

# Advanced Data Analysis for Earth Science

## What is this course?

---

“Advanced Data Analysis for Earth Science” is a graduate-level programming and data analysis course aimed at giving students in environment-related disciplines—like Geology, Atmospheric Science, Physical Geography, Environmental Science, etc.—the technical skills necessary to do high-quality science.

## What will I learn?

---

The main goal of the course is that you leave the course with a strong understanding of how to do rigorous, reproducible, and high-quality programming and data analysis in support of research. This involves developing a combination of technical skills and programming habits.

Specifically, you will be able to:

1. summarize fundamental programming concepts and how to apply them with Python
2. implement common environmental data analysis tasks, such as reading and writing data, computing basic statistics, curve fitting, and spectral analysis
3. apply machine learning libraries to real data
4. parallelize *embarrassingly parallel* data processing tasks
5. generate informative publication-quality plots
6. write code in a modular, reusable way
7. design and employ unit tests to minimize the chances that buggy data makes its way into peer-reviewed publications
8. utilize version-control software to maintain research code libraries
9. write data analysis code both in scripts and in Jupyter Notebooks

Note: while the above may sound like a lot for one course, the last four items are topics that will be interwoven through every class.

## How will I learn?

---

This is a lab-based course, in which you will do a combination of in-class exercises to reinforce concepts and also work to apply concepts learned in class to a final data analysis project (ideally one that is related to your own research!). The class will be one three-hour lab session each week, which will include a small amount of lecture and live demonstration time.

Learning will be reinforced by weekly readings and/or videos. There will also be weekly homework assignments that (a) prompt you to summarize the weekly readings/videos, and (b) have you apply concepts from class to real data (in most cases this will be your own research data, but I will provide alternatives if that turns out not to be practical).

## Who is teaching it and how can I reach them?

---

Prof. Travis A. O'Brien (he/him/his) - I'm a scientist who leads a group of undergraduate, graduate, and postdoctoral researchers focused on understanding what controls year-to-year variability in extreme weather and how this is being affected by climate change (see [my Google Scholar page](#)). I have been programming in general

since 1995 and programming in the sciences since 2005, learning many programming languages along the way: BASIC, C, C++, Fortran, Perl, NCL, Matlab/Octave, BASH, HTML, LaTeX (believe it or not, LaTeX is [Turing complete](#)!), Javascript, Julia, and of course Python. If it is not obvious from that list, I love programming—it's sort of meditative for me, and it's something I do as a hobby in addition to what I do for work. Along the way, I have written a Python package that is used in numerous fields (see [fastKDE](#)) and contributed to other software packages (see [my github page](#))

I'll post formal office hours at the beginning of the semester, and in general, if my office door is open, you're welcome to drop by to chat any time: Geology 4037.

## How will I be graded?

---

There will be three primary areas that contribute to your grade:

- (40%) active participation in lab sessions
- (40%) weekly homework assignments
- (20%) final project

### Active participation in lab sessions

One of the primary ways that I will assess your learning is through direct interactions with you in weekly lab sessions. In general, I'll assign full points for each day that you attend and actively participate in the lab: as long as you're making a clear attempt to learn the material. I may deduct points for a given day if it seems that substantial time was spent not engaged in learning for the course (e.g., if you're doing homework for another class, etc.).

If for some reason you are unable to attend lab on a given day, reach out to me in advance and we will work out an alternative.

### Weekly homework assignments

Weekly homework assignments will generally have two parts: (1) a reading summary, and (2) application of new concepts to real data and code. The homework assignments will always be distributed with a rubric, so the basis for assigning a grade will be clear.

The assignments will have a clearly-stated set of objectives for you to meet: e.g., “generate a multipanel figure containing a contour plot and a timeseries”, or “create a git repository and push at least two commits”. The rubric will be based on assessing the extent to which those objectives are met.

In some instances, I may also request peer review, which will also count for points in the homework category.

### Final project

The final project will consist of research code and Jupyter notebooks in a git repository, along with a well-written ‘README’ file that orients the viewer to the contents of the repository. Specific requirements for the final project will be given later in the course, but the general goal of the project is to (a) demonstrate that concepts in the course have been applied in your research, and (b) demonstrate that all of the learning goals of the course (see “what will I learn?”) have been met. A rubric will be provided when the final project is assigned, and peer reviews may be requested.

The final project will be due no later than the end-time of the final exam that would be scheduled for this class, and it will be due no earlier than midnight of the Sunday before finals week.

## How are assignments submitted?

Specific submission instructions will be provided for each assignment; (I'll typically use Canvas).

In-class discussion exercises will be collected and evaluated in class.

## How will I know how I'm doing in this course?

Grades will usually be posted in Canvas within two weeks of the due date. Since Canvas keeps track of all of your grades, you should always be able to calculate your current grade in the course. If you need assistance, please contact me.

## What about late work?

All late assignments will be subject to a 40% penalty unless arrangements are made in advance of the due date: the grade for late assignments will therefore be calculated as 0.6 times the actual grade given. Assignments turned in after the last day of the semester will not be counted unless arrangements have been made for an incomplete, consistent with [IU policies on incompletes](#).

## What are the prerequisites?

---

This course has no formal prerequisites, though it would probably be best if you have at least some programming experience (even if, for example, its just that you used matlab in an engineering course once): this will be a fast-paced course, and it may be difficult to keep up if this is the first time you've done any programming. Reach out to me if you have any concerns.

## What materials do I need for this course?

---

- \* A laptop, for use in lab
- the course textbook: [An Introduction to Python Programming for Scientists and Engineers](#), ISBN 9781108701129 or 9781108571531 (digital)

\* if you don't have a laptop, reach out to me and we will figure something out.

## Why Python?

---

*Why Python and not [Lisp, Fortran, Matlab, C++, Julia, etc.]?* There are a number of reasons I'm choosing Python as the language for this class:

1. It is free
2. It is used by many scientists and engineers
3. It could help you get a data scientist job later on
4. There is an enormous community of people who offer code—for free—that you can easily make use of
5. Once you know one programming language, you can pick up another pretty easily, so why not start here?
6. I have about 12 years of experience with Python on a nearly daily basis
7. It can be used as a 'glue' language - e.g., making use of external code written in Fortran, C/C++, R, Julia, and probably some others I'm not immediately thinking of.

## What is expected of me?

---

In this class, communication and consistent effort are absolutely important.

**Communication:** Please keep me informed if you are facing any challenges, if something about the class is not working well for you, if you have questions, if you want to talk about your career, etc. The pandemic continues to be challenging, and I welcome (and urge) you to contact me at any time. My preferred mode of contact is canvas e-mail, and I'm generally very fast to respond. I'll work with you to set up a time for a meeting.

**Consistency:** Doing well in this class requires consistently engaging in the class: doing homework assignments, attending labs, engaging in discussions, etc. The readings and homework assignments are generally moderate, but they're frequent and can really add up. If you start to fall behind, don't worry: I want to make sure you can succeed in this class. Reach out to me and we can work on a plan to get back on track.

**Effort:** This class will require consistent effort. You should anticipate needing to spend at 6–8 hours \* each week in addition to lecture.

\* If you are finding that you're consistently needing to spend more than 6–8 hours each week on work outside lecture time, please let me know!

## What if I need accommodations?

---

Every attempt will be made to accommodate students with disabilities (e.g. mental health, learning, chronic health, physical, hearing, vision, neurological, etc.). You must have established your eligibility for support services through the appropriate office that services students with disabilities. Note that services are confidential, may take time to put in place, and are forward moving. Captions and alternate media for print materials may take three or more weeks to get produced. Please contact the Disability Services for Students (DSS) at 812-855-7578 as soon as possible if accommodations are needed. The office is located on the third floor, west tower, of the Wells Library, Room W302. Walk-ins are welcome, 8 a.m. to 5 p.m. Monday through Friday. You can also locate a variety of campus resources for students and visitors who need assistance at [iu.edu/~ada](http://iu.edu/~ada).

## What is the grading policy?

---

The official breakdown of percentages is given in “How will I be graded?” and the following details the policy for how a final grade will be assigned. It is written in great detail so that there is no question about how/when grades are assigned, augmented, or reduced. At my discretion, there may be exceptions to the following policies to accommodate extenuating circumstances, but exceptions will only be granted if they are requested before a given assignment is due.

During the semester, I may consider modify the basis of the grading for the course, but I will only do so after talking with you as a class, allowing some form of anonymous input (e.g., a google form), and holding a formal and unanimous vote. If I seek to change the grading basis, it will only be to benefit you in some way, so it is important to me that I have your input to make sure that any change is in fact beneficial.

### Final Grade

The final grade for the course will be the sum of percentages from each of the above categories, with any extra credit applied. No individual assignment or exam will be curved. At the end of the course, and at my discretion, the grading scale may be adjusted (only downward) such that the threshold for receiving a given grade is lowered. The grading scale will not be shifted to the detriment of any student. Percentages will be rounded prior to assigning grades. The following table provides the highest thresholds that will be applied to achieve a given letter grade for the course:

98-100% A+

93-97% A

90-92% A-

87-89% B+

83-86% B

80-82% B-

77-79% C+

73-76% C

70-72% C-

67-69% D+

63-66% D

60-62% D-

0-59% F

## Are there any class policies?

---

Students are expected to adhere to the IU [Code of Student Rights, Responsibilities, and Conduct](#). Any work in this course that is deemed to be a product of academic dishonesty will receive an automatic 0, and I will refer any involved students to the Office of Student Conduct. The following outlines some of the course-specific policies regarding academic honesty.

**Academic Integrity:** Academic misconduct is defined as any activity that tends to undermine the academic integrity of the institution and includes: cheating, fabrication, plagiarism, interference, violation of course rules and facilitating academic dishonesty. When you submit an assignment with your name on it, you are signifying that the work contained therein is yours, unless otherwise cited or referenced. Any ideas or materials taken from another source for either written or oral use must be fully acknowledged. Sharing or receiving information about the content of assignments via apps such as GroupMe constitutes academic misconduct. In addition, posting answers to quizzes/exams or assignments or downloading/copying them from online sources is considered academic misconduct. All suspected violations of the Code will be reported to the Dean of Students (Office of Student Conduct) and handled according to University policies. Sanctions for academic misconduct may include a failing grade on the assignment, reduction in your final course grade, and a failing grade in the course, among other possibilities. If you are unsure about the expectations for completing an assignment or taking a test or exam, be sure to seek clarification from your instructor in advance.

**Homework:** Students are strongly encouraged to work in groups on homework assignments, however all work that is turned in must be entirely the student's own. Specifically, I draw the ethical/unethical line for homework assignments at direct copying of assignments. If I find that answers from two or more homework assignments are implausibly identical, all assignments involved will receive an automatic 0 and the student(s) will be referred to the Office of Student Conduct.

**Written Assignments:** The same policies for homework assignments apply to any course projects. In addition, plagiarism should be strictly avoided. Students should be sure to cite any sources used and to quote any text (paraphrased or direct quoted) that is not the student's own. Students should consult with me if uncertain about what constitutes [plagiarism](#). If a written assignment is found to be plagiarized in any way, that assignment will receive an automatic 0 and the student(s) will be referred to the Office of Student Conduct.

Note selling: Several commercial services have approached students regarding access to class notes/study guides/exams. Providing instructor's notes/study guides or uploading course assignments to these sites in exchange for access to materials for other courses is not permitted. Violations of this policy will be reported to the Dean of Students as academic misconduct (violation of course rules). Sanctions for academic misconduct may include a failing grade on the assignment for which the notes/study guides or assignments are being uploaded, a reduction in your final course grade, or a failing grade in the course, among other possibilities. Additionally, you should know that selling a faculty member's notes/study guides individually or on behalf of one of these services using IU email or via Canvas may also constitute a violation of IU information technology and IU intellectual property policies; additional consequences may result.

Online course materials: The faculty member teaching this course holds the exclusive right to distribute, modify, post, and reproduce course materials, including all written materials, study guides, lectures, assignments, exercises, and exams. Some of the course content may be downloadable for students who may only have intermittent access to the internet, but you should not distribute, post, or alter the faculty member's intellectual property. While you are permitted to take notes on the online materials and lectures posted for this course for your personal use, you are not permitted to re-post in another forum, distribute, or reproduce content from this course without the express written permission of the faculty member. Online class or discussion sessions should not be recorded individually by students and recordings of class sessions should be reposted or distributed in any other forum. Violations of this course rule will be reported to the appropriate university offices and officials, including to the Dean of Students as academic misconduct.