Week 7

Kriging and model diagnostic

Overview

Welcome to Week 7 of Spatial Analysis.

This week we will cover two topics:

- Kriging

- Model diagnostic

You will start this week learning several methods to predict spatial values at unsampled locations. These methods found numerous applications in data science and are called kriging. We will demonstrate them using several datasets, in particular, predicting the zinc concentration in the Meuse River region.

Then you will learn how kriging methods can be applied to spatial models' diagnostics.

You practise conducting an analysis of real spatial data.

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

Topic 1: Kriging

- Introduction to kriging methods.
- Three types of kriging.
- Kriging with geoR package.
- Example of kriging for Meuse data.

Topic 2: Model diagnostic

- Introduction to model diagnostic.
- Leave-one-out cross-validation.
- N-fold cross-validation.

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

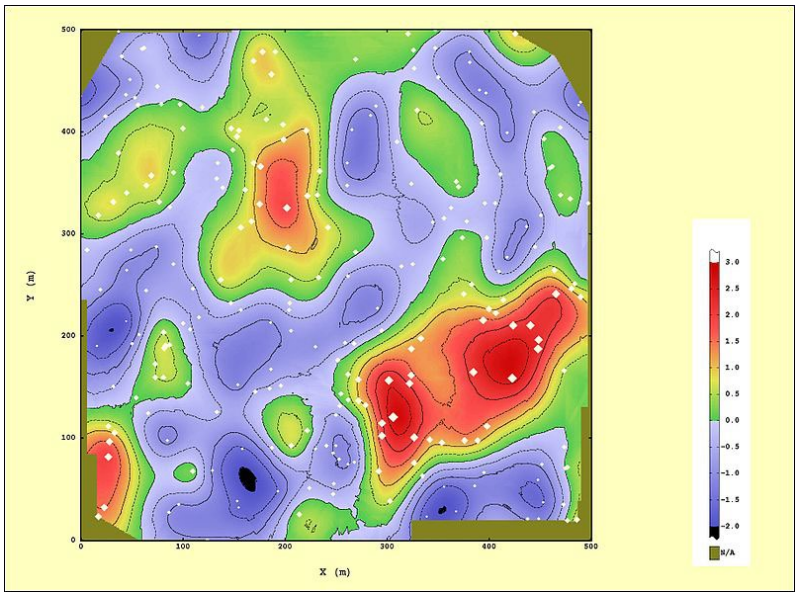
1. Formulate purposeful questions to explore new statistical ideas and subsequently design valid statistical experiments.
2. Present clear, well-structured analysis of important statistical model results.
3. Creatively find solutions to real-world problems consistent with those commonly faced by practising 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: Kriging and model diagnostic

This topic discusses the basic methods of predicting values at unsampled spatial locations. Such methods are called kriging and are based on covariance functions. A predicted value is a weighted sum of the known values at the sampled locations. The weights are determined by using the covariance function estimated from the data. Apart from extrapolation, kriging also found numerous applications to other problems. In this topic, we consider how to use it for spatial model diagnostics.

Figure 7.1

Kriging spatial prediction



https://commons.wikimedia.org/wiki/File:Exemple_basique_de_carte_krig%C3%A9e.jpg

Kriging methods

In this part, you will learn three types of kriging:

- Simple kriging
- Ordinary kriging
- Universal kriging

Each of these methods uses different assumptions about spatial trends to predict values at unsampled spatial locations. We first formulate the basic requirements of spatial prediction models. Then, we will give a solution to the prediction problem that satisfies these requirements assuming that the covariance function is known. In practical data analysis, instead of the known theoretical covariance function, one uses the estimated covariance function (which can be obtained from one of the previously discussed methods).

Read

First, in the reading Kriging methods, you will learn the definitions of the three kriging methods and will find the corresponding equations to compute predicted values and their variance.

From the folder, open the document titled **Week_7_Topic_1** and read **slides 1–6**.

Kriging with R

In this part, we consider kriging with the geoR and gstat packages. The packages have options for:

- Simple kriging

- Ordinary kriging.

- Universal kriging.

To perform kriging we can use the estimated parameters of covariance functions that were fitted using one of the methods from the previous parts. The krige and krige.conv commands choose the kriging method depending on the provided information. The results include predicted values and corresponding variances at the specified locations.

If one would like to visualise the obtained results, then a dense prediction grid covering the area of interest is selected and values are predicted to and visualised at each point of this grid.

We illustrate kriging applications to real data by using s100 and Meuse datasets.

Read

You will read the second part of Kriging methods and will learn about how to use R software for the kriging of real spatial data.

From the folder, open the document titled **Week_7_Topic_1** and read **slides 7–17**.

Topic 2: Model diagnostic

Spatial model diagnostics

Choosing the best spatial model for given data is a complex task. If the selection process involves several stages, it is almost impossible to use standard statistical tests of hypothesis or similar methods to validate the final results. In such cases, data analysts often use cross-validation methods. The cross-validation splits the existing data into two sets: a modelling (training) set and a validation (test) set. First, the modelling set is used for building and fitting a model. Then, the obtained model and kriging are performed at the locations of the validation set. Finally, the actual measurements at the validation locations are compared to their kriging predictions.

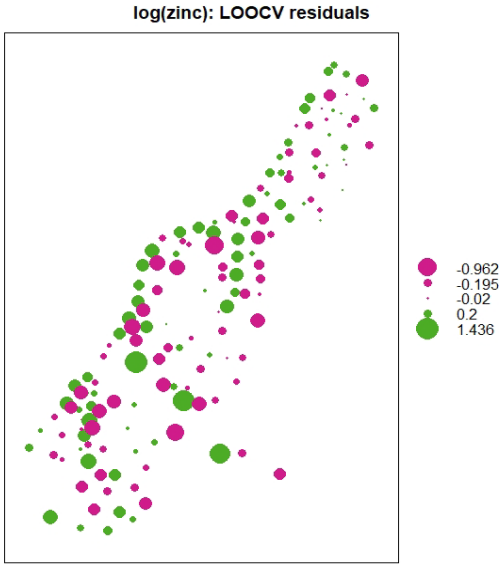
This approach can be used to compare several models. The model with the smallest total error between actual and predicted values at the validation locations is chosen. The simplest baseline model for any type of spatial prediction is the mean of all observations. One can consider the ratio of squared residuals for an alternative model and this mean prediction. If that ratio is smaller than 1, then the alternative model performs better. For values greater than 1, the mean prediction is better. In practical data analysis, one usually subtracts that ratio from 1 and checks whether it is greater than 0.

Also, it is useful to change the validation set several times in a such way that the combined set covers all spatial locations. Then, plotting the corresponding residuals can help to identify areas of concern with large residuals. The spatial plot of residuals can be used to find inadequate spatial model performance and suggest ways to improve them.

When you apply this approach to the Meuse river data, you will see (Fig. 7.2) that the obtained residuals are rather large in the central part of the Meuse region. It can be explained as the Meuse river changed its direction in the past and the distance to the current river banks does not include this information.

Figure 7.2

Kriging for spatial cross-validation



Read

In the third part of the reading Kriging methods, you will learn about applications of the kriging to model diagnostics via spatial cross-validation.

From the folder, open the document titled **Week_7_Topic_2** and read **slides 1–9**.

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_7_Topic_2** and read **slide 10**.

Workshop

Activity

Workshop

This activity will be completed in R. Repeat the R programming content covered in Week 7. Modify the code and understand the impact of different R parameters on changes in results.

Your task

- Repeat R commands learnt in Week 7 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.
- 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 spend around two hours on this activity.

Summary

This week, we studied kriging methods that use the estimated covariance functions to predict spatial values at unsampled locations. We considered two kriging approaches and their realisation in R. We practised performing kriging with the s100 and Meuse data.

Finally, we learned about cross-validation spatial methods which employ kriging.

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
- CRAN documentation for the gstat package: https://cran.r-project.org/web/packages/gstat/index.html
- Meuse river data set https://rsbivand.github.io/sp/reference/meuse.html

Theoretical concepts:

- Kriging: https://desktop.arcgis.com/en/arcmap/10.3/tools/3d-analyst-toolbox/how-kriging-works.htm https://link.springer.com/chapter/10.1007/978-3-662-03098-1_11

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