INFERENCE AND VISUALIZATION FOR SPATIAL POINT PROCESSES

A Dissertation

by

NGOC ANH DAO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Huiyan Sang
Co-Chairs of Committee,	Marc G. Genton
Committee Members,	Michael T. Longnecker
	Ramalingam Saravanan
Head of Department,	Brani Vidakovic

December 2021

Major Subject: Statistics

ABSTRACT

Spatial point processes are statistical models that describe the arrangement of objects that are randomly distributed in the plane or in space. In recent years, they have received sustained attention because researchers use them to model objects in ecology, biology, medicine and material science, to name a few. Inevitably, a goodness-of-fit test is needed to assess the fit of these models and to justify their choice. In this thesis, I propose a method, consisting of nested Monte Carlo simulations, which removes the bias of the resulting empirical level of the test. As a further contribution to statistical inference for the spatial point processes in this thesis, I introduce skew-elliptical cluster processes, where the clusters can have an anisotropic structure allowing the choice of a flexible covariance matrix and incorporating skewness or ellipticity parameters into the structure. Theses processes help to tackle the challenge arising with non-circular clusters, e.g., induced by a wind direction in the pattern. In particular, I formulate the construction of skew-elliptical-normal and skew-elliptical-t cluster processes. For the parameter estimation, I propose the minimum contrast method using an approximating pair correlation function to circumvent the complicated derivation of the maximum- or pseudo-likelihood and the computational complexity of the Bayesian approach or MCMC algorithm. The last contribution in this thesis is in diagnostics and influential measures for spatial point processes. I describe a method to define influential events of a spatial point pattern based on a parametric likelihood model or a second-order summary characteristic function if the likelihood model is difficult to derive. In particular, instead of deleting one observation/event at a time like in commonly-used approaches in detecting influential events, I add some noise to one event at a time. The perturbation provides a whole course of change of estimators based on which I quantify the influence. To visualize influential events, I use hair-plots and disc-plots to display the influence of each event. Those events with significantly high magnitude of influence can be considered as influential.

DEDICATION

To Simon, Johann, Tobias & Elisabeth, my mother Thu Hiền and my father Tám.

ACKNOWLEDGMENTS

My first ‘thank you’ goes to my advisor, Dr. Marc Genton for his support and belief in my abilities. His mentorship and encouragement helped me persevere through the various points in which I was tempted to forgo pursuit of my Ph.D.. I appreciate the knowledge and insights he provided and the role model served as a Statistics Professor.

I want to thank my husband, Simon, for his support, understanding and tolerance of the time and effort I have expended pursuing my degree, often at the short-term expense of family. I want to thank my children, Johann, Tobias and Elisabeth who have inspired me to become a better person; in particular, with regard to completing worthwhile pursuits, once started. At last, following a ten year break from my graduate study, my Ph.D. is soon to be achieved.

Lastly, I want to thank my mother, Thu Hiền, and my father, Tám. My parents have always given me the feeling that I am special to them. Because of their unfailing belief in me, I believe I can achieve any high goal through hard work and persistence. Because of them, I started my Ph.D. and because of them, I now want to complete it.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Huiyan Sang [chair], Professor Michael T. Longnecker of the Department of Statistics, Professor Ramalingam Saravanan of the Department of Atmospheric Sciences of the Texas A&M University and Professor Marc G. Genton [co-chair and dissertation-advisor] of CEMSE Division, King Abdullah University of Science and Technology (KAUST), Saudi Arabia.

Dr. Naisyin Wang suggested the topic of Chapter 2.

The data analyzed for Chapters 2, 3 and 4 were provided by `R`-package `spatstat`.

The former Texas A&M University Brazos HPC cluster contributed to the computation carried out in Chapter 2.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University. In particular, Chapter 2 is based in part on work supported by award no. KUS-C1-016-04 made by KAUST and by NSF grants DMS-1007504 and DMS-1106494.

NOMENCLATURE

AGOF	adjusted goodness-of-fit
AML	approximate maximum likelihood
cdf	cumulative distribution function
CP	cluster process
CSR	complete spatial randomness
df	degrees of freedom
EST	extended skew-t
GOF	goodness-of-fit
MCM	minimum contrast method
MPL	maximum pseudo-likelihood
MPPL	maximum profile pseudo-likelihood
pcf	pair correlation function
PGOF	plug-in Monte Carlo GOF test
pdf	probability density function
SAUC	squared areas under the curve
SPP	spatial point pattern
SUE	unified skew-elliptical
SUN	unified skew-normal
TP	Thomas process

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION AND LITERATURE REVIEW

A spatial point process is a random pattern of points in d-dimensional space where usually $d = 2$ or $d = 3$ in applications. Spatial point processes are useful as statistical models in the analysis of observed patterns of points, where the points represent the locations of some object of study, e.g., locations of trees or bird nests (statistical ecology), the positions of stars and galaxies (astrostatistics), the locations of point-like defects in a silicon crystal wafer (materials science), petty crimes (social science), the locations of neurons in brain tissue (medical science), the home addresses of individuals diagnosed with a rare disease (spatial epidemiology), or locations of active agents of some drugs coupled to antigen sites [Illian et al., 2008]. To evaluate whether a certain spatial point process deems to be a good fit for a spatial point pattern (SPP), a goodness-of-fit (GOF) test is needed. Introduced by Ripley [1977] and later widely applied in statistical ecology, the envelop test - a Monte Carlo test [Barnard, 1963] - is based on computing a summary function of the SPP, such as Ripley's K-function [Ripley, 1976], and comparing it with the envelope of the same functions obtained from several simulations of the null model. Diggle [2003] and Baddeley et al. [2014] noted that such a test works correctly if the null model is simple or the null model is a complete spatial randomness. However, it is conservative if the null hypothesis is composite or the parameters are estimated from the observed SPP. To overcome this shortcoming, I propose a test in Chapter 2 that adjusts for the bias of the estimated p-value and achieves higher power through nested Monte Carlo simulations. This Monte Carlo adjusted GOF test has now already been adopted under the name *Dao-Genton* test [Dao and Genton, 2014] in the spatial point process community [Baddeley et al., 2017, 2020]. Chapter 2 of this thesis is a complete re-print of [Dao and Genton, 2014] with permission being requested from Taylor & Francis Journal. Here, the correctness of this test is proven and the test itself is evaluated via simulations for an inhibitory, completely random, and clustered process. Then it is applied to a Phlebocarya filifolia plants dataset [Illian et al., 2008] to study the propagation mechanism and to a micro-anatomy dataset to learn about the developmental growth of immature retina cells in rabbits [Diggle, 2003].

Beyond elaborating the GOF test in this thesis, I also make contributions in statistical inference for spatial point processes. Chapter 3 is a complete re-print of [Dao and Genton, 2021] with permission from Copyright Clearance Center's RightsLink. Here, I introduce the skew-elliptical-normal cluster processes (CP) and the skew-elliptical-t CP [Dao and Genton, 2021]. Their introduction is motivated by the challenge in modeling datasets which have clusters of non-circular pattern. In nature, wind or the slope of a location can cause the positions of the cluster events to be skew-elliptical distributed. For example, if wind usually blows in a particular direction, it would carry plant seeds in this direction and plants would propagate according to the assistance of this wind direction. Hence the shape of that particular forest area would not be circular but non-circular, for example skewed or elliptical. Although patterns having skew-elliptical clusters occur quite often in the nature, statistical models for skewed-elliptical CP are not yet common. To tackle this challenge, I introduce a generalization of the Thomas process (TP) [Thomas, 1949]. The application of TP is widely seen in the field of spatial point processes because it has the intrinsic statistical ability to model propagation or clustering in nature. However, the limitation of TP lies in its ability of modeling only circular clusters. That means characteristics like ellipticity and skewness of clusters cannot be modeled. To obtain more flexibility, I impose a unified skew-elliptical distribution [Arellano-Valle and Genton, 2010a] on the distribution of the cluster point/event, whereas in the TP case, the distribution of the cluster event is a bivariate normal distribution with the (circular) covariance matrix as a diagonal matrix with the diagonal entries $\sigma_{11}^2 = \sigma_{22}^2 = \sigma^2$. The clustered redwoods dataset representing the locations of 195 Californian red seedlings shows pattern of non-circular clusters. This dataset was first described and analyzed by Strauss [1975] and later reapplied in many works in spatial point processes [Diggle, 2003, Illian et al., 2008]. However, according to my knowledge, modeling this non-circular characteristic has not been addressed in the literature of spatial point processes. So I apply a skew-elliptical normal CP and a range of skew-elliptical t CPs to this dataset. The GOF test confirms that an elliptical normal CP among the aforementioned models provides the best fit and the conclusion supports the visual observation that the clusters are not circular.

In the diagnostic data analysis for spatial point processes, I introduce a method in Chapter 4 which defines and visualizes influential points/events of a SPP. For a given estimator, I say that a point or an event is influential whenever a change of its position in the pattern leads to a radical change in the estimate or in the estimate of a functional, e.g., second-order summary characteristic such as F-, G-, K- and pair correlation functions [Diggle, 2003, Illian et al., 2008]. There are techniques for residual analysis [Baddeley et al., 2005], detecting outliers, criticizing model-based outliers [Illian et al., 2008], identifying leverage and influential points [Baddeley et al., 2013]. Especially, in the latter work, the authors established measures of leverage and influence for the dependence of a point process model on covariates by direct analogy with standard techniques for generalized linear models. The technique developed by Baddeley et al. [2013] requires, however, the likelihood or composite likelihood to be formally equivalent to a Poisson likelihood and that terms in the model must be available in closed form. This excludes a wide range of models including, for example, Cox processes, Neyman-Scott cluster processes and other models which are effectively hierarchical or mixed Poisson models.

In this thesis, I introduce a concept of detecting influential events of a SPP, which does not require the maximum likelihood or pseudo-likelihood in closed form. In particular, I let a second-order summary characteristic such as F-, G-function [Diggle, 2003, Sec. 4.3], Ripley's K-function [Ripley, 1976], and pair correlation function [Illian et al., 2008] to describe the SPP and define a measure of departure based on the second-order summary characteristic of one's choice. This approach allows many more spatial point processes to be studied.

In the setting of generalized linear models, the standard method of identifying influential observations is to quantify the radical change of the statistical inference when carrying out statistical analysis without that observation. In my opinion, this "deleting" method is only justified for the cases underlying the assumption that the observations are independent and identically distributed (iid). In the field of spatial point processes, Baddeley et al. [2013, 2019] still kept this "deleting" method to study the leverage and influence of the data. However, in the case of dependent data such as in spatial statistics or spatial point processes, deleting events may not be a sensible

practice due to the dependence structure of the events. As noted by Cook and Weisberg [1994], deleting cases is only a way of introducing small changes in the data and there are others. As in Genton and Ruiz-Gazen [2010], I introduce some noise to the data by perturbing additively one event at a time. Unlike deleting a point/event, the *perturbing* method provides a whole course of change of estimators or conclusions on GOF due to a set of possible amounts of perturbation. I adopt and extend the approach by Genton and Ruiz-Gazen [2010] to develop a method to detect influential events of SPPs. Via perturbing events and defining discrepancy between the perturbed and observed functionals, I define influential events and use graphical tools such as hair-plots and disc-plots to visualize them. The method is applied to a Queensland copper cores dataset [Berman, 1986] and to the Swedish pines dataset [Strand, 1972] to study influential events which might yield radical changes in statistical inference.

Finally, I draw conclusions on the contributions of this thesis and discuss possible future work directions for each topic in Chapter 5.

2. A MONTE CARLO-ADJUSTED GOODNESS-OF-FIT TEST FOR PARAMETRIC MODELS DESCRIBING SPATIAL POINT PATTERNS*

2.1 Chapter Overview

Spatial point processes have received sustained attention in recent years in many disciplines such as ecology, biology, medicine and material science, to name a few. For example, using spatial point processes, ecologists model the propagation mechanism of forests, biologists model the developmental growth of immature cells, and pharmacologists model the spatial structure of active agents of some drugs coupled to antigen sites [Illian et al., 2008]. Inevitably, a goodness-of-fit (GOF) test is needed to assess the fit of these models. Up to now, GOF tests were dominantly available for assessing the fits of simple Poisson processes. A large number of testing methods for assessing homogeneous Poisson processes, otherwise known as complete spatial randomness (CSR), are referenced and described in Cressie [1993, Sec. 8] and Diggle [2003, Sec. 2]. Notable examples include a test based on the approximate distribution of the mean of the nearest neighbor distances [Donnelly, 1978] and quadrat count tests [Diggle, 2003, Sec. 2.5]. For sparsely sampled point patterns, Besag and Gleaves [1973] and Hines and Hines [1979] developed distance-based tests, whereas Assunção [1994a] and Assunção and Reis [2000] developed an angle-based test. When the fitted model is an inhomogeneous Poisson process, Guan [2008] derived a GOF test with a statistic based on a discrepancy measure function that is constructed from residuals obtained from the fitted model. The test statistic has a limiting standard normal distribution, so the test can be performed by comparing the test statistic with available critical values. While the aforementioned methods are suitable for Poisson spatial point processes, they do not extend trivially to the fits of more complicated processes. An exploratory analysis of such data involves a diagnostic assessment of the residuals that can be handled by the graphical methods of Baddeley et al. [2005] and Baddeley et al. [2008]. This diagnostic approach, however, does not serve as a formal GOF

*Reprinted with permission being requested from "A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns" by Ngoc Anh Dao and Marc G. Genton, 2014. *Journal of Computational and Graphical Statistics*, 23:2, pages 497 - 517, Copyright 2021 by Taylor & Francis Academic Journals.

test. A possible complement is the Monte Carlo GOF test [Ripley, 1988, Diggle, 2003], which is valid when, for example, the data consist of multiple spatial point patterns. Analogous to training and validation datasets in regression analysis, one can obtain the parameter estimate for the null model from one, or from some, pattern(s) and apply the GOF test to the remaining patterns. This traditional Monte Carlo GOF test and its graphical version, the envelope test, are described in Sec. 2.2.2.

When the data consist of a single spatial point pattern (SPP), the parameter estimate has to come from the only pattern available, which suggests the concept of *the plug-in Monte Carlo GOF test* (PGOF), which is widely applied in the literature. However, in the general statistical context, Robins et al. [2000] noted that the p-value resulting from the PGOF test does not necessarily follow the standard uniform distribution, not even in the asymptotic sense. In particular, in the context of spatial point processes, Diggle [2003, Sec. 6.2] noted that "Such tests are strictly invalid, and probably conservative, if parameters have been estimated from the data. To some extent, this problem can be alleviated if I use a goodness-of-fit statistic which is only loosely related to the estimation procedure". Similarly, in the general statistical context, Bayarri and Berger [2000] suggested that if the distribution of the GOF test statistic is conditioned upon a sufficient statistic, then this distribution does not depend on the true parameter and therefore the p-value calculated with respect to this conditional distribution has a standard uniform distribution. However, in the context of spatial point processes, except for the homogeneous Poisson process, one cannot find closed forms for sufficient statistics. Hence, the approach motivated by Bayarri and Berger [2000] cannot be applied. Diggle [2003, p. 10] also noted that "An inherent weakness of the Monte Carlo approach is its restriction to simple hypotheses[...]. Composite hypotheses can be tested if pseudo-random sampling is made conditional on the observed values of sufficient statistics for any unknown parameters, but this is seldom practicable. Note that a goodness-of-fit test which ignores the effects of estimating parameters will tend to be conservative". Diggle [2003, p. 10] noted further that this particular conservativeness does not arise with tests of CSR for mapped data, because the observed number of events, N, is sufficient for the intensity, λ, and, conditional on

N, CSR is a simple hypothesis. But ignoring the effect of estimating parameters does affect the assessment of goodness-of-fit for more general stochastic models. Figure 2.1 shows how far off the empirical level can be from the nominal level, $\alpha = 0.05$. The histograms show from left to right the distributions of the estimated p-values from the PGOF test assessing the fits of Strauss, CSR, and Thomas processes, which are described in Sec. 2.3.1. CSR is a simple hypothesis; hence, the distribution of the estimated p-values is standard uniform. However, by letting $\mathbb{X}$ denote the process underlying the observed spatial point pattern defined in Sec. 2.2.1, the Strauss process ($H_0 : \mathbb{X} \sim f_{\beta,\gamma,R}$, where f denotes the probability density function) and the Thomas process ($H_0 : K(\mathbb{X}) = K_{\kappa,\sigma,\mu}$, where K denotes the K-function) are composite hypotheses because there is no knowledge about the existence or the form of sufficient statistics. The distributions of the estimated p-values are not standard uniform. The PGOF test is too conservative for these two processes. It is important to correct for the bias of its empirical level because if the type-II error rate were pre-specified, the PGOF test would suffer from power loss as noted by Marriott [1979]: The Monte Carlo test gives a "blurred" critical region, which is not of the usual form. A conventional significance test would reject H_0 if the statistic lays in some well-defined critical region. However, for the Monte Carlo test, there is a range of values of the statistic over which there is a varying probability of rejecting H_0. This blurring of the critical region leads to a loss of power. This weakness of the PGOF test still remains when using the envelope test [Ripley, 1977]. If the Monte Carlo simulation is not made conditional on an observed value of a sufficient statistic, there is an undesirably high probability that the upper and lower envelopes of the null model would contain the empirical summary or second-order characteristic function of the observed SPP, which does not underlie the null model. To overcome the weaknesses with respect to the conservativeness and power loss of PGOF tests, I propose a test that adjusts for the bias of the empirical level through nested Monte Carlo simulations to reach the nominal level correctly.

The remainder of this chapter is organized as follows. Sec. 2.2 describes the methodology of the proposed GOF procedure. Extensive simulation studies demonstrate in Sec. 2.3 that my method gives more accurate empirical levels than those provided by the existing PGOF test. Sec. 2.4

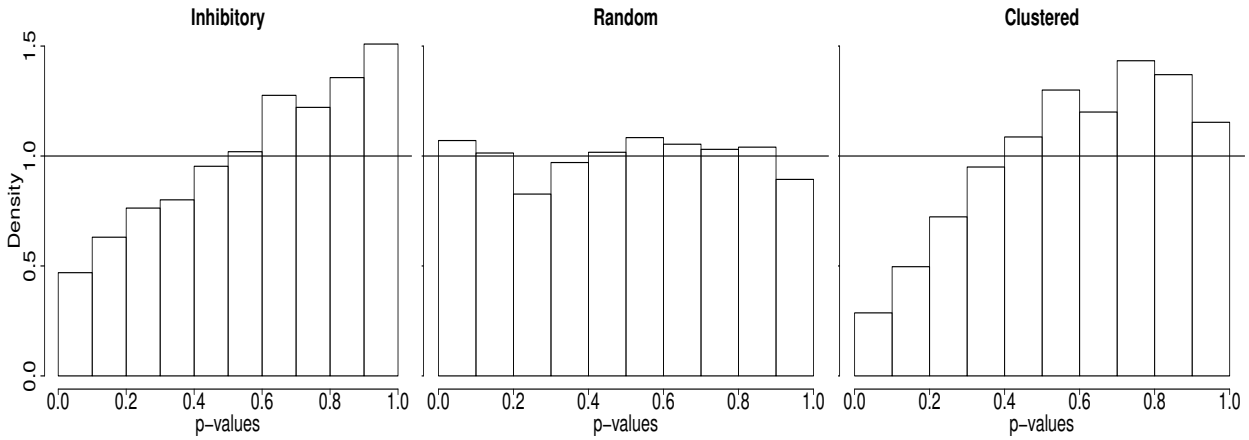


Figure 2.1: Distributions of the estimated p-values $\hat{p}$'s resulting from PGOF tests. The Strauss processes (left), CSR (center), Thomas processes (right) are defined in Sec. 2.3.1. The nearest neighbor distance distribution function, G-function, defined in Sec. 2.2.2, is used to compute the test statistic for the PGOF tests. The horizontal solid lines stand for the probability density function of the standard uniform distribution.*

provides two data applications, one to a forest dataset to study the propagation mechanism of Phlebocarya filifolia plants, and the other to micro-anatomy data to learn about the developmental growth of immature retina cells in rabbits. Finally, Sec. 2.5 discusses two computational techniques that make the computations more efficient.

2.2 Monte Carlo Goodness-of-fit Tests

2.2.1 Definitions

Hereafter, I adopt the notation given in Baddeley et al. [2005, Sec. 5]. An SPP is a dataset, $X = \{x_1, \ldots, x_N\}$, where the x_i's are unordered locations observed in a bounded region, W of $\mathbb{R}^2$. I let f_θ denote the parametric model (a parametric spatial point process) fitted to X, where θ is an arbitrary finite-dimensional vector of parameters. I assume that $f_\theta(X)$ is a probability density function with respect to the unit rate Poisson process on the window W, such that f_θ satisfies the positivity condition: if $f_\theta(X) > 0$ and $Y \subset X$, then $f_\theta(Y) > 0$ for any finite point patterns,

*Reprinted with permission being requested from "A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns" by Ngoc Anh Dao and Marc G. Genton, 2014. *Journal of Computational and Graphical Statistics*, 23:2, pages 497 - 517, Copyright 2021 by Taylor & Francis Academic Journals.

$X, Y \subset W$. Under this setup, I am interested in testing the null hypothesis:

$$H_0 : \quad \mathbb{X} \sim f_\theta, \tag{2.1}$$

where $\mathbb{X}$ is the spatial point process from which the observed SPP, X, is generated. The common p-value for the hypothesis above is $p = 1 - \Pr\{t(\mathbb{X}) < t(X)\}$ for some test statistic, $t(\cdot)$, i.e., this definition of the p-value is based on the assumption that a small value of $t(X)$ supports the hull hypothesis, H_0. If the probability density of $t(\cdot)$ is tractable, then a Monte Carlo GOF test is not necessary because the computation of the p-value is with respect to the probability density of $t(\cdot)$, and the distribution of the p-value is exactly standard uniform for a known θ, and approximately standard uniform for an unknown θ under some conditions [Bayarri and Berger, 2000]. When, however, the probability density of $t(\cdot)$ is analytically unavailable, a Monte Carlo GOF test is an alternative to the classical methods to estimate the p-value to assess the fit.

2.2.2 A Traditional Monte Carlo Goodness-of-Fit Test

As a basis for assessing (2.1), I first briefly describe the test from Diggle [2003, Sec. 6], who proposed a test statistic, u, and the pseudo-statistics, u_i, using the summary characteristic G-function, which will be described after introducing the estimated p-value in (2.5):

$$u = \int_0^\infty \left\{\hat{G}(h) - \overline{G}(h)\right\}^2 dh; \quad u_i = \int_0^\infty \left\{\hat{G}_i(h) - \overline{G}_i(h)\right\}^2 dh, \; i = 1, \ldots, n, \tag{2.2}$$

where $\hat{G}(h)$ is the empirical G-function of the observed SPP, X, $\hat{G}_i(h)$ is the empirical G-function of the simulated SPP, X_i, $i = 1, \ldots, n$, under (2.1) with $\theta = \hat{\theta}$ *for some* $\hat{\theta}$ *not necessarily from the data*, and n is the number of Monte Carlo simulations. Also,

$$\overline{G}(h) = \frac{1}{n}\sum_{i=1}^{n} \hat{G}_i(h); \quad \overline{G}_i(h) = \frac{1}{n}\left\{\hat{G}(h) + \sum_{l=1, l\neq i}^{n} \hat{G}_l(h)\right\}, \; i = 1, \ldots, n. \tag{2.3}$$

For a graphical depiction of the Monte Carlo GOF test under the plug-in fashion, see the "Plug-

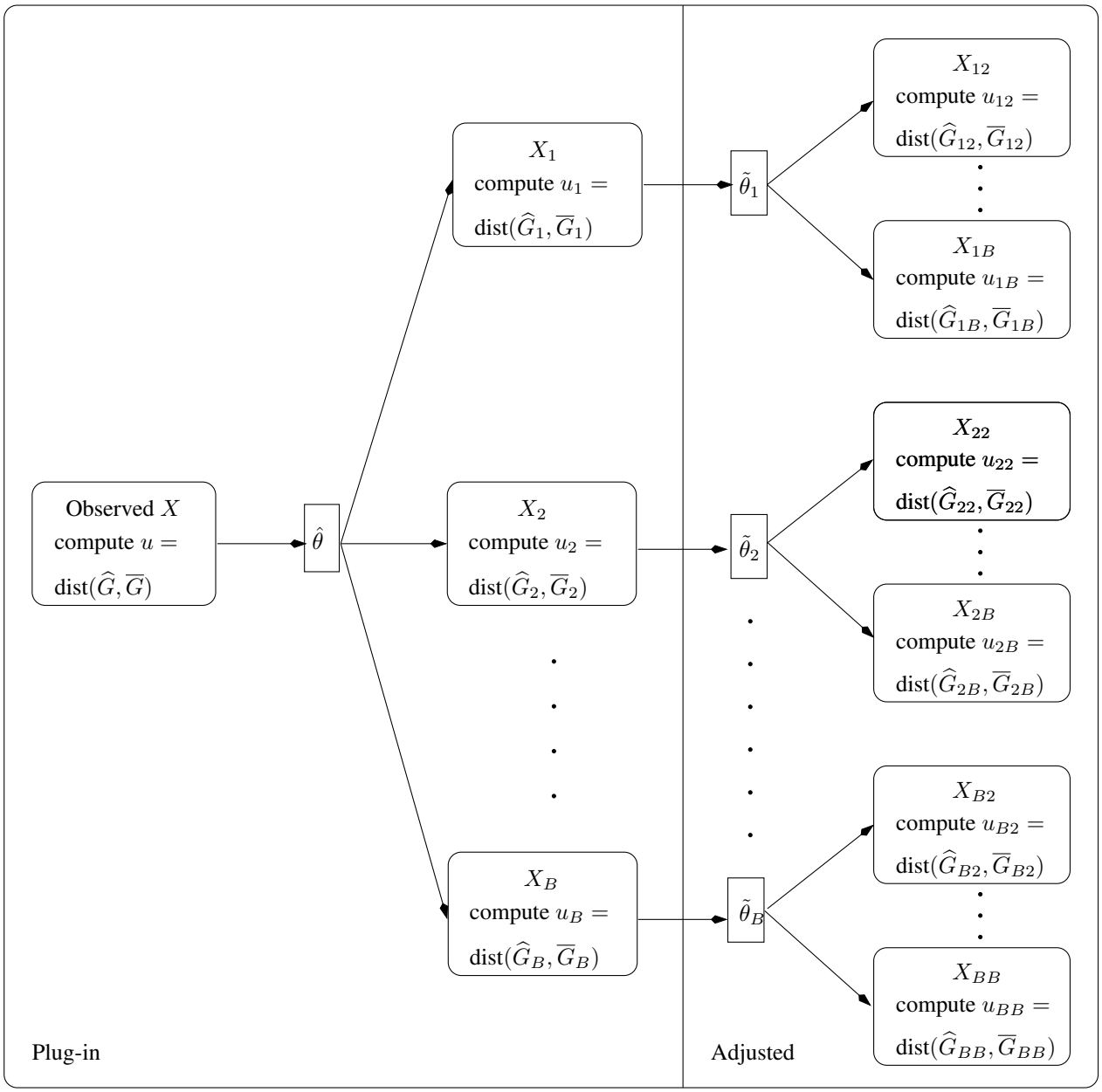


Figure 2.2: The test statistic, u, and its estimated p-value, $\hat{p}$, in (2.5) are computed using the Monte Carlo simulations represented in the "Plug-in" part. $\text{dist}(s,t) = \int_0^\infty \{s(h) - t(h)\}^2 dh$ is the squared L_2-norm of the difference between the functions s and t. The PGOF test rejects H_0 if $\hat{p} \leq \alpha$, where α is the nominal level. The AGOF test would reject H_0 if $\hat{p} \leq \hat{\alpha}^*$, where $\hat{\alpha}^*$ is the estimated adjusted level, which is computed in the "Adjusted" step. The Monte Carlo simulation is applied to each X_i, $i = 1, \ldots, n$, to obtain the estimated p-values $\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n$, to solve for $\hat{\alpha}^*$ in (2.11).*

*Reprinted with permission being requested from "A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns" by Ngoc Anh Dao and Marc G. Genton, 2014. *Journal of Computational and Graphical Statistics*, 23:2, pages 497 - 517, Copyright 2021 by Taylor & Francis Academic Journals.

in" part in Figure 2.2. The p-value for the Monte Carlo GOF test in (2.1) is

$$p \;=\; 1 - \Pr(U_{\hat{\theta}} < u), \tag{2.4}$$

where $U_{\hat{\theta}}$ is the random variable taking u_i, $i = 1, \ldots, n$, as realizations. Practically, this p-value can be estimated from

$$\hat{p} \;=\; 1 - \frac{1 + \sum_{i=1}^{n} I(u_i < u)}{n+1}, \tag{2.5}$$

where $I(\cdot)$ is the indicator function. The PGOF test would reject H_0 if $\hat{p} \leq \alpha$. Let $\hat{\mathcal{P}}$ denote the random variable corresponding to $\hat{p}$. Its distribution as noted in Sec. 2.1 is not standard uniform, except for $\mathbb{X}$ underlying a CSR.

Analogously, u, u_i, $i = 1, \ldots, n$, and $\hat{p}$ can also be computed using the other summary characteristic, the F-function, the second-order characteristic, K-, or the inhomogeneous K-function, for instance. The terminology of the summary characteristic and that of the second-order characteristic function are adopted from Illian et al. [2008]. The G-function, also known as the nearest neighbor distance distribution function, is the distribution function of the distance from an arbitrary *event* in X to the nearest other event in X, and the F-function, otherwise known as the empty space function, is the distribution function of the distance from an arbitrary *point* in the chosen window, W, to the nearest event in X. Both are estimated empirically using $\widehat{G}(h) = \sum_{i=1}^{N} I(h_i \leq h)/N$ and $\hat{F}(h) = \sum_{i=1}^{m} I(h_i^* \leq h)/m$, where $h > 0$, $N \equiv N(X)$ is the number of events in X, h_i denotes the distance from the ith event in X to the nearest other event in X, and h_i^* is the distance from the ith point of a sample of m selected points in W to the nearest of N events in X. The theoretical K-function of a stationary spatial point process is defined as $K(h) = \lambda^{-1} E(\text{the number of extra events within distance } h \text{ of a randomly chosen event})$, where $h \geq 0$ and λ is the intensity of the process [Bartlett, 1964, Ripley, 1977]. Ripley's estimator of the K-function is, for instance, in Cressie [1993, p. 640]. The inhomogeneous K-function as the K-function for non-stationary processes is introduced by Baddeley et al. [2000].

The envelope test is a graphical view of the Monte Carlo GOF test. Using the G-function, for instance, it declares a model as a good fit if the upper envelope $U(h)$ and the lower envelope $L(h)$ envelop the $\hat{G}$-function of the observed SPP X, where $U(h) = \max_{i=1,\ldots,n} \hat{G}_i(h)$, $L(h) = \min_{i=1,\ldots,n} \hat{G}_i(h)$, with typically $n = 100$.

To overcome the weaknesses with respect to the conservativeness and the power loss of PGOF tests, I propose a test that adjusts for the bias of its empirical level through nested Monte Carlo simulations to reach the nominal level correctly.

2.2.3 The Monte Carlo Adjusted Goodness-of-Fit Test

The motivation for my proposed test arises as follows. To make the PGOF test *correctly sized*, i.e., its empirical level, $\hat{\alpha}$, reaches the nominal level, α, my goal is to find $\hat{\alpha}^*$ such that

$$\Pr\{\hat{\mathcal{P}} < \hat{\alpha}^*\} \quad = \quad \alpha, \tag{2.6}$$

where $\hat{\alpha}^*$ is an estimate of α^*, which is referred to as *the adjusted level*. Usually α^* is unknown, and $\alpha^* = \alpha$ only in case of CSR. The decision rule of rejecting H_0 if $\hat{p} \leq \hat{\alpha}^*$ yields a correctly sized test because $\Pr\{\text{Reject (2.1) at level } \hat{\alpha}^* | \text{ (2.1) is true}\} = \alpha$ by (2.6).

My proposed Monte Carlo adjusted GOF (AGOF) test consists of three steps. The first step is to determine $\hat{p}$. For that u, $u_1, \ldots, u_n$ are computed, based, for instance, on the G-function, as given in (2.2) with

$$\overline{G}(h) = \frac{1}{n-1} \sum_{j=2}^{n} \hat{G}_j(h); \qquad \overline{G}_i(h) = \frac{1}{n-1} \sum_{j=1, j\neq i}^{n} \hat{G}_j(h), \;\; i = 1, \ldots, n. \tag{2.7}$$

The above $\overline{G}(h)$ and $\overline{G}_i(h)$, $i = 1, \ldots, n$, deliberately do not include $\hat{G}(h)$ like in (2.3). Thus, the $\overline{G}_i(h)$'s are not contaminated with the observed information through $\hat{G}(h)$ but are computed purely based on patterns generated under H_0. Also, I choose the above $\overline{G}(h) = \sum_{j=2}^{n} \hat{G}_j(h)/(n-1)$ as opposed to $\sum_{j=1}^{n} \hat{G}_i(h)/n$ because the $\overline{G}_i(h)$'s are normalized with the constant $(n-1)$. Simulations using Strauss and Thomas processes with $n = 50$ in Figure 2.3 demonstrate that, except for

one case, the empirical levels resulting from the modification in (2.7) are higher than those resulting from the traditional GOF test as proposed in (2.3). Only for the case of the Thomas process using the GOF-G test, the empirical levels resulting from (2.7) are not higher than that from (2.3). Overall, this modification makes the traditional GOF test become slightly more sensitive, which can be advantageous for small n. Eventually, the advantage of this modification will diminish as n increases, in particular for $n \geq 100$.

The second step is to find $\hat{\alpha}^*$. To achieve that, I have to reconstruct the distribution of $\hat{\mathcal{P}}$ via pseudo-random sampling using a Monte Carlo technique. I claim that the following procedure can provide the pseudo-values, $\hat{p}_1, \ldots, \hat{p}_n$ for $\hat{\mathcal{P}}$. They are estimates of the unknown plug-in p-values, $p_1, \ldots, p_n$, of the following hypotheses

$$H_{i,0} \quad : \quad \mathbb{X}_i \sim f_\theta, \tag{2.8}$$

with $\mathbb{X}_i$, $i = 1, \ldots, n$, as the underlying spatial point processes corresponding to the simulated X_i, $i = 1, \ldots, n$, under (2.1). In Figure 2.2, the AGOF test is illustrated as the extension of the PGOF test: I apply the PGOF test to each X_i, i.e. I simulate $X_{i,j}$, $j = 2, \ldots, n$, under (2.8) with $\theta = \widetilde{\theta}_i$, where $\widetilde{\theta}_i$ is the parameter estimate obtained from X_i, which is also $X_{i,1}$ in my notation. To compute $\hat{p}_1, \ldots, \hat{p}_n$, I define

$$u_{i,j} \quad = \quad \int_0^\infty \left\{ \hat{G}_{i,j}(h) - \overline{G}_{i,j}(h) \right\}^2 dh, \; i, j = 1, \ldots, n,$$

where

$$\overline{G}_{i,1}(h) = \frac{1}{n-2} \sum_{j=3}^{n} \hat{G}_{i,j}(h); \quad \overline{G}_{i,j}(h) = \frac{1}{n-2} \sum_{l=2, l \neq j}^{n} \hat{G}_{i,l}(h), \; j = 2, \ldots, n.$$

Then, the plug-in p-value ,

$$p_i \quad = \quad 1 - \Pr(U_{\widetilde{\theta}_i} < u_{i,1}), \tag{2.9}$$

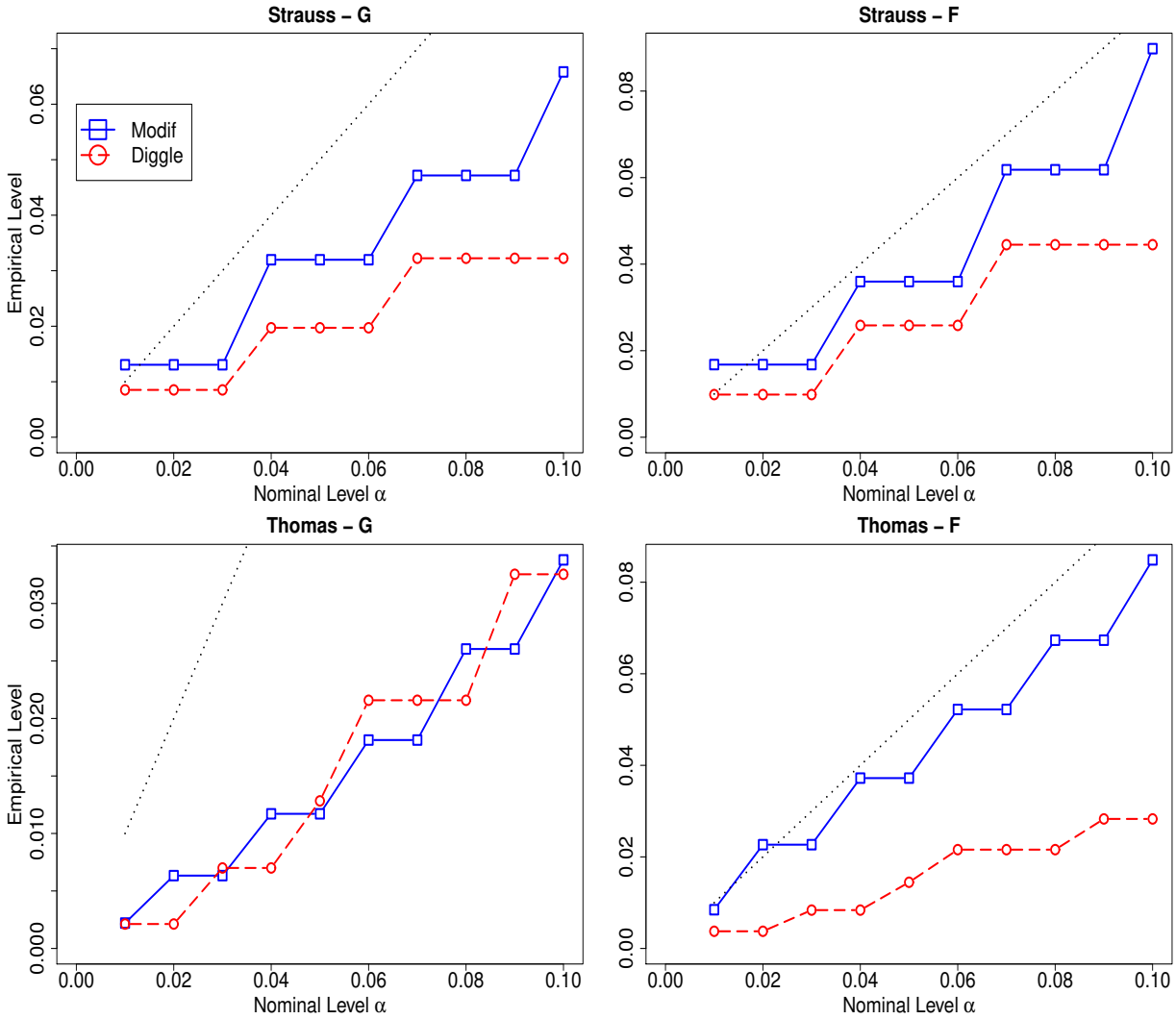


Figure 2.3: Empirical level curves resulting from the traditional Monte Carlo GOF tests in Sec. 2.2.2, represented by the dashed lines going through circles labeled as "Diggle", and their modified version through (2.7), represented by the solid lines going through squares labeled as "Modif", are shown; the dotted line is the reference 45° line. The tests use $n = 50$ based on $M = 3000$ generated patterns of Thomas and Strauss processes.*

*Reprinted with permission being requested from "A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns" by Ngoc Anh Dao and Marc G. Genton, 2014. *Journal of Computational and Graphical Statistics*, 23:2, pages 497 - 517, Copyright 2021 by Taylor & Francis Academic Journals.

for (2.8) with $U_{\widetilde{\theta}_i}$ as the random variable corresponding to $u_{i,2}, \ldots, u_{i,n}$ can be estimated by

$$\hat{p}_i \quad = \quad 1 - \frac{1 + \sum_{j=2}^{n} I(u_{i,j} \leq u_{i,1})}{n}. \tag{2.10}$$

I claim that the distribution of $\hat{\mathcal{P}}$ is herewith reconstructed via the pseudo-values, $\hat{p}_1, \ldots, \hat{p}_n$. Now, I find $\hat{\alpha}^*$ such that

$$\sum_{i=1}^{n} I\{\hat{p}_i < \hat{\alpha}^*\} \quad = \quad \alpha n. \tag{2.11}$$

Via the quantile function, $\hat{\alpha}^*$ can be obtained from the α-quantile of $\hat{p}_1, \ldots, \hat{p}_n$. The third, and also the last, step is making the conclusion about the fit, i.e. reject H_0 if $\hat{p} \leq \hat{\alpha}^*$. The validity of the AGOF test follows with the proof and pseudo-code given in the online supplementary material.

Proposition 1. *Let $\hat{\mathcal{P}}$ denote the random variable corresponding to the estimated plug-in p-value, $\hat{p}$, resulting from (2.5), and let $\hat{\alpha}^*$ be that found in (2.11). Then, I have (2.6). In particular when the null model is true, the probability of rejecting the null model to the significance level of $\hat{\alpha}^*$ is α, i.e., Pr $\{\hat{\mathcal{P}} < \hat{\alpha}^*\}$ = Pr $\{$Reject H_0 at level $\hat{\alpha}^* | \ H_0$ is true$\} = \alpha$.*

Remark 1. *The "*traditional*" Monte Carlo GOF test means that the Monte Carlo GOF test is applied to: (i) data consisting of at least two datasets; or (ii) data consisting of a single dataset with the parameter estimate depending only "loosely" on the estimation procedure. The PGOF test is the Monte Carlo GOF test with a plug-in, i.e., when the test is applied to the data consisting of a single dataset and the parameter estimate can depend on the estimation procedure. The AGOF test is the Monte Carlo adjusted GOF test, which is an extension of the PGOF test.*

2.3 Simulation Studies

From a process of interest, I generate $M = 3000$ patterns to compute the size of a PGOF or AGOF test. As in Sec. 2.2, n gives the number of Monte Carlo simulations within a single Monte Carlo GOF test. M, however, sets the number of replications of the Monte Carlo GOF tests. I use various processes to represent four main types of patterns, which are random (independent),

inhibitory (regular), clustered (aggregated), and inhomogeneous. To compare fairly my simulation results with those in Guan [2008], Illian et al. [2008], and Diggle [2003], I generate processes with patterns having a comparable number of events as described below. However, other simulations confirm that the performance of the PGOF and AGOF tests are not sensitive to the choice of specific parameter values. Computational work is done using the `spatstat` package [Baddeley and Turner, 2005b] in `R` [R Core Team, 2014].

2.3.1 Generating Patterns

In my simulation studies, the observation window of all point patterns is the unit square. In the first three types of patterns, the average number of events of each pattern is 100. First, for the random pattern, I choose the CSR process with intensity $\lambda = 100$. Second, for the inhibitory pattern, I choose a Strauss process with $(\beta, \gamma, R) = (200, 0.25, 0.05)$, where β is the chemical activity parameter, γ the interaction parameter, R the interaction radius, and the probability density function $f(\mathbb{X}) = \nu\beta^n\gamma^s$, with ν as the normalizing constant and s as the number of distinct unordered pairs of events closer than R units apart [Møller and Waagepetersen, 2003]. Third, for the clustered pattern, I consider a Thomas process as representative of this group. Here, a CSR with intensity κ is generated to obtain the so-called "parent" points. Each parent point is replaced by a random cluster of "children" points that are Poisson distributed with intensity μ. The positions of the children points are distributed about the parent location according to a bivariate Gaussian distribution with covariance $\sigma^2 I_2$, where I_2 is the 2×2 identity matrix [Møller and Waagepetersen, 2003]. Here the likelihood is intractable, but the theoretical K-function is available, $K(h) = \pi h^2 + \kappa^{-1}\left[1 - \exp\left\{-h^2/(4\sigma^2)\right\}\right]$. For my purposes, I chose a Thomas process with $(\kappa, \sigma, \mu) = (20, 0.05, 5)$. Finally, for the inhomogeneous Poisson processes, I adopt the models in the simulation studies in Guan [2008]. That is, I consider the following intensity functions:

$$\lambda(x) = \exp(\beta_0 + \beta_1 x), \tag{2.12}$$

$$\lambda(x) = \exp\{\beta_0 + \beta_1 \sin(2\pi x)\}, \tag{2.13}$$

where β_0 is the normalizing constant and β_1 specifies the inhomogeneity in the data. For each intensity function, I consider the cases $\beta_1 = -1$ and $\beta_1 = -2$. The bigger the absolute value of β_1, the more inhomogeneous the pattern. Here, β_0 is manipulated to obtain $\mu = 100$ and $\mu = 400$ number of events for each combination of β_1 and the intensity functions. I refer to the process with the intensity function (2.12) as linear and (2.13) as sine models.

2.3.2 Parameter Estimation and Performance of GOF Tests

I use various methods to obtain the estimates. For the CSR, the number of events is the estimate of the intensity parameter. For the Strauss process, I use the maximum profile pseudo-likelihood (MPPL) estimator [Baddeley and Turner, 2000] to obtain the estimate of R via `profilepl` and the maximum pseudo-likelihood (MPL) estimators [Besag, 1978, Berman and Turner, 1992, Baddeley and Turner, 2000] to obtain the estimates of β and γ via `ppm`. For the Thomas process, I use a minimum contrast method using the K-function (MCM-K) [Diggle, 2003, Sec. 6] via `thomas.estK`. For the inhomogeneous Poisson processes, I obtain the maximum likelihood estimate for β_0 and β_1 via `ppm`.

In the following, GOF-G, -F, -K or -K^{in} tests denote GOF tests using a G-, F-, K- or K^{in}-function to compute the test statistics. If the K-function is used in the estimation procedure, Diggle [2003, Sec. 6.2] recommended not using it again in computation of the statistic and the pseudo-statistics in (2.2) and (2.7). Hence, I have just GOF-G and -F tests for the Thomas process, and also have the GOF-K test for the Strauss and CSR processes. Additionally, for the inhomogeneous Poisson processes, I also employ the GOF-K^{in} test, where the K^{in}-function denotes the *inhomogeneous* K-function, which does not assume stationarity as the G-, F-, and K-functions do. With respect to the significance level, I choose the common nominal levels $\alpha = 0.05$, and $\alpha = 0.10$, due to Guan [2008]. A GOF test is correctly sized if its empirical level, $\hat{\alpha}$, reaches the nominal level, α, correctly according to the usual definition, i.e., the interval $(\hat{\alpha} - 2se, \hat{\alpha} + 2se)$, where $se = \sqrt{\hat{\alpha}(1-\hat{\alpha})/M}$, contains α.

Table 2.2 presents two facts. First, the PGOF test is correctly sized only for the CSR. In contrast, my proposed AGOF test is correctly sized for all processes, except for the AGOF-G test

applied in the Strauss case. Second, the estimated adjusted level, $\hat{\alpha}^*$, can serve as an indication for some patterns, as opposed to randomness, in the data. The AGOF test would provide $\hat{\alpha}^*$ strongly deviating from α if the SPP comes from a process that is very different from CSR.

For CSR, as explained in Sec. 2.1, the AGOF test is not needed. Here, the PGOF test coincides with the traditional Monte Carlo GOF test; thus, $\alpha^* = \alpha$. Nevertheless, the simulations in Table 2.2 validate the correctness of my method as $\widehat{\alpha}_{\text{AGOF}}$ conforms with $\widehat{\alpha}_{\text{PGOF}}$, $\overline{\alpha}^* = \sum_{k=1}^{M} \hat{\alpha}_k / M$ and α. For the Strauss process, the MPPL $\hat{R}$ of the irregular parameter, R, is a poor estimator, which strongly affects the estimation of the other two regular parameters, β and γ, which can be obtained via MPL or approximate maximum likelihood (AML) estimation [Huang and Ogata, 1999]. Due to the high computational complexity of the latter method, I use the MPL estimators of β and γ. Simulations show that $\hat{\gamma}$ tends to be greater than the true $\gamma = 0.25$. In fact, the median of $\hat{\gamma}$ is 0.277 and the mean is 0.289. The difficulty in parameter estimation carries on in the estimation of α^*. However, increasing n from 100 to 150 makes the conclusion of the AGOF test more accurate. I also conjecture that using the AML instead of the MPL estimators of β and γ would lead to a correctly sized AGOF-G test for the Strauss process.

Table 2.1 shows that the PGOF-G tests are correctly sized in two cases of the model (2.12) when $\beta_1 = -1$ with $\mu = 100$ and $\mu = 400$. This seems to contradict my observation made in Sec. 2.1 and the finding in the previous paragraph that the PGOF test is correctly sized only for CSR. My other simulations show that if the pattern and its H_0 are not too inhomogeneous as in the case of (2.12) with $\beta_1 = -1$, the PGOF test can, but does not have to, be correctly sized as the G-function of a CSR and of those models can be similar. However, the PGOF-F and -K^{in} tests remain incorrectly sized. Table 2.1 shows that the PGOF tests are overall incorrectly sized and their empirical levels deviate away from the nominal levels in two situations: (i) when the inhomogeneity factor, $|\beta_1|$, increases; and (ii) when the intensity function changes from linear, (2.12), to non-linear, (2.13). In contrast, the AGOF tests are correctly sized throughout. Overall, the AGOF tests clearly outperform the PGOF tests and do not underperform the test by Guan [2008]. Here, I did not implement the test of Guan [2008], which was described in Sec. 2.1. The

empirical levels, $\hat{\alpha}_{\text{Guan}}$, are extracted from $t = 0.2$ in Guan [2008, Tables 1 and 2].

Table 2.1: For nominal level α, $\widehat{\alpha}_{\text{PGOF}}$ and $\widehat{\alpha}_{\text{AGOF}}$ are empirical levels from the PGOF and AGOF tests using $n = 100$, $M = 3000$ replicates. $\overline{\alpha}^* = \sum_{k=1}^{M} \hat{\alpha}_k^*/M$. Linear and sine models and β_0 and β_1 are defined in (2.12) and (2.13) in Sec. 2.3.1. Notations -G, -F, and -K^{in} label computations of the GOF test statistics using the G-, F-functions, and the *inhomogeneous* K-function, respectively. From Guan [2008, Tables 1 and 2], $t = 0.2$, $\hat{\alpha}_{\text{Guan}}$ is extracted. Estimated standard errors multiplied by 100 are in parentheses.*

				$\alpha = 0.05$			$\alpha = 0.10$			
Model	# events	β_0	$-\beta_1$	$\widehat{\alpha}_{\text{PGOF}}$	$\overline{\alpha}^*$	$\widehat{\alpha}_{\text{AGOF}}$	$\widehat{\alpha}_{\text{PGOF}}$	$\overline{\alpha}^*$	$\widehat{\alpha}_{\text{AGOF}}$	$\hat{\alpha}_{\text{Guan}}$
Linear-G	100	5.06	1	.048	.054	.052	.095	.107	.104	.096
				(.390)		(.404)	(.535)		(.557)	
Linear-F	100	5.06	1	.032	.077	.055	.072	.139	.102	.096
				(.323)		(.417)	(.473)		(.552)	
Linear-K^{in}	100	5.06	1	.034	.072	.063	.077	.138	.109	.096
				(.331)		(.443)	(.486)		(.570)	
Linear-G	100	5.45	2	.039	.062	.052	.075	.124	.099	.130
				(.353)		(.405)	(.480)		(.546)	
Linear-F	100	5.45	2	.015	.128	.051	.035	.213	.101	.130
				(.112)		(.345)	(.155)		(.469)	
Linear-K^{in}	100	5.45	2	.035	.069	.046	.074	.132	.105	.130
				(.333)		(.385)	(.478)		(.560)	
Linear-G	400	6.45	1	.051	.049	.052	.094	.107	.099	.094
				(.349)		(.357)	(.478)		(.497)	
Linear-F	400	6.45	1	.033	.087	.052	.059	.161	.102	.094
				(.327)		(.408)	(.431)		(.552)	
Linear-K^{in}	400	6.45	1	.043	.063	.053	.083	.128	.107	.094
				(.359)		(.409)	(.502)		(.565)	
Linear-G	400	6.85	2	.041	.061	.052	.083	.124	.108	.098

Continued on next page

Table 2.1 – continued from previous page

				$\alpha = 0.05$			$\alpha = 0.10$			
Model	# events	β_0	$-\beta_1$	$\widehat{\alpha}_{\text{PGOF}}$	$\overline{\alpha}^*$	$\widehat{\alpha}_{\text{AGOF}}$	$\widehat{\alpha}_{\text{PGOF}}$	$\overline{\alpha}^*$	$\widehat{\alpha}_{\text{AGOF}}$	$\hat{\alpha}_{\text{Guan}}$
				(.365)		(.405)	(.503)		(.567)	
Linear-F	400	6.85	2	.012	.146	.052	.027	.238	.101	.098
				(.201)		(.405)	(.297)		(.550)	
Linear-K^{in}	400	6.85	2	.025	.086	.049	.053	.160	.093	.098
				(.283)		(.395)	(.408)		(.531)	
Sine-G	100	4.37	1	.026	.081	.044	.054	.149	.090	.100
				(.288)		(.374)	(.412)		(.523)	
Sine-F	100	4.37	1	.004	.176	.045	.012	.272	.100	.100
				(.115)		(.378)	(.204)		(.548)	
Sine-K^{in}	100	4.37	1	.023	.076	.044	.056	.145	.096	.100
				(.275)		(.373)	(.420)		(.537)	
Sine-G	100	3.80	2	.019	.096	.046	.048	.169	.093	.108
				(.253)		(.382)	(.393)		(.532)	
Sine-F	100	3.80	2	.004	.195	.053	.011	.293	.108	.108
				(.115)		(.409)	(.193)		(.567)	
Sine-K^{in}	100	3.80	2	.048	.048	.048	.093	.109	.091	.108
				(.393)		(.393)	(.532)		(.527)	
Sine-G	400	5.80	1	.029	.076	.049	.058	.142	.094	.082
				(.329)		(.412)	(.447)		(.558)	
Sine-F	400	5.80	1	.007	.178	.046	.015	.278	.097	.082
				(.162)		(.400)	(.237)		(.566)	
Sine-K^{in}	400	5.80	1	.037	.075	.054	.064	.143	.100	.082
				(.361)		(.432)	(.465)		(.573)	

Continued on next page

Table 2.1 – continued from previous page

				$\alpha = 0.05$			$\alpha = 0.10$			
Model	# events	β_0	$-\beta_1$	$\widehat{\alpha}_{\text{PGOF}}$	$\overline{\alpha}^*$	$\widehat{\alpha}_{\text{AGOF}}$	$\widehat{\alpha}_{\text{PGOF}}$	$\overline{\alpha}^*$	$\widehat{\alpha}_{\text{AGOF}}$	$\hat{\alpha}_{\text{Guan}}$
Sine-G	400	5.15	2	.028	.093	.057	.063	.166	.106	.114
				(.325)		(.452)	(.476)		(.602)	
Sine-F	400	5.15	2	.004	.189	.056	.014	.288	.104	.114
				(.125)		(.428)	(.230)		(.596)	
Sine-K^{in}	400	5.15	2	.042	.064	.052	.072	.128	.101	.114
				(.394)		(.435)	(.506)		(.589)	

2.3.3 Effective Simulation Size and Computational Time

Marriott [1979] proposed to consider $m/n = \alpha$, where m is chosen in a way that if u is among the m largest values of $u_1, \ldots, u_n$, H_0 is rejected. Besag and Diggle [1977] suggested that $m = 5$ might be a suitable value for the traditional Monte Carlo tests for spatial patterns. Consequently, $n = 100$ should be used for $\alpha = 0.05$ and $n = 500$ for $\alpha = 0.01$. Hence, the smaller the α, the bigger the n. My simulation studies shown in Tables 2.2 and 2.1 are run using $n = 100$ to compare fairly the performance of the PGOF and AGOF tests at the nominal level $\alpha = 0.05$. However, Table 2.3 shows that except for AGOF-G and -F tests for the Strauss process, the AGOF tests need only $n = 20$ to be correctly sized at the $\alpha = 0.05$ level, and $n = 100$ to be correctly sized at the $\alpha = 0.01$ level. That is, the effective simulation size is reduced by a factor of 5 in the setting of the AGOF test, except for the AGOF-G and -F tests for processes having parameters being estimated by the MPPL method, including Strauss processes. For CSR, while Table 2.3 shows that the AGOF tests just need $n = 20$, other simulations show that the PGOF tests need $n \geq 100$ to be correctly sized at $\alpha = 0.05$.

Table 2.4 shows my investigation of the computational time of the PGOF and AGOF tests. The

*Reprinted with permission being requested from "A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns" by Ngoc Anh Dao and Marc G. Genton, 2014. *Journal of Computational and Graphical Statistics*, 23:2, pages 497 - 517, Copyright 2021 by Taylor & Francis Academic Journals.

Table 2.2: For nominal level α, $\widehat{\alpha}_{\text{PGOF}}$ and $\widehat{\alpha}_{\text{AGOF}}$ are empirical levels from the PGOF and AGOF tests using $n = 100$, $M = 3000$ replicates. $\overline{\alpha}^* = \sum_{k=1}^{M} \hat{\alpha}_k^*/M$. Notations -$G$, -$F$, and -$K$ label computations of the GOF test statistics using the G-, F-, K-functions, respectively. The CSR, Strauss and Thomas processes are defined in Sec. 2.3.1. Estimated standard errors multiplied by 100 are in parentheses.*

			$\alpha = 0.05$			$\alpha = 0.10$		
Model	# events	Pattern	$\widehat{\alpha}_{\text{PGOF}}$	$\overline{\alpha}^*$	$\widehat{\alpha}_{\text{AGOF}}$	$\widehat{\alpha}_{\text{PGOF}}$	$\overline{\alpha}^*$	$\widehat{\alpha}_{\text{AGOF}}$
CSR-G	100	Random	.053	.049	.054	.096	.106	.105
			(.410)		(.414)	(.538)		(.560)
CSR-F		Random	.049	.049	.048	.095	.107	.099
			(.396)		(.389)	(.536)		(.545)
CSR-K		Random	.051	.048	.050	.093	.099	.091
			(.401)		(.398)	(.529)		(.524)
Strauss-G	100	Inhibitory	.026	.141	.068	.044	.233	.125
			(.292)		(.459)	(.374)		(.603)
Strauss-F		Inhibitory	.020	.106	.047	.043	.184	.090
			(.226)		(.385)	(.371)		(.522)
Strauss-K		Inhibitory	.040	.062	.053	.078	.124	.100
			(.356)		(.410)	(.488)		(.548)
Thomas-G	100	Clustered	.008	.153	.050	.027	.241	.107
			(.153)		(.374)	(.277)		(.517)
Thomas-F		Clustered	.037	.073	.055	.074	.140	.100
			(.322)		(.388)	(.446)		(.512)

*Reprinted with permission being requested from "A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns" by Ngoc Anh Dao and Marc G. Genton, 2014. *Journal of Computational and Graphical Statistics*, 23:2, pages 497 - 517, Copyright 2021 by Taylor & Francis Academic Journals.

Table 2.3: The empirical levels $\widehat{\alpha}_{\text{AGOF}}$ result from AGOF tests using $n = 20, 50, 100$ to assess the effective simulation size n to nominal levels $\alpha = 0.01, 0.05, 0.10$. sine, Strauss, and Thomas processes are defined in Sec. 2.3.1.*

	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
Model	20	50	100	20	50	100	20	50	100
CSR-G	.049	.024	.012	.049	.057	.054	.110	.111	.105
CSR-F	.054	.025	.008	.054	.058	.048	.112	.111	.099
CSR-K	.056	.023	.009	.056	.058	.049	.109	.111	.091
Sine-G	.054	.024	.011	.054	.051	.051	.105	.105	.095
Sine-F	.051	.022	.010	.051	.050	.052	.101	.100	.100
Sine-K	.048	.020	.011	.048	.051	.049	.100	.105	.094
Strauss-G	.076	.026	.015	.076	.065	.068	.101	.123	.125
Strauss-F	.069	.036	.010	.069	.079	.047	.133	.133	.090
Strauss-K	.051	.017	.010	.051	.051	.053	.109	.101	.100
Thomas-G	.053	.024	.012	.053	.046	.050	.099	.100	.092
Thomas-F	.053	.023	.011	.053	.057	.049	.106	.110	.100

computation was done on compute nodes that have 8 CPU cores, 32GB of RAM and the CPU processors clocked at 2.4GHz or faster. For each of the CSR, sine, Strauss and Thomas processes defined in Sec. 2.3.1, $M = 3000$ patterns are generated to provide the average computational time given in seconds. The time is to obtain $\hat{p}$ from a PGOF test, $\hat{p}$ and $\hat{\alpha}^*$ from an AGOF test using G-, F-, K- or K^{in}- and pc- or pc^{in}-functions altogether. The pc-function denotes the pair correlation function $pc(h) = K'(h)/(2\pi h)$, [Illian et al., 2009]. The pc^{in}-function results analogously from the K^{in}-function.

Correctly sizing the AGOF test does come with a cost as the computational time is always longer than that of the PGOF test. While the Monte Carlo simulations are of the size n for the PGOF test, they are of the size n^2 for the AGOF test as illustrated in Figure 2.2, which also shows that the AGOF test is an extension of the PGOF test.

*Reprinted with permission being requested from "A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns" by Ngoc Anh Dao and Marc G. Genton, 2014. *Journal of Computational and Graphical Statistics*, 23:2, pages 497 - 517, Copyright 2021 by Taylor & Francis Academic Journals.

Table 2.4: For each of CSR, sine, Strauss and Thomas processes defined in Sec. 2.3.1, $M = 3000$ patterns are generated to provide the average computational time given in seconds. The time in seconds is to obtain results from a GOF test using G-, F-, K- or K^{in}- and pc- or pc^{in}-functions altogether.*

	$n = 20$		$n = 50$		$n = 100$	
Model	PGOF	AGOF	PGOF	AGOF	PGOF	AGOF
CSR	2	42	7	227	9	835
Sine	7	87	19	541	36	1,897
Strauss	269	5,138	642	31,907	1,288	121,671
Thomas	2	52	5	313	9	1,076

2.3.4 Statistical Power

In the following, I show that the power of the AGOF test is at least as good as the power of the PGOF test via a simulation study. I simulate $M = 3000$ patterns under the sine model in (2.13), and fit wrongly a Thomas process (null model) to each of the patterns. In my opinion, this example is realistic because patterns underlying both (2.13) and a Thomas process contain clusters. In a real scenario, there might be a pattern underlying (2.13) that is incorrectly categorized as a Thomas process. After fitting a Thomas process to these datasets, I obtained 3000 $\hat{p}$'s from the PGOF-G and -F tests. The type-II error rate, β, was set at 0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40. The power was the rejection rate based on α for the PGOF tests and based on $\hat{\alpha}^*$ for the AGOF tests. In Figure 2.4, the power curve of the AGOF test is the solid curve going through 9 squares and the one of the PGOF test is the dashed curve going through 9 circles. The power curve of the AGOF test is clearly higher than the one of the PGOF at all given β.

Remark 2. *My simulations show that there can be contradictory conclusions for the GOF-G, - F, -K, and -K^{in} tests based on one spatial pattern although fitting the correct model. Indeed, each of the G-, F-, K-, and K^{in}-functions measures completely different things. Thus, disagreement*

*Reprinted with permission being requested from "A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns" by Ngoc Anh Dao and Marc G. Genton, 2014. *Journal of Computational and Graphical Statistics*, 23:2, pages 497 - 517, Copyright 2021 by Taylor & Francis Academic Journals.

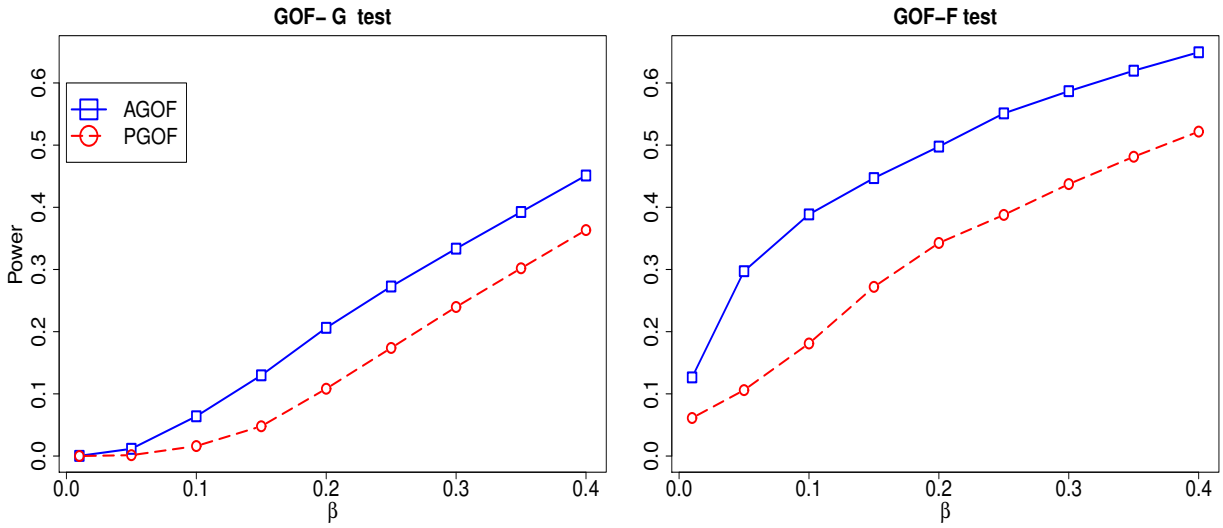


Figure 2.4: Power curves of AGOF-G, -F and PGOF-G, -F tests, using $n = 100$, are evaluated based on $M = 3000$ simulated patterns of the sine model in (2.13) when wrongly fitting a Thomas process. The power curves of the AGOF tests are the solid curves going through 9 squares, and of the PGOF tests are the dashed curves going through 9 circles. β denotes the type-II error rate at 0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40.*

among the AGOF tests can occur, but their performances can still be correctly sized.

2.4 Data Applications

2.4.1 Phlebocarya Filifolia Plants

Figure 2.5 (left) displays, in a 22m$\times$22m window, positions of 207 Phlebocarya filifolia plants that are typically located in scattered positions throughout Western Australia. Illian et al. [2008, Example 4.19] used this dataset as an example of an homogeneous clustered pattern by demonstrating the occurrence of aggregation via a large value of the Clark-Evans index and a small value of the mean-direction index. Illian et al. [2008, Example 7.2] concluded that the Matérn cluster process provides an acceptable fit via the envelope test with a plug-in as the envelopes using the G- and F-functions with $n = 100$ envelop the empirical G- and F-functions, respectively. I would like to re-evaluate the GOF of a process with an homogeneous clustered pattern with my AGOF-G

*Reprinted with permission being requested from "A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns" by Ngoc Anh Dao and Marc G. Genton, 2014. *Journal of Computational and Graphical Statistics*, 23:2, pages 497 - 517, Copyright 2021 by Taylor & Francis Academic Journals.

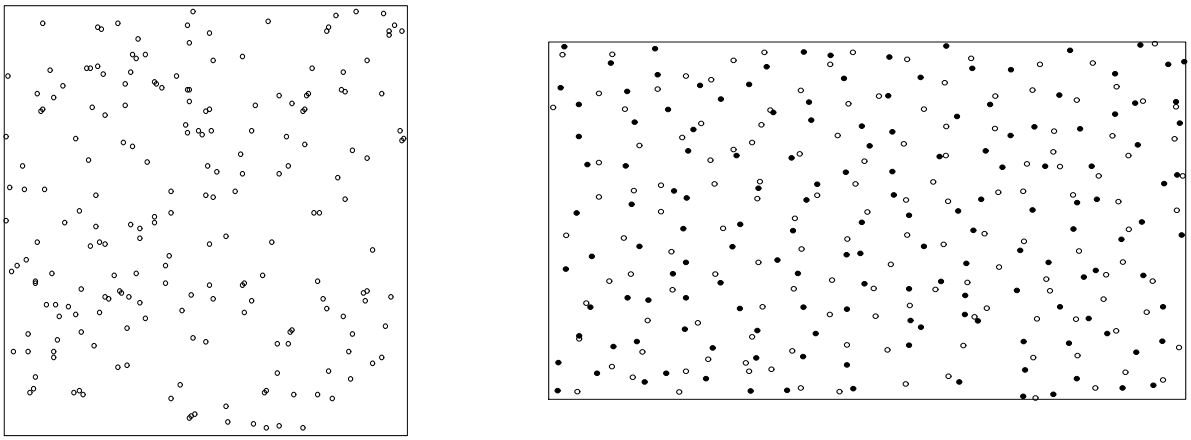


Figure 2.5: Left: Positions of 207 Phlebocarya filifolia plants in a 22m$\times$22m window at Cooljarlo near Perth, Australia. Right: Positions of 294 displaced amacrine cells in the retina of a rabbit. Solid and open circles represent on and off cells, respectively. The observation window for the data is $(0, 1.62) \times (0, 1)$.*

and -F tests.

As clustering is not immediately obvious, I first used the PGOF test to assess the fit of CSR to detect whether actual modeling is necessary. The PGOF-G, -F, and -K tests yielded estimated p-values of 0, 0.515, and 0.01, respectively, suggesting that the fit of CSR is disputable. Figure 2.6 shows the empirical G-, F-, and K-function (dash-dotted curves); other references are the functions of a CSR (solid curve) representing a random pattern, of a Thomas process (dashed curve) representing a clustered pattern, and of a Strauss process (green dotted curve) representing an inhibitory pattern. These processes have approximately 200 events. While the left column of plots shows these functions on their complete domains, the right column shows these functions on a much smaller domain to scrutinize their behavior for distances shorter than 0.01. While the empirical F-function aligning with the one of the CSR can suggest random behavior in agreement with the conclusion of the PGOF-F test, the empirical G-function lying above the one of the CSR indi-

*Reprinted with permission being requested from "A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns" by Ngoc Anh Dao and Marc G. Genton, 2014. *Journal of Computational and Graphical Statistics*, 23:2, pages 497 - 517, Copyright 2021 by Taylor & Francis Academic Journals.

cates clustering, except for distances shorter than 0.0044. The empirical K-function significantly signals slight clustering as it lies completely above the one of the CSR, except for distances shorter than 0.0047. At this distance, the behavior of this dataset and that of the reference processes seem to be similar because the biggest discrepancy is less than 10^{-4}. While the G- and F-functions are believed to be "short-sighted" as they describe the distances to nearest neighbors of reference events or points and say little or nothing about the spatial dependence of the events beyond the nearest neighbor [Illian et al., 2008, Sec. 4.3.1], the K-function can be an effective summary of spatial dependence over a wide range of scales [Cressie, 1993, Sec. 8.4.3]. Thus, I ascribe clustering to this dataset due to the plots of the K-function. I consider a Matérn cluster process [Illian et al., 2008] and a Thomas process as reasonable models because they are representatives of processes with homogeneous clustered patterns.

Assuming that the pattern comes from a Matérn cluster process [Møller and Waagepetersen, 2003], an homogeneous Poisson point process with intensity κ is generated to obtain the "parent" points. Each parent point is replaced by a random cluster of "children" points that are Poisson distributed with intensity μ. The positions of the children points are placed independently and uniformly inside a disc of radius R centered at the parent point. Neither likelihood nor pseudo-likelihood is available, but the K-function is $K(h) = \pi h^2 + s(h)\{h/(2R)\}/\kappa$, where $s(h) = 2+(1/\pi)\left\{(8h^2-4)\arccos(h) - 2\arcsin(h) + 4h\sqrt{(1-h^2)^3} - 6h\sqrt{1-h^2}\right\}$ for $h \leq 1$, and $s(h) = 1$ for $h > 1$. The MCM-K provides the estimate $(\hat{\kappa}, \hat{R}, \hat{\mu}) = (132.996/\{22m \times 22m\}, 3.482m, 1.556)$. The AGOF-$G$ and -F tests give the $\hat{p}$'s of 0.054, 0.634 and $\hat{\alpha}^*$'s of 0.123, 0.449, respectively. The AGOF-G test does not support the fit as $\hat{p} < \hat{\alpha}^*$ $(0.023 < 0.123)$. However, the AGOF-F test supports the good fit of the Matérn cluster model because $\hat{p} > \hat{\alpha}^*$ $(0.634 > 0.449)$.

Now, suppose the pattern comes from a Thomas process, where the MCM-K provides the estimate $(\hat{\kappa}, \hat{\sigma}, \hat{\mu}) = (111.455/\{22m \times 22m\}, 2.099m, 1.857)$. The AGOF-$G$ and -F tests give $\hat{p}$'s of 0.044, 0.761 and $\hat{\alpha}^*$'s of 0.115, 0.620, respectively. The AGOF-G test does not support the fit as $\hat{p} < \hat{\alpha}^*$ $(0.044 < 0.115)$. However, the AGOF-F supports the fit of the Thomas process because

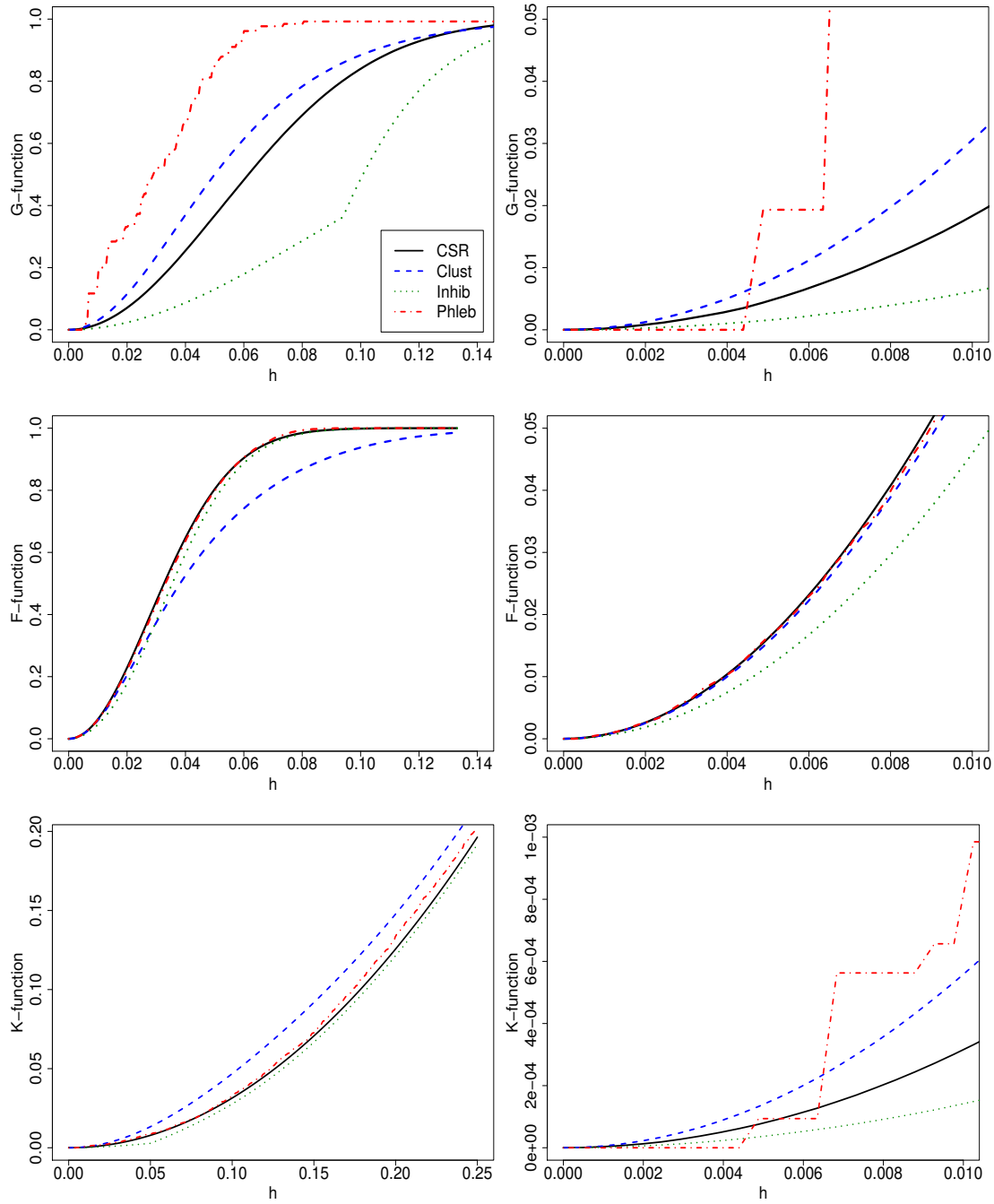


Figure 2.6: Study of the summary characteristics based on the G- and F-functions and second-order characteristic based on the K-function of the Phlebocarya filifolia plants. Left column: complete domain; Right column: smaller domain. The empirical K-function in dashed curve indicates slightly clustering.*

*Reprinted with permission being requested from "A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns" by Ngoc Anh Dao and Marc G. Genton, 2014. *Journal of Computational and Graphical Statistics*, 23:2, pages 497 - 517, Copyright 2021 by Taylor & Francis Academic Journals.

$\hat{p} > \hat{\alpha}^*$ $(0.761 > 0.620)$.

I think that the AGOF-G tests reject the fits of the Matérn cluster and the Thomas processes because the G-function in Figure 2.6 indicates, on the most part, clustering, but it suggests inhibition on distances, h, less than 0.0044. Thus, I suspect that a fit of any clustered or inhibitory model would unlikely gain support by the AGOF-G test. Overall, I would support either the fit of the Matérn cluster or the Thomas process because the AGOF-F test in each case supports the fit and the plot of the empirical K-function significantly indicates slight clustering. This finding is in agreement with the conclusion made by Illian et al. [2008, Example 7.2]. One might understand the propagation mechanism of Phlebocarya filifolia plants better: there is indeed clustering, although just slightly. One might also prefer a Thomas to a Matérn cluster process as the Thomas process models more "children" points at closer distances to parent points; this characteristic might be more natural for forest data.

2.4.2 Amacrine Cells

The data displayed in Figure 2.5 (right) show the bivariate pattern of amacrine cells within the retina of a rabbit. Interest lies in distinguishing between two developmental hypotheses in studying the retinas of rabbits. The observation window for the data is the rectangle $(0, 1.62) \times (0, 1)$ according to Diggle [2003]. The two types of cells are responses to a light being switched *on* and *off*. The *separate layer* hypothesis is that the on and off cells are initially formed in two separate layers that later fuse to form a mature retina. The *single layer* hypothesis is that the two types of cells are initially undifferentiated in a single layer and acquire their separate functions at a later stage. Via a Monte Carlo test of independence, Diggle [2003] failed to reject the on and off cells being independent processes, supporting the separate layer hypothesis [Diggle, 2003, Sec. 4.7]. To answer whether they would fuse to form a mature retina, one might consider fitting a statistical model to one and run a GOF test on the other dataset.

Diggle [2003, Sec. 4.7] showed that these processes have very similar second-order properties including inhibitory behaviors at small distances. Subsequently, a model is fitted to the on cells only and the off cells are reserved for a GOF assessment. A pairwise interaction point process, i.e.,

a Markov point process, was chosen due to the inhibitory behavior at small distances, together with the absence of any obvious longer-range heterogeneity or other form of aggregation. Ripley [1977] formalized the class of pairwise interaction processes, $f(X) = \nu\beta^n \prod_i \prod_{j\neq i} d\{\|x_i - x_j\|\}$, where ν is the normalizing constant, β reflects the intensity of the process, $d(h)$ is non-negative for all $h > 0$ and the product is over all pairs of distinct points in the SPP, X. The parametric model with the interaction function for amacrine cells is taken from Diggle and Gratton [1984]:

$$d(h) = \begin{cases} 0, & h < \delta, \\ \{(h-\delta)/(\rho-\delta)\}^\kappa, & \delta \leq h \leq \rho, \\ 1, & h > \rho. \end{cases}$$

The distance from δ to ρ can be interpreted as an interaction distance. The parameters δ, ρ, and κ are nonnegative, and $\delta < \rho$. The strength of inhibition increases with κ. For $\kappa = 0$, the model is a hard-core process with radius δ, whereas for $\kappa = \infty$ it has radius ρ.

Diggle [2003, Sec. 7.2] obtained $(\hat{\delta}, \hat{\rho}, \hat{\kappa}) = (0.020, 0.12, 4.90)$ from the on cells via the MPL estimation. Then, to assess the GOF to the off cells, a traditional GOF-G test and a traditional GOF-F test were employed. The p-values were reported to be 0.01 and 0.37, respectively. Since these two tests give opposite conclusions with respect to any commonly used nominal level, Diggle [2003] was then motivated to obtain different estimates $(\hat{\delta}, \hat{\rho}, \hat{\kappa}) = (0.016, 0.12, 1.96)$ via AML estimation. Diggle [2003, Figure. 7.7] showed that there is agreement between the F-function of the off cells and its corresponding envelopes. However, the GOF-G test still shows a poor fit of the pairwise interaction point process to the off cells. Diggle [2003] did not give a clear evaluation of the GOF for the amacrine cells data.

I fit the pairwise interaction process to the off cells dataset (training set) and obtained estimates of δ, ρ via the MPPL, and of κ via the AML or the MPL estimation. Then I ran AGOF-G, -F, and -K tests. If they concluded a good fit, these parameter estimates were used for the PGOF-G, -F, -K tests applied to the on cells data (validation set).

Via MPPL and AML estimation I obtained the estimates $(0.015, 0.11, 3.435)$ from the off cells dataset, and $(0.026, 0.28, 1.083)$ from the on cells dataset for (δ, ρ, κ). Due to a very high computational intensity, I used $n = 50$ for the AGOF-G, -F, and -K tests, which provided p-values of 0.868, 0.647, 0.092 and $\hat{\alpha}^* = 0.022, 0, 0.091$, respectively. All AGOF tests indicated a good fit of the pairwise interaction process to the off cells dataset because all $\hat{p}$'s are greater than the corresponding $\hat{\alpha}^*$'s. Then I used the estimate of the off cells dataset as the parameter for the pairwise interaction process to fit to the on cells dataset. The PGOF-G, -F, and -K tests provided p-values of 0.200, 0.519 and 0.340, which indicated a good fit because the $\hat{p}$'s were greater than $\alpha = 0.05$. The AGOF tests drew the same conclusions as the PGOF tests as the $\hat{p}$'s were greater than the corresponding $\hat{\alpha}^* = 0.032, 0.061, 0.191$.

Via MPPL and MPL estimation, I obtained the estimates $(0.015, 0.11, 2.44)$ from the off cells dataset, and $(0.026, 0.28, 1.062)$ from the on cells dataset for (δ, ρ, κ). The AGOF-G, -F, and -K tests provided $\hat{p}$'s of 0.203, 0.011, 0.478 and $\hat{\alpha}^* = 0.015, 0.052, 0.052$ respectively. Only the AGOF-G and AGOF-K tests indicated a good fit of the pairwise interaction process to the off cells dataset because only their $\hat{p}$'s were greater than the corresponding $\hat{\alpha}^*$'s. I could have stopped here, but for curiosity I proceeded to run PGOF tests using the estimate of the off cells dataset as the parameter for the pairwise interaction process to fit to the on cells dataset. The PGOF-G, -F, and -K test tests provided p-values of 0.685, 0.092 and 0.606, which indicated a good fit because the $\hat{p}$'s were greater than $\alpha = 0.05$. The AGOF tests drew the same conclusions as the PGOF tests as the $\hat{p}$'s were greater than the corresponding $\hat{\alpha}^* = 0.050, 0.050, 0.174$.

The conclusions of the AGOF-F tests for the fit of the off cells dataset differed, but the conclusions of the PGOF and AGOF tests for the fit of the on cells dataset agreed. Since the AML estimators had better statistical properties [Huang and Ogata, 1999], I relied on the finding using AML estimates and concluded that the separate layer hypothesis holds [Diggle, 2003, Sec. 4.7], i.e., these two independent processes later fuse to form a mature retina.

2.5 Discussion

This chapter discussed a method using nested Monte Carlo simulations to obtain the estimated adjusted level, $\hat{\alpha}^*$, corresponding to a prespecified nominal level, α. I rejected the null model if $\hat{p} \leq \hat{\alpha}^*$, where $\hat{p}$ is the estimated p-value. The nested Monte Carlo method was, however, computationally intensive, and I now discuss two techniques to improve the computational complexity, the first of which also estimates α^* more efficiently.

The first technique uses the idea of interpolation. Due to the computational intensity, I reduce the number of replications of Monte Carlo simulations, n, and compensate that by estimating the new u_i and $u_{i,j}$ with interpolation techniques, such as kernel density estimation. From the Monte Carlo simulations in Sec. 2.2, I obtain two sets, $S = \{u_1, \ldots, u_n\}$ and $S^* = \{u_{i,j}, i, j = 1, \ldots, n\}$. Via kernel density estimation, I estimate the new u_i's augmenting S to $S_{\text{aug}} = \{u_1, u_2, \ldots, u_n, \hat{u}_1, \hat{u}_2, \ldots, \hat{u}_\tau\}$ to obtain a better $\hat{p}$. Analogously, I estimate u_{ij}'s augmenting S^* to $S^*_{\text{aug}} = \{u_{1,2}, \ldots, u_{1,n}, \hat{u}_{1,1}, \ldots, \hat{u}_{1,\tau}; u_{2,2}, \ldots, u_{2,n}, \hat{u}_{2,1}, \ldots, \hat{u}_{2,\tau}; \ldots; u_{n,2}, \ldots, u_{n,n}, \hat{u}_{n,1}, \ldots, \hat{u}_{n,\tau}\}$ to obtain better $\hat{p}_1, \ldots, \hat{p}_n$, which leads to estimating a better $\hat{\alpha}^*$. The reason is that $\hat{\alpha}^*$ is the α-quantile of $\hat{p}_1, \ldots, \hat{p}_n$ as formalized in (2.11). Here, the quantities $\hat{u}_i$'s and $\hat{u}_{i,j}$'s denote the new estimated pseudo-statistics, and τ^{-1} specifies the interpolation factor. Using Taylor expansions, I can show that the new $\hat{p}$ is similar to the one in Sec. 2.2 minus a term of order, $O(\tau^{-1})$. Since τ^{-1} can be chosen arbitrarily small, this term is negligible and the new $\hat{p}$ possesses the same properties as discussed in Sec. 2.2. The simulation studies use CSR, sine, Strauss, and Thomas processes defined in Sec. 2.3.1 and $n = 20$ and $\tau = 100$. Table 2.5 demonstrates that interpolation techniques can make the AGOF tests more accurate as the mean squared errors are smaller in 15 out of 22 studied cases. The strength of interpolation might be increased or optimized by a different choice of τ.

The second technique considers the characteristics of sequential Monte Carlo p-values. Besag and Clifford [1991] introduced several ways of calculating exact Monte Carlo p-values by sequential sampling. Instead of fixing the sample size, n, sampling is continued until a prespecified number, H, of values larger than the value u of the test statistic, U, has been observed. Let l be the value of the random sample size, L, at termination. Besag and Clifford [1991] showed that a

Table 2.5: To nominal level α, the empirical level $\hat{\alpha} = \widehat{\alpha}_{\text{AGOF}}$ using $\tau = 100$, $n = 20$ and $M = 5000$ is shown. $\overline{\alpha}^* = \sum_{k=1}^{M} \hat{\alpha}_k^*/M$. MSE is the mean squared error. The "Standard" columns display empirical levels from the AGOF tests from Sec. 2.2.3. The "Interpolation" columns display empirical levels from the AGOF tests incorporating interpolation as described in Sec. 2.5. Notations -G, -F, and -K label computations of the GOF test statistics using the G-, F-, K-functions, respectively. The CSR, sine, Strauss, Thomas processes are defined in Sec. 2.3.1.*

	$\alpha = 0.05$						$\alpha = 0.10$					
	Standard			Interpolation			Standard			Interpolation		
	$\overline{\alpha}^*$	$\hat{\alpha}$	MSE	$\overline{\alpha}^*$	$\hat{\alpha}$	MSE	$\overline{\alpha}^*$	$\hat{\alpha}$	MSE	$\overline{\alpha}^*$	$\hat{\alpha}$	MSE
CSR-G	.029	.051	.103	.025	.046	.254	.077	.110	1.117	0.72	.098	.210
CSR-F	.027	.055	.334	.023	.046	.233	.072	.115	2.576	.069	.103	.263
CSR-K	.028	.057	.656	.023	.049	.100	.074	.115	2.394	.069	.099	.183
Sine-G	.058	.064	2.024	.052	.052	.279	.119	.116	2.765	.121	.103	.287
Sine-F	.109	.049	.103	.116	.044	.444	.192	.102	.223	.224	.095	.463
Sine-K	.031	.056	.419	.027	.048	.124	.079	.119	3.974	.078	.103	.287
Strauss-G	.088	.076	7.005	.086	.066	2.683	.169	.134	12.066	.182	.122	5.323
Strauss-F	.062	.062	1.605	.060	.054	.246	.127	.112	1.591	.136	.099	.197
Strauss-K	.036	.053	.217	.031	.043	.519	.088	.107	.626	.086	.092	.839
Thomas-G	.096	.045	.317	.100	.044	.493	.172	.091	1.048	.196	.092	.745
Thomas-F	.046	.051	.107	.044	.043	.545	.106	.105	.399	.115	.094	.485

*Reprinted with permission being requested from "A Monte Carlo-Adjusted Goodness-of-Fit Test for Parametric Models Describing Spatial Point Patterns" by Ngoc Anh Dao and Marc G. Genton, 2014. *Journal of Computational and Graphical Statistics*, 23:2, pages 497 - 517, Copyright 2021 by Taylor & Francis Academic Journals.

p-value equal to H/l has properties similar to those obtained by sampling with a fixed sample size, n. With this method, Monte Carlo simulations may become less computationally intensive as they can terminate much earlier.

Closing up this chapter, I continue to contribute to statistical inference in the next chapter by introducing a new family of clustered processes.

3. SKEW-ELLIPTICAL CLUSTER PROCESSES*

3.1 Chapter Overview

The Thomas process (TP) [Thomas, 1949] is very important in the field of spatial point processes because it has the intrinsic statistical ability to model propagation or clustering in nature. In particular, the TP is widely used in astronomy, biology and forestry, to name a few areas. In this chapter, I introduce a class of skew-elliptical cluster processes, which includes the (traditional) TP and offers the possibility of modeling the ellipticity, skewness, and, in some situations, information in the tail of the distribution of the "children" events. These characteristics would otherwise remain unknown if the (traditional) TP were used to model the data.

The TP is a special case of the Neymann–Scott cluster point process [Neyman and Scott, 1952], which is a specific type of homogeneous, independent clustering applied to a stationary Poisson process. Neyman and Scott [1952, 1958] and Neyman et al. [1953] used this process to model patterns formed by the locations of galaxies in space. Neyman and Scott [1972] gave further examples of such processes to model the distributions of insect larvae in fields and the geometry of bombing patterns. Penttinen et al. [1992] and Tanaka et al. [2008] used Neymann–Scott cluster point processes to model patterns of trees such as pines in natural forests. Similarly, Illian et al. [2008] used the TP to model 207 Phlebocarya filifolia plants and Diggle [2003, Chap. 6.3] used the TP to model 62 redwood seedlings.

I now look at how Neymann–Scott cluster point processes and the TP are defined. A Neymann–Scott cluster point process is constructed by letting unobservable, so-called "parent" events form a stationary Poisson process with intensity κ. The "children" events in a cluster are random in number and scattered independently and with identical distribution around each "parent" event. To construct a TP, a complete spatial randomness (CSR) with intensity κ is generated to obtain

*Reprinted with permission from "Skew-Elliptical Cluster Processes" by Ngoc Anh Dao and Marc G. Genton in I. Ghosh, N. Balakrishnan, and H. Ng, editors, *Advances in Statistics - Theory and Applications: Honoring the Contributions of Barry C. Arnold in Statistical Science*, pages 365-393. Springer, New York, 2021. Copyright 2021 by Copyright Clearance Center's RightsLink.

the "parent" events. Each "parent" event is replaced by a random cluster of "children" events, the number of which is Poisson distributed with intensity μ. The positions of the "children" events are distributed around the "parent" event according to a bivariate normal distribution with circular covariance matrix $\sigma^2\mathbf{I}_2$, where $\mathbf{I}_2$ is the 2×2 identity matrix [Thomas, 1949, Møller and Waagepetersen, 2003]. Stoyan and Stoyan [2006] introduced a generalized TP with small and large clusters and Tanaka et al. [2008] proposed a generalized Thomas model of type A, in which the probability density function (pdf) of the distance between the "children" events and their "parent" event corresponds to a mixture of distances from two Gaussian distributions with two different dispersion parameters.

Another generalizing work on TP was done by Castelloe [1998] by extending an isotropic bivariate normal offspring distribution to the case of a general bivariate normal offspring distribution. The extended process is no longer isotropic but anisotropic. The pair correlation function (pcf), a concept borrowed from physics, physical chemistry and statistical mechanics, is also commonly called a radial distribution function [McQuarrie, 1976, Chap. 13] and it describes how the density of points changes with the distance from a reference point. For the aforementioned processes, it is complicated and analytically incomplete. For the estimation, Castelloe [1998] considered a Bayesian approach. Further studies on extension of TP was done by Møller and Toftaker [2014] where anisotropic spatial point processes were introduced. There, Cox, shot noise Cox and log Gaussian Cox processes having elliptical pcf were studied. In this context, the TP was presented as a limiting case of the Whittle-Matérn shot noise Cox process. Møller and Toftaker [2014] applied a more sophisticated MCMC algorithm to the anisotropic cluster process proposed by Castelloe [1998]. However, the estimation still remains complicated and computationally intensive [Møller and Toftaker, 2014, p. 426].

Unlike Stoyan and Stoyan [2006] and Tanaka et al. [2008], but like Castelloe [1998] I generalize the TP in my approach not by introducing the pcf first but by presenting the general distribution of the "children" events. In particular, I impose a unified skew-elliptical (SUE) distribution [Arellano-Valle and Genton, 2010a] on them. The SUE family is a member of skewed multivariate models

[Arnold and Beaver, 2000] among which there are some other members with certain characteristics such as skewed multivariate models related to hidden truncation [Arnold and Beaver, 2002] and multivariate skew-normal distributions [Azzalini and Dalla Valle, 1996] to name a few. Although the SUE family includes a wide range of continuous distributions, I focus on only two representatives of this family here. They are the unified skew-normal (SUN) distribution [Arellano-Valle and Azzalini, 2006] and the extended skew-t (EST) distribution [Arellano-Valle and Genton, 2010b]. The reason for my focus on these two distributions is that, in contrast to other continuous distributions in the SUE family, their probabilistic properties have previously been intensively studied. With these results I can therefore carry out explicit theoretical derivations for approximation. If the "children" events are SUN distributed, then the process would be named a *skew-elliptical-normal* cluster process (CP). If they are EST distributed, then it would be called a *skew-elliptical-t* CP. These two classes of processes together give us *skew-elliptical* CP. It is obvious that a TP is simply a special case of the skew-elliptical-normal CP. Due to its circular shape of the "children" clusters induced by the dispersion matrix $\sigma^2\mathbf{I}_2$ of the bivariate normal distribution, a TP can also be called a circular-normal CP.

The introduction of the skew-elliptical-normal CP and the skew-elliptical-t CP is natural because datasets sometimes have non-circular patterns which need to be statistically modeled. If wind or the slope of a location caused the positions of the "children" events to be skew-elliptical distributed, the circular-normal CP (TP) would apparently be inferior to a skew-elliptical-normal or skew-elliptical-t CP. Without going into great details about these models, I motivate my approach by showing graphical representations of "children" events of skew-elliptical-normal CPs in Figure 3.1 and skew-elliptical-t CPs in Figure 3.2. The spatial point patterns (SPPs) are generated via `R` [R Core Team, 2019] using the same seed, 999, and all have $\kappa = 5$ and $\mu = 25$. The meanings of the dispersion parameters, σ_1, σ_2, and the skewness parameters, $\boldsymbol{\alpha}$ or $\boldsymbol{\alpha}_{\mathbf{Y}}$, of the skew-elliptical CPs are presented in Sections 3.2 and 3.3. In each of the first rows of Figures 3.1 and 3.2, the patterns of a circular-normal and a circular-t CP (left) have clusters in a circular shape induced by the isotropic dispersion matrix, $\sigma^2\mathbf{I}_2$, of the bivariate normal and t-distributions of the "children"

events. The patterns of an elliptical-normal and an elliptical-t CP (middle) have elliptically shaped clusters with the vertical dispersion double the horizontal one induced by the anisotropic dispersion matrix diag(σ_1^2, σ_2^2), with $\sigma_2 = 2\sigma_1$, of the bivariate normal and t-distribution of the "children" events. Castelloe [1998] dealt with the elliptical-normal CP. The patterns of a skew-normal and a skew-t CP (right) have clusters which are relocated further away from the diagonal reference line and skewed toward the upper-right corner. This shape is induced by the isotropic dispersion matrix, $\sigma^2\mathbf{I}_2$, the skewness parameter, $\boldsymbol{\alpha} = \alpha(1,1)^T$, of the bivariate skew-normal distribution according to Azzalini and Capitano [1999] and the skewness parameter, $\boldsymbol{\alpha}_{\mathbf{Y}}$ (Section 3.3), of the bivariate EST distribution according to Arellano-Valle and Genton [2010b]. The simulated patterns of skew-elliptical-t CPs with four degrees of freedom (df) in the first row of Figure 3.2 have more dispersed clusters than do those of the skew-elliptical-normal CPs, in Figure 3.1. This distinction is clearer in the second rows where the corresponding contour plots of the distribution of the "children" events are shown. In general, regardless of the df, the "children" events of skew-elliptical-t CPs are more dispersed than those of skew-elliptical-normal CP. The second rows also show the shapes of the clusters: circular (left), elliptical (middle), and skewed or squeezed (right), indicating that the "children" events are not symmetrically distributed around the "parent" event, but have fewer events in one particular quarter. In this example, the lower left quarter has fewer events, compared with the number of events in the other three quarters.

The theoretical summary descriptions, in particular pcfs, of the skew-elliptical CP's, except for the TP, are all analytical incomplete. As Castelloe [1998] and Møller and Toftaker [2014] I face the challenge in estimation using the Bayesian approach. However, if I relax the anisotropy condition to the assumption of isotropy, approximation of the pcf is analytical complete. Then, I make use of the minimum contrast method (MCM) for estimation because it is computationally easy, allowing quick exploration of a range of possible models. An estimation via MCM minimizes the discrepancy between the approximating pcf and the empirical pcf of the process. In my case, the minimizer of the discrepancy is the estimator of the parameters of the approximating pcf, but it is also good enough to be considered as the estimator of true parameters.

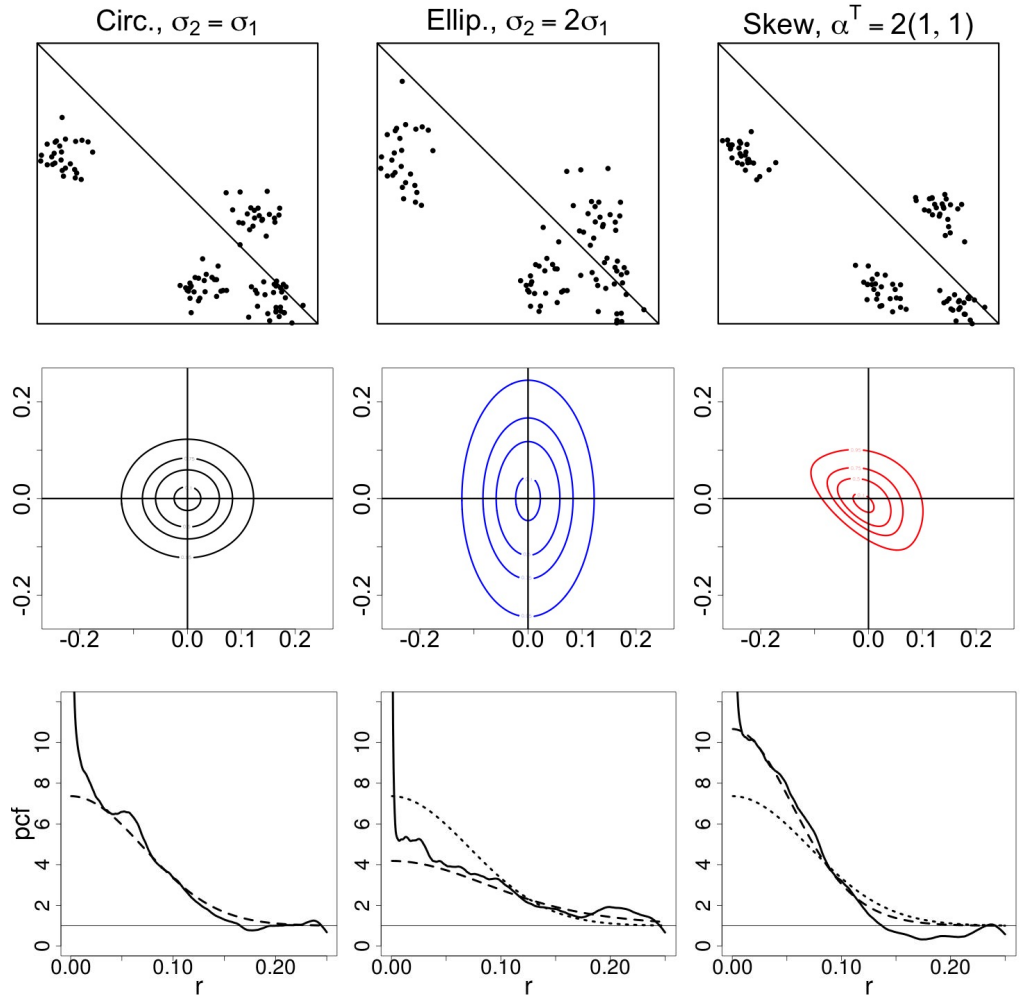


Figure 3.1: To generate spatial point patterns (SPPs), $\kappa = 5$, $\mu = 25$ and the same random seed were used. The first row shows a pattern of a circular-normal CP (TP) (left) with "children" events, $\mathbf{Y}$, being bivariate normal distributed with the isotropic dispersion matrix $\sigma^2\mathbf{I}_2$, $\sigma^2 = 0.05^2$; one of an elliptical-normal CP (middle) with $\mathbf{Y}$ being bivariate normal distributed with the anisotropic dispersion matrix with $\sigma_1^2 = 0.05^2$, $\sigma_2^2 = 0.10^2$ in the diagonal; one of a skew-normal CP (right) with $\mathbf{Y}$ being bivariate skew-normal distributed with the isotropic dispersion matrix $\sigma^2\mathbf{I}_2$, $\sigma^2 = .05^2$, and skewness parameter $\boldsymbol{\alpha} = 2(1,1)^T$. The parameters are described in Section 3.2. The diagonal line serves as a reference to better identify the difference in the cluster shape of $\mathbf{Y}$. In the corresponding column, the second row shows the contour plots of the distribution of $\mathbf{Y}$ of the CPs, the SPPs of which are shown in the first row: circular (black), elliptical (blue), skewed (red). The four contour levels from the most outer to the most inner level correspond to the 95th-, 75th-, 50th- and 10th- percentile of the distribution of $\mathbf{Y}$. The origin in the second row serves as an unobservable "parent" event. The third row shows the empirical pcf (solid) of the observed SPP from the corresponding first row, the theoretical pcf (dashed) of each model, and the theoretical pcf (dotted) of the circular-normal CP (TP) as a reference.*

*Reprinted with permission from "Skew-Elliptical Cluster Processes" by Ngoc Anh Dao and Marc G. Genton in I. Ghosh, N. Balakrishnan, and H. Ng, editors, *Advances in Statistics - Theory and Applications: Honoring the Contributions of Barry C. Arnold in Statistical Science*, pages 365-393. Springer, New York, 2021. Copyright 2021 by Copyright Clearance Center's RightsLink.

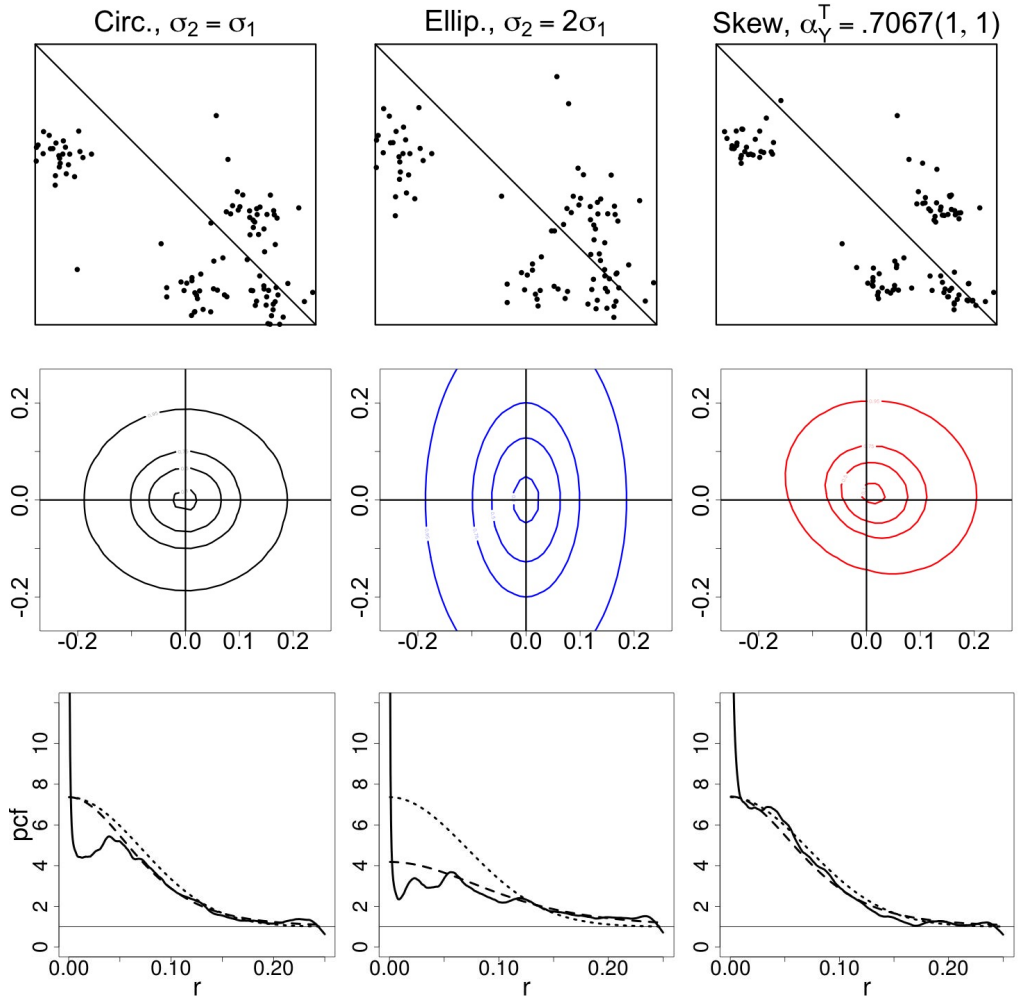


Figure 3.2: As in Figure 3.1, to generate SPPs of skew-elliptical-t CPs with four df ($\nu = 4$), $\kappa = 5$, $\mu = 25$ and the same random seed were used. The first row shows a pattern of a circular-t CP (left) with "children" events, $\mathbf{Y}$, being bivariate t-distributed with dispersion matrix $\sigma^2\mathbf{I}_2$ with $\sigma^2 = 0.05^2$, one of a elliptical-t CP (middle) with $\sigma_1^2 = 0.05^2$, $\sigma_2^2 = 0.10^2$, and one of a skew-t CP (right) with $\tau = 1$, $\sigma^2 = 0.05^2$, $\boldsymbol{\alpha}_{\mathbf{Y}}^T = (\alpha_1/\sqrt{1+\alpha_1^2+\alpha_2^2}, \alpha_2/\sqrt{1+\alpha_1^2+\alpha_2^2}) = (0.7067, 0.7067)$, where $\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2) = (20, 20)$. The roles of these parameters are described in Section 3.3. The diagonal line serves as a reference to better identify the difference in the cluster shape of $\mathbf{Y}$. In the corresponding column, the second row shows the contour plots of the distribution of $\mathbf{Y}$ of the CPs, the SPPs of which are shown in the first row: circular (black), elliptical (blue), skewed (red). The four contour levels from the most outer to the most inner level correspond to the 95th-, 75th-, 50th- and 10th- percentile of the distribution of $\mathbf{Y}$. The origin in the second row serves as an unobservable "parent" event. The third row shows the empirical pcf (solid) of the observed SPP from the corresponding first row, the theoretical pcf (dashed) of each model, and the theoretical pcf (dotted) of the circular-normal CP (TP) as a reference.*

*Reprinted with permission from "Skew-Elliptical Cluster Processes" by Ngoc Anh Dao and Marc G. Genton in I. Ghosh, N. Balakrishnan, and H. Ng, editors, *Advances in Statistics - Theory and Applications: Honoring the Contributions of Barry C. Arnold in Statistical Science*, pages 365-393. Springer, New York, 2021. Copyright 2021 by Copyright Clearance Center's RightsLink.

This chapter is organized with the following structure. Sections 3.2 and 3.3 present the pcfs of the skew-elliptical-normal and the corresponding skew-elliptical-t CPs. Some analytical derivations were carried out with `Mathematica` [Wolfram Research, 2020]. The intermediate derivation steps are given in the Appendix. Section 3.4 demonstrates the performance of parameter estimation via the MCM using the function `optim` available in `R` [R Core Team, 2019]. Section 3.5 provides a data application of these skew-elliptical CPs on a fraction of the dataset called `fullredwood` available in the `R`-library `spatstat` [Baddeley and Turner, 2005a, Baddeley et al., 2015]. Finally, Section 3.6 introduces alternative probability distributions to extend my work on TP, suggests to generalize a similar clustered spatial point process, and raises a possible exploration for an adjustment of the MCM.

3.2 Skew-Elliptical-Normal Cluster Processes

3.2.1 Distributions of "Children" Events

Let $\mathbf{Y}$, the random vector representing the position of the "children" event in a cluster, be bivariate skew-normal distributed with skewness parameter vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$, location parameter $-\boldsymbol{\omega\delta}\sqrt{2/\pi}$ where $\boldsymbol{\delta} = \boldsymbol{\alpha}/\sqrt{1+\boldsymbol{\alpha}^T\boldsymbol{\alpha}}$, dispersion matrix $\boldsymbol{\Omega} = \text{diag}(\sigma_1^2, \sigma_2^2)$ with $\sigma_1 > 0$, $\sigma_2 > 0$, and $\boldsymbol{\omega} = \text{diag}(\boldsymbol{\Omega}^{1/2})$. In short, $\mathbf{Y} \sim \text{SN}_2(-\boldsymbol{\omega\delta}\sqrt{2/\pi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$. In particular, its pdf is $f_{\mathbf{Y}}(\mathbf{y}) = 2\phi_2(\mathbf{y} + \boldsymbol{\omega\delta}\sqrt{2/\pi}; \boldsymbol{\Omega})\ \Phi\{\boldsymbol{\alpha}^T\boldsymbol{\omega}^{-1}(\mathbf{y} + \boldsymbol{\omega\delta}\sqrt{2/\pi})\}$, where $\phi(\cdot)$, $\Phi(\cdot)$ denote the pdf and cumulative distribution function (cdf) of the univariate standard normal distribution, $\phi_2(\cdot;)$ and $\Phi_2(\cdot;)$ the corresponding functions of the bivariate normal distribution and $\mathbf{y} = (y_1, y_2)^T$ [Azzalini and Dalla Valle, 1996, Arellano-Valle and Azzalini, 2006]. Then $\text{E}(\mathbf{Y}) = \mathbf{0}$ and $\text{Var}(\mathbf{Y}) = \boldsymbol{\Omega} - \frac{2}{\pi}\boldsymbol{\omega\delta\delta}^T\boldsymbol{\omega}$ [Azzalini and Capitano, 1999, Gupta et al., 2013]. Then $\mathbf{Y}$ is also a unified skew-normal (SUN) random vector [Arellano-Valle and Azzalini, 2006, Azzalini and Capitano, 2014]. It is important to state here that the SUN distribution has the additive property. In general, the SUN distribution introduced by Arellano-Valle and Azzalini [2006] generalizes the parametrization of several variants of the original multivariate skew-normal distribution developed by Azzalini and Dalla Valle [1996]. To name a few of these variants, there are the closed skew-normal of González-

Farías et al. [2004], the hierarchical skew-normal of Liseo and Loperfido [2003], the fundamental skew-normal of Arellano-Valle and Genton [2005], and the multivariate skew-normal of Gupta et al. [2004].

It is advantageous to use the notation according to Azzalini and Capitano [2014]: $\mathbf{Y}$ has distribution denoted by $\text{SUN}_{2,1}(-\boldsymbol{\omega}\boldsymbol{\delta}\sqrt{2/\pi}, \boldsymbol{\Omega}, \boldsymbol{\delta}, 0, 1)$. For $\mathbf{X} \overset{d}{=} \mathbf{Y}_1 - \mathbf{Y}_2$, where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are two independent "children" events within a cluster, due to the additive property, $\mathbf{X} \sim \text{SUN}_{2,2}\left(\mathbf{0}, 2\boldsymbol{\Omega}, \boldsymbol{\Delta}, 0, \mathbf{I}_2\right)$ [Azzalini and Capitano, 2014, Ch. 7], where $\boldsymbol{\Delta} = \boldsymbol{\delta}/\sqrt{2}\,(1,\ -1)$, i.e., the pdf of $\mathbf{X}$ is $f_{\mathbf{X}}(\mathbf{x}) = 4\phi_2\left(\mathbf{x}; 2\boldsymbol{\Omega}\right)\ \Phi_2(\boldsymbol{\Delta}^T\boldsymbol{\omega}^{-1}\mathbf{x}/\sqrt{2}; \mathbf{I}_2 - \boldsymbol{\Delta}^T\boldsymbol{\Delta})$. Explicitly,

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\exp\left(-\frac{\sigma_2^2x_1^2+\sigma_1^2x_2^2}{4\sigma_1^2\sigma_2^2}\right)}{\pi\sigma_1\sigma_2} \quad \Phi_2\left\{\frac{\left(\frac{\alpha_1x_1}{\sigma_1}+\frac{\alpha_2x_2}{\sigma_2}\right)\begin{pmatrix}1\\-1\end{pmatrix}}{2\sqrt{1+\alpha_1^2+\alpha_2^2}}; \frac{\begin{pmatrix}2+\alpha_1^2+\alpha_2^2 & \alpha_1^2+\alpha_2^2\\ \alpha_1^2+\alpha_2^2 & 2+\alpha_1^2+\alpha_2^2\end{pmatrix}}{2(1+\alpha_1^2+\alpha_2^2)}\right\},$$

where $\mathbf{x} = (x_1, x_2)^T$. Note that the distribution of $\mathbf{X}$ shown above is centrally symmetric. The reason for the symmetry is that $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are identically distributed. Hence, $\mathbf{X} = \mathbf{Y}_1 - \mathbf{Y}_2$ and $-\mathbf{X} = \mathbf{Y}_2 - \mathbf{Y}_1 = -(\mathbf{Y}_1 - \mathbf{Y}_2)$ have the same distribution.

3.2.2 Approximation of the Pair Correlation Function

The usual way of defining the pcf of an anisotropic spatial point process is $g(\mathbf{u}, \mathbf{v}) = \lambda^{(2)}(\mathbf{u}, \mathbf{v})/[\lambda(\mathbf{u})\lambda(\mathbf{v})]$ where $\lambda^{(2)}(\mathbf{u}, \mathbf{v})$ is the second-order product density and λ is the intensity function. In my setting, g is anisotropic but translation invariant, $g(\mathbf{u}, \mathbf{v}) = g(\mathbf{v} - \mathbf{u})$, I obtain

$$K(r) = \int_{\mathbb{R}^2} 1_{[\|\mathbf{h}\| \le r]} g(\mathbf{h}) d\mathbf{h},$$

where $r > 0$ and $1_{[\|\mathbf{h}\| \le r]}$ is an indicator function. I will then approximate g by g_d where the subscript d stands for distance and where g_d will be an isotropic function, i.e., $g_d(r)$ with $r = \|\mathbf{h}\|$. Then $g_d(r) = K_d'(r)/(2\pi r)$ where $K_d'(r) = \partial K_d(r)/\partial r$.

For MCM, the most popular choice for theoretical summary description is the second-order characteristic known as Ripley's K-function [Ripley, 1976]. The information from Ripley's K-function is the expected number of events found within a distance r from an event of interest, $K(r) = \mathrm{E}[N\{b(\mathbf{o}, r)\}]/\lambda$, where N denotes the number of events within a disc, $b(\mathbf{o}, r)$, of radius $r \geq 0$ at the event of interest $\mathbf{o}$, and λ denotes the global intensity of the process. However, according to Illian et al. [2008, Chap. 4.3.1], the pcf offers the best statistical way to represent the distributional information contained in the point patterns. Additionally, the advantage of using g_d here is that while most approximating pcfs g_d are analytically complete, their corresponding K_d-functions are not. I therefore focus on deriving g_d and provide K_d only if they are analytically complete.

Under the relaxed assumption of isotropy, to derive K_d and g_d, I calculate the distribution of the Euclidean distance, or lag, $R = \sqrt{(\mathbf{Y}_1 - \mathbf{Y}_2)^T(\mathbf{Y}_1 - \mathbf{Y}_2)} = \sqrt{\mathbf{X}^T\mathbf{X}}$, where the $\mathbf{Y}_i$'s represent two independent "children" events within a cluster. They are independently and identically distributed bivariate random vectors and R is the random variable representing the lag between two randomly distributed "children" events in a cluster under the assumption of isotropy. I first derive its cdf, $F_d(r)$, since $K_d(r) = \pi r^2 + F_d(r)/\kappa$ [Cressie, 1993]. Then the pcf is $g_d(r) = 1 + F_d'(r)/(2\pi\kappa r) = 1 + f_d(r)/(2\pi\kappa r)$, where $f_d(r)$ is the pdf of R.

I consider the following transformation with $R \geq 0$, $0 \leq \Theta \leq 2\pi$, $\mathbf{X} = (X_1, X_2)^T$,

$$\begin{aligned} X_1 = R\cos\Theta, \qquad & X_2 = R\sin\Theta, \text{ and} \\ R = \sqrt{\mathbf{X}^T\mathbf{X}} = \sqrt{X_1^2 + X_2^2}, \qquad & \Theta = \arctan(X_2/X_1). \end{aligned} \tag{3.1}$$

The determinant of the Jacobian matrix is $|\partial(r,\theta)/\partial(x_1,x_2)| = 1/r$. Thus, $f_{R,\Theta}(r,\theta) = r f_{X_1,X_2}(r\cos\theta, r\sin\theta)$, $f_d(r) = \int_0^{2\pi} f_{R,\Theta}(r,\theta)d\theta$, and $F_d(r) = \int_0^r f_d(t)dt$. From (3.1), the joint distribution, $f_{R,\Theta}(r,\theta)$, is

derived in (B.1). The pdf of R follows easily and, from $g_d(r) = 1 + f_d(r)/(2\pi\kappa r)$, the pcf is

$$g_d(r) = 1 + \int_0^{2\pi} \frac{\exp\left(-\frac{\sigma_2^2 r^2 \cos^2\theta + \sigma_1^2 r^2 \sin^2\theta}{4\sigma_1^2\sigma_2^2}\right)}{2\pi^2\kappa\sigma_1\sigma_2} \times \Phi_2\left\{\frac{\left(\frac{\alpha_1 r\cos\theta}{\sigma_1} + \frac{\alpha_2 r\sin\theta}{\sigma_2}\right)\begin{pmatrix}1\\-1\end{pmatrix}}{2\sqrt{1+\alpha_1^2+\alpha_2^2}}; \frac{\begin{pmatrix}2+\alpha_1^2+\alpha_2^2 & \alpha_1^2+\alpha_2^2\\ \alpha_1^2+\alpha_2^2 & 2+\alpha_1^2+\alpha_2^2\end{pmatrix}}{2(1+\alpha_1^2+\alpha_2^2)}\right\} d\theta. \quad (3.2)$$

For $\boldsymbol{\alpha} \neq \mathbf{0}$, the pcf is analytically incomplete since the integration over the analytically incomplete function, $\Phi_2(\cdot;)$, is analytically incomplete. In particular, the pcf becomes analytically complete if $\alpha_1 = \alpha_2 = 0$, i.e., $\Phi_2(\cdot;) = \Phi_2\left\{(0,0)^T; \mathbf{I}_2\right\} = 1/4$.

3.2.3 The Elliptical-Normal Cluster Process

Now assume that $\sigma_1 \neq \sigma_2$ and $\boldsymbol{\alpha} = \mathbf{0}$. That is, $\mathbf{Y}$ is bivariate normal distributed, i.e., $\mathbf{Y} \sim \mathrm{N}_2(\mathbf{0}, \boldsymbol{\Omega})$. Here, the distribution of the "children" events is elliptical around the "parent" event and the skewness parameter, $\boldsymbol{\alpha}$, is not present. Then, from (3.2), the approximating pcf is

$$g_d(r) = 1 + \frac{1}{4\pi\kappa\sigma_1\sigma_2}\exp\left\{-\frac{(\sigma_1^2+\sigma_2^2)r^2}{8\sigma_1^2\sigma_2^2}\right\}\mathrm{BesselI}_0\left\{\frac{(\sigma_1^2-\sigma_2^2)r^2}{8\sigma_1^2\sigma_2^2}\right\},$$

where $\mathrm{BesselI}_0(x) = \sum_{n=0}^{\infty}(x/2)^{2n}/(n!)^2$ is a modified Bessel function of the first kind. A different parametrization, $\sigma_1 \equiv \sigma$ and $\sigma_2 = c_\sigma\sigma$ with $c_\sigma > 0$, can be beneficial in parameter estimation with respect to identifiability because I no longer have two dispersion parameters as above but have one dispersion and its scaling parameter instead,

$$g_d(r) = 1 + \frac{1}{4\pi\kappa c_\sigma\sigma^2}\exp\left\{-\frac{(1+c_\sigma^2)r^2}{8c_\sigma^2\sigma^2}\right\}\mathrm{BesselI}_0\left\{\frac{(1-c_\sigma^2)r^2}{8c_\sigma^2\sigma^2}\right\}.$$

K_d of the elliptical-normal CP is not analytically complete. I estimate κ, σ^2, and c_σ^2 using g_d via the MCM.

As mentioned in Section 3.1, Stoyan and Stoyan [2006] and Tanaka et al. [2008] introduced different models with more than one dispersion parameters to generalize the (traditional) TP. For comparison, I provide the pdf of R of my model in (B.2) and (B.3) in the Appendix.

3.2.4 The Circular-Normal Cluster Process

Assume that $\boldsymbol{\alpha} = \mathbf{0}$ and $\sigma_1 = \sigma_2 = \sigma$ for the distribution of $\mathbf{Y}$. That is, "children" events are distributed symmetrically circular around their "parent" event. The corresponding process is the traditional TP and is isotropic. For completeness, $f_d(r) = f_R(r)$ is provided in (B.4) in the Appendix. From (3.2), the true pcf is

$$g(r) \;=\; 1 + \frac{\exp\{-r^2/(4\sigma^2)\}}{4\pi\kappa\sigma^2}. \tag{3.3}$$

The true K-function can be computed as $K(r) = \int_0^r 2\pi t\; g(t)\; dt = \pi r^2 + [1 - \exp\{-r^2/(4\sigma^2)\}]/\kappa$. This formula of the K-function has been widely used prior to this work; e.g., it can be found in Cressie [1993]. To estimate κ and σ^2, the MCM can use either the pcf or the K-function.

3.2.5 The Skew-Normal Cluster Process

Let the distribution of $\mathbf{Y}$ be a special case of the SUN distribution mentioned earlier in Section 3.2.1 with $\sigma_1 = \sigma_2 = \sigma$. For a scalar $\sigma > 0$ and a bivariate vector $\boldsymbol{\delta} = \boldsymbol{\alpha}/\sqrt{1 + \boldsymbol{\alpha}^T\boldsymbol{\alpha}}$ with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$, assume that $\mathbf{Y} = -\boldsymbol{\delta}\sigma\sqrt{2/\pi} + \boldsymbol{\delta}\sigma V_0 + \sigma\mathbf{V}_1$, where V_0 and $\mathbf{V}_1$ are an independent random variable and vector, respectively. Here, V_0 follows the univariate standard normal distribution truncated below 0 with $E(V_0) = \sqrt{2/\pi}$, $\text{Var}(V_0) = 1 - 2/\pi$, and $\mathbf{V}_1$ is bivariate normal distributed, $\text{N}_2(\mathbf{0}, \boldsymbol{\Psi})$, where

$$\boldsymbol{\Psi} = \mathbf{I}_2 - \boldsymbol{\delta}\boldsymbol{\delta}^T = \mathbf{I}_2 - \boldsymbol{\alpha}\boldsymbol{\alpha}^T/(1 + \boldsymbol{\alpha}^T\boldsymbol{\alpha}) = \begin{pmatrix} 1 + \alpha_2^2 & -\alpha_1\alpha_2 \\ -\alpha_1\alpha_2 & 1 + \alpha_1^2 \end{pmatrix} /(1 + \alpha_1^2 + \alpha_2^2)$$

is a correlation matrix. Under this setting, according to Arellano-Valle and Azzalini [2006, Sec. 2.1.], $\mathbf{Y}$ is bivariate SUN distributed, in particular $\text{E}(\mathbf{Y}) = \mathbf{0}$, $\text{Var}(\mathbf{Y}) = \sigma^2(\mathbf{I}_2 - 2/\pi\boldsymbol{\delta}\boldsymbol{\delta}^T)$. This

distribution is purely skewed and does not have any elliptical property. For $\mathbf{Y}_1$ and $\mathbf{Y}_2$ representing two independent positions of the "children" events in a cluster, $\mathbf{X} \stackrel{d}{=} \mathbf{Y}_1 - \mathbf{Y}_2$ has the pdf f_X in Section 3.2.1 with $\sigma_1 = \sigma_2 = \sigma$. Since it is centrally symmetric, I approximate it with a $\mathrm{N}_2\left(\mathbf{0}, 2\sigma^2(\mathbf{I}_2 - 2/\pi\boldsymbol{\delta}\boldsymbol{\delta}^T)\right)$ distribution, i.e., bivariate normal with pdf

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2c_0\sqrt{\pi^2 c_1 c_2 - 4\alpha_1^2\alpha_2^2}} \exp\left[-\frac{\pi\{\pi(c_2x_1^2 + c_1x_2^2) + 4\alpha_1\alpha_2x_1x_2\}}{2c_0(\pi^2c_1c_2 - 4\alpha_1^2\alpha_2^2)}\right],$$

where $c_0 = 2\sigma^2/(1 + \alpha_1^2 + \alpha_2^2)$, $c_1 = 1 + \alpha_1^2(1 - 2/\pi) + \alpha_2^2$, and $c_2 = 1 + \alpha_1^2 + \alpha_2^2(1 - 2/\pi)$. The joint distribution, $f_{R,\Theta}(r, \theta)$, is given in (B.5) in the Appendix. The pcf, $g(r)$, is analytically complete only in the following two cases. First, assume that $\alpha_1^2 = \alpha_2^2$, i.e., (i) $\boldsymbol{\alpha} = \alpha(1, 1)^T$, (ii) $\boldsymbol{\alpha} = \alpha(-1, -1)^T$, (iii) $\boldsymbol{\alpha} = \alpha(1, -1)^T$, or (iv) $\boldsymbol{\alpha} = \alpha(-1, 1)^T$, for $\alpha > 0$. Then, $f_d(r)$ is given in (B.6) in the Appendix. Consequently, $g_d(r)$ is

$$\begin{aligned} g_d(r) &= 1 + \frac{\sqrt{1 + 2\alpha^2}}{4\kappa\sigma^2\sqrt{\pi\{\pi(1 + 2\alpha^2) - 4\alpha^2\}}} \exp\left[-\frac{\{\pi + 2\alpha^2(\pi - 1)\}r^2}{4\sigma^2\{\pi(1 + 2\alpha^2) - 4\alpha^2\}}\right] \\ &\quad \times \text{BesselI}_0\left[\frac{\alpha^2r^2}{2\sigma^2\{\pi(1 + 2\alpha^2) - 4\alpha^2\}}\right]. \end{aligned}$$

Second, suppose that (i) $\boldsymbol{\alpha} = (0, \alpha)^T$ or (ii) $\boldsymbol{\alpha} = (\alpha, 0)^T$. Then, $f_d(r)$ is given in (B.7) in the Appendix. Consequently, $g_d(r)$ is

$$\begin{aligned} g_d(r) &= 1 + \frac{r(1 + \alpha^2)}{4\pi\kappa\sigma^2\sqrt{(1 + \alpha^2)\{1 + \alpha^2(1 - 2/\pi)\}}} \exp\left[-\frac{r^2\{1 + \alpha^2(1 - 1/\pi)\}}{4\sigma^2\{1 + \alpha^2(1 - 2/\pi)\}}\right] \\ &\quad \times \text{BesselI}_0\left[\frac{\alpha^2r^2}{4\pi\sigma^2\{1 + \alpha^2(1 - 2/\pi)\}}\right]. \end{aligned}$$

K_d of the above scenarios are analytically incomplete. I estimate κ, σ^2, and α^2 via MCM using g_d. The complete determination of $\boldsymbol{\alpha}$ results from choosing the optimal $\hat{\boldsymbol{\alpha}}$ from the above possibilities such that the cluster shape of simulated SPP can illustrate that of the observed SPP as best as possible.

Remark 3. *So far I have emphasized on presenting CPs having the approximating pcf g_d as being*

analytically complete because they are advantageous in MCM. In practice, however, for CP having only analytically incomplete pcfs or K-functions, the parameter estimation can still be carried out, for example with a Bayesian approach but the computation is more intensive.

3.3 Skew-Elliptical-t Cluster Processes

3.3.1 General Scenario and Relaxing Independence

Let $\mathbf{Y}$ be the bivariate random vector representing the position of a "children" event in a cluster and let $(\mathbf{Y}^T, \mathbf{Y}^{*T})^T$ be four-variate extended skew-t (EST) distributed, i.e., $\text{EST}_4(\mathbf{0}, \text{diag}(\boldsymbol{\Omega}, \boldsymbol{\Omega}), (\boldsymbol{\alpha}^T, \boldsymbol{\alpha}^T)^T, \nu, \tau)$ [Arellano-Valle and Genton, 2010b] with a 4×4 dispersion matrix $\text{diag}(\boldsymbol{\Omega}, \boldsymbol{\Omega})$, four variate shape parameter $(\boldsymbol{\alpha}^T, \boldsymbol{\alpha}^T)^T$, ν df, and extension parameter $\tau \in \mathbb{R}$, where the 2×2 matrix $\boldsymbol{\Omega} = \text{diag}(\sigma_1^2, \sigma_2^2)$ and the bivariate vector $\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2)$. According to Arellano-Valle and Genton [2010b, Prop. 3], the marginal distribution of $\mathbf{Y}$ is also EST distributed: $\mathbf{Y} \sim \text{EST}_2(\mathbf{0}, \boldsymbol{\Omega}, \boldsymbol{\alpha}_\mathbf{Y}, \nu, \tau_\mathbf{Y})$ where $\boldsymbol{\alpha}_\mathbf{Y} = \boldsymbol{\alpha}/\sqrt{1 + \boldsymbol{\alpha}^T\boldsymbol{\alpha}}$ is termed as marginal shape parameter and $\tau_\mathbf{Y} = \tau/\sqrt{1 + \boldsymbol{\alpha}^T\boldsymbol{\alpha}}$ is termed as marginal extension parameter. Note that (i) $\boldsymbol{\alpha}_\mathbf{Y}$ is not necessary in the setting of skew-elliptical-normal CPs, because there $\boldsymbol{\alpha} = \boldsymbol{\alpha}_\mathbf{Y}$ and (ii) the statistical characteristic of $\boldsymbol{\alpha}_\mathbf{Y}$ of a skew-t CP is equivalent to that of $\boldsymbol{\alpha}$ of a skew-normal CP. Moreover, for simplicity I have set the location parameter to zero but it could be adjusted to yield $\text{E}(\mathbf{Y}) = \mathbf{0}$ with the results of Section 2.3 in Arellano-Valle and Genton [2010b]. From Proposition 5 of the same paper, I derive that $\mathbf{X} = \mathbf{Y} - \mathbf{Y}^* \sim \text{EST}_2(\mathbf{0}, 2\boldsymbol{\Omega}, \mathbf{0}, \nu, \tau/\sqrt{1 + 2\boldsymbol{\alpha}^T\boldsymbol{\alpha}})$.

Although $\boldsymbol{\alpha}$ is neither the shape parameter of the distribution of $(\mathbf{Y}^T, \mathbf{Y}^{*T})^T$ nor of $\mathbf{Y}$, it is important in the setting of skew-elliptical-t CPs. First, it contributes to the shape of the distribution of $(\mathbf{Y}^T, \mathbf{Y}^{*T})^T$ and of $\mathbf{Y}$; thus I know how the "children" events are distributed and know how the process is constructed. Second, $\boldsymbol{\alpha}$, but not $\boldsymbol{\alpha}_\mathbf{Y}$, appears in the formulas of the pcfs of skew-t CPs; hence, I can use the theoretical pcf to estimate the parameter $\boldsymbol{\alpha}$ and then compute $\boldsymbol{\alpha}_\mathbf{Y}$. The task of $\boldsymbol{\alpha}_\mathbf{Y}$ is to describe the shape of the marginal distribution of $\mathbf{Y}$: the cluster shape of the process.

For independent and identically distributed children $\mathbf{Y}_i, \mathbf{Y}_j$ with $i \neq j$ in a cluster, $\mathbf{X}_{\text{true}} \stackrel{d}{=} \mathbf{Y}_i - \mathbf{Y}_j$ is not bivariate EST, unified skew-t, or bivariate t distributed. In fact, its distribution is

unknown. The only sub-family of the skew-elliptical distributions that has the additive property is the SUN family [Arellano-Valle and Genton, 2010b,a, González-Farías et al., 2004]. I chose to approximate the distribution of $\mathbf{X}_{\text{true}}$ by $\mathbf{X}$, i.e., $\mathbf{X}_{\text{true}} \overset{d}{\approx} \mathbf{X} \sim \text{EST}_2(\mathbf{0}, 2\boldsymbol{\Omega}, \mathbf{0}, \nu, \tau/\sqrt{1+2\boldsymbol{\alpha}^T\boldsymbol{\alpha}})$. Note that the distribution of $\mathbf{X}$ loses the information about $\boldsymbol{\alpha}$ of the distribution of $\mathbf{Y}$ if $\tau = 0$. The pdf of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{T_1\left\{\frac{\tau}{\sqrt{1+2\boldsymbol{\alpha}^T\boldsymbol{\alpha}}}\left(\frac{\nu+2}{\nu+\mathbf{x}^T\boldsymbol{\Omega}^{-1}\mathbf{x}/2}\right)^{1/2}; \nu+2\right\}}{2\pi|2\boldsymbol{\Omega}|^{1/2}\left(1+\frac{\mathbf{x}^T\boldsymbol{\Omega}^{-1}\mathbf{x}/2}{\nu}\right)^{(\nu+2)/2} T_1\left(\frac{\tau}{\sqrt{1+2\boldsymbol{\alpha}^T\boldsymbol{\alpha}}}; \nu\right)},$$

where $T_1(\cdot;\nu)$ denotes the cdf of the univariate t-distribution with ν degrees of freedom. The explicit form of $f_{\mathbf{X}}(\mathbf{x})$ is given in (B.8) and under the isotropy assumption, the joint distribution function, $f_{d,R,\Theta}(r,\theta)$, is provided in (B.9). If $\boldsymbol{\alpha} \neq \mathbf{0}$ and $\sigma_1 \neq \sigma_2$, the approximating pcf is analytically incomplete:

$$g_d(r) = 1 + \frac{1}{8\pi^2\kappa\sigma_1\sigma_2 T_1\left\{\frac{\tau}{\sqrt{1+2(\alpha_1^2+\alpha_2^2)}}; \nu\right\}} \times \int_0^{2\pi} \frac{T_1\left[\frac{\tau}{\sqrt{1+2(\alpha_1^2+\alpha_2^2)}}\left\{\frac{\nu+2}{\nu+\left(r^2\cos^2\theta/\sigma_1^2+r^2\sin^2\theta/\sigma_2^2\right)/2}\right\}^{1/2}; \nu+2\right]}{\left(1+\frac{r^2\cos^2\theta/\sigma_1^2+r^2\sin^2\theta/\sigma_2^2}{2\nu}\right)^{(\nu+2)/2}} d\theta. \tag{3.4}$$

For MCM, I use a sequence of ν. For each value of ν, I estimate the other parameters. Then I choose the set of estimates and corresponding ν that provide the smallest discrepancy between the approximating and the empirical pcfs.

3.3.2 The Skew-t Cluster Process

If $\boldsymbol{\alpha} \neq \mathbf{0}$ and $\sigma_1 = \sigma_2$, i.e., $(\mathbf{Y}^T, \mathbf{Y}^{*T})^T \sim \text{EST}_4(\mathbf{0}, \sigma^2\mathbf{I}_4, (\boldsymbol{\alpha}^T, \boldsymbol{\alpha}^T)^T, \nu, \tau)$ [Arellano-Valle and Genton, 2010b] and hence $\mathbf{Y} \sim \text{EST}_2(\mathbf{0}, \sigma^2\mathbf{I}_2, \boldsymbol{\alpha}_{\mathbf{Y}}, \nu, \tau_{\mathbf{Y}})$ where $\boldsymbol{\alpha}_{\mathbf{Y}} = \boldsymbol{\alpha}/\sqrt{1+\boldsymbol{\alpha}^T\boldsymbol{\alpha}}$ and $\tau_{\mathbf{Y}} = \tau/\sqrt{1+\boldsymbol{\alpha}^T\boldsymbol{\alpha}}$, I obtain the following approximating pcf under isotropy assumption from

(3.4):

$$g_d(r) \;=\; 1 + \frac{T_1\left[\frac{\tau}{\sqrt{1+2(\alpha_1^2+\alpha_2^2)}}\left\{\frac{\nu+2}{\nu+r^2/(2\sigma^2)}\right\}^{1/2};\nu+2\right]}{4\pi\kappa\sigma^2\, T_1\left\{\frac{\tau}{\sqrt{1+2(\alpha_1^2+\alpha_2^2)}};\nu\right\}\left(1+\frac{r^2}{2\nu\sigma^2}\right)^{(\nu+2)/2}}.$$

The previous formula has two parameters, α_1 and α_2, of the same role: they both contribute to the skewness of the distribution of $\mathbf{Y}$. A different parametrization, $\alpha_1 = \alpha$ and $\alpha_2 = c_\alpha\alpha$ with c_α being a real constant, can be useful for parameter estimation.

K_d of the skew-t CP is analytically incomplete. For the parameter estimation, I estimate κ, σ^2, α^2 and c_α^2 via MCM using the pcf. Then, I can compute the estimates of $\alpha_{\mathbf{Y},1}$ and $\alpha_{\mathbf{Y},2}$, because eventually I am interested in knowing the estimate of the skewness parameter of the skew-t CP, which is $\boldsymbol{\alpha}_{\mathbf{Y}}$, not $\boldsymbol{\alpha}$.

Remark 4. *Recall that $\alpha_{\mathbf{Y},i} = \alpha_i/\sqrt{1+\alpha_1^2+\alpha_2^2}$, $i = 1, 2$. Thus, they have absolute values less than 1, i.e. $|\alpha_{\mathbf{Y},1}| < 1$, $|\alpha_{\mathbf{Y},2}| < 1$, although the absolute values of α_1 and α_2 can be large. For $|\alpha_{\mathbf{Y},1}| = |\alpha_{\mathbf{Y},2}|$, their absolute values can be at most $1/\sqrt{2} \approx 0.7071$. Consequently, only skew-normal CPs with skewness parameters having absolute values smaller than $1/\sqrt{2}$ can be considered to be approximated by a skew-t CP with large df. A demonstration of this statement is given in Section 3.5.*

3.3.3 The Elliptical-t Cluster Process

If $\tau = 0$, $\boldsymbol{\alpha} = \mathbf{0}$, but $\sigma_1 \neq \sigma_2$, i.e., $(\mathbf{Y}^T, \mathbf{Y}^{*T})^T \sim \mathrm{t}_\nu((\mathbf{0}^T, \mathbf{0}^T)^T, \mathrm{diag}(\boldsymbol{\Omega}, \boldsymbol{\Omega}))$ where $\boldsymbol{\Omega} = \mathrm{diag}(\sigma_1^2, \sigma_2^2)$, and hence $\mathbf{Y} \sim \mathrm{t}_\nu(\mathbf{0}, \boldsymbol{\Omega})$ where t_ν is the multivariate Student t-distribution with ν df, then the approximating pcf is

$$g_d(r) \;=\; 1 + \frac{1}{8\pi^2\kappa\sigma_1\sigma_2}\int_0^{2\pi}\left(1+\frac{r^2\cos^2\theta/\sigma_1^2 + r^2\sin^2\theta/\sigma_2^2}{2\nu}\right)^{-(\nu+2)/2} d\theta.$$

For $\nu = 1$, $g(r) \equiv 1$. Only for $\nu = 2k$, where $k \in \mathbb{N}$, is $g(r)$ analytically complete. For each even df, I have to compute the approximating pcf individually since there is no general formula for the

pcf. `Mathematica` can compute up to 26 df analytically. For df greater than 6, the formulas of pcfs are very cumbersome and can take several rows to be displayed. I choose to represent the pcfs only for $\nu = 2$, 4 and 6 in the following. For $\nu = 2$,

$$g_d(r) = 1 + \frac{2\{(\sigma_1^2 + \sigma_2^2)r^2 + 8\sigma_1^2\sigma_2^2\}}{\pi\kappa\{(r^2 + 4\sigma_1^2)(r^2 + 4\sigma_2^2)\}^{3/2}},$$

$\nu = 4$,

$$g_d(r) = 1 + \frac{16\left\{512\sigma_1^4\sigma_2^4 + 64\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)r^2 + (3\sigma_1^4 + 2\sigma_1^2\sigma_2^2 + 3\sigma_2^4)r^4\right\}}{\pi\kappa\{(r^2 + 8\sigma_1^2)(r^2 + 8\sigma_2^2)\}^{5/2}},$$

and $\nu = 6$,

$$g_d(r) = 1 + \frac{324\left[\{24\sigma_1^2\sigma_2^2 + (\sigma_1^2 + \sigma_2^2)r^2\}\{1152\sigma_1^4\sigma_2^4 + 96\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)r^2 + (5\sigma_1^4 - 2\sigma_1^2\sigma_2^2 + 5\sigma_2^4)r^4\}\right]}{\pi\kappa\{(r^2 + 12\sigma_1^2)(r^2 + 12\sigma_2^2)\}^{7/2}}.$$

K_d of the elliptical-t CP are analytically incomplete regardless of df. I can estimate κ, $\sigma_1^2 = \sigma^2$ and c_σ^2, where $\sigma_2^2 = c_\sigma^2\sigma^2$, via MCM using the pcf.

3.3.4 The Circular-t Cluster Process

If $\tau = 0$, $\boldsymbol{\alpha} = \mathbf{0}$ and $\sigma_1 = \sigma_2$, i.e., $(\mathbf{Y}^T, \mathbf{Y}^{*T})^T \sim \mathrm{t}_\nu((\mathbf{0}^T, \mathbf{0}^T)^T, \sigma^2\mathbf{I}_4)$, hence $\mathbf{Y} \sim \mathrm{t}_\nu(\mathbf{0}, \sigma^2\mathbf{I}_2)$, the pcf is

$$g(r) = 1 + \frac{1}{4\pi\kappa\sigma^2}\left(1 + \frac{r^2}{2\nu\sigma^2}\right)^{-(\nu+2)/2} \tag{3.5}$$

and the K-function is

$$K(r) = \pi r^2 + \frac{1 - \{1 + r^2/(2\sigma^2\nu)\}^{-\nu/2}}{\kappa}. \tag{3.6}$$

I can estimate κ and σ^2. The results presented in Section 3.4 are from an estimation using the pcf; however, the K-function could be employed just as well.

It is important to note that, in this setting, $\tau = 0$, $\boldsymbol{\alpha} = \mathbf{0}$ and $\sigma_1 = \sigma_2$, the exact distribution of $\mathbf{X}$ is multivariate Behrens-Fisher [Dickey, 1966] with pdf [Dickey, 1968]:

$$f_{\mathbf{X}}(\mathbf{x}) \;=\; C\, B\left(\frac{\nu+2}{2}, \frac{\nu+2}{2}\right) F_1\left(\frac{\nu+2}{2}; \nu+1, \nu+1; \nu+2; s_1, s_2\right),$$

where the constant $C = \Gamma(\nu+1)/[\pi\nu\{\Gamma(\nu/2)\}^2]$ for $\Gamma(\cdot)$ denoting the Gamma-function, $B\{(\nu+2)/2, (\nu+2)/2\} = \{\Gamma(\nu/2)\}^2/\Gamma(\nu+2)$ and F_1 is Appell's hypergeometric function. In particular [Erdélyi et al., 1953]:

$$\begin{aligned} &F_1\left((\nu+2)/2, \nu+1, \nu+1, \nu+2; s_1, s_2\right) = \\ &[B\{(\nu+2)/2, (\nu+2)/2\}]^{-1} \times \int_0^1 \{t(1-t)\}^{\nu/2} \{(1-ts_1)(1-ts_2)\}^{-\nu-1}\, dt, \end{aligned}$$

and s_1, s_2 are the two real roots of the equation $s^2 + (s-1)\mathbf{x}^T\mathbf{x}/(2\sigma^2\nu) = 0$. According to the transformation in (3.1),

$$f_d(r) \;=\; \frac{2\Gamma(\nu+1)\, r}{\nu\{\Gamma(\nu/2)\}^2} \int_0^1 \frac{\{t(1-t)\}^{\nu/2}}{\{(1-ts_1)(1-ts_2)\}^{\nu+1}}\, dt,$$

where $s_{1,2} = -r^2/(4\sigma^2\nu) \pm \sqrt{r^2/(2\sigma^2\nu) + \{r^2/(4\sigma^2\nu)\}^2}$. The notation $s_{1,2}$ denotes s_1 and s_2. The pcf, $g(r)$, can be derived from $f_d(r)$. However, the computation of $f_d(r)$ is computationally intensive and does not yield any advantage for the parameter estimation, since $g(r)$ remains analytically incomplete from this approach. This again confirms that using the approximation distribution of $\mathbf{X}_{\text{true}}$ is computationally advantageous.

3.3.5 The Case of Orthogonality

If $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are orthogonal, i.e., E $(\mathbf{Y}_1^T\mathbf{Y}_2) = 0$, and if they are jointly scale mixtures of bivariate normals, i.e., $\mathbf{Y}_i = V^{-1/2}\mathbf{Z}_i$, $i = 1, 2$, where the $\mathbf{Z}_i$'s are independently and identically $\mathrm{N}_2(\mathbf{0}, \boldsymbol{\Sigma})$ distributed, which are independent of $V \sim G$ and have a cdf with $G(0) = 0$, then $\mathbf{X} = V^{-1/2}\mathbf{Z}$ with $\mathbf{Z} \sim \mathrm{N}_2(\mathbf{0}, 2\boldsymbol{\Sigma})$ is independent of V. In particular, for $V \sim$ Gamma$(\nu/2, \nu/2)$, $\mathbf{Y}_i$ follows the bivariate Student t-distribution mentioned above with $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_2$, and the exact pcf

and K-function are given in (3.5) and (3.6).

3.4 Parameter Estimation by Minimum Contrast

Diggle [2003, Sec. 6] defined the *minimum contrast method* (MCM) using the K-function. MCM minimizes discrepancy between the theoretical K-function, $K(r;\boldsymbol{\theta}) \equiv K(r)$, of the assumed model and the empirical K-function, $\hat{K}(r)$, of the observed pattern. In particular, the discrepancy is defined as $D(\boldsymbol{\theta}) = \int_0^{r_0} w(r) \left[\{\hat{K}(r)\}^{c_{\text{stabil}}} - \{K(r;\boldsymbol{\theta})\}^{c_{\text{stabil}}}\right]^2 dr$, where the constants, r_0 and c_{stabil}, and the weighting function, $w(r)$, are to be chosen. Here, c_{stabil} acts as a variance-stabilizing transformation, and $\boldsymbol{\theta}$ is the vector comprising the parameters of the K-function, $K(r)$, or of the pcf, $g(r)$. The estimator, $\hat{\boldsymbol{\theta}}$, is the minimizer of $D(\boldsymbol{\theta})$.

In my setting, using the approximating pcf $g_d(r;\boldsymbol{\theta})$ and the empirical pcf $\hat{g}(r)$, I redefine the discrepancy,

$$D_{d,g}(\boldsymbol{\theta}) \;=\; \int_0^{r_0} w(r) \left[\{\hat{g}(r)\}^{c_{\text{stabil}}} - \{g_d(r;\boldsymbol{\theta})\}^{c_{\text{stabil}}}\right]^2 dr. \tag{3.7}$$

For the data simulation, I want to work with spatial point patterns (SPPs) having approximately 200 events on a unit square. Consequently, the dispersion parameters σ_1, σ_2 should not be larger than 0.10, otherwise the data generation cannot produce enough events, because the cluster dispersion is too large. Additionally, I want the number of "parent" and of "children" events to be between 10 and 20, so that the parameter estimation can be stable. Thus, I chose $\kappa = 20$, $\sigma_1 = \sigma = 0.04$, $\mu = 10$. In Table 3.1, the models of interest are given in the first column and the parameters are given in the second column. For the elliptical-normal CP, $\sigma_2 = 0.08$ or $c_\sigma = 2$ were chosen. For the skew-normal CP, $\alpha_1 = \alpha_2 = \alpha = 2$ were chosen. For elliptical-t CP with $\nu = 6$, $\sigma_2 = 0.08$ or $c_\sigma = 2$, and for the skew-t CP with the same df, $\alpha_1 = \alpha_2 = \alpha = 20$, $\tau = 1$ were selected. Thus, the skewness parameter of the skew-t CP is $\boldsymbol{\alpha}_{\mathbf{Y}}^T = (\alpha_1/\sqrt{1+\alpha_1^2+\alpha_2^2}, \alpha_2/\sqrt{1+\alpha_1^2+\alpha_2^2}) = (0.7067, 0.7067)$.

The `R`-package `spatstat` computes the empirical pcf with an isotropic-corrected estimator [Ripley, 1988] and a translation-corrected estimator [Ohser, 1983]. My experience shows that the

empirical pcf according to the former sometimes has NA (not-available) values, which can stop the computation of the estimation. Hence, for the parameter estimation in this section as well as for the data application in Section 3.5, I use the empirical pcf according to the translation-corrected estimator. The computation was done on compute nodes that have 8 CPU cores, 32GB of RAM and the CPU processors clocked at 2.4GHz or faster.

Since I generated the SPPs on a unit square, $r_0 = 0.25$ and $c_{\text{stabil}} = 0.25$ were chosen [Diggle, 2003, Chap. 6.1] for the parameter estimation. Additionally, $w(r) = 1$ was set due to clustered patterns [Diggle, 2003, Chap. 6.3]. I used the function `optim` available in `R` to minimize (3.7).

The logarithms of the starting values needed for the function `optim` are given in the second column of Table 3.1 since I estimated the logarithms of κ, σ^2, c_σ^2, α^2. For the parameter estimation, I just need to estimate α^2 since I set $\alpha_1 = \alpha_2 = \alpha$, i.e., $c_\alpha = 1$, for the skew-normal and skew-t CP. Additionally, I set $\nu = 6$ and $\tau = 1$ for a simple computation for the skew-t CP. In practice, however, the parameter estimation is done differently: one sequence of ν and one of τ are considered, the parameter estimation is done given a pair of (ν, τ). Among these possible combinations, a set of values is chosen as a set of estimates when it delivers the smallest discrepancy between the approximating and the empirical pcf. Table 3.1 provides the average computational time, $\bar{T}$, in seconds in the third column, and provides information to determine whether or not the choice of MCM and the function `optim` make sense in the last three columns. Let $\overline{\text{Dis}^2_{d,g}}(\hat{\boldsymbol{\theta}})$ denote the average of *bilateral discrepancy*,

$$\text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}) \;=\; \int \{\hat{g}(r) - g_d(\hat{\boldsymbol{\theta}}, r)\}^2 + \{g_d(\hat{\boldsymbol{\theta}}, r) - g_d(\boldsymbol{\theta}, r)\}^2 dr, \tag{3.8}$$

where the d, g-subscript shows the involvement of the approximating pcf, $g_d(r)$, and $\hat{\boldsymbol{\theta}}$ denotes the estimate. $\overline{\text{Dis}^2_{d,g}}(\hat{\boldsymbol{\theta}}_n)$ is the average of bilateral discrepancy $\text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}_n)$, where $\hat{\boldsymbol{\theta}}_n$ denotes the estimate resulting from the assumption of the (true) novel (skew-elliptical) CP. Similarly, $\overline{\text{Dis}^2_{d,g}}(\hat{\boldsymbol{\theta}}_t)$ is the average of bilateral discrepancy $\text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}_t)$, where $\hat{\boldsymbol{\theta}}_t$ denotes the estimate resulting from the assumption of the (wrong) traditional TP. For each of 3000 SPPs, I could compute $\text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}_n)$ and

Table 3.1: 3,000 SPPs were generated from each skew-elliptical CP. The first column provides the model specification. The second column gives information about the parameters of the model and, in the second row of each cell in this column, the logarithms of starting values for my estimation are provided. In the third column, the average computational time in seconds is represented by $\bar{T}$. $\overline{\text{Dis}^2_{d,g}}(\hat{\boldsymbol{\theta}}_n)$ denotes the average of $\text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}_n)$ according to (3.8), where $\hat{\boldsymbol{\theta}}_n$ denotes the MCM-estimate from the (true) novel (skew-elliptical) CP. $\overline{\text{Dis}^2_{d,g}}(\hat{\boldsymbol{\theta}}_t)$ is the average of $\text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}_t)$ according to (3.8), where $\hat{\boldsymbol{\theta}}_t$ denotes the MCM-estimate from the (wrong) traditional TP. In the sixth column, $\%$ provides the percentage of how often $\text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}_n)$ is smaller than $\text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}_t)$.*

	Parameters/ Starting values	$\bar{T}$	$\overline{\text{Dis}^2_{d,g}}(\hat{\boldsymbol{\theta}}_n)$	$\overline{\text{Dis}^2_{d,g}}(\hat{\boldsymbol{\theta}}_t)$	$\%$
Elliptical-normal	$(\kappa, \sigma_1, c_\sigma)^T = (20, .04, 2)^T$ $\log(\kappa_0, \sigma^2_{1,0}, c^2_{\sigma,0})^T = (0, -4, 3.5)^T$	.164	.029	.031	96.0
Skew-normal	$(\kappa, \sigma, \alpha_1 = \alpha_2 = \alpha)^T = (20, .04, 10)^T$ $\log(\kappa_0, \sigma^2_0, \alpha^2_0)^T = (0, -4, 5.5)^T$	.177	.176	.179	83.9
Elliptical-t, df $= 6$	$(\kappa, \sigma_1, c_\sigma)^T = (20, .04, 2)^T$ $\log(\kappa_0, \sigma^2_{1,0}, c^2_{\sigma,0})^T = (0, -4, 3.3)^T$	.115	.035	.038	85.3
Skew-t, df $= 6$	$(\kappa, \sigma, \alpha_1 = \alpha_2 = \alpha)^T = (20, .04, 20)^T$ $\log(\kappa_0, \sigma^2_0, \alpha^2_0)^T = (0, -5, 5)^T$	.082	.098	.101	84.7
Circular-t, df $= 6$	$(\kappa, \sigma)^T = (20, .04)^T$ $\log(\kappa_0, \sigma^2_0)^T = (0, -4)^T$	.035	.077	.085	88.1

$\text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}_t)$. The percentage in the last column shows how often $\text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}_n) < \text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}_t)$; i.e., if the correct model is assumed, the MCM using the approximating pcf can provide better estimates than assuming a TP. It shows that in the most cases $\text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}_n) < \text{Dis}^2_{d,g}(\hat{\boldsymbol{\theta}}_t)$. Additional information, $\overline{\text{Dis}^2_{d,g}}(\hat{\boldsymbol{\theta}}_n) < \overline{\text{Dis}^2_{d,g}}(\hat{\boldsymbol{\theta}}_t)$, also supports this statement.

The estimate of the mean number of "children" events, μ, does not come from the MCM directly, since μ does not appear in the pcf and hence is not involved in the minimization of $D_{d,g}(\boldsymbol{\theta})$ in (3.7). The estimator of μ is, in fact, $\hat{\mu} = n/\hat{\kappa}$, where n is the number of events of the observed pattern and $\hat{\kappa}$ denotes the estimate of κ and can be obtained via MCM.

The left column of Table 3.2 displays the choice of models and statistical information of the

*Reprinted with permission from "Skew-Elliptical Cluster Processes" by Ngoc Anh Dao and Marc G. Genton in I. Ghosh, N. Balakrishnan, and H. Ng, editors, *Advances in Statistics - Theory and Applications: Honoring the Contributions of Barry C. Arnold in Statistical Science*, pages 365-393. Springer, New York, 2021. Copyright 2021 by Copyright Clearance Center's RightsLink.

Table 3.2: The information about the models is given in Table 3.1. std = se $\times \sqrt{3000}$, where std is the estimate of standard deviation and se is the standard error. 2.5% gives the 2.5-percentile and 97.5% gives the 97.5-percentile of the distribution of the estimates, respectively. $\text{Bias}^2 = \{\text{E}(\hat{\theta}) - \theta\}^2$, where $\text{E}(\cdot)$ denotes the expectation and is approximated by average of the 3000 estimates. MSE = $\text{Bias}^2 + \text{Var}(\hat{\theta})$ where $\text{Var}(\hat{\theta})$ is the variance of $\hat{\theta}$ and is approximated by std^2. The four columns under Skew-elliptical Cluster Processes show the estimates and the statistical properties under the true models and the two columns under Thomas Process provide the ones under the (traditional) TP, the wrong model.*

	Skew-elliptical Cluster Processes				Thomas Process	
	$\hat{\kappa}$	$\hat{\sigma} = \hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\alpha} = \hat{\alpha}_1 = \hat{\alpha}_2$	$\hat{\kappa}$	$\hat{\sigma}$
Elliptical-Normal	21.854	.042	.098		26.485	.052
std	12.686	.013	.077		12.522	.011
2.5%	4.711	.020	.038		9.374	.035
97.5%	52.724	.068	.313		56.785	.077
Bias^2	3.436	29×10^{-7}	31×10^{-5}		42.060	14×10^{-5}
MSE	164.366	16×10^{-5}	.006		198.867	26×10^{-5}
Skew-Normal	23.175	.034		2.355	23.593	.028
std	7.486	.007		4.716	7.611	.004
2.5%	11.255	.023		.066	11.503	.022
97.5%	40.024	.049		10.789	40.912	.037
Bias^2	10.081	4×10^{-5}		20.471	12.907	15×10^{-5}
MSE	66.125	9×10^{-5}		42.711	70.831	16×10^{-5}
Elliptical-*t*, df $= 6$	20.470	.046	.094		26.625	.053
std	12.168	.017	.064		13.469	.012
2.5%	4.988	.018	.037		9.126	.035
97.5%	51.158	.084	.268		61.441	.080
Bias^2	.221	4×10^{-5}	19×10^{-5}		43.897	18×10^{-5}
MSE	148.270	32×10^{-5}	431×10^{-5}		225.319	32×10^{-5}
Skew-*t*	19.596	.040		19.245	22.244	.037
std	7.952	.010		8.624	8.399	.007
2.5%	7.174	.027		6.848	9.108	.028
97.5%	7.403	.065		41.041	40.904	.056
Bias^2	.163	1×10^{-7}		297.381	5.035	73×10^{-7}
MSE	63.403	5×10^{-5}		371.755	75.584	6×10^{-5}
Circular-*t*	22.030	.039			22.820	.039
std	8.446	.006			8.703	.006
2.5%	10.253	.029			10.538	.029
97.5%	41.927	.054			43.340	.053
Bias^2	4.119	3×10^{-7}			7.952	8×10^{-7}
MSE	75.446	35×10^{-6}			83.702	39×10^{-6}

*Reprinted with permission from "Skew-Elliptical Cluster Processes" by Ngoc Anh Dao and Marc G. Genton in I. Ghosh, N. Balakrishnan, and H. Ng, editors, *Advances in Statistics - Theory and Applications: Honoring the Contributions of Barry C. Arnold in Statistical Science*, pages 365-393. Springer, New York, 2021. Copyright 2021 by Copyright Clearance Center's RightsLink.

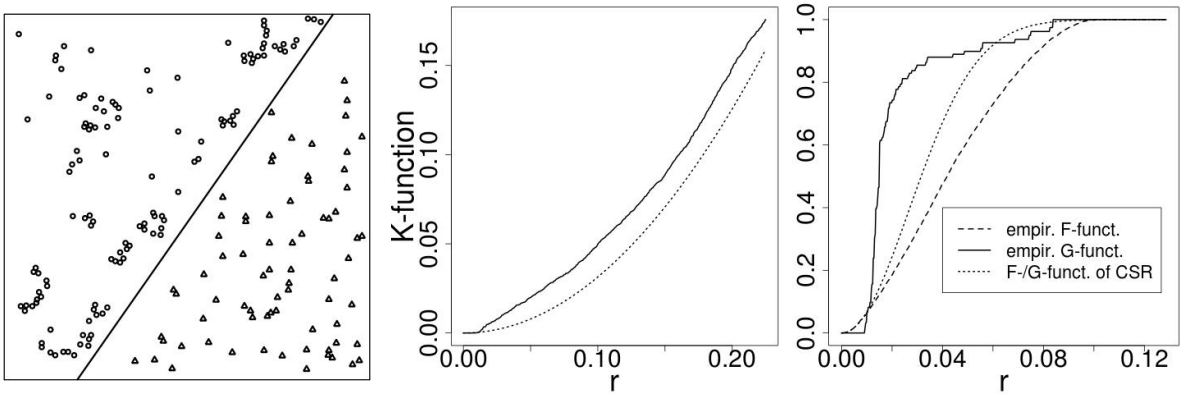


Figure 3.3: On the left, the locations of 195 Californian redwood seedlings and saplings in a square sampling region, 130×130 feet, are shown. They are displayed in two partitions: circles represent the clustered redwoods and triangles the inhibitory ones, respectively. In the middle, the empirical K-function (solid line) of the clustered redwoods and the theoretical one of a CSR of the same global intensity (dotted line) are shown. Here, the global intensity, λ, over the polygon containing the circles is approximately 221. On the right, the empirical F-function (dashed line) and G-function (solid line) are plotted along with the theoretical F- and G-functions of a CSR of 221 events (dotted line). Note that for a CSR, the theoretical F-function $\equiv$ G-function.*

estimates. If the hypothesized model is correctly assumed, the MCM-estimators of κ and $\sigma = \sigma_1$ outperform the ones under TP, the wrong model, with respect to MSE. The estimators of c_σ or σ_2 of elliptical-normal and -t CP, and of α of skew-normal and skew-t CP seem to be very reasonable since they are relatively unbiased and have tolerable variance. Overall, MCM provided reasonable estimates with respect to minimizing the discrepancy in (3.7).

3.5 The Clustered Redwoods Dataset

The `redwoodfull` dataset, available from the library `spatstat` and representing the locations of *195 Californian redwood seedlings and saplings* in a square sampling region, was first described and analyzed by Strauss [1975]. Additionally, according to Baddeley and Turner [2005a], it has never been subjected to a comprehensive analysis. In fact, only a small subset of it, known as the dataset `redwood` and consisting of only 62 trees, was analyzed in many works of spatial point

*Reprinted with permission from "Skew-Elliptical Cluster Processes" by Ngoc Anh Dao and Marc G. Genton in I. Ghosh, N. Balakrishnan, and H. Ng, editors, *Advances in Statistics - Theory and Applications: Honoring the Contributions of Barry C. Arnold in Statistical Science*, pages 365-393. Springer, New York, 2021. Copyright 2021 by Copyright Clearance Center's RightsLink.

processes, e.g., Diggle [2003]. The `redwoodfull` dataset appears to be interesting because it has many clusters that display non-circular shapes. I plotted `redwoodfull` in Figure 3.3 (left) in clustered and inhibitory partitions, represented by circles and triangles, respectively. In my opinion, 73 trees represented by triangles cannot be reasonably described by a clustered spatial point process since they follow an inhibitory pattern. I am interested only in analyzing the *clustered redwoods*, especially in finding out which skew-elliptical CPs can best model the process generating it. Figure 3.3 (middle) shows the empirical K-function (solid line) of the clustered redwoods and, for reference, the theoretical K-function (dotted line) of a CSR with the same intensity over the polygon. Here, r is the Euclidean distance from the event of interest. Figure 3.3 (right) displays the empirical F-function (dashed line), and G-function (solid line). For reference, the theoretical F- and G-functions of a CSR of the same intensity over the polygon are shown with a dotted line. Note for CSR, $F \equiv G$. Two facts indicating clustering are given in the following: (i) the empirical F-function lies below the theoretical F-function of a CSR and (ii) the K-function progresses above the theoretical K-function of a CSR. The empirical G-function also suggests clustering although not as clearly as the empirical F and K functions do. Sometimes it lies below, indicating inhibition, and sometimes above the reference line (the theoretical G-function of a CSR) over the domain of r approximately from 0.01 to 0.05, suggesting clustering. Overall, there are graphical hints that the redwoods of interest are clustered. I have, however, to investigate statistically whether this is truly the case. First, I test whether CSR can provide a good fit to the redwoods of interest. For that, the plug-in goodness-of-fit test using the G- and the F-function [Diggle, 2003, Chapt. 1.7] are employed and the resulting estimated p-values, $\hat{p}$, are all 0. Since the p-values are smaller than the nominal significance level of $\alpha_{\text{GOF}} = 0.05$, I reject that CSR provides a good fit and conclude that the redwoods of interest are clustered. Second, for the circular-, elliptical-, skew-normal CPs and the corresponding -t CPs with a certain df, I compute the estimates and the corresponding discrepancy between the empirical pcf and the theoretical one of the underlying model. For elliptical-t CPs, I choose 2, 6, 10, 20, 26 df for simplicity since their pcfs can be computed analytically with `Mathematica`. For skew- and circular-t CPs, I consider all even df up to 30, although I display

results only for 2, 10, 20, 30 df. The estimates such as $\hat{\kappa}$, $\hat{\sigma}^2$, $\hat{c}_\sigma^2$, $\hat{\alpha}^2$ and $\hat{c}_\alpha^2$ are obtained directly from the MCM (Table 3.3) except the one of μ, the mean of children number per cluster, which is absent in the pcf and hence irrelevant in this context.

The empirical pcf (solid line, right plot in the first row) in Figure 3.4 takes small values for small r, increases over the domain $0.001 < r < 0.0123$ and decreases for $r > 0.0123$. This observation is unlike how the pcf of a cluster process should progress. Illian et al. [2008, Sec. 4.3.1, 4.3.4] state that for a cluster process, the pcf takes large values for small r and decreases as r increases. This empirical pcf is indeed problematic at small r and I am aware that "the estimation of the pair correlation function is more delicate and complicated than that of K due to the serious issues of bandwidth choice and estimation for small r" [Illian et al., 2008, p. 227, Sec. 4.3.2]. I believe that using the complete curve of the empirical pcf would produce misleading estimates of κ, σ^2, c_σ^2, α^2 and c_α^2. Thus, two estimation possibilities should be investigated. The first data analysis uses the empirical pcf, $\hat{g}(r)$, completely. The second data analysis discards the first 28 pairs from 512 pairs of data (r_i, g_i), $i = 1, \ldots, 512$, where r_i denotes one of 512 grid points representing the domain of r and g_i the value of the empirical pcf at r_i. Estimates and the discrepancies, $\text{Dis}_g^1(\hat{\boldsymbol{\theta}})$ in (3.9), of the corresponding models from both analyses are listed in Table 3.3. I define the discrepancy between the empirical and the approximating pcfs at $\hat{\boldsymbol{\theta}}$ as follows

$$\text{Dis}_g^1(\hat{\boldsymbol{\theta}}) \;=\; \int_0^{r_0} \{\hat{g}(r) - g_d(\hat{\boldsymbol{\theta}}, r)\}^2 \, dr, \tag{3.9}$$

where $r_0 = 0.25$ is chosen for analysis of datasets on a unit square.

The first analysis using the empirical pcf completely assigns the smallest discrepancy to a skew-normal CP. The second analysis using the empirical pcf partially assigns the smallest discrepancy, however, to an elliptical-normal CP. Before the goodness-of-fit (GOF) of these two models are tested, I want to clarify a point that might appear to be an inconsistency in my calculation. Under the column "Using $\hat{g}(r)$ partially", the discrepancies, $\text{Dis}_g^1(\hat{\boldsymbol{\theta}})$, of the skew-$t$ CPs

Table 3.3: In the first column, the models applied to the clustered redwoods are shown. The resulting MCM-estimates are given in the second and third column. Here, $\hat{g}(r)$ is the empirical pcf. The discrepancy, $\text{Dis}_g^1(\hat{\boldsymbol{\theta}})$, between the empirical pcf and the theoretical one of certain underlying model is defined in (3.9). The smallest discrepancy in each column is displayed boldly.*

CP	**Using $\hat{g}(r)$ completely**		**Using $\hat{g}(r)$ partially**	
	Estimates	$\text{Dis}_g^1(\hat{\boldsymbol{\theta}})$	Estimates	$\text{Dis}_g^1(\hat{\boldsymbol{\theta}})$
circular-normal	$\hat{\kappa} = 60.910$, $\hat{\sigma} = .022$	.04824	$\hat{\kappa} = 60.082$, $\hat{\sigma} = .018$	.007080
elliptical-normal	$\hat{\kappa} = 60.880$, $\hat{\sigma}_1 = .022$, $\hat{\sigma}_2 = .022$	.04823	$\hat{\kappa} = 53.098$, $\hat{\sigma}_1 = .009$, $\hat{\sigma}_2 = .030$	**.004028**
skew-normal	$\hat{\kappa} = 60.873$, $\hat{\sigma} = .022$, $\hat{\alpha}_1 = -.185$, $\hat{\alpha}_2 = .185$	**.04822**	$\hat{\kappa} = 58.647$, $\hat{\sigma} = .023$ $\hat{\alpha}_1 = -20.254$, $\hat{\alpha}_2 = 20.254$	.006032
circular-t, df=2	$\hat{\kappa} = 47.182$, $\hat{\sigma} = .0242$	.05329	$\hat{\kappa} = 49.164$, $\hat{\sigma} = .0172$	.004481
circular-t, df=10	$\hat{\kappa} = 57.988$, $\hat{\sigma} = .0223$	.04899	$\hat{\kappa} = 57.754$, $\hat{\sigma} = .0177$	.005673
circular-t, df=20	$\hat{\kappa} = 59.369$, $\hat{\sigma} = .0222$	.04857	$\hat{\kappa} = 58.845$, $\hat{\sigma} = .0178$	.006274
circular-t, df=30	$\hat{\kappa} = 59.852$, $\hat{\sigma} = .0222$	.04846	$\hat{\kappa} = 59.283$, $\hat{\sigma} = .0179$	.006513
elliptical-t, df=2	$\hat{\kappa} = 47.148$, $\hat{\sigma}_1 = .0242$, $\hat{\sigma}_2 = .0242$	.05328	$\hat{\kappa} = 49.158$, $\hat{\sigma}_1 = .0172$, $\hat{\sigma}_2 = .0172$	.004481
elliptical-t, df=10	$\hat{\kappa} = 57.949$, $\hat{\sigma}_1 = .0223$, $\hat{\sigma}_2 = .0223$	.04899	$\hat{\kappa} = 52.712$, $\hat{\sigma}_1 = .0095$, $\hat{\sigma}_2 = .0282$	.004154
elliptical-t, df=20	$\hat{\kappa} = 59.042$, $\hat{\sigma}_1 = .0223$, $\hat{\sigma}_2 = .0223$	.04862	$\hat{\kappa} = 52.845$, $\hat{\sigma}_1 = .0092$, $\hat{\sigma}_2 = .0292$	.004112
elliptical-t, df=26	$\hat{\kappa} = 59.732$, $\hat{\sigma}_1 = .0222$, $\hat{\sigma}_2 = .0222$	.04847	$\hat{\kappa} = 52.959$, $\hat{\sigma}_1 = .0091$, $\hat{\sigma}_2 = .0296$	.004078
skew-t, df=2	$\hat{\kappa} = 47.174$, $\hat{\sigma} = .0243$, $\hat{\alpha}_1 = -5.178$, $\hat{\alpha}_2 = 62.388$ $(\hat{\alpha}_{\mathbf{Y},1} = -.083, \hat{\alpha}_{\mathbf{Y},2} = .996)$	.053	$\hat{\kappa} = 49.180$, $\hat{\sigma} = .0172$, $\hat{\alpha}_1 = -12.368$, $\hat{\alpha}_2 = 84.517$ $(\hat{\alpha}_{\mathbf{Y},1} = -.145, \hat{\alpha}_{\mathbf{Y},2} = .989)$	.004480
skew-t, df=10	$\hat{\kappa} = 57.963$, $\hat{\sigma} = .0224$, $\hat{\alpha}_1 = -3.356$, $\hat{\alpha}_2 = 49.077$ $(\hat{\alpha}_{\mathbf{Y},1} = -.068, \hat{\alpha}_{\mathbf{Y},2} = .997)$	.049	$\hat{\kappa} = 57.732$, $\hat{\sigma} = .0177$, $\hat{\alpha}_1 = -2.595$, $\hat{\alpha}_2 = 28.391$ $(\hat{\alpha}_{\mathbf{Y},1} = -.091, \hat{\alpha}_{\mathbf{Y},2} = .995)$	.005671
skew-t, df=20	$\hat{\kappa} = 59.427$, $\hat{\sigma} = .0222$, $\hat{\alpha}_1 = -8.095$, $\hat{\alpha}_2 = 51.185$ $(\hat{\alpha}_{\mathbf{Y},1} = -.156, \hat{\alpha}_{\mathbf{Y},2} = .988)$	.0486	$\hat{\kappa} = 58.890$, $\hat{\sigma} = .0178$, $\hat{\alpha}_1 = -4.079$, $\hat{\alpha}_2 = 45.750$ $(\hat{\alpha}_{\mathbf{Y},1} = -.089, \hat{\alpha}_{\mathbf{Y},2} = .996)$	.006273
skew-t, df=30	$\hat{\kappa} = 59.915$, $\hat{\sigma} = .0222$, $\hat{\alpha}_1 = -8.646$, $\hat{\alpha}_2 = 18.942$ $(\hat{\alpha}_{\mathbf{Y},1} = -.415, \hat{\alpha}_{\mathbf{Y},2} = .909)$	.048	$\hat{\kappa} = 59.270$, $\hat{\sigma} = .0179$, $\hat{\alpha}_1 = -4.048$, $\hat{\alpha}_2 = 45.369$ $(\hat{\alpha}_{\mathbf{Y},1} = -.089, \hat{\alpha}_{\mathbf{Y},2} = .996)$	.006519

*Reprinted with permission from "Skew-Elliptical Cluster Processes" by Ngoc Anh Dao and Marc G. Genton in I. Ghosh, N. Balakrishnan, and H. Ng, editors, *Advances in Statistics - Theory and Applications: Honoring the Contributions of Barry C. Arnold in Statistical Science*, pages 365-393. Springer, New York, 2021. Copyright 2021 by Copyright Clearance Center's RightsLink.

do not converge to the one (0.006032) of the skew-normal CP when the df increase. The reason for it is the absolute values of the skewness parameters of the skew-normal CP are really large $|\alpha_1| = |\alpha_2| = 20.254$, while the absolute values of the skewness parameters of the skew-t CP are much smaller, $|\alpha_{\mathbf{Y},i}| < 1$, $i = 1, 2$. On the contrary, under the column "Using $\hat{g}(r)$ completely", the discrepancies, $\text{Dis}_g^1(\hat{\boldsymbol{\theta}})$, of the skew-$t$ CPs do converge to the one (0.04822) of the skew-normal CP when the df increase. The reason for it is the absolute values of the skewness parameters of the skew-normal CP are smaller than 1, in particular $|\alpha_1| = |\alpha_2| = 0.185$, and the absolute values of the skewness parameters of the skew-t CP are also smaller than 1, $|\alpha_{\mathbf{Y},i}| < 1$, $i = 1, 2$. These phenomena can serve as demonstrations of a statement in Remark 4.

Now, the adjusted goodness-of-fit (AGOF) test [Dao and Genton, 2014] is applied since the plug-in GOF test is not appropriate because only one dataset is available. The AGOF test is also termed as the Dao-Genton test in Baddeley et al. [2015] and is made available in the `R`-library `spatstat`. Diggle [2003, Sec. 6.2.] recommended not to use the GOF test based on the K-function if the K-function was used for parameter estimation. Since I used the empirical pcf (originating from the K-function) for the parameter estimation, I could rely on the GOF conclusion from the AGOF-G or -F test. I expect, however, that the AGOF-G tests would not support the fit of any CP due to the limited support of the clustering of the empirical G-function. Therefore, I decided to rely mainly on the conclusion from the AGOF-F test. For the testing, the nominal significance level is $\alpha_{\text{GOF}} = 0.05$ and $\hat{\alpha}^*_{\text{AGOF}}$ denotes the estimated adjusted level [Dao and Genton, 2014]. For completeness, I run AGOF-G tests which rejected all the models to be a good fit. This is expected due to the limited support of the clustering of the empirical G-function explained previously. The AGOF-F test, the only test to be relied on, provided (i) $\hat{p} = 0.025$ which is greater than $\hat{\alpha}^*_{\text{AGOF}} = 0.005$ for the skew-normal CP model and (ii) $\hat{p} = 0.035$ which is greater than $\hat{\alpha}^*_{\text{AGOF}} = 0.004$ for the elliptical-normal CP model. For the latter model, I used the empirical pcf partially as described previously for the parameter estimation, but used the empirical pcf completely for the computation of the $\hat{p}$-value. The AGOF-F test provided $\hat{p} > \hat{\alpha}^*_{\text{AGOF}}$ for both models. Hence, I conclude that these models provide a good fit, statistically speaking.

Now, I examine these models graphically. The right plot (Figure 3.4) shows the empirical pcf (solid), the approximating pcf of the circular-normal CP (TP) (thin), the theoretical skew-normal CP (dotted, red) and the approximating elliptical-normal CP (dashed, blue). The two approximating pcfs of the circular- and the skew-normal are very similar due to the negligible estimate of the skewness-parameter, $\hat{\alpha} = 0.185$. The approximating pcf of the skew-normal CP does not represent well the empirical pcf, neither at short nor at middle distance, i.e. $r < 0.066$. The approximating pcf of the elliptical-normal CP, however, does represent the empirical pcf well from the middle distance, $r > 0.0123$. The left plot (Figure 3.4) shows that the simulated events (triangles) of the skew-normal CP do not mimic the clustered redwoods (circles) well because while the cluster shape of the clustered redwoods is oblong, that of the simulated data is fairly circular. On the contrary, the middle plot shows that the simulated events (triangles) of the elliptical-normal CP have oblong cluster shape which is similar to the cluster shape of the clustered redwoods. One can see this more clearly if one turns the simulated data by an angle of approximately $40°$.

Overall, I think that the elliptical-normal CP represents the data better than does the skew-normal CP. This data application also confirms that the introduction of skew-elliptical CPs is necessary, otherwise the ellipticity of the cluster shape could not be modeled.

3.6 Discussion

There are a few robustness problems in estimation. First, the MCM uses the approximating instead of the theoretical pcf. The approximating pcf results from isotropy assumption of the CP to achieve the analytical completeness, easy to be incorporated in MCM. The isotropy leads to a significant loss of information, and therefore the results of estimates need to be carefully verified.

Second, there is sensitivity towards starting values of the empirical pcf under isotropy assumption. It is usually poorly estimated at a short distance, i.e., r close to 0. I encountered this problem in my data application: the empirical pcf does not decrease throughout although it should be strictly decreasing since the assumed model is clustered [Illian et al., 2008, Sec. 4.3.4]. It even increases over a short domain close to 0. Using the complete curve of the empirical pcf might produce misleading estimates, but at the same time, ignoring the poorly estimated part of the empirical pcf

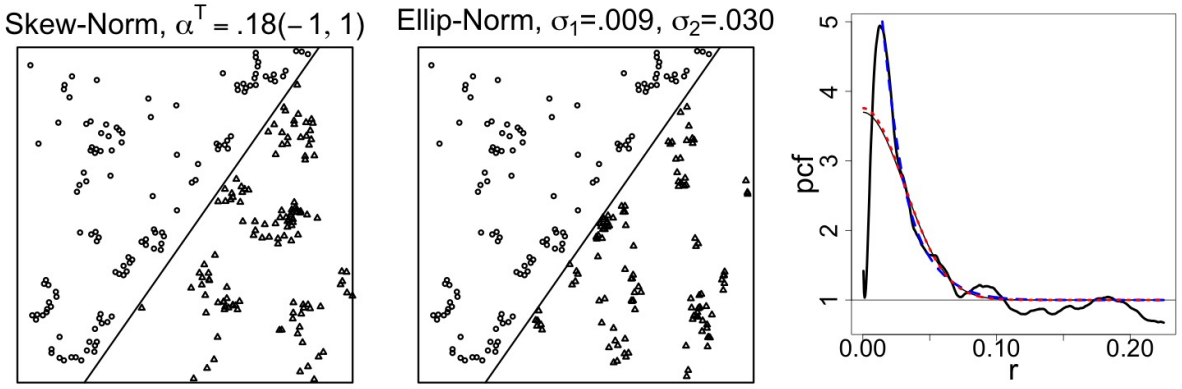


Figure 3.4: The 122 clustered redwoods (circles) are displayed in the upper left polygon of the left and middle plots. There, in the lower right polygon, the left plot shows the simulated events (triangles) of the skew-normal CP with parameters $\kappa = 60.873, \sigma = 0.022, \alpha_1 = -0.185, \alpha_2 = 0.185, \mu = 3.632$ and similarly, the middle plot shows the simulated events (triangles) of the elliptical-normal CP with parameters $\kappa = 53.098$, $\sigma_1 = 0.009$, $\sigma_2 = 0.030$, $\mu = 4.163$. The right plot shows the empirical pcf (solid), the theoretical pcf of the circular-normal CP (TP) (thin), the approximating pcf of skew-normal CP (dotted, red) and the approximating pcf of elliptical-normal CP (dashed, blue). The theoretical pcf of TP and the approximating pcf of skew-normal are very similar due to the negligible estimate, $\hat{\alpha} = 0.185$. Here, these two pcfs overlay each other. For the simulations, the random seed, 999, as in Figure 3.1 was used.*

might cause overfitting. It may be possible to come up with a cut-off point from which data of the pcf can be used.

Third, the more parameters the model has, the more sensitive the estimation can become with respect to the starting values. In general, it is usually difficult to estimate high-dimensional parameters. One can try to improve the robustness by estimating certain parameters at a time. For example, assuming a few parameters, say $\boldsymbol{\theta}_1$, are given, one estimates the remaining parameters, say $\boldsymbol{\theta}_2$, where $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)$. Then, I plug in the estimates $\boldsymbol{\theta}_2 = \hat{\boldsymbol{\theta}}_2$ in the pcf and estimate $\boldsymbol{\theta}_1$. The estimation continues until the discrepancy (3.9) goes below a pre-set limit. According to my limited estimation studies, parameters such as c_α or α could be treated as $\boldsymbol{\theta}_1$, and κ, c_σ or σ could be treated together as $\boldsymbol{\theta}_2$.

*Reprinted with permission from "Skew-Elliptical Cluster Processes" by Ngoc Anh Dao and Marc G. Genton in I. Ghosh, N. Balakrishnan, and H. Ng, editors, *Advances in Statistics - Theory and Applications: Honoring the Contributions of Barry C. Arnold in Statistical Science*, pages 365-393. Springer, New York, 2021. Copyright 2021 by Copyright Clearance Center's RightsLink.

A possible extension of generalizing the TP is to consider enlarging the choice of the distribution that is imposable on $\mathbf{Y}$, which is the location of a "children" event in a cluster. Besides the SUN and EST classes, there may be other distributions of the unified skew-elliptical families. One of the requirements for the distribution of $\mathbf{Y}$ is that it has the additive property because the distribution of $\mathbf{X}$ has to be established where $\mathbf{X} \stackrel{d}{=} \mathbf{Y}_1 - \mathbf{Y}_2$ and $\mathbf{Y}_i$, $i = 1, 2$, representing two independent positions of the "children" events in a cluster.

In this chapter, I generalized the TP to some extent. However, I can shift the focus to the Matérn process, the role of which is very similar to that of the TP in the field of spatial point processes. Both are special cases of the Neymann–Scott cluster point process. A Matérn process is constructed similar to a TP except that the positions of the "children" events are distributed independently and uniformly inside a disc with the "parent" event as the center. Similar to this work, it is possible to establish some variations of the Matérn process with respect to the circular, elliptical and skew properties of the distribution of "children" events.

Closing up this chapter and my contribution to statistical inference for spatial point processes in this thesis, I will spend my attention on deriving a method to identify influential events of a SPP in the next chapter.

4. VISUALIZING INFLUENTIAL EVENTS IN SPATIAL POINT PATTERNS

4.1 Chapter Overview

In the field of spatial point patterns (SPPs), it is now possible to carry out accurate statistical inference, in particular fitting complex statistical models [Møller and Waagepetersen, 2003, Diggle, 2003, Illian et al., 2008, Dao and Genton, 2021]. Goodness-of-fit (GOF) tests for an homogeneous spatial Poisson process, or otherwise known as complete spatial randomness (CSR), are envelope tests or quadrat count tests [Diggle, 2003]. For testing GOF of inhomogeneous spatial Poisson processes, Guan [2008] developed a GOF test with a statistic based on a discrepancy constructed from residuals obtained from the fitted model. For many other intricate parametric models, the envelope test validates GOF when observing several SPPs [Diggle, 2003] and the Monte Carlo-adjusted GOF test [Dao and Genton, 2014] can be employed when observing only a single SPP. There are techniques for residual analysis [Baddeley et al., 2005], detecting outliers, criticizing model-based outliers [Illian et al., 2008], identifying leverage and influential points [Baddeley et al., 2013]. Especially, in the latter work, the authors established measures of leverage and influence for the dependence of a point process model on covariates by direct analogy with standard techniques for generalized linear models. These concepts are suitable for Poisson point processes fitted by maximum likelihood [Rathbun and Cressie, 1994, Kutoyants, 1998], or by robust M-estimators [Assunção, 1994b, Assunção and Guttorp, 1999] and to Gibbs point processes fitted by maximum pseudo-likelihood [Baddeley and Turner, 2000]. A recent work on studying leverage and influence diagnostics for Gibbs spatial processes [Baddeley et al., 2019] introduced graphical tools and a new diagnostic analogous to the effect measure difference in fit, also known as DFFIT, in the regression [Belsley et al., 1980].

The technique developed by Baddeley et al. [2013] requires, however, the likelihood or composite likelihood to be formally equivalent to a Poisson likelihood and that term in the model must be available in closed form. This excludes a wide range of models including, for example, Cox

processes, Neyman-Scott cluster processes and other models which are effectively hierarchical or mixed Poisson models.

In this chapter, I introduce a concept of detecting influential events of a SPP, which does not require the maximum likelihood or pseudo-likelihood in closed form. In particular, I let a second-order summary characteristic such as F-, G-function [Diggle, 2003, Sec. 4.3], Ripley's K-function [Ripley, 1976], and pair correlation function [Illian et al., 2008] to describe the SPP and define a measure of departure based on the second-order summary characteristic of one's choice. This approach allows many more spatial point processes to be studied.

In the setting of generalized linear models, the standard method of identifying influential observations is to quantify the radical change of the statistical inference when carrying out statistical analysis without that observation. I would call this approach a "deleting" method. In my opinion, this is only justified for the cases underlying the assumption that the observations are independent and identically distributed (iid). In the field of spatial point processes, Baddeley et al. [2013, 2019] still kept this "deleting" method to study the leverage and influence of the data. However, in the case of dependent data such as in spatial statistics or spatial point processes, I think that deleting event(s) does not seem as a sensible practice due to the dependence structure of the events. As noted by Cook and Weisberg [1994], deleting cases is only a way of introducing small changes in the data and there are others. As in Genton and Ruiz-Gazen [2010], I introduce some noise to the data by perturbing additively one event at a time. I call this a *perturbing* method. There are two advantages using it. First, deleting events might be questionable, if, for example, events represent plants, galaxies, or sites in cells, to name a few. Second, unlike deleting an observation/event, the perturbing method provides a whole course of change of estimators or conclusions on GOF due to a set of possible amounts of perturbation.

I adopt the approach by Genton and Ruiz-Gazen [2010] to develop a method to detect influential events. Via perturbing events and defining discrepancy between the perturbed and observed functionals, I define influential event and use graphical tools such as hair-plots and disc-plots to visualize them.

The remainder of this chapter is organized as follows. Section 4.2 describes the methodology to define and visualize the influential and local influential event. Simulation studies demonstrate in Section 4.3 that my method can detect and visualize the influential events well and the method seems not to suffer from edge effect. Section 4.4 provides two data applications. One application is the southern half of the Queensland copper cores dataset. The other application is the Swedish pines dataset. The detected influential events are compared with the ones which were detected by other works. Finally, Section 4.5 discusses an alternative discrepancy to define influential events. This discrepancy can take the relevant domain or range of the K-function into account.

4.2 Perturbation Method

4.2.1 Setting

Hereafter, I adopt the notation given in Baddeley et al. [2005, Sec. 5]. A SPP is a dataset, $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, where the $\mathbf{x}_i$'s are unordered locations observed in a bounded region, $\mathbf{W}$, of $\mathbb{R}^2$. I let $f_{\boldsymbol{\theta}}$ denote the parametric model (a parametric spatial point process) fitted to $\mathbf{X}$, where $\boldsymbol{\theta}$ is an arbitrary finite-dimensional vector of parameters. I assume that $f_{\boldsymbol{\theta}}(\mathbf{X})$ is a probability density function with respect to the unit rate Poisson process on the window $\mathbf{W}$, such that $f_{\boldsymbol{\theta}}$ satisfies the positivity condition: if $f_{\boldsymbol{\theta}}(\mathbf{X}) > 0$ and $\mathbf{Y} \subset \mathbf{X}$, then $f_{\boldsymbol{\theta}}(\mathbf{Y}) > 0$ for any finite point patterns, $\mathbf{X}, \mathbf{Y} \subset \mathbf{W}$. Under this setup, I am interested in identifying influential events of $\mathbf{X}$ under the following parametric assumption:

$$H_0 : \quad \mathbb{X} \sim f_{\boldsymbol{\theta}}, \tag{4.1}$$

or under the following nonparametric assumption:

$$H_0 : \quad \mathbb{X} \sim f_{\text{non}}, \tag{4.2}$$

where $\mathbb{X}$ is the spatial point process from which the observed SPP, $\mathbf{X}$, is generated and f_{non} is a statistical model which is not a maximum likelihood, pseudo-likelihood, pseudo-profile-likehood

in closed form

4.2.2 Perturbation

In the following, I assume the SPP is on a unit square. Let $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ denote the SPP of interest and $\mathbf{x}_i = (x_{i1}, x_{i2})^\top$ represent the xy-coordinates of the event $\mathbf{x}_i$. I perturb the location of one event at a time to find the most radical change in the inference. For an event $\mathbf{x}_i$, its perturbation is regulated by an angle γ, $0 \leq \gamma < 2\pi$, and by a radius ζ, $0 \leq \zeta \leq \sqrt{2}$.

The function $\mathbf{T}(\mathbf{x}_i; \gamma, \zeta) = [x_{i,1} + \zeta\cos(\gamma), x_{i,2} + \zeta\sin(\gamma)]^\top$ regulates the perturbation of the event $\mathbf{x}_i$. Let $\mathbf{X}(i, \gamma, \zeta) = \mathbf{X}\backslash\{\mathbf{x}_i\} \cup \mathbf{T}(\mathbf{x}_i; \zeta, \gamma)$.

4.2.3 Hair-functions, Hair-plots, Disc-plots

Like in Genton and Ruiz-Gazen [2010], the concepts of *hair-plot* and *disc-plot* will be applied here to the setting of SPPs.

4.2.3.1 Parametric Models

For the parametric model in (4.1), I introduce two discrepancies which can individually or mutually be considered to identify influential events. The first discrepancy is

$$\hat{D}_i(\zeta_j, \gamma_m) \quad = \quad \int \left[\hat{K}_{i,j,m}(h) - K_{\hat{\boldsymbol{\theta}}}(h)\right]^2 \mathrm{d}h, \tag{4.3}$$

where $\hat{K}_{i,j,m}(h)$, $i = 1, \ldots, n, j = 1, \ldots, S, m = 1, \ldots, M$, is the empirical K-function of the perturbed spatial pattern $\mathbf{X}(i, \zeta_j, \gamma_m)$, $\zeta_1 = 0, \ldots, \zeta_j, \ldots, \zeta_S = \sqrt{2}$, $\gamma_1 = 0, \ldots, \gamma_m, \ldots, \gamma_M = 2\pi$ and $K_{\hat{\boldsymbol{\theta}}}(h) = K_{\boldsymbol{\theta}}(h)|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ where $K_{\boldsymbol{\theta}}(h)$ is the theoretical K-function derived from the hypothesized model in (4.1) and $\hat{\boldsymbol{\theta}}$ is the plug-in estimate computed from the observed pattern $\mathbf{X}$ under (4.1). If the K-function is not analytically complete, an approximating K-function, $\tilde{K}$, can be used instead of the true K-function.

The discrepancy in (4.3) measures the departure of the empirical K-function of the perturbed pattern and the theoretical K-function under H_0 at the plug-in estimate $\hat{\boldsymbol{\theta}}$ from the observed (unperturbed) pattern. Intentionally, I do not consider the quantity $\int \left[K_{\hat{\boldsymbol{\theta}}_{i,j,m}}(h) - \hat{K}(h)\right]^2 \mathrm{d}h$ where

$K_{\hat{\boldsymbol{\theta}}_{i,j,m}}(h)$ is the K-function of the hypothesized model at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{i,j,m}$ which is computed from the perturbed spatial pattern $\mathbf{X}(i, \zeta_j, \gamma_m)$. In my opinion, the perturbation can destroy the parametric structure of the observed pattern. Hence using $K_{\hat{\boldsymbol{\theta}}_{i,j,m}}$ is not reasonable.

For the introduced discrepancy (4.3), the definition of the hair-functions are established as follows. Any or both hair-functions can be applied to identify influential events:

$$\text{Hair}_i^{\text{mean}}(\zeta) = \frac{1}{M}\sum_{m=1}^{M}\frac{\hat{D}_i(\zeta,\gamma_m)}{\hat{D}(0,0)} - 1, \text{ or} \tag{4.4}$$

$$\text{Hair}_i^{\text{max}}(\zeta) = \max_{m=1,\ldots,M}\frac{\hat{D}_i(\zeta,\gamma_m)}{\hat{D}(0,0)} - 1, \text{ where} \tag{4.5}$$

$$\hat{D}(0,0) = \int \left[\hat{K}(h) - K_{\hat{\boldsymbol{\theta}}}(h)\right]^2 \mathrm{d}h. \tag{4.6}$$

where $\hat{K}(h)$ is the empirical K-function of the originally observed spatial pattern. An event is declared as *influential* if its corresponding A_i is an outlier among other squared areas under the curve (SAUC):

$$A_i^{\text{mean}}(\zeta) = \int [\text{Hair}_i^{\text{mean}}(\zeta)]^2 \ \mathrm{d}\zeta, \text{ or} \tag{4.7}$$

$$A_i^{\text{max}}(\zeta) = \int [\text{Hair}_i^{\text{max}}(\zeta)]^2 \ \mathrm{d}\zeta. \tag{4.8}$$

Note that SAUC quantifies the *influence* of an event in this context.

The *second* discrepancy can be viewed as a *parametric* discrepancy:

$$\hat{D}_{\text{par},i}(\zeta_j,\gamma_m) = \int \left[K_{\hat{\boldsymbol{\theta}}_{i,j,m}}(h) - K_{\hat{\boldsymbol{\theta}}}(h)\right]^2 \mathrm{d}h. \tag{4.9}$$

It measures how sensitive the hypothesized model can become towards perturbation.

The hair-functions can be defined for the parametric discrepancy as follows,

$$\text{Hair}^{\text{mean}}_{\text{par},i}(\zeta) = \frac{1}{M}\sum_{m=1}^{M}\hat{D}_{\text{par},i}(\zeta,\gamma_m), \text{ or} \tag{4.10}$$

$$\text{Hair}^{\text{max}}_{\text{par},i}(\zeta) = \max_{m=1,\ldots,M}\hat{D}_{\text{par},i}(\zeta,\gamma_m). \tag{4.11}$$

For identifying influential events, the corresponding SAUC can be defined as in (4.7) and (4.8), in particular

$$A^{\text{mean}}_{\text{par},i}(\zeta) = \int \left[\text{Hair}^{\text{mean}}_{\text{par},i}(\zeta)\right]^2 \,\mathrm{d}\zeta, \text{ or} \tag{4.12}$$

$$A^{\text{max}}_{\text{par},i}(\zeta) = \int \left[\text{Hair}^{\text{max}}_{\text{par},i}(\zeta)\right]^2 \,\mathrm{d}\zeta. \tag{4.13}$$

For parametric models, I list below the theoretical K-functions of the CSR and the Thomas process as examples for my simulations studies. The theoretical K-function of CSR is

$$K_{\boldsymbol{\theta}}(h) = \pi h^2,$$

where $\boldsymbol{\theta} = 0$ in this case.

For a clustered pattern, I consider a Thomas process as a representative of this group. Here, a CSR with intensity κ is generated to obtain the so-called "parent" points. Each parent point is replaced by a random cluster of "children" points that are Poisson distributed with intensity κ. The positions of the children points are distributed about the parent location according to a bivariate Gaussian distribution with covariance $\sigma^2 I_2$, where I_2 is the 2×2 identity matrix [Møller and Waagepetersen, 2003]. Then

$$K_{\boldsymbol{\theta}}(h) = \pi h^2 + \kappa^{-1}[1 - \exp -h^2/(4\sigma^2)]$$

is the theoretical K-function of a Thomas process, where $\boldsymbol{\theta} = (\kappa, \sigma)^\top$

4.2.3.2 *No Model Assumption*

For the nonparametric model in (4.2), I introduce the *third* discrepancy which can be viewed as a *nonparametric* discrepancy:

$$\hat{D}_{\text{non},i}(\zeta_j, \gamma_m) \;=\; \int \left[\hat{K}_{i,j,m}(h) - \hat{K}(h)\right]^2 \mathrm{d}h, \tag{4.14}$$

where $\hat{K}_{i,j,m}(h)$ is the empirical K-function of $\mathbf{X}(i, \zeta_j, \gamma_m)$. This discrepancy is of advantage if one does not want to make a parametric assumption but still want to identify influential events.

The hair functions can be defined for the parametric discrepancy as follows,

$$\text{Hair}^{\text{mean}}_{\text{non},i}(\zeta) \;=\; \frac{1}{M}\sum_{m=1}^{M} \hat{D}_{\text{non},i}(\zeta, \gamma_m), \text{ or} \tag{4.15}$$

$$\text{Hair}^{\text{max}}_{\text{non},i}(\zeta) \;=\; \max_{m=1,\ldots,M} \hat{D}_{\text{non},i}(\zeta, \gamma_m). \tag{4.16}$$

For identifying influential events, I define the corresponding SAUC as follows

$$A^{\text{mean}}_{\text{non},i}(\zeta) \;=\; \int \left[\text{Hair}^{\text{mean}}_{\text{non},i}(\zeta)\right]^2 \mathrm{d}\zeta, \text{ or} \tag{4.17}$$

$$A^{\text{max}}_{\text{non},i}(\zeta) \;=\; \int \left[\text{Hair}^{\text{max}}_{\text{non},i}(\zeta)\right]^2 \mathrm{d}\zeta. \tag{4.18}$$

The hair-functions, $\text{Hair}^{\text{mean}}_{i}(\zeta)$, $\text{Hair}^{\text{max}}_{i}(\zeta)$, $\text{Hair}^{\text{mean}}_{\text{par},i}(\zeta)$, $\text{Hair}^{\text{max}}_{\text{par},i}(\zeta)$, $\text{Hair}^{\text{mean}}_{\text{non},i}(\zeta)$ and $\text{Hair}^{\text{max}}_{\text{non},i}(\zeta)$, can be visualized in hair-plots. For each event i, $i = 1, \ldots, n$, the SAUCs (influence) such as A^{mean}_{i}, A^{max}_{i}, $A^{\text{mean}}_{\text{par},i}$, $A^{\text{max}}_{\text{par},i}$, $A^{\text{mean}}_{\text{non},i}$ and $A^{\text{max}}_{\text{non},i}$ can be computed and displayed in disc-plots. The radii of the discs are proportional to these quantities in a disc-plot. Based on the boxplot of these n SAUCs, I declare the events to be *influential* if their SAUCs are outliers according to the boxplot.

4.2.4 Local Influence

Like in Genton and Ruiz-Gazen [2010], the concept of *local influence* can be applied to SPPs. Let τ^{mean}_{i} denote the local influence from the average approach, correspondingly τ^{max}_{i} the local

influence from the maximum approach. The local influence quantifies the rate of change of $\text{Hair}_i^{\text{mean}}(\zeta)$ in (4.4), correspondingly $\text{Hair}_i^{\text{max}}(\zeta)$ in (4.5), at $\zeta = 0$. In particular,

$$\tau_i^{\text{mean}} = \frac{\partial}{\partial\zeta}\text{Hair}_i^{\text{mean}}(\zeta)|_{\zeta=0}, \tag{4.19}$$

$$\tau_i^{\text{max}} = \frac{\partial}{\partial\zeta}\text{Hair}_i^{\text{max}}(\zeta)|_{\zeta=0}. \tag{4.20}$$

Here τ_i^{mean}, or respectively τ_i^{max}, quantifies the rate of change from the hair-function, $\text{Hair}_i^{\text{mean}}(\zeta)$ or respectively $\text{Hair}_i^{\text{max}}(\zeta)$, due to a very small perturbation around the event i.

Then $\tau_{\text{par},i}^{\text{mean}}$, $\tau_{\text{par},i}^{\text{max}}$ and $\tau_{\text{non},i}^{\text{mean}}$, $\tau_{\text{non},i}^{\text{max}}$ can be defined as similarly as above.

For each event i, $i = 1, \ldots, n$, the *local influence* such as τ_i^{mean}, τ_i^{max}, $\tau_{\text{par},i}^{\text{mean}}$, $\tau_{\text{par},i}^{\text{max}}$, $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\text{max}}$ can be computed and displayed in disc-plots. The radii of the discs are proportional to the local influence in a disc-plot. Based on the boxplot of these local influences, I declare the events to be *locally influential* if their local influence are outliers according to the boxplot.

These hair-plots and disc-plots are shown in Section 4.3.

4.3 Simulations

Figure 4.1 displays three simulated patterns on the unit square from a CSR, Thomas process and a process mixing the previous two. The hair-functions, the SAUC, the local influence, and the resulting influential events will be computed based on the discrepancy in (4.3), the parametric discrepancy in (4.9) and the nonparametric discrepancy in (4.14).

The domain of γ was chosen as $(0, 1/120, 2/120, \ldots, 119/120)^\top \times 2\pi$ with $M = 120$. At each γ_m, $m = 1, \ldots, M = 120$, a domain of ζ-values is $(\zeta_1 = 0, \ldots, \zeta_S = \sqrt{2})$ with $S = 50$ applied in Sec. 4.2.2.

The simulation results for the three simulated patterns can be found in Tables 4.2, 4.3, 4.4. The overlapped influential events resulting from the mean and the maximum approaches based on a specific discrepancy in either (4.3), (4.9), or (4.14) are displayed in bold numbers. The size of the overlap will change if I used more perturbations arisen from combinations from finer grids for γ and ζ. Data from more perturbations definitely help to quantify the SAUC more precisely.

Colors	Percentile Bandwidths
purple	$C_{25} := (\text{Hair}_{37.5}; \text{Hair}_{62.5})$
magenta	$C_{50} := (\text{Hair}_{25}; \text{Hair}_{75}) \backslash C_{25}$
pink	$C_{75} := (\text{Hair}_{12.5}; \text{Hair}_{87.5}) \backslash C_{50}$
orange	$C_{95} := (\text{Hair}_{2.5}; \text{Hair}_{97.5}) \backslash C_{75}$
brown	$C_{100} := (\text{Hair}_{1}; \text{Hair}_{100}) \backslash C_{95}$
black	median hair-function
red, dashed	influential hair-function

Table 4.1: Color labels of the hair-functions according their SAUCs to form the bandwidths

The overlap size would not necessarily increase because the mean and the maximum approaches describe different characteristics of the datasets.

In the following, the hair-plots display the hair-functions in different colors. The colors label the different ranges of percentiles of the hair-functions. The hair-plots have a vertical dashed line at $\sqrt{2}/2$ as a reference. The introduction of this reference line is to assist the visualization because the influence is usually higher in the interval $[0, \sqrt{2}/2)$ than in the interval $[\sqrt{2}/2, \sqrt{2})$.

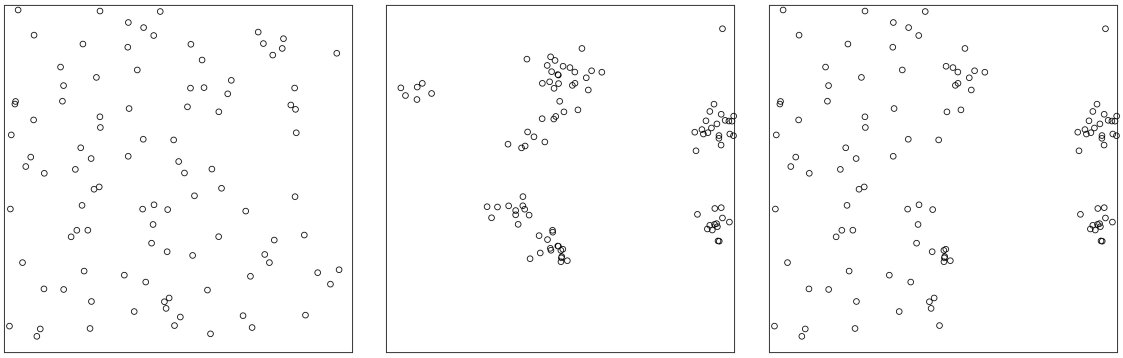
From Table 4.1, the percentile bandwidth of C_{25} includes the hair-functions which rank $37.5th$ to the $62.5th$ hair-function. The hair-function of C_{25} are shown in purple. Then C_{50} includes the hair-functions which rank $25th$ to $75th$ excluding C_{25}. Similarly, C_{50}, C_{75}, C_{95}, C_{100} are defined.

To assist the readability I gather a list of definitions of the hair-functions, their influences, their local influences and their equation numbers in Table 4.7.

4.3.1 Complete Spatial Randomness

A CSR SPP of 100 events, in Figure 4.1a, was generated. The p-value from the GOF-K test is 0.60 [Diggle, 2003, Dao and Genton, 2014]. The influential events are reported in Table 4.2.

I apply the GOF-K test from Diggle [2003] to verify whether the simulated SPP represents the attributes of the CSR process specified above well. The estimated p-value from the GOF-K test is $\hat{p} = 0.60$. Since $\hat{p} > \alpha$, where $\alpha = 0.05$, I do not reject the hypothesis that the SPP represents the CSR process well.



(a) CSR (b) Thomas process (c) CSR mixing with Thomas

Figure 4.1: Studying a pattern of a CSR, a Thomas process, and the mixing of the previous two processes in Sec. 4.3.1, 4.3.2 and 4.3.3

The influential events are displayed in Tables 4.2. The visualization is shown via hair-plots and disc-plots which follow hereafter.

4.3.1.1 Hair-plots and Disc-plots

In the first row of Figure 4.2, the hair-functions are plotted. For the left, the definitions of $\text{Hair}_i^{\text{mean}}(\zeta)$ are in (4.4), for the right $\text{Hair}_i^{\text{max}}(\zeta)$ in (4.5) based on the discrepancy in (4.3). In the second row, the radii of the discs are proportional to the A_i^{mean} and A_i^{max}. Influential events are displayed in triangles, their corresponding discs are red and the area inside the disc is shaded.

The parametric discrepancy in (4.9) is not applicable for this case because $K_{\hat{\boldsymbol{\theta}}_{i,j,m}}(h) = K_{\hat{\boldsymbol{\theta}}}(h) = \pi h^2$.

Without any model assumption and using the discrepancy in (4.14), Figure 4.3 shows in the first row the hair-functions $\text{Hair}_{\text{par},i}^{\text{mean}}(\zeta)$ defined in (4.15) on the left and $\text{Hair}_{\text{par},i}^{\text{max}}(\zeta)$ defined in (4.16) on the right.

4.3.1.2 Local Influences and Disc-plots

In Figure 4.4, on the left, local influence according to τ_i^{mean} from (4.19) and on the right, local influence according to τ_i^{max} from (4.20) are displayed.

In Figure 4.5, on the left, local influence according to $\tau_{\text{non},i}^{\text{mean}}$ from (4.19) and on the right, local

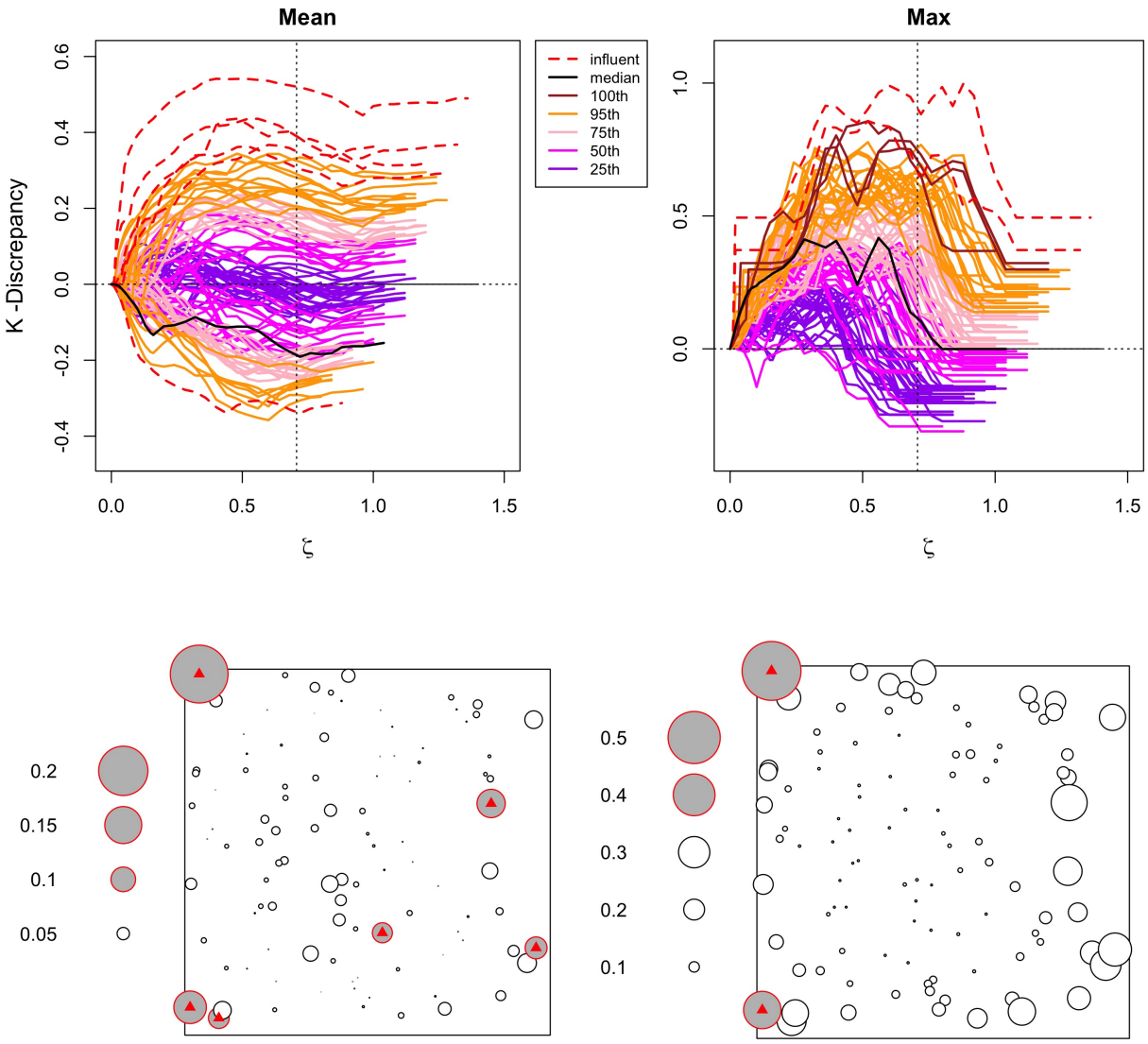


Figure 4.2: The CSR pattern is defined in Sec. 4.3.1. Hair-functions are $\text{Hair}_i^{\text{mean}}(\zeta)$ defined in (4.4) for the left and $\text{Hair}_i^{\max}(\zeta)$ in (4.5) for the right. The colors of the hair-functions are explained in Table 4.1 and in the legend. In the second row, the radii of the discs are proportional to the A_i^{mean} in (4.7) and $A_i^{\max}$ in (4.8). Influential events are displayed in triangles, their corresponding discs are red and the area inside the disc is shaded. These influential events are shown in Table 4.2.

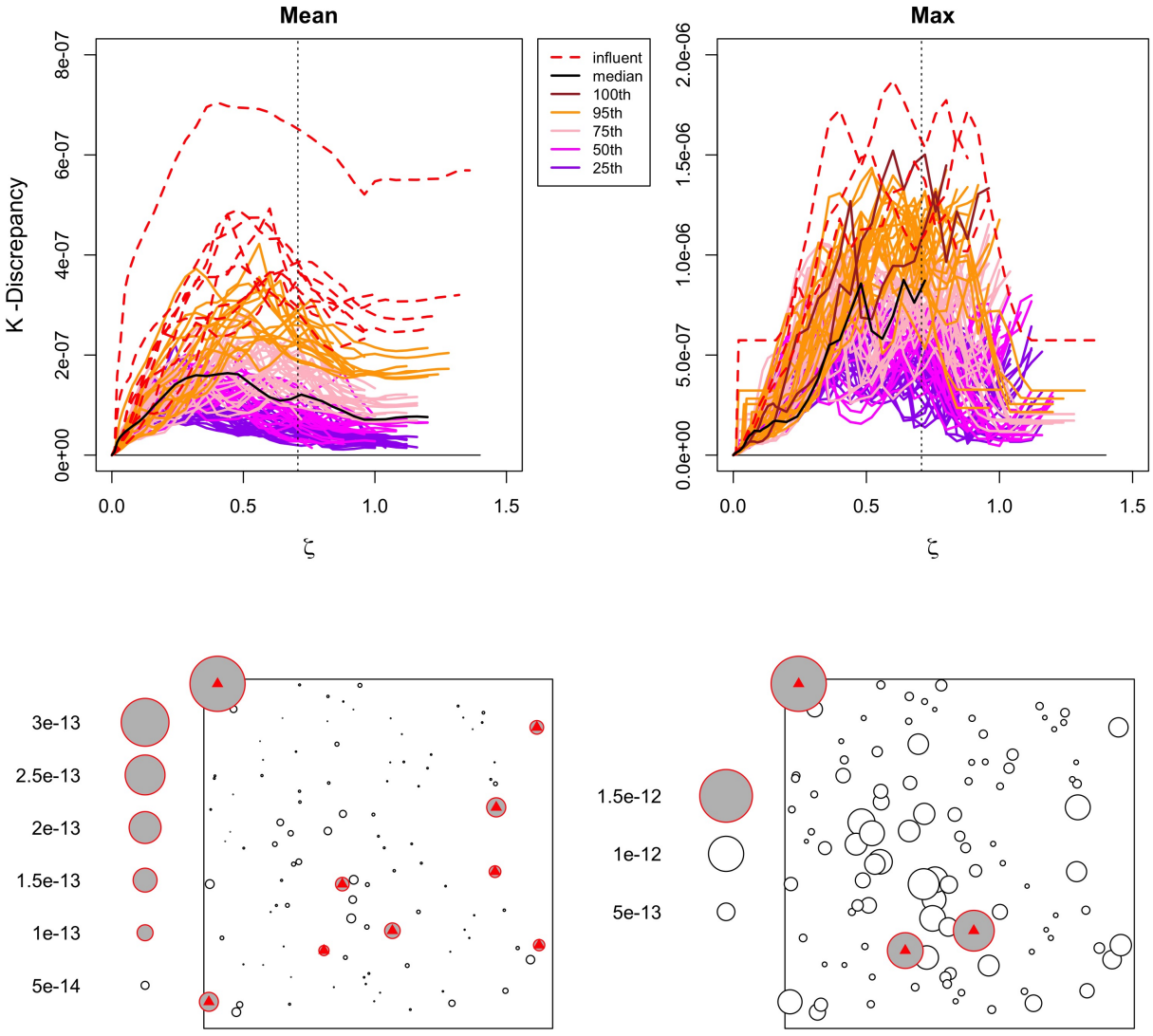


Figure 4.3: The CSR pattern is defined in Sec. 4.3.1. Hair-functions are $\text{Hair}^{\text{mean}}_{\text{non},i}(\zeta)$ defined in (4.15) for the left and $\text{Hair}^{\text{max}}_{\text{non},i}(\zeta)$ (4.16) for the right. The colors of the hair-functions are explained in Table 4.1 and in the legend. In the second row, the radii of the discs are proportional to the $A^{\text{mean}}_{\text{non},i}$ from (4.17) and $A^{\text{max}}_{\text{non},i}$ from (4.18). Influential events are displayed in triangles, their corresponding discs are red and the area inside the disc is shaded. These influential events are shown in Table 4.2.

Fig. 4.2		Fig. 4.3		Fig. 4.4		Fig. 4.5	
A_i^{mean}	$A_i^{\max}$	$A_{\text{non},i}^{\text{mean}}$	$A_{\text{non},i}^{\max}$	τ_i^{mean}	$\tau_i^{\max}$	$\tau_{\text{non},i}^{\text{mean}}$	$\tau_{\text{non},i}^{\max}$
							22
25		25					
30							
		36					
41	**41**	41					
					49		
		62					
				65			
73	**73**	**73**	**73**				
76		76					
		94		94			
96		**96**	**96**				
		99	**99**				99

Table 4.2: Identifying influential events of the CSR process from Sec. 4.3.1. The overlapped influential events are displayed in bold numbers. They are from the mean and the maximum approaches based on a specific discrepancy in either (4.3) for A_i^{mean}, $A_i^{\max}$, τ_i^{mean} and $\tau_i^{\max}$ or (4.14) for $A_{\text{non},i}^{\text{mean}}$, $A_{\text{non},i}^{\max}$, $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\max}$.

influence according to $\tau_{\text{non},i}^{\max}$ in (4.20) are displayed.

The mean and the maximum approach provide an overlap of the influential events 41 and 73 for A_i^{mean} and $A_i^{\max}$ based on the discrepancy in (4.3) and 73, 96, and 99 for $A_{\text{non},i}^{\text{mean}}$ and $A_{\text{non},i}^{\max}$ based on the discrepancy in (4.14). The event 73 which is located at the upper left corner exhibit the biggest influence in all influence types. While the quantity $A_i^{\max}$ seems to suffer from edge effect as the discs are bigger at the margins of the Figure 4.2, $A_{\text{non},i}^{\max}$ does not seem to suffer from edge effect.

For this simulation case, the influential events according to the local influences like τ_i^{mean}, $\tau_i^{\max}$, $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\max}$ do not have much in common with the ones according to the influence A_i^{mean}, $A_i^{\max}$, $A_{\text{non},i}^{\text{mean}}$ and $A_{\text{non},i}^{\max}$. While the quantity $\tau_i^{\max}$ seems to suffer slightly from edge effect as the discs are bigger at the left edge of the Figure 4.3, $\tau_{\text{non},i}^{\max}$ does not seem to suffer from an edge effect at all.

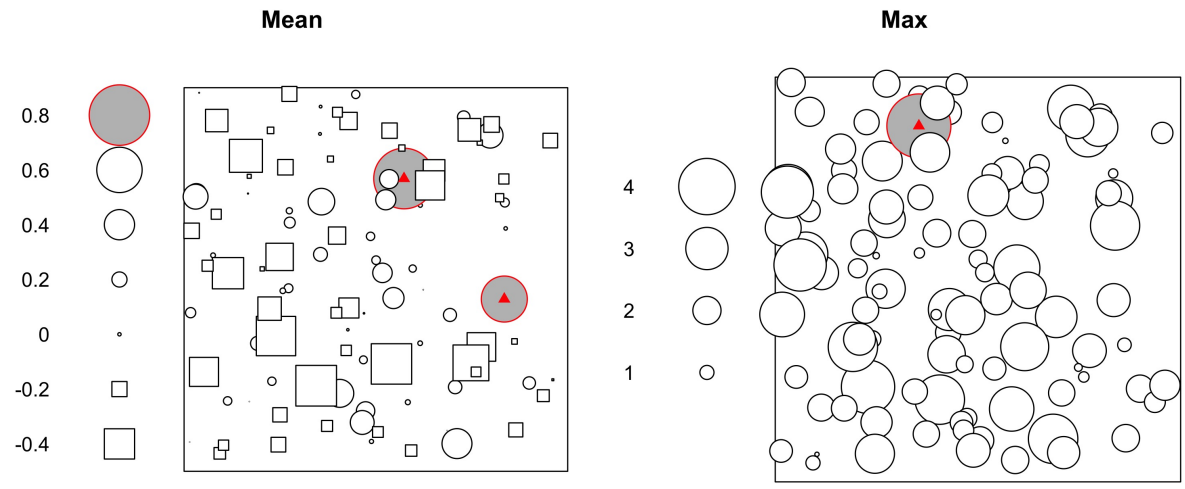


Figure 4.4: The CSR pattern is defined in Sec. 4.3.1. The radii of the discs are proportional to τ_i^{mean} defined in (4.19) for the left and τ_i^{max} in (4.20) for the right. These quantities are the rates of change at $\zeta = 0$ of $\text{Hair}_i^{\text{mean}}(\zeta)$ from (4.4) and $\text{Hair}_i^{\text{max}}(\zeta)$ from (4.5), respectively.

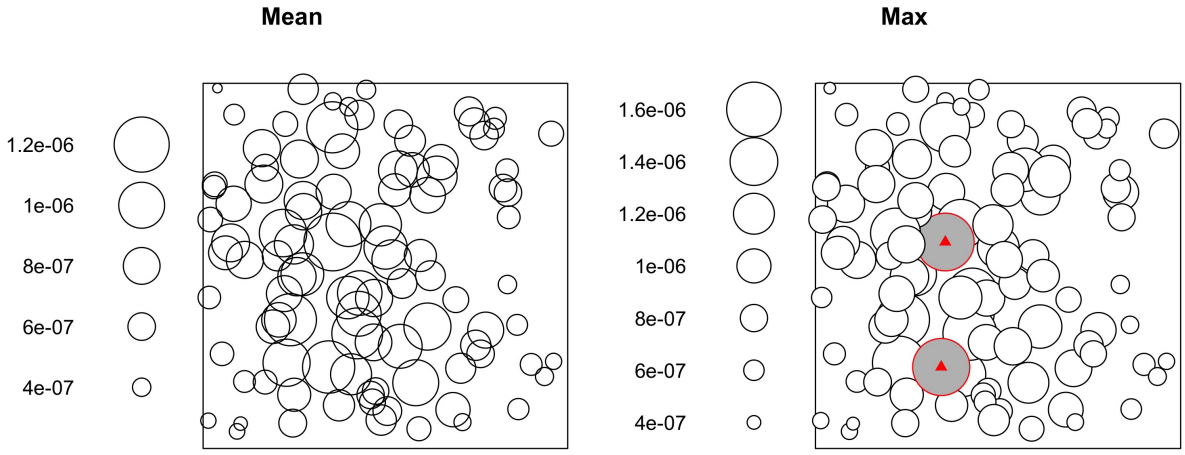


Figure 4.5: The CSR pattern is defined in Sec. 4.3.1. The radii of the discs are proportional to $\tau_{\text{non},i}^{\text{mean}}$ defined in (4.19) for the left and $\tau_{\text{non},i}^{\text{max}}$ in (4.20) for the right. These quantities are the rates of change at $\zeta = 0$ of $\text{Hair}_{\text{non},i}^{\text{mean}}(\zeta)$ from (4.15) and $\text{Hair}_{\text{non},i}^{\text{max}}(\zeta)$ from (4.16), respectively. The discrepancy in (4.14) is applied here.

4.3.2 Thomas Process

A SPP of 100 events is generated with $\kappa = 10$, $\sigma = 0.03$, $\mu = 10$, in Figure 4.1b. I apply the AGOF-K test from Dao and Genton [2014] to verify whether the simulated SPP represents the attributes of the Thomas process specified above well. The estimated p-value from the AGOF-K test is $\hat{p} = 0.977$ and the adjusted α-level is $\hat{\alpha}^* = 0.26$ at the nominal significance level $\alpha = 0.05$. Since $\hat{p} > \hat{\alpha}^*$, I do not reject the hypothesis that the SPP represents the Thomas process well.

The influential events are reported in Table 4.3. The visualization is shown via hair-plots and disc-plots which follow hereafter.

4.3.2.1 Hair-plots and Disc-plots

In the first row of Figure 4.6, the hair-functions are plotted. For the left, the definitions of $\text{Hair}_i^{\text{mean}}(\zeta)$ are in (4.4), for the right, $\text{Hair}_i^{\text{max}}(\zeta)$ in (4.5) based on the discrepancy in (4.3). Here, A_i^{mean} and A_i^{max} produce the same sets of influential events, $17, 19, \ldots, 24$.

In the second row Figure 4.6, the radii of the discs are proportional to the A_i^{mean} and A_i^{max}. Influential events are displayed in triangles, their corresponding discs are red and the area inside the disc is shaded.

In the first row of Figure 4.7, the hair-functions are plotted. For the left, the definitions of $\text{Hair}_{\text{par},i}^{\text{mean}}(\zeta)$ are in (4.10), for the right $\text{Hair}_{\text{par},i}^{\text{max}}(\zeta)$ in (4.11) based on the discrepancy in (4.9). Here, $A_{\text{par},i}^{\text{mean}}$ and $A_{\text{par},i}^{\text{max}}$ lead to disjoint sets of influential events.

Without any model assumption, in the first row of Figure 4.8, the hair-functions are plotted. For the left, the definitions of $\text{Hair}_{\text{non},i}^{\text{mean}}(\zeta)$ are in (4.15), for the right $\text{Hair}_{\text{non},i}^{\text{max}}(\zeta)$ in (4.16) based on the discrepancy in (4.14). $A_{\text{non},i}^{\text{mean}}$ produces only a single influential event and $A_{\text{non},i}^{\text{max}}$ produces no influential events.

4.3.2.2 Local Influences and Disc-plots

In Figure 4.9, on the left, local influence according to τ_i^{mean} from (4.19) and on the right, local influence according to τ_i^{max} from (4.20) are displayed. On the left, the local influential events are 19, 21, 52, 57 and 58 according to τ_i^{mean}. On the right, there is no influential events according to

Fig. 4.6		Fig. 4.7		Fig. 4.8		Fig. 4.9		Fig. 4.10		Fig. 4.11	
A_i^{mean}	A_i^{max}	$A_{\text{par},i}^{\text{mean}}$	$A_{\text{par},i}^{\text{max}}$	$A_{\text{non},i}^{\text{mean}}$	$A_{\text{non},i}^{\text{max}}$	τ_i^{mean}	τ_i^{max}	$\tau_{\text{par},i}^{\text{mean}}$	$\tau_{\text{par},i}^{\text{max}}$	$\tau_{\text{non},i}^{\text{mean}}$	$\tau_{\text{non},i}^{\text{max}}$
		11									
			14								
17	**17**	17		17							
19	**19**		19			19					
20	**20**		20								
21	**21**					21					
22	**22**		22								
23	**23**		23								
24	**24**		24								
						26	**26**				
						37	**37**				
						38	**38**				
			39								
			51								
		52				52		**52**	**52**	**52**	**52**
			57			57		**57**	**57**	**57**	**57**
						58		**58**	**58**	**58**	**58**
			64								
			68								
			85								
			89								
			96					96			

Table 4.3: Identifying influential events of the Thomas process from Sec. 4.3.2. The overlapped influential events are displayed in bold numbers. They are from the mean and the maximum approaches based on a specific discrepancy in either (4.3) for A_i^{mean}, A_i^{max}, τ_i^{mean} and τ_i^{max}, or (4.9) for $A_{\text{par},i}^{\text{mean}}$, $A_{\text{par},i}^{\text{max}}$, $\tau_{\text{par},i}^{\text{mean}}$ and $\tau_{\text{par},i}^{\text{max}}$, or (4.14) for $A_{\text{non},i}^{\text{mean}}$, $A_{\text{non},i}^{\text{max}}$, $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\text{max}}$.

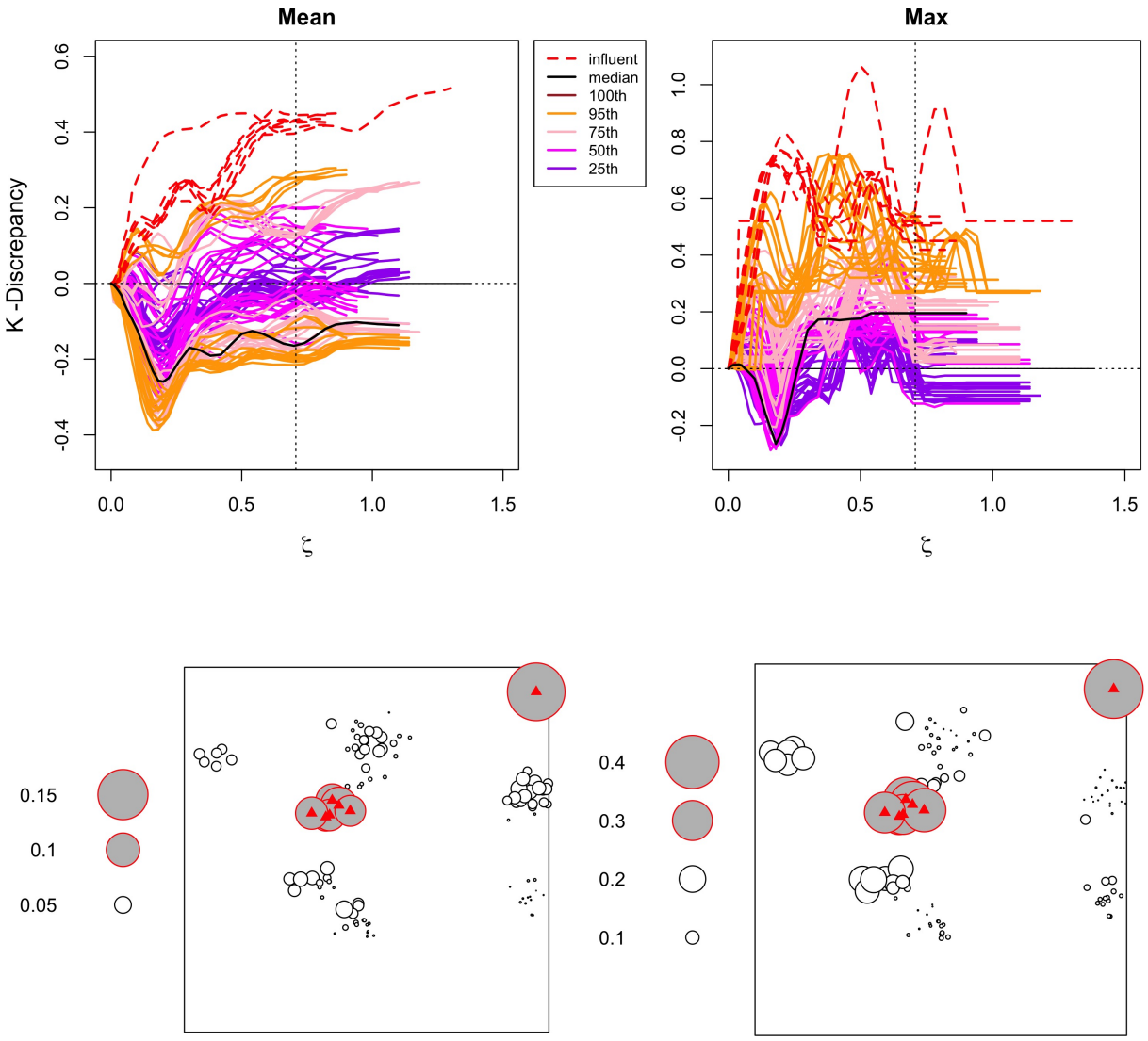


Figure 4.6: The Thomas pattern is defined in Sec. 4.3.2. The discrepancy is defined in (4.3). There are hair-functions $\text{Hair}_i^{\text{mean}}(\zeta)$ and $\text{Hair}_i^{\max}(\zeta)$ are defined in (4.4) on the left and in (4.5) on the right. The colors of the hair-functions are explained in Table 4.1 and in the legend. In the second row, the radii of the discs are proportional to A_i^{mean} and $A_i^{\max}$. Influential events are displayed in triangles, their corresponding discs are red and the area inside the disc is shaded.

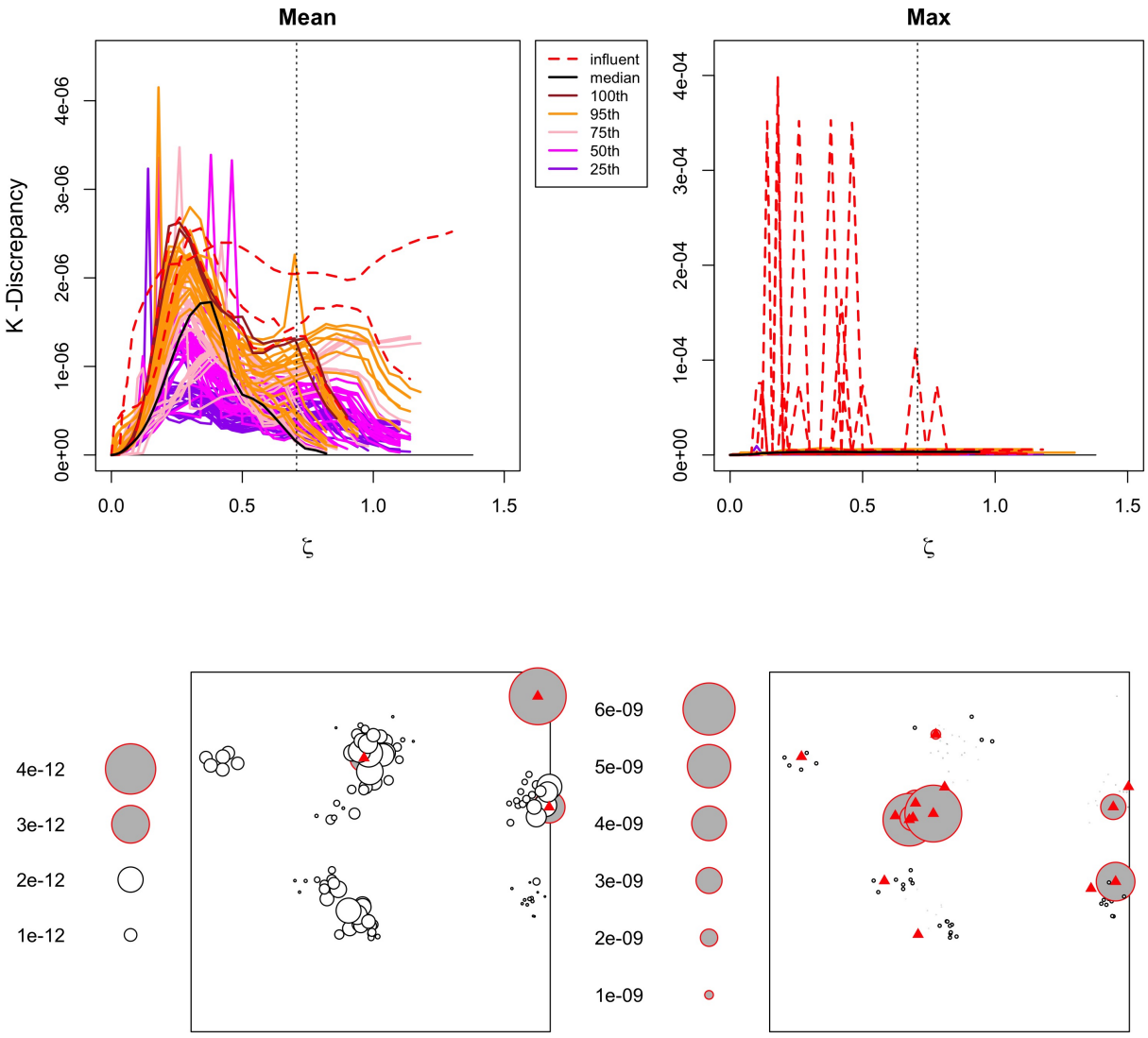


Figure 4.7: The Thomas pattern is defined in Sec. 4.3.2. The parametric discrepancy is defined in (4.9). There are hair-functions $\mathrm{Hair}^{\mathrm{mean}}_{\mathrm{par},i}(\zeta)$ and $\mathrm{Hair}^{\mathrm{max}}_{\mathrm{par},i}(\zeta)$ are defined in (4.10) on the left and in (4.11) on the right. The colors of the hair-functions are explained in Table 4.1 and in the legend. In the second row, the radii of the discs are proportional to $A^{\mathrm{mean}}_{\mathrm{par},i}$ and $A^{\mathrm{max}}_{\mathrm{par},i}$. Influential events are displayed in triangles, their corresponding discs are red and the area inside the disc is shaded.

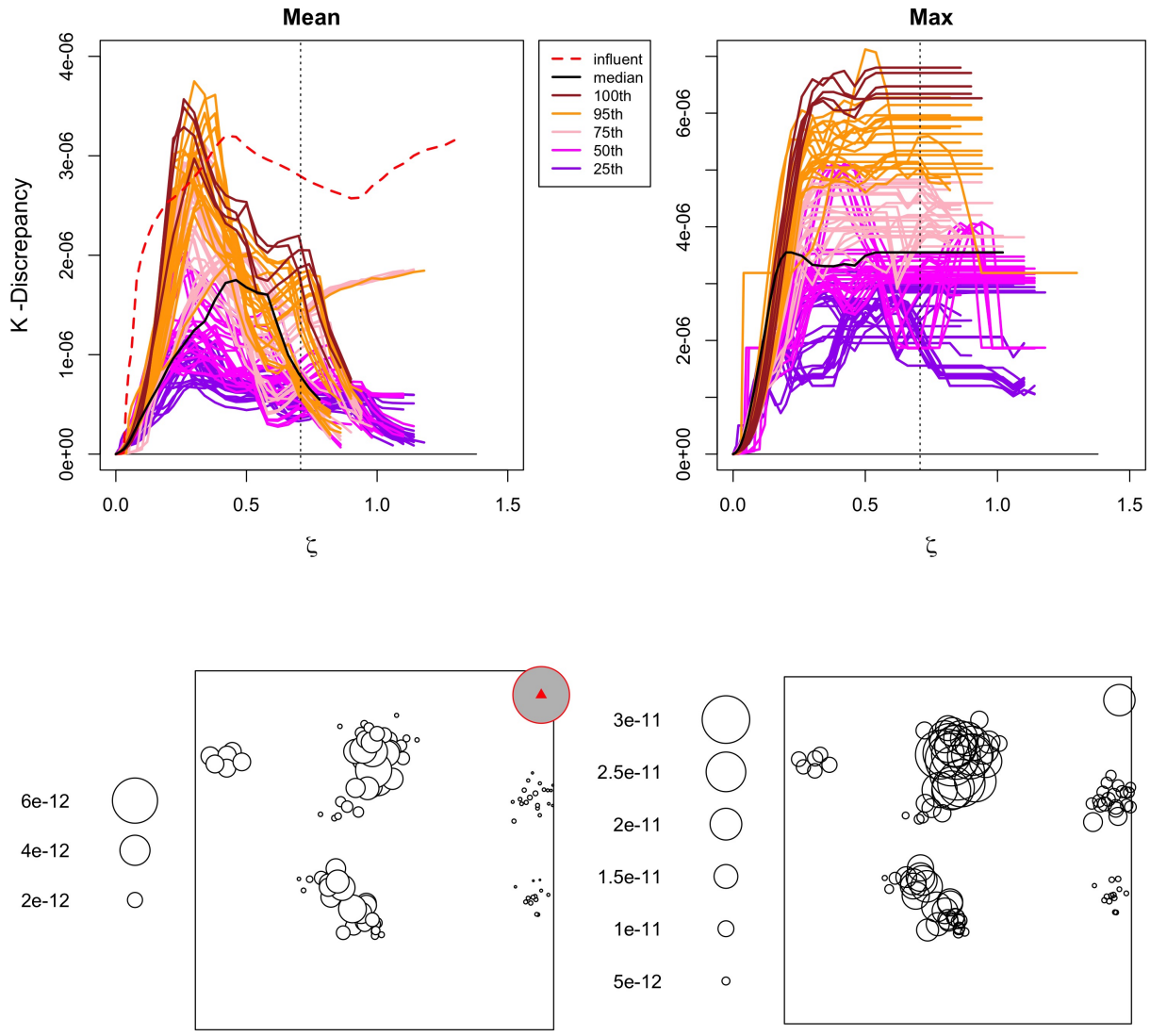


Figure 4.8: The Thomas pattern is defined in Sec. 4.3.2. The nonparametric discrepancy is defined in (4.14). There are hair-functions are defined in (4.15) on the left and in (4.16) on the right. The colors of the hair-functions are explained in Table 4.1 and in the legend. In the second row, the radii of the discs are proportional to $A^{\text{mean}}_{\text{non},i}$ and $A^{\text{max}}_{\text{non},i}$. Influential events are displayed in triangles, their corresponding discs are red and the area inside the disc is shaded.

$\tau_i^{\max}$.

In Figure 4.10, on the left, local influence according to $\tau_{\text{par},i}^{\text{mean}}$ from (4.19) and on the right, local influence according to $\tau_{\text{par},i}^{\max}$ from (4.20) are displayed. The two approaches produce a joint set of influential events: 26, 37, 38, 52, 57 and 58. Beside these influential events, on the left, the local influential events have 96 as another event according to τ_i^{mean}.

In Figure 4.11, on the left, local influence according to $\tau_{\text{non},i}^{\text{mean}}$ from (4.19) and on the right, local influence according to $\tau_{\text{non},i}^{\max}$ from (4.20) are displayed. The two approaches produce the same set of influential events: 52, 57 and 58.

In the case of exploring influential events of the Thomas SPP of 100 events, Table 4.3 and Figures 4.6 to 4.11 show that there are more influential events than in the case of exploring influential events of the CSR SPP also of 100 events documented in Table 4.2, Figures 4.2 to 4.5 in Section 4.3.1. I think it is understandable as the clustered structure of the Thomas SPP is more complicated than the complete randomness in the case of the CSR SPP. In general, the events of a clustered pattern have more relevance and hence are more influential if they are perturbed because the perturbation can break down the data structure in a highly dependent data structure but not so in a complete randomness (CSR) case.

4.3.3 Mixing of Complete Spatial Randomness and Thomas Process

I study a union of two sub-patterns described as follows. A sub-pattern with 59 events from the window $[0; 0.5) \times [0; 1]$ of the CSR SPP studied in the Section 4.3.1 and a sub-pattern with 52 events from the window $[0.5; 1] \times [0; 1]$ of the Thomas SPP generated with $\kappa = 10$, $\sigma = 0.03$, $\mu = 10$ as studied in the Section 4.3.2. Now, I determine whether the mixed SPP has the CSR structure. I apply the GOF-K test from Diggle [2003] and the p-value is 0. Hence I reject the hypothesis that CSR provides a good fit for the mixed SPP. Then I apply the AGOF test from Dao and Genton [2014] to test the hypothesis that a Thomas process can provide a good fit for the mixed SPP. The estimated p-value is 0.55 and the adjusted α-level is $\hat{\alpha}^* = 0.22$ when $\alpha = 0.05$ is considered. I fail to reject the hypothesis that a Thomas process can provide a good fit to the mixed SPP. In the following, I explore whether there are any influential events existent for this mixed SPP.

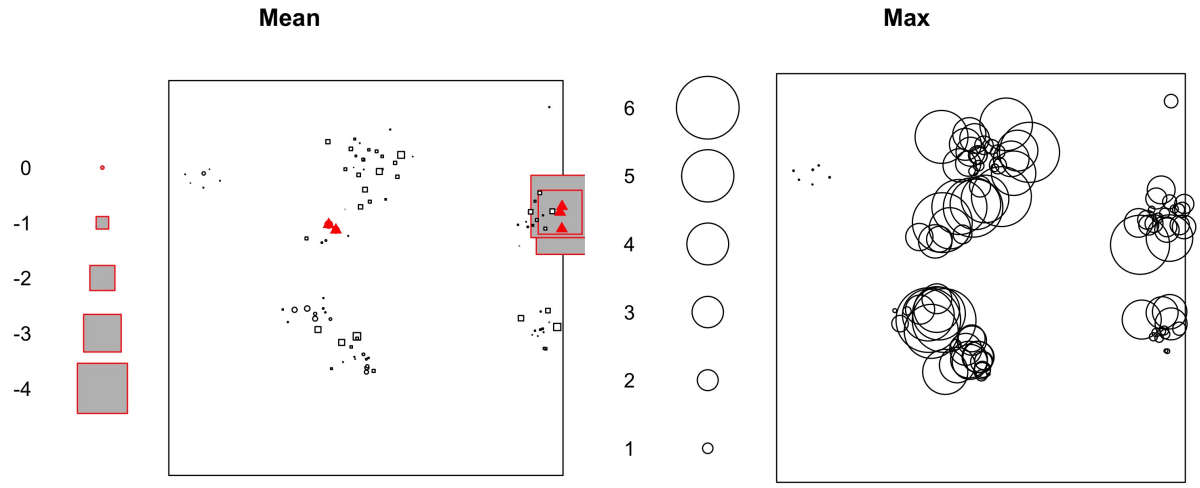


Figure 4.9: The Thomas pattern is defined in Sec. 4.3.2. The radii of the discs are proportional to τ_i^{mean} defined in (4.19) for the left and τ_i^{max} in (4.20) for the right. These quantities are the rates of change at $\zeta = 0$ of $\text{Hair}_i^{\text{mean}}(\zeta)$ from (4.4) and $\text{Hair}_i^{\text{max}}(\zeta)$ from (4.5), respectively. The discrepancy in (4.3) is applied here.

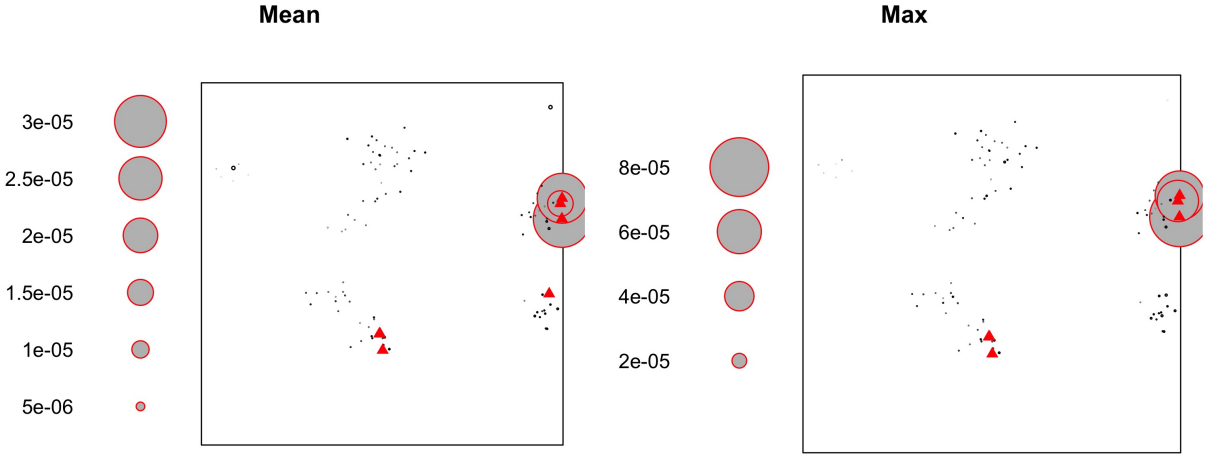


Figure 4.10: The Thomas pattern is defined in Sec. 4.3.2. The radii of the discs are proportional to $\tau_{\text{par},i}^{\text{mean}}$ defined in (4.19) for the left and $\tau_{\text{par},i}^{\text{max}}$ in (4.20) for the right. These quantities are the rates of change at $\zeta = 0$ of $\text{Hair}_{\text{par},i}^{\text{mean}}(\zeta)$ from (4.10) and $\text{Hair}_{\text{par},i}^{\text{max}}(\zeta)$ from (4.11), respectively. The discrepancy in (4.9) is applied here.

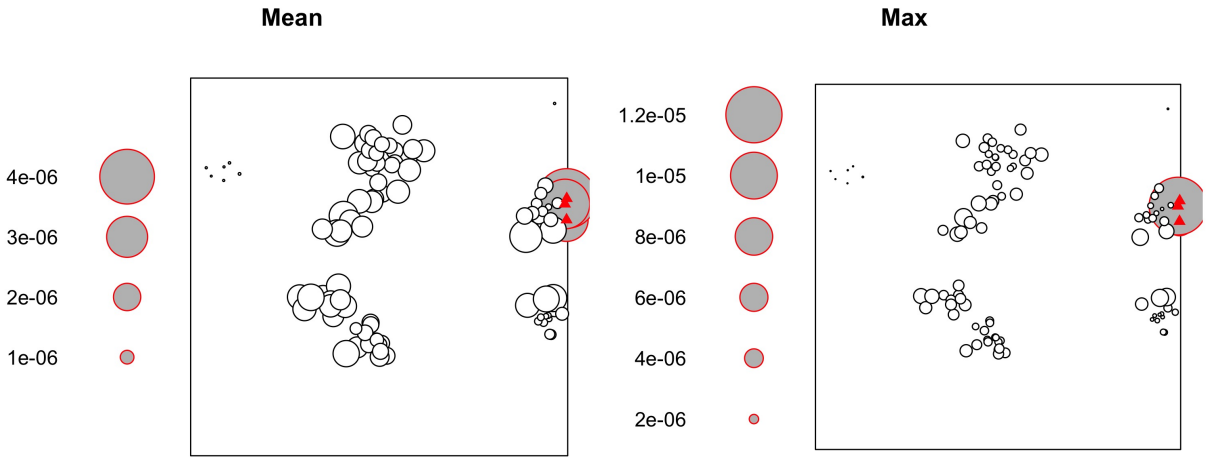


Figure 4.11: The Thomas pattern is defined in Sec. 4.3.2. The radii of the discs are proportional to $\tau^{\text{mean}}_{\text{non},i}$ defined in (4.19) for the left and $\tau^{\text{max}}_{\text{non},i}$ in (4.20) for the right. These quantities are the rates of change at $\zeta = 0$ of $\text{Hair}^{\text{mean}}_{\text{non},i}(\zeta)$ from (4.15) and $\text{Hair}^{\text{max}}_{\text{non},i}(\zeta)$ from (4.16), respectively. The discrepancy in (4.14) is applied here.

The influential events are displayed in Tables 4.4. The visualization is shown via hair-plots and disc-plots which follow hereafter.

Table 4.4: Identifying influential events of the mixed process from Sec. 4.3.3. The overlapped influential events are displayed in bold numbers. They are from the mean and the maximum approaches based on a specific discrepancy in either (4.3) for A^{mean}_i, A^{max}_i, τ^{mean}_i and τ^{max}_i, or (4.9) for $A^{\text{mean}}_{\text{par},i}$, $A^{\text{max}}_{\text{par},i}$, $\tau^{\text{mean}}_{\text{par},i}$ and $\tau^{\text{max}}_{\text{par},i}$, or (4.14) for $A^{\text{mean}}_{\text{non},i}$, $A^{\text{max}}_{\text{non},i}$, $\tau^{\text{mean}}_{\text{non},i}$ and $\tau^{\text{max}}_{\text{non},i}$.

Fig. 4.12		Fig. 4.13		Fig. 4.14		Fig. 4.15		Fig. 4.16		Fig. 4.17	
A^{mean}_i	A^{max}_i	$A^{\text{mean}}_{\text{par},i}$	$A^{\text{max}}_{\text{par},i}$	$A^{\text{mean}}_{\text{non},i}$	$A^{\text{max}}_{\text{non},i}$	τ^{mean}_i	τ^{max}_i	$\tau^{\text{mean}}_{\text{par},i}$	$\tau^{\text{max}}_{\text{par},i}$	$\tau^{\text{mean}}_{\text{non},i}$	$\tau^{\text{max}}_{\text{non},i}$
								4	**4**		
		11						**11**	**11**		
			14								
								16	**16**		
17		17									

Continued on next page

Table 4.4 – continued from previous page

Fig. 4.12		Fig. 4.13		Fig. 4.14		Fig. 4.15		Fig. 4.16		Fig. 4.17	
A_i^{mean}	A_i^{max}	$A_{\text{par},i}^{\text{mean}}$	$A_{\text{par},i}^{\text{max}}$	$A_{\text{non},i}^{\text{mean}}$	$A_{\text{non},i}^{\text{max}}$	τ_i^{mean}	τ_i^{max}	$\tau_{\text{par},i}^{\text{mean}}$	$\tau_{\text{par},i}^{\text{max}}$	$\tau_{\text{non},i}^{\text{mean}}$	$\tau_{\text{non},i}^{\text{max}}$
			19								
			20								
22			22								
			23								
			24								
								33	**33**		
								36	**36**		
			39								
								42	**42**		
43				**43**	**43**						
			51								
		52									
55											
			57								
58	**58**					58					
								60	**60**		
								62	**62**		
			64								
									66		
								67	**67**		
			68								
								69	**69**		
71				**71**	**71**						

Continued on next page

Table 4.4 – continued from previous page

Fig. 4.12		Fig. 4.13		Fig. 4.14		Fig. 4.15		Fig. 4.16		Fig. 4.17	
A_i^{mean}	A_i^{max}	$A_{\text{par},i}^{\text{mean}}$	$A_{\text{par},i}^{\text{max}}$	$A_{\text{non},i}^{\text{mean}}$	$A_{\text{non},i}^{\text{max}}$	τ_i^{mean}	τ_i^{max}	$\tau_{\text{par},i}^{\text{mean}}$	$\tau_{\text{par},i}^{\text{max}}$	$\tau_{\text{non},i}^{\text{mean}}$	$\tau_{\text{non},i}^{\text{max}}$
								73	**73**		
				78	**78**	**78**	**78**	**78**	**78**	**78**	**78**
				79	**79**						
				80	**80**						
				81	**81**						
				82	**82**	82		**82**	**82**		82
						83	**83**			**83**	**83**
				84		**84**	**84**	**84**	**84**	**84**	**84**
			85	**85**	**85**			**85**	**85**		
				87	**87**						
			89	**89**	**89**						
90				**90**	**90**		90			**90**	**90**
				91	**91**						
				92	**92**						
				93	**93**			**93**	**93**		
94	**94**			**94**	**94**		94			**94**	**94**
				95	**95**						
			96	**96**	**96**						
				97	**97**						
				98	**98**						
									99		
							100	**100**	**100**	**100**	**100**

Continued on next page

Table 4.4 – continued from previous page

Fig. 4.12		Fig. 4.13		Fig. 4.14		Fig. 4.15		Fig. 4.16		Fig. 4.17	
A_i^{mean}	A_i^{max}	$A_{\text{par},i}^{\text{mean}}$	$A_{\text{par},i}^{\text{max}}$	$A_{\text{non},i}^{\text{mean}}$	$A_{\text{non},i}^{\text{max}}$	τ_i^{mean}	τ_i^{max}	$\tau_{\text{par},i}^{\text{mean}}$	$\tau_{\text{par},i}^{\text{max}}$	$\tau_{\text{non},i}^{\text{mean}}$	$\tau_{\text{non},i}^{\text{max}}$
								105	**105**		
						107	**107**			**107**	**107**
							108	**108**	**108**	**108**	**108**
								109	**109**		
						110					

4.3.3.1 Hair-plots and Disc-plots

In the first row of Figure 4.12, the hair-functions are plotted. For the left, the definitions of $\text{Hair}_i^{\text{mean}}(\zeta)$ are in (4.4), for the right $\text{Hair}_i^{\text{max}}(\zeta)$ in (4.5) based on the discrepancy in (4.3). Here, A_i^{mean} and A_i^{max} produce the different sets of influential events. However, they share two common influential events 58 and 94. In the second row Figure 4.12, the radii of the discs are proportional to the A_i^{mean} and A_i^{max}. Influential events are displayed in triangles, their corresponding discs are red and the area inside the disc is shaded.

In the first row of Figure 4.13, the hair-functions are plotted. For the left, the definitions of $\text{Hair}_{\text{par},i}^{\text{mean}}(\zeta)$ are in (4.10), for the right $\text{Hair}_{\text{par},i}^{\text{max}}(\zeta)$ in (4.11) based on the discrepancy in (4.9). Here, $A_{\text{par},i}^{\text{mean}}$ and $A_{\text{par},i}^{\text{max}}$ lead to disjoint sets of influential events.

Without any model assumption, in the first row of Figure 4.14, the hair-functions are plotted. For the left, the definitions of $\text{Hair}_{\text{non},i}^{\text{mean}}(\zeta)$ are in (4.15) and for the right, $\text{Hair}_{\text{non},i}^{\text{max}}(\zeta)$ in (4.16) based on the discrepancy in (4.14). $A_{\text{non},i}^{\text{mean}}$ and $A_{\text{non},i}^{\text{max}}$ produce a large joint set of influential events: $43, 71, \ldots, 82, 85, 87, 89, \ldots, 98$. Beside this set, $A_{\text{non},i}^{\text{mean}}$ produces 84 as another influential event.

4.3.3.2 Local Influences and Disc-plots

In Figure 4.15, on the left, local influence according to τ_i^{mean} from (4.19) and on the right, local influence according to τ_i^{max} from (4.20) are displayed. τ_i^{mean} and τ_i^{max} have different sets influential events. However, they have 4 influential events in common 78, 83, 84 and 107. The radii of the

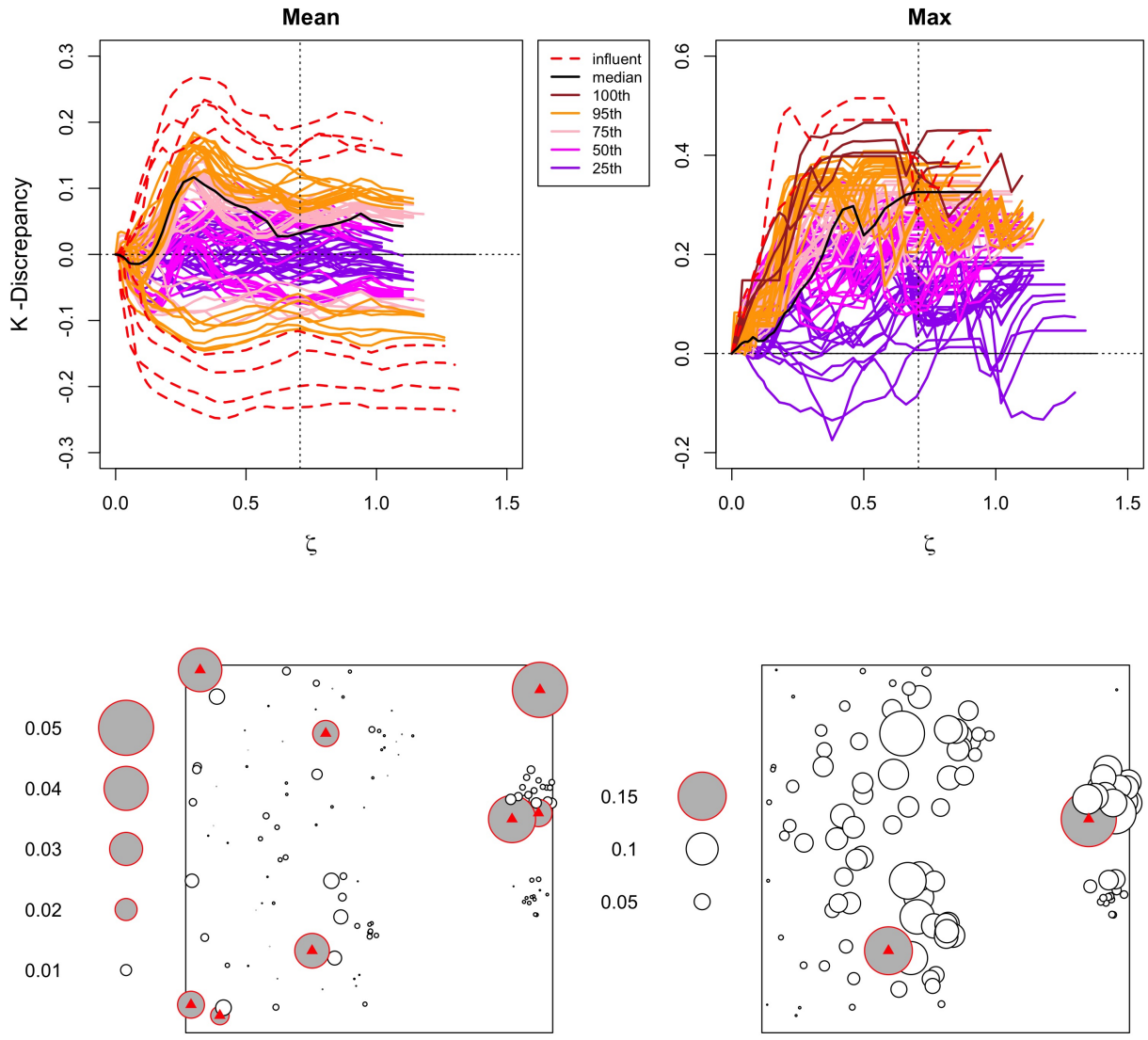


Figure 4.12: The pattern of the mixing of the CSR from Sec. 4.3.1 and the Thomas pattern from Sec. 4.3.2 is defined in Sec. 4.3.3. The discrepancy in (4.3) is applied here. In the first row, there are hair-functions defined in (4.4) for the left and in (4.5) for the right. The colors of the hair-functions are explained in Table 4.1 and in the legend. In the second row, the radii of the discs are proportional to A_i^{mean} and A_i^{max}. Influential events are displayed in triangles, their corresponding discs are red and the area inside the disc is shaded.

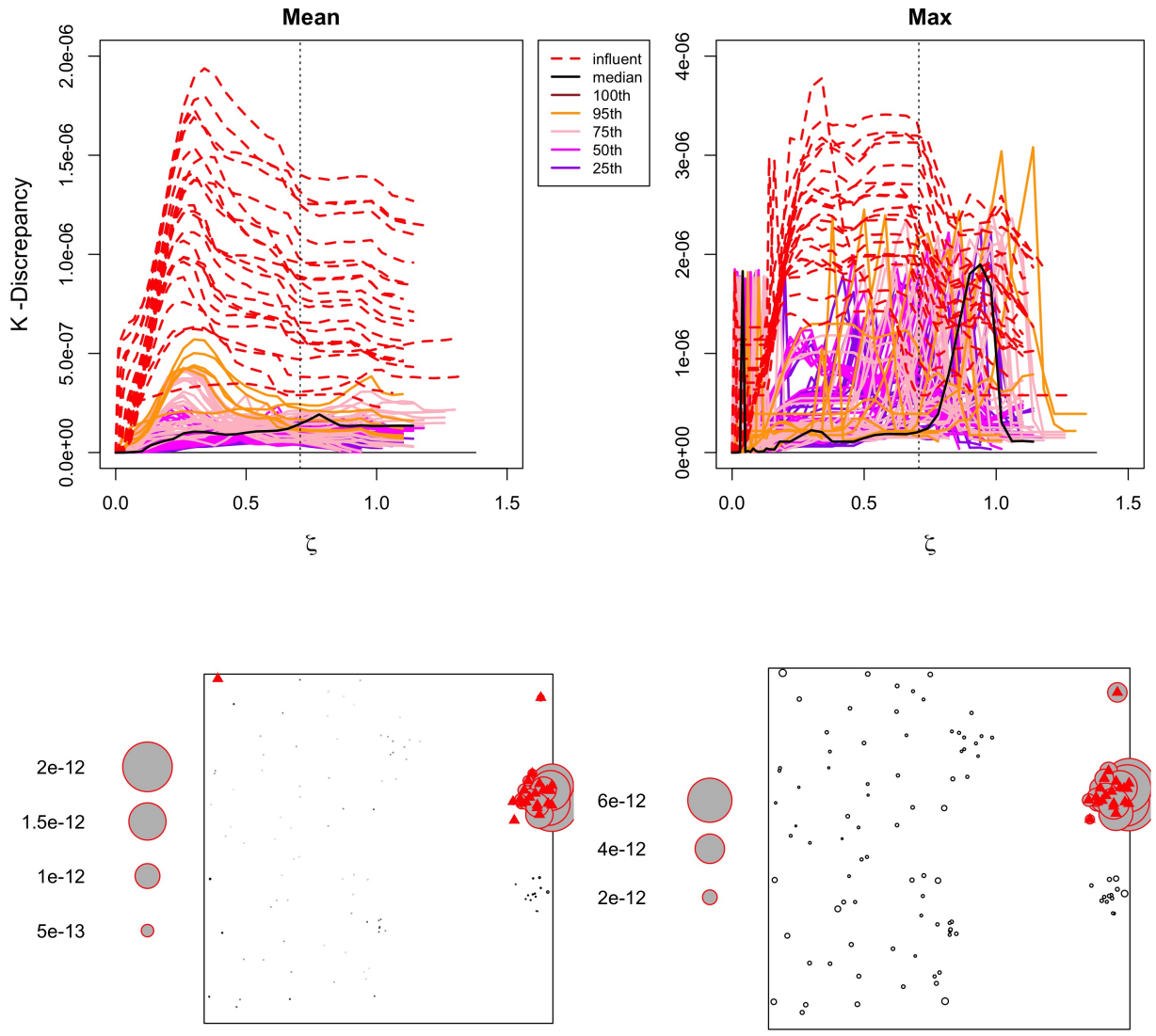


Figure 4.13: The pattern of the mixing of the CSR from Sec. 4.3.1 and the Thomas pattern from Sec. 4.3.2 is defined in Sec. 4.3.3. The parametric discrepancy in (4.9) is applied here. In the first row, there are hair-functions defined in (4.10) an (4.11). The colors of the hair-functions are explained in Table 4.1 and in the legend. In the second row, the radii of the discs are proportional to $A^{\text{mean}}_{\text{par},i}$ and $A^{\max}_i$. Influential events are displayed in triangles, their corresponding discs are red and the area inside the disc is shaded.

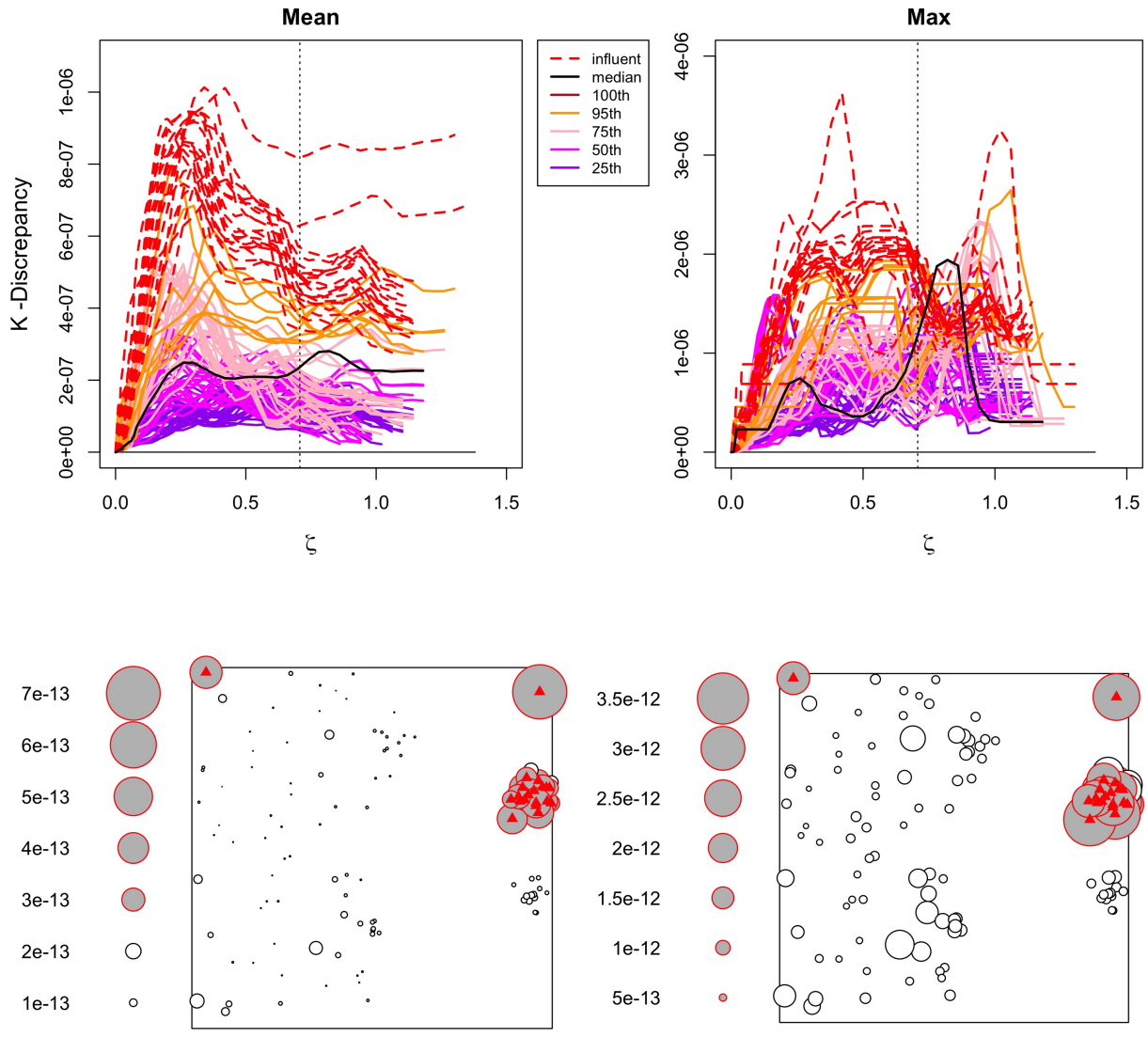


Figure 4.14: The pattern of the mixing of the CSR from Sec. 4.3.1 and the Thomas pattern from Sec. 4.3.2 is defined in Sec. 4.3.3. The nonparametric discrepancy in (4.14) is applied here. In the first row, there are hair-functions defined in (4.15) for the left and (4.16) for the right. The colors of the hair-functions are explained in Table 4.1 and in the legend. In the second row, the radii of the discs are proportional to $A^{\text{mean}}_{\text{non},i}$ and $A^{\text{max}}_{\text{non},i}$. Influential events are displayed in triangles, their corresponding discs are red and the area inside the disc is shaded.

discs are proportional to the τ_i's. Influential events are displayed in triangles, their corresponding discs are red and the area inside the disc is shaded.

In Figure 4.16, on the left, local influence according to $\tau_{\text{par},i}^{\text{mean}}$ from (4.19) and on the right, local influence according to $\tau_{\text{par},i}^{\text{max}}$ from (4.20) are displayed. The two approaches produce a large joint set of influential events. On the right, $\tau_{\text{par},i}^{\text{max}}$ produces additionally 66, 76 and 99 as another influential events.

In Figure 4.17, on the left, local influence according to $\tau_{\text{non},i}^{\text{mean}}$ from (4.19) and on the right, local influence according to $\tau_{\text{non},i}^{\text{max}}$ from (4.20) are displayed. The two approaches produce a large joint set of influential events. On the right, $\tau_{\text{non},i}^{\text{max}}$ produces additionally 82 as another influential event.

The pattern from Sec. 4.3.3 did not reject the AGOF-K test for a Thomas process. That means that the Thomas model contributes a more relevant role to the model structure of the mixing pattern. Subsequently, most scenarios in Sec 4.3.3 studying influential events confirmed that the majority of influential events come from the Thomas sub-pattern. This validates the conclusion of AGOF-K test for a Thomas process for this SPP.

I observe that influential events resulting from τ_i^{mean}, τ_i^{max}, $\tau_{\text{par},i}^{\text{mean}}$, $\tau_{\text{par},i}^{\text{max}}$, $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\text{max}}$ do not coincide with the influential events derived from the SAUC's such as A_i^{mean}, A_i^{max}, $A_{\text{par},i}^{\text{mean}}$, $A_{\text{par},i}^{\text{max}}$, $A_{\text{non},i}^{\text{mean}}$ and $A_{\text{non},i}^{\text{max}}$. Here, while the latter quantify the influence of those events, the corresponding τ_i^{mean} and τ_i^{max} describe the rates of change in the corresponding SAUC's.

In the case of exploring influential events of the mixed SPP of 111 events, Table 4.4 and Figures 4.12 to 4.17 show that there are more influential events than in the case of exploring influential events for the Thomas SPP of 100 events documented in Table 4.3, Figures 4.6 to 4.11 in Section 4.3.2. I think it is understandable as the mixed structure of a Thomas and a CSR SPPs is more complicated than the clustered pattern in the case of the Thomas SPP. In general, the events of a complicated pattern (mixing of two SPPs) have more relevance and hence are more influential if they get perturbed because the perturbation can change the model assumptions quickly for a inhomogeneous data structure but not so in a homogeneous (Thomas) case.

Overall, I observe that the nonparametric discrepancy (4.14) facilitates the mean and maximum

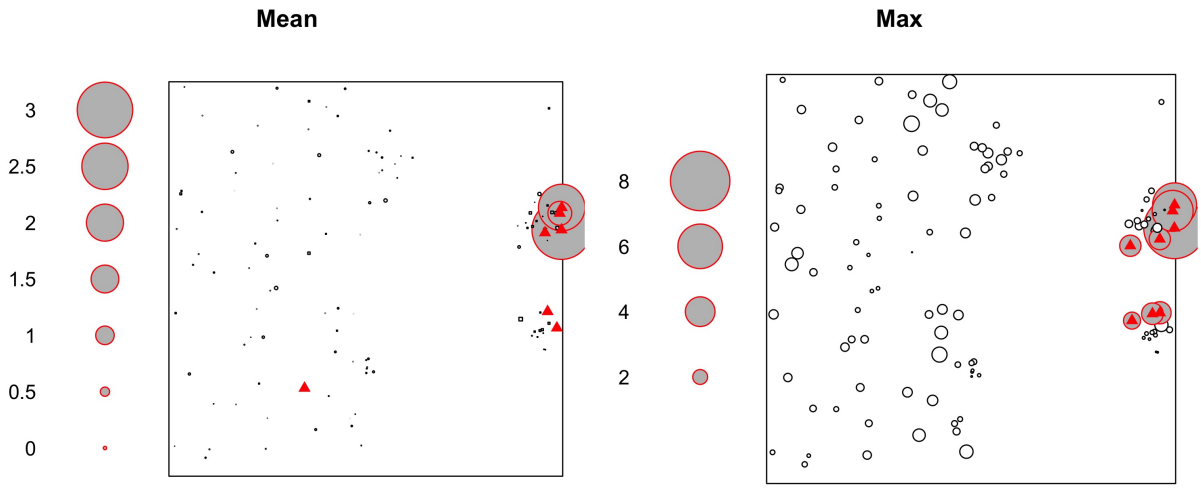


Figure 4.15: The mixed pattern is defined in Sec. 4.3.3. The radii of the discs are proportional to τ_i^{mean} defined in (4.19) for the left and τ_i^{max} in (4.20) for the right. These quantities are the rates of change at $\zeta = 0$ of $\text{Hair}_i^{\text{mean}}(\zeta)$ from (4.4) and $\text{Hair}_i^{\text{max}}(\zeta)$ from (4.5), respectively. The discrepancy in (4.3) is applied here.

approaches to detect a more consistent sets of the influential events, i.e., according to $A_{\text{non},i}^{\text{mean}}$, $A_{\text{non},i}^{\text{max}}$, $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\text{max}}$.

4.4 Data Applications

Figure 4.18 demonstrates two datasets, copper ores and Swedish pines, which are subjected to testing my method. Both were studied in the context of influential diagnostics for spatial point processes in Baddeley et al. [2013] and Baddeley et al. [2019]. With my new approach, I aim at adding more insights to the datasets.

In the following subsections, the theoretical K-function is not numerically complete. Hence, I used $\tilde{K}$-function, the approximation of the theoretical K-function for a Gibbs process. The computation of $\tilde{K}$-function uses the second-order Poisson-saddlepoint approximation [Baddeley and Nair, 2012].

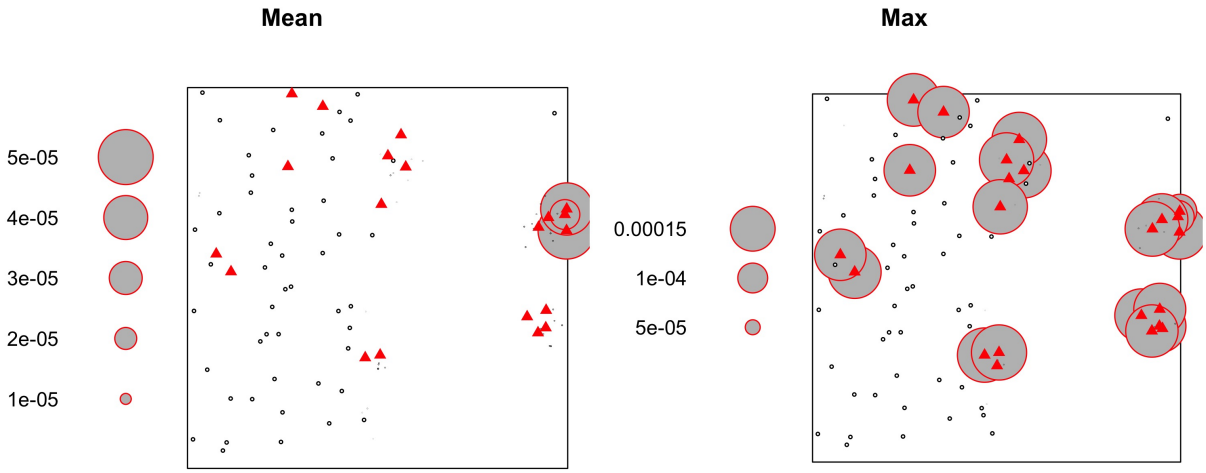


Figure 4.16: The mixed pattern is defined in Sec. 4.3.3. The radii of the discs are proportional to $\tau^{\text{mean}}_{\text{par},i}$ defined in (4.19) for the left and $\tau^{\text{max}}_{\text{par},i}$ in (4.20) for the right. These quantities are the rates of change at $\zeta = 0$ of $\text{Hair}^{\text{mean}}_{\text{par},i}(\zeta)$ from (4.10) and $\text{Hair}^{\text{max}}_{\text{par},i}(\zeta)$ from (4.11), respectively. The discrepancy in (4.9) is applied here.

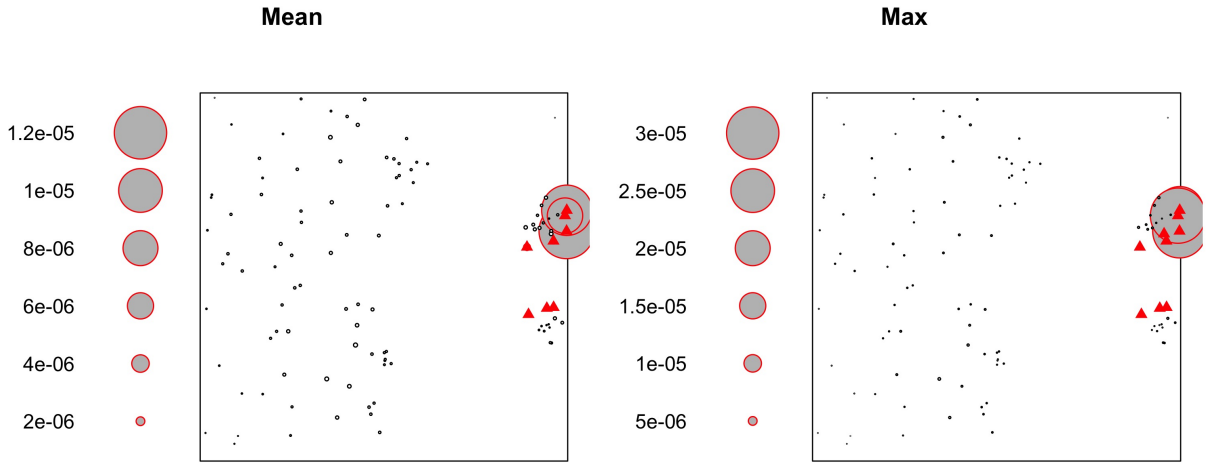
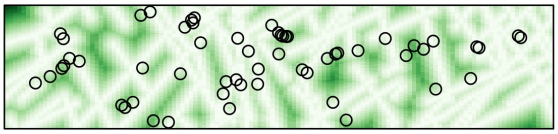
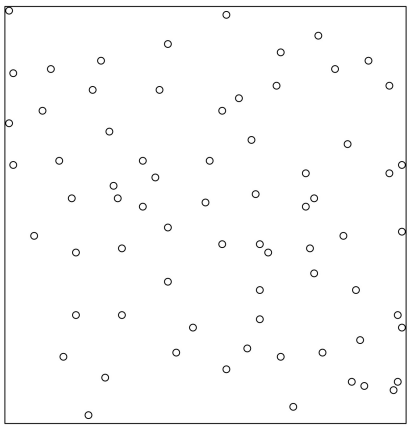


Figure 4.17: The mixed pattern is defined in Sec. 4.3.3. The radii of the discs are proportional to $\tau^{\text{mean}}_{\text{non},i}$ defined in (4.19) for the left and $\tau^{\text{max}}_{\text{non},i}$ in (4.20) for the right. These quantities are the rates of change at $\zeta = 0$ of $\text{Hair}^{\text{mean}}_{\text{non},i}(\zeta)$ from (4.15) and $\text{Hair}^{\text{max}}_{\text{non},i}(\zeta)$ from (4.16), respectively. The discrepancy in (4.14) is applied here.



(a) Copper ores



(b) Swedish pines

Figure 4.18: Data applications in Sec. 4.4.1 and 4.4.2

4.4.1 Copper Ores

Figure 4.18a shows the southern half of the Queensland copper data introduced and analysed by Berman [1986] and again studied by Baddeley et al. [2013]. The dataset consists of 57 copper ore deposits, represented by small black circles, and 90 line segments, shown as white lines, representing geological lineaments such as faults from the geological survey of a 35×158 km region in central Queensland, Australia. The copper ores and lineaments are shown on the distance map. The further the distance of a location to its nearest lineament, the darker (greener) the area. In the initial works, it was of interest to predict the occurrence of copper ores from the lineament pattern. The authors employed an often used model: a loglinear model fitted to 57 locations of copper ore deposits. The covariate of the model is the distance in kilometres from an event to the nearest lineament. The intensity function is $\lambda(x, y) = \exp\{\beta_0 + \beta_1 d(x, y)\}$, where $d(x, y)$ is the distance in kilometres from the location (event) (x, y) to the nearest lineament. I applied this model and obtained the fitted slope $\hat{\beta}_1 = 0.055$ km^{-1} with the asymptotic standard error 0.089 and the 95% confidence interval $[-0.121, 0.227]$. I concluded that there is no evidence of dependence on the lineament. To test for the GOF, I applied the Monte-Carlo AGOF-K test [Dao and Genton, 2014]. With the estimated p-value, $\hat{p} = 0.11$, and the adjusted level, $\hat{\alpha}^* = 0.06$, at the significance level

$\alpha = 0.05$, I do not reject the hypothesis that the model can provide a good fit.

4.4.1.1 Hair-plots and Disc-plots

In the first row of Figure 4.19, the hair-functions are plotted. For the left, the definitions of $\text{Hair}_i^{\text{mean}}(\zeta)$ are in (4.4), for the right $\text{Hair}_i^{\text{max}}(\zeta)$ in (4.5) based on the discrepancy in (4.3). Here, A_i^{mean} and A_i^{max} produce different sets of influential events. However, the joint set is the set of influential events, 40 and 41, resulting from A_i^{mean}.

Studying influence on the basis of the parametric discrepancy is not applicable because $\tilde{K}_{\hat{\boldsymbol{\theta}}_{i,j,m}}$ is numerically equal to $\tilde{K}_{\hat{\boldsymbol{\theta}}}$ in (4.9). Here, $\tilde{K}$ is an approximation to the theoretical K-function of a Gibbs process. Hence I go on to studying influence on the basis of the nonparametric discrepancy in (4.14).Without any model assumption, in the first row of Figure 4.20, the hair-functions are plotted. For the left, the definitions of $\text{Hair}_{\text{non},i}^{\text{mean}}(\zeta)$ are in (4.15), for the right $\text{Hair}_{\text{non},i}^{\text{max}}(\zeta)$ in (4.16) based on the discrepancy in (4.14). While $A_{\text{non},i}^{\text{mean}}$ produces many influential events, $A_{\text{non},i}^{\text{max}}$ has a single influential event, 40, which is also the joint influential event with the mean approach.

The discrepancies in (4.3) and (4.14) produce similar sets of influential events. Mostly, they overlap 40 and 41. Altogether, one can say, that at least these two events can be considered to be two influential events for this dataset according to A_i^{mean}, A_i^{max}, $A_{\text{non},i}^{\text{mean}}$ and $A_{\text{non},i}^{\text{max}}$

4.4.1.2 Local Influences and Disc-plots

In Figure 4.21, on the left, local influence according to τ_i^{mean} from (4.19) and on the right, local influence according to τ_i^{max} from (4.20) are displayed. τ_i^{mean} and τ_i^{max} produce different sets of influential events but overlap two events 40 and 41. The radii of the disc-plots correspond to τ_i^{mean} and τ_i^{max}. In the second row, the local influential events in red triangles and the regular events are displayed on the distance map.

As mentioned above studying influence on the basis of the parametric discrepancy in (4.9) is not applicable because $\tilde{K}_{\hat{\boldsymbol{\theta}}_{i,j,m}}$ is numerically equal to $\tilde{K}_{\hat{\boldsymbol{\theta}}}$ in (4.9).

In Figure 4.22, on the left, local influence according to $\tau_{\text{non},i}^{\text{mean}}$ from (4.19) and on the right, local influence according to $\tau_{\text{non},i}^{\text{max}}$ from (4.20) are displayed. Here, $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\text{max}}$ produce very similar

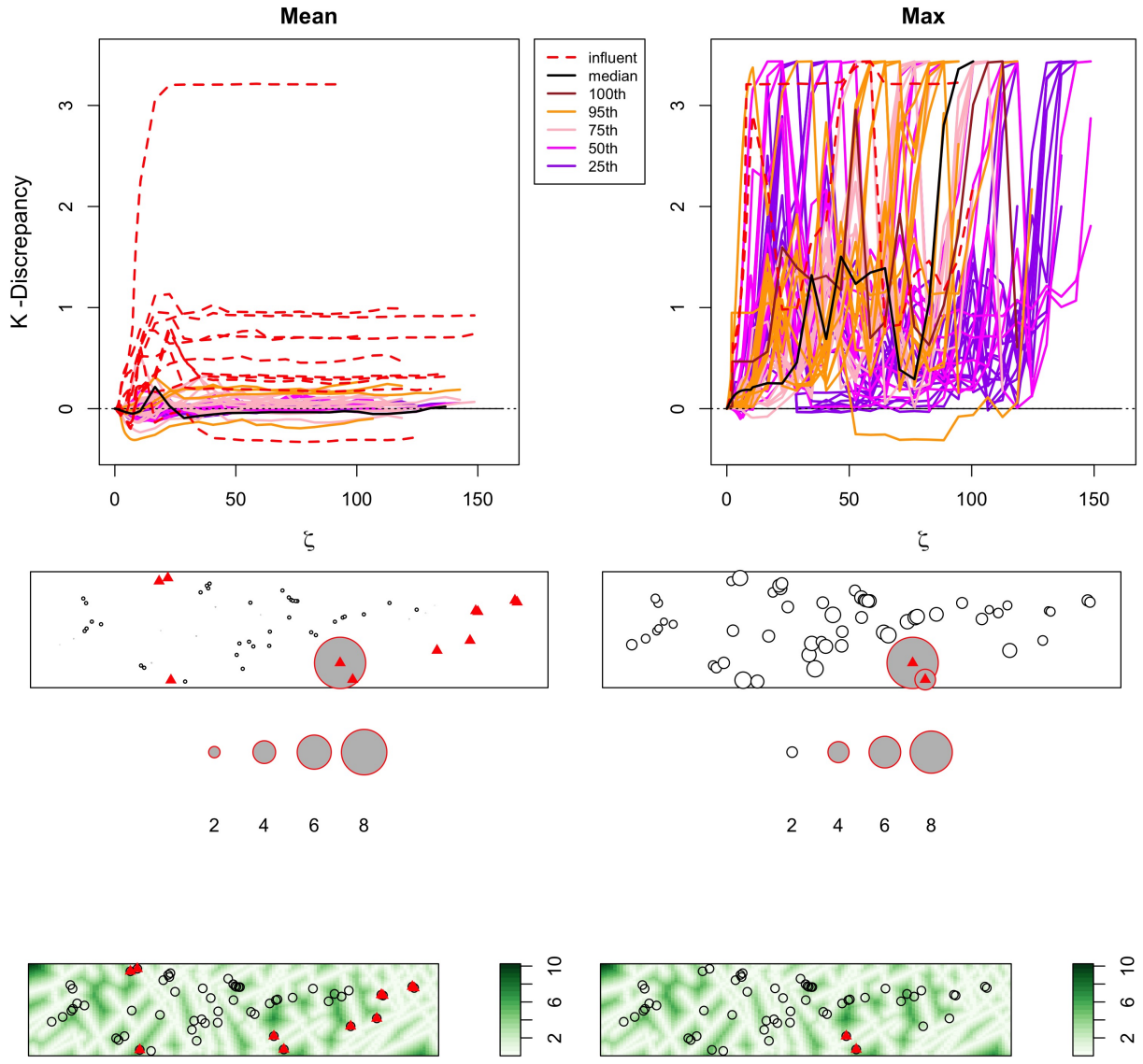


Figure 4.19: The copper ores dataset is described in Sec. 4.4.1. The discrepancy is from (4.3). In the first row, the hair-functions, $\text{Hair}_i^{\text{mean}}(\zeta)$, on the left, and $\text{Hair}_i^{\text{max}}(\zeta)$, on the right are plotted. The colors of the hair-functions are explained in Table 4.1 and in the legend. In the second row, the radii of the discs are proportional to τ_i^{mean} and τ_i^{max}. In the third row, the positions of the influential events are marked red on the distance map of the copper ores to their nearest faults.

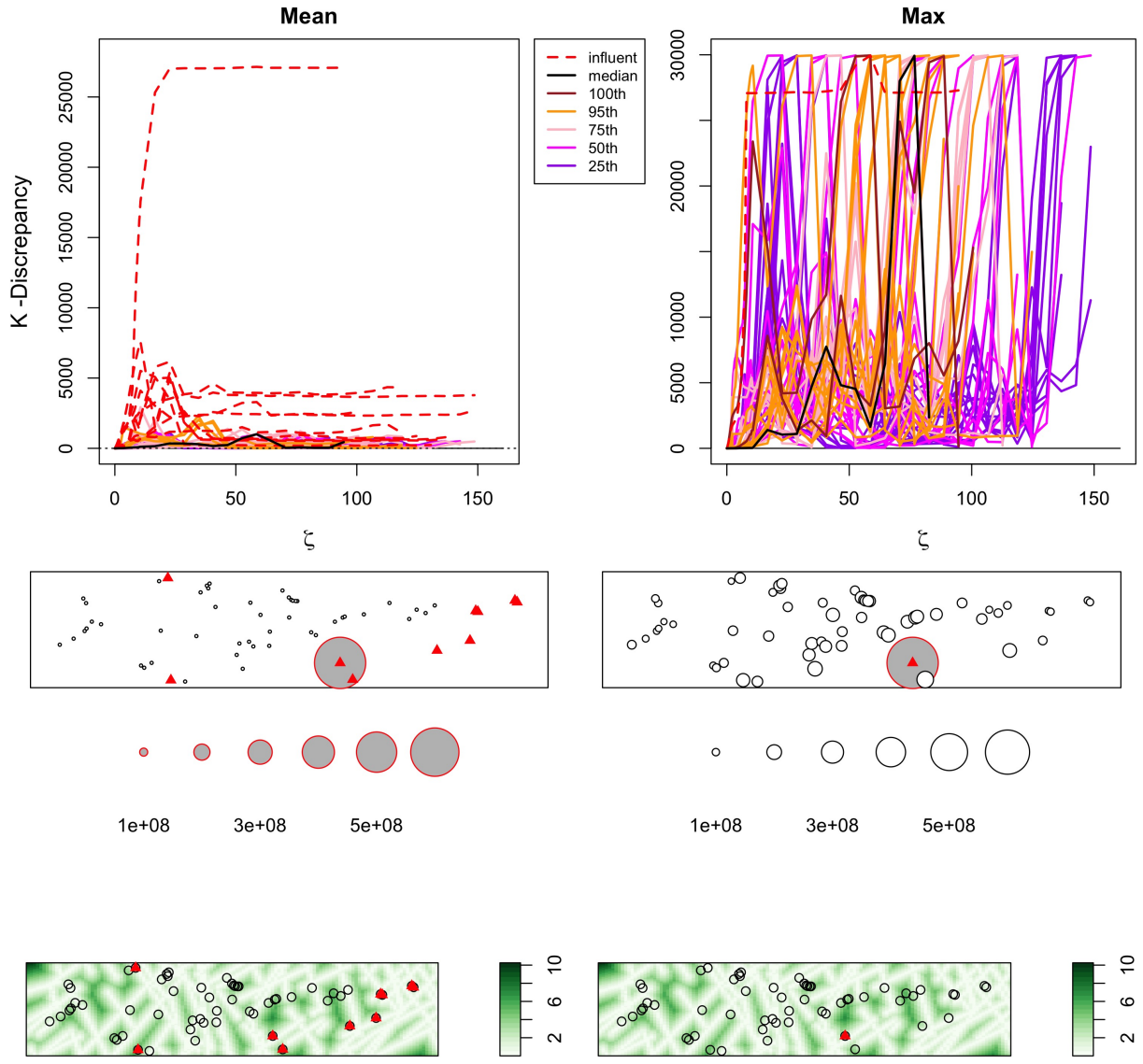


Figure 4.20: The copper ores dataset is described in Sec. 4.4.1. The copper ores dataset is described in Sec. 4.4.1. The discrepancy is from (4.14). In the first row, the hair-functions, $\mathrm{Hair}^{\mathrm{mean}}_{\mathrm{non},i}(\zeta)$, on the left, and $\mathrm{Hair}^{\mathrm{max}}_{\mathrm{non},i}(\zeta)$, on the right, are plotted. The colors of the hair-functions are explained in Table 4.1 and in the legend. In the second row, the radii of the discs are proportional to $\tau^{\mathrm{mean}}_{\mathrm{non},i}$ and $\tau^{\mathrm{max}}_{\mathrm{non},i}$. In the third row, the positions of the influential events are marked red on the distance map of the copper ores to their nearest faults.

Fig. 4.19		Fig. 4.20		Fig. 4.21		Fig. 4.22	
A_i^{mean}	A_i^{max}	$A_{\text{non},i}^{\text{mean}}$	$A_{\text{non},i}^{\text{max}}$	τ_i^{mean}	τ_i^{max}	$\tau_{\text{non},i}^{\text{mean}}$	$\tau_{\text{non},i}^{\text{max}}$
				1			
				3			
					11		
12		12					
					14		
					15		
16							
17		17					
					18		
				26			
40	**40**	**40**	**40**	**40**	**40**	**40**	**40**
41	**41**	41		**41**	**41**	**41**	**41**
					45	**45**	**45**
					47	**47**	**47**
				48		**48**	**48**
				49		49	
					50		
				51			
52	52				52		
53	53						
54	54						
55	55				55		
56	56						
57	57						

Table 4.5: Identifying influential events of the copper ores dataset. The overlapped influential events are displayed in bold numbers. They are from the mean and the maximum approaches based on a specific discrepancy in either (4.3) for A_i^{mean}, A_i^{max}, τ_i^{mean} and τ_i^{max}, or (4.14) for $A_{\text{non},i}^{\text{mean}}$, $A_{\text{non},i}^{\text{max}}$, $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\text{max}}$. The underlined events, 12 and 49, are also detected by Baddeley et al. [2013].

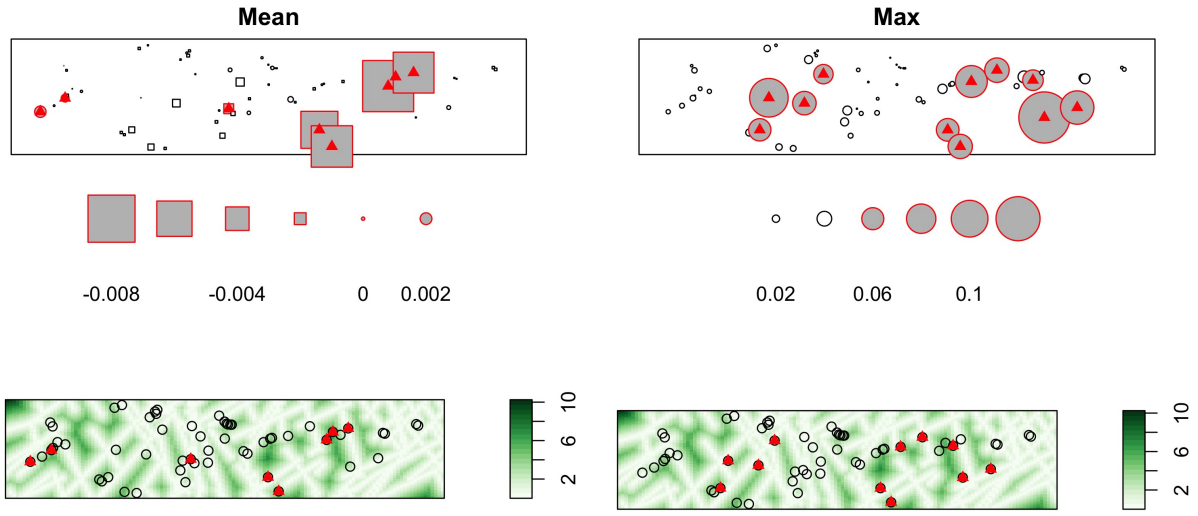


Figure 4.21: The copper ores dataset is described in Sec. 4.4.1. The radii of the discs are proportional to τ_i^{mean} defined in (4.19) for the left and τ_i^{max} in (4.20) for the right. These quantities are the rates of change at $\zeta = 0$ of $\text{Hair}_i^{\text{mean}}(\zeta)$ from (4.4) and $\text{Hair}_i^{\text{max}}(\zeta)$ from (4.5), respectively. The discrepancy in (4.3) is applied here. In the second row, events are displayed on the distance map. The influential events are marked red.

sets of influential events overlapping five events 40, 41, 45, 47, 48. Additionally, $\tau_{\text{non},i}^{\text{mean}}$ produces 49 as another influential event. The radii of the disc-plots correspond to $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\text{max}}$. In the second row, the local influential events are in red triangles and the regular events are displayed on the distance map.

The discrepancies in (4.3) and (4.14) produce similar sets of influential events. Mostly, they overlap 40 and 41. Altogether, one can say that at least these two events can be considered to be two influential events for this dataset according A_i^{mean}, A_i^{max}, $A_{\text{non},i}^{\text{mean}}$ and $A_{\text{non},i}^{\text{max}}$. Possibly, these four events, 40, 41, 45 and 47, can be considered to be influential events for this dataset according to A_i^{max}, $A_{\text{non},i}^{\text{mean}}$ and $A_{\text{non},i}^{\text{max}}$.

The influential points resulting from Baddeley et al. [2013, Figure 3] appeared to be located at the darkest (greenest) area on the distance map, or in other words, they are furthest from their nearest lineament. The presence of these events tends to raise and their absence tends to lower the estimate of β_1. Baddeley et al. [2013, Figure 3] show that the two largest values of influence occur at copper deposit, 12 and 49, with largest values of the distance to the nearest lineament. These

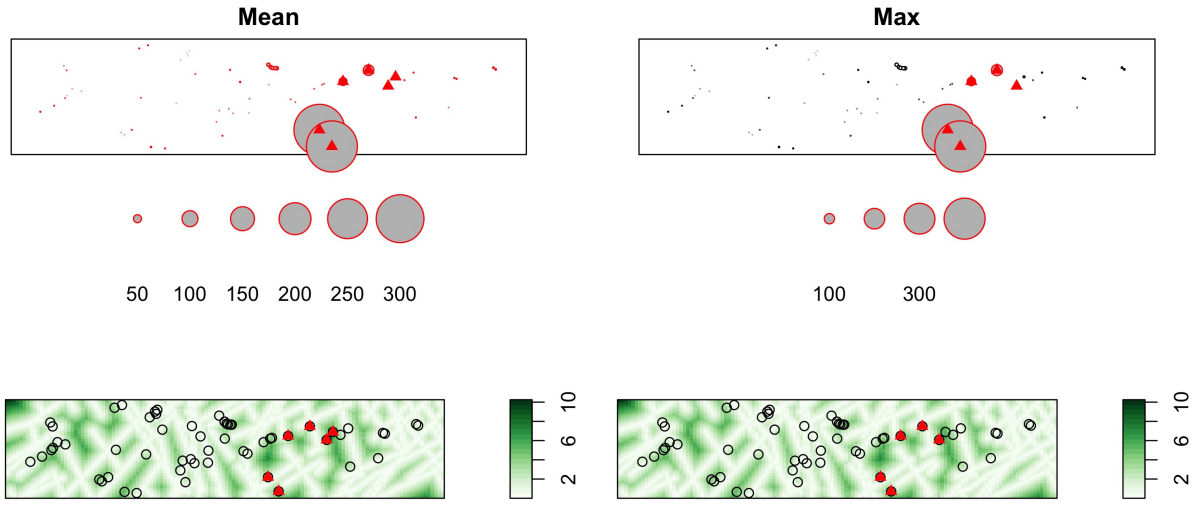


Figure 4.22: The copper ores dataset is described in Sec. 4.4.1. The radii of the discs are proportional to $\tau^{\text{mean}}_{\text{non},i}$ defined in (4.19) for the left and $\tau^{\text{max}}_{\text{non},i}$ in (4.20) for the right. These quantities are the rates of change at $\zeta = 0$ of $\text{Hair}^{\text{mean}}_{\text{non},i}(\zeta)$ from (4.15) and $\text{Hair}^{\text{max}}_{\text{non},i}(\zeta)$ from (4.16), respectively. The discrepancy in (4.14) is applied here. In the second row, events are displayed on the distance map. The influential events are marked red.

two values are located in the peripheral parts of the dataset and hence seem to suffer from edge effect.

My influence method also detected the event 12 and my local influence method also chose the event 49 as influential events among other influential events, see Table 4.5. Beyond these two influential events, I also found many more influential events. In principle, the influential events from my method tend to be located in the somewhat green areas and at the edge of the dark green areas on the distance maps, or in other words, their distance to the lineament is not furthest. It means, however, that their perturbation is relevant to the statistical model. If the influential event is re-located (perturbed) into the green area, the estimate of β_1 can be increased or into the lighter (whiter) area, the estimate of β_1 can be decreased. Here, the lighter areas present fairly short distance of the location to the nearest lineament. Influential events according to my method are not affected by the edge effect because the conclusion is drawn from the findings of many perturbations.

Influential events from my method do not only measure the effect on the estimate of β_1 but also

measure how sensitive this event is for the model, or in different words, how easy it is to affect the model structure by perturbing this event.

4.4.2 Swedish Pines

Another example shown in Figure 4.18b is the dataset called Swedish pines of Strand [1972]. It gives the locations of 71 pine saplings in a 9.6×10 metre survey quadrat. This dataset was often studied, recently also by Baddeley et al. [2019]. Here, the authors described diagnostic quantities analogous to the classical regression's diagnostics of leverage and influence. Since the pattern appears to be spatially inhomogeneous, the authors used Poisson point process models for simplicity to demonstrate their implementation of diagnostic quantities. They fitted a Poisson process to the dataset in which the intensity $\lambda(x, y)$ at spatial location (x, y) is a log-quadratic function of the Cartesian coordinates:

$$\lambda_{\boldsymbol{\theta}}(x, y) \quad = \quad \exp(\theta_0 + \theta_1 x + \theta_2 y + \theta_3 x^2 + \theta_4 xy + \theta_5 y^2).$$

Baddeley et al. [2019, Figure 4] show that large influential values occur at some data events near the corners of the survey rectangle. They showed that the fitted model is highly sensitive to the observed data in the corners of the survey region.

To verify that the Poisson point process suggested by Baddeley et al. [2019] provides a good fit to the data, I again applied the Monte Carlo AGOF-K test. For $\hat{p} = 0.13$, the estimated p-value, and $\hat{\alpha}^* = 0.08$ at the significance level $\alpha = 0.05$, I fail to reject the hypothesis that the model provides a good fit to the dataset. These influential quantities were computed based on the discrepancies from (4.3) and in (4.14). The discrepancy from (4.9) is not applicable because $\tilde{K}_{\hat{\boldsymbol{\theta}}_{i,j,m}}$ and $\tilde{K}_{\hat{\boldsymbol{\theta}}}$ are numerically very close, where $\tilde{K}$ is the approximating K-function, hence the discrepancy from (4.9) will be 0 numerically. The hair-functions and disc-plots are visualized in Figures 4.23 and in Figure 4.24, respectively. The disc-plots according to the local influence τ_i^{mean} in (4.19) and $\tau_i^{\max}$ in (4.20) derived from the hair-functions in (4.4), (4.5) and the discrepancy in (4.3) are shown in Figure 4.25 and from the hair-functions in (4.15), (4.16) and the discrepancy in (4.14) are shown in

Figure 4.26. While some of the influential or local influential events do appear at the edge, some also appear in the inner area of the survey region.

4.4.2.1 Hair-plots and Disc-plots

In the first row of Figure 4.23, the hair-functions are plotted. For the left, the definitions of $\text{Hair}_i^{\text{mean}}(\zeta)$ are in (4.4), for the right $\text{Hair}_i^{\text{max}}(\zeta)$ in (4.5) based on the discrepancy in (4.3). Here, A_i^{mean} and A_i^{max} produce different sets of influential events. However, there is a joint influential event, 63, at the right upper corner. In the second row, the radii of the discs are proportional to the A_i^{mean} and A_i^{max}.

Like in Section 4.4.1.1, studying influence on the basis of the parametric discrepancy is not applicable because $\tilde{K}_{\hat{\boldsymbol{\theta}}_{i,j,m}}$ is numerically equal to $\tilde{K}_{\hat{\boldsymbol{\theta}}}$ in (4.9). Hence I go on to studying influence on the basis of the nonparametric discrepancy in (4.14), i.e., without any model assumption. In the first row of Figure 4.24, the hair-functions are plotted. For the left, the definitions of $\text{Hair}_{\text{non},i}^{\text{mean}}(\zeta)$ are in (4.15), for the right $\text{Hair}_{\text{non},i}^{\text{max}}(\zeta)$ in (4.16) based on the discrepancy in (4.14). Here, A_i^{mean} and A_i^{max} produce a large joint set of six influential events. They are 1, 9, 16, 21, 25, 56. This is also the set of influential events produced by A_i^{max}. In addition to this set, A_i^{mean} also yields three more influential events: 13, 35, 58. In the second row, the radii of the discs are proportional to the A_i^{mean} and A_i^{max}.

4.4.2.2 Local Influences and Disc-plots

In Figure 4.25, on the left, local influence according to τ_i^{mean} from (4.19) and on the right, local influence according to τ_i^{max} from (4.20) are displayed via disc-plots. Here, τ_i^{mean} and τ_i^{max} produce different sets of influential events but overlap one event, 11. The radii of the disc-plots correspond to τ_i^{mean} and τ_i^{max}.

In Figure 4.26, on the left, local influence according to $\tau_{\text{non},i}^{\text{mean}}$ from (4.19) and on the right, local influence according to $\tau_{\text{non},i}^{\text{max}}$ from (4.20) are displayed. $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\text{max}}$ produce very similar sets of influential events overlapping 5 events 9, 11, 16, 28, 35. Additionally, $\tau_{\text{non},i}^{\text{mean}}$ produces 21 and $\tau_{\text{non},i}^{\text{max}}$ produces 13 as another influential event. The radii of the disc-plots correspond to $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\text{max}}$.

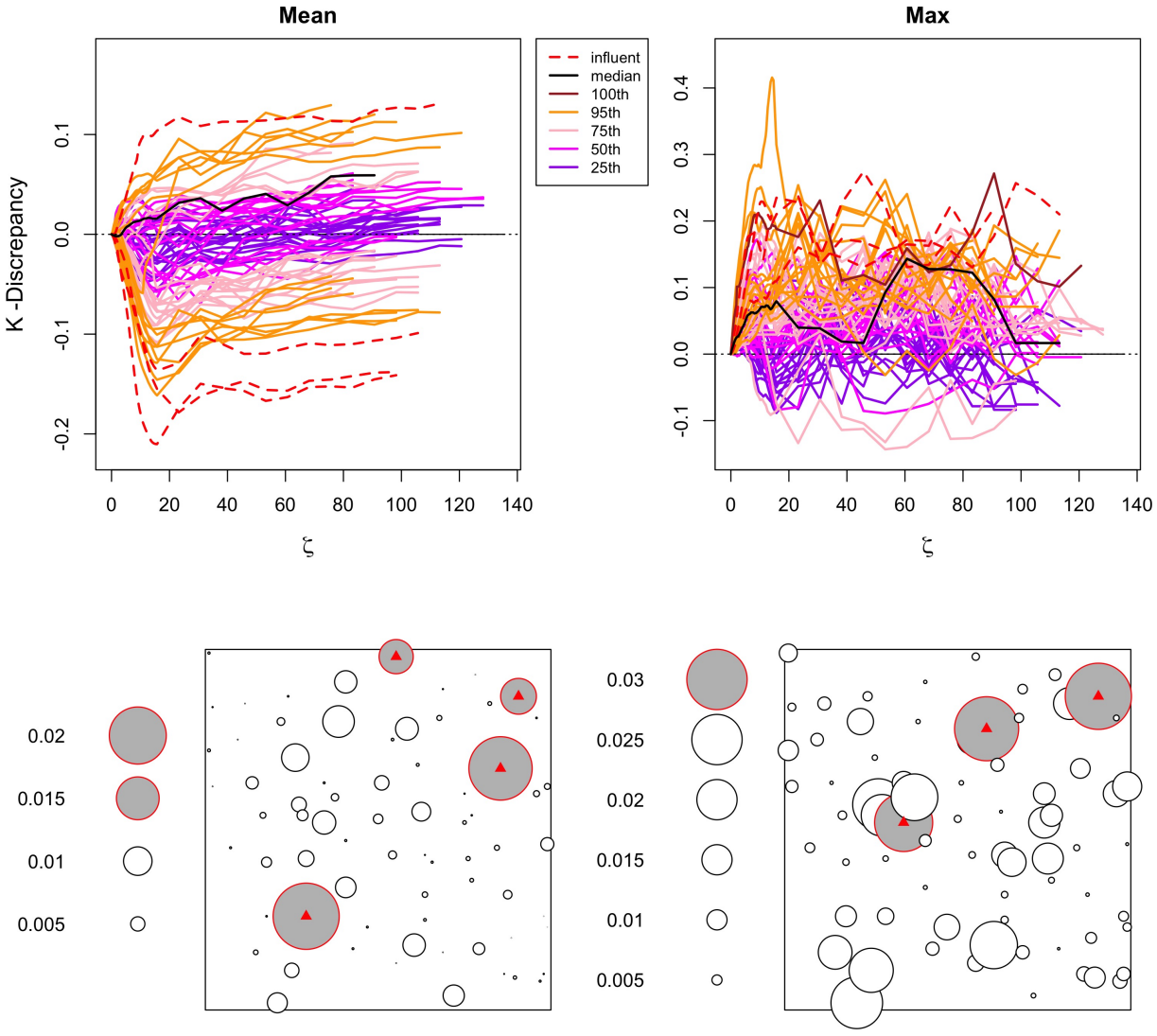


Figure 4.23: The Swedish pines dataset is described in Sec. 4.4.2. The discrepancies are defined in (4.4) on the left and in (4.5) on the right. The colors of the hair-functions are explained in Table 4.1 and in the legend. In the second row, influential events are displayed in red triangles. The radii of the discs are proportional to A_i^{mean} from (4.7) and $A_i^{\max}$.

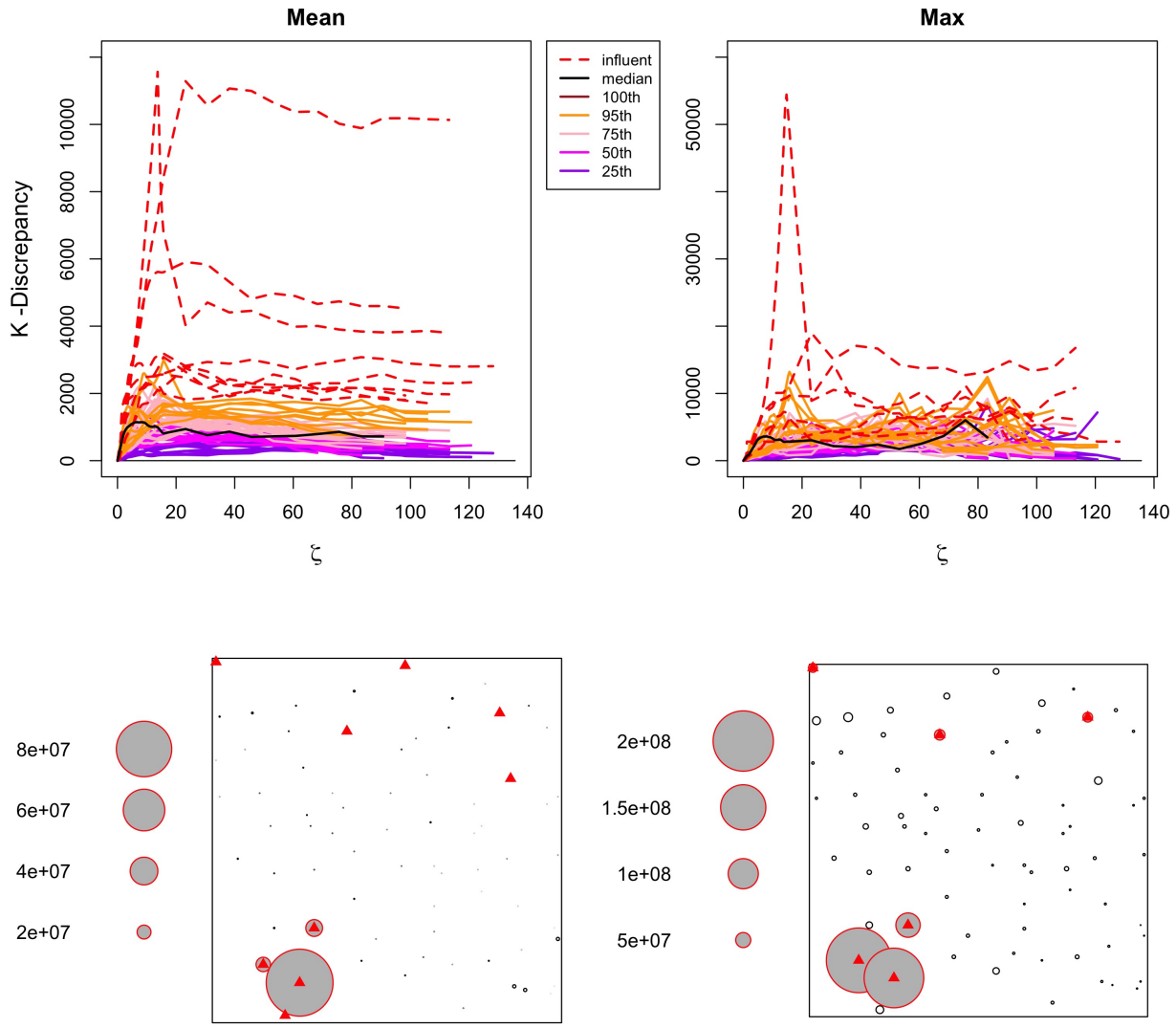


Figure 4.24: The Swedish pines dataset is described in Sec. 4.4.2. The discrepancies are defined in (4.15) for the left and in (4.16) for the right. The colors of the hair-functions are explained in Table 4.1 and in the legend. In the second row, influential events are displayed in red triangles. The radii of the discs are proportional to $A^{\text{mean}}_{\text{non},i}$ and $A^{\text{max}}_{\text{non},i}$.

Fig. 4.23		Fig. 4.24		Fig. 4.25		Fig. 4.26	
A_i^{mean}	$A_i^{\max}$	$A_{\text{non},i}^{\text{mean}}$	$A_{\text{non},i}^{\max}$	τ_i^{mean}	$\tau_i^{\max}$	$\tau_{\text{non},i}^{\text{mean}}$	$\tau_{\text{non},i}^{\max}$
		1	**1**				
		9	**9**			**9**	**9**
				11	**11**	**11**	**11**
		13			13		13
		16	**16**		16	**16**	**16**
21		**21**	**21**			21	
	22						
		25	**25**				
						28	**28**
35		35				**35**	**35**
	37			37			
		56	**56**	56			
58		58					
63	**63**						

Table 4.6: Identifying influential events of the Swedish pines dataset. The overlapped influential events are displayed in bold numbers. They are from the mean and the maximum approaches based on a specific discrepancy in either (4.3) for A_i^{mean}, $A_i^{\max}$, τ_i^{mean} and $\tau_i^{\max}$, or (4.14) for $A_{\text{non},i}^{\text{mean}}$, $A_{\text{non},i}^{\max}$, $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\max}$.

In the second row, the local influential events in red triangles and the regular events are displayed on the distance map.

As Baddeley et al. [2019, Sec. 2.2.2] also studied this dataset, they found almost all peripheral events have larger or much larger influence values. I'd say that this signals some edge effect. The influential and local influential events from my method are not usually peripheral but also inside the spatial point pattern. They do not appear to suffer from edge effect. Since my method draws conclusion from a range of perturbations, it can avoid edge effect.

The meaning of influence according to Baddeley et al. [2019, Sec. 2.2.2] is to quantify the magnitude the influential events have on the parameter estimation. The meaning of my method is, however, to identify the ones which can, via perturbation, break down the model structure most effectively.

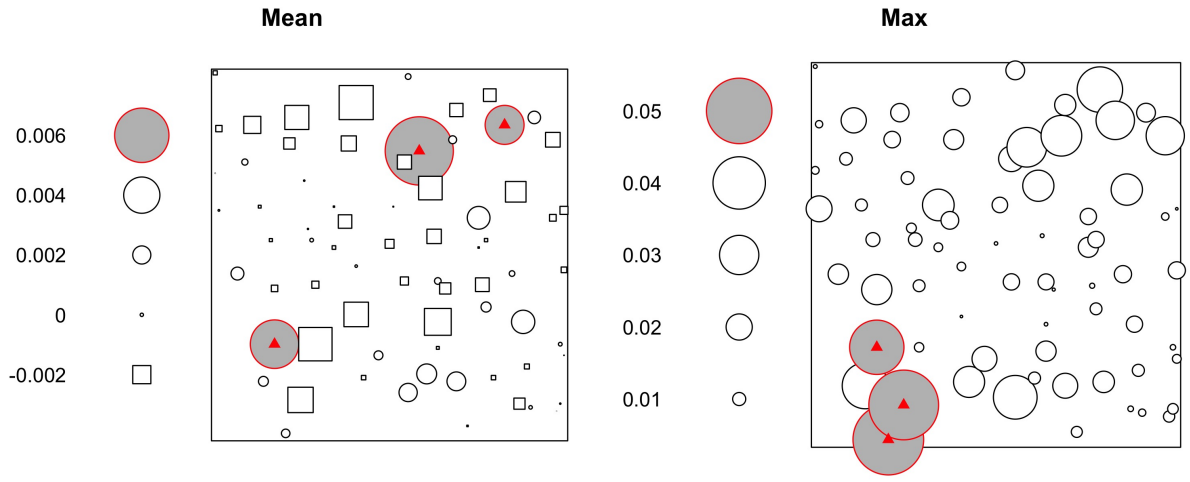


Figure 4.25: The Swedish pines dataset is described in Sec. 4.4.2. The radii of the discs are proportional to τ_i^{mean} defined in (4.19) for the left and τ_i^{max} in (4.20) for the right. These quantities are the rates of change at $\zeta = 0$ of $\text{Hair}_i^{\text{mean}}(\zeta)$ from (4.4) and $\text{Hair}_i^{\text{max}}(\zeta)$ from (4.5), respectively. The discrepancy in (4.3) is applied here. Here, τ_i^{mean} and τ_i^{max} have a joint influential event.

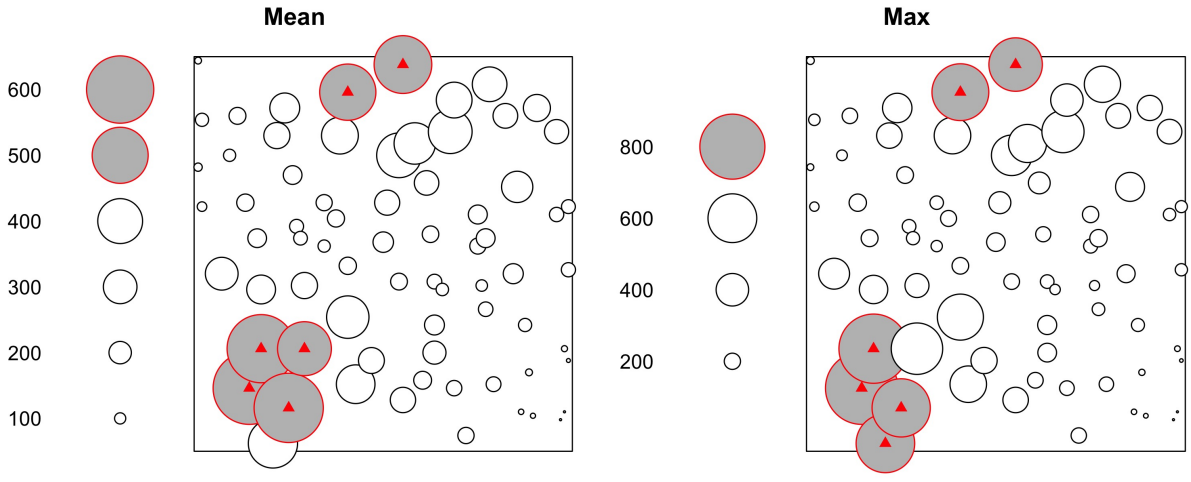


Figure 4.26: The Swedish pines dataset is described in Sec. 4.4.2. The radii of the discs are proportional to $\tau_{\text{non},i}^{\text{mean}}$ defined in (4.19) for the left and $\tau_{\text{non},i}^{\text{max}}$ in (4.20) for the right. These quantities are the rates of change at $\zeta = 0$ of $\text{Hair}_{\text{non},i}^{\text{mean}}(\zeta)$ from (4.15) and $\text{Hair}_{\text{non},i}^{\text{max}}(\zeta)$ from (4.16), respectively. The discrepancy in (4.14) is applied here.

4.5 Discussion

This chapter introduced a method to quantify the magnitude of influence of an event of a SPP. The magnitude of influence means in this context the measure of departure referenced to the observed. The computation of the influence does not necessarily require a maximum, maximum pseudo- or maximum pseudo-profile-likelihood in closed form, in contrast to Baddeley et al. [2013, 2019]. The computation can alternatively use a second-order summary characteristic function. This flexibility allows many more spatial point processes to be studied.

I introduced three discrepancy measures, one in (4.3), a parametric one in (4.9) and a nonparametric one in (4.14). I observe that the latter leads to a more consistent finding of the influential and locally influential events from the mean and maximum approaches: from A_i^{mean} and A_i^{max}, $A_{\text{par},i}^{\text{mean}}$ and $A_{\text{par},i}^{\text{max}}$, $A_{\text{non},i}^{\text{mean}}$ and $A_{\text{non},i}^{\text{max}}$, and respectively, τ_i^{mean} and τ_i^{max}, $\tau_{\text{par},i}^{\text{mean}}$ and $\tau_{\text{par},i}^{\text{max}}$ and $\tau_{\text{non},i}^{\text{mean}}$ and $\tau_{\text{non},i}^{\text{max}}$. The measure of consistence is based on the size of the joint set of the influential events from the mean and maximum approaches. The size of the joint set of influential events seems to be bigger if the nonparametric discrepancy in (4.14) is used.

The more complicated the model is the more influential events my method discovers. The size of influential events increases from studying the CSR SPP in Section 4.3.1, Table 4.2, to studying the Thomas SPP in Section 4.3.2, Table 4.3, to studying the mixed SPP of the previous two SPPs in Section 4.3.3, Table 4.4.

In the following, I will discuss a few future work possibilities. The discrepancies defined in (4.3), (4.9) and (4.14) did not take into account the relevance of certain domain of h on $K(h)$. To integrate this role into the discrepancy one can consider an approach of the tilting technique introduced by Choi et al. [2000]. Versus the discrepancy introduced in (4.3), one can call the below as the *tilted* discrepancy

$$\hat{D}_i(\zeta_j, \gamma_m) \;=\; \int \omega(h) \left[\hat{K}_{i,j,m}(h) - K_{\hat{\boldsymbol{\theta}}}(h)\right]^2 \mathrm{d}h, \tag{4.21}$$

where $\omega(h)$ represents a weighting function. The discrepancy in (4.3) contains $\omega(h) \equiv 1$. Equiva-

lent quantities like the discrepancies in (4.9) and (4.14) can be defined accordingly as in (4.21). In general, a non-constant weighting function allows different weights on values of $K(h)$ at certain range of h which might be more relevant to the question of interest. For example, if h values within $[0, c]$ are more relevant than within $(c, h_{\max}]$, $\omega(h)$ would put more weight on the range $[0, c]$.

A measure of influence can be a discrepancy between the observed and perturbed F-, G-, K- and pair correlation functions as shown in my work but also potentially between the estimates of the perturbed and the observed SPP. In particular, the discrepancy can be defined as $(\hat{\boldsymbol{\theta}}_{i,\gamma,\zeta} - \hat{\boldsymbol{\theta}})^\top (\hat{\boldsymbol{\theta}}_{i,\gamma,\zeta} - \hat{\boldsymbol{\theta}})$, where $\hat{\boldsymbol{\theta}}_{i,\gamma,\zeta}$ is the estimate of the perturbed and $\hat{\boldsymbol{\theta}}$ of the observed SPP.

4.6 List of Definitions

In Table 4.7, a list of many definitions was compiled to assist the readability.

Definitions	Equation number
$\hat{D}_i(\zeta_j, \gamma_m)$	(4.3)
$\hat{D}_{\text{par},i}(\zeta_j, \gamma_m)$	(4.9)
$\hat{D}_{\text{non},i}(\zeta_j, \gamma_m)$	(4.14)
$\text{Hair}_i^{\text{mean}}(\zeta)$	(4.4)
$\text{Hair}_i^{\max}(\zeta)$	(4.5)
$\text{Hair}_{\text{par},i}^{\text{mean}}(\zeta)$	(4.10)
$\text{Hair}_{\text{par},i}^{\max}(\zeta)$	(4.11)
$\text{Hair}_{\text{non},i}^{\text{mean}}(\zeta)$	(4.15)
$\text{Hair}_{\text{non},i}^{\max}(\zeta)$	(4.16)
A_i^{mean}	(4.7)
$A_i^{\max}$	(4.8)
$A_{\text{par},i}^{\text{mean}}$	(4.12)
$A_{\text{par},i}^{\max}$	(4.13)
$A_{\text{non},i}^{\text{mean}}$	(4.17)
$A_{\text{non},i}^{\max}$	(4.18)
τ_i^{mean}	(4.19)
$\tau_i^{\max}$	(4.20)
$\tau_{\text{par},i}^{\text{mean}}$	(4.19)
$\tau_{\text{par},i}^{\max}$	(4.20)
$\tau_{\text{non},i}^{\text{mean}}$	(4.19)
$\tau_{\text{non},i}^{\max}$	(4.20)

Table 4.7: Definitions of the Chapter 4

5. CONCLUSIONS AND FUTURE WORK DIRECTIONS

This thesis focused on inference and visualization for spatial point processes. Research from this thesis has important implications in various aspects.

Chapter 2 presented a Monte Carlo-adjusted goodness-of-fit test for parametric models describing spatial point patterns. This test has been known as the Dao-Genton test [Baddeley et al., 2017, 2020]. Prior to this test, envelope tests in the field of spatial point processes usually tend to be conservative. Using the nested, or otherwise termed as two-stage, Monte Carlo simulation, the Dao-Genton test estimates the adjusted level, α^*, such that $\Pr\{\hat{\mathcal{P}} < \hat{\alpha}^*\} = \alpha$, where Pr denotes probability, $\mathcal{P}$ the p-value, $\hat{\alpha}^*$ an estimate of α^* and α the nominal significance level, usually $\alpha = 0.05$. Normally, α^* is unknown, and $\alpha^* = \alpha$ only in case of complete spatial randomness. The Dao-Genton test allows the GOF test to correctly achieve the pre-specified α. The downside is that it demands high computational burden. This issue can motivate future works to either reduce the computational complexity or to derive the asymptotic distribution of $\hat{\alpha}^*$. For the former possibility, I proposed an interpolation approach to estimate α^* faster and showed via a simulation study that it is beneficial. Another possibility is to apply the sequential Monte Carlo p-values [Besag and Clifford, 1991]. To circumvent the high burden of computation completely, deriving the asymptotic distribution of $\hat{\alpha}^*$ can become an effective solution because it is computationally simple and it helps to make a more accurate conclusion of the GOF. I conjecture that the asymptotic distribution is related to a gamma distribution or a mixture of gamma distributions.

Chapter 3 introduced skew-elliptical cluster processes to model SPPs having non-circular clusters. In particular, I used only two representatives of the unified skew-elliptical distribution [Arellano-Valle and Genton, 2010a] to describe the non-circular clusters. They are the unified skew-normal and extended skew-t distributions. This approach leads to establishing skew-normal, elliptical-normal, skew-t, elliptical-t cluster processes. Prior to the introduction of theses processes, it has not been widely common to model patterns having non-circular clusters and estimation of the parameters has not been easy due to the difficulty to derive the maximum-likelihood, or pseudo-

likelihood. Additionally, the computation can be intensive like in the case of Castelloe [1998] using a Bayesian approach. Although the challenge of deriving the maximum- and/or pseudo-likelihood still remains for the new introduced processes, I circumvented this by using the approximation of a second-order summary characteristic function such as the pair correlation function of these processes and employed the minimum contrast method [Diggle, 2003] to estimate the parameters. The upside of this approach is that this estimation is computationally easy. The downside is that the bias can be large if the initial value is not wisely chosen. Also the approximation of the pair correlation cannot provide a good estimate as an estimator derived from the true pair correlation function. Possible future works are to derive a maximum-, pseudo-, or profile likelihood for theses processes or to derive the true pair correlation function.

In this chapter, I presented CPs with clusters showing either only skew or elliptical shape. From here, one could further explore an opportunity to come up with a CP with clusters of both skew and elliptical shape.

A possible extension of generalizing the TP is to consider enlarging the choice of the distribution that is imposable on the location of a "children" event in a cluster. Besides the SUN and EST classes, there may be other distributions of the unified skew-elliptical families. One of the requirements is the additive property because the distribution of a linear combination of the positions of two "children" events has to be established.

In this chapter, I attempted to generalize the TP to some extent. For a future work, one could apply the same approach to the Matérn point process, the role of which is very similar to that of the TP in the field of spatial point processes. Both are special cases of the Neymann–Scott cluster point process. A Matérn process is constructed similar to a TP except that the positions of the "children" events are distributed independently and uniformly inside a disc with the "parent" event as the center. Similar to this work, it is possible to establish some variations of the Matérn process with respect to the circular, elliptical and skew properties of the distribution of "children" events.

Then in Chapter 4, I introduced various possible definitions of influential events and visualization tools via hair-plots and disc-plots. The approach in this chapter makes possible for any SPP

underlying a likelihood or just a second-order summary characteristic function, if the likelihood is undefined, to identify influential events effectively.

I presented a method where the perturbation is carried out via a range of the angle γ and the radius ζ. In this Chapter, I used equidistant grids for γ and ζ. It is reasonable to expect that the more γ and ζ values are taken into account into the computation of the influence, the more representative or accurate the influence quantities are. This can lead to a high computational complexity. A future work can explore a possibility to determine (i) how many γ and ζ values are needed and/or (ii) what are the relevant ranges of γ and ζ before starting the computation of the influence quantities.

The discrepancies defined in Chapter 4 did not take into account the relevance of certain domain of h on $K(h)$. To integrate this role into the discrepancy I can consider an approach of the tilting technique introduced by Choi et al. [2000], where a weighting function can tilt the discrepancy towards certain chosen domains of h.

Parallel to the definition of influence presented via discrepancies between the observed and perturbed second-order summary characteristic functions, especially the pair correlation function as shown in the chapter, potential definitions of influence in a future work can consider the dot product or scalar product of the discrepancy between the parameter estimates of the perturbed and the observed SPP.

REFERENCES

R. B. Arellano-Valle and A. Azzalini. On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, 33(3):561–574, 2006.

R. B. Arellano-Valle and M. G. Genton. On fundamental skew distributions. *Journal of Multivariate Analysis*, 96:93–116, 2005.

R. B. Arellano-Valle and M. G. Genton. Multivariate unified skew-elliptical distributions. *Chilean Journal of Statistics*, 1:17–33, 4 2010a.

R. B. Arellano-Valle and M. G. Genton. Multivariate extended skew-t distributions and related families. *Metron*, 68(3):201–234, 10 2010b.

B. C. Arnold and R. J. Beaver. Some skewed multivariate distributions. *American Journal of Mathematical and Management Sciences*, 20(1–2):27–38, 2000.

B. C. Arnold and R. J. Beaver. Skewed multivariate models related to hidden truncation and/or selective reporting. *TEST*, 11(1):7–54, 2002.

R. M. Assunção. Testing spatial randomness by means of angles. *Biometrics*, 50(2):531–537, 1994a.

R. M. Assunção. *Robust estimation on Point Processes*. Ph.d. thesis, University of Washington, Seattle, 1994b.

R. M. Assunção and P. Guttorp. Robustness for inhomogeneous Poisson point processes. *Annals of the Institute of Statistical Mathematics*, 51:657–678, 1999.

R. M. Assunção and I. A. Reis. Testing spatial randomness: A comparison between t^2 methods and modifications of the angle test. *Brazilian Journal of Probability and Statistics*, 14:71–86, 2000.

A. Azzalini and A. Capitano. Statistical applications of the multivariate skew normal distributions. *Journal of the Royal Statistical Society B*, 61:579–602, 1999.

A. Azzalini and A. Capitano. *The Skew-normal and Related Families*. Cambridge University Press, 2014.

A. Azzalini and A. Dalla Valle. The multivariate skew-normal distribution. *Biometrika*, 83:715–726, 1996.

A. Baddeley and R. Turner. spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42, 2005a. URL `http://www.jstatsoft.org/v12/i06/`.

A. Baddeley, P. Diggle, A. Hardegen, T. Lawrence, R. Milne, and G. Nair. On tests of spatial pattern based on simulation envelopes. *Ecological Monographs*, 83(3):477–489, 2014.

A. Baddeley, A. Hardegen, T. Lawrence, R. Milne, G. Nair, and S. Rakshit. On two-stage Monte Carlo test of composite hypotheses. *Computational Statistics & Data Analysis*, 114:75–87, 2017.

A. J. Baddeley and G. Nair. Approximating the moments of a spatial point process. *Stat*, 1:18–30, 2012.

A. J. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns. *Australian & New Zealand Journal of Statistics*, 42:283–322, 2000.

A. J. Baddeley and R. Turner. Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12:1–42, 2005b.

A. J. Baddeley, J. Møller, and R. P. Waagepetersen. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54:329–350, 2000.

A. J. Baddeley, R. Turner, J. Møller, and M. Hazelton. Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society B*, 67:617–666, 2005.

A. J. Baddeley, J. Møller, and A. G. Pakes. Properties of residuals for spatial point processes. *Annals of the Institute of Statistical Mathematics*, 60:629–649, 2008.

A. J. Baddeley, Y. Chang, and Y. Song. Leverage and influence diagnostics for spatial point processes. *Scandinavian Journal of Statistics*, 40:86–104, 2013.

A. J. Baddeley, E. Rubak, and R. Turner. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London, 2015. URL `http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/`. In press.

A. J. Baddeley, E. Rubak, and R. Turner. Leverage and influence diagnostics for gibbs spatial point

processes. *Spatial Statistics*, 29:15–48, 2019.

A. J. Baddeley, R. Turner, and E. Rubak. *Package 'spatstat'*. Online, 2020. URL `http://spatstat.org/resources/spatstatManual.pdf`.

G. A. Barnard. Discussion of paper by M. S. Bartlett. *Journal of the Royal Statistical Society B*, 25:294, 1963.

M. S. Bartlett. The spectral analyis of point processes. *Biometrika*, 51:299–311, 1964.

M. J. Bayarri and J. O. Berger. P values in composite null models. *Journal of the American Statistical Association*, 95:1127–1142, 2000.

D. A. Belsley, E. Kuh, and R. E. Welsh. Regression diagnostics: Identifying influential data and sources of collinearity. In *Wiley Series in Probability and Mathematical Statistics*, pages 11–16, New York, 1980. John Wiley & Sons.

M. Berman. Testing for spatial association between a point process and another stochastic process. *Applied Statistics*, 35:54–62, 1986.

M. Berman and T. R. Turner. Approximating point process likelihoods with `glim`. *Applied Statistics*, 41:31–38, 1992.

J. E. Besag. Some methods of statistical analysis of spatial data. *Bulletin of the International Statistical Institute*, 47:77–92, 1978.

J. E. Besag and P. Clifford. Sequential Monte Carlo p-values. *Biometrika*, 78(2):301–304, 1991.

J. E. Besag and P. J. Diggle. Simple monte carlo tests for spatial patterns. *Applied Statistics*, 26: 327–333, 1977.

J. E. Besag and J. T. Gleaves. On the detection of spatial pattern in plant communities. *Bulletin of the International Statistical Institute*, 45:153–158, 1973.

J. Castelloe. *Issues in Reversible Jump Markov Chain Monte Carlo and Composite EM Analysis, Applied to Spatial Poisson Cluster Processes*. PhD thesis, University of Iowa, 1998.

E. Choi, P. Hall, and B. Presnell. Rendering parametric procedures more robust by empirically tilting the model. *Biometrika*, 87:453–465, 2000.

R. D. Cook and S. Weisberg. *An Introduction to Regression Graphics*. Wiley, New York, 1994.

N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, New York, revised edition, 1993.

N. A. Dao and M. G. Genton. A Monte Carlo adjusted goodness-of-fit test for parametric models describing spatial point patterns. *Journal of Computational and Graphical Statistics*, 23:497–517, 2014.

N. A. Dao and M. G. Genton. Skew-elliptical cluster processes. In I. Ghosh, N. Balakrishnan, and H. Ng, editors, *Advances in Statistics - Theory and Applications: Honoring the Contributions of Barry C. Arnold in Statistical Science*, pages 365–393. Springer, New York, 2021.

J. M. Dickey. On a multivariate generalization of the Behrens-Fisher distribution. *The Annals of Mathematical Statistics*, 37:763, 1966.

J. M. Dickey. Three multidimensional-integral identities with Bayesian applications. *The Annals of Mathematical Statistics*, 39(5):1615–1627, 1968.

P. J. Diggle. *Statistical Analysis of Spatial Point Patterns*. Arnold, London, 2nd edition, 2003.

P. J. Diggle and R. J. Gratton. Monte Carlo methods of inference for implicit statistical models (with discussions). *Journal of the Royal Statistical Society B*, 46:193–227, 1984.

K. Donnelly. Simulations to determine the variance and edge-effect of total nearest neighbor distance. In I. Hoddler, editor, *Simulation Methods in Archaeology*, pages 91–95. Cambridge University Press, London, 1978.

A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi. *Higher Transcendental Functions*. McGraw-Hill, New York, 1953.

M. G. Genton and A. Ruiz-Gazen. Visualizing influential observations in dependent data. *Journal of Computational and Graphical Statistics*, 19(4):808–825, 2010.

G. González-Farías, J. A. Domíniguez-Molina, and A. K. Gupta. The closed skew-normal distribution. In M. G. Genton, editor, *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, pages 25–42. Chapman & Hall/CRC, Boca Raton, FL, 2004.

Y. Guan. A goodness-of-fit test for inhomogeneous spatial Poisson processes. *Biometrika*, 95(4): 831–845, 2008.

A. K. Gupta, G. González-Farías, and J. A. Domíniguez-Molina. A multivariate skew normal

distribution. *Journal of Multivariate Analysis*, 89:181–190, 2004.

A. K. Gupta, M. A. Aziz, and W. Ning. On some properties of the unified skew normal distribution. *Journal of Statistical Theory and Practice*, 7(3):480–495, 2013.

W. G. S. Hines and R. J. O. Hines. The Eberhardt index and the detection of non-randomness of spatial point distributions. *Biometrika*, 66:73–80, 1979.

F. Huang and Y. Ogata. Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *Journal of Computational and Graphical Statistics*, 8:510–530, 1999.

J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*. Statistics in Practice. Wiley, 2008.

J. B. Illian, J. Møller, and R. P. Waagepetersen. Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics*, 16(3):389–405, 2009.

Y. Kutoyants. *Statistical Inference for Spatial Poisson Processes*. No. 134 in Lecture Notes in Statistics. Springer, New York, 1998.

B. Liseo and N. Loperfido. A Bayesian interpretation of the multivariate skew-normal distribution. *Statistics and Probability Letters*, 61:395–401, 2003.

F. H. C. Marriott. Monte Carlo tests: How many simulations. *Applied Statistics*, 28:75–77, 1979.

D. A. McQuarrie. *Statistical Mechanics*. Harper & Row, New York, 1976.

J. Møller and H. Toftaker. Geometric anisotropic spatial point pattern analysis and cox processes. *Scandinavian Journal of Statistics*, 41:414–435, 2014.

J. Møller and R. P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC Press, 2003.

Y. Neyman and E. L. Scott. A theory for the spatial distribution of galaxies. *Astrophysics Journal*, 116:144–163, 1952.

Y. Neyman and E. L. Scott. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society B*, 20:1–43, 1958.

Y. Neyman and E. L. Scott. Processes of clustering and applications. In P. A. W. Lewis, editor,

Stochastic Processes, pages 646–681. J. Wiley & Sons, Inc., New York, 1972.

Y. Neyman, E. L. Scott, and C. D. Shane. On the spatial distribution of galaxies: A specific model. *Astrophysics Journal*, 117:92–133, 1953.

J. Ohser. On estimators for the reduced second moment measure of point processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 14:63–71, 1983.

A. Penttinen, D. Stoyan, and H. Henttonen. Marked point processes in forest statistics. *Forest Science*, 38:806–824, 1992.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL `https://www.R-project.org/`.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL `https://www.R-project.org/`.

S. L. Rathbun and N. Cressie. Asymptotic properties of estimators of the parameters of spatial inhomogeneous Poisson point processes. *Advances in Applied Probability*, 26(1):122–154, 1994.

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

B. D. Ripley. Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society B*, 39:172–212, 1977.

B. D. Ripley. *Statistical Inference for Spatial Processes*. Cambridge University Press, Australia, 1988.

J. M. Robins, A. Van der Vaart, and V. Ventura. Asymptotic distribution of p values in composite null models. *Journal of the American Statistical Association*, 95(452):1143–1156, 2000.

D. Stoyan and H. Stoyan. Estimating pair correlation functions of planar cluster processes. *Biometrical Journal*, 38:259–271, 2006.

L. Strand. A model for stand growth. In *IUFRO Third Conference Advisory Group of Forest Statisticians*, pages 207–216. Institut National de la Recherche Agronomique, Paris, 1972.

D. J. Strauss. A model for clustering. *Biometrika*, 63:467–475, 1975.

U. Tanaka, Y. Ogata, and D. Stoyan. A model selection and estimation of the Neyman-Scott type

spatial cluster models. *Biometrical Journal*, 50:43–57, 2008.

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

I. Wolfram Research. *Mathematica Version 12.1*. Wolfram Research, Inc., Champaign, IL, USA, 2020. URL `https://www.wolfram.com/mathematica/`.

APPENDIX A

CHAPTER 2: A MONTE CARLO-ADJUSTED GOODNESS-OF-FIT TEST FOR PARAMETRIC MODELS DESCRIBING SPATIAL POINT PATTERNS

A.1 PROOF OF PROPOSITION 1

Let P_i and $\hat{P}_i$ denote the random variables corresponding to p_i in (9) and $\hat{p}_i$ in (10), $i = 1, \ldots, n$, respectively. Since $\hat{\alpha}^*$ is the solution of the estimating equation (11), I have $\alpha n = E\left\{\sum_{i=1}^{n} I(\hat{p}_i < \hat{\alpha}^*)\right\} = \sum_{i=1}^{n} E\left\{I(\hat{p}_i < \hat{\alpha}^*)\right\} = \sum_{i=1}^{n} \text{Pr}(\hat{P}_i < \hat{\alpha}^*)$ and since the $\hat{P}_i$'s are independent and identically distributed I have $\sum_{i=1}^{n} \text{Pr}(\hat{P}_i < \hat{\alpha}^*) = n\text{Pr}(\hat{P}_1 < \hat{\alpha}^*)$. Thus $\text{Pr}\{\hat{P}_1 < \hat{\alpha}^*\} = \alpha$. By the strong law of large numbers, $\hat{P}_1 \xrightarrow{a.s.} P_1$. Consequently, I have

$$\text{Pr}\{P_1 < \hat{\alpha}^*\} \;=\; \alpha. \tag{A.1}$$

In the following, I will show that the asymptotic distributions of $\mathcal{P}$ and P_1 are the same.

Let F_X denote the cumulative distribution function of a random variable, X. From (9), I have $p_1 = 1 - F_{U_{\tilde{\theta}_1}}(u_{1,1})$ where $U_{\tilde{\theta}_1}$ is the random variable corresponding to $u_{1,2}, \ldots, u_{1,n}$. According to the Taylor expansion,

$$p_1 \;=\; 1 - F_{U_{\tilde{\theta}_1}}(u_{1,j^*}) - F'_{U_{\tilde{\theta}_1}}(u_{1,j^*}) \underbrace{(u_{1,1} - u_{1,j^*})}_{O(n^{-1})} - F^{(2)}_{U_{\tilde{\theta}_1}}(u_{1,j^*}) \frac{(u_{1,1} - u_{1,j^*})^2}{2} \;-\; \text{Remainder},$$

where u_{1,j^*} is a value of the set $u_{1,2}, \ldots, u_{1,n}$ and is within a neighborhood of $u_{1,1}$. Since u_{1,j^*} is a realization of $U_{\tilde{\theta}_1}$, $F_{U_{\tilde{\theta}_1}}(u_{1,j^*})$ follows the standard uniform distribution. Thus, $F'_{U_{\tilde{\theta}_1}}(u_{1,j^*}) = 1$ and $F^{(k)}_{U_{\tilde{\theta}_1}}(u_{1,j^*}) = 0$ for $k \geq 2$. That is,

$$p_1 \;=\; 1 - F_{U_{\tilde{\theta}_1}}(u_{1,j^*}) - O(n^{-1}). \tag{A.2}$$

From (4), $p = 1 - F_{U_{\hat{\theta}}}(u)$, where $U_{\hat{\theta}}$ is the random variable corresponding to $u_1, \ldots, u_n$, I can

decompose p as a realization of $\mathcal{P}$ as follows

$$p \;=\; 1 - F_{U_{\hat{\theta}}}(u_{i^*}) - F'_{U_{\hat{\theta}}}(u_{i^*}) \underbrace{(u - u_{i^*})}_{O(n^{-1})} - F^{(2)}_{U_{\hat{\theta}}}(u_{i^*}) \frac{(u - u_{i^*})^2}{2} \;-\; \text{Remainder},$$

where u_{i^*} is a value of the set $u_1, \ldots, u_n$ and is in a neighborhood of u. Thus, $F_{U_{\hat{\theta}}}(u_{i^*})$ follows a uniform distribution. Consequently, $F'_{U_{\hat{\theta}}}(u_{i^*}) = 1$, $F^{(k)}_{U_{\hat{\theta}}}(u_{i^*}) = 0$ for $k \geq 2$, and

$$p \;=\; 1 - F_{U_{\hat{\theta}}}(u_{i^*}) - O(n^{-1}). \tag{A.3}$$

From (A.2) and (A.3), the asymptotic distributions of P_1 and $\mathcal{P}$ are the same. By (A.1), $\Pr\{\mathcal{P} < \hat{\alpha}^*\} = \alpha$. It is left to show that $\Pr\{\hat{\mathcal{P}} < \hat{\alpha}^*\} = \alpha$. Now assume either $\Pr\{\hat{\mathcal{P}} < \hat{\alpha}^*\} > \alpha$ or $\Pr\{\hat{\mathcal{P}} < \hat{\alpha}^*\} < \alpha$, which leads to either $\Pr\{\mathcal{P} < \hat{\alpha}^* + (\mathcal{P} - \hat{\mathcal{P}})\} > \alpha$ or $\Pr\{\mathcal{P} < \hat{\alpha}^* + (\mathcal{P} - \hat{\mathcal{P}})\} < \alpha$, i.e. either $\Pr\{\mathcal{P} > \hat{\mathcal{P}}\} = 1$ or $\Pr\{\mathcal{P} < \hat{\mathcal{P}}\} = 1$, which leads to a contradiction with $\hat{\mathcal{P}} \xrightarrow{a.s.} \mathcal{P}$ due to the strong law of large numbers. Hence, $\Pr\{\hat{\mathcal{P}} < \hat{\alpha}^*\} = \alpha$.

A.2 PSEUDO-CODE OF THE MONTE CARLO ADJUSTED GOF TEST

From the only observed SPP, X, I obtain the parameter estimate, $\hat{\theta}$, for the true but unknown θ, and additionally obtain $\hat{G}(h)$, $\hat{F}(h)$, $\hat{K}(h)$, or $\hat{K}^{in}(h)$.

For (i in $1 : n$) {

- Generate a SPP, X_i, with $\hat{\theta}$ as the parameter. Estimate $\tilde{\theta}_i$, compute and store $\hat{G}_i(h)$, $\hat{F}_i(h)$, $\hat{K}_i(h)$, or $\hat{K}^{in}_i(h)$. (In my notation, $\hat{G}_i(h) \equiv \hat{G}_{i,1}(h)$, $\hat{F}_i(h) \equiv \hat{F}_{i,1}(h)$, $\hat{K}_i(h) \equiv \hat{K}_{i,1}(h)$, or $\hat{K}^{in}_i(h) \equiv \hat{K}^{in}_{i,1}(h)$)

- For (j in $2 : n$) {

  – Generate a SPP, $X_{i,j}$, using $\tilde{\theta}_i$ as the parameter. Compute and store $\hat{G}_{i,j}(h)$, $\hat{F}_{i,j}(h)$, $\hat{K}_{i,j}(h)$, or $\hat{K}^{in}_{i,j}(h)$

  } end of the j-loop.

- Use all stored quantities $\hat{G}_{i,j}(h)$, $\hat{F}_{i,j}(h)$, $\hat{K}_{i,j}(h)$, or $\hat{K}^{in}_{i,j}(h)$, $j = 1, \ldots, n$ to compute $\overline{G}_{i,j}(h)$, $\overline{F}_{i,j}(h)$, $\overline{K}_{i,j}(h)$, or $\overline{K}^{*}_{i,j}(h)$, $j = 1, \ldots, n$. Subsequently, compute $u_{i,1}, u_{i,2}, \ldots, u_{i,n}$ and $\hat{p}_i$ in (10) for the hypothesis $H_{i,0}$ in (8).

} end of the i-loop

The three steps of the AGOF test described in Sec. 2.3 are as follows:

1st Step Compute u, $u_1, \ldots, u_n$ from the stored quantities $\hat{G}(h)$, $\hat{F}(h)$, $\hat{K}(h)$, or $\hat{K}^{in}(h)$; $\hat{G}_i(h)$, $\hat{F}_i(h)$, $\hat{K}_i(h)$, or $\hat{K}^{in}_i(h)$, $i = 1, \ldots, n$. Subsequently, compute $\hat{p}$ in (5).

2nd Step Estimate $\hat{\alpha}^*$ from the α-quantile of $\hat{p}_1, \ldots, \hat{p}_n$ to fulfill $\sum_{i=1}^{n} I\{\hat{p}_i < \hat{\alpha}^*\} = \alpha n$.

3rd Step Reject H_0 if $\hat{p} \leq \hat{\alpha}^*$.

APPENDIX B

CHAPTER 3: SKEW-ELLIPTICAL CLUSTER PROCESSES

B.1 SKEW-ELLIPTICAL-NORMAL CLUSTER PROCESSES

According to the transformation in (3.1), the joint distribution $f_{R,\Theta}(r,\theta)$ of (R,Θ) is

$$f_{R,\Theta}(r,\theta) = \frac{r\exp\left(-\frac{\sigma_2^2 r^2\cos^2\theta+\sigma_1^2 r^2\sin^2\theta}{4\sigma_1^2\sigma_2^2}\right)}{\pi\sigma_1\sigma_2} \times \Phi_2\left\{\frac{\left(\frac{\alpha_1 r\cos\theta}{\sigma_1}+\frac{\alpha_2 r\sin\theta}{\sigma_2}\right)\begin{pmatrix}1\\-1\end{pmatrix}}{2\sqrt{1+\alpha_1^2+\alpha_2^2}}; \frac{\begin{pmatrix}2+\alpha_1^2+\alpha_2^2 & \alpha_1^2+\alpha_2^2\\ \alpha_1^2+\alpha_2^2 & 2+\alpha_1^2+\alpha_2^2\end{pmatrix}}{2(1+\alpha_1^2+\alpha_2^2)}\right\}. \tag{B.1}$$

Elliptical-Normal Cluster Process

$$f_d(r) = \frac{r}{2\sigma_1\sigma_2}\exp\left\{-\frac{(\sigma_1^2+\sigma_2^2)r^2}{8\sigma_1^2\sigma_2^2}\right\}\text{BesselI}_0\left\{\frac{(\sigma_1^2-\sigma_2^2)r^2}{8\sigma_1^2\sigma_2^2}\right\}. \tag{B.2}$$

For a different parametrization, $\sigma_1 \equiv \sigma$ and $\sigma_2 = c_\sigma\sigma$ with $c_\sigma > 0$, the pdf $f_d(r)$ can be rewritten as follows:

$$f_d(r) = \frac{1}{2c_\sigma\sigma^2}\exp\left\{-\frac{(1+c_\sigma^2)r^2}{8c_\sigma^2\sigma^2}\right\}\text{BesselI}_0\left\{\frac{(1-c_\sigma^2)r^2}{8c_\sigma^2\sigma^2}\right\}. \tag{B.3}$$

Circular-Normal Cluster Process

Here, $\sigma_1 = \sigma_2$. I provide the pdf of R, $f_R(r) = f_d(r)$ in the following:

$$f_R(r) \;=\; \frac{r}{2\sigma^2}\exp\left(-\frac{r^2}{4\sigma^2}\right). \tag{B.4}$$

Skew-Normal Cluster Process

Following the transformation defined in (3.1),

$$\begin{aligned} f_{R,\Theta}(r,\theta) \;=\; & \frac{r}{2c_0\sqrt{\pi^2 c_1 c_2 - 4\alpha_1^2\alpha_2^2}} \\ & \times\; \exp\left(-\frac{\pi r^2[\pi\{\cos^2\theta(c_2-c_1)+c_1\}+4\alpha_1\alpha_2\cos\theta\sin\theta]}{2c_0(\pi^2 c_1 c_2 - 4\alpha_1^2\alpha_1^2)}\right), \end{aligned} \tag{B.5}$$

where $c_0 = 2\sigma^2/(1+\alpha_1^2+\alpha_2^2)$, $c_1 = 1+\alpha_1^2(1-2/\pi)+\alpha_2^2$, and $c_2 = 1+\alpha_1^2+\alpha_2^2(1-2/\pi)$. The pdf, $f_d(r)$, is analytically complete only in the following two cases. First, $\alpha_1^2 = \alpha_2^2$, i.e., (i) $\boldsymbol{\alpha}^T = \alpha(1,1)$, (ii) $\boldsymbol{\alpha}^T = \alpha(-1,-1)$, (iii) $\boldsymbol{\alpha}^T = \alpha(1,-1)$, or (iv) $\boldsymbol{\alpha}^T = \alpha(-1,1)$, assuming that $\alpha > 0$. Consequently, $c_1 = c_2$,

$$\begin{aligned} f_d(r) \;=\; & \frac{\pi\sqrt{1+2\alpha^2}r}{2\sigma^2\sqrt{\pi\{\pi(1+2\alpha^2)-4\alpha^2\}}}\exp\left[-\frac{\{\pi+2\alpha^2(\pi-1)\}r^2}{4\sigma^2\{\pi(1+2\alpha^2)-4\alpha^2\}}\right] \\ & \times \text{BesselI}_0\left[\frac{\alpha^2 r^2}{2\sigma^2\{\pi(1+2\alpha^2)-4\alpha^2\}}\right], \end{aligned} \tag{B.6}$$

where $\text{BesselI}_0(x) = \sum_{n=0}^{\infty}(x/2)^{2n}/(n!)^2$ is a modified Bessel function of the first kind.

Second, suppose that $\boldsymbol{\alpha} = (0,\alpha)^T$ or $\boldsymbol{\alpha} = (\alpha,0)^T$. Then,

$$\begin{aligned} f_d(r) \;=\; & \frac{r(1+\alpha^2)}{2\sigma^2\sqrt{(1+\alpha^2)\{1+\alpha^2(1-2/\pi)\}}}\exp\left[-\frac{r^2\{1+\alpha^2(1-1/\pi)\}}{4\sigma^2\{1+\alpha^2(1-2/\pi)\}}\right] \\ & \times \text{BesselI}_0\left[\frac{\alpha^2 r^2}{4\pi\sigma^2\{1+\alpha^2(1-2/\pi)\}}\right]. \end{aligned} \tag{B.7}$$

B.2 SKEW-ELLIPTICAL-t CLUSTER PROCESSES

For $\mathbf{x} = (x_1, x_2)^T$,

$$f_{\mathbf{X}}(x_1, x_2) = \frac{T_1\left[\frac{\tau}{\sqrt{1+2(\alpha_1^2+\alpha_2^2)}}\left\{\frac{\nu+2}{\nu+\left(x_1^2/\sigma_1^2+x_2^2/\sigma_2^2\right)/2}\right\}^{1/2}; \nu+2\right]}{4\pi\sigma_1\sigma_2\left(1+\frac{x_1^2/\sigma_1^2+x_2^2/\sigma_2^2}{2\nu}\right)^{(\nu+2)/2} T_1\left\{\frac{\tau}{\sqrt{1+2(\alpha_1^2+\alpha_2^2)}}; \nu\right\}}, \tag{B.8}$$

where $T_1(\cdot; \nu)$ denotes the cdf of the univariate t-distribution with ν degrees of freedom. According to the transformation in (3.1), the joint distribution of (R, Θ) is

$$\begin{aligned} & f_{R,\Theta}(r, \theta) \\ &= \frac{r\, T_1\left[\frac{\tau}{\sqrt{1+2(\alpha_1^2+\alpha_2^2)}}\left\{\frac{\nu+2}{\nu+\left(r^2\cos^2\theta/\sigma_1^2+r^2\sin^2\theta/\sigma_2^2\right)/2}\right\}^{1/2}; \nu+2\right]}{4\pi\sigma_1\sigma_2\left(1+\frac{r^2\cos^2\theta/\sigma_1^2+r^2\sin^2\theta/\sigma_2^2}{2\nu}\right)^{(\nu+2)/2} T_1\left\{\frac{\tau}{\sqrt{1+2(\alpha_1^2+\alpha_2^2)}}; \nu\right\}}. \end{aligned} \tag{B.9}$$