MPO/ATM/OCE 624: Applied Data Analysis, Spring 2022

Instructor: Brian Mapes, office MSC 366, x5-4275 (305-421-4275)

Description (UM Bulletin): The course covers data and computer practices for careers in geosciences. Academic topics include probability & statistics, linear models and regression, spectral analysis, matrix decomposition, and machine learning basics. Strategic topics include the nature of scientific estimation and inference. Practical topics include the contemporary open-source software stack for collaborative open science. Students do projects on data from their research or interests.

Text: None required.
All materials on github.com/MPOcanes/MPO624-2022

Learning objectives for students:

(1) appreciation, appropriate level of confidence, and positive feeling for the lifetime project of mastering ADA and evidence-based reasoning

(2) vocabulary about data analysis concepts and tools, from probability (frequentist and Bayesian) to statistics (correlation and regression; Fourier analysis; hypothesis testing and 'significance') to new methods (information theory, machine learning)

(3) software skills: a can-do feeling about accessing and understanding the open-source ecosystem of computational tools and data about Earth systems

(4) a personal success (your project)

Prerequisites:
College physics and math, or permission of instructor.

Course Policies

Class Attendance:
Attendance is essential, we are a learning family sharing our progress. Absences should be communicated in advance, or the day of class if emergencies arise.

Honor Code:
Collaboration and peer learning are actively encouraged, but students are expected to follow the University of Miami's honor code (https://www.grad.miami.edu/_assets/pdf/graduate_student_honor_code_2016_2017.pdf).

Course Structure:
Three main topical units will divide the course into sections. Each class session will have a Zoom session (recorded) for screen-share purposes, consisting of:

* *Three 15-minute student-led screen presentation & discussions*
* 30 minutes of instructor-presented material

Student presentations will be assigned in advance on a rotating schedule. With 6 students, *everyone will do one 15-minute screen share per week, summarizing the most interesting results of 3-6 hours of exploration or effort*, in 3 categories:

* Foundations: computers and programming, up to Python-numpy-pandas
* Strategic: how data analysis and display are used in scientific discourse
* Middleware: cool high-level packages, examples, galleries, web sites, etc.

Grading:
1. 50% Projects: result notebook(s) from your research, internship, curiosity
2. 50% Class participation: sharing of your learning, readings, Web scoutings

Week	Topics (subject to adjustment and change)
1	Intros and software setup: github, jupyter, python (hello world)
2	Computer spinup: Jupyter notebooks and collaboration tools Readings: 1. How science works? 2. p-values explained simply.
3	Computer spinup: Python basics Readings: 2. Excerpt on the history of statistics from The Book of Why, 3. The method of multiple working hypotheses.
4	Computer spinup: Packages and galleries, working from examples Reading, The nature of quantitative models in science.
5-7	Unit 1: Probability and statistics. Estimation theory and goals, minimization of some objective function. Regression models, univariate and multivariate. Matrix algebra tools (MCA, SVD). Standard significance testing approaches and concepts. Bootstraps, jack-knives, Monte Carlo.
8-10	Unit 2: Fourier analysis. Power spectra of 1D series and sound/acoustics. Spectrograms and wavelet analysis. Spectra of 2D imagery. Fourier-space manipulations and filtering theory.
11-13	Unit 3: Generalized (nonclassical) machine learning. Terminology, examples, applications.
14	Project-specific special topics
15 + final	Projects: notebooks and presentations