Introduction:
It is your first day of work and your supervisor comes to you and says, “It’s great to finally have working for us someone who knows something about environmental data. What I’d like from you is a map showing maximum hourly 10-year return temperatures for the 1950-2005 period for all weather stations reporting in North America.”

There are various points and questions associated with this task:
- There are almost half a million hourly observations for the period 1950-2005
- There are thousands of weather stations
- What data set(s) are there to handle this? – Where to get them?
- How to read them in?
- How is a return frequency analysis performed?
- How do we ramp it up to run for so many weather stations?
- How do we present the results?

And where or how do you start? This is what we will explore.

This course is an introduction to the access and analysis of the large data sets that are often encountered in atmospheric sciences, and many other allied disciplines, including geography. We will cover these broad areas:
- where to find data
- how to get it into a form that can be analyzed (programming)
- various standard analysis methods, including extreme event work and multivariate methods
We will be learning Python for this course. There are many computer analysis options out there, both programmatic and turnkey. I prefer programmatic solutions, because it is often the case that you need to modify and tailor an analysis solution for your own needs. Python combines turnkey analysis packages and options with the flexibility of a powerful programming environment. Python is also the programming language behind ArcGIS and it is widely used for internet programming solutions. It is free. It has many packages for various types of analyses and display. It has no end of support/help/tutorial support online. It is a good language to learn.

Other types of systems you may have heard of include SPSS, SAS, IDL, MatLab, R, C, Fortran, NCL, amongst many others. I’ve used most of them.

This course is also a forum to assist you with your own research needs. Assignments build specific competencies linked (mostly) to coursework, and as we get later in the course you will commence a larger analysis project linked hopefully to your thesis. We will discuss this individually.

We do have a mandatory text. The name of the book is presented below. The previous edition (2nd) is available in PDF format and we will use that. Readings from the text will be regularly assigned. The course will follow these readings because the textbook was designed for an atmospheric analysis type of course, and you should keep up with them. Of course in class we will emphasize certain topics.

Course Objectives:

1. **Learn about large environmental data storage formats.** This includes various regular and non-regular ascii formats for observational data and a range of self-describing gridded binary formats, including HDF, GRIB, netCDF. Understand how their dimensionality works.

2. **Learn to manipulate Python to tackle large data sets.** This is central to analysis work; familiarity with their operation is essential.

3. **Learn about statistical methods to reduce and analyze data sets.** Analysis methods frequently utilized in weather and climate research will be explored. This includes standard aggregate reduction statistics, time series work, eigen methods, and multi-variate work.

4. **Learn about extreme event analysis.** A frequently requested output is some idea of return intervals of extreme events. We will explore this topic.

5. **Learn about model output analysis.** Weather and climate models are heavily utilized for many disciplines. Assessment of their output is a critical first step before their results can be folded into research.

6. **Results presentation.** Essential to any analytical work is the presentation of your results, both in oral format as well as in written/journal format. We will work on issues of presentation via submitted work and class presentations.
Student learning objectives

By the end of this course students should be able to:

- Describe what a self-describing gridded binary format is and identify major formats currently in use.
- Know how to gain information about such a dataset when presented with one.
- Delve into a dataset to reduce and prepare it for further analysis.
- Perform a variety of major types of data analysis.
- Present project results in a clear and concise manner. This includes readying results for print publication.

General Course Subjects:

1. Data storage methods
2. Aggregate statistics
3. Time-series statistics
4. Eigenmode analysis methods
5. Multi-variate methods
6. Mixed models, Gaussian and other types
7. Parametric vs non-parametric concepts
8. Extreme value analysis
9. Model output analysis (time permitting)
10. Spectrum analysis (time permitting)

Textbook/readings:

The textbook for this course is:

/statistical methods in the atmospheric sciences 2nd ed., by daniel wilks (2005), academic press, isbn 978-0127519661. (i have placed the PDF on coursespaces)

Material will be drawn from (but you don’t have to buy the book):


Research papers will be placed on coursespaces as PDFs.

Assignments

There will be a number of general homework assignments, usually one per week (55% of the grade). These will always consist of some sort of computer work with an annotated output. There will also be one large assignment (20% of the grade) that can be student guided, that is, some analysis problem related to your research, or I have plenty of things I can suggest. This will be submitted as a written analysis and as a presentation to the class.

Late policy: Late assignments will be reduced by 10% per day.
Paper discussion
In certain weeks I will assign a paper for us to read. A student will be selected to present a summary of the paper. This will be followed by a “roundtable” discussion. Participation is mandatory. There will be enough papers that each student has an opportunity to present one paper. Papers will be placed on coursespaces. Marks will be given for in-class participation and clarity of presentations.

Web page
You will find the course on coursespaces. Verify you can see it; if not, tell me and I will put you on.

Evaluation: The course grade will be based on the following:

<table>
<thead>
<tr>
<th>Element</th>
<th>Weight</th>
<th>Grading considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Assignments</td>
<td>55%</td>
<td>Accurate numerical or graphical solution, correct steps followed and presented, or if written, thorough assessment, clearly expressed. Emphasis will be placed on clarity of expression because of the crucial role communication will play in your futures.</td>
</tr>
<tr>
<td>2 Large project</td>
<td>20%</td>
<td>Grading will center around style and clarity of presentation. Students will participate in assessment process.</td>
</tr>
<tr>
<td>3 Paper discussion (participation)</td>
<td>10%</td>
<td>Active engagement (and presence) in in-class discussions.</td>
</tr>
<tr>
<td>4 Paper discussion (presentation)</td>
<td>15%</td>
<td>Present synthesis of paper. Relevant findings highlighted, problems identified. Marks also given for style and clarity of presentation.</td>
</tr>
</tbody>
</table>

Attendance, plagiarism
Attendance is mandatory; participatory marks reflect this. Plagiarism is not acceptable and will be dealt with as per UVic regulations. Discuss all missed work with me, preferably before it is missed. I am less helpful when something has been let slide with latitude being sought only after the fact. Emergencies of course are handled appropriately.
Tentative course outline
This is our objective but timings and topics may change as we see how rapidly we progress.

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Lecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Th Jan 7</td>
<td>Intro and data structures</td>
</tr>
<tr>
<td>2</td>
<td>Th Jan 14</td>
<td>Aggregate statistics (moment descriptors), gaussian distribution and nature</td>
</tr>
<tr>
<td>3</td>
<td>Th Jan 21</td>
<td>Time series, correlation, autocorrelation, regression</td>
</tr>
<tr>
<td>4</td>
<td>Th Jan 28</td>
<td>Time series (continued)</td>
</tr>
<tr>
<td>5</td>
<td>Th Feb 4</td>
<td>Eigenmode analysis – EOF, PCA</td>
</tr>
<tr>
<td></td>
<td>Th Feb 11</td>
<td>Spring break – no class</td>
</tr>
<tr>
<td>6</td>
<td>Th Feb 18</td>
<td>Eigenmode analysis (continued)</td>
</tr>
<tr>
<td>7</td>
<td>Th Feb 25</td>
<td>Mixed models</td>
</tr>
<tr>
<td>8</td>
<td>Th Mar 3</td>
<td>Multi-variate techniques</td>
</tr>
<tr>
<td>9</td>
<td>Th Mar 10</td>
<td>Multi-variate techniques (continued)</td>
</tr>
<tr>
<td>10</td>
<td>Th Mar 17</td>
<td>No class; I’m away</td>
</tr>
<tr>
<td>11</td>
<td>Th Mar 24</td>
<td>Parametric vs non-parametric</td>
</tr>
<tr>
<td>12</td>
<td>Th Mar 31</td>
<td>Extreme value analysis</td>
</tr>
</tbody>
</table>
“The University of Victoria is committed to promoting, providing and protecting a positive and safe learning and working environment for all its members.”

**Geography Departmental web site:** [http://www.uvic.ca/socialsciences/geography/](http://www.uvic.ca/socialsciences/geography/)

**GEOGPLAN planning guide:** [http://www.geog.uvic.ca/moodle/](http://www.geog.uvic.ca/moodle/)  [Login as a guest]

**Graduate Advisor:** Dr. Dennis Jelinski (jelinski@office.geog.uvic.ca)

Students with diverse learning styles and needs are welcome in this course. In particular, if you have a disability/health consideration that may require accommodations, please feel free to approach me and/or the Resource Centre for Students with a Disability (RCSD) as soon as possible. The RCSD staff are available by appointment to assess specific needs, provide referrals and arrange appropriate accommodations [http://rcsd.uvic.ca/](http://rcsd.uvic.ca/). The sooner you let us know your needs the quicker we can assist you in achieving your learning goals in this course.

*The University of Victoria is committed to promoting, providing and protecting a positive and safe learning and working environment for all its members.*