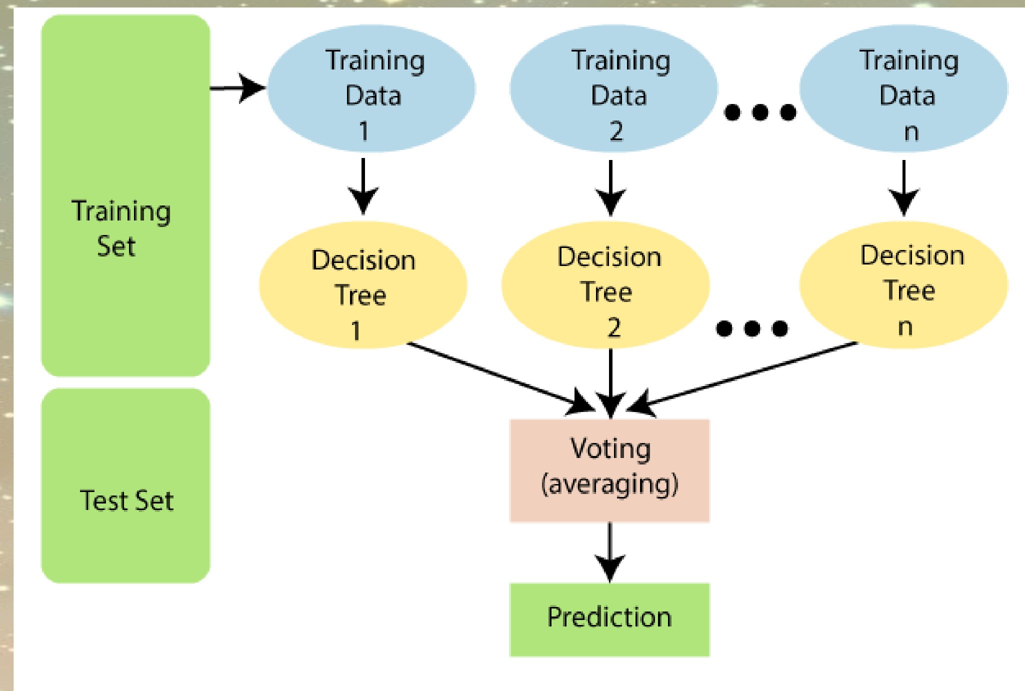
Using Machine Learning Techniques to Determine Photometric Redshifts for Gravitational wave Cosmology

Lara Janiurek , Supervised by Prof Martin Hendry

The inference of the Hubble constant using gravitational wave data has allowed for a new way for the expansion of the universe to be probed. The use of binary black hole merger events to measure the Hubble constant may shed considerable light on the current Hubble tension. Galaxy redshift surveys are a key ingredient for the application of these dark sirens in the measurement of H_0.

- Improving the performance of dark sirens requires a better understanding of the photometric redshift errors.
- Current redshift values used by LIGO for cosmological inference are assumed to have an associated Gaussian error.
- Quantification of the redshift posteriors would give a more accurate result in the overall.
- Photometric redshift surveys often contain significant errors.
- Spectroscopic surveys are expensive and rely on cosmological models
- Machine learning techniques are advantageous in that they don't rely on these models.

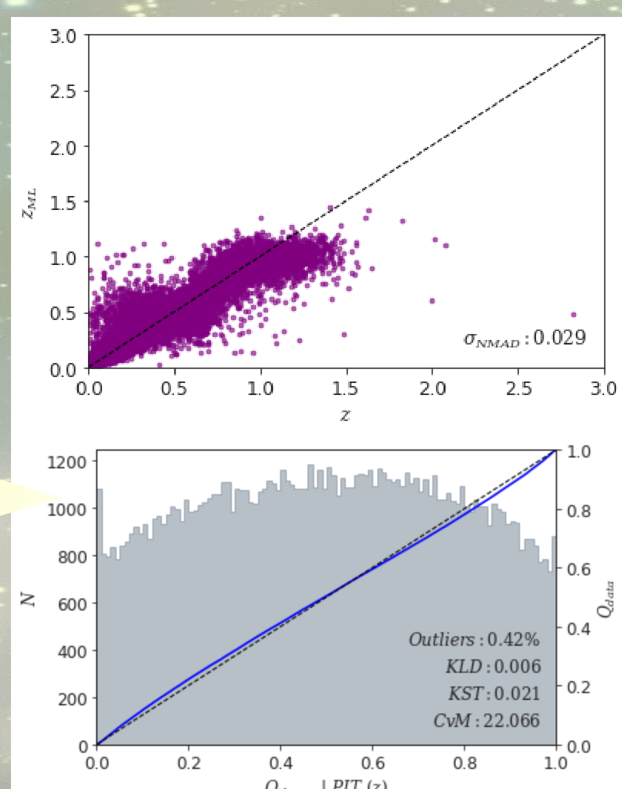
Contact
lara.Janiurek@strath.ac.uk
Supervisor: Martin Hendry
Martin.Hendry@glasgow.ac.uk



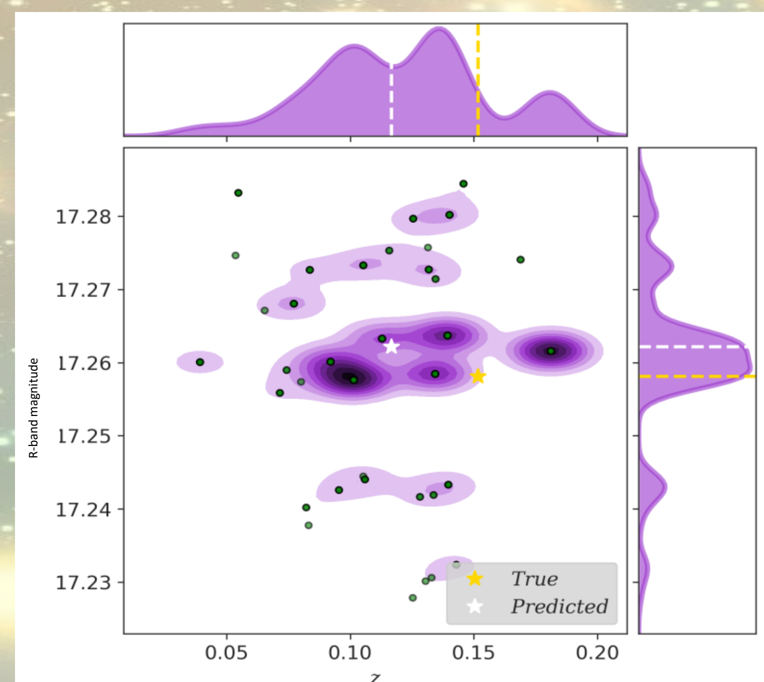
A simplified diagram of how a random forest algorithm learns mapping and predicts outcomes [1].

In this work, the random forest algorithm GALPRO is implemented to generate photometric redshift posteriors. It is initially calibrated and optimized using a truth dataset compiled by Zhou et al.

The random forest algorithm learns the mapping between the photometry data and spectroscopic redshift of each galaxy. This learnt mapping can then be applied to a new dataset which only has photometry data.



The photometric versus spectroscopic redshift and PIT plot of GALPRO when trained and tested using subsets of the Zhou et al dataset.

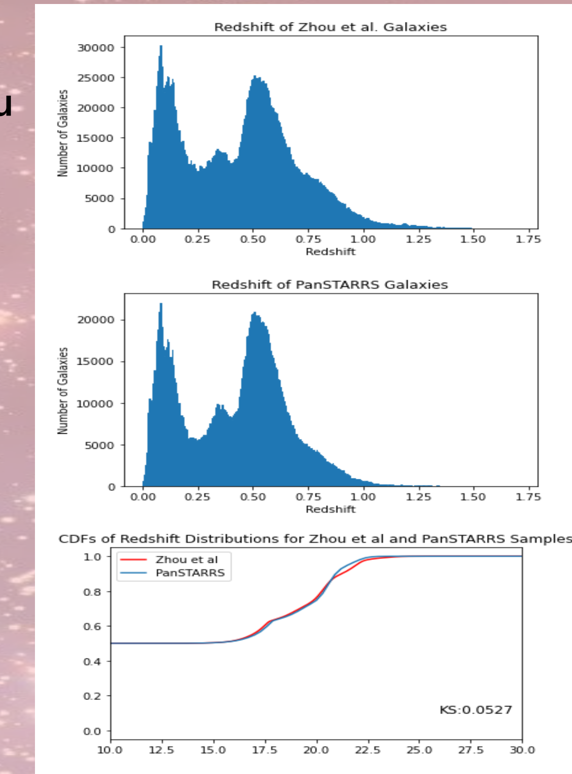


A joint posterior distribution of a randomly selected galaxy from the Zhou et al dataset. The redshift posterior is shown on the top and the r-band magnitude posterior is shown on the right.

The analysis suggests that the redshift posterior distributions are largely non-Gaussian, reinforcing the need for a reliable method to generate redshift posteriors to better represent these photometric errors.

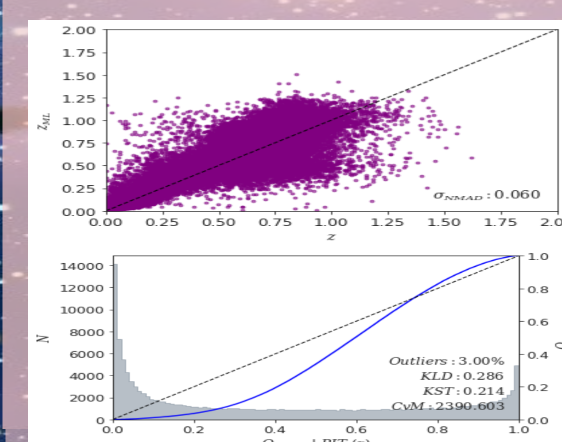
The initial calibration showed that GALPRO is very useful when provided with an incomplete galaxy survey with missing spectroscopic values, as missing redshift can be accurately predicted.

- Tests were run using the Zhou et al. dataset to determine how statistically similar the training and testing datasets must be for GALPRO to be applicable.
- The training and testing datasets were found to require similar redshift distributions and overlap by at least 90% in the band ranges to give accurate results.
- GALPRO was then trained using the Zhou et al. dataset and applied to a sample from the PanSTARRS survey to explore if GALPRO could be trained using a trusted dataset and applied to a general, new survey.



The redshift distributions and their compared CDFs for the PanSTARRS and Zhou datasets.

Despite this, application of the algorithm still resulted in a catastrophic failure, indicating that there must be some underlying fundamental difference between the two surveys that cause the program to not learn the correct mapping. This serves as a cautionary tale in the application of random forests to new surveys when generating photometric redshift posteriors.



The photometric versus spectroscopic redshift plot and PIT of the PanSTARRS sample when trained using the Zhou et al dataset. The scatter and PIT demonstrate a failure in the learnt mapping of the algorithm.

References
[1]https://www.javatpoint.com/machine-learning-random-forest-algorithm
[2] R Gray. Gravitational wave cosmology: measuring the Hubble constant with dark standard sirens. PhD thesis, University of Glasgow, 2021.

- Producing galactic spectrums is expensive and uses a lot of energy.
- These posteriors are used in many areas of astronomy.
- Being able to produce them using machine learning algorithms would decrease the energy cost of production and therefore help astronomy and the planet!