

Math and Biology in Trinity University's New Scientific Computing Minor



TRINITY UNIVERSITY
SAN ANTONIO, TEXAS 78269

Nataša Macura¹ and Mark Brodl²

¹Department of Mathematics and ²Department of Biology, Trinity University, One Trinity Place, San Antonio, TX 78212

Abstract

Educating 21st-century science students so that they can easily traverse traditional biology-math-computer science disciplines presents significant challenges. For both faculty and students in biology, quantitative skills are often insufficient, while for both faculty and students in mathematics (and computer science) biological understanding and insights can be limiting. Trinity University's 2004 HHMI grant has helped us to address this issue through building a minor in scientific computing that features three new math and computer science classes with attached investigative laboratory sections. The labs provide both a training ground and a data source. The lectures introduce students to mathematical concepts and build critical quantitative skills. The courses are staffed by collaborating faculty in math/computer science and biology. Students begin the minor with a basic computer science course and basic calculus. They then enroll in CSCI 2121 – Introduction to Scientific Computing, followed by MATH 1310 – Mathematical Models in the Life Sciences and MATH 3311 – Probabilistic Models in the Life Sciences. The minor culminates with a one-credit, student-generated research project that uses modeling tools. This project is done in conjunction with an upper-level biology course that the student selects from a menu of options, with the biologist teaching the course and a math faculty member serving as co-advisers. Syllabi for the courses are presented along with some details of the MATH 3311 course.

MATH 3310 – Mathematical Models in the Life Sciences

Course Description: The course is designed to introduce basic tools to study mathematical models in life sciences including their practical applications. The focus will be on understanding the processes, implications and results of modeling phenomena in life sciences in the laboratory setting. The course investigates exponential growth and logistic models, cooperative, competitive, and predator-prey models, harvesting models, and epidemiological models. The integrated laboratory experience consists of several experiments on model organisms such as bacteria, protists, and duckweed. In addition to laboratory experiments, human epidemiological data will be utilized to study epidemic models.

Module 1: Discrete models for a single species.

- Laboratory experiments: duckweed, protists, and bacteria
- Simple models including exponential growth and logistic models.
- Graphical analysis.
- Introduction to curve fitting.

Module 2: Age-structured models

- Linear systems
- Stability analysis
- Laboratory experiments: plants

Module 3: Competition models of two species (4 weeks).

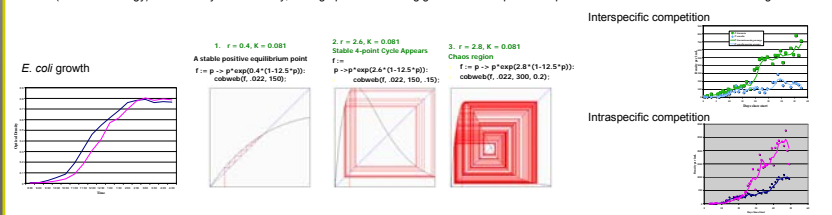
- Laboratory experiments include all of the above mentioned experiments. In addition, human epidemiological data are used to enhance understanding of epidemic models.
- Phase space, stability and linearization analysis.
- Laboratory experiments include bacteria on model organisms, protists.

Approach within modules (with Module 1 as example)

Introduction to the problem and the principal mathematical models employed (logistical, Ricker model, Beverton-Holt model). Graphical Analysis of the Models: Time Series and Cob-Web Diagrams Conducting in-class laboratory experiments for data Plotting the raw data and parameter estimation Comparing the data from the laboratory and the data obtained from the mathematical models Modifying the models

Analyses and student data from the teaching lab

Students (math and biology) meet weekly for laboratory, setting up and monitoring growth and competition experiments. The data are used in modelling.



Trinity University's Minor in Scientific Computing

SCIENTIFIC COMPUTING

SCIENTIFIC COMPUTING

FACULTY ADVISORY COMMITTEE:

FARZAN AHMAD, Ph.D., Professor, Engineering Science
SARIS BLAVIS, Ph.D., Professor, Mathematics
JULIO ROBERTO HARSFURA-BUENAGA, Ph.D., Associate Professor, Mathematics
MARK LEWIS, Ph.D., Assistant Professor, Computer Science
KEVIN LIVINGSTONE, Ph.D., Assistant Professor, Biology
NATASA MACURA, Ph.D., Assistant Professor, Mathematics
DAVID O. RISBLE, Ph.D., Professor, Biology
LISE F. SCHEITING, Ph.D., Assistant Professor, Psychology
DANIEL W. LOCKMAN, Ph.D., Assistant Professor, Physics
ADAM URSBACH, Ph.D., Assistant Professor, Chemistry

THE MINOR

Scientific Computing is a multi-disciplinary program designed for science majors who wish to explore the application of computing within the natural sciences. It is intended primarily for students majoring in one of the natural sciences or engineering. The program of study brings together courses from Computer Science, Mathematics, and disciplines in the natural sciences. Each student takes the required core courses in computer science and calculus. After the core, two advanced-level courses selected from five options further hone quantitative skills needed for scientific computing. Students should work with their advisors to identify courses that are most appropriate for their backgrounds and goals. Finally students bring their computation skills to their minor by selecting one upper-level course from a list of approved disciplinary courses that make use of computation. Concurrent with this course students enroll in SCOM 3199 – Scientific Computing Project. Note: these disciplinary courses have several prerequisites that are not listed among the requirements for a minor in Scientific Computing – students majoring in a discipline among the natural sciences will have already fulfilled these prerequisites as a part of their major.

Students interested in a Scientific Computing minor will submit an application to the Chair of the Committee, who will assign a faculty advisor to the student. Completion of the program will be indicated on the student's transcript with the notation "Minor in Scientific Computing."

A minor in Scientific Computing will consist of a total of 19 to 20 semester hours, depending on the upper-level course selected in the major. The minor must include at least one hour of upper-division courses in mathematics or science. The requirements for a minor in scientific computing are as follows:

SCIENTIFIC COMPUTING

SCIENTIFIC COMPUTING

A. The Core (9 hours)

• CSCI 1320 Principles of Algorithm Design I
• CSCI 2323 Scientific Computing
Calculus
• MATH 1307 Calculus I for the Life and Social Sciences
OR
• MATH 1311 Calculus I

B. Advanced Quantitative Skills (6 hours)

Take two of the following five courses

- CSCI 3352 Simulation
- MATH 2310 Mathematical Models in Life Sciences
- MATH 2311 Probabilistic Models in Life Sciences
- MATH 3320 Probability and Statistics for Engineers and Scientists
- MATH 3338 Mathematical Modeling

C. Disciplinary Practice (4.5 hours depending on the chosen elective)

• SCOM 3199 Scientific Computing Project
(Must be taken in conjunction with one of the courses listed below)

And one of the following courses:

- BIOL 3434 Ecology
- BIOL 3440 Animal Behavior
- BIOL 3444 Molecular Biology
- CHEM 3334 Physical Chemistry (includes lab CHEM 3135)
- CHEM 3420 Analytical Chemistry
- MATH 4364 Senior Project
- PHYS 3321 Statistical Physics and Thermodynamics
- PHYS 3322 Classical Mechanics and Nonlinear Dynamics
- PHYS 3325 Optical Physics
- PHYS 3336 Advanced Theoretical Physics
- PHYS 4343 Quantum Physics II
- PHYS 4345 Advanced Modern Physics
- PSYC 3311 Sensation and Perception
- PSYC 3331 Memory and Cognition
- PSYC 3333 Simulation of Neural and Cognitive Processes

*This course requires completion of a 3000-level course in Biology or Psychology

SCOM 3199 Scientific Computing Project

The course involves undertaking a project that expands one of the laboratory or classroom exercises to make significant use of computers as a research tool. Students must work with a faculty mentor to develop an appropriate project. Usually the instructor in the course.

Co-requisite: Enrollment in this course requires concurrent enrollment in one other disciplinary practice courses approved for section C – Disciplinary Practice of the requirements for a minor in Scientific Computing.

MATH 3311 – Probabilistic Models in the Life Sciences

Topics covered

- Basic concepts of probability, with an emphasis on conditional probability and its applications in probabilistic modeling, especially in biology and medicine.
- Discrete and continuous random variables along with expectation, variance, and standard deviation of random variables. A number of distribution functions that appear often in applications is carefully studied.
- Markov models, with emphasis on applications of discrete Markov chains in biology, especially in biological sequence analysis.
- Probabilistic models used in biological sequence comparison, are thoroughly studied, combining biological data, mathematical models and algorithm design. DNA and protein sequences from public databases are used to derive and evaluate probabilistic models used in pairwise sequence comparison. Subtopics include global and local sequence alignments, scoring matrices, alignment algorithms, and evolutionary distances.
- Other possible mathematical topics include Poisson process, branching process, and diffusion.

What makes this course different from a standard applied probability courses?

- Throughout the course the emphasis is on comprehending the biological examples and the mathematical concepts introduced are motivated by such examples and ideas.
- Models are applied to problems in biology and medicine or life sciences in general.
- The knowledge and understanding of the models and tools are used to evaluate the modeling results and interpret biological significance.
- The course includes computational labs. During the labs students use tools based on modeling approaches and mathematical techniques learned in the class to search through databases and compare biological sequences. They learn about a variety of tools developed in the field as well as about their strengths and disadvantages.
- Prerequisites and approach to new material throughout the course is adjusted to the existing knowledge of a typical student. For example, when introducing random variables, strong emphasis is put on discrete random variables since students' integration skills are limited to those gained during the first semester of Calculus series.

Probabilistic modeling in life sciences: discrete random variables through examples from biology

The negative binomial model

The employees of a firm that manufactures insulation are being tested for indications of asbestos in their lungs. The firm is asked to send three employees who have positive indications of asbestos on to a medical center for further testing.

a) If 40% of the employees have positive indications of asbestos in their lungs, find the probability that ten employees must be tested to find three positives.

b) If each test costs \$20, find the expected value and the variance of the total cost of conducting tests to locate three positives.

The hyper-geometric model

Consider a gene composed of 5 subunits. Consider a cell with a gene composed of m , mS , mutant subunits and $5-m$ normal subunits. Before the cell divides into two daughter cells, the gene duplicates. The corresponding gene of one of the daughter cells is composed of S units chosen at random from the $2m$ mutant subunits and the $2(5-m)$ normal subunits. Suppose that we follow a fixed line of descent from a given gene. Let X_0 be the number of mutant subunits initially present and let X_n be the number present in the n -th descendant gene. Explain why X_n is a Markov chain, list its states and find the transition probabilities.

The binomial random variable

In testing the lethal concentration of a chemical found in polluted water, researchers have determined that a certain concentration will kill 20% of the fish that are subjected to it for 24 h. If 20 fish are placed for 24 h in the tank containing this concentration of the chemical, find the probability that at least 10 will survive. Find the number of fish expected to die.

Lab 1

- Goal: get familiar with biosequence databases, understand what sequence database accessions look like (FASTA formats, etc.) and how to query databases.
- First, go through [NCBI](#) (What does NCBI stand for and where is it?)
- [Entrez tutorial](#) Make sure that you understand some terminology, like, for example: database (what kind of databases do you know about), "tools," search engines, gene, annotated sequence, FASTA format. Write a list of terms that you did not understand. When you are done with the tutorial, take a break at [NCBI](#) coffee break. Then visit the [BLAST Tutorial](#). Summarize what you understood about BLAST and what you did not understand. In which flavors does BLAST come? Take a quick look at the link below, but do not spend much time studying it. We will talk in a great detail similarity matrices and alignment scores.
- [BLAST](#) Look at the page below, visit some of the links and make sure that you understand what they are; if a database, what kind of database is it, if a sequence comparison or analysis server, what's happening there?
- [Bioinformatics links](#)

Lab 2

- Here are some links to information on myosin, a protein in muscles. Read the info about the gene and retrieve the identifiers and the sequence.
- [Wiki on myosin](#)
- [Myosin](#) Find the top hits for this gene in several different species and then use i-clone or ClustalW to do a multiple alignment.
- [Clustalw: multiple sequence alignments](#)
- [i-clone server](#) Write a summary of your findings, including the biological implications and print outs of your work.

Lab 3

- Align and compare the myosin gene in different species that you found in the previous lab. Add genes from more distant species if you already haven't! Do local and global alignments and compare the results. Use different scoring matrices (PAM and BLOSUM with different numbers) and gap penalties and compare the outputs. Use different alignment tools and compare the outputs. Write a summary of your findings, including the biological implications and print outs of your work. Below are links to the two sequence comparison servers which are based on slower algorithm which guarantees to find an optimal alignment.
- [Needle: Needleman-Wunsch global alignment](#)
- [Water: Smith-Waterman local alignment](#)
- [Pairwise Alignment and Database Searching](#)
- [LAGAN/PLALIGN find internal duplications by calculating non-intersecting local alignments of protein or DNA](#)
- [Clustalw: multiple sequence alignments](#)
- [i-clone server](#)

Students taking the course include students in scientific computation minor as well as biology, neuroscience, chemistry and math majors.

Examples positive comments from course evaluations

- Teaches you to think mathematically.
- With a little more bio or neuroscience, this course could be one of the most interesting at Trinity.
- Great course! I loved the applications to life sciences!
- Very helpful for math-majors that want to continue taking probability classes. Also helpful for students in natural sciences.
- Because of this course I gained a lot of probability knowledge and strong foundations.
- If you enjoy math and you love natural sciences you should take this course.

Only one concern expressed on course evaluations:

- Students who came with the only the required prerequisite (1st semester Calculus) felt the course should not be available to students taking more advanced probability courses. They found it unfair that math majors in the class dominated the curve.