OCEA 260/EART 260Data Analysis in Ocean and Earth Sciences Winter 2016

Instructor: Christopher Edwards
Office: A447
Phone: (831) 459-3734
email: cedwards@ucsc.edu
Class location: E&MS D250
Class time: TTh 10-11:45

Office Hours: TTh 12-1.
Other times are also possible if I'm in my office and available, or by appointment.

An introduction to data analysis methods regularly encountered within the ocean and earth sciences. Topics include error propagation, least squares analysis, data interpolation methods, empirical orthogonal functions, and Monte Carlo methods applied to problems drawn from oceanographic and earth sciences datasets. A student project consists of analysis of their own dataset.

The course will introduce and use a high level computing and visualization package, Matlab, though students are welcome to use Python instead for their homework. Both Matlab and Python are outstanding research tools. You will benefit from knowledge of either one. Matlab has the easier learning curve. Python has a steep learning curve but is ultimately more powerful (not necessarily for statistical analysis but for other broader tasks, such as accessing and processing the computer operating system, the web, and innumerable other tasks).

Topics covered approximately by week:

1. Matlab basics and linear algebra.
2. Systematic and random errors, propagation of error, probability distributions, mean, variance, standard deviation.
3. Central limit theorem, confidence intervals, significance tests.
4. Correlation, autocorrelation, integral time-scale, degrees of freedom.
5. Interpolation/Extrapolation
6. Fourier Analysis (may swap this out for another topic)
7. Empirical Orthogonal Functions/Principal components
8. Least-squares regression
9. Student projects

Project: Students are expected to carry out analysis of a dataset of their choosing. For those students who do not have a dataset of interest at hand, we will together to find a suitable analysis project. Discussion of the data and desired analysis will be carried out in the third week of class.

This course introduces methods frequently applied in geophysical sciences. But, it is only an introduction and general in its overview. The material covered should enable you to calculate preliminary results and prepare you to read literature on more advanced methods and refinements appropriate for your specific field. It is up to you to continue to read that literature and determine the exact best method for your specific research problem.

OCEA 260/EART 260Data Analysis in Ocean and Earth Sciences Winter 2016

Grading:
Probably 6 homework sets will be assigned, due 1 week after they are distributed. In addition, there will be an independent project, which counts as 1 homework (0.25 for oral presentation, 0.75 for written report). Homework assignment #3 and the final homework assignment will be considered as take-home midterm and final exams.
The final grade will be the result of the average of the scores.

Example:
HW1: 78%
HW2: 75%
Midterm/HW3: 90%
HW4: 70%
HW5: 80%
Final/HW6: 75%
Project: 82%

Total: 550/700 = 78% (final percentage)

Letter grades will be scaled relative to class.

Homework policy
Students often learn effectively through talking with other students. Discussion about the homework is encouraged, but the work you turn in must be (1) in your own words and (2) reflect the degree to which you understand or think you understand the material. Verbatim copying of homework is NOT allowed. It is all right to work through a problem with other students, including understanding which equation is used and why and how it is solved. Communication is an important part of any profession. Much of the homework in this course is matlab programming. It is NOT acceptable to cut and paste computer code to turn in. You can show fellow students your code, and you can look at other people's code, but I expect that the results that you turn in to be the result of your own typing. Do not email around snippets of code, though you can email function names that are useful and their syntax.

There will be no discussions between students about the midterm and final homework assignments.

Note on matlab toolboxes or externally supplied scripts (e.g., from matlab central): The problems in this course are designed to be carried out with the standard matlab distribution and NOT the statistics (or other) toolbox. You may use the toolbox (or other scripts) to learn from, and it's a great way to check your code (in fact, you often learn more when there's a difference between your code and that supplied by matlab or someone else!). But since part of the purpose of the class is to learn and practice matlab programming, homework that uses the statistics (or other) toolbox routines will be marked down. Also, if you use scripts from elsewhere, please tell me on the homework what you used. The same spirit applies to python. Please, whenever possible, write the code to calculate statistics we develop in class.

Late Homework: Homeworks are due as listed on the assignments. Late homeworks will be accepted up to one class beyond the due date with no penalty. Homework turned in after that day will receive a zero.

Books

An introduction to error analysis, J. R. Taylor, University Science Books, 1997. An amazingly readable, simple (physics) level textbook. Really a great resource. We will use quite a bit of this in the first few weeks of the class. Required.

Statistical methods in the atmospheric sciences, 3rd edition, D. S. Wilks, Academic Press, 2011. This is another amazingly readable, very complete statistics and data analysis book. This book is written at a higher level than Taylor, and it includes many more topics. We will only use some parts in the class, though I highly recommend it as a reference on your shelf if you will be using the techniques discussed. Highly recommended.

Other useful books:

Random Data Analysis and Measurement Procedures, J. S. Bendat and A. G. Piersol.

Data Reduction and Error Analysis for the physical sciences, P. R. Bevington and D. K. Robinson.

Numerical Recipes, W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling.

The ocean circulation inverse problem, C. Wunsch, Cambridge University Press, 1996. This is an excellent text, though it covers material that is likely more useful to PO students than other areas of ocean sciences.

Modeling Methods for Marine Science, Glover, D.M., W.J. Jenkins, and S.C. Doney, 2011, Cambridge University Press. A fairly new text and written by oceanographers. It includes lots of useful material, but not as completely as Wilks, so I do not use it for this class.

Data analysis methods in physical oceanography, W.J. Emery and R.E. Thomson, Elevier, 2nd Ed, 2001. This is also an excellent book from which we will draw a few examples, but not as useful as Wilks.

Matlab access: Matlab can be accessed in several ways:

1) Buy the student version of matlab (I think ~$100) and use it on your own pc/mac.
2) Access matlab on your advisor's machines.
3) You can set up matlab on your machine of choice to access a department/building license server.
4) Go to the UCSC computer centers (BE 109, Cowell Apt, Kresge, Ming Ong, SS1 all have matlab).

Here is a link to all the IT-supported computer centers available at UCSC: http://its.ucsc.edu/computer-labs/descriptions/all-labs-summary.html

Python access: Python is a free language. In addition to the basic distribution, you'll need numpy, scipy, and matplotlib as well. My recommendation is to install "anaconda" which is an integrated development environment (IDE) including a collection of python modules that are useful for scientific applications, an editor and a way to run your code.