Naples group

Valeria Amaro, Massimo Brescia, Stefano Cavuoti, Giuseppe Longo, Civita Vellucci

Testing of MLPQNA on different datasets (while waiting for the second data cahllenge)

PDFRAPTOR: a Pipeline for the production of PDFs with machine learning methods

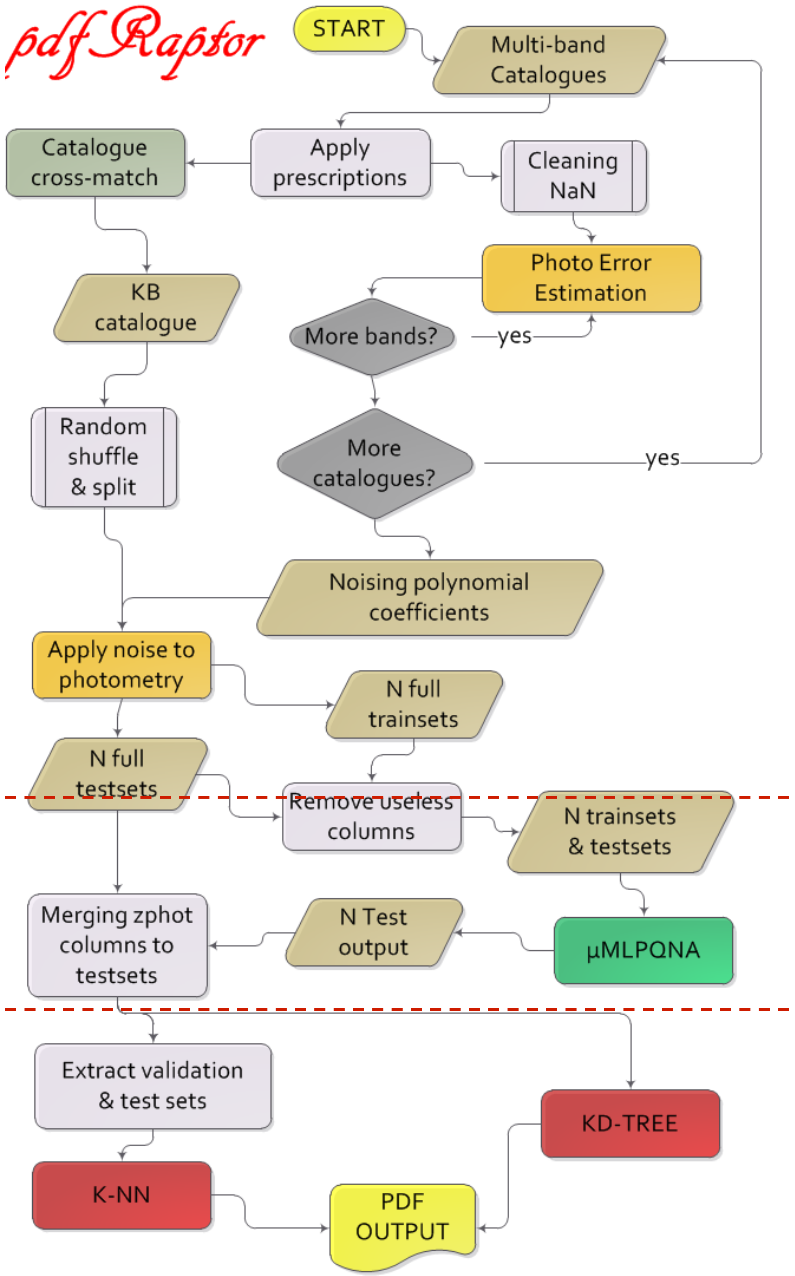
Comparison of systematics between SED (Le Phare) fitting and interpolative (MLPQNA) methods

The model is MLPQNA (Multi Layer Perceptron trained by the Quasi Newton Algorithm), is being validated on several real cases.

Photo-z with MLPQNA

- **PHAT1 Contest** (*Cavuoti et al. 2012, A&A, 546, A13*)
- **GALEX+SDSS+UKIDSS+WISE QSOs** (*Brescia et al. 2013, ApJ, 772, 2, 140*)
- **CLASH-VLT** (*Biviano et al. 2013, A&A, 558, A1*)
- **EUCLID PHZ** (*Coupon et al. 2014, Challenge #1 internal report*)
- **SDSS DR9** (*Brescia et al. 2014, A&A, 568, A126*)
- **KiDS DR2** (*Cavuoti et al. 2015, MNRAS, accepted, in press*)
- **VST VOICE** (*Covone et al. 2015, in prep.*)
- **XMM** (*Vaccari et al. 2015, in prep.*)

pdfRaptor pipeline architecture



Data Pre-processing: photometric evaluation and error estimation of the multi-band catalogue used as KB of the photo-z experiment.

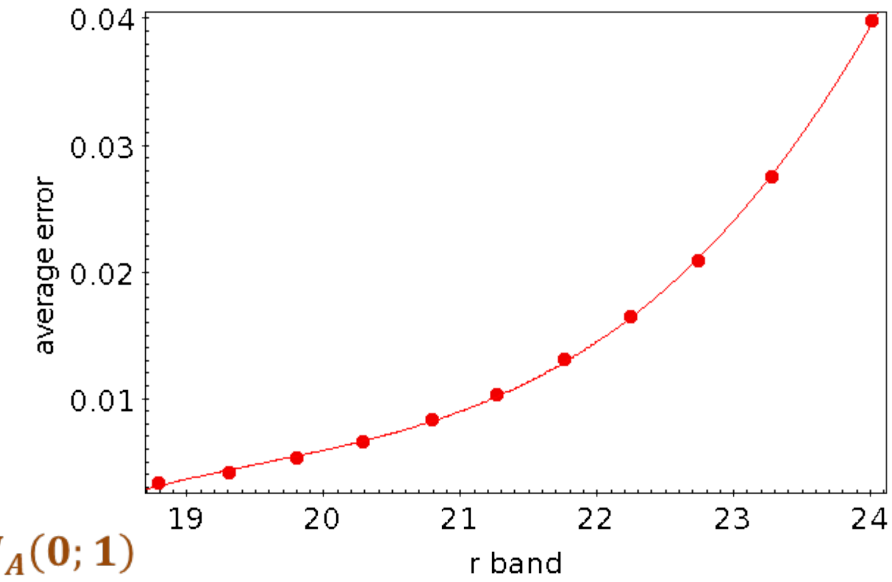
Photo-z calculation: training/test phase to be performed through the selected interpolative method (in this case µMLPQNA, which stands for multi-thread MLPQNA).

PDF calculation: methods designed and implemented to furnish a PDF evaluation for the photo-z produced.

Photometry perturbation

Given a dataset A, a normal distribution on A, and

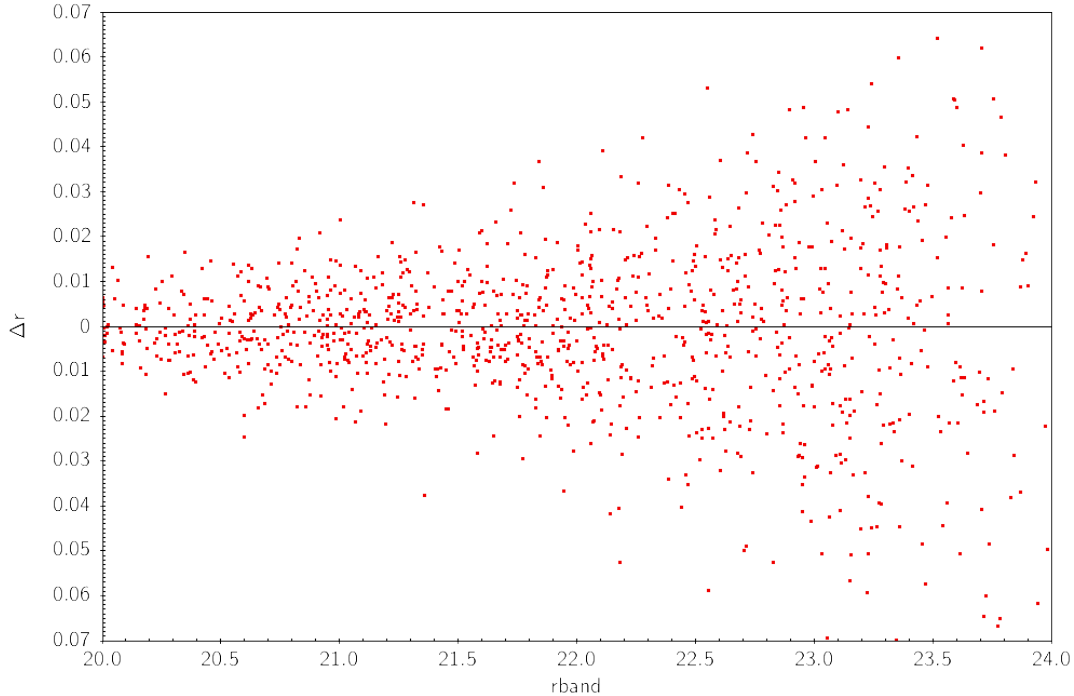
$N_{samples}$	number of objects in a given dataset A
$N_{perturb}$	number of perturbations to be done
N_{mags},	number of affected magnitudes
p_b	polynomial used to perturb mag of band b
$alpha_b$	perturbation constant for the band b
$mag_b(o_i)$	mag value of the band b for the object o_i



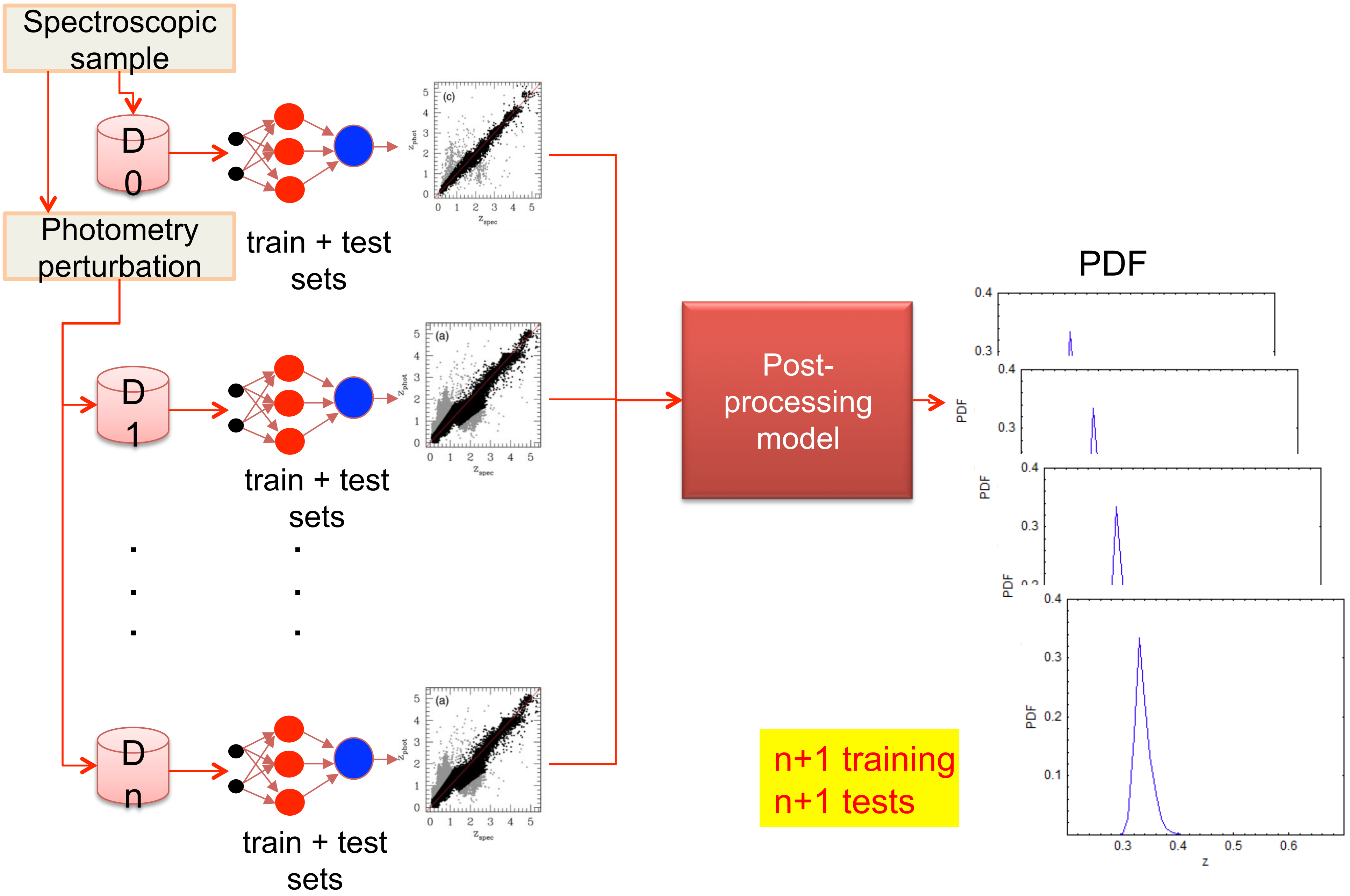
$$\boldsymbol{m_{ijperturbed}(o_i) = m_{ij} + alpha_b * p_b \circ (mag(o_i)) * N_A(0;1)}$$

where the symbol "∘" stays for the scalar product,
$N_A(0;1)$ is a normal distribution with the dimension of the dataset A to be perturbed, i.e. a distribution of a number $N_{samples}$ of values in the interval (-1,1).

The variation of the percentage of noise is ensured by the randomly generated normal distribution at each step.



pdfRaptor processing flow

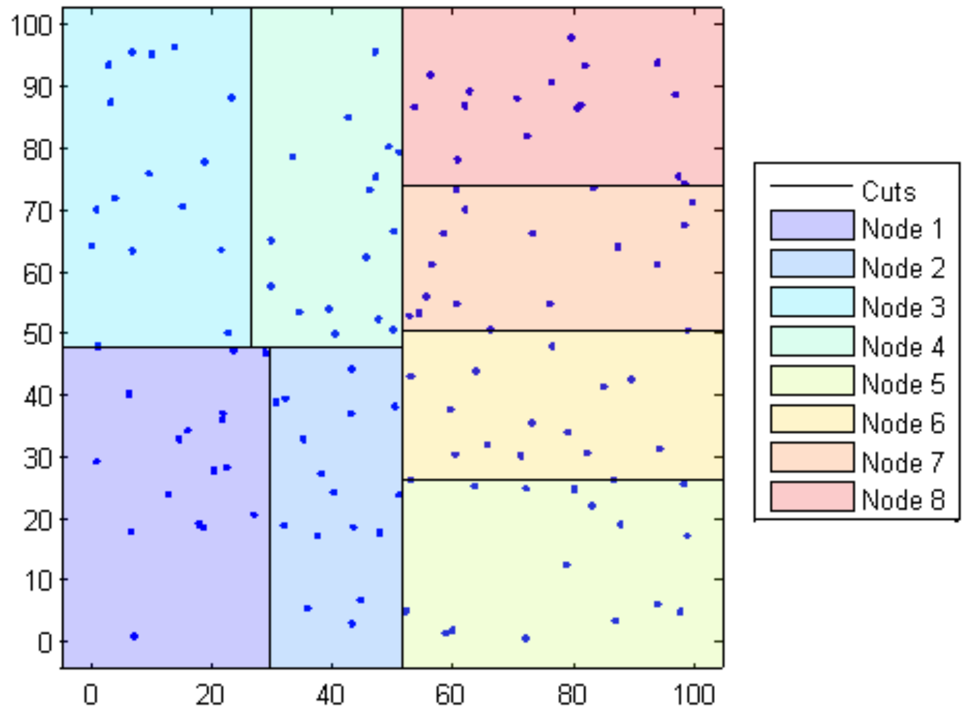


Post-processing for PDF (KD-TREE)

Post-processing Model
KD-TREE

At a high level, a KD-TREE is a generalization of a binary search tree that stores points in *k*-dimensional space.

The method uses the well-known KD-TREE algorithm to partition the photometric and spectroscopic Parameter Space on the base of, respectively, the photometric magnitudes and the zspec present within the used data.



The partitioning produces a series of bins and through the analysis of the associated standard deviations it could be possible to evaluate the trend of photometric error vs the spectroscopic one, giving the possibility to estimate the error distribution and to correlate both types of error trends.

Post-processing for PDF (K-NN)

Post-processing Model **K-NN**

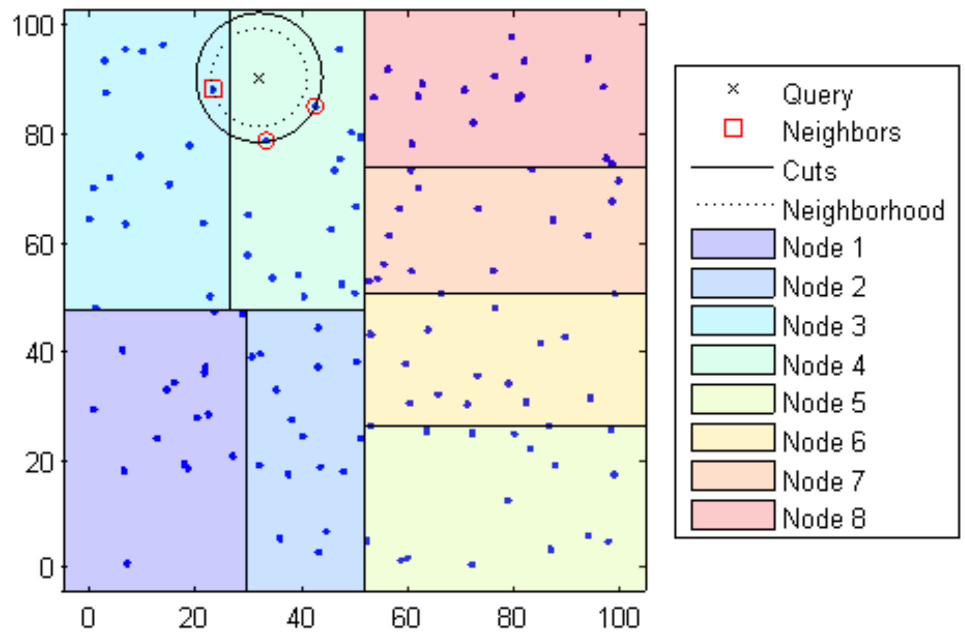
At a high level, in a K-NN (K Nearest Neighbours) the input consists of the *k* closest training examples in the feature space. An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its *k* nearest neighbors.

The **K-NN** method is based on the extraction of arbitrary N objects within the test set closest to each single object selected within the *Evaluation set*.

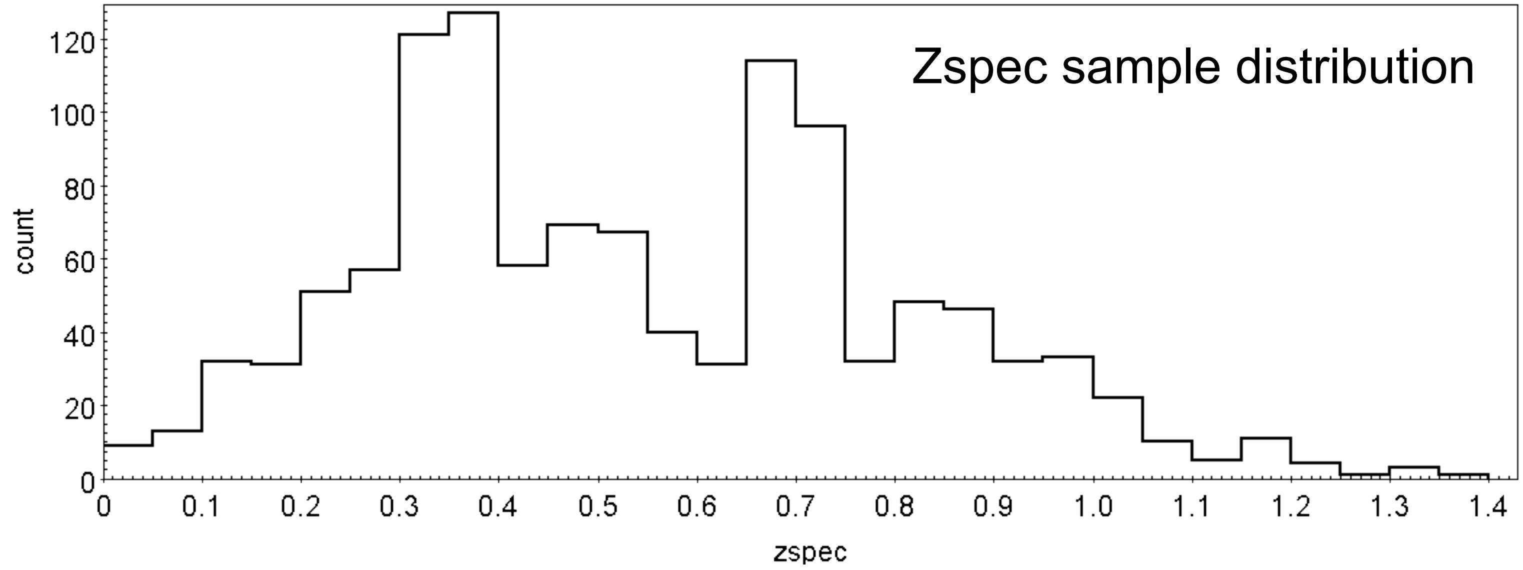
Here closest has to be intended in terms of euclidean distance among all photometric features of the objects.

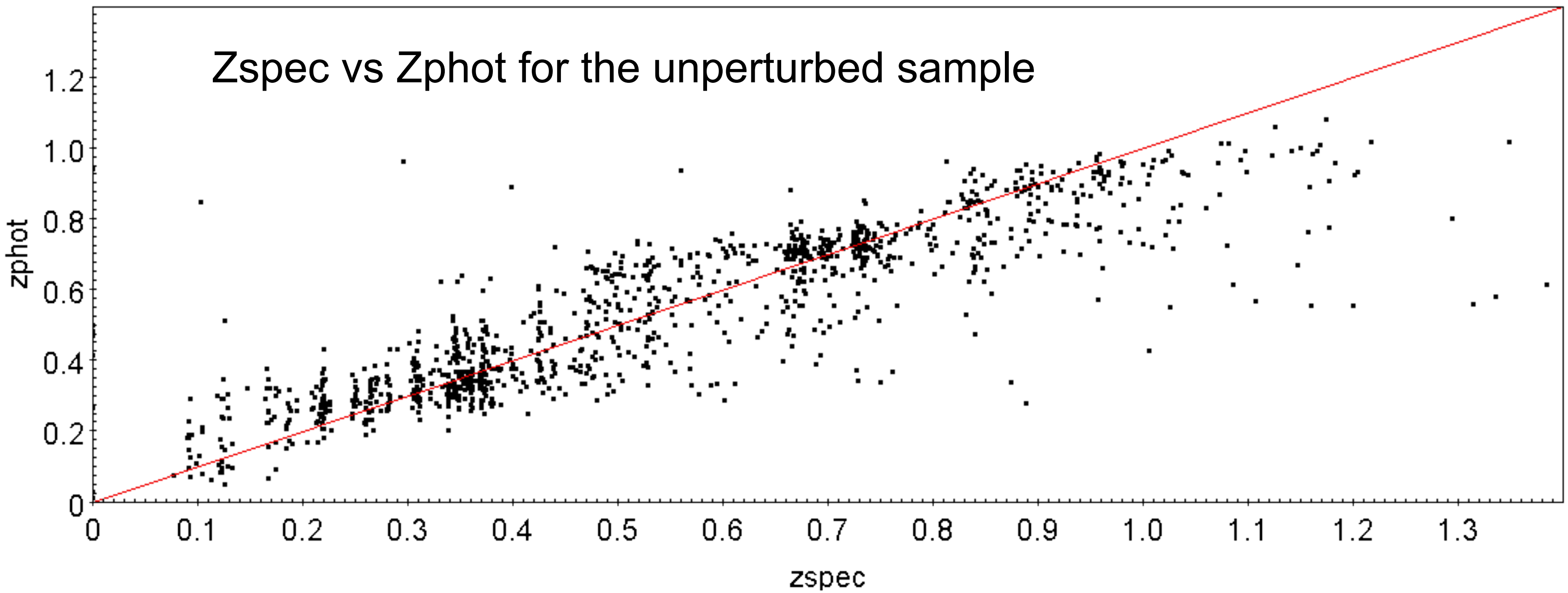
The resulting distribution of the Δz is obtained by considering the N values for each object of the *Evaluation set*.

The associated error (bias ± σ) is the PDF estimation.

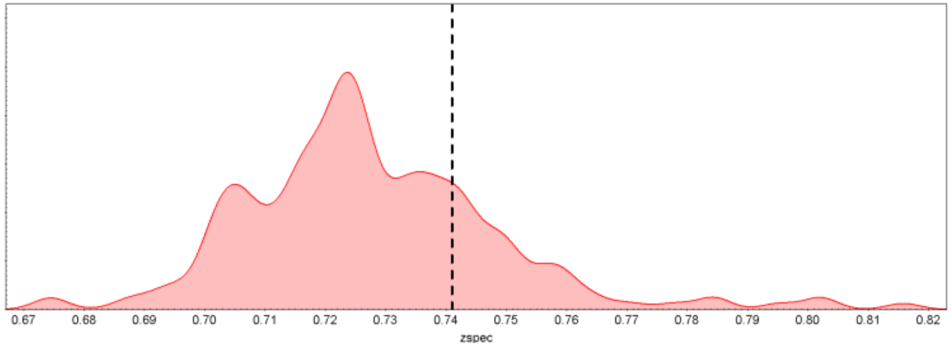


Knowledge base I/O

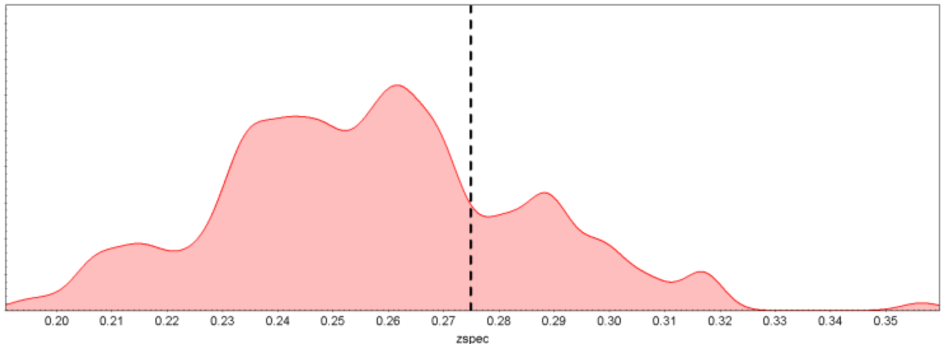




Base algorithm – PDF examples



zCosmos + Cosmos + Ukidss
Bands: R, I, Z, K
objects:2,887
(60%) 1,723 training set
(40%) 1,164 test set



Tests	Bias	σ
Unperturbed test set	0.0014	0.0087
Average on 200 perturbed test sets + unperturbed test set	0.0004	0.0085