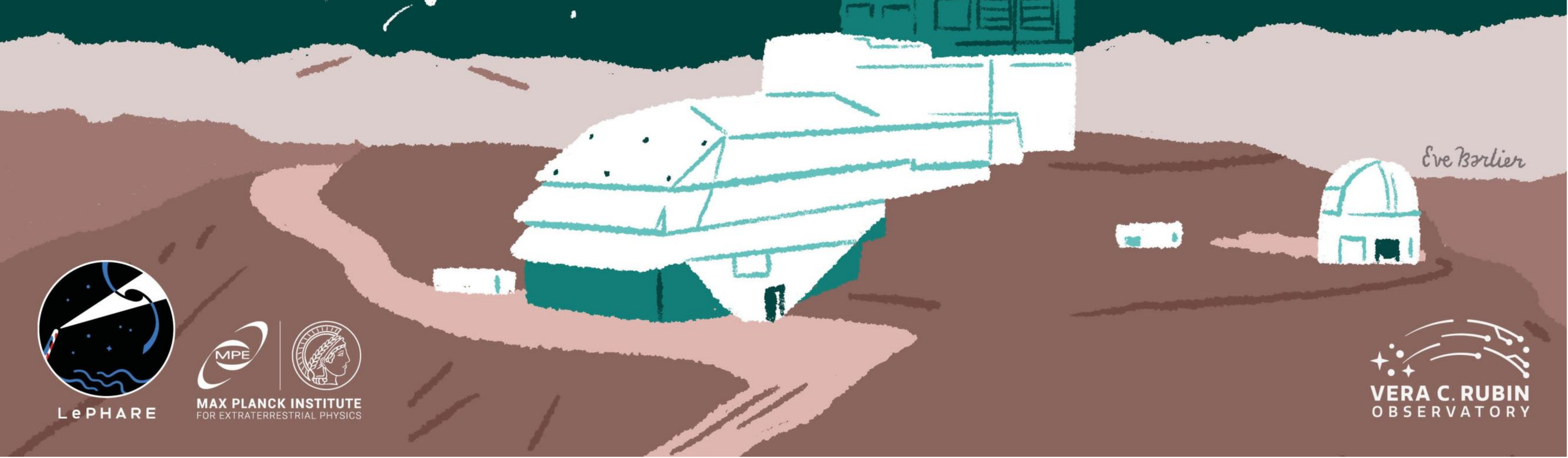
Photometric redshifts for active galactic nuclei with LePHARE for the Vera C. Rubin Observatory

Raphael Shirley

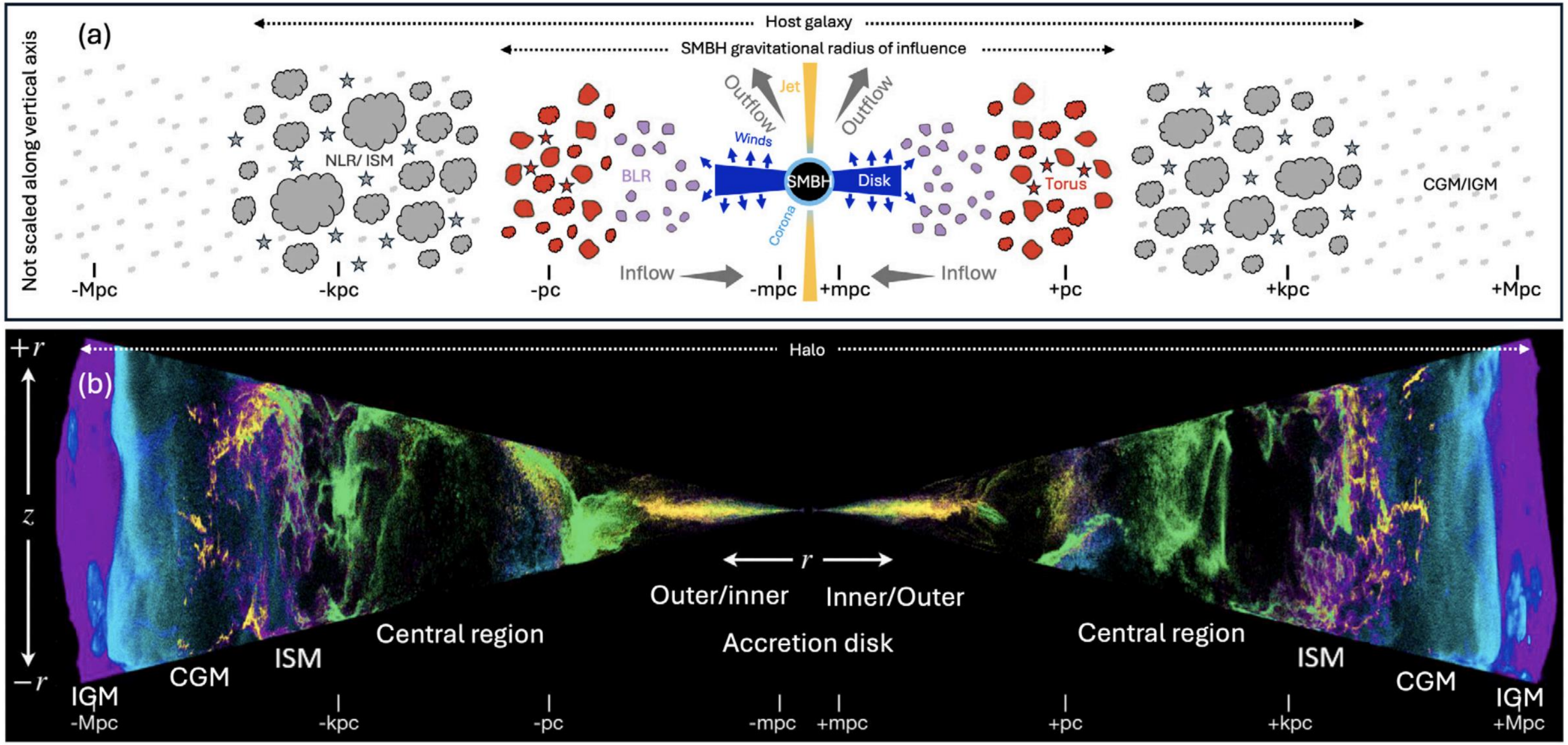
Max Planck Institute for Extraterrestrial Physics

With Mara Salvato, Johann Cohen-Tanugi, and Olivier Ilbert.

17 September 2025

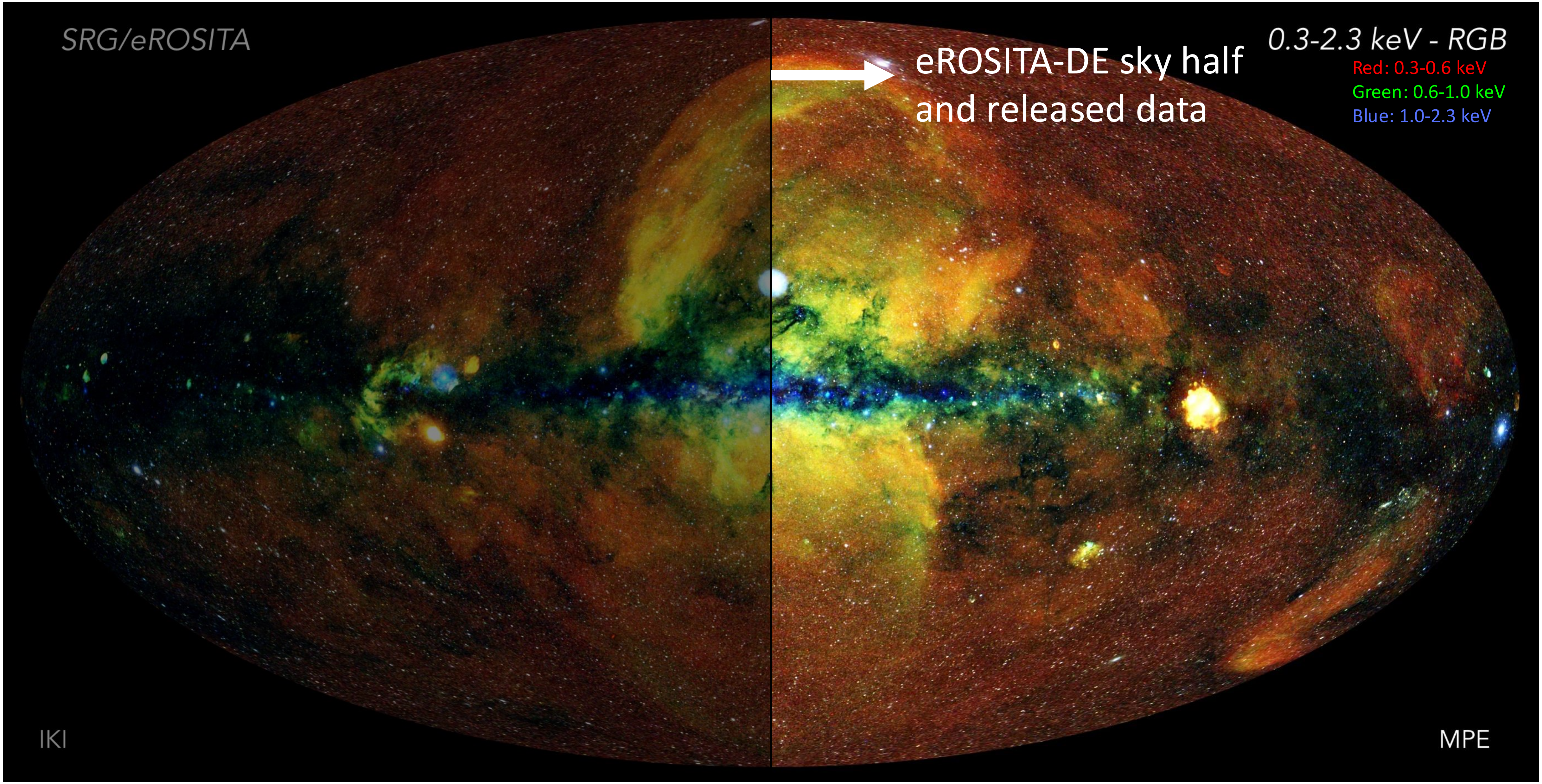


AGN in host-galaxy and halo environment



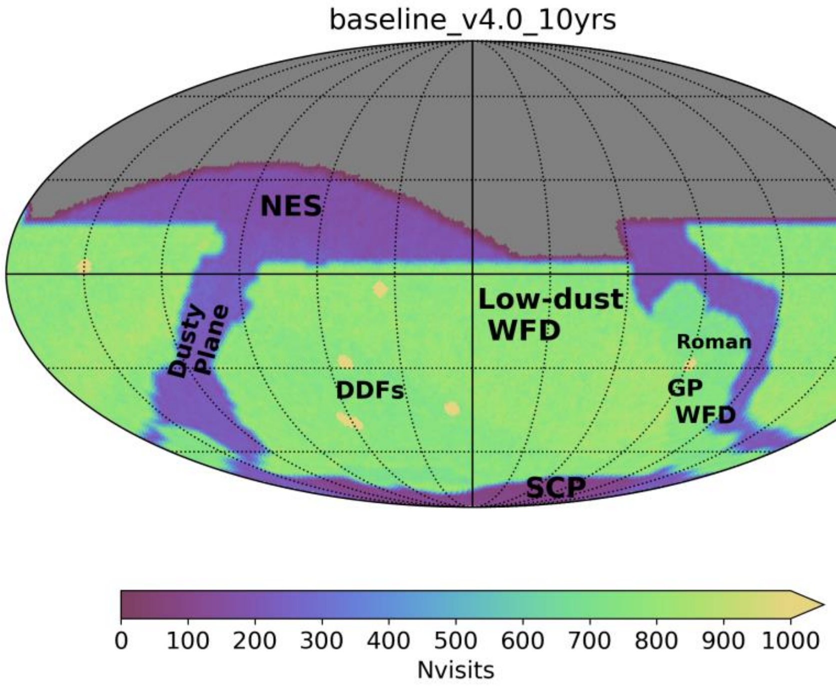
Alexander et al. 2025

The eROSITA eRASS1 sky

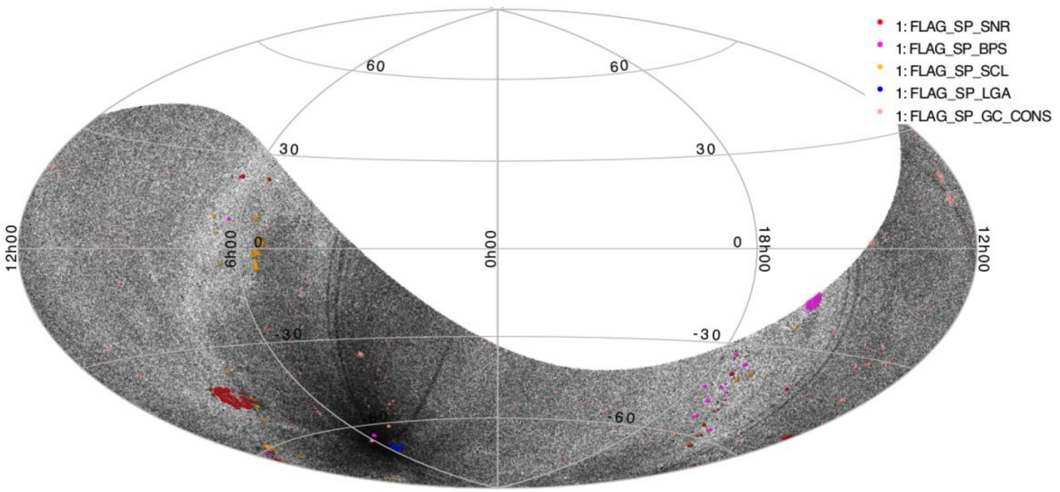


Merloni et al. (2024), Salvato et al. (2025)

LSST/eROSITA overlap

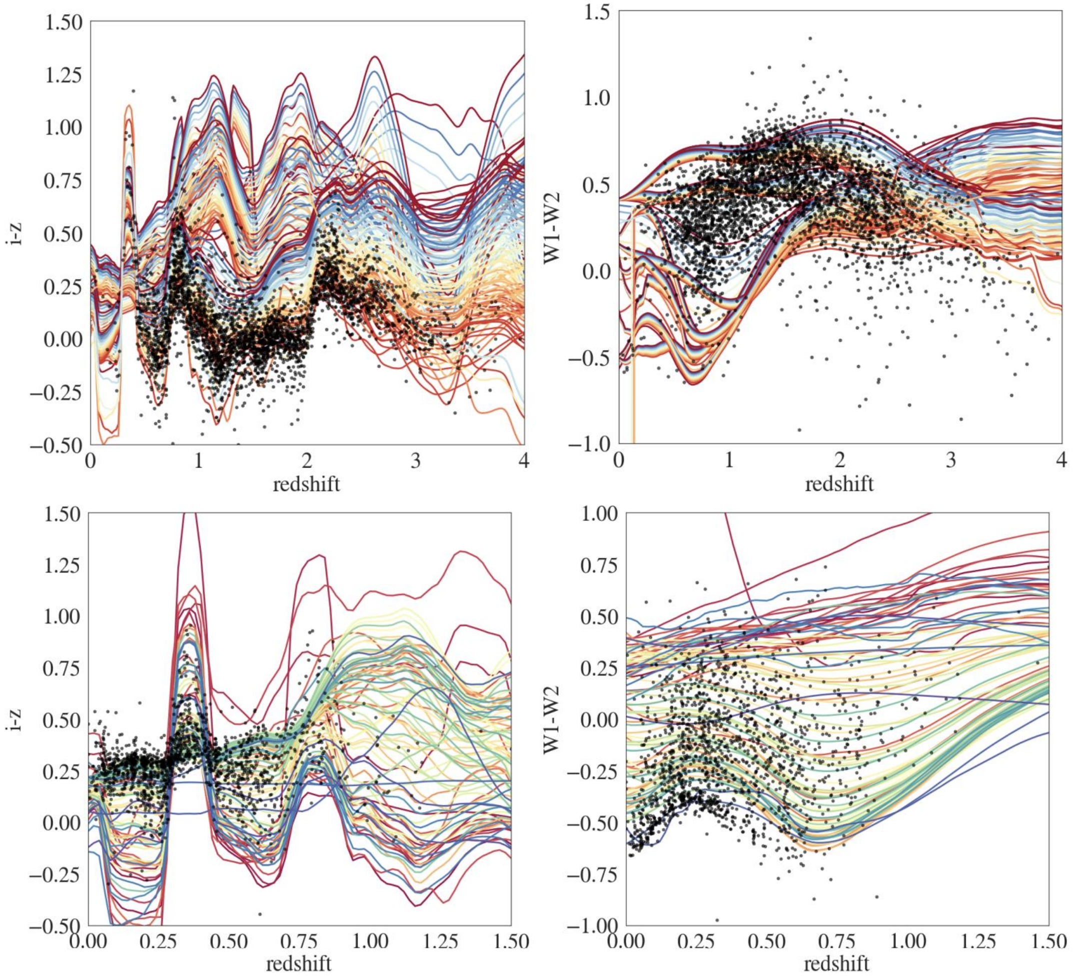


PSTN-056



Merloni et al. (2024)

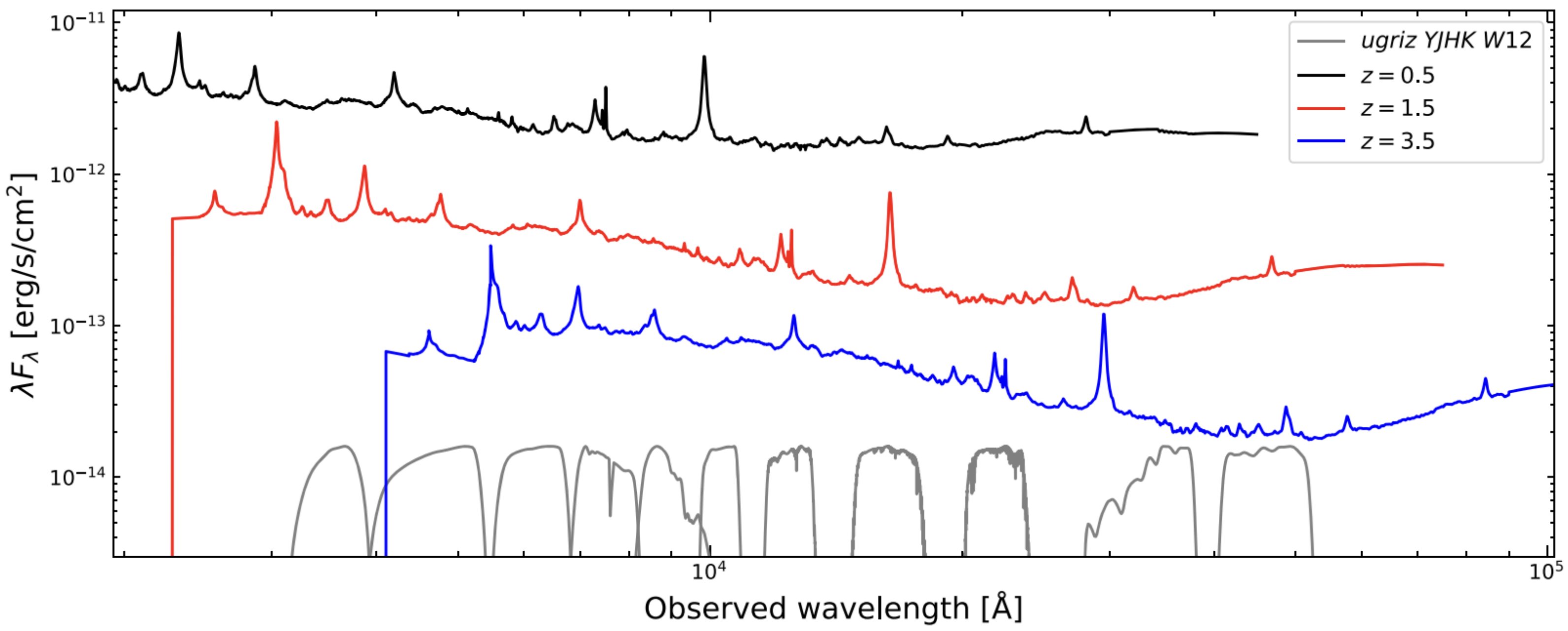
Colour redshift relations



Salvato et al. (2022)

- Mapping between colour space and redshift space
- Template fitting predicts colour redshift relations at all redshifts
- Machine learning trains on colours at fixed spectroscopic redshifts
- Some parameter space not in training may be real

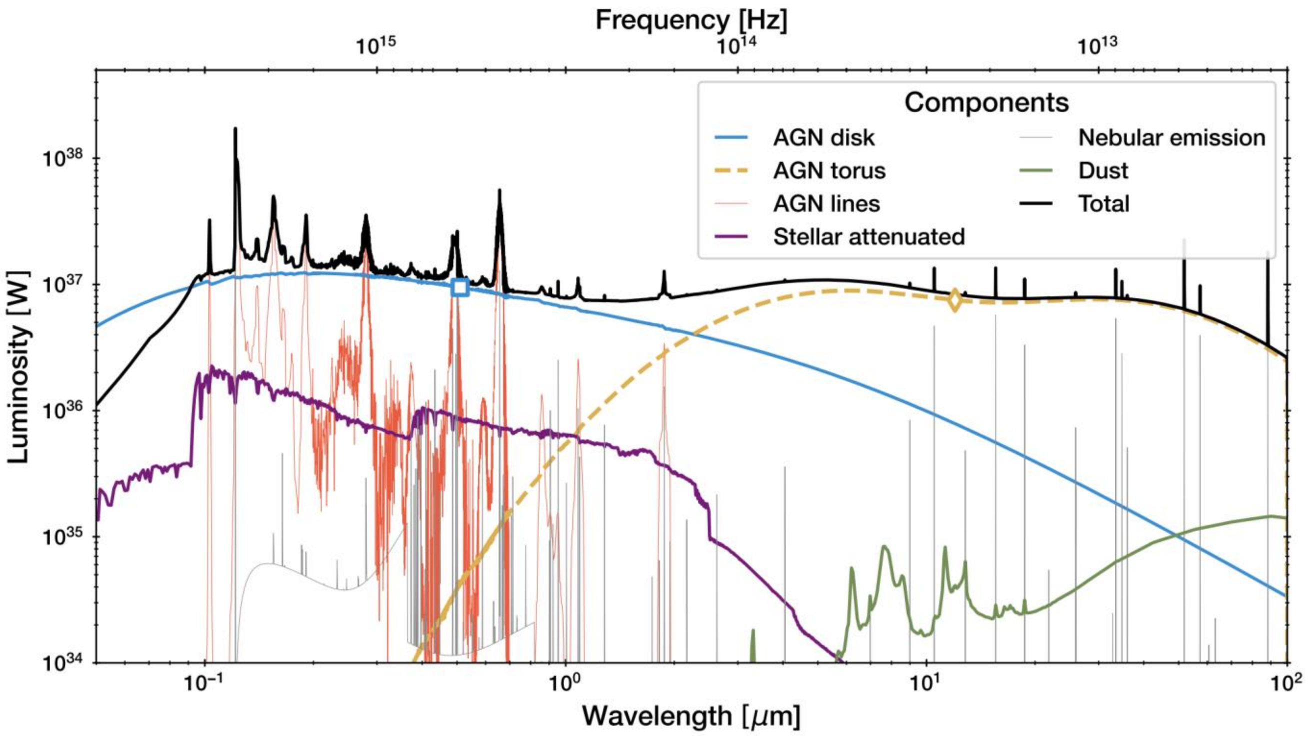
Challenges with photoz for AGN



Temple et al. (2021)

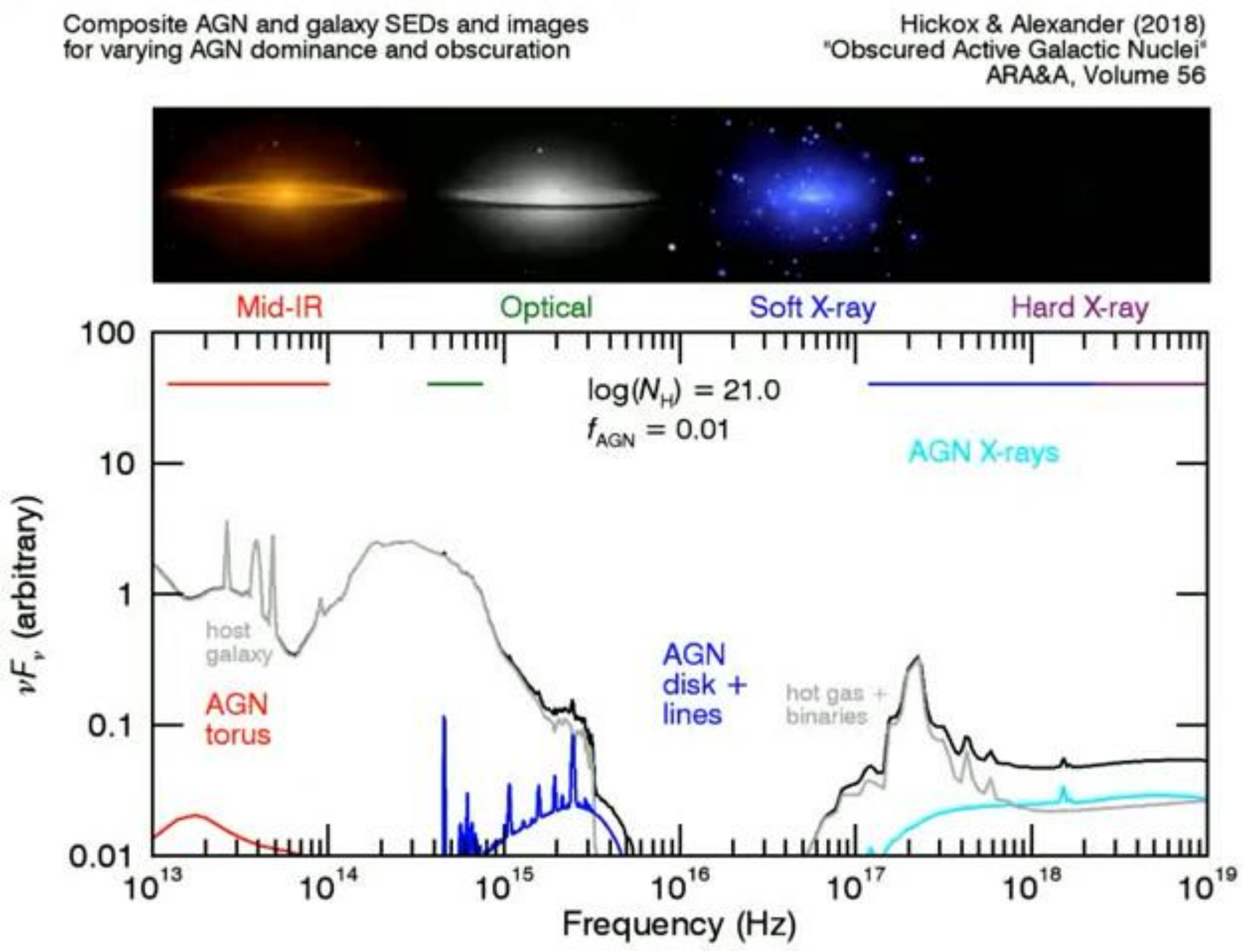
- Negative power law continuum degenerate in normalisation/redshift
- Most colour information reflects lyman break moving through filters or broad lines.
- Lyman break only useful above redshift 2

Host vs AGN properties



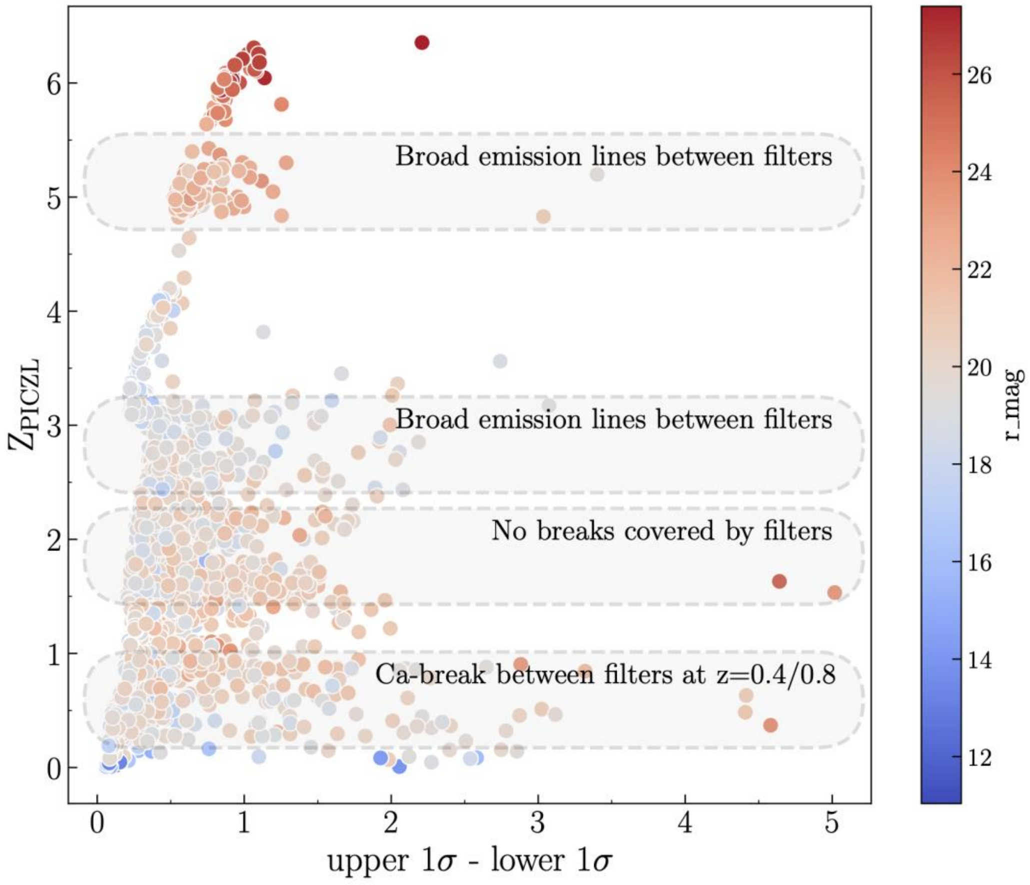
Buchner et al. (2024) GRAHSP

Host vs AGN properties



Hickox and Alexander (2018)

Error has redshift dependence



Roster et al. (2024)

- Above redshift 3 Lyman break moves into u band
- Full posteriors can account for this
- Sample from posteriors in binning to propagate errors

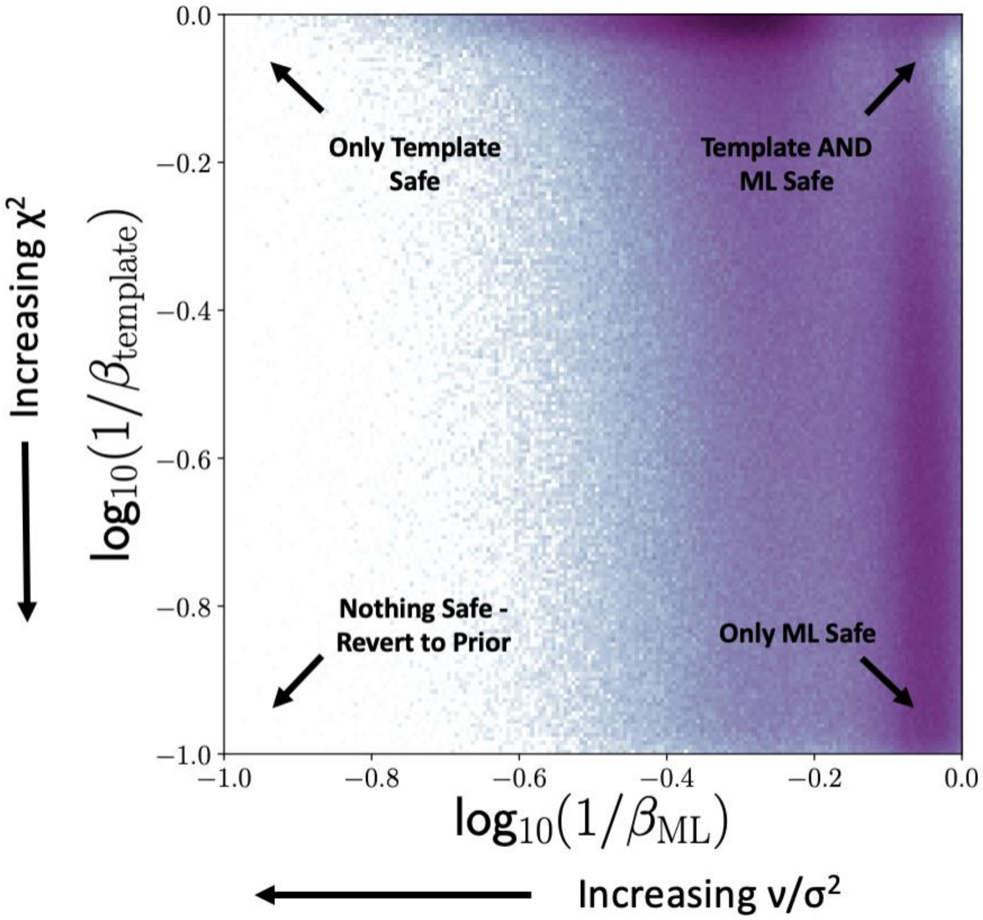
Templates or machine learning?

- Machine learning:
  - High performance inside training space
  - Easy to use additional inputs
  - Fast to run after training
  - Poor handling of rare objects
  - Requires retraining on new data

- Template fitting:
  - Physically motivated
  - No training set required
  - No training set bias
  - Can probe new regimes
  - Statistically robust
  - Only method for small samples
  - Requires accurate measurements and errors
  - Physical properties

Templates **and** machine learning

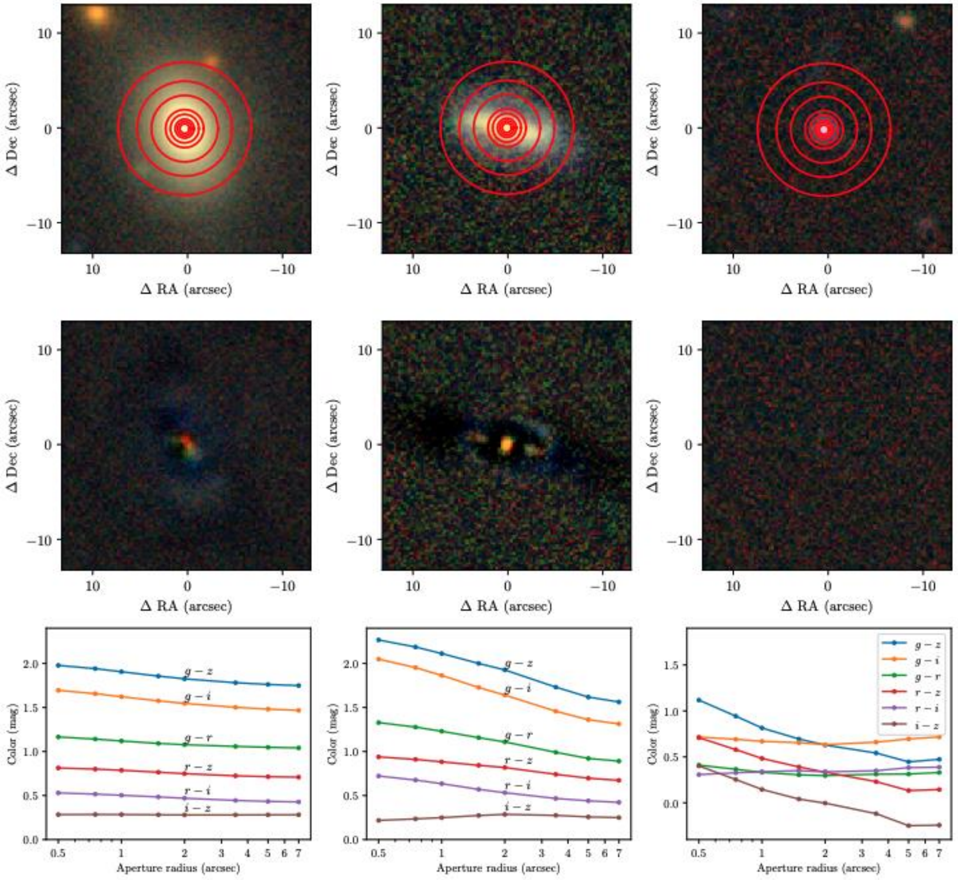
Bayesian hierarchical model

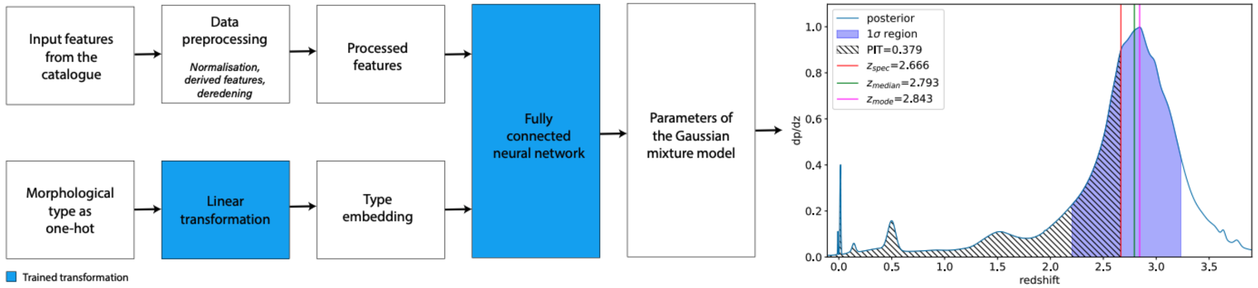


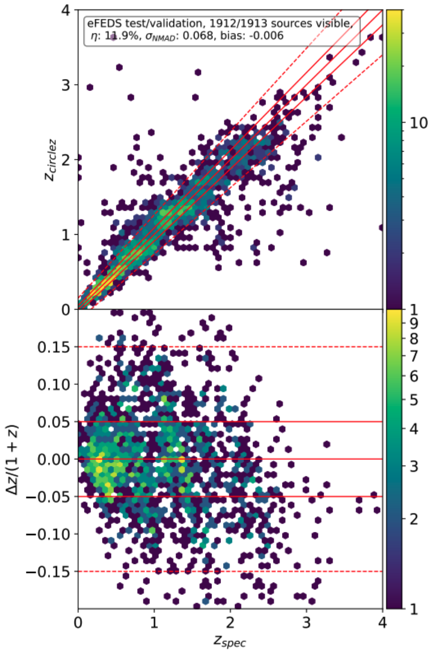
Hatfield et al. (2021)

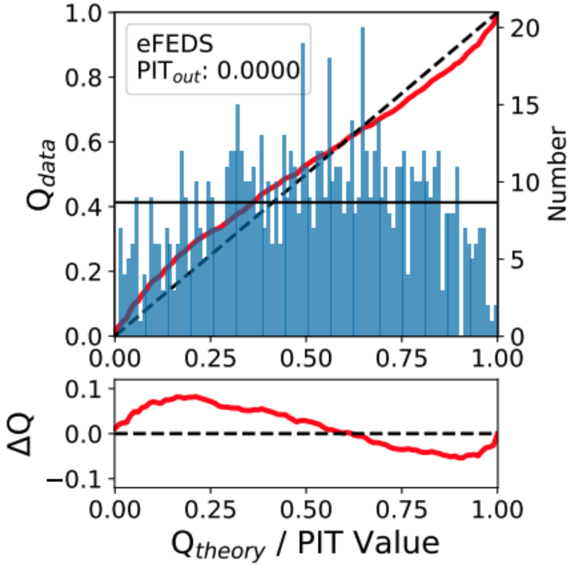
- Combing posteriors from two methods using weights
- Can also simply flag objects with disagreements
- Regions where templates contribute significantly

CIRCLEZ: Using aperture morphology



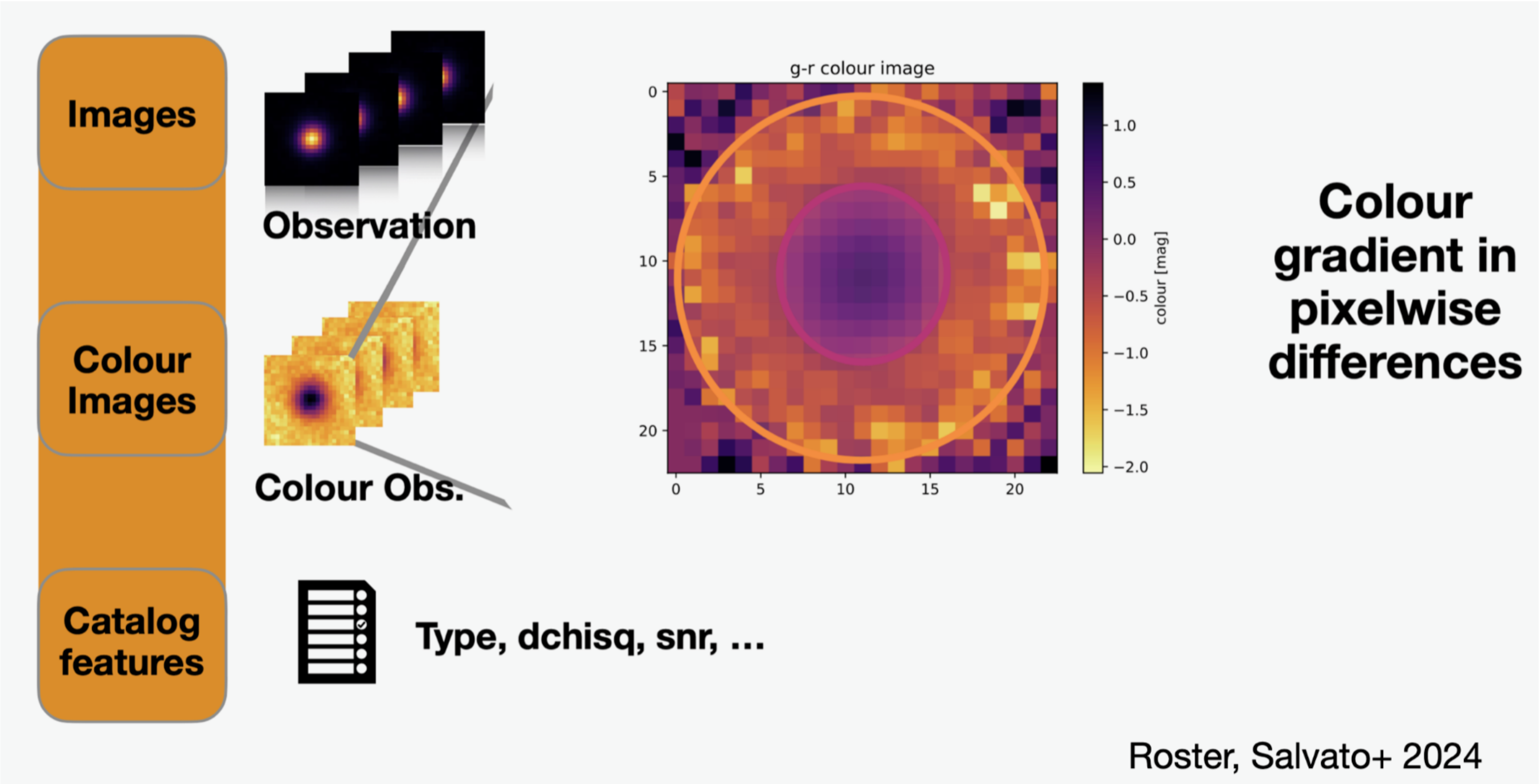




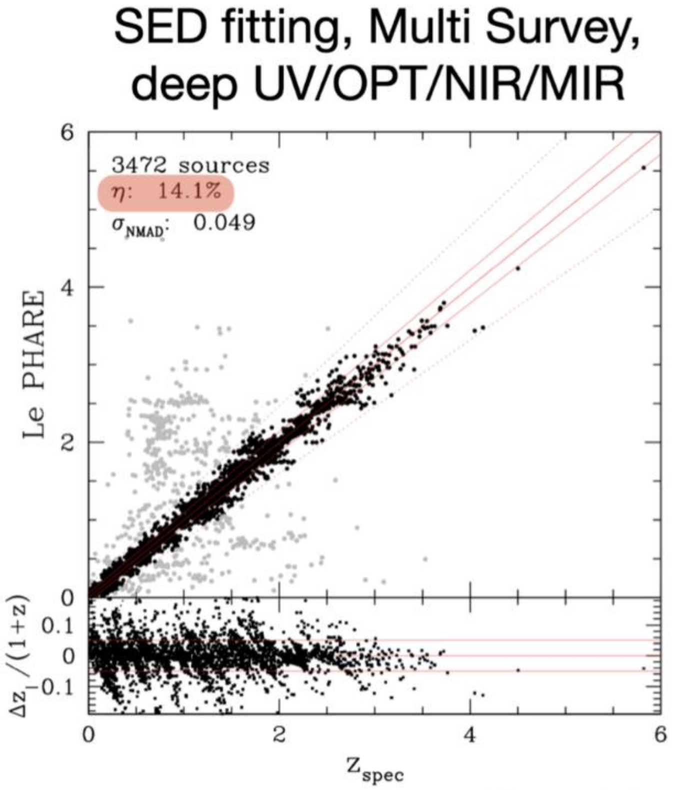


Saxena et al. (2024)

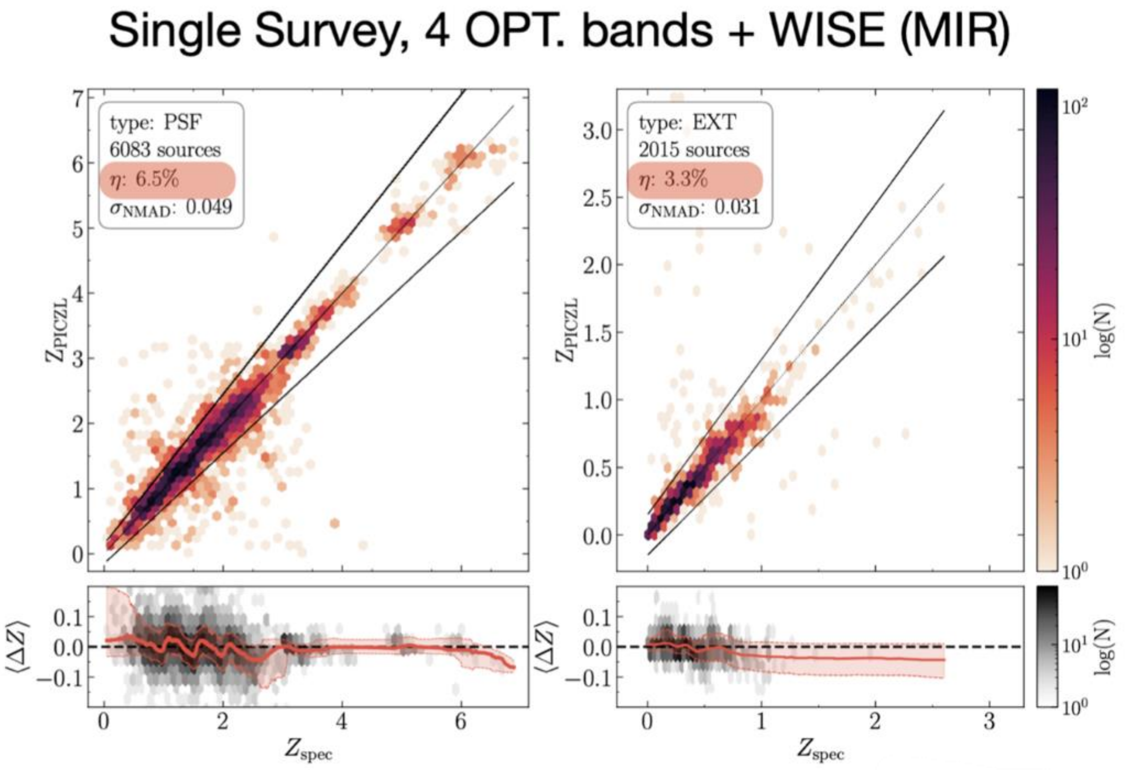
PICZL: Image-based Photometric Redshifts for AGN



Performance with multiwavelength coverage



Salvato et al. (2022)



Roster et al. (2025)

LePHARE

- Originally a fortran code (Arnouts and Ilbert 2011).
- Spectral Energy Distribution (SED) fitting using set of templates from spectroscopic measurements.
- New c++ version https://github.com/lephare-photoz/lephare
- Python interface – pip install lephare
- RAIL interface - https://github.com/LSSTDESC/rail_lephare
- pip install rail_lephare
- Stars, galaxies, and AGN fit seperately.
- Chi squared fitting for each template selecting model with minimum chi squared overall or marginalising over templates.
- Many thanks to LINCC engineers!

Splitting samples for AGN

- Many studies have shown the importance of using different templates for different sources based on X-ray detection and flux, point sourcedness, and other factors (Salvato et al. 2009, 2011, 2018, 2022).
- In the example notebook we will extend the galaxy example and look at AGN samples.

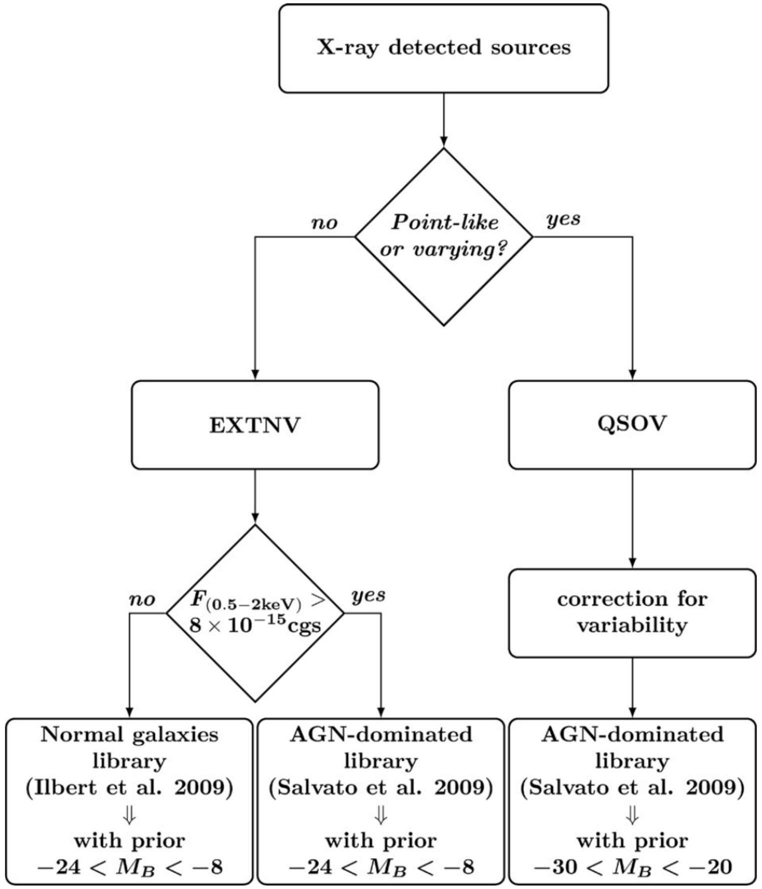
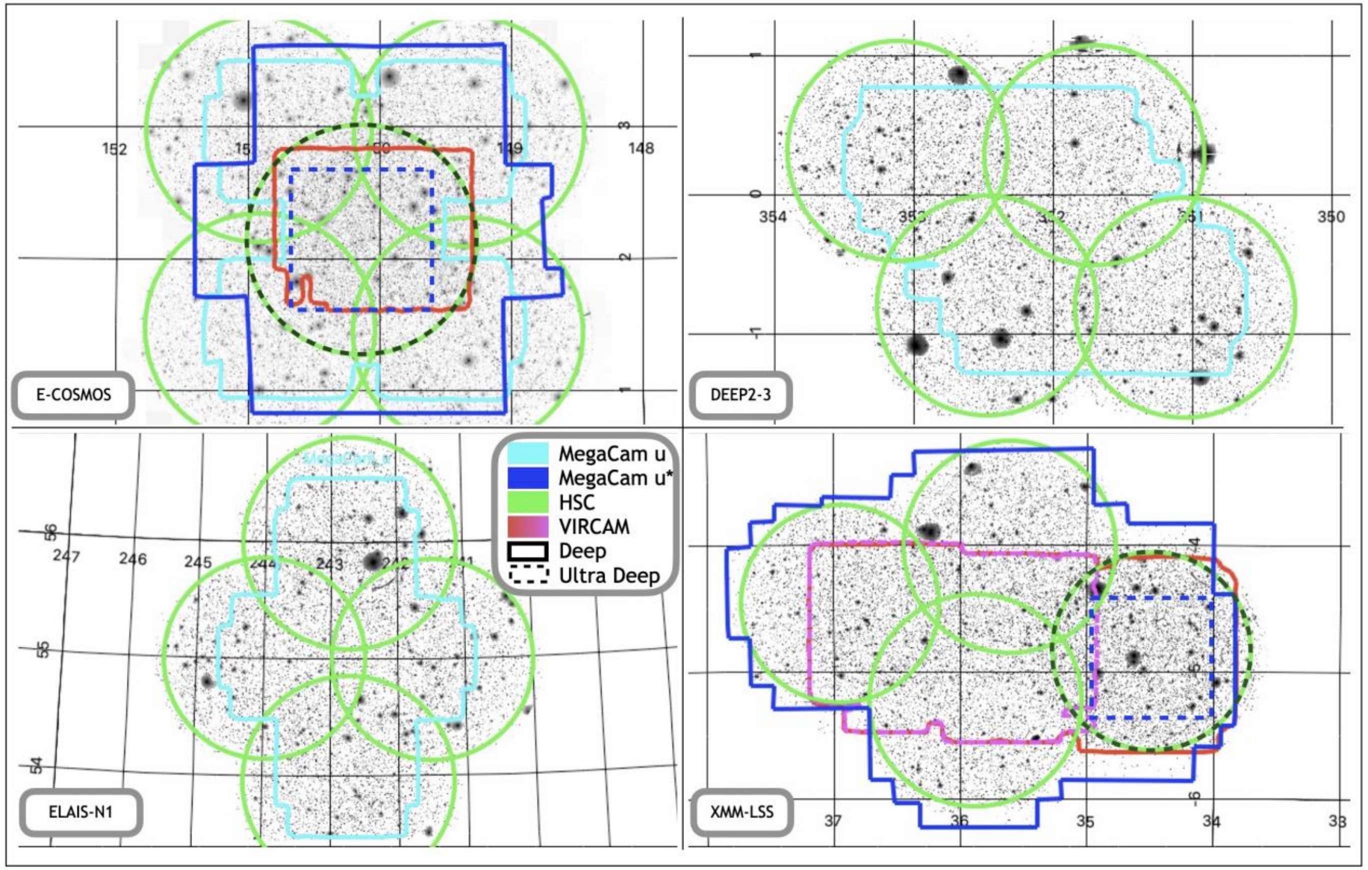


Figure 8. Flow chart of the procedure adopted to compute photo-z for X-ray-detected sources.

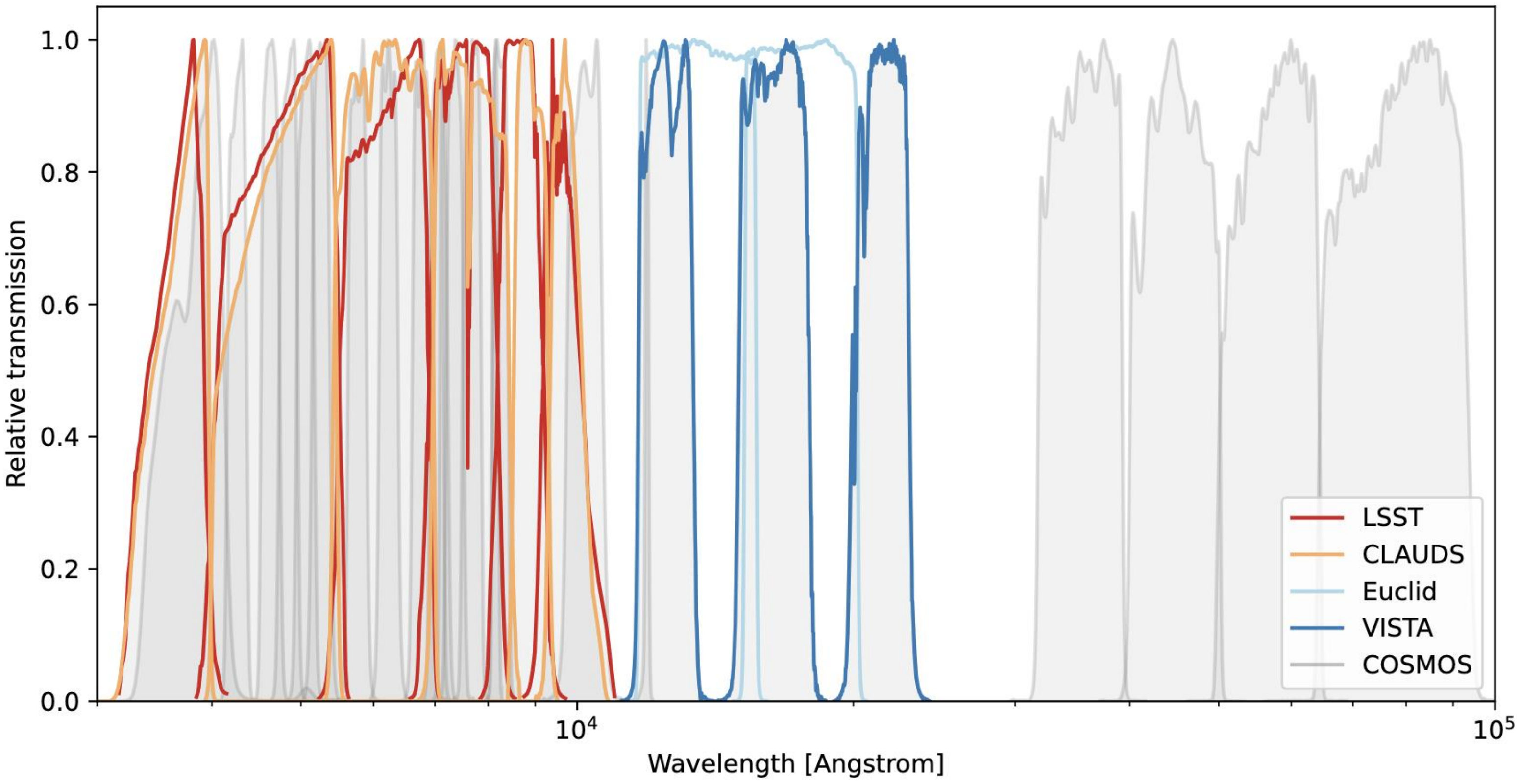
Salvato et al. (2011)

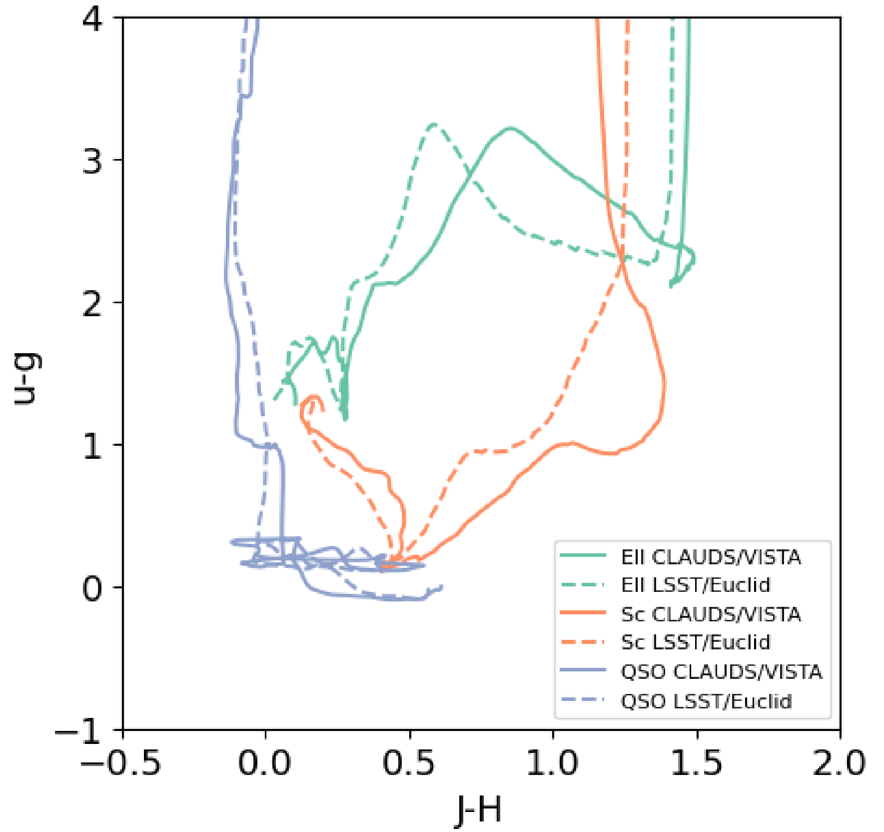
CLAUDS (Desprez et al. 2023)



- CLAUDS (Canada-France-Hawaii Telescope CFHT Large Area U-band survey, Desprez 2023) *ugrizy* data
- We concentrate on the COSMOS field for testing
- Deep Chandra x-ray sample from Marchesi et al. (2016)
- Deep spectroscopic sample from Khostovan et al. (2025)

CLAUDS vs LSST/Euclid





Shirley et al. (in prep)

- Colour information is comparable
- Work shows importance of near-infrared
- Euclid JH significantly different to VISTA
- Photometric redshift accuracy and classification power

AGN samples

- We have provided a broadline sample from Khostovan et al. (2025) and an X-ray sample from Marchesi et al. (2016).
- The X-ray sample is very deep and contains low luminosity AGN.
- We will look at the performance of each template set and the effect of absolute magnitude prior on performance.

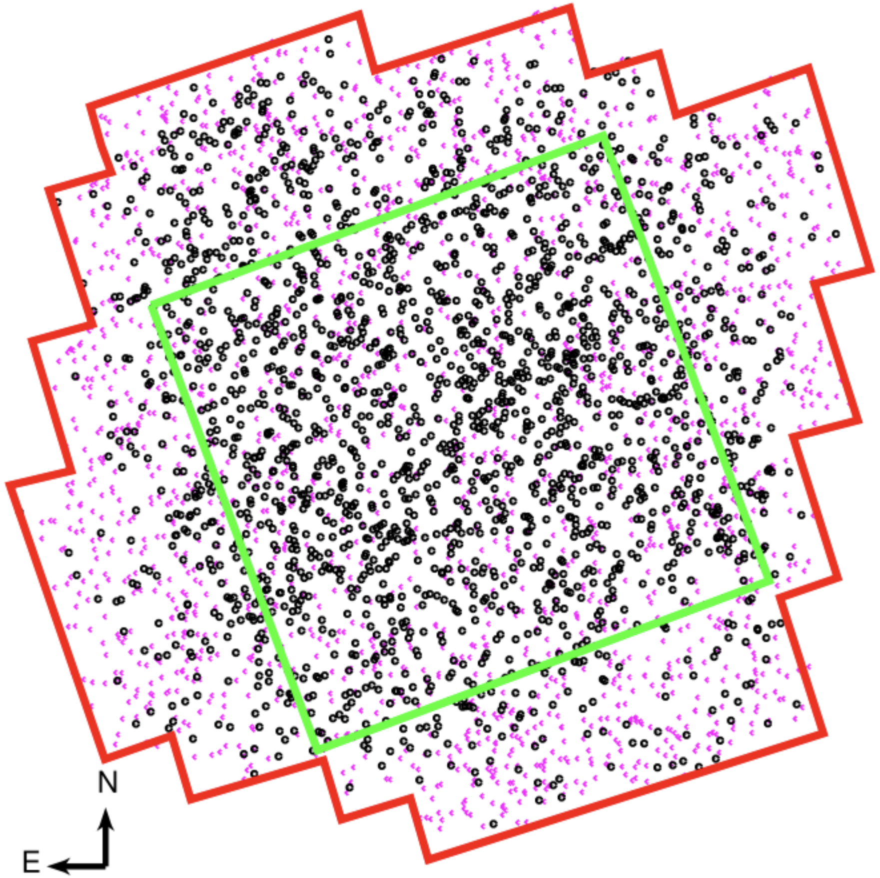


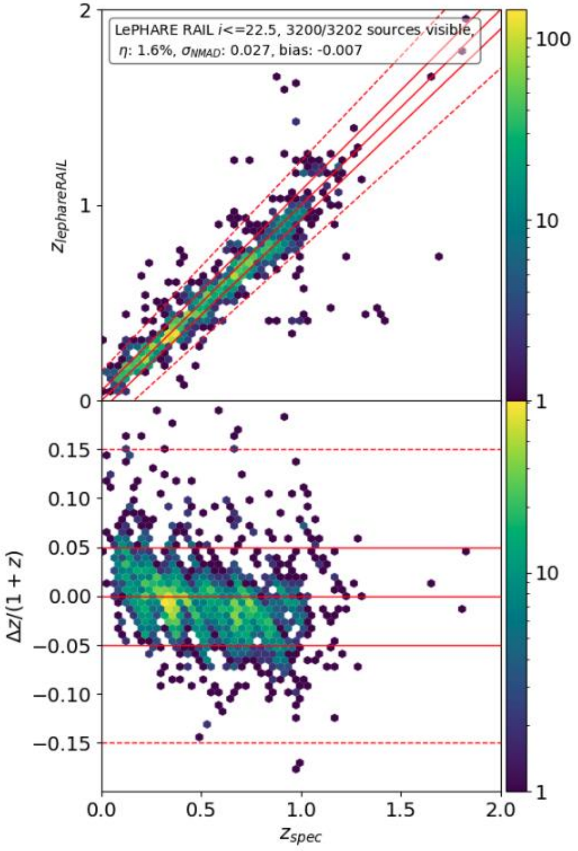
Figure 5. Sources with (black circles) and without (magenta circles) spectroscopic redshift in the *Chandra COSMOS-Legacy* area (red solid line). The C-COSMOS area is also plotted (green solid line). A significant fraction of sources in the external part of the field has not been spectroscopically followed-up yet.

Marchesi et al. (2016)

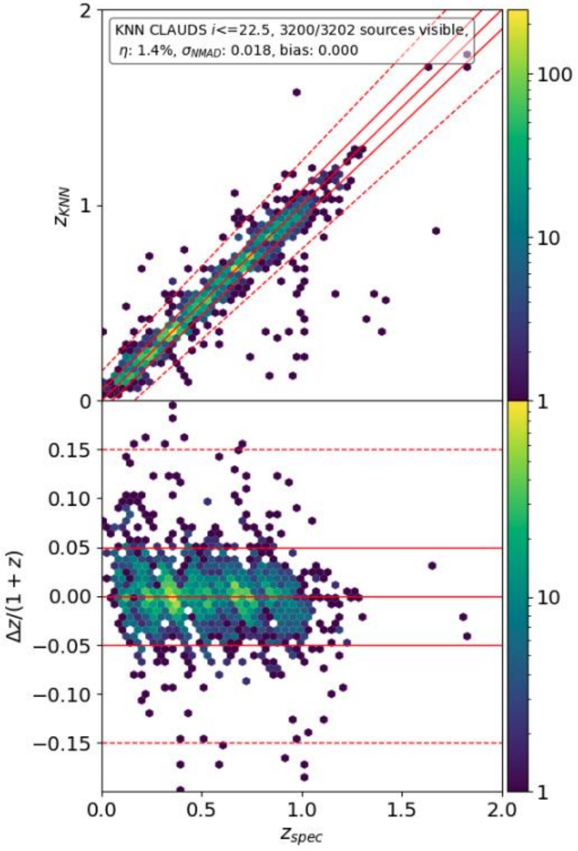
HSC-CLAUDS galaxy performance

All ML codes trained on all spectroscopic objects regardless of broadline features.

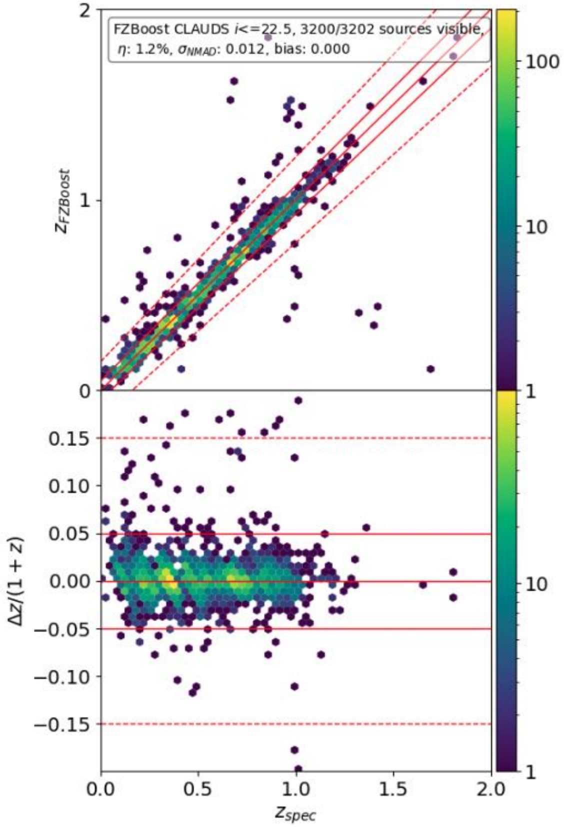
LePHARE



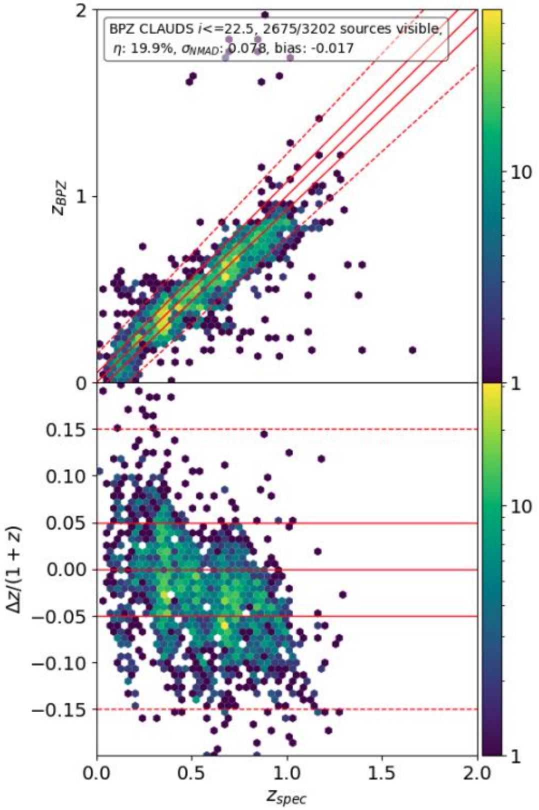
KNN



FZBoost



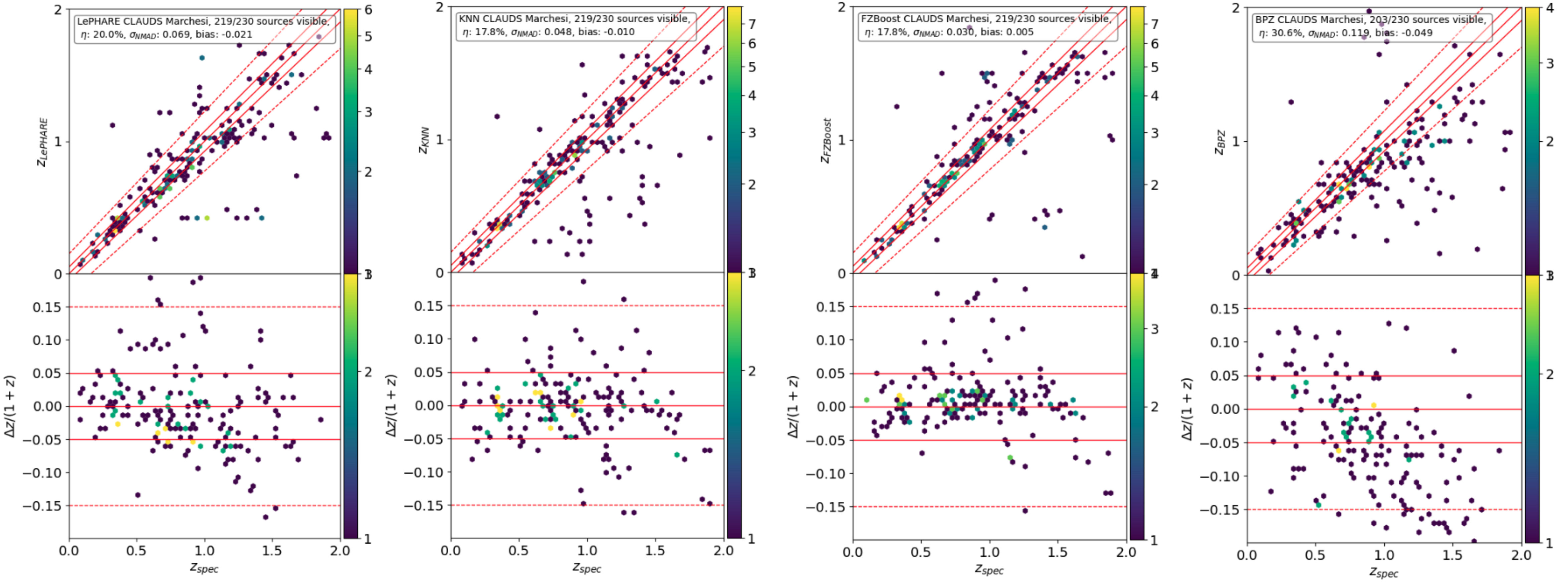
BPZ



Shirley et al. (in prep)

HSC-CLAUDS X-ray AGN performance (Marchesi et al. 2016)

LePHARE KNN FZBoost BPZ



Shirley et al. (in prep)

Exercises 1 – Check you can run the notebooks

- The notebook is similar to the Galaxy example. We use a sample from HSC-CLAUDS.
- pip install lephare
- git clone https://github.com/raphaelshirley/lephare-examples.git
- Conda environement issues?

```
[1]: # lephare must be installed if not already
     #!pip install lephare
```

```
[2]: import os
     import lephare as lp
     from astropy.table import Table
     import numpy as np
     from matplotlib import pylab as plt
     import yaml
```

```
LEPHAREDIR is being set to the default cache directory:
/Users/rshirley/Library/Caches/lephare/data
More than 1Gb may be written there.
LEPHAREWORK is being set to the default cache directory:
/Users/rshirley/Library/Caches/lephare/work
Default work cache is already linked.
This is linked to the run directory:
/Users/rshirley/Library/Caches/lephare/runs/20240516T005248
```

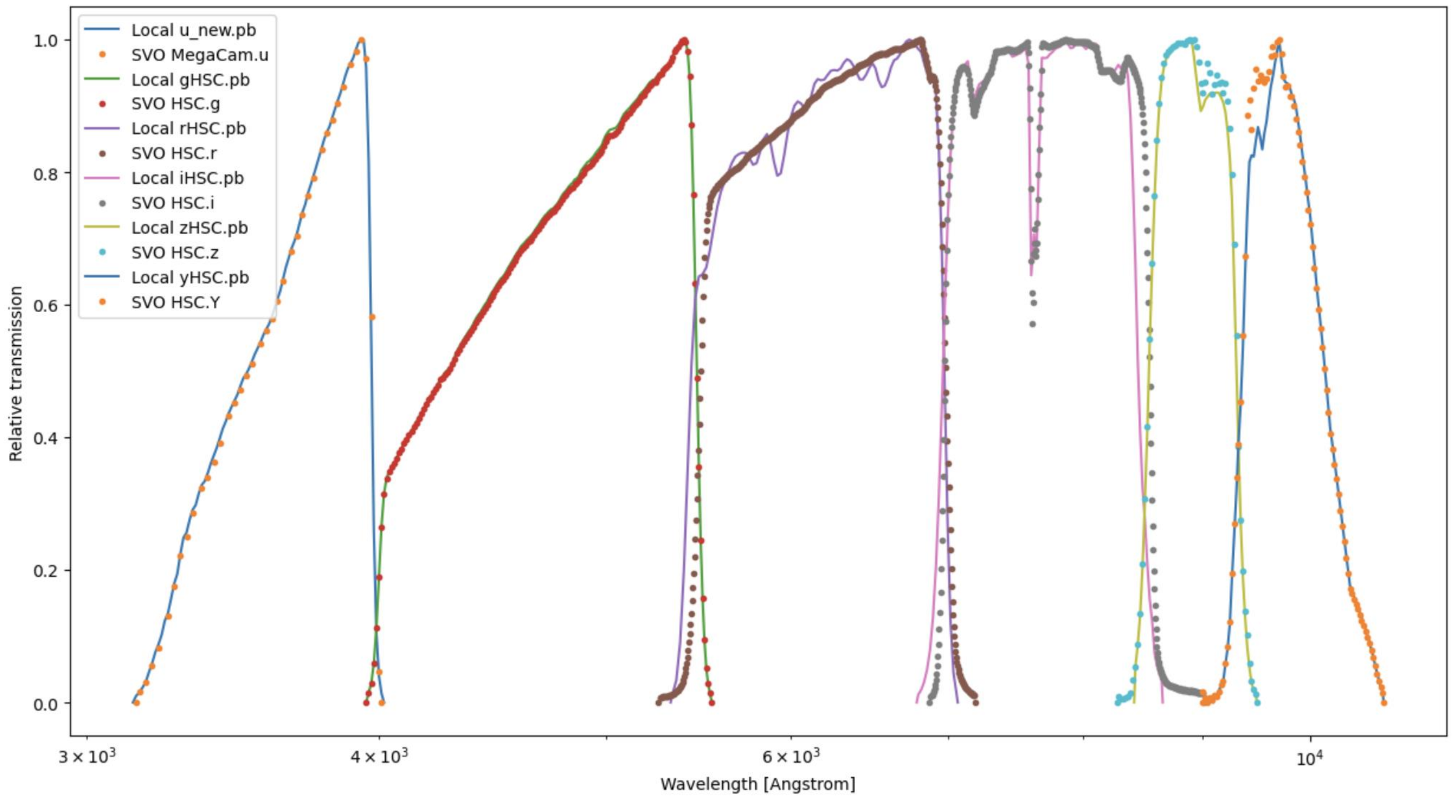
Load the example data

In the documentation example we were looking at COSMOS data. Here we are looking at a d
the COSMOS data so some parts of the configuration are identical.

```
[3]: input_lp=Table.read('./data/input_lp_agnsample_20251030.fits')
```

Exercises 2 – Compare filters from lephare data and SVO

- Can you get some more filters from the SVO?
- https://svo2.cab.inta-csic.es/theory/fps/index.php?mode=browse
- How different are the HSC-CLAUDS filters to LSST?



Exercise 3 – Run lephare and check the outputs

- How do the point estimates differ for different samples and different lephare config values?
- Can you recommend a given estimate for each sample?
- How do the point estimates relate to the posterior?

$$\left|\frac{z_{\rm phot} - z_{\rm spec}}{1 + z_{\rm spec}}\right| > 0.15, \tag{1}$$

giving the outlier fraction, η, as

$$\eta = n_{\rm outliers}/n_{\rm total} \tag{2}$$

We also look at the standard deviation estimated from the normalized median absolute deviation, (NMAD; Hoaglin et al. 2000), because it is less sensitive to outliers than the typical definition:

$$\sigma_{\rm nmad} = 1.48 \times {\rm median}\left|\frac{z_{\rm phot} - z_{\rm spec}}{1 + z_{\rm spec}}\right| \tag{3}$$

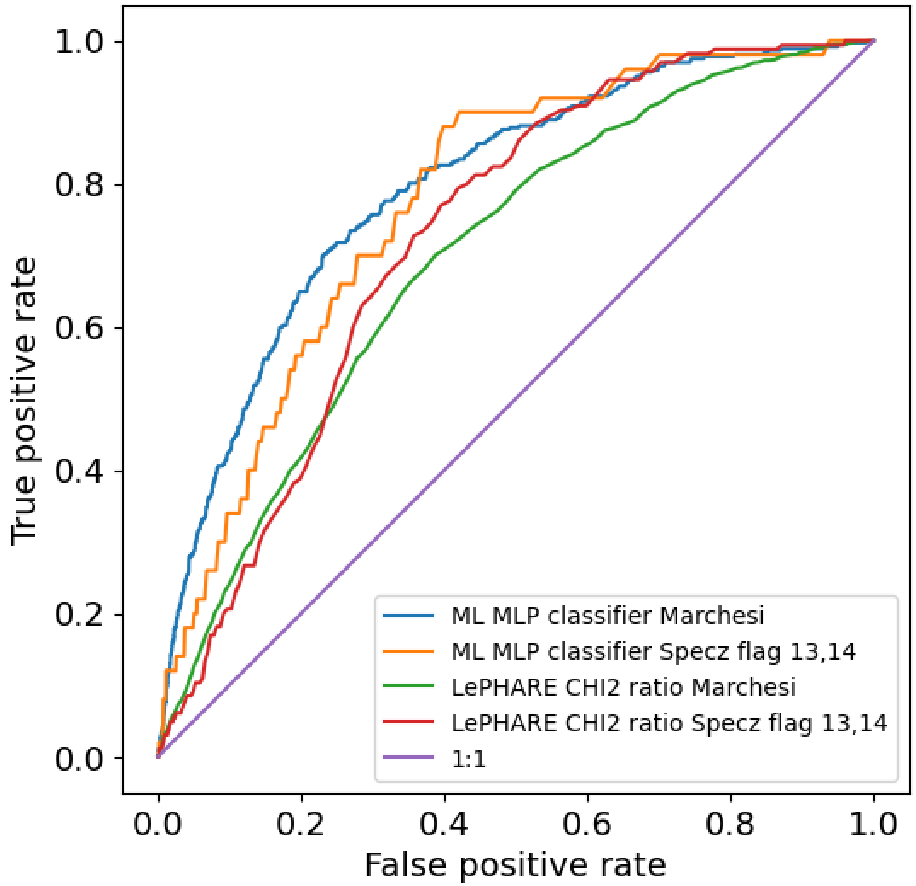
and the bias

$$\beta = {\rm median}\left(\frac{z_{\rm phot} - z_{\rm spec}}{1 + z_{spec}}\right) \tag{4}$$

Because $\sigma_{\rm nmad}$ will be impacted by β we can also compute an unbiased NMAD estimator as

$$\sigma_{\rm nmad,unbiased} = 1.48{\times}{\rm median}\left|\frac{z_{\rm phot} - z_{\rm spec} - {\rm median}(z_{\rm phot} - z_{\rm spec})}{1 + z_{\rm spec}}\right| \tag{5}$$

AGN classification – ROC curves



Shirley et al. (in prep)

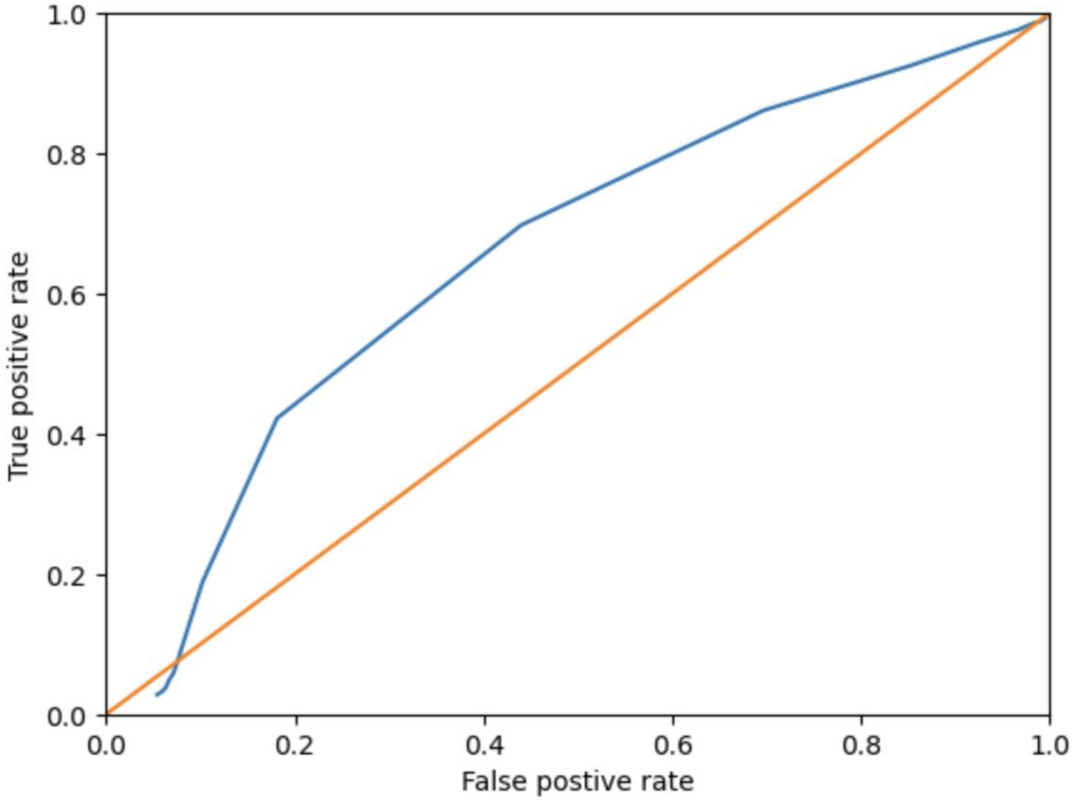
- We have two AGN samples:
  - Khostovan broadline
  - Marchesi x-ray selected
  - Some overlap
- LePHARE produces comparable classification performance to a multilayer perceptron.
- Additional utility of template fitting run

AGN classification

- We have implemented a simple ROC curve in the notebook
- Calculate True Positive Rate and False Positive Rate
- If there is time - How does it compare to a Multi Layer Perceptron classifier?
- https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

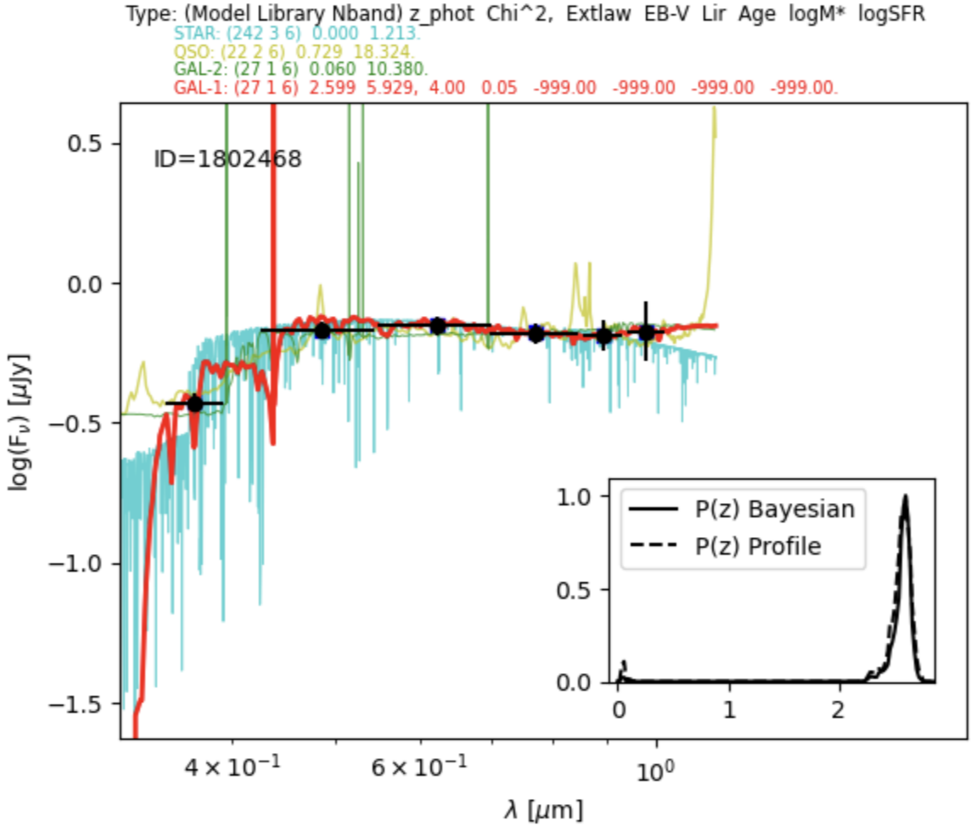
$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

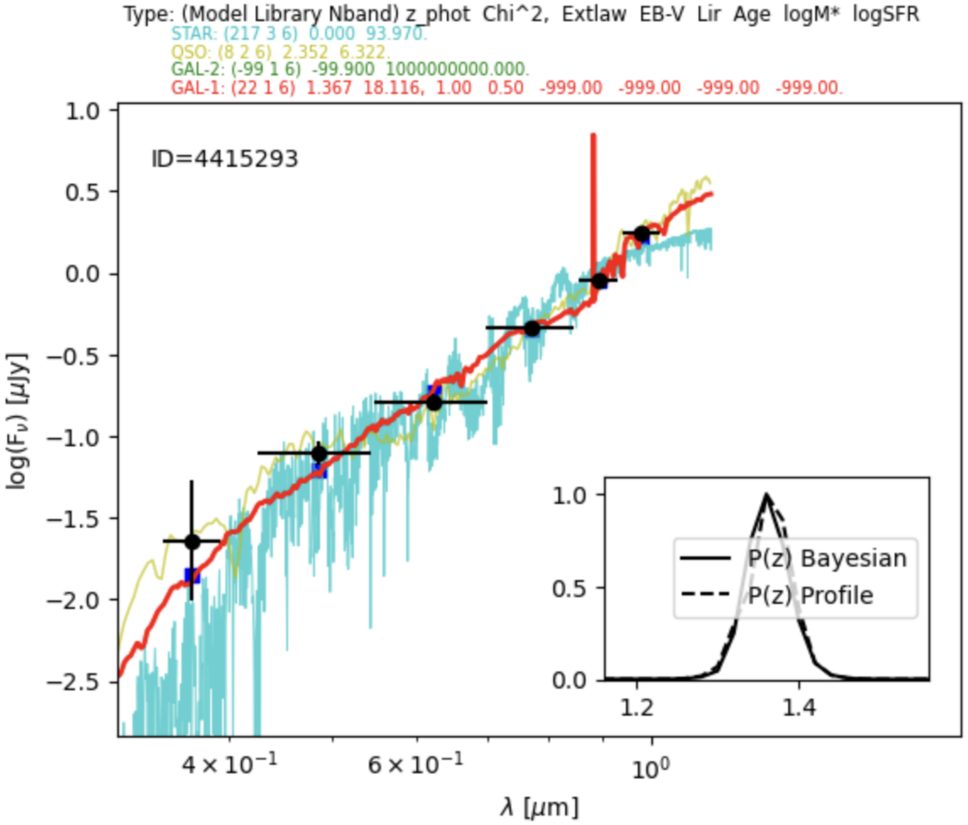


Example SEDs

Galaxy (Khostovan et al. 2025)



X-ray AGN (Marchesi et al. 2016)



Conclusions

- Machine learning and template fitting will both be useful for AGN science.
- Template fitting requires careful division of samples and comparisons of performance of different template sets
- New version of LePHARE https://github.com/lephare-photoz/lephare
- pip install lephare
- Experiment with configuration parameters
- Investigate the classification power
- Chi squared can be used to determine which template set to use.