Academic Year	AY2018/19	**Semester**	2
Course Coordinator	Asst. Professor David LALLEMANT		
Course Code	ES0002		
Course Title	Fundamentals of Data Science for Earth and Environmental Systems Science		
Pre-requisites	MH1800 Calculus for the Sciences I ES2001 Computational Earth Systems Science		
No of AUs	3		
Contact Hours	Total hours – 39 (Lecture – 18; Lab – 21)		
Proposal Date	14 May 2018		

Course Aims

The goal of this class is to develop a working knowledge of data science and its use in earth systems research and practice. The course is split evenly between key concepts / theory and practical experience writing code and analyzing outputs. Some introductory knowledge of statistics is assumed, and some familiarity with programming is helpful but not required.

Intended Learning Outcomes (ILO)

By the end of this course, you (as a student) will be able to:

1. Develop rationally designed statistical learning and machine learning models to extract meaningful insight from data.
2. Write logical expressions in R.
3. Generate meaningful visualizations of data and data analysis results.
4. Design appropriate experiments for data collection and apply statistical tests to answer research questions.
5. Illustrate theoretical knowledge of key concepts in statistical learning and data science.
6. Develop communication, creative and critical thinking skills.
7. Discuss the role of big data, statistical learning, machine learning and AI in earth and environmental systems sciences.
8. Discuss the social-ethical implications of data science technologies

Course Content

The course will focus on fundamental processes for scientific modelling, inference and data-science and their application in earth and environmental system science. The class will cover (1) basic concepts and tools in data science, (2) brief review of statics, (3) goals and principles of scientific modeling, (4) model development, (5) model calibration and selection, (6) sensitivity analysis, (7) model evaluation, (8) model predictions, (9) results visualization and communication.

Assessment (includes both continuous and summative assessment)

Component	Course LO Tested	Related Programme LO or Graduate Attributes	Weighting	Team/ Individual	Assessment Rubrics
1. Contributions	5,6,7,8	2,3,8,9,10	10%	Individual	Appendix 1

during class time & online Q&A forum					
2. Assignments 1-2 – Coding	1,2,4,5	1,2,3,4,5,8,9	20%	Individual	Appendix 2
3. Assignments 3-6 – (Coding + reports)	1,2,3,4,5, 6,7,8	1,2,3,4,5,8,9	70%	Individual	Appendix 3
Total			100%		

Formative feedback

You will receive oral feedback for Component 1, written feedback for Component 2 & 3.

Learning and Teaching approach

Approach	How does this approach support students in achieving the learning outcomes?
Lecture	Lectures will pass on the theoretical knowledge required to understand the different components of data analysis and statistical learning process.
Lab	Labs sessions will: • Demonstrate practical applications of statistical learning and data science in the field of earth and environmental system science. • Enable you to code using R and address any coding issues.

Reading and References

This is course aims to encourage you to think critically, and solve practical problems with a series of tools. The following books will be used as the main references/textbooks:

1. Hastie, Trevor J, Robert J Tibshirani, and J Jerome H Friedman. The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer, 2017; e-book ISBN 978-0-387-84858-7 [https://web.stanford.edu/~hastie/ElemStatLearn/]
2. Kabacoff, Robert. R in Action, Second Edition: Data analysis and graphics with R. Manning, 2015; ISBN-13: 9781617291388
3. Pebesma E., Nüst D., Bivand R. (2012) The R software environment in reproducible geoscientific research. EOS, Transactions American Geophysical Union 93:163-163. [Forum]
4. Chamberlin, T.C., 1890. The method of multiple working hypotheses: Science, v. 15. doi, 10: 92–96. [Essay]
5. Jackson, L.J., Trebitz, A.S., Cottingham, K.L., 2000. An introduction to the practice of ecological modeling. Bioscience. 50, 694-706. [Article]
6. Strayer, D.L., Ewing, H.A., Bigelow, S., 2003. What kind of spatial and temporal details are required in models of heterogeneous systems? Oikos. 102, 654-662. [Article]
7. Schewe J., Levermann A. (2012) A statistically predictive model for future monsoon failure in India. Environmental Research Letters 7:044023. [Article]
8. Saltelli, A., K. Chan, and E.M. Scott, (eds.) 2000. Sensitivity Analysis. John Wiley and Sons, New York. First chapter; ISBN-13: 978-0-470-74382-9

Course Policies and Student Responsibilities

(1) General

You are expected to complete all assigned pre-class readings and activities on time, attend all lectures, tutorial and class discussions, and submit all scheduled assignments by due dates. You are expected to take responsibility to follow up with course notes, assignments and course related announcements for sessions they have missed.

(2) Absenteeism

Absence from scheduled lectures and discussion without a valid reason will affect your overall course grade. Valid reasons include falling sick supported by a medical certificate.
If you miss a lecture or tutorial exercise you must inform me via email (dlallemant@ntu.edu.sg) prior to the start of the class.

(3) Compulsory Assignments

You are required to submit compulsory assignments on due dates, unless a valid reason is provided. Valid reasons include falling sick supported by a medical certificate. If you will miss a deadline for a valid reason you must inform me via email (dlallemant@ntu.edu.sg) prior to the deadline, and as soon as is possible.

Academic Integrity

Good academic work depends on honesty and ethical behavior. The quality of your work as a student relies on adhering to the principles of academic integrity and to the NTU Honor Code, a set of values shared by the whole university community. Truth, Trust and Justice are at the core of NTU's shared values.

As a student, it is important that you recognize your responsibilities in understanding and applying the principles of academic integrity in all the work you do at NTU. Not knowing what is involved in maintaining academic integrity does not excuse academic dishonesty. You need to actively equip yourself with strategies to avoid all forms of academic dishonesty, including plagiarism, academic fraud, collusion and cheating. If you are uncertain of the definitions of any of these terms, you should go to the academic integrity website for more information. Consult your instructor(s) if you need any clarification about the requirements of academic integrity in the course.

Course Instructors

Instructor	Office Location	Phone	Email
David LALLEMANT	N2-01c-45	6592-3199	dlallemant@ntu.edu

Planned Weekly Schedule

Week	Topic	Instructor	Course LO	Readings/ Activities
1	Introduction to course, algorithms, programing, statistical learning	D.L	1,2,4,5,6, 7,8	
2	Review of statistics, introduction to R	D.L	1,2,4,5,6,7	Kabacoff, 2015
3	Goal and principles of scientific modeling; modeling examples.	D.L	1,2,3,4,5,6,7,8	Pebesma, 2012; Assignment 1
4	Designing and programing models (statistical and machine learning) – Part 1	D.L	1,2,4,5,6,7	Assignment 2
5	Designing and programing models (statistical and machine learning) – Part 2	D.L	1,2,4,5,6,7	Schewe and Leverman, 2012
6	Model Calibration and Selection	D.L	1,2,4,5,6,7,8	Assignment 3
7	Model Calibration and Selection	D.L	1,2,4,5,6,7,8	Saltelli, 2000
8	Sensitivity Analysis	D.L	1,2,4,5,6,7,8	Assignment 4
9	Model Evaluation	D.L	1,2,4,5,6,7,8	
10	Model Predictions	D.L	1,2,4,5,6,7,8	Assignment 5
11	Model Predictions and uncertainties	D.L	1,2,3,4,5,6,7,8	
12	Visualization	D.L	1,2,3,4,5,6,7,8	
13	Ethical data science and AI applications	D.L	1,2,3,4,5,6,7,8	Assignment 6

Appendix 1: Assessment Criteria for Contributions during class time & online Q&A forum

Standards	Criteria
A+ (Exceptional) A (Excellent)	Important contributions to class discussion and piazza online forum; uses piazza to ask coding questions and respond to questions of other students; capacity to articulate and present points of view very clearly; participates in a meaningful and constructive manner including enabling other students to contribute and not dominating; evidence of having read and assimilated class material beyond the assigned reading; strong signs of evidence-based formation of points of view on the topics.
A- (Very good) B+ (Good)	Meaningful contributions to class discussion and piazza online forum; uses piazza to ask coding questions and respond to questions of other students; capacity to articulate and present points of view clearly; participates in a meaningful and constructive manner; evidence of having read and assimilated the class material; some signs of evidence-based formation of points of view on the topics.
B (Average) B- (Satisfactory) C+ (Marginally satisfactory)	Some contributions to class discussion and piazza online forum; uses piazza to ask coding questions; some evidence of constructive engagement during discussion; some familiarity with the assigned reading; some evidence of having thought about controversial topics.
C (Bordering unsatisfactory) C- (Unsatisfactory)	Minimal contributions to class discussion or piazza online forum; limited capacity to articulate and present points of view; limited evidence of constructive engagement during discussion; little or no familiarity with the assigned reading.
D, F (Deeply unsatisfactory)	Very minimal or no contributions to class discussion and piazza online forum; no questions; no evidence of an individual viewpoint; failure to read the assigned reading; unexplained or unjustified absences from class activities.

Appendix 2: Assessment Criteria for Coding Assignments 1-2

Standards	Criteria
A+ (Exceptional) A (Excellent)	Provides clear, efficient, working and well-documented code; evidence of programing understanding and concern for code efficiency beyond getting correct solution. Demonstrated ability to develop multiple approaches to programming task, and understanding of their respective advantages.
A- (Very good) B+ (Good)	Provides clear, efficient, working and well-documented code; evidence of programing understanding.
B (Average) B- (Satisfactory) C+ (Marginally satisfactory)	Working but limited documentation of code.
C (Bordering unsatisfactory) C- (Unsatisfactory)	Write the code with lots of help from TA and instructor. Limited code documentation or demonstration of conceptual understand.
D (Deeply unsatisfactory) F (0-44)	Lack of demonstrated conceptual understanding. Non-function code.

Appendix 3: Assessment Criteria for Coding Assignments 3-6

Standards	Criteria
A+ (Exceptional) A (Excellent)	Takes an original approach to the questions; very well structured reports with good interpretations of results; evidence of excellent ability to apply knowledge taught in the course while thinking outside the box; provides clear, efficient, working and well-documented code
A- (Very good) B+ (Good)	Takes a conventional approach to the question; good interpretation of results; evidence of ability to apply knowledge taught in the course; provides clear, efficient, working and well-documented code;
B (Average) B- (Satisfactory) C+ (Marginally satisfactory)	Takes a conventional approach to the question; limited interpretation of results; evidence of some (but not significant) ability to apply knowledge taught in the course; working but limited documentation of code.
C (Bordering unsatisfactory) C- (Unsatisfactory)	Limited understanding of process; incorrect or miss-interpreted results; limited evidence of ability to apply knowledge taught in the course. Non-functional or limited code documentation.
D, F (Deeply unsatisfactory)	Inadequate in addressing the question; incorrect and/or miss-interpretation of results; lacks structure and focus, and is mostly or wholly off topic; inadequate capacity to apply knowledge taught in the course; non-functional code. OR failure to submit the report.