AT 655: Objective Analysis in Atmospheric Sciences
Course Syllabus for Spring 2020

Class: 9:00AM - 9:50AM MWF, ATS 101
Course Webpage: `https://colostate.instructure.com/courses/100690`

Objective analysis of geophysical data: basic statistics; matrix methods; filtering; time series analysis; basic machine learning methods. Emphasis on applications to real world data.

1 Instructor

Prof. Elizabeth A. Barnes

Office:	ATS 402
email:	eabarnes@atmos.colostate.edu
Office Hours:	W 10:00-11:00 and other times by appointment

2 Teaching Assistant

Ben Toms

Office:	ATS 403
email:	batoms@rams.colostate.edu
Office Hours:	??????? and other times by appointment

3 Course Focus

The course provides an overview of the methods used to interpret data sets in the atmospheric and oceanic sciences. This is a tools class: the objective is to provide a working knowledge of the statistical tools most commonly used in the field, as well as those that are up-and-coming. However, additional time will be spent on how to professionally present and objectively review scientific results and how the tools are appropriately and inappropriately used in the field. During this course you will:

- become comfortable with basic statistical techniques,
- learn to break-down real science problems into testable hypotheses,
- learn to review and assess the analysis techniques presented in the literature,
- practice writing-up scientific results in a professional manner.

4 Course Expectations

The following list presents the minimum requirements for passing this course:

- show-up to class and ask questions,
- keep up with the reading (when applicable),
- submit all assignments on time and at an acceptable level of quality,
- satisfactorily complete all quizzes.

5 Course Prerequisites

You are expected to be familiar with basic high-school and college-level mathematical concepts. Minimal time will be spent in lecture reviewing these topics:

- algebra (e.g. equations for lines, solving basic algebraic equations)
- basic calculus (e.g. how to take a derivative and an integral)
- basic matrix algebra (e.g. addition, subtraction, multiplication)
- sine and cosine functions

If you are concerned about your background in these areas, please speak with me. While the concepts, tools, and techniques explored in this course will be taught within the context of atmospheric science, there are no atmospheric science prerequisites.

You will be expected to write and implement computer code throughout this course. I do not care what software you use, but note, neither I nor the TA will not spend office hours debugging your code for you. Luke Davis will be serving as the Department Programming TA over the semester and he will be available to assist with coding questions. Please speak with Luke for specific office hours (lukelbd@gmail.com).

6 Course Web Page

The course web site will be used for posting resources and homework assignments. The course web site is through CSU Canvas and is listed at the top of this syllabus. All students enrolled in the course should have access to the Canvas material.

You will submit your assignments via Canvas by uploading a *.pdf* file of your final report for each assignment by the due date. Canvas will also allow you to keep track of your grade/points in the course.

7 Grading

7.1 Grade Break-down

Your course grade will be made up of homeworks only. That is, homeworks and class participation will together cover 100% of your grade.

7.2 Homework

There will be approximately 7 homeworks throughout this course (although I maintain the right to increase or decrease this number), each of varying length.

If you need help in completing the assignment, first ask your peers for assistance and request help from your TA and instructor second. You are *encouraged* to interact with your classmates by sharing ideas and discussing the specifics of the material and homeworks. You are, however, expected to hand-in your own homework assignment, and it should not be a direct copy of your classmate's.

Your homework assignments must be typed-up and clearly written. Figures should be of publication quality - no low-resolution figures. I repeat, *no low-resolution figures.* By doing this, you are not just being nice to the me and the TA, who have to grade your work, but you will gain practice in presenting your results clearly and professionally as required for your careers as scientists.

In addition to being clear and neat, I expect all figures to include a descriptive caption, legend (if applicable) and *labeled axes with units*! Having axes without labels or units will automatically deduct points from your homework. (You may be surprised to learn that many papers sent out to review by scientists in the field do not satisfy these basic requirements.)

7.3 Midterm & Final Exams

There will not be a midterm or a final exam in this course. However, I will give a handful of *ungraded* quizzes throughout the semester so that you can check-in on your understanding.

8 Text Books & Resources

There is no required textbook for this course. Many of the materials, notes, homeworks, etc. in this course are borrowed from the course materials of Prof. Mike Wallace and Prof. Dennis Hartmann at the University of Washington and Prof. Dave Thompson at Colorado State University. We will make extensive use of the my own typed notes, but you can also refer to those of Prof. Hartmann (which I have linked to on the course webpage). I advise you to save/print all chapters of these notes and store them somewhere easily accessible. You will likely use them long after this course is done.

There is one required resource in this course - **the internet**. Google is amazing - use it. In addition, the statistics pages on Wikipedia are very good, both the standard and advanced pages - you should use them often. One of the most important things to learn in graduate school is "how to look it up." In my own research, I use most of the techniques we will discuss, but I have very few of them memorized. By the end of this course, you should aim to be self-sufficient in finding the analysis techniques you need. You should not care whether you have a specific derivation or formula memorized, but whether you know how to find it.

Additional references you may find useful are listed on the course webpage.

9 Software

9.1 Analysis & Plotting Software

You are required to have an analysis and plotting software package (often they are one in the same) with which you can do the homeworks. I do not care what you use, but keep in mind you will want it to have basic statistical algorithms including (e.g. correlation, fast fourier transform "fft", singular value decomposition "svd")[1]. Keep in mind that all course examples will be coded in Python. 3.x. For Python, the Anaconda Scientific Python Distribution is free and easy to setup. If you are interested in using *MATLAB*, it is free to students for their home and personal computers through the College of Engineering.

9.2 *Optional: LaTeX*

LaTeX[2] is a type-set program that takes macro code and formats it into a final (often pdf) document. For example, this syllabus was written with *LaTeX*. The end result is a clean, consistently formatted document. Many scientists use *LaTeX* to write-up their research, and journals are increasingly preferring *LaTeX* files to Mircrosoft Word files for manuscript submission.

A main reason to use *LaTeX* is the ease with which mathematical symbols, equations, etc. are formatted. In addition, including figures is efficient: the user does not "cut and paste" the figure into the text, but rather places the actual document path of the figure in the *LaTeX* code. Thus, whenever the figure is changed, it is automatically updated in the manuscript file. *LaTeX* is free and can be used on all common operating systems (e.g. Linux, Mac, Windows). In addition, Overleaf.com is a fantastic site for typesetting and collaborating with *LaTeX* in the could (i.e. does not require you install anything on your computer).

I will not require that you use *LaTeX* for your homeworks, however, I highly encourage you to do so, and a handout will be provided at the beginning of the semester. While the initial learning curve is rather steep, I think that the payoff is worth it. Equation type-setting is easy and always neat, figures will be easily updated, and references are straight-forward to handle with *BibTeX*[3].

As an incentive for you to try *LaTeX*, I will provide extra credit points for those that write-up at least one of their homeworks in *LaTeX*.

10 CSU Honor Pledge

This course will adhere to the CSU Academic Integrity Policy as found in the General Catalog (`http://catalog.colostate.edu/general-catalog/policies/students-responsibilities/#academic-integrity`) and the Student Conduct Code (`http://www.conflictresolution.colostate.edu/conduct-code`). At a minimum, violations will result in a grading penalty in this course and a report to the Office of Conflict Resolution and Student Conduct Services."

[1] often, if your software package is widely available, the code will exist somewhere on the internet.
[2] pronounced "LAY-tek" or "LAH-tek".
[3] *LaTeX*'s bibliography manager.

11 Tentative Outline

The following is a tentative outline for the class. Reality will almost surely deviate from this outline, especially since I can shorten some topics if the class has topic requests.

1. **Probability & statistics**
    1.1. Basic statistics, probability distributions, Bayes Theorem
    1.2. Statistical decision making; hypothesis testing
    1.3. Basic nonparametric statistics
    1.4. Monte Carlo techniques; bootstrap; jackknife; field significance
    1.5. Compositing
2. **Regression & correlation**
    2.1. Linear regression
    2.2. Theory of correlation
    2.3. Autocorrelation/autoregressive methods; estimating the number of independent samples
    2.4. Multiple regression
    2.5. Granger Causality
3. **Time series analysis**
    3.1. Harmonic analysis; power spectra; methods of computing power spectra
    3.2. Significance testing of spectral peaks
    3.3. Data windows
    3.4. Filtering; filter design; recursive/nonrecursive filters; response functions
    3.5. Cross spectrum analysis
    3.6. Mixed Space-Time Analysis
4. **Seeking Structure in Data with Principal Components**
    4.1. Review of linear algebra
    4.2. Empirical orthogonal functions (EOF)
    4.3. Application of EOFs to real data
5. **Basic Machine Learning Methods**
    5.1. Loss functions and optimization
    5.2. Training, testing and validation data sets
    5.3. Cluster Analysis (e.g. k-means, self-organizing maps)
    5.4. Neural networks; gradient descent; backpropagation
    5.5. Convolutional neural networks
    5.6. Overview of neural network visualization
    5.7. Decision trees; Random forests
6. **Effective figures**

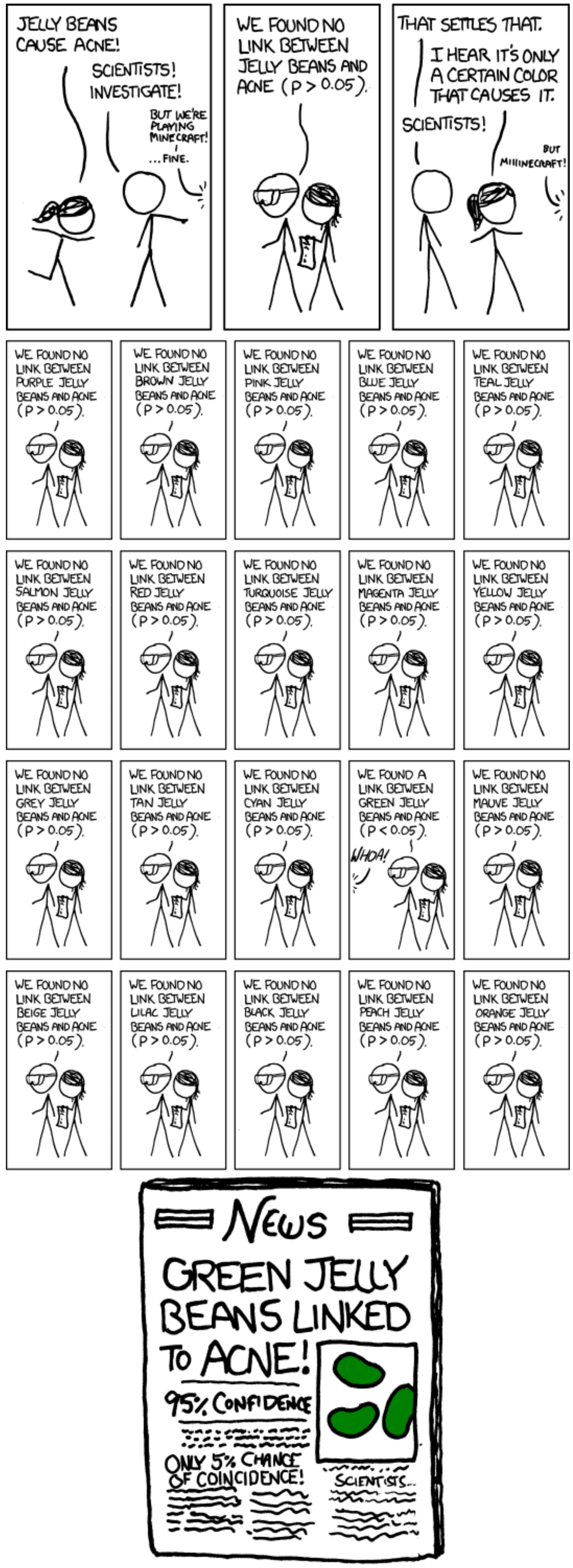

Figure 1: https://xkcd.com/882/

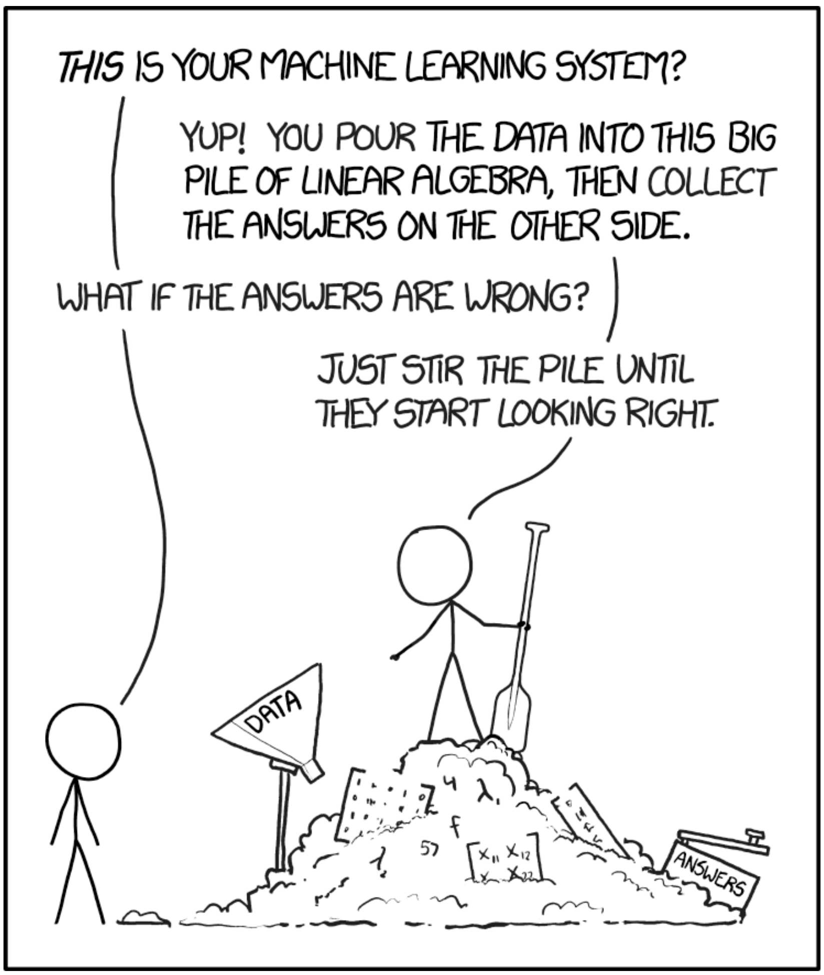

Figure 2: https://xkcd.com/1838/