ATM/ENV 315 Environmental Statistics and Computation Fall 2015

Professor: Dr. Paul E. Roundy ES339A 442-4476
Office hour 1:30-2:30 p.m. Tuesday or by appointment
proundy@albany.edu
Graduate Student Assistant: Jennifer Gahtan (jgahtan@albany.edu, ES337, Office hour Monday 10-11 am or by appointment)
Meeting Time: TH 11:45-1:05. Classroom ES232

Students must download and install Enthought canopy from https://www.enthought.com/products/canopy/ or another Python analysis environment for use on their own computers or devices, as the in-class and project exercises will be done via this software. Pre-requisites: ATM210; MAT112. Recommended Course: MAT220.

GRADING: 2 Exams, 30% each. Participation and quizzes- 30%. Participation includes brief daily oral or written summary reports in which the students explain exercises from the previous class and interpret the results physically (See page 2 of this syllabus). Class project- 10%.

94-100 A, 90-93 A-, 87-89 B+, 83-87 B, 80-83 B-, etc.

Improvement greater than 15% from the first exam to the last yields 3% added to the final grade.

PURPOSE: This course is designed to prepare students from various backgrounds to succeed in applying statistical methods of data analysis for operational and research activities with a particular emphasis on *physical interpretation* of the results.

The class will include practice with Python programming. To gain basic programming experience, students are required to take the free online Python course at codeacademy.com within the first week of class. During the second week, students should listen to the "Let's learn Python-Basic" series on YouTube. Data analysis applications require specialized Python modules not included in these online courses. Students are required to familiarize themselves with the numerical Python (numpy) http://www.numpy.org numerical analysis tools and MatPlotlib http://matplotlib.org plotting libraries. Use of technological devices is necessary for the course and must be brought to class, but students should be respectful to other students in how these devices are used.

Course content will include (in the context of the Python programming environment):

1. Characterization and interpretation of atmospheric and environmental datasets, going beyond the basics of mean, median, range, standard deviation, variance, etc.
2. Spatial statistics of environmental data (correlation, autocorrelation, scaling, sampling)
3. Hypothesis testing
4. Understanding and communicating forecast probabilities for Weather or Climate events
5. Basic vector and matrix methods for discrete environmental data and a simple empirical orthogonal function (EOF) analysis (also known as Principal Component analysis)
6. Statistical modeling and forecasting of environmental data (e.g., representing Mona Loa CO_2, MOS (Model Output Statistics) temperature forecasting, and assessing forecast skill
7. Basic power spectrum analysis

and related topics, developed in the context of *real* atmospheric/environmental datasets and associated problems. Students are encouraged to create exercises of their own to better learn the material.

Exams
Exam 1 is tentatively scheduled for October 15. Exam 2 is scheduled November 24. Exams are fully electronic. Students fulfill exam exercises on their own computers and can use any previously written code or the Internet as needed, but students may not cooperate with each other on exams, either directly or through the Internet or electronic devices. Exam papers must be typed either in Word or in a PDF document, including any needed figures, and e-mailed immediately at the end of the exam session of class to proundy@albany.edu and jgahtan@albany.edu with "exam 315" in the subject line.

Class Meeting Summaries

No formal graded homework exercises will be assigned, but randomly selected students will present roughly 5-10-minute oral summaries of the previous class at the beginning of each class, with special emphasis on the projects carried out in class. While reviewing their in class activities after each class meeting, all students should each prepare a brief presentation about the in-class project(s), including a demonstration of the algorithms applied and the physical interpretation thereof. Students are encouraged to demonstrate the techniques on datasets different from those used in class, obtained from the Internet.

Each student is expected to fulfill at least four oral summaries in cooperation with one other student. *The discussion should include a demonstration of a solution to the problem presented in the previous class.* Every student should prepare to present each day, and students will be selected at random, in pairs, for the actual presentations. The first person selected is considered primary, and will connect his or her computer to the class projector. The second student will assist the first. At the end of each oral summary, the discussion is opened up to the class to revise the summary content and to discuss points that might have been left out.

Students are not expected to provide summaries for the class day before, the day of, or the class day after exams.

Summaries should include the following, dependent on the specific topic of the day:

1. Explain the problem that was the focus of the class.
2. Explain any theoretical support for the algorithms applied.
3. Explain how you solved the problem.
4. Interpret physically your results.
5. Discuss caveats of your approach and your interpretation (i.e., how might you misinterpret your results).

In the event that these oral summaries are deemed insufficient, 1-page written summaries might occasionally also be required.

Should written reports be required, they should be submitted electronically by 10:00 a.m. on the day before the next class meeting, since these are used to enhance the discussion during the next

class session. When requested, written summaries should be submitted to each of proundy@albany.edu and jgahtan@albany.edu for review, with the word "Homework" in the subject line. The professor and teaching assistant will usually return them with feedback to assist students in their studies.

Written summaries should be as *brief as possible*, but should provide enough detail to demonstrate topics that were well understood and those that need further clarification. Students should justify use of more than one face of one page.

These summaries are not graded in a strict sense, but completed summaries will be recorded in the grade book and count toward the participation grade. Summaries should show depth and insight. Each student will be granted 2 "buys" wherein they can decline the invitation to present on a given day. Although individual summaries are not graded, a pattern of apparent lack of attention to the summaries over the term will result in lower homework grades. However, wrong interpretations at the learning stage will not necessarily yield reduction in homework grades, because mistakes are a necessary part of the learning process, and student's benefit from taking chances. The instructor and teaching assistant will frequently provide feedback to help students correct poor understanding.

Friday class meetings:

Friday class meetings are intended to build student experience in programming and data analysis, and to address their questions. Graduate assistant Nick Schiraldi, who has extensive programming experience with Python language, will lead most of these 50-minute sessions.

Class Project:

As part of the departmental assessment of the success of the course, students will carry out a class project in which they analyze datasets of their own choosing, preferably connected to the student's own area of expertise. Students will present their projects in 10-minute summaries at the end of the semester. Projects should include a combination of statistical techniques applied to gain understanding of the phenomena represented in the datasets. Students will procure data from the Internet or from direct observations or models.