GG691 DATA EXPLORATION & PROCESSING [3 CR]

Prerequisites:
GG 413, or the instructor's consent.

Course Purpose and Content:
Quantitative skills are becoming more and more important in the natural sciences. With the rapid development of remote sensing from satellites and remotely operated vehicles, the amount of data that engineers and ocean, Earth, and environmental scientists must process and interpret is overwhelming; being able to analyze data on a computer becomes a necessity and often a job requirement. This course is the second (and more advanced) of two GG courses on how to quantitatively analyze data in the sciences†. Topics that might be covered include the analysis of sequences, time series and spectral (Fourier) analysis, filtering of data sets in the time and frequency domains, fractals and wavelets and their use in the natural sciences, triangulation, gridding, and interpolation, factor and cluster analysis, empirical orthogonal function (EOF) analysis, map patterns and crossover analysis. Computer applications of the various techniques will be implemented and applied to typical data sets in the natural sciences. There will be several laboratory sessions and a class project, in which the graduate students analyze a data set of their choice using the techniques discussed in lectures and labs. All computing will be done with MATLAB (or a student’s preferred software) on the GG department's workstations or student’s laptops.

Text:
Recommended (optional) text: John C. Davis, Statistics and Data Analysis in Geology, 3rd edition, John Wiley. The instructor’s extensive notes will supplement the optional text.

GG Student Learning Objectives (SLOs):
GG department has provisionally defined 4 and 5 learning objectives, respectively, for the MS and PhD graduate degree programs, relating to technical knowledge, the conduct of science, oral and written skills, and professional skills. For MS students, this course addresses:

MS. 1. Technical knowledge. M.S. graduates are proficient in applying technical knowledge of theory, laboratory methods, field methods, computer applications, and the supporting disciplines (math, physics, chemistry, biology) to help advance the fields of geology and geophysics.

MS 2. Scientific method M.S. graduates are able to (a) construct scientific hypotheses, (b) define and carry out research to evaluate them in a timely manner, (c) analyze and synthesize the

† The foundation for this class is laid in GG413 – Introduction to Statistics & Data Analysis, which presents elementary probability and statistics, statistical tests, curve fitting, simple and multiple regression, basic spectral analysis, and analysis of directional data.
results of their research, and (d) derive conclusions that help advance the fields of geology and geophysics.

For Ph.D. students, the course addresses:

**PhD 1.** Technical knowledge Ph.D. graduates are proficient in applying technical knowledge of relevant theory, laboratory methods, field methods, computer applications, and the supporting disciplines (math, physics, chemistry, biology) to advance the fields of geology and geophysics.

**PhD 3.** Scientific method Ph.D. graduates are able to independently (a) construct scientific hypotheses, (b) design and carry out research to evaluate them in a timely manner, (c) analyze and synthesize the results of their research, and (d) derive conclusions that advance the fields of geology and geophysics.

**Course Goals:**

This course will enable students to perform more advanced data analysis of typical natural science data sets using computer software such as MATLAB or the free clone Octave. The specific student learning outcomes of the lectures are to:

- Understand the basics of auto-regressive, moving average methodology.
- Become familiar with filtering both as convolution and geospatial operation
- Learn how gridding works and gain insight into proper selection of gridding algorithms
- Become familiar with methods used in cluster and factor analysis.
- Explore the analysis of lines in the plane, their distribution and intersections.
- Explore matrix methods as used in empirical orthogonal function analysis
- Gain familiarity with pattern recognition techniques suitable for map analysis
- Understand more advanced aspects of spectral analysis in 1 and 2 dimensions
- Become acquainted with modern analytical techniques using fractals or wavelets.

Students will reach the course SLO’s by reading assigned notes before attending the lectures, participating in laboratory sessions, and analyzing their own data sets in a class project.

**Assessment and Grading:**

This graduate course requires motivated students to get the most out of it. The course covers many topics in detail. Course attendance is mandatory. Lab exercises constitute 50% of the course grade; the final project contributes the other 50%. Grading will account for overall class performance, individual effort in completing the assignments, and attendance. If a student is concerned about his or her grade they may ask the instructor at any time about their standing.

**Class Format:**

This is mostly a lecture course. Students are encouraged to actively ask questions in class, particularly if they do not understand the material being discussed. Students will get the most out of the lectures if they keep up with the reading. The course will have 4–5 laboratory sessions using MATLAB in the department’s computer room (or on your own device). Students are expected to be fully fluent in MATLAB or an equivalent prior to enrolling in the course.

**Lecture Notes:**

These will be available for students to purchase ($15), with additional material available on Laulima. These are required readings optionally supplemented by the recommended text. Likewise, lab exercises will be distributed via Laulima and students will be required to submit lab responses via the Laulima course website.