Likelihood Ratio Method

JS Ricol DESC Meeting Pittsburgh

Outlook

- Main ingredients of the simulation:
  - Galaxy catalog
  - SEDs: interpolated templates

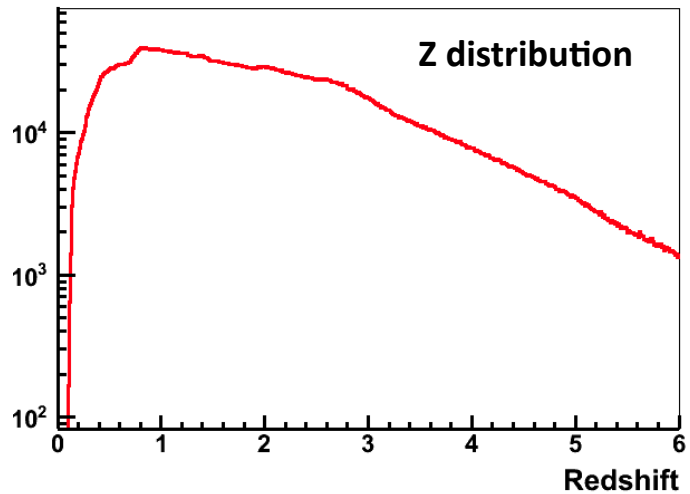
- Template fitting PhotoZ reconstruction

- Likelihood Ratio method

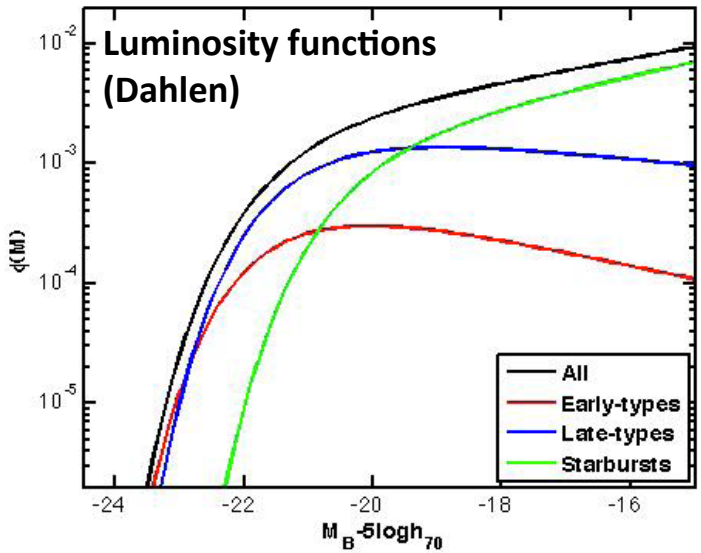
- Neural network and template fitting correlation

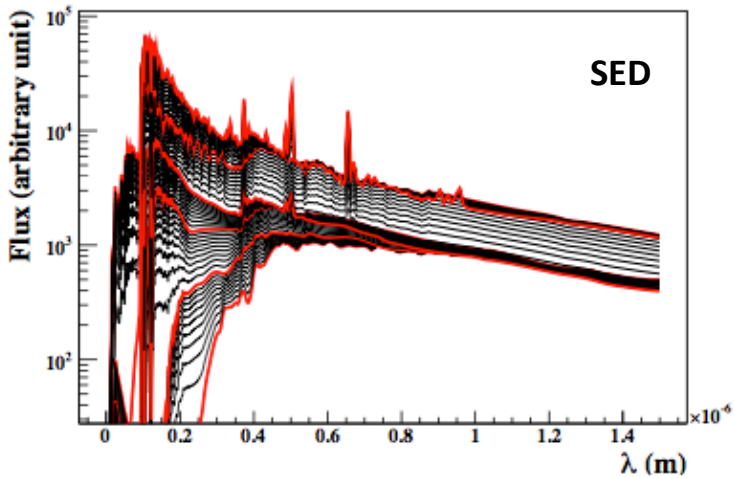
Galaxies catalog

Home made catalog of 8.10^9 galaxies
Fits files produced by A. Abate
Theoretical early density fluctuations + standard cosmology
-> galaxies distribution: ra, dec, z



GOODS luminosity functions: M, BType





SEDs Libray of 51 templates interpolated from 6 SEDs
El (type=0), Sbc (type=10), Scd (type=20), Im (type=30) - *Coleman et al*,
SB1 (type=40), SB2 (type=50) - *Kinney et al*

SEDs are extrapolated into UV using GISSEL synthetic spectra (*Bruzual, Charlot*)

Fits to Root : 5000 root subcatalogs of 10,000 galaxies each

Apparent magnitude

- Dust reddening: random uniform distribution

  E(B-V) = [0, 0.1] for El, [0, 0.3] for others

  Attenuation law : Cardelli (El, Sbc, Scd), Calzetti (Im, SB)

- IGM extinction : Madau

- Apparent magnitude errors

  LSST : $\sigma_X^2 = (0.04 - \gamma_X)x + \gamma_X x^2 + \sigma_{syst}^2$

  $$x = 10^{0.4(m_X - m_{5,X})}$$

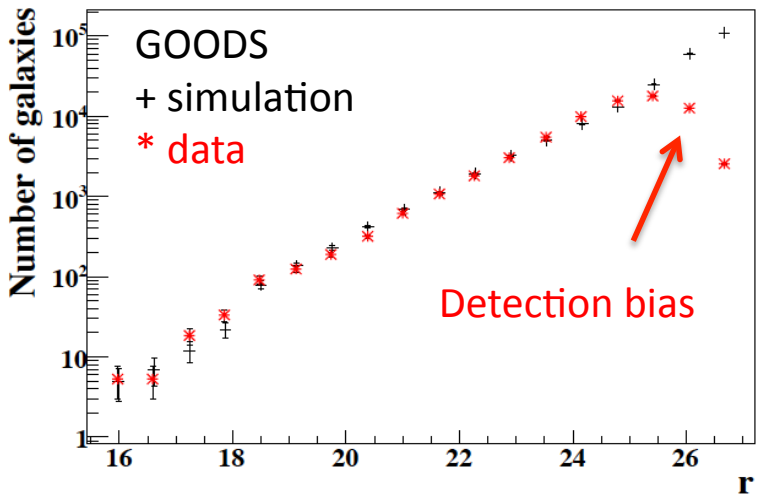
  $$\sigma_{syst} = 0.005$$

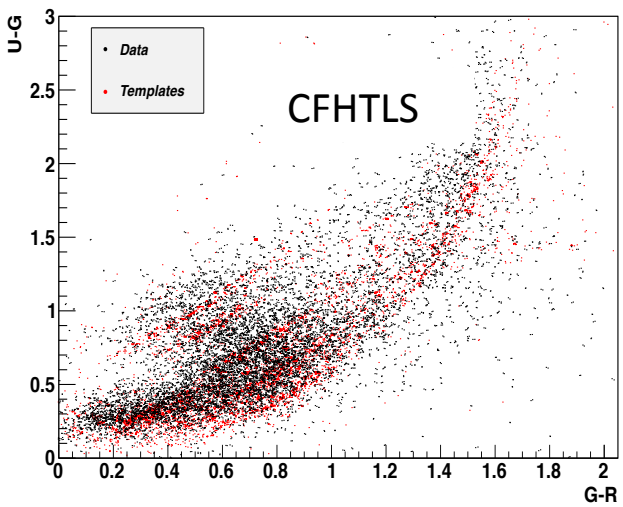
  Previous surveys : error from data

LSST 10 years of observation

	u	g	r	i	z	y
Exp. :	56	80	184	184	160	160
$m_{5,X}$:	26.1	27.4	27.5	26.8	26.1	24.9

Good agreement with data





Template fitting method

Reconstruct θ={z, type, E(B-V), N} from **m** = {m_X}

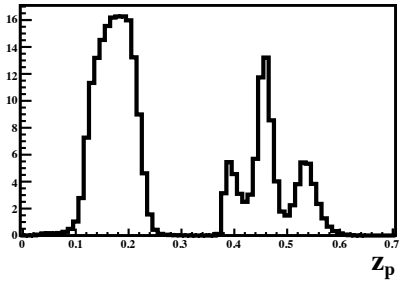
χ^2 computed on a 3D grid (z, type, E(B-V))

$$\chi^2(z,T,E(B-V),N) = \sum_{i=1}^{N_{bandes}} \left(\frac{F_i^{obs}(m_i) - N F_i^{mod}(z,T,E(B-V))}{\sigma\left(F_i^{obs}(m_i,\sigma(m_i))\right)} \right)^2$$

Prior (Benitez) : $P(z,T|i) = P(z|T,i) \times P(T|i)$

3D posterior pdf

$\mathcal{L}$ = exp-(χ^2/2)



z_s = 0.161599

T_s = 45

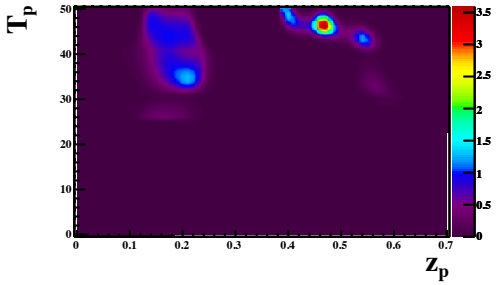
$E(B-V)_s$ = 0.244462

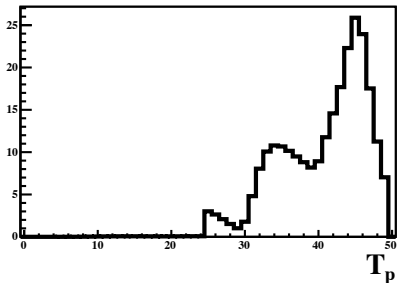
z_p^{grid} = 0.46748

T_p^{grid} = 46.0017

$E(B-V)_p^{grid}$ = 0.119983

One bad exemple :

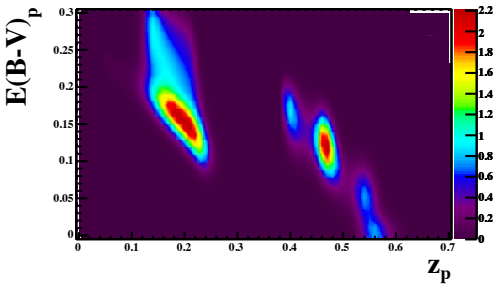


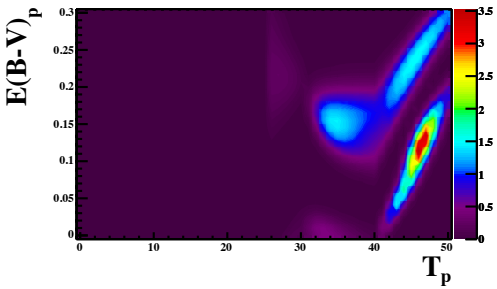


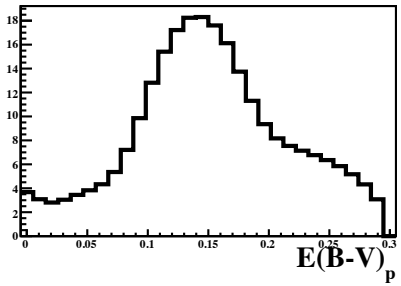
z_p^{marg} = 0.2

T_p^{marg} = 46

$E(B-V)_p^{marg}$ = 0.15



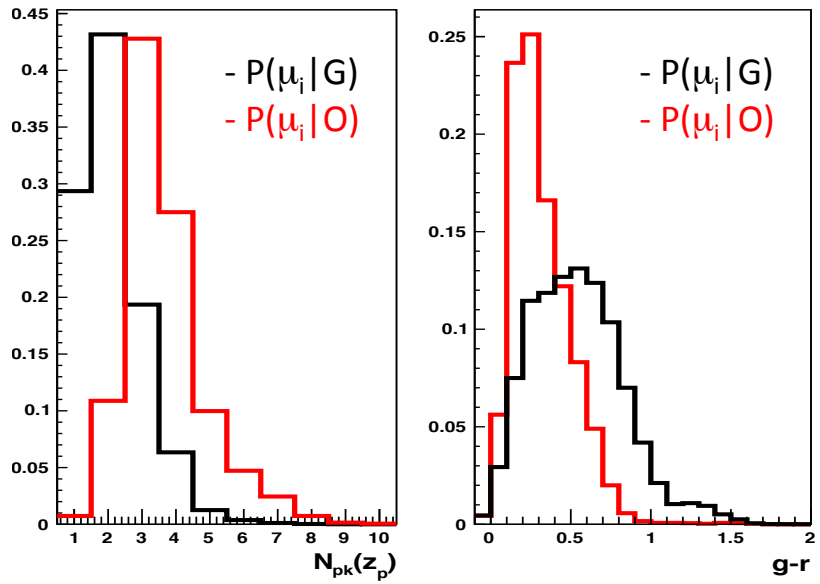




Likelihood Ratio

LR statistical test on 16 parameters μ_i :

- number of peaks in the marginalized 1D PDF: $N_{pk}(z)$, $N_{pk}(T)$, $N_{pk}(E(B-V))$,
- relative width/height between main and 2nd peak,
- χ^2 value,
- $z_{\chi 2} - z_{marg}$
- color terms (u-g, g-r, r-i, i-z, z-y)

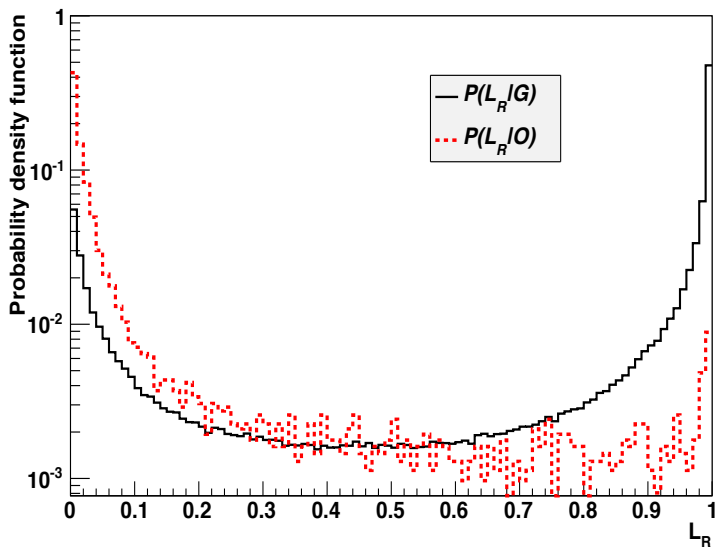


$P(\mu_i|G)$, $P(\mu_i|O)$ computed on a training subsample

O : |zp-zs|/(1+zs) > 0.15
G : |zp-zs|/(1+zs) < 0.15

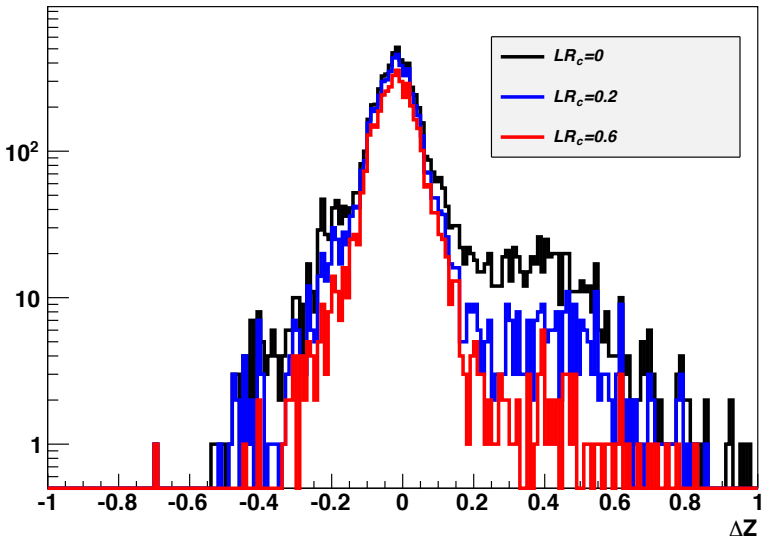
Then we can calculate LR for each galaxy in the remaining catalog and use a quality cut on this parameter $LR>LR_c$ to reject outliers

Likelihood ratio : $$L_R(\mu) = \frac{P(\mu|G)}{P(\mu|G)+P(\mu|O)}$$



Outliers rejection

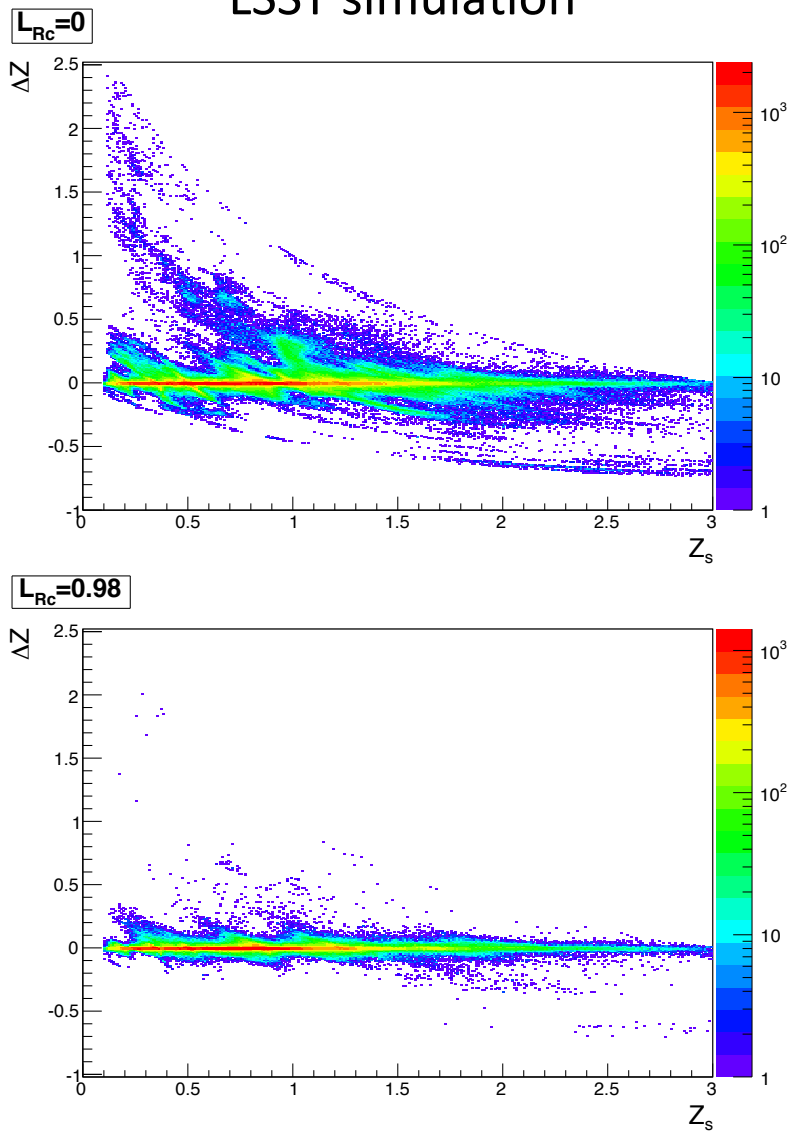
CFHTLS data



$\Delta z = (zp - zs) / (1+zs)$

LRc	Mean	RMS	Outliers
0	0.01	0.16	12%
0.2	-0.006	0.12	11%
0.6	-0.01	0.08	2.4%

LSST simulation



LR cut (LR ≥ LRc) is very powerful to reject outliers
Our study on CFHTLS data showed that LR could be trained either on data or simulation with similar results.

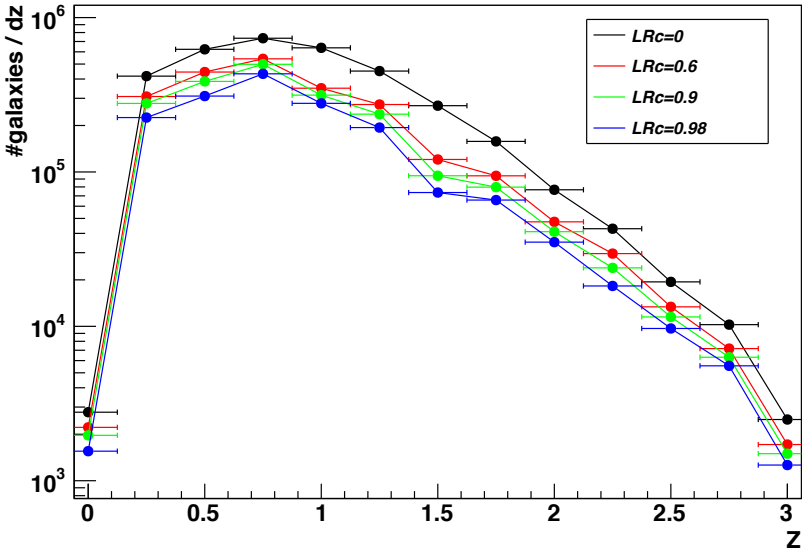
PhotoZ analysis for LSST

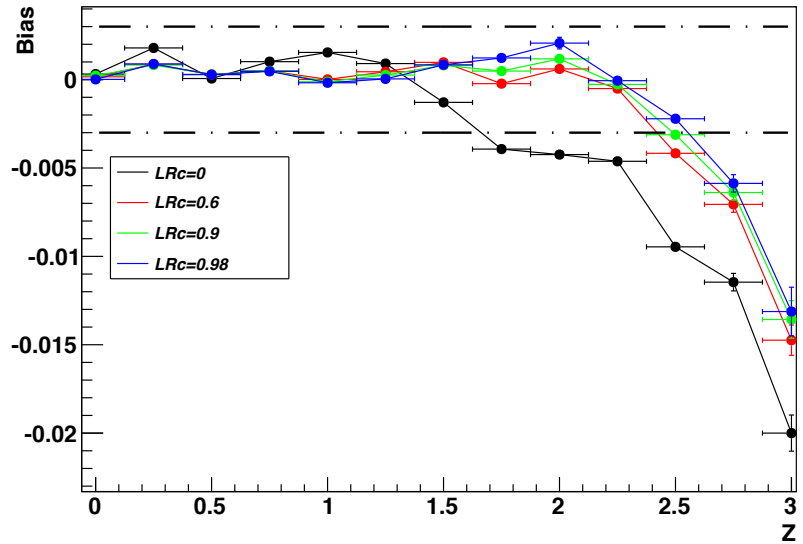
Catalog:
10^7 galaxies
~ 10^6 galaxies with S/N>5 in at least 5 bands
~ 2.10^5 galaxies for training ($P(\mu_i|G)$, $P(\mu_i|O)$ computation)
~ 8.10^5 galaxies for test (LR calculation)

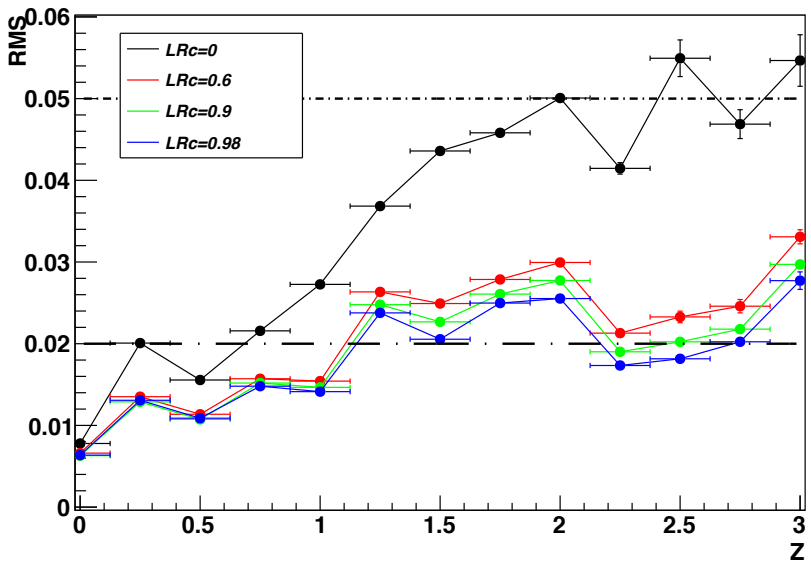
Parameters:
For different zbins we compute
- ngal/dz = the number of galaxies / zbin
- the bias = median
- the RMS = interquartile range
- the percentage of outliers (η)

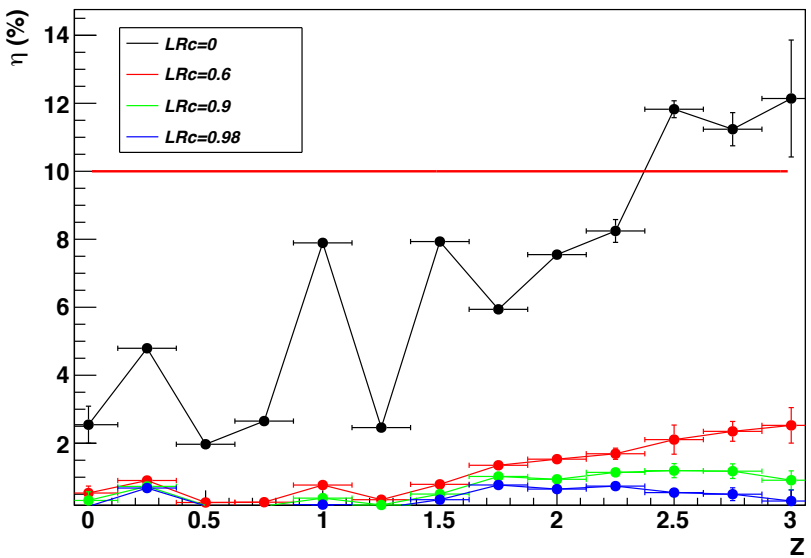
Errors on parameters estimated by comparing the results of 5 different analysis on 5 independent catalogs

PhotoZ performances for LSST



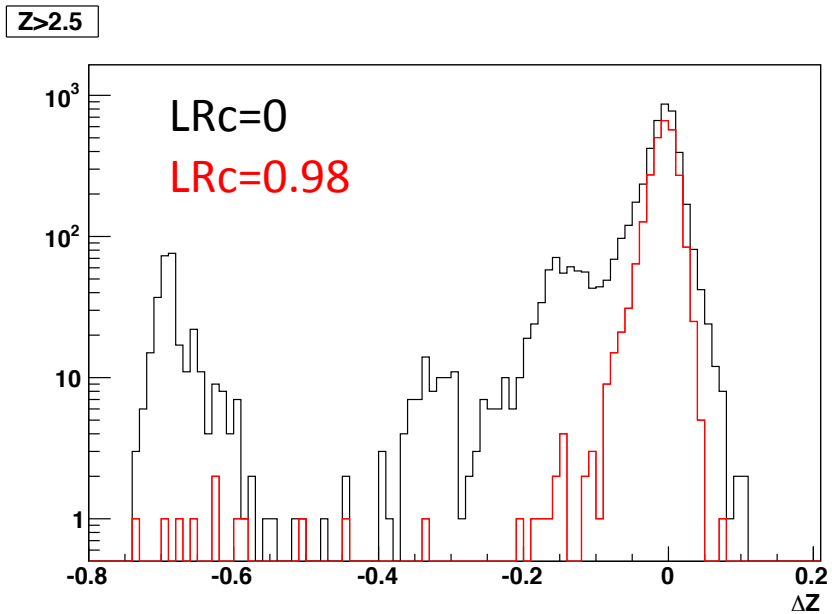






For the galaxies satisfying LR>0.9, we reach LSST goals (almost for bias) up to z = 2.5

PhotoZ performances for LSST z > 2.5

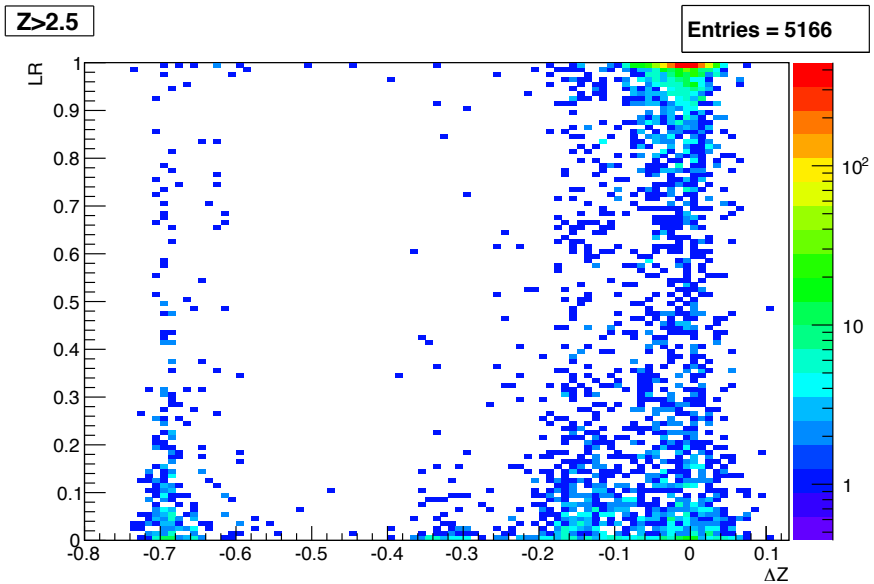


The reconstruction is not bad after LRcut but slightly shifted

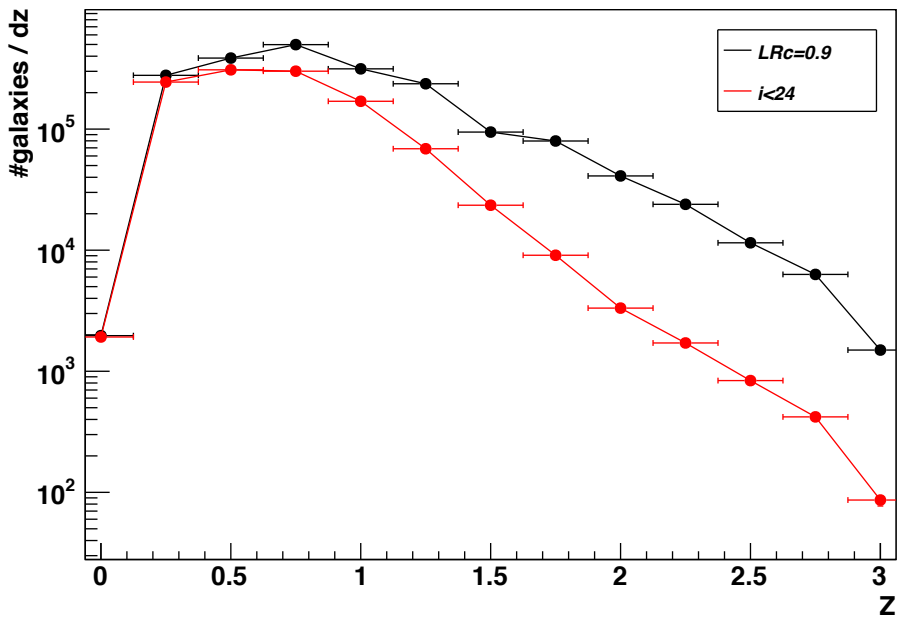
LR mostly trained on z ~ 1 galaxies

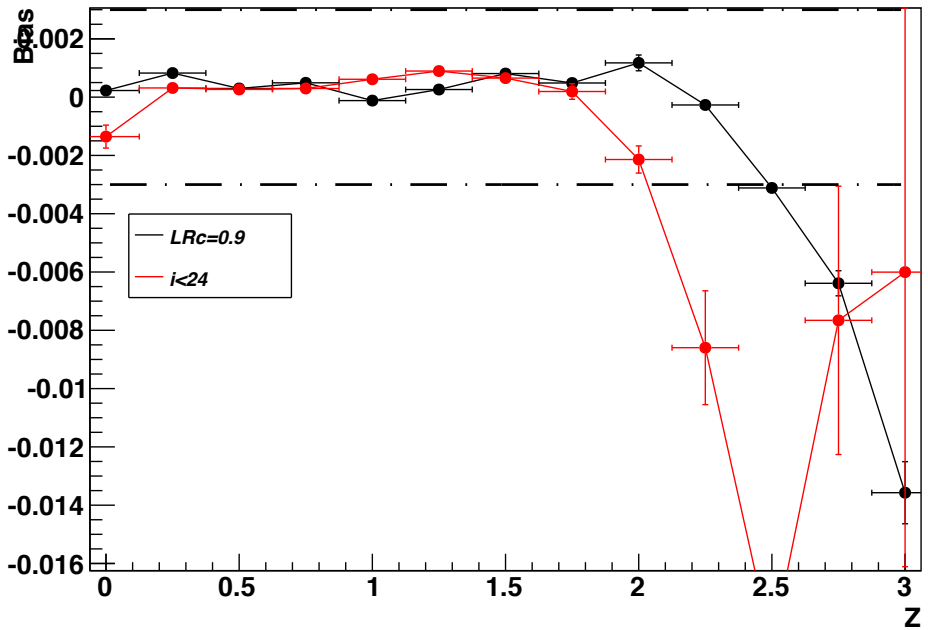
Training LR in different z bins didn't improve the results.
For some outliers LR(ztrue)~0 but LR(zphot)~1

Need more stats at high z



LRcut vs icut





LR gives much better results than a cut on magnitude

Template fitting versus Neural Network

Pros and Cons

Template fitting :

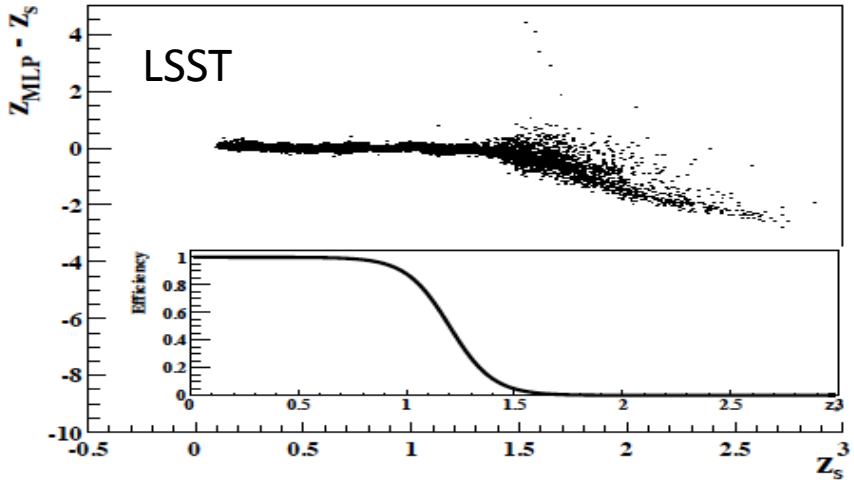
Pros : no selection bias ?

Cons : need a complete template library

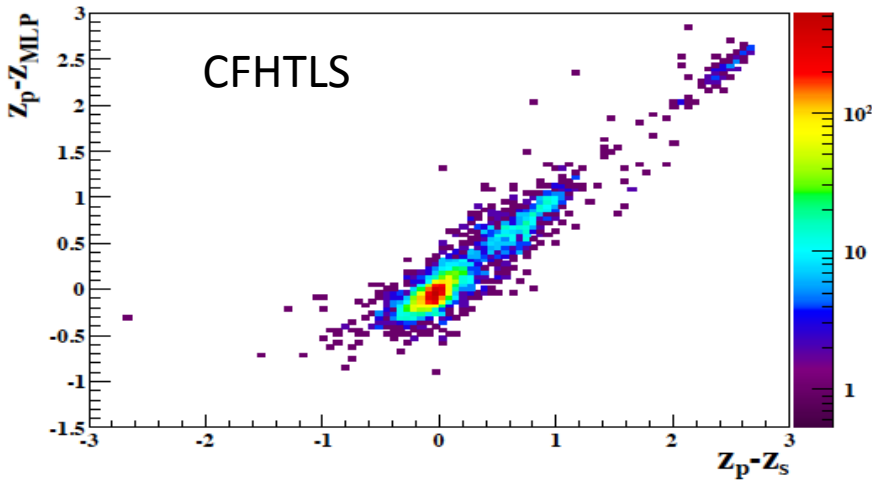
Neural network :

- Pros : no model
- Cons : selection bias

Neural network selection bias



Interesting correlated results

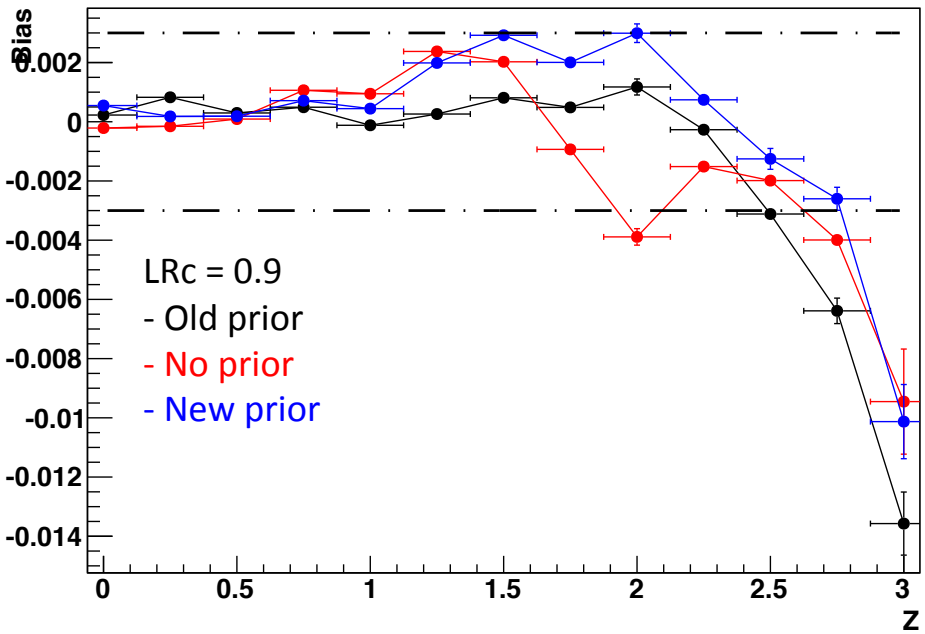


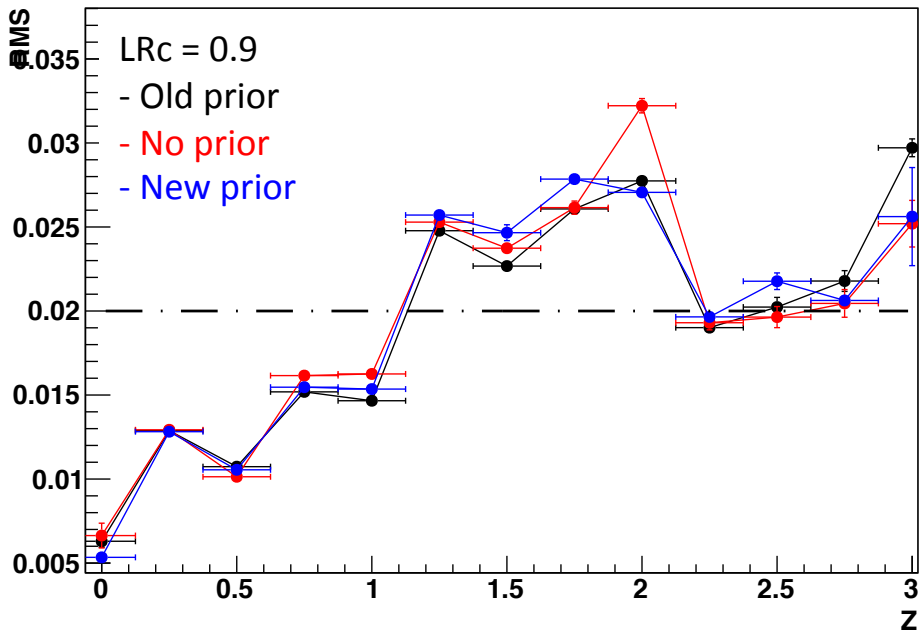
Correlation may be usefull to tag outliers
The final PhotoZ reconstruction method could be a hybrid tool taking advantages of both method

Prior preliminary study

$$P(z,T|i) = P(z|T,i) \times P(T|i)$$

Comparison between our old prior parametrization, no prior and new computation





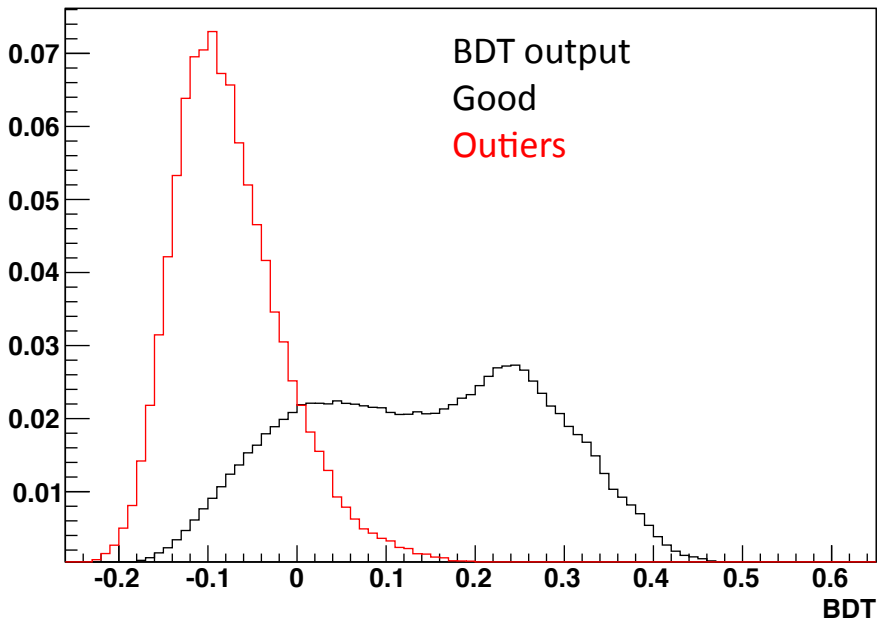
Clear improvement at high z
Need further study

ROOT TMVA Factory

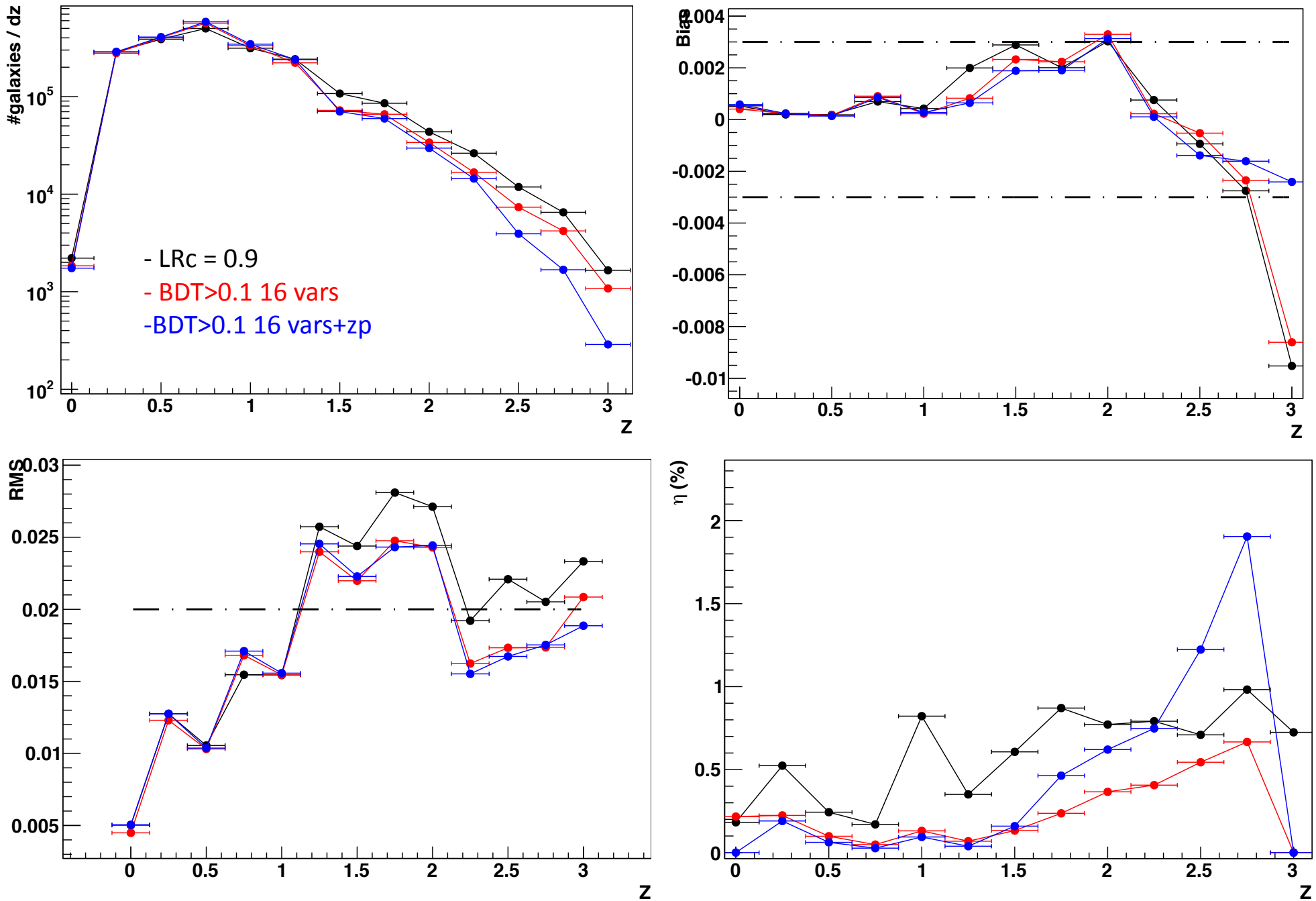
ROOT TMVA::Factory
Lighter use than LR

16 variables

```
---------------------------------------------------------------------------
Rank : Variable                          : Variable Importance
---------------------------------------------------------------------------
   1 : log(Likelihood2[0])-log(Likelihood1[0]) : 1.056e-01
   2 : mag1-mag2                          : 1.040e-01
   3 : mag2-mag3                          : 1.006e-01
   4 : mag3-mag4                          : 9.469e-02
   5 : npeak[0]                           : 9.148e-02
   6 : mag4-mag5                          : 8.732e-02
   7 : mag0-mag1                          : 7.484e-02
   8 : raportInt2                         : 7.085e-02
   9 : npeak[1]                           : 5.053e-02
  10 : log(Likelihood2[1])-log(Likelihood1[1]) : 4.786e-02
  11 : logImax                            : 4.344e-02
  12 : npeak[2]                           : 4.150e-02
  13 : log(Likelihood2[2])-log(Likelihood1[2]) : 4.023e-02
  14 : raportInt1                         : 2.633e-02
  15 : raportInt0                         : 2.069e-02
  16 : TMath::Abs(varp[0]-varp_marg[0])       : 0.000e+00
---------------------------------------------------------------------------
```



BDT preliminary results



Conclusions

- LR / BDT method using posterior PDF characteristics of template fitting method is a very useful tool to reject outliers and improve photoZ reconstruction

- Correlation between neural network and template fitting could be used to tag outliers

- More details in Gorecki *et al* published in Astronomy and Astrophysics
http://arxiv.org/abs/1301.3010v2