

Non-parametric analysis of spatial and spatio-temporal point patterns

Point-by-point answers to the reviewer's comments

General comment

The authors present an example analysis of spatio-temporal point pattern data in R. While there are many different tools to analyze spatio-temporal point data, the authors argue that there is not clear, practical examples of how to perform such an analysis. The authors do a good job of communicating spatio-temporal point process models and motivating what they are attempting to implement with their code. The authors guide the reader through some typical scenarios one might encounter when modelling spatio-temporal data. Overall, the authors present a good tutorial for exploratory spatio-temporal point-process modelling.

I want to thank the authors on their improvements to the manuscript. The current version is much easier to read and follow. In addition, the authors have taken the time to be more precise with their language and to think more about what the reader may or may not already know which improves both the readability and the clarity of communication with the intended audience (those that are interested in point process models but don't know how to fit them). The addition of approximate runtime of the code is also a nice addition that helps the user manage expectations for how long one should wait. I believe the code is much better than the previous submission; however, I can't run the code without the data (`AegissData.RData` and `Results.RData`), and thus the code is not reproducible. This should be a minimum requirement for an R journal submission, especially an article that is purporting to be a tutorial for an applied user. I was able to use `AegissData.RData` from the prior submission, but this isn't a solution. Perhaps the data objects are large, but these should be able to be hosted online (perhaps GitHub or on a file-sharing site like Google Drive) for download if need be, as I don't think the data should be that large.

In addition, the Discussion needs a minor revision. Currently, the discussion reads as four unconnected paragraphs and, at times, is either an introduction to the paper (first half of paragraph 3) or only tangentially related to the bulk of the manuscript. What did you accomplish and what value have you added to other users who want to implement these models. Answer these questions, and the discussion will be more cohesive.

Authors answer

We are very grateful for the detailed report the referee provided us on this occasion, which allowed us to improve our manuscript further.

In the submission, together with the manuscript and the R code, we are attaching the file that contains the data we describe and analyse: `AegissData.RData`. All the methods and code in the paper apply to this dataset. Therefore, loading the dataset as a starting point is required to reproduce the analyses. `AegissData.RData` includes two fundamental variables: the point pattern of gastrointestinal disease

reports and the images of the population in the years 2001, 2002 and 2003. This file only occupies 238 KB of memory.

We are also attaching the file `Results.RData` responding to the referee's requirement to provide the variables that take longer to be calculated in a single document. It can facilitate the reproducibility of the codes in case the reader does not have time to wait for all routines. If this file is not loaded, everything will have to be computed, which will take more than 6 hours on a conventional laptop, just as we indicated in the discussion.

Finally, we have rewritten the discussion. We have tried to keep our original ideas but connect them better with what we have done in the paper and follow a guideline that allows the reader to understand what they could do with what we offer in our work.

Specific comments

1. **Reviewer:** Page 2: the `foreach`, `GET`, and `doParallel` packages are in a different font.

Authors: These packages are in different fonts as the `foreach` package is not a CRAN package; it is a package included in R, whereas the other two are packages available on CRAN. Therefore, we used different template functions (`\pkg{}`, `\CRANpkg{}`) for referencing them.

2. **Reviewer:** Page 8: “Usually, this distribution is assumed in quite simple possible ways” – What are you trying to say here? Are you trying to say that a simplifying assumption of separability is usually made?

Authors: We have rewritten the sentence: “Usually, this distribution is simplified by assuming, for example, separability.”

3. **Reviewer:** Page 8: “where $\lambda_1()$ and $\lambda_2()$ are non-negative functions”
– Add “of space and time, respectively”.

Authors: Done.

4. **Reviewer:** Page 8: “...it can be tested to gain a fast preliminary perception by a simple χ^2 -test...”
– Delete “gain a fast preliminary perception” as this is unclear. You are just doing a hypothesis test, so stick with that.

Authors: Done.

5. **Reviewer:** Page 9: [Ghorbani et al. \(2021\)](#) reference should be in parentheses.

Authors: There was a typo in this sentence that prevented the proper display of the reference; now it is: ([Choi and Hall, 1999](#); [González et al., 2019](#); [Ghorbani et al., 2021](#)).

6. **Reviewer:** Page 9: 1047 intensity measurements? Earlier in the paper, you list 1095 measurements. Where are some of these duplicates/missing? Can you be more clear here?

Authors: We note that there are 1047 different times v_i , as said in this paragraph, i.e., 1047 in the cardinal of the temporal set. However, the values of that time points range in the interval

[0, 1095], i.e., there are some temporal instants where there were no disease reports. Therefore, we added this sentence: “Note that 1047 is the size of the temporal coordinates set and not the maximum temporal value (1095).”

7. **Reviewer:** Page 9: (Figure 4) - missing right parenthesis; same for Figure 2 (left).

Authors: Fixed.

8. **Reviewer:** Page 9: second to last paragraph discussing Figure 4: Are the intensity estimates really much lower, or are they just plotted on a different scale in Figure 4 than in Figure 2? Why are the scales off by orders of magnitudes/scaling of units? This needs to be made clear if you are making this comparison here.

Authors: We have removed the logarithm scale of Figure 2 (left) to facilitate the comparison. In addition, we have added the following sentence to the interpretation: “In the spatio-temporal case (Figure 7), we can see a behaviour in the intensity of cases much lower than in the spatial case (Figure 2 (left)) due to the temporal granularity. This scale difference is due to the number of events per unit of time, which is not considered in the only spatial estimation. When considering the temporal dimension, the total number of points must be divided by the number of bins in the temporal grid where the estimation is made. Therefore, there are fewer points per unit area at each time than only when spatial coordinates are considered.”

9. **Reviewer:** Page 11: In the last code chunk, the indentation seems misaligned. Double-check this.

Authors: Fixed.

10. **Reviewer:** Pages 13 and 14: Figures 6 and 7 are not referred to in the text. Re-write the text to make sure each figure is referred to in the main document.

Authors: We included both figure references in the text.

11. **Reviewer:** Page 13: Is $\hat{g}(r, t)$ shown in Figure 6?

Authors: Yes, it is. We have modified the paragraph: “The $\hat{g}(r, t)$ surface, shown in Figure 6, describes the spatio-temporal interaction structure of the disease... ”

12. **Reviewer:** Page 13: “ $K(r, t)$ is simply the expected number of further points”
– change “further” to “additional”. “Further” can imply distance, which can be confusing as distance is a part of the equation. If you really mean “additional” or “more than expected”, further is potentially confusing.

Authors: We understood the potential confusion and changed “further” by “additional”.

13. **Reviewer:** Page 13-14: No reference to Figure 7. I assumed this is $\hat{K}(r, t)$, but this isn’t clear.

Authors: We have added a reference to Figure 7.

14. **Reviewer:** Page 14: the last paragraph before the discussion needs 2-3 more sentences of explanation.

Authors: We have expanded the explanation: “Figure 7 shows the estimated K -function. Since the K -function is centred on its theoretical value ($2\pi r^2 t$), we can easily interpret it; positive deviations suggest clustering, while negative deviations represent regularity or inhibition. In our case, we always appreciate positive deviations that become larger as the spatial and temporal distances grow together. Therefore, as time passes and spatial distances become larger, the level of clustering also increases in our point pattern, at least in the range we have considered, that is, up to 7km in space and up to 5 months in time.”

Comments on code/script

1. **Reviewer:** Missing the datasets.

Authors: The dataset is already included in the manuscript material.

2. **Reviewer:** The code:

```
bwG <- bw.pcf(X, cv.method = "compLik", divisor = "d", lambda = MP)
```

throws 23 warnings. Why? Mention this as either a comment or in the text if this is expected.

Authors: We had a comment for these warnings already included in the text just after the code line: “Note that this function can warn of undetermined contributions to the pair correlation. These caveats come from the divisor, which, when too small, can conflict with the numerical tolerance of R and eventually lead to indeterminacy.”

References

- Choi, E. and Hall, P. (1999). Nonparametric approach to analysis of space-time data on earthquake occurrences. *Journal of Computational and Graphical Statistics*, 8:733–748.
- Ghorbani, M., Vafaei, N., Dvořák, J., and Myllymäki, M. (2021). Testing the first-order separability hypothesis for spatio-temporal point patterns. *Computational Statistics & Data Analysis*, 161:107245.
- González, J. A., Hahn, U., and Mateu, J. (2019). Analysis of tornado reports through replicated spatiotemporal point patterns. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(1):3–23.