Discussion of "Marked Spatial Point Processes: Current State and Extensions to Point Processes on Linear Networks"

Mari MYLLYMÄKI, Tuomas RAJALA, and Aila SÄRKKÄ

We congratulate the authors for their comprehensive review of the statistical summary characteristics for marked spatial point processes. These characteristics are valuable for understanding spatial dependencies between points, marks, and points and marks. It is essential that summary characteristics are developed and made available for more general marks, such as object-valued marks, and for non-Euclidean spaces, such as linear networks.

To add to the already impressive list of summary characteristics, we would like to mention the work Rajala and Illian (2012), where they studied indices for patterns with a high component count (e.g. rainforest data), including the mingling function mentioned here. There are also some third-order point process characteristics available for multitype point processes (Ayala and Simó 2020; Comas et al. 2010). Moreover, the authors call for characteristics to inspect local behaviour of marks. We note that the mark sum measures, as considered in Illian et al. (2008) and Myllymäki (2009), may be viewed as simple local characterisations of the potentially spatially varying mark distribution.

Marked summary characteristics play a pivotal role in the preliminary data analysis of marked point patterns. To interpret the information contained in a chosen characteristic, it is necessary to compare the empirical characteristic to its counterpart under a specific hypothesis, such as random labelling or random superposition as considered by the authors. In addition to the analysis performed by the authors, formal statistical tests, e.g. the traditional deviation tests (Diggle 2013; Myllymäki et al. 2015) and the global envelope tests, which aid in interpreting the test results through a graphical interpretation (Myllymäki et al. 2017; Myllymäki and Mrkvička 2023), can be performed. Here, when testing hypotheses with complex statistical characteristics, it is worth bearing in mind that even the very simplest

This is a commentary to the article https://doi.org/10.1007/s13253-024-00605-1.

M. Myllymäki (✉)· T. Rajala, Natural Resources Institute Finland (Luke), Latokartanonkaari 9, 00790 Helsinki, Finland
(E-mail: *mari.myllymaki@luke.fi*).
A. Särkkä, Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 29, Number 2, Pages 399–404
https://doi.org/10.1007/s13253-024-00610-4

characteristics, such as the cross-K function, might require a surprisingly large amount of data (e.g. hundreds of points per component) to provide sufficient statistical power, especially in real-world systems of weak interactions such as diverse plant communities (Rajala et al. 2019).

Sometimes summary characteristics combined with a formal test of a particular hypothesis provide a sufficient answer to the scientific question at hand. In practice however, if some correlation between the marks is detected, it is often important to proceed from exploratory analysis to modelling the interaction using marked point process models. In the second part of our comment, Sect. 1, we will discuss and provide some examples of such models in the traditional Euclidean space. Even though a reasonably large amount of models can be found for marked spatial point patterns in the literature, it appears that applications of marked point processes on linear networks are still rare. While some marking models might generalise to linear networks rather straightforwardly, others might face significant challenges, e.g. when defining for Markov processes a reasonable neighbourhood relation that depends not only on distance between points but also on marks.

1. MARKED POINT PROCESS MODELS

If a random superposition of two or more components of the point process is a reasonable assumption, each component can be modelled separately using unmarked point processes. Furthermore, under the random labelling assumption, the unmarked point pattern could be modelled first and then, the mark for each point could be drawn from a mark distribution independent of the points and of other marks. We note that the random labeling hypothesis and model are valid alternatives both for qualitative and quantitative marks (e.g. Grabarnik et al. 2011). Some examples of models are given below for the cases of qualitative and quantitative marks where there may be correlations between the marks or between the marks and points.

We would like to emphasise that marks can be either observed properties of the studied objects, such as the species or height of a tree, or other type of measurements made at the object locations such as concentration of some important substance in the soil (e.g. Stoyan and Penttinen 2000; Diggle et al. 2010). Furthermore, the marks can sometimes be constructed from the point pattern such as distance to the nearest neighbouring point (see e.g. Konstantinou et al. 2023).

1.1. QUALITATIVE MARKS

The two model classes that have frequently been used for multitype point patterns are the Cox and Markov point processes. The log Gaussian Cox processes (Møller et al. 1998), in particular, have gained popularity. These processes are advantageous when dealing with small-scale clustering of points that cannot be accounted for by the available spatial covariates. When it comes to multitype point patterns, the latent multivariate Gaussian field that describes the intensity is divided into random fields that are common to all point types and random fields that are type-specific (Waagepetersen et al. 2016). Another variant of the Cox process is the random set generated Cox process, which has been employed with set-marks (Penttinen and Niemi 2007; Myllymäki and Penttinen 2010).

In a Gibbs process (a special case of Markov processes), a potential function describes the interaction between the points. In the multitype case, a potential function is defined for each type of points to model the interaction within the type as well as the interactions between the types. Simple pairwise interaction models for multitype forest data were presented, for example, in Ogata and Tanemura (1985); Goulard et al. (1996); Reich et al. (2009); Stoyan and Penttinen (2000). However, pairwise interaction models are typically not suitable for clustered patterns; for clustered patterns, other Markov processes, such as Geyer's saturation process or the area-interaction process, have been used. The saturation process is similar to the well known Strauss process, a pairwise-interacting process with a weight on the number of close-by neighbours any point has. In the Strauss model, a positive weight leads to realisations with all points collapsed together, but the saturation process avoids this by setting additionally an upper limit to the neighbour-counts. In the area-interaction process, each point is replaced by a disc and the strength of interaction is measured by the area covered by the discs. Multitype versions of the saturation process have been suggested in Eckel et al. (2009), Lee et al. (2017) and Rajala et al. (2018), and of the area-interaction model e.g. in Picard et al. (2009). In Rajala et al. (2018), for example, first order non-stationarity, within-type point-to-point interactions and cross-type point-to-point interactions were all modelled simultaneously.

Sometimes, there can be a hierarchical structure between the different types of points. For example in forestry, large trees affect the locations of small trees but not vice versa. Different summary characteristics may not be able to give information on the hierarchy but it can be included in the modelling approach. Hierarchical multitype Gibbs models with non-symmetric interaction between different types of points were introduced by Högmander and Särkkä (1999) and Grabarnik and Särkkä (2009).

1.2. QUANTITATIVE MARKS

If the marks are correlated but independent of point locations, they may be described through geostatistical marking (Mase 1996; Schlather et al. 2004). Here, the marks are drawn from a random field $\{U(s)\}$ that is independent of the point process $\check{X}$: the marks are simply $m(x_i) = U(x_i)$ for all $x_i \in \check{X}$. Due to the independence of $\check{X}$ and marks, the marks are a representative sample from $\{U(s)\}$ and properties of $\{U(s)\}$ can be deduced from the marks using standard geostatistical techniques (e.g. Chilés and Delfiner 1999). Ho and Stoyan (2008); Myllymäki and Penttinen (2009) and Diggle et al. (2010) proposed marked point process models with intensity-dependent marks in a similar 'marking' manner: All proposed models assumed a log Gaussian Cox process as the unmarked point process $\check{X}$, having the random intensity $\Lambda(s) = \exp\{Z(s)\}$ for $s \in \mathbb{R}^d$ and a stationary Gaussian random field $\{Z(s)\}$. The mark of point $x \in \check{X}$ was then considered to stem from the conditional mark distribution of $m(x)$ given $\Lambda(x)$ (or $Z(x)$), either the mean (Ho and Stoyan 2008; Diggle et al. 2010) or both the mean and the variance of the distribution (Myllymäki and Penttinen 2009) depending on $\Lambda(x)$. Illian et al. (2012) further used similar types of models for two different quantitative marks, and Malinowski et al. (2015) presented a generalisation based on a bivariate Gaussian random field, with one of the two components driving the intensity and the other used to construct the marks. As an alternative to conditional marking, Myllymäki (2009) suggested using point process transformations such as thinning to generate intensity-dependent marks.

Pairwise Gibbs point processes have also been used as a model for point patterns with continuous marks. Ogata and Tanemura (1985) were among the first to suggest such models. They included the marks in the model by dividing the distance between two points by some function of the marks connected to the points. A similar type of construction can be found in Reich et al. (2009). A generalisation of the Strauss process, called Strauss disc process, was given in Goulard et al. (1996). Each point is replaced by a disc, where the size of the disc describes the size of the object at that point, e.g. a diameter of a tree. Two points interact with constant strength if the discs connected to them overlap. Møller and Waagepetersen (2004) introduced a model similar to the Strauss disc model, where the interaction is not constant but depends on the area of overlap between two discs. Marked generalisations of the area-interaction processes, such as the quermass-interaction process, have also been introduced (Kendall et al. 1999).

Funding Open access funding provided by Natural Resources Institute Finland.

[Accepted February 2024. Published Online March 2024.]

REFERENCES

Ayala G, Simó A (2020) Measuring interaction in bivariate point patterns with applications. Preprint No. 30, Mathematics Department, Universitat Jaume I

Chilés J-P, Delfiner P (1999) Geostatistics: modeling spatial uncertainty. Wiley, New York

Comas C, Mateu J, Särkkä A (2010) A third-order point process characteristic for multi-type point processes. Stat Neerl 64(1):19–44

Diggle PJ (2013) Statistical analysis of spatial and spatio-temporal point patterns, 3rd edn. CRC Press, Boca Raton

Diggle PJ, Menezes R, Su T-L (2010) Geostatistical inference under preferential sampling. J R Stat Soc: Ser C (Appl Stat) 59(2):191–232

Eckel S, Fleischer F, Grabarnik P, Kazda M, Särkkä A, Schmidt V (2009) Modelling tree roots in mixed forest stands by inhomogeneous marked Gibbs point processes. Biom J 51:522–539

Goulard M, Särkkä A, Grabarnik P (1996) Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. Scand J Stat 23:365–379

Grabarnik P, Myllymäki M, Stoyan D (2011) Correct testing of mark independence for marked point patterns. Ecol Model 222(23–24):3888–3894

Grabarnik P, Särkkä A (2009) Modelling the spatial structure of forest stands by multivariate point processes with hierarchical interactions. Ecol Model 220(9–10):1232–1240

Ho LP, Stoyan D (2008) Modelling marked point patterns by intensity-marked Cox processes. Stat Probab Lett 78(10):1194–1199

Högmander H, Särkkä A (1999) Multitype spatial point patterns with hierarchical interactions. Biometrics 55(4):1051–1058

Illian J, Penttinen A, Stoyan H, Stoyan D(2008). Statistical analysis and modelling of spatial point patterns (1 ed.). Statistics in Practice. Chichester: Wiley

Illian JB, Sørbye SH, Rue H (2012) A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). Ann Appl Stat 6(4):1499–1530

Kendall W, van Lieshout M, Baddeley A (1999) Quermass-interaction processes: conditions for stability. Adv Appl Probab 31:315–342

Konstantinou K, Ghorbanpour F, Picchini U, Loavenbruck A, Särkkä A (2023) Statistical modeling of diabetic neuropathy: exploring the dynamics of nerve mortality. Stat Med 42(23):4128–4146

Lee A, Särkkä A, Madhyastha TM, Grabowski TJ (2017) Characterizing cross-subject spatial interaction patterns in functional magnetic resonance imaging studies: a two-stage point-process model. Biom J 59(6):1352–1381

Malinowski A, Schlather M, Zhang Z (2015) Marked point process adjusted tail dependence analysis for high-frequency financial data. Stat Interface 8(1):109–122

Mase S (1996) The threshold method for estimating total rainfall. Ann Inst Stat Math 48(2):201–213

Møller J, Syversveen AR, Waagepetersen RP (1998) Log Gaussian Cox processes. Scand J Stat 25(3):451–482

Møller J, Waagepetersen R. P(2004). Statistical inference and simulation for spatial point processes (1 ed.). Chapman & Hall/CRC, Boca Raton

Myllymäki M (2009). Statistical models and inference for spatial point patterns with intensity-dependent marks. Ph.D. thesis, University of Jyväskylä, Jyväskylä

Myllymäki M, Grabarnik P, Seijo H, Stoyan D (2015) Deviation test construction and power comparison for marked spatial point patterns. Spatial Stat 11:19–34

Myllymäki M, Mrkvička T (2023) GET: global envelopes in R. arXiv:1911.06583 [stat.ME]

Myllymäki M, Mrkvička T, Seijo H, Grabarnik P, Hahn U (2017) Global envelope tests for spatial processes. J R Stat Soc: Ser B (Stat Methodol) 79(2):381–404

Myllymäki M, Penttinen A (2009) Conditionally heteroscedastic intensity-dependent marking of log Gaussian Cox processes. Stat Neerl 63(4):450–473

Myllymäki M, Penttinen A (2010) Bayesian inference for Gaussian excursion set generated Cox processes with set-marking. Stat Comput 20(3):305–315

Ogata Y, Tanemura M (1985) Estimation of interaction potentials of marked spatial point patterns through the maximum likelihood method. Biometrics 41(2):421–433

Penttinen A, Niemi A (2007) On statistical inference for the random set generated Cox process with set-marking. Biom J 49(2):197–213

Picard N, Bar-Hen A, Mortier F, Chadoeuf J (2009) The multi-scale marked area-interaction point process: a model for the spatial pattern of trees. Scand J Stat 36(1):23–41

Rajala T, Illian J (2012) A family of spatial biodiversity measures based on graphs. Environ Ecol Stat 19(4):545–572

Rajala T, Murrell DJ, Olhede SC (2018) Detecting multivariate interactions in spatial point patterns with Gibbs models and variable selection. J R Stat Soc: Ser C (Appl Stat) 67(5):1237–1273

Rajala T, Olhede SC, Murrell DJ (2019) When do we have the power to detect biological interactions in spatial point patterns? J Ecol 107(2):711–721

Reich R, Bonham C, Metzger K (2009) Modeling small-scale spatial interaction of shortgrass prairie species. Ecol Model 101:163–174

Schlather M, Ribeiro Jr PJ, Diggle PJ (2004) Detecting dependence between marks and locations of marked point processes. J R Stat Soc Ser B (Stat Methodol) 66(1):79–93

Stoyan D, Penttinen A (2000) Random recent applications of point process methods in forestry statistics. Stat Sci 15:61–78

Waagepetersen R, Guan Y, Jalilian A, Mateu J (2016) Analysis of multispecies point patterns by using multivariate log-Gaussian Cox processes. Appl Stat 65:77–96