An Undergraduate Curriculum in Computational and Data Sciences

John Wallin
Undergraduate Coordinator
Department of Computational and Data Sciences
George Mason University

November 15, 2009
Computational Science at Mason

The Data Sciences

The Curriculum

CDS 101

CDS 130

The Mentorship Program

---

Theory

Experiment

Computational Sciences

---
Arctic Ice Cap
Data Sciences

Satellite Data - 1979-2003 SSMI Composite Data
See http://www.everybodysweather.com/Static_Media/Polar_Ice_Cap_Melter/index.htm
The Problem

Data rates for future missions will overwhelm Scientists!

- The Sloan Digital Sky Survey Atlas Images are 1.5 TB
  - The 150,000+ Citizen Scientists in GalaxyZoo have classified the galaxies from these images in 18 months
- The LSST project will generate 30 Terabytes of data per day!
  - This is about 5000 DVD’s per day!
- We need to develop new ways to deal with data.
Computational and Data Sciences at Mason
A Comprehensive Curriculum

- Ph.D. program
- Masters program
- Undergraduate major and minor
- Training for science majors
- Courses for non-science majors
- A structured mentorship program for undergradautes
Undergraduate Major and Minor
A Curriculum in the Computational and Data Sciences

- Mathematics (calculus, ODE, finite math, linear algebra, numerical methods)
Undergraduate Major and Minor
A Curriculum in the Computational and Data Sciences

- Mathematics (calculus, ODE, finite math, linear algebra, numerical methods)
- Computer Science (programming, computer systems, algorithms)
Undergraduate Major and Minor
A Curriculum in the Computational and Data Sciences

- Mathematics (calculus, ODE, finite math, linear algebra, numerical methods)
- Computer Science (programming, computer systems, algorithms)
- Core Sciences (concentrations in physics, chemistry, biology)
Undergraduate Major and Minor
A Curriculum in the Computational and Data Sciences

- Mathematics (calculus, ODE, finite math, linear algebra, numerical methods)
- Computer Science (programming, computer systems, algorithms)
- Core Sciences (concentrations in physics, chemistry, biology)
- Synthesis Courses
  - Computational Science (introduction, simulations)
  - Data Sciences (introduction, statistics, visualization, databases, data mining)
Computational Science- Modeling Courses

- **CDS 410 - Modeling and Simulation I**
  Numerical differentiation and integration, initial-value and boundary-value problems for ordinary differential equations, methods of solution of partial differential equations, iterative methods of solution of nonlinear systems, and approximation theory.

- **CDS 411 - Modeling and Simulation II**
  This course covers the application of modeling and simulation methods to various scientific applications, including fluid dynamics, solid mechanics, materials science, molecular mechanics, and astrophysics. It will also provide an introduction to modeling and simulation software, as well as high-performance computing.
Data Sciences- Data Courses

- CDS 301 - Scientific Information and Data Visualization
  The techniques and software used to visualize scientific simulations, complex information, and data visualization for knowledge discovery.

- CDS 302 - Scientific Data and Databases
  Data and databases used by scientists. Includes basics about database organization, queries, and distributed data systems. Student exercises will include queries of existing systems, along with basic design of simple database systems.

- CDS 401 - Scientific Data Mining
  Data mining techniques from statistics, machine learning, and visualization to scientific knowledge discovery.
CDS 101
Introduction to the Computational and Data Sciences

Course Description
“Introduction to the use of computers in scientific discovery through simulations and data analysis. Covers historical development and current trends in the field.”
http://cds101.org

- serves as a General Education Natural Science Elective
- designed for non-science majors and our majors
- assumes mathematical background of high school algebra
CDS 101
Introduction to the Computational and Data Sciences - Objectives

Students will be able to

- Understand and use the scientific method
- Use computers to investigate simple scientific problems
- Describe how data is acquired, processed, analyzed, and visualized in a variety of scientific domains
- Explain how simulations are used across the natural sciences, and understand their limitations
- Describe the connections between advances in computing and advances in the natural sciences
CDS 101
Introduction to the Computational and Data Sciences
CDS 101
Introduction to the Computational and Data Sciences - Schedule

- Week 1: The Scientific Method
- Week 2: Computer Internals and Scientific Applications
- Week 3: Computer Algorithms and Tools
- Week 4: Data acquisition
- Week 5: Signal Processing
- Week 6: Scientific Databases
- Week 7: Data Reduction and Analysis - from data to knowledge
- Week 8: Data Mining
- Week 9: Computer Models
- Week 10: Computer Simulations - solving linear systems
- Week 11: Computer Simulations - applications
- Week 12: Computer Visualization - seeing experiments as images
- Week 13-14: High Performance Computing and Future Directions
Course Description

“In this course, students will learn how to use computers to solve practical scientific problems. Topics will include creating effective scientific presentations, analysis of experimental data, on-line literature, data/information ethics, scientific modeling, and communication/collaboration tools. Beyond just teaching tools, this course will equip students with the knowledge and confidence they need to use future hardware in software both as students and throughout their career.”

- serves as a General Education Information Technology Elective
- designed for science majors in their first semester
- assumes mathematical background of high school algebra
CDS 130
Computing for Scientists - Objectives

Students will be able to

- Use computational tools to develop effective scientific presentations
- Use computers in the experimental sciences
- Become effective, skeptical, ethical consumers of information
- Build computational models that solve scientific problems
- Effectively use on-line tool to collaborate on projects
- Be equipped to learn new computational tools that will be used in the sciences
CDS 130
Computing for Scientists - Schedule

- Weeks 1-2: How Computers Work
- Week 3: Measuring data in the laboratory
- Week 4: Basic Data Analysis
- Week 5: Graphics and Visualization
- Week 6: Basic modeling of data
- Weeks 7-8: On-line information systems
- Week 9: Data Ethics
- Weeks 10-11: Scientific simulation
- Week 12: Effective scientific publications and presentations
- Week 13: Collaborative tools
- Week 14: The future of scientific computing
URCM
Undergraduate Research in Computational Mathematics Program

- NSF funded year-long mentoring program for Math students
- Student work in groups of two with a faculty member
- Summer course + laboratory combined with a weekly seminar
- Students receive a generous stipend for participation and present their work at national and regional conferences
- 3rd year of this program - about 7 students per year
Projected Impact

We anticipate that

- 5-10 students per year will Major or Minor in CDS
  - The primary issue is recruiting
Projected Impact

We anticipate that

- 5-10 students per year will Major or Minor in CDS
  - The primary issue is recruiting
- 10-20 undergraduate students per year in our CDS mentorship program
  - Most funded from by individual PI grants
Projected Impact

We anticipate that

- 5-10 students per year will Major or Minor in CDS
  - The primary issue is recruiting
- 10-20 undergraduate students per year in our CDS mentorship program
  - Most funded from by individual PI grants
- 50-100 students per year will take CDS 101, most of them will be non-science majors
  - An companion lab course is being added next semester
  - Northern Virginia Community College has plans to offer this course as well
Projected Impact

We anticipate that

- 5-10 students per year will Major or Minor in CDS
  - The primary issue is recruiting
- 10-20 undergraduate students per year in our CDS mentorship program
  - Most funded from by individual PI grants
- 50-100 students per year will take CDS 101, most of them will be non-science majors
  - An companion lab course is being added next semester
  - Northern Virginia Community College has plans to offer this course as well
- 200-300 science majors per year will take CDS 130
  - about 80% of our science majors
  - many will take additional CDS course, some will change majors or minors
Summary

- George Mason has developed a comprehensive curriculum in the Computational and Data Sciences
  - Freshman non-majors through advanced graduate students
- Our goal is to move interdisciplinary Computational Science fully into the main stream of the university at all levels
Collaborators

- Dr. Kirk Borne
- Dr. Igor Griva
- Dr. Dimitris Papaconstantopoulus
- Dr. Tim Sauer
- Dr. Robert Weigel
- Dr. Jie Zhang

We gratefully acknowledge the NSF for their support:

- NSF DMS #107716 “A Curriculum for and Undergraduate Program in the Data Sciences :CUPIDS”
- NSF DMS #0639300 “Undergraduate Research in Computational Mathematics”