Syllabus
Scientific Data Analysis
(Physics 4750)
Spring Semester 2026

Instructor Information:

Instructor:	Prof. Jens Oberheide http://globaldynamics.sites.clemson.edu/index.html
Office:	103 Kinard Lab
Department:	Dept. of Physics and Astronomy, CU
Phone:	864-656-5163
Email:	joberhe@clemson.edu
Office Hours:	Tuesday 2–3 pm. Additional appointments can be made through email on an as needed basis.
Class Hours:	MWF 9:05–9:55, Kinard 223; you may leave class if Prof. Oberheide has not arrived after 15 minutes

Course Information

Course Rationale: A critical element of a robust education in STEM disciplines is an understanding of sets of numbers and how to process, plot, and compare them. This concept is usually first taught in introductory laboratory courses and reinforced in the advanced lab classes. These labs, however, are usually focused on the scientific concept being explored by the experiment as well as the methodologies of setting up the equipment and making the measurements. These classes often only give a brief glimpse into the many techniques for analyzing the obtained number sets, and these courses often devote less attention to how one should visualize the number sets and compare number sets. Uncertainties surrounding observations and model results govern the interpretation of the values and especially the comparison of several values. A related concept is uncertainty propagation, keeping track of the uncertainties as the values are processed (e.g., used as a value in an equation to yield a new number). Without a grasp on the uncertainty of a given number, its comparison with other numbers is meaningless.

This course will provide a comprehensive introduction to data analysis, visualization, and data-model comparisons and metrics, within the framework of the uncertainty around the values. Using Python for analysis, a lot of the examples come from Earth, atmosphere, space and planetary sciences but the tools are directly applicable to all fields of physics. While no prior experience with Python is required, basic familiarity with programming is highly recommended.

Objectives/Learning Outcomes: By the completion of the course, students will be able to: read and write data sets using Python, perform large data set analysis, hypothesis testing and model goodness-of-fit quantification, and produce publication-ready scientific data visualization.

Course Outline: The course follows a new textbook and will cover the twelve chapters.

- Chapter 1: Uncertainty around data, comparing a single number to a group, and the Gaussian "normal" distribution
- Chapter 2: Visualizing a data set, elements to consider when making a plot, and best practices for conveying a message with a figure
- Chapter 3: Calculating the uncertainty of a processed data set
- Chapter 4: Quantifying the centroid and spread of a number set, Poisson counting statistics in relation to centroids and spreads
- Chapter 5: Methods for determining if a number set follows a Gaussian distribution, including the chi-squared test and Kolmogorov-Smirnov test, rejecting single data points from a set, what to do if it isn't Gaussian
- Chapter 6: Comparing two data sets, t tests, covariance, correlation coefficients, and calculating an uncertainty of a metric with the jackknife and bootstrap methods
- Chapter 7: Fitting a line between two paired data sets, including polynomial fitting and nonlinear curve fitting techniques
- Chapter 8: Visualization techniques for comparing a data set with a model trying to reproduce it, approaches to data-model comparison metrics, categories of metrics
- Chapter 9: Fit performance metrics (those that use the continuous nature of the two number sets)
- Chapter 10: event detection metrics (those that categorize the data and model values into event/non-event status)
- Chapter 11: techniques for sliding the threshold of event identification
- Chapter 12: summary of best options for data-model comparison metrics choices for certain applications, maximizing interpretive value with combinations of metrics, binomial distribution and decisions with metrics, and introductions to additional statistical topics

Method of Teaching: This is a lecture course based upon a new textbook. You will prepare by reading the chapters. We will also engage in a number of exercises. Assignments are administered through Canvas.

Grading: Assignments (all through Canvas) in this course are divided into these general categories, which carry the following weight in your final grade calculations:

Category	Weight
Homework assignments – mostly weekly	40%
Project – a long, free-form homework at the end of the class	20%
Midterm exam – in-class exam	20%
Final exam – in-class exam	20%

I expect to have 9 homework assignments. One will be by-hand problems, on uncertainty propagation, but the others are programming assignments, submitting Jupyter notebooks with explanatory text and Python code. You may have some extra work to do in the "text" sections of the notebook. The project is essentially a long homework, but this time with slide-set report (in addition to the Notebook) and flexibility on the topic and data set. The lowest homework grade will be dropped before the final grade is computed.

Student collaboration: When doing homework assignments and the project, I encourage collaboration and peer tutoring. Please help each other learn the material and get through the work. You can even help each other edit and hone your Python code. When it comes to actually typing up the submission, though, I expect each of you to do your own work. You learn very little by copying another's answers. For the midterm and final exams, you will do your own independent work and submission without any input or aid from others.

Use of AI: For homeworks and projects, you are allowed to use generative large language model services (such as ChatGPT) to produce work products. Please provide, as an appendix to any assignment (separate doc or PDF from the code file, please), a log of all inputs and relevant outputs to the text/code generation engine.

Grades: A: 90-100%; B: 80-89%; C: 70-79%; D: 60-69%; F:<60%

This class will not be curved. Rounding will be applied at the two-decimal-place digit, so a 89.50% is an A and a 89.49% is a B.

Late policy: Homework assignments and the project are expected to be **submitted by 11:59 pm** on the listed due date. Assignments submitted until 11:59 pm the next evening will be considered **late and reduced by 10% (off the full possible value)**. After 1 more day, the grade is **reduced another 10%.** After two days, the assignment will not be graded. I will be reasonable with no-reduction extensions but, except for very rare cases, excused late submissions **must** be requested ***before*** the due date and time (except for emergencies).

Extra credit: There will be two opportunites for optional extra credit near the end of the course.

One is turning in the receipt acknowledging that you filled out the course evaluation. If you upload a screen shot/pic/PDF of the page showing that you submitted it, then you will receive **2% extra towards your overall course grade**. This will not be reflected in Canvas but will be added afterwards. Make sure that your submission does not show ratings/comments. This is a new course and I am considering to make this a regular elective. I thus highly value your feedback about the course and look forward to reading your comments to make it better.

The second is presenting your project slide set to the class. As of the plan right now, we will do this on the second-to-last day of class. The content itself is not graded or assessed, but feedback might be provided, by either me or students. This is also worth **up to 2% extra towards your overall course grade**, depending on how it satisfies the required elements as defined in the project description.

Web Sites: The main course web site is on Canvas, accessible at https://clemson.instructure.com/courses/286452. Announcements, and assignments will be posted here.

Required Textbook:

- Michael W. Liehmohn. Data Analysis for the Geosciences: Essentials of Uncertainty, Comparison, and Visualization, Wiley, E-book version: ISBN: 978-1-119-74789-5 ($120), Paperback version: ISBN: 978-1-119-74787-1 ($150)

Course Policies

The following policies are the standard syllabus material. Please read completely.

Prerequisites: No formal prerequisite. However, we will do derivatives and some programming experience is expected.

Attendance Policy: I won't check attendance but expect regular attendance. I reserve the right to drop students with excessive absences (more than 4) from the course.

University Policies

Academic Integrity: The Clemson University statement on academic integrity reads: "As members of the Clemson University community, we have inherited Thomas Green Clemson's vision of this institution as a "high seminary of learning." Fundamental to this vision is a mutual commitment to truthfulness, honor and responsibility, without which we cannot earn the trust and respect of others. Furthermore, we recognize that academic dishonesty detracts from the value of a Clemson degree. Therefore, we shall not tolerate lying, cheating or stealing in any form."

All infractions of academic dishonesty by undergraduates must be reported to Undergraduate Learning for resolution through that office. In cases of plagiarism instructors may use the Plagiarism Resolution Form.

Plagiarism, which includes the intentional or unintentional copying of language, structure, or ideas of another and attributing the work to one's own efforts. Graded works generated by artificial intelligence or ghostwritten (either paid or free) are expressly forbidden. See the Undergraduate Academic Integrity Policy website for additional information and the current catalog ("Academic Regulations" section) for the policy. Send questions to UGSintegrity@clemson.edu.

Student Accessibility Statement: Clemson University values the diversity of our student body as a strength and a critical component of our dynamic community. Students with disabilities or temporary injuries/conditions may require accommodations due to barriers in the structure of facilities, Students who experience a barrier to full access to this class should let the instructor know and are encouraged to request accommodations through SAS (Student Accessibility Services) as soon as possible. To request accommodations through SAS, please see this link: www.clemson.edu/academics/student-accessibility-

services/how-to-register/requesting-accommodations. You can also reach out to SAS with questions by calling 864-656-6848, email CUSAS@clemson.edu or visiting SAS at the ASC Suite 239. Contact the office for the most updated drop-in schedule if you would prefer not to schedule an appointment.

Clemson University Title IX Statement: Clemson University is committed to a policy of equal opportunity for all persons and does not discriminate on the basis of race, color, religion, sex, sexual orientation, gender, pregnancy, national origin, age, disability, veteran's status, genetic information or protected activity (e.g., opposition to prohibited discrimination or participation in any complaint process, etc.) in employment, educational programs and activities, admissions and financial aid. This includes a prohibition against sexual harassment and sexual violence as mandated by Title IX of the Education Amendments of 1972. This policy is located at http://www.clemson.edu/campus-life/campus-services/access/title-ix/. Ms. Alesia Smith is the Clemson University Title IX Coordinator. She is also the Assistant Vice President of Equity Compliance. Her office is located at 223 Brackett Hall, 864.656.3181. Remember, email is not a fully secured method of communication and should not be used to discuss Title IX issues.

Emergency Preparedness Statement: Emergency & Safety Procedures have been posted in all buildings and on elevators. Students should be reminded to review these procedures for their own safety. All students and employees should be familiar with guidelines from Clemson University Public Safety. Clemson University is committed to providing a safe campus environment for students, faculty, staff, and visitors. As members of the community, we encourage you to take the following actions to be better prepared in case of an emergency: 1. Ensure you are signed up for emergency alerts (CU Alerts). Alerts are only sent whenthere is a potential threat to safety, a major disruption to campus services, and for once-monthly tests. 2. Familiarize yourself with all possible exits, safer locations, and other key information onthe emergency evacuation maps in this building and those that you visit regularly. 3. Make a plan for how you would Run, Hide, and Fight in case of an active threat in thisbuilding and those that you visit regularly. Run – What are all the possible exits in this building, and the routes to them? Hide – What are the potential hiding locations in this room and building that are outof sight of doors and windows, how do you lock the door(s), how would youbarricade the door(s) and windows, and where do you turn off the lights? Fight – What tools are available in this room and building, should you have to fight? 4. Learn what you can do to prepare yourself for the hazards that affect our locations(https://www.clemson.edu/cusafety/emergency-management/emergency-procedures/index.html). 5. Download the Rave Guardian app to your phone (https://www.clemson.edu/cusafety/cupd/rave-guardian/).