COURSE OF DOCTORAL STUDIES

Course title	Field of science (branch) code	University / Faculty	Institute / Department
Numerical and mathematical methods of the geological data analysis	Natural Sciences (Geology) N 005	Vilnius University / Faculty of Chemistry and Geosciences	Institute of Geosciences /

Study methods	Number of credits allocated	Study methods	Number of credits allocated
Lectures		Seminars	
Individual work	**10**	Consultations	

Course annotation

The aim of the subject is to acquaint the doctoral student with the basic concepts and methods of mathematical modeling and statistical data analysis applied in geology. Also, the PhD student must master at least one of the widely used scientific programming languages, e.g. R, Python, MATLAB, or Mathematica.

Mathematical models and their classes. Analytical models: models of ordinary differential equations, partial differential equations, models of stochastic processes. Differential equation and numerical analysis of models. Observational models, application of functions, numerical optimization in parameter search, filtering and state search.

Measurement scales. Probability, combinatorics, conditional probabilities and their use in solving geological problems. The central limit theorem and its meaning. The most important distributions, their properties and examples of their application in geology. Descriptive statistics and graphical representation of data. Elementary univariate statistics. Data transformations, conditions of their application and effects of transformations for statistical analysis. Composite ("closed space") data and peculiarities of their analyses.

Sequence and time series analysis. Markov chains and hidden Markov chains and their applications in lithofacies analysis and paleogeographic modeling. Autocorrelation and cross-correlation of sequences. Recurrence, cross and joint recurrence plots and statistical analysis of recurrence. Spectral analysis: Fourier transform, wavelets and evolutionary spectra with applications in quantitative stratigraphy.

Statistical regression models: linear regression, nonlinear regression, multivariate regression. Statistical significance of regression coefficients. Selection of regression models using information criteria. Model averaging. Bayes theorem and Bayesian models. Structural equation and causally explicit models.

Geostatistics: concept of regional variable, kriging, semi-variance, semi-variograms and their application in geological field analysis. Spectral analysis of 2D data. Identification of geological tendencies (trends) and anomalies. Methods of studying the distribution of points and lines in space. Methods of directional data analysis in space, von Mises distribution. Theoretical distributions of geological bodies on the surface of the sphere. Statistical testing of spherical distributions.

Fractals and multifractals. (In)dependence of geometric properties as a function of scale. Fractal dimensionality calculations and application to solving geological problems. Fractals of low dimensionality and estimating the internal complexity of 3D bodies.

Statistical analysis of multivariate data. Discriminant functions with examples of application in paleontology and petrology. Factor analysis, analysis of principal components, analysis of principal coordinates.

Methods of machine learning and artificial intelligence. Basics of the theory of classification and regression trees with applications in geology, paleontology, taxonomy. Random forest models. Artificial Neural Networks and Deep Neural Networks. Advantages and disadvantages of machine learning methods compared to classical statistical methods in the context of geology.

After choosing a certain topic, the doctoral student makes a half-hour long presentation, during which (s)he reveals the working principles of the method or their class and provides examples of independent analysis of geological data. The analysis must be performed using one of the numerical analysis languages.

Required readings

Davis, J.C. and Sampson, R.J., 2002. Statistics and data analysis in geology, Third Edition. New York: Wiley, 638 p.

McElreath, R., 2020. Statistical rethinking: A Bayesian course with examples in R and Stan. CRC press, 612 p.

Haneberg, W., 2012. Computational geosciences with Mathematica. Springer Science & Business Media, 394 p.

Shi, G., 2013. Data mining and knowledge discovery for geoscientists. Elsevier. 376 p.
Trauth, M. and Sillmann, E., 2012. MATLAB® and design recipes for earth sciences: how to collect, process and present geoscientific information. Springer Science & Business Media, 304 p.
Moseley, B. and Krischer, L., 2020. Machine Learning and Artificial Intelligence in Geosciences. Academic Press, 316 p.
Estabrook, G.F., 2011. A Computational Approach to Statistical Arguments in Ecology and Evolution. Cambridge University Press, 257 p.
Gershenfeld, N.A. and Gershenfeld, N., 1999. The nature of mathematical modeling. Cambridge university press, 356 p.
Harper, D. A. T. (ed.) 1999. Numerical Palaeobiology. Computer-Based Modelling and Analysis of Fossils and their Distributions.: x+468 pp. Chichester, New York, Weinheim, Brisbane, Singapore, Toronto: John Wiley & Sons, 468 p.
Weedon, G.P., 2003. Time-series analysis and cyclostratigraphy: examining stratigraphic records of environmental cycles. Cambridge University Press, 259 p.

Consulting lecturers Name, surname	Degree	The most important works in the field of science (branch) have been published during the last 5 years
Andrej Spiridonov	**Dr.**	**Spiridonov A**, Balakauskas L, Lovejoy S. 2022. Longitudinal expansion fitness of brachiopod genera controlled by the Wilson cycle. Global and Planetary Change, 103926 **Spiridonov A.** , Lovejoy S. 2022. Life rather than climate influences diversity at scales greater than 40 million years. Nature, 607, 307–312 **Spiridonov A** , Stankevič R, Gečas T, Brazauskas A, Kaminskas D, Musteikis P, Kaveckas T, Meidla T, Bičkauskas G, Ainsaar L, Radzevičius S. 2020. Ultra-high resolution multivariate record and multiscale causal analysis of Pridoli (late Silurian): implications for global stratigraphy, turnover events, and climate-biota interactions. Gondwana Research,Volume 86, 222-249 **Spiridonov A**., Samsonė J, Brazauskas A, Stankevič R, Meidla T, Ainsaar L, Radzevičius S. 2020. Quantifying the community turnover of the uppermost Wenlock and Ludlow (Silurian) conodonts in the Baltic Basin. Palaeogeography, Palaeoclimatology, Palaeoecology, Volume 549, 109128 **Spiridonov A**, Balakauskas L, Stankevič R, Kluczynska G, Gedminienė L, Stančikaitė M. 2019. Holocene vegetation patterns in the southern Lithuania indicate astronomical forcing on the millennial and centennial time scales. Scientific Reports, 9, 14711

Approved by the doctoral committee of Geology (N 005) on 1st of December 2022 (No. (7.17 E) 15600-KT-467).

Committee Chairman prof. dr. Sigitas Radzevičius