# Imperial College London

Masters in Computational Methods
in Ecology and Evolution

# Programme Guidebook

# 2016 – 2017

# Contents

Along with this Course guidebook, you will receive a copy of the Student Masters Student Guidebook (AKA Student Guidebook) for the Silwood Park Campus Masters Courses, containing (among other things) the following important information:

## Introduction to Silwood Park and the Department
Including Key contacts and information on the library, IT, safety and seminars.

## Academic regulations
CMEE course-specific regulations are provided in this guidebook. The Student Guidebook provides information about the general regulations that apply across courses. This includes academic integrity, plagiarism, employment during your studies and complaint and appeals procedures.

## Welfare and Advice
Imperial has a wide support network for students. The Student Guidebook provides details of the available support and key contacts and links.

## Student Feedback and Representation
We are very grateful for feedback on the course and will ask you for it at regular intervals! However, there are a range of options for providing feedback and getting support on your academic studies and the Silwood Masters Guidebook provides details.

## Thesis Guidelines
How to prepare you thesis, including word limits, formatting, etc.

## Project and Supervison Guidelines
How to choose a project, student reserach budgets, what to expect (and not to expect!) from supervisors.

## The FrEE Symposium
All about the student-run Frontiers in Ecology and Evolution (FrEE) summer symposium at Silwood.

## Key Dates
Dates for various Silwood-wide student activities and events

Electronic copies of both of these gudebooks are available on the course website as well as Blackboard. A copy can also be obtained by emailing the Course Administrator Mrs. Amanda Ellis (*amanda.ellis@imperial.ac.uk*).

# Course Overview

Welcome to the Masters programmes in Computational Methods in Ecology and Evolution (CMEE) at Silwood Park!

The frontiers of Biology are increasingly at the interface between mathematics, computing, and large empirical datasets ("big data"). Answering important problems about issues ranging from disease dynamics and epidemiology to the effects of climate climate and over-fishing on aquatic ecosystems, require computationally sophisticated approaches for management & analysis of big data, theoretical modelling, and fitting models to the data. Indeed, biology is *the* new frontier for applied computer science and mathematics. Donald Knuth, an eminent mathematician and computer scientist has said: "*It is hard for me to say confidently that, after fifty more years of explosive growth of computer science, there will still be a lot of fascinating unsolved problems at peoples' fingertips, that it won't be pretty much working on refinements of well-explored things. Maybe all of the simple stuff and the really great stuff has been discovered. It may not be true, but I can't predict an unending growth. I can't be as confident about computer science as I can about biology. Biology easily has 500 years of exciting problems to work on, it's at that level.*".

MSc and MRes CMEE aim to teach computational approaches towards the empirical and theoretical study of ecological and evolutionary systems. We focus on Computational *Ecology and Evolutionary Biology* in particular because some of the most complex and important dynamics in human-dominated as well as natural environments arise from ecological and evolutionary processes. Therefore, computational skills needed to analyse data and model ecological and evolutionary systems are very relevant to other disciplines. For example, it is now generally recognized that a major new frontier in understanding disease dynamics and epidemiology is the need to consider the *Ecology* of infectious diseases. Furthermore, ecological and evolutionary models and computational tools are also relevant for non-biological disciplines — for example, ecosystem models are increasingly being considered relevant for understanding microeconomic systems.

Both MSc and MRes CMEE run for one year commencing the first week in October through to the end of September. The first nine weeks of the first term is shared between both courses and includes a combination of lectures, workshops and practicals to cover a range of fundamental skills. The MRes option then focuses on a substantial research project for the second and third terms while the MSc option continues with more advanced taught modules followed by a shorter project. Switching between the courses is straightforward during the initial part of the first term, and possible at later stages. MRes students are welcome to choose further MSc lectures that are relevant to their research project.

Research projects are undertaken from the 10th week of the course (MRes) or from the 23rd week (MSc) (see timetables in section 2.2).

*Both MSc and MRes CMEE projects must entail significant amounts of mathematical theory or statistical analyses of large datasets, or a combination of these.*

Daily lectures and practicals, unless otherwise stated, commence at 10:00 and would normally finish by 1700, incorporating breaks. Additional independent work is also required. On Wednesday the afternoon is normally, but not always, reserved for private study, sports and leisure activities for students.

In addition to the formal taught and research components of the programme, there is a research seminar series that runs every Thursday at 1pm, presented by local and visiting academics. In addition, the Graduate-student organized Frontiers in Ecology and Evolution Symposium (FrEE!) in September

## Course Overview

provides students with an opportunity to present their work and see what their peers have been up to.

Teaching materials and other course materials will be provided using the online Blackboard virtual learning environment http://bb.imperial.ac.uk, and/or for certain modules, using a git version control repository to provide code and data files. Paper copies of lecture notes and handouts are not normally provided but you will receive printing credit for use during the course on your security card.

It is anticipated that reading and coursework will require additional study in your own time. During research projects, you are expected to work full time on the project, but with flexible hours, and in general, students who work extra hours do get more out of the course. Some projects may require out-of-hours work, for example maintaining greenhouse experiments.

The following sections provide an overview of the programme and assessment structure for the two courses. Aims, objective and learning outcomes for the course are in section 1. The full programme specifications for the MSc and MRes are available on Blackboard and from the course website (url below).

Course website
    https://www.imperial.ac.uk/life-sciences/postgraduate/masters-courses/-msc-in-computational-methods-in-ecology-and-evolution-cmee/

## Course Administration

Please see the Student Guidebook for more descriptions of the roles of the Postgraduate Administrator & Tutor, and Director of Postgraduate Studies.

| | |
|---|---|
| *MSc & MRes CMEE Course Director* | Dr. Samraat Pawar (ext. 42213, s.pawar@imperial.ac.uk) |
| *MSc & MRes CMEE Course Co-Director* | Dr. James Rosindell (ext. 42242, j.rosindell@imperial.ac.uk) |
| *Postgraduate Administrator* | Mrs. Amanda Ellis (ext. 42251, amanda.ellis@imperial.ac.uk) |
| *Postgraduate Tutor* | Dr. Julia Schroeder (julia.schroeder@imperial.ac.uk) |
| *Director of Postgraduate Studies* | Dr. Niki Gounaris (ext. 4 5209, k.gounaris@imperial.ac.uk) |
| *Course Tutor* | Mr. Samual D Thompson (samuel.thompson14@imperial.ac.uk) |
| *Course Representative* | Up to you (see Silwood Masters Guidebook) |

(Add 020 759 to extension numbers to call from external phones)

## Course Aims

Both MSc and MRes CMEE aim to:

- Provide training in key biological, statistical, computational, and mathematical topics and how they are integrated.

- Provide students with a background to computational biology with particular focus on theoretical and empirical approaches towards the study of ecological and evolutionary systems.
- Teach state-of-art, scientific computing techniques for the management and analyses of big data ad simulation as well as mathematical modelling.
- Provide a thorough understanding of a range of modern techniques in bioinformatics, genomics and ecoinformatics.
- Show students how these principles and skills can be applied to solve real world problems and make an informed choice of research topic.
- Prepare students for PhD studies and other appropriate career paths in industry or NGOs focusing on any combinations of biology, maths and computing skills.

The main rationale for having an MSc as well as MRes CMEE is to give students the choice to find a right balance between learning technical skills and learning research skills. With this flexibility, students can choose a course and a project that best suits their personal interests and future plans.

## Learning outcomes

Students will develop:

- Competence in computational methods required to address a range of topical 'real-world' questions in a range of ecological and evolutionary topics, including population biology, population genetics, genomics, and complex systems
- An ability to choose an appropriate quantitative method, possibly including statistical and mechanistic mathematical modelling, for answering a particular biological question
- An ability to develop, analyse, and numerically simulate theoretical models for ecological and evolutionary systems and fit empirical data to these models

## Transferable Skills

A central objective of the CMEE Masters' is to prepare students for PhD studies and/or a career in computational biology by teaching a suite of transferable skills, be it within academic institutions, government, or industry, by developing a professional approach towards developing and delivering high-quality science along with qualitative, critical thinking and problem solving skills. Students will learn a unique set of transferrable skills in computation relevant not just to biology but also other fields that involve large datasets and complex dynamics and patterns, such as economics, medicine, or sociology. Students will be able to:

- apply computational, statistical and modelling skills;
- communicate effectively through oral presentations, written reports and scientific publications;
- management skills: decision making, problem definition, project design and evaluation, risk management, teamwork and coordination;
- integrate and evaluate information from a variety of sources;
- transfer techniques and solutions from one discipline to another;
- use Information and Communications Technology;
- manage resources and time;
- learn independently with open-mindedness and critical enquiry;
- learn effectively for the purpose of continuing professional development;
- depending upon choice of taught modules and research project, learn lab and field techniques.

# Course Overview

The course objectives will be achieved by providing:

- A course of lectures, seminars and practicals within distinct modules, linked to cutting edge academic research and research groups in Silwood Park and collaborator institutions elsewhere. Both MSc and MRes students will attend the first 9 weeks of modules where they will learn fundamental concepts and techniques in computational biology, with particular focus on ecological and evolutionary theory, data, and systems.
- Hands-on experience of a wide repertoire of methods and techniques involved in the application of computational techniques to ecological and evolutionary systems.
- A research project ($\sim$9 months for MRes, $\sim$5 months for MSc) based upon theoretical/analytical work, possibly including field and/or laboratory study, on an advanced and original topic from systems biology, ecology, evolution, conservation or another biological field, supervised by academics from either biological, computational or mathematical fields, or combinations of these.
- Strategically timed workshops on publishing scientific manuscripts, selecting a research project, and selecting Job/PhD opportunities and applying to them.

# Course activities and assessment overview

# Assessment Overview

|  | MSc CMEE | | MRes CMEE | |
| --- | --- | --- | --- | --- |
| **Component** | % of Course | % of Component | % of Course | % of Component |
| **Coursework** | | | | |
| *Computing* | 13.75 | 55 | 13.75 | 55 |
| *CMEE Mini-project* | 6 | 24 | 6 | 24 |
| *HPC Long Practical* | 5 | 20 | 5 | 20 |
| *Seminar Diary* | 0.25 | 1 | 0.25 | 1 |
| **Coursework Total** | 25 | 100 | 25 | 100 |
| **Exams** | | | | |
| *Exam 1* | 10 | 40 | – | – |
| *Exam 2* | 15 | 60 | – | – |
| **Exam Total** | 25 | 100 | – | – |
| **Project** | | | | |
| *Final Report + Presentation* | 35 | 70 | 52.5 | 70 |
| *Viva* | 12.5 | 25 | 18.75 | 25 |
| *Supervisor mark* | 2.5 | 5 | 3.75 | 5 |
| **Project Total** | 50 | 100 | 75 | 100 |

## Course structure overview

| Activity | MSc CMEE | MRes CMEE |
|---|---|---|
| *Lectures + practicals, with assessment* | Required for first 20 weeks | Required for first 9 weeks, optional attendance in MSc modules within reason thereafter |
| *Exams* | Required | Not required |
| *Project report (Dissertation)* | Required | Required |
| *Seminars* | Required, seminar diary required for a minimum 16 weeks | Required, seminar diary required for a minimum 16 weeks |
| *Workshops* | All optional | All optional |

## MSc CMEE

MSc CMEE students will attend 20 weeks of taught modules over the Autumn and Spring Terms followed by a 5-month research project leading to a dissertation (Thesis report). There are three main components to the course assessment:

i) *The coursework*: 25% of final mark, made up a number of components and modules, assessed by weekly problem sets and practical submissions as well as a seminar diary (see below)

ii) *Two examinations*: 25% of final mark. January exam covers Autumn material (10%), April exam (15%), mainly covers Spring term material, but assumes knowledge of previous material

iii) *The research project*: 50% of final mark, including + final viva + thesis (oral presentation + written report)

## MRes CMEE

MREs CMEE students will attend 9 weeks of taught modules over the Autumn Term followed by a 9-month research project leading to a dissertation (thesis) report). There are two main components to the course assessment:

i) *The coursework*: 25% of final mark, made up a number of components and modules, assessed by weekly problem sets and practical submissions as well as a seminar diary (see below)

ii) *The research project*: 75% of final mark, including final viva + thesis (oral presentation + written report)

*Both MSc and MRes CMEE projects must entail significant amounts of mathematical theory or statistical analyses of large datasets, or a combination of these.*

## External vivas and examiners

Details in External Examination policies and schedules can be found in the Student Guidebook. The current External Examiners are:

## Course Overview

| Prof. Robert Freckleton | University of Sheffield |
| Prof. Andrew Leitch | Queen Mary University of London |

## Weekly Seminars and Seminar Diary

Both MSc and MREs students must attend all the Thursday seminars unless told otherwise, even if one is rescheduled to a day other than Thursday. You will write a half-page on at least 16 of these seminars and create a "seminar journal" for submission as part of their coursework mark (see section on the Coursework Element). The Seminar diary is due at the end of the Spring term, the week before your final (project) dissertation submission 2.2.

## Workshops

We will organize a series of skills workshops in the induction week and then certain Wednesday afternoons (see the timetable in section 2.2), such as those involved in choosing a project, writing skills, choosing and applying for PhD places, etc.

## Computing and your computers

There are three levels of computing solutions available to CMEE masters students.

At the lowest and most immediately available level are your student laptops, issued in the Autumn (and returned in September at the end of the course). A 64-bit Linux operating system (most likely Ubuntu 14.04) will be available on the laptops (that's all you should need for the course). Necessary software will be installed and more can be installed by the students as and when needed. Part of being a good quantitative/computational biologist is achieving a level of mastery of management of your computer hardware and software – these are your main tools of trade. Much more information will be provided on using your computer in many ways, but you should also take it upon yourself to develop expertise in this area beyond what is taught.

In addition, the Hamilton Computer Room is on the first floor of the Hamilton Building and provides 44 desktop computers (with Ubuntu on virtual machines or dual boot), charging and network points for laptops and printing facilities. Certain computer practicals will be held in the Hamilton Computer Room.

At the intermediate level are two local, multi-core linux machines in the Pawar lab, called "William" and "Harvey". These make it relatively easy to learn to run large-ish parallel jobs. If you want access, please ask Samraat Pawar. You would be able to log onto one of them remotely and run single-core or multi-core simulations, statistical fitting, and other computing jobs, with a few rules or constraints to be followed out of respect for other users. William has 16 cores and Harvey has 12. They each have around 50Gb memory. Documentation for these can be found at https://bitbucket.org/mhasoba/pawarlab/raw/8d814b4571d265d3939b84f9d4fb793e2c716fce/Computing/LabServers/LabServers.pdf At the highest level is the Imperial College High Performance Computing (HPC) Cluster, which puts over 10,000 cores at the disposal of the researcher. For most purposes in quantitative biology this is essentially infinite computing power, if you learn to harness it. You will be taught how to harness it during the CMEE HPC module.

Information about more general computing and IT support are provided in the Student Guidebook.

## The Illumina Harvey Prizes

The student with highest final mark in MSc as well as MRes CMEE will be awarded a Illumina Harvey Prize each. Each award comes with a cash award of £500. But more importantly, it is an honour for a job very well done in a difficult course.

The cash prize is sponsored by Illumina, is a global company that develops innovative array-based solutions for DNA, RNA, and protein analysis; they are also some of the world-leaders in Next Generation DNA sequencing Technologies (`www.Illumina.com`).

The award is partly named after William Harvey (1578–1657), who provided a very early and very compelling example of combined quantitative and biological reasoning, leading to the first proof that blood circulates in the human body, and capillaries must exist connecting the arteries and veins in the circulatory system. Before the invention of the microscope, scientists were aware of arteries and veins but could not see the connections between them (capillaries), and so assumed they were not connected. They knew the heart was a pump, and dominant theories were that blood ebbed and flowed, back and forth like the tides in the arteries and veins, and that blood went out from the heart, was absorbed by the body, was regenerated in the liver which was thought to feed the veins, and the new blood flowed back to the heart. Harvey used biological reasoning, by observing that veins have valves, so blood can only flow through veins toward the heart, showing that the ebb-and-flow theory cannot be correct. Harvey used quantitative reasoning, by calculating the volume of the ventricles of the heart (about 1.5 imperial ounces) and the fraction of blood expelled from the ventricles on each heartbeat (about $/frac18$), and thereby calculating that if your heart beats 1000 times every half hour, the liver would have to produce 540 pounds of blood in a day for the dominant theory to be correct. He then developed the hypothesis of capillaries and blood circulation, later confirmed by Harveys experimental work and by the invention of the microscope.

## Background Reading

These books are reccommended to all students for wading into quantitative thinking in ecology and evolution. Module-specific readings can be found later on in this Guidebook.

- Kingsland, S. E. 1995 Modeling nature. Episodes in the history of population ecology. 2nd edn. Chicago, Illinois: University of Chicago Press.

- Otto, S. and Day, T. 2007 A biologists guide to mathematical modeling in ecology and evolution. New Jersey: Princeton University Press.

- Levins, R. 1966 The strategy of model building in population biology. Am. Sci. 54, 421–431.

- In addition, some might find a essay written by our very own Samraat Pawar when he was a young and energetic PhD student interesting!: `https://dl.dropboxusercontent.com/u/44578955/The_Turbulent_Marriage01.pdf`

# Course details, timetables, and module descriptions

## Teaching Staff

| | |
|---|---|
| Dr Alberto Pascual-Garcia | Theoretical ecology and complex networks (*a.pascual-garcia@imperial.ac.uk*) |
| Dr Matteo Fumagalli | Computational Genomics; Methods for DNA sequencing data analysis; Genetic adaptation and human disease susceptibility; Past demography and natural selection (*m.fumagalli@ucl.ac.uk*) |
| Mr Tin-Yu Hui | Statistical genetics (*tin-yu.hui11@imperial.ac.uk*) |
| Prof Tim Barraclough | Evolution of species diversity; asexual evolution; experimental evolution. (*t.barraclough@imperial.ac.uk*) |
| Prof Austin Burt | Evolutionary biology of selfish genetic elements (*a.burt@imperial.ac.uk*) |
| Dr Samraat Pawar | Theoretical Ecology; Ecological Systems Biology; Ecoinformatics; Metabolic theory and physiological ecology. (*s.pawar@imperial.ac.uk*) |
| Dr James Rosindell | Theoretical Ecology; Biodiversity theory; Island biogeography; Ecological neutral theory; Scientific data visualisation. (*j.rosindell@imperial.ac.uk*) |
| Dr Julia Schroeder | Genetic control of social behavior and interactions, quantitative genetics (*julia.schroeder@imperial.ac.uk*) |
| Prof Vincent Jansen | Mathematical biology (*01784 443179, Vincent.Jansen@rhul.ac.uk*) |
| Dr Isabel Rosa | Stochastic models of land cover change (*i.rosa09@imperial.ac.uk*) |
| Dr Jason Hodgson | Evolutionary genomics and bioinformatics in primate evolution (*j.hodgson@imperial.ac.uk*) |
| Dr Rob Ewers | Spatial patterns of forest biodiversity (*r.ewers@imperial.ac.uk*) |
| Dr David Orme | Spatial patterns of biodiversity; Macroecology; Ecoinformatics (*d.orme@imperial.ac.uk*) |
| Dr Maria Dickinson | Spatial patterns of extinction Risk; Climate change (*m.dickinson@imperial.ac.uk*) |
| Dr Martin Brazeau | Palaeontology; Morphology and evolution; Phylogenetics (*m.brazeau@imperial.ac.uk*) |

## Outline timetable and important dates

| Week | Dates | MSc CMEE | MRes CMEE |
|---|---|---|---|
| | | **Autumn term** | |
| 1 | 3 – 7 Oct | Induction & Foundations of Biological Computing | |
| 2 | 10–14 Oct | Biological Computing in Python I | |
| 3 | 17–21 Oct | Biological Computing in R | |
| 4 | 24–28 Oct | Statistics in R | |
| 5 | 31 Oct–4 Nov | Spatial Analyses & Geographic Information Systems (GIS) | |
| 6 | 7–11 Nov | Genomics and Bioinformatics | |
| 7 | 14–18 Nov | Biological Computing in Python II | |

| 8  | 21–25 Nov     | CMEE Miniproject | |
| 9  | 28 Nov–2 Dec  | High Performance Computing | |
| 10 | 5–9 Dec       | Population Genetics | Project |
| 11 | 12–16 Dec     | Reading | |

**Spring term**

| 15 | 9–13 Jan      | Exam 1 | |
| 16 | 16–20 Jan     | Generalised Linear Modelling | |
| 17 | 23–27 Jan     | Maths for Biology | |
| 18 | 30 Jan–3 Feb  | Maths for Biology | |
| 19 | 6–10 Feb      | Maximum Likelihood | Project |
| 20 | 13–17 Feb     | Bayesian Statistics | |
| 21 | 20–24 Feb     | Evolutionary Modelling | |
| 22 | 27 Feb–3 Mar  | Ecological Modelling | |
| 23 | 6–10 Mar      | Biological Data Structures and C | |
| 24 | 13–17 Mar     | Reading | |
| 25 | 20–24 Mar     | Reading | |
| 26 | 27–31 Mar     | Exam 2 | |

*For later weeks and other dates see below*

**Other important dates and deadlines:**

| Date          | Activity/Item due |
|---------------|-------------------|
| 9 Dec, 5PM    | MRes: Project proposal |
| 15 Dec, 5PM   | MSc, MRes: HPC Long Practical |
| 17 Feb, 5PM   | MSc, MRes: CMEE Miniproject |
| 7 April, 5PM  | MSc, Project proposal |
| 25 Aug, 5pm   | MSc, MRes: Seminar Diary |

**Further important dates, including thesis submission are same across all Silwood Masters courses – please refer to the Silwood Student Guidebook**

**CMEE Masters project proposals should be written in LATEX, and the source code submitted (pushed), in pdf format to your bitbucket git repository into a directory called Proposal inside a Project directory**. Please ask Samraat if you have any questions about this, and refer to the Silwood Student Guidebook for proposal preparation guidelines.

**Workshops:**

In additon, there are a number of professional skill development workshops, the dates for which are given in the detailed weekly timetables below.

## Taught Module Descriptions

All teaching activities are typically scheduled from 10.00am until 5.00pm, except Wednesday afternoon. The morning lectures in the modules/weeks on computing (UNIX/Linux, Python, R, etc.) are inteactive, will all require you to use your laptop — so please bring you laptop to every class and session! The afternoon sessions of these modules will involve practicals, and you will typically be assisted by demonstrators (typically one per 6 students).

# Course details, timetables, and module descriptions

Also, please note that most if not all the recommended readings and resources in the following module descriptions are available in Central or Silwood Libraries, and often also as e-books.

## Course Induction & Foundations of Computing

**Week:** October 03 2016–October 07 2016

**Convenor:** Samraat Pawar

This module introduces the philosophy of the CMEE Masters course and focuses on training in fundamentals of scientific computing, including an introduction to UNIX and Linux, shell scripting, LaTeX, and version control. Along with technical training, The week begins with a number of important induction events. All lectures and practicals will be computer (laptop)-based.

**Aims:**

- Learn how to use the UNIX environment and the terminal for tasks ranging from data exploration to simple calculations and data processing.

- Learn about basic data types and structures.

- Learn how to write shell scripts for automating computing tasks.

- Learn how to LaTeXfor tysetting documents in a reproducible, consistent (and elegant!) way.

- Learn how to version control computer code with git.

**Timetable:**

| Monday, October 03 2016 | |
|---|---|
| 09:00 – 10:00 | *Welcome to Silwood Park* (Lecture, Fisher+Haldane) |
| 10:00 – 10:30 | *Computer and key issue, meet each other* (Social, Wallace) |
| 10:30 – 11:00 | *Introduction to the CMEE Masters* (Presentation, Wallace) |
| 11:00 – 13:00 | *Silwood Treasure Hunt* (Social, Silwood Field) — In teams with other new Masters and PhDs, you will scour Silwood for answers to Tim Barracloughs fiendish geographically and ecologically based questions. Meet at the Hamilton building foyer and be prepared in case of bad weather. |
| 13:00 – 14:00 | *Buffet Lunch* (Social, Hamilton Foyer) |
| 14:00 – 16:00 | *The Big Picture: Group Discussion* (Discussion, Hamilton Foyer) |
| Tuesday, October 04 2016 | |
| 09:30 – 10:00 | *Computer setup* (Practical, Wallace) |
| 10:00 – 12:30 | *Intro to UNIX and Linux* (Lecture+Exercises, Wallace) |
| 13:30 – 17:00 | *How to get addicted to the terminal and shell scripting* (Lecture+Exercises, Wallace) |
| 17:00 – 19:00 | *Welcome Reception* (Social, CPB Seminar Room) |
| Wednesday, October 05 2016 | |
| 09:30 – 11:00 | *Version control with Git* (Lecture+Exercises, Wallace) |
| 11:15 – 12:30 | *Scientific typesetting using LaTeX* (Lecture+Exercises, Wallace) |
| 15:30 – 17:00 | *Introduction to the Library* (Induction, Hamilton Computer Room) |
| Thursday, October 06 2016 | |
| 09:30 – 13:00 | *Professional Skills Development Programme* (Workshop, Fisher/Haldane/Darwin/CPB Seminar Room) |
| 14:00 – 17:30 | *Professional Skills Development Programme* (Workshop, Fisher/Haldane/Darwin/CPB Seminar Room) |
| 15:45 – 17:15 | *Silwood Labs & Projects Talks* (Presentation, Fisher+Haldane) |
| 17:30 – 18:30 | *Silwood Labs & Projects + Networking Plenary and Refreshments* (Workshop, Hamilton Foyer) |

| Friday, October 07 2016 | |
|---|---|
| 10:00 – 11:00 | *Provost's Welcome* (Lecture, Fisher/Haldane) |
| 11:00 – 13:00 | *Professional Skills Development Programme* (Workshop, Fisher/Haldane/Wallace) |
| 14:00 – 15:00 | *Safety Induction* (Lecture, Fisher/Haldane) |
| 15:00 – 17:00 | *Professional Skills Development Programme* (Workshop, Fisher/Haldane/Wallace/CPB Seminar Room) |

**Readings and Resources:**

- http://software-carpentry.org (Chapter "shell"). Watch the video tutorials and/or read pdfs of the presentations.

- The Imperial College library provides you with access to several e-books on UNIX, some specific to Mac OSX or Ubuntu, and some more general. Go through the http://www3.imperial.ac.uk/library website.

- An extensive list of UNIX commands can be found here (along with their man page): www.oreillynet.com/linux/cmd/.

- http://www.git-scm.com/book/en/v2 hosts a great book on git.

- There are also several tutorials on the internet: especially look at http://www.sbf5.com/~cduan/technical/git/ and https://www.atlassian.com/git/

- (Health?) benefits of using latex: http://www.andy-roberts.net/writing/latex/benefits

- Word vs. LaTeX: http://openwetware.org/wiki/Word_vs._LaTeX

- Leslie Lamport, LaTeX: A document preparation system, users guide and reference manual, 1994, Addison-Wesley.

- Myriad other online resources for LaTeX, including www.http://en.wikibooks.org/wiki/LaTeX/Introduction, www.ctan.org/tex-archive/info/lshort/english/, and http://ftp.uni-erlangen.de/mirrors/CTAN/info/lshort/english/lshort.pdf

- Bibliographies in LaTeX: https://schneider.ncifcrf.gov/latex.html

# Additonal Graduate School Workshops - Professional Skills Development Programme

**Convenor:** Janet De Wilde

**Timetable:**

| Friday, January 13 2017 | | |
|---|---|---|
| 10:00 – 13:00 | *Professional Skills Development Programme* Fisher/Haldane/Wallace/CPB Seminar Room) | (Workshop, |
| 14:00 – 17:00 | *Professional Skills Development Programme* Fisher/Haldane/Wallace/CPB Seminar Room) | (Workshop, |

## Biological Computing in Python I

**Week:** October 10 2016–October 14 2016

**Convenor:** Samraat Pawar

This is the first of a two-week module on biological computing in the Python language. With the profusion of genomic, environmental and ecological information, the ability to develop automated, reproducible analyses of massive datasets using computer scripts and programs is an essential skill for any research student. The aim of this module is to introduce the basics of programming using Python, which is a modern, easy-to-write, interpreted (semi-compiled) language that was conceived with readability of script in mind. It has a feature-rich set of packages that can be used for a wide variety of applications and analyses. The approach will be hands-on and informal, involving lectures interspersed with short exercises in class. There will be demonstrator-aided longer exercises (and an opportunity for you to catch up!) during the last three hours at the end of each day.

**Aims:**

To learn principles of computer program design and scientific computing, using Python. Specifically,

- Learn to write simple functions and programs in Python.

- Understand of the basics of Python program testing, debugging and documentation.

- Learn to use Python for retrieving, managing, and analyzing data from local and remote databases.

- Learn to automate file handling, string manipulation, and run shell scripts.

- Learn to use Python for efficient numerical analyses.

- Learn to run analyses by patching together `R` or `R` + Python scripts and functions.

**Timetable:**

| Monday, October 10 2016 | |
|---|---|
| 09:30 – 10:00 | *Git and LATEX Q&A* (Lecture, Wallace) |
| 10:00 – 11:00 | *Why write computer programs?* (Lectures+Exercises, Wallace) |
| 11:30 – 12:30 | *OK, but why Python?* (Lectures+Exercises, Wallace) |
| 14:00 – 17:00 | *Programming in Python - basics* (Lecture+Short Practicals, Wallace) |
| **Tuesday, October 11 2016** | |
| 10:00 – 11:00 | *Python basics continued* (Lectures+Exercises, Wallace) |
| 11:30 – 12:30 | *Python basics continued* (Lectures+Exercises, Wallace) |
| 14:00 – 17:00 | *Python basics continued* (Lecture+Practicals, Wallace) |
| **Wednesday, October 12 2016** | |
| 10:00 – 11:00 | *Writing Python Code* (Lecture+Short Practicals, Wallace) |
| 11:30 – 12:30 | *Writing Python Code* (Lecture+Short Practicals, Wallace) |
| **Thursday, October 13 2016** | |
| 10:00 – 11:00 | *Writing and testing Python functions* (Lectures+Exercises, Wallace) |
| 11:30 – 12:30 | *Writing and testing Python functions* (Lectures+Exercises, Wallace) |
| 14:00 – 17:00 | *Writing and testing Python functions* (Lecture+Practicals, Wallace) |
| **Friday, October 14 2016** | |
| 10:00 – 11:00 | *Writing and testing Python functions* (Lectures+Exercises, Wallace) |
| 11:30 – 12:30 | *Writing and testing Python functions* (Lectures+Exercises, Wallace) |

14:00 – 17:00 | *Writing and testing Python functions* (Lecture+Practicals, Wallace)

**Readings and Resources:**

- Browse the Python tutorial: https://docs.python.org/2/tutorial/

- For functions and modules: www.learnPythonthehardway.org/book/ex40.html

- For IPython: http://iPython.org/documentation.html and https://github.com/iPython/iPython/wiki?path=Cookbook

## Biological Computing in R

**Week:** October 17 2016–October 21 2016

**Convenor:** Samraat Pawar

In this module, you will learn how to use this freely available statistical software with strong programming capabilities. R has become tremendously popular in Biology due to several factors: (i) many packages are available to perform all sorts of statistical and mathematical analysis, (ii) it can produce beautiful graphics, and (iii) it has a very good support for matrix-algebra (you might not know it, but you use it!). So with R, you have an expanded and versatile suite of biological computing tools at your fingertips, especially for automating statistical analysis and the generation of figures. Therefore, R should become an indispensable component of your biological research workflow. This week is shared with other MSc/MRes courses. CMEE students should bring their laptops to all sessions and use them instead of the desktops in the Computer Room. The membership of groups A & B will announced the week before.

**Aims:**

You will learn:

- how to use R for data exploration
- how to use R for data visualization and producing elegant, intuitive, and publication quality graphics.
- R data types & structures and control flows.
- how to write and debug efficient R scripts and functions.
- how to use R packages.

**Timetable:**

| Monday, October 17 2016 | |
|---|---|
| 09:00 – 09:30 | *Why R? Why us? (Group A+B)* (Lecture, Fisher/Haldane) |
| 09:30 – 11:00 | *Easing into R (Group A)* (Lectures + Exercises, Hamilton Computer Room) |
| 11:00 – 12:30 | *Useful R commands and functions (Group A)* (Lectures + Exercises, Hamilton Computer Room) |
| 13:30 – 15:00 | *Easing into R (Group B)* (Lectures + Exercises, Hamilton Computer Room) |
| 15:00 – 16:30 | *Useful R commands and functions (Group B)* (Lectures + Exercises, Hamilton Computer Room) |
| **Tuesday, October 18 2016** | |
| 09:15 – 10:45 | *Programming in R (Group B)* (Lectures + Exercises, Hamilton Computer Room) |
| 11:00 – 12:30 | *Programming in R (Group B)* (Lectures + Exercises, Hamilton Computer Room) |
| 13:30 – 15:00 | *Programming in R (Group A)* (Lectures + Exercises, Hamilton Computer Room) |
| 15:15 – 16:45 | *Programming in R (Group A)* (Lectures + Exercises, Hamilton Computer Room) |
| **Wednesday, October 19 2016** | |
| 09:15 – 10:45 | *R Help Sessions* (Lectures + Exercises, Hamilton Computer Room) |
| 11:00 – 12:30 | *R Help Sessions* (Lectures + Exercises, Hamilton Computer Room) |
| **Thursday, October 20 2016** | |
| 09:15 – 10:45 | *Numerical R (Group A)* (Lectures + Exercises, Hamilton Computer Room) |
| 11:00 – 12:30 | *Numerical R (Group B)* (Lectures + Exercises, Hamilton Computer Room) |
| 14:00 – 15:30 | *Efficient R (Group B)* (Lectures + Exercises, Hamilton Computer Room) |

| | |
|---|---|
| 15:15 – 16:45 | *Efficient R (Group B)* (Lectures + Exercises, Hamilton Computer Room) |
| **Friday, October 21 2016** | |
| 09:15 – 10:45 | *Graphics in R (Group B)* (Lectures + Exercises, Hamilton Computer Room) |
| 11:00 – 12:30 | *Advanced graphics in R (Group B)* (Lectures + Exercises, Hamilton Computer Room) |
| 13:30 – 15:00 | *Graphics in R (Group A)* (Lectures + Exercises, Hamilton Computer Room) |
| 15:15 – 16:45 | *Advanced graphics in R (Group A)* (Lectures + Exercises, Hamilton Computer Room) |

**Readings and Resources:**

- The Use R! series (the yellow books) by Springer are really good. In particular, consider: 'A Beginner's Guide to R', 'R by Example', 'Numerical Ecology With R', 'ggplot2' (we'll see this in another week), 'A Primer of Ecology with R', 'Nonlinear Regression with R', 'Analysis of Phylogenetics and Evolution with R'.

- Ben Bolker's 'Ecological Models and Data in R'

- For more focus on dynamical models: Soetaert & Herman. 2009 'A practical guide to Ecological Modelling: using R as a simulation platform'.

- There are excellent websites. Besides CRAN (containing all sorts of guides and manuals), you should check out www.statmethods.net and en.wikibooks.org/wiki/R_Programming and google 'R Graph Gallery' for various sites showing graphing options and code.

## Statistics in R

**Week:** October 24 2016–October 28 2016

**Convenor:** Julia Schroeder

In this module we will build upon the introduction to R you received in "Biological Computing in R" and review a core set of statistical methods that are of wide use in research projects. These statistical tests will form the basis for many data analyses you will do in the future. This module is shared with most courses and runs in two blocks A and B like the previous module. This week is shared with other MSc/MRes courses. CMEE students should bring their laptops to all sessions and use them instead of the desktops in the Computer Room. The membership of groups A & B will be the same as in the Biological Computing in R week.

**Aims:**

We will cover:

- the difference between response and explanatory variables and between ordinal, categorical and continuous variables;

- the underlying structure of statistical testing using both parametric and non-parametric approaches;

- tests for assessing differences between samples and correlation between samples;

- analysis of categorical data;

- and fitting and assessing linear models of continuous response variables.

**Timetable:**

| Monday, October 24 2016 | |
|---|---|
| 09:15 – 12:15 | *Probability theory and distributions (Group A)* (Lectures + Exercises, Hamilton Computer Room) |
| 13:30 – 16:30 | *Probability theory and distributions (Group B)* (Lectures + Exercises, Hamilton Computer Room) |
| **Tuesday, October 25 2016** | |
| 09:15 – 12:15 | *Comparing two samples and correlations (Group A)* (Lectures + Exercises, Hamilton Computer Room) |
| 13:30 – 16:30 | *Comparing two samples and correlations (Group B)* (Lectures + Exercises, Hamilton Computer Room) |
| **Wednesday, October 26 2016** | |
| 09:15 – 10:45 | *Inference (A+B)* (Lecture, Fisher/Haldane) |
| 11:00 – 12:00 | *Inference (A+B)* (Discussion, Fisher/Haldane) |
| 13:30 – 15:30 | *Choosing and Designing a Research Project* (Workshop, Fisher+Haldane) |
| **Thursday, October 27 2016** | |
| 09:15 – 12:15 | *Linear regression analysis (Group A)* (Lectures + Exercises, Hamilton Computer Room) |
| 13:30 – 16:30 | *Linear regression analysis (Group B)* (Lectures + Exercises, Hamilton Computer Room) |
| **Friday, October 28 2016** | |
| 09:15 – 12:15 | *Multiple linear regression (Group A)* (Lectures + Exercises, Hamilton Computer Room) |

| 13:30 – 16:30 | *Multiple linear regression (Group B)* (Lectures + Exercises, Hamilton Computer Room) |
|---|---|

**Readings and Resources:**

There are a wide range of introductory books for R. See later statistics and computing modules for more specialist texts but, for this week, the following are good introductory and reference texts that are available in Silwood library and as an e-book through Imperial: Main references:

- Crawley, Michael J (2012) Statistics: An Introduction Using R. John Wiley. A gentler introduction:

- Beckerman, Andrew P. and Petchey, Owen (2012) Getting Started with R : An introduction for biologists Oxford University Press.

## Spatial Analyses and Geographic Information Systems (GIS)

**Week:** October 31 2016–November 04 2016

**Convenor:** Robert Ewers

This week will teach key skills in using and handling GIS data, along with basic remote sensing to generate GIS data and the use of GIS data in a range of applications. We will use the open source GIS program QGIS (`http://www.qgis.org/`). We will look at creating and georeferencing both vector and raster data and how to use GIS tools to create a workflow to carry out simple analyses. This week is shared with other MSc/MRes courses. CMEE students should bring their laptops to all sessions and use them instead of the desktops in the Computer Room.

**Aims:**

At the end of this module you should have:

- Familiarity with a range of GIS data types

- Confidence in obtaining and handling GIS data

- Familiarity with open source tools for GIS

- Practical experience in applying GIS to ecological and evolutionary questions

**Timetable:**

| Monday, October 31 2016 | |
|---|---|
| 10:00 – 11:00 | *Introduction to GIS*  (Lecture, Fisher+Haldane) |
| 11:00 – 12:00 | *Remote sensing and classification*  (Lecture, Fisher+Haldane) |
| 13:00 – 17:00 | *Supervised and unsupervised classification*  (Practical, Hamilton Computer Lab) |
| **Tuesday, November 01 2016** | |
| 10:00 – 11:00 | *Species distribution models*  (Lecture, Fisher+Haldane) |
| 11:00 – 12:00 | *Assessing model accuracy*  (Lecture, Fisher+Haldane) |
| 13:00 – 17:00 | *Species distribution modelling*  (Practical, Fisher+Haldane) |
| **Wednesday, November 02 2016** | |
| 10:00 – 11:00 | *Climate change and species distributions 1*  (Lecture,Fisher+Haldane) |
| 11:00 – 12:00 | *Climate change and species distributions 2*  (Lecture,Fisher+Haldane ) |
| **Thursday, November 03 2016** | |
| 10:00 – 11:00 | *Essentials of land use change modelling*  (Lecture, Fisher+Haldane) |
| 11:00 – 12:00 | *Types of land use change models*  (Lecture, Fisher+Haldane) |
| 13:00 – 17:00 | *Modelling tropical deforestation*  (Practical, Fisher+Haldane) |
| **Friday, November 04 2016** | |
| 10:00 – 11:00 | *Spatial modelling*  (Lecture, Fisher+Haldane) |
| 13:00 – 17:00 | *Spatial modelling*  (Practical, Hamilton Computer Lab) |

**Readings and Resources:**

- GIS overview: Longley, PA (2011) Geographical information systems and science. Wiley.

- Coordinate systems: Van Sickle, G (2010) Basic GIS coordinates. CRC Press `https://www.dawsonera.com/abstract/9781420092325`.

- Land use change modelling: Rosa et al. (2013) Pedictive modelling of contagious deforestation in the Brazilian Amazon. PLoS ONE 8:e77231.

## Genomics and Bioinformatics

**Week:** November 07 2016–November 11 2016

**Convenor:** Jason Hodgson

Genetic data contain information about who organisms are, their relationships to other organisms, their population histories, and their histories of adaptation. Thus, genetic data and genetic techniques are central to addressing many questions in evolution, ecology, and conservation. New technologies allow for genetic characterization at the genomic level, and these data allow for an understanding of population processes at resolutions not possible in the past. The goal of this module is to introduce students to the types of questions that can be addressed with genomic data, and the methodologies that are available for answering these questions. Learning will be accomplished through a mix of lectures, computer practicals and group discussions. This week is shared with other MSc/MRes courses. CMEE students should bring their laptops to all sessions and use them instead of the desktops in the Computer Room.

**Aims:**

- An understanding of genomic data collection methods, and how to choose the data collection technique most appropriate to your question.

- An understanding of the wealth of data available to biologists in public genomic databases.

- An understanding of how genetic structure develops within and between populations, how to characterise it, and how to interpret the results of common analyses such as STRUCTURE and PCA.

- An understanding of how demographic history affects genomic variation, and how to infer past population expansions and contractions from genomic data.

- An understanding of how migration affects genomic variation, and how patterns of gene flow can be inferred from genomic data.

- An understanding of how natural selection affects genomic variation, and how selection can be identified from genomic data.

- An understanding of how phylogenetic relationships among species can be inferred, and what this information can tell us about evolution and conservation efforts.

**Timetable:**

| Monday, November 07 2016 | |
|---|---|
| 10:00 – 11:00 | *Intro* (Lecture, Fisher+Haldane) |
| 11:30 – 12:30 | *Population Structure* (Lecture, Fisher+Haldane) |
| 13:30 – 16:30 | *Data and databases practical* (Computer Practical, Hamilton Computer Lab) |
| **Tuesday, November 08 2016** | |
| 10:00 – 11:00 | *Demographic history* (Lecture, Fisher+Haldane) |
| 11:30 – 12:30 | *Gene flow and migration* (Lecture, Fisher+Haldane) |
| 13:30 – 16:30 | *Population structure and migration practical* (Computer Practical, Hamilton Computer Lab) |
| **Wednesday, November 09 2016** | |
| 09:30 – 10:30 | *Natural Selection* (Lecture, Fisher+Haldane) |
| 11:00 – 12:00 | *Genomic approaches to ecology and conservation* (Lecture, Fisher+Haldane) |
| **Thursday, November 10 2016** | |

| 10:00 – 11:00 | *Phylogenomics* (Lecture, Fisher+Haldane) |
| 11:30 – 12:30 | *Detecting selection practical* (Computer Practical, Hamilton Computer Lab) |
| 14:30 – 17:00 | *Phylogenomics practical* (Computer Practical, Hamilton Computer Lab) |
| **Friday, November 11 2016** | |
| 10:30 – 11:30 | *Study design discussion* (Discussion, Fisher+Haldane) |
| 13:30 – 15:30 | *Study design presentations* (Discussion, Fisher+Haldane) |

**Readings and Resources:**

- Novembre, J. & Ramachandran, S. Perspectives on human population structure at the cusp of the sequencing era. Annual review of genomics and human genetics 12, 245-274, doi:10.1146/annurev-genom-090810-183123 (2011).

- Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. Genetics 155, 945-959 (2000).

- Bertorelle, G., Benazzo, A. & Mona, S. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. Molecular ecology 19, 2609-2625, doi:10.1111/j.1365-294X.2010.04690.x (2010).

- Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. Nature 475, 493-496, doi:10.1038/nature10231 (2011).

- Pozzi, L. et al. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. Molecular phylogenetics and evolution, doi:10.1016/j.ympev.2014.02.023 (2014).

- Purvis, A., Agapow, P. M., Gittleman, J. L. & Mace, G. M. Nonrandom extinction and the loss of evolutionary history. Science 288, 328-330 (2000).

- Jarvis, E. D. et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346, 1320-1331, doi:10.1126/science.1253451 (2014).

- Zhang, G. et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346, 1311-1320, doi:10.1126/science.1251385 (2014).

- Barreiro, L. B., Laval, G., Quach, H., Patin, E. & Quintana-Murci, L. Natural selection has driven population differentiation in modern humans. Nat Genet 40, 340-345 (2008).

- Sabeti, P. C. et al. Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913-918 (2007).

- Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. K. & Hedrick, P. W. Conservation genetics in transition to conservation genomics. Trends Genet 26, 177-187, doi:10.1016/j.tig.2010.01.001 (2010).

- Lopes, C. M. et al. DNA metabarcoding diet analysis for species with parapatric vs sympatric distribution: a case study on subterranean rodents. Heredity (Edinb) 114, 525-536, doi:10.1038/hdy.2014.109 (2015).

# Biological Computing in Python II

**Week:** November 14 2016–November 18 2016

**Convenor:** Samraat Pawar

This is the second of the two-week module on biological computing in the Python language. The aims, format, and venue of the lectures and practicals are same as same as *Biological computing in Python I*.

**Aims:**

Same as *Biological computing in Python I*

**Timetable:**

| Monday, November 14 2016 | |
|---|---|
| 10:00 – 11:00 | *Profiling and debugging in Python* (Lectures+Exercises, Wallace) |
| 11:30 – 12:30 | *Profiling and debugging in Python* (Lectures+Exercises, Wallace) |
| 14:00 – 17:00 | *Profiling and debugging in Python* (Lecture+Practicals, Wallace) |
| **Tuesday, November 15 2016** | |
| 10:00 – 11:00 | *Useful Python packages and tools* (Lectures+Exercises, Wallace) |
| 11:30 – 12:30 | *Useful Python packages and tools* (Lectures+Exercises, Wallace) |
| 14:00 – 17:00 | *Useful Python packages and tools* (Lecture+Practicals, Wallace) |
| **Wednesday, November 16 2016** | |
| 10:00 – 11:00 | *Scipy and scientific computing (number crunching!) in Python* (Lecture+Practicals, Wallace) |
| 11:30 – 12:30 | *Scipy and scientific computing (number crunching!) in Python* (Lecture+Practicals, Wallace) |
| **Thursday, November 17 2016** | |
| 10:00 – 11:00 | *Databases and Python* (Lectures+Exercises, Wallace) |
| 11:30 – 12:30 | *Databases and Python* (Lectures+Exercises, Wallace) |
| 14:00 – 17:00 | *Databases and Python* (Lecture+Practicals, Wallace) |
| **Friday, November 18 2016** | |
| 10:00 – 11:00 | *Build and run your workflow with Python* (Lectures+Exercises, Wallace) |
| 11:30 – 12:30 | *Build and run your workflow with Python* (Lectures+Exercises, Wallace) |
| 14:00 – 17:00 | *Build and run your workflow with Python* (Lecture+Practicals, Wallace) |

**Readings and Resources:**

- `www.docs.Python.org/2/howto/regex.html`

- Googles short class on regex in Python: `www.code.google.com/edu/languages/` and `www.google-Python-class/regular-expressions.html`

- `www.regular-expressions.info` has a good intro, tips and a great array of canned solutions

- For SciPy and Matplotlib, the official documentation is good: `www.docs.scipy.org/doc/scipy/reference/` and `www.http://matplotlib.org/`

- "The Definitive Guide to SQLite" is a pretty complete guide and freely available from `http://evalenzu.mat.utfsm.cl/Docencia/2012/SQLite.pdf`

## CMEE Miniproject

**Week:** November 21 2016–November 25 2016

**Convenor:** Samraat Pawar

In this week, you will choose and start working on your miniproject assignment. The week will start off with an introduction to your miniproject assignment and objectives, after which you will choose your miniproject, and work on it intensively during thg the week, with the aim of getting a significant amount of work done. You will then continue working in subsequent weeks on the miniproject. The deadline for submission is given in the key dates table. The overall aim of the miniproject is to give you an opportunity to try a "dry run" of your masters project, with focus on the computing (especially developing reproducible workflows) and write up. You may choose a project that is related to your masters dissertation project! Further details, gudelines, and marking criteria can be found in the CMEE Masters computing course notes.

**Aims:**

The specific aims of the miniproject assignment are:

- To carry out a computationally intensive analysis that includes elements of shell scripting, R, & Python

- To learn to address a question involving data processing and model fitting (e.g., using non-linear-least squares fitting).

- To learn to write up and compile a meaningful report on the analsysis / study using LaTeX

- To learn to keep the project workflow under version control and to be able to run the whole analysis and produce writeup reproducibly and automatically

**Timetable:**

| Monday, November 21 2016 | |
|---|---|
| 10:00 – 11:00 | *Introduction to the miniproject* (Lecture, Wallace) |
| 11:30 – 12:30 | *Introduction to the miniproject* (Lecture, Wallace) |
| 14:00 – 17:00 | *Choose miniproject* (Lecture, Wallace) |
| **Tuesday, November 22 2016** | |
| 10:00 – 12:30 | *Work on miniproject!* (Lecture, Wallace) |
| 14:00 – 17:00 | *Work on miniproject!* (Lecture, Wallace) |
| **Wednesday, November 23 2016** | |
| 10:00 – 12:30 | *Work on miniproject!* (Lecture, Wallace) |
| **Thursday, November 24 2016** | |
| 10:00 – 12:30 | *Work on miniproject!* (Lecture, Wallace) |
| 14:00 – 17:00 | *Work on miniproject!* (Lecture, Wallace) |
| **Friday, November 25 2016** | |
| 10:00 – 12:30 | *Work on miniproject!* (Lecture, Wallace) |
| 14:00 – 17:00 | *Work on miniproject!* (Lecture, Wallace) |

**Readings and Resources:**

- Depends upon your choice of project!

## High Performance Computing

**Week:** November 28 2016–December 02 2016

**Convenor:** James Rosindell

The use of high performance computing is becoming increasingly important in biology. For certain computational tasks we can use large numbers of CPUs in parallel to get numerical results in days that would otherwise have taken years. This module will introduce students to the tools and techniques of high performance computing for biological problems using R. It will also be a good opportunity to develop practical programming skills in the R programming language. However, there will also be a Python HPC session at the end of the week. This module has a greater proportion of practical than usual because the best way to learn about programming is to try it, the convener will help students individually and be available to answer questions throughout all practical sessions. The biological topics covered will include individual based models, ecological neutral theory and fractals in biology but the techniques learned will be useful much more generally. A significant number of CMEE students typically end up using high performance computing as a tool for their research projects.

**Aims:**

Learning objectives are:

- Develop an advanced understanding of programming in R by tackling some more difficult problems.

- Learn the principles of using High Performance Computing (HPC) to crack otherwise intractable computational problems.

- The computational problems studied in class are chosen to have relevance to ecology and evolution so students will also learn about some new areas of biology along the way:

  - Applying ecological neutral theory and individual based models of community assembly

  - Understanding fractal geometry and its relevance in biology

**Timetable:**

| Monday, November 28 2016 | |
|---|---|
| 10:00 – 11:00 | *Introduction and individual based models* (Lecture, Wallace) |
| 11:30 – 12:30 | *Introduction and individual based models* (Lecture, Wallace) |
| 14:00 – 17:00 | *IB models practical* (Practicals, Wallace) |
| **Tuesday, November 29 2016** | |
| 10:00 – 11:00 | *Using HPC* (Lecture, Wallace) |
| 11:30 – 12:30 | *Using HPC* (Practicals, Wallace) |
| 14:00 – 17:00 | *Using HPC* (Practicals, Wallace) |
| **Wednesday, November 30 2016** | |
| 10:00 – 11:00 | *Coalescence methods* (Lecture, Wallace) |
| 11:30 – 12:30 | *Coalescence methods* (Practicals, Wallace) |
| **Thursday, December 01 2016** | |
| 10:00 – 11:00 | *Fractal geometry* (Lecture, Wallace) |
| 11:30 – 12:30 | *Fractal geometry* (Practicals, Wallace) |
| 14:00 – 17:00 | *Fractal geometry* (Practicals, Wallace) |
| **Friday, December 02 2016** | |
| 10:00 – 11:00 | *Fractals in nature* (Lecture, Wallace) |

| | |
|---|---|
| 11:30 – 12:30 | *Fractals in nature*  (Practicals, Wallace) |
| 14:00 – 17:00 | *HPC with Python*  (Lecture+Practicals, Wallace) |

**Readings and Resources:**

- The best preparation for this module is to be well practiced at programming, especially in R.

## Population Genetics

**Week:** December 05 2016–December 09 2016

**Convenor:** Julia Schroeder

A better understanding of population genetics and ecology is crucial for work in conservation, but also for basic ecology and evolution. This one-week course introduces students to classical population ecology, demography, population modelling, survival models, and quantitative genetics in wild populations.

**Aims:**

- Understand dynamics of wild populations

- Apply population dynamical models to own data

- Understand how to apply knowledge from population dynamical models to real world problems

- Learn how to do quantitative genetics in the wild

**Timetable:**

| Monday, December 05 2016 | |
|---|---|
| 10:00 – 12:00 | *Demography and population ecology*  (Lecture, Hamilton Computer Room) |
| 13:30 – 16:30 | *Demography*  (Practical, Hamilton Computer Room) |
| **Tuesday, December 06 2016** | |
| 10:00 – 11:00 | *Wrapping up population dynamics*  (Lectures, Hamilton Computer Room) |
| 11:15 – 12:15 | *Case study*  (Lectures, Hamilton Computer Room) |
| 13:00 – 14:30 | *Repeatability and Heritability*  (Lecture and practical, Hamilton Computer Room) |
| **Wednesday, December 07 2016** | |
| 10:00 – 12:30 | *Introduction to miniprojects*  (Discussion, Hamilton Computer Room) |
| **Thursday, December 08 2016** | |
| 10:00 – 11:30 | *Quantitative genetics in the wild*  (Lecture, Hamilton Computer Room) |
| 13:30 – 15:30 | *Estimating heritability*  (Practical, Hamilton Computer Room) |
| **Friday, December 09 2016** | |
| 10:00 – 11:00 | *Integral projection models, case study*  (Lecture, Hamilton Computer Room) |
| 13:30 – 16:00 | *Selection gradients, predicting response to selection*  (Practical, Hamilton Computer Room) |

**Readings and Resources:**

- Introduction to population Ecology. Rockwood L. 2015. Wiley Blackwell. Chichester

- Population ecology. Begon M, Mortimer M and Thomson D. J. 1996. Blackwell, Oxford.

- Evolutionary quantitative genetics. Roff D. A. 1997. Chapman & Hall, New York.

- Introduction to quantitative genetics. Falconer D S. & Mackay T. F. C. 1996. Longman, Harlow.

- Quantitative genetics in the wild. Charmantier A., Garant D., Kruuk L. E. B. 2014 Oxford University Press, Oxford.

## Generalised Linear Models

**Week:** January 16 2017–January 20 2017

**Convenor:** Julia Schroeder

This module builds on the basic linear models introduced in the previous term to introduce some key concepts that allow linear models to be applied to a wider range of research problems. This will include using generalised linear models to handle count and binomial data - where residuals are not expected to follow a normal distribution - and the use of structured models to allow for nonindependence in data and to control for known sources of variation in data.

**Aims:**

To understand, apply and interpret GLMs and GLMMs

**Timetable:**

| Monday, January 16 2017 | |
|---|---|
| 09:15 – 12:15 | *Likelihood, deviance, and AIC* (Lectures + Exercises, Hamilton Computer Room) |
| 13:30 – 16:30 | *Count data and poisson models* (Lectures + Exercises, Hamilton Computer Room) |
| **Tuesday, January 17 2017** | |
| 09:15 – 12:15 | *Proportions and the binomial distribution* (Lectures + Exercises, Hamilton Computer Room) |
| 13:30 – 16:30 | *Contrasts in linear models* (Lectures + Exercises, Hamilton Computer Room) |
| **Wednesday, January 18 2017** | |
| 13:30 – 16:30 | *Quasi-poisson models* (Lectures + Exercises, Hamilton Computer Room) |
| **Thursday, January 19 2017** | |
| 09:15 – 12:15 | *Workshop and troubleshooting* (Lectures + Exercises, Hamilton Computer Room) |
| 13:30 – 16:30 | *Variance and random effects* (Lectures + Exercises, Hamilton Computer Room) |
| **Friday, January 20 2017** | |
| 09:15 – 12:15 | *Fixed and random effects* (Lectures + Exercises, Hamilton Computer Room) |
| 13:30 – 16:30 | *Mixed effects modeling* (Lectures + Exercises, Hamilton Computer Room) |

**Readings and Resources:**

- Zuur A Ieno E Walker N Saveliev A Smith G et. al. (2009) Mixed effects models and extensions in ecology with R, Springer New York

- Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

## Maths for Biology

**Week:** January 23 2017–February 03 2017

**Convenor:** Alberto Pascual-Garcia

This is a two-week module aimed to provide a primer in core mathematical topics that will be of use in the rest of this course and indeed, hopefully, throughout students careers. You will receive an introduction to the fundamental concepts of calculus, linear and matrix algebra, probability theory, and their application to ecological and evolutionary prblems and systems. Scheduling for this module in both weeks will be (all activities in Wallace):

10:00–11:00: Lecture

11:30–12:30: Lecture

1400–1700: Practicals, except Wednesday

**Aims:**

You will learn to:

- Apply linear algebra methods to solve equations arising from biological models
- Solve and integrate ordinary differential equation models of biological systems
- Use differentiation to find maxima and inflection points of functions
- Calculate and sample from probability distribtions relevant to biology
- Solve high-dimensional biological problenms using matrix methods

**Readings and Resources:**

- Sarah P Otto and Troy Day, A Biologists Guide to Mathematical Modelling in Ecology and Evolution, Princeton University Press, 2007.
- C. Neuhauser, Calculus for Biology and Medicine, 3rd edition (2009)
- D. C. Lay, Linear Algebra and its Applications, 4th edition (2010)
- Linear Algebra, Schaum's Outlines, 4th edition (2008)
- Any of numerous basic texts in calculus, linear algebra, and probability theory. There are many of these and students prefer different ones, so try a few and choose one. many are available in the Silwood or Central libraries.

## Maximum Likelihood

**Week:** February 06 2017–February 10 2017

**Convenor:** Tin-Yu Hui

Maximum likelihood estimation (MLE) plays a key role in statistical estimation. It provides a framework to obtain the "best" set of parameters given the observation with an associated statistical model. Many statistical methods used in ecology and evolution, including most of the general and generalised linear models described in the statistics module, are consequences of maximum likelihood under specific applications.

**Aims:**

This module aims to give a formal definition to MLE, and to apply the technique to problems in ecology. By the end of this module, student will be able to understand the concepts of MLE, and implement MLE in R for standard and non-standard problems.

**Timetable:**

| Monday, February 06 2017 | |
|---|---|
| 10:00 – 11:00 | *Background probability theory* (Lecture, Wallace) |
| 11:30 – 12:30 | *Background probability theory (Contd.)* (Lecture, Wallace) |
| 14:00 – 17:00 | *Background probability theory (Contd.)* (Lecture+Practicals, Wallace) |
| Tuesday, February 07 2017 | |
| 10:00 – 11:00 | *Probability and Likelihoods* (Lecture, Wallace) |
| 11:30 – 12:30 | *Probability and Likelihoods (Contd.)* (Lecture, Wallace) |
| 14:00 – 17:00 | *Probability and Likelihoods (Contd.)* (Lecture+Practicals, Wallace) |
| Wednesday, February 08 2017 | |
| 10:00 – 11:00 | *Properties of maximum likelihood estimators* (Lecture, Wallace) |
| 11:30 – 12:30 | *Properties of maximum likelihood estimators (Contd.)* (Lecture, Wallace) |
| Thursday, February 09 2017 | |
| 10:00 – 11:00 | *Likelihood-ratio test* (Lecture, Wallace) |
| 11:30 – 12:30 | *Confidence interval estimation* (Lecture, Wallace) |
| 14:00 – 17:00 | *Likelihood-ratio test and Confidence interval estimation* (Lecture+Practicals, Wallace) |
| Friday, February 10 2017 | |
| 10:00 – 11:00 | *MLE examples* (Practicals, Wallace) |
| 11:30 – 12:30 | *MLE examples* (Practicals, Wallace) |
| 14:00 – 17:00 | *MLE examples* (Practicals, Wallace) |

**Readings and Resources:**

- Millar, R.B., Maximum Likelihood Estimation and Inference – With Examples in R, SAS and ADMB. Wiley-Blackwell, 2011.

- Casella, G. & Berger, R.L., Statistical Inference, second edition, Cengage Learning, 2001.

## Bayesian Statistics

**Week:** February 13 2017–February 17 2017

**Convenor:** Matteo Fumagalli

This module will introduce students to background theory and computational practice of using Bayesian statistics. Topics to be covered include the Bayes theorem, fitting by Bayesian methods, MCMC, and other basics of Bayesian statistics as applied to biology.

**Timetable:**

(All activities will be in Wallace)

1000–1100: Lecture

1130–1230: Lecture

1400-1700: Practicals, except Wednesday

**Readings and Resources:**

- R. Christensen et al, Bayesian Ideas and Data Analysis (Chapman & Hall/CRC Texts in Statistical Science)

## Evolutionary Modelling

**Week:** February 20 2017–February 24 2017

**Convenor:** Tim Barraclough & Austin Burt

This module will give an introduction to the classic models in population genetics that have been used to study the action and interaction of mutation, drift, migration selection, and transmission ratio distortion.The theory will be integrated with practicals involving microbial population genetics/genomics data. Scheduling for this module will be (all activities in Wallace):

10:00–11:00: Lecture

11:30–12:30: Lecture

1400–1700: Practicals, except Wednesday

**Readings and Resources:**

- TBD

# Ecological Modelling

**Week:** February 27 2017–March 03 2017

**Convenor:** Vincent Jansen

Dynamical systems theory plays a major role in modern theoretical approaches to ecological concepts and phenomena such as competition, predation, metapopulation dynamics, diversity and evolution, and disease spread. This module will introduce some of the key basics of dynamical systems theory in application to these topics. We will look at ordinary differential equations and difference equation models and will use stability analysis and bifurcation analysis as tools to understand the qualitative behaviour of ecological models.

**Aims:**

- To be able to interpret and classify the qualitative behaviour of mathematical models in ecology
- To be able to apply dynamical systems theory and bifurcation analysis to ordinary differential equation models in ecology

**Timetable:**

Scheduling for this module in both weeks will be (all activities in Wallace):

10:00–11:00: Lecture

11:30–12:30: Lecture

1400–1700: Practicals, except Wednesday

**Readings and Resources:**

Core:

- Bart Ermentrout: Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XP-PAUT for Researchers and Students. SIAM. 2002
- Nicholas Gotelli, A Primer of Ecology, Sinauer Associates, 4th Edn, 2008

Additional:

- Sarah P Otto and Troy Day, A Biologists Guide to Mathematical Modelling in Ecology and Evolution, Princeton University Press, 2007.
- Bart Ermentrout: Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XP-PAUT for Researchers and Students. SIAM. 2002
- Bifurcation analyses using Python http://www.ni.gsu.edu/~rclewley/PyDSTool/FrontPage.html
- mathcont: https://sourceforge.net/projects/matcont/
- Kuznetsov, Y.A., 2013. Elements of applied bifurcation theory (Vol. 112). Springer Science & Business Media.

## Biological Data Structures and C

**Week:** March 06 2017–March 10 2017

**Convenor:** Martin Brazeau

This module will introduce basics of procedural programming in C and applications related to biological data structures such as (phylogenetic) trees and other networks. C is a small and extremely flexible programming language, but is not for the faint of heart. In this module, you will learn basic elements of C that can be used to improve performance of computationally intensive tasks common in biological computing. We will explore low-level methods for representing (phylogenetic) trees and networks at the machine level, analyse algorithms for working with these structures, and learn how to implement basic tricks for speeding up calculations in comparative biological methods.

**Aims:**

In this module, you will learn:

- The very basics of C programming

- How to implement C subroutines within R and Python code

- Memory management techniques for larger applications

- Bitwise representations of categorical data (e.g. DNA, amino acid, trait variables)

- Storing, manipulating, and working with phylogenetic trees: (e.g. pointers and records, edge tables, Newick vs. XML formats, tree traversal operations)

- Program optimization techniques

**Timetable:**

| Monday, March 06 2017 | |
|---|---|
| 10:00 – 11:00 | *Introduction to C; the C compiler; first C program* (Lecture, Wallace) |
| 11:30 – 12:30 | *Variables, data types, and memory* (Lecture, Wallace) |
| 14:00 – 17:00 | *Arrays, strings; control flow, looping, and scope; writing first programs.* (Lecture+Practicals, Wallace) |
| **Tuesday, March 07 2017** | |
| 10:00 – 11:00 | *Functions and scope* (Lecture, Wallace) |
| 11:30 – 12:30 | *The stack, the heap, lower-level analysis of memory* (Lecture, Wallace) |
| 14:00 – 17:00 | *Advanced data types; bitwise operations* (Lecture+Practicals, Wallace) |
| **Wednesday, March 08 2017** | |
| 10:00 – 11:00 | *Pointers; pointers and advanced data types* (Lecture, Wallace) |
| 11:30 – 12:30 | *Memory management* (Lecture, Wallace) |
| **Thursday, March 09 2017** | |
| 10:00 – 11:00 | *Practical pointers: trees and linked lists with biological applications* (Lecture, Wallace) |
| 11:30 – 12:30 | *Other representations of (phylogenetic) trees and networks* (Lecture, Wallace) |
| 14:00 – 17:00 | *Packing discrete biological data into bits to speed up your code.* (Lecture+Practicals, Wallace) |
| **Friday, March 10 2017** | |
| 10:00 – 11:00 | *Headers, multiple files, working with other languages (R and Python)* (Lecture, Wallace) |

| | |
|---|---|
| 11:30 – 12:30 | *Optimising your Python/R code with C (or your knowledge of C)* (Lecture, Wallace) |
| 14:00 – 17:00 | *Safe programming practices; partial object-orientation in C* (Lecture+Practicals, Wallace) |

**Readings and Resources:**

- The C standard: Kernighan, B. W., & Ritchie, D. M. 1988. The C programming language. Prentice Hall, 2nd Ed. 274 pp.

- Useful instructional: Kochan, S. G. 2005. Programming in C. Sams Publishing, 3rd Ed. 543 pp.

- Helpful tutorials: http://www.cprogramming.com/

**Don't forget to check the Silwood Masters Student Gudebook for additional and important information and dates!**