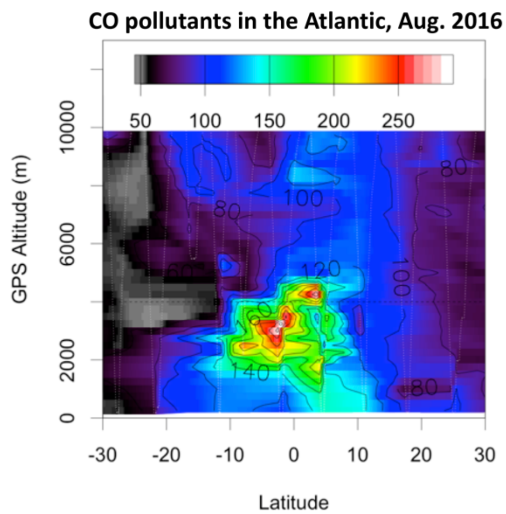
EPS 236 Environmental Modeling and Data Analysis. Spring Term 2019

Prof. Steven C. Wofsy, Geo Museum Room 453. Telephone: 617 495-4566; email: swofsy@seas.harvard.edu
Prof. Daniel J. Jacob, Pierce Hall Room 110C. Telephone: 617 495-1794; email: djacob@seas.harvard.edu

Teaching Fellow:
Tianning Zhao zhaot@g.harvard.edu



Location: Geo Museum 375 (behind the globe in the climate exhibit)
Time: Wed. and Fri., 1500 – 1615.
Office hours and section times TBD
1st meeting ***Wednesday January 30, 2019***.

Course Website:
https://canvas.harvard.edu/courses/47513

Course overview

EPS 236 is a project-oriented, hands-on course that provides a graduate-level introduction to modeling, data analysis, and data visualization suitable for students from many fields, with applications drawn from the atmospheric science, environmental engineering, and climate science. It is suitable for graduate students and advanced undergraduates in Earth and Planetary Sciences, Engineering Sciences/Environmental Science and Engineering, and allied natural science departments (e.g. Organismic and Evolutionary Biology, Chemistry and Chemical Biology, Physics; at MIT, students from EAPS and Civil &. Environmental have often enrolled).

Prerequisite: Applied Mathematics 105b or equivalent (may be taken concurrently); a course in atmospheric chemistry (EPS 133 or 200 or equivalent) is helpful, but not required; or permission of the instructors.

The course is divided into two parts:

1. **Data analysis** focusing on *understanding the science content* and *quantifying sources of error* in analysis of complex data sets from environmental networks, satellite sensors, or individual instruments. Model concepts that underlie statistical inference, data analysis, and application of basic analysis principles to real data, are emphasized. *R will be used as a tool for visualization, time series analysis, Monte Carlo methods,* and *statistical assessment.*
2. **Models in environmental science** emphasizing *(a) linear models* (mathematical principles, time evolution operator, eigenvalues and eigenvectors; Markov chains), (b) *chemical transport models* including basic principles and numerical methods, and (c) *inverse modeling* (optimal estimation, Kalman filter, adjoint methods).

Course requirements

- **Credit:** Homework (bi-weekly, 40%), Projects (40%); Oral presentations 20%.
- **Recommended**: Dalgaard, P. (2008) Introductory Statistics with R; *or similar*
- **Collaboration policy and electronic devices:**

For assignments in this course, you are encouraged to consult with classmates as you work on problem sets. However, after discussions with peers, make sure that you can work through the problem yourself and ensure that *answers you submit are the result of your own efforts*. You must cite any books, articles, websites, lectures, etc that have helped you with your work using appropriate citation practices. Access to solutions from previous years is strictly forbidden. Note: *Use of laptop computers during class should be exclusively for in-class work.*

Lecture topics for Part 1.

Part 1a: Linear models, Markov chains, analysis with eigenvectors/eigenvalues. (Steve Wofsy)

Linear models provide a widely used, basic conceptual framework for modeling many types of data. In Part 1 of the course, linear systems are examined to illustrate the fundamental properties of mass-conserving and non-conserving systems simulating chemical species in the environment, including stochastic models, inverse and adjoint models.

Topics include the following: Setting up the conceptual model-how do we structure the model and obtain estimates for the magnitudes of the parameters (the simplest "inverse modeling")? *Solving the mode*l-eigenvalues and eigenvectors, the importance of non- orthogonality, the time-evolution operator, transient and steady-state behavior, tangent linear approximations for non-linear systems. *Applying the model*-how do we use these models as tools to improve our understanding?

Students will receive training to use R, which will be utilized in problems focused on applications to global chemical cycles, urban atmospheric structure and chemistry, etc. (Students already proficient in Matlab, Python, or similar applications may use one of those applications, but R will provide the course reference material for Part Ib, data visualization). Note: *Excel and similar spreadsheet applications are not permitted*.

Part 1b. Statistical inference; Time series, spatial data, and visualization (Steve Wofsy)

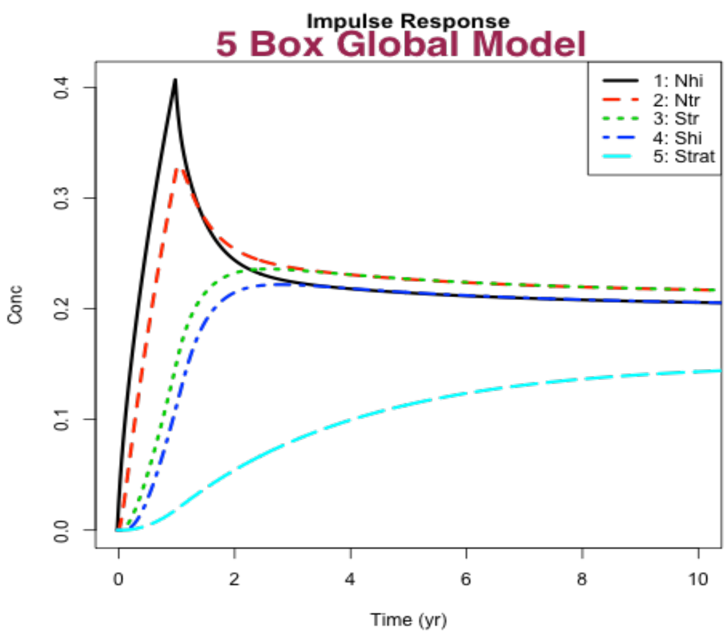
Statistical inference is an essential part of the earth and environmental sciences. Probability distributions, variance, errors, and estimation of "uncertainty" of our models and data analysis are requirements for almost every paper in the field. We start this discussion with a very close look at how statistical inference is conceptualized and applied in a field where controlled experiments and repeated resampling of the same system are impossible *a priori*.

Part 1b topics: Statistical inference in the earth and environmental sciences. Probability distributions, variance, errors, and "uncertainty".
Distributions and t tests; parametric and non-parametric regression. Analysis of data: linear regression, regressions with errors in dependent and independent variables, transformations of data; time series analysis, autocorrelated time series; error estimation: bootstrapping, correlated errors, bias, conditional sampling. Visualization of data: time series, scatter plots, missing data; smoothing and filling data using basic and advanced methods (interpolation, weighted least squares, the Savistky-Golay filter, Haar and Gaussian wavelets).

Part 1c. Hands-on Class Projects (Steve Wofsy and Teaching Fellow)

Environmental data often consist of a large number disparate observations directed towards understanding a particular phenomenon or set of phenomena. The data are often strictly incomparable in that they sample different spatial and/or temporal scales and different processes and attributes of the physical system. Examples include atmospheric trace gases measured from an aircraft, fluxes of these gases observed at points on the surface, long-term data acquired are remote stations on a weekly basis, and winds and temperatures obtained from radiosondes. We will use case studies to learn about data visualization and statistical inference in analysis of real data sets. *Class projects will be selected from various topics, using real data sets. Data sets for class projects in 2019 focus on*:

- Distributions of reactive and greenhouse gases as observed from aircraft and surface networks using a linear model framework (data from the NOAA global network) and a Lagrangian Particle Dispersion Model at global scale (data from a recent global aircraft program, the Atmospheric Tomography Experiment, see figure at the beginning of the syllabus).



Lecture topics for Part 2.
Chemical Transport Models and Inverse Modeling (Daniel Jacob)

Lectures in Part 2 focus on the construction of chemical transport models (CTMs) in the atmosphere and oceans. Topics will start with the mass continuity equation, Eulerian and Lagrangian model frameworks, numerical solution of the advection equation and of chemical mechanisms, simulation of turbulence. The second set of lectures will focus on inverse modeling methods. Topics will include the general philosophy of inverse modeling, Bayes' theorem, optimal estimation, Kalman filters, adjoint methods.

Text: G.P. Brasseur and D.J. Jacob (2015), *Mathematical Modeling of Atmospheric Chemistry* (http://acmg.seas.harvard.edu/education/brasseur_jacob/index.html)

Lecture Schedule.

Part 1a. Linear modeling of environmental systems ("box models")
Note: There will be a training session in the use of the R programming language.

Jan 30. Introduction to EPS 236 – what we will be doing and why we will be doing it
Linear Modeling Part I: time evolution operator, solutions to the general problem.

Feb 1 and Feb 6. Linear Modeling; part II Analytical properties of linear systems; Markov chain equiv. Mean Age and Age spectrum; time evolution operators, tangent linear approximations for non-linear dynamical systems; application to estimating global fluxes of atmospheric tracers.

Feb 8. Introduction to the 1st Class Data Project:
A 5-box model of greenhouse gases in the atmosphere (Workshop 1). Application of linear modeling methods, R coding skills development

Part 1b. Statistical inference, regressions, curve fitting, confidence intervals, bootstrapping, MCMC *—focus on concepts and advanced applications*

Feb 13. Statistical Inference – a close look at the fundamentals from the point of view of atmospheric, marine, and environmental Science

Feb 15. Linear regressions: Fitting a line (curve) to data;
Correlated parameters, degrees of freedom, overfitting.

Feb 20. Type II regressions, York regressions, Fitexy (Chi-sq fitting)

Feb 22. Confidence intervals, t-tests,
bootstrap error estimates, non-parametric assessment of data

1c. Time series and wavelet methods

Feb 27. Data Filtering; Classifying data smoothing methods

Mar 1. Modeling and analyzing atmospheric time series data (Workshop 2)

Mar 6. Autoregressive data; systems with serial correlation

Mar 8. Filtering and interpolation of data: wavelets and image processing

Mar 13. Filtering and interpolation: Frequency domain, FFT, spectral decomposition

Mar 13, 15. The global Atmospheric Tomography Mission
Introduction; examples of using our tools
Data Workshop 3 (short): students apply the tools
2nd Class Data Project: Emissions of CO and other pollutants from Africa

Mar 21, 23. Spring Recess

Mar 28. Special topics, tools; continuation of 2nd Class Project

Mar 30. Student presentations of 2nd Class Project
Selection and initiation of individual/group projects

Apr 3, Apr 5, Apr 10.
Hands-on class projects in groups of 2 or 3, interacting with instructors Wofsy and Zhao;
Group presentations: oral, and annotated electronic media; poster preparation.
Lectures will be interspersed to include:

- Introduction to machine learning in R.
- Continuity equation, Eulerian and Lagrangian models (introductory for Part 2a, Chemical Transport Models).

Apr 12. *Presentations of individual and group projects*

Part 2a. Chemical transport models

Apr 17. Numerical methods for advection

Apr 19. Numerical solution of chemical mechanisms

2b. Inverse modeling (Daniel Jacob)
Apr 24. Applications of inverse modeling to atmospheric problems, Bayes' theorem

Apr 26. Vector-matrix tools for inverse modeling
Analytical solution of the inverse problem

May 1. Kalman filtering and 3-DVAR data assimilation
Adjoint methods and 4-DVAR data assimilation