Maths and Computing training

for Biology Researchers



Module 1:

Learning

Objectives

SysMIC is developed by a consortium of leading UK universities that have world-class expertise in bioscience research and distance learning.



Since 2012 we have welcomed over 1,500 UK researchers who collectively have completed over 50,000 hours of training with us.

website: [sysmic.ac.uk](http://sysmic.ac.uk)

email: [sysmic-team@sysmic.ac.uk](mailto:sysmic-team@sysmic.ac.uk)

location: University College London

G03 Anatomy Building

Gower Street

London, WC1E 6BT

Introduction

SysMIC Module 1 covers the essential skills that are required to create and work with models of biological systems and analyse the related experimental data.

It involves twelve sessions which are delivered online through our website. Each session includes in-depth reading materials, video tutorials, quizzes, exercises access to help from SysMIC tutors. Typically the module takes 6 months to complete requiring around 5 hours study per week.

This booklet lists the Learning Objectives covered by the sessions in SysMIC Module 1.

# Intro to MATLAB

## Biological contexts:

Fibonacci population growth model

## Mathematical topics:

Arrays; Sequences; Functions

## Summary:

This session provides an introduction to computer programming with MATLAB using it to investigate the Fibonacci series.

## Learning objectives:

* understand how to work with pieces of data in MATLAB and the meanings of variables, arrays and types
* understand how to use a script (code file) to run multiple commands
* be able to use MATLAB to carry out mathematical calculations
* be able to use "for loops"
* understand how to create plots of simple functions

# Networks I

## Biological contexts:

Food webs; relationships between species

## Mathematical topics:

Networks/Graphs

## Summary:

This session introduces the theory and terminology used in the study of biological networks aka “graphs", and shows how to create, display and manipulate them using the MATLAB Bioinformatics toolbox.

## Learning objectives:

* understand the meanings of graph, node, vertex, edge, arc, degree
* understand how to represent a network as an adjacency matrix
* be able to create a graph in MATLAB by entering its adjacency matrix
* be able to draw, edit and work with graphs using the MATLAB biograph function

# Networks II

## Biological contexts:

Glycolysis

## Mathematical topics:

Analysing networks; bipartite graphs

## Summary:

This session builds on the previous one. It introduces basic quantifications and illustrations of network properties. Bipartite graphs are discussed as a means of representing biochemical reactions. We demonstrate how network information can be found and retrieved from an online database.

## Learning objectives:

* Understand the degree and clustering coefficient as quantification of network properties
* Be able to set up network with specified properties
* Learn to write scripts to deal with network visualisation
* Access and download network information from the Reactome database
* Import downloaded network information into MATLAB and work on it

# Vectors and Matrices

## Biological contexts:

Plasmid replication; Protein synthesis

## Mathematical topics:

Vectors and matrices; linear algebra; Petri nets

## Summary:

In this session we cover the basics of working with vectors and matrices in MATLAB, and introduce the use of Petri nets as a means to analyse network properties of some simple biological models.

## Learning objectives:

* understand the meaning of matrix, row vector, column vector and how to work with them in MATLAB
* understand the principles of matrix operations and be able to carry them out in MATLAB, and the concepts of transpose, null, identity and inverse matrices
* be able to interpret a Petri net diagram in terms of the set of reactions it describes
* be able to construct a stoichiometry matrix from a given of a set of reactions or Petri net diagram
* understand how to use a Petri net to find the final state of a system, given initial conditions and a set of reactions
* understand how to use a stoichiometry matrix to determine conservation laws associated with the system (T and P invariants)

# Functions and Calculus I

## Biological contexts:

Bacterial population models

## Mathematical topics:

Iterative methods, exponential functions

## Summary:

In this session we examine population models for the evolution of a bacterial colony. We look at discrete and continuous changes of the population size using exponential growth/decay and logistic growth models.

## Learning objectives:

* understand that mathematical functions can be used to model the dynamics of biological systems
* understand how to work with iterative formula in MATLAB
* be able to solve simple problems working with iterative formula (e.g. how many steps needed to reach given condition)
* understand how to work with mathematical functions in MATLAB (creating them, defining parameter values, plotting them)
* be able to use MATLAB to solve symbolic equations (will be further developed in next session)
* understand how exponential and logistic growth functions arise in population models

# Functions and Calculus II

## Biological contexts:

Enzymatic reaction kinetics (Michaelis-Menten rate law and the Hill function); bacterial populations

## Mathematical topics:

Rate equations, differentiation, investigating functions, integration

## Summary:

In this session we examine the Michaelis-Menten and Hill equations used to model enzymatic reaction kinetics. In particular we use plots to investigate their properties, and look how to investigate them analytically using MATLAB.

## Learning Objectives:

* understand that differentiation gives the gradient or rate of change of a process
* understand the key properties of the Michaelis-Menten and Hill function models for enzymatic reaction kinetics, and be able to define their associated functions in MATLAB
* be able to use MATLAB to create an overlaid plot that illustrates the relationship between a function and its parameters
* be able to use MATLAB work with symbolic expressions including: simplifying; differentiating; solving analytically e.g. to find maxima and minima; conversion into MATLAB functions; and evaluation using numerical substitution
* be able to use MATLAB to integrate symbolic expressions using the int function
* be able to use MATLAB’s dsolve function to find the behaviour that associated with a simple rate equations

# Functions and Calculus III

## Biological contexts:

Models of gene expression and genetic regulatory networks (the toggle switch)

## Mathematical topics:

Systems of rate equations; numerical solution of differential equations.

## Summary:

In this session we examine models of gene expression and how to simulate them in MATLAB. We then look at how to take rate equations and numerically simulate systems that cannot be solved analytically, and examine the genetic toggle switch as an example of a bistable system.

## Learning Objectives:

* understand how functions that can be used to model gene expression in the presence of activators and repressors, and how to modify these to include cooperativity
* understand how to model degradation processes
* use dsolve to symbolically solve systems of rate equations
* be able to produce overlay plots showing modelled system behaviour for a range of input parameter values
* understand that numerical simulation can be used to determine qualitative and quantitative system behaviour when rate equations can not be solved as symbolic equations
* understand how to use MATLAB’s ode solver to numerically solve systems of rate equations
* be able create plots using the output from MATLAB’s ode solver to illustrate system behaviour

# Building Models

## Biological contexts:

Modelling of erythropoietin (EPO) receptor signalling; modelling blood glucose level response to glucagon

## Mathematical topics:

Differential equations, Petri nets, negative feedback loops

## Summary:

In this session we look at the role of model building in the scientific process. We then consider how dynamic models can be developed and simulated starting from a Petri net representation. Finally we look at how to model observed behaviour when we do not have exact information about the reactions taking place. Starting from initial biological knowledge we show how a model can be developed using assumptions of mass action or enzymatic reaction kinetics, and then

investigated and improved.

## Learning objectives:

* understand the role modelling can play in the scientific process
* understand how mass action rate equations can be used to model reactions
* be able to translate a Petri net model into a system of differential equations that can be modelled in MATLAB (assuming mass action kinetics)
* be able to produce a MATLAB function file containing a set of species rate equations when provided with a proposed set of reactions and information on their rates
* be able to run simulations of a model, interpret their outcome and compare the outcome to experimental recordings

# Introduction to R

## Biological contexts:

Allometry; working with experimental data

## Mathematical topics:

Statistics (mean, standard deviation, normal distribution, plotting data, linear fits)

## Summary:

In this session we first run through the installation of the R statistics package. Then we use the R coding language for data filtering, various types of analysis, and plotting of results.

## Learning objectives:

* Be able to work with pieces of data in variables, vectors, matrices and lists, and understand data can be stored as numeric, character, or logical types
* be able to perform simple mathematical operations in R ( + - / \* exp log sqrt pow )
* be able to enter a data and calculate its associated statistics (mean, standard deviation, sum, median)
* be able to create a pseudo-random set of numbers from a given normal distribution
* be able to work with data tables using R data frames: creating data frames, selecting rows/columns, creating sub frames, filtering using a condition or combination of conditions using OR and AND operators
* be able to load data from a text file into a data frame, and save a data frame into a text file
* be able to plot histograms and scatterplots of a given data set be able to apply a straight line fit to a scatterplot

# Statistics I

## Biological contexts:

Quantitative fluorescent microscopy using GFP cells

## Mathematical topics:

Sampling and the central limit theorem

## Summary:

In this session we show how modelling a procedure can help when designing an experiment using a hypothetical example based on measurement of cell fluorescence.

## Learning objectives:

* be able to generate test data by using R to create pseudo-random datasets
* be able to use R to sample a dataset
* be able to write loops and functions in R
* understand the errors that are associated with sampling measurements, and how to estimate them using the central limit theorem
* get accustomed to plotting and interpreting results in R

# Statistics II

## Biological contexts:

Gene expression measurements using DNA microarrays

## Mathematical topics:

Hypothesis tests (Students t-test and p values, Bonferroni correction)

## Summary:

In this session we look at DNA microarray data from an experiment investigating the effect of cadmium toxicity on *C. elegans*, and show how to analyse the data using R and select up and down regulated genes using the Students t-test with an appropriate p value to minimise false positives.

## Learning objectives:

* be able to load DNA microarray data from the GEO database into R
* be able to plot histograms and density plots of data
* be able to use a quantile-quantile plot to test whether a dataset is approximately normally distributed
* know how to run a t-test on data using R to test a null hypothesis and interpret the resulting p value
* understand the principles of identifying significant up/down regulation in microarray experiments, including accounting for multiple tests

# Miniprojects

## Biological contexts:

Predator prey interactions; models of neural firing; models of allosteric enzyme regulation; the SIR model of infectious diseases; the lactose operon dynamics

## Mathematical topics:

Investigation and analysis of mathematical models

## Summary:

In this session students take a simple model of a biological system, and work through a series of steps to investigate the model using MATLAB. Students then write a report that puts the model into the biological context and includes a critical discussion of their findings.

## Learning objectives:

* be able to code a simple model of a biological system given the rate equations that describe it
* learn to describe the components of a model and give justifications for their inclusion
* be able to investigate the effect of changing model parameters and produce plots that illustrate the outcome of numerical simulations
* be able to critically evaluate a model in terms of its assumptions, limitations and successes in the biological context



Find out more about our training modules at:

sysmic.ac.uk