Training Sample Machine for Euclid

Jean Coupon

Photo-z workshop, Sendai, May 19th 2017

Overview

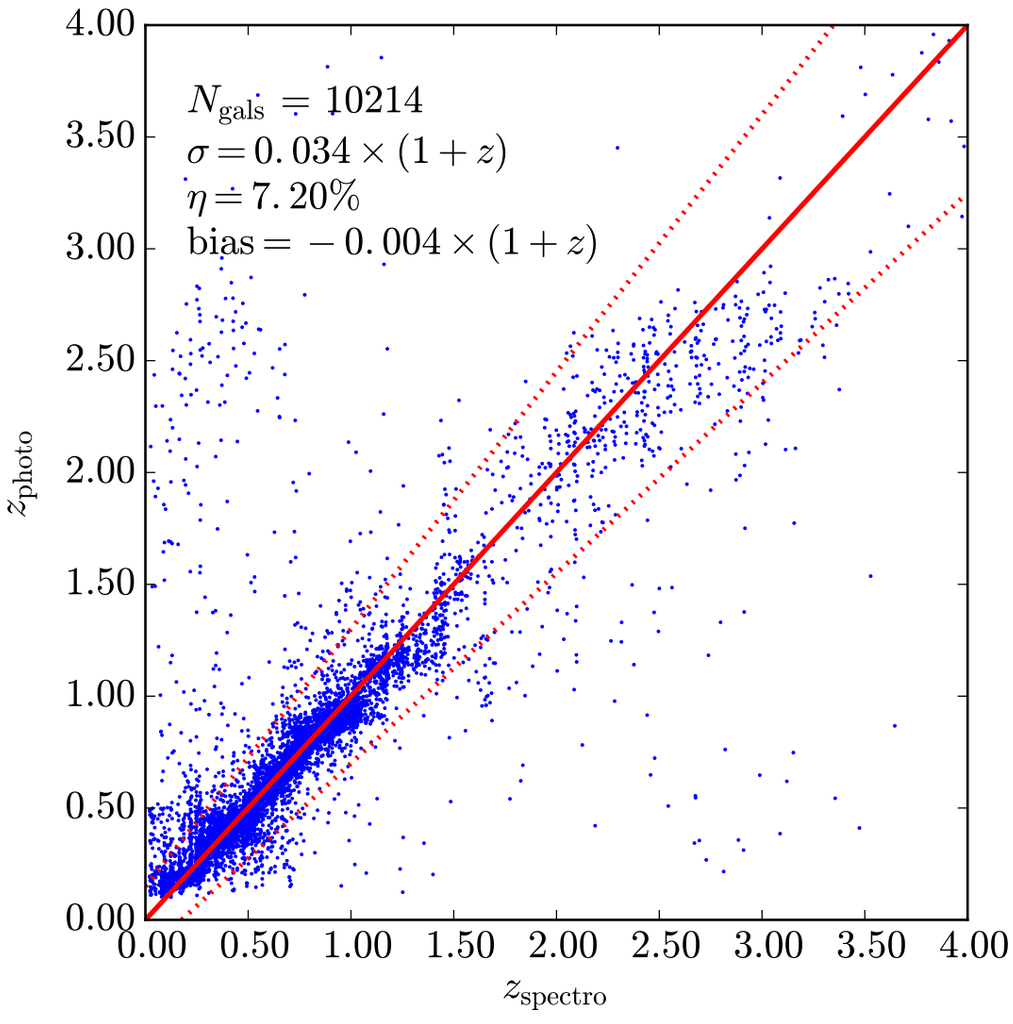
- Euclid: 0.2% precision needed on <z>
- under 1%-level photometry is required
- for machine learning: need large training sample, unbiased, and in same photometric system as unknown sample
- but:
  - many systematics: extinction, filter response, sky
  - not all training redshift fluxes are observed in the same response function

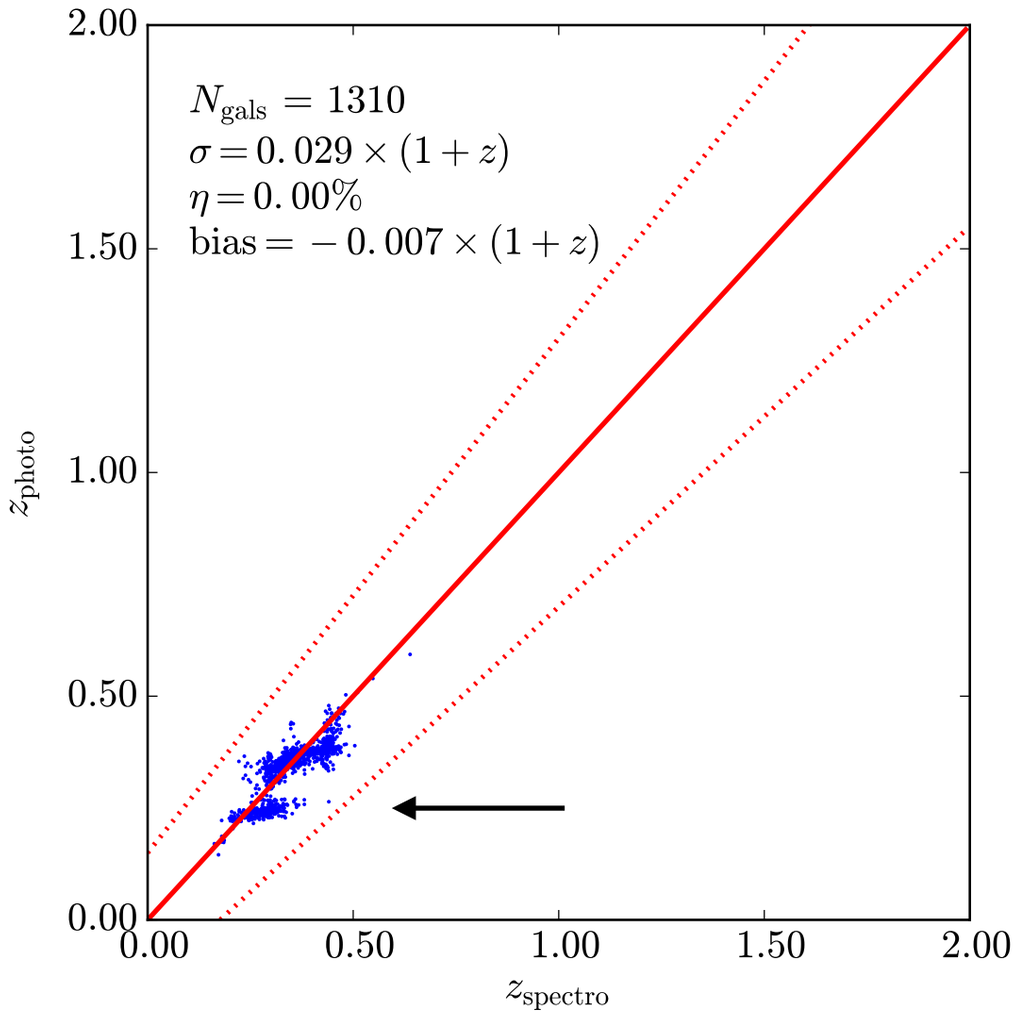
Building a proper training sample

Example of a biased training sample

trained with COSMOS

tested in BOSS (other field)





Other sources of bias

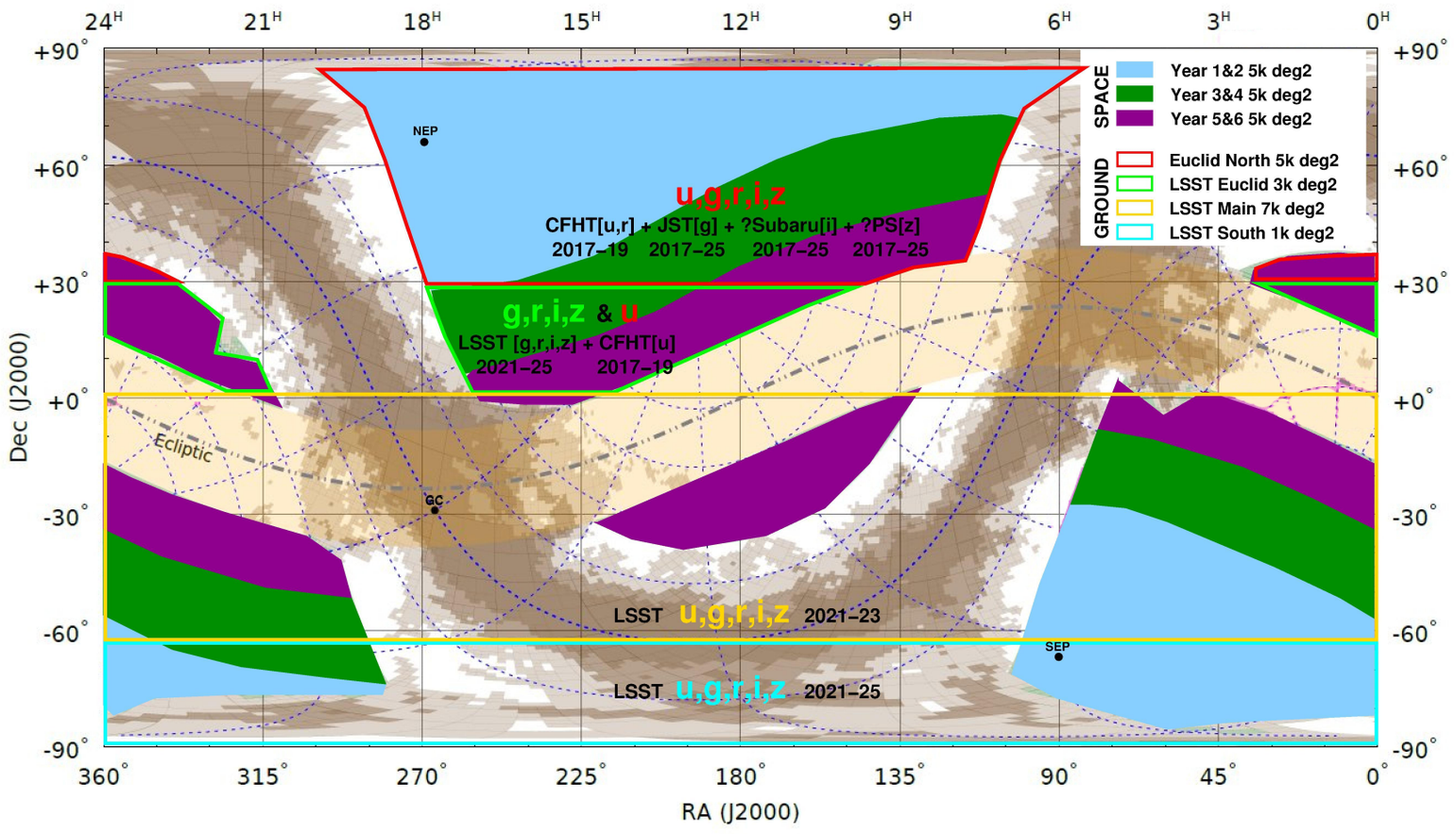
Unknown-source colors may differ from reference sample:

- telescope visibility (different photometric system)
- Galactic extinction
- photometric calibration (including sky absorption)
- filter/optics color terms on focal plane

Different photometric systems

- simple case: AEGIS galaxies' fluxes measured with HSC, re-generated in LSST filter system

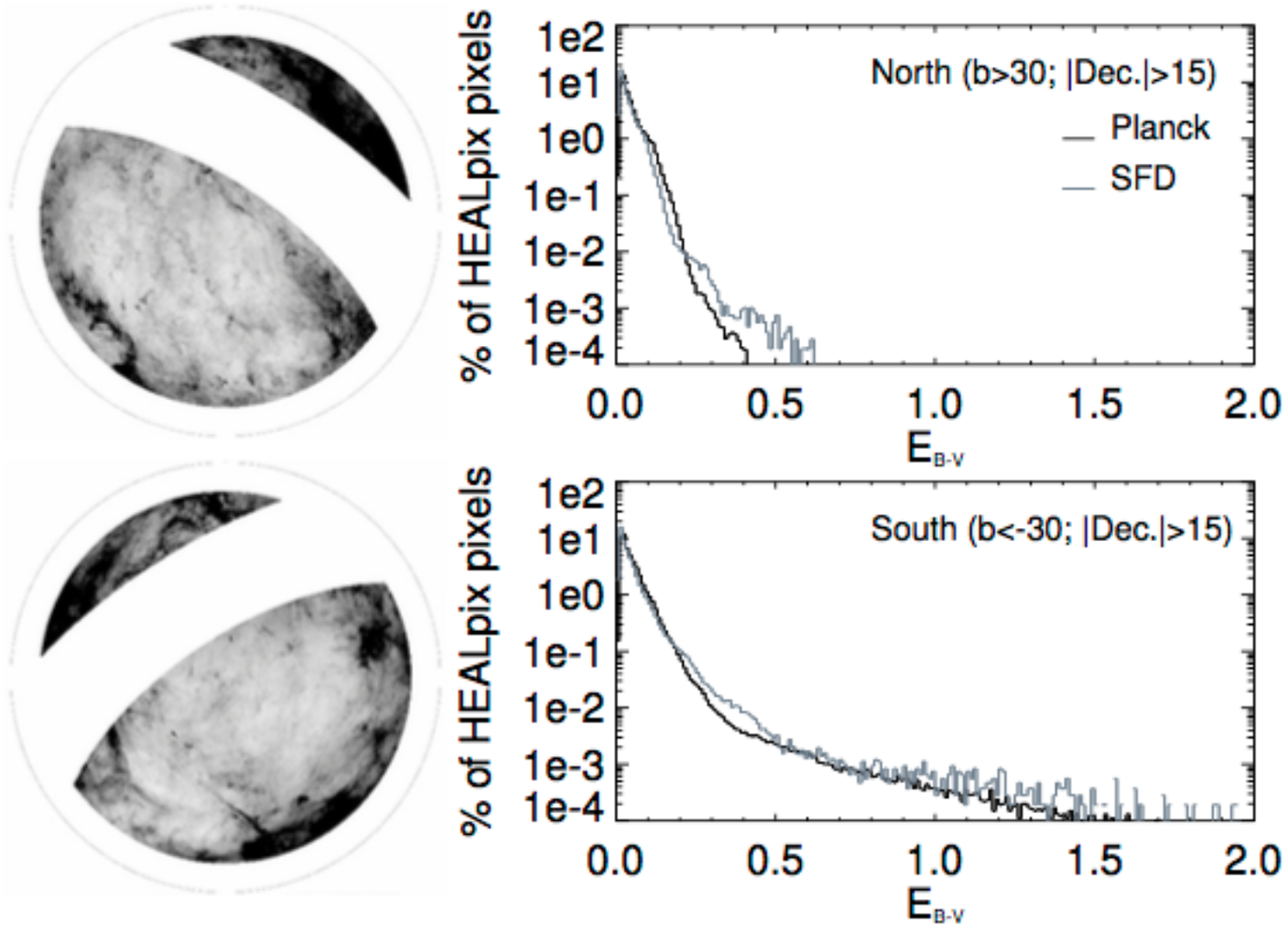
too far north for LSST



J.-C. Cuillandre

Galactic extinction

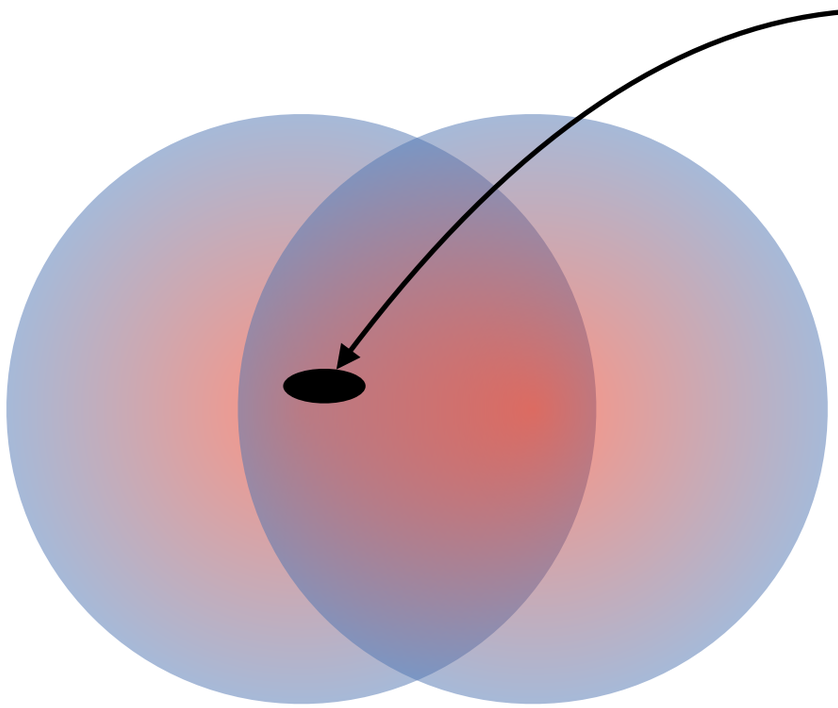
- complicated case: a training sample for each extinction value?



Galametz et al. (2017)

Color terms in filter transmission

- even more complicated case: instrumental color-term on the focal plane



filters have color terms, image co-addition makes it worse

Training sample machine

- idea: collect one calibrated SED for each reference redshift (spectroscopy or template fitting)

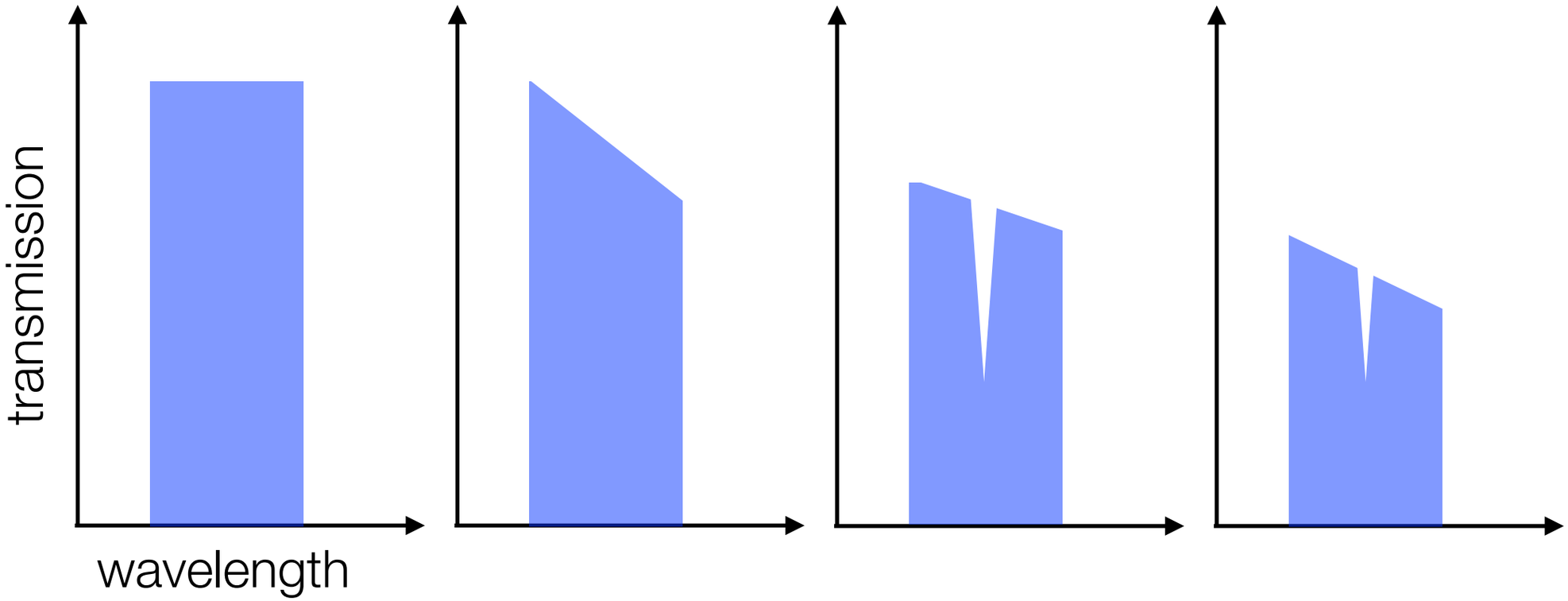
- re-compute fluxes of training sample for the response function used to observe unknown source

- (here response function means all physical and instrumental effects: filters, extinction, sky, etc.)

- run machine-learning/nnpz algorithms

A concrete example

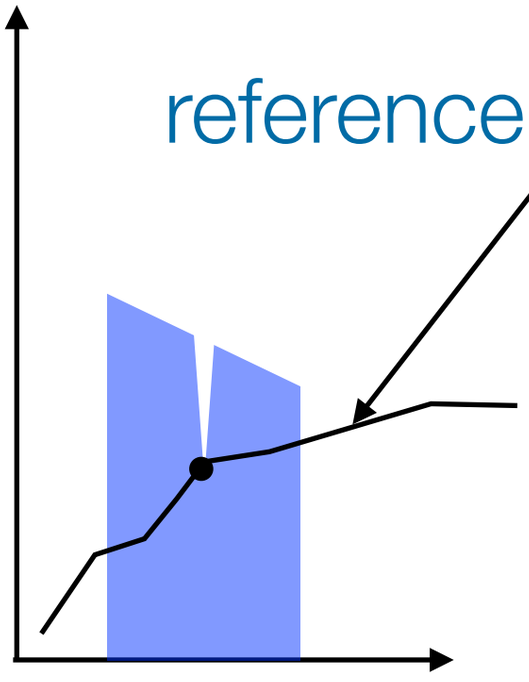
- gathering everything into one single response function



filter +extinction +sky +filter color term

A concrete example

- and evaluating fluxes



final response function

training sample fluxes in proper response function

training sample table

z	mag1	mag2	mag3	…
0.23	24.25	25.23	26.01	
0.56	23.78	24.56	24.12	

What if we don't know the response function?

- treat it as unknown:
  - assuming we can derive bias = f(response function):
  - predict it (simulations)!
    - derive distributions (as a function of position, mainly)
    - add uncertainties to response function: extinction, filter variation
- but whenever possible, use the proper response function

Numerical considerations

- extreme case: one training sample per object (= per position on the sky and on the focal plane)?

- not feasible

- trade off: build a number of training samples and interpolate in color space, re-compute fluxes only for neighbours

- number and sizes of training samples?

Conclusions

- training sample evaluated in same proper response function as unknown source
- key to beat biases and variations in filter transmissions
- need template fitting code and large reference sample to build reference spectra
- challenges:
  - SEDs (what precision?)
  - knowledge of the response function (to be investigated)
  - what precision for the training sample / how many samples?