GEOSTATISTICS

MS in Earth Science Engineering, Geophysical Engineering specialization

First semester 2018/2019

COURSE COMMUNICATION DOCUMENT

University of Miskolc
Faculty of Earth Science and Engineering
Institute of Geophysics and Geoinformatics

Course datasheet

<table>
<tr><td>Course title: Geostatistics
Responsible professor:
Norbert Péter Szabó Dr., PhD, dr. habil., associate professor</td><td>Code: MFGFT730017
Responsible Institute/Department:
Institute of Geophysics and Geoinformatics / Department of Geophysics</td></tr>
<tr><td>Semester: third</td><td>Pre-requisites: -</td></tr>
<tr><td>Number of Contact Hours per Week:
2 lec. + 2 lab.</td><td>Type of Assessment (exam. / pr. mark. / other):
exam (oral)</td></tr>
<tr><td>Credits: 4</td><td>Type of Program: full time
Program and Specializations: MS in Earth Science Engineering, Geophysical Engineering specialization</td></tr>
<tr><td colspan="2">Study goals:
The subject deals with the theoretical description and practical issues of mathematical statistical methods used in earth sciences.
Competencies to be developed:
Knowledge: T3, T4, T5, T6
Ability: K1, K2
Attitude: A1, A2, A3, A4, A5, A7
Autonomy and responsibility: F1, F2, F3, F4, F5</td></tr>
<tr><td colspan="2">Course content:
Data distributions, the probability density function (pdf) and the cumulative distribution function (cdf). The calculation of modal value. The characterization of uncertainty. Robust estimations, the most frequent value method. The linear and rank correlation coefficient. Covariance and correlation matrices. Linear and non-linear regression analysis. Spatial correlation of petrophysical parameters, kriging. Multidimensional scaling and data analysis. Reduction of dimensionality by means of principal component and factor analysis. Hierarchical and non-hierarchical cluster analysis. Multidimensional modeling and parameter estimation. Discrete inverse theory and its application to geophysical datasets. Linearized and global optimization methods. The calculation of estimation error of the model parameters. Characterization of accuracy and reliability of the inversion result. Genetic algorithms and Simulated Annealing methods. Analysis of multidimensional relationships by means of neural networks.
Education method: Lectures with projected MS-PowerPoint presentation. Demonstration of statistical methods using own developed MATLAB codes (recipes) and the MATLAB Statistical Toolbox.</td></tr>
<tr><td colspan="2">Type of assessment:
Attendance at lectures is regulated by the university code of education and examination. Two writing tests with satisfactory results, and one assignment (MS-PowerPoint presentation) during the semester is the requirement of signature.
Grading scale: >86 %: excellent, 71-85 %: good, 61-70 %: medium, 46-60 %: satisfactory, <45 %: unsatisfactory.</td></tr>
</table>

Compulsory and recommended literature resources:

- Edward H. Isaacs, R. Mohan Srivastava, 1989. An introduction to applied geostatistics. Oxford University Press.
- Troyan V., Kiselev J., 2010. Statistical methods of geophysical data processing. World Scientific Publishing Co.
- Clark I., 1979: Practical geostatistics. Elsevier Applied Science.
- Steiner F., 1991: The most frequent value – Introduction to modern conception of statistics. Akadémiai Kiadó.
- Szabó N. P., 2017. Geostatistics. Electronic course material. http://www.uni-miskolc.hu/~geofiz/education.html

Course schedule

Week	Lecture
1	Data distributions. Datasets, histograms, pdf and cdf types. Determination of the modal value. Characterization of uncertainty. The Steiner's most frequent value method as robust statistical estimator.
2	The Maximum Likelihood method. Confidence intervals. Skewness and kurtosis. Propagation of error.
3	The Pearson's and the Spearman's correlation coefficient. Multivariate linear functional relationships. Covariance and correlation matrices. Linear and non-linear regression analysis. Robust regression methods. Earth science examples.
4	Spatial correlation of geophysical parameters, variogram models and kriging.
5	Writing test 1. Introduction to multivariate statistical methods. Data matrices. Multi-dimensional scaling. Multidimensional modeling and data analysis.
6	Hierarchical and non-hierarchical cluster analysis, the K-means clustering method. Rock typing and other earth science examples.
7	Reduction of dimensionality. Principal component analysis, factor analysis and their applications in geosciences. Lithology determination, estimation of petrophysical parameters.
8	Discrete inverse theory. Linearized and global optimization methods. Geophysical inverse problems.
9	Linear regression using inversion tools. The Gaussian Least Squares method. Weighted norms to be minimized. Well-logging applications.
10	The quality check of inversion results. Statistical backgrounds. The relation between the data and model covariance matrices. Error propagation.
11	Simulated Annealing methods. Classical and float-encoded genetic algorithm. Artificial neural networks. Earth science applications.

12	Simulated conference I.
13	Writing test 2. Simulated conference II.
14	Repeating the writing tests.

Week	**Practice**
1	Computer practice using MATLAB recipes.
2	Computer practice using MATLAB recipes.
3	Computer practice using MATLAB recipes.
4	Computer practice using MATLAB recipes.
5	Writing test 1.
6	Computer practice using MATLAB recipes.
7	Computer practice using MATLAB recipes.
8	Computer practice using MATLAB recipes.
9	Computer practice using MATLAB recipes.
10	Computer practice using MATLAB recipes.
11	Computer practice using MATLAB recipes.
12	Simulated conference I. Students deliver powerpoint presentations on the assigned topics. Evaluation of presentations.

13	Writing test 2. Simulated conference II. Students deliver powerpoint presentations on the assigned topics. Evaluation of presentations.
14	Repeating the writing tests.

Sample of writing test II

1. What kind of multidimensional scaling methods do you know?
2. How do we extract the principal components from a multivariate data set?
3. How does K-means cluster analysis work?
4. What are the operators of the float-encoded genetic algorithm?
5. Please characterize the training phase of ANN models. What is called a perceptron?

Solution

The answers can be found in the course material "Geostatistics" (and the recommended literature) uploaded to the site of the Department of Geophysics:
http://www.uni-miskolc.hu/~geofiz/education.html

1. See the slide titled "Scaling of observations" in the above course material.
2. See the slide titled "Estimation of principal components" in the above course material.
3. See the slides titled "Non-Hierarchical Clustering" and "Initial Model Dependence" in the above course material.
4. See the slide titled "Genetic operators" in the above course material.
5. See the slides titled "Training of ANNs" and "Perceptron" in the above course material.

Exam questions

1. Data distributions. Datasets, histograms, pdf and cdf types. Determination of the modal value. Characterization of uncertainty. The Steiner's most frequent value method as robust statistical estimator.
2. The Maximum Likelihood method. Confidence intervals. Skewness and kurtosis. Propagation of error.
3. The Pearson's and the Spearman's correlation coefficient. Multivariate linear functional relationships. Covariance and correlation matrices. Linear and non-linear regression analysis. Robust regression methods. Earth science examples.
4. Hierarchical and non-hierarchical cluster analysis, the K-means clustering method. Rock typing and other earth science examples.
5. Reduction of dimensionality. Principal component analysis, factor analysis and their applications in geosciences. Lithology determination, estimation of petrophysical parameters.
6. Discrete inverse theory. Linearized and global optimization methods. Geophysical inverse problems.
7. Linear regression using inversion tools. The Gaussian Least Squares method. Weighted norms to be minimized. Well-logging applications.

8. The quality check of inversion results. Statistical backgrounds. The relation between the data and model covariance matrices. Error propagation.
9. Simulated Annealing methods. Classical and float-encoded genetic algorithm.
10. Artificial neural networks. Earth science applications.