Data Science for Astronomy

Course Structure and Philosophy

Walter Silima

Staff

Lecturers

- Patric Woudt
- Russ Taylor
- Mattia Vaccari
- Jordan Collier (remote)
- Nathan Deg (remote)
- Walter Silima
- Sushant Dutta
- Mayhew Steyn

External Examiner

- Tapiwa Gundu (NMU)

Lecture Schedule

17 February – 23 April 2025

28 Lessons: Mondays 2-4 pm and 3-5 pm, Tuesdays 2-4 pm, Wednesdays 10-12 pm and 2-4 pm, and Thursdays 3-5 pm

Venue: Astronomy Seminar Room (5.40) and IDIA Visualisation Lab (5.20), RW James Building, Upper Campus

Snippet from Science Faculty Handbook

AST5004Z DATA SCIENCE FOR ASTRONOMY

12 NQF credits at HEQSF level 9

Convener: Professor P Woudt

Course entry requirements: Core modules of the NASSP Master's coursework.

Course outline:

This course introduces students to various aspects of data-intensive astrophysics, ranging from data visualization and complex databases to advanced statistical tools for astronomical data analysis and computational astrophysics. At the core of this module are examples of modern data-intensive astrophysics derived from the global data challenges around MeerKAT, the Square Kilometre Array (SKA), associated projects in radio astronomy, and other large multi-wavelength surveys. Students will be introduced to the use of Bayesian statistics in astronomy, the complexity of visualizing large data cubes, optimizing database operations in the presence of multi-dimensional data, data mining, and discovery tools, and the role of large-scale simulations in interpreting the significance of astronomical observations.

DP requirements: 50% average for the two projects.

Assessment: Two projects: 25% each. Practical 'take-home' data science examination: 50%. A sub-minimum of 50% for each of the projects, and examination component will be required.

Preamble

The course does not require a previous background in astronomy, so the start will include a brief survey of astronomy, astronomical observation, and astronomical data. Following this introductory material, the specifics of astronomical data sets and databases, software systems and environments, and algorithms and modalities of data processing, visualisation, and data fusion will be covered.

Students should emerge with a basic understanding of

- astronomical data and data systems
- experience with astronomical software and program environments
- appreciation of the state of the discipline globally, major developments in SA, and the global online data and tool environment

- what are the characteristics of astronomical data for the image domain, time domain, and catalogue domain
- understanding of basics for
    - data processing and pipelines
    - image processing
    - time series analysis
    - visualization and visual analytics
    - data mining and applications of automation and machine learning
    - data challenges
- specialist experience in a specific area based on a course project

Program Elements

1. Introduction to astronomy, telescopes, and observations. (Lecture)
    a. The universe in a nutshell for the layman - planets, stars, galaxies, large-scale
    structure, cosmology
    b. Radio and optical telescopes
    c. International facilities/projects on the ground and in space
    d. Astrophysical quantities, radiation, emission mechanisms, spectral lines
    e. What data looks like in image domain, time domain, catalogue domain
2. Intro to Radio Astronomy (Lecture)
    a. Radio waves, single dishes, interferometry
    b. MeerKAT, ASKAP, SKA, science
3. Data structures & tools / Data inspection and analysis (Practical)
    a. Python, Jupyter notebooks
    b. File formats: measurement sets, cubes, HDF5, FITS, etc
    c. Containers, parallelism
    d. Visualisation tools: CARTA, DS9, Aladin, CASA, etc
4. Large-scale simulations (Lecture)
    a. Basic concepts, 2-body problem, gravity, motion
    b. Different algorithms, code, and examples
5. Data to information I: Time domain astrophysics (Lecture)

    a. Time series data, light curves, from synoptic wide-field sky survey
    b. Fourier analysis of a small sample of light curves: period analysis
6. Statistical tools in astronomy I: Intro and application (Practical)
    a. scipy, numpy and matplotlib libraries, curve_fit (Jordan)
    b. Radio spectra (Jordan)
    c. Statistical methods for RFI detection (Walter)
7. Statistical tools in astronomy II: Bayesian statistics (Lecture + practical)
    a. Parameter search, estimation, and MCMC
    b. Bayes' theorem
8. Data to information II: Multi-wavelength data, data fusion and machine learning (Lecture +
practical)
    a. Galaxy formation and evolution, SFRs, SEDs, multi-wavelength data
    b. Data fusion, HELP, cross-matching and classification
    c. Source characterisation, HIPPO, SDSS, LSST, SKA, IDIA
    d. ML with SDSS, SciServer, KNN classification and regression, photo-z
9. Data to information III: Radio Data processing (Lecture)
    a. Calibration and imaging
    b. Processing environments such as CASA, automation and pipelines
    c. Polarisation, Stokes parameters, frequency
10. Data to information IV: Radio source finding and characterisation (Lecture + practical)
    a. Radio sources and galaxies
    b. Source finding, source characterisation
11. Visualisation and Visual Analytics (Lecture + demo)
    a. Visualising large data in the cloud
    b. Cubes, HDF5, and tiling
    c. Rendering N-dimensional catalogues in real time
    d. Volume rendering of 3D data
    e. Virtual Reality for data visualisation
12. Astronomical data and databases. (Lecture)
    a. Multi-wavelength data, missions, observatories, surveys
    b. Major astronomical databases: NED, WISE, SDSS, LSST, IVOA

c. WISE: IRSA, databases, imaging services, catalogue services, interoperability

d. SDSS: magnitudes, schema, SQL, file formats

Assessment

Project 1: 25%

Tentative release 12 March, hand in 26 March

Project 2: 25%

Tentative release 28 March, hand in 07 April

Exam: 50%

Computational Environment

JupyterHub access, and the relevant software containers.