Discussion of the Paper by Eckardt and Moradi

Dietrich STOYAN

I appreciate the editor's invitation to discuss Eckardt and Moradi's paper and to delve into the statistics of marked point processes, a field of spatial statistics in which I have been engaged for years. Although I am not as active in this field as I used to be and may not be aware of its latest developments, I will approach this discussion as an experienced user of the relevant statistical methods. My tendency is to question the 'why' when analyzing data.

A crucial inquiry in utilizing statistics for marked point processes is determining how the marks or marked points were generated and which models can be employed to comprehend this process. The paper lacks a thorough discussion of this issue. As we are all aware, the simplest scenario is independent marking, which the authors use as a null model. A more complex model is the random field model, where the marks come from an independent random field and are therefore spatially correlated. Maybe it could be an improved null model for the second example with the street trees. Even this simple model is not considered by the authors. Are there really no other interesting modern models for marked point processes?

I noticed the idea of labeling the protein points with the values of their local intensity. This reminds me of a model considered in Wälder and Stoyan (1997), where, for a homogeneous Poisson point process, the number of points in a circle around a point x was used as its mark $m(x)$. Of course, it is not at all surprising that for intensity marking the function $\kappa_{mm}(r)$, as used by the authors, takes large values (for small r), since in intensity marking large marks appear where the points are dense and close together.

Equation (12) with the definition of $\kappa_{t_f}(r)$ caught my interest. Why is there a κ? The books of Baddeley et al. (2015), Illian et al. (2008) and Wiegand and Moloney (2014) use k instead. Why this change that might confuse beginners?

I had a look at old papers and found a $\kappa(r)$ in Stoyan (1984) for the non-normalized version of $k_{mm}(r)$. However, already in Penttinen and Stoyan (1989) there is then the usual notation with $k_f(r)$. By the way, I would rewrite equation (12) as

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

D. Stoyan (✉) Technische Universität Bergakademie Freiberg, Freiberg, Germany.
(E-mail: *stoyan@math.tu-freiberg.de*).

Journal of Agricultural, Biological, and Environmental Statistics
https://doi.org/10.1007/s13253-024-00608-y

Published online: 26 March 2024

$$k_{t_f}(r) = \frac{\mathbb{E}[t_f(m(x), m(y)|x, y \in X \text{ with } d(x, y) = r]}{c_{t_f}}. \tag{1}$$

It is potentially interesting that the authors present a large collection of mark-correlation functions, perhaps all of them which have been used so far. However, for my taste, this is only half the job; in a paper like the one discussed here, one expects that these functions are compared with respect to their ability to describe mark correlations. Perhaps two well-chosen (maybe artificial) point-pattern examples could have been used to compare the forms of all these functions. This would have helped newcomers to point process statistics to find the right correlation functions for their applications. Also the corresponding recommendations of Illian et al. (2008) and Baddeley et al. (2015) could be commented. I noticed that the two examples in the paper use exactly the two functions I had preferred, $K_{ij}(r)$ and $k_{mm}(r)$.

The two examples are disappointing to me as examples, they look like statistical finger exercises. In both cases, I miss the scientific question that led to the analysis. In the case of the street trees it could be: How can we help the city gardener to plant trees in an optimal way?

In the protein example, I miss an explanation of why a change in the indicated distributional behavior occurs at $r = 400$ when using the K function, but at $r = 200$ when using the J function. The results for the street tree example made me doubt. My experience with forest statistics tells me that interaction between trees usually ends at distances of 10 or 20 ms. I cannot believe that street trees interact over distances of kilometers. If I had made the analysis and observed the strange result for the *Populus* trees at $r = 50$ m, the rather small value of $\kappa_{mm}(50)$, I would have considered all tree pairs with an inter-tree distance of around 50 m in order to find an explanation for this strange value. To me the result smells like a statistical artifact, perhaps caused by a single unusual tree. Finally, also the planting dates of the trees may be available and could provide helpful information.

Yes, it was a big event when Brian Ripley introduced the K function for planar point processes. But we should not forget that physicists had introduced the pair-correlation function for three-dimensional data decades earlier, see Illian et al. (2008), p. 226. Modern point process statistics prefers whenever possible (if the number of points is large enough) to use the pair-correlation function instead of the K function, which may be used in goodness-of-fit tests.

A warning at the end. The authors write in the Introduction that the mark variogram $\gamma_m(r)$ has similarities to the (semi-)variogram commonly used in geostatistical contexts". However, the mark variogram $\gamma_m(r)$ is a characteristic of a conditional nature (it considers as $k_{mm}(r)$ the marks of the members of a point pair under the condition that their distance is r), while the geostatistical variogram has something of the nature of a correlation function of a random field and is therefore harmless. Furthermore, the mark variogram can take forms that are impossible for geostatistical variograms, as shows already the paper Wälder and Stoyan(1996). However, $\gamma_m(r)$ is indeed a valuable summary property for marked point processes. More can be said than the colorless words in the discussed paper: "...which can provide valuable insights into the spatial relationship and variability of marks within a given range of distances". It helps to determine the range of correlations of the marks, and for small distances between points, it helps to find out the degree of similarity of the marks.

The paper by Bonneau and Stoyan (2022) demonstrates this and recommends a parallel use of $\gamma_m(r)$ and $k_{mm}(r)$. Furthermore, that paper gives an example of the use of directional marks in a study of patterns of geological fracture networks.

Funding Open Access funding enabled and organized by Projekt DEAL.

[Accepted February 2024.]

REFERENCES

Bonneau F, Stoyan D (2022) Directional pair-correlation analysis of fracture networks. J. Geophys. Res. Solid Earth 127:e2022JB024424

Stoyan D (1984) On correlations of marked point processes. Math. Nachrichten 116:197–207

Wälder O, Stoyan D (1997) Models of markings and thinnings of Poisson processes. Statistics 29:179–202