Course Proposal EAS 8803 Special Topic
Advanced Environmental Data Analysis

INSTRUCTOR:

Prof. Emanuele Di Lorenzo

phone 404-894-3994
office ES&T 3252
email edl@eas.gatech.edu

Course website: http://ocean.eas.gatech.edu/eas-8803

TEXTS:
- *Statistical methods in the Atmospheric Sciences*, Wilks, Academic Press,1995.
- *Discrete Inverse and State Estimation Problems*, Carl Wunsch, Cambridge Press.
- *Lecture notes* and additional reading material provided on website.

COURSE PHILOSOPHY AND GOALS: This course is an advanced introduction to environmental data analysis and modeling. It is intended for first year graduate students and senior undergraduates in environmental sciences and engineering. The goal of this class is to provide a deeper understanding of the theory underlying the statistical analysis of environmental data, both in the space and time domain, and to provide the students with a hands on experience. Ideally at the end of this class the students will have developed a series of computer programming tool boxes and theoretical skills that should immediately be available for analyzing and modeling data in their own research.

Although some preview knowledge of probability and statistics is required, a background review will be provided. Concepts and notation will be reintroduced as needed. In this class students will learn how to combine models, which quantify statistical or dynamical relationships, with observations, time series analysis, hindcasting/forecasting and interpolation/extrapolation and signal decomposition. A more detail description of these topics is appended in the **LECTURE TOPICS (1-5)** below.

HOMEWORK: There will be 4 homework assignments, one for each of the core topics. You will be required to learn some computer programming skills with either MATLAB or the R – software (http://www.r-project.org/). If you do not have access to a computer with these software you will be provided with an account at the beginning of the semester. The type of data to be analyzed in the homework will vary depending on the interest of the attending students.

EXAMS: There is going to be a short review MIDTERM.

PRESENTATION/STUDENT PROJECT: To help you put into practice your data analysis skills you will be asked to choose a data analysis project, possibly involving your own research data, that you will present at the end of the semester. The project is to be chosen based on a set of questions that you would like to answer rather than the type of data analysis technique you would like to apply, and it may require the use of one or more data analysis techniques.

GRADING: 35% Homework, 25% Midterm, 40% Presentation.

LECTURE TOPICS:

1) BACKGROUND REVIEW: The goal of this section is to introduce relevant notation, review fundamental statistics used in environmental modeling (e.g. air-quality, ocean, atmosphere and climate sciences).

- How statistical analysis relates to environmental processes.
- A unifying view of statistical methods in environmental sciences.
- Review of statistics fundamentals, univariate and multivariate statistics and PDFs.
- Statistically Optimal Linear Estimators and Bayesian estimation fundamentals.

2) COMBINING MODELS AND OBSERVATIONS: The goal of this section is to introduce and apply different flavors of *Least Squares* (LSQ) methods to combine observations with models. LSQ is the basis for more advanced estimation and data assimilation techniques used in environmental modeling (see section 4). In this section we will understand how the role of the *model* is to (1) provide a testable hypothesis and (2) constraint the relationships between the observations and the parameters to be estimated, in other words the *covariance* between data and parameters. Examples of models explored in this section include empirical, statistical and dynamical models.

- Review of Linear Algebra and Vector Calculus.
- Least Squares (LSQ), weighted and tapered LSQ, uncertainty estimates.
- Empirical Models: Interpolation and function fitting with LSQ.
  - *(*) The CO_2 curve and global warming*
  - *(*) Mapping sea surface temperature in the Pacific Ocean*
- Dynamical Models: LSQ and Inverse Modeling.
  - *(*) Reconstructing the source of a pollutant with an advection diffusion model*
  - *(*) Adaptive Sampling to monitor pollutant over a specific location*
- Dynamical Constraint in LSQ, Lagrange Multipliers and Adjoints
  - *(*) Adjoint sensitivity analysis*
  - *(*) Inverse/Adjoint modeling of advection/diffusion of tracers (e.g. O_2, NO_X)*

3) TIME SERIES ANALYSIS :Time series analysis can be understood in the context of combining models and observations. We will study time and frequency domain models, auto-regressive and moving average models. These models will enable us to perform spectral coherence analysis, trend analysis with significance testing, and error estimations when reconstructing time series.

- Frequency domain models, spectrum and autocovariance functions, convolutions and cross-correlation, aliasing, Discrete Fourier Transform (DFT), spectral estimates with LSQ, and tapering
  - *(*) El Niño periodicities and significance of climate change effects*
  - *(*) The spectrum of oceanic and atmospheric waves*
- Time domain, autoregressive, and moving average models
  - *(*) The Hasselmann 1D model of the climate system*
  - *(*) Autoregressive processes in a stratified fluid (e.g. vortices, waves, mixing)*
- Cross-spectra and coherence analysis of multiple timeseries and space/time filters (e.g. high-pass, low-pass, band-pass)
  - *(*) Mapping the propagation and spatial structure of Rossby and Kelvin waves*

- ❖ Wavelet analysis
  - *(*) Stationarity of El Niño and tropical sea surface temperature statistics*
  - *(*) Intraseasonal rain events over the Indian subcontinent*

4) INTERPOLATION/EXTRAPOLATION and HINDCASTING/FORECASTING: In this section we will move beyond LSQ towards a more advanced estimation and data assimilation technique that make use of mulitvariate covariances between data and parameters. These methods range from *optimal interpolation, kriging* and *objective mapping* to *3D* and *4D variational data assimilation* used in weather forecasting and also sequential *Kalman ensemble filters* and *Monte Carlo methods*. We will show how the main differences between these various methods is based on how they define/approximate the covariance between the observations and the parameters to be estimated (e.g. covariance derived using dynamical models or built using statistics from previous observations, or simply assumed to have a functional form). These multivariate covariances are the foundation of interpolation/extrapolation (in the spatial domain) and hindcasting/forecasting (in the temporal domain).

- ❖ Covariance Modeling, Theory and Applications.
  - *(*) ENSO Forecasting (time domain application)*
  - *(*) Reconstructing sea surface temperature from precipitation data (space/time domain application)*
- ❖ Optimal Interpolation and Objective Mapping.
  - *(*) Mapping sparse data using empirical anisotropic covariances*
- ❖ 4D Variational Data Assimilation.
  - *(*) The ECMWF weather forecasting system*
  - *(*) IOM, a modular variational assimilation platform*
- ❖ Sequential Methods and Montecarlo Sampling.
  - *(*) Comparing the extended and ensemble Kalman Filter*
  - *(*) A Bayesian Hierarchical Model (BHM) example*

5) SIGNAL DECOMPOSITION: Covariance matrices are central in many statistical applications (see previous sections). By decomposing these covariance matrices we are able to extract and synthesize essential information about a dynamical system. This section covers most common decomposition methods used in environmental sciences (e.g. SVD, EOFs, PCA, MCA, CCA, POP)

- ❖ Singular Value Decomposition (SVD)
- ❖ Empirical Orthogonal Functions (EOFs) and Principal Component Analysis (PCA)
  - *(*) Atmospheric teleconnection patterns from 1000 mb geopotential heigth*
- ❖ Maximum Covariance Analysis (MCA)
  - *(*) Modes of covariability between sea surface temperature and precipitation*
- ❖ Combined EOFs (using SVD) and Canonical Correlation Analysis (CCA)
  - *(*) Impacts of global warming on the coupled variability of ocean/atmosphere*
- ❖ Principal Oscillation Patterns (POP)
  - *(*) Using statistics to build a low order dynamical operator of the tropics*

6) STUDENT PRESENTATION - Presentation Guidelines are posted on class website: http://ocean.eas.gatech.edu/eas-8803