ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Implementing probabilistic photometric redshifts on DES SV data

Matías Carrasco Kind

Department of Astronomy
University of Illinois

DES collaboration meeting LBNL April 8 - 12, 2013

DES SV data (fixed!)

VVDS Deep 02hr (3117)

CDFS (3721)

VVDS Wide 14hr (2970)

COSMOS (6148)

Secure redshifts $3 \leq \mathrm{ZFLAG} \leq 4$

$0.01 \leq Z \leq 1.5$

15956 galaxies

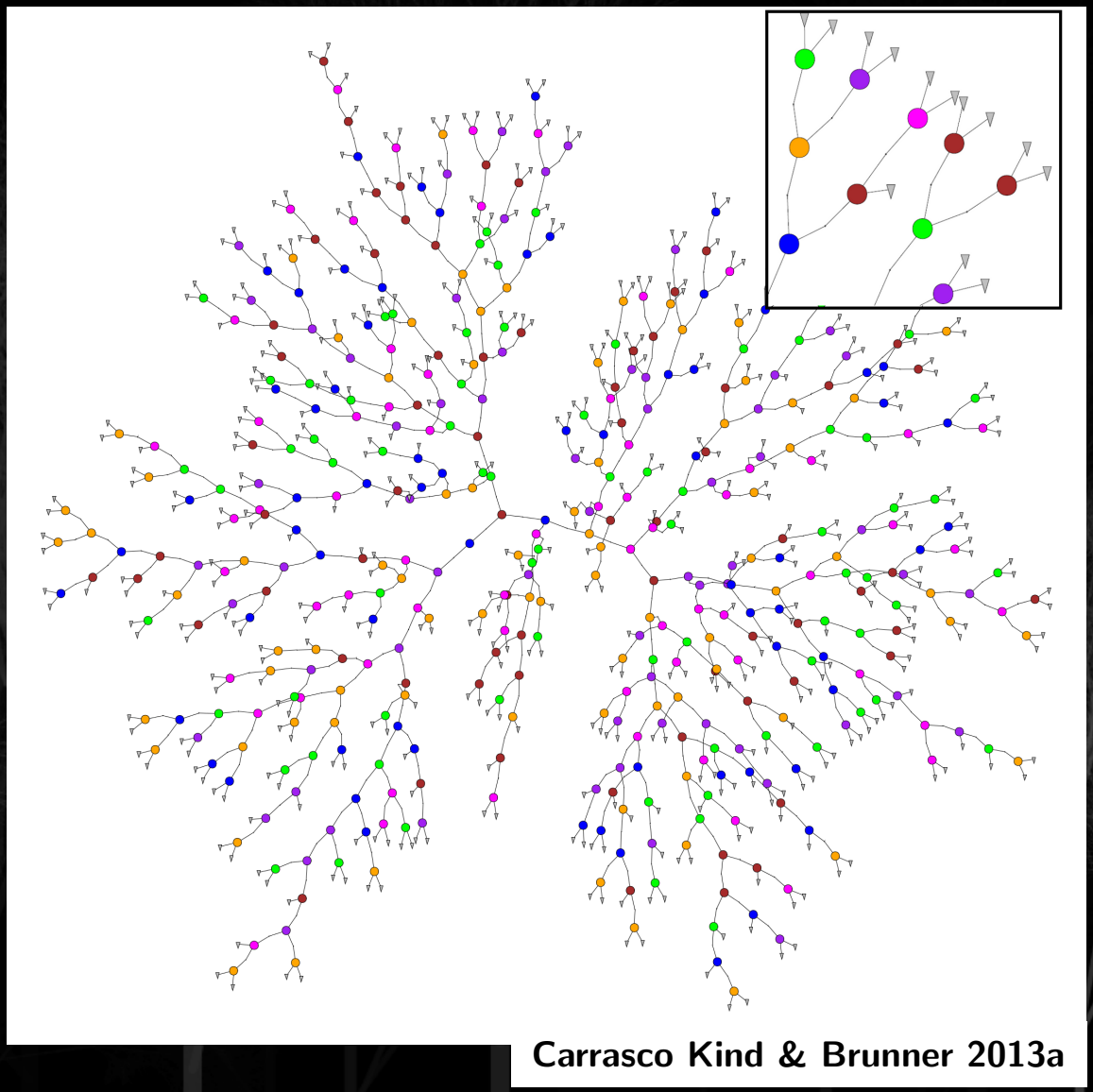
7978 for training and 7978 for testing

Use 5 bands from MAG_AUTO and MAG_DETMODEL

Use colors (8)

TPZ: Trees for Photo-Z

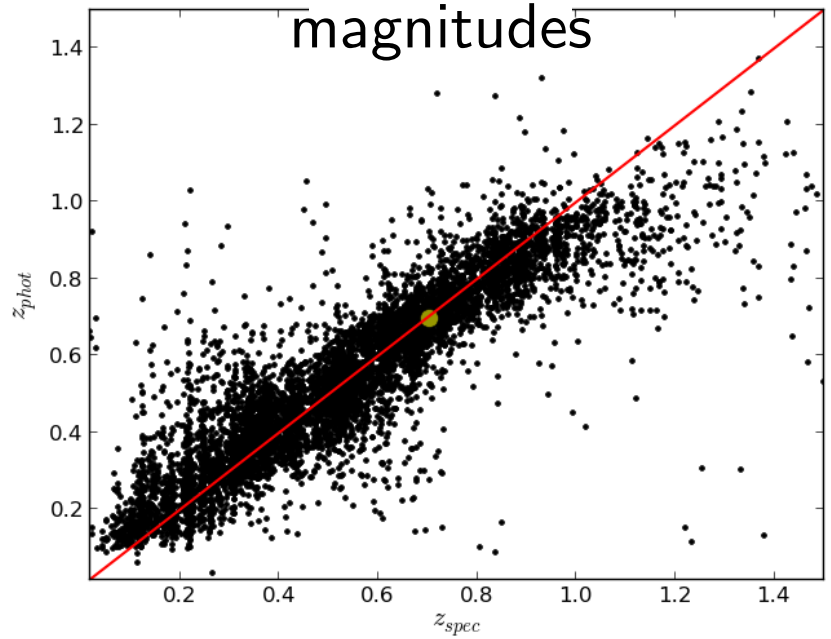
- Provides photo-z PDF and confidence values
- Deals with missing data
- Includes measurements errors
- Provides useful ancillary information
- *Out-of-Bag* data for unbiased errors
- No need for validation set

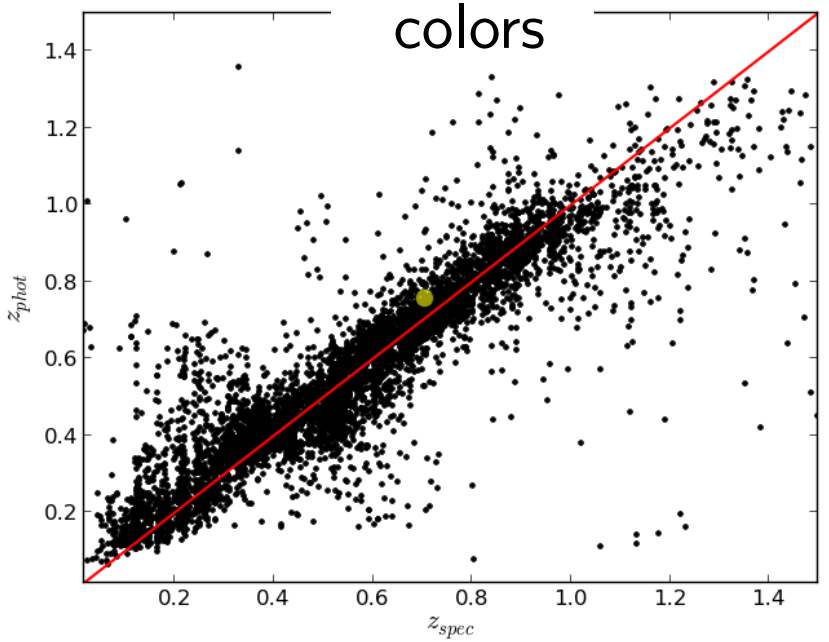


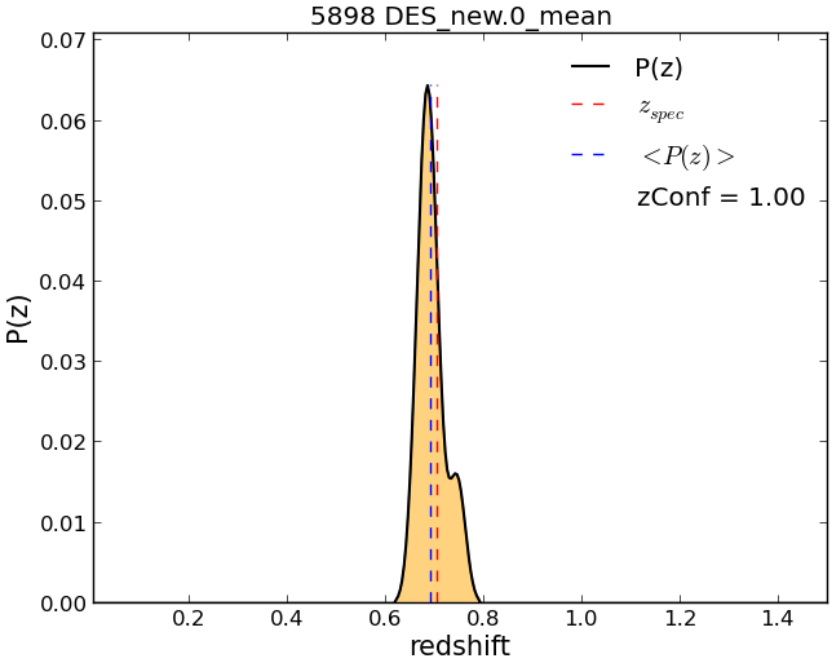
Carrasco Kind & Brunner 2013a

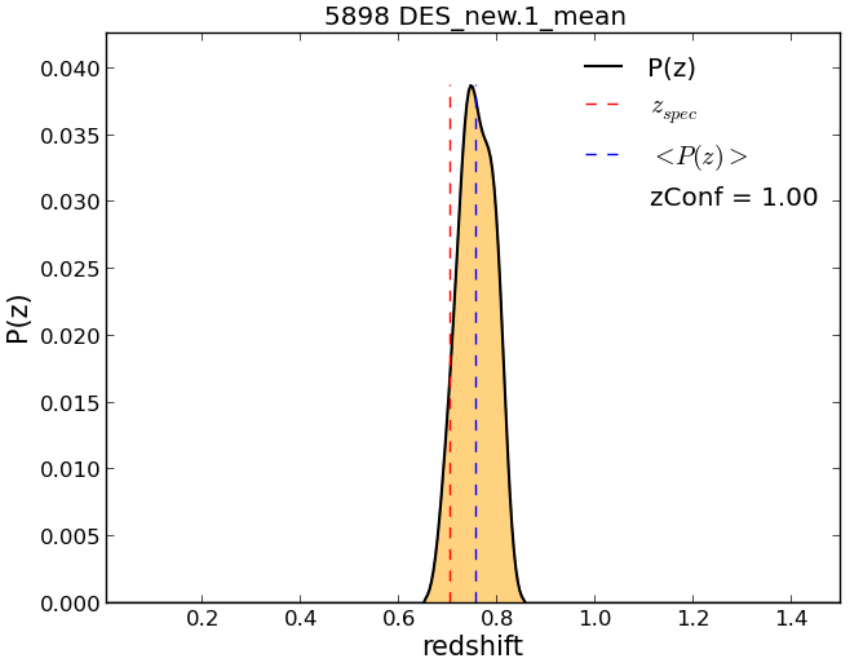
http://lcdm.astro.illinois.edu/research/TPZ.html

Preliminary results : True plots and PDF

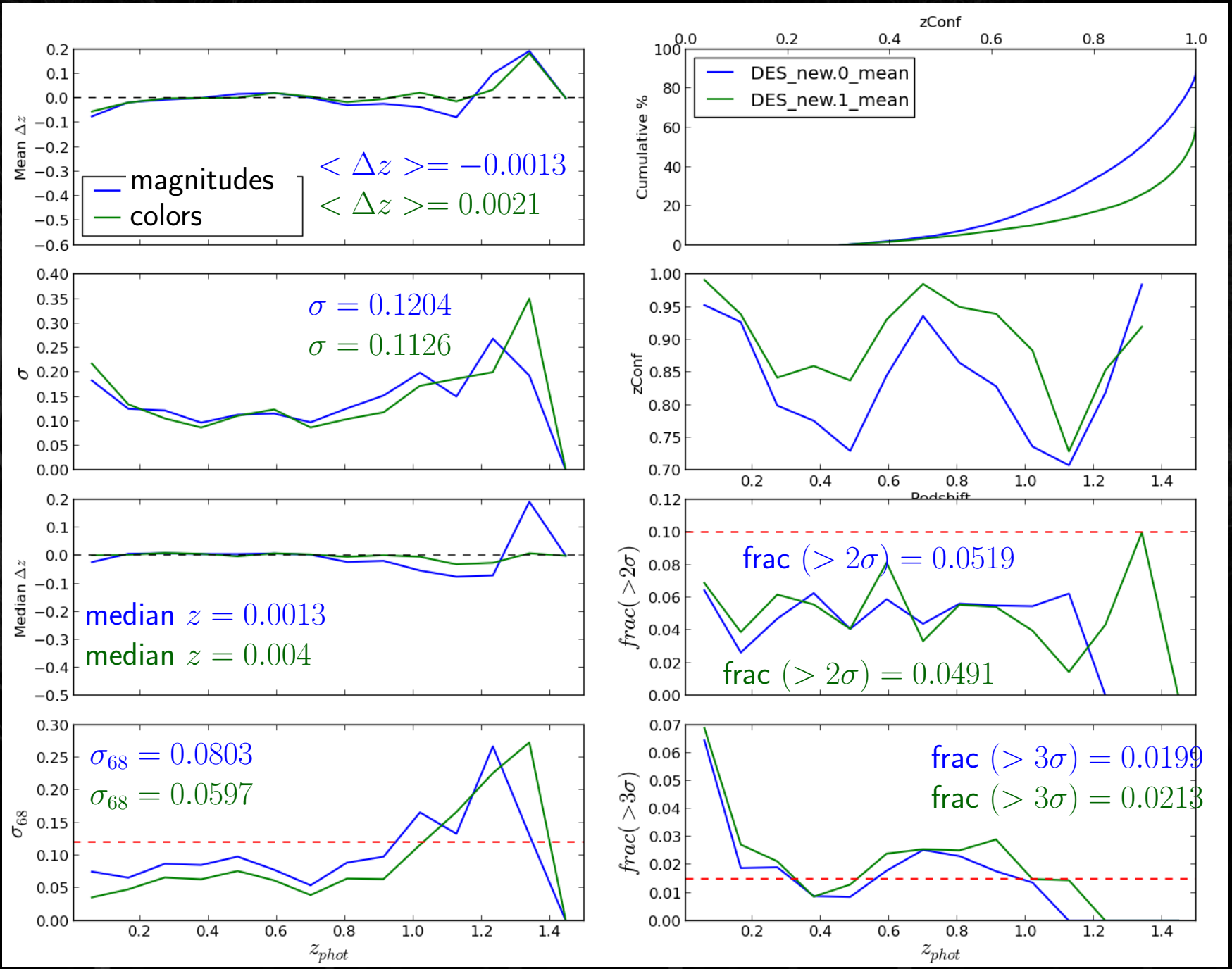




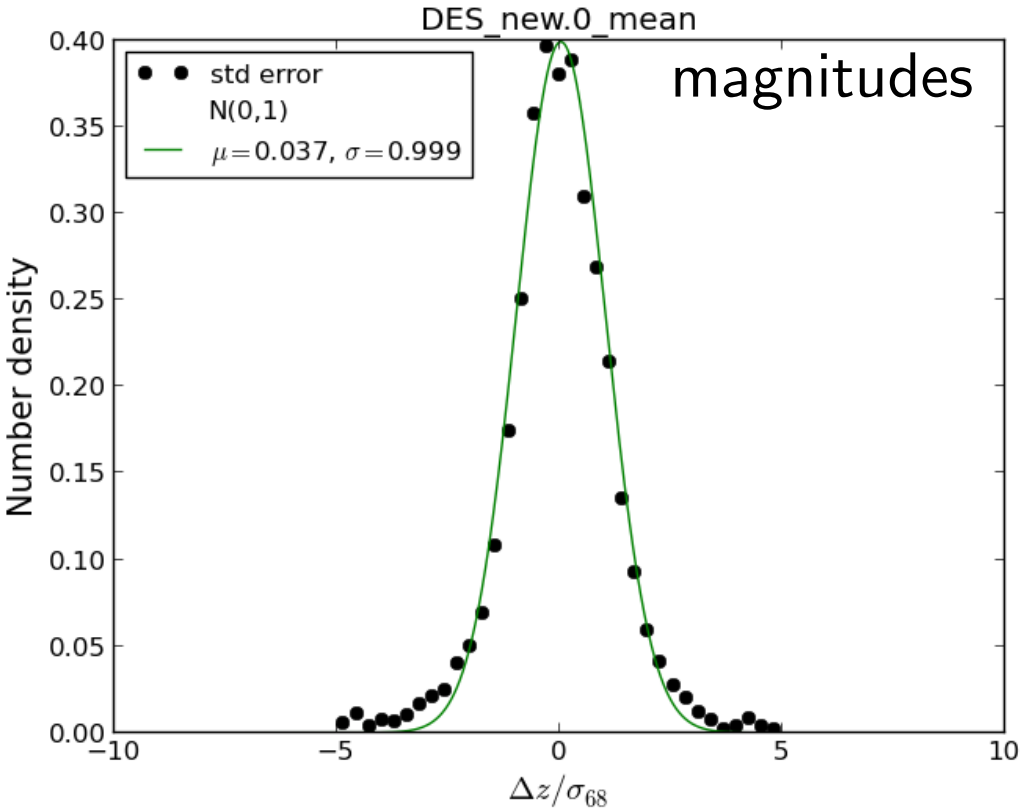


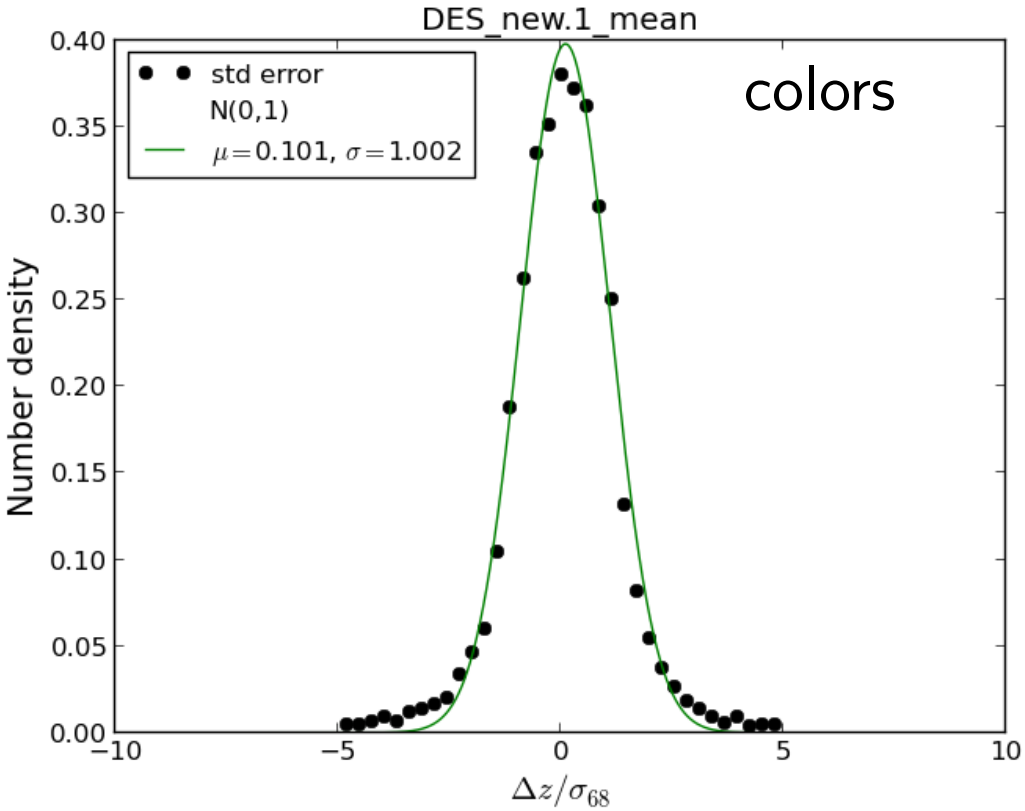


Preliminary results : Metrics



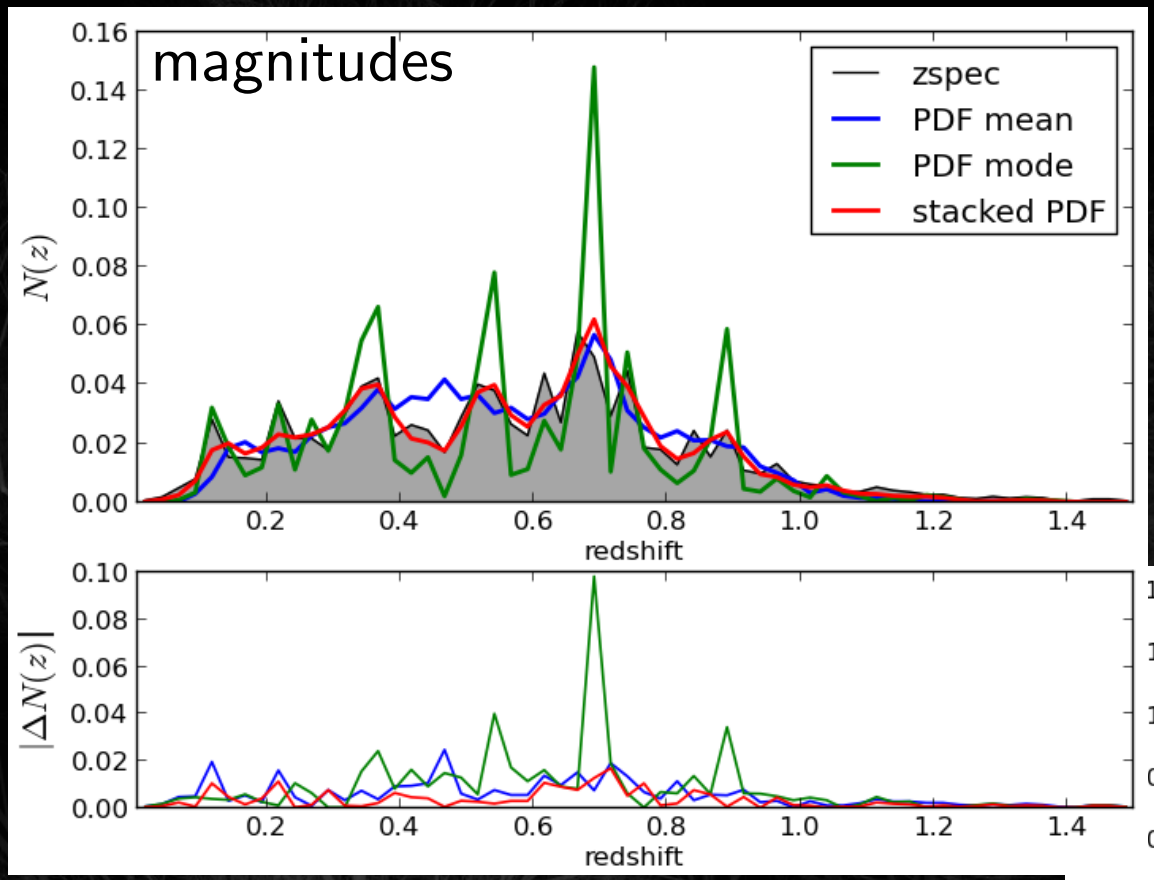
Preliminary results: Error distribution



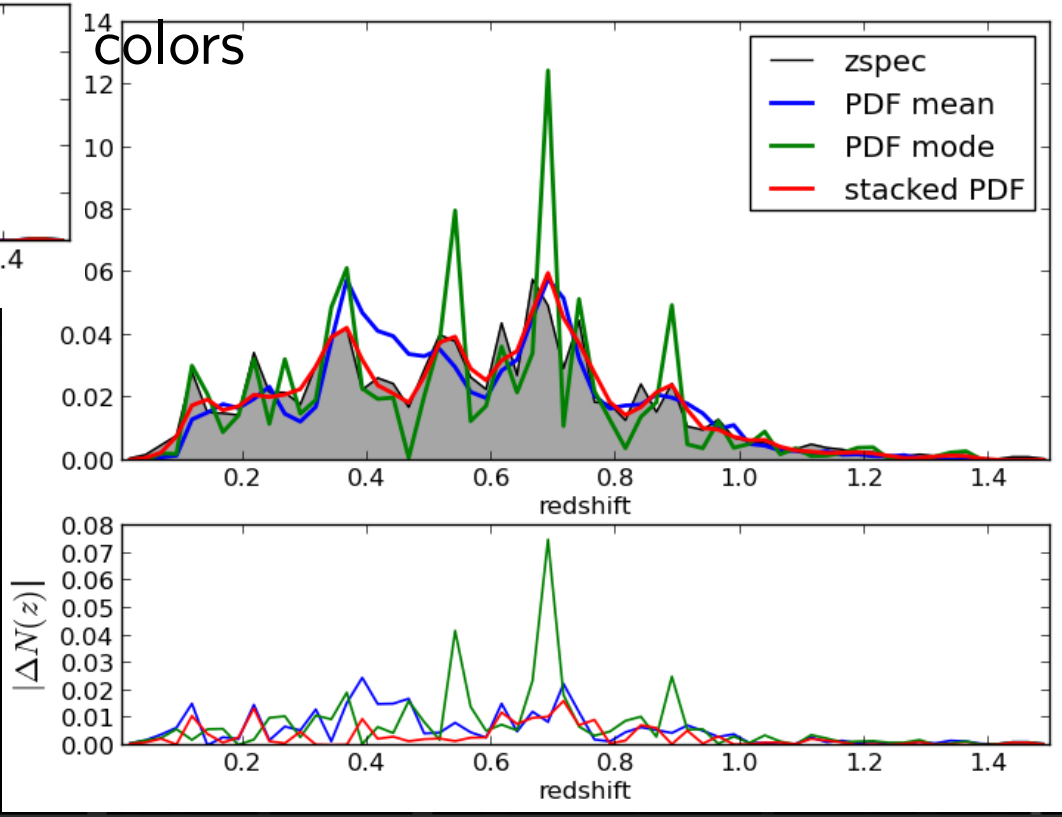


All these metrics were calculated using the mean of the probability density function

Preliminary results: Using PDFs for N(z)

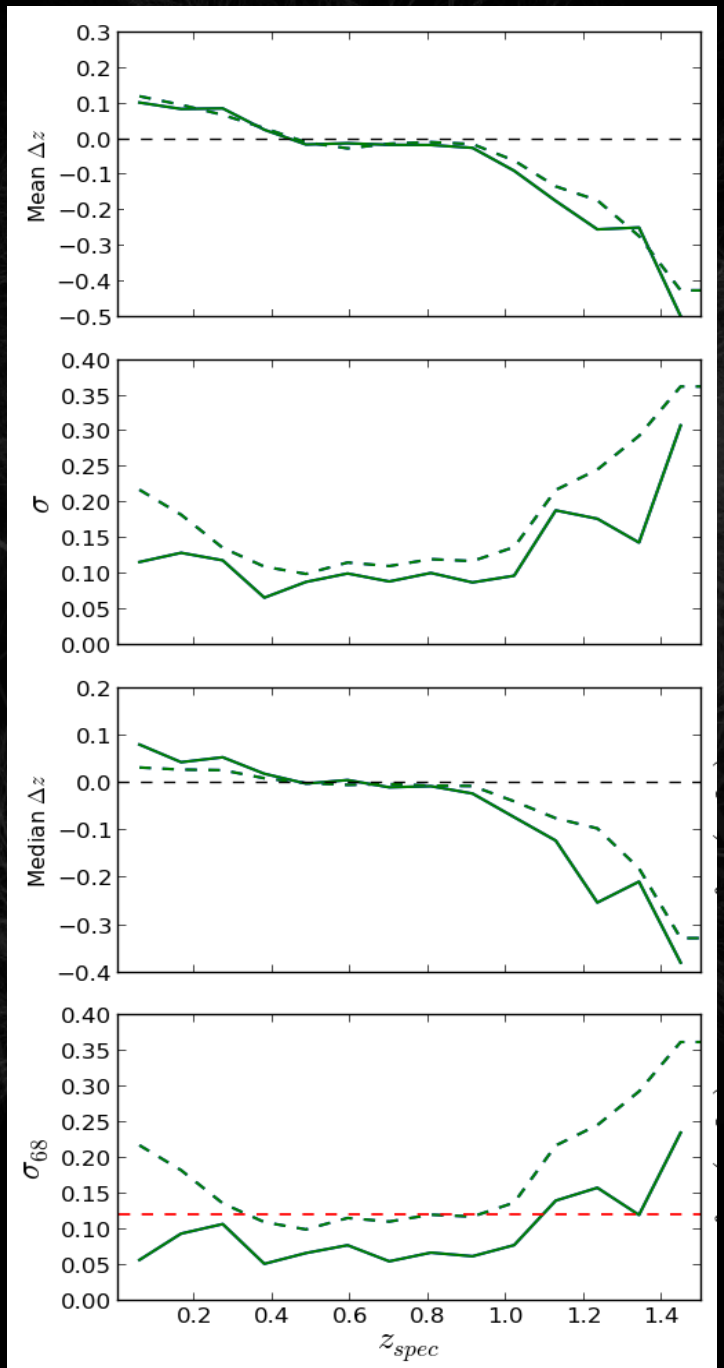


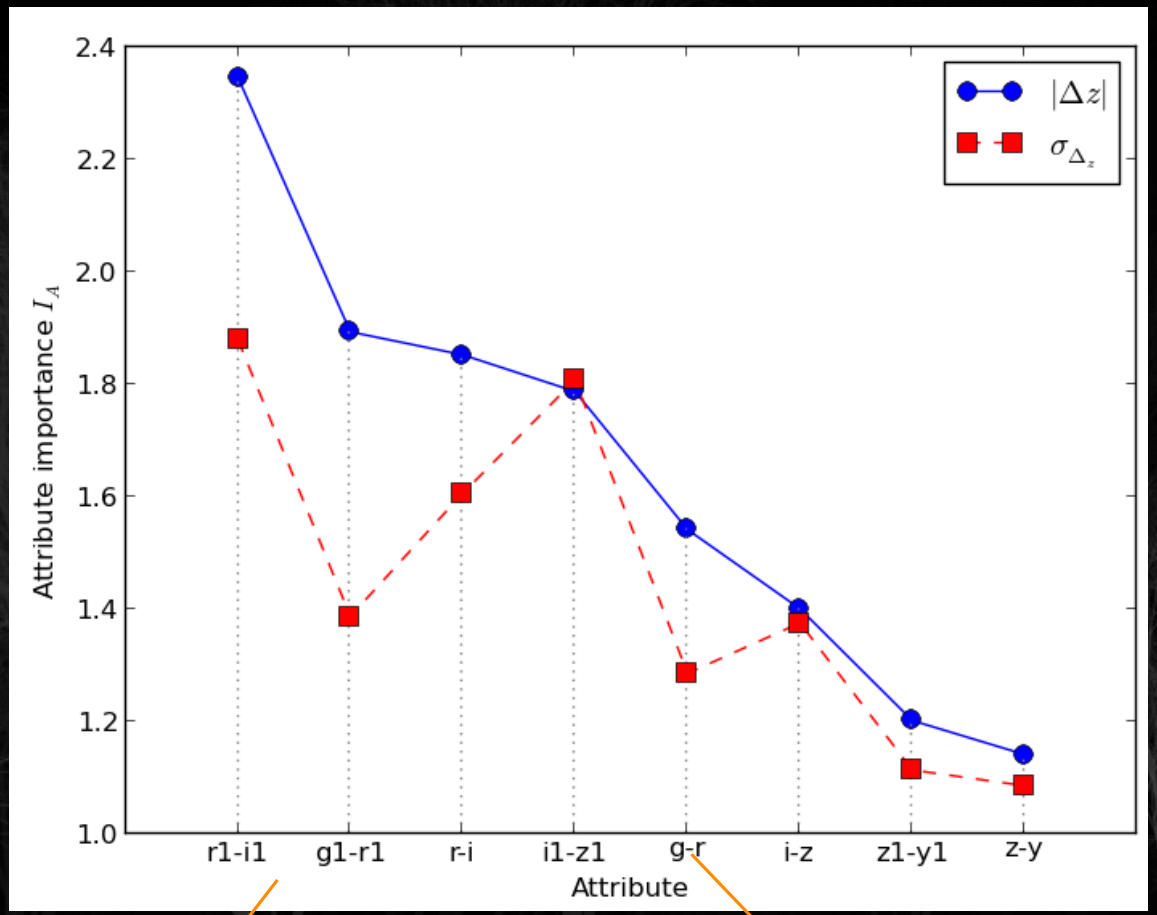
N(z) from PDFs reproduces better the underlying distribution.



Also seen in SDSS, DEEP2 and zCOSMOS among others.

Ancillary examples: OOB error and importance





MAG_DETMODEL

MAG_AUTO

Combining techniques

Multiple "independent" techniques can be combined into a powerful one (Carrasco Kind & Brunner, 2013b, c in prep.)

We use a modified and parallel version of BPZ (Benitez, 2000) with a prior from Random Naïve Bayes Classifier method

Extend it using a 3rd technique from Self-Organized-Maps

Our approach

Template fitting

$+$

Prior information

$+$

Empirical method

$+$

Weigthing scheme

$\Downarrow$

photo-z PDF

Template fitting photo-z

Suppose a set of templates T and n magnitudes $m_1, m_2, ..., m_n$, the probability is:

$$P(z|\vec{m}) = \sum_T P(z, T|\vec{m}) \propto \sum_T P(z, T|\vec{m}) P(\vec{m}|z, T)$$

where $\vec{m} = (m_1, m_2, ..., m_n)$

Template fitting photo-z

Suppose a set of templates T and n magnitudes $m_1, m_2, ..., m_n$, the probability is:

$$P(z|\vec{m}) = \sum_T P(z, T|\vec{m}) \propto \sum_T P(z, T|\vec{m}) P(\vec{m}|z, T)$$

where $\vec{m} = (m_1, m_2, ..., m_n)$

Prior

Likelihood

Template fitting photo-z

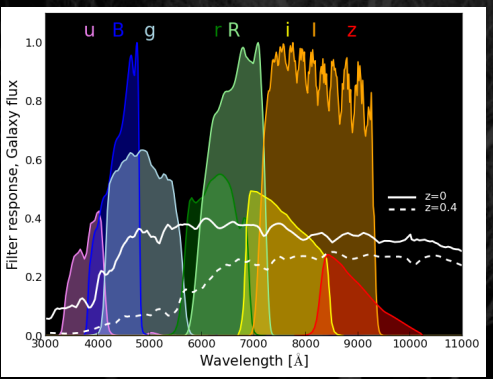
Suppose a set of templates T and n magnitudes $m_1, m_2, ..., m_n$, the probability is:

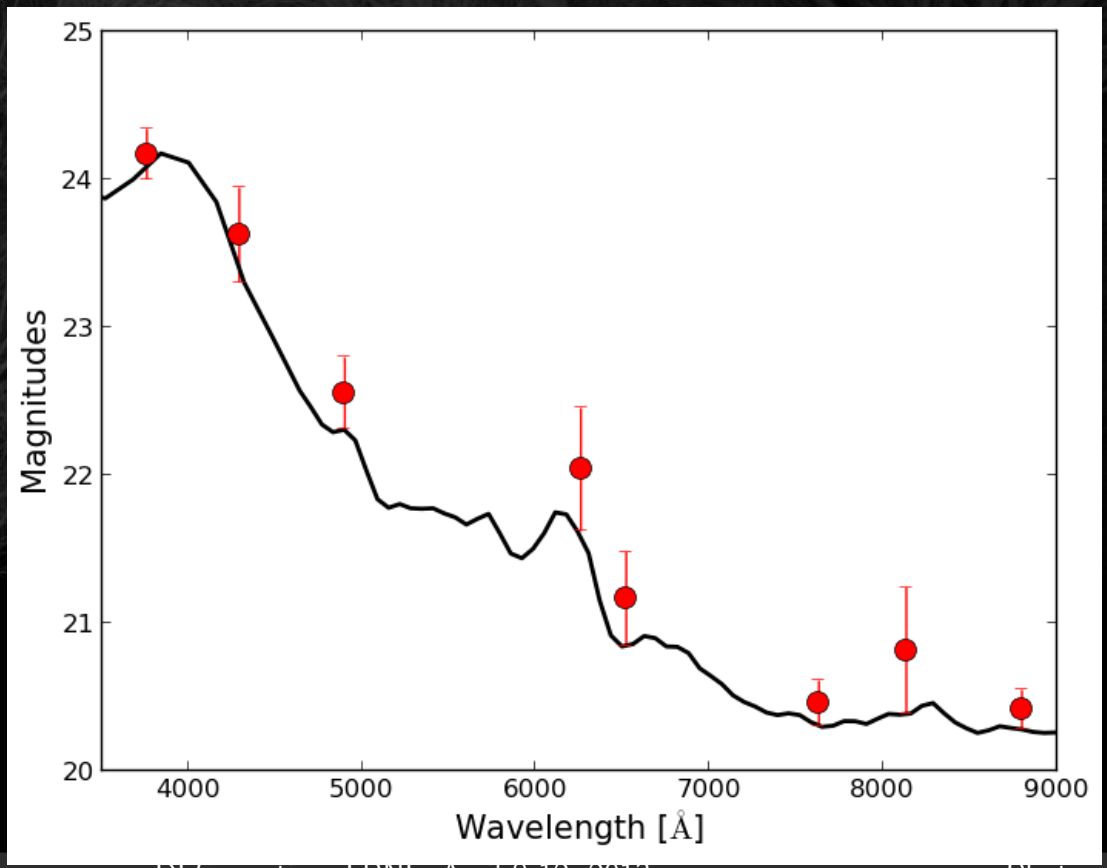
$$P(z|\vec{m}) = \sum_T P(z, T|\vec{m}) \propto \sum_T P(z, T|\vec{m}) P(\vec{m}|z, T)$$

where $\vec{m} = (m_1, m_2, ..., m_n)$

Prior

Likelihood





Random Naïve Bayes Classifier prior

$$P(z|\vec{m}) = \sum_T P(z,T|\vec{m}) \propto \sum_T P(z,T|\vec{m}) P(\vec{m}|z,T)$$

Prior

Likelihood

RNBC needs training set

Assume *naïvely* $m_i \perp m_j$

Bootstrap samples, random subset of magnitudes

Random Naïve Bayes Classifier prior

$$P(z|\vec{m}) = \sum_T P(z,T|\vec{m}) \propto \sum_T P(z,T|\vec{m}) P(\vec{m}|z,T)$$

Prior

Likelihood

RNBC needs training set

Assume *naïvely* $m_i \perp m_j$

Bootstrap samples, random subset of magnitudes

$$\text{RNBC prior} \Rightarrow P(z,T|\vec{m}) \propto P(z)P(T|z) \prod_{i=1}^{n} P(m_i|z,T)$$

Random Naïve Bayes Classifier prior

$$P(z|\vec{m}) = \sum_T P(z,T|\vec{m}) \propto \sum_T P(z,T|\vec{m}) P(\vec{m}|z,T)$$

Prior

Likelihood

RNBC needs training set

Assume *naïvely* $m_i \perp m_j$

Bootstrap samples, random subset of magnitudes

$$\text{RNBC prior} \Rightarrow P(z,T|\vec{m}) \propto P(z)P(T|z) \prod_{i=1}^{n} P(m_i|z,T)$$

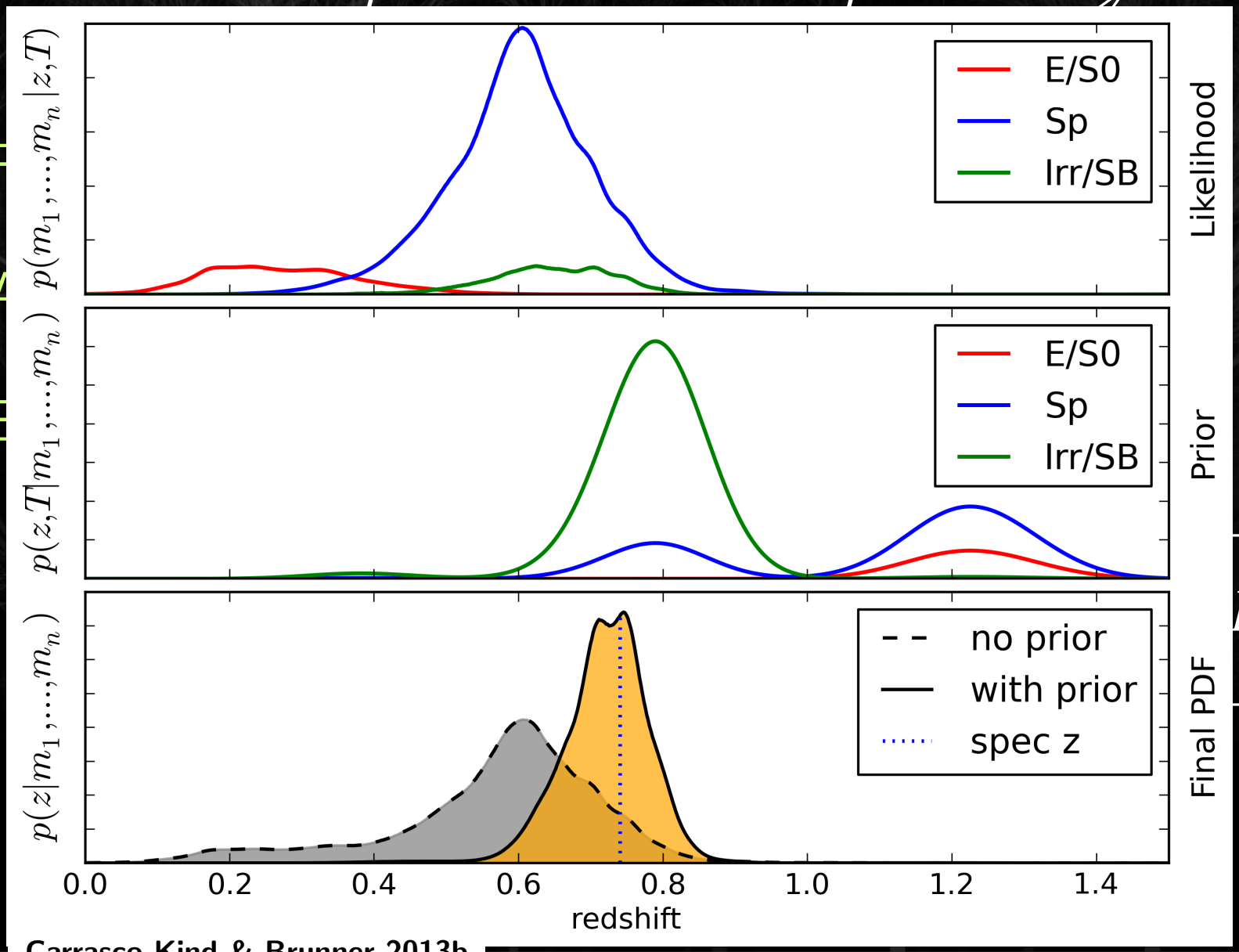
PCA transformations

A model is assumed for this term

Random Naïve Bayes Classifier prior

$$P(z|\vec{m}) = \sum_T P(z,T|\vec{m}) \propto \sum_T P(z,T|\vec{m}) P(\vec{m}|z,T)$$

Likelihood

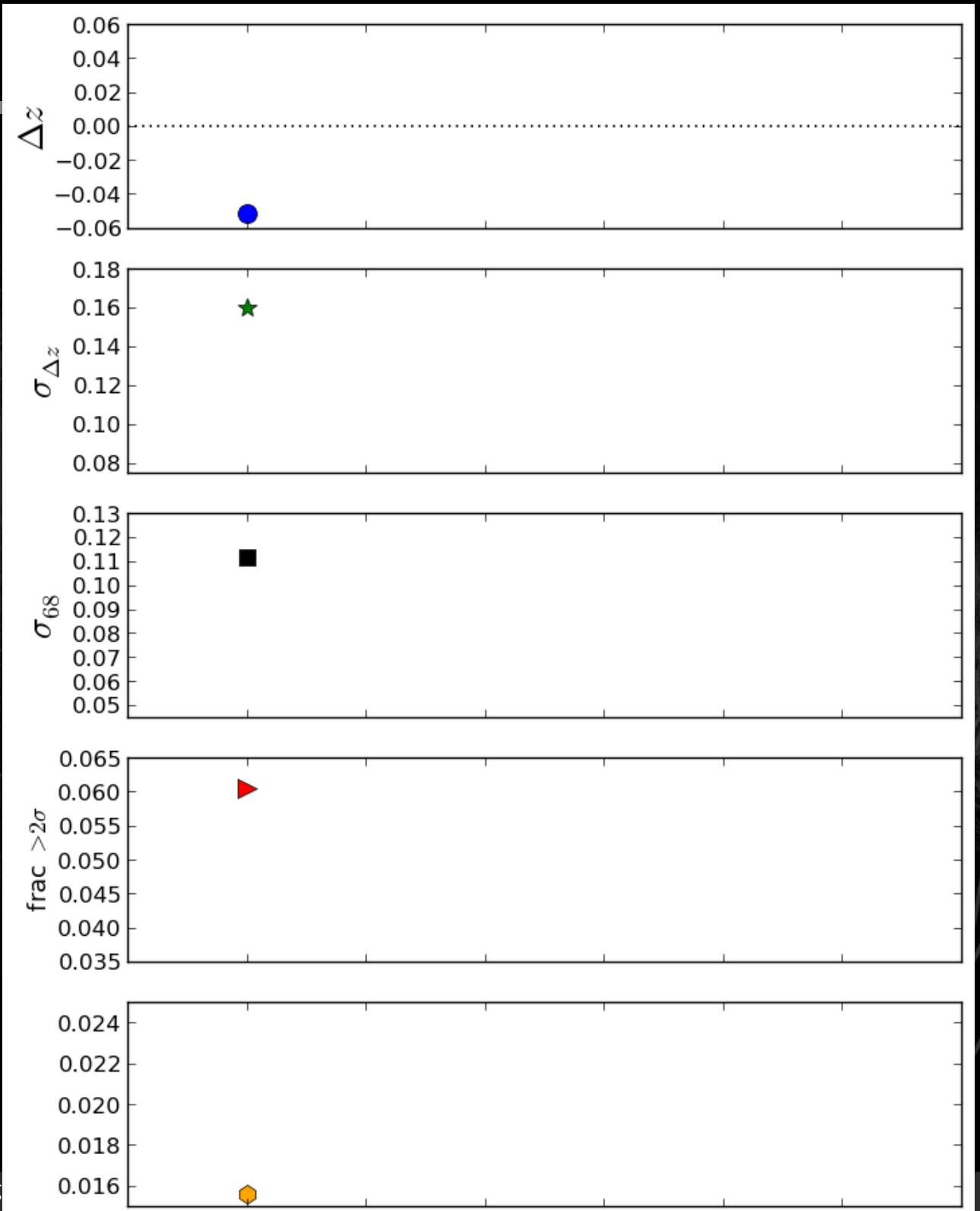


$$P(m_i|z,T)$$

Carrasco Kind & Brunner 2013b

Combining techniques

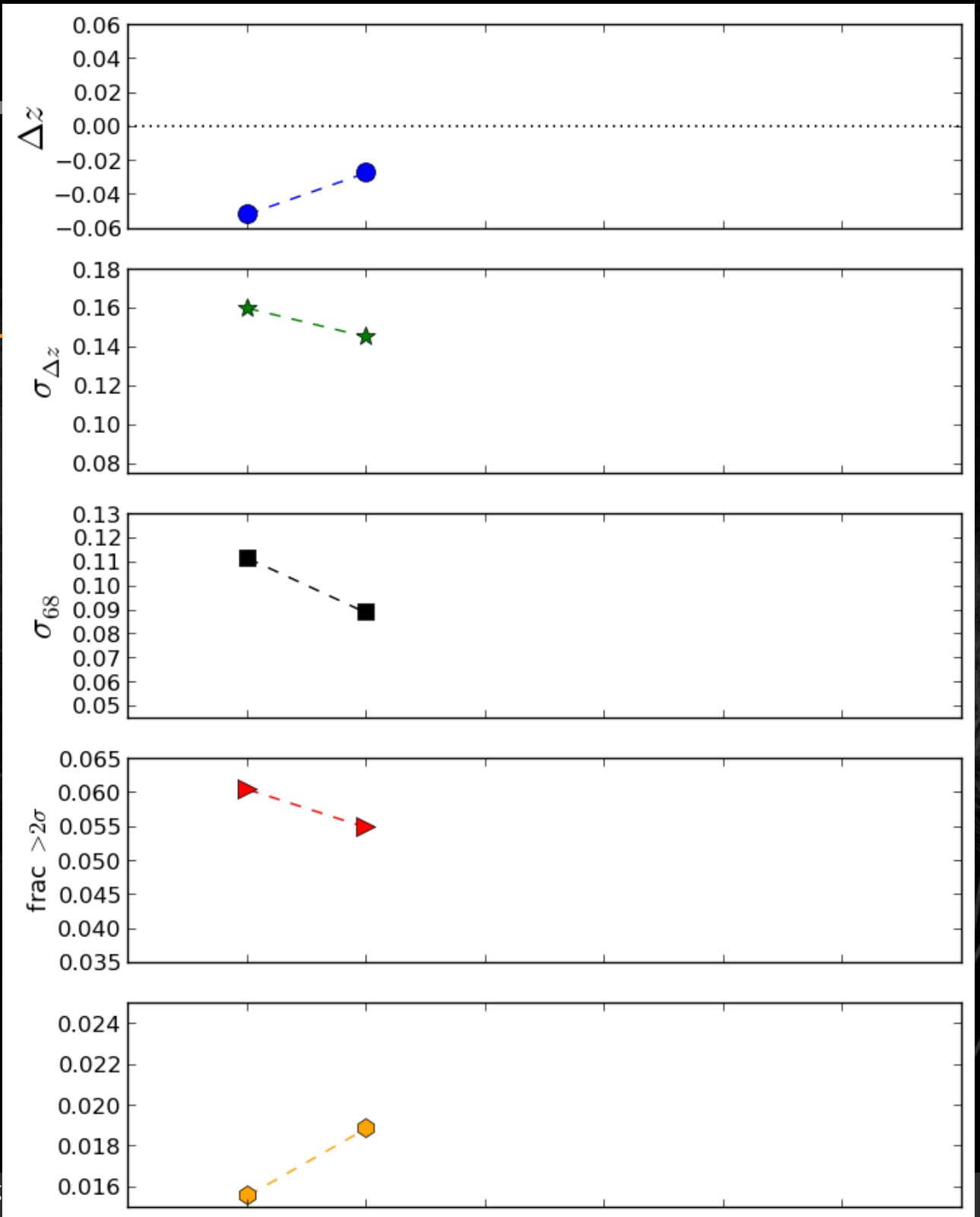
Mod BPZ



Combining techniques

Mod BPZ

Mod BPZ + RNBC prior

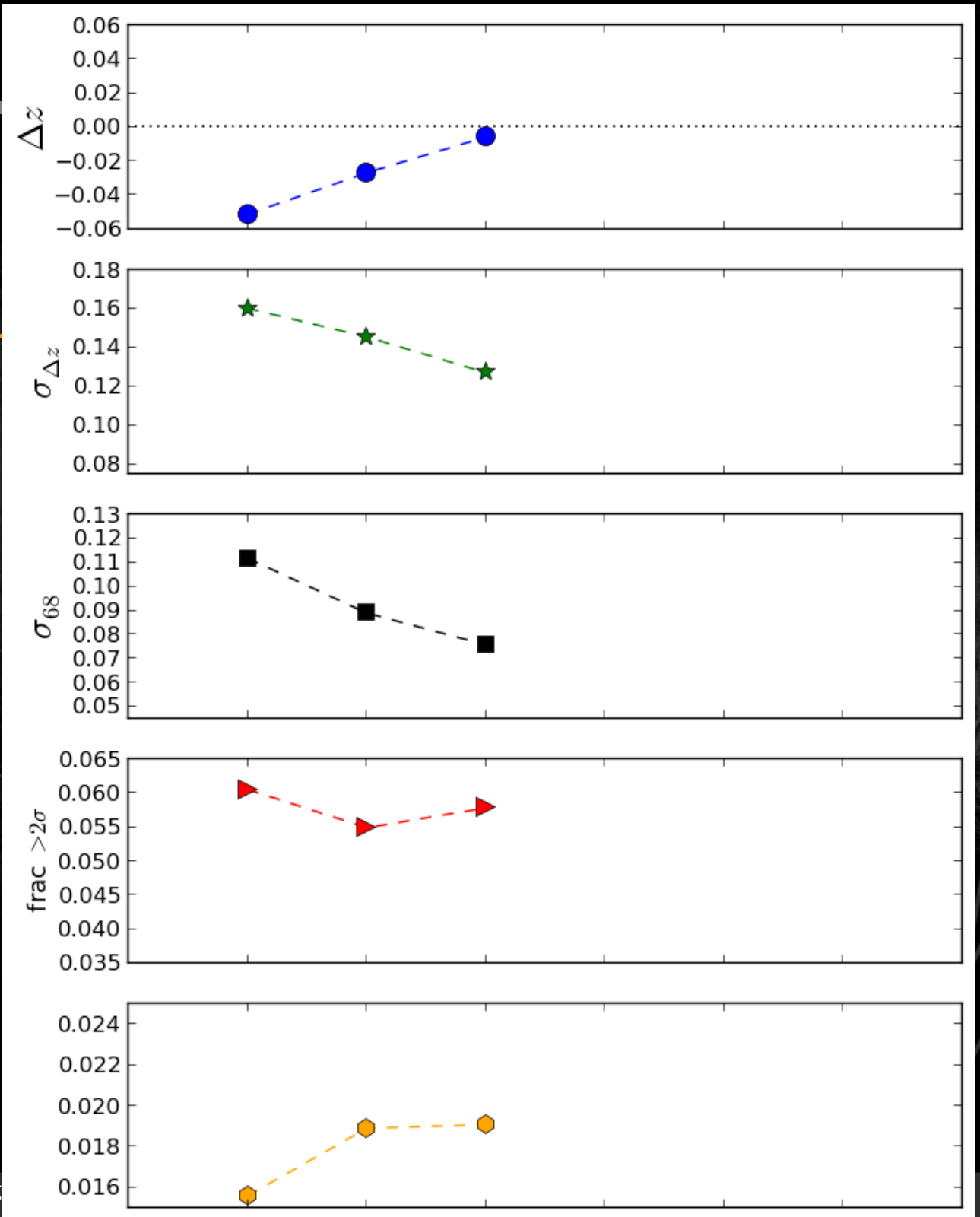


Combining techniques

Mod BPZ

Mod BPZ + RNBC prior

TPZ



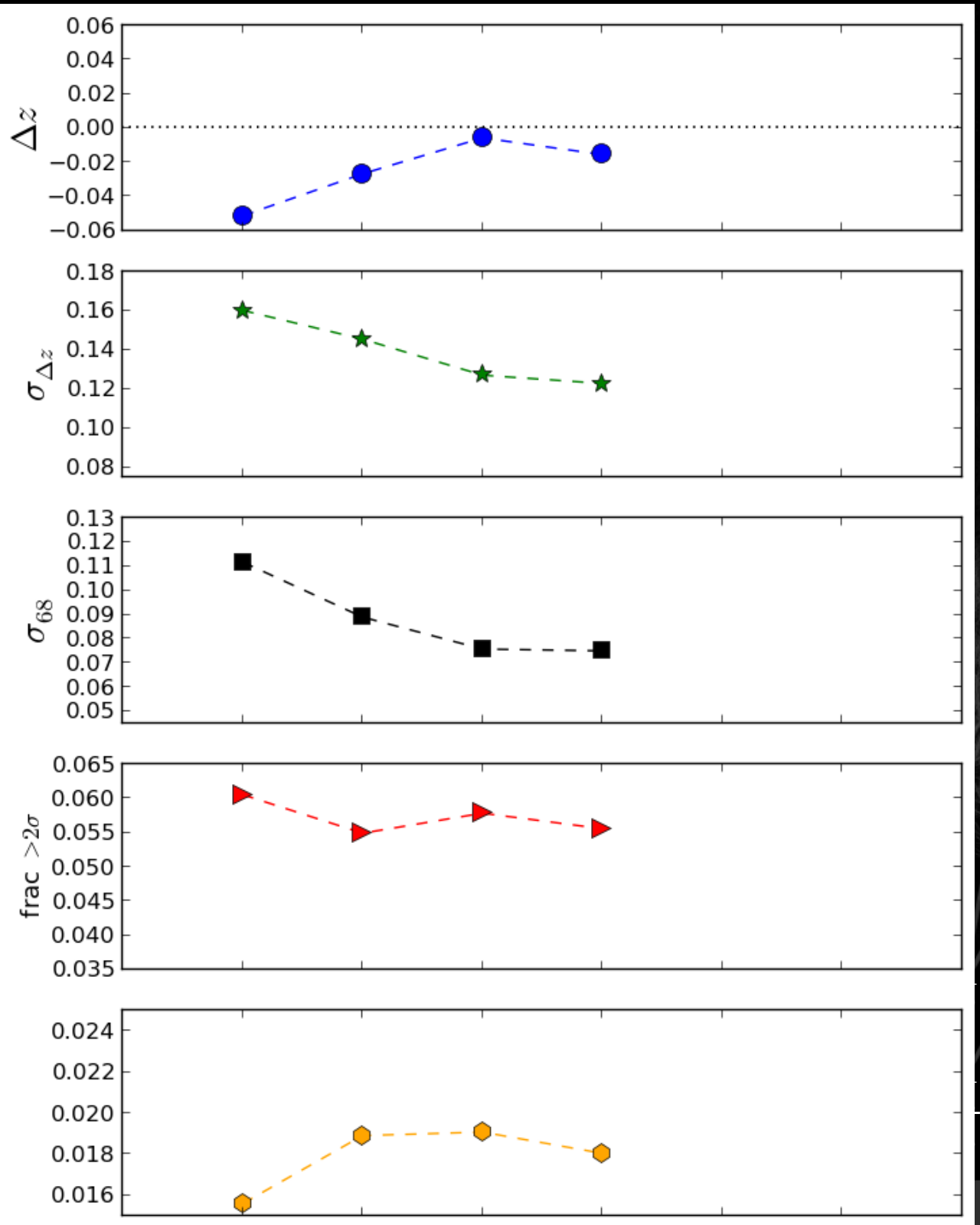
Combining techniques

Mod BPZ

Mod BPZ + RNBC prior

TPZ

TPZ+Mod BPZ + prior



Combining techniques

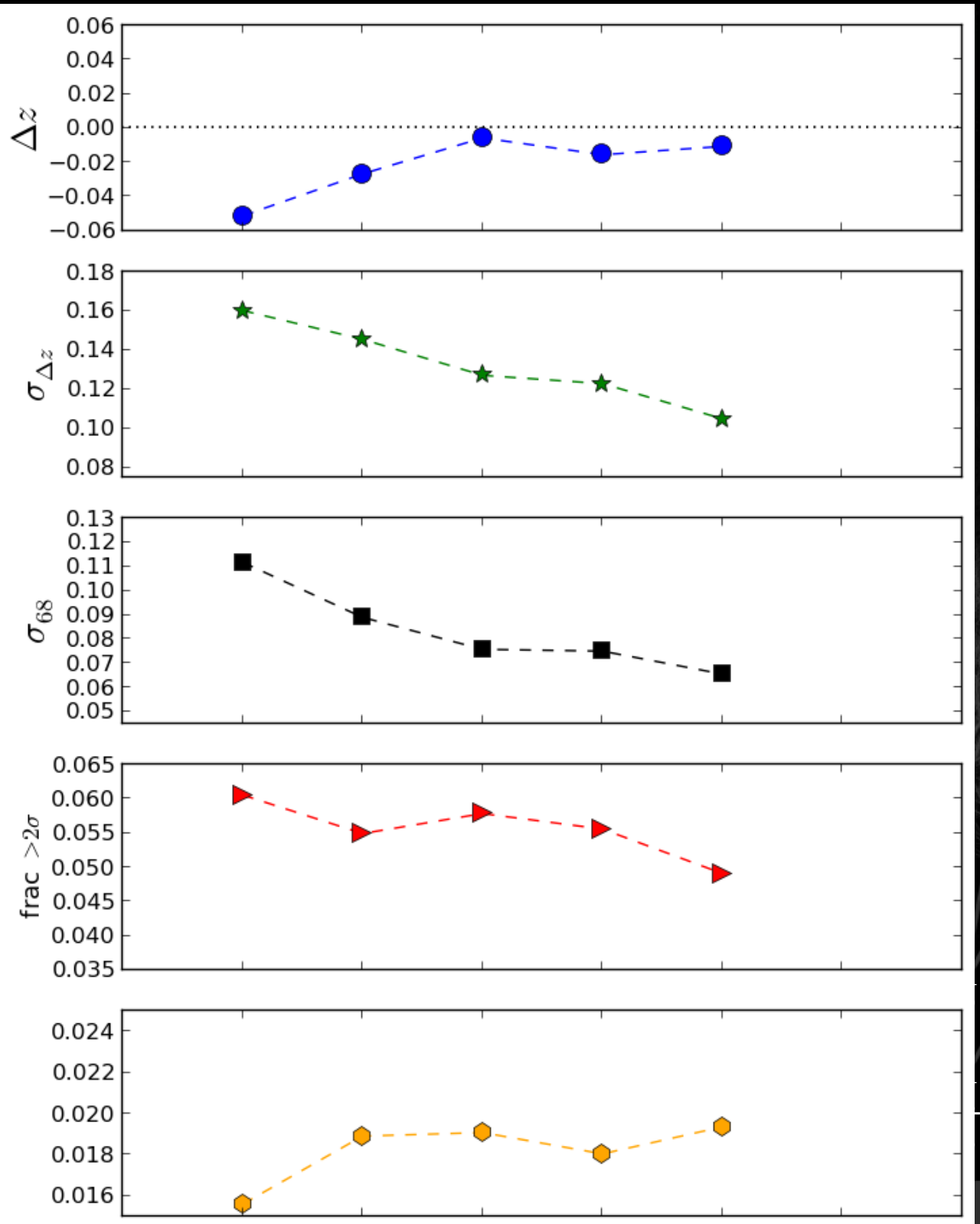
Mod BPZ

Mod BPZ + RNBC prior

TPZ

TPZ+Mod BPZ + prior

TPZ+Mod BPZ+prior
zConf > 0.5



Combining techniques

Mod BPZ

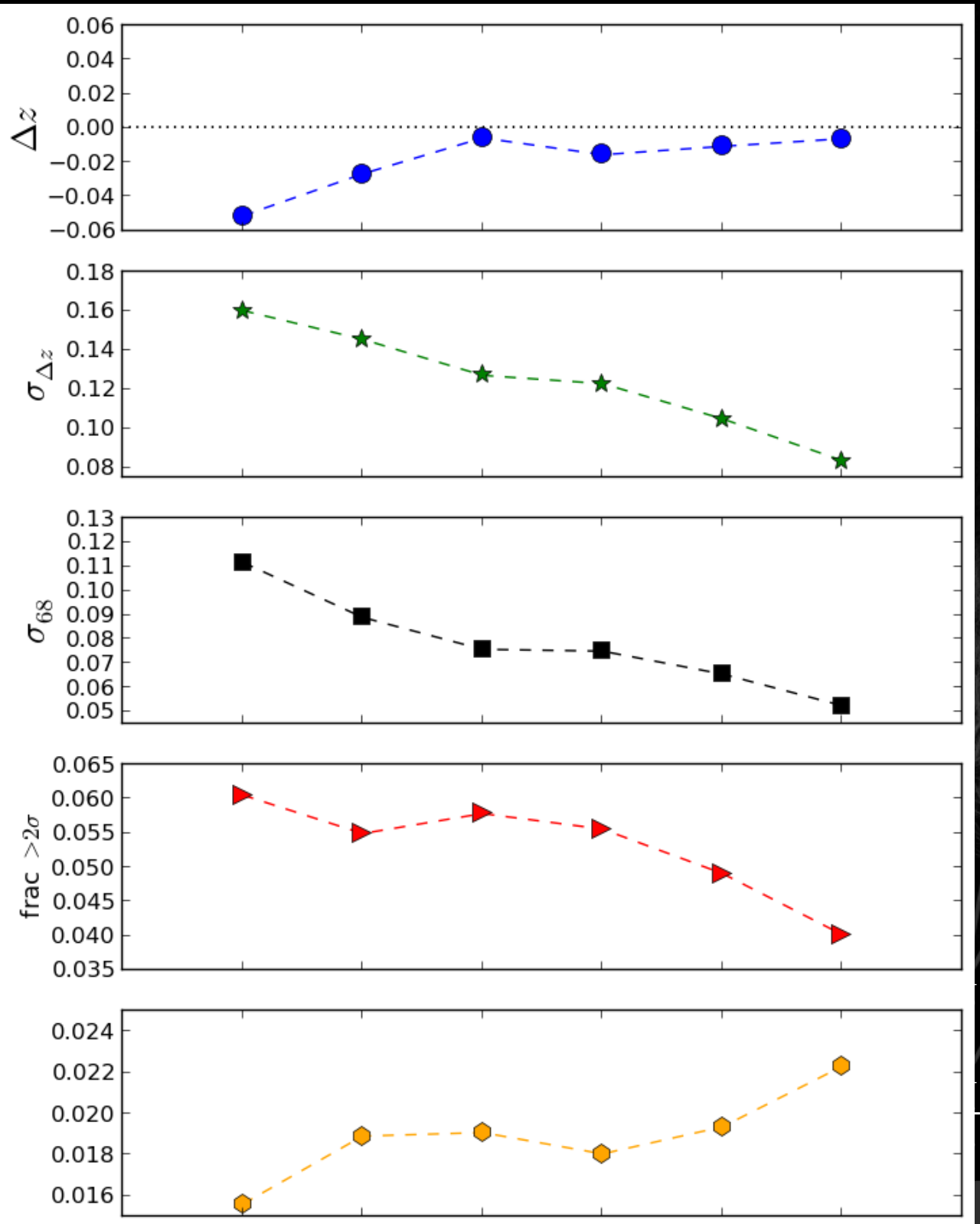
Mod BPZ + RNBC prior

TPZ

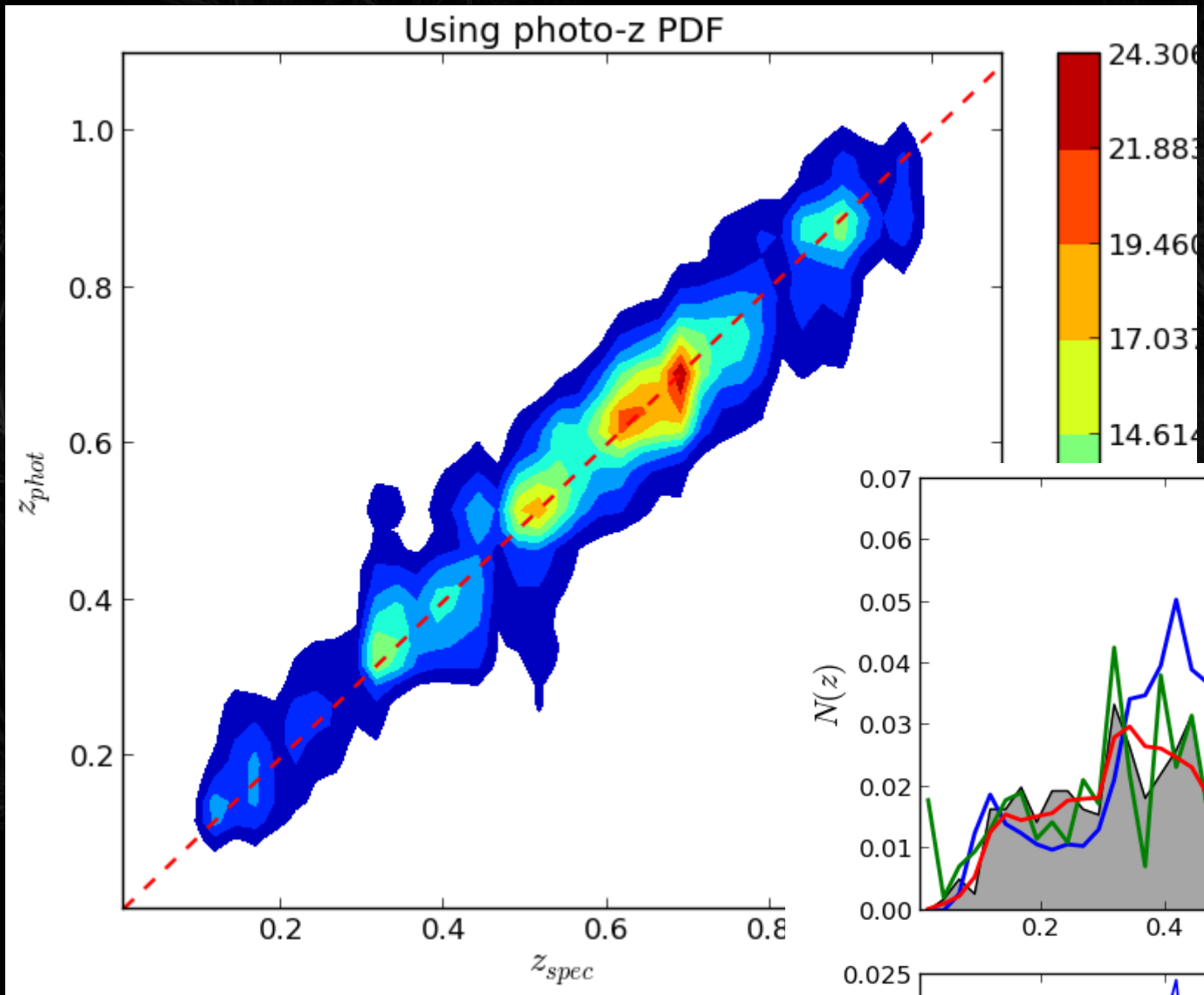
TPZ+Mod BPZ + prior

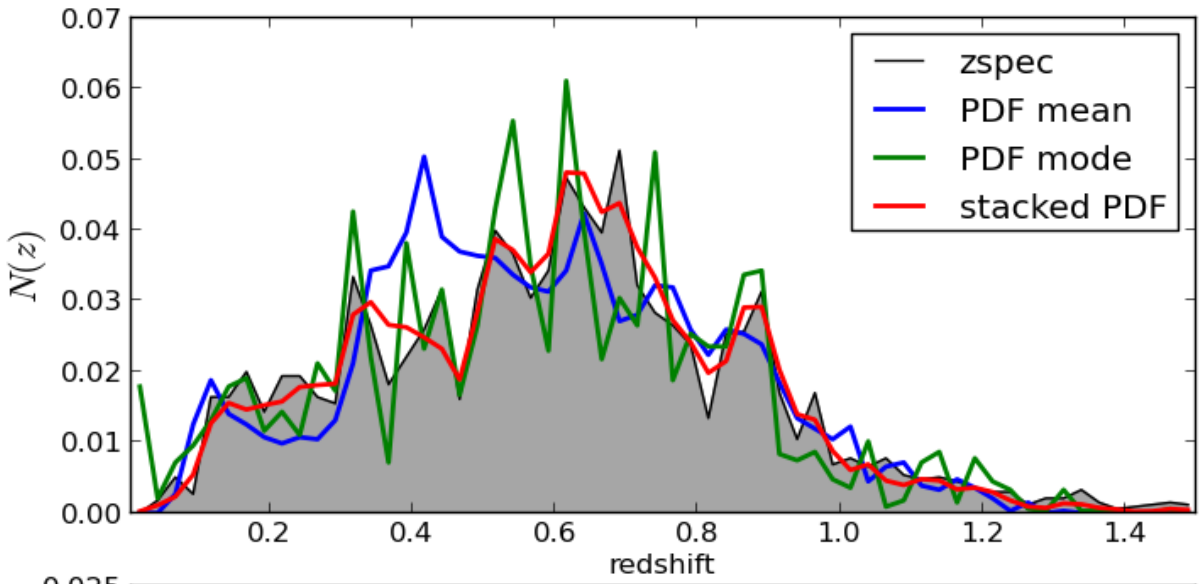
TPZ+Mod BPZ+prior
zConf > 0.5

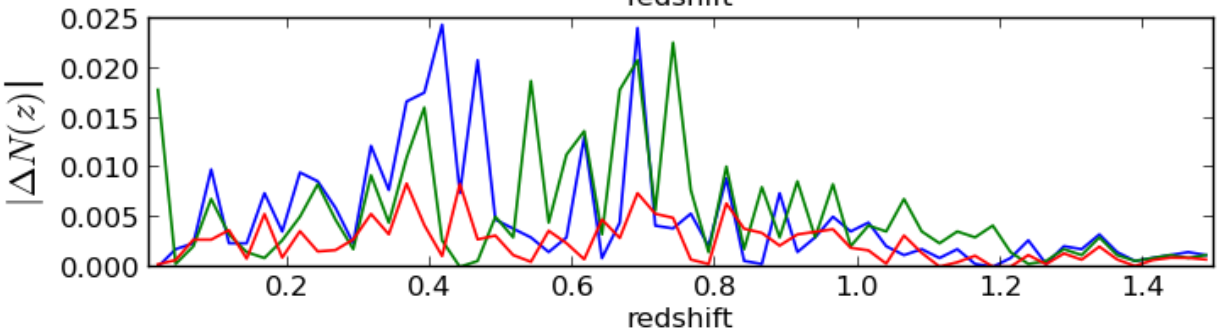
TPZ+Mod BPZ+prior
zConf > 0.9



Combining techniques







Simulated DES

We use TPZ (Carrasco Kind & Brunner, 2013a) to generate photo-z for all galaxies.

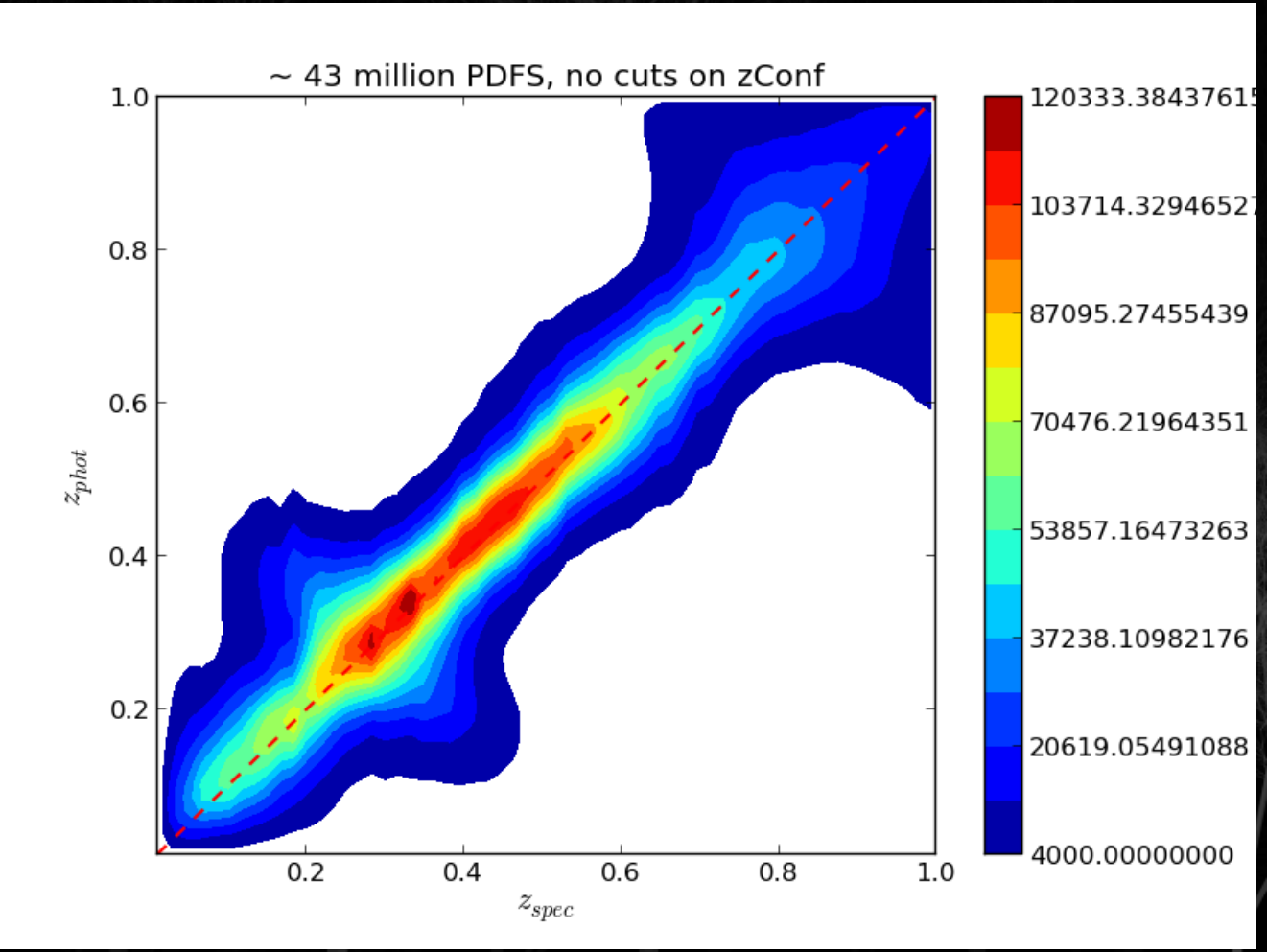
100,00 for training

5 magnitudes only

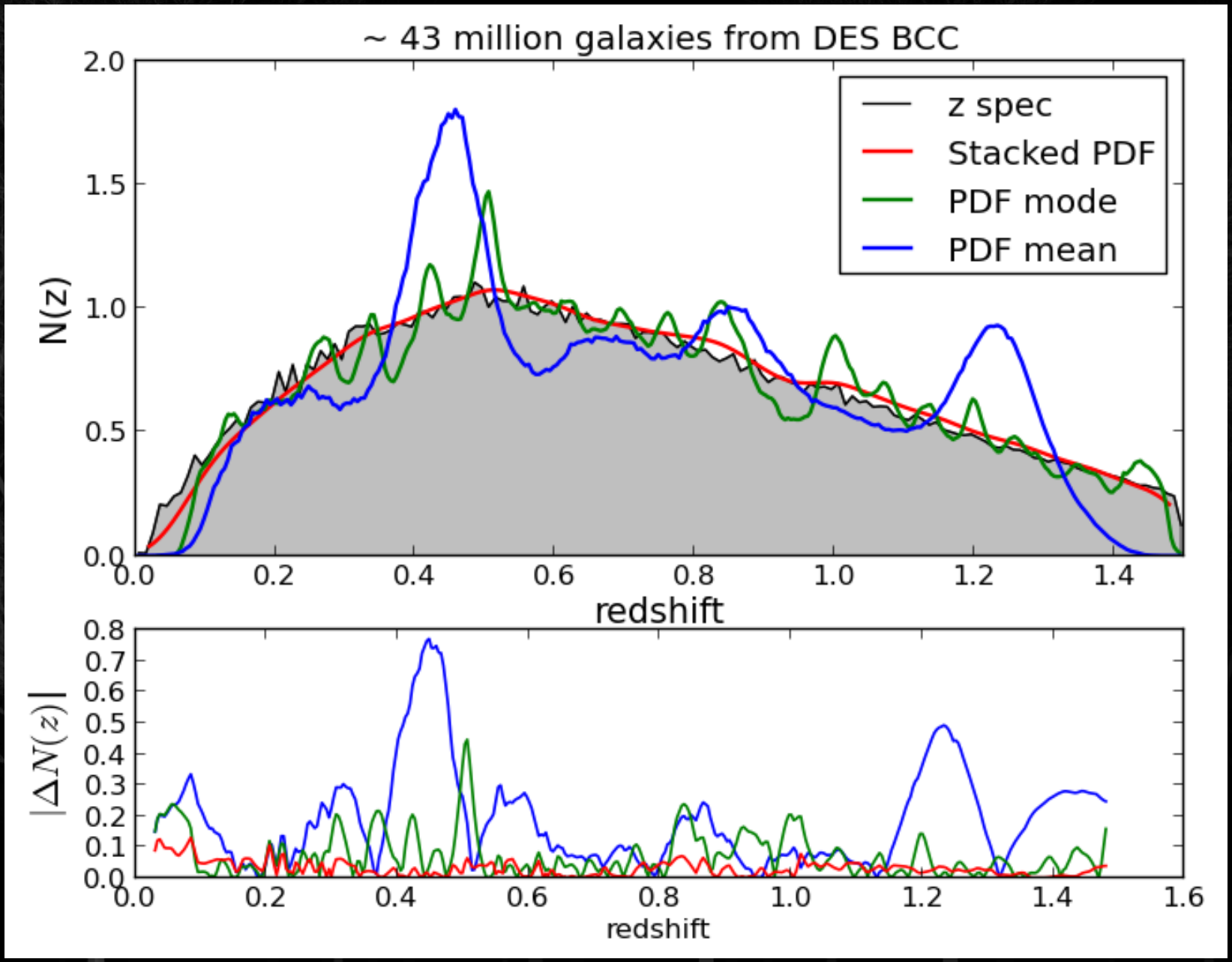
$\sim$ 0.17 sec per PDF

Store 43 million PDFs for analysis

No outlier removal



Simulated DES

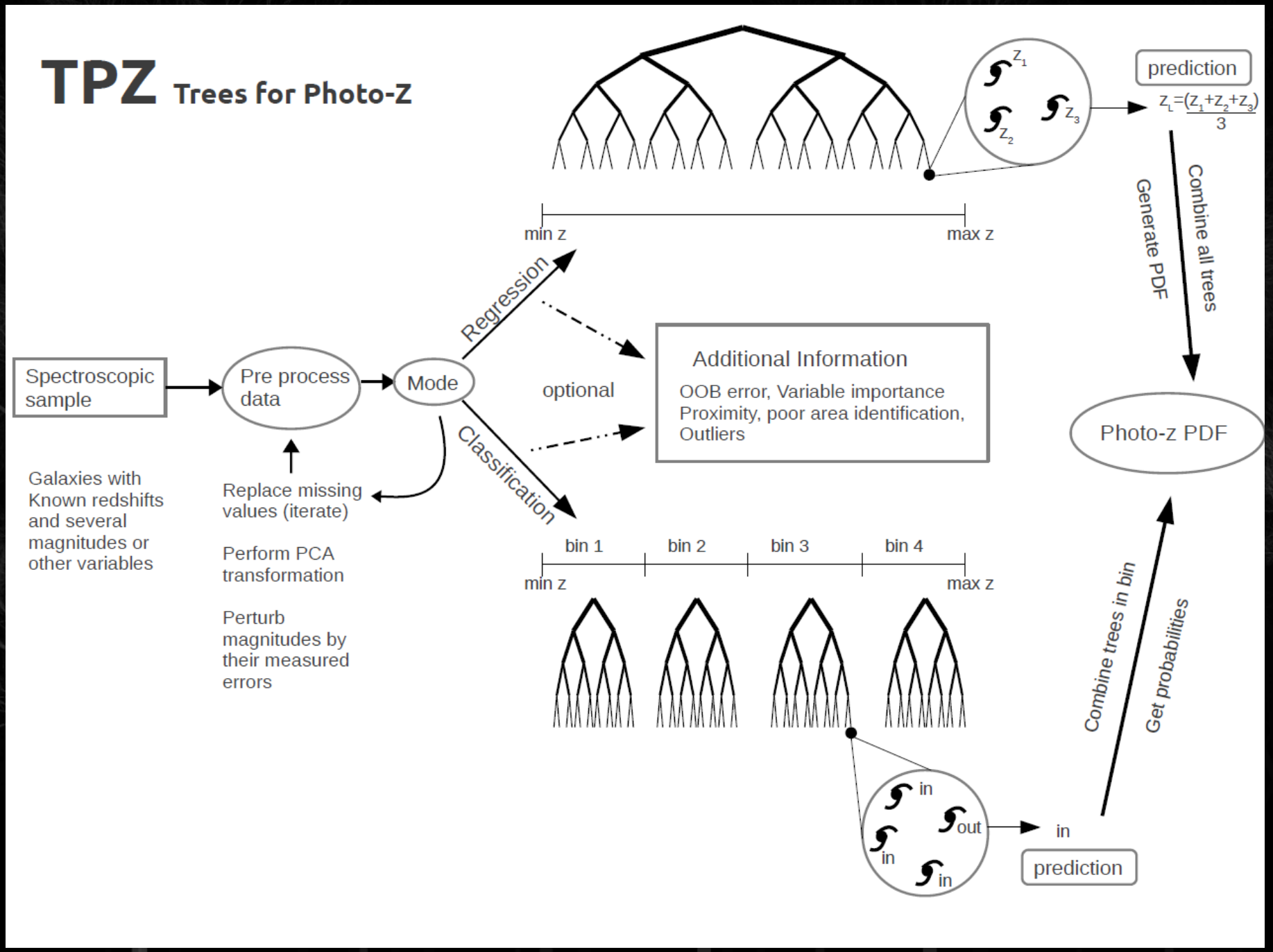


Thanks!

Questions?

Matias Carrasco Kind
University of Illinois
mcarras2@illinois.edu

TPZ scheme


TPZ Trees for Photo-Z
prediction
$z_L=\frac{(z_1+z_2+z_3)}{3}$
z_1
z_2
z_3
min z
max z
Generate PDF
Combine all trees
Spectroscopic sample
Pre process data
Mode
Regression
Classification
optional
Additional Information
OOB error, Variable importance
Proximity, poor area identification,
Outliers
Photo-z PDF
Galaxies with Known redshifts and several magnitudes or other variables
Replace missing values (iterate)
Perform PCA transformation
Perturb magnitudes by their measured errors
bin 1
bin 2
bin 3
bin 4
min z
max z
Combine trees in bin
Get probabilities
in
out
in
in
in
prediction

Performance tests