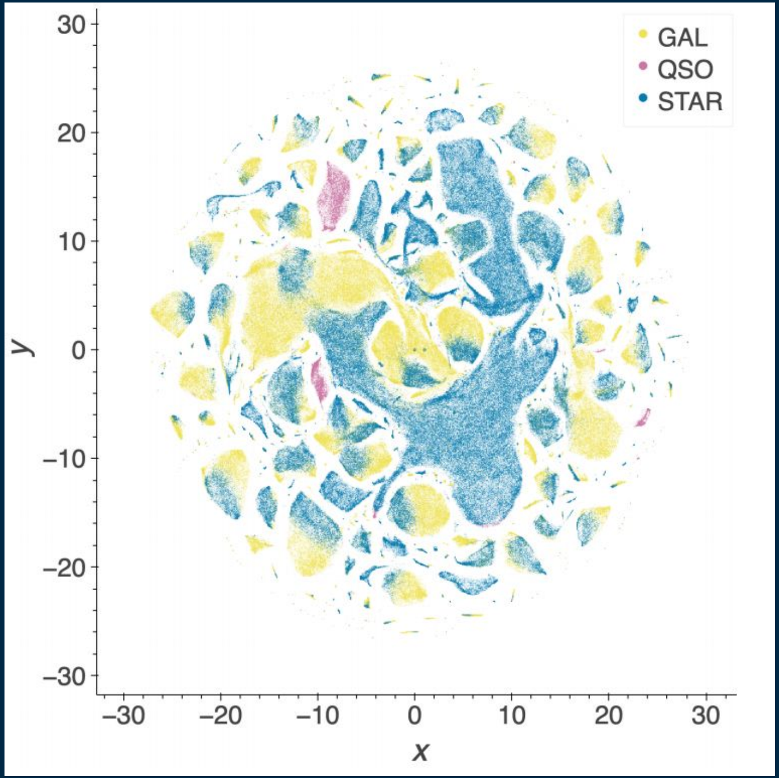
Can we extend photo-z estimations to quasars?

Raquel Ruiz Valença
Collaborators: Lilianne Nakazono and Erik Vinicius
Advisor: Claudia L. M. de Oliveira
IAG/USP

Conselho Nacional de Desenvolvimento Científico e Tecnológico

INTRODUCTION

- Photometric redshifts have been obtained for S-PLUS galaxies in the Local Universe (z < 0.7) [Lima et al. 2021] with an error of 1.7%.
- Our goal is to extend the analysis to quasars (z < 5), also using a machine learning approach.
- With the S-PLUS classification by Nakazono et al. 2021, we will be able to build a catalog of quasar candidates with their redshift predictions.



Nakazono et al. 2021

Which questions do we want to answer?

1. Can we improve quasar photo-z estimations using the S-PLUS narrow bands?

2. Can we improve quasar photo-z estimations at high redshift?

DATA

CROSS-MATCHES

S-PLUS × SDSS × GALEX

(iDR3 - Stripe 82) **(DR16Q)** **(GR6+7)**

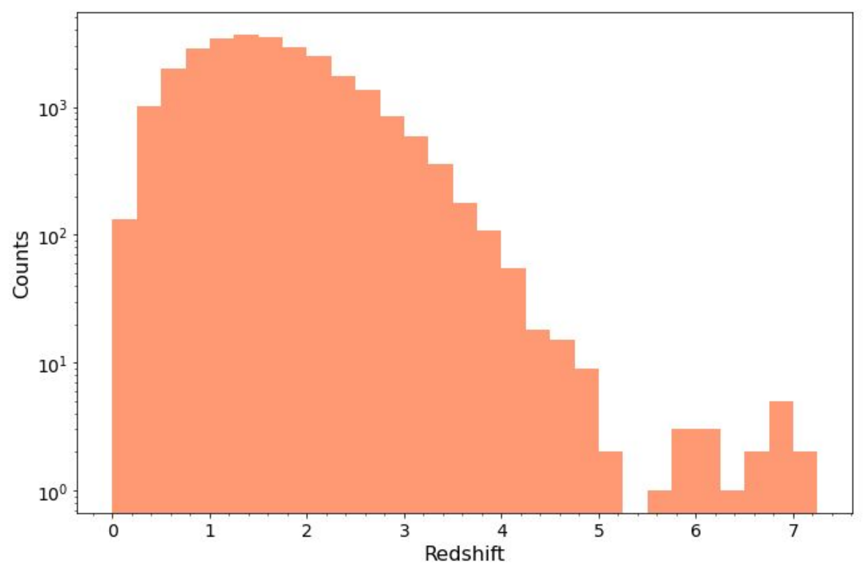
<u>1'' matching-radius</u>
Quasars
2 WISE magnitudes

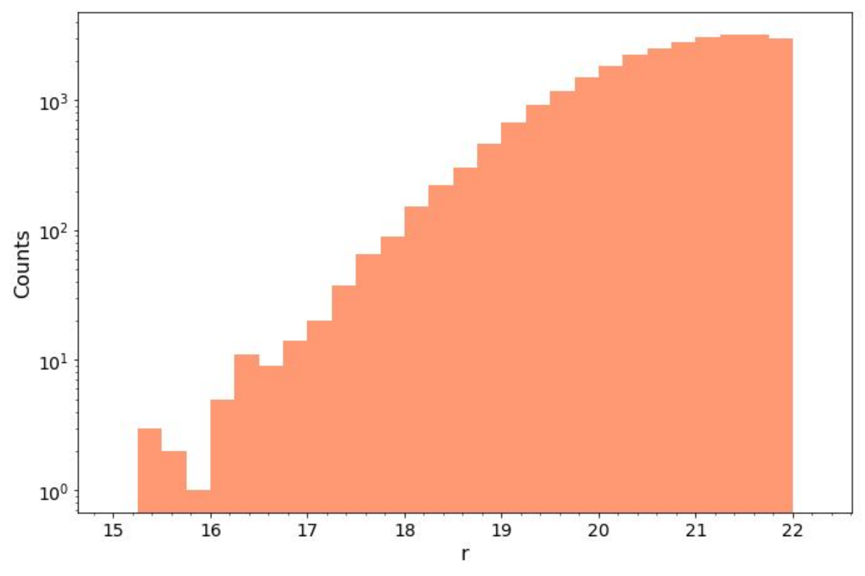
<u>2'' matching-radius</u>
2 GALEX magnitudes

SAMPLES

Total sample: 27,337 quasars, r ≤ 22

Without missing bands: 3,506 quasars





METHODS

ALGORITHMS

01

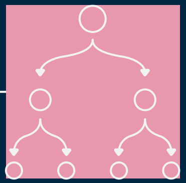
Linear Regression

Finds coefficients for a function (line or polynomial) by Ordinary Least Squares.

02

Lasso Regression

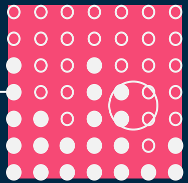
Linear Regression but with AIC or BIC penalties on the residual sum of squares.



03

Random Forest

Based on the mean of multiple decision trees.
[Breiman, 2001]



04

K-Nearest Neighbors

Finds the k closest neighbors in an euclidian space.
[Fix and Hodgers, 1961]

ALGORITHMS

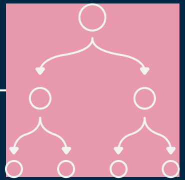
01
Linear Regression

$y = a_0 + \sum_{i=1}^{n} a_i \; x_i$

02
Lasso Regression

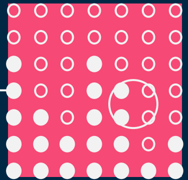
$y = a_0 + \sum_{i=1}^{n} a_i \; x_i$

criterion = AIC, BIC



03
Random Forest

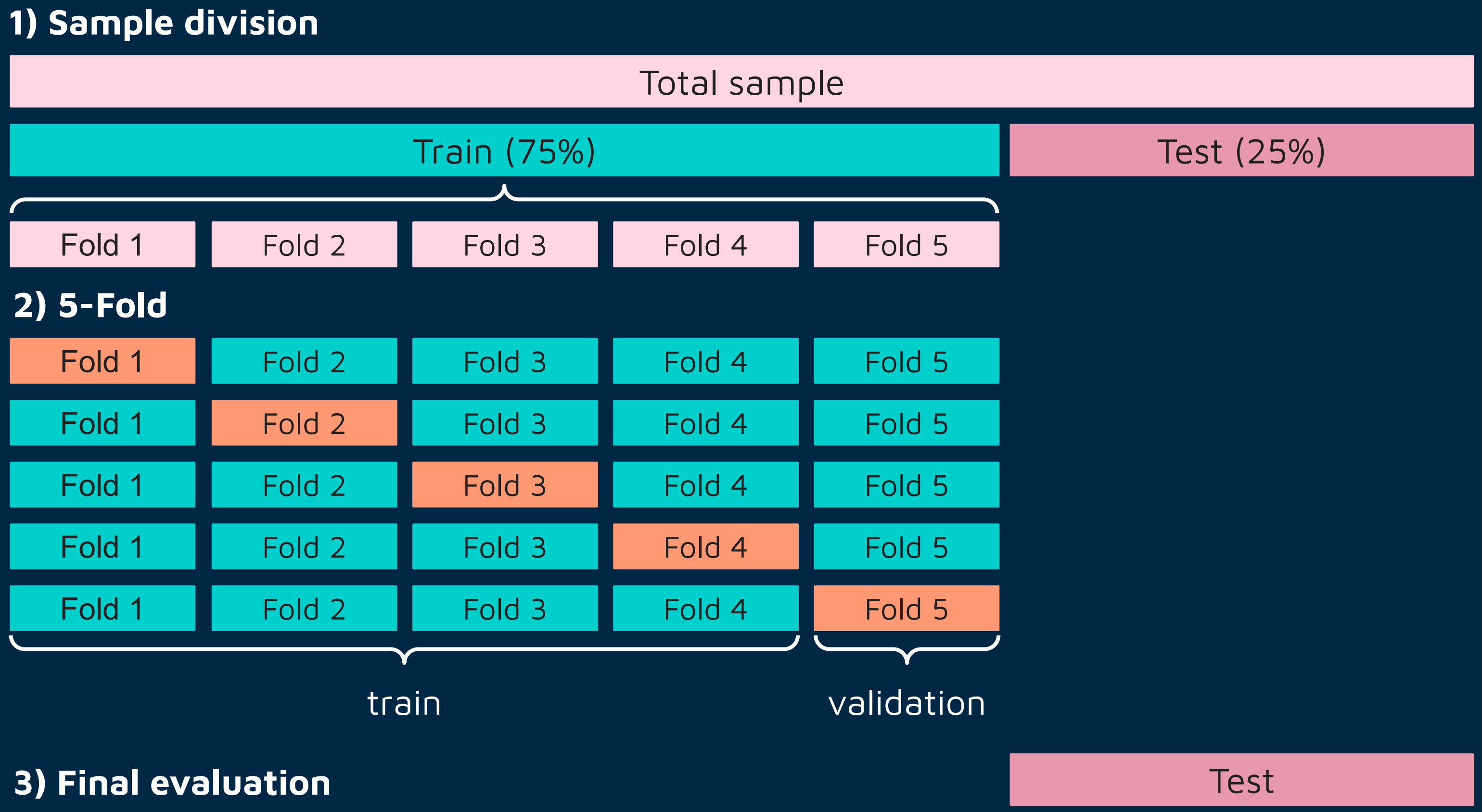
n_estimators = 100



04
K-Nearest Neighbors

n_neighbors = 3,5,7
distance = euclidian, uniform

CROSS-VALIDATION



METRICS

Root mean square error

$$\sigma_{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \delta_{z_i}{}^2}$$

Normalized median absolute deviation

$$\sigma_{NMAD} = 1.48 \cdot \mathrm{median}\left(\frac{|\delta z - \mathrm{median}(\delta z)|}{1 + z_{spec}}\right)$$

FEATURES

- **Broad bands:** u, g, r, i, z
- **Narrow bands:** J0378, J0395, J0410, J0430, J0515, J0660, J0861
- **WISE bands:** W1, W2
- **GALEX bands:** FUV, NUV

	BROAD	NARROW	WISE	GALEX
	✓	✓	✓	✓
	✓	✓	✓	
	✓	✓		✓
	✓	✓		
	✓		✓	✓
	✓		✓	
	✓			✓
	✓	✓	✓	
		✓	✓	✓
		✓		
		✓		✓

FEATURES

- **Broad bands:** u, g, r, i, z
- **Narrow bands:** J0378, J0395, J0410, J0430, J0515, J0660, J0861
- **WISE bands:** W1, W2
- **GALEX bands:** FUV, NUV

BROAD	NARROW	WISE	GALEX
✓	✓	✓	✓
✓	✓	✓	
✓	✓		✓
✓	✓		
✓		✓	✓
✓		✓	
✓			✓
✓	✓	✓	
	✓	✓	✓
	✓		
	✓		✓

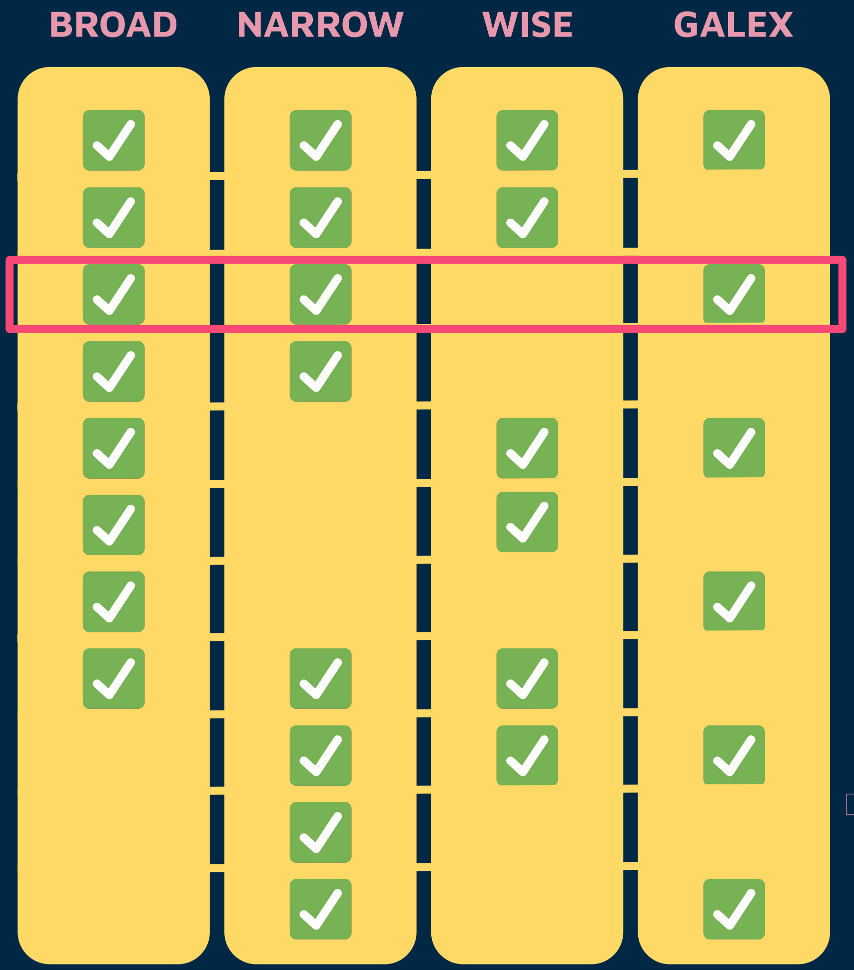
FEATURES

- **Broad bands:** u, g, r, i, z
- **Narrow bands:** J0378, J0395, J0410, J0430, J0515, J0660, J0861
- **WISE bands:** W1, W2
- **GALEX bands:** FUV, NUV

BROAD	NARROW	WISE	GALEX
✓	✓	✓	✓
✓	✓	✓	
✓	✓		✓
✓	✓		
✓		✓	✓
✓		✓	
✓			✓
✓	✓	✓	
	✓	✓	✓
	✓		
	✓		✓

FEATURES

- **Broad bands:** u, g, r, i, z
- **Narrow bands:** J0378, J0395, J0410, J0430, J0515, J0660, J0861
- **WISE bands:** W1, W2
- **GALEX bands:** FUV, NUV



RESULTS

BEST 5 MODELS (OUT OF 816)

	σ_{NMAD}	σ_{RMSE}	Broad	Narrow	WISE	GALEX	Colors
792	0.097	0.434	✅	✅	✅	✅	✓
796	0.106	0.448	✅		✅	✅	✓
768	0.114	0.449	✅	✅	✅	✅	
794	0.105	0.454	✅	✅	✅		✓
772	0.119	0.459	✅	✅	✅		

BEST 5 MODELS (OUT OF 816)

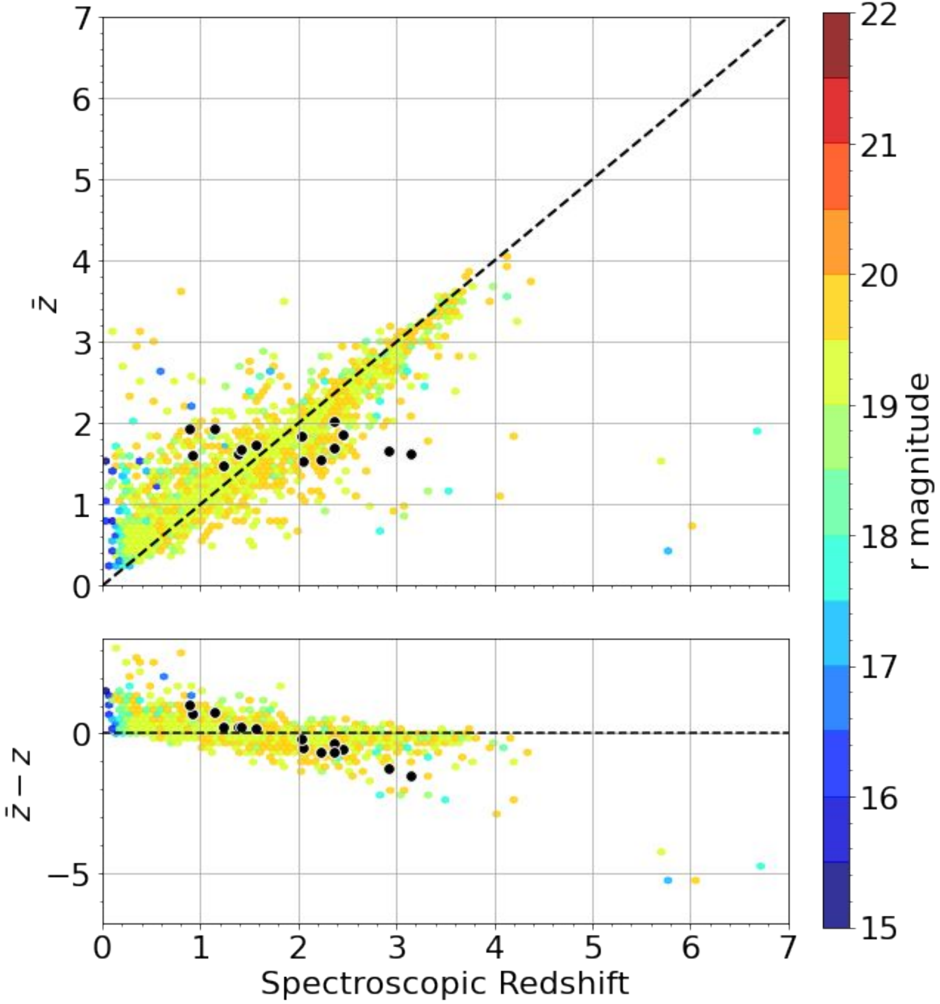
	σ_{NMAD}	σ_{RMSE}	Broad	Narrow	WISE	GALEX	Colors
792	0.097	0.434	✓	✓	✓	✓	✓
796	0.106	0.448	✓		✓	✓	✓
768	0.114	0.449	✓	✓	✓	✓	
794	0.105	0.454	✓	✓	✓		✓
772	0.119	0.459	✓	✓	✓		

BEST 5 MODELS (OUT OF 816)

	σ_{NMAD}	σ_{RMSE}	Broad	Narrow	WISE	GALEX	Colors
792	0.097	0.434	✅	✅	✅	✅	✓
796	0.106	0.448	✅		✅	✅	✓
768	0.114	0.449	✅	✅	✅	✅	
794	0.105	0.454	✅	✅	✅		✓
772	0.119	0.459	✅	✅	✅		

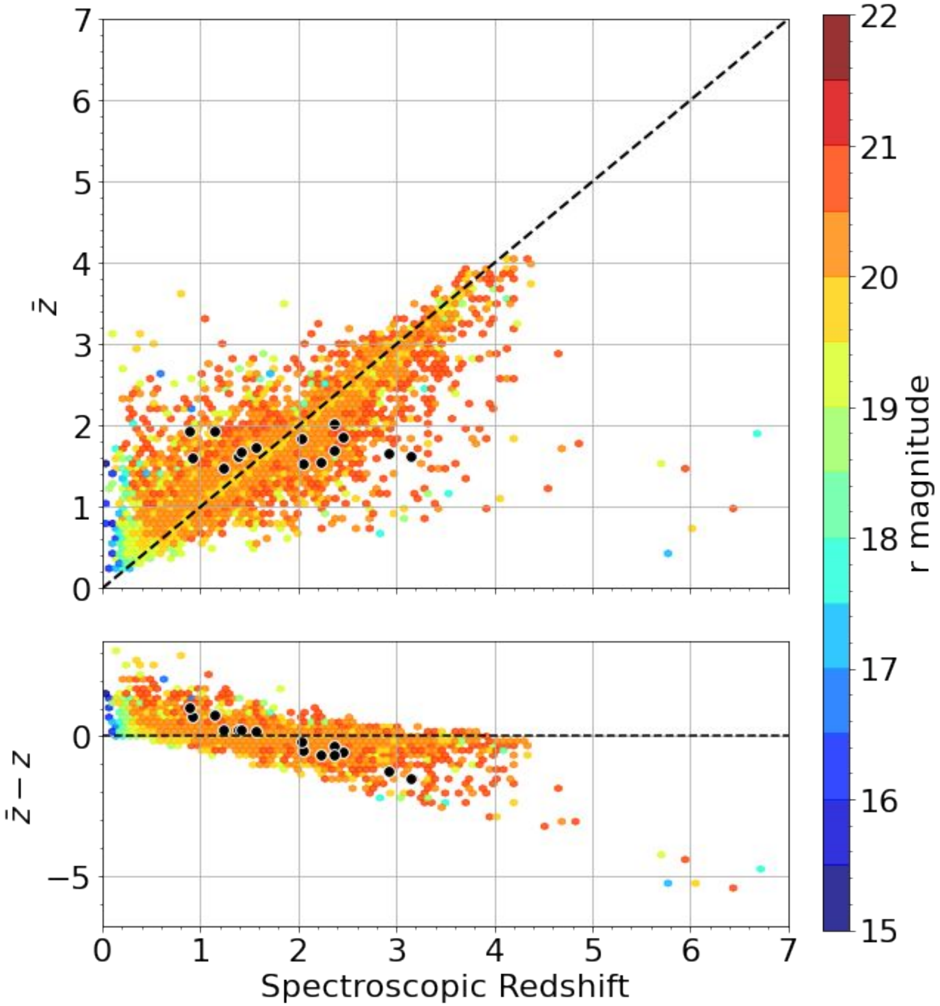
Best model (792)

↪ Residuals



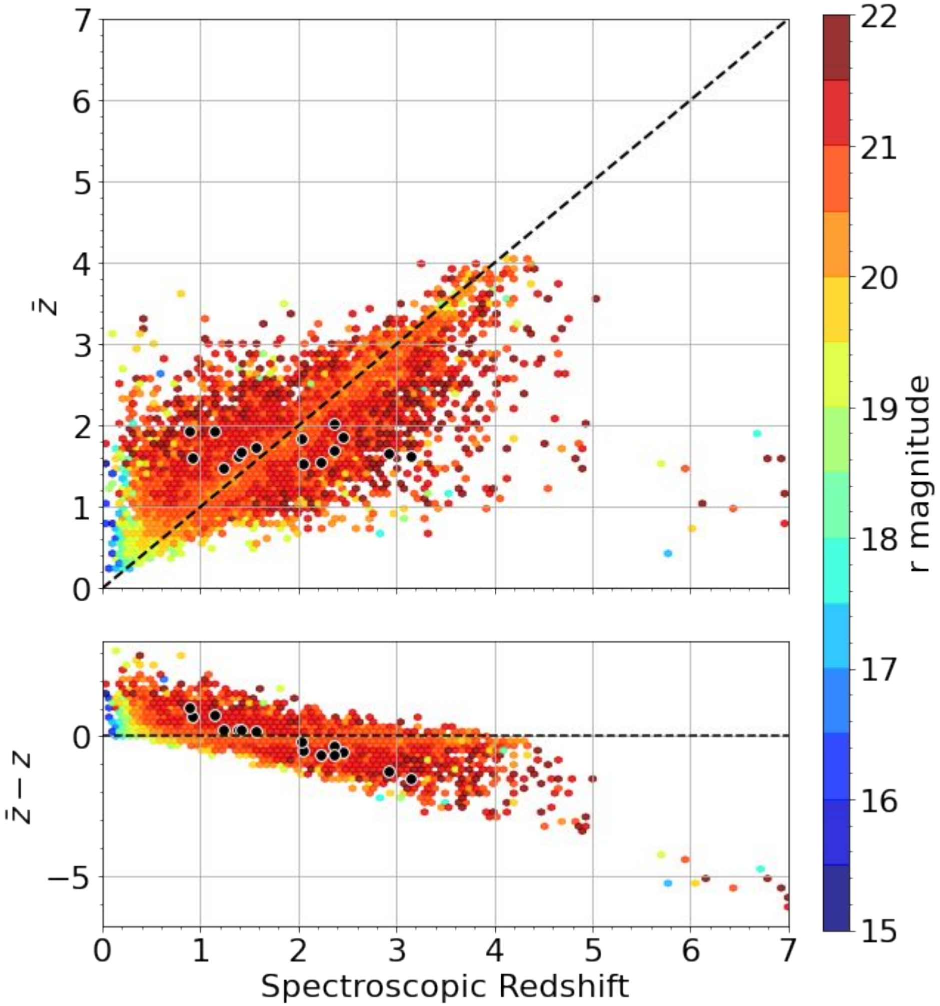
Best model (792)

↪ Residuals



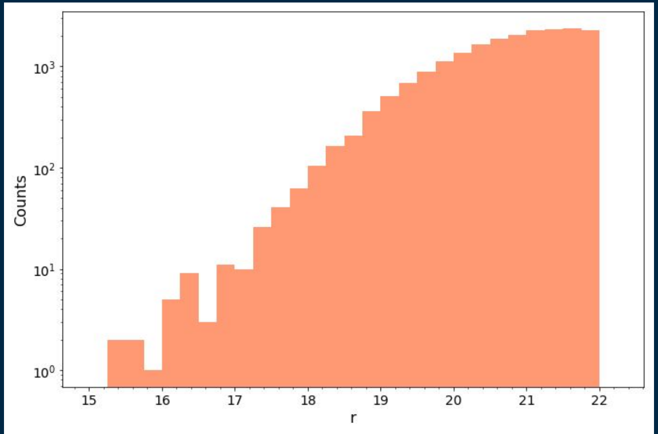
Best model (792)

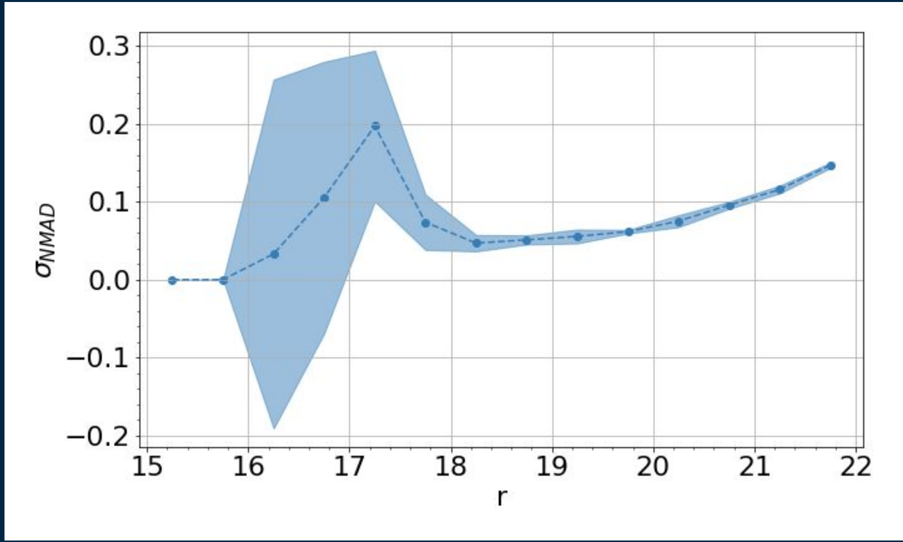
↪ Residuals

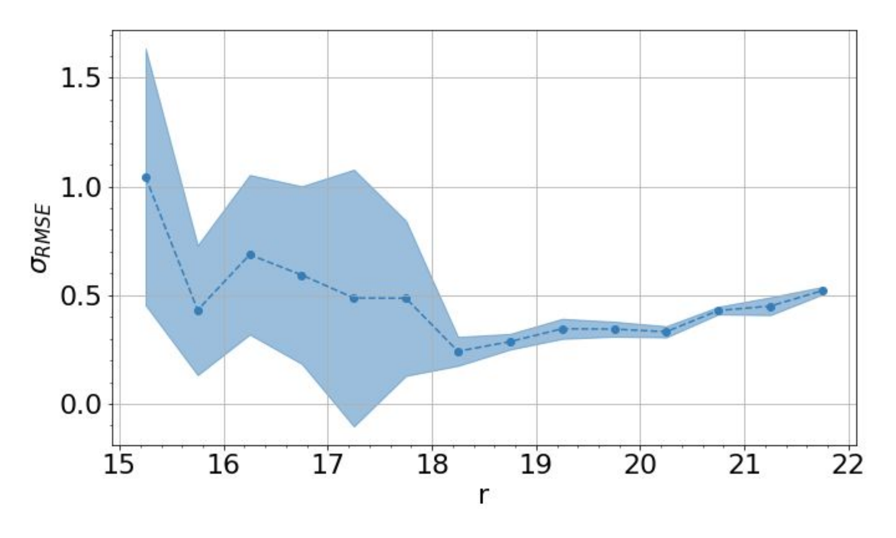


Best model (792)

↪ Error per r bin

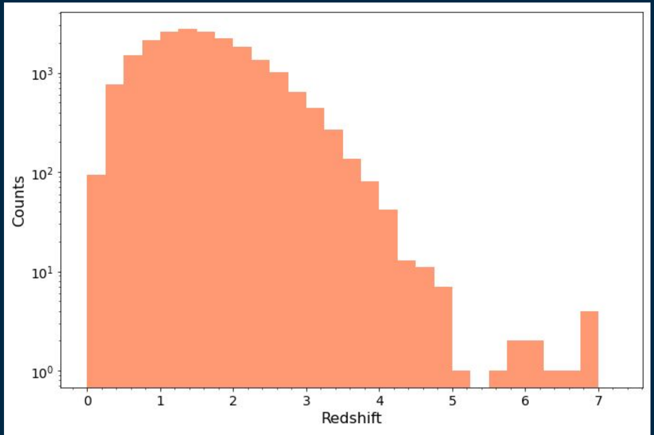


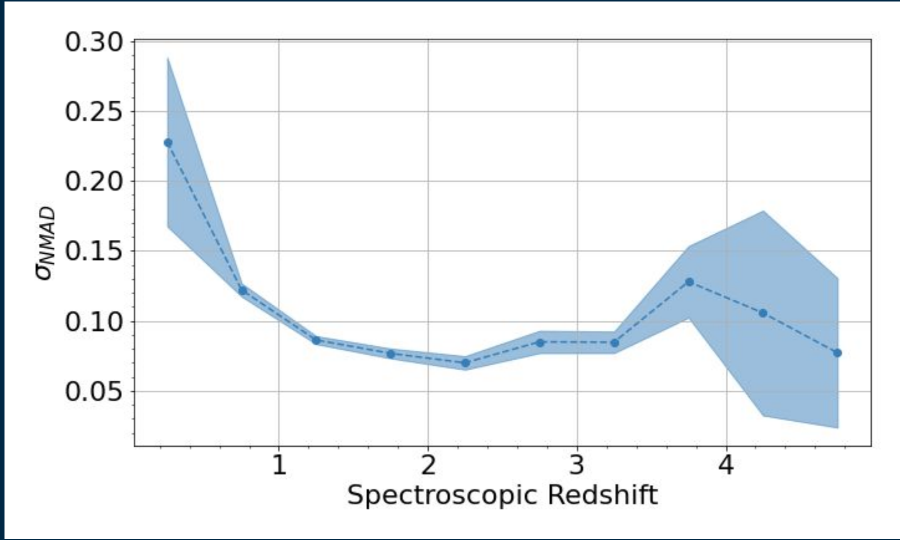


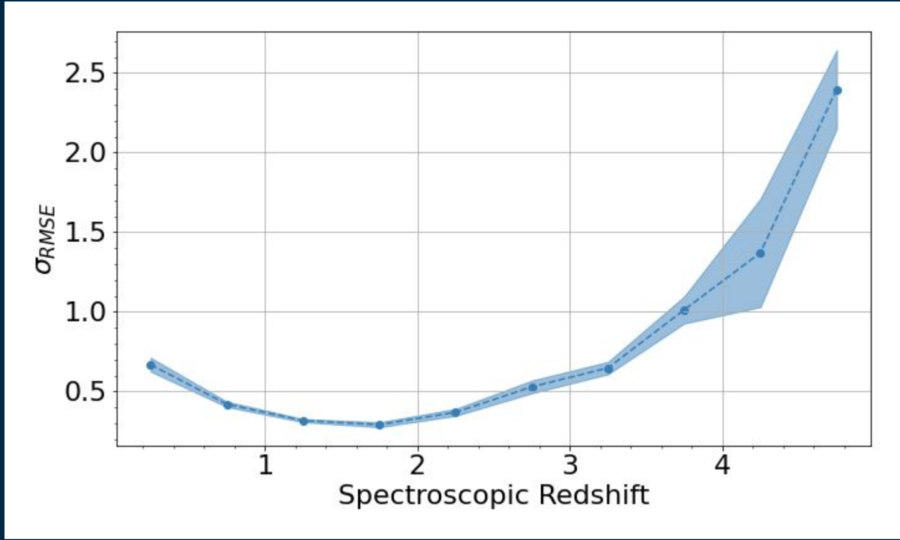


Best model (792)

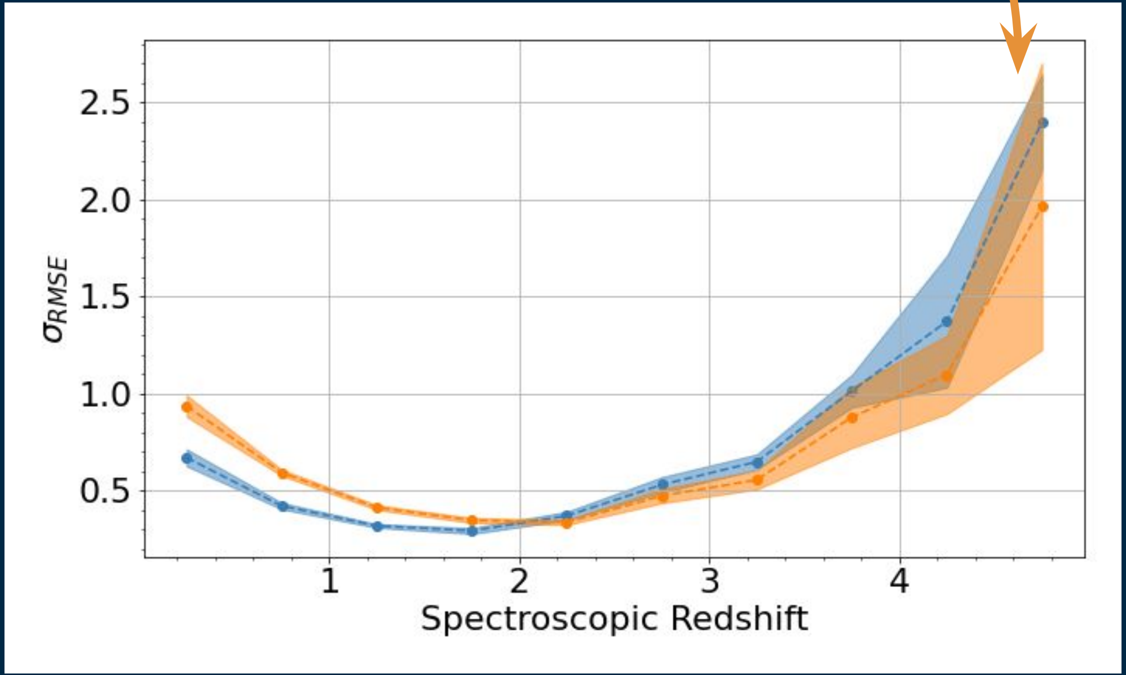
↪ Error per redshift bin







With data augmentation
(using XDGMM, Bovy et al. 2011)



FUTURE PLANS

- We will study changes in the Random Forest hyperparameter space in order to reduce the error in the predictions.
- We intend to test the Bayesian Neural Network architecture from Lima et al. 2021 and then compare all the models.
- Once the best model is chosen, we will evaluate its performance in the test sample and build a VAC with photo-zs of all the objects classified as quasars in S-PLUS.

Do you have any questions?

vruizraquel@usp.br

THANKS