COLUMBIA UNIVERSITY
School of Professional Studies

Master of Science in Sustainability Science

SUSC 5060 Statistics, Data Analysis, and Coding for Sustainability Science
3 credits
Wednesdays 4:10-6PM, 407 Mathematics Building

Instructor (Fall 2022): James L. Davis, Lamont Research Professor in the Lamont-Doherty Earth Observatory of Columbia University, jdavis@ldeo.columbia.edu, (845) 365-8425;

Instructor (Spring 2023): Michael Previdi, Lamont Associate Research Professor in the Lamont-Doherty Earth Observatory of Columbia University, mprevidi@ldeo.columbia.edu, (845) 365-8631

Office Hours: By appointment

Response Policy: Instructor is available for short discussions after class or on e-mail, with response within one business day. Longer discussions should be scheduled during office hours (Zoom or phone by appointment).

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 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 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

Anaconda Navigator documentation: https://docs.anaconda.com/anaconda/navigator/

Ryan Abernathey, *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/

Resources

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: http://library.columbia.edu/.

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

Programming/Data Analysis
Students will download the free-access application Anaconda-Navigator, which they will need to install on their computers prior to the second class. In addition, students should be familiar with standard office software to support their completion of course assignments.

Course Requirements (Assignments)

Class Participation (20%) (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. Students will make short group presentations reporting on progress on the Coding Application.

In-Class Coding Exercises (20%) (L1, L2, L3, L4, L5)

Some lectures will include 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 (30%) (L1, L2, L3, L4)

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

Coding Application in Sustainability Science (30%) (L1, L2, L3, L4, L5)

Students will be assigned as homework a computer coding problem of greater complexity than the in-class exercises or the Problem Sets. The Coding Application will involve reading and manipulating multiple data sets, application of the mathematical, statistical, and data analysis skills learned in class, and visualization of the data sets and derived computer models. For example, students may be provided with global temperature data sets in a standard climate-community format and be asked to develop code to provide an analysis of trends, and to produce computer-generated graphical images. Five weeks will be allotted for completion of the assignment.

Evaluation/Grading

Participation (20%)

Participation includes class attendance, performing assigned reading, and class discussion. The students are expected to show critical thinking, respectful interactions with classmates and the instructor, and a positive attitude towards learning and freely discussing the topics proposed.

In-Class Coding Exercises (20%)

Some lectures will include 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 (30%)

Two problem sets, each having 15% weight, will be each graded on a scale of 0–100. The students will be asked to write code to solve a problem that will exercise the mathematical concepts covered in the course. The students will be graded on the degree to which they appropriately translate the mathematical problem to computer code (50%), the use of proper syntax in their code (30%), and on arriving at the correct answer (20%).

Coding Application in Sustainability Science (30%)

One take-home assignment will be graded on a scale of 0–100. The assignment will be graded based on degree of correct completion of the assigned task, and on the degree to which the developed code adheres to best practices of computer programming as developed in the class.

Final Grading Scale

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
In-class coding exercises	20
Problem set #1	15
Problem set #2	15
Coding Application	30
Class participation	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 of 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 1/3 of a grade for every day it is late (e.g., from a B+ to a B for 1 day 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 must be worked on outside of class. Collaboration on these assignments is encouraged, but students must turn in their own work.

School Policies

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 http://sps.columbia.edu/student-life-and-alumni-relations/academic-integrity-and-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.

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: http://health.columbia.edu/services/ods/support.

Master of Science in Sustainability Science

Spring 2023 Course Schedule/Course Calendar

Class Date	Topics and Activities *Math/Stats/Data analysis*	*Coding*	Readings (due before class on this day)	Assignment handed out	Assignments due midnight Friday following Class Date
1/18	Course overview				
1/25	Mathematics: Calculus, vectors, trigonometry	Anaconda-Navigator environment, online documentation	Hill, Chapter 1		
2/1	Mathematics: Spherical geometry, linear algebra	Numbers, variables, precision, relational expressions and assignments, Python Objects	Hill, §2.1–2.4		
2/8	Overview of probability and statistics	Flow control, input/output, functions	Hill, §2.5–2.7	**Problem Set #1**	
2/15	Data analysis: Models, parameters	Basic graphs and plots	Hill, Chapter 3		
2/22	Data analysis: Linear least squares (setting up the problem)	NumPy: Arrays and matrices, linear algebra	Hill, §6.1.1–6.1.9, §6.5	**Problem Set #2**	**Problem Set #1**
3/1	Data analysis: Linear least squares (solution & uncertainty, analysis of fit)	Dictionaries and sets, Python data structures	Hill, §4.2, §6.1.10–6.1.12		
3/8	Data analysis: Linear least squares (example)	Debugging with Spyder, Python idioms, interaction with the OS, modules and packages	Hill, §4.1, §4.3–4.5		
3/15	No class: spring recess				
3/22	xarray and NetCDF files for climate and other data sets		Abernathey section on xarray	**Coding Application**	**Problem Set #2**
3/29	Discussion of Coding Application		Coding Application assignment		
4/5	Basic and geographic plotting with PyGMT		Browse pygmt.org		
4/12	Jupyter notebooks		Hill, §5		
4/19	Tabular data, Pandas		Hill, §9, Abernathey section on pandas		
4/26	Python classes		Hill, §4.6		**Coding Application**