

ASTR4004

COMPUTATIONAL ASTRONOMY

Week 4: https://github.com/svenbuder/astr4004_2025_week4

Spiral galaxy M74 in face-on view. Figure credit: Gemini Observatory, GMOS Team



Simulated spiral galaxy in face-on view.

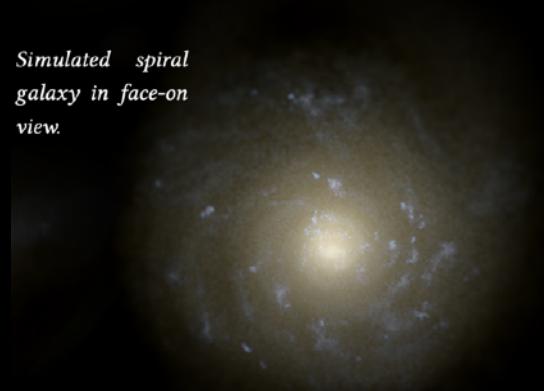
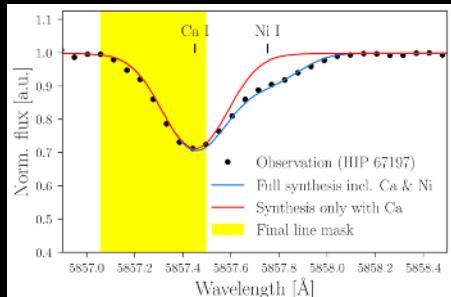


Figure credit: Tobias Buck

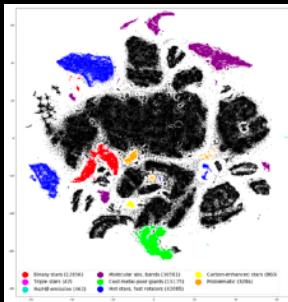


Australian
National
University

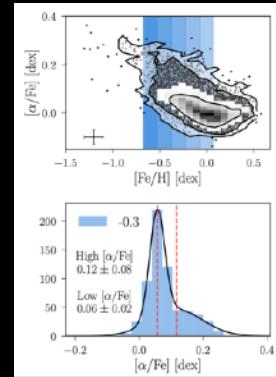
How I got into Computational Astronomy



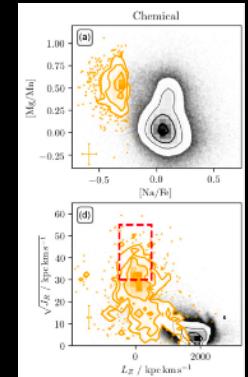
1 Mio. Spectrum analyses
Physics- & Data-driven



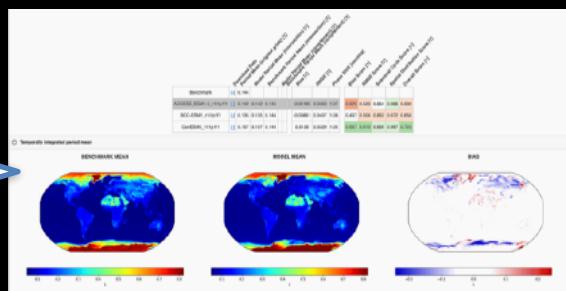
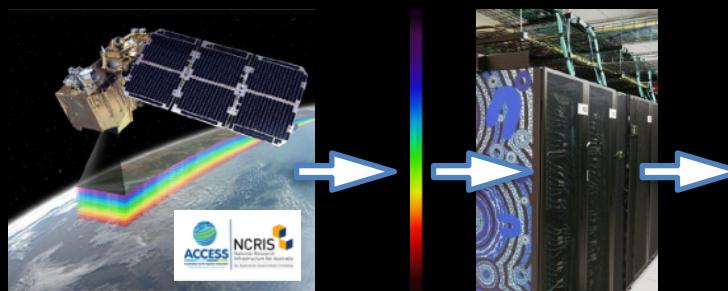
Dimensionality Reduction
(Sorting Spectra)



(Gaussian) Mixture Models of 2 or more components in N dimensions



Secondment
to Climate
Modellers:
git



My idea for this week:

Week	Summary	What I actually plan to talk about
4	Data Processing	git, csv/FITS Files, ADQL/SQL, joining & cleaning catalogues, ...
	Statistics	How to calculate and report uncertainties
	Plot Clinic	How to plot well & better -with my (and hopefully your) examples, ...

What I will also teach you later in the semester:

How to fit data *properly*,

How to reduce dimensionality of data

How to identify clustering in data

How to judge different models and confirm/reject hypotheses

How to apply what you learned outside of astronomy (other science or "industry")



Disclaimer: I can only cover the basics

There is a lot of great introductions to computational astronomy out there - including great tutorials!

I have selected what I found most useful in my personal experience.

I am more than happy to adjust the course and go more into depth for specific parts if you want to.

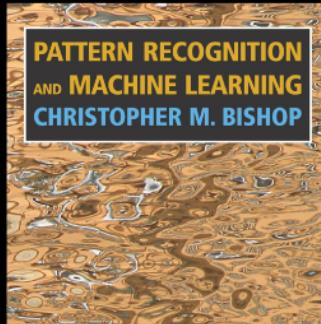
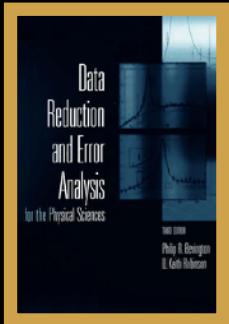
The image shows a grid of 12 cards from the scikit-learn website, each illustrating a different machine learning concept:

- Classification:** Identifying which category an object belongs to. Examples include a grid of astronomical images (galaxies and stars) and a scatter plot of handwritten digits.
- Regression:** Predicting a continuous-valued attribute associated with an object. Examples include a line graph showing predicted average energy transfer over time.
- Clustering:** Automatic grouping of similar objects into sets. Examples include a scatter plot of handwritten digits grouped into clusters.
- Dimensionality reduction:** Reducing the number of random variables to consider. Examples include a 3D scatter plot of astronomical data labeled "Virgo", "Ursa Major", and "Ursa Minor".
- Model selection:** Comparing, validating and choosing parameters and models. Examples include a line graph showing the effect of different parameters on model performance.
- Preprocessing:** Feature extraction and normalization. Examples include a grid of astronomical images showing data transformation.

<https://scikit-learn.org/stable/index.html>

Some useful online references

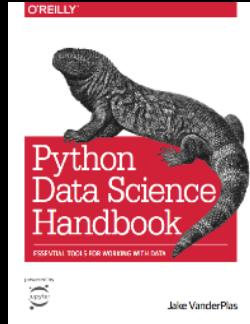
Bevington & Robinson: Data Reduction and Error Analysis for Physical Sciences



Bishop: Pattern Recognition and Machine Learning

GitHub tutorials such as big-data-tutorial
Chat GPT is also helpful (if you reflect on answers critically!!)

VanderPlas: Python Data Science Handbook



<https://scikit-learn.org/>
Preprocessing, Model selection, Regression, Classification, Clustering, Dimensionality Reduction

<https://www.astropy.org/>



<https://docs.scipy.org/>
integrate, optimize, interpolate,
Fast Fourier Transforms: fftsignal
Linear algebra: linalg



What I want you to get out of the course

A better understand of how to think critically

Something useful for your research + your career

How can we use computing to enable (better) science?

A toolset for computing
in and outside of astronomy

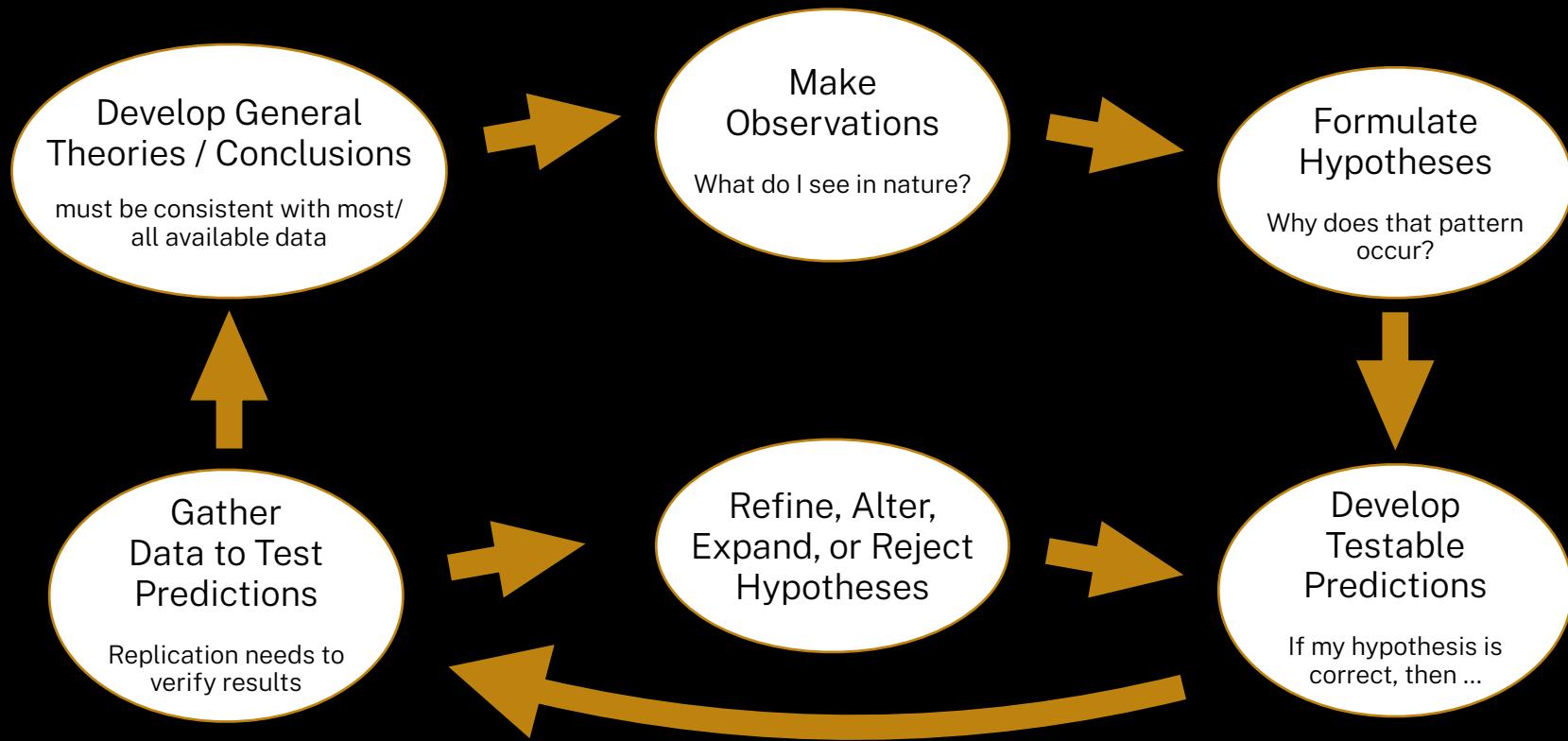


BEFORE YOU START CODING...



Australian
National
University

There is a method for how to do science!



Responsible Research Practices: Three Rs

- Repeatability: same results when repeated under same conditions by same team using same software using same method: consistency within original setting
- Reproducibility: same results when repeated under same conditions by different team using different software with same method: sound & transparent methodology
- Replicability: consistent results when repeated under different conditions in a different context: generalisability of the methodology



Rationality, (Un)Certainty, and Open Science

Claims of scientific truth can be opposed by:

- falsifying them
- questioning their certainty
- asserting the claim to be incoherent

Problems that might get in the way of doing exactly that:

- You do not publish your data (you may have your reasons) -or nobody can actually open your data files: **Preprocessing**
- You are not calculating or reporting your uncertainties: **Statistics**
- Nobody understands what you are showing or claiming: **Plot Clinic**
- Nobody can reproduce your results: **Codesharing on GitHub**



In practice: Your "marking guideline" in journals



Is the subject appropriate for the journal?	<input type="radio"/> Yes	<input type="radio"/> No	<input type="radio"/> Unsure
Is the content sufficiently significant to warrant publication in MNRAS? Please comment on the significance and originality of the paper in the comments box below.	<input type="radio"/> Yes	<input type="radio"/> No	<input type="radio"/> Unsure
Are the methods and results set out clearly?	<input type="radio"/> Yes	<input type="radio"/> No	<input type="radio"/> Unsure
Are all the necessary and appropriate references given?	<input type="radio"/> Yes	<input type="radio"/> No	<input type="radio"/> Unsure
Is the title appropriate and sufficiently informative?	<input type="radio"/> Yes	<input type="radio"/> No	<input type="radio"/> Unsure
Is the abstract sufficiently informative?	<input type="radio"/> Yes	<input type="radio"/> No	<input type="radio"/> Unsure
Can the paper be shortened without loss of clarity?	<input type="radio"/> Yes	<input type="radio"/> No	<input type="radio"/> Unsure
Are all the figures and tables necessary and adequate?	<input type="radio"/> Yes	<input type="radio"/> No	<input type="radio"/> Unsure
Is the paper in reasonable English?	<input type="radio"/> Yes	<input type="radio"/> No	<input type="radio"/> Unsure
If the paper contains long tables, finding charts etc., can these be published as supplementary material?	<input type="radio"/> Yes	<input type="radio"/> No	<input type="radio"/> Unsure

Referee questionnaire from MNRAS

Please address the following questions in your report to the Scientific Editor

- 1. Why should this paper be published?*
- 2. Are the assumptions spelled out clearly?*
- 3. Are the methods fully described?*
- 4. Are the new results adequately emphasized?*
- 5. Are all the figures and tables necessary and properly laid out?*
- 6. Which material (sections, tables, figures) should be published in electronic form only?*
- 7. Is the designation of objects according to IAU rules?*

Referee questionnaire from A&A:
Bertout & Schneider (2004), A&A, 420, E1

“The best reviews are those that are thorough, identify strengths as well as weaknesses, are critical yet constructive, and are respectful to the authors.”

Schultz (2022), Monthly Weather Review, 150, 6

CODESHARING



Australian
National
University

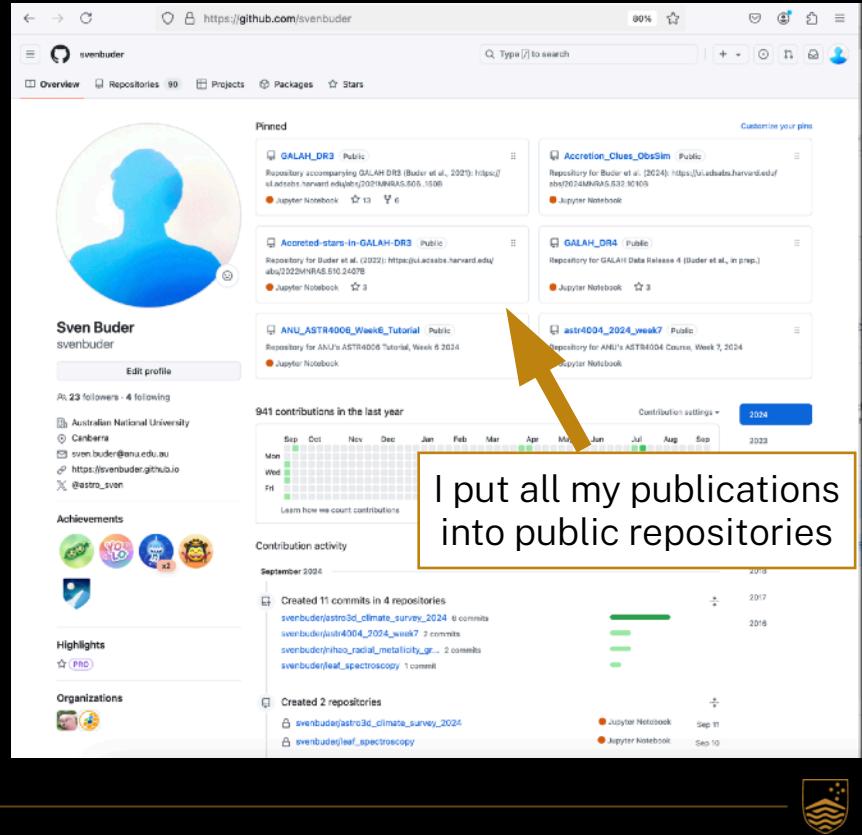
github.com: upload your code and share it

Redundancy & Flexibility:

- Large data files (>100MB) should be stored on servers
- As much of your data & code as possible should be stored in multiple places (redundancy)
- using git allows you to save multiple versions (in parallel) via branches.

A very real example: Your computer is stolen or breaks... what now?

Another reason: Easily test the same code on your Apple computer & the Linux supercomputer!



github.com: upload your code and share it

If you have not signed up to GitHub
or have never used it on any computer

<https://swcarpentry.github.io/git-novice/>

(intro to git)

<https://ivastar.github.io/advanced-git/>

(advanced git developed by StSci)



The screenshot shows a GitHub repository page for 'accretion_clues_ObsSim'. The repository has 354 commits and was last updated on June 18, 2024. It includes sections for About, Releases, Packages, and Contributors. A large orange circular icon with a white arrow is overlaid in the center.

The screenshot shows a LaTeX code editor with a file named 'Accretion_GALAHvsNIHAO.tex'. The code includes various packages like 'tikz', 'amsmath', and 'natbib'. A large orange circular icon with a white arrow is overlaid in the center.

The screenshot shows a file browser displaying a directory structure for '2404_Accretion_Clues_ObsSim'. It contains files like 'accretion_clues_simulation.ipynb', 'Accretion_GALAHvsNIHAO_240330.pdf', and 'Accretion_GALAHvsNIHAO_240404_MR.pdf'. A large orange circular icon with a white arrow is overlaid in the center.

The screenshot shows a terminal window with a command-line interface. The user is in a directory 'rsaa-067686/2404_Accretion_Clues_ObsSim'. They run the command 'ls -l' which lists several files and folders. Then they run 'git pull' and receive a message 'Already up to date.' A large orange circular icon with a white arrow is overlaid in the center.

Connect git (coding) & overleaf (writing)

The screenshot shows the Overleaf interface with a floating 'GitHub Sync' dialog box and a sidebar menu.

Sidebar Options:

- Code Editor
- Visual Editor
- C C B I
- Recom
- LEGACY
- example.bib
- example.eps
- example.png
- example.ps
- mnras_guide.pdf
- mnras_guide.tex
- mnras_te...** (highlighted in green)
- mnras.bst
- mnras.cls
- readme.txt
- File outline
- Introduction

GitHub Sync Dialog:

Export Project to GitHub

This project is not linked to a GitHub repository. You can create a repository for it in GitHub:

Owner: svenbuder **Repository Name:** Monthly Notices of the Roy

Description (Optional): (empty field)

Visibility:

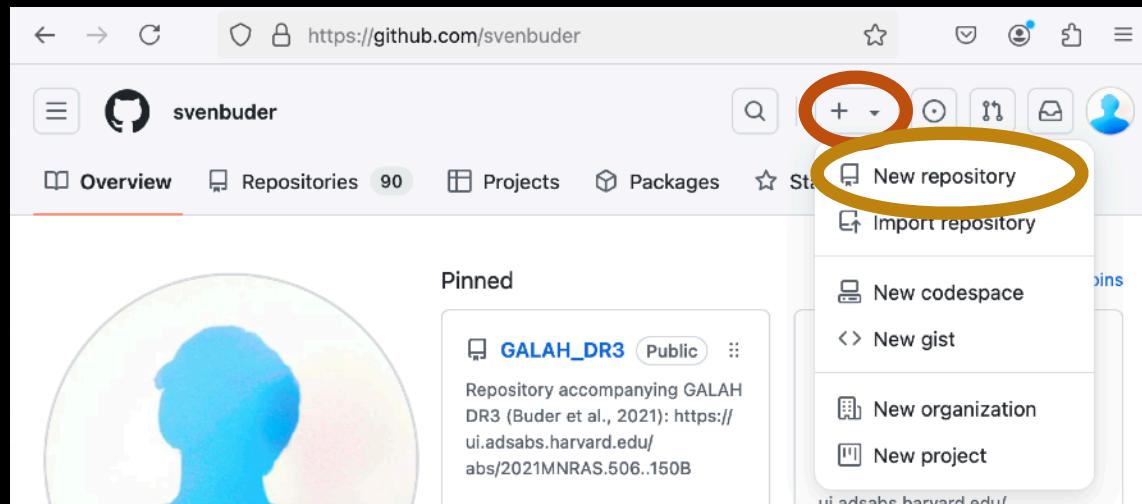
- Public** Anyone can see this repository. You choose who can commit.
- Private** You choose who can see and commit to this repository.

Create a GitHub repository



Creating a new repository

If starting a PhD: setup a journal template in overleaf once then clone it for new projects + link it to GitHub
If just starting a new coding project: create a private GitHub Repo



Creating a new repository

The screenshot shows a GitHub repository page for user 'astr4004_svenbuder'. At the top, there are buttons for 'Unwatch' (1), 'Fork' (0), and 'Star' (0). Below the header, there are two main sections: 'Set up GitHub Copilot' and 'Add collaborators to this repository'. The 'Set up GitHub Copilot' section includes a link to 'Get started with GitHub Copilot'. The 'Add collaborators' section has a search bar and a 'Invite collaborators' button. A large blue box titled 'Quick setup — if you've done this kind of thing before' contains the following information:

- Links for 'Set up in Desktop' (with icons for Mac, Windows, and Linux), 'HTTPS', and 'SSH', along with the URL https://github.com/svenbuder/astr4004_svenbuder.git.
- A note: 'Get started by [creating a new file](#) or [uploading an existing file](#). We recommend every repository include a [README](#), [LICENSE](#), and [.gitignore](#)'.
- A section titled '...or create a new repository on the command line' with the following git commands:

```
echo "# astr4004_svenbuder" >> README.md
git init
git add README.md
git commit -m "first commit"
git branch -M main
git remote add origin https://github.com/svenbuder/astr4004_svenbuder.git
git push -u origin main
```
- A section titled '...or push an existing repository from the command line' with the following git commands:

```
git remote add origin https://github.com/svenbuder/astr4004_svenbuder.git
git branch -M main
git push -u origin main
```

There are a lot of tutorials out there on how to set your repository up nicely.

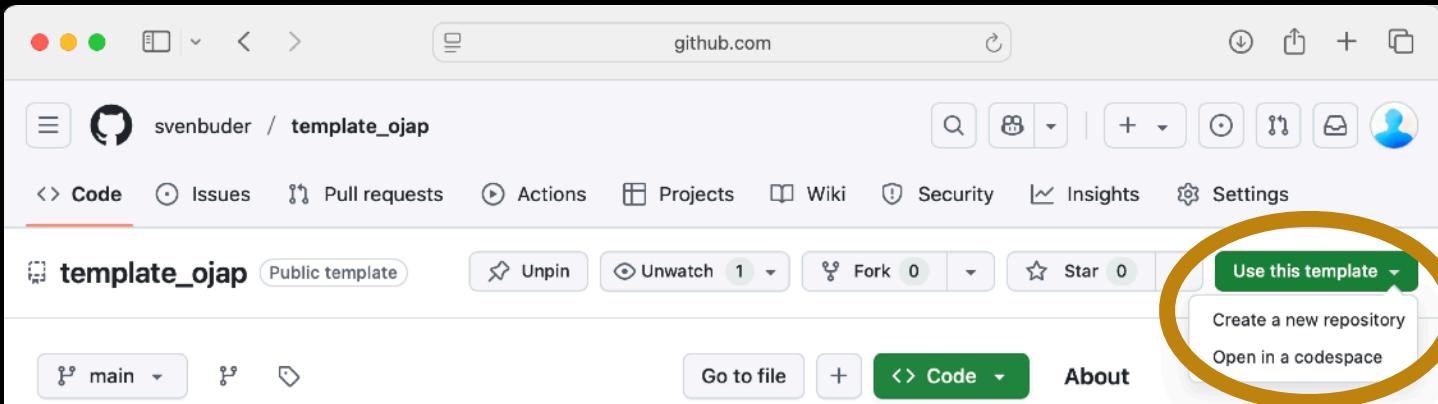
create .gitignore with everything want git to ignore:
*.file_ending
directory/
specific_file

and then use
git add .



Creating a new repository

Use someone else's template, e.g. https://github.com/svenbuder/template_ojap



The screenshot shows a GitHub repository page for 'template_ojap'. The repository is public and has 1 watch. It features a 'Code' tab, which is currently selected. A green button labeled 'Use this template' is highlighted with a yellow circle. A dropdown menu from this button contains the options 'Create a new repository' and 'Open in a codespace'. To the right of the repository details, there is a sidebar listing files and folders: 'data', 'figures', 'tables', 'tex_text', '.gitignore', 'README.md', 'bib.bib', 'mnras.bst', 'openjournal.cls', 'research_paper.ipynb', and 'research_paper.tex'.

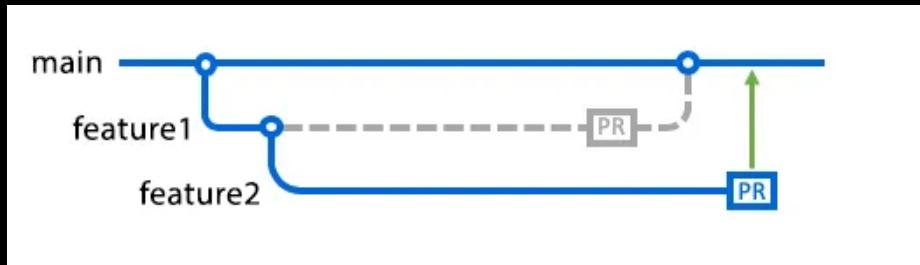
Step-by-step guide on how to set up a research paper and data structure

Creating a new repository branch

This can be extremely important for collaborative coding!

main branch:

the one that is "live" for example a website or a computer game program



development branch, e.g. feature1 or feature2:

the one you actively code on that can break without causing disaster



Creating a new repository branch

You can create branches online:

The screenshot shows a 'Branches' page with three sections: 'Default', 'Your branches', and 'Active branches'. In the 'Default' section, there is one branch named 'main'. In the 'Your branches' section, there are two branches: 'fit_line' (updated 'now') and 'fit_line' (updated '15 minutes ago'). In the 'Active branches' section, there is one branch named 'fit_line'.

But you can also do it in bash:

```
git checkout -b fit_line  
git push --set-upstream origin fit_line  
git pull origin main
```

Check with branch you're on locally,
see all local+remote branches,
and switch to another branch:

```
git status  
git branch -a  
git checkout my-branch-name
```



Merge into another branch

The screenshot shows a pull request interface on a web application. At the top, there's a yellow banner with the message "fit_line had recent pushes less than a minute ago" and a "Compare & pull request" button. Below the banner are navigation links: Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. A search bar contains the query "is:pr is:open". To the right of the search bar are buttons for "Labels 9", "Milestones 0", and "New pull request".

The main area is titled "Open a pull request" and includes instructions: "Create a new pull request by comparing changes across two branches. If you need to, you can also compare across forks. Learn more about diff comparisons here." It shows the base branch is "main" and the compare branch is "fit_line", with a note "Able to merge. These branches can be automatically merged."

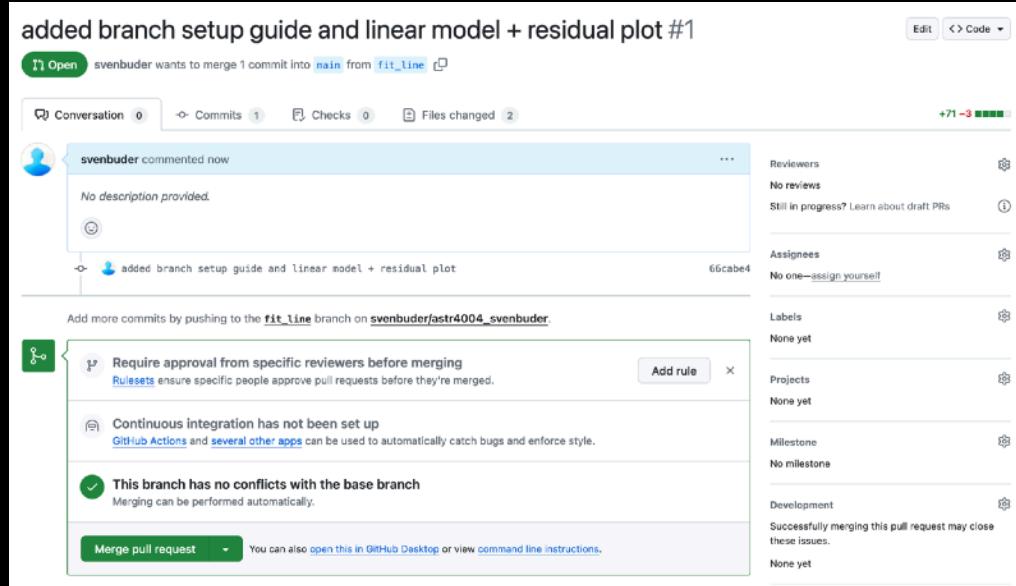
The pull request form has fields for "Add a title" (containing "added branch setup guide and linear model + residual plot") and "Add a description" (with a rich text editor placeholder "Add your description here..."). Below the editor are buttons for "Markdown is supported" and "Paste, drop, or click to add files". A "Create pull request" button is at the bottom.

On the right side, there are sections for "Reviewers" (No reviews), "Assignees" (No one—assign yourself), "Labels" (None yet), "Projects" (None yet), "Milestone" (No milestone), and "Development" (Use Closing keywords in the description to automatically close issues). A "Helpful resources" link is also present.

For large projects, you can add a description of what you actually changed



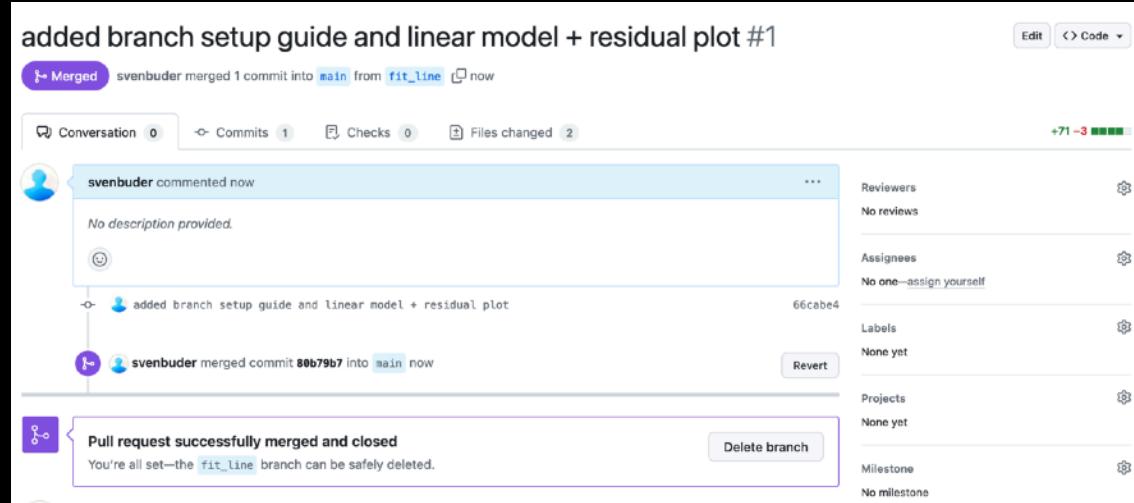
Merge into another branch



For professional coding projects, teams often work with a peer review

We will not discuss it here, but you can also include "actions" (like github running code online if you open a merge request or push content)

Merge into another branch



This is exactly what I did with a team of software engineers when I was working for "industry", that is, a climate modelling institute
git's version control also allows you to publish packages and releases



For science: don't forget to synchronise git and overleaf

The screenshot shows the Overleaf web interface for a LaTeX project titled "mnras_template.tex". A red arrow points from the sidebar menu to the "GitHub" sync option in the "Sync" section of the sidebar. A yellow arrow points from the main workspace towards the "GitHub Sync" modal window.

GitHub Sync

This project is synced with the GitHub repository at [svenbuder/survey_abundance_validation](#)

Recent commits in GitHub

- with example survey 056467
- by svenbuder <sven.buder@anu.edu.au>
- zwischenspeichern 7ac3b3
- by svenbuder <sven.buder@anu.edu.au>
- add first literature homogenisation example 3603c2
- by svenbuder <sven.buder@anu.edu.au>

Pull GitHub changes into Overleaf

Push Overleaf changes to GitHub

Code Editor Visual Editor **Record**

Download

Source PDF

Actions

Copy Project Word Count

Sync

Dropbox Git GitHub

Settings

Compiler pdfLaTeX

MONTHLY NOTICES
of the Royal Astronomical Society

LEGACY example.bib example.eps example.png example.ps mnras_guide.pdf mnras_guide.tex mnras_te... mnras.bst mnras.cls readme.txt

File outline Introduction

DATA PROCESSING

- 1) Writing/Reading files and catalogues:
Excel, CSV, FITS, HDF5, Xarray
- 2) Working with catalogues:
Download with ADQL, joining & cleaning catalogues
- 3) Interpolation

For this part, we will turn to the jupyter notebook
[anu_astr4004_2025_week4a_preprocessing.ipynb](#)



Australian
National
University

STATISTICS

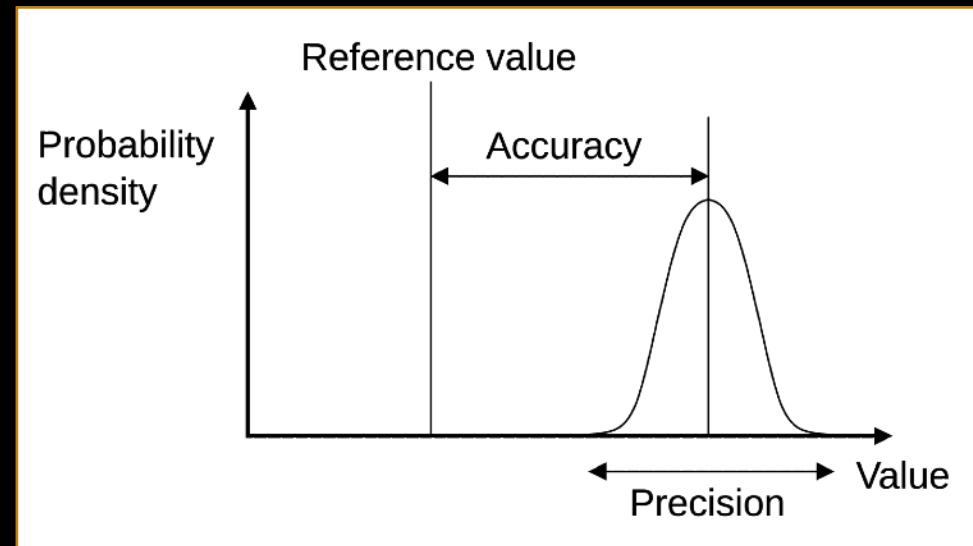
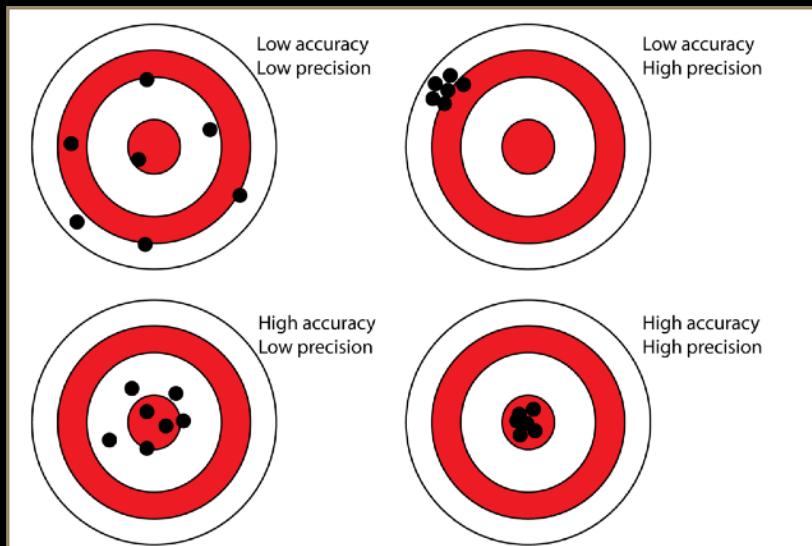


Australian
National
University

Measurement Uncertainties: Accuracy & Precision

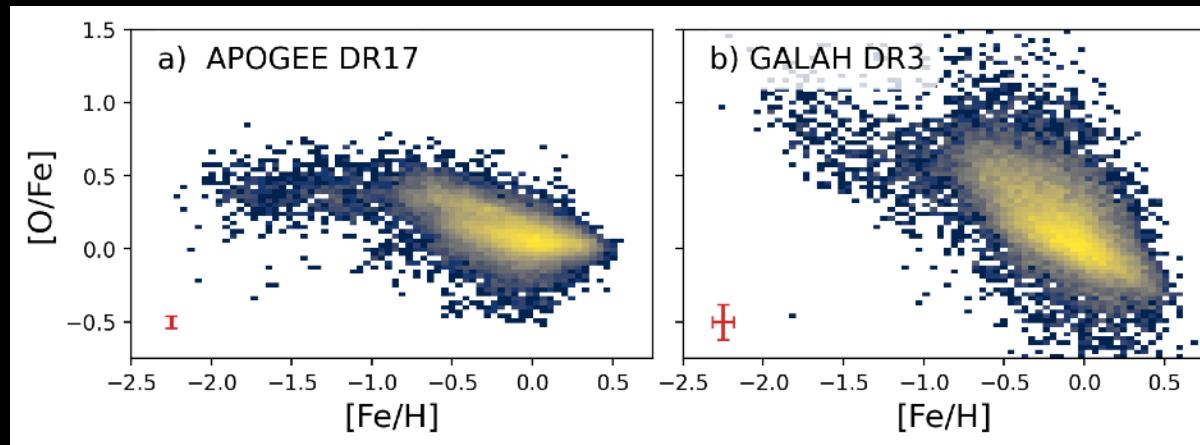
Accuracy is a description of *systematic uncertainty*, a measure of **bias**.

Precision is a description of *random uncertainties*, a measure of **variability**.



Measurement Uncertainties: Accuracy & Precision

An example: the amount of oxygen and iron in the same 26,000 stars
Error bars indicate the measurement precision (APOGEE is more precise)
But even with more precise measurements, the trends would not agree!
One of the two measurements (or both) are inaccurate!



Measurement Uncertainties: Accuracy

Estimating the accuracy of measurements can often be quite hard.
You need a reference that you can compare yourself with.

We do not "measure" oxygen, but we measure the influence from oxygen on light of stars (absorption lines in spectra) and then use models to estimate what amount of oxygen would cause such a shape.

Also note: one person's analysis output
can be another one person's analysis input



Measurement Uncertainties: Accuracy

Also note: one person's analysis output
can be another one person's analysis input

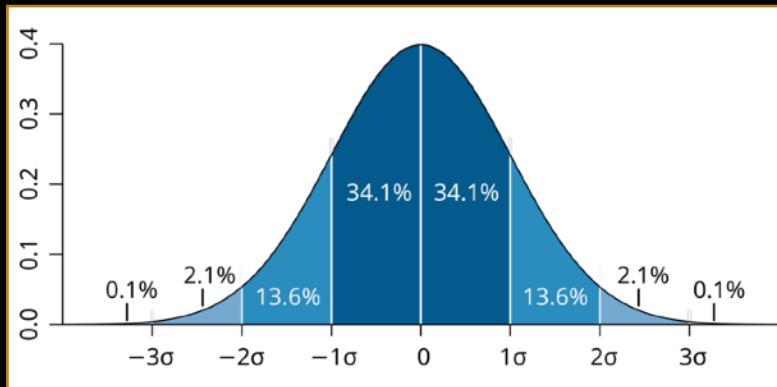
Chemical evolution modellers then use these "measurements" as their input to create models that predict how oxygen is created.

To be able to make any progress, we sometimes neglect measurement inaccuracies (but need to always keep them in mind and discuss them)

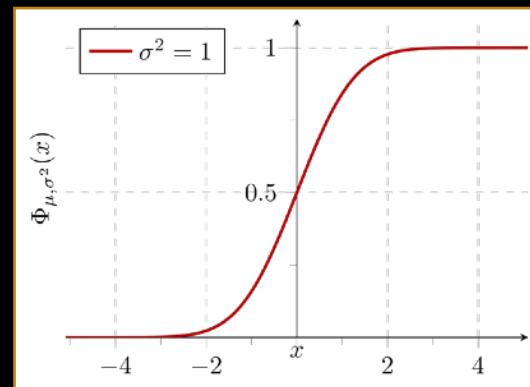


Measurement Uncertainties: Precision

For simplicity: assume we have perfect accuracy around true value 0
The only uncertainty we have is a Gaussian precision uncertainty



Distribution Function:
68.3% within $\pm 1 \sigma$, 13.7% outside
 $2 \sigma : 94.4\% / 5.6\%$; $3 \sigma : 99.7\% / 0.3\%$



**Cumulative
Distribution Function (CDF)**



Statistics: Significance

Extraordinary claims require
extraordinary evidence!

3 sigma for most measurements

5 or even 7 sigma for planet
detections or cosmology

You can also turn this around with
upper limits: I would have detected
it, if the signal would have been
that strong

