ERTH 4004 – Data Analysis and Statistics – on the use of statistical techniques in the Earth Sciences (and beyond)

Quantitative skills are becoming more and more important in Earth Sciences. With the rapid development of remote sensing from satellites, vast online public databases, the amount of data an Earth scientist must process and interpret is overwhelming. Being able to analyze data on a computer becomes a necessity and often a job requirement. In order to handle large data sets, the ability to program, even in its simplest form, is imperative.

In this course, students will be introduced to the fundamentals and hone their coding skills for plotting, analyzing geologic data and modelling geologic systems. The objective is not to necessarily train professional programmers, but instead to develop geochemists, geophysicists, paleontologists, etc., who are also able to skillfully use technology in their chosen field. Thus, the ultimate goal is to shift from pure computer science to an informatics focus. In the assignments, students are encouraged to use data from their own research, or industry-grade datasets will be provided for analysis.

To get the most out of their data, students must also be comfortable with statistics. This course is an introductory class on how to quantitatively analyze data in the geosciences, and presents the fundamentals of exploratory data analysis, elementary probability theory and statistics, interpolation, curve fitting, regression, missing data and outliers, model uncertainty, timeseries (spectral) analysis, principal component analysis, least squares, data visualization, etc.

By the end of the course, students will realize that writing programs is a creative and rewarding activity. A primary motivation of the course, and in their future careers, is for students to gain the confidence to write programs to solve difficult data analysis problems, to perform statistical analyses, and to create clear visualizations of their data and results.

Instructor: Mareike Adams
Herzberg Laboratories 2249
E-mail: mareike.adams@carleton.ca

Office hours: By appointment, if you have any questions please reach out to me and we can set up a time to zoom.

Lectures: 11:35 am – 2:25 pm on Wednesday, via zoom. The class will start with a lecture period (1 – 1.5 hours), followed by an in-class lab where we will practice the techniques learned in the first half of the class.

Zoom lecture information:
Topic: Mareike Adams' Personal Meeting Room
Link to join Zoom Meeting:
https://zoom.us/j/5086475674?pwd=RUN1YXdIUjR6RlJQdUFQYWFCYjNOUT09

Meeting ID: 508 647 5674
Passcode: catsRkool

Course Texts (Optional):

Matlab Recipes for Earth Sciences
Martin H Trauth, 3rd edition, 2010

PDF of book available on the course website.

Course Topics (Tentative order):

- Introduction to Matlab
- Univariate statistics (Gaussian distributions, hypothesis testing, etc.)
- Bivariate statistics
- Linear, weighted, polynomial regression
- Data and parameter uncertainty, interpolation, missing data, error bars
- Model uncertainty – evaluating residuals, Akaike information criterion, etc.
- Multivariate analysis – plotting and visual representations
- Principal component analysis
- Multivariate linear regression, partial least squares

Learning Outcomes

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

- Acquire, load, manipulate and visualize 3D data
- Apply a firm understanding of statistical and numerical methods to numerical data
- Utilize relevant computer software packages to analyze the acquired data
- Confidently write their own codes to load, analyze, and create clear visualizations of data
- Create descriptive models that "predict" data, then evaluate model and parameter uncertainty
- Identify and test what type of probability density function describes the data
- Use the results to answer scientific questions or test hypotheses, make interpretations and explain the meaning/implications of results

Grading scheme:

Laboratory attendance and assignments	40%
3 Major Projects	40%
Take-home final exam	10%
Participation	10%

The labs contain a few theoretical problems, to ensure a firm understanding of the fundamentals, but are largely computer-based analytical problems to practice data analysis and statistical techniques. There is no coding experience expected, and all techniques, tools, and programming skills will be taught from the ground up.

All Carleton students can download Matlab to their own computers for free by following the instructions at this website:

https://carleton.ca/its/all-services/computers/site-licensed-software/matlab-students/

The labs will take place on Zoom right after the lecture portion of the course, and will either be handed in at the end of class, or at the end of the week if more time is required. This course is designed to provide interactive, collaborative hands-on learning – you can't learn programming without actually doing it – and so a large portion of class time will be designated to working on the techniques together. Furthermore, I encourage you to discuss the problems with your classmates, but it is absolutely imperative that any work you submit is your own. This means you must very clearly attribute any quotations or copied figures (citing name + year + publication of any sources). You should always mention any classmates with whom you have collaborated (a brief marginal note will suffice), and it is not EVER permitted to copy another student's work. If you are found to be in violation of this policy, there are very serious consequences. The instructor *is required* to report all incidents (or suspected incidents) of plagiarism to the Dean.

There will be three main projects throughout the term that focus on a more in-depth analysis of a particular dataset. You are welcome to use your own data from a research project, otherwise large datasets will be provided that students can pick from. Each project will involve applying different statistical techniques to the data, analyzing, plotting and discussing the results.

For all laboratory exercises and projects you must attach your well-documented code, and everything will be handed in on CuLearn.

For all labs and exams *always* show your full working for mathematical problems. As well as making it much easier to judge where/if you made any errors, I will not award full marks if the logic and work-flow of the answer is not clear. Make sure to properly highlight your final answer to each problem. Answers should be mathematically *correct*, i.e. if you write an "equals sign", both things on either side of it must be equal. This sounds totally obvious, but it is often not done, leading to avoidable errors and marks deducted. **Get in the practice of being meticulous with your mathematics!**

Please note that 10% of your grade is devoted to participation. This involves participation in class discussions, laboratory sections, and any literature reviews that we conduct. Participation has nothing to do with being the most correct, or the most profound. It's really about encouraging students to ask questions, to make comments, and to think out loud about what they are reading and learning.

Course Requirements:

- Attendance in class is mandatory. It is required that you email the instructor to advise of absences due to illness or emergencies.
- Labs must be handed in on time. Late labs will be accepted in the instance of illness, with a medical note, or in the instance of emergencies, by consultation with your instructor.
- It is the student's responsibility to come to classes and labs prepared. Reading assignments are mandatory.
- It is your responsibility to refer regularly to the course website for lecture topics, reading assignments, laboratory topics and pre-lab review or homework.

- Regularly log onto the CuLearn course website to check for announcements, course information, laboratory assignments and lecture material.
- Lab exercises will be posted on CuLearn, and you will hand in your course work on CuLearn as well.

Tentative Schedule (note: it is subject to change):

Week	Class Date	Topic	Lab
1	Jan. 13	Welcome	Introduction to Matlab
2	Jan. 20	Univariate Statistics I	Standard deviation, variance, Gaussian distributions, PDF's, CLT, probability distributions, hypothesis testing
3	Jan. 27	Univariate Statistics II	
4	Feb. 3	Bivariate Statistics I	Correlation coefficient, covariance, linear regression, Monte Carlo methods, data and parameter uncertainty, interpolation, missing data, error bars, weighted regression
5	Feb. 10	Bivariate Statistics II	
6	Feb. 17	**WINTER BREAK**	
7	Feb. 24	Bivariate Statistics III	Polynomial regressions, model uncertainty – evaluating residuals, Akaike information criterion, autoregressive models & random walks, correlated and autocorrelated time series
8	Mar. 3	Bivariate Statistics IV	
9	Mar. 10	Multivariate Analysis I	Plotting multivariate data, correlation matrix, Principal Component Analysis, multivariate linear regression, partial least squares
10	Mar. 17	Multivariate Analysis II	
11	Mar. 24	Multivariate Analysis III	
12	Mar. 31	Multivariate Analysis IV	
13	Apr. 7	Review	No Lab
14	Apr. 14	Last day of classes – Friday Schedule – No class	
15	Apr. 16 - 27	**Final exam period**	

ACADEMIC INTEGRITY

It is your responsibility to review Carleton's policy on Academic Integrity - Section 14 of the Calendar.
http://calendar.carleton.ca/undergrad/regulations/academicregulationsoftheuniversity/acadregsuniv14/

Plagiarism

The instructor is required to report all incidents (or suspected incidents) of plagiarism to the Dean. **All work handed in must be your own.** Plagiarism and cheating are viewed as being particularly serious and the sanctions imposed are accordingly severe. Students are expected to familiarize themselves with and follow the Carleton University Student Academic Integrity Policy. The Policy is strictly enforced and is binding on all students. Plagiarism and cheating – presenting another's ideas, arguments, words or images as your own, using unauthorized material, misrepresentation, fabricating or misrepresenting research data, unauthorized co-operation or collaboration or completing work for another student – weaken the quality of the graduate degree. Academic dishonesty in any form will not be tolerated. Students who infringe the Policy may be subject to one of several penalties including: expulsion; suspension from all studies at Carleton; suspension from full-time studies; a refusal of permission to continue or to register in a specific degree program; academic probation; or a grade of Failure in the course.

REQUESTS FOR ACADEMIC ACCOMMODATION

Please review the Carleton's Student Guide to Academic Accommodations at
http://carleton.ca/equity/wp-content/uploads/Student-Guide-to-Academic-Accommodation.pdf, and the websites therein.

For Students with Disabilities:

"The Paul Menton Centre for Students with Disabilities (PMC) provides services to students with Learning Disabilities (LD), psychiatric/mental health disabilities, Attention Deficit Hyperactivity Disorder (ADHD), Autism Spectrum Disorders (ASD), chronic medical conditions, and impairments in mobility, hearing, and vision. If you have a disability requiring academic accommodations in this course, please contact PMC at 613-520-6608 or pmc@carleton.ca for a formal evaluation. If you are already registered with the PMC, contact your PMC coordinator to send me your ***Letter of Accommodation*** *at the beginning of the term, and no later than two weeks before the first in-class scheduled test or exam requiring accommodation*. After requesting accommodation from PMC, meet with me to ensure accommodation arrangements are made. Please consult the PMC website (www.carleton.ca/pmc) for the deadline to request accommodations for the formally-scheduled exam . "

For Religious Observance:

1. *As soon as you receive your course syllabus*, identify any potential conflicts between your religious obligations and course requirements. **2.** Make a formal written request to your instructor indicating the nature of the religious obligation and suggest possible alternative dates and/or means of satisfying the academic requirements. *NOTE: Such request should be made during the first two weeks of the term*, or as soon as possible after a need for accommodation is known to exist, but in no case later than the second last week of classes for that term. For

detailed information on Religious Obligations please visit our website at: carleton.ca/equity/accommodation/academic.

For Pregnancy:
A. For final exams. Identify and discuss your needs for final examinations with your professors. When an agreement is reached fill out and submit the online **Pregnancy Accommodation Final Exam Request Form** at: carleton.ca/equity/ accommodation. Equity Services will forward the request to Exam Services to coordinate the accommodation. **B.** For in-class accommodations ONLY. If you anticipate you will only require in-class accommodations, discuss them directly with your course instructor. This request should be made in the first two weeks of the academic term. For detailed information on pregnancy and parental leave policies please visit the website at: carleton.ca/equity/accommodation/academic/

Equity and PMC Contact information:

- Department of Equity and Inclusive Communities
  613-520-5622
  3800 Carleton Technology & Training Centre
  equity@carleton.ca
  Website: carleton.ca/equity
- Paul Menton Centre for Students with Disabilities
  613-520-6608
  pmc@carleton.ca
  500 University Centre
  Website: carleton.ca/pmc