Week 8

Spatial Point Processes

[icon] Overview

Welcome to Week 5 of Spatial Analysis.

This week we will cover two topics:

- Monitoring Network Optimization

- Introduction to Spatial Point Processes

The first topic of this week concludes our studies of random field models. You will learn how to use kriging for monitoring network optimisation. Such spatial optimisation approaches are used to determine sites to be included or excluded from the set of locations where the data are collected. It is an important economic decision-making problem. We will find that different criteria can give different monitoring network configurations. You will practise these approaches with R for the Meuse dataset.

Then we will consider the second large class of models, spatial point processes. They are used to represent spatial data collected at random locations. You will see how to model these processes by using their intensity. We practise with one of the most popular R packages, spatstat, and formats, ppp, for such type of data. You will learn about the structure of R objects in this format and how to create ppp objects from other objects.

Finally, you will participate in a workshop, in which you can practise conducting an analysis of real spatial data.

By the end of this week, you will learn about:

Topic 1: Spatial Monitoring Network Optimization

- Introduction to monitoring network optimization.
- An example based on mean kriging variance.
- An example based on delineating contours.
- Applications to the Meuse data.

Topic 2: Introduction to Spatial Point Processes

- Introduction to spatial point processes.
- The R package spatstat.
- Investigating spatial intensity in R.
- Covariates in spatial analysis.
- Class of ppp objects.
- Converting between ppp and sp formats.
- Creating ppp objects from csv files.

By completing this module, you will be working towards the following subject-intended learning outcomes:

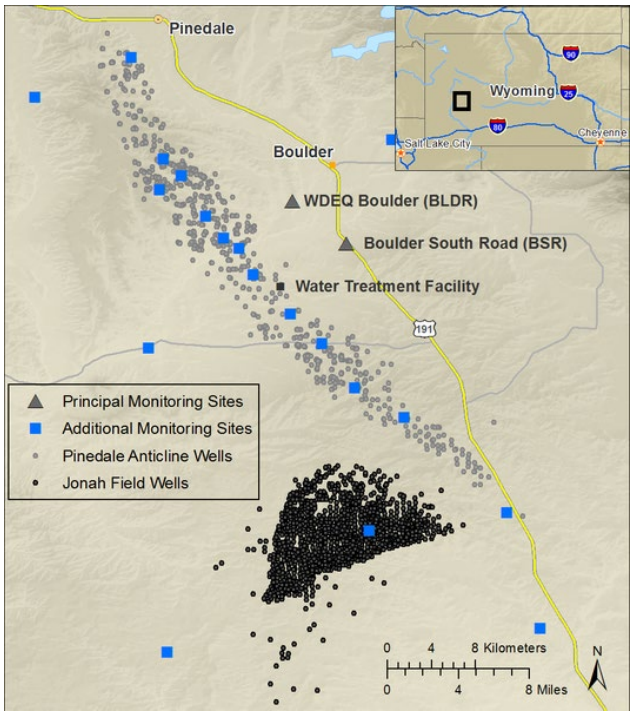
1. Formulate purposeful questions to explore new statistical ideas and subsequently design valid statistical experiments.
2. Present clear, well-structured analysis of important statistical model results.
3. Creatively find solutions to real-world problems consistent with those commonly faced by practising statisticians.
4. Professionally defend or question the validity of existing statistical analyses and associated evidence-based conclusions that are derived via application of sound spatial statistical methodology.

Topic 1: Spatial Monitoring Network Optimization

This is the last topic on random field models. You will learn and practise several methods of optimising spatial monitoring networks. We will use the kriging and the Meuse data. You will see how to apply different criteria to include or exclude sites in a spatial monitoring network.

Figure 5.1

Study area including monitoring sites and oil and gas wells



https://commons.wikimedia.org/wiki/File:Map-of-study-area-including-principal-monitoring-sites-and-locations-of-oil-and-natural-gas-wells-within-the-Jonah-Field.png

Upon completion of this topic, you will be able to decide which monitoring stations can be removed with a minimal impact on a quantity of interest, or where to add new monitoring stations to improve results. You will further develop your spatial computational skills using the kriging methods from R.

In this topic's examples, we consider the situation when one would like to determine and remove some monitoring stations that do not substantially change the specified quality criteria.

Example 1

A very simple common approach towards monitoring network optimization is to determine monitoring stations whose removal leads to the smallest increase in the mean of overall kriging variance. So, the mean-square error for the optimised small monitoring network will remain rather similar to the original full network. It can be done by the exhaustive search when we remove monitoring sites one by one and compute the mean kriging variance each time. The site with the smallest increase is the first candidate for removal. After that one can repeat the process again and find the next site, etc. until required.

Example 2

Then, we will find the optimal monitoring configuration to determine the lines with a specified level of data values. Such criteria are rather common when one explores ore richness, pollution levels, or areas with similar hose prices. To deal with such problems one can consider the normalised deviations from the specified level. We will use the CDF of the standard normal distribution and compare these deviations with the probability 0.5, which corresponds to the zero-argument of the CDF.

If one would like to add new monitoring sites, then a grid covering the region of interest can be used. The locations of the grid should be checked one by one using a specified criterion.

Read

In the reading Monitoring Network Optimization, you will explore several methods to optimise spatial monitoring networks depending on specified criteria. You will apply them to Meuse data.

From the folder, open the document titled **Week_8_Topic_1** and read **slides 1–8**.

Revise key R commands used in the first topic.

Read

In the reading Key R commands, you will revise some of the key R commands that were used in this topic's materials.

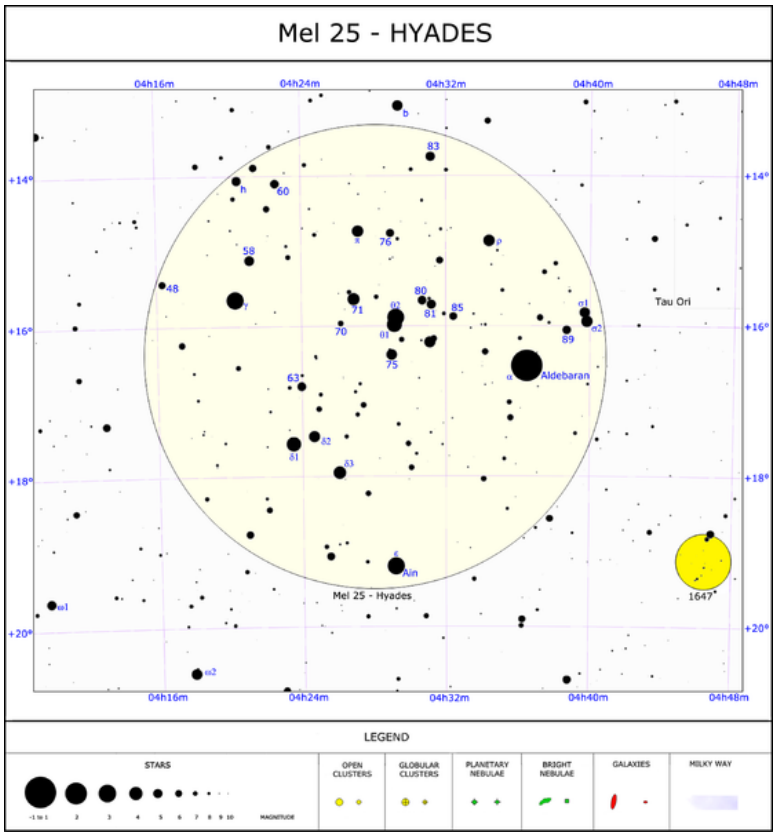
From the folder, open the document titled **Week_8_Topic_1** and read **slide 9**.

Topic 2: Introduction to Spatial Point Processes

This topic introduces spatial point processes. This statistical model is used to describe randomly located points in time or space. We will start by discussing the basic theoretical foundations of this model. Then we will learn several methods to investigate various properties of spatial processes with the R software. Finally, we will consider how to convert data to the ppp format, which is used in R to represent spatial point datasets.

Figure 5.2

Hyades star cluster



https://commons.wikimedia.org/wiki/File:Hyades_cluster_map.png

Spatial point processes and their intensities

First, in this part, we will introduce basic definitions and notations used to describe the statistical model called a spatial point process. Then, you will learn about the intensity of a spatial point process. This function is one of the main characteristics of a point process that describes the "local" density of its points. For stationary point processes, that have the same properties in all locations, the intensity is constant.

We will consider several methods to estimate the intensity from realisations of a point process. These estimates can be used as empirical evidence or refutation of stationarity.

We also will introduce the kernel estimator of intensities. This type of non-parametric estimators is very popular in applications as it does not require an explicit parametric model for points' distribution.

Finally, we will consider the modelling of the intensity by using covariates. Covariates are data that are treated as explanatory, rather than as part of the response.

All approaches will be illustrated by using the data bei about the locations of trees in a tropical rainforest.

Read

First, in the reading Spatial Point Processes, you will learn the definitions of the spatial point process and its intensity. Then you will practice estimating intensities with R packages spatstat and the bei dataset.

From the folder, open the document titled **Week_8_Topic_2** and read **slides 1–16**.

ppp objects

In this part, you will learn about the structure of ppp objects that are used in R to work with spatial point data. These objects include information about the locations of points and their attributes, i.e. measurements at these points. These attributes are called marks. You will learn how to work with ppp objects and their marks.

Then, you will study how to convert objects between ppp and sp formats. Finally, you will see how to create ppp objects from usual text files that contain spatial coordinates and corresponding measured values.

We illustrate these methods by using the data bei and the dataset with information about world cities.

Read

You will read the second part of Spatial Point Processes and will learn how to use the R object format ppp for representing spatial point data.

From the folder, open the document titled **Week_8_Topic_2** and read **slides 17–26**.

Revise key R commands used in the second topic.

Read

In the reading Key R commands, you will revise some of the key R commands that were used in this topic's materials.

From the folder, open the document titled **Week_8_Topic_2** and read **slide 27**.

Workshop

Activity

Workshop

This activity will be completed in R. Repeat the R programming content covered in Week 8. Modify the code and understand the impact of different R parameters on changes in results.

Your task

- Repeat R commands learnt in Week 8 prior to the two-hour workshop session.
- Try to modify the code and understand the impact and meaning of different R function parameters. Interpret the observed changes in plots and analysis results.
- Feel free to discuss questions with other students as you go in the forum, and please also take the time to help others. It is amazing how much we all can learn from each other's questions, and how in helping others we strengthen our own understanding.
- Come along to the two-hour scheduled workshop session and discuss any challenges, seek advice and work through some problems with your peers and facilitator.
- Revisit these problems in later weeks and challenge yourself to get a deeper understanding to build on what you learn later.

Guidelines

- This activity is not graded but is an essential part of your learning. It will be held synchronously and facilitated by your instructor.
- You don't need to submit your R code; however, to be successful in this subject it is necessary to work through all R coding materials from this week and understand how to apply the corresponding R commands.
- You should repeat all R commands in this week's materials before the workshop. This will give you an opportunity to efficiently work with the facilitator during the workshop and get your questions answered.
- You should spend around two hours on this activity.

Workshop solutions

The R code for Workshop is provided for your reference. Click this link to access the R code. From the folder, please open the document titled – **Week_8_Workshop_8_RCode**.

Summary

This week, we finished the studies of spatial models based on random field approaches. The last random field topic was the methods for optimising monitoring network configurations. Then, we considered another type of spatial data and the methodology based on point processes. You learned how to work with such data in R and investigate their properties. We practised these point processes methods with the spatstat package.

Next week, we will continue learning about other applications of the obtained results for several problems in spatial data analysis.

Here's a list of tasks that you should be working on or have completed:

- **Required readings**
- **Workshop**

The following resources provide you with this week's references and additional suggested readings.

Additional suggested readings and resources

While these readings and resources are not essential, they provide greater insight into the concepts covered in the week's lectures and give you the choice to enhance your learning or pursue an area of interest in greater detail.

Software and data:

- CRAN documentation for the gstat package: https://cran.r-project.org/web/packages/gstat/index.html
- CRAN documentation for the spatstat package: https://cran.r-project.org/web/packages/spatstat/index.html
- Meuse river data set https://rsbivand.github.io/sp/reference/meuse.html
- Tropical rain forest trees data http://127.0.0.1:22118/library/spatstat.data/html/bei.html

Theoretical concepts:

- Spatial point processes: https://www.apps.stat.vt.edu/leman/VTCourses/BaddeleyPointProcesses.pdf

Books:

- Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). Applied spatial data analysis with R (2nd ed.). Springer. https://doi.org/10.1007/s12061-014-9118-y Avalilable on-line in La Trobe EBL ebook Library