ASTR 502: Astronomical Discovery in the Era of Big Data

Professor: Prof. Nicholas Law (nmlaw@physics.unc.edu)
Class Location: Phillips 222
Class Hours: Tuesday & Thursday, 3:30pm – 4:45pm
Office Hours: TBD in class
Email: nmlaw@physics.unc.edu

COURSE OVERVIEW

UNC's Evryscope sky-survey telescopes (right) image the entire accessible sky every two minutes and have the largest light-collecting power (étendue) of any currently-operating sky survey. The systems, based in Chile and California, produce hundreds of terabytes of data per year, monitoring tens of millions of astronomical objects. In this class, we will explore the Evryscope dataset using signal processing and big-data tools to discover new astronomical phenomena.

The Southern Evryscope deployed in the Chilean Andes. The system is a 23-telescope array mounted into a hemisphere which tracks the sky.

The course is largely independent research, although everyone will work on the same topic and collaboration is very strongly encouraged. Class sessions will each start with formal instruction but will mostly consist of interactive and/or collaborative work.

LOGISTICS

Textbook: There is no required textbook.

Class attendance: The class periods are designed to a) give you the information and techniques you need for your research, and b) the opportunity to interact one-on-one with your classmates and Prof. Law to solve the research challenges. Each week we will start with a discussion of a new technique, often with an in-class exercise. To avoid students slipping behind, **each class is mandatory – let me know if you have to miss a class.**

Computing hardware and software: For the initial in-class exercises you will need a computer capable of running python in some form (for undergraduates: your ASTR202/ASTR301 setup is fine; grad students should ask me if you're not sure). Once we start the full analysis, the vast majority of the computations will be performed on the class server. You will need a computer with the ability to run an SSH client to connect to it (we will discuss in-class what SSH means, but almost anything should be able to do it).

COURSE STRUCTURE & ASSESSMENT

Because there is an element of luck in astronomical discovery, **you will not be assessed on whether you find something or not, just on your performance with the class material.**

The grading is divided as follows:

On-going research:	60%
Paper presentations:	5%
Final write-ups:	20%
Final presentations:	15%

On-going research assessment & homeworks: The primary method of assessment is the research log you will keep on Sakai. You are expected to update this log each week; your research and this update forms the class homework. The blog should contain a record of your research, in sufficient detail that another researcher would be able to duplicate and verify your results (we will discuss in class the detailed expectations).

Like most research, you are expected to manage your time effectively without routine formal assignments. I will give you regularly-updated estimated timelines during class (both on the semester scale and short-term projects for the next week), as well as regular individual grades and feedback. If you are struggling to keep up or manage your time, you should contact me for help.

Paper presentations: to gain familiarity with the astronomical literature and practice presenting research results, on Thursdays we will have a 5-minute presentation from a student about a recent large-survey astronomy paper selected from any of that week's releases on http://arxiv.org/archive/astro-ph. This assignment will rotate though the semester.

Final: the class will conclude with writing up your results in the form of a short paper, and presenting your results in a presentation (the final presentations are also the final exam). Depending on how well our research goes (it's always uncertain), the best papers may be invited to be submitted for publication as an American Astronomical Society Research Note, or be collated for publication in a refereed journal.

Collaboration policy: As a research-based course, collaboration is **strongly** encouraged – you will work individually, but attacking the same problems as your classmates. However, *each student must do their own work – this means that, although it is great to ask your classmates for help, and figure out problems together, you must write and execute all your own code (much the same as when you prepare solutions in a more mathematical class)*. This will be assessed by the code and results you post on your research blog. ***Copying code between students is not allowed, and will be checked for.***

CLASS SCHEDULE

Formal lecture components in normal text; *in-class research work in italics.*

Class 1: Introduction to Evryscope
- Intro talk & syllabus
- Science goals

2: Evryscope light curves
- Magnitudes, MJD, etc.
- Expectations for research blogs
- *1st exercise: load in light curves, plot w. flags culled out*

3: Initial variability searching
- RMS variability search, benefits and pitfalls
- *Perform RMS search and interpret results*

4: Matched filter searches 1
- How BLS works
- *Start writing BLS code*

5: Matched filter searches 2
- How to interpret BLS periodograms
- *Continue making BLS code work*
- *Try BLS with example light curves*

6: Matched filter searches 3
- BLS SNR & false positive levels
- *Continue working on BLS code with example LCs*

7: Astrophysical interpretations
- Types of astrophysical variability & what we're interested in
- *Continue working on BLS code with example LCs; find interesting variability*

8: Systematics interpretations
- Types of non-astrophysical variability
- Pre-whitening
- *Code pre-whitening*
- *Complete BLS search of example LCs*

9: Finishing off example LC search
- *Day to complete LC search*

10: Target selection & literature searches 1

Discuss star formation regions, surveys, target areas

Select groups & targets (start downloads from CTIO)

Discuss searching astronomical literature

Each group learns everything they can about their star formation region and presents next class

11: Target selection & literature searches 2

Preparation for presentations

Group presentations on regions (10 minutes each)

12: Server processing, multiplexing data analysis and efficiency

How to access class server

Efficient code in python (exercise)

Efficient multiprocessing in python

13: Starting target searches

Querying for LCs

Starting the search!

14-19: initial search through dataset for interesting objects

generating astrophysical interpretations of discoveries

cross-validation between students

19-20: *schedule follow-up for most interesting targets (w. SkyNet; may be dropped in interests of time)*

20-22: *collating results & writing up discoveries w. simple interpretations*

22-25: *model discoveries*

25+: *writing up results as papers + contingency*