Skew-Elliptical Cluster Processes

Ngoc Anh Dao and Marc G. Genton

Abstract This paper introduces skew-elliptical cluster processes. In contrast to the simple Gaussian isotropic structure of the distribution of the "children" events of a Thomas process, we propose an anisotropic structure by allowing the choice of a flexible covariance matrix and incorporating skewness or ellipticity parameters into the structure. Since the theoretical pair correlation functions of these processes are complex and analytically incomplete, and therefore the estimation of the parameters is computationally intensive, we propose reasonable approximations of the theoretical pair correlation functions of these cluster processes, which allow for a simpler parameter estimation. We present the estimation of their parameters using the minimum contrast method. For a data application, we use a fraction of the full redwood dataset. Our analysis shows that an elliptical cluster process can describe this point pattern better than a common Thomas process, since it is able to statistically model the non-circular shapes of the clusters in the data. The skew-elliptical cluster processes can be very meaningful for analyzing complex datasets in the field of spatial point processes since they provide more flexibility to detect interesting characteristics of the data.

1 Introduction

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 work, we introduce a class of skew-elliptical cluster processes that includes the (traditional) TP and offers the possibility

N. A. Dao
Department of Statistics, Texas A&M University, College Station, TX, USA
e-mail: dao@stat.tamu.edu

M. G. Genton (✉)
Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
e-mail: marc.genton@kaust.edu.sa

I. Ghosh et al. (eds.), *Advances in Statistics - Theory and Applications*, Emerging Topics in Statistics and Biostatistics, https://doi.org/10.1007/978-3-030-62900-7_18

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 Neyman–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 Neyman–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.

We now look at how Neyman–Scott cluster point processes and the TP are defined. A Neyman–Scott cluster point process is constructed by letting unobservable, the 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 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 (Møller and Waagepetersen 2003; Thomas 1949). 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 (2006) and Tanaka et al. (2008), but like Castelloe (1998) we generalize the TP in our approach not by introducing the pcf first but by presenting the general distribution of the "children" events. In particular, we 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, we 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 our 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, we 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 that 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, we motivate our approach by showing graphical representations of "children" events of skew-elliptical-normal CPs in Fig. 1 and skew-elliptical-t CPs in Fig. 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 Sects. 2 and 3. In each of the first rows of Figs. 1 and 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 $\text{diag}(\sigma_1^2, \sigma_2^2)$, 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 that 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 Capitanio (1999), and the

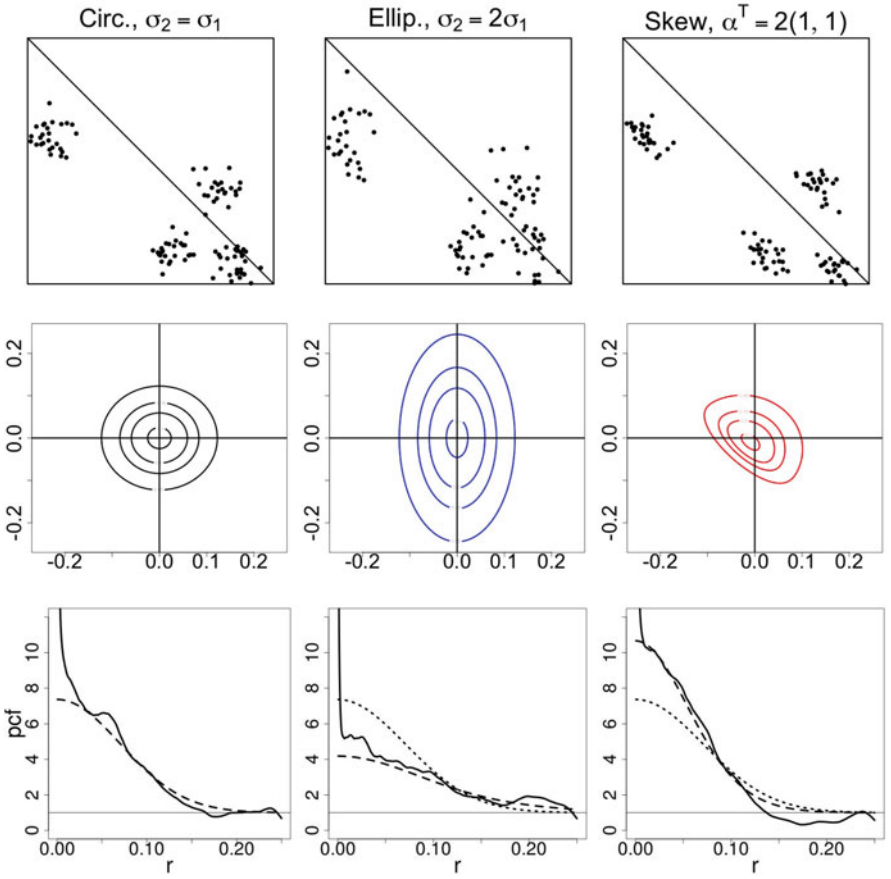

Fig. 1 To generate spatial point patterns (SPPs), $\kappa = 5$, $\mu = 25$, and the same random seed were used. The first row shows a pattern of a circular-normal CP (TP) (left) with "children" events, $\mathbf{Y}$, being bivariate normal distributed with the isotropic dispersion matrix $\sigma^2\mathbf{I}_2$, $\sigma^2 = 0.05^2$; 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 $\sigma^2\mathbf{I}_2$, $\sigma^2 = 0.05^2$, and skewness parameter $\boldsymbol{\alpha} = 2(1, 1)^T$. The parameters are described in Sect. 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), and 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 approximating pcf (dashed) of each model, and the theoretical pcf (dotted) of the circular-normal CP (TP) as a reference

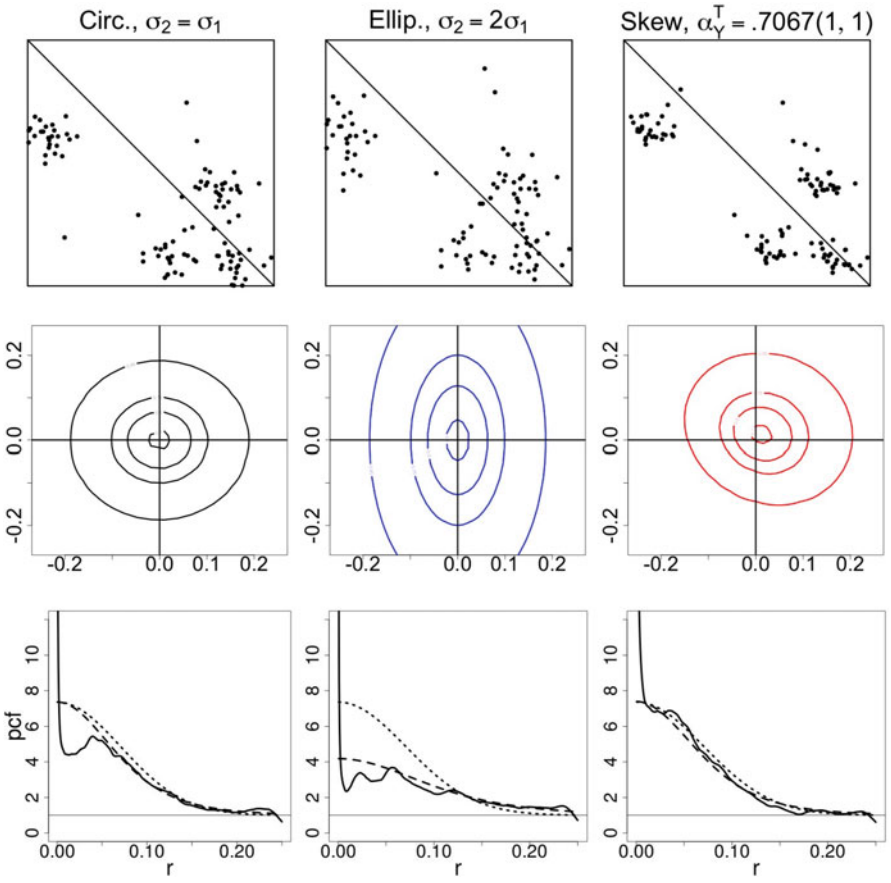

Fig. 2 As in Fig. 1, to generate SPPs of skew-elliptical-t CPs with four df ($\nu = 4$), $\kappa = 5$, $\mu = 25$, and the same random seed were used. The first row shows a pattern of a circular-t CP (left) with "children" events, $\mathbf{Y}$, being bivariate t-distributed with dispersion matrix $\sigma^2\mathbf{I}_2$ with $\sigma^2 = 0.05^2$, one of an 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 Sect. 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), and 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 approximating pcf (dashed) of each model, and the theoretical pcf (dotted) of the circular-normal CP (TP) as a reference

skewness parameter, $\boldsymbol{\alpha}_{\mathbf{Y}}$ (Sect. 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 Fig. 2 have more dispersed clusters than do those of the skew-elliptical-normal CPs, in Fig. 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 CPs, except for the TP, are all analytically incomplete. In Castelloe (1998) and Møller and Toftaker (2014), we face the challenge in estimation using the Bayesian approach. However, if we relax the anisotropy condition to the assumption of isotropy, approximation of the pcf is analytically complete. Then, we 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 our 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.

This paper is organized with the following structure. Sections 2 and 3 present the approximating pcfs of the skew-elliptical-normal and the corresponding skew-elliptical-t CPs. Some analytical derivations were carried out with `Mathematica` (Wolfram Research, Inc. 2020). The intermediate derivation steps are given in the Appendix. Section 4 demonstrates the performance of parameter estimation via the MCM using the function `optim` available in `R` (R Core Team 2019). Section 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 2005; Baddeley et al. 2015). Finally, Sect. 6 introduces alternative probability distributions to extend our work on TP, suggests to generalize a similar clustered spatial point process, and raises a possible exploration for an adjustment of the MCM.

2 Skew-Elliptical-Normal Cluster Processes

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 $-\boldsymbol{\omega}\boldsymbol{\delta}\sqrt{2/\pi}$, 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, $\mathbf{Y} \sim \mathrm{SN}_2(-\boldsymbol{\omega\delta}\sqrt{2/\pi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$. 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 $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cumulative distribution function (cdf) of the univariate standard normal distribution, $\phi_2(\cdot;)$ and $\Phi_2(\cdot;)$ denote the corresponding functions of the bivariate normal distribution and $\mathbf{y} = (y_1, y_2)^T$ (Arellano-Valle and Azzalini 2006; Azzalini and Dalla Valle 1996). Then, $\mathrm{E}(\mathbf{Y}) = \mathbf{0}$ and $\mathrm{Var}(\mathbf{Y}) = \boldsymbol{\Omega} - \frac{2}{\pi}\boldsymbol{\omega\delta\delta}^T\boldsymbol{\omega}$ (Azzalini and Capitanio 1999; Gupta et al. 2013). Then, $\mathbf{Y}$ is also a unified skew-normal (SUN) random vector (Arellano-Valle and Azzalini 2006; Azzalini and Capitanio 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 Capitanio (2014): $\mathbf{Y}$ has distribution denoted by $\mathrm{SUN}_{2,1}(-\boldsymbol{\omega\delta}\sqrt{2/\pi}, \boldsymbol{\Omega}, \boldsymbol{\delta}, 0, 1)$. 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, $\mathbf{X} \sim \mathrm{SUN}_{2,2}(\mathbf{0}, 2\boldsymbol{\Omega}, \boldsymbol{\Delta}, 0, \mathbf{I}_2)$ (Azzalini and Capitanio 2014, Ch. 7), where $\boldsymbol{\Delta} = \boldsymbol{\delta}/\sqrt{2}\,(1, -1)$, i.e., the pdf of $\mathbf{X}$ is $f_{\mathbf{X}}(\mathbf{x}) = 4\phi_2(\mathbf{x}; 2\boldsymbol{\Omega})\ \Phi_2(\boldsymbol{\Delta}^T\boldsymbol{\omega}^{-1}\mathbf{x}/\sqrt{2}; \mathbf{I}_2 - \boldsymbol{\Delta}^T\boldsymbol{\Delta})$. Explicitly,

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

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.

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 our setting, g is anisotropic but translation invariant, $g(\mathbf{u}, \mathbf{v}) = g(\mathbf{v} - \mathbf{u})$, we 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. We 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 $r \ge 0$ 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. We 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, we 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. We 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.

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

$$X_1 = R\cos\Theta, \quad X_2 = R\sin\Theta, \text{ and} \tag{1}$$
$$R = \sqrt{\mathbf{X}^T\mathbf{X}} = \sqrt{X_1^2 + X_2^2}, \quad \Theta = \arctan(X_2/X_1).$$

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 (1), the joint distribution, $f_{R,\Theta}(r,\theta)$, is derived in (A.1). The pdf of R follows easily, and from $g_d(r) = 1 + f_d(r)/(2\pi\kappa r)$, the pcf is

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

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., $\Phi_2(\cdot;) = \Phi_2\left\{(0,0)^T; \mathbf{I}_2\right\} = 1/4$.

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 (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 we 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. We estimate κ, σ^2, and c_σ^2 using g_d via the MCM.

As mentioned in Sect. 1, Stoyan (2006) and Tanaka et al. (2008) introduced different models with more than one dispersion parameters to generalize the (traditional) TP. For comparison, we provide the pdf of R of our model in (A.2) and (A.3) in the Appendix.

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 (A.4) in the Appendix. From (2), the true pcf is

$$g(r) = 1 + \frac{\exp\{-r^2/(4\sigma^2)\}}{4\pi\kappa\sigma^2}. \tag{3}$$

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; for example, it can be found in Cressie (1993). To estimate κ and σ^2, the MCM can use either the pcf or the K-function.

2.5 The Skew-Normal Cluster Process

Let the distribution of $\mathbf{Y}$ be a special case of the SUN distribution mentioned earlier in Sect. 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 $\mathrm{E}(V_0) = \sqrt{2/\pi}$, $\mathrm{Var}(V_0) = 1 - 2/\pi$, and $\mathbf{V}_1$ is bivariate normal distributed, $\mathrm{N}_2(\mathbf{0}, \boldsymbol{\Psi})$, 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, Sect. 2.1.), $\mathbf{Y}$ is bivariate SUN distributed, in particular, $\mathrm{E}(\mathbf{Y}) = \mathbf{0}$, $\mathrm{Var}(\mathbf{Y}) = \sigma^2(\mathbf{I}_2 - 2/\pi\boldsymbol{\delta}\boldsymbol{\delta}^T)$. 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 Sect. 2.1 with $\sigma_1 = \sigma_2 = \sigma$. Since it is centrally symmetric, we 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

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2c_0\sqrt{\pi^2 c_1 c_2 - 4\alpha_1^2\alpha_2^2}} \exp\left[- \frac{\pi\{\pi(c_2 x_1^2 + c_1 x_2^2) + 4\alpha_1\alpha_2 x_1 x_2\}}{2c_0(\pi^2 c_1 c_2 - 4\alpha_1^2\alpha_2^2)} \right],$$

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 (A.5) in the Appendix. The approximating pcf, $g_d(r)$, 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 (A.6) in the Appendix. Consequently, $g_d(r)$ is

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

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

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

K_d of the above scenarios is analytically incomplete. We 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 1 So far we 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 Skew-Elliptical-*t* Cluster Processes

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, we 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, we derive that $\mathbf{X} = \mathbf{Y} - \mathbf{Y}^* \sim \mathrm{EST}_2(\mathbf{0}, 2\boldsymbol{\Omega}, \mathbf{0}, \nu, \tau/\sqrt{1+2\boldsymbol{\alpha}^T\boldsymbol{\alpha}})$.

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, we 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, we 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 2010a,b; González-Farías et al. 2004). We chose to approximate the distribution of $\mathbf{X}_{\text{true}}$ by $\mathbf{X}$, i.e., $\mathbf{X}_{\text{true}} \stackrel{d}{\approx} \mathbf{X} \sim \mathrm{EST}_2(\mathbf{0}, 2\boldsymbol{\Omega}, \mathbf{0}, \nu, \tau/\sqrt{1+2\boldsymbol{\alpha}^T\boldsymbol{\alpha}})$. 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 (A.8), and under the isotropy assumption, the joint distribution function, $f_{d,R,\Theta}(r, \theta)$, is provided in (A.9). If $\boldsymbol{\alpha} \neq \mathbf{0}$ and $\sigma_1 \neq \sigma_2$, the approximating pcf is analytically incomplete:

$$g_d(r) = 1 + \frac{1}{8\pi^2\kappa\sigma_1\sigma_2 T_1\left\{\frac{\tau}{\sqrt{1+2(\alpha_1^2+\alpha_2^2)}}; \nu\right\}} \times \int_0^{2\pi} \frac{T_1\left[\frac{\tau}{\sqrt{1+2(\alpha_1^2+\alpha_2^2)}}\left\{\frac{\nu+2}{\nu+(r^2\cos^2\theta/\sigma_1^2+r^2\sin^2\theta/\sigma_2^2)/2}\right\}^{1/2}; \nu+2\right]}{\left(1+\frac{r^2\cos^2\theta/\sigma_1^2+r^2\sin^2\theta/\sigma_2^2}{2\nu}\right)^{(\nu+2)/2}} d\theta. \tag{4}$$

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

3.2 The Skew-t Cluster Process

If $\boldsymbol{\alpha} \neq \mathbf{0}$ and $\sigma_1 = \sigma_2$, i.e., $(\mathbf{Y}^T, \mathbf{Y}^{*T})^T \sim \mathrm{EST}_4(\mathbf{0}, \sigma^2\mathbf{I}_4, (\boldsymbol{\alpha}^T, \boldsymbol{\alpha}^T)^T, \nu, \tau)$ (Arellano-Valle and Genton 2010b) and, hence, $\mathbf{Y} \sim \mathrm{EST}_2(\mathbf{0}, \sigma^2\mathbf{I}_2, \boldsymbol{\alpha}_{\mathbf{Y}}, \nu, \tau_{\mathbf{Y}})$, where $\boldsymbol{\alpha}_{\mathbf{Y}} = \boldsymbol{\alpha}/\sqrt{1+\boldsymbol{\alpha}^T\boldsymbol{\alpha}}$ and $\tau_{\mathbf{Y}} = \tau/\sqrt{1+\boldsymbol{\alpha}^T\boldsymbol{\alpha}}$, we obtain the following approximating pcf under isotropy assumption from (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, we estimate κ, σ^2, α^2, and c_α^2 via MCM using the pcf. Then, we can compute the estimates of $\alpha_{\mathbf{Y},1}$ and $\alpha_{\mathbf{Y},2}$ because eventually we are interested in knowing the estimate of the skewness parameter of the skew-t CP, which is $\boldsymbol{\alpha}_{\mathbf{Y}}$, not $\boldsymbol{\alpha}$.

Remark 2 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 Sect. 5.

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 $\nu = 1$, $g_d(r) \equiv 1$. Only for $\nu = 2k$, where $k \in \mathbb{N}$, is $g_d(r)$ analytically complete. For each even df, we 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. We 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 is analytically incomplete regardless of df. We 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.

3.4 The Circular-t Cluster Process

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

$$g(r) = 1 + \frac{1}{4\pi\kappa\sigma^2}\left(1 + \frac{r^2}{2\nu\sigma^2}\right)^{-(\nu+2)/2}, \tag{5}$$

and the K-function is

$$K(r) = \pi r^2 + \frac{1 - \left\{1 + r^2/(2\sigma^2\nu)\right\}^{-\nu/2}}{\kappa}. \tag{6}$$

We can estimate κ and σ^2. The results presented in Sect. 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 $\sigma_1 = \sigma_2$, and 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 the Appell's hypergeometric function. In particular (Erdélyi et al. 1953),

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

and s_1 and 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 (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.

3.5 *The Case of Orthogonality*

If $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are orthogonal, i.e., $\mathrm{E}(\mathbf{Y}_1^T\mathbf{Y}_2) = 0$, 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 $V \sim \text{Gamma}(\nu/2, \nu/2)$, $\mathbf{Y}_i$ 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 (5) and (6).

4 Parameter Estimation by Minimum Contrast

Diggle (2003, Sect. 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 our setting, using the approximating pcf $g_d(r;\boldsymbol{\theta})$ and the empirical pcf $\hat{g}(r)$, we redefine the discrepancy,

$$D_{d,g}(\boldsymbol{\theta}) = \int_0^{r_0} w(r)\left[\{\hat{g}(r)\}^{c_{\text{stabil}}} - \{g_d(r;\boldsymbol{\theta})\}^{c_{\text{stabil}}}\right]^2 dr. \tag{7}$$

For the data simulation, we want to work with spatial point patterns (SPPs) having approximately 200 events on a unit square. Consequently, the dispersion parameters σ_1 and 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, we want the number of "parent" and of "children" events to be between 10 and 20, so that the parameter estimation can be stable. Thus, we chose $\kappa = 20$, $\sigma_1 = \sigma = 0.04$, $\mu = 10$. In Table 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

Table 1 3000 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 our 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 (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 (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	$(\kappa, \sigma_1, c_\sigma)^T = (20, 0.04, 2)^T$ $\log(\kappa_0, \sigma^2_{1,0}, c^2_{\sigma,0})^T =$ $(0, -4, 3.5)^T$	0.164	0.029	0.031	96.0
Skew-normal	$(\kappa, \sigma, \alpha_1 = \alpha_2 = \alpha)^T =$ $(20, 0.04, 10)^T$ $\log(\kappa_0, \sigma^2_0, \alpha^2_0)^T =$ $(0, -4, 5.5)^T$	0.177	0.176	0.179	83.9
Elliptical-t, df $= 6$	$(\kappa, \sigma_1, c_\sigma)^T = (20, 0.04, 2)^T$ $\log(\kappa_0, \sigma^2_{1,0}, c^2_{\sigma,0})^T =$ $(0, -4, 3.3)^T$	0.115	0.035	0.038	85.3
Skew-t, df $= 6$	$(\kappa, \sigma, \alpha_1 = \alpha_2 = \alpha)^T =$ $(20, 0.04, 20)^T$ $\log(\kappa_0, \sigma^2_0, \alpha^2_0)^T =$ $(0, -5, 5)^T$	0.082	0.098	0.101	84.7
Circular-t, df $= 6$	$(\kappa, \sigma)^T = (20, 0.04)^T$ $\log(\kappa_0, \sigma^2_0)^T = (0, -4)^T$	0.035	0.077	0.085	88.1

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, $\alpha_1 = \alpha_2 = \alpha = 20$, and $\tau = 1$ 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). Our 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 Sect. 5, we use the empirical pcf according to the translation-corrected estimator. The computation was done on compute nodes that have 8 CPU cores, 32 GB of RAM, and the CPU processors clocked at 2.4 GHz or faster.

Since we 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). We used the function `optim` available in `R` to minimize (7).

The logarithms of the starting values needed for the function `optim` are given in the second column of Table 1 since we estimated the logarithms of κ, σ^2, c_σ^2, and α^2. For the parameter estimation, we just need to estimate α^2 since we set $\alpha_1 = \alpha_2 = \alpha$, i.e., $c_\alpha = 1$, for the skew-normal and skew-t CP. Additionally, we 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, and 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 empirical pcfs. Table 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` makes sense in the last three columns. Let $\overline{\text{Dis}_{d,g}^2}(\hat{\boldsymbol{\theta}})$ denote the average of *bilateral discrepancy*,

$$\text{Dis}_{d,g}^2(\hat{\boldsymbol{\theta}}) = \int \{\hat{g}(r) - g_d(\hat{\boldsymbol{\theta}}, r)\}^2 + \{g_d(\hat{\boldsymbol{\theta}}, r) - g_d(\boldsymbol{\theta}, r)\}^2 dr, \tag{8}$$

where the d, g subscript shows the involvement of the approximating pcf, $g_d(r)$, and $\hat{\boldsymbol{\theta}}$ denotes the estimate. $\overline{\text{Dis}_{d,g}^2}(\hat{\boldsymbol{\theta}}_n)$ is the average of bilateral discrepancy $\text{Dis}_{d,g}^2(\hat{\boldsymbol{\theta}}_n)$, where $\hat{\boldsymbol{\theta}}_n$ denotes the estimate resulting from the assumption of the (true) novel (skew-elliptical) CP. Similarly, $\overline{\text{Dis}_{d,g}^2}(\hat{\boldsymbol{\theta}}_t)$ is the average of bilateral discrepancy $\text{Dis}_{d,g}^2(\hat{\boldsymbol{\theta}}_t)$, where $\hat{\boldsymbol{\theta}}_t$ denotes the estimate resulting from the assumption of the (wrong) traditional TP. For each of 3000 SPPs, we could compute $\text{Dis}_{d,g}^2(\hat{\boldsymbol{\theta}}_n)$ and $\text{Dis}_{d,g}^2(\hat{\boldsymbol{\theta}}_t)$. The percentage in the last column shows how often $\text{Dis}_{d,g}^2(\hat{\boldsymbol{\theta}}_n) < \text{Dis}_{d,g}^2(\hat{\boldsymbol{\theta}}_t)$; 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 (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 2 displays the choice of models and statistical information of the 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 (7).

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 (2005), 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 processes, e.g., Diggle (2003). The `redwoodfull` dataset appears to be interesting because it has many clusters that display non-circular shapes. We plotted `redwoodfull` in Fig. 3 (left) in clustered and inhibitory partitions, represented by circles and triangles, respectively. In our opinion, 73 trees represented by triangles cannot be reasonably described by a clustered spatial point process since they follow an inhibitory pattern. We are interested only in analyzing the *clustered redwoods*, especially in finding out which skew-elliptical CPs can best model the process generating it. Figure 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 (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

Table 2 The information about the models is given in Table 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. $\text{MSE} = \text{Bias}^2 + \text{Var}(\hat{\theta})$, 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	0.042	0.098		26.485	0.052
Std	12.686	0.013	0.077		12.522	0.011
2.5%	4.711	0.020	0.038		9.374	0.035
97.5%	52.724	0.068	0.313		56.785	0.077
Bias^2	3.436	29×10^{-7}	31×10^{-5}		42.060	14×10^{-5}
MSE	164.366	16×10^{-5}	0.006		198.867	26×10^{-5}
Skew-normal	23.175	0.034		2.355	23.593	0.028
Std	7.486	0.007		4.716	7.611	0.004
2.5%	11.255	0.023		0.066	11.503	0.022
97.5%	40.024	0.049		10.789	40.912	0.037
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	0.046	0.094		26.625	0.053
Std	12.168	0.017	0.064		13.469	0.012
2.5%	4.988	0.018	0.037		9.126	0.035
97.5%	51.158	0.084	0.268		61.441	0.080
Bias^2	0.221	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	0.040		19.245	22.244	0.037
Std	7.952	0.010		8.624	8.399	0.007
2.5%	7.174	0.027		6.848	9.108	0.028
97.5%	7.403	0.065		41.041	40.904	0.056
Bias^2	0.163	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	0.039			22.820	0.039
Std	8.446	0.006			8.703	0.006
2.5%	10.253	0.029			10.538	0.029
97.5%	41.927	0.054			43.340	0.053
Bias^2	4.119	3×10^{-7}			7.952	8×10^{-7}
MSE	75.446	35×10^{-6}			83.702	39×10^{-6}

of r approximately from 0.01 to 0.05, suggesting clustering. Overall, there are graphical hints that the redwoods of interest are clustered. We have, however, to investigate statistically whether this is truly the case. First, we test whether CSR can provide a good fit to the redwoods of interest. For that, the plug-in goodness-of-fit

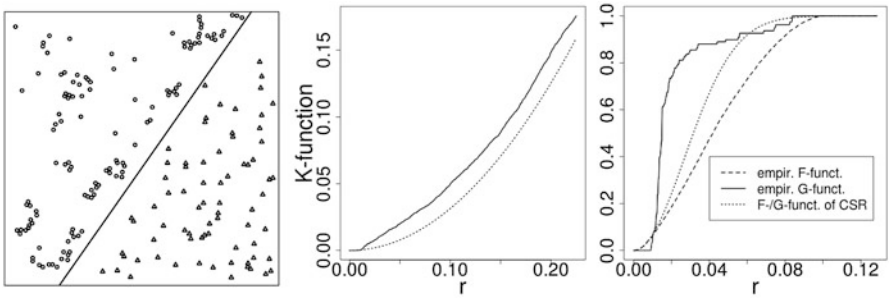

Fig. 3 On the left, the locations of 195 Californian redwood seedlings and saplings in a square sampling region, 130 × 130 feet, are shown. They are displayed in two partitions: circles represent the clustered redwoods and triangles the inhibitory ones, respectively. In the middle, the empirical K-function (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

test using the G- and F-functions (Diggle 2003, Chap. 1.7) is employed, and the resulting estimated p-values, $\hat{p}$, are all 0. Since the p-values are smaller than the nominal significance level of $\alpha_{\text{GOF}} = 0.05$, we 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, we compute the estimates and the corresponding discrepancy between the empirical pcf and the theoretical one of the underlying model. For elliptical-t CPs, we choose 2, 6, 10, 20, and 26 df for simplicity since their pcfs can be computed analytically with `Mathematica`. For skew- and circular-t CPs, we consider all even df up to 30, although we display results only for 2, 10, 20, and 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) 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 Fig. 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, Sect. 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 we are 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, Sect. 4.3.2). We 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

Table 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 columns. Here, $\hat{g}(r)$ is the empirical pcf. The discrepancy, $\mathrm{Dis}_g^1(\hat{\boldsymbol{\theta}})$, between the empirical pcf and the theoretical one of certain underlying model is defined in (9). The smallest discrepancy in each column is displayed boldly

	Using $\hat{g}(r)$completely		Using $\hat{g}(r)$ partially	
CP	Estimates	$\mathrm{Dis}_g^1(\hat{\boldsymbol{\theta}})$	Estimates	$\mathrm{Dis}_g^1(\hat{\boldsymbol{\theta}})$
Circular-normal	$\hat{\kappa} = 60.910$, $\hat{\sigma} = 0.022$	0.04824	$\hat{\kappa} = 60.082$, $\hat{\sigma} = 0.018$	0.007080
Elliptical-normal	$\hat{\kappa} = 60.880$, $\hat{\sigma}_1 = 0.022$, $\hat{\sigma}_2 = 0.022$	0.04823	$\hat{\kappa} = 53.098$, $\hat{\sigma}_1 = 0.009$, $\hat{\sigma}_2 = 0.030$	**0.004028**
Skew-normal	$\hat{\kappa} = 60.873$, $\hat{\sigma} = 0.022$, $\hat{\alpha}_1 = -0.185$, $\hat{\alpha}_2 = 0.185$	**0.04822**	$\hat{\kappa} = 58.647$, $\hat{\sigma} = 0.023$ $\hat{\alpha}_1 = -20.254$, $\hat{\alpha}_2 = 20.254$	0.006032
Circular-t, df $= 2$	$\hat{\kappa} = 47.182$, $\hat{\sigma} = 0.0242$	0.05329	$\hat{\kappa} = 49.164$, $\hat{\sigma} = 0.0172$	0.004481
Circular-t, df $= 10$	$\hat{\kappa} = 57.988$, $\hat{\sigma} = 0.0223$	0.04899	$\hat{\kappa} = 57.754$, $\hat{\sigma} = 0.0177$	0.005673
Circular-t, df $= 20$	$\hat{\kappa} = 59.369$, $\hat{\sigma} = 0.0222$	0.04857	$\hat{\kappa} = 58.845$, $\hat{\sigma} = 0.0178$	0.006274
Circular-t, df $= 30$	$\hat{\kappa} = 59.852$, $\hat{\sigma} = 0.0222$	0.04846	$\hat{\kappa} = 59.283$, $\hat{\sigma} = 0.0179$	0.006513
Elliptical-t, df $= 2$	$\hat{\kappa} = 47.148$, $\hat{\sigma}_1 = 0.0242$, $\hat{\sigma}_2 = 0.0242$	0.05328	$\hat{\kappa} = 49.158$, $\hat{\sigma}_1 = 0.0172$, $\hat{\sigma}_2 = 0.0172$	0.004481
Elliptical-t, df $= 10$	$\hat{\kappa} = 57.949$, $\hat{\sigma}_1 = 0.0223$, $\hat{\sigma}_2 = 0.0223$	0.04899	$\hat{\kappa} = 52.712$, $\hat{\sigma}_1 = 0.0095$, $\hat{\sigma}_2 = 0.0282$	0.004154
Elliptical-t, df $= 20$	$\hat{\kappa} = 59.042$, $\hat{\sigma}_1 = 0.0223$, $\hat{\sigma}_2 = 0.0223$	0.04862	$\hat{\kappa} = 52.845$, $\hat{\sigma}_1 = 0.0092$, $\hat{\sigma}_2 = 0.0292$	0.004112
Elliptical-t, df $= 26$	$\hat{\kappa} = 59.732$, $\hat{\sigma}_1 = 0.0222$, $\hat{\sigma}_2 = 0.0222$	0.04847	$\hat{\kappa} = 52.959$, $\hat{\sigma}_1 = 0.0091$, $\hat{\sigma}_2 = 0.0296$	0.004078
Skew-t, df $= 2$	$\hat{\kappa} = 47.174$, $\hat{\sigma} = 0.0243$, $\hat{\alpha}_1 = -5.178$, $\hat{\alpha}_2 = 62.388$ ($\hat{\alpha}_{\mathbf{Y},1} = -0.083$, $\hat{\alpha}_{\mathbf{Y},2} = 0.996$)	0.053	$\hat{\kappa} = 49.180$, $\hat{\sigma} = 0.0172$, $\hat{\alpha}_1 = -12.368$, $\hat{\alpha}_2 = 84.517$ ($\hat{\alpha}_{\mathbf{Y},1} = 0 - 0.145$, $\hat{\alpha}_{\mathbf{Y},2} = 0.989$)	0.004480

(continued)

Table 3 (continued)

CP	Using $\hat{g}(r)$completely		Using $\hat{g}(r)$ partially	
	Estimates	$\text{Dis}_g^1(\hat{\boldsymbol{\theta}})$	Estimates	$\text{Dis}_g^1(\hat{\boldsymbol{\theta}})$
Skew-t, df $= 10$	$\hat{\kappa} = 57.963$, $\hat{\sigma} = 0.0224$, $\hat{\alpha}_1 = -3.356$, $\hat{\alpha}_2 = 49.077$ ($\hat{\alpha}_{\mathbf{Y},1} = 0 - 0.068$, $\hat{\alpha}_{\mathbf{Y},2} = 0.997$)	0.049	$\hat{\kappa} = 57.732$, $\hat{\sigma} = 0.0177$, $\hat{\alpha}_1 = -2.595$, $\hat{\alpha}_2 = 28.391$ ($\hat{\alpha}_{\mathbf{Y},1} = 0 - 0.091$, $\hat{\alpha}_{\mathbf{Y},2} = 0.995$)	0.005671
Skew-t, df $= 20$	$\hat{\kappa} = 59.427$, $\hat{\sigma} = 0.0222$, $\hat{\alpha}_1 = -8.095$, $\hat{\alpha}_2 = 51.185$ ($\hat{\alpha}_{\mathbf{Y},1} = 0 - 0.156$, $\hat{\alpha}_{\mathbf{Y},2} = 0.988$)	0.0486	$\hat{\kappa} = 58.890$, $\hat{\sigma} = 0.0178$, $\hat{\alpha}_1 = -4.079$, $\hat{\alpha}_2 = 45.750$ ($\hat{\alpha}_{\mathbf{Y},1} = 0 - 0.089$, $\hat{\alpha}_{\mathbf{Y},2} = 0.996$)	0.006273
Skew-t, df $= 30$	$\hat{\kappa} = 59.915$, $\hat{\sigma} = 0.0222$, $\hat{\alpha}_1 = -8.646$, $\hat{\alpha}_2 = 18.942$ ($\hat{\alpha}_{\mathbf{Y},1} = 0 - 0.415$, $\hat{\alpha}_{\mathbf{Y},2} = 0.909$)	0.048	$\hat{\kappa} = 59.270$, $\hat{\sigma} = 0.0179$, $\hat{\alpha}_1 = -4.048$, $\hat{\alpha}_2 = 45.369$ ($\hat{\alpha}_{\mathbf{Y},1} = 0 - 0.089$, $\hat{\alpha}_{\mathbf{Y},2} = 0.996$)	0.006519

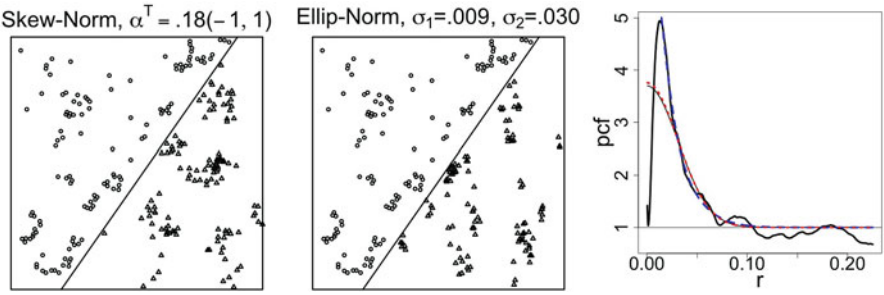

Fig. 4 The 122 clustered redwoods (circles) are displayed in the upper left polygon of the left and middle plots. There, in the lower right polygon, the left plot shows the simulated events (triangles) of the skew-normal CP with parameters $\kappa = 60.873$, $\sigma = 0.022$, $\alpha_1 = -0.185$, $\alpha_2 = 0.185$, and $\mu = 3.632$, and similarly, the middle plot shows the simulated events (triangles) of the elliptical-normal CP with parameters $\kappa = 53.098$, $\sigma_1 = 0.009$, $\sigma_2 = 0.030$, and $\mu = 4.163$. 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 Fig. 1 was used

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 (9), of the corresponding models from both analyses are listed in Table 3. We define the discrepancy between the empirical and approximating pcfs at $\hat{\boldsymbol{\theta}}$ as follows

$$\text{Dis}_g^1(\hat{\boldsymbol{\theta}}) = \int_0^{r_0} \{\hat{g}(r) - g_d(\hat{\boldsymbol{\theta}}, r)\}^2 \, dr, \tag{9}$$

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 is tested, we want to clarify a point that might appear to be an inconsistency in our calculation. Under the column "Using $\hat{g}(r)$ partially", the discrepancies, $\text{Dis}_g^1(\hat{\boldsymbol{\theta}})$, of the skew-$t$ CPs do not converge to the one (0.006032) of the skew-normal CP when the df increases. 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 increases. 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 2.

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, Sect. 6.2.) recommended not to use the GOF test based on the K-function if the K-function was used for parameter estimation. Since we used the empirical pcf (originating from the K-function) for the parameter estimation, we could rely on the GOF conclusion from the AGOF-G or $-F$ test. We 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, we 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, we run AGOF-G tests that 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, we 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, we conclude that these models provide a good fit, statistically speaking.

Now, we examine these models graphically. The right plot (Fig. 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 (Fig. 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 that 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, we 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.

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 toward starting values of the empirical pcf under isotropy assumption. It is usually poorly estimated at a short distance, i.e., r close to 0. We encountered this problem in our 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, Sect. 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 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 that 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, we 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 (9) goes below a pre-set limit. According to our 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 $\boldsymbol{\theta}_2$.

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 work, we generalized the TP to some extent. However, we 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 Neyman–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.

Acknowledgments The authors would like to thank Professor Reinaldo Arellano-Valle and a reviewer for their useful suggestions and comments. The authors also acknowledge the Texas A&M University Brazos cluster that contributed to the research reported here. The work of the second author was supported by King Abdullah University of Science and Technology (KAUST).

Appendix

Table of Acronyms

Acronym	Complete words
AGOF	Adjusted goodness-of-fit
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
pcf	Pair correlation function
pdf	Probability density function
SPP	Spatial point pattern
SUE	Unified skew-elliptical
SUN	Unified skew-normal
TP	Thomas process

Skew-Elliptical-Normal Cluster Processes

According to the transformation in (1), the joint distribution $f_{R,\Theta}(r,\theta)$ of (R,Θ) is

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

Elliptical-Normal Cluster Process

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

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:

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

Circular-Normal Cluster Process

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

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

Skew-Normal Cluster Process

Following the transformation defined in (1),

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

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, $c_1 = c_2$,

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

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,

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

Skew-Elliptical-t Cluster Processes

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

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

where $T_1(\cdot;\nu)$ denotes the cdf of the univariate t-distribution with ν degrees of freedom. According to the transformation in (1), the joint distribution of (R,Θ) is

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

References

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

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

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

Arellano-Valle, R. B., & Genton, M. G. (2010b). Multivariate extended skew-*t* distributions and related families. *Metron, 68*, 2001–234

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

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

Azzalini, A., & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distributions. *Journal of the Royal Statistical Society B, 61*, 579–602.

Azzalini, A., & Capitanio, A. (2014). *The skew-normal and related families*. Cambridge: Cambridge University Press.

Azzalini, A., & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika, 83*, 715–726.

Baddeley, A. J., & Turner, R. (2005). SPATSTAT: An R package for analyzing spatial point patterns. *Journal of Statistical Software, 12*, 1–42.

Baddeley, A. J., Rubak, E., & Turner, R. (2015). *Spatial point patterns: Methodology and applications with R*. London: Chapman and Hall/CRC Press.

Castelloe, J. M. (1998). Issues in reversible jump Markov chain Monte Carlo and composite EM analysis, applied to spatial Poisson cluster processes. Ph.D. thesis, University of Iowa (1998)

Cressie, N. A. C. (1993). *Statistics for spatial data*. New York, NY: Wiley (1993)

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

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

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

Diggle, P. J. (2003). *Statistical analysis of spatial point patterns*. London: Arnold.

Erdélyi, A., Magnus, W., Oberhettinger, F., & Tricomi, F. G. (1953). *Higher transcendental functions*. New York, NY: McGraw-Hill.

González-Farías, G., Domíniguez-Molina, J. A., & Gupta, A. K. (2004). The closed skew-normal distribution. In M. G. Genton (Ed.), *Skew-elliptical distributions and their applications: A journey beyond normality* (pp. 25–42). Boca Raton, FL: Chapman & Hall/CRC.

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

Gupta, A. K., González-Farías, G., & Domíniguez-Molina, J. A. (2004). A multivariate skew normal distribution. *Journal of Multivariate Analysis, 89*, 181–190.

Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). Statistical analysis and modelling of spatial point patterns. Wiley, Statistics in Practice.

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

McQuarrie, D. A. (1976). *Statistical mechanics*. New York, NY: Harper & Row.

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

Møller, J., & Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. London: Chapman and Hall/CRC Press.

Neyman, Y., Scott, E. L.: A theory for the spatial distribution of galaxies. *Astrophysics Journal, 116*, 144–163 (1952)
Neyman, Y., & Scott, E. L. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society B, 20*, 1–43.
Neyman, Y., Scott, E. L. (1972). Processes of clustering and applications. In P. A. W. Lewis (Ed.), *Stochastic processes* (pp. 646–681). New York, NY: Wiley.
Neyman, Y., Scott, E. L., & Shane, C. D. (1953). On the spatial distribution of galaxies: A specific model. *Astrophysics Journal, 117*, 92–133.
Ohser, J. (1983). On estimators for the reduced second moment measure of point processes. *Statistics: A Journal of Theoretical and Applied Statistics, 14*, 63–71.
Penttinen, A., Stoyan, D., & Henttonen, H. (1992). Marked point processes in forest statistics. *Forest Science, 38*, 806–824.
R Core Team. (2019). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/
Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability, 13*, 255–266.
Ripley, B. D. (1988). *Statistical inference for spatial processes*. Australia: Cambridge University Press.
Stoyan, D., & Stoyan, H. (2006). Estimating pair correlation functions of planar cluster processes. *Biometrical Journal, 38*, 259–271.
Strauss, D. J. (1975). A model for clustering. *Biometrika, 63*, 467–475.
Tanaka, U., Ogata, Y., & Stoyan, D. (2008). A model selection and estimation of the Neyman–Scott type spatial cluster models. *Biometrical Journal, 50*, 43–57.
Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika, 36*, 18–25.
Wolfram Research. (2020). *Mathematica version 12.1*. Champaign, IL, USA (2020). https://www.wolfram.com/mathematica