Photometric Redshifts for Euclid

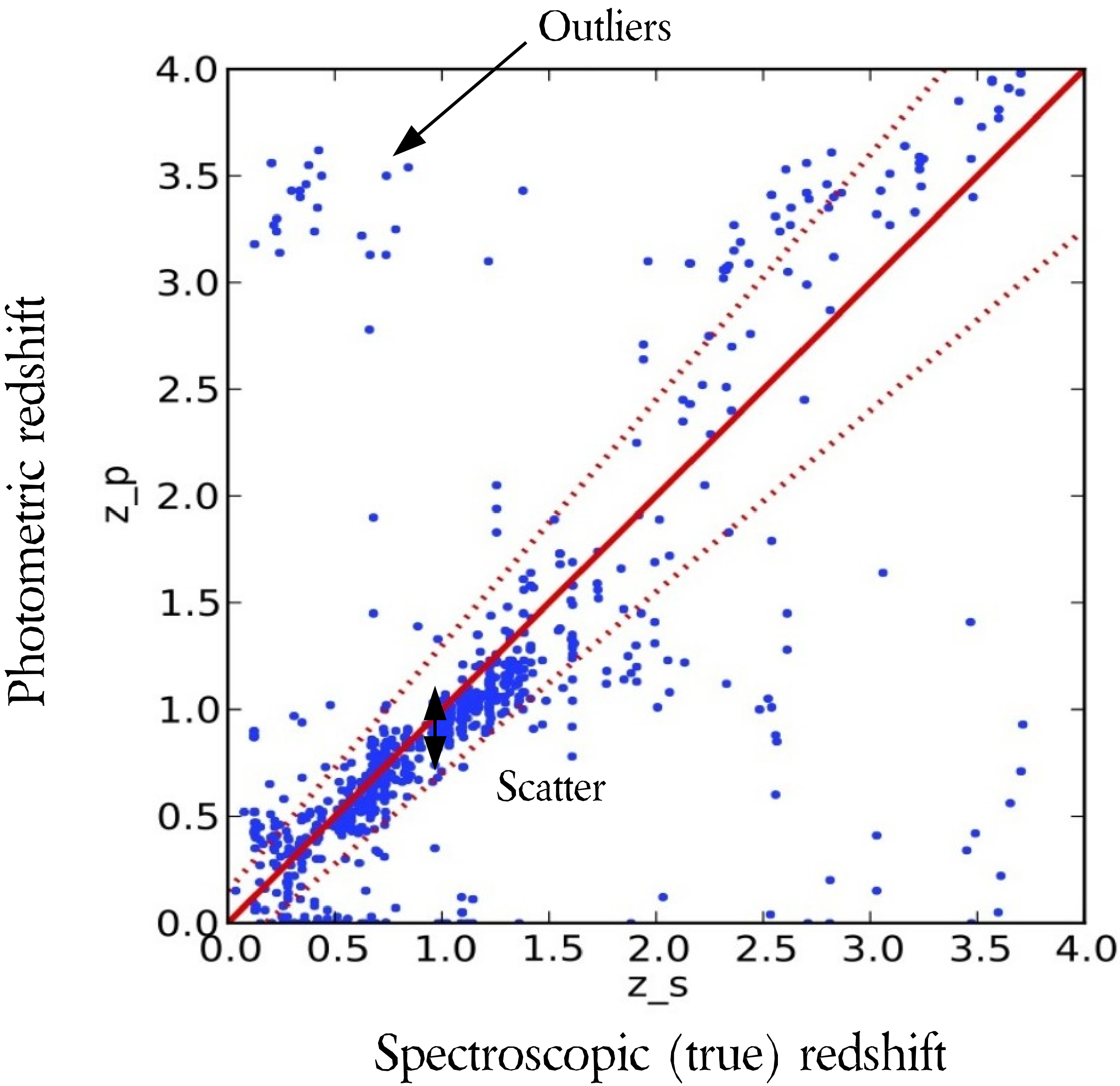
Stéphane Paltani
University of Geneva

With 75 more people from OU-PHZ and CH-SDC

Photo-z Workshop for Large Surveys, 仙台 , May 2017

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES
Département d'astronomie

Euclid photo-z requirements



Euclid requirements :

Scatter : $\sigma_z < 0.05(1+z)$

Outlier fraction : < 10 % beyond 0.15(1+z)

Bias: knowledge of $\bar{z}$ in any tomographic bin better than $0.002(1+z)$

Photo-z additional requirements

- For each object in the WL sample, provide the PDF of the redshift
- Perform star/galaxy(/QSO) separation
- Provide (observed) SEDs of the stars (for PSF determination). What about galaxies?

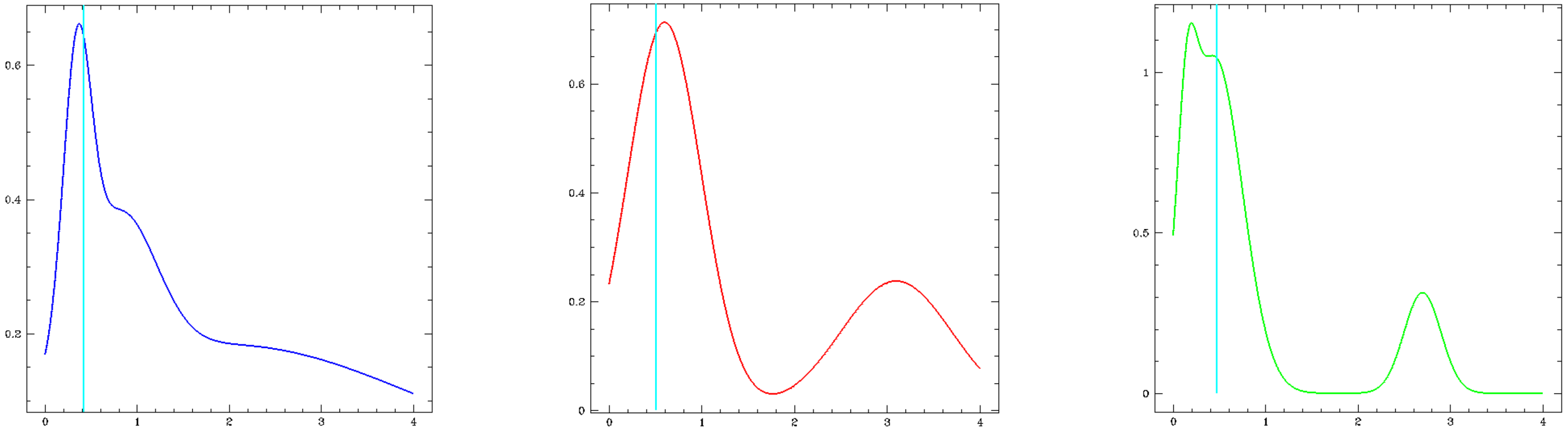
- Plus a whole lot of legacy science requirements

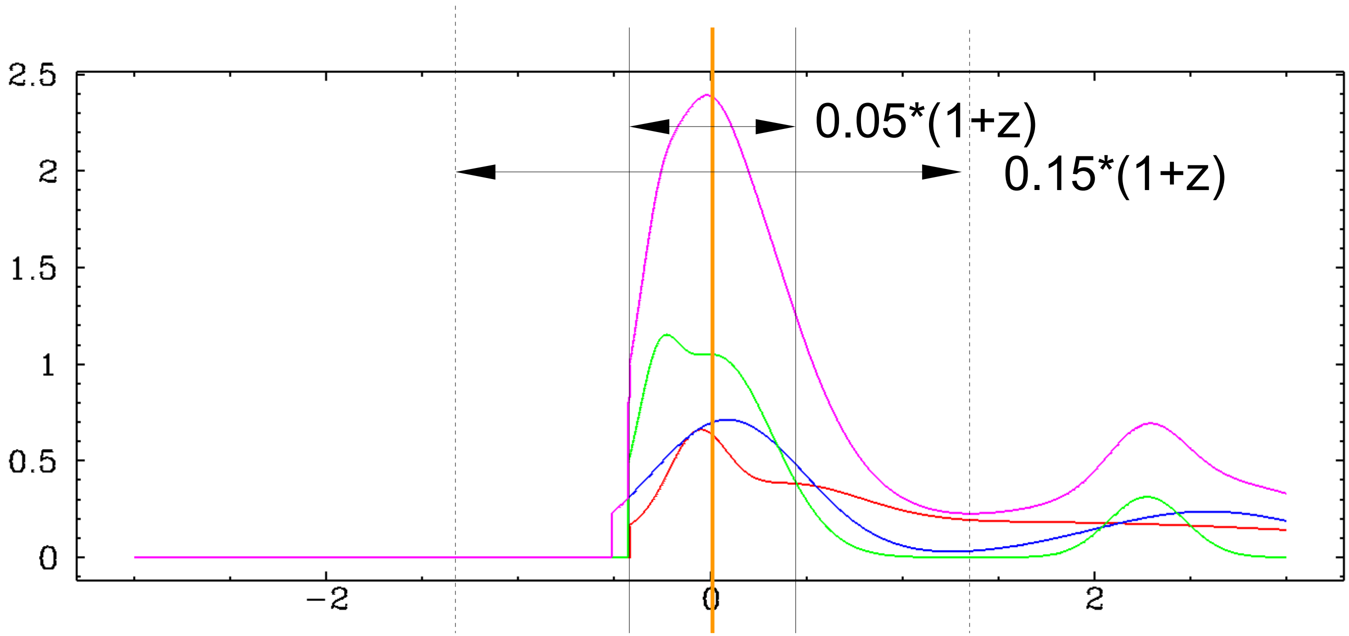
Photo-z requirement based on PDF

In each subset (bin) used for the weak-lensing analysis, the average of the true_z-subtracted PDF(z) (PDF(z-true_z)) shall meet the following cumulative probability requirements:

Within ABS(z-true_z)/(1+z)	Fraction of probability
0.05	68%
0.15	90%

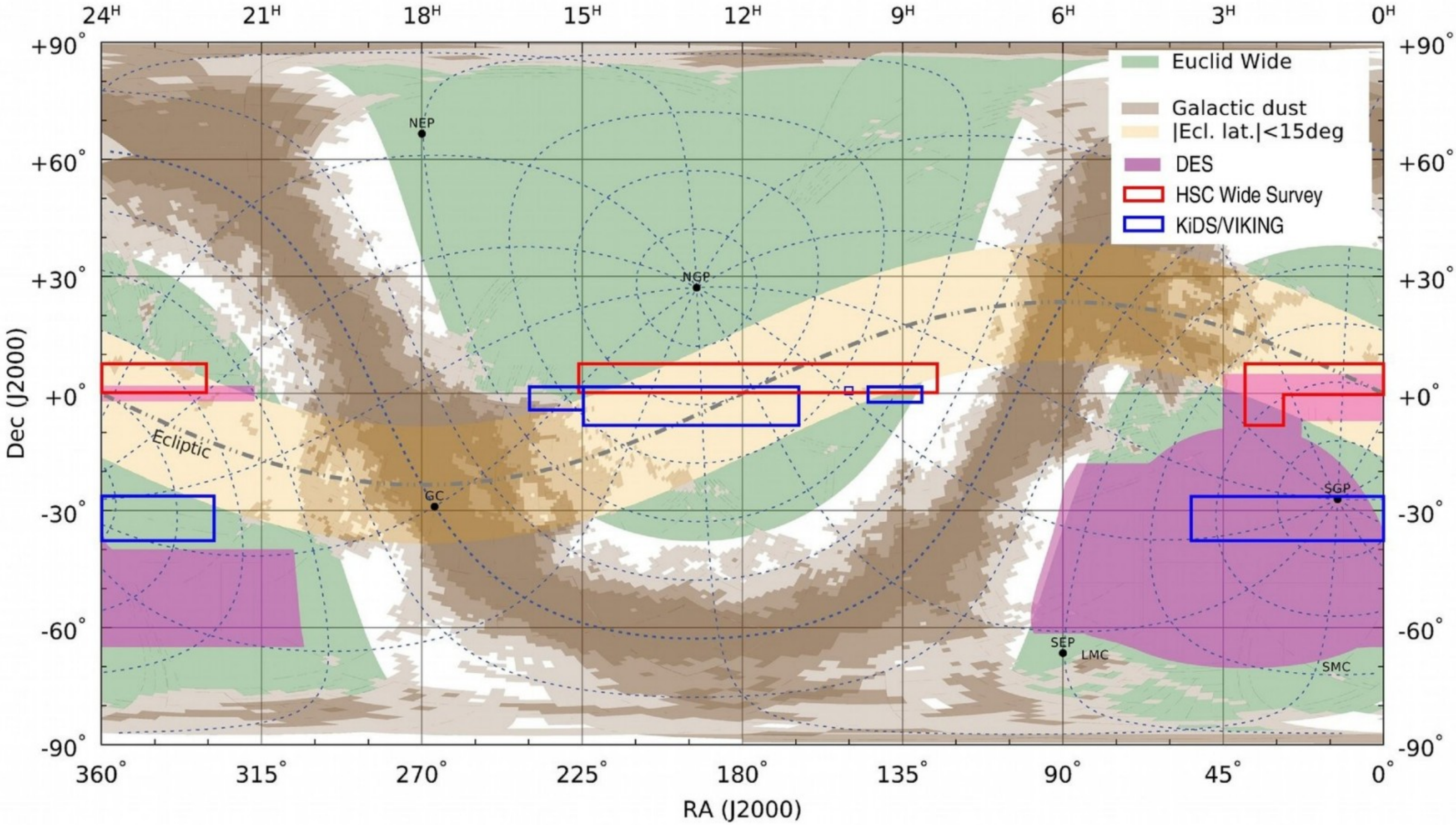
Photo-z requirement based on PDF



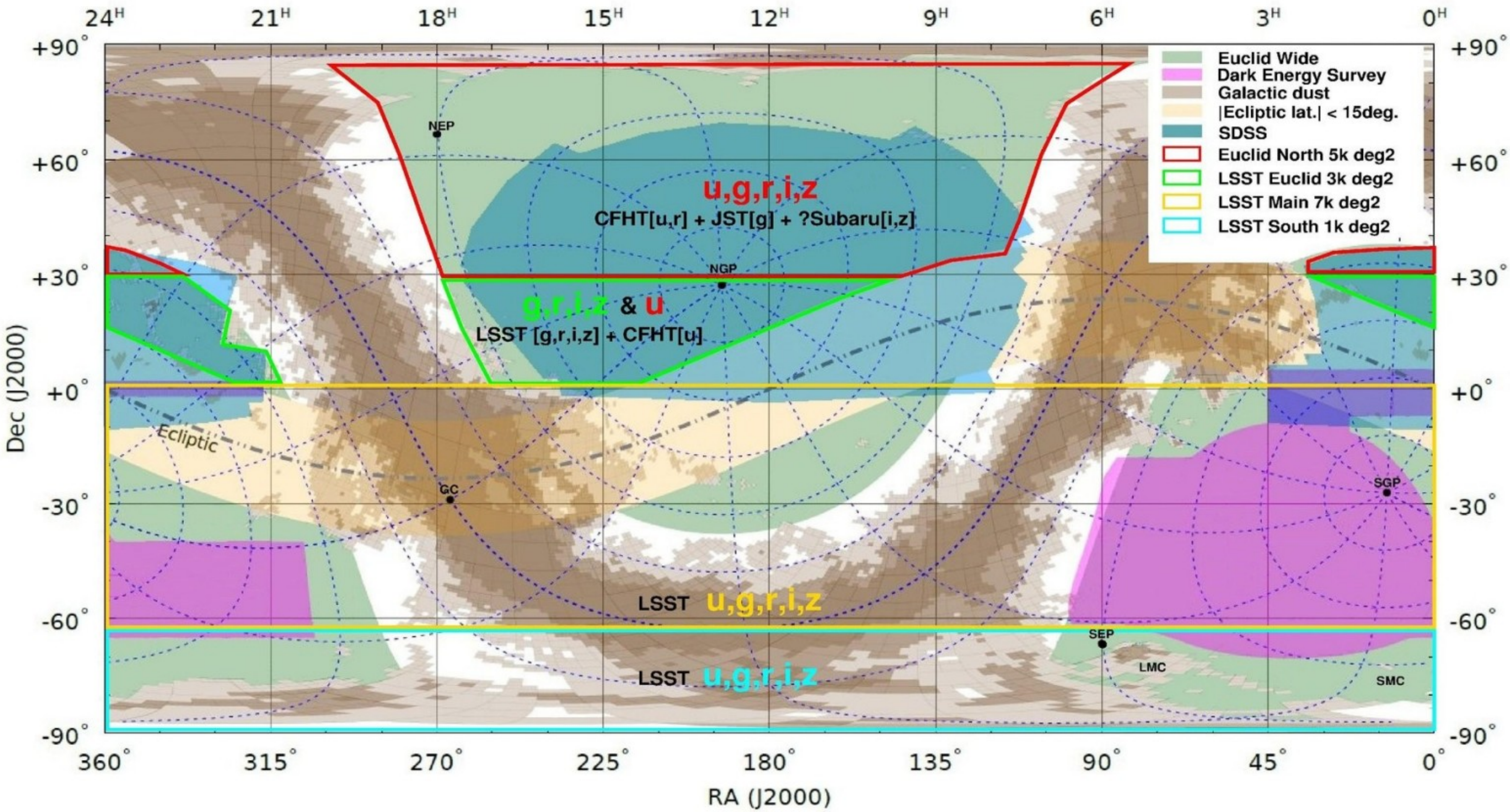


: 68 %
: 90 %

The Euclid wide survey



Optical ground-based data

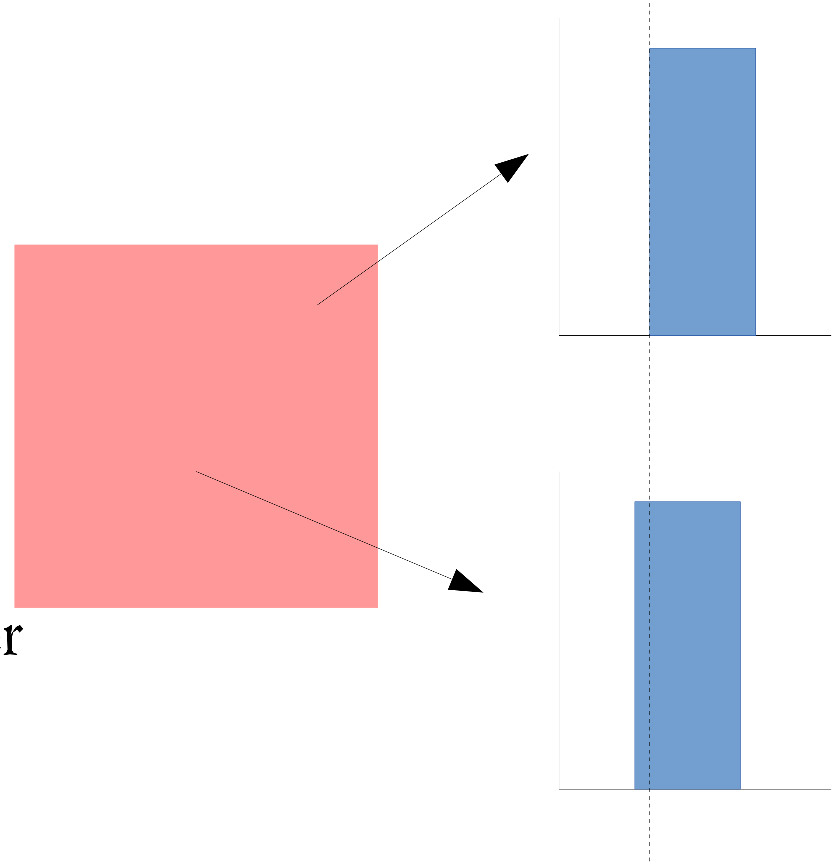


Calibration fields

- Not the Euclid Deep Fields!
- 4 ~1 deg^2 equatorial fields: COSMOS, SXDS, VVDS 2h, E-CDFS + 2 (GOODS-N, EGS) not so equatorial
- 25x wide exposure with Euclid, but also from optical surveys (?)
- Used to study the color distribution of galaxies
- And to build the color-redshift calibration relation (Dan's talk)
- Secondary calibration using astrometric redshifts (Vivien's talk)

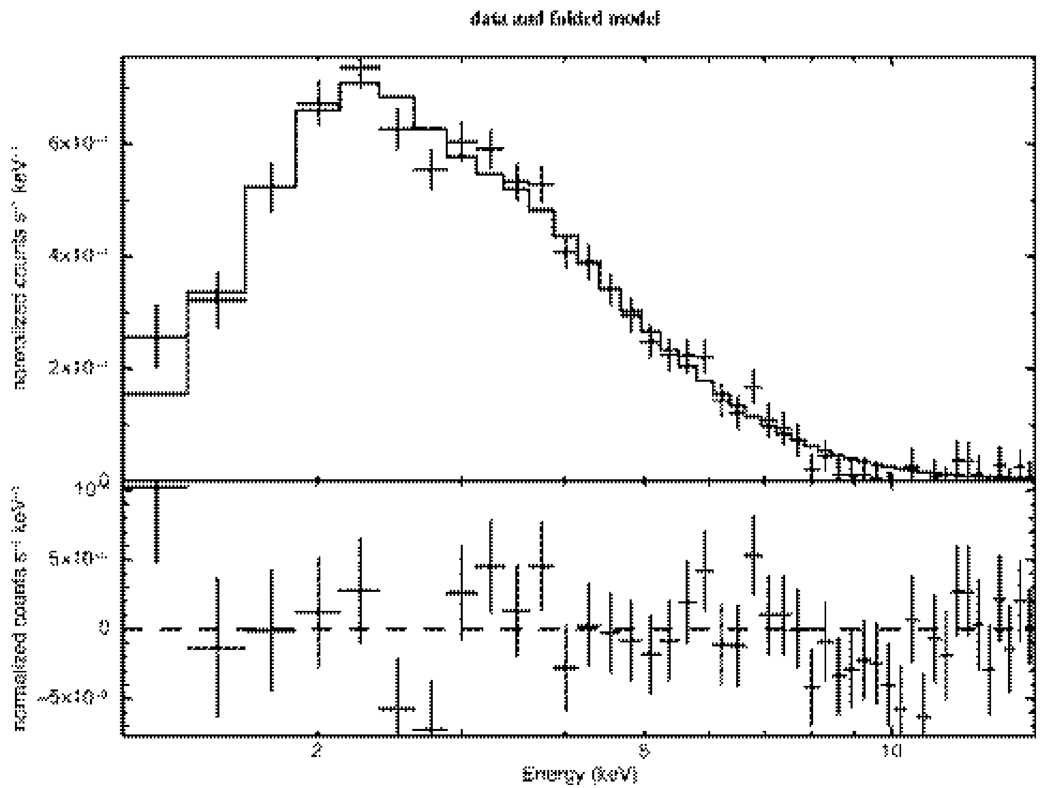
Variable filters

- Filter transmission is location dependent
- In the best case this introduces a scatter
- But probably a bias, if the filter shifts in wavelength



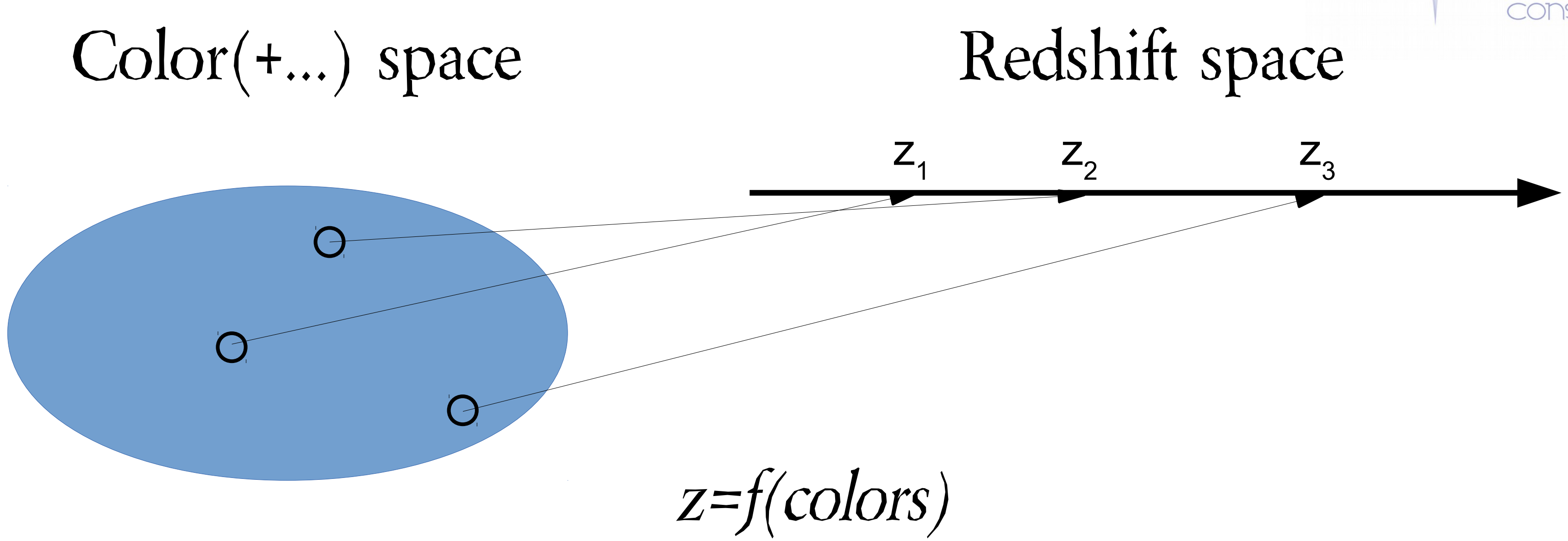
- Galactic absorption (Audrey's talk) is another source of fluctuation in the transmission
- And actually, don't forget the atmosphere… but we may be unable to do anything about that...

Photometry, the X-ray way



- In X-ray astronomy, each observation comes with its own response. The source spectral properties are then obtained by forward fitting an emission model through the response to the count rates
- This can be “easily” treated with template-fitting algorithms
- Can we fix the colors for ML? (Jean's talk)

Photometric-redshift algorithms



Mapping f can be constructed based on prior knowledge :

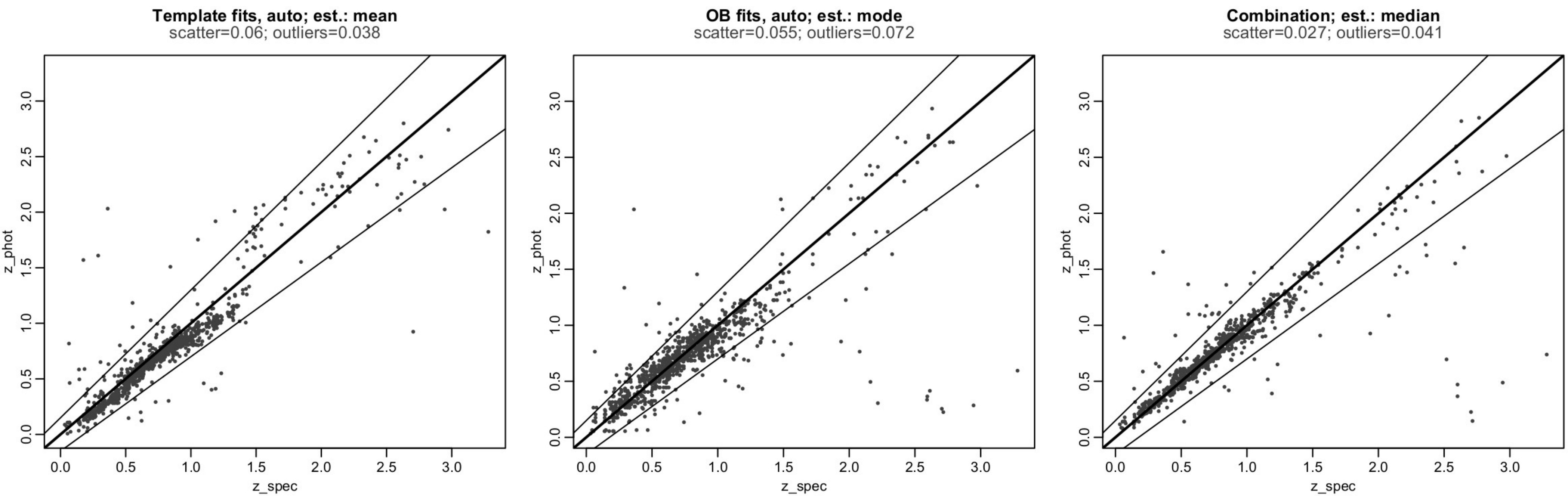
- Template-fitting: Hyper-Z, Le Phare, BpZ, Phosphoros,...

Or it can be discovered:

- Machine-learning: Nearest neighbors, Perceptron, Support vector regression, Random Forest, Adaboost, Gaussian Processes, …

Both have advantages and disadvantages; we probably want to use both

Combining TF and ML (with ML)

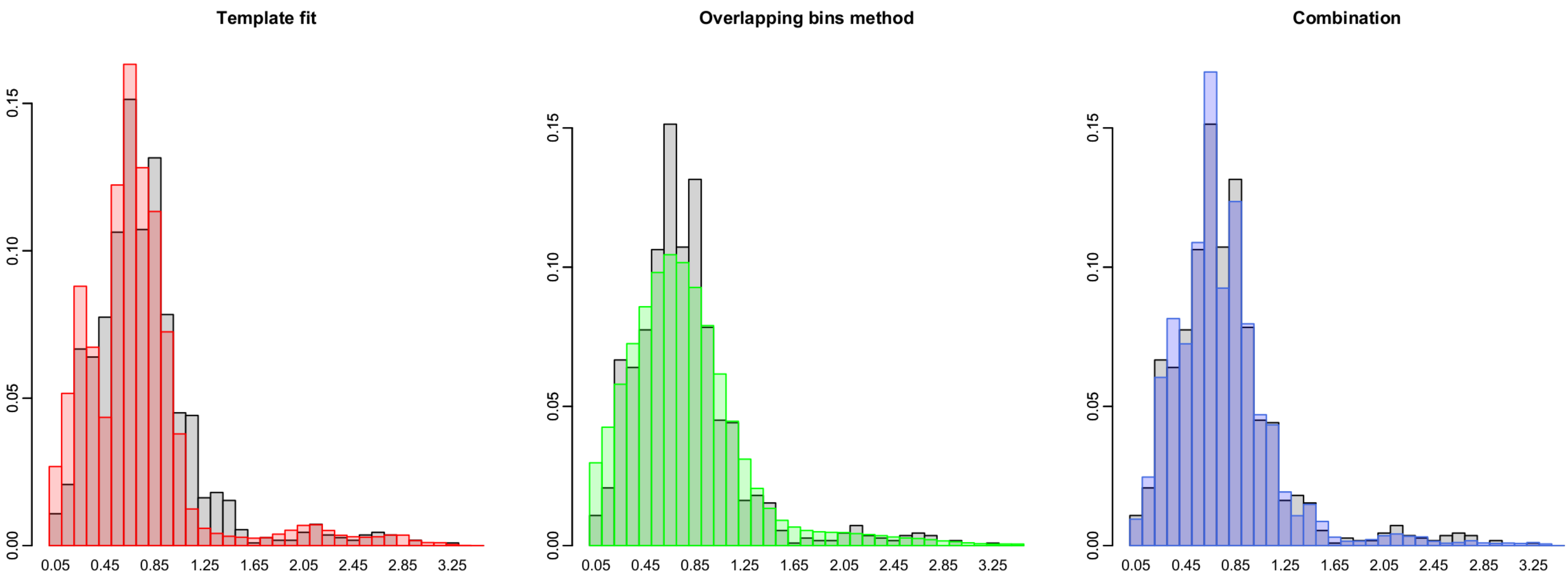


Le Phare TPZ Combination

Classifier-based combination (Random Forest)

Süveges et al. in prep.

N(z) Reconstruction



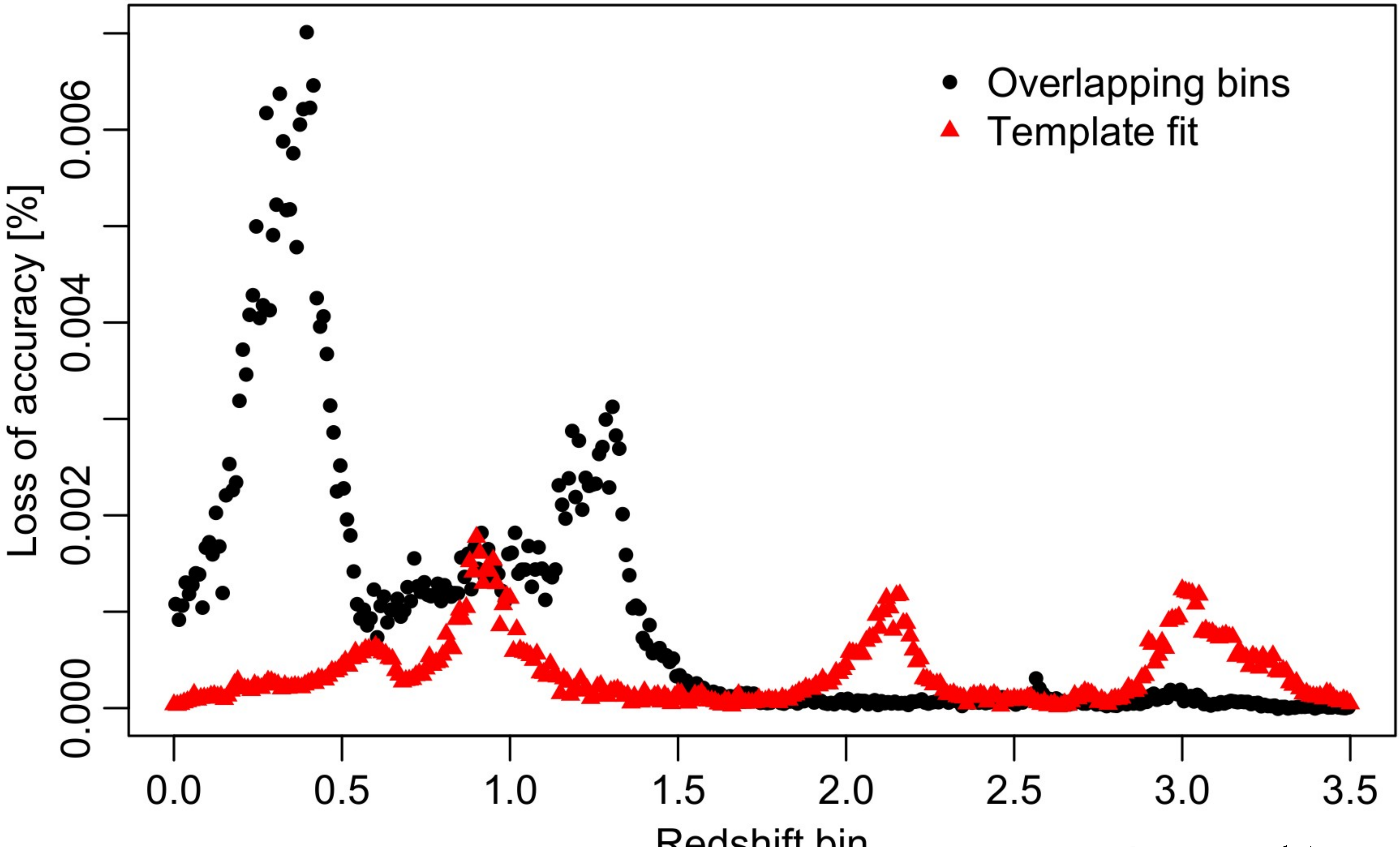
Le Phare

TPZ

Combination

Süveges et al. in prep.

Feature importance



Süveges et al. in prep.

Conclusions

- Euclid will use inhomogeneous optical survey
- Outlier fraction and scatter requirements not so hard, if one has good photometry
- But very stringent requirements on the bias
- Ready to cope with variable transmissions ??
- We are probably using more than one photo-z algorithm

Template-Fitting Advantages

- ✔ Based on astrophysical knowledge; the better the knowledge, the better the algorithm
- ✔ Any physical process that is understood can be modeled explicitly (e.g., Galactic absorption)
- ✔ Constructs naturally a likelihood, and can be turned into a fully Bayesian approach
- ✔ Can cope with informative priors in a very natural way, e.g. luminosity function, cosmological volume

Template-Fitting Disadvantages

- ✗ Knowledge of the sky is imperfect (wrong templates) and incomplete (lack of templates)
- ✗ No clear guidelines on the number of templates (not a continuous quantity)
- ✗ Computationally intensive
- ✗ Cannot easily cope with additional features (galaxy shape, etc. ; but is it useful ?)
- ✗ Link between photometry and galaxy properties not so clear (e.g., aperture effects)

Machine-Learning Advantages

- ✔ No need to understand the astrophysics or to model any physical process

- ✔ Can easily incorporate additional features, e.g., different types of photometry; good ML algorithms can do it without loss of stability

- ✔ A sound ML algorithm will be optimal where training set is "good"

- ✔ Not linked to galaxy properties, so photometry does not really matter

Machine-Learning Limitations

- ✗ Many algorithms cannot produce naturally a PDF
- ✗ There are “hidden priors” in the selection of the training set
- ✗ The training set must be “good” whatever that means
- ✗ There might be over- or under-fitting if the model complexity is not chosen properly
- ✗ Extrapolations might/will occur if the training set is incomplete

But is ML better ?

- Results depend strongly on the quality of the training set
- Training set and test set generally come from the same population
  - Meaningful comparison must at least use a weighting scheme (e.g., Lima et al. 2008)
  - Any missing population will probably be better characterized with template-fitting
- Template-fitting involves some “black magic”, so the result depends a lot on fine tuning