



Journal of Statistical Software

December 2016, Volume 75, Book Review 2.

doi: 10.18637/jss.v075.b02

Reviewer: Virgilio Gómez-Rubio
Universidad de Castilla-La Mancha

Spatial Point Patterns: Methodology and Applications with R

Adrian Baddeley, Ege Rubak, Rolf Turner
Chapman & Hall/CRC, Boca Raton, 2015.
ISBN 978-1-4822-1020-0. 810 pp. USD 99.95 (H).
<http://www.spatstat.org/>

A spatial pattern records the locations of events produced by an underlying spatial process in a study region. Examples include the location of trees in a forest, cases of a disease in a country, and many more. This book describes how to perform an analysis of a spatial point pattern with R and focuses on the **spatstat** package that has been developed by the authors for more than 20 years.

The book aims at linking statistical theory and methods for the analysis of spatial point patterns to mainstream statistical analysis, e.g., summary statistics, model fitting, model assessment, analysis of the residuals, etc. For many years, the analysis of point patterns has been conducted in a way that allowed specific methods to be developed, but the authors' focus is on linking these methods to more typical statistical approaches. The book is structured in four separate blocks (*Basics*; *Exploratory data analysis*; *Statistical inference*; and *Additional structure*) that describe the different steps in a typical data analysis (not necessarily of a point pattern).

In Chapter 1, an introduction to spatial point patterns and a summary of the book is presented. The authors make an excellent job of introducing the different types of spatial point patterns and their main differences. They also lay out the different steps required in any analysis of spatial point patterns and how this should be done by proposing and testing hypotheses. In particular, they clearly explain why models should be at the core of any statistical analysis.

Chapter 2 is devoted to giving a short introduction to the R language and the **spatstat** package. Although the introduction to the R language is short, it highlights some of the topics required in the introduction to **spatstat**, such as data types, formulas, etc. The introduction to **spatstat** shows some short examples on the use of the package for the analysis of a spatial point pattern, and then describes the different classes of objects implemented in the package. This is important as they are closely related to the different formats in which spatial data are available and the output generated by the functions in **spatstat**.

In Chapter 3, the authors provide some important ideas on how the data collection should be done. They stress that not only the presence of points is informative, but also the absence

of them, and how gathering data about the covariates where the observed points are absent is crucial in order to make a valid statistical analysis. They also state the importance of correcting for bias and recording missing data, as these two situations can be taken into account with appropriate statistical modeling. The authors also provide important hints on how to enter spatial data in R and **spatstat**, and provide a detailed description on the main data types in the package, such as **ppp** (planar point pattern), **owin** (observation window) and **im** (pixel image), among others. Finally, they describe how to import data in several GIS formats, by first importing them to **sp** classes and then into one of the **spatstat** formats. This is important because the main packages to import GIS data (in particular, **rgdal** and **mapproj**) will automatically create one of the classes in the **sp** package.

Once the data are in R, the next step is to visualize and explore them. This is described in Chapter 4. **spatstat** implements a number of functions to visualize and create plots from spatial data. This includes point patterns, windows (i.e., polygons) and pixel images, as well as layered plots by combining some of these data types in a single plot. Several functions to provide summary statistics on the spatial data are also available in the package. The authors describe several functions to manipulate point patterns, polygons and images, and how to obtain new objects from them by subsetting the original data, producing geometrical alterations (such as, shifting, etc.). For point patterns, visualization involves displaying the locations and, possibly, associated marks. For windows, several polygon operations are described to produce new windows, and how to subset the point pattern to a redefined boundary. Operations on pixel images are also described and how to produce heat maps with them. At the end of the chapter, the authors introduce tessellations and discuss how to create them in a number of ways.

An introduction to the mathematical foundations of point processes is given in Chapter 5. Homogeneous and inhomogeneous Poisson processes and their properties are described here avoiding unnecessary mathematical complexity, followed by other simple models for point processes such as several simple Matérn processes. A discussion on the goals of analysis ensues, with an interesting discussion on how to differentiate between inhomogeneous intensity and interaction in the analysis of a point process. Given that this chapter introduces basic concepts, the novice reader would have preferred to see this earlier in the book, but I must say that I really enjoyed going through it as the contents are clearly presented.

This makes the first part of the book on the basics of point pattern data analysis. The next block of the book covers exploratory data analysis by means of the intensity and interaction between points.

Chapter 6 describes several ways to estimate the intensity (i.e., the average number of points per unit area) using non-parametric methods. This includes estimating the intensity directly from the observed points and/or covariates. Among all non-parametric methods, kernel smoothing is the most widely used, and important advice is given on how to choose the best bandwidth and provide estimation of the intensity at the observed points. When using covariates to estimate the intensity of a point pattern, it is necessary to test whether there is any significant dependence on the covariate. This is also discussed, and several tests are shown. Because of the actual difficulty in differentiating inhomogeneous intensity from clustering, the last part of the chapter is focused on detecting anomalies in the intensity, that could lead to hot-spots or unusual aggregation or dispersion. Finally, the chapter shows how to provide spatial smoothing on the marks attached to the point pattern, which is useful to show the spatial variation of the marks.

Although the intensity of a point pattern gives a description of the spatial variation of a point pattern, it says nothing about how points depend on each other, or how they are correlated. Investigating dependence between points is addressed in Chapter 7. After introducing simple manual methods to assess correlation, the authors describe Ripley's K -function, which is probably the most widely used method to assess dependence in a point pattern. The L -function, a transformation of the K -function, is also discussed, as well as different edge corrections. As the output of an estimation of these functions is an object of class `fv` (a function object in **spatstat**) the authors explain how these objects work, as they are used to store estimates of, for example, the K -function for different edge corrections that are often plotted together in a single plot. These plots can be used to assess clustering or regularity in a point pattern. The pair correlation function, a measure of spatial correlation between points, is also discussed. Estimation of standard errors and confidence intervals for the K -function are introduced to compute confidence bands that can be used to assess departures from complete spatial randomness. When correlation among points occurs in a given direction, the process is anisotropic, and this can be detected by using a version of the K -function and the pair correlation function. Finally, the authors explain how to account for inhomogeneous intensity when assessing dependence between points, as most of the methods to detect correlation will be sensitive to inhomogeneous patterns.

As stated by the authors, a point pattern provides information in two ways: the location of the observed points and the locations where no points have been observed. Both are crucial in the analysis of point patterns, and this is why the authors devote Chapter 8 to the analysis of spacing or shortest distances between points in the pattern. Here, the authors primarily focus on different functions that measure distance from a point to its nearest neighbor (the G -function) and distance from an arbitrary point in the observation window to the nearest observed point (the F -function), and the derived J -function (that is equal to 1 for an homogeneous Poisson point process and provides yet another baseline to assess complete spatial randomness). They discuss how these functions can be estimated accounting for edge corrections and inhomogeneity, and how envelopes and tests can be performed to assess complete spatial randomness, and identify clustering or regularity.

In its third block, the book moves to issues related to statistical modeling. Chapter 9 describes Poisson models and how they can be fitted with **spatstat**. After describing Poisson processes and their properties in more detail than in Chapter 5, the authors describe how to fit Poisson models using the `ppm` function. To me, this is one of the many wonders in the package: a function that behaves like any other function to fit models (such as, for example, `glm`) but to fit a point pattern to a point process! The model can be specified via a formula, to include the point pattern and covariates to be used to estimate the intensity, so that different types of models can be built in a very flexible way. The output returned by `ppm` can then be used to provide a summary of the fitted model, plots, and even predictions of the intensity. Generic functions to extract information from the fitted model (such as, estimated coefficients, standard errors, log-likelihood, etc.) are also provided in **spatstat**. Different fitted models can be compared using an analysis of the deviance or AIC, and simulating from a fitted model is also implemented. Hence, the authors have successfully developed a model fitting framework for spatial processes that follows other mainstream functions for model fitting. The computational tools used in the implementation of the `ppm` function are explained by the authors, but I believe that this will be advanced material for many readers. Finally, conditional logistic regression, and model fitting using approximate Bayesian inference, profile

likelihood (for non-loglinear models) and local likelihood are discussed. All these approaches have specific functions for model fitting, but they are very similar to `ppm`.

After model fitting, Chapter 10 covers a number of tests for statistical significance of the model components and computation of simulation envelopes for hypothesis testing. The focus is now on comparing models (using, for example, a likelihood ratio test and others) as well as testing for the significance of individual covariates in the model. The authors have included an interesting discussion on how these tests should be proposed, executed and their outcomes interpreted in the context of the analysis of point patterns. Monte Carlo tests are introduced here as well. Monte Carlo tests are important to test whether the observed pattern may have been generated under a given model such as, for example, complete spatial randomness or an inhomogeneous Poisson process with a known intensity. When the statistic used in the Monte Carlo test is a summary function on the point pattern, computing envelopes is useful to estimate the variation of the function under the null model. This is thoroughly explained, with an important discussion on the interpretation of pointwise and simultaneous (or global) envelopes. `spatstat` provides a very flexible framework to compute envelopes, in the sense that the null model can take many different forms and the function used in the test (and for which the envelopes are computed) can also be a very general function.

Chapter 11 considers validation of a fitted model. Here the reader will find a description of the different functions provided in `spatstat` to validate specific components of the model such as, for example, the fitted intensity, effects of the covariates and independence. It is worth noting that the authors devote a good deal of the chapter to the analysis of the residuals of Poisson processes, which has been a notable addition to the field, in which the authors have been involved. By having properly defined residuals in the context of point processes, the authors are able to link (again) the theory of point processes to the mainstream statistical analysis of model assessment and validation. They also discuss leverage and influence in the context of the analysis of point patterns, so that the impact of individual observations on the fitted model can be assessed. All over the chapter the authors have included details on the theoretical foundations of the concepts on residuals and leverage but, as already noted in the book, some readers will skip this.

Chapters 12 and 13 deal with Cluster and Cox models, and Gibbs models, respectively. These types of models are important because they can be used to model different types of dependence among points. Cluster and Cox models are particularly interesting to model positive dependence or clustering between points, and they can be fitted using function `kppm` (which works similarly as `ppm`). Gibbs processes can be used to model negative dependence or competition between points, and they can be fitted using the `ppm` function with an extra argument for the type of interaction. Hence, `spatstat` will be able to fit models where points are independent or they have different types of interactions. Simulation and prediction using these models is also covered in the book, as well as ways of validating and assessing the validity of the fitted models.

Despite the plethora of models and methods presented so far in the book, only point patterns with a single type of points have been covered in depth. Chapter 14 introduces the main features of multitype point patterns (e.g., patterns with different types of events). These are useful to model, for example, the location of different diseases in a region or different types of crimes in a city. When analyzing multitype point patterns the aim is often to assess whether the different types have been generated by the same point process and, if not, determine the different processes behind each point type. The standard methods for analysis of point

patterns are extended to the multitype case (e.g., estimation of different intensities for each point type), as well as testing for independence and dependence between the different point types. Fitting multitype Poisson, cluster, Cox and Gibbs models is also described here, as well as simulating from these models.

The last part of the book is about point patterns with specific structures that makes them more complex than the point patterns presented so far, and these should be considered more advanced topics. Chapter 15 describes point patterns in higher dimensions and multivariate marks. This includes space-time point patterns as a particular case. Chapter 16 studies point patterns that are the outcome of an experiment and that come as replications of the same experiment. Similarly to standard methods for the analysis of this type of data, available models include Poisson models, this time with random effects to take into account the possible differences between the replicates. Replicated data are particularly interesting for disentangling intensity and correlation between points. For this reason, the authors also discuss how the use of Gibbs models can benefit from the use of replicated data. However, the authors make it clear that different interaction mechanisms may be affecting the different replicates in the data. Finally, Chapter 17 deals with point patterns that occur on a linear network, for example, the locations of car accidents in the streets of a city. In this case, the structure of the network needs to be taken into account and properly stored and manipulated in R, for which **spatstat** offers several functions and data types. Estimation of the intensity in the linear network as well as Poisson models are covered. Dependence of the points in the network can be studied by the corresponding pair correlation function and several specific K -functions defined for this particular problem.

In a nutshell, this book covers a large portion of the methods for the analysis of spatial point patterns and their implementation in the **spatstat** package. I would have liked a larger portion on the analysis of spatio-temporal point patterns because of the increasing availability of this type of data. As **spatstat** has evolved with help from its users and the community, a list of frequently asked question (FAQ) is included at the end of most chapters. This will help to clarify some of the contents and guide the user in the data analysis by pointing at different important points to consider. The book is also full of tips, clarifications and discussions on how to conduct the analysis, which clearly will benefit practitioners. It presents and discusses many applications from different fields, so that it will be of interest to a wide range of researchers.

I really enjoyed reading this book and it has changed my views on **spatstat**. In addition to a package for the analysis of point patterns, I now regard this package as a toolbox that will allow the development of further methods and software for the analysis of point patterns, as the package provides a number of functions to rely on when developing new methods.

Reviewer:

Virgilio Gómez-Rubio
Department of Mathematics
Universidad de Castilla-La Mancha

Avda. España s/n
02071 Albacete, Spain
E-mail: Virgilio.Gomez@uclm.es
URL: <http://www.uclm.es/profesorado/vgomez/>