PHZ Processing Function

SC*5* Kick-Off

**Stéphane Paltani, Nikos Apostolakos,
Jean Coupon, Pierre Dubath,
OU-PHZ, SDC-CH**

PHZ Pipeline

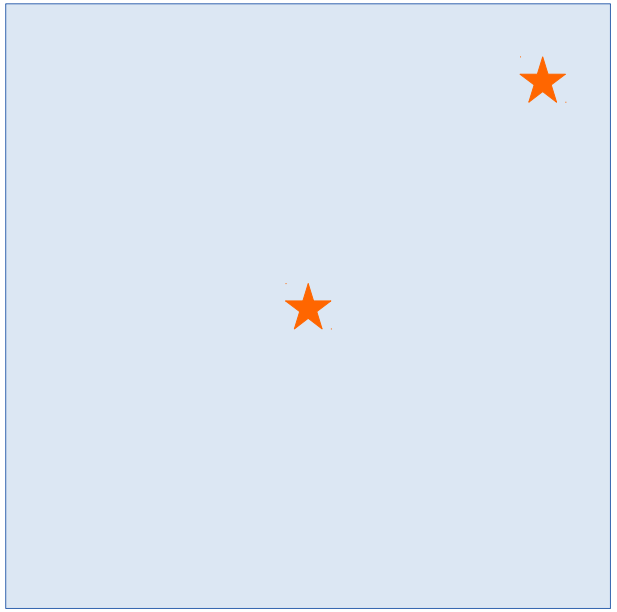
The PHZ pipeline baseline has evolved significantly to take into account extrinsic color variations due to:

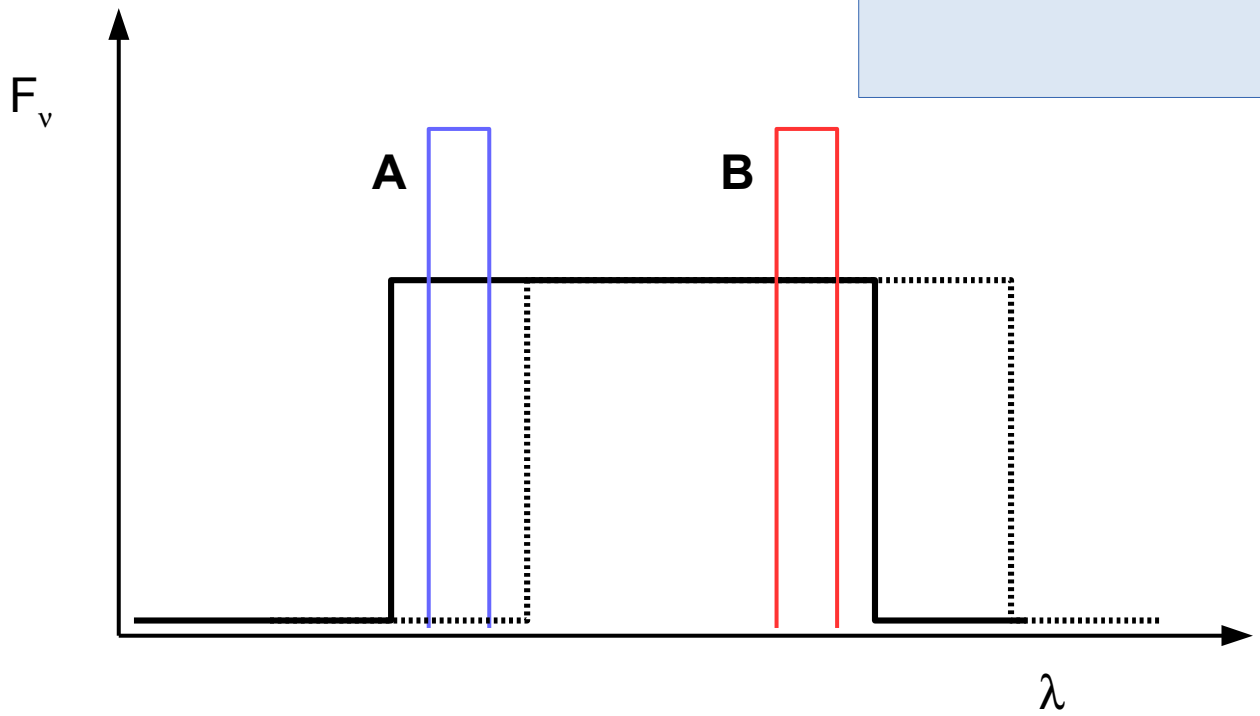
- Galactic extinction
  - This is the only effect we expect in SC4/5/6
- Location in the FOV (CCD QE, filter non-uniformity, off-axis angle)
- Time? (CCD and filter aging? Replacement of filter?)
- Atmospheric extinction (cannot be dealt with?)

Effect of variable transmissions

Differences in transmissions lead to inhomogeneous colors

The effect is SED dependent





Dealing with variable transmissions

For a given Euclid object, we can in principle know the effective filters. To determine photo-z:

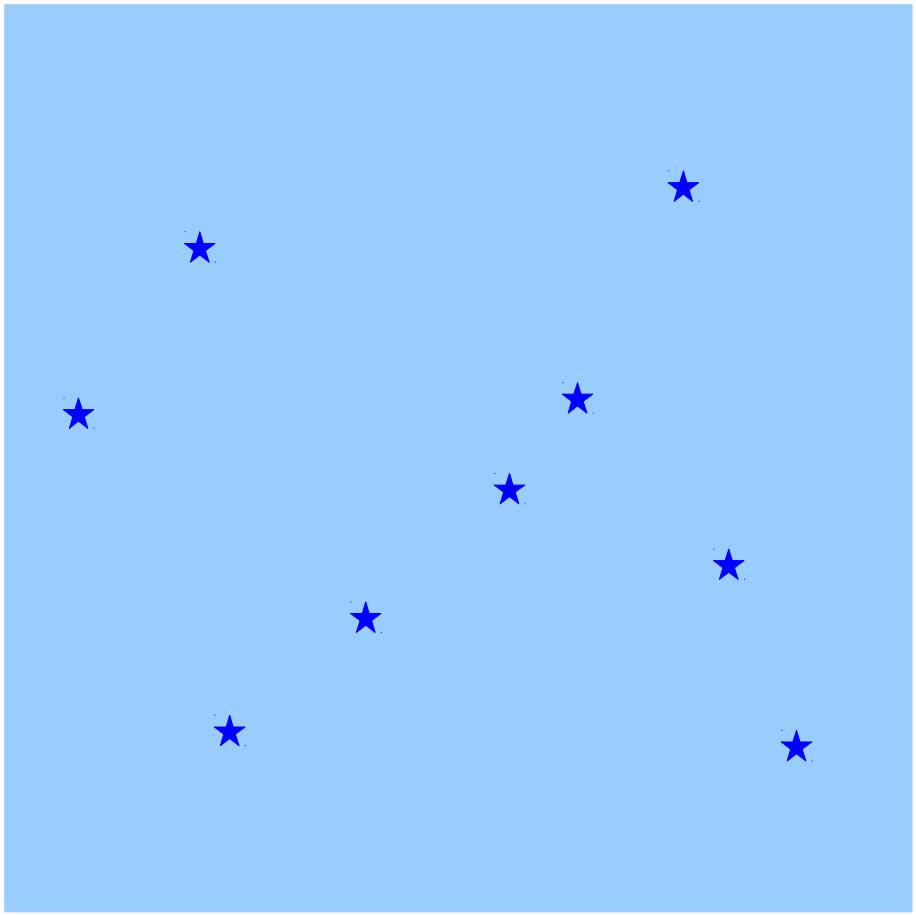
- Assuming a set of templates, we can compute likelihoods for the fluxes using template-fitting
  - But it would be computationally impossible, because each object could in principle have its own grid
- Provided we have an adequate training set with known SEDs, we can map the objects in the training set to the given Euclid object and perform the training of a ML algorithm
  - But the mapping would in general be extremely costly
  - And the training phase would have to be repeated for each object; in general this is impossible

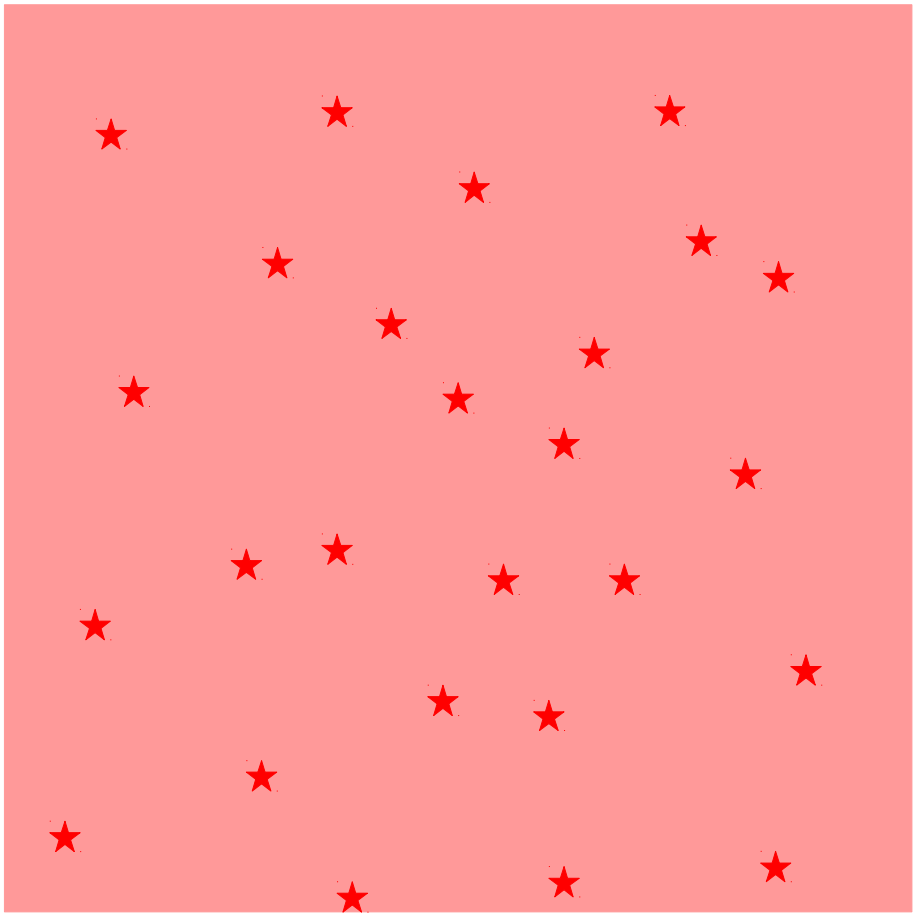
Reference Sample

Reference Sample
10^6 objects
Known SEDs and PDZs

COLOR SPACE

Euclid Sample
Few 10^9 objects
Unknown PDZs



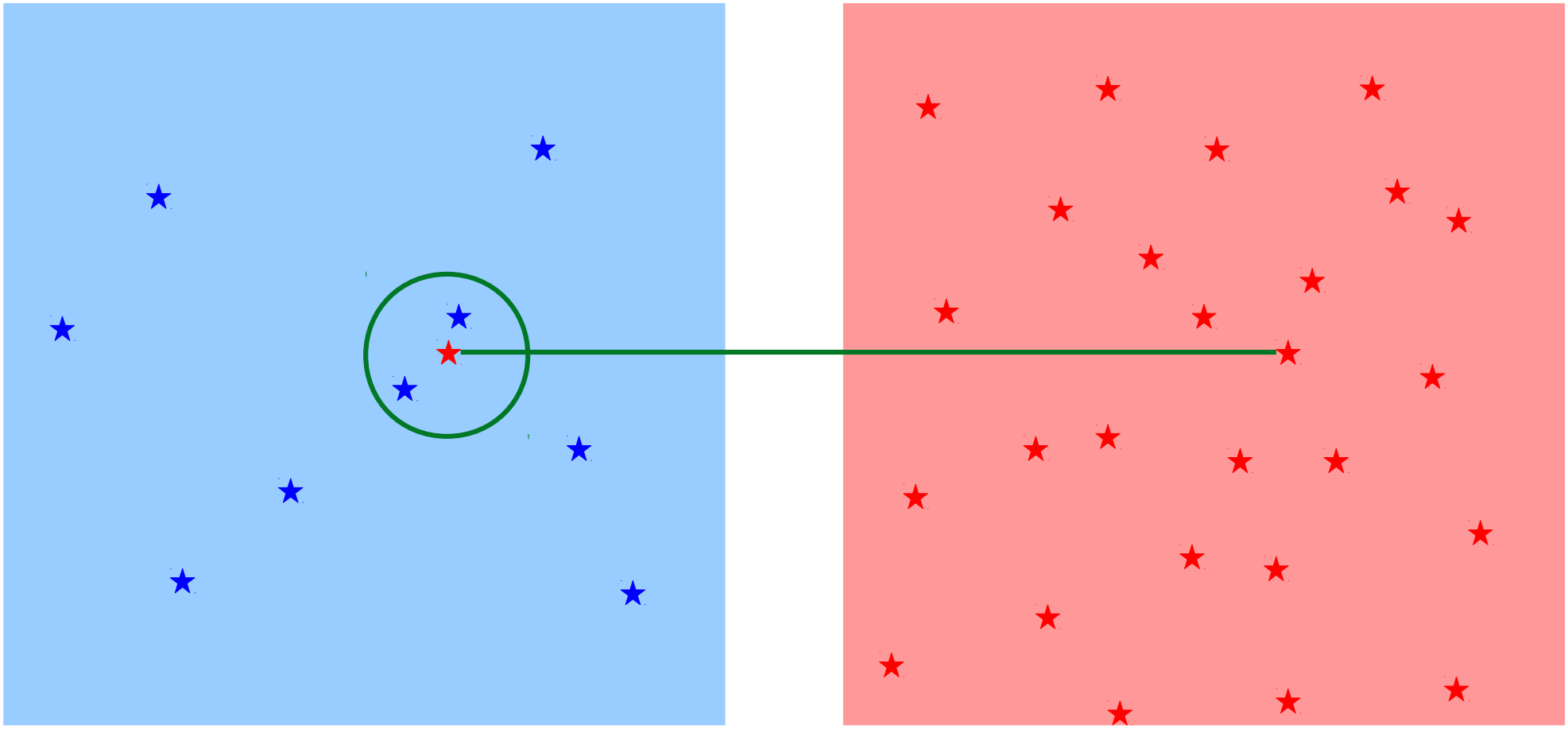


Nearest-neighbor approach

Reference Sample
10^6 objects
Known SEDs and PDZs

COLOR SPACE

Euclid Sample
Few 10^9 objects
Unknown PDZs

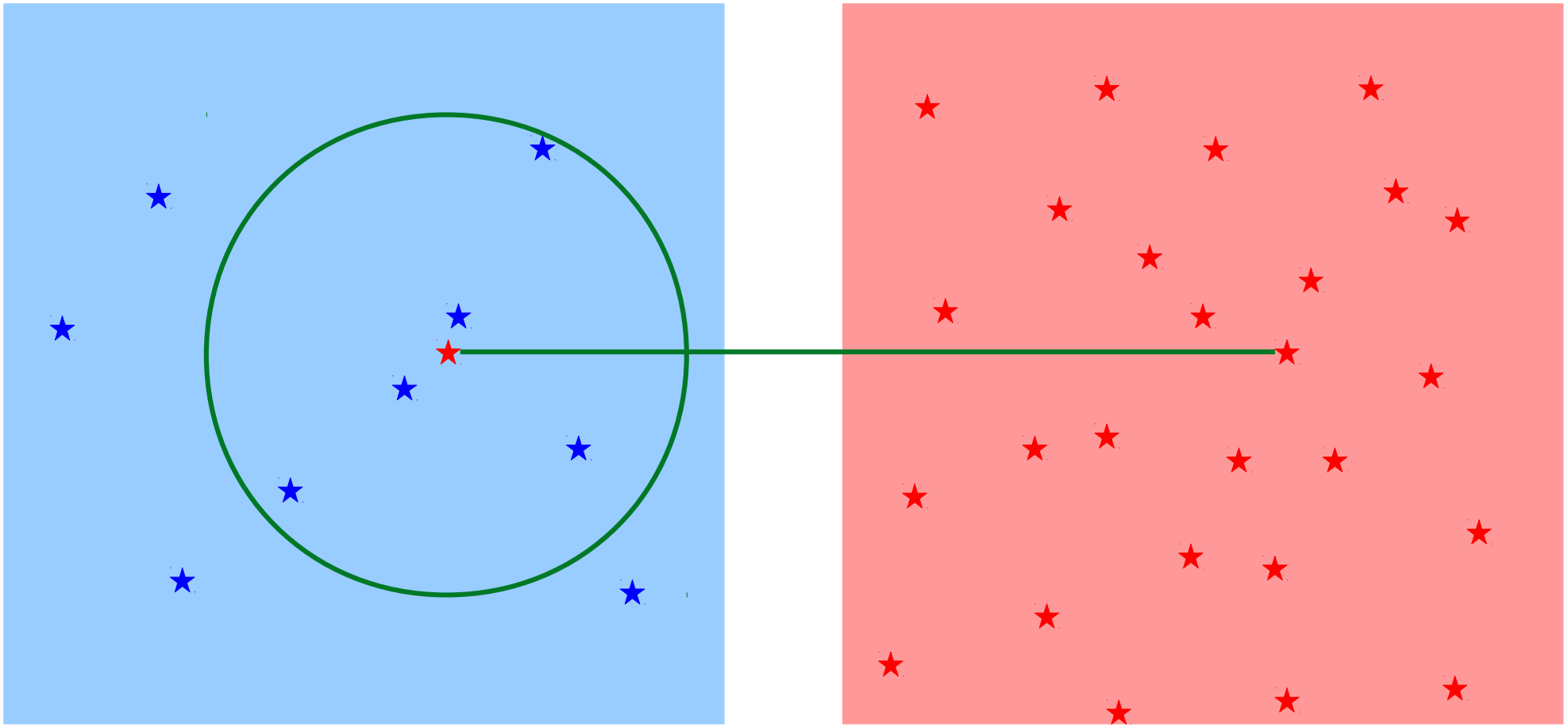


Corrected nearest-neighbor approach

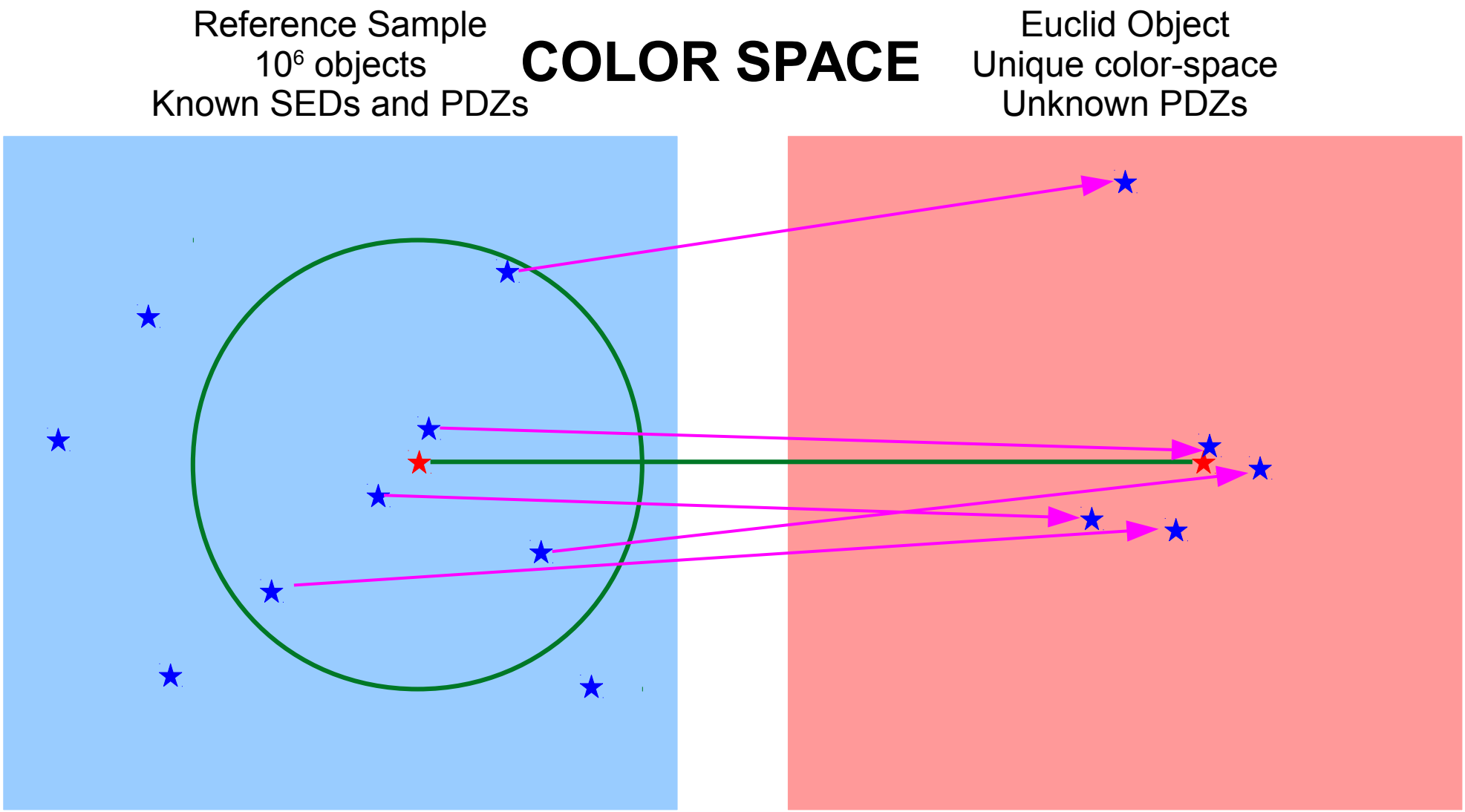
Reference Sample
10^6 objects
Known SEDs and PDZs

COLOR SPACE

Euclid Sample
Few 10^9 objects
Unknown PDZs



Corrected nearest-neighbor approach

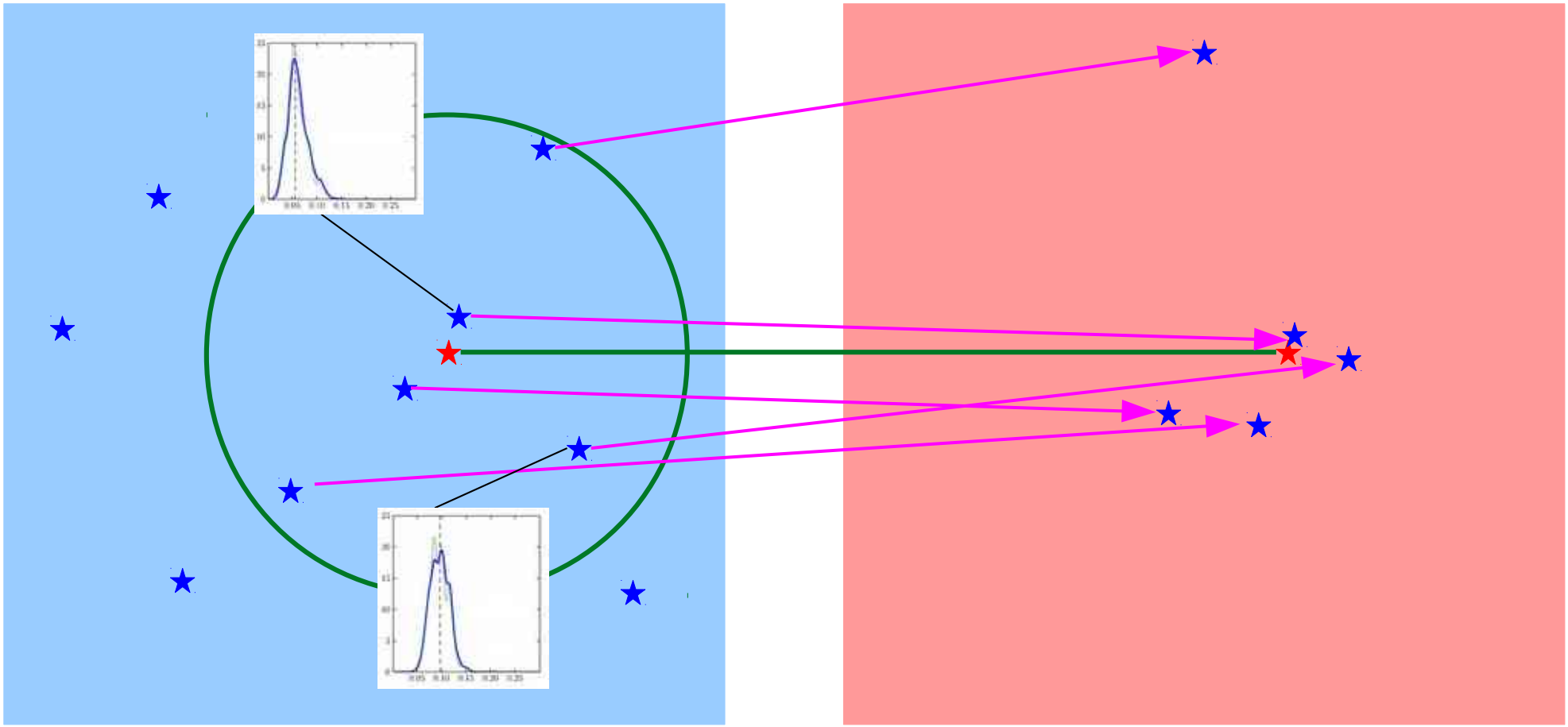


Corrected nearest-neighbor approach

Reference Sample
10^6 objects
Known SEDs and PDZs

COLOR SPACE

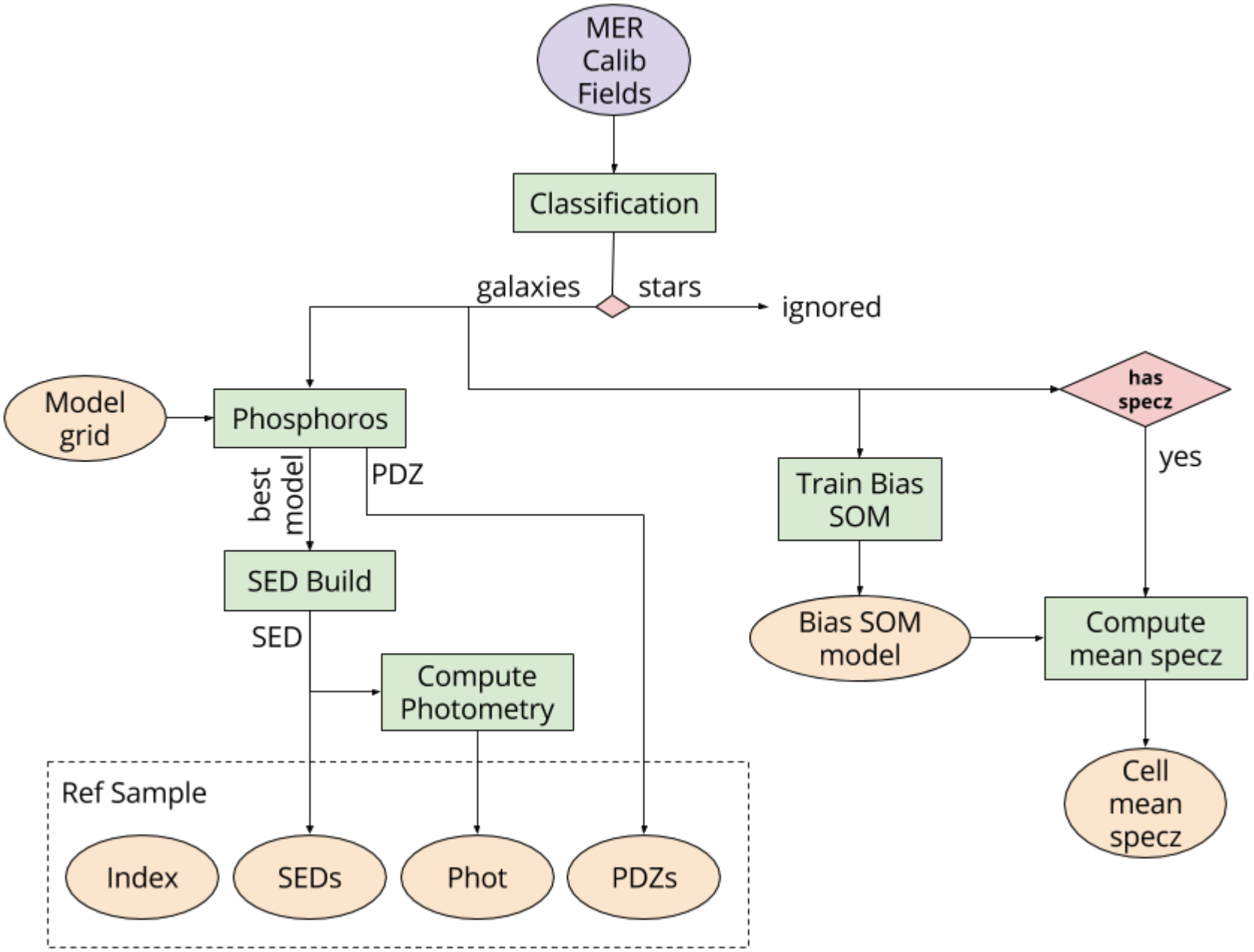
Euclid Object
Unique color-space
Unknown PDZs



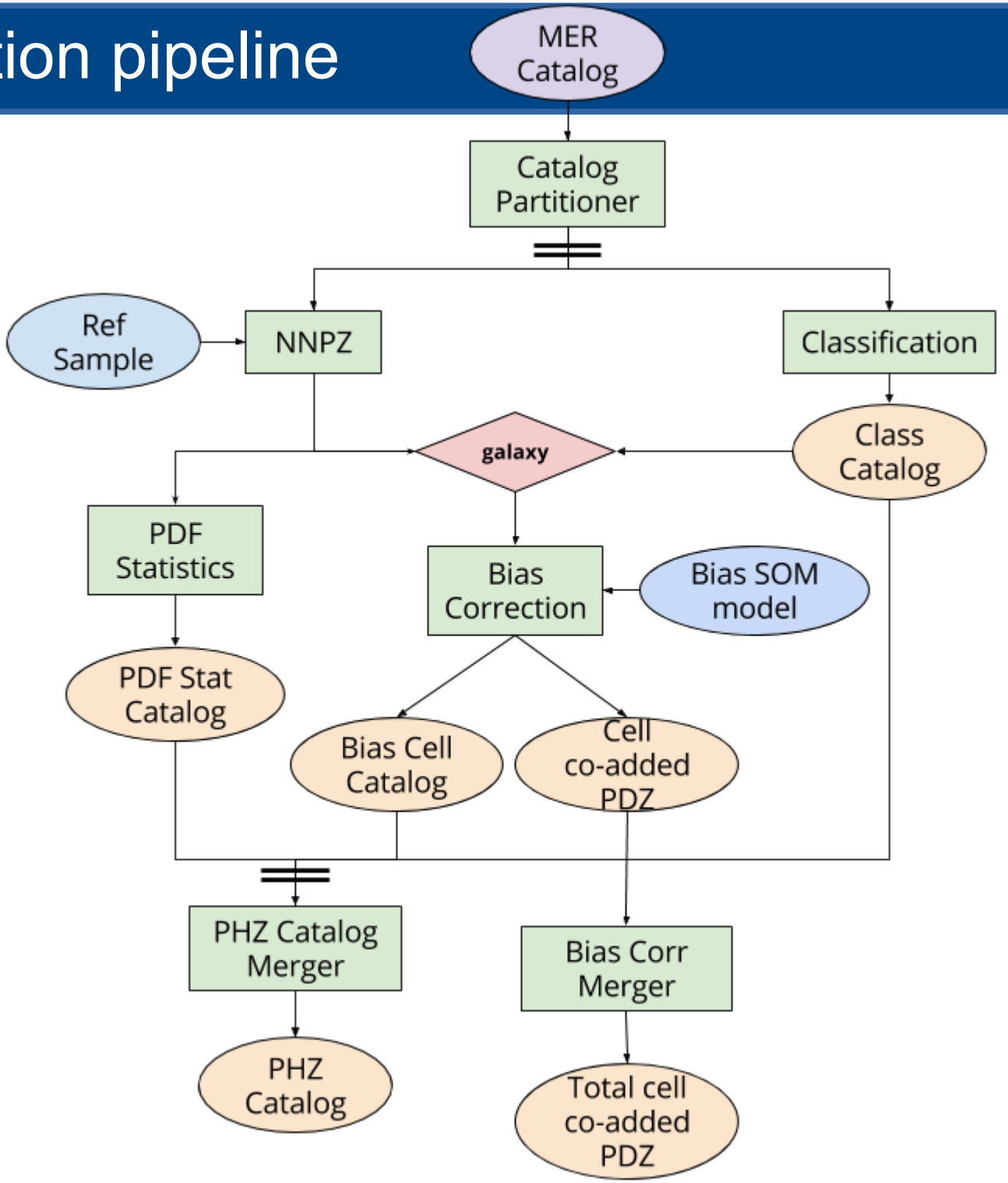
Requirement on SC4/5/6 input data

- Calibration fields: COSMOS, SXDS, *VVDS, ECDFS,…* need to be simulated at 5x Euclid depth
- Enough true redshifts need to be provided on the calibration fields for calibration (TBD)
- We do not request a simulation of the effect of variable transmissions in SC4/5/6 (impact on SIM, EXT, MER)
  - We propose to include this in SC7
- No color term should be applied in the photometric calibration
  - Actually, we could have both with and without color term, as this may speed up PHZ computation
- Galactic extinction should be provided in the catalogue, but colors should not be corrected for this effect
  - This extrinsic color variation will be taken care of by our NN pipeline, pre-validation of the method

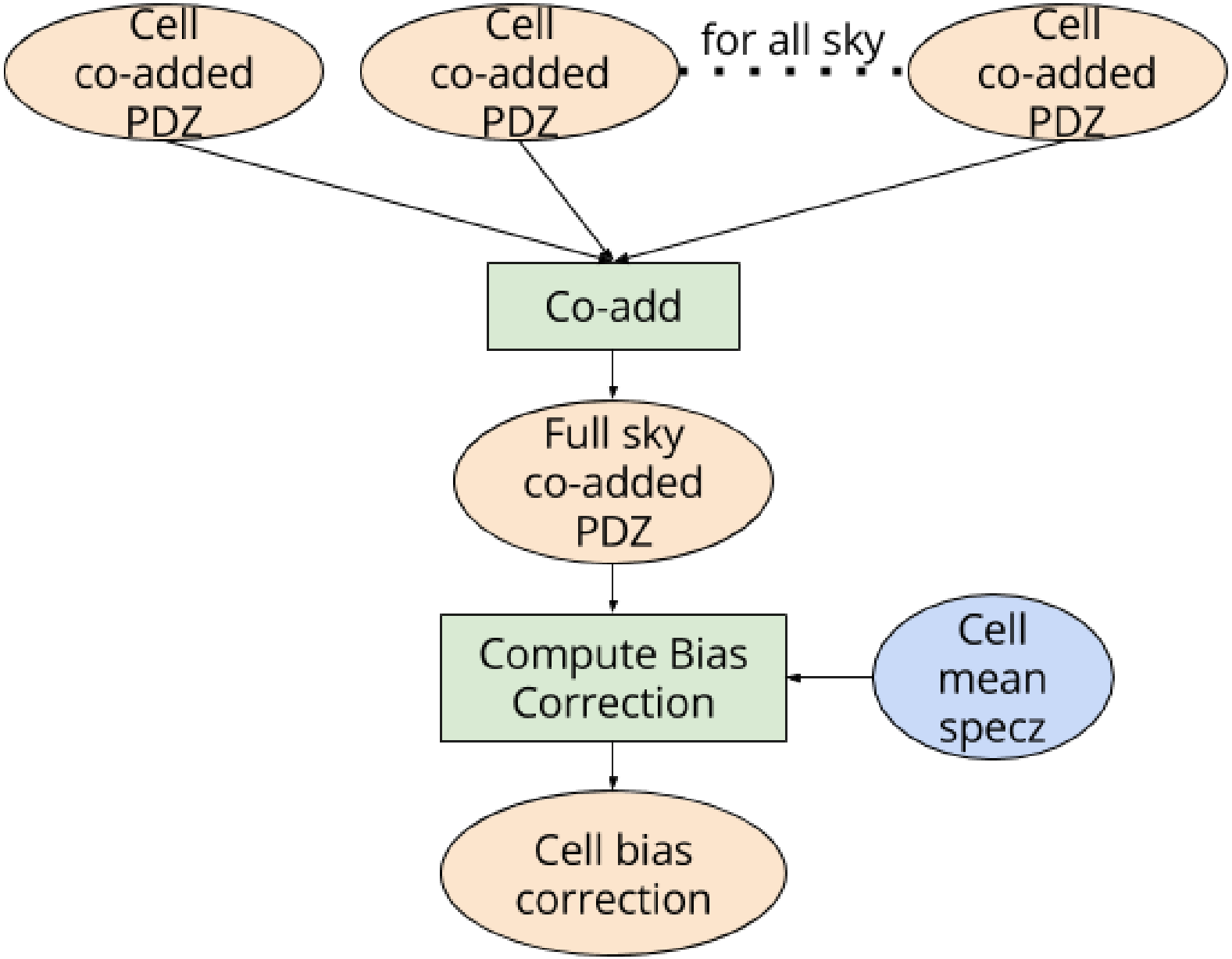
Reference sample pipeline



Production pipeline



Bias correction pipeline



DM status

- Reference sample product is defined, but will be split in four files
- Bias correction products need to be defined
  - SOM model DM (internal, input to the Production pipeline)
  - Co-added PDZs for each SOM cells (input for the Bias correction pipeline)
  - Mean spectroscopic redshift for each SOM cell (input for the Bias correction pipeline)
- Output catalog products already defined; we will merge them into one data product
- Additional products for LE3 to be defined
  - Bias correction per SOM cell lookup-table
  - Extra column in the photo-z catalogue for the SOM cell ID

MDB needs

- Reference filter transmissions
- Galaxy SEDs
- Star SEDs
- Galactic extinction curve(s?)
- Intrinsic extinction curves
- (Luminosity) priors(?)

Infrastructure needs

For the Production pipeline:

- With 1GB memory for PDZs, the input catalogue should be 500’000 objects
- Overhead for reading the SEDs (10^6 SEDs, 100 GB) for the reference sample: 12 min (0.001s per object if input catalog is 500’000 objects)
  - This is an input data product, which needs to be copied by IAL for each instance of the Production pipeline!
- Finding the neighbors takes 0.1s per source with brute force
- We need 0.05s to compute the colors for 30 neighbors
- Bottleneck is the finding the neighbors (KD-tree?)

- We can process on average about 10 objects per seconds per core
- To process SC5 in 10 days (10^6 seconds), assuming 10^7 objects, we need 1 core...

Validation of scientific requirements

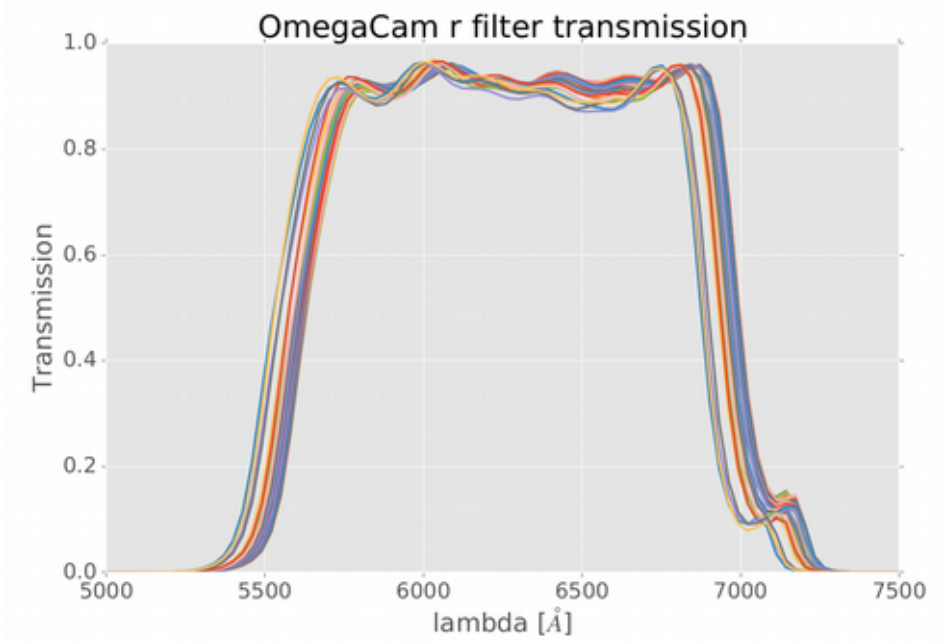
- Most functional requirements (see below)
- True redshifts can be used to test performance requirements, but scope limited: z<2.3, no AGN, not fully realistic SEDs, ...
  - R-PHZ-PRD-P-010 PHZ Performance (scatter and outlier fraction)
  - R-PHZ-PRD-P-020 PHZ Bias
- Excluded
  - R-PHZ-PRD-F-030 Spectroscopic redshift
  - R-PHZ-PRD-F-050 Star SED
  - R-PHZ-PRD-F-060 Flags
  - R-PHZ-PRD-F-060 WL subset definition
  - R-PHZ-PRD-F-100 Star classification knowledge (partially)
  - R-PHZ-PRD-P-030 Star classification performance

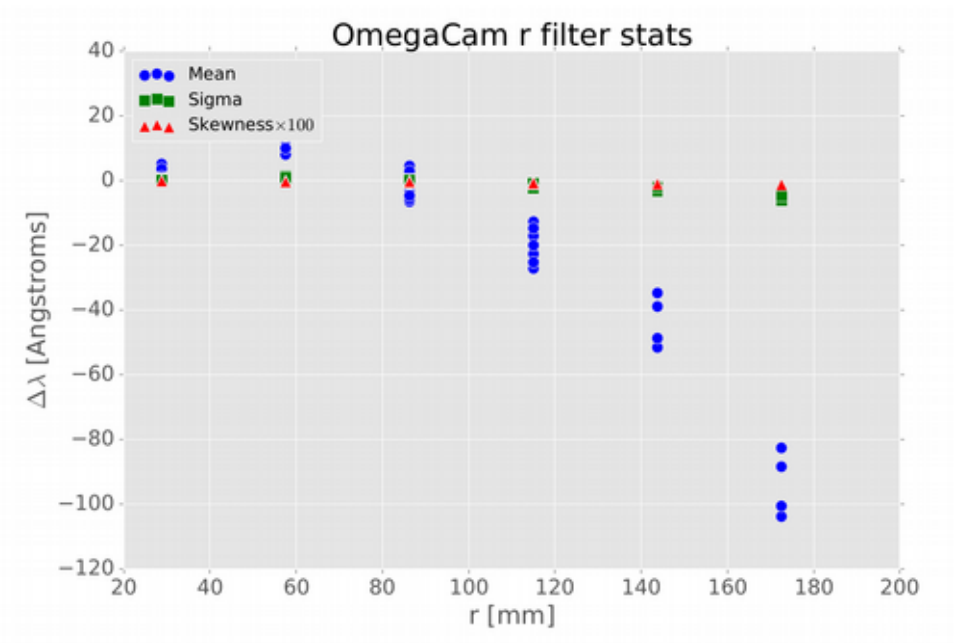
The effect of variable transmissions on photometric redshifts

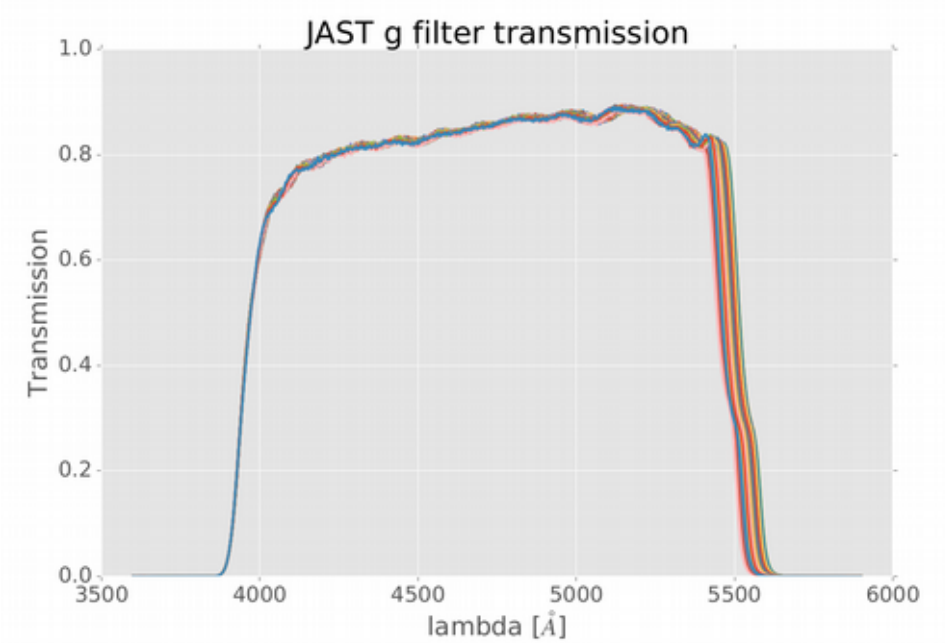
Jean Coupon

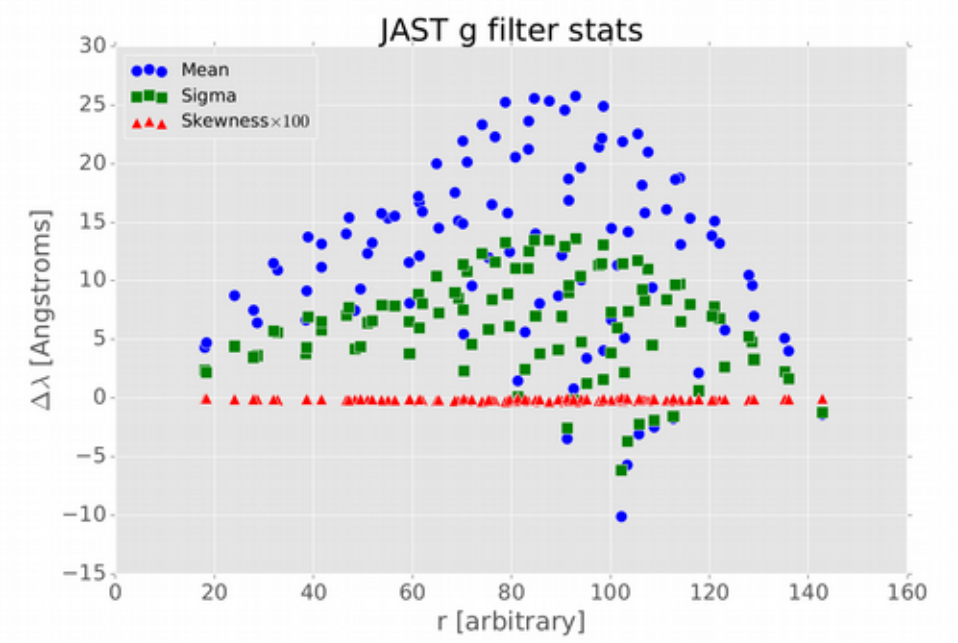
(University of Geneva) et al.
January 15, 2018

Filter variation examples

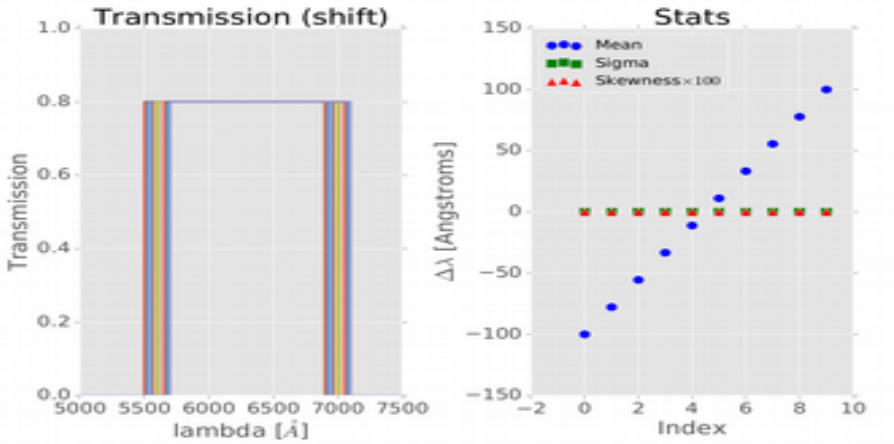


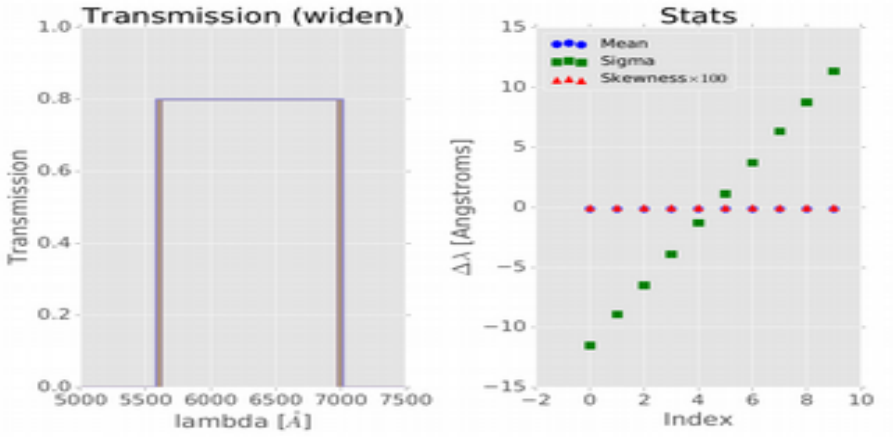


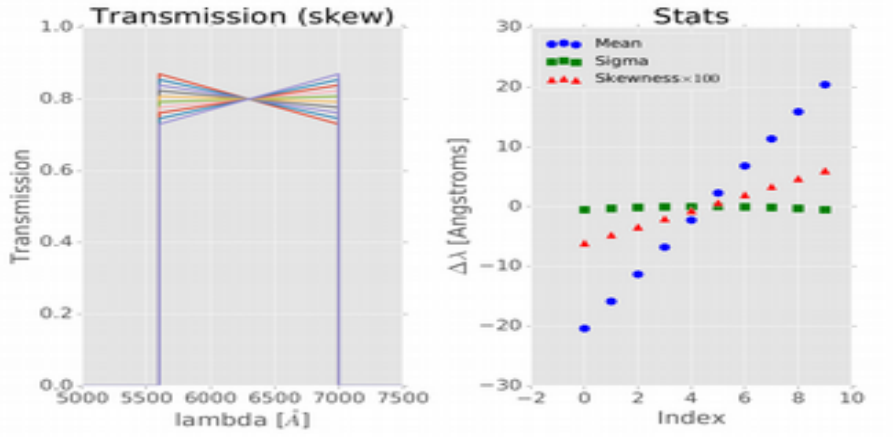


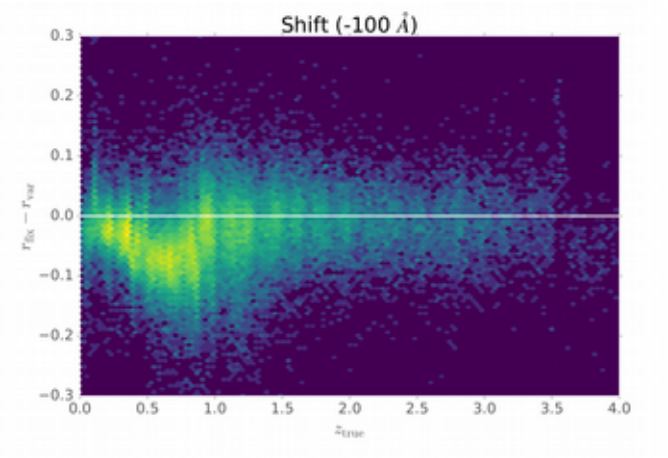


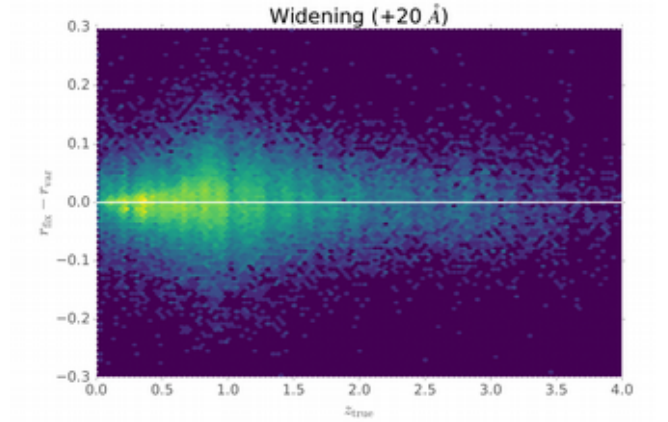
Simulating filter variations











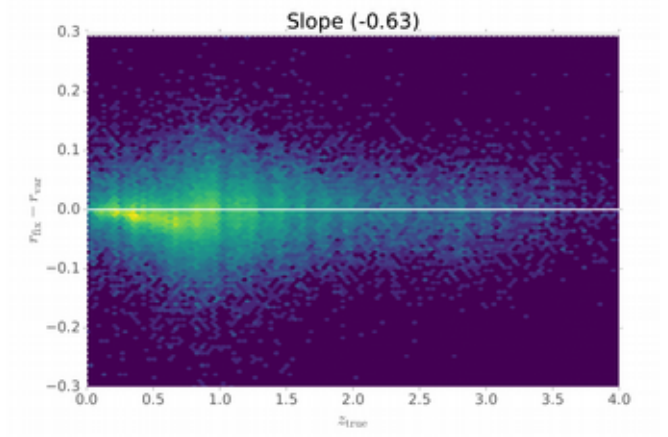
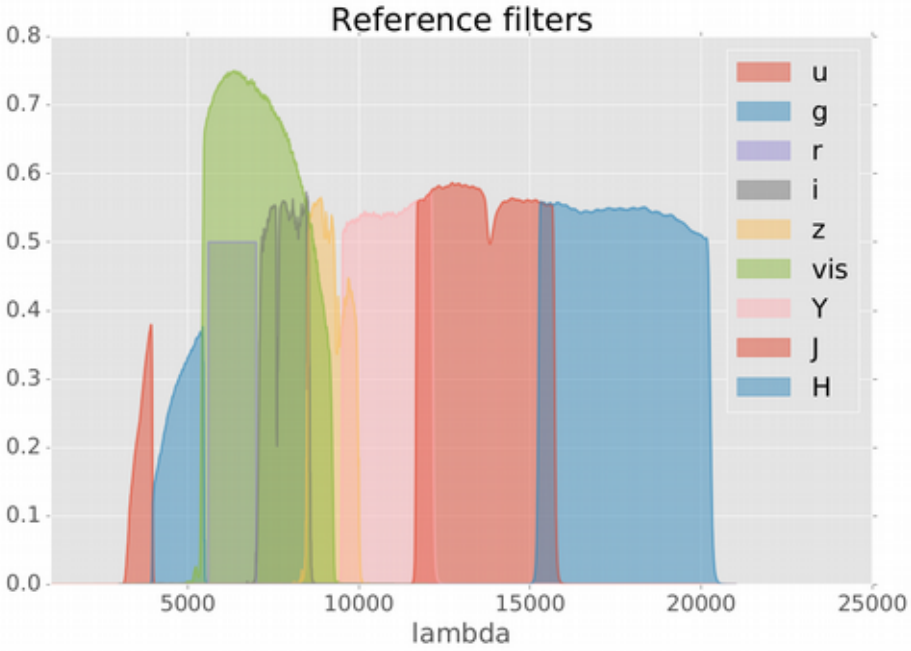
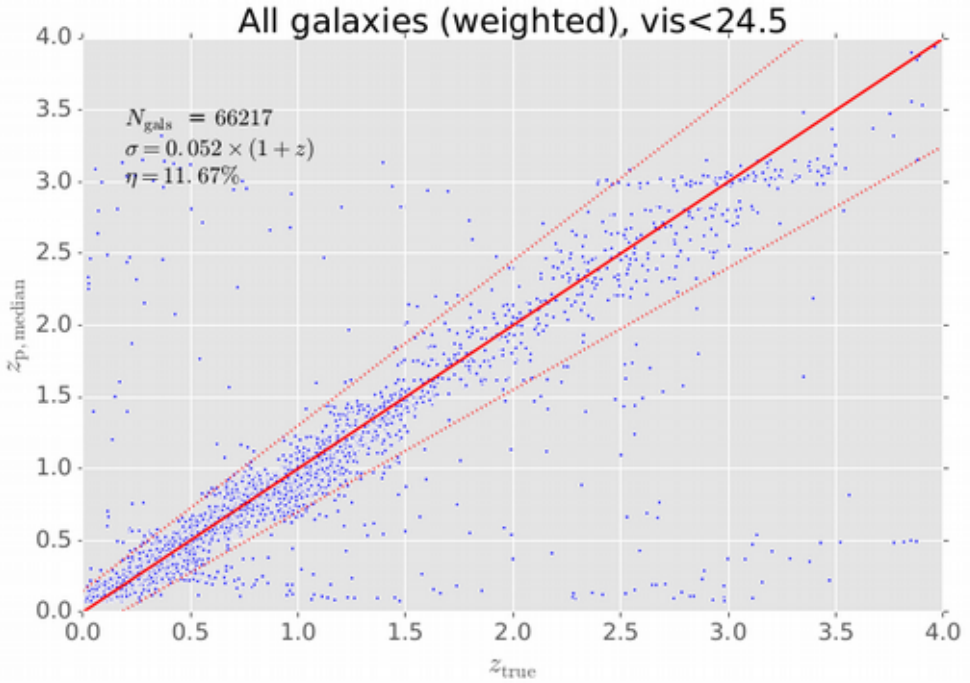


Photo-z's with fixed filters





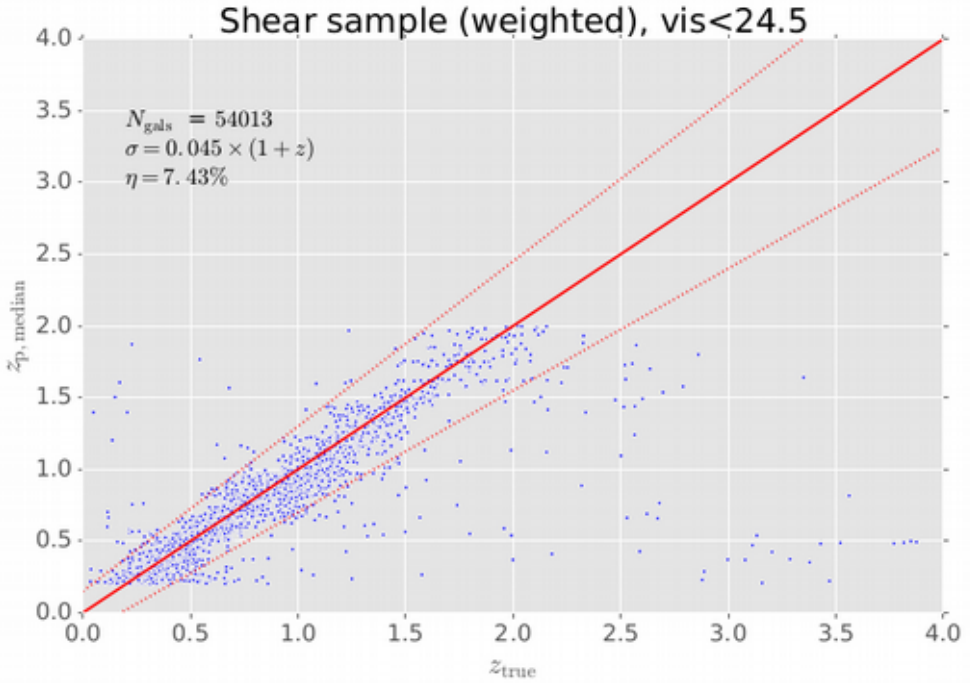
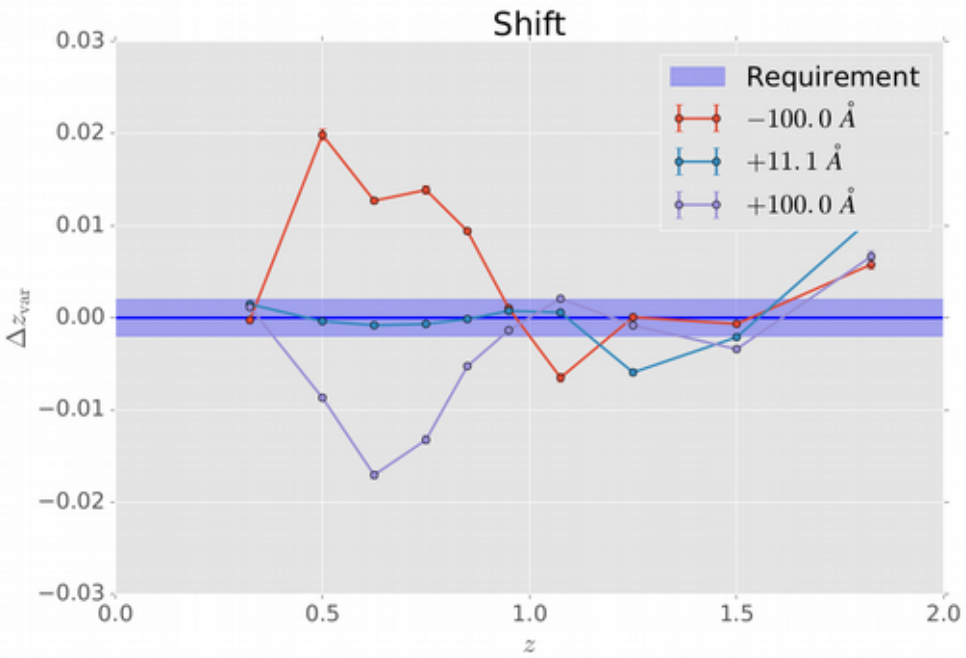
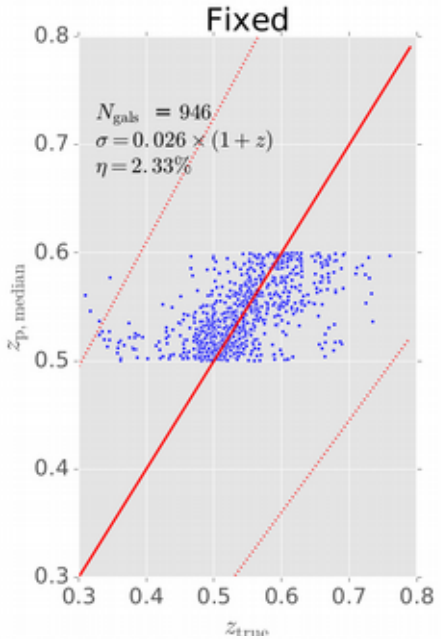
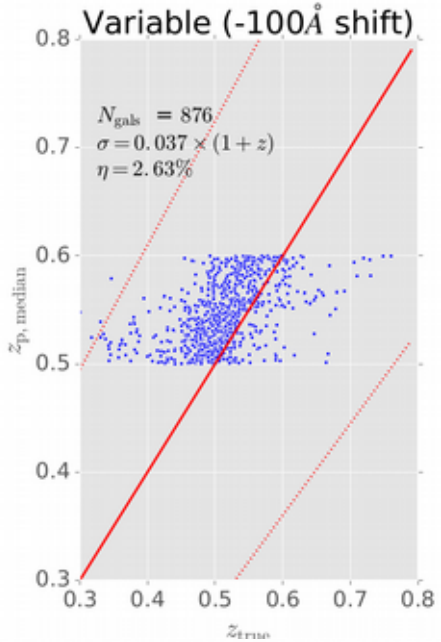
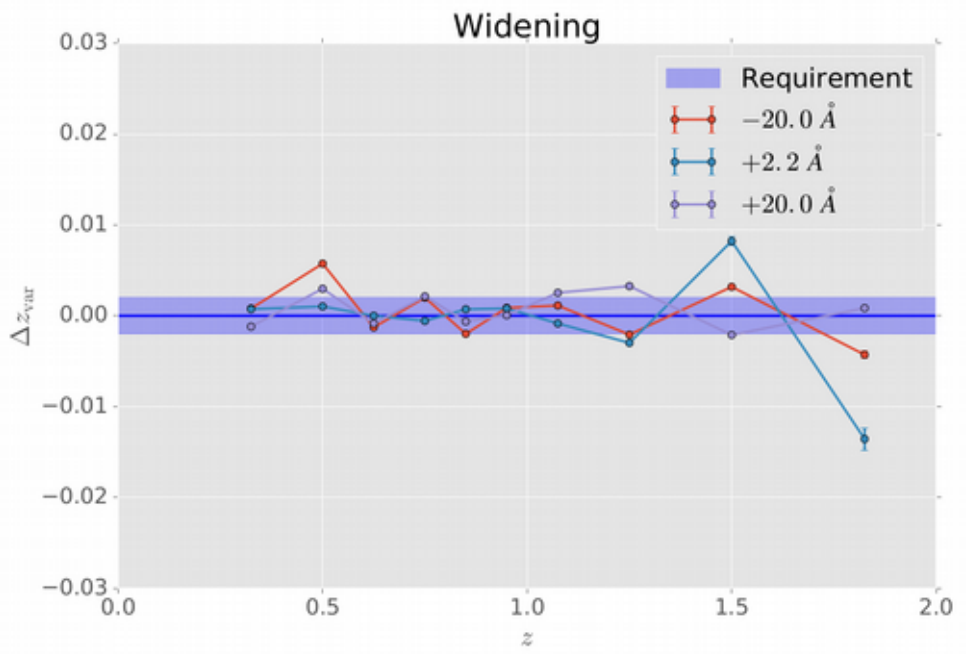


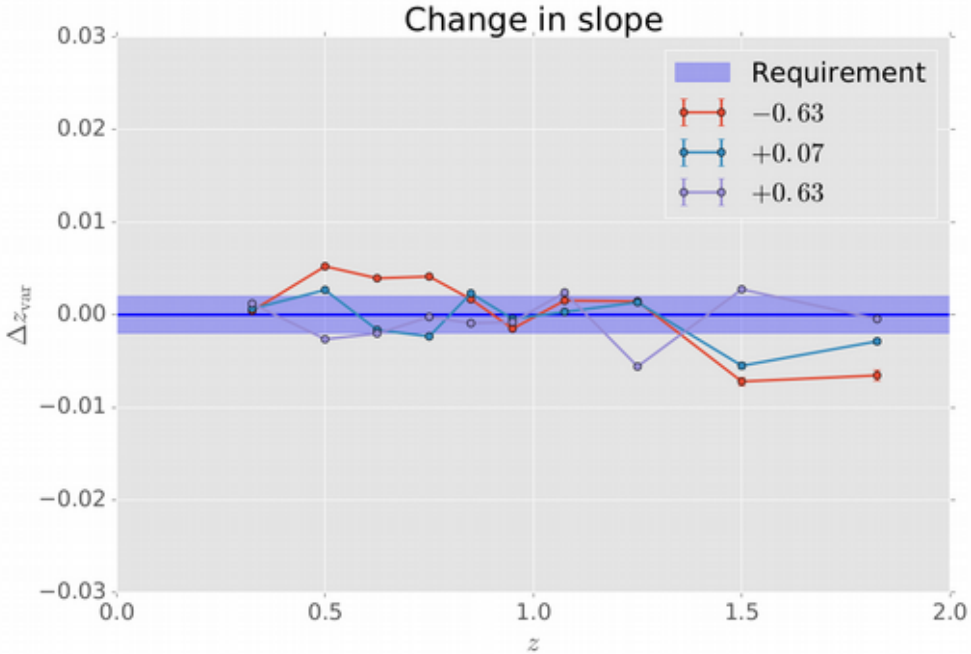
Photo-z's with variable filters



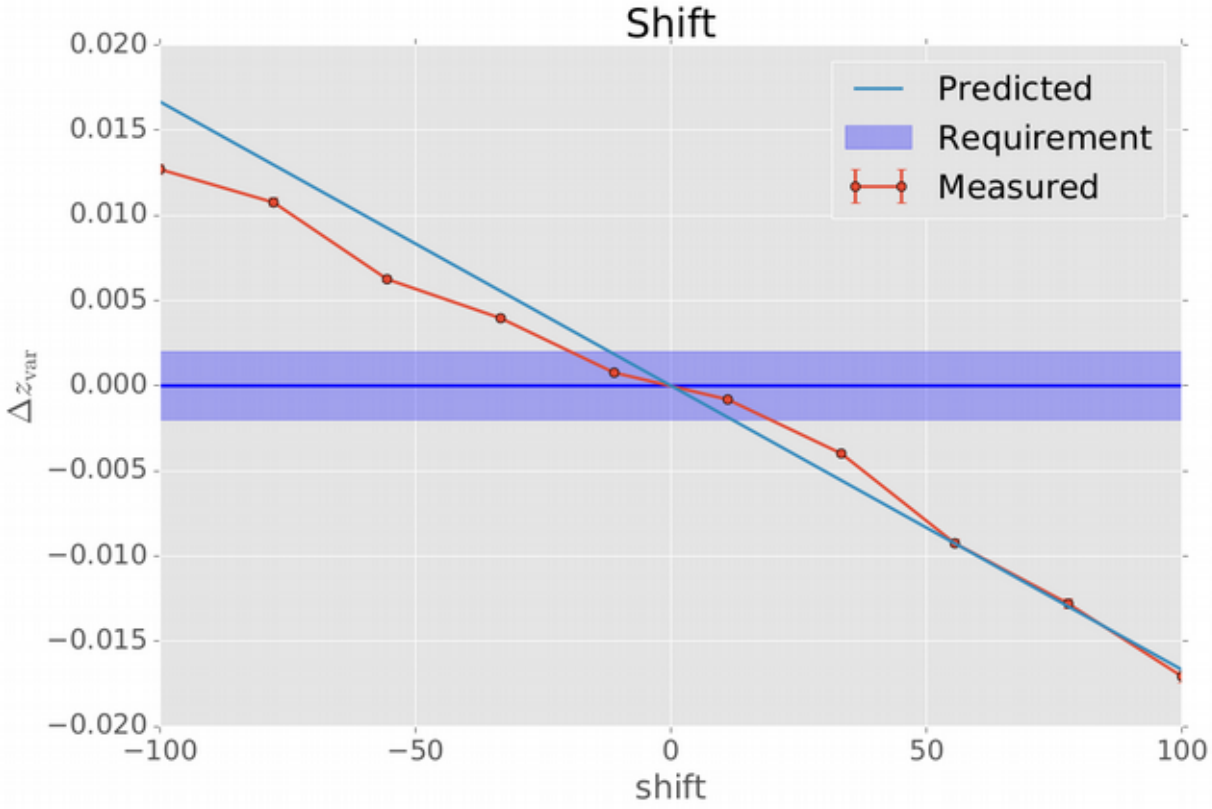


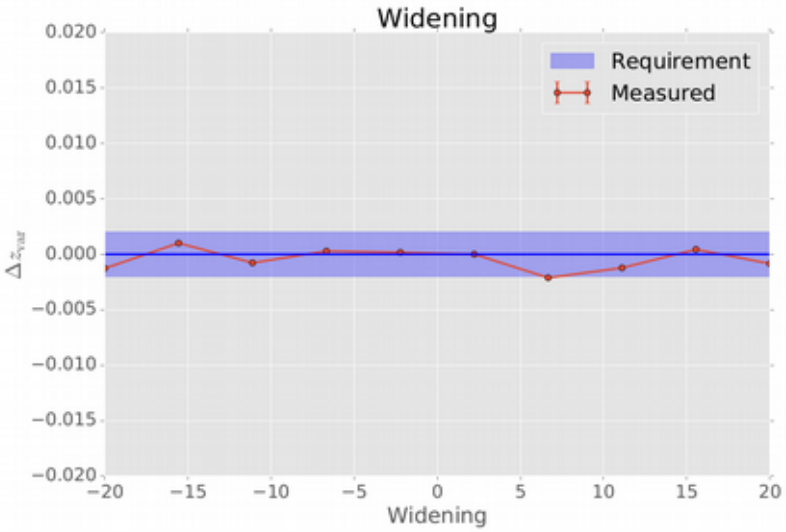


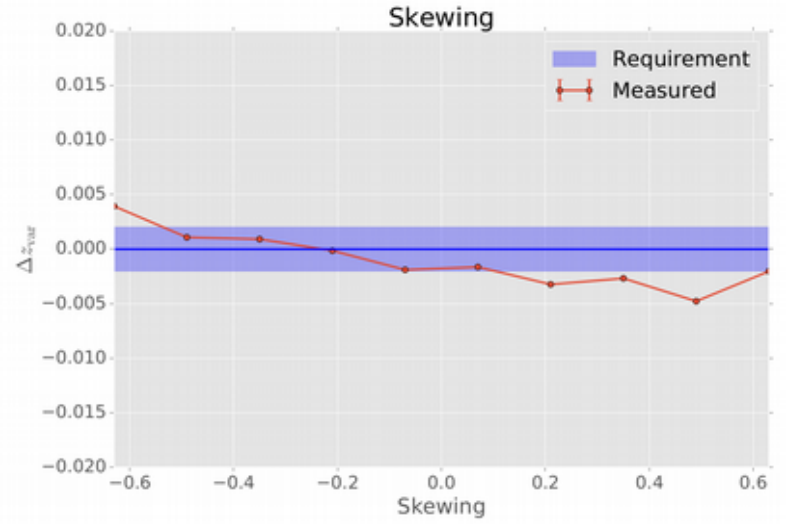




Bias: measured vs predicted







Conclusions

- Major effect is the change in central wavelength
- Width and skewness have much smaller effect
- Filter central wavelength shifts up to 10 nm observed
- Very good agreement with naive predictions (i.e., entirely due to the location of Balmer break)
- Color correction is redshift-dependent, so better not to perform any at the photometric extraction level

→ Proposal: MER provides PHZ with an effective central wavelength for each filter and for each object, with a precision of 0.2 nm