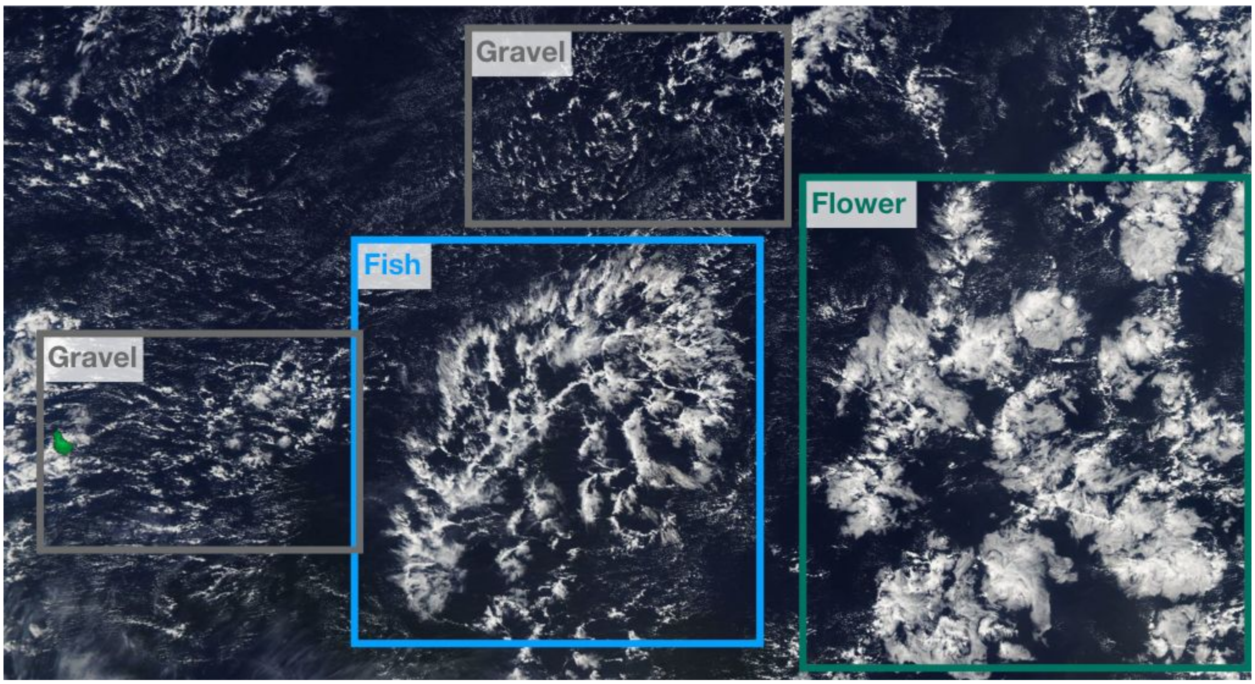
ATMS 597 SN

SYLLABUS

Weather & Climate Data Analysis (4 HOURS)



Clouds classification from a machine learning algorithm that automates the detection of the cloud patterns by learning from human labels. Image credit: [Max Planck Institute for Meteorology](#)

Instructors

Prof. Stephen Nesbitt
3060 Natural History Building
1301 W. Green Street
Urbana, IL 61801
snesbitt@illinois.edu
+1 217.244.3740
Office Hours: W, 2:00-4:30 PM

Location and Times

1071 Natural History Building
T TH, 12:30 – 1:50 PM

Prerequisites

Graduate student status.

Description

In this course, graduate students will gain experience using python language-based tools to analyze weather and climate datasets, and learn tools that will allow for the development of statistical representations of data and their uncertainties. Topics covered will include (1) object-oriented python programming, (2) version control, open source, and open data best practices, (3) using pandas and xarray for wrangling data, (4) parallelization in python and big data, (5) supervised learning, including linear, logistic, Bayesian, random forest regression techniques, and (6) unsupervised learning, including clustering, empirical orthogonal functions, and neural and deep learning.

Learning Outcomes

Upon completing this course, you will be able to …

- Perform a wide variety of data analysis, visualization, and machine learning workflows in geosciences using the open source programming language using python, with a focus on available weather and climate datasets.
- Develop end-to-end data analysis tools from scratch: locate relevant data sources and appropriate references, develop open-source analysis software, display tools in order to enable scientific hypotheses to be addressed and present information to scientists and decision makers.
- Create open source coding projects from scratch, enabling the student to learn industry standards in coding/curation practices, thus students will learn to create

fully transparent open-source data science projects, a valuable skill in data science careers.

- Understand and use a wide variety of datasets available from open data repositories, generate and curate locally-generated datasets, carrying through traceability through the analysis and publication process.
- Use open source software packages in python, use version control software to obtain, curate, and document code that you write in order to perform reproducible scientific analysis of data

Textbooks (free with open source code examples)

1. VanderPlas, Jake, 2017 *Python Data Science Handbook.* O'Reilly Publishing. Available free online: https://jakevdp.github.io/PythonDataScienceHandbook/
   Code repository here: https://github.com/jakevdp/PythonDataScienceHandbook

2. Raschka, Sebastian, and Vahid Mirjalili, 2019: Python Machine Learning, 3rd Edition, Packd Publishing. Available for free with UIUC login credentials:
   https://learning.oreilly.com/library/view/-/9781789955750/?ar
   Code repository here:
   https://github.com/rasbt/python-machine-learning-book-3rd-edition

Course Website and Tools

The Moodle learning management system will be used for this course, and is available at https://learn.illinois.edu. We will use this site for course announcements, posting course documentation, submitting course assignments, and posting

We will also use Microsoft Teams facilitate coordination of experiments and lab write-ups. The link to our Teams group is here: ATMS 597 Teams Group

Course Expectations

This is a graduate level course, and I expect you to show up prepared to learn, experiment, break and fix stuff in your code, etc. We will be spending time 'learning through doing' in much of the class, so you will have the opportunity to leave the course with tested tools in hand that you can hopefully apply to your own research. This course will be conducted to ensure that all students have an equal opportunity to learn, ask questions, and succeed. Discussions in class should be conducted in a manner to “criticize the ideas, not the person.” If you have any concerns about anything related to the class, please feel free to contact me!

Assessments and Grading

Data Science Projects and Presentation (16.67% each)	100%

Rubric for each project

Code notebook	50%
-does it function as intended?	*50%*
-is the code documented, object oriented, and efficient?	*20%*
-does it follow PEP8?	*10%*
*-project reflection write-up**	*20%*
GitHub usage	10%
Project presentation	25%
Literature review and citations	15%

*Individual for each student

Data Science Projects

Each 2-3 week course module will consist of a group project. These projects will use datasets that will be provided, and an objective to use a tool discussed in class to perform a specific type of data analysis. The groups of 2 or 3 will be rotated throughout the semester to allow for students to gain teamwork skills and share knowledge from a variety of students in the class.

More About Projects

The purpose of the guided projects and presentations in class is to give students the opportunity to gain hands on experience in data science projects, applying knowledge of applying algorithms, coding practices, and assembling code into a product, and communicating results that will prepare them for employment opportunities in scientist and data science-related positions in the atmospheric and related sciences. Projects will be built on the open version control repository GitHub. Your projects will be submitted using GitHub and showcased on your GitHub site and should include the following deliverables: 1) a documented Jupyter Notebook including all code generated for your project and a description of the data products used and created, results, and conclusions; 2) a slide deck based presentation describing the background, methodology, and results that you will present to our class.

Grading Scale

A > 95.0 > A- > 90.00 > B+ > 86.67 > B > 83.34 > B- > 80.00 > C+ > 76.67 > C > 73.34 > C- > 70.00 > D+ > 66.67 > D > 63.34 > D- > 60.00 > F; the grade of A+ is reserved for the top 5% of the class.

Course Policies

Contesting Grades

Balancing timely feedback with accurate grading can be challenging. We may make mistakes and we may even miscommunicate on the finer details of the course grading rubrics. Therefore, we encourage you to petition us with a regrading request if you feel your grade does not reflect your work quality or if there is a discrepancy between the grading rubric and your grade.

All initial re-grading requests must be made via email with a specific request (i.e. Please reexamine question 2, as I believe my explanation of the greenhouse effect was accurate.) rather than a general request (i.e. Please regrade assignment 2, I feel that I deserve a better grade based on the exemplary quality of my work.)

Submitting Late Assignments

Your late assignments will be accepted up to three days late with a late penalty of 10% per day. If an unforeseen personal event disrupts your progress in ATMS 597, please see me as soon as possible so that we can discuss your options including penalty free extensions, a withdrawal or an incomplete. You have options, and I urge you to be proactive in your pursuit of knowledge and skills and your self-advocacy.

Reporting Suspicions of Cheating

My primary goal is to ensure a fair, respectful and stimulating learning environment. Witnessing cheating is demoralizing and demotivating. If you suspect or have evidence of cheating in ATMS 597, we want to know. Students suspected of cheating will be referred to the College of LAS on a case dependent basis.

Student Accommodations

To obtain disability-related academic adjustments and/or auxiliary aids, please contact me, snesbitt@illinois.edu and/or the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES you may visit 1207 S. Oak St., Champaign, call 333-4603 (V/TTY), or e-mail a message to disability@illinois.edu. Please also inform us if you are under evaluation by DRES. We can likely implement accommodations prior to your completed evaluation.

Academic Integrity

In order to maximize your learning experience, it is best that you submit your own/your groups' own original work for a grade. In the event that you plagiarize another person's work without citation or permission, your case will be submitted to the College of LAS in accordance with LAS protocols.

The Student Code at the University of Illinois is a document that specifies your rights and responsibilities. The current version of this document is available on the web at: http://admin.illinois.edu/policy/code/.

Course Calendar

Module Name	Topics
Module 1: Data stewardship, formats; introduction to python, NumPy Weeks 1 - 2	Reproducible science, units, DOI numbers Self-describing data formats Version control, using github & creating a simple github repository Local vs. cloud computing, introduction to python Data integrity Writing python code; PEP8 Learning how to summarize literature and examine for reproducibility Review of numpy, matplotlib, cartopy Also: Gezelter (2009): http://openscience.org/what-exactly-is-open-science/

	Watch: NOVA segment on reproducibility crisis: https://www.youtube.com/watch?v=NGFO0kdbZmk&feature=youtu.be **Project 1:** Due January 30, 2020
Module 2: pandas Weeks 3 - 4	Introducing Pandas Objects Data Indexing and Selection Operating on Data in Pandas Handling Missing Data Hierarchical Indexing Combining Datasets: Concat and Append Combining Datasets: Merge and Join Aggregation and Grouping Pivot Tables Vectorized String Operations Working with Time Series High-Performance Pandas: eval() and query() **Project 2:** Due February 13, 2020
Module 3: xarray Weeks 5 - 6	Plotting Data structures Indexing and selecting data Interpolating data Computation GroupBy Reshaping and reorganizing data Combining data Time series data Weather and climate data Working with pandas Reading and writing files Parallel computing with dask **Project 3:** Due February 27, 2020
Module 4: Introduction to machine learning, supervised learning Weeks 7 - 8 - 9	Linear regression Logistic regression Decision tree regression Random forest regression Support vector machines Empirical orthogonal functions Using regression and EOFs for physical interpretation

	Project 4: Due March 26, 2020
March 16-20	**SPRING BREAK**
Module 5: Machine Learning 2 Weeks 10 – 11 - 12	Bayes Theorem Naive Bayes classification k-Means clustering Gaussian mixture models Artificial neural networks Deep learning networks **Project 5:** Due April 16, 2020
Module 6: Machine learning 3 Weeks 13 - 14 - 15	Image classification and detection Convolutional neural networks General adversarial networks **Project 7:** Due May 5
Final Examination	**None**