MODULE TITLE	**Fundamentals of Data Science**	**CREDIT VALUE**	**30**
MODULE CODE	**MTHM601**	**MODULE CONVENER**	**Dr Tim Hughes (Coordinator)**

DURATION: TERM	**1**	**2**	**3**
DURATION: WEEKS	**11**	**0**	**0**

Number of Students Taking Module (anticipated)	**50**

DESCRIPTION - summary of the module content

This module develops core skills in data science, modelling, and essential programming skills. The ability to extract information from data as a basis for evidence-based decision making and policy is becoming increasingly important across a wide variety of sectors in the world of big data, including climate, health, technology, and the environment. This module will equip you with the tools required to collate, import and manipulate data, together with methods for inference. You will be introduced to different types and sources of data and the tools for performing data analysis, from producing informative graphical summaries to generating sophisticated visualisations. These techniques are crucial both as the basis for communication and for informing complex modelling. This will be placed in a contemporary and cutting edge setting through the use of locally curated and global open source datasets, and will draw on the flexible and freely available programming environments of Python and R.

AIMS - intentions of the module

This module aims to equip you with the skills that are required to collect, collate, process, manipulate, analyse and interpret data effectively and efficiently. You will be introduced to techniques for importing data from a range of sources into the format that is appropriate for many data types and their further processing and analysis. You will learn how to merge information from multiple sources to develop greater insight, and you will learn how to pre-process data to enable the effective application of analysis techniques. This will include data cleansing, handling of missing, corrupted, uncertain and/or biased data, and the graphical representation of data. You will develop an appreciation of these concepts, and the ways in which their effects might be mitigated. This will enable you to communicate possible issues with the analysis of data when writing reports and making recommendations based on statistical analyses.

This module will also equip you with the skills that are needed to perform a range of data science and statistical analysis techniques, and to understand and interpret their outputs. This will include an introduction to the mathematical and statistical techniques underpinning data science, familiarisation with the open source scientific computing languages R and Python, and an overview of supervised and unsupervised machine learning methods.

You will be encouraged and supported to develop your data science skills alongside your specialism, exploring datasets relevant to ecology; evolution; environment; sustainability; and/or renewable energy. Activities will include data wrangling, data analysis, report writing and presentation. Assessments will be based on a series of practical examples using real-world data examples that aim to demonstrate the full range of skills required to make effective use of data.

INTENDED LEARNING OUTCOMES (ILOs) (see assessment section below for how ILOs will be assessed)

On successful completion of this module ***you should be able to***:

Module Specific Skills and Knowledge

1. Demonstrate the ability to import, manipulate and summarise data, including an understanding of the relative merits of different methods of formatting;
2. Demonstrate an understanding of how data source and way of collection effect subsequent data analyses;
3. Demonstrate effective use of Python and/or R/RStudio to facilitate data wrangling, unsupervised and supervised data analyses;

Discipline Specific Skills and Knowledge

4. Demonstrate effective and efficient data processing and programming skills;
5. Demonstrate competencies of data visualization;
6. Demonstrate an understanding of the methodology and practical use of a range of data analysis techniques, including unsupervised and supervised machine learning and statistical modelling methods;
7. Demonstrate an understanding of common pitfalls in data processing and analysis and how to avoid them;
8. Demonstrate appreciation and understanding of relevant datasets in application areas;

Personal and Key Transferable / Employment Skills and Knowledge

9. Data and statistical analysis skills;
10. Use of Python, R/RStudio and other software;
11. Effective use of learning resources;
12. Report writing and presentation.

SYLLABUS PLAN - summary of the structure and academic content of the module

The precise syllabus may vary slightly from year to year, and the below is provided asan indication of the typical content.

Data collection, pre-processing and communication:
- Cleansing;
- Visualisation;
- Handling missing, corrupted, uncertain and/or biased data;

Effective programming:
- Coding in R/R Studio and Python;
- Computer Hardware;
- Version control, collaborative and high performance computing;
- Reproducible programming;

Analysis:
- Fundamentals of probability, linear algebra and calculus;
- Fundamentals of statistical modelling;
- Sampling and sampled data;
- Inference, confidence intervals, and hypothesis testing;
- Regression analysis and model selection;
- Spatial-temporal and hierarchical models;
- Introduction to machine learning: supervised methods (e.g., classification and regression) and unsupervised methods (e.g., clustering and dimensionality reduction);

Application areas:
- Datasets for ecology and evolution: populations, infectious diseases, biodiversity, genetics;
- Datasets for renewable energy: solar, wind, marine (resource and generation data), electricity/heat consumption, smart grid;
- Datasets for environment and sustainability: sustainable development indices, health, weather and climate, land and marine pollution.

The assessment structure on this module is subject to review and may change before the start of the new academic year. Any changes will be clearly communicated to you before the start of term and if you wish to change module as a result of this you can do so in the module change window.

LEARNING AND TEACHING

LEARNING ACTIVITIES AND TEACHING METHODS (given in hours of study time)

Scheduled Learning & Teaching Activities	60.00	Guided Independent Study	240.00	Placement / Study Abroad	0.00

DETAILS OF LEARNING ACTIVITIES AND TEACHING METHODS

Category	Hours of study time	Description
Scheduled Learning and Teaching Activities	30	Lectures and tutorials
Scheduled Learning and Teaching Activities	30	Hands-on practical sessions
Guided Independent Study	120	Self-study and background reading
Guided Independent Study	120	Assessed data analyses, quizzes, report writing and preparation for presentations

ASSESSMENT

FORMATIVE ASSESSMENT - for feedback and development purposes; does not count towards module grade

Form of Assessment	Size of Assessment (e.g. duration/length)	ILOs Assessed	Feedback Method
Exercises	Several quizzes/exercise sheets	1-11	Oral, during tutorial sessions
Practicals	Several practical sheets for self-directed and guided learning	1-11	Oral, during tutorial sessions

SUMMATIVE ASSESSMENT (% of credit)

Coursework	100	Written Exams	0	Practical Exams	0

DETAILS OF SUMMATIVE ASSESSMENT

Form of Assessment	% of Credit	Size of Assessment (e.g. duration/length)	ILOs Assessed	Feedback Method
Exercises	50	Several quizzes/ exercise sheets (4 expected)	1-11	Written, oral or automated feedback
Report	50	Approx. 10-15 pages	1-12	Written

DETAILS OF RE-ASSESSMENT (where required by referral or deferral)

Original Form of Assessment	Form of Re-assessment	ILOs Re-assessed	Time Scale for Re-assessment
Exercises	Coursework (100%)	1-11	To be agreed by consequences of failure meeting
Report	Coursework (100%)	All	To be agreed by consequences of failure meeting

RE-ASSESSMENT NOTES

Deferral – if you miss an assessment for certificated reasons judged acceptable by the Mitigation Committee, you will normally be either deferred in the assessment or an extension may be granted. The mark given for a re-assessment taken as a result of deferral will not be capped and will be treated as it would be if it were your first attempt at the assessment.

Referral – if you have failed the module overall (i.e. a final overall module mark of less than 50%) you will be required to resubmit the original assessment as necessary. The mark given for a re-assessment taken as a result of referral will be capped at 50%.

RESOURCES

INDICATIVE LEARNING RESOURCES - The following list is offered as an indication of the type & level of information that you are expected to consult. Further guidance will be provided by the Module Convener

Web-based and electronic resources:
ELE – https://vle.exeter.ac.uk/
Other resources:
Recent articles and open-source codes provided by the tutors.
Reading list for this module:

Type	Author	Title	Edition	Publisher	Year	ISBN	Search
Set	James, G., Witten, D., Hastie, T., Tibshirani, R.	An Introduction to Statistical Learning: with Applications in R		Springer	2013	978-1461471370	[Library]
Set	Simon Rogers & Mark Girolami	A First Course in Machine Learning	2nd	CRC Press	2016	B01N7ZEBK8	[Library]
Set	Murphy, K.	Machine Learning: A Probabilistic Perspective	1st	MIT Press	2012	978-0-262-018029	[Library]
Set	Hastie T., Tibshirani R. & Friedman J.	The Elements of Statistical Learning: Data Mining, Inference, and Prediction	2nd	Springer	2009	978-0387848587	[Library]
Set	Bishop, C.	Pattern Recognition and Machine Learning	1	Springer	2006	978-0387310732	[Library]
Set	Aurelien Geron	Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow		O'Reilly	2019	978-1492032649	[Library]
Set	Sebastian Raschka, Vahid Mirjalili	Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-learn, and TensorFlow	2nd	Packt Publishing	2017	978-1787125933	[Library]

CREDIT VALUE	30	ECTS VALUE	15
PRE-REQUISITE MODULES	None		
CO-REQUISITE MODULES	None		
NQF LEVEL (FHEQ)	7	AVAILABLE AS DISTANCE LEARNING	No
ORIGIN DATE	Monday 14 December 2020	LAST REVISION DATE	Tuesday 17 October 2023
KEY WORDS SEARCH	Data processing; Data visualisation; Programming; Statistical modelling; Machine learning; Applied data analysis		