Week 4

Modelling Spatial Objects and Descriptive spatial statistics

Overview

Welcome to Week 4 of Spatial Analysis.

This week we will cover two topics:

- Random fields

- Descriptive spatial statistics

This week we will revise the basic theory of random variables and show how it can be generalised to more complete statistical models. Namely, we will discuss random fields and some related concepts, which will be used in the following weeks as the main statistical models for raster data.

You will continue this week learning about several methods of descriptive statistical analysis for spatial data. You will use the geoR package and get experience in producing several plots to visualise different properties of spatial data. Then, we will discuss the fitting linear models and the parametric approach to fitting spatial trends. You will continue practising with the Meuse dataset.

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:

Topic 1: Spatial models

- Revise the basic theory of random variables
- Lear about stationary time series
- Learn about random fields
- Explore the concept of positive definiteness and its properties

Topic 2: Descriptive spatial statistics

- Plotting with the package geoR.
- Basic analysis with the package sp.
- Visualising the Meuse data.
- Fitting spatial trends.

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.

Topic 1: Spatial models

In this topic, you will start with a short revision of a few key concepts from the probability theory. Then, you will learn some extensions of these concepts that are used for modelling temporal and spatial data. We will see applications and some specific examples of these models in practical analysis of spatial data in the following weeks.

Random variables

Random variables are random functions that variable that take numerical values for depending on possible outcomes of a statistical experiment. In spatial applications researchers use the following two most popular types of random variables: discrete and continuous.

A discrete random variable is defined by a list of its possible values and a list of its corresponding probabilities (its probability distribution). Most of the numeric properties of such variables can be computed via summations over a range of the values of interest.

A continuous random variable is defined by intervals of its possible values and a non-negative function defined on these intervals. The function is called a probability density and defines the “local” probability of each possible value. Most of the numeric properties of these variables can be computed via integrations over a range of values of interest.

The two numerical values which are of main interest to us are the expected value and the variance. The expected value characterises the center of distribution and the variance is used to measure deviations from this center.

Read

First, in the reading Random fields, you will revise basic probability theory.

From the folder, open the document titled **Week_4_Topic_1** and read **slides 1–5**.

Time Series

A collection of random variables indexed by time is called a time series. Their expectations are used to represent trends in data and covariance/correlation functions describe dependencies of observations at two time moments.

The class of time series is very big, and it is difficult to study its general properties. Introducing its specific subclasses lets researchers use more specific techniques and obtain concrete results that can be applied to corresponding models and data. For example, weakly stationary time series have properties that do not change under time shifts. In particular, their trend is constant and covariance/correlation functions depend only on the distance between time moments.

Read

First, in the reading Random fields, you will revise basic probability theory.

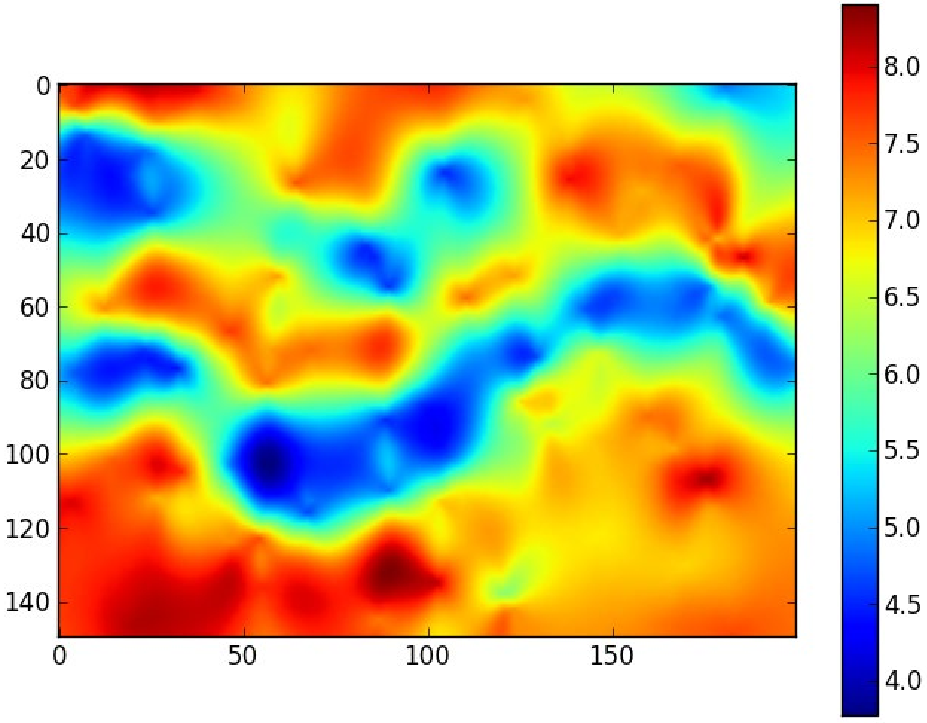
From the folder, open the document titled **Week_4_Topic_1** and read **slides 6–7**.

Random fields

A more general concept of random fields refers to a collection of random variables indexed by any space or surface indices, see, for example, Fig. 2.5. Thus, time series are a particular example of random fields.

Figure 2.5

Example of a realisation of a random field



https://commons.wikimedia.org/wiki/File:Simple_kriging_140_70.png

The concept of weakly stationary is extended to random fields but instead of time distances (lags) one uses spatial distances.

The class of covariance functions coincides with a class of positive definite functions. The concept of positive definiteness has numerous applications in data analysis, statistics and mathematics. fundamental in many areas. We consider 5 properties of positive definite functions which can be used to construct new covariance functions using the known ones.

Read

First, in the reading Random fields, you will revise basic probability theory.

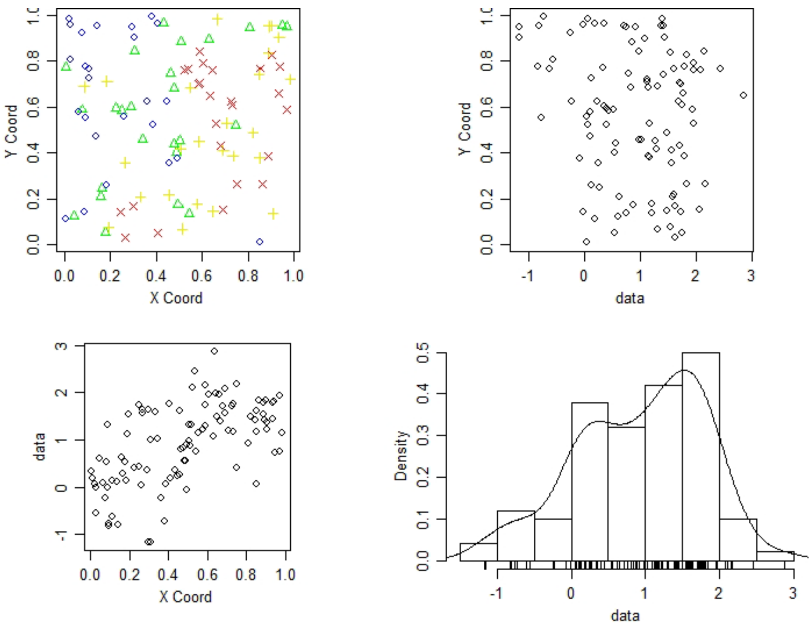
From the folder, open the document titled **Week_4_Topic_1** and read **slides 8–12**.

Topic 2: Descriptive spatial statistics

We will begin this week by discussing several methods of spatial exploratory data analysis. We will use the geoR package and plot various maps that help to visualise different properties of spatial data.

Figure 3.1

The descriptive plot of spatial data with its attributes



Then, you will learn how to use linear models to fit spatial trends. We will use the parametric approach when a trend is a function known up a finite number of numeric parameters. These parameters must be estimated to fit the trend to real data.

Upon completion of this topic, you will be able to distinguish between different types of spatial descriptive plots, use basic methods of the geoR package, and estimate spatial trends.

Visualization of spatial objects

The geoR package provides several functions for geostatistical analysis and visualisation of spatial objects. We will use the simulated data s100 from this package to illustrate the methods introduced in this week's lectures. As this dataset is simulated from a known statistical model, using it provides a convenient way to confirm the obtained results. The considered methods and estimators should give results that are close to the parameters of that model. First, we consider several methods to visualise the spatial locations of the points and their attributes. In particular, we will use options that specify point sizes, patterns and colours to be proportional to the data values or specified quantiles.

Read

In the reading Exploratory spatial data analysis, you will explore several methods to visualise spatial data. You will apply them to the s100 data from geoR. In this part, from the information provided for this dataset, we will use only the coordinates of data locations and the numeric data values at these locations.

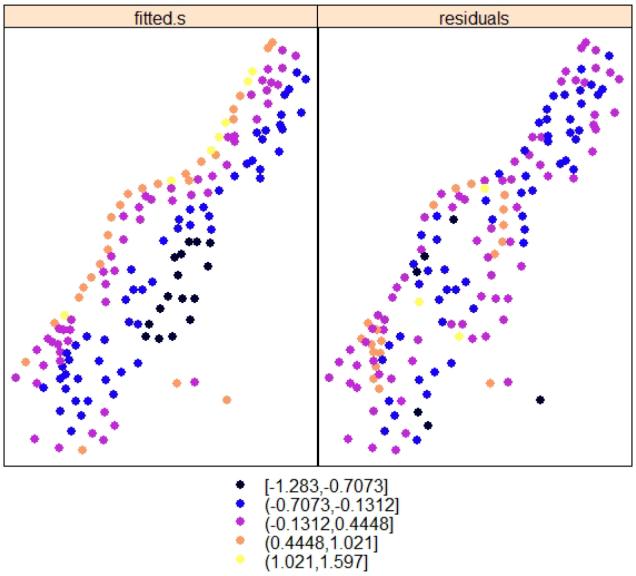
From the folder, open the document titled **Week_4_Topic_2** and read **slides 1–5**.

Basic analysis of Meuse data with sp

After introducing basic descriptive plots, you will apply them and use other basic methods to investigate zinc concentration in the Meuse river area. You will find that the concentration increases closer to the river. We suggest a model that links the transformed concentration and the distance to the river. Then, we will fit this model to the data and visualise the predicted result and the corresponding residuals, see Fig 3.2. For this fitting, we will employ the linear model, which can be used via the lm(…) R functions and the commands predict(…) and residuals(…). These commands are often used in conjunction with liner models in R.

Figure 3.2

The plot of predicted zinc concentrations and model residuals for the Meuse data



From the residual plot, you will see that the residuals do not appear to behave as spatially unstructured (residuals with similar values occur regularly close to each other). It suggests that we need more advanced models to improve the prediction.

Read

In the second part of the reading Exploratory spatial data analysis, you will apply several visualisation approaches to the zinc concentration data at the Meuse river area. You will suggest a spatial trend model and will analyse it using the linear model approach and its residuals.

From the folder, open the document titled **Week_4_Topic_2** and read **slides 6–12**.

Fitting spatial trends

A special form of linear regression can be used when a parametric function is used for spatial prediction. This form is called trend surface analysis.

We will continue using the meuse data introduced in the previous topics and consider the example of a second-order polynomial trend.

Read

In the third part of the reading Exploratory spatial data analysis, you will learn about the fitting parametric spatial trends in R.

From the folder, open the document titled **Week_4_Topic_2** and read **slides 13–14.**

Revise key R commands used in this week's materials.

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_4_Topic_2** and read **slide 15**.

Workshop

Activity

Workshop

This activity will be completed in R. Repeat the R programming content covered in Week 4. Modify the code and understand the impact of different R parameters on changes in results.

Your task

- Repeat R commands learnt in Week 4 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.
- If you can, 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.

Workshop solutions

The R code for Workshop is provided for your reference. From the LMS, please open the document titled – **Week_4_Workshop_4_RCode_Solutions**.

Summary

This week, we started with revising some concepts from the probability theory and extended them for modelling spatial data via random field models.

Then we considered exploratory spatial data analysis. This stage helps to formulate hypotheses, models and approaches for further steps of 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 geoR package: https://cran.r-project.org/web/packages/geoR/index.html
- Meuse river data set https://rsbivand.github.io/sp/reference/meuse.html

Theoretical concepts:

- Random variables: http://www.stat.yale.edu/Courses/1997-98/101/ranvar.htm
- Discrete random variables: https://amsi.org.au/ESA_Senior_Years/SeniorTopic4/4c
- Continuous random variables: https://amsi.org.au/ESA_Senior_Years/SeniorTopic4/4e
- Positive-definite kernels: https://en.wikipedia.org/wiki/Positive-definite_kernel

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
- Cressie, N.A.C (1993) Statistics for spatial data. Wiley. https://onlinelibrary.wiley.com/doi/book/10.1002/9781119115151 Avalilable on-line in La Trobe EBL ebook Library