COLUMBIA UNIVERSITY
School of Professional Studies

Master of Science in Sustainability Science

SUSC 5060 Statistics, Data Analysis, and Coding for Sustainability Science
Wednesdays, 6:10–8:00 PM
3 credits
Core course in Area 1 (Integrative Courses in Sustainability)

Instructor:	James L. Davis, Lamont Research Professor, Lamont-Doherty Earth Observatory, jdavis@ldeo.columbia.edu
Office Hours:	By appointment
Response Policy:	Instructor is available for short discussions on e-mail, with response within 1–2 business days. Longer discussions should be scheduled.
Teaching Assistant:	Storm Heidinger, csh2177@columbia.edu
Office Hours:	By appointment
Response Policy:	Will respond to request for scheduled Zoom meeting within 1–2 business days.

Course Overview

Students in the Master of Science in Sustainability Science will encounter a range of scientific problems throughout their Science-specific courses that require a strong foundational level of mathematical and statistical knowledge. In addition, coursework will involve computer coding to read, analyze, and visualize data sets. This course provides an overview of essential mathematical concepts, an introduction to new concepts in statistics and data analysis, and provides computer coding skills that will prepare students for coursework in the Master of Science in Sustainability Science program as well as to succeed in a career having a sustainability science component. In addition to an overview of essential mathematical concepts, the skills gained in this course include statistics, and coding applied to data analysis in the Sustainability Sciences. Many of these skills are broadly applicable to science-related professions, and will be useful to those having careers involving interaction with scientists, managing projects utilizing scientific analysis, and developing science-based policy. Students enrolled in this course will learn through lectures, class discussion, and hands-on exercises that address the following topics:

1. Mathematical concepts in calculus, trigonometry, and linear algebra.
2. Mathematical concepts related to working on a spherical coordinate system (such as that for the Earth).
3. Probability and statistics, including use of probability density functions to calculate expectations, hypothesis testing, and the concept of experimental uncertainty.
4. Concepts in data analysis, including linear least squares, time-series analysis, parameter uncertainties, and analysis of fit.
5. Computer coding skills, including precision of variables, arrays and data structures, input/output, flow control, and subroutines, and coding tools to organize and manipulate data, produce basic X-Y plots as well as images of data fields on a global map.

Instruction and coding assignments will utilize Python, but the basic coding concepts taught in the course are of wide applicability.

An undergraduate background in any field of science or engineering is required, as is expected for students in the MS in Sustainability Science Program. This course is approved to satisfy the Area 1 – Integrative Courses in Sustainability curriculum area requirement for the M.S. in Sustainability Science program.

Learning Objectives

By the end of this course, students will be able to:

L1: Utilize basic mathematical skills for use in solving scientific problems in sustainability science.

L2: Understand basic concepts in probability and statistics and their relationship to real-world data.

L3: Use the linear least-squares technique to estimate model parameters, evaluate parameter uncertainties, and assess post-fit models.

L4: Apply basic coding techniques to read a variety of data files in common scientific formats.

L5: Perform data analysis on complex data sets and produce plots and images of calculated computer models and data sets.

Readings

Core Text:

Hill, C. (2020), *Learning Scientific Programming with Python, Second Edition*, Cambridge University Press, ISBN 978-1-108-74591-8.

On-line resources:

Spyder documentation: https://docs.spyder-ide.org/current/index.html

Ryan & Abernathy, *An Introduction to Earth and Environmental Data Science*: https://earth-env-data-science.github.io/intro

Online Python reference: https://docs.python.org/3/reference/

NumPy Documentation: https://numpy.org

PyGMT Documentation: https://www.pygmt.org/latest/

Assignments and Assessments

Class Participation (5%) (L1, L2, L3, L4, L5)

Class participation, including oral and written communication, exercises important job and life skills. Assigned readings must be completed before class. Classes will begin with an interactive overview lecture and include class discussions.

In-Class Coding Exercises (15%) (L1, L2, L3, L4, L5)

Some lectures will be followed by in-class coding exercises. Students will be expected to participate in the coding exercises and to contribute to the class discussions that follow.

Problem Sets (80%) (L1, L2, L3, L4)

Students will be assigned four take-home coding problem sets, which will enable the students to exercise their problem-solving abilities using the mathematical and coding concepts covered in the course.

Grading

The final grade will be calculated as described below:

Grade	Percentage
A+	98–100 %
A	93–97.9 %
A-	90–92.9 %
B+	87–89.9 %
B	83–86.9 %
B-	80–82.9 %
C+	77–79.9 %
C	73–76.9 %
C-	70–72.9 %
D	60–69.9 %
F	59.9% and below

ASSIGNMENT	% Weight
Class participation	5
In-class coding exercise participation	15
Problem set 1	20
Problem set 2	20
Problem set 3	20
Problem set 4	20

Course Policies

Participation and Attendance
You will be expected to come to class on time and prepared. It is understandable that new concepts to which you are introduced in the readings may not be fully absorbed; you will not be penalized for having an imperfect understanding of the concepts in the readings but you will be expected to ask questions during class discussion. Attendance is thus very important, and more than one absence will affect your grade. Lecture notes will be made available on the day after class.

Late work
Work that is not submitted on the due date noted in the course syllabus without advance notice and permission from the instructor will be graded down 3% for every day it is late.

Classwork/Labs
On some class days, students will be assigned coding problems in class and have class time to work on these prior to group discussion. Students should bring their laptops to class.

Homework
Problem sets and the coding application problems must be worked on outside of class. Collaboration on problem sets among students is encouraged, but students must turn in their own assignments.

Use of AI, instructor-provided code, and on-line resources
An important aspect of coding is the adaptation of existing code to address an assigned problem. Students may use snippets of code obtained from on-line sites, AI-developed code, examples of code provided in the lectures, and other sources. Students are ultimately responsible for meeting the rubric of assigned problems, and are subject to the Academic Integrity policies of Columbia University.

School and University Policies and Resources

Copyright Policy
Please note—Due to copyright restrictions, online access to this material is limited to instructors and students currently registered for this course. Please be advised that by clicking the link to the electronic materials in this course, you have read and accept the following:

> The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted materials. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

Academic Integrity
Columbia University expects its students to act with honesty and propriety at all times and to respect the rights of others. It is fundamental University policy that academic dishonesty in any guise or personal conduct of any sort that disrupts the life of the University or denigrates or endangers members of the University community is unacceptable and will be dealt with severely. It is essential to the academic integrity and vitality of this community that individuals do their own work and properly acknowledge the circumstances, ideas, sources, and assistance upon which that work is based. Academic honesty in class assignments and exams is expected of all students at all times.

SPS holds each member of its community responsible for understanding and abiding by the SPS Academic Integrity and Community Standards posted at https://sps.columbia.edu/students/student-support/academic-integrity-community-standards. You are required to read these standards within the first few days of class. Ignorance of the School's policy concerning academic dishonesty shall not be a defense in any disciplinary proceedings.

Diversity Statement
It is our intent that students from all diverse backgrounds and perspectives be well-served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that the students bring to this class be viewed as a resource, strength and benefit. It is our intent to present materials and activities that are respectful of diversity: gender identity, sexuality, disability, age, socioeconomic status, ethnicity, race, nationality, religion, and culture.

Accessibility
Columbia is committed to providing equal access to qualified students with documented disabilities. A student's disability status and reasonable accommodations are individually determined based upon disability documentation and related information gathered through the intake process. For more information regarding this service, please visit the University's Health Services website: https://health.columbia.edu/content/disability-services.

Class Recordings
All or portions of the class may be recorded at the discretion of the Instructor to support your learning. At any point, the Instructor has the right to discontinue the recording if it is deemed to be obstructive to the learning process. If the recording is posted, it is confidential and it is prohibited to share the recording outside of the class.

SPS Academic Resources
The Division of Student Affairs provides students with academic counseling and support services such as online tutoring and career coaching: https://sps.columbia.edu/students/student-support/student-support-resources.

Columbia University Information Technology
Columbia University Information Technology (CUIT) provides Columbia University students, faculty and staff with central computing and communications services. Students, faculty and staff may access University-provided and discounted software downloads.

Columbia University Library
Columbia's extensive library system ranks in the top five academic libraries in the nation, with many of its services and resources available online.

The Writing Center
The Writing Center provides writing support to undergraduate and graduate students through one-on-one consultations and workshops. They provide support at every stage of your writing, from brainstorming to final drafts. If you would like writing support, please visit the following site to learn about services offered and steps for scheduling an appointment. This resource is open to Columbia graduate students at no additional charge. Visit http://www.college.columbia.edu/core/uwp/writing-center.

Career Design Lab
The Career Design Lab supports current students and alumni with individualized career coaching including career assessment, resume & cover letter writing, agile internship job search strategy, personal branding, interview skills, career transitions, salary negotiations, and much more. Wherever you are in your career journey, the Career Design Lab team is here to support you. Link to https://careerdesignlab.sps.columbia.edu/

Fall 2023 Course Schedule/Course Calendar

Class Date	Topics and Activities	Readings (due before class on this day)	Assignment handed out
9/6	Directory organization. Numbers, variables, types, precision, relational expressions, assignments, methods, and attributes. Setting up and using the Spyder IDE.	Hill, Chapter 1, §2.1–2.2	**Problem Set 1 (due 9/24)**
9/13	Objects: Strings, lists, tuples, loops, list comprehension, control flow.	Hill, §2.3–2.5	
9/20	Input/output, intro to functions, basic graphs and plots with Matplotlib.	Hill, §2.5–2.7	
9/27	Errors, exceptions, dictionaries, Python packages; Classes and object-oriented programming. Using on-line documentation.	Hill, Chapter 4	**Problem Set 2 (due 10/15)**
10/4	Linear algebra and basic calculus. NumPy: Numerical arrays, the ndarray type, array math, polynomials, calculus.	Hill Chapter 6	
10/11	Data structures and data organization in Python. xarray and NetCDF files for climate and other global data sets.	Ryan & Abernathy section on xarray	
10/18	Models, parameters linear least squares, analysis of fit, uncertainties.		**Problem Set 3 (due 11/5)**
10/25	Debugging with Spyder. Tabular data, Pandas and Geopandas.	Hill Chapter 9; Ryan & Abernathy, section on pandas	
11/1	Probability and statistics with Python. More on classes, functions, and plotting.	Hill Chapter 7	
11/8	Plotting with PyGMT	PyGMT website	**Problem Set 4 (due 12/10)**
11/15	SciPy	Hill Chapter 8	
11/22	No class		
11/29	Time series analysis		
12/6	Jupyter notebooks	Hill §5.2	