and report the locations of only those microtubules which belong to a single K-fiber. An estimate to the observation window may be recreated using the observed data, resulting in $W(\underline{x})$ given by

$$W(\underline{x}) = \cup_{x_j \in \underline{x}} b(x_j, 105), \tag{17}$$

where $b(x_j, 105)$ is the two dimensional ball centred at x_j with radius 105.

Paired coordinate locations for the TACC3 data represent the endpoints of microtubules which may be modelled as straight lines at the scale of our observations. Point locations x_j specify microtubule locations, while marks v_j specify microtubule directions. Combining these pieces of information, a parametric specification of the j^{th} microtubule is given by $L_j(t) = x_j + tv_j$ for $t \in [0, 1]$.

The purpose of the K-fiber is the separation of chromosome pairs during mitosis, with each K-fiber forming a physical link between a chromosome and the mitotic spindle pole, transferring the force required to physically pull apart the chromosome pair. Under the simplifying assumption that force is applied through the K-fiber via the application of forces of identical magnitude, F, in the direction of each of the microtubules, v_j, the resultant force is given by

$$\sum_{j=1}^{n(\underline{y})} F\hat{v}_j = F\beta.$$

The magnitude of the resultant force is $F||\beta||$ and, if we further assume the K-fiber to be optimally directed such that all of this resultant force is used to separate the chromosome pair, the proportion of input force effectively transferred is

$$\frac{F||\beta||}{F \times n(\underline{y})} = \frac{||\beta||}{n(\underline{y})} = EFT(\underline{y}).$$

Therefore, the effective force transference summary statistic intuitively relates to the performance of the K-fiber, with larger values of effective force transference indicating that the K-fiber is more effective at transferring force as its component microtubules are better aligned.

The TACC3 paired point pattern data arises from parallel imaging planes which are oriented approximately perpendicular to the axis of the K-fiber. As the K-fiber axis is not uniquely defined and its features more generally are obscured before imaging takes place, there is the possibility of variation in the orientation of the imaging planes between samples. Imaging a single sample at different orientations results in changes in the length of marks, $|v_j|$, but does not affect the

trajectory of marks relative to each other. The effective force transference test statistic has been proposed with this particular property in mind and, as it is calculated using only relative mark trajectories, it is independent of the orientation of the imaging planes (provided imaging planes are not parallel to any of the microtubule directions).

5.3 Assumption checking

Coordinate locations reported for TACC3 data represent the centres of microtubules. Under the assumption that microtubules have radii of 12.5 and are connect by a mesh structure which extends a distance of no further than 80, we expect

$$\text{nnd}(x_j) \in [25, 105], \tag{18}$$

for all points x_j in all point patterns $\underline{x}$. In practice 1811 of the 1824 total points satisfy (18), with two values of $\text{nnd}(x_j) = 23.5$ and 11 values of $\text{nnd}(x_j) > 105$, the largest of which is 127.3.

Extending the idea of a nearest neighbour distance to marked point patterns through the expression of lines $L_j(t)$, we have

$$\text{nnd}_{\text{min}}(L_j) = \min_{k \in \{1,2,\ldots,j-1,j+1,\ldots,n(\underline{y})\}} \min_{t_1,t_2 \in [0,1]} ||L_j(t_1) - L_k(t_2)||,$$

and expect that $\text{nnd}_{\text{min}}(L_j) \in [25, 105]$. In practice, for all but two patterns values of $\text{nnd}_{\text{min}}(L_j^i)$ lie within the slightly extended range $[20, 110]$.

Marked point pattern $\underline{y}^{13} \in \underline{y}^{I_0}$ has a large number of points with $\text{nnd}_{min}(L_j^{13}) > 110$ and pattern $\underline{y}^{17} \in \underline{y}^{I_1}$ has a large number of points with $\text{nnd}_{min}(L_j^{17}) < 20$. Individual investigation of these patterns indicates that the discrepancy in $\underline{y}^{13}$ may be explained by incorrect scaling from pixel coordinates to physical coordinates and the discrepancy in $\underline{y}^{17}$ may be caused by imaging planes significantly far from perpendicular to the K-fiber axis. Despite these deviations from the assumptions, we continue with analysis of the entire data set as the observations are only slightly outside of the expected range and because the amount of data is already limited due to the time and expertise required to collect the images. A sensitivity analysis of the data after the removal of these two marked point patterns indicates that although there is a marginal change in p-value, the significance of results remains unchanged.

Each of the proposed summary statistics is calculated under the assumption that point patterns are stationary. Visualisation of the estimated intensity of point patterns, obtained using the approach described by Diggle (1985), supports this assumption. Point patterns are further assumed to be generated by a process other than the homogeneous Poisson point process, as a result of which we propose alternative methods. Consider x_j a randomly selected point from a homogeneous Poisson point process with intensity ρ. We then have that

$$\mathbb{P}(25 \leq \text{nnd}(x_j) \leq 105) = \exp(-\rho\pi 25^2) - \exp(-\rho\pi 105^2),$$

which is maximised by the value of $\rho = \rho^*$ given by

$$\rho^* = \frac{2\log 105 - 2\log 25}{\pi(105^2 - 25^2)} = 8.8 \times 10^{-5}, \tag{19}$$

at which its value is 0.79. As over 99 percent of observed points x_j do satisfy $\text{nnd}(x_j) \in [25, 105]$, there is strong evidence to support the assumption that point patterns are not realisations of homogeneous Poisson point processes.

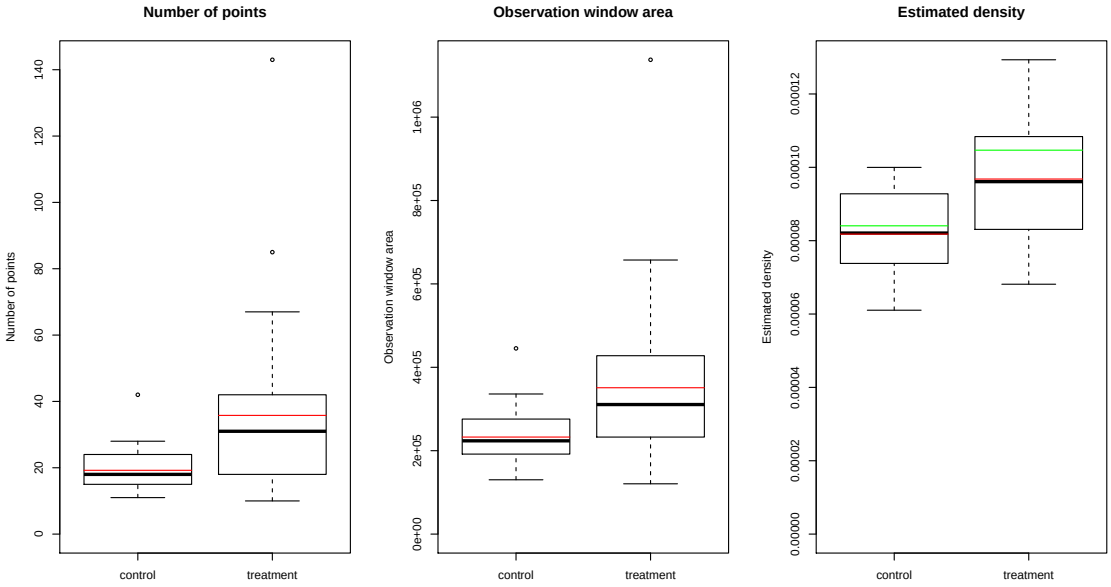

Figure 5: Boxplots of $n(\underline{x})$, $|W(\underline{x})|$ and $\hat{\rho}$ for TACC3 point patterns divided by controls, indexed by I_0, and treatments, indexed by I_1. Horizontal red lines indicates means, horizontal green lines indicate weighted means.

5.4 Exploratory data analysis

Under the reconstruction of observation windows given by (17), the area of the observation window, $|W(\underline{x})|$, may be taken as a surrogate for the cross sectional area of the K-fiber. There is then a clear dependence between the observation window area $|W(\underline{x})|$ and the number of points $n(\underline{x})$. This dependence impacts the accuracy of the estimated density $\hat{\rho} = n(\underline{x})/|W(\underline{x})|$.

Figure 5 displays boxplots of the number of points, observation window area and estimated density for the TACC3 point patterns separated into collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$ according to whether they were obtained under the control or treatment experimental regime. Means and medians are greater for treatment observations for each of the first order test statistics, indicating that K-fibers are made up of a greater number of microtubules which are more closely separated within thicker K-fibers when the TACC3 protein is overexpressed. Weighted mean densities over the collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$ are greater than unweighted means, indicating that K-fibers with greater numbers of microtubules are more tightly packed.

Figure 6 displays boxplots of the mean nearest neighbour distance and mean minimum spanning distance for the TACC3 point patterns divided into two collections for control, $\underline{x}^{I_0}$, and treatment, $\underline{x}^{I_1}$, observations. In each case the average separation distance is reduced for treatment observations, indicating that TACC3 may limit the distance at which microtubules can be connected. Each weighted mean is also less than the corresponding unweighted mean, indicating that K-fibers with more microtubules are more tightly packed.

The assumption that $\text{nnd}(\underline{x}) \leq 105$, verified in Section 5.3, in conjunction with the specification of observation windows $W(\underline{x})$ in (17), ensures that $\text{nn}(x_j) \in \underline{x}$ for all points $j \in \{1, 2, \ldots, n(\underline{x})\}$ in each of the patterns. The nearest neighbour function, G, may therefore be calculated without the

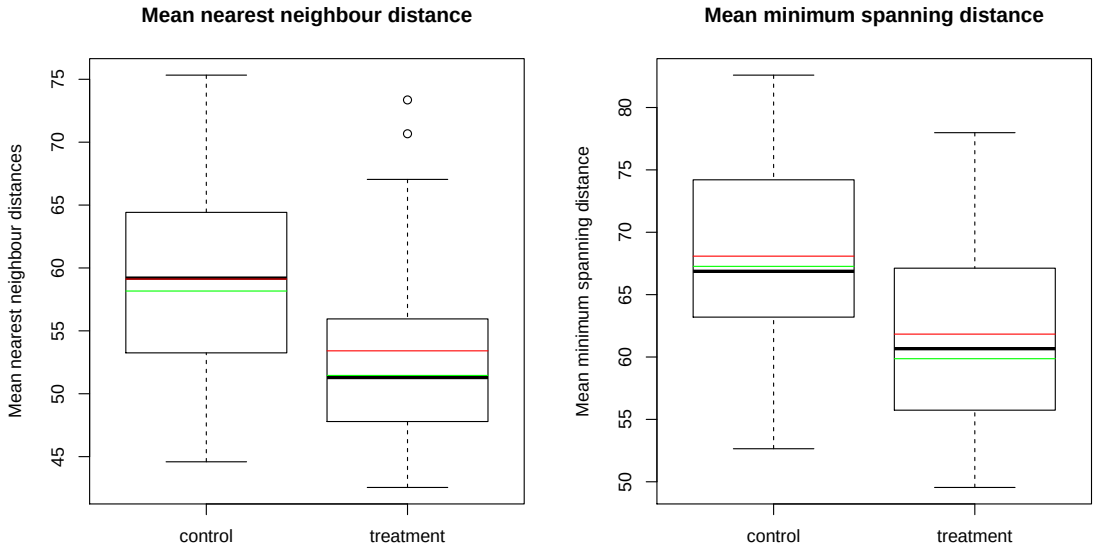

Figure 6: Boxplots of $\overline{\text{nnd}}(\underline{x})$ and $\overline{\text{msd}}(\underline{x})$ for TACC3 point patterns divided by controls, indexed by I_0, and treatments, indexed by I_1. Horizontal red lines indicates means, horizontal green lines indicate weighted means.

need for edge correction.

Plots of the estimated nearest neighbour function, $\hat{G}(\underline{x}, r)$, for TACC3 point patterns typically lie below the theoretical form for the Poisson point process with intensity ρ^*, from (19) $G(r) = 1 - \exp(-\rho^* \pi r^2)$, for small values of r less than approximately 50. There is therefore evidence of regularity on short length scales, although it is unclear the extent to which this is caused by the restriction of $\text{nnd}(x_j) > 25$ and the cumulative nature of the nearest neighbour function obscures the exact scale on which there is regularity. At values of r greater than approximately 50, plots of the nearest neighbour function exceed their expected theoretical form. There is therefore evidence of clustering at large length scales, although it is once again unclear the extent to which this is caused by the restriction of $\text{nnd}(x_j) < 105$.

Evidence of clustering and regularity can be observed in Figure 7 which includes a plot of $\hat{G}(\underline{x}^{I_0}, r)$ for control observations against $\hat{G}(\underline{x}^{I_1}, r)$ for treatment observations. Across the entire range of distances r we observe $\hat{G}(x^{I_0}, r) \leq \hat{G}(x^{I_1}, r)$, indicating that $\mathbb{P}(\text{nnd}(x_j) < r)$ is greater for treatment observations than control observations, further supporting the previous assertion that microtubules are more tightly packed for observations collected under the treatment regime. It can also be observed that weighted mean estimates of the nearest neighbour function lie below unweighted mean estimates across the entire range of distances r, again suggesting that K-fibers with more microtubules are packed more tightly.

Evidence of clustering and regularity can also be observed in the plots of scaled neighbourhood count functions, which support the interpretation suggested by the nearest neighbour function. Values of $\hat{K}(r)$ lie below their expected theoretical form of πr^2 for small values of r less than approximately 60, indicating regularity. On larger length scales, for values of r greater than approximately 70, values of $\hat{K}(r)$ exceed πr^2 indicating clustering. The scaled neighbourhood

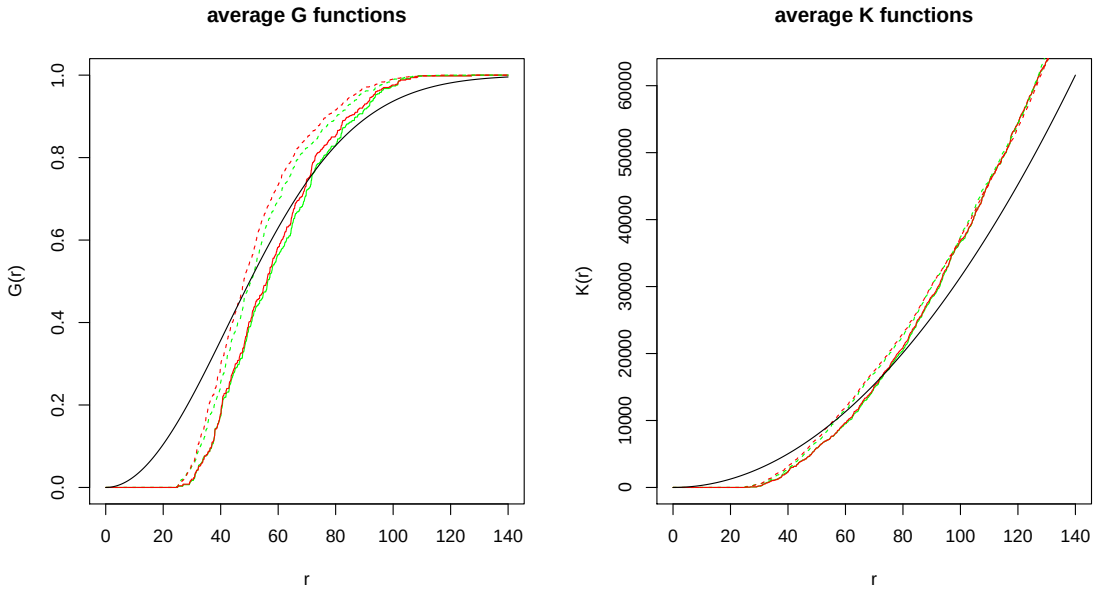

Figure 7: Plots of average nearest neighbour functions, G (left), and scaled neighbourhood count functions, K (right), for TACC3 point patterns. Green lines indicate means, red lines indicate weighted means. Solid lines indicate the average over controls, indexed by I_0, dashed lines indicate average over treatments, indexed by I_1. Black lines are theoretical values for Poisson point processes with intensity ρ^*.

count function is again a cumulative function, making it difficult to accuracy specify the distance of r at which the transition from regularity to clustering occurs.

Figure 7 includes plots of estimates $\bar{K}(\underline{x}^{I_0}, r)$ for control observations, $\bar{K}(\underline{x}^{I_1}, r)$ for treatment observations and the theoretical value πr^2. Over short distances r, $\bar{K}(\underline{x}^{I_0}, r) < \bar{K}(\underline{x}^{I_1}, r)$ indicating that microtubules are more tightly packed when TACC3 is overexpressed even when normalising for differences in point density. The difference between the weighted and unweighted means of estimated scaled neighbourhood count functions is negligible.

Boxplots of the effective force transference for control observations and treatments observations are provided in Figure 8. Across all observations the effective force transference is above 0.9, suggesting that microtubules are generally consistently aligned. The effective force transference is on average greater for control observations, indicating greater agreement in microtubule directions in comparison to treatment observations. Application of the testing methodology is required to assess the significance of the observed difference.

5.5 Permutation testing results

Table 3 summarises the p-values for approximate permutation testing of the TACC3 data using 10^4 permutations for each of the 15 proposed test statistics. In every case other than the scaled neighbourhood count test statistic, δ_K, the difference between $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$ or $\underline{y}^{I_0}$ and $\underline{y}^{I_1}$ is reported as significant at the five percent significance level, with some reported as significant at the 0.5 and even 0.05 percent significance levels.

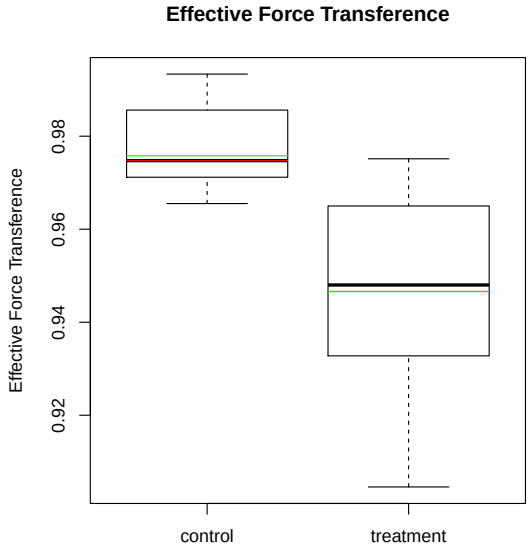

Figure 8: Boxplots of $EFT(\underline{x})$ for TACC3 marked point patterns divided by controls, indexed by I_0, and treatments, indexed by I_1. Horizontal red lines indicates means, horizontal green lines indicate weighted means.

δ_N	0.0005	δ_{nnd}	0.0057	δ_K	0.1092	δ_{EFT}	0.0011
δ_W	0.0018	$\delta_{\text{nnd},\omega}$	0.0005	$\delta_{G,1}$	0.0061	$\delta_{EFT,\omega}$	0.0005
δ_ρ	0.0001	δ_{msd}	0.0019	$\delta_{G,1,\omega}$	0.0005		
$\delta_{\rho,\omega}$	0.0002	$\delta_{\text{msd},\omega}$	0.0005	$\delta_{G,\infty}$	0.0087		
				$\delta_{G,\infty,\omega}$	0.0013		

Table 3: Reported p-values for TACC3 data.

Point pattern size, area and density are significantly greater for $\underline{x}^{I_1}$ than for $\underline{x}^{I_0}$, with little difference between the results for weighted and unweighted intensity statistics. Biologically, this may be taken as evidence that overexpression of TACC3 causes K-fibers to contain significantly greater numbers of microtubules, have significantly larger cross-sectional area and significantly higher density of microtubules than K-fibers under control conditions. The difference between weighted and unweighted density test statistics is negligible.

Mean nearest neighbour distance and mean minimum spanning distance are significantly greater for $\underline{x}^{I_0}$ than for $\underline{x}^{I_1}$. Biologically, this may be taken as evidence that overexpression of TACC3 causes microtubules to be more closely spaced within K-fibers than under control conditions. Weighted mean nearest neighbour distance and mean minimum spanning distance test statistics result in noticeably smaller p-values than their unweighted alternatives.

Tighter packaging of microtubules within K-fibers following overexpression of TACC3 is further supported by all of the nearest neighbour function test statistics. The greater degree of significance for L_1 nearest neighbour distribution test statistics than for L_∞ nearest neighbour distribution test statistics indicates that the average difference between $\hat{G}(\underline{x}^{I_0}, r)$ and $\hat{G}(\underline{x}^{I_1}, r)$ is more significant than the maximum difference. Although both weighted and unweighted test statistics are significant at the five percent level, p-values for weighted test statistics are noticeably smaller.

The scaled neighbourhood count test statistic, δ_K, is the only test statistic which fails to reject the null hypothesis at the five percent level. This result may be explained by the formulation of the scaled neighbourhood count function or the accuracy of its estimation. Failure to reject the null hypothesis for δ_K, combined with rejection of the null hypothesis for density test statistics, δ_ρ and $\delta_{\rho,\omega}$, could be evidence that differences in the generating processes $\underline{X}^0$ and $\underline{X}^1$ are limited to the density of points. A difference in point density would also be expected to result in differences in point separation distances, which have been detected for the TACC3 data. Alternatively, difficulties accurately estimating the scaled neighbourhood count function are caused by the requirement of edge correction and uncertain estimation of the density due to the non-standard specification of the observation windows. These difficulties limit the reliability of results obtained using the scaled neighbourhood count function in comparison to the presented alternatives.

Both the weighted and unweighted effective force transference test statistics result in rejection of the null hypothesis at the upper 0.5 percent level, indicating that the effective force transference is significantly reduced in samples overexpressing TACC3 in comparison to control samples. Biologically, this may be taken as evidence that increased expression of TACC3 causes microtubules in K-fibers to be less well aligned, potentially reducing their capacity to accurately divide chromosomes.

Taken in combination, the results for point patterns $\underline{x}$ indicate that overexpression of TACC3 results in larger K-fibers which are made up of more tightly packed microtubules. The results for marked point patterns $\underline{y}$ indicate further that overexpression of TACC3 results in K-fibers comprised of less well aligned microtubules. The protein TACC3 is believed to influence changes in the microtubule structure indirectly, by altering properties of the mesh structure binding microtubule into K-fibers. On the evidence of our analysis an interpretation may be that increased expression of TACC3 alters the mesh structure by limiting the distance at which it may bind microtubules. Limiting mesh connection distances may require microtubules to be closer and pull microtubules out of a natural alignment, resulting in reduced force transference properties which are then com-

pensated for by increased numbers of microtubules per K-fiber. All of these effects have been observed in the data set studied. Further biological experimentation would clearly be needed to confirm or contradict this hypothesis.

Considering a number of tests applied to the same set of data and the resulting p-values, it is clearly necessary to account for multiple testing. We have proposed 13 test statistics for point data and two statistics for marked point data. A simple, if somewhat conservative, Bonferroni correction (Dunn, 1961) to the five percent significance level would have us reduce the critical p-value to $0.05/13 = 0.0038$ for point patterns and $0.05/2 = 0.025$ for marked point patterns. Observed p-values are less than this adjusted critical value for both marked point pattern tests and nine of the 12 point pattern tests, supporting our assertion that overexpression of TACC3 has a significant impact on microtubule structure within K-fibers.

5.6 Sensitivity analysis

5.6.1 Sensitivity analysis design

The are multiple possibilities for errors when imaging biological samples and reporting microtubule centre coordinate locations. Errors can be minimised, but are difficult to completely remove. As a result, the proposed methodology should be robust to perturbations which are small on a scale determined by the application. To ensure that this is the case we carry out a sensitivity analysis. Each of the perturbations are simulated and tested using each of the appropriate test statistics using 1000 random permutations. Repeating this process 100 times produces 100 p-values for each test statistic.

Relabelling perturbations consider scenarios in which observations are incorrectly assigned to collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$, potentially due to human error. Perturbed index sets are denoted by I_0^q and I_1^q, for $q \in [0, 0.5]$ the probability of incorrect assignment. That is

$$\mathbb{P}(i \in I_0^q) = \begin{cases} 1-q & i \in I_0 \\ q & i \in I_1 \end{cases} \qquad \mathbb{P}(i \in I_1^q) = \begin{cases} q & i \in I_0 \\ 1-q & i \in I_1. \end{cases}$$

Relabelling perturbations are tested over values of $q \in \{0, 0.05, 0.1, 0.2, 0.5\}$ where $q = 0$ leaves collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$ unchanged and $q = 0.5$ results in completely random allocation of point patterns to collections, expected to produce uniformly distributed p-values.

Additional points perturbations consider scenarios in which there exist additional microtubule centres which are not reported, potentially due to poor image quality. For each point pattern $\underline{x}$, a homogeneous Poisson point pattern $\underline{x}'$ is simulated according to $HPPP(\alpha\rho_0, W(\underline{x}))$ and $\underline{x} \cup \underline{x}'$ reported. The parameter value $\rho_0 = 10^{-4}$ is an approximation to the density of TACC3 point patterns. For each marked point pattern $\underline{y}$, a marked point pattern $\underline{y}'$ is simulated according to $MPP(n, u, \phi)$ and $\underline{y} \cup \underline{y}'$ reported. The number of marked points n is simulated as $\text{po}(30\alpha)$, with reference direction $u = \beta$ and maximum angle $\phi = 20$ degrees. The parameter α is varied over the set $\{0.2, 1, 3\}$ to represent increases in the number of points of approximately 20 percent, 100 percent and 300 percent. As the additional points in $\underline{x}'$ and $\underline{y}'$ are identically distributed across I_0 and I_1, increasing values of α are expected to reduce the reported significance of differences between the collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$.

Point location perturbations consider scenarios in which the location of microtubule centres are

	Relabelling, α					Additional points, α			Point location, ϵ	
	0	0.05	0.1	0.2	0.5	0.2	1	3	5	20
δ_N	1.00	0.96	0.79	0.50	0.06‡	1.00	1.00	1.00	1.00	1.00
δ_W	1.00	0.92	0.73	0.44	0.06‡	1.00	1.00	1.00	1.00	1.00
δ_ρ	1.00	1.00	0.89	0.60	0.05‡	1.00	0.96	0.90	1.00	1.00
$\delta_{\rho,\omega}$	1.00	0.92	0.72	0.46	0.04‡	1.00	1.00	0.99	1.00	1.00
δ_{nnd}	1.00	0.86	0.68	0.34	0.05‡	0.68	0.36	0.23	1.00	0.99
$\delta_{\text{nnd},\omega}$	1.00	0.90	0.73	0.39	0.06‡	0.96	0.52	0.28	1.00	1.00
δ_{msd}	1.00	0.94	0.80	0.39	0.06‡	0.93	0.60	0.37	1.00	1.00
$\delta_{\text{msd},\omega}$	1.00	0.93	0.77	0.41	0.04‡	1.00	0.77	0.54	1.00	1.00
δ_K	0.00	0.04	0.15	0.11	0.08‡	0.02	0.04	0.51	0.00	0.03
$\delta_{G,1}$	1.00	0.83	0.66	0.33	0.05‡	0.74	0.38	0.23	1.00	1.00
$\delta_{G,1,\omega}$	1.00	0.90	0.73	0.40	0.06‡	0.99	0.64	0.30	1.00	1.00
$\delta_{G,\infty}$	1.00	0.79	0.59	0.27	0.05‡	0.66	0.24	0.12	1.00	0.89
$\delta_{G,\infty,\omega}$	1.00	0.89	0.64	0.38	0.08‡	0.98	0.46	0.17	1.00	1.00
δ_{EFT}	1.00	0.95	0.82	0.39	0.06‡	1.00	0.87	0.05‡	1.00	1.00
$\delta_{EFT,\omega}$	1.00	0.87	0.69	0.32	0.04‡	1.00	0.81	0.05‡	1.00	1.00

Table 4: Proportion of p-values in the range $[0, 0.05]$, ‡ indicates failure to reject uniformity of observed p-values under the Kolmogorov-Smirnov test at the Bonferroni corrected (Dunn, 1961) $5/13 = 0.38$ (for point patterns) or $5/2 = 2.5$ (for marked point patterns) percent significance level.

recorded with some degree of error. Point locations x_j ($x_{j,1}$ in the case of marked point patterns) are each shifted in a random direction by a distance uniformly sampled between zero and ϵ. The parameter ϵ takes values in the set $\{5, 20\}$, in comparison to the typical microtubule radius of 12.5.

5.6.2 Sensitivity analysis results

A summary of the 100 p-values obtained for each test statistic under each perturbation is provided by Table 4.

For relabelling perturbations the proportion of p-values reported significant at the five percent level reduces with increasing probability of mislabelling, α. The impact when $\alpha = 0.05$ is generally small, indicating that minor errors in allocation to collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$ do not render the results useless. For $\alpha = 0.5$ a Kolmogorov-Smirnov test at the five percent significance level reports insufficient evidence to reject the uniform distribution of p-values on the interval $[0, 1]$ across all test statistics, as expected. Increasing values of α for point patterns $\underline{x}$ appear to have a greater impact when unweighted test statistics are used, while the converse is true for marked point patterns $\underline{y}$.

In the case of additional points perturbations, the number of p-values reported as significant at the five percent level reduces as the number of additional points increases with α. The impact when $\alpha = 0.2$ is generally small, indicating that the methodology is robust so long as a large proportion of points are identified. First order statistics report very little change as more points are

added because original and perturbed observation window areas and number of points are strongly correlated. Increasing values of α again have a greater impact on unweighted test statistics for point patterns, $\underline{x}$.

Point location perturbations have a negligible impact on the proportion of p-values reported significant at the five percent level, indicating that the methodology is robust to small errors in point locations. For first order statistics this behaviour is expected as they are calculated independently of point locations. While perturbations to point locations do impact the value of second order summary statistics, the random perturbations do not make collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$ more similar and thus the difference between them remains consistently significant.

6 Conclusions

This report presents a number of test statistics for the comparison of point patterns and marked point patterns of a particular form based upon existing and novel summary statistics. The suitability of each for a variety of scenarios has been highlighted by application to simulated data sets, the results of which indicate that the methodology is adept at detecting minor differences, smaller than those which may be detected by eye. Application to a biological data set has provided positive results, indicating a significant difference between control and treatment observations which matches and extends a previous analysis of the same data. A sensitivity analysis further reports that the testing procedure is not unduly affected by variation in the data on a scale consistent with measurement error.

We have proposed a total of 13 test statistics for application to 2D point pattern data and two further statistics for application to marked point pattern data for the comparison of collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$ assumed to represent independent realisations of point processes $\underline{X}^0$ and $\underline{X}^1$. Alongside these statistics we propose the use of a nonparametric permutation hypothesis testing procedure to determine the p-value of the proposed test statistic for a given collection of observed data $\underline{x}^I$ under the null hypothesis that $\underline{X}^0$ and $\underline{X}^1$ are identically distributed.

The test statistics are designed to test for differences in various features of the point patterns under comparison, for example the number of points or the average distance between nearest neighbouring points. The particular feature tested will depend upon the data set being analysed and the questions of interest, but our suggestions cover a variety. Test statistics also differ in the format of the comparison, for example the difference between average nearest neighbour distances or through comparisons between functions which quantify nearest neighbour distances. Some of the proposed comparisons are more straightforward to present and interpret, while others are more sensitive to differences between the collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$.

A detailed simulation study has been used to highlight the various scenarios in which each of the proposed test statistics are effective. Manipulating the size of the difference between simulated collections $\underline{x}^{I_0}$ and $\underline{x}^{I_1}$ allows the sensitivity of each of the proposed approaches to be quantified, the results of which support the use of the proposed techniques to be able to detect even minor differences between collections. In the case of no difference between simulated data sets, there is insufficient evidence to reject the uniform distribution of p-values on the interval $[0, 1]$, indicating that the testing approach has good specificity.

The proposed methodologies are also applied to a collection of real biological data in which

points represent microtubules under control, $\underline{x}^{I_0}$, and treatment, $\underline{x}^{I_1}$, conditions where TACC3 has been overexpressed. For this data set, investigation in particular of the distances between points is of interest as it is understood that nearby microtubules are bound together in K-fibers by a mesh-like structure and that TACC3 may have an impact on this mesh. Investigation by eye is insufficient for this application as the data sets appear visually very similar, so we apply the proposed methodology.

Testing of the TACC3 data using all of the proposed test statistics results in a number of rejections of the null hypothesis. Summarising these results, we learn that treatment K-fibers contain significantly more microtubules, the constituent microtubules are more tightly packed and variation in the orientation of microtubules is greater for treatment K-fibers. A possible explanation for this may be that overexpression of TACC3 limits the distance over which the mesh may connect microtubules, bringing them closer together and pulling them out of line with one another. Due to the vital importance of K-fibers during mitosis, the results obtained by the analysis support and suggest useful directions for more targeted investigation of the impact of TACC3.

The TACC3 data analysed in this report is also analysed by Nixon et al. (2015), for which we had some input. There is an agreement in results between our works, although ours extends further by introducing test statistics based upon different features and with different measures of comparison. The largest difference between our work comes in the testing procedure, with Nixon et al. (2015) quantifying the significance of observed differences using t-tests reliant upon the assumption of normality, while we carry out permutation tests that require no similar assumptions. A significant contribution of our work is to consider the analysis of marked point pattern data to make comment on the 3D structure of microtubules within K-fibers, with results indicating that TACC3 has a significant effect on this structure.

The results of an in-depth sensitivity analysis of the proposed methodology suggest it to be robust to the small errors in data recording which are expected to be most likely. This is particularly valuable for our application, as images are produced through a human controlled imaging process from which point locations are manually reported, with the possibility for errors at each step.

References

A. J. Baddeley and B. W. Silverman. A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics*, 40(4):1089, dec 1984. doi: 10.2307/2531159.

D. G. Booth, F. E. Hood, I. A. Prior, and S. J. Royle. A TACC3/ch-TOG/clathrin complex stabilises kinetochore fibres by inter-microtubule bridging. *The EMBO Journal*, 30(5):906–919, feb 2011. doi: 10.1038/emboj.2011.15.

D. R. Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 17(2):129–164, 1955. ISSN 00359246. URL `http://www.jstor.org/stable/2983950`.

P. Diggle. A kernel method for smoothing point process data. *Applied Statistics*, 34(2):138, 1985. doi: 10.2307/2347366.

P. J. Diggle. *Statistical Analysis of Spatial Point Patterns*. Hodder Education Publishers, 2003. ISBN 978-0340740705.

P. J. Diggle. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Taylor & Francis Inc, 2013. ISBN 1466560231.

P. J. Diggle, N. Lange, and F. M. Benes. Analysis of variance for replicated spatial point patterns in clinical neuroanatomy. *Journal of the American Statistical Association*, 86(415):618–625, 1991. ISSN 01621459. URL http://0-www.jstor.org.pugwash.lib.warwick.ac.uk/stable/2290390.

P. J. Diggle, J. Mateu, and H. E. Clough. A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns. *Advances in Applied Probability*, 32(02):331–343, jun 2000. doi: 10.1017/s0001867800009952.

O. J. Dunn. Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64, mar 1961. doi: 10.1080/01621459.1961.10482090.

E. S. Edgington. Randomization tests. *The Journal of Psychology*, 57(2):445–449, apr 1964. doi: 10.1080/00223980.1964.9916711.

E. S. Edgington. Approximate randomization tests. *The Journal of Psychology*, 72(2):143–149, jul 1969. doi: 10.1080/00223980.1969.10543491. URL http://dx.doi.org/10.1080/00223980.1969.10543491.

C. Gaetan and X. Guyon. *Spatial Statistics and Modeling*. Springer New York, 2009. ISBN 978-0-387-92257-7.

E. Grochowski and R. Hoyt. Future trends in hard disk drives. *IEEE Transactions on Magnetics*, 32(3):1850–1854, may 1996. doi: 10.1109/20.492876.

A. J. Holland and D. W. Cleveland. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nature Reviews Molecular Cell Biology*, 10(7):478–487, jul 2009. doi: 10.1038/nrm2718.

K.-H. Jöckel. Finite sample properties and asymptotic efficiency of monte carlo tests. *The Annals of Statistics*, 14(1):336–347, 1986. ISSN 00905364. URL http://www.jstor.org/stable/2241285.

O. Kanoun and H.-R. Trankler. Sensor technology advances and future trends. *IEEE Transactions on Instrumentation and Measurement*, 53(6):1497–1501, dec 2004. doi: 10.1109/tim.2004.834613.

J. F. C. Kingman. *Poisson Processes (Oxford Studies in Probability)*. Clarendon Press, 1993. ISBN 9780198536932.

F. J. Massey. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46(253):68–78, mar 1951. doi: 10.1080/01621459.1951.10500769. URL http://dx.doi.org/10.1080/01621459.1951.10500769.

J. Mateu, F. P. Schoenberg, D. M. Diez, J. A. González, and W. Lu. On measures of dissimilarity between point patterns: Classification based on prototypes and multidimensional scaling. *Biometrical Journal*, 57(2):340–358, dec 2014. doi: 10.1002/bimj.201300150.

J. Møller. Shot noise cox processes. *Advances in Applied Probability*, 35(03):614–640, sep 2003. doi: 10.1017/s0001867800012465.

J. Neyman and E. L. Scott. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(1):1–43, 1958. ISSN 00359246. URL http://www.jstor.org/stable/2983905.

F. M. Nixon, C. Gutiérrez-Caballero, F. E. Hood, D. G. Booth, I. A. Prior, and S. J. Royle. The mesh is a network of microtubule connectors that stabilizes individual kinetochore fibers of the mitotic spindle. *eLife*, 4, jun 2015. doi: 10.7554/elife.07635. URL http://dx.doi.org/10.7554/eLife.07635.

B. D. Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(2):172–212, 1977. ISSN 00359246. URL http://www.jstor.org/stable/2984796.

M. N. M. van Lieshout. *Markov Point Processes and Their Applications*. World Scientific Pub Co Pte Lt, jul 2000. doi: 10.1142/p060.