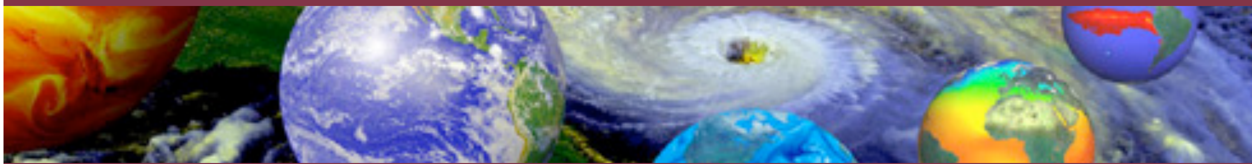
GEOL425: Data Analysis in the Earth & Environmental Sciences



Department of Earth Sciences, University of Southern California

Synopsis

This problem-based, upper division course will introduce students to elementary techniques of data analysis as integral tools to the study of the Earth & Environmental Sciences. Since students may not be familiar with mathematical parlance, we follow a two-tiered approach. The first tier serves as an introduction (and/or refresher) to the fundamental principles of calculus, linear algebra and probability theory. In the second tier, students will learn how to apply these foundations to univariate and multivariate datasets taken from classic problems in the Earth & Environmental Sciences (geophysics, seismology, paleoclimatology, paleobiology, ecology, oceanography, atmospheric sciences). The class will consist in two 1.5 hour lectures and one 2 hour lab per week, culminating with a term paper where students will apply the acquired tools to a research project discussed with their advisor (or suggested by the instructors).

Instructors:	Prof. J. Emile-Geay (julieneg@usc.edu) Prof. T. W. Becker (twb@usc.edu)
Assessment:	Term paper 30%; Mid term exam 20% Lab exercises 40%; Class participation 10%
Dates:	Fall semester, 2009.
Class meeting times:	*Tue* 1 – 2:20 pm (lecture) *Wed* 12 :30 – 1:50om (lecture) and 2:00 – 4:00 (lab)
Location	Lectures ZHS 264. Lab ZHS 130.
Recommended Preparation	MATH 125-126, Matrix Algebra
Units	4 (w/ TA).

Objectives and approach

The objective of this class is to enable students in the Earth, Environmental and Biological Sciences to perform essential tasks in data analysis, include :

- Visualizing timeseries and multivariate datasets with a number of common software packages, notably Matlab.
- Plotting data with error estimates and perform basic error propagation analyses.
- Computing, and correctly interpret correlation coefficients.
- Computing, represent and correctly interpret timeseries power spectra and cross-spectra using several methods.
- Performing basic linear regressions and least-squares fits.
- Solving a wide class of inverse problems.
- Being conversant with classic parametric and non-parametric statistical tests.
- Applying a number of these tools to their own research.

To this end, we begin each section of the class with an appropriate "refresher" in the underlying mathematical foundation (calculus, linear algebra and probability theory) . We then describe the theory behind quantitative tools and have students apply them to real-world problems from the solid and fluid Earth. In addition, the class will cover advanced topics, as deemed appropriate by instructors given the students' level of interest.

Lecture Texts

- Required reading: Gubbins, D. *Time Series Analysis and Inverse Theory for Geophysicists,* Cambridge University Press, 2004
- Recommended reading:
  - Stoica,P. & Moses R., *Spectral Analysis of Signals,* Prentice Hall, New York, 2005
  - Wilks, D., *Statistical methods in the atmospheric sciences*, Academic Press, 2006.
  - Boas, M, *Mathematical Methods in the Physical Sciences, 2nd Edition,* Wiley&Sons , 1983
  - Talyor, J.R., *An Introduction to Error Analysis*, University Science Books, 1997

Syllabus

Calculus I Week 1 (Aug 24 - Aug 28) BECKER
Exponentials and logarithms. Trigonometry. Complex numbers (Amplitude and phase). Derivatives. Ordinary Differential Equations.
Reading: Boas, Chap 1. Gubbins, Chap 2.1
Lab: Introduction to Matlab. Tutorials, Demos.

Calculus II Week 2 (Aug 31 - Sep 4) BECKER
Integrals, Numerical quadrature. Special functions. Taylor expansions. Series expansions. Energy (L^2 norm).
Reading: Boas. Chap 2. Taylor, Chap 3.
Lab: Numerical Integration, Polynomial Approximations.

Statistics I Week 3 (Sep 7 - Sep 11) EMILE-GEAY
Notion of probability. Bayesian vs. frequentist interpretations.
Probability law. Discrete and continuous distributions. Normal distribution.
Poisson's distribution. Random variables. Moments. Median. Expectance operator.
Reading: Taylor, Chap 4, 5 & 10
Lab: Binomial and normal distributions. Error propagation in a mass-spectrometer.

Statistics II Week 4 (Sep 14 - Sep 18) EMILE-GEAY
Gaussian statistics. Significance tests: Student's T, Fischer's Z, and F-test.
Correlation (Pearson's rho, Kendall's tau). Goodness-of-fit tests (KS, Chi squared).
Reading: Taylor, Chap 9, 12. Wilks, Chap 5
Lab: Correlation is not causation.

Timeseries I Week 5 (Sep 21 - Sept 25) BECKER
Notion of timeseries. Fourier series, Fourier transform. Parseval's theorem.
Time and frequency representation. Wavenumbers. Spherical harmonics.
Reading: Gubbins Appendix 1,2 ,3
Lab: Space-time decomposition of atmosphere, ocean and lithospheric waves.

Statistics III Week 6 (Sep 28 - Oct 2) EMILE-GEAY
Histograms. Kernel density estimation. Long-tailed distributions.
Non-parametric statistics : Bootstrap, Jacknife, Monte-Carlo methods.
Reading: Wilks, Chap 3
Lab: Significance of correlation coefficients (parametric vs nonparametric tests) (geology example)

Timeseries II Week 7 (Oct 5 Oct 9) EMILE-GEAY
FFT. Convolution. Linear filters. Sampling theorem. Spectral estimation : Aliasing. Leakage. Bias vs variance tradeoff. Significance of spectral peaks.
Reading: Gubbins, Chap 3-5
Lab exercise: Seismological instruments as filters.

Timeseries III Week 8 (Oct 12 Oct 16) EMILE-GEAY
Auto-correlation, Cross-correlation/Cross-spectrum. Coherency. Timeseries modeling. AR(n) processes. Advanced spectral methods : Singular spectrum analysis. Multi-Taper Method. Maximum Entropy Method.
Reading: Ghil, M, 2002 : "Advanced Spectral Method for climatic timeseries", Rev Geophys.
Lab exercise:. Spectral analysis of Vostok Ice Core data

Linear Algebra I Week 9 (Oct 19 -Oct 23) BECKER
Linear Spaces. Vectors and matrices. Tensors. Matrix algebra. Basis. Projections. Systems of linear equations.
Reading: Boas, Chap 2
Lab: Mapping, projections, rotations.

Linear Algebra II Week 10 (Oct 26 -Oct 30) EMILE-GEAY
Eigensystems. Diagonalization. Singular Value Decomposition. Principal Component analysis.
Reading: Boas, Chap 3, Wilks, Chap 11.
Lab: Stress tensor/Principal stresses. EOF analysis

Inverse Theory I Week 11 (Nov 2 - Nov 6) BECKER
Least-squares fitting : Ordinary, weighted, total least squares. Linear regression

Reading: Wilks, Gubbins Appendix 4&5
Lab exercise: Fourier analysis as an inverse problem (Gubbins, Chap 10)

Inverse Theory II Week 12 (Nov 9 - Nov 13) BECKER
Geophysical Inverse Theory. Deconvolution.
Reading: Gubbins, Chap 6 to 9
Lab: Seismic tomography problem

Interpolation Week 13 (Nov 16- Nov 20) EMILE-GEAY
Univariate : Linear interpolation. Spline interpolation. Lagrange polynomials.
Multivariate : Kriging. Optimal Interpolation.
Reading: Kaplan, JGR 1998.
Lab: Infilling a sea-surface temperature data or a geologic map.

Special topics Week 14 (Nov 30- Dec 4) EMILE-GEAY
Wavelets. RegEM, and more (time-permitting)...

Academic Accommodations:

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to us as early in the semester as possible. DSP can be reached at ability@usc.edu and is open 8:30am-5:00pm Monday through Friday. The phone number for DSP is 213-740-0776.

Synopsis

(25 words for USC Catalogue) : Introduction to mathematical methods giving insight into Earth and Environmental data. Topics include : probability & statistics, timeseries analysis, spectral analysis, inverse theory, interpolation.

Lab sessions

Each week, students will be invited to apply the concepts learned in class using a real-world geoscience dataset and Matlab code provided by the instructors. During the lab sessions, they will be asked to load, analyze and plot data before answering questions designed to gauge their understanding of theoretical concepts as viewed through that particular application. A T.A. from the Earth Science department will take care of the important tasks of attending weekly lab sessions, providing student support during labs and office hours, and grading the weekly lab reports. The practical work involved in lab sessions is absolutely central to the acquisition of the notions presented in the class, is very time-intensive for both students and instructors, and will thus require TA participation. In addition, she/he will proctor the midterm exam and ensure that course material gets uploaded to the class webpage in a timely manner. We thus request 0.5 T.A. units for this course.