Contents

1 Course Overview 1
  1.1 Course Administration .................................................. 2
  1.2 Course Aims ................................................................. 3
  1.3 Learning outcomes ......................................................... 3
  1.4 Transferable Skills .......................................................... 3
  1.5 Course activities and assessment overview .............................. 4
    1.5.1 Assessment Overview ............................................... 4
    1.5.2 Course structure overview ....................................... 5
    1.5.3 MSc CMEE ............................................................... 5
    1.5.4 MRes CMEE ............................................................... 5
  1.6 External vivas and examiners ............................................. 5
  1.7 Weekly Seminars and Seminar Diary .................................... 6
  1.8 Workshops ................................................................. 6
  1.9 Computing ................................................................. 6
  1.10 The Illumina Harvey Prizes .............................................. 7
  1.11 Background Reading ..................................................... 7

2 Course details, timetables, and module descriptions 8
  2.1 Teaching Staff ............................................................. 8
  2.2 Outline timetable and important dates .................................. 8
  2.3 Taught Module Descriptions ............................................... 9
    2.3.1 Course Induction & Q/CMEE Bootcamp Intro ....................... 11
    2.3.2 Q/CMEE Bootcamp: Intro to Biological Computing in Python ........ 13
    2.3.3 Q/CMEE Bootcamp: Biological Computing in R .................... 14
    2.3.4 Statistics in R + Miniproject: Intro ................................ 15
    2.3.5 Statistics in R + Spatial Analyses and Geographic Information Systems (GIS) + CMEE Miniproject: Intro, project selection and preliminary work .................................................. 16
    2.3.6 Genomics and Bioinformatics ........................................ 17
    2.3.7 Biological Computing in Python II .................................. 18
    2.3.8 CMEE Miniproject: Hackathon ...................................... 19
    2.3.9 High Performance Computing ........................................ 20
    2.3.10 Biological Data Structures and C .................................. 21
    2.3.11 Generalised Linear Models ........................................... 22
    2.3.12 Maths for Biology .................................................... 23
    2.3.13 Ecological Modelling ................................................ 24
    2.3.14 Evolutionary Modelling ............................................... 25
    2.3.15 Maximum Likelihood .................................................. 26
    2.3.16 Bayesian Statistics .................................................... 27
    2.3.17 CMEE Miniproject: Final push ...................................... 28
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 your thesis, including word limits, formatting guidelines, etc.

**Project and Supervision Guidelines**
How to choose a project, student research budgets, what to expect (and not to expect!) from supervisors.

**The FrEEC Symposium**
All about the student-run Frontiers in Ecology, Evolution and Conservation (FrEE) summer symposium.

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

Electronic copies of both of these guidebooks 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).
Welcome to the Masters programmes (MSc and MRes) 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 (increasingly often, truly “big data”). Answering important problems about issues ranging from disease dynamics and epidemiology to the effects of climate change and over-fishing on aquatic ecosystems require computationally sophisticated approaches for management & analysis of large amounts of data, building theoretical models, and fitting models to these 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. Similarly, scientists are increasingly beginning to appreciate that the ecology of, and evolution in, agricultural fields is crucial for maintaining crop yields. 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 sit in MSc lectures relevant to their research project.

Research projects are undertaken from the 10th week of the course (MRes) or from the 27th week (MSc) (see timetables in section 2.2).

*Both MSc and MRes CMEE projects must entail significant amounts of one (or a combination of) the following: mathematical theory, computational simulations and analyses (may include machine-learning approaches) of large datasets.*

Daily lectures and practicals, unless otherwise stated, commence at 10:00 and finish by 17:00, incorporating breaks. Additional independent work is also required. Wednesday afternoons are normally reserved for private study, sports and leisure activities for students.
Course Overview

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

- The Graduate-student organized Frontiers in Ecology and Evolution Symposium (FrEE!) in September (freesymp.org), which provides students with an opportunity to present their work and see what their peers have been up to.

- A Monday pizza and beer-fueled social seminar series where anybody can give a short, informal talk about their research.

Teaching materials and other course materials will be provided through a git repository (https://mhasoba.github.io/TheMulQuaBio/) and the online Blackboard virtual learning environment http://bb.imperial.ac.uk. 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. 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 or debugging the code that runs or collates data from the 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
www.cmee.co.uk

Course Administration

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

<table>
<thead>
<tr>
<th>MSc &amp; MRes CMEE Course Director</th>
<th>Dr. Samraat Pawar (ext. 42213, <a href="mailto:s.pawar@imperial.ac.uk">s.pawar@imperial.ac.uk</a>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSc &amp; MRes CMEE Course Co-Director</td>
<td>Dr. James Rosindell (ext. 42242, <a href="mailto:j.rosindell@imperial.ac.uk">j.rosindell@imperial.ac.uk</a>)</td>
</tr>
<tr>
<td>Postgraduate Administrator</td>
<td>Mrs. Amanda Ellis (ext. 42251, <a href="mailto:amanda.ellis@imperial.ac.uk">amanda.ellis@imperial.ac.uk</a>)</td>
</tr>
<tr>
<td>Postgraduate Tutor</td>
<td>Dr. Julia Schroeder (<a href="mailto:julia.schroeder@imperial.ac.uk">julia.schroeder@imperial.ac.uk</a>)</td>
</tr>
<tr>
<td>Director of Postgraduate Studies</td>
<td>Dr. Niki Gounaris (ext. 4 5209, <a href="mailto:k.gounaris@imperial.ac.uk">k.gounaris@imperial.ac.uk</a>)</td>
</tr>
<tr>
<td>Course Tutor</td>
<td>Katie Willis (<a href="mailto:katie.willis16@imperial.ac.uk">katie.willis16@imperial.ac.uk</a>) &amp; Tom Clegg (<a href="mailto:t.clegg17@imperial.ac.uk">t.clegg17@imperial.ac.uk</a>)</td>
</tr>
<tr>
<td>Course Representative</td>
<td>Up to you (see Silwood Masters Guidebook)</td>
</tr>
</tbody>
</table>

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

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 large data sets and 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 their preferred 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 course 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 graduating with a CMEE Masters Degree 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;
Course Overview

- 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.

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 (~9 months for MRes, ~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, selecting a research project, and selecting Job/PhD opportunities and applying to them.

Course activities and assessment overview

Assessment Overview

*Note: Please refer to the Silwood Masters Guidebook for the criteria for degree classifications (i.e., obtaining a Distinction, Merit, etc.)*

<table>
<thead>
<tr>
<th>Component</th>
<th>MSc CMEE % of Course</th>
<th>MSc CMEE % of Component</th>
<th>MRes CMEE % of Course</th>
<th>MRes CMEE % of Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computing</td>
<td>13.75</td>
<td>55</td>
<td>13.75</td>
<td>55</td>
</tr>
<tr>
<td>CMEE Mini-project</td>
<td>6</td>
<td>24</td>
<td>6</td>
<td>24</td>
</tr>
<tr>
<td>HPC Long Practical</td>
<td>5</td>
<td>20</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>Seminar Diary</td>
<td>0.25</td>
<td>1</td>
<td>0.25</td>
<td>1</td>
</tr>
<tr>
<td>Coursework Total</td>
<td>25</td>
<td>100</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>Exams</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam 1</td>
<td>10</td>
<td>40</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Exam 2</td>
<td>15</td>
<td>60</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Exam Total</td>
<td>25</td>
<td>100</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Project</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Report</td>
<td>30</td>
<td>60</td>
<td>45</td>
<td>60</td>
</tr>
<tr>
<td>Viva</td>
<td>12.5</td>
<td>25</td>
<td>18.75</td>
<td>25</td>
</tr>
<tr>
<td>Presentation</td>
<td>5</td>
<td>10</td>
<td>7.5</td>
<td>10</td>
</tr>
<tr>
<td>Supervisor mark</td>
<td>2.5</td>
<td>5</td>
<td>3.75</td>
<td>5</td>
</tr>
<tr>
<td>Project Total</td>
<td>50</td>
<td>100</td>
<td>75</td>
<td>100</td>
</tr>
</tbody>
</table>
Course Overview

Course structure overview

<table>
<thead>
<tr>
<th>Activity</th>
<th>MSc CME</th>
<th>MRes CME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lectures + practicals, with assessment</td>
<td>Required for 19 weeks</td>
<td>Required for first 9 weeks, optional attendance in MSc modules within reason thereafter</td>
</tr>
<tr>
<td>Exams</td>
<td>Required</td>
<td>Not required</td>
</tr>
<tr>
<td>Project report (Dissertation)</td>
<td>Required</td>
<td>Required</td>
</tr>
<tr>
<td>Seminars</td>
<td>Required, seminar diary to be submitted based on the attendance of a minimum of 12 seminars</td>
<td>Required, seminar diary to be submitted based on the attendance of a minimum of 12 seminars</td>
</tr>
<tr>
<td>Workshops</td>
<td>All optional</td>
<td>All optional</td>
</tr>
</tbody>
</table>

MSc CME

MSc CME students will attend 19 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 CME

MRes CME 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 CME 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:
Weekly Seminars and Seminar Diary

Both MSc and MREs students must try to 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 12 of these seminars and create a “seminar journal” for submission as part of the coursework mark (see section on the Coursework Element). The Seminar diary is due at the end of the Spring term (email a pdf to course director), the week before your final (project) dissertation submission (see section 2.2).

Workshops

We will organize a series of skills workshops in the induction week and then certain Wednesday afternoons, such as those involved in choosing a project, writing skills, and choosing and applying for PhD places. The dates for these will be provided through your iCalendar.

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). You will be taught to install a 64-bit Linux operating system (that’s all you should need for the course), and necessary software. 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 and Seminar Room 1 on the first floor of the Hamilton Building provide charging and network points for your laptop, some desktop computers, 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 www.pawarlab.org.

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 Silwood Masters 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 \( \frac{1}{8} \)), 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 Harvey’s experimental work and by the invention of the microscope.

Background Reading

These readings are recommended to all students for wading into quantitative thinking in ecology and evolution. Module-specific readings and resources can be found later on in this Guidebook, with additional ones being provided as part of the learning materials used in each module/week.

Course details, timetables, and module descriptions

Teaching Staff

Dr Alberto Pascual-Garcia  
Theoretical ecology and complex networks  
(a.pascual-garcia@imperial.ac.uk)

Dr Bhavin Khatri  
Theoretical ecology and population genetics  
(b.khatri@imperial.ac.uk)

Dr Lucas Dias-Fernandes  
Theoretical Ecology; Biodiversity theory; Networks  
(l.dias-fernandes@imperial.ac.uk)

Dr Matteo Fumagalli  
Population genetics, bioinformatics  
(m.fumagalli@imperial.ac.uk)

Dr Tim-Yu Hui  
Statistical genetics  
(tim-yu.hui11@imperial.ac.uk)

Dr Josh Hodge  
Quantitative Ecology and Statistics  
(j.hodge@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 David Orme  
Spatial patterns of biodiversity; Macroecology; Ecoinformatics  
(d.orme@imperial.ac.uk)

Dr Emma Cavan  
Climate change; Aquatic Ecosystems and Biogeochemical Cycles  
(e.cavan@imperial.ac.uk)

Dr Martin Brazeau  
Palaeontology; Morphology and evolution; Computational Phylogenetics  
(m.brazeau@imperial.ac.uk)

Outline timetable and important dates

<table>
<thead>
<tr>
<th>Week</th>
<th>Start date</th>
<th>MSc CMEE</th>
<th>MRes CMEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30 Sep</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>7 Oct</td>
<td>Q/CxEE Induction + Bootcamp Intro</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>14 Oct</td>
<td>Q/CxEE Bootcamp: Intro to Biological Computing in Python</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>21 Oct</td>
<td>Q/CxEE Bootcamp: Biological Computing in R</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>28 Oct</td>
<td>Statistics in R + CxEE Miniproject Start</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4 Nov</td>
<td>Statistics in R + Spatial Analyses and GIS + CxEE Miniproject Start</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>11 Nov</td>
<td>Genomics and Bioinformatics</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>18 Nov</td>
<td>Biological Computing in Python II</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>25 Nov</td>
<td>CxEE Miniproject: Hackathon</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2 Dec</td>
<td>Biological Data Structures and C</td>
<td></td>
</tr>
</tbody>
</table>

Autumn term
Course details, timetables, and module descriptions

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity/Item due</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 9 Dec</td>
<td>Reading</td>
</tr>
<tr>
<td>Spring term</td>
<td></td>
</tr>
<tr>
<td>15 7 Jan</td>
<td>Exam 1</td>
</tr>
<tr>
<td>16 13 Jan</td>
<td>Generalised Linear Modelling</td>
</tr>
<tr>
<td>17 20 Jan</td>
<td>Maths for Biology</td>
</tr>
<tr>
<td>18 27 Jan</td>
<td>Maths for Biology</td>
</tr>
<tr>
<td>19 3 Feb</td>
<td>Ecological Modelling</td>
</tr>
<tr>
<td>20 10 Feb</td>
<td>Evolutionary Modelling</td>
</tr>
<tr>
<td>21 17 Feb</td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td>22 24 Feb</td>
<td>Bayesian Statistics</td>
</tr>
<tr>
<td>23 2 Mar</td>
<td>CMEE Miniproject: Final push</td>
</tr>
<tr>
<td>24 9 Mar</td>
<td>Reading</td>
</tr>
<tr>
<td>25 16 Mar</td>
<td>Reading</td>
</tr>
<tr>
<td>26 24 Mar</td>
<td>Exam 2</td>
</tr>
</tbody>
</table>

For later weeks and other dates see below

Other important dates and deadlines:

<table>
<thead>
<tr>
<th>Date</th>
<th>Activity/Item due</th>
</tr>
</thead>
<tbody>
<tr>
<td>09 Dec, 5PM</td>
<td>MRes: Project Proposal</td>
</tr>
<tr>
<td>13 Dec, 5PM</td>
<td>MSc, MRes: HPC Long Practical</td>
</tr>
<tr>
<td>6 March, 5PM</td>
<td>MSc, MRes: CMEE Miniproject Submission</td>
</tr>
<tr>
<td>3 April, 5PM</td>
<td>MSc, Project Proposal</td>
</tr>
<tr>
<td>1 July, 5pm</td>
<td>MSc, MRes: Seminar Diary</td>
</tr>
</tbody>
</table>

Other 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 git repository into a directory called Proposal inside a Project directory. Please also email the pdf to your Course Director. Please ask the Course Director if you have any questions about this, and refer to the Silwood Student Guidebook for proposal preparation guidelines.

Workshops:

In addition, there are a number of professional skill development workshops, the dates for which will be provided through your iCalendar.

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 interactive, and will typically require you to use your laptop. So please bring you laptop to every class and session. The afternoon sessions of these modules will involve computing practicals. In most of the first nine weeks, you will be assisted by demonstrators (typically, one per 6 students).

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 details, timetables, and module descriptions

The following descriptions of the content and learning objectives of the weekly lectures typically exclude day-level timetables because these will be available through the iCalendar (AKA iCal) service; please see http://www.imperial.ac.uk/timetabling/view/icalendar. Materials for most of the computing weeks in the first term are available at https://mhasoba.github.io/TheMulQuaBio (including readings and resources for each topic).
Course details, timetables, and module descriptions

Course Induction & Q/CMEE Bootcamp Intro

Week of: September 30 2019
Convener: Samraat Pawar

This module introduces the philosophy and structure of the CMEE Masters course and the Quantitative and Modelling-skills in Ecology and Evolution Centre for Doctoral Training (QMEE CDT), and kicks off the 2.5-week Bootcamp in Biological Computing shared by these two student cohorts. Over about 2.5 weeks, the Bootcamp will cover UNIX (and Shell scripting), LaTeX, Version control (using Git), Python and R. 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 the Computing Bootcamp is to introduce the basics of computing with focus on biological applications. The Bootcamp Intro component (this week) focuses on training in fundamental tools of scientific computing: UNIX and Linux, shell scripting, \LaTeX, and Git. Throughout the 2.5 weeks of the Bootcamp, the teaching will be hands-on and informal, involving lectures interspersed with short exercises in class to reinforce concepts and techniques. You will also tackle longer exercises/problems during long computing practical sessions, some independently, and some in groups. These sessions will be demonstrator-aided and are also an opportunity for you to catch up. Along with the Bootcamp, the week includes a number of induction events. You will also be provided with your course laptop (if needed). All lectures and practicals will be computer (laptop)-based.

Aims:

- Learn to format a computer, install Linux, and set it up.
- Learn basic hardware/software concepts in computing.
- Learn to use the UNIX environment and the terminal for tasks ranging from data exploration to simple calculations and data processing.
- Learn to use \LaTeX for typesetting documents in a reproducible, consistent and elegant way.
- Learn how and why to version control computer code with git.

Timetable:

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday, September 30</td>
<td>09:00 – 10:00</td>
<td>Welcome to Silwood Park (Lecture, Fisher+Haldane)</td>
</tr>
<tr>
<td></td>
<td>10:00 – 11:00</td>
<td>Introduction to CMEE Masters and QMEE CDT + issue of computers (Presentation, CPB Common Room)</td>
</tr>
<tr>
<td></td>
<td>11:00 – 13:00</td>
<td>Silwood Treasure Hunt (Social, Silwood Field) — In teams with other new Masters and PhDs, you will scour Silwood for answers to fiendish geographically and ecologically based questions. Meet at the Hamilton building foyer and be prepared in case of bad weather.</td>
</tr>
<tr>
<td></td>
<td>13:00 – 14:00</td>
<td>Buffet Lunch (Social, Hamilton Foyer)</td>
</tr>
<tr>
<td></td>
<td>14:00 – 16:00</td>
<td>The Big Picture: Group Discussions (Discussion, Hamilton Foyer)</td>
</tr>
<tr>
<td>Tuesday, October 01</td>
<td>09:30 – 10:00</td>
<td>Computer setup (Practical, CPB Common Room)</td>
</tr>
<tr>
<td></td>
<td>10:00 – 12:30</td>
<td>Intro to UNIX and Linux (Lecture+Exercises, CPB Common Room)</td>
</tr>
<tr>
<td></td>
<td>13:30 – 17:00</td>
<td>How to get addicted to the terminal and shell scripting (Lecture+Exercises, CPB Common Room)</td>
</tr>
<tr>
<td></td>
<td>17:00 – 19:00</td>
<td>Welcome Reception (Social, CPB Common Room)</td>
</tr>
</tbody>
</table>

11
### Course details, timetables, and module descriptions

<table>
<thead>
<tr>
<th>Wednesday, October 02 2019</th>
<th>09:30 – 12:30</th>
<th>Version control with Git (Lecture+Exercises, CPB Common Room)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14:00 – 16:00</td>
<td>Scientific typesetting using \LaTeX\ (Lecture+Exercises, CPB Common Room)</td>
</tr>
<tr>
<td></td>
<td>16:00 – 17:00</td>
<td>Introduction to the Library (Induction, Hamilton Computer Lab)</td>
</tr>
<tr>
<td></td>
<td>17:00 – 19:00</td>
<td>C/QMEE mixer (Social, Refectory area)</td>
</tr>
<tr>
<td>Thursday, October 03 2019</td>
<td>09:30 – 15:30</td>
<td>Professional Skills Development Programme (Workshop, Fisher+Haldane)</td>
</tr>
<tr>
<td></td>
<td>15:45 – 17:15</td>
<td>Silwood Labs &amp; Projects Talks (CPB Common Room)</td>
</tr>
<tr>
<td></td>
<td>17:30 – 18:30</td>
<td>Silwood Labs &amp; Projects + Networking Plenary and Refreshments (CPB Common Room)</td>
</tr>
<tr>
<td>Friday, October 04 2019</td>
<td>10:00 – 11:00</td>
<td>Provost’s Welcome (Lecture, Fisher+Haldane)</td>
</tr>
<tr>
<td></td>
<td>11:00 – 12:00</td>
<td>Graduate School Induction Talks (Workshop, Fisher+Haldane/Wallace)</td>
</tr>
<tr>
<td></td>
<td>14:00 – 15:00</td>
<td>Safety Induction (Lecture, Fisher+Haldane)</td>
</tr>
<tr>
<td></td>
<td>15:00 – 15:30</td>
<td>Research Ethics (Lecture, Fisher+Haldane)</td>
</tr>
</tbody>
</table>

**Readings and Resources:**

Check the relevant chapters of [https://mhasoba.github.io/TheMulQuaBio/](https://mhasoba.github.io/TheMulQuaBio/)
Q/CMEE Bootcamp: Intro to Biological Computing in Python

Week: October 07 2019

Convener: Samraat Pawar

This is the second week of the Bootcamp in Biological Computing. It builds on the training you received in the Intro component (UNIX / Linux, shell scripting, \LaTeX, and Git). This week you will be introduced to the Python programming language. Python is a modern, easy-to-write, interpreted (semi-compiled) language that was conceived with readability of script in mind. It has a numerous of feature-rich packages that can be used for a wide variety of biological applications and analyses. Bring your laptops to all sessions.

Aims: To learn principles of computer program design and scientific computing. Specifically, you will learn

- about basics of Python as a programming language.
- about basic Python data types and structures.
- how to write clean and well-annotated Python scripts for automating computing tasks.
- to write Python functions and programs.

Timetable:

All Lectures and Practicals will be in the CPB Common Room. Timetable is available through ICL Celcat (iCal app).

Readings and Resources:

Check the relevant chapters of https://mhasoba.github.io/TheMulQuaBio
Q/CMEE Bootcamp: Biological Computing in R

Week: October 14 2019

Convener: Samraat Pawar

In this week, you will build on the skills learned in the first two weeks of the Q/CMEE Computing Bootcamp. Here, you will learn how to use R, a freely available statistical software with strong programming capabilities. R has become tremendously popular in data analytics and visualization (and not just in biology) due to several factors: (i) many packages are available to perform a wide range of exploratory, as well as statistical/mathematical analysis, (ii) it can produce beautiful visualizations/graphics, and (iii) it has a very good support for matrix-algebra (you might not know it, but you do 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 visualizations/figures. Therefore, R should become an indispensable component of your biological research workflow. Bring your laptops to all sessions.

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 and applications in certain areas (e.g., Genomics, Population biology).

Timetable:

All Lectures and Practicals will be in the CPB Common Room

Readings and Resources:

Check the relevant chapters of https://mhasoba.github.io/TheMulQuaBio
Statistics in R + Miniproject: Intro

Week: October 21 2019

Convener: Julia Schroeder & Josh Hodge (Stats) + Samraat Pawar (Miniproject)

Stats and R Component: In this module we will build upon the introduction to R you received in the Biological Computing in R week (or the Q/CMEE Bootcamp: Biological Computing in R week) and learn 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 across multiple courses and runs in two blocks A and B.

Aims:

Basic statistics for ecology and evolution, with a focus on applicability. Mostly parametric tests (descriptive statistics, t-test, ANOVA, correlations, linear models, hypothesis testing).

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:


Miniproject Component: You will be introduced to your miniproject assignment and objectives. You will also be given primers on model fitting, including Non-Linear Least Squares fitting. You will then choose your miniproject, and start working on it. You will continue working on it in subsequent weeks, with another two weeks to follow reserved for this (see your schedule). The deadline for submission is given in the key dates table. Further details, guidelines, and marking criteria can be found in the Silwood/CMEE Masters computing course notes.

Aims:

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. 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 analysis / 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
Statistics in R + Spatial Analyses and Geographic Information Systems (GIS) + CMEE Miniproject: Intro, project selection and preliminary work

Week: October 28 2019

Convener: Julia Schroeder (Stats) + Emma Cavan (GIS) + Samraat Pawar (Miniproject)

Stats and R Component: See previous week’s description

Spatial Analyses and GIS Component: 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 using R. 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. All students should bring their laptops to all sessions.

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
- Practical experience in creating maps
- Be able to perform basic data analyses and hypotheses testing in the spatial domain

Miniproject: See previous week’s description
Genomics and Bioinformatics

Week: November 04 2019
Convener: Matteo Fumagalli

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. Students should bring their laptops to all sessions.

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.
Biological Computing in Python II

Week: November 11 2018

Convener: Samraat Pawar

In this week, you will build on what you learned in the Intro Python Week. The overall aims, format, and venue of the lectures and practicals are same as same as that Python-focused Bootcamp week.

Aims:

To learn more advanced Python topics; specifically, you will learn

- Python program testing, debugging and documentation.
- to use Python for retrieving, managing, and analyzing data from local and remote databases.
- to automate file handling, string manipulation, and run shell scripts.
- to use Python for efficient numerical analyses.
- to run analyses by patching together R or R + Python scripts and functions.
CMEE Miniproject: Hackathon

**Week:** November 18 2019

**Convener:** Samraat Pawar

In this week, you will work on your miniproject assignment in a hackathon format. More details will be given in the Miniproject Intro week, and please refer to that week’s description for the Miniproject objectives and learning goals.
High Performance Computing

**Week:** November 25 2019

**Convener:** 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
Biological Data Structures and C

Week: December 02 2019

Convener: 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:

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
Generalised Linear Models

Week: January 13 2020

Convener: Julia Schroeder & Josh Hodge

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.

This week will be shared with other Masters Courses. Please bring your laptops.

Aims:

- To understand, apply and interpret GLMs and GLMMs
- Revisiting linear models
- Interactions and squared terms
- Model selection
- Poisson models
- Logistic regression
- Bi- and multivariate GL(M)Ms
- Variance-covariance analysis
- Linear mixed models
Maths for Biology

Week: January 20 & 27, 2020

Convener: Alberto Pascual-Garcia, Lucas Dias-Fernandes, & Bhavin Khatri

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 problems and systems.

Aims:

You will learn:

- Basic mathematical definitions
- to use differentiation and understand its basic applications
- to use integration methods and understand their basic applications
- to solve and integrate ordinary differential equation models of biological systems
- to apply linear algebra methods to solve equations arising from biological models
- to solve high-dimensional biological problems using matrix methods

Readings and Resources (B - Basic, A - Advanced):

- (B) C. Neuhauser, Calculus for Biology and Medicine, 3rd edition (2009)
- (B) 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.
Ecological Modelling

**Week:** February 3 2020

**Convener:** Vincent Jansen

In this course we will look at a number of ecological models, understand how they behave, and learn to interpret and classify their behaviour using dynamical systems theory and bifurcation analysis. Dynamical systems theory plays a major role in modern theoretical approaches to ecological concepts and phenomena such as competition, predation, metapopulation dynamics 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

**Readings and Resources:**

**Core:**


**Additional:**

- Bifurcation analyses using Python [http://www.ni.gsu.edu/~rclewley/PyDSTool/FrontPage.html](http://www.ni.gsu.edu/~rclewley/PyDSTool/FrontPage.html)
- mathcont: [https://sourceforge.net/projects/matcont/](https://sourceforge.net/projects/matcont/)
Evolutionary Modelling

**Week:** February 10 2020

**Convener:** Austin Burt & Tin-Yu Hui

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.

**Readings and Resources:**

Maximum Likelihood

Week: February 17 2020

Convener: 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.

Readings and Resources:

Bayesian Statistics

**Week:** February 24 2020

**Convener:** Matteo Fumagalli

In this week, you will be introduced to a wide range of Bayesian methods to solve statistical problems in the field of evolution and ecology. Topics include Bayes theorem, fitting methods, point estimation and hypothesis testing, approximated and sampling methods. Relevance to the field of evolution and ecology will be enforced with an appropriate guest lecture and/or journal club.

**Aims:** At the end of this module you will be able to:

- critically discuss advantages and disadvantages of Bayesian data analysis;
- illustrate Bayes’ Theorem and concepts of prior probability, posterior distribution and Bayes factors;
- implement Bayesian methods in R, including sampling and approximated techniques;
- apply Bayesian methods to solve problems in ecology and evolution.

**Readings and Resources:**

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

**Week:** March 02 2020

**Convener:** Samraat Pawar

In this week, you will finish your miniproject assignment, focusing mainly on the report writing. More details will be given in the Miniproject Intro week. Please refer to that week’s description for the Miniproject objectives and learning goals.
Don’t forget to check the Silwood Masters Student Guidebook for additional and important information and dates!