ENHANCING QUANTITATIVE TRAINING IN SOLID EARTH GEOPHYSICS

DR. CRAIG PARTRIDGE

CHAIR, COMPUTER SCIENCE, COLORADO STATE UNIVERSITY

MEMBER, COMPUTER SCIENCE AND TELECOMMUNICATIONS BOARD, NASEM

WHY WE ARE HERE

We are not solving the problem: rather we are scoping it

…a scoping meeting for the development of a quantitative training framework in solid Earth geophysics. A quantitatively-literate workforce is needed to investigate the multi-scale, multi-physics solid Earth science processes at the core of many scientific and societally-relevant problems. However, such skills are in short supply in some Earth science disciplines. While there exist training networks in academia, national laboratories, and industry, the education and career pathways that lead to skilled quantitative workers and researchers are not robust nor coordinated, and individual academic departments are generally ill-equipped to tackle these challenges on their own.

The meeting objective is to develop a new NASEM activity to create a community-driven framework of programming, data analysis, mathematics, statistics, potential fields, continuum mechanics, and computational modules for solid Earth geophysics. The framework could then be refined through discipline-specific applications (e.g., geodesy).

WHY A SCOPING MEETING?

- To determine if there's a potential path to success
- Giving a committee an impossible assignment is not the goal
- And integrating quantitative methods needed for success with geophysics is **not** easy

AN EXTENDED THOUGHT EXERCISE: ENHANCING QUANTITATIVE METHODS IN A GEOPHYSICS MASTERS DEGREE

START WITH A GEOPHYSICS MASTERS PROGRAM…

- 24-30 credits of coursework
  - 8 to 10 classes (3 credits, with occasional 4 credit course)
- Has core component of numerical methods and signal processing
  - Students lacking differential equations and two semesters calculus typically make up w/o credit
  - Note that often those who do have this math background took courses 4 years prior and a refresher needed

"TYPICAL" GEOPHYSICS MS CURRICULUM

Required Courses	Geophysics Courses (typically 3-4 courses)	Non-Geophysics Distribution Requirements (typically 2 courses)	Quantitative Courses (2–3 courses)
Scientific Writing / Presentation Skills	Seismology / Seismic Methods	Plate Tectonics	Inverse Theory
Graduate Seminar	Solid Earth Geophysics	Geomorphology	Signal Processing
	Field / Environmental Geophysics	Geochemistry	Numerical Methods
	Potential Field Theory	Climate Change	Scientific Programming
	GIS / Remote Sensing	Oceanography/ Atmospheric Sciences	*Data Science*

This typically results in 8-10 classes taken over 4 semesters + thesis research credits.

Additional Prerequisites: Calculus I & II + Differential Equations; 2 semesters of college physics; 2 semesters of college chemistry

GOAL: BE ABLE TO DO ANALYSIS ON A DATA SET COLLECTED AND VALIDATED BY OTHERS

Courses Needed

Statistics 300 (Theory and Practice – regression, ANOVA, Bayes)

Programming In Python (nonmajors)

Skills acquired: basic quantitative skills and the ability to program data-specific analysis in Python (and probably R). Students can start doing useful analysis after 1st semester

6 new credits

---

18 Geophysics course credits left (out of 24)

GOAL: COLLECT YOUR OWN DATA AND/OR COMBINE DATA SETS WITH THOSE OF OTHERS

Courses Added

Data Sci: Data Wrangling, Data Quality

Ethics of Data Collection?

Skills acquired: Ability to determine the suitability of data sets for combination, ability to discover and remove bad or anomalous data, understanding of ethical issues related to data collection

13 credits (+7 from last slide)

---

11 Geophysics course credits left

GOAL: ABILITY TO USE MACHINE LEARNING ON THE DATA

Courses Added

Linear Algebra

Programming for Majors (2 semesters)

Data Structures

Machine Learning Foundations and Practical Applications

Skills acquired: Sturdy programming and data structures skills, combined with linear algebra skills, all required to use basic machine learning algorithms appropriately

28 credits (+16 from last slide)

-4 Geophysics credits

OOPS!

TWO TAKEAWAYS

1. If you enhance quantitative methods naively, you overstuff the intellectual knapsack

2. Less obvious, but embedded in this exercise
   - Much of that additional quantitative coursework is real
   - There is not an intense summer bootcamp or 4 credit quant course that will somehow teach only the "good stuff" that solves the problem

OTHER ISSUES

- The departments teaching the data science courses may not be able to open them to geophysics students
  - Due to demand, access to the data science/CS courses is restricted at some campuses
  - Other campuses seek to accept all students
- But even if the students are welcome in the courses, the courses may be an imperfect fit in some cases
  - Data science curriculum is evolving (R vs. Python; balance of stats and CS)
  - Geophysics is a not big enough source of students that it can ask for on-campus customizations
- Have not discussed more specialized data science skills
  - E.g. high velocity data (data coming so fast and in such quantity you can't capture all of it – think Facebook click streams, particle accelerators, and large sensor arrays) requires a set of Stats and CS skills not included here
- That 300-level Machine Learning Foundations course? Not offered on most campuses yet…
- A lot of the courses the Masters students need are undergrad courses
  - There are fixes to this problem (e.g. creating graduate variants)

BUT DON'T DESPAIR

THINGS WE MIGHT DO…

- Make some things pre-requisites
  - NOTE: has harmful effect on attracting diverse students (forces them to take courses in anticipation of enjoying geophysics rather than concurrently, when they can see benefit)
- Trim quantitative courses
  - Maybe Geophysics doesn't need Ethics of Data Collection (no moral agents or patients impacted by studies?)
  - Skip Python if they are getting the core CS programming courses
- Combine courses in new forms
  - A new Geophysics course that combines some courses (Data Wrangling and Data Quality?) with a geophysics bent?
  - A bootcamp that covers another 6 to 8 credits?
- The ideas above might save 7 credits and move another 6 to 8 quantitative credits to summer.
- Could also encourage students to take certificates in quantitative topics (i.e. move out of core geophysics)

THE REST OF TODAY'S MEETING

- Recall, we're not solving problems (that's for a NASEM activity)
- Today our goal is to get a rough sense of what might have to be fit into in a next generation geophysics program – what skills (rather than courses, think course modules) do we think a student in geophysics should have – both quantitative and core geophysics
  - This should give us a better sense of just how hard an NASEM activity will be…
- The underlying logic is that we're not going to get to just add (a lot of) quantitative skills courses and trim geophysics courses. Rather an integrated curriculum is going to be needed
  - that's hard, as geophysics is a small major, so it can't get other programs to change offerings just for geophysics
  - BUT… many other STEM programs have similar curriculum challenges, so this is likely a time of curricular evolution, where if geophysics knows what it wants, it may be in a position to help/lead a campus effort