ATOC 557 - Research Methods: Atmospheric and Oceanic Sciences

Instructor: Robert Fajber

Email: robert.fajber@mcgill.ca

Office: BH824

Overview

- This class is mostly about statistical methods and how to use them to interpret real world data. In particular, we're going to study how to work with noisy data that is ordered in time and space. The goal of this class is to give you as much hands-on experience as possible, which is the only real way to learn how to analyze data.
- We're going to use a flipped classroom approach (described below). You will have course notes to read at home, and in class assignments to do during class. The assessment for the course will be based on these in class assignments as well as a final project.
- The class is "math-y" but not mathematical. By that I mean that you will see a lot of math, but not the details of a bunch of proofs. Instead, the focus will be on developing an intuitive understanding of what the different methods are doing, and how to practically implement them (i.e. code them up). Often the course notes will cover a particular case (e.g. a scalar instead of a vector, or a real variable instead of a complex variable), and then state what the more general results are.
- This course is designed to be more of a "survey" course. Instead of diving deep into one area we are going to cover multiple areas at a level that will allow you to build intuition about how the different methods work so that you can recognize them in the literature and at talks, and implement them yourself. I have included references which provide more details for the methods that you are most interested in, and I hope that you will use the course project to dive more deeply into an area of particular interest to you.

Flipped Classroom

The goal of a flipped classroom is to do the part of a course where you don't really need any help (reading some notes) at home, and do the part where you probably will need help (the assignments) during class time. My hope is that this format will allow you to get help and feedback on the hands on part of the course, rather than just listening to me talk. The in-class labs that you do will make up the course assignments that you need to submit.

The flow

- before class: read the notes for the upcoming lab activity, email in a reading response before 10AM
- during class: discussion and Q/A on the notes, followed by the lab activity. Discuss project or ask any other questions if there is leftover time.
- after class: finish any parts of the lab activity not done in class.

Grading

Reading Responses	5%
Assignments	45%
Final Project Presentation	10%
Final Project Report	40%

Reading Responses

- Before 10:00 AM the day of that lecture you will email me a "Reading Response" which can (and should) be as short as 1 sentence containing either a question or comment about the reading for that lecture. This is a great chance to ask a question without actually having to ask it in class.
- Please make the subject of the emails : ATOC557-[LASTNAME]-[READINGRESPONSE#]. The reading response number will be in the notes. If you do not format your subject like this I may lose the email.
- The marks for the reading responses are only participation marks.

Assignments

- For the first 20(ish) classes, there will be an in class lab Each of these labs is going to convert into 1 assignment question.
- After each section of the course, you will have to submit an assignment, which will be a compilation of the in class labs in that section.
- Each question will be marked out of 5:
  - 2 marks for the code completed and working
  - 1 mark for the code being commented
  - 2 marks for the analysis of the results

Missed Assignments Policy

- Things happen, it's ok. Without having to ask I will allow you to submit 18/20 in class labs/assignment questions, and 18/20 reading responses. I would encourage you to do them all, but I understand that things happen. If you need further extensions or exemptions please reach out to me. If you do not contact me and hand in less than 18/20 of the assignments you will get a 0 on them.

Class Project

- You will write a short paper, with 8-10 pages of double-spaced text, and 4-6 figures/tables. This will be formatted like a research paper, but will have less than usual emphasis on the introduction, and more than usual emphasis on the methods and discussion. Additionally, you will also have to submit your code with documentation. A more formal set of requirements and rubric will be distributed in the course.
- The project will focus on applying one of the methods in the course (or closely related) to a dataset or toy model of your choice. -If you have started any other kind of research project or thesis you can use a dataset from your research, *however*, the method must be new to you.

- About half-way through the course you will have to submit a half page proposal detailing the data and/or model, and method, that you will use in your project. This is mainly to check that you have chosen something feasible and that you have checked to make sure that you have access to data.

Outline

Section 1 - Basics and Hypothesis Testing

1. Data Handling
2. Probability distributions and moments
3. Hypothesis testing
4. Linear regression and trend analysis

Section 2 - Stochastic Processes and Autocorrelation

1. Univariate stochastic processes and feedbacks
2. Multivariate stochastic processes and the stochastic oscillator
3. Autocovariance and autocorrelation, cross correlation.

Section 3 - Time Series Analysis

1. Fourier transforms and the Weiner-Kinchin Theorem
2. Convolution theorem and windowing functions, periodograms.
3. Filtering time series
4. Cross spectral analysis
5. Wavelett analysis

Section 4 - Spatial Analysis

1. Point correlation and field significance
2. Principal component analysis
3. PCA with a stochastic diffusion model, sampling.
4. Singular value decomposition analysis and canonical correlation analysis

Section 5 - Data Assimilation

1. Optimal Interpolation
2. Kalman filtering
3. Ensemble Kalman filtering
4. State augmentation methods

Extra Topics

Note - we will discuss this in class, but there will be time at the end of class to go through some extra topics in a more conventional lecture format. We will decide as a group in the first couple of weeks which ones we want to cover. There might also be guest speakers who cover other topics as well. There will not be an in-class lab for these topics, and instead the time can be used to work on your project. If there's a topic you

would like to cover that isn't on this list let me know and I will look into either covering it myself or finding a guest speaker for it. Note I make no promises about any of this material

1. Stochastic double well and early warning signs of critical change
2. Extreme Value Theory
3. Variational Data Assimilation (3d/4d Var)
4. Spherical Harmonics and Spectra on a sphere
5. Multiplicative noise processes
6. Analysis of Variance (ANOVA) / Predictable Component Analysis
7. Stochastic Resonance
8. Singular Spectrum Analysis
9. Rotated EOFs
10. Diffusive Null Hypothesis for EOFs
11. Lyapunov exponents and forecast initialization
12. Non-gaussian statistics and the Weibull distribution
13. Multi-taper spectral analysis

Logistics

Computing

Most of the classes will need a computer. You are welcome to bring your laptop to class and use that, I have also reserved the computer lab for the same time, so you can use a desktop computer as well.

Jupyter Hub

In order to do the labs in the course there is a jupyterhub available to use: https://jhub.science.mcgill.ca/ You should use the ATOC 557 environment, which has all the packages you will need for the course as well as some that might be useful for your course project as well. Your home environment on jupyterhub is persistent, so you can access your files from any computer. You're welcome to do the labs on your laptop as well, but it might be tricky to move the data from jupyter hub to your laptop, and it won't be backed up in the same way.

Schedule

I will release a draft schedule in the first week. This is subject to change with either guest lectures and/or the pace of in class assignments.

Course Materials

The course notes are written as ipython notebooks and will be available on Jupyterhub. I will also upload them to mycourse. All the other material (updated schedule, project rubric, etc.) will also be uploaded to mycourse.

Office Hours

We will discuss this and pick a time in class that works for as many people as possible. I am also available by appointment through email.

Policies

Academic Honesty

My expectations are that you will discuss the assignments with each other during class, that is what the class is for. However I expect that you will each code your own solutions and write your own analysis without copying from each other. You are also free to discuss the final project with me and other people in class, but I expect that you will write your own code and paper by yourself. I also expect that you do the work yourself and do not use ChatGPT or any other AI tools.

McGill's full policy on academic honesty can be found: https://www.mcgill.ca/students/srr/academicrights/integrity/cheating

Classroom code of conduct

McGill's full policy on A big part of this class is working and discussing the lab activities with myself and your peers. I will make every effort to treat you with respect, and I expect that you will do the same with your peers. Harassment or bullying will not be tolerated.

All people have the right to be addressed and referred to in accordance with their personal identity. In this class, we will have the chance to indicate the name that we prefer to be called and, if we choose, to identify pronouns with which we would like to be addressed. I will do my best to address and refer to all students accordingly and support classmates in doing so as well.

McGill's full policy on personal rights and responsibilities can be found: https://www.mcgill.ca/students/srr/personalrights

Language Policy

This course will be taught in English, however all students have the right to submit work in English or French.

McGill's full policy can be found: https://www.mcgill.ca/students/srr/academicrights/course/french

References

There is no required course text, but much of the course roughly follows the book:

- DelSole, Timothy, and Michael Tippett. Statistical methods for climate scientists. Cambridge University Press, 2022.

with some supplements from the books:

- Jacobs, Kurt. Stochastic processes for physicists: understanding noisy systems. Cambridge University Press, 2010.
- Bendat, Julius S., and Allan G. Piersol. Random data: analysis and measurement procedures. John Wiley & Sons, 2011.

all of which can both be downloaded as a pdf from the McGill library website. However there were other textbook chapters and papers that I referenced while developing the course. If you're interested in a given topic I would strongly encourage you to look at the referenced papers. To be clear, you are not responsible for any of the material included here beyond the course notes, but if you are interested in digging deeper, or in look of ideas for a course project, these resources might be useful. I might update this with more papers as the course goes on, but here is the current list:

Section 1

- DelSole, Timothy, and Michael Tippett. Statistical methods for climate scientists. Cambridge University Press, 2022. Chapters 1-3, 8-9
- Jacobs, Kurt. Stochastic processes for physicists: understanding noisy systems. Cambridge University Press, 2010. Chapter 1.

Section 2

- DelSole, Timothy, and Michael Tippett. Statistical methods for climate scientists. Cambridge University Press, 2022. Chapters 5,6
- Jacobs, Kurt. Stochastic processes for physicists: understanding noisy systems. Cambridge University Press, 2010. Chapter 2-3.

This book is a classic reference, and the more general results in this section come from Chapter 4. This chapter goes into *much* greater detail on stochastic processes than we will cover in this course.

- Gardiner, Crispin W. Handbook of stochastic methods. Vol. 3. Berlin: springer, 1985. Chapter 4.

This paper won 1/4 of the 2020 Nobel Prize in physics using the tools that you will learn in this section:

- Hasselman, K. "Stochastic climate model. Part I: Theory." Tellus 28 (1976): 289-305.

Section 3

- DelSole, Timothy, and Michael Tippett. Statistical methods for climate scientists. Cambridge University Press, 2022. Chapters 6

This book is a classic on time series analysis, chapter 11 is a great overview of how to practically analyze time series data:

- Bendat, Julius S., and Allan G. Piersol. Random data: analysis and measurement procedures. John Wiley & Sons, 2011. Chapter 5,8,9,11

If you ever use waveletts this is a great reference:

- Torrence, Christopher, and Gilbert P. Compo. "A practical guide to wavelet analysis." Bulletin of the American Meteorological society 79.1 (1998): 61-78.

Section 4

- DelSole, Timothy, and Michael Tippett. Statistical methods for climate scientists. Cambridge University Press, 2022. Chapters 12-15

These papers do a great job explaining field significance, most of the textbook chapters are cribbed from these:

- Wilks, D. S. ""The stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it." Bulletin of the American Meteorological Society 97.12 (2016): 2263-2273.
- Wilks, D. S. "On "field significance" and the false discovery rate." Journal of applied meteorology and climatology 45.9 (2006): 1181-1189.

This is a classic paper which discusses EOF sampling and uses a 2D version of the example from the class:

- North, Gerald R., et al. "Sampling errors in the estimation of empirical orthogonal functions." Monthly weather review 110.7 (1982): 699-706.

These 2 papers show some of the problems with EOF analysis:

- Monahan, Adam H., et al. "Empirical orthogonal functions: The medium is the message." Journal of Climate 22.24 (2009): 6501-6514.
- Chen, Ge, and Xuan Wang. "Impact of domain geometry on the results of empirical orthogonal function analysis." Annals of GIS 20.3 (2014): 205-216.

This paper outlines how SVD works:

- Cherry, Steve. "Singular value decomposition analysis and canonical correlation analysis." Journal of Climate 9.9 (1996): 2003-2009.

Section 5

- DelSole, Timothy, and Michael Tippett. Statistical methods for climate scientists. Cambridge University Press, 2022. Chapters 20-21

This is the paper that introduced linear inverse modelling, originally applied to ENSO:

- Penland, Cécile, and Prashant D. Sardeshmukh. "The optimal growth of tropical sea surface temperature anomalies." Journal of climate 8.8 (1995): 1999-2024.

The next 2 papers discuss the Ensemble Kalman Filter compared to other types of filters, which I found very clarifying:

- Katzfuss, Matthias, Jonathan R. Stroud, and Christopher K. Wikle. "Understanding the ensemble Kalman filter." The American Statistician 70.4 (2016): 350-357.
- Stroud, Jonathan R., et al. "An ensemble Kalman filter and smoother for satellite data assimilation." Journal of the american statistical association 105.491 (2010): 978-990.

The next 2 papers use variants of the Ensemble Kalman Filter (a little different from what we use in class) to do parameter estimation:

- Kwasniok, Frank. "Estimation of noise parameters in dynamical system identification with Kalman filters." Physical Review E 86.3 (2012): 036214.
- Yang, Xiaosong, and Timothy Delsole. "Using the ensemble Kalman filter to estimate multiplicative model parameters." Tellus A: Dynamic Meteorology and Oceanography 61.5 (2008): 601-609.