The Hong Kong University of Science and Technology

Data Analysis in Ocean Science

OCES 3301

3 credits

Pre-requisites: COMP 1021 OR OR COMP 1023 OR COMP 1029P

Name: Julian Mak

Email: jclmak@ust.hk

Consultation Hours: in class as part of the computer workshops

Course Description

This course provides the basic tools and skills for data analysis within ocean sciences. The focus is on the quantitative aspect, and really on the tools, though the examples and context will be largely from the ocean science context. Familiarity with the relevant problems in ocean science to be tackled would be beneficial but not completely necessary; if anything, one can potentially learn a bit more about the topic concerned by actively looking through the data itself rather than passively taking on content. A particular focus of this course is the active/experimental/hands-on type learning, provided through guided computer workshops, and supplemented with lectures.

Intended Learning Outcomes (ILOs)

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

1. Describe and explain data relevant to ocean sciences.

2. Perform basic and standard analysis techniques for data using a programming language.

3. Discuss data formats and possible issues with quality within ocean sciences.

4. Perform data exploration, as an avenue for improving scientific understanding.

5. Formulate and execute scientifically informed presentation via presentation of data and scientific writing.

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided below, outlining the criteria used for evaluation.

Assessments:

All written assessments are to be sent in as a Jupyter notebook (possibly including the relevant files) and submitted on Canvas. All Jupyter notebooks need to be able to run from scratch in a Conda environment at least; we will maximally spend 10-15 mins on trying to fix bugs, but after that there will be penalisation on the coding and scientific content criteria (see below in grading rubric).

For all written assessments, students are allowed a 1 week grace period with no questions asked (but you need to let us know on or before the assignment due time). Any further requests for extensions will require some proof (e.g. medical note, proof of internet going down such as electrical outage so Canvas submission was not possible). Measure of lateness will be done via the Canvas timestamp, and will be at 1% per minute penalty (i.e. don't bother handing anything in after 100 mins, because you already got zero).

Assessment Task	Contribution to Overall Course grade (%)	Due date
Assignment 1	25%	Friday 7th Mar
Assignment 2	25%	Friday 4th Apr
Assignment 3	25%	Friday 25th Apr
Assignment 4	25%	Friday 16th May

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
All assignments	ILO1, ILO2, ILO3, ILO4, ILO5	All handed in reports (ILO5) require students to provide an outline and explanation of the context (ILO1), perform their own calculations (ILO2), evaluate their results (ILO3) and explore the topic and tools(ILO4) in a coherent manner fitting of a scientific report, and in line with expectations for capstone and FYPs. Additionally provides practice with report writing skills; appropriate feedback will be given.

Grading Rubrics

All assessments are marked out of 110%, with the breakdown as scientific content (50%), writing and presentation (incl. grammar, figure quality etc.; 35%), coding and use of Jupyter notebooks functionalities (15%), and extras (going above and beyond the course material and approach; 10%). Most of the marks are given in terms of the scientific content to emphasise the focus, and the extras allows opportunity for further exploration. Any mark above 100% still only counts as 100%.

For class attendance, 13 (and sometimes 12) sessions are scheduled, discounting the add/drop period, out of the remaining 11 classes, students should turn up to at least 8 (or 7 for certain years) of these for full credit (zero credit for those who turn up for 3 or less, going up as a linear function of number of classes attended). Class attendance will be taken around two hours into the three hour session.

Final Grade Descriptors:

See also provided model good and bad hand-ins provided on the course GitHub page.

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance (>90% in the course)	Shows mastery of knowledge and understanding of the main subject matter, can problem-solve and critically evaluate approach, strong ability in communicating scientific and

		technical content.
B	Good Performance (75 – 90% in the course)	Shows good knowledge and understanding of the main subject matter, competence in problem-solving and some evaluation of approach, and the ability to communicate scientific and technical content.
C	Satisfactory Performance (60 – 75% in the course)	Shows adequate knowledge and understanding of the main subject matter, some issues with problem-solving, some ability to communicate scientific and technical content.
F	Fail (<60% in the course)	Shows poor knowledge and understanding of the main subject matter, struggles with problem-solving, unable to communicate scientific and technical content.

Course AI Policy

Use of AI is allowed and encouraged and by all means use it to help you code, but it is somewhat irrelevant in that most of the content is marked according to the scientific content anyway.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include a marked up PDF report with marked up comments, and a breakdown of the marks. Students who have further questions about the feedback including marks should consult the instructor within five working days after the feedback is received.

Resubmission Policy

Given the arrangements already for no-questions asked extensions and the use of continuous assessment, no resubmission or alternative assignments will be provided under normal circumstances.

Required Texts and Materials

All Jupyter notebooks should be self-contained; further reading is given in the notebooks accordingly.

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to Academic Integrity | HKUST – Academic Registry for the University's definition of plagiarism and ways to avoid cheating and plagiarism.

[Optional] Additional Resources

All Jupyter notebooks should be self-contained; further resources are given in the notebooks accordingly.

[Optional] Spring 25 Schedule (Fridays, 1330 to 1630, Room 4402)

- 7th February: introduction to python, anaconda, JupyterLab, basics of plotting and data
- 14th February: data reading, further plotting, basic manipulations
- 21th February: regression, concept of mismatch, linear regression, polynomial fitting
  [ASSIGNMENT 1 SET]
- 28th February: multilinear regression, A/BIC, principal component analysis,
- 7th March (NO CLASS, cancelled)
  [ASSIGNMENT 1 DUE]
- 14th March: basics of probability, statistical tests
- 21st March: more statistical tests, other concepts in probability (e.g. Shannon index)
  [ASSIGNMENT 2 SET]
- 28th March: time series data, basic manipulations, down-sampling, filtering, power spectrum
- 4th April (NO CLASS, public holiday)
  [ASSIGNMENT 2 DUE]
- 11th April: missing data, interpolation and extrapolation
  [ASSIGNMENT 3 SET]
- 18th April (NO CLASS, public holiday)
- 25th April: spatial data, plotting, analogous manipulations
  [ASSIGNMENT 3 DUE]
- 2nd May: maps, interpolation/extrapolation, empirical orthogonal functions
  [ASSIGNMENT 4 SET; due 16th May]
- 9th May: float lecture, probably machine learning or satellite data demonstration