ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

How to produce, combine, store and use photo-z PDFs

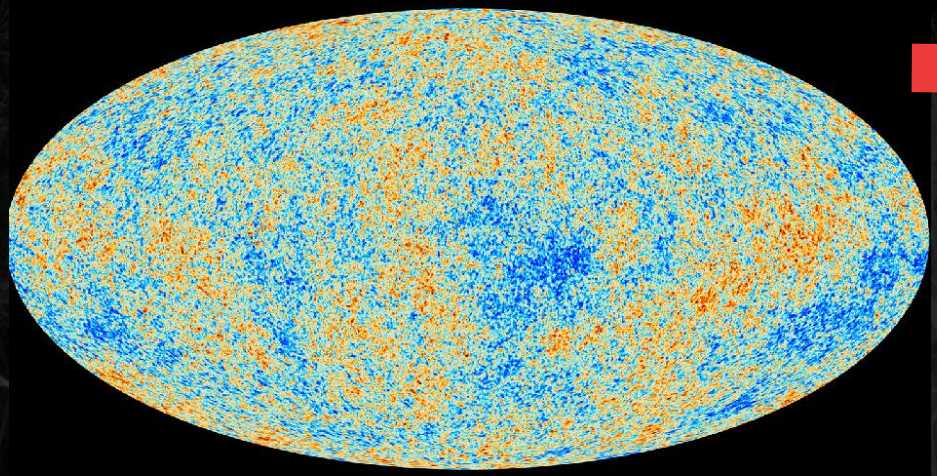
Matías Carrasco Kind

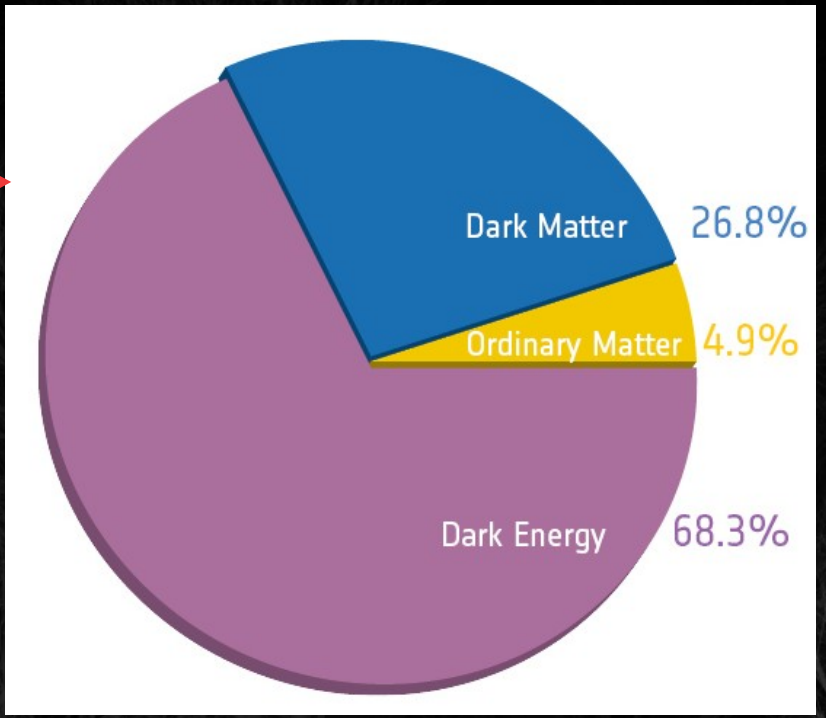
Department of Astronomy
University of Illinois

LIneA Seminar May 8, 2014

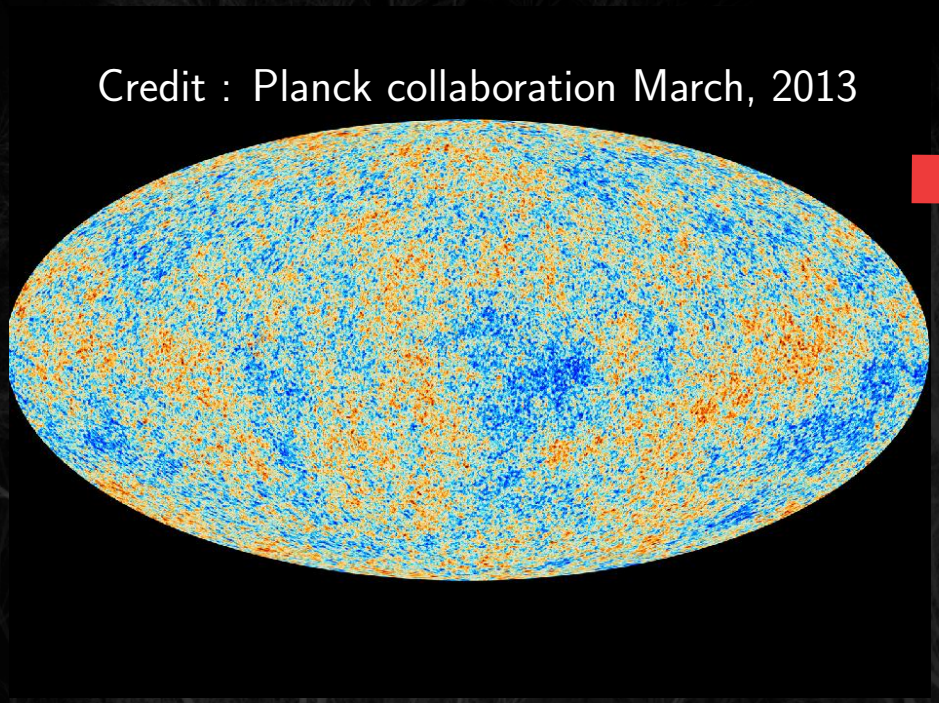
Current picture of the Universe

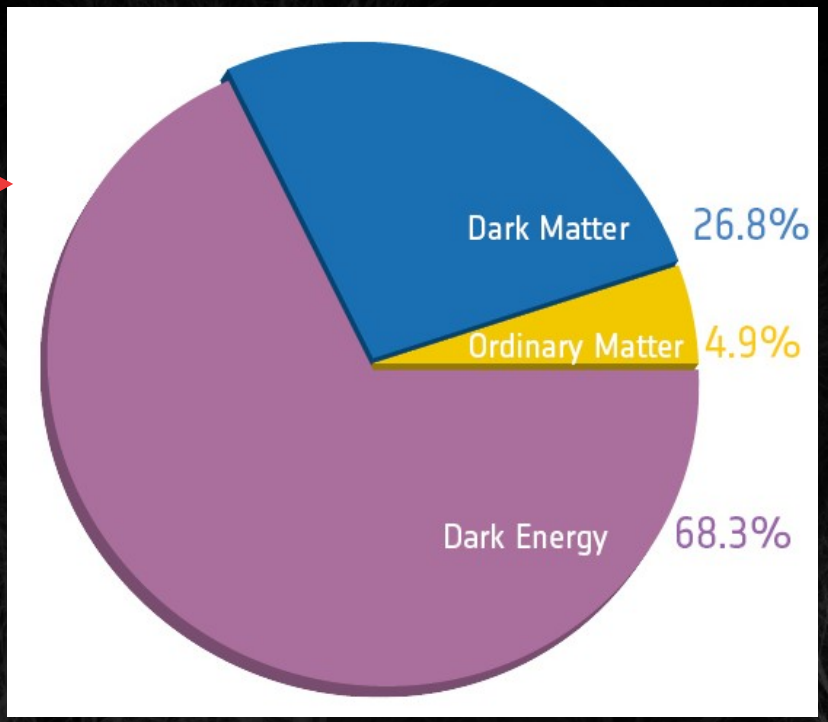
Credit : Planck collaboration March, 2013





Current picture of the Universe





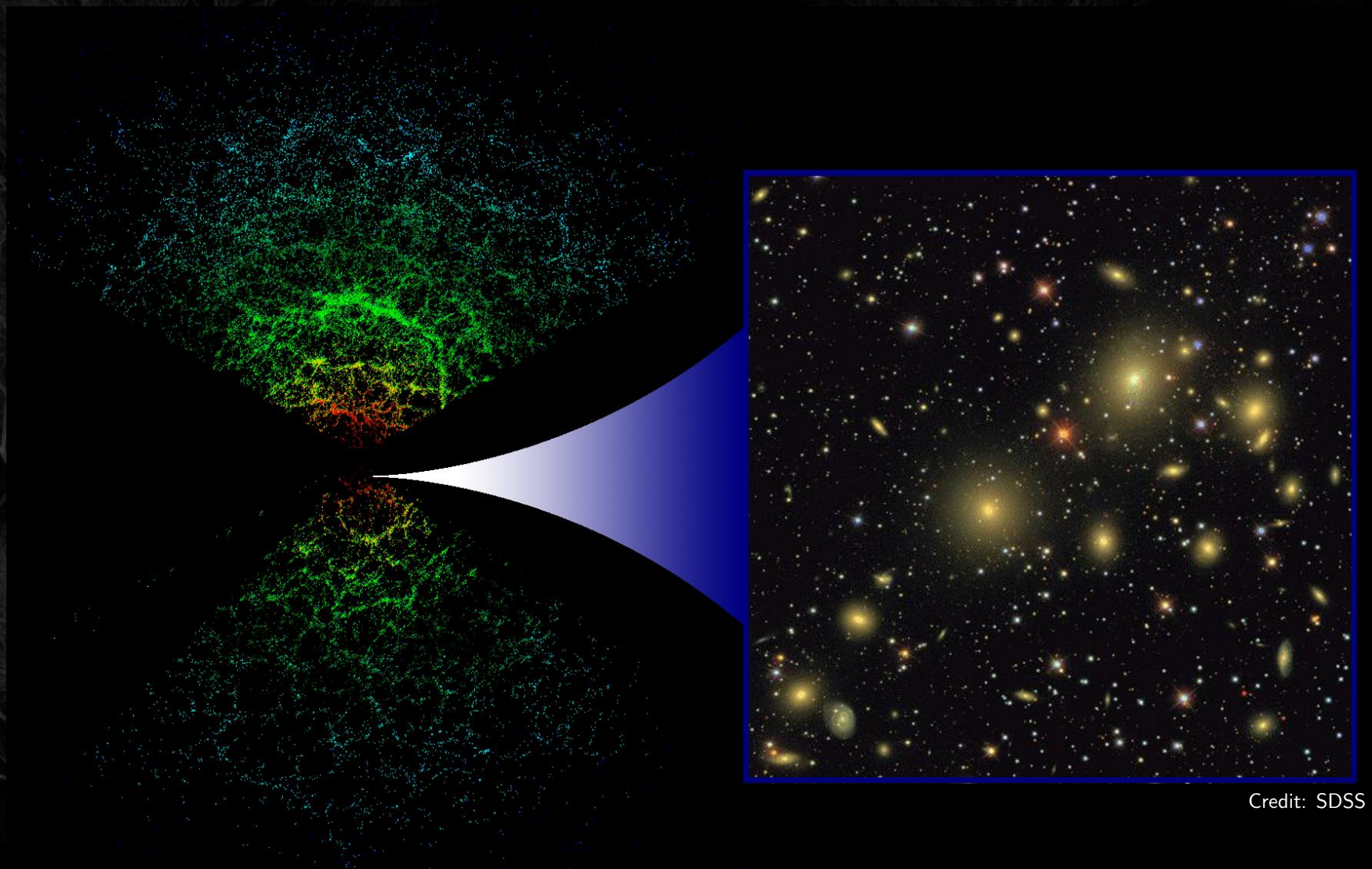
Still many questions about the Universe

Described by a set of cosmological parameters

Several methods to answer these question / constrain these

Statistical analysis of the spatial distribution of galaxies

The need of distances in cosmology



Credit: SDSS Collaboration

3D Clustering of galaxies as a probe in cosmology, e.g., 2 point correlation function, power spectrum of the galaxy distribution, etc.

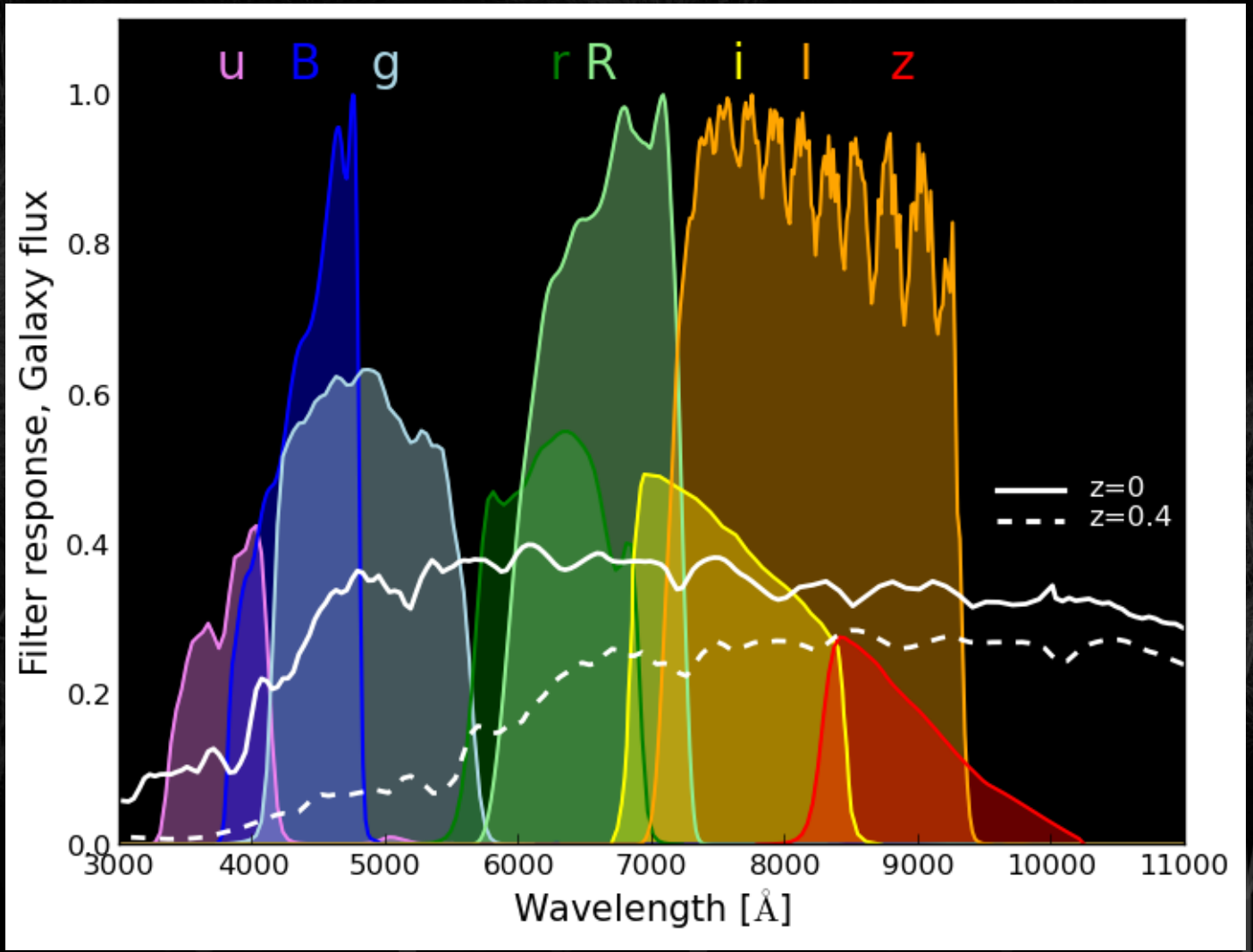
Photometric redshift (photo-z)

Examples of an elliptical spectra at z=0 and z=0.4

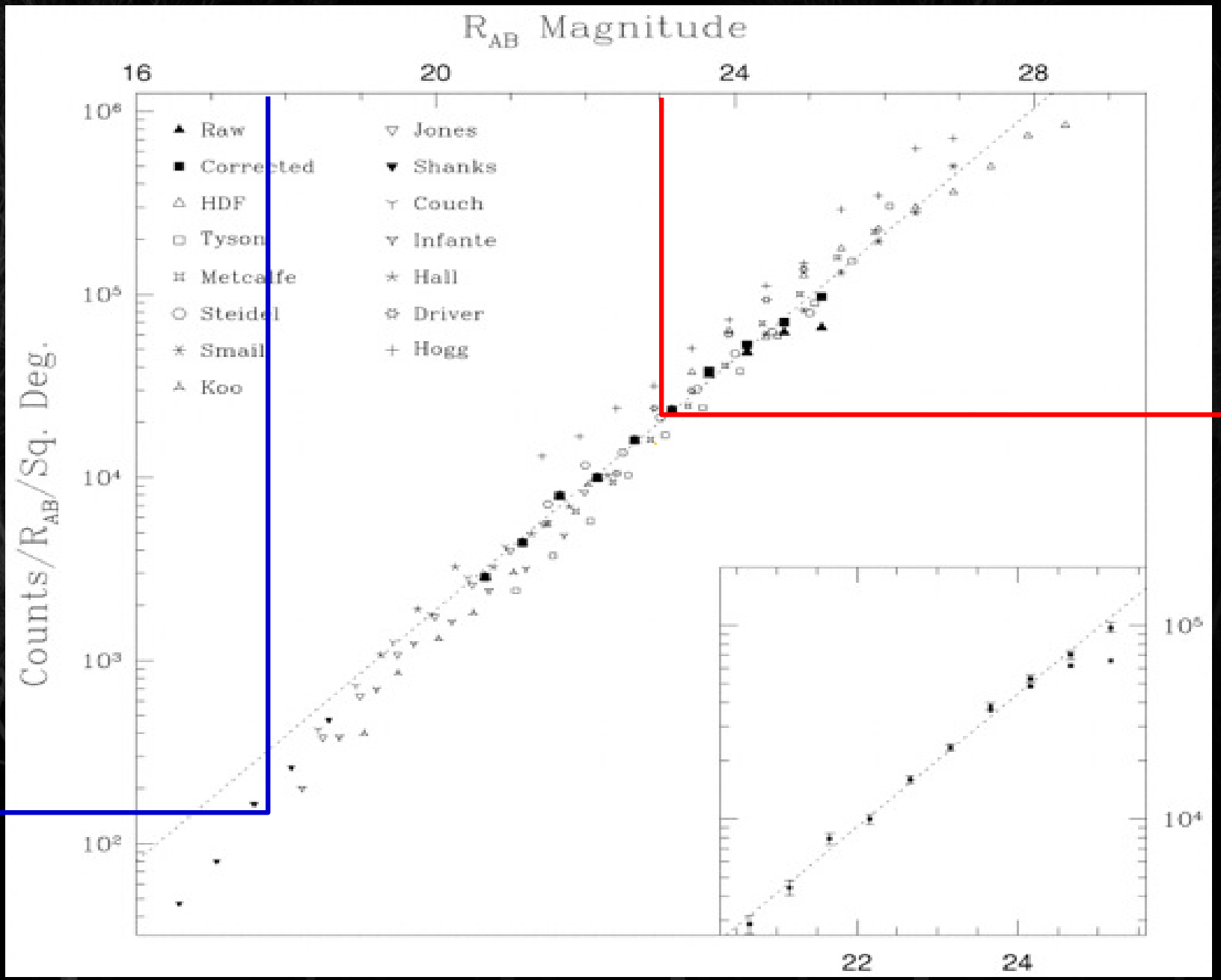
8 optical filters

Convolve spectrum with filter curves

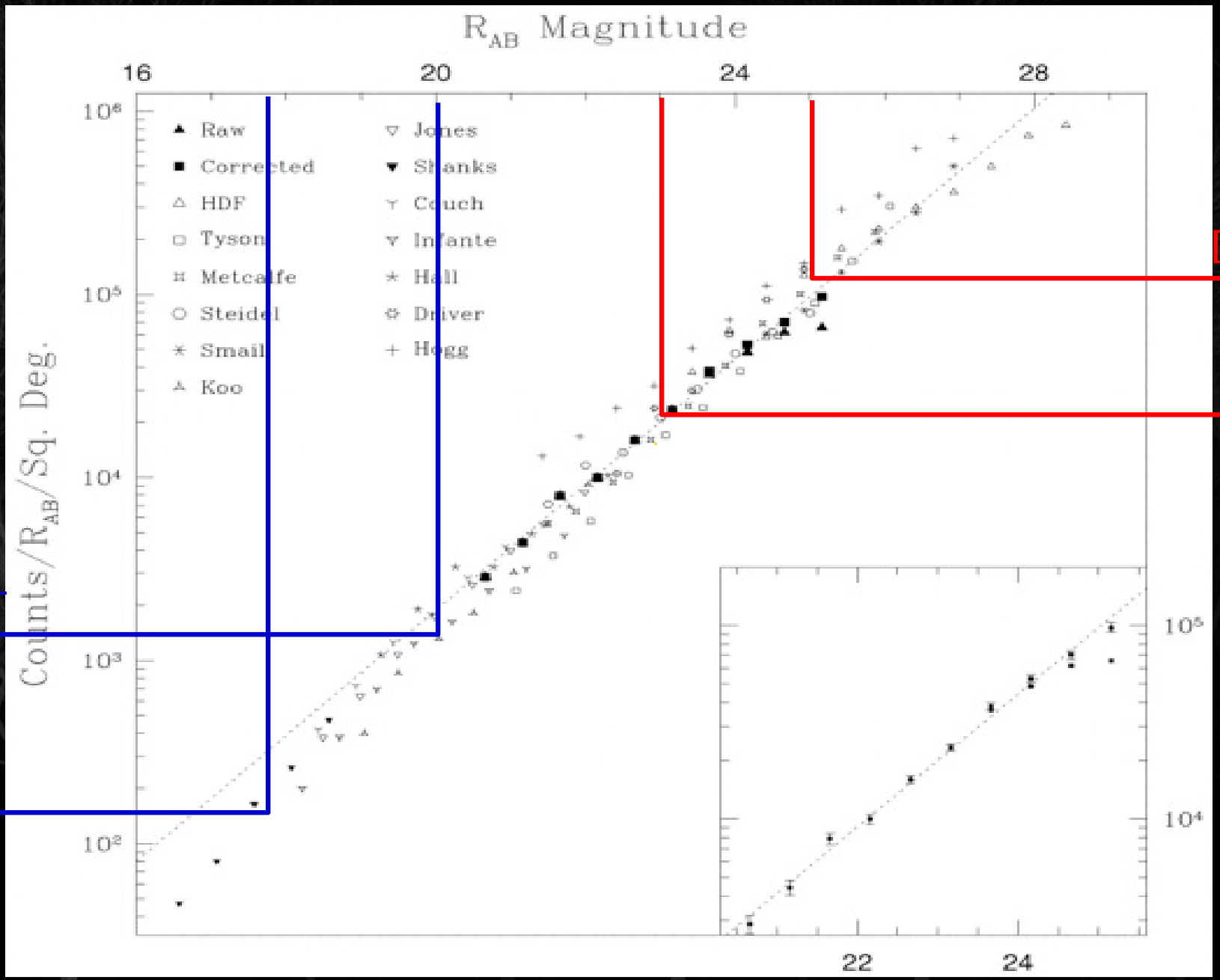
8 points instead of 5000 or more



Photometric surveys



Photometric surveys



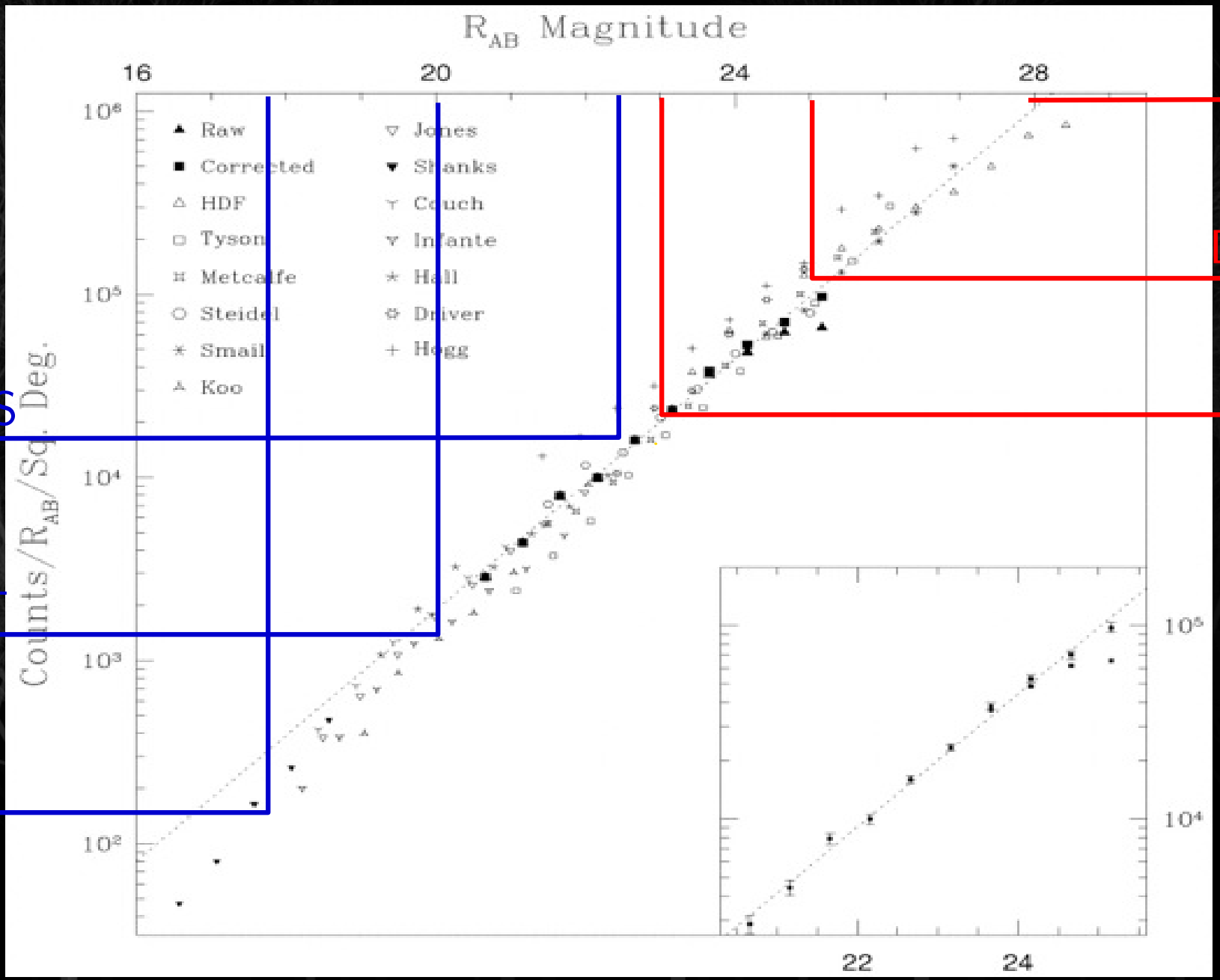
LAMOST

SDSS

DES final

SDSS

Photometric surveys



LSST

DES final

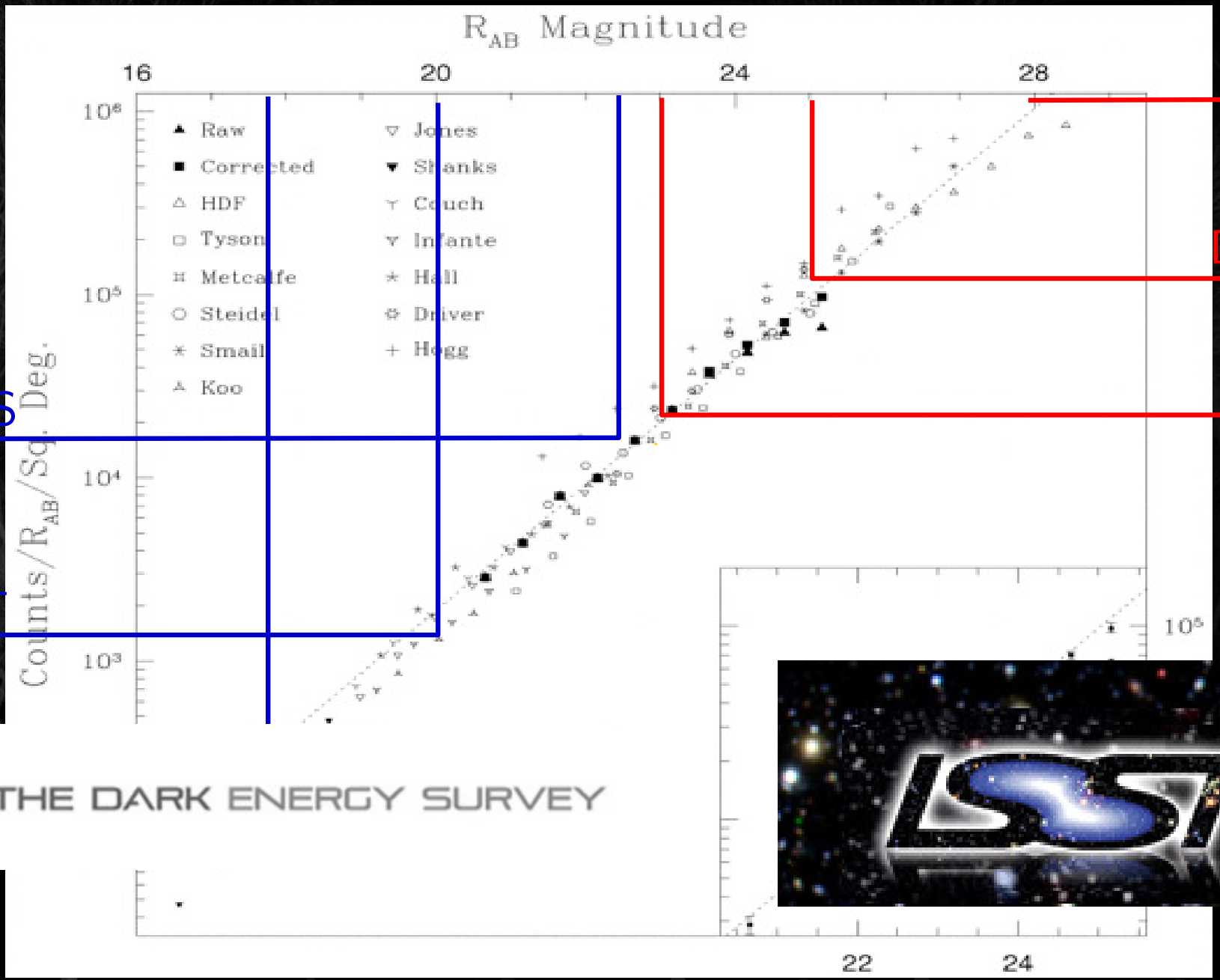
SDSS

Big BOSS

LAMOST

SDSS

Photometric surveys



LSST

DES final

SDSS

Big BOSS

LAMOST

Motivation

- Photo-z PDF are important in cosmology
- Several! methods/codes to compute photo-z
- Need for a meta-algorithm that combines multiple techniques
- PDF are good **but** for large datasets, storage and I/O will be an issue
- Machine Learning and statistical tools

Motivation

- Photo-z PDF are important in cosmology
- Several! methods / codes to compute photo-z
- Need for a meta-algorithm that combines multiple techniques
- PDF are good **but** for large datasets, storage and I/O will be an issue
- Machine Learning and statistical tools

Motivation

- Photo-z PDF are important in cosmology
- Several! methods/codes to compute photo-z
- Need for a meta-algorithm that combines multiple techniques
- PDF are good **but** for large datasets, storage and I/O will be an issue
- Machine Learning and statistical tools

Motivation

- Photo-z PDF are important in cosmology
- Several! methods/codes to compute photo-z
- Need for a meta-algorithm that combines multiple techniques
- PDF are good **but** for large datasets, storage and I/O will be an issue
- Machine Learning and statistical tools

Motivation

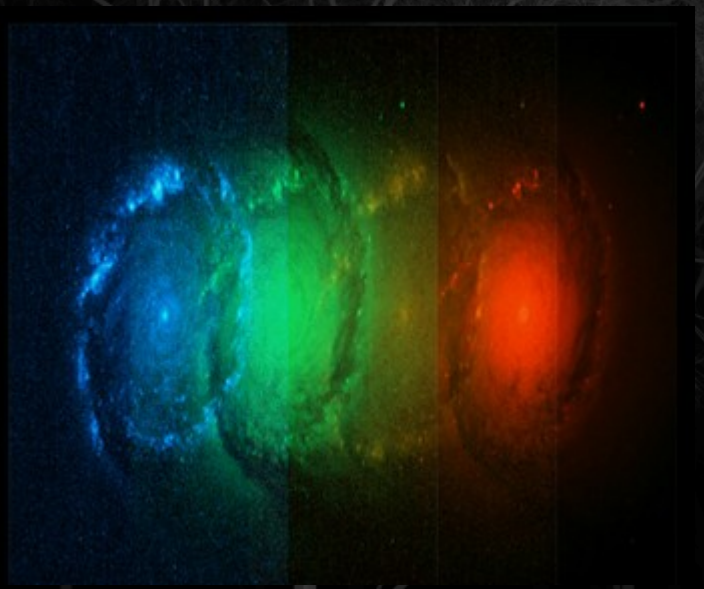
- Photo-z PDF are important in cosmology
- Several! methods/codes to compute photo-z
- Need for a meta-algorithm that combines multiple techniques
- PDF are good **but** for large datasets, storage and I/O will be an issue
- Machine Learning and statistical tools

Motivation

- Photo-z PDF are important in cosmology
- Several! methods/codes to compute photo-z
- Need for a meta-algorithm that combines multiple techniques
- PDF are good **but** for large datasets, storage and I/O will be an issue
- Machine Learning and statistical tools

How to produce, combine, store and use P(z)

Photo-z PDF estimation



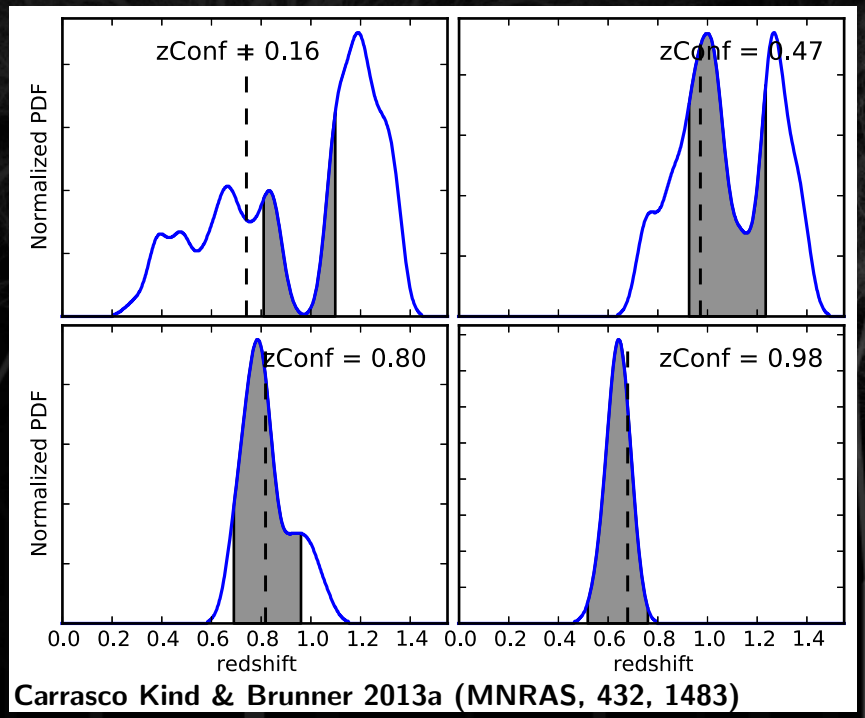
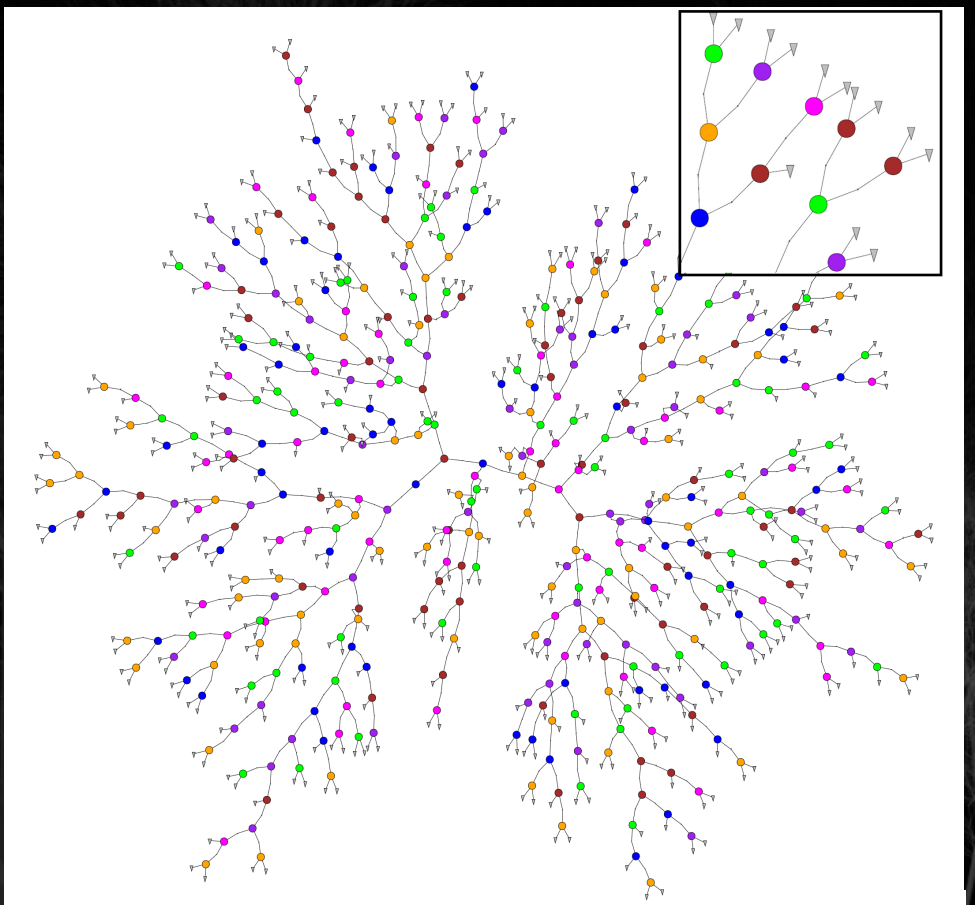


Photo-z PDF estimation: TPZ

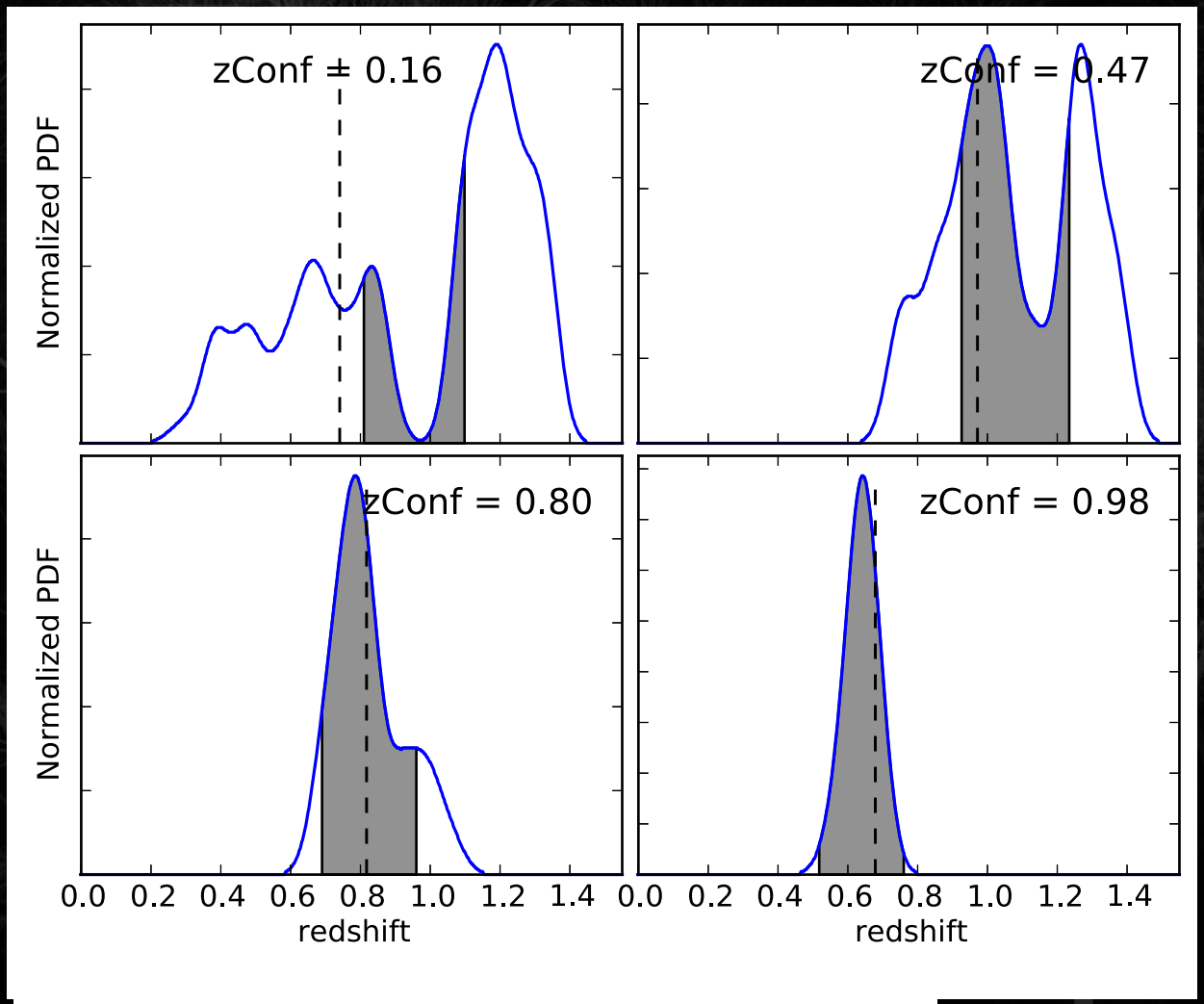
- TPZ (Trees for Photo-Z) is a supervised machine learning code
- Prediction trees and random forest
- Incorporate measurements errors and deals with missing values
- Ancillary information: expected errors, attribute ranking and others
- Application to the S/G



Carrasco Kind & Brunner 2013a (MNRAS, 432, 1483)

http://lcdm.astro.illinois.edu/code/mlz.html

Photo-z PDF estimation: TPZ



Carrasco Kind & Brunner 2013a (MNRAS, 432, 1483)

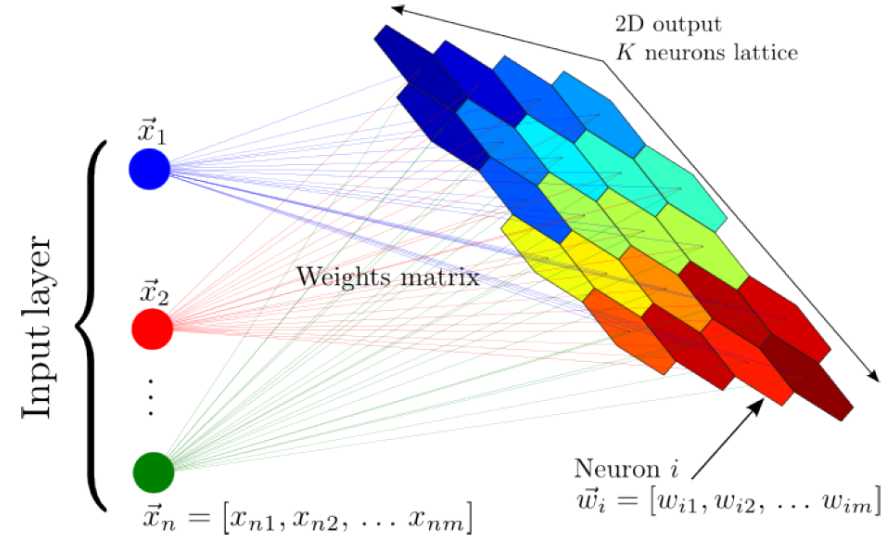
4 PDFs from the DEEP2 catalog with dfferent $zConf$ levels

$$zConf = \int_{z_1}^{z_2} P(z)dz$$

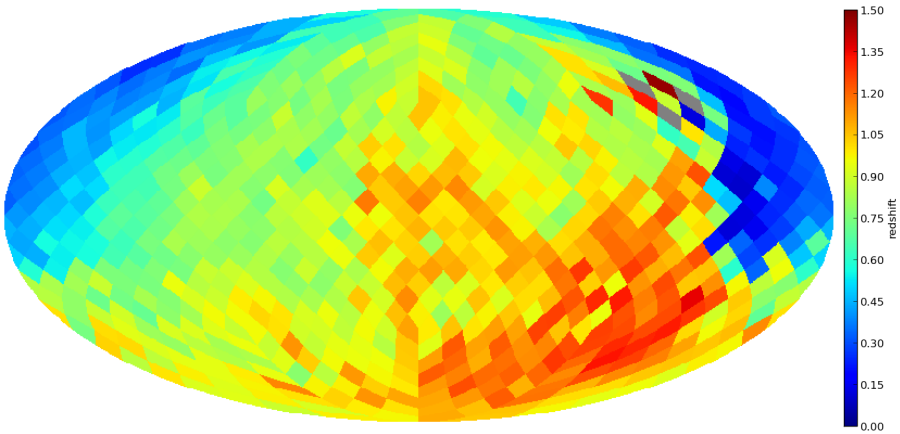
$$z_1, z_2 = z_{\rm phot} \pm \sigma_{TPZ}(1 + z_{\rm phot})$$

Photo-z PDF estimation: SOM

- SOM(Self Organized Map) is a unsupervised machine learning algorithm
- Competitive learning to represent data conserving topology
- 2D maps and *Random Atlas*
- Framework inherited from TPZ
- Application to the S/G



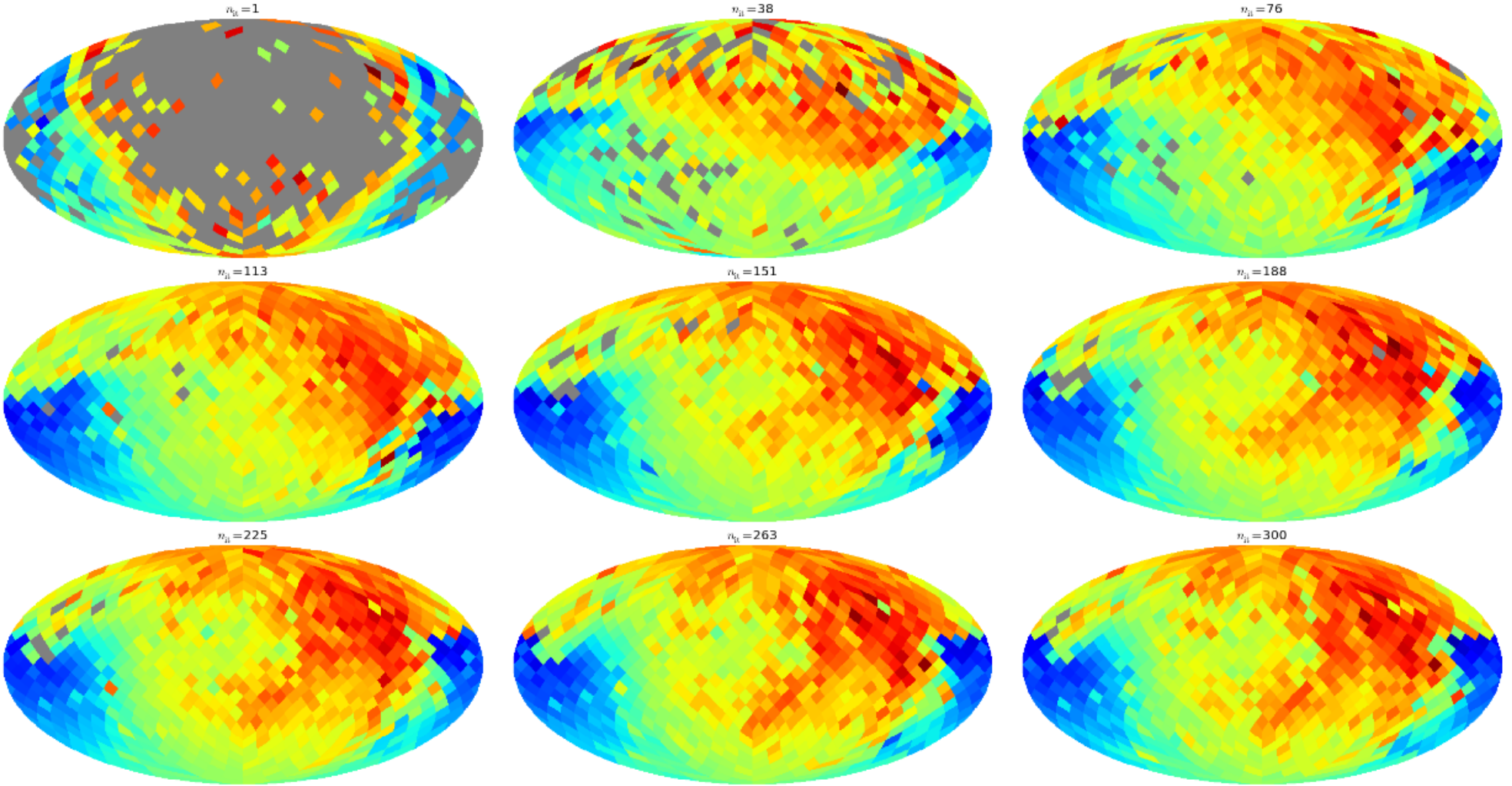
Carrasco Kind & Brunner 2014a (MNRAS, 438, 3409)



Carrasco Kind & Brunner 2014a (MNRAS, 438, 3409)

Photo-z PDF estimation: SOM

Self organize map construction, colors indicate median redshift of each cell



Carrasco Kind & Brunner 2014a (MNRAS, 438, 3409)

Photo-z PDF estimation: BPZ

- BPZ (Benitez, 2000) is a Bayesian template fitting method to obtain PDFs
- Set of calibrated SED and filters
- Doesn't need training data
- Priors can be included

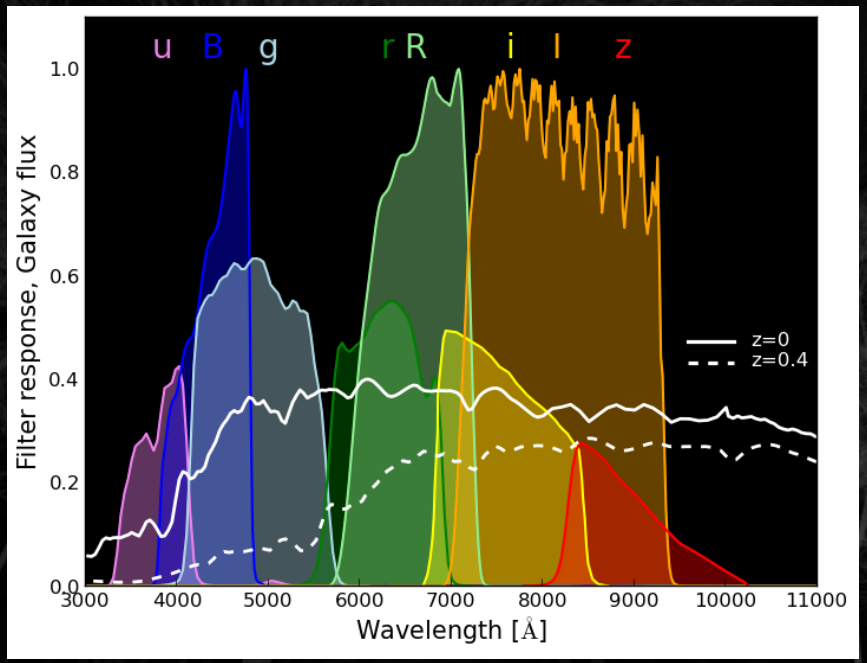


Photo-z PDF estimation: BPZ

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

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

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

Photo-z PDF estimation: BPZ

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

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

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

Prior

Likelihood

Photo-z PDF estimation: BPZ

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

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

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

Prior

Likelihood

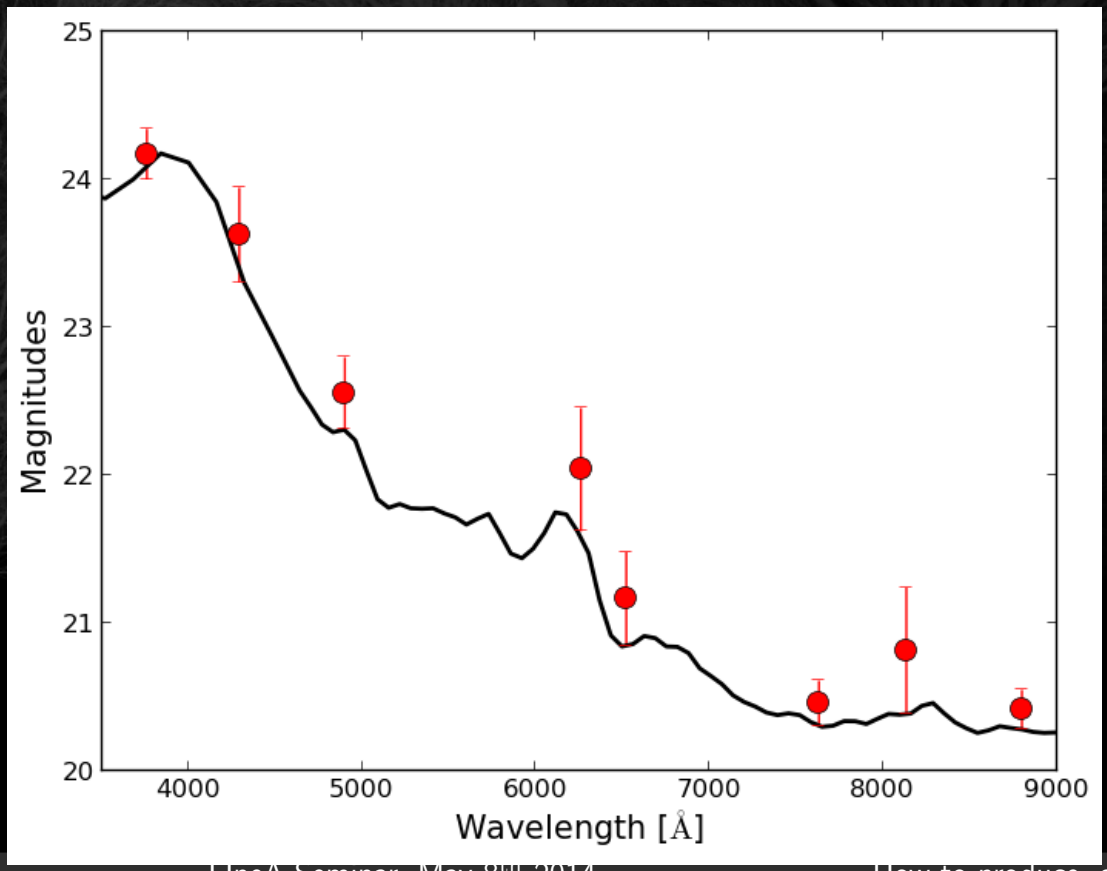


Photo-z PDF estimation: MLZ

MLZ :Machine Learning for photo-Z

http://lcdm.astro.illinois.edu/code/mlz.hml

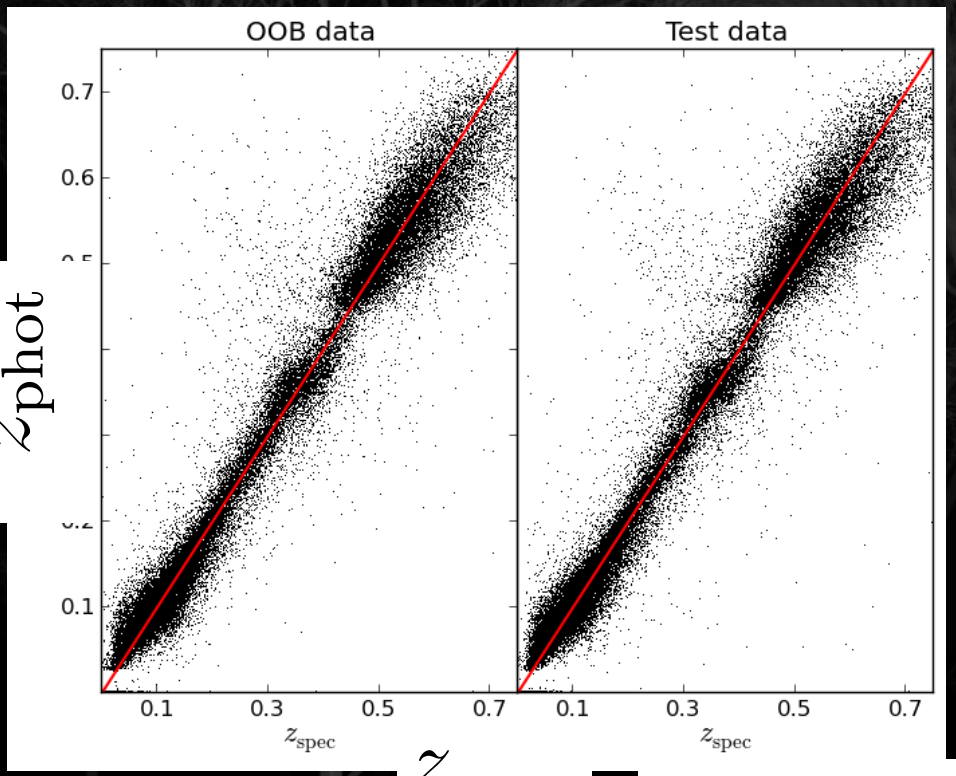
- TPZ, SOM and BPZ incorporated in one python framework
- Public, parallel and easy to use
- PDF Sparse representation included
- Current version 1.1, GitHub repository
- pycuda and numba in folder

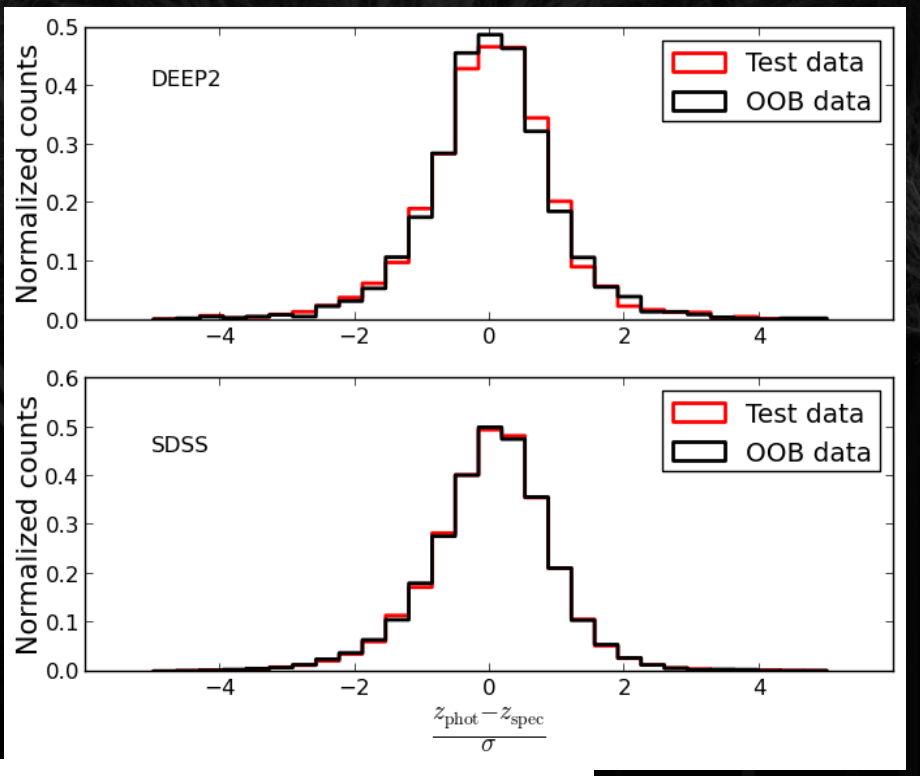
Photo-z PDF estimation: Error and validation

Out of Bag data used to validate trees/maps

Changes for every tree/map and is not used during training

We can learn from the cross-validation data!



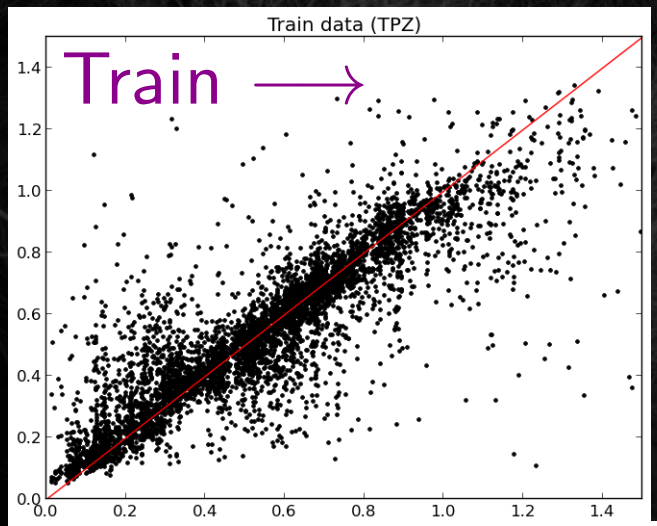


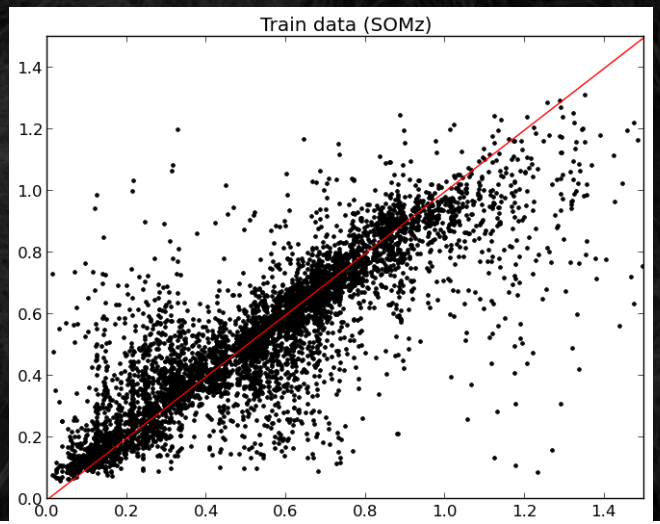
Carrasco Kind & Brunner 2014c (MNRAS submitted)

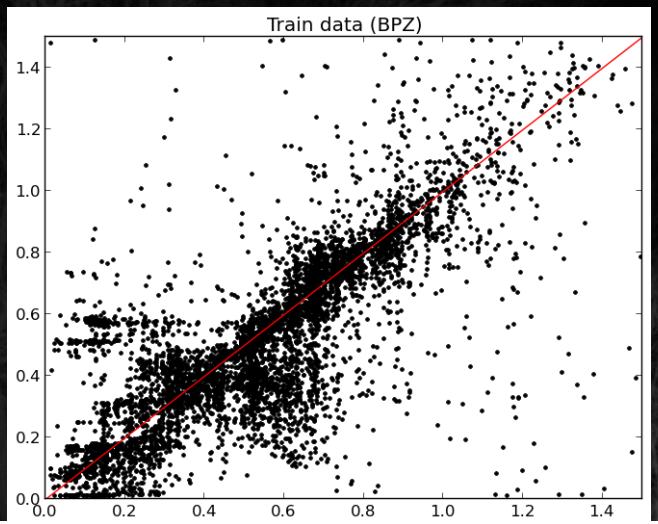
Photo-z PDF estimation: Error and validation

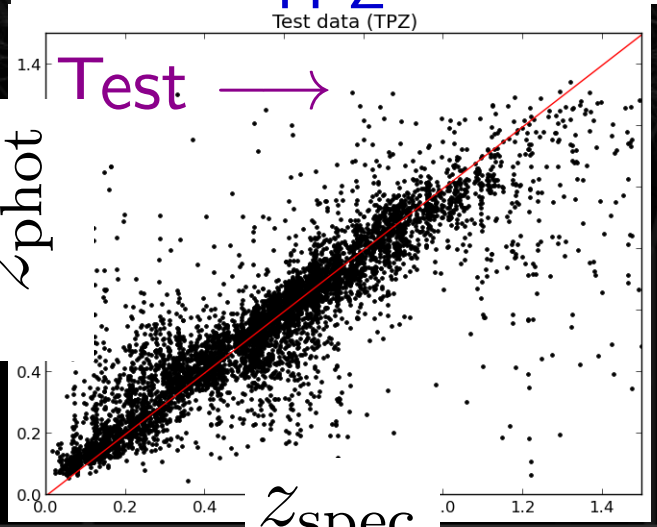
We can improve the estimation but combining information

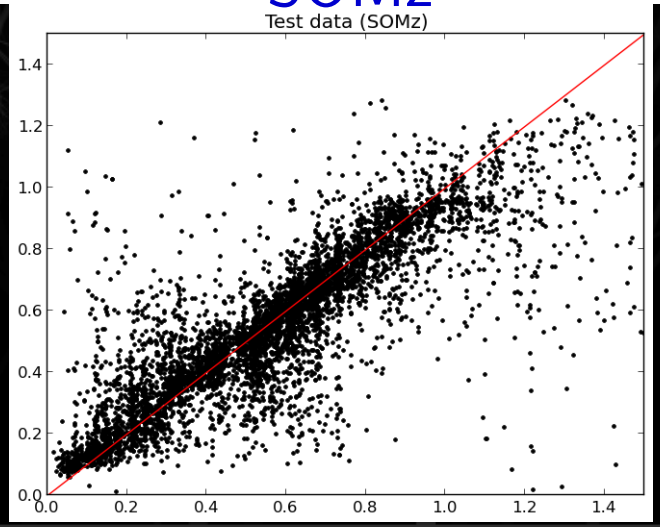
These are plots from DES SV data

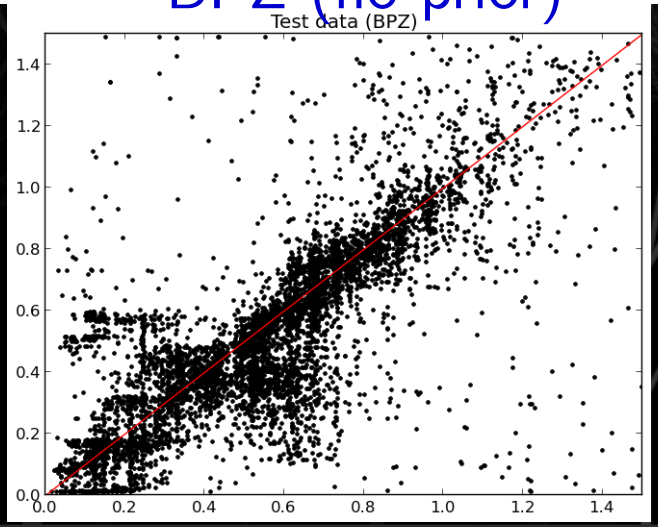






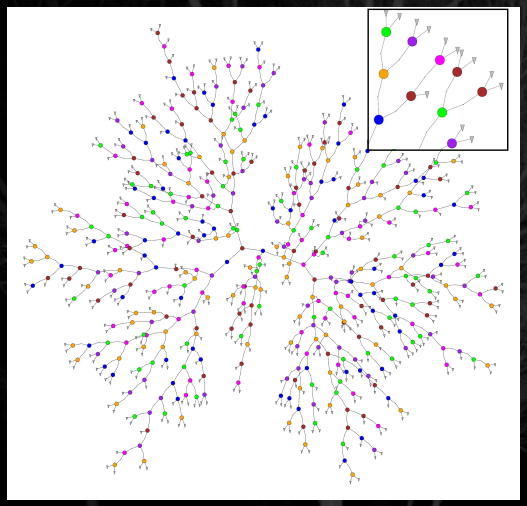




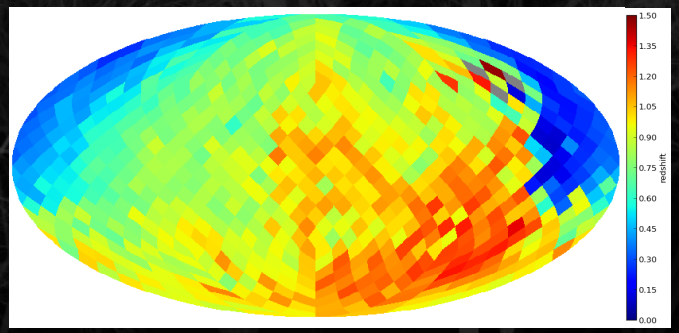


How to produce, combine, store and use P(z)

Photo-z PDF combination



+



+

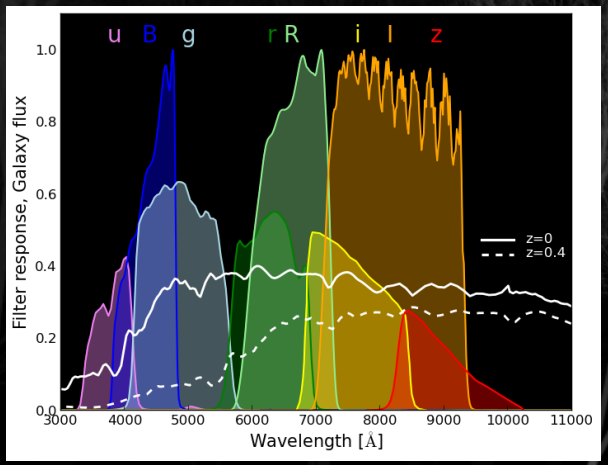


Photo-z PDF combination: Bayesian framework

- Use random naïve bayes model to compute individual priors (Carrasco Kind & Brunner, 2013b)

- We explored different models such as: (Carrasco Kind & Brunner, 2014c, submitted arxiv: 1403.0044)

  - Hierarchical Bayes model (Dahlen et al., 2013)
  - Bayesian model averaging and combination
  - MCMC parameter estimation
  - Use machine learning to learn from outliers and errors

Photo-z PDF combination: Bayesian framework

- Use random naïve bayes model to compute individual priors (Carrasco Kind & Brunner, 2013b)

- We explored different models such as:
(Carrasco Kind & Brunner, 2014c, submitted arxiv: 1403.0044)

  - Hierarchical Bayes model (Dahlen et al., 2013)
  - Bayesian model averaging and combination
  - MCMC parameter estimation
  - Use machine learning to learn from outliers and errors

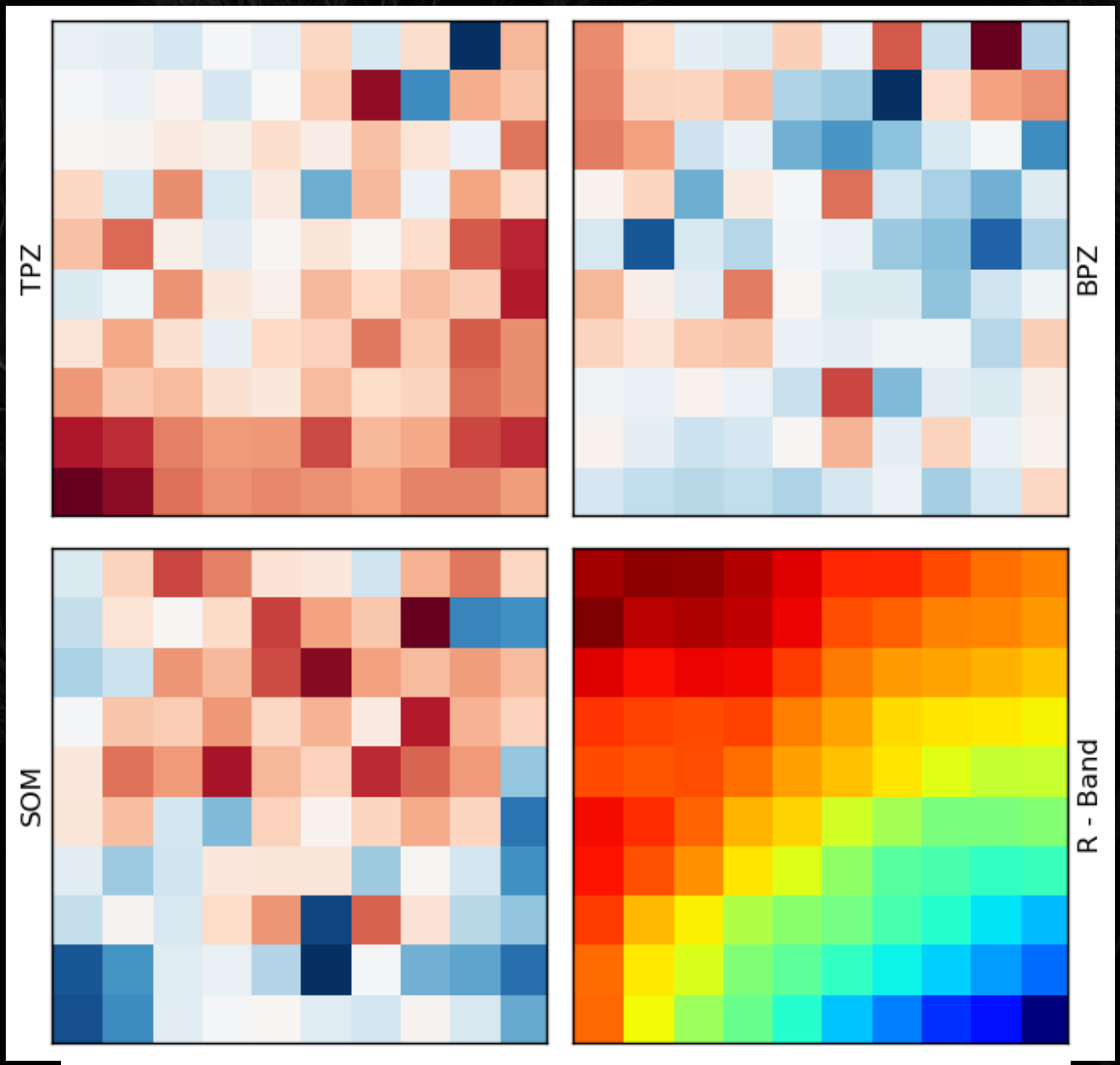
Photo-z PDF combination: Bayesian framework

- Use random naïve bayes model to compute individual priors (Carrasco Kind & Brunner, 2013b)

- We explored different models such as: (Carrasco Kind & Brunner, 2014c, submitted arxiv: 1403.0044)

  - Hierarchical Bayes model (Dahlen et al., 2013)
  - Bayesian model averaging and combination
  - MCMC parameter estimation
  - Use machine learning to learn from outliers and errors

Photo-z PDF combination: Bayesian framework



Carrasco Kind & Brunner 2014c (MNRAS submitted)

Our approach

Supervised method

+

Unsupervised method

+

Template fitting

+

Weigthing scheme

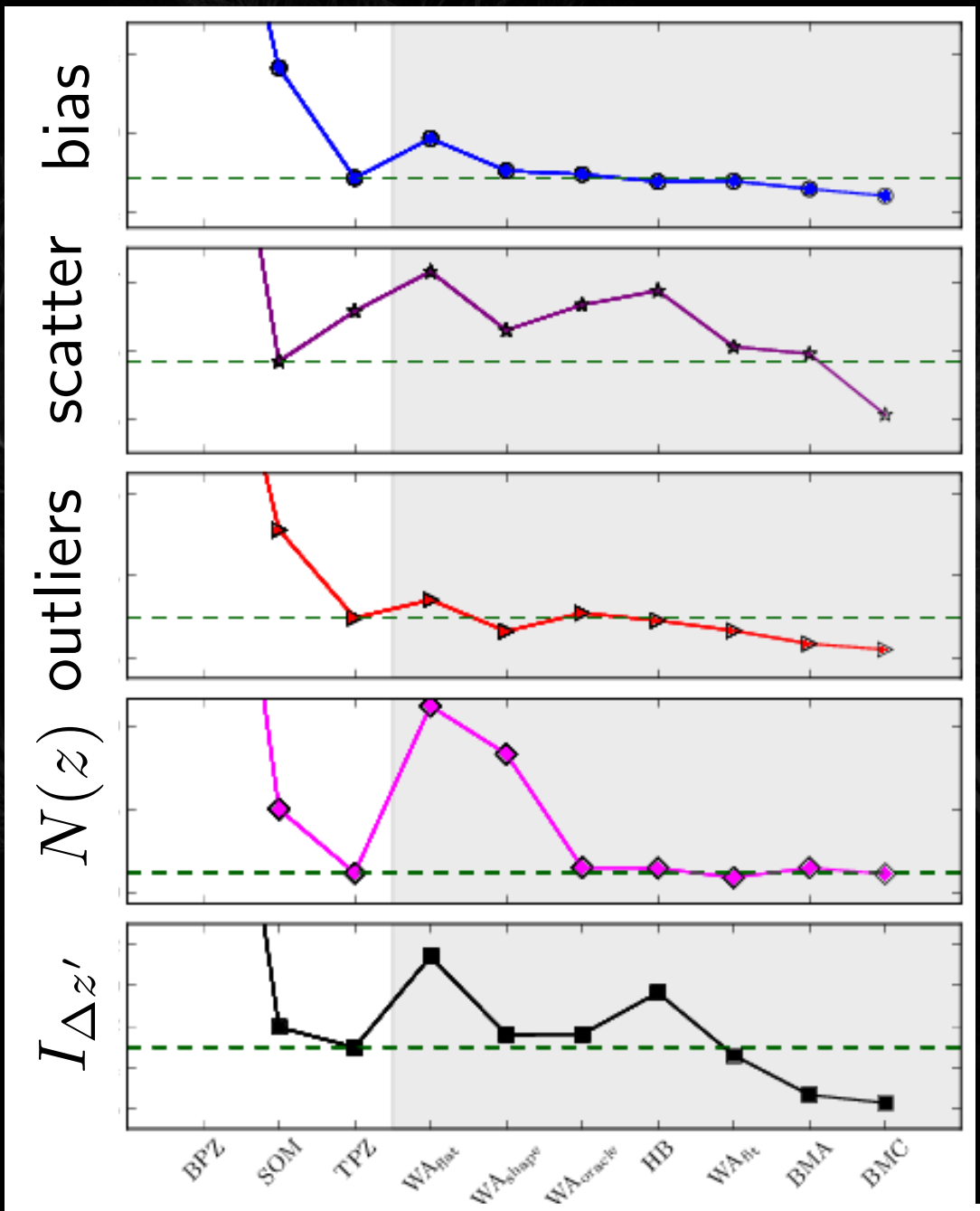
⇓

photo-z PDF

+

Outliers

Photo-z PDF combination: Results



Carrasco Kind & Brunner 2014c (MNRAS submitted)

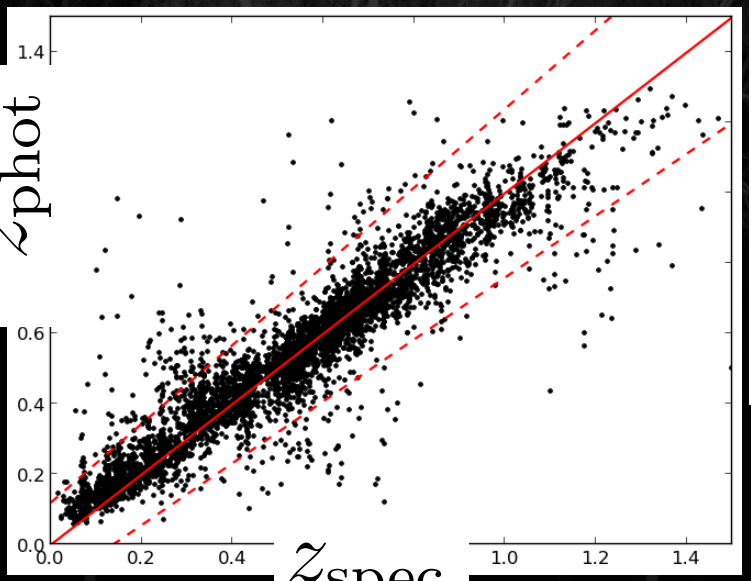
- Several combination methods
- Bayesian model averaging (BMA) and combination (BMC) are the best
- We introduce the I-score which combine multiple metics after being rescaled to compare different methods and/or codes

$$I_{\Delta z'} = \sum w_i M_i$$

Photo-z PDF combination: DES Results

Averaged metrics for all test galaxies

$$\Delta z = \frac{|z_{\rm spec} - z_{\rm phot}|}{1 + z_{\rm spec}}$$



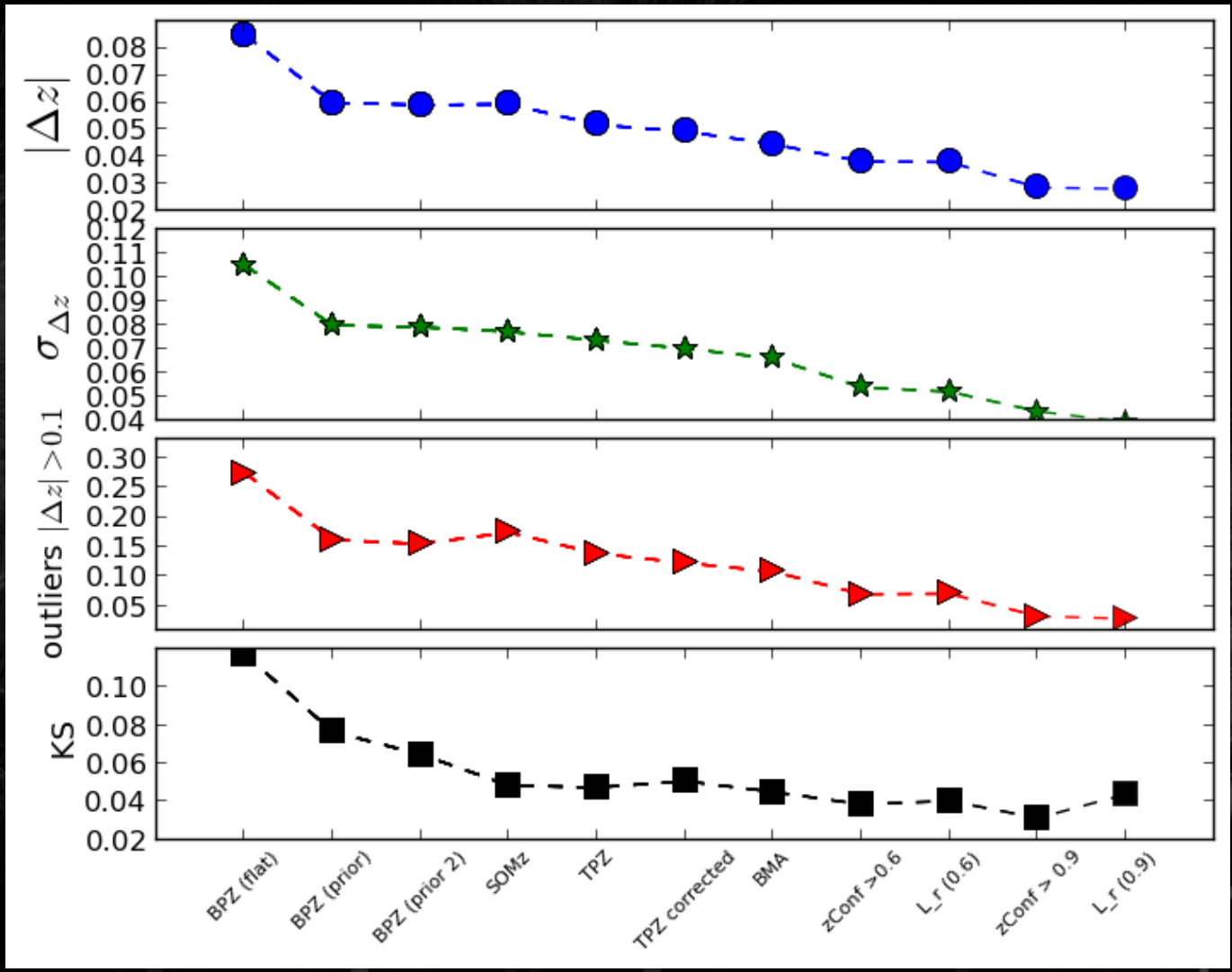


Photo-z PDF combination: Outliers

Naïve Bayes Classifier (same used for spam emails) to identify "spam" galaxies using information from multiple techniques

The prob. given a set of N_θ "features" $\boldsymbol{\theta}$ is:

$$P(\text{out} \mid \boldsymbol{\theta}) = \frac{P(\text{out})P(\boldsymbol{\theta} \mid \text{out})}{P(\boldsymbol{\theta})}$$

Naïvely the Likelihood is given assuming independence:

$$P(\boldsymbol{\theta} \mid \text{out}) = P(\theta_1, \theta_2, \ldots, \theta_{N_\theta} \mid \text{out}) = \prod_{i=1}^{N_\theta} P(\theta_i \mid \text{out})$$

then:

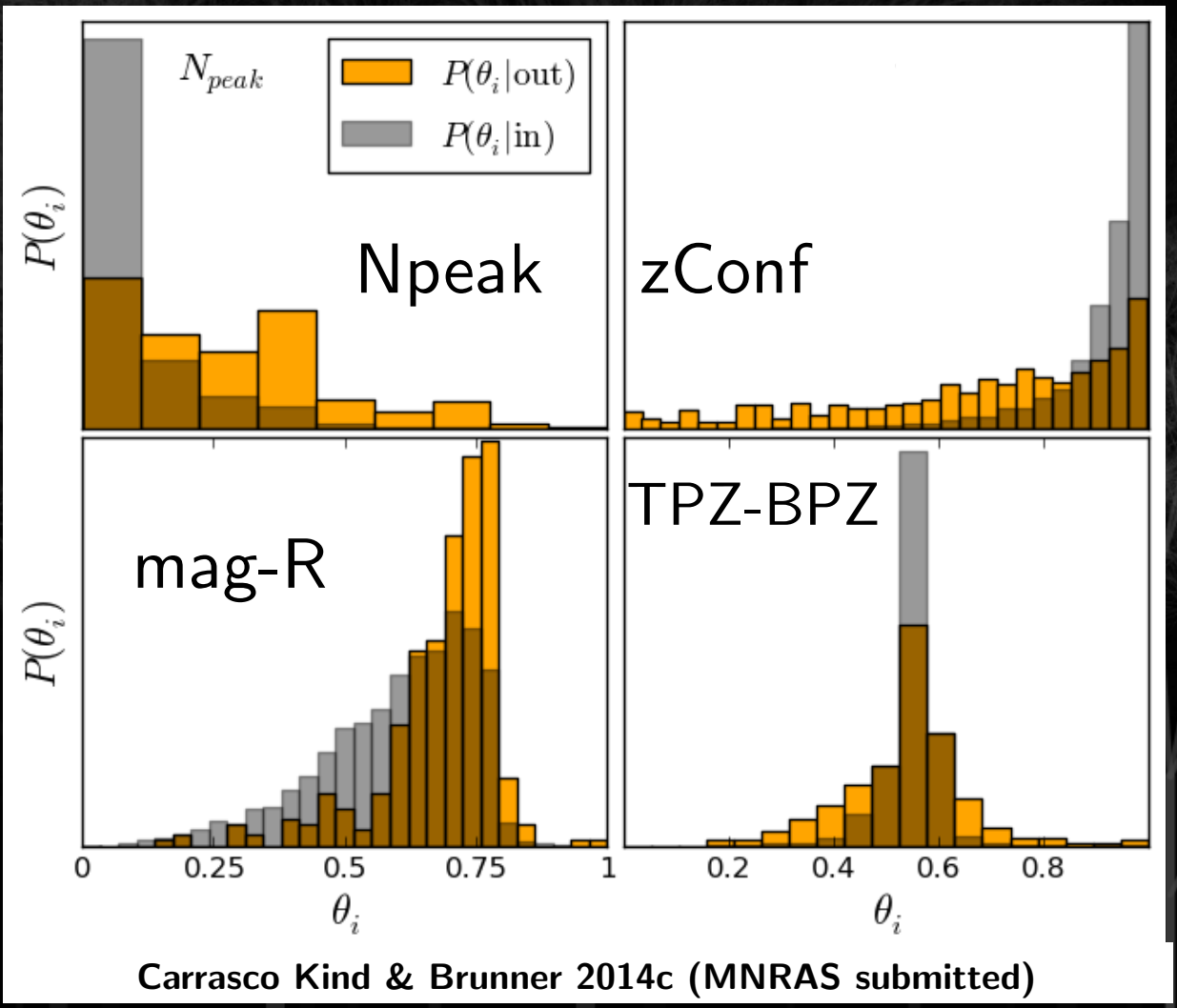
$$P(\text{out} \mid \boldsymbol{\theta}) = \frac{P(\text{out}) \prod P(\theta_i \mid \text{out})}{\prod P(\theta_i \mid \text{out}) + \prod P(\theta_i \mid \text{in})}$$

$\boldsymbol{\theta}$ includes: number of peaks, magnitudes, shape of PDF, differences, etc...

Photo-z PDF combination: Outliers

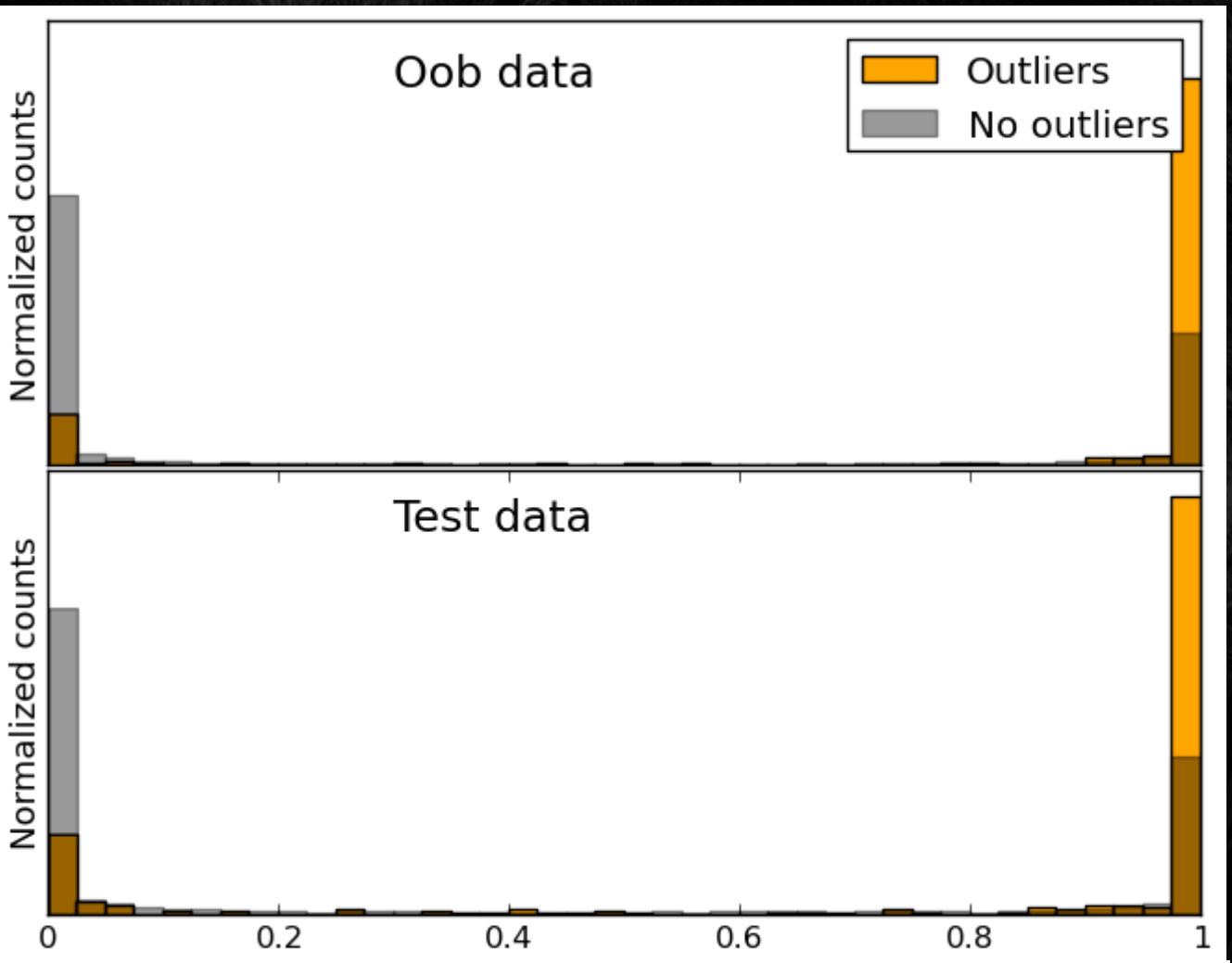
Naïve Bayes Classifier (same used for spam emails) to identify "spam" galaxies using information from multiple techniques

Each feature provides information about these two classes, and can be combined to make a stronger classifier



Carrasco Kind & Brunner 2014c (MNRAS submitted)

Photo-z PDF combination: Outliers



$P(\text{out} \mid \boldsymbol{\theta})$

Carrasco Kind & Brunner 2014c (MNRAS submitted)

- Highly bimodal
- Little contamination
- Good discriminant
- Consistent between samples

Photo-z PDF storage

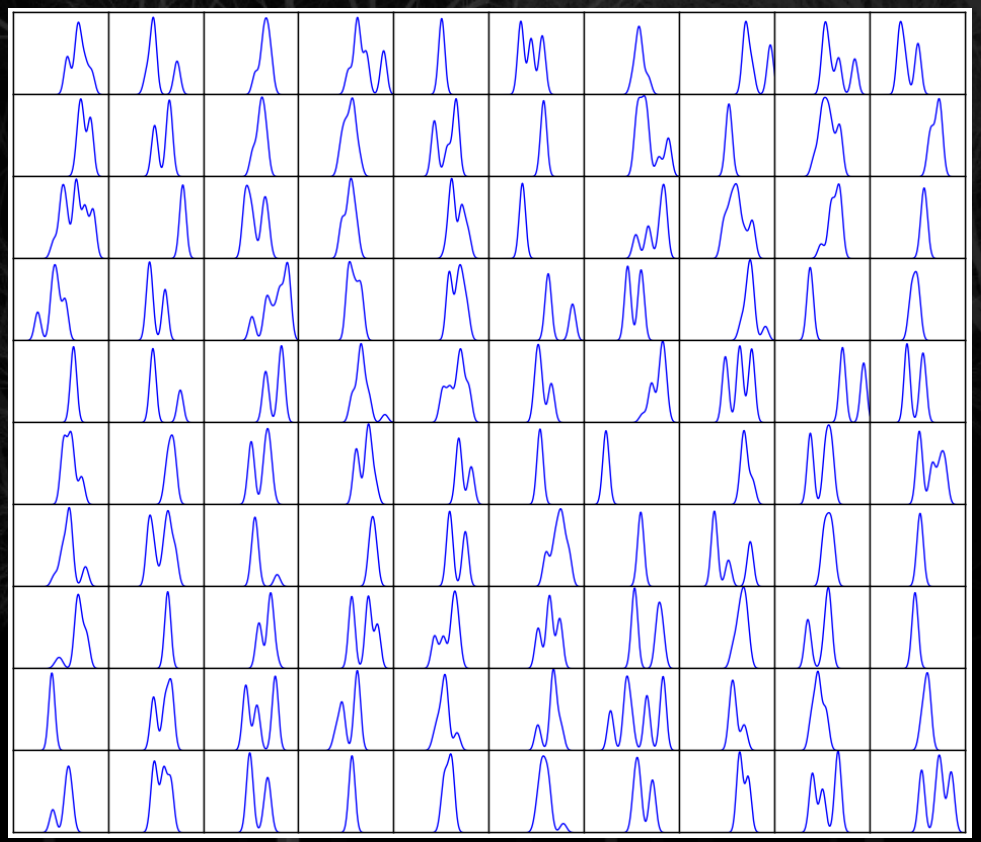


Photo-z PDF storage: Strategies

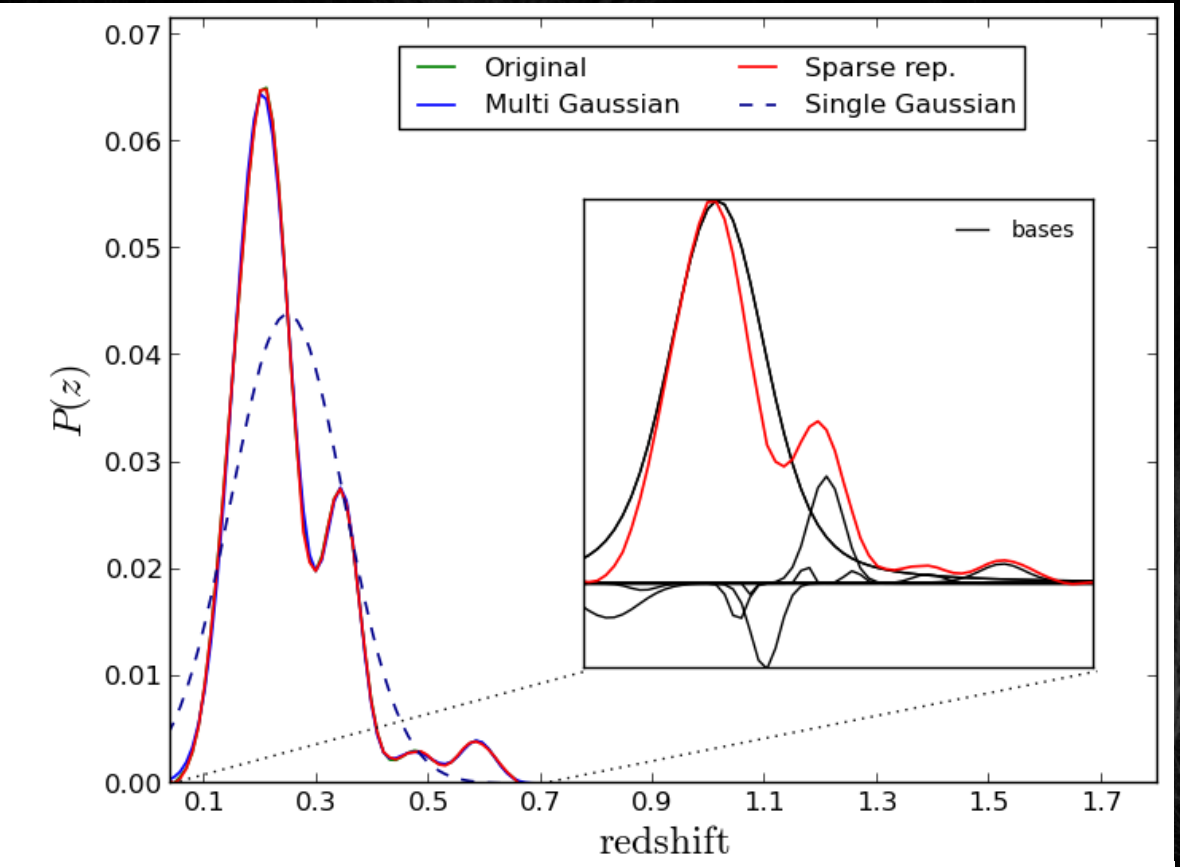
Single Gaussian fit

Multi-Gaussan fit

Monte Carlo sampling

Sparse representation techniques

Carrasco Kind & Brunner 2014b, MNRAS in press, arxiv: 1404.6442



Carrasco Kind & Brunner 2014b (MNRAS, in press.)

Photo-z PDF storage: Sparse representation

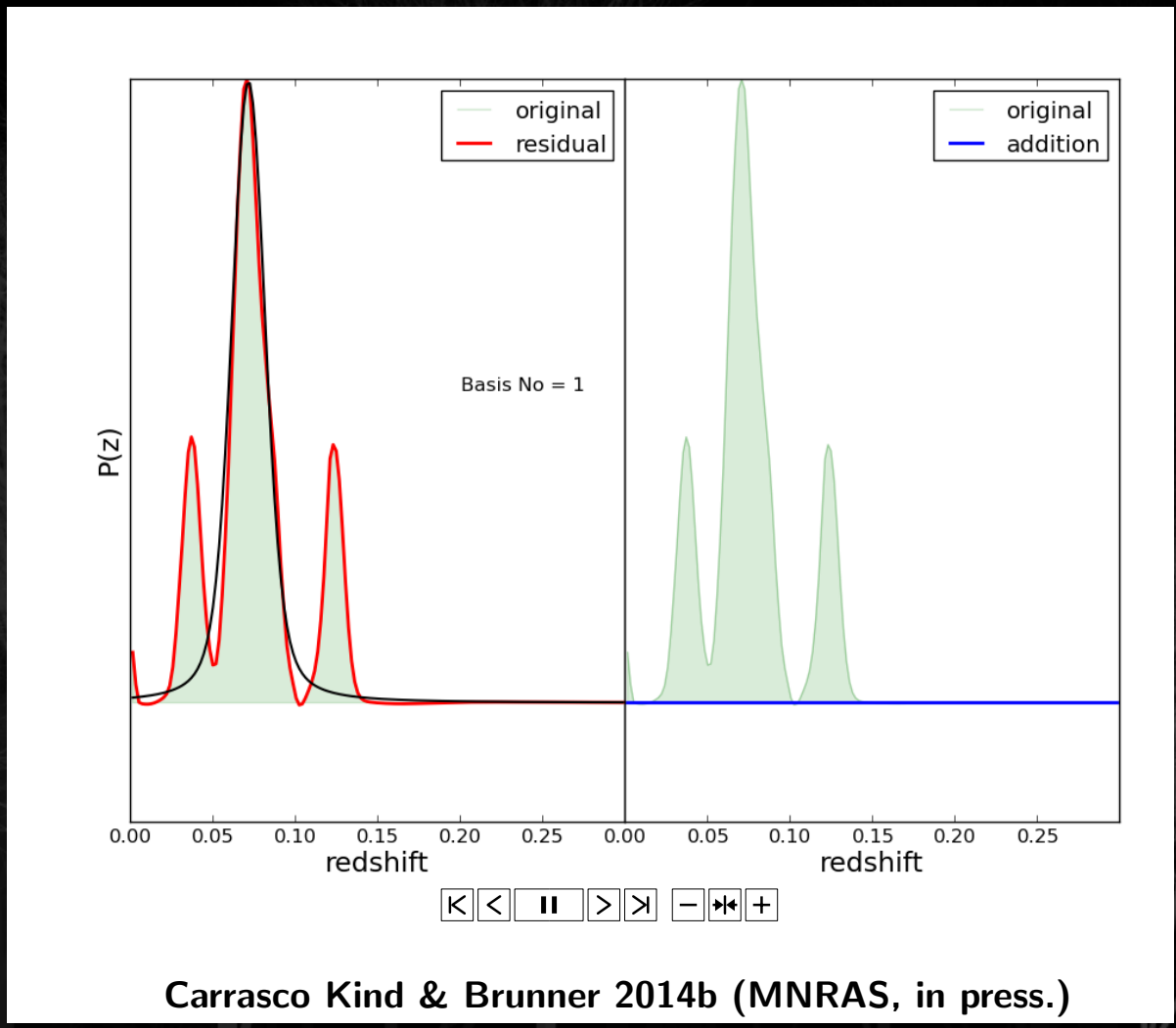
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitud to reduce residual on each step

With only 10-20 bases achieve 99.9 % accuracy

Use 32-bits integer per basis, compression

Store Multiple PDFs



Carrasco Kind & Brunner 2014b (MNRAS, in press.)

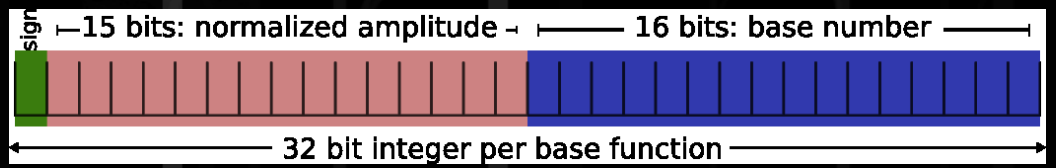
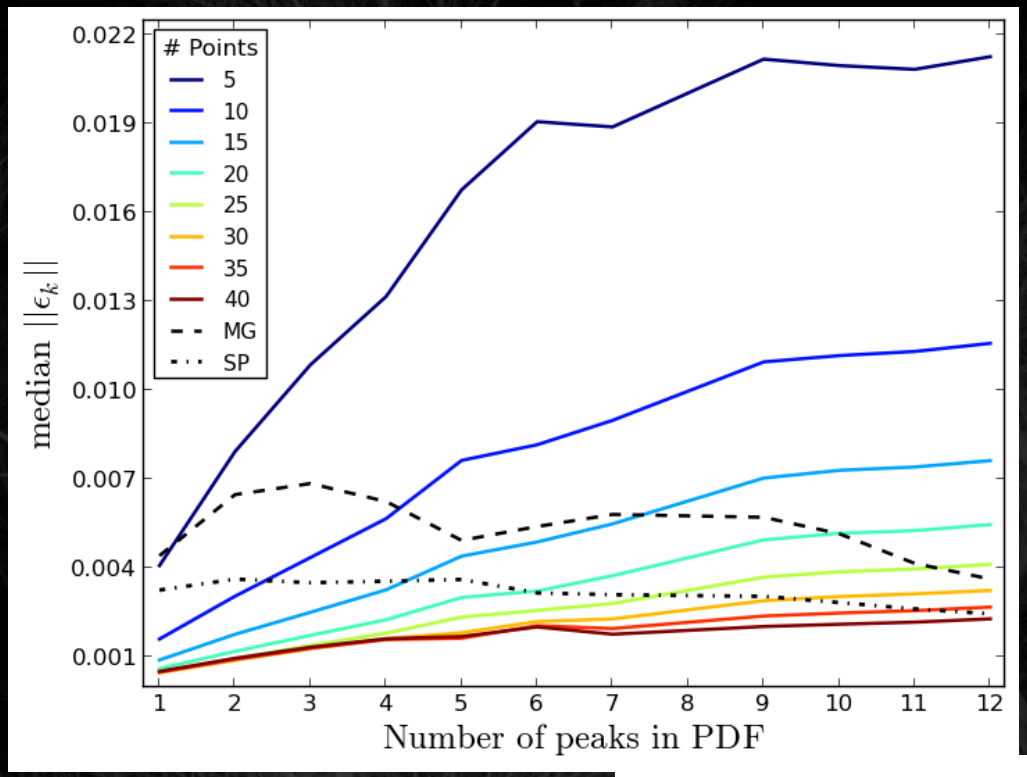
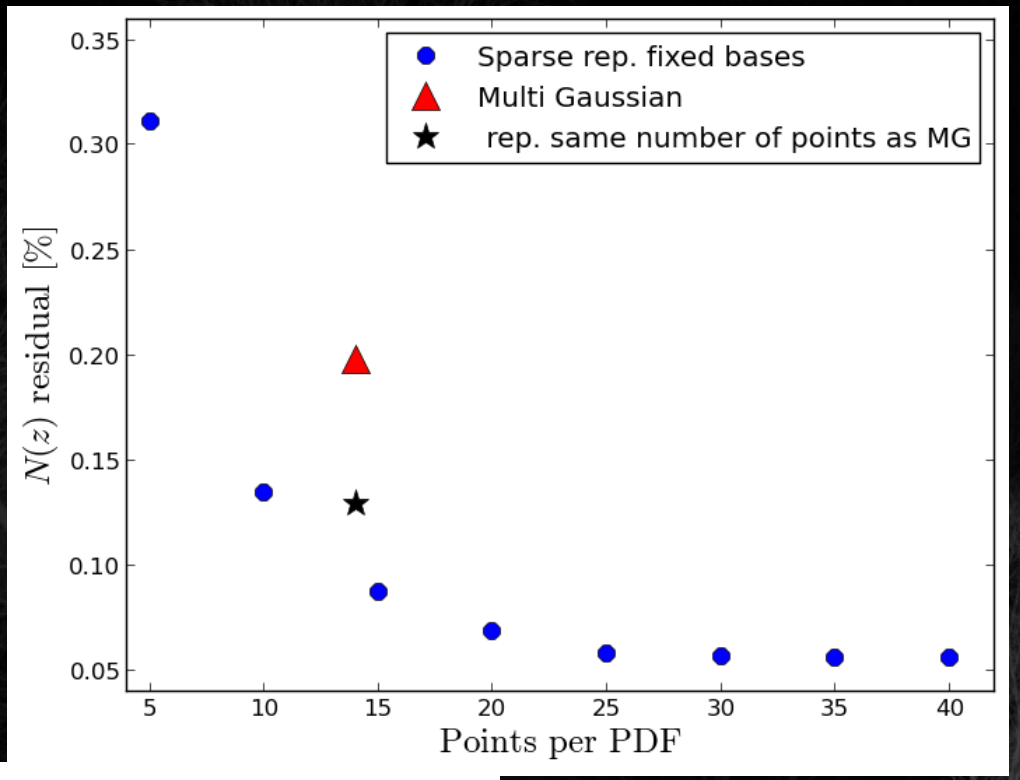


Photo-z PDF storage: Results





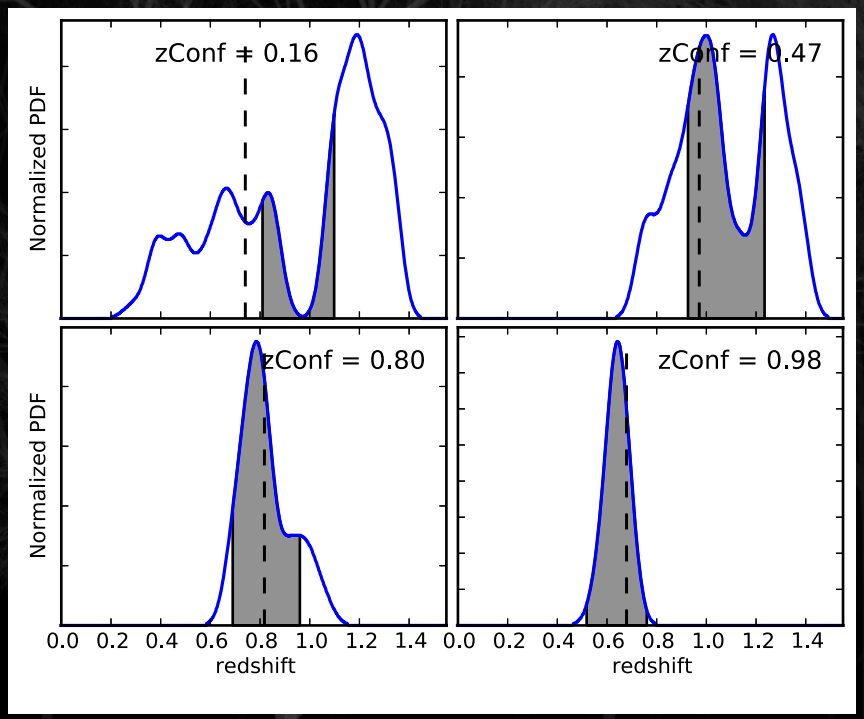
Carrasco Kind & Brunner 2014b (MNRAS, in press.)

For PDFs with less than 4 peaks 5-10 points should be sufficient

Sparse representation gives more accurate and more compressed representation for N(z), 99.9% accuracy with 15 points (200 points originally)

How to produce, combine, store and use P(z)

Photo-z PDF applications



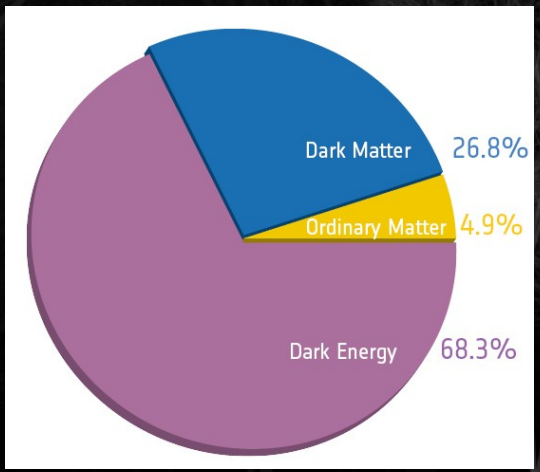


Photo-z PDF application: $N(z)$

$N(z)$ distribution of galaxies, simple yet important feature

Stacked PDF produces better distribution than taken the mean of the PDF

Very important for clustering and weak lensing studies

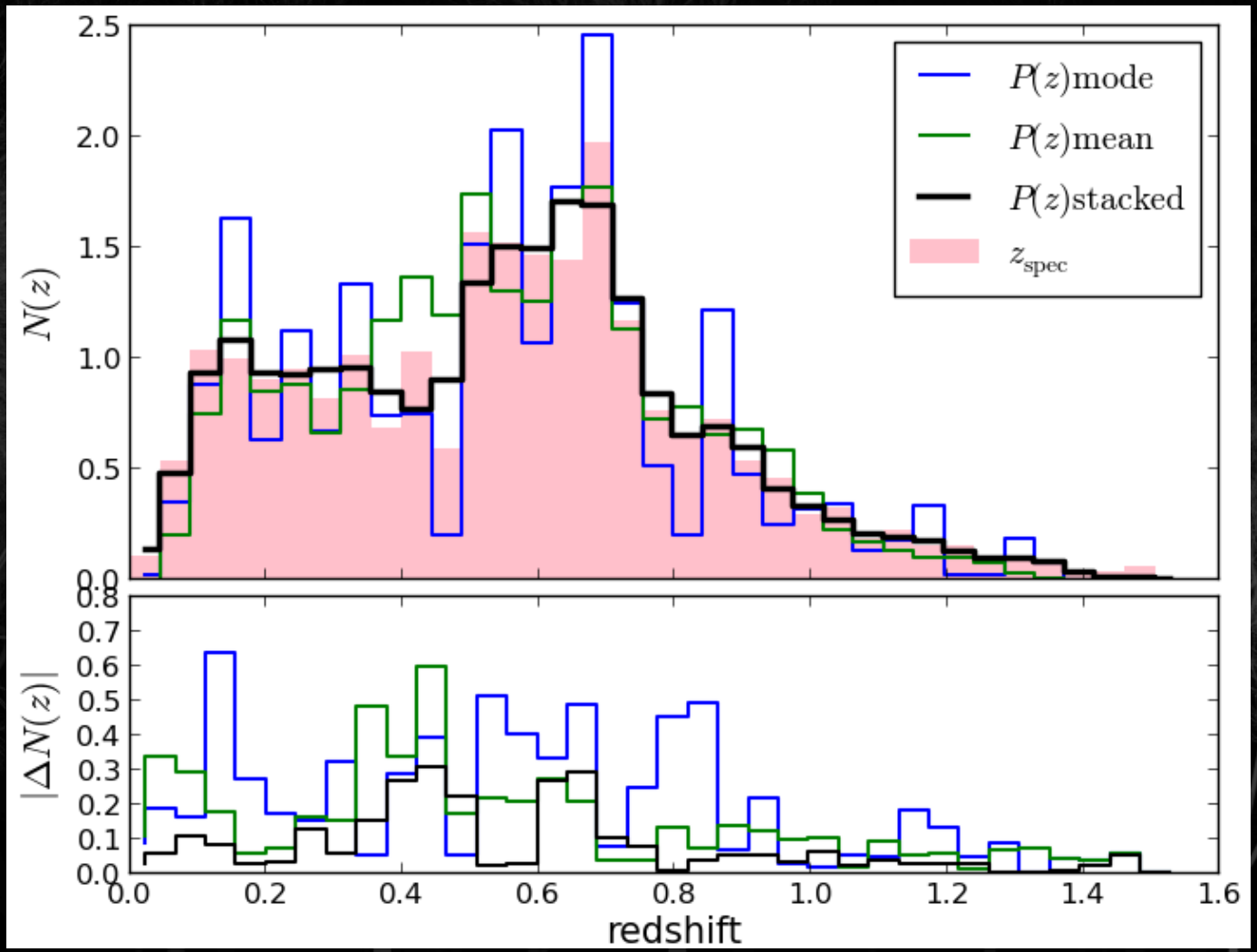


Photo-z PDF application: $N(z)$

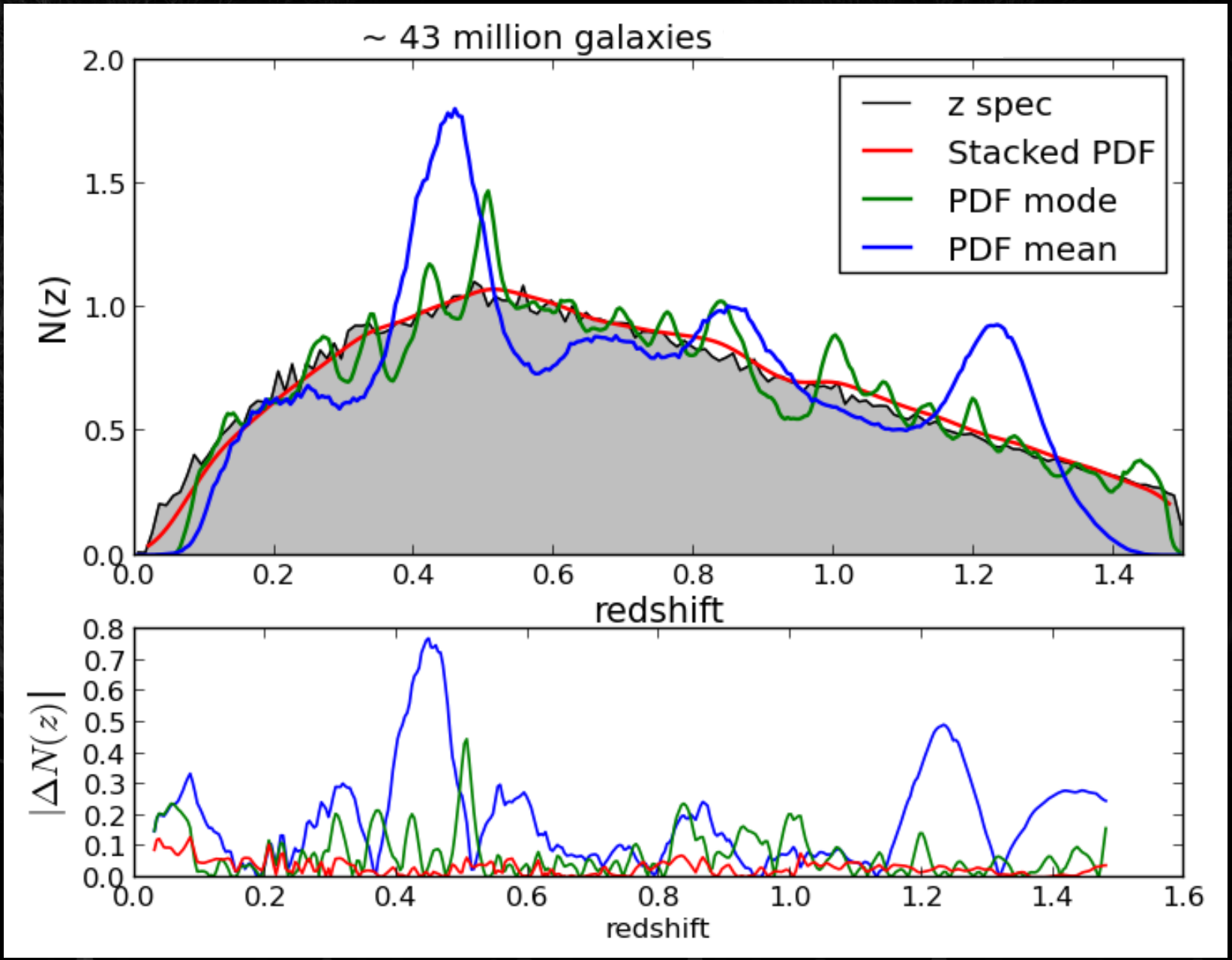


Photo-z PDF application: $N(z)$

By definition:

$$N(z) = \sum_{k=1}^{N} \int_{z-\Delta z/2}^{z+\Delta z/2} P_k(z) dz$$

Photo-z PDF application: $N(z)$

By definition:

$$N(z) = \sum_{k=1}^{N} \int_{z-\Delta z/2}^{z+\Delta z/2} P_k(z) dz$$

Using sparse representation, we represent each PDF pz_k as:

$\mathbf{pz}_k \approx \mathbf{D} \cdot \boldsymbol{\delta}_k$ $\mathbf{D}$ is the dictionary, $\boldsymbol{\delta}_k$ is the sparse vector, then

Photo-z PDF application: $N(z)$

By definition:

$$N(z) = \sum_{k=1}^{N} \int_{z-\Delta z/2}^{z+\Delta z/2} P_k(z) dz$$

Using sparse representation, we represent each PDF pz_k as:

$\mathbf{pz}_k \approx \mathbf{D} \cdot \boldsymbol{\delta}_k$ $\mathbf{D}$ is the dictionary, $\boldsymbol{\delta}_k$ is the sparse vector, then

$$N(z) = \sum_{k=1}^{N} \boldsymbol{\delta}_k \cdot \int_{z-\Delta z/2}^{z+\Delta z/2} \mathbf{D} dz$$

Only bases are integrated

Photo-z PDF application: $N(z)$

By definition:

$$N(z) = \sum_{k=1}^{N} \int_{z-\Delta z/2}^{z+\Delta z/2} P_k(z) dz$$

Using sparse representation, we represent each PDF pz_k as:

$\mathbf{pz}_k \approx \mathbf{D} \cdot \boldsymbol{\delta}_k$ $\mathbf{D}$ is the dictionary, $\boldsymbol{\delta}_k$ is the sparse vector, then

$$N(z) = \sum_{k=1}^{N} \boldsymbol{\delta}_k \cdot \int_{z-\Delta z/2}^{z+\Delta z/2} \mathbf{D} dz$$

Only bases are integrated

by precomputing:

$$\boldsymbol{\delta}_N = \sum_{k=1}^{N} \boldsymbol{\delta}_k \qquad \mathbf{I_D}(z) = \int_{z-\Delta z/2}^{z+\Delta z/2} \mathbf{d}_j dz \qquad j = 1, 2, \ldots, m$$

Photo-z PDF application: $N(z)$

By definition:

$$N(z) = \sum_{k=1}^{N} \int_{z-\Delta z/2}^{z+\Delta z/2} P_k(z) dz$$

Using sparse representation, we represent each PDF pz_k as:

$\mathbf{pz}_k \approx \mathbf{D} \cdot \boldsymbol{\delta}_k$ $\mathbf{D}$ is the dictionary, $\boldsymbol{\delta}_k$ is the sparse vector, then

$$N(z) = \sum_{k=1}^{N} \boldsymbol{\delta}_k \cdot \int_{z-\Delta z/2}^{z+\Delta z/2} \mathbf{D} dz$$

Only bases are integrated

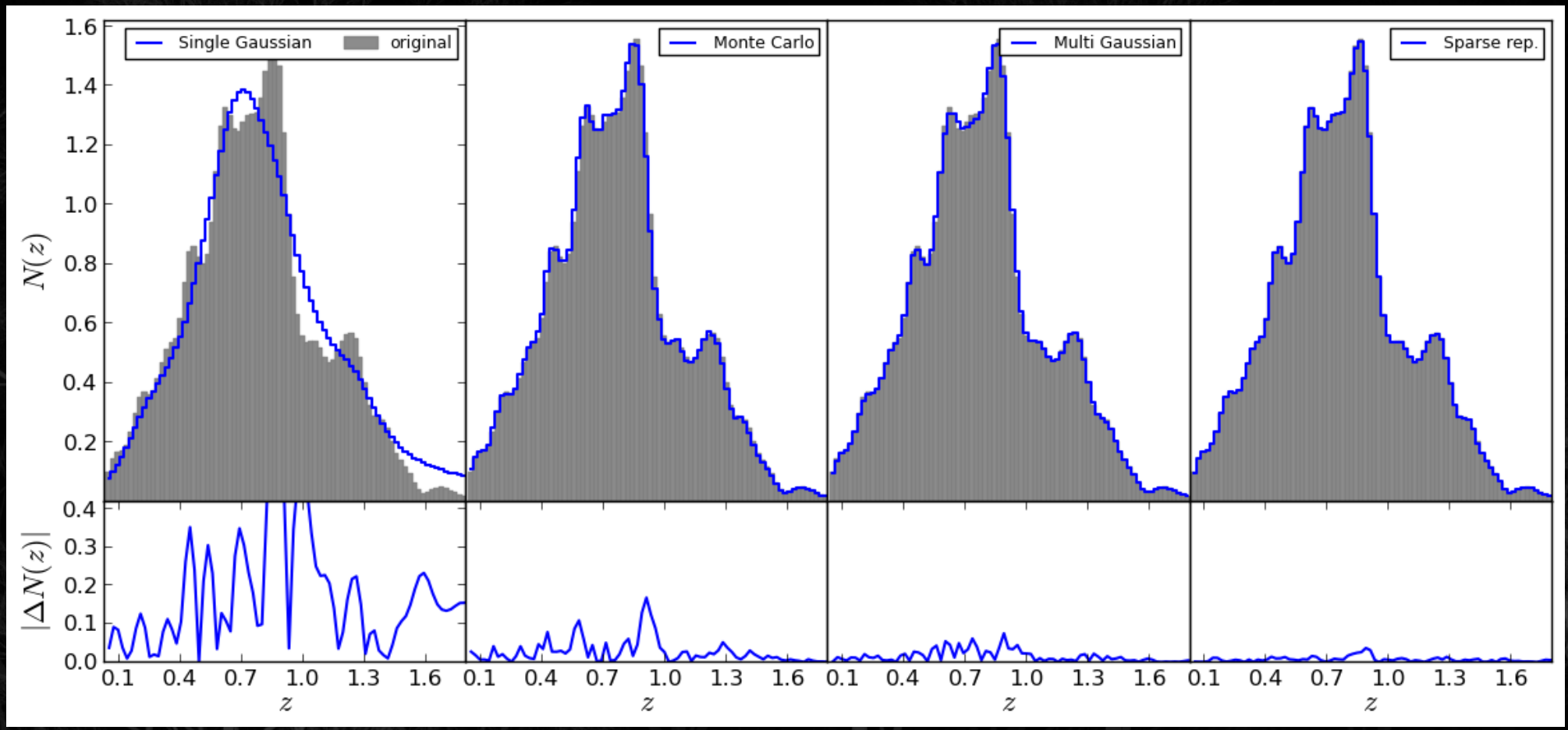
by precomputing:

$$\boldsymbol{\delta}_N = \sum_{k=1}^{N} \boldsymbol{\delta}_k \qquad \mathbf{I_D}(z) = \int_{z-\Delta z/2}^{z+\Delta z/2} \mathbf{d}_j dz \qquad j = 1, 2, \ldots, m$$

$N(z)$ is reduce to a simple dot product

$$\boxed{N(z) = \mathbf{I_D}(z) \cdot \boldsymbol{\delta}_N}$$

Photo-z PDF application: $N(z)$



$N(z)$ original (gray) compared to 4 PDF representation methods, Single Gaussian, MonteCarlo, Multi Gaussian, Sparse rep.

Photo-z PDF application: Angular Power Spectrum

- The angular power spectrum (APS) contains important information about the matter density field
- 2D projection of $P(k)$ using $N(z)$in the kernel
- Constrains cosmological models. Could be used to resolve BAOs
- Use photo-z PDF in overdensities

$$\delta_i = \frac{\Omega_{survey} \sum_j^{N_{in}} \int_{z_1}^{z_2} P_{ij}(z) dz}{\Omega_i \sum_j^{N_{tot}} \int_{z_1}^{z_2} P_j(z) dz} - 1$$

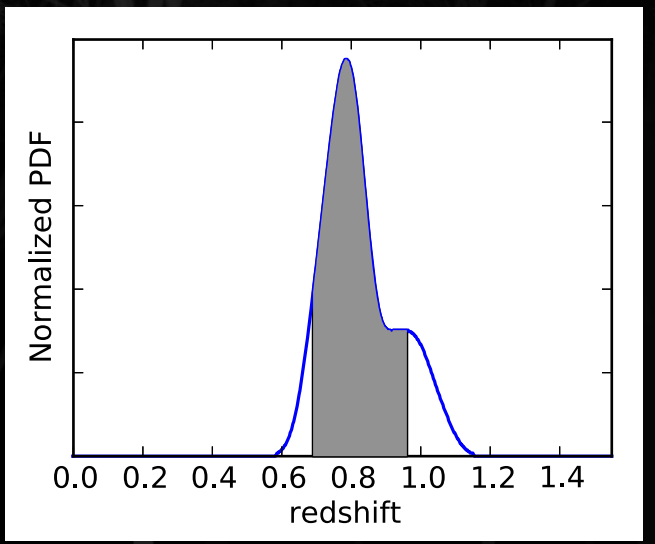


Photo-z PDF application: C_ℓ and $\omega(\theta)$

Limber approximation with no redshift-space distortions and scale-independent bias b:

$$C_\ell = \frac{\ell(\ell+1)}{2\pi} b^2 \int dz \phi^2(z) \frac{H(z)}{r^2(z)} P\left(\frac{\ell+1/2}{r(z)}, z\right)$$

CAMB and HALOFIT for non linear $P(k,z)$

$\phi(z)$ is the galaxy distribution $N(z)$

Fitting using Monte Carlo Markov Chain methods

$$\chi^2(a_p) = \sum_{bb'} (\ln \mathcal{C}_b - \ln \mathcal{C}_b^T) \mathcal{C}_b F_{bb'} \mathcal{C}_{b'} (\ln \mathcal{C}_{b'} - \ln \mathcal{C}_{b'}^T)$$

Photo-z PDF application: C_ℓ and $\omega(\theta)$

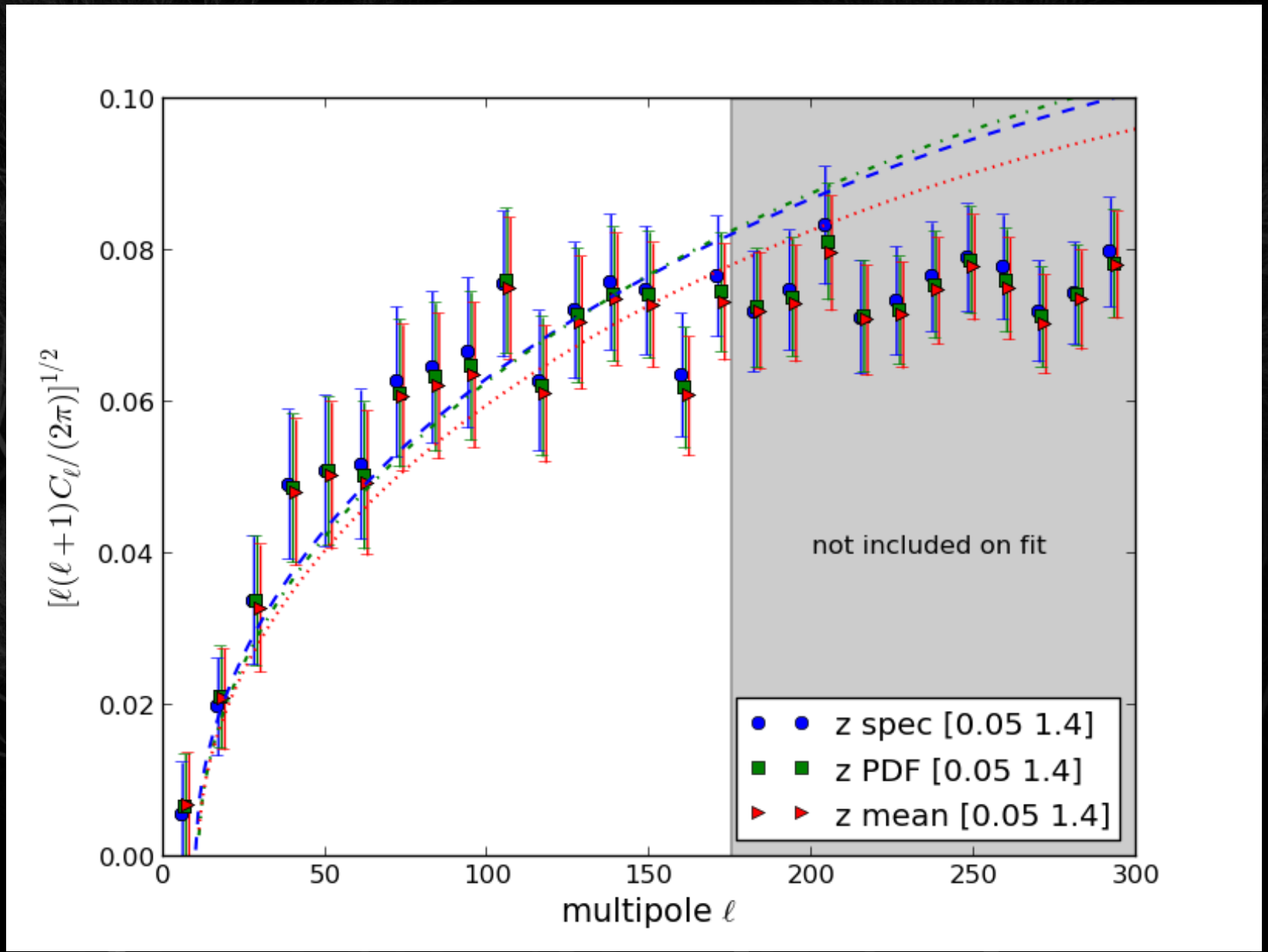
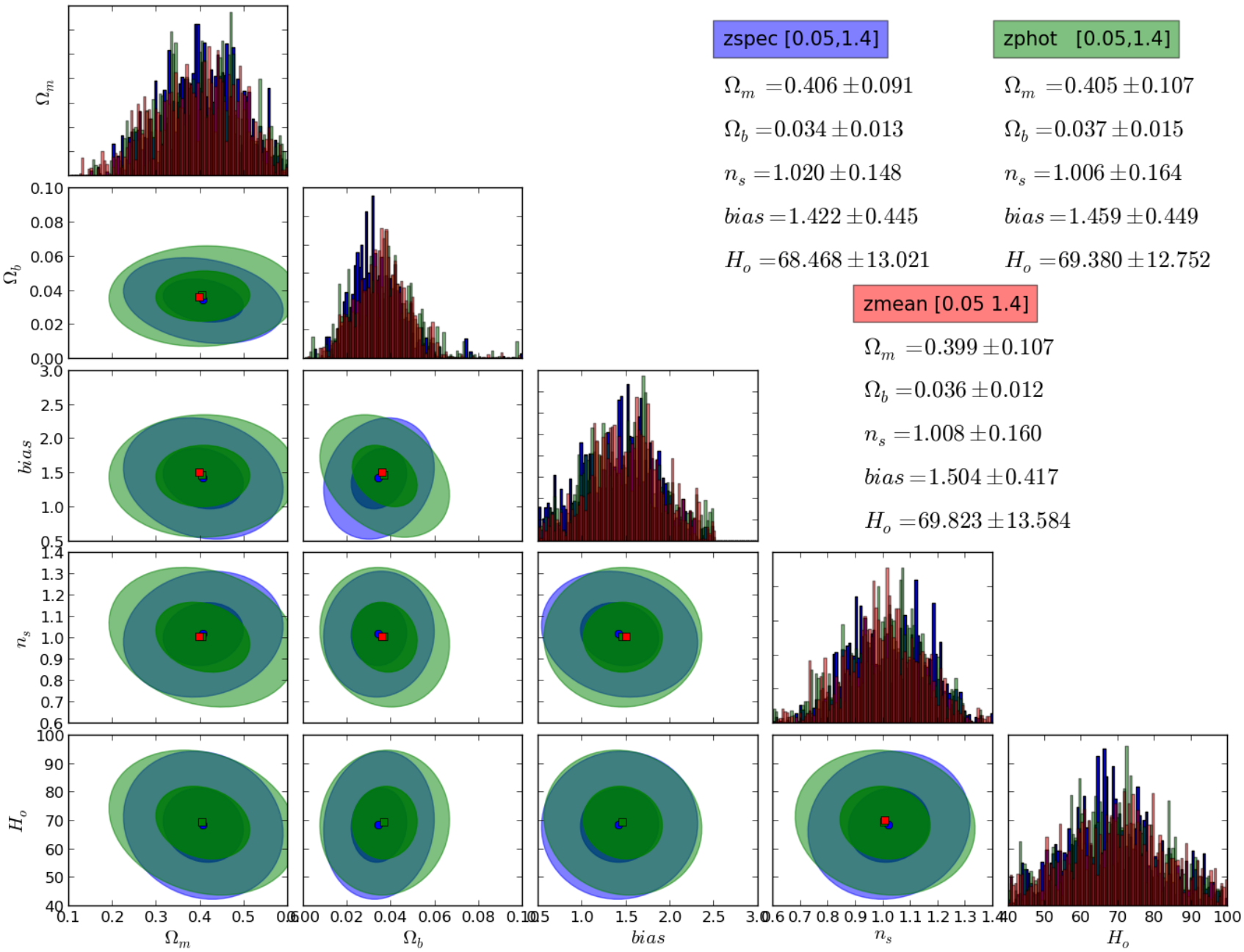


Photo-z PDF application: Fiting cosmologies



Conclusions

* Individual techniques: good information, MLZ framework to compute photo-z PDFs using Machine Learning

* Combination technique: more and better information for computing photo-z PDFs and outlier identification

* Sparse representation reduce the storage by a order of magnitude with 99.9% accuracy

* Sparse representation can be incorporate in theoretical framework to reduce computational time

* Photo-z PDF in cosmological analysis to enhance signals

THANKS!

Questions?

Matias Carrasco Kind
University of Illinois
mcarras2@illinois.edu
https://sites.google.com/site/mgckind/

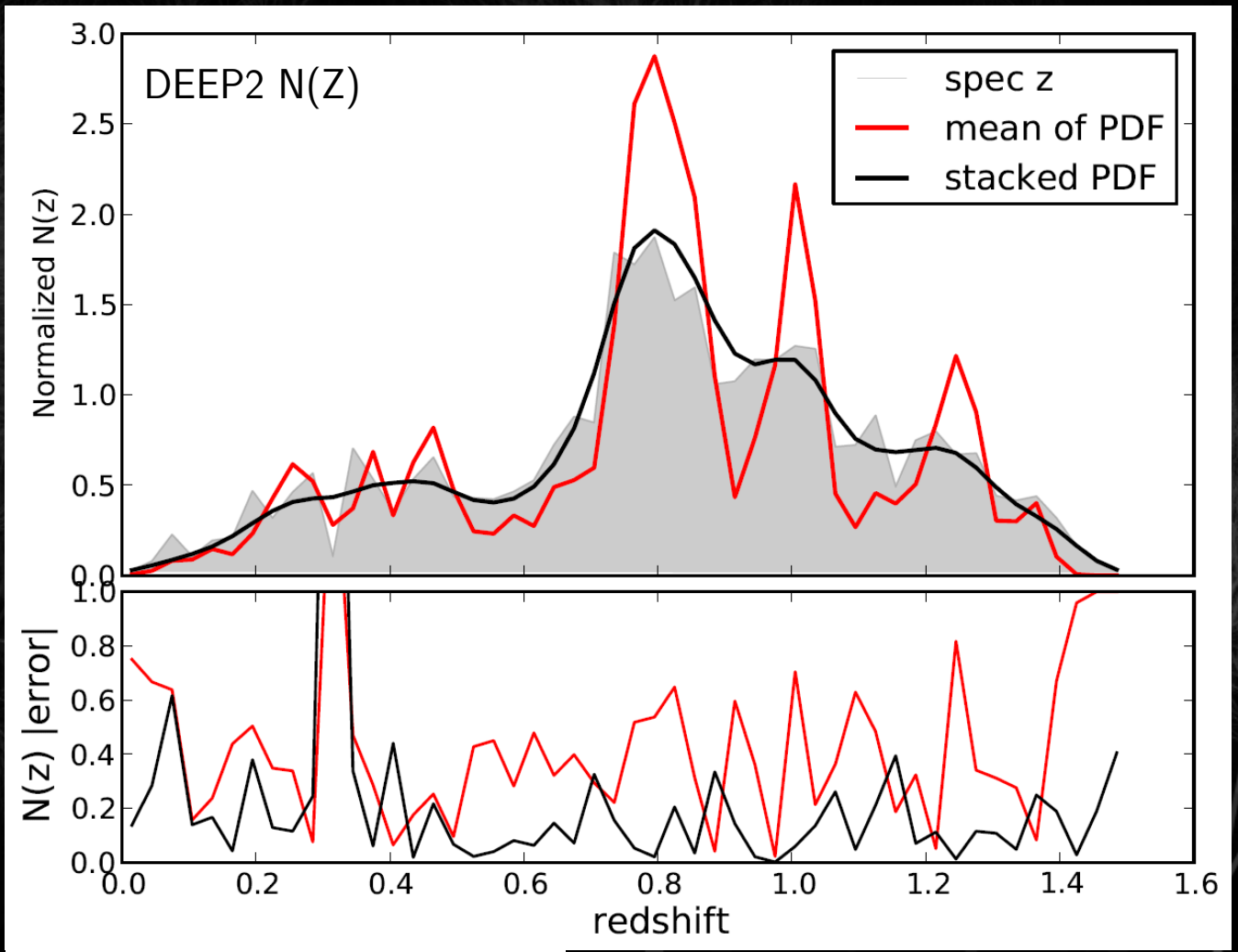
EXTRA SLIDES

Using photo-z PDF in cosmological anaysis

$N(z)$ distribution of galaxies, simple yet important feature

Stacked PDF produces better distribution than taken the mean of the PDF

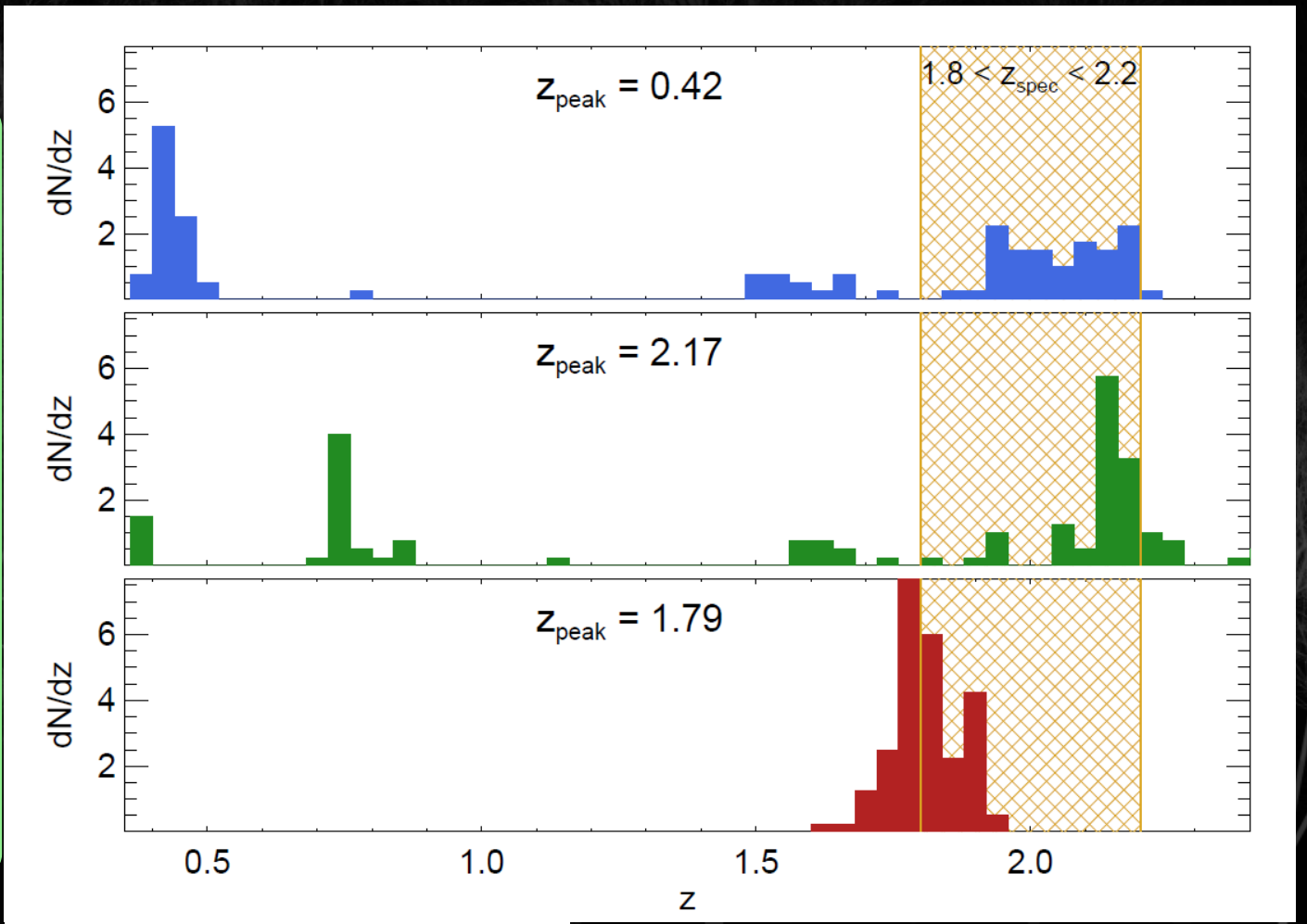
Very important for clustering and weak lensing studies



Carrasco Kind & Brunner 2013a

Example application of photo-z PDF

- Incorporating PDF on clustering measurements
- Problems of using mode of photo-z PDF
- Extend to other measurements



Myers, White & Ball 2009

Photometric redshift PDFs using TPZ

We use TPZ to generate photo-z for all galaxies.

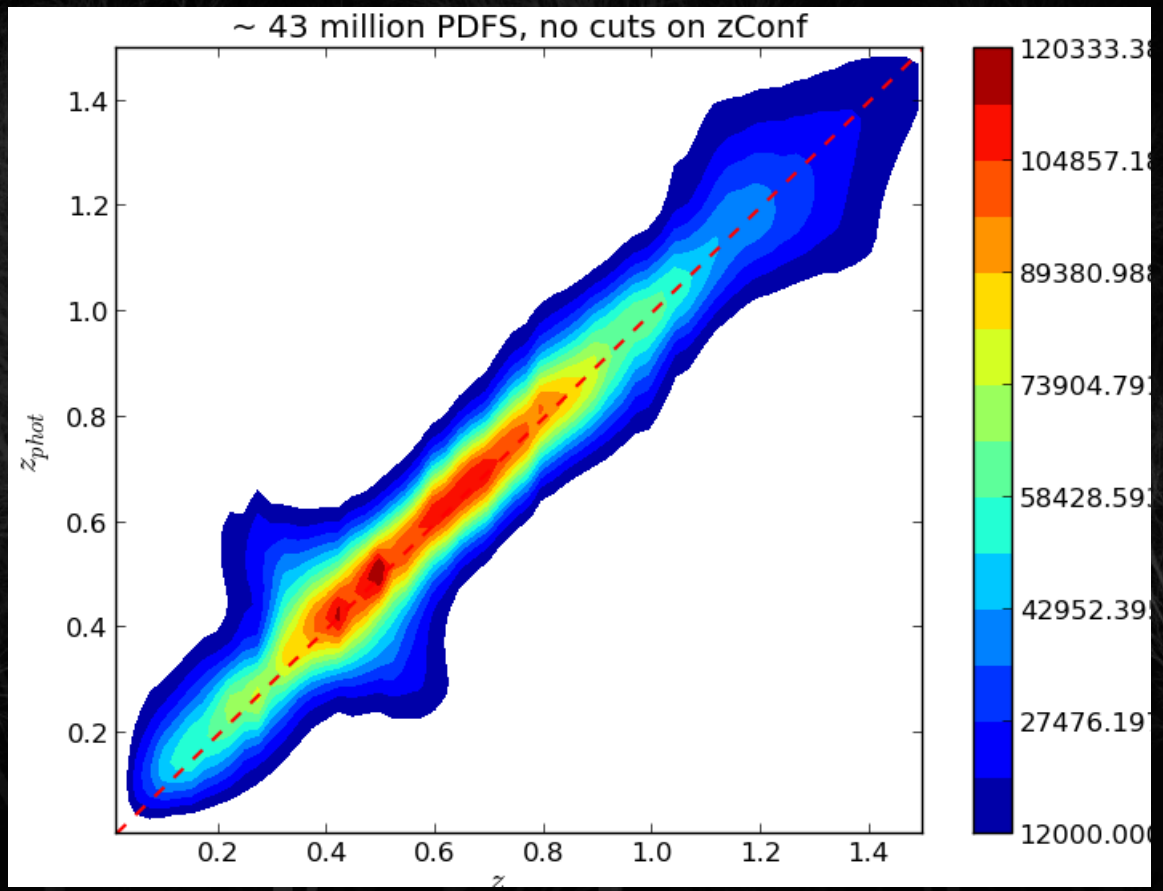
100,00 for training

5 magnitudes only

$\sim$ 0.17 sec per PDF

Store 43 million PDFs for analysis

No outlier removal



Photometric redshift PDFs using TPZ

Metrics

$(\Delta z = z_{phot} - z_{spec})$

$< \Delta z >= 0.0088$

$< |\Delta z| >= 0.089$

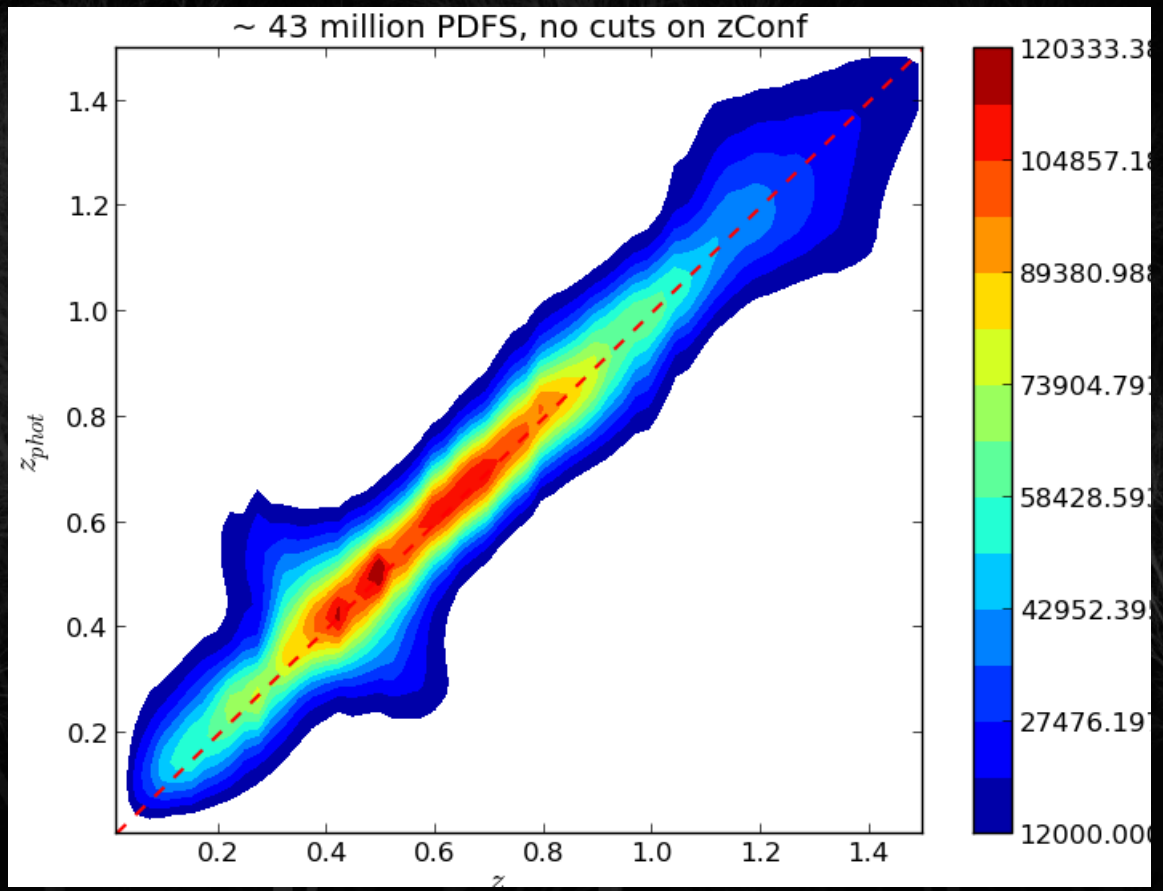
$\sigma_{\Delta z} = 0.1421$

$\sigma_{|\Delta z|} = 0.1109$

$\sigma_{68} = 0.0885$

$frac > 2\sigma = 0.0531$

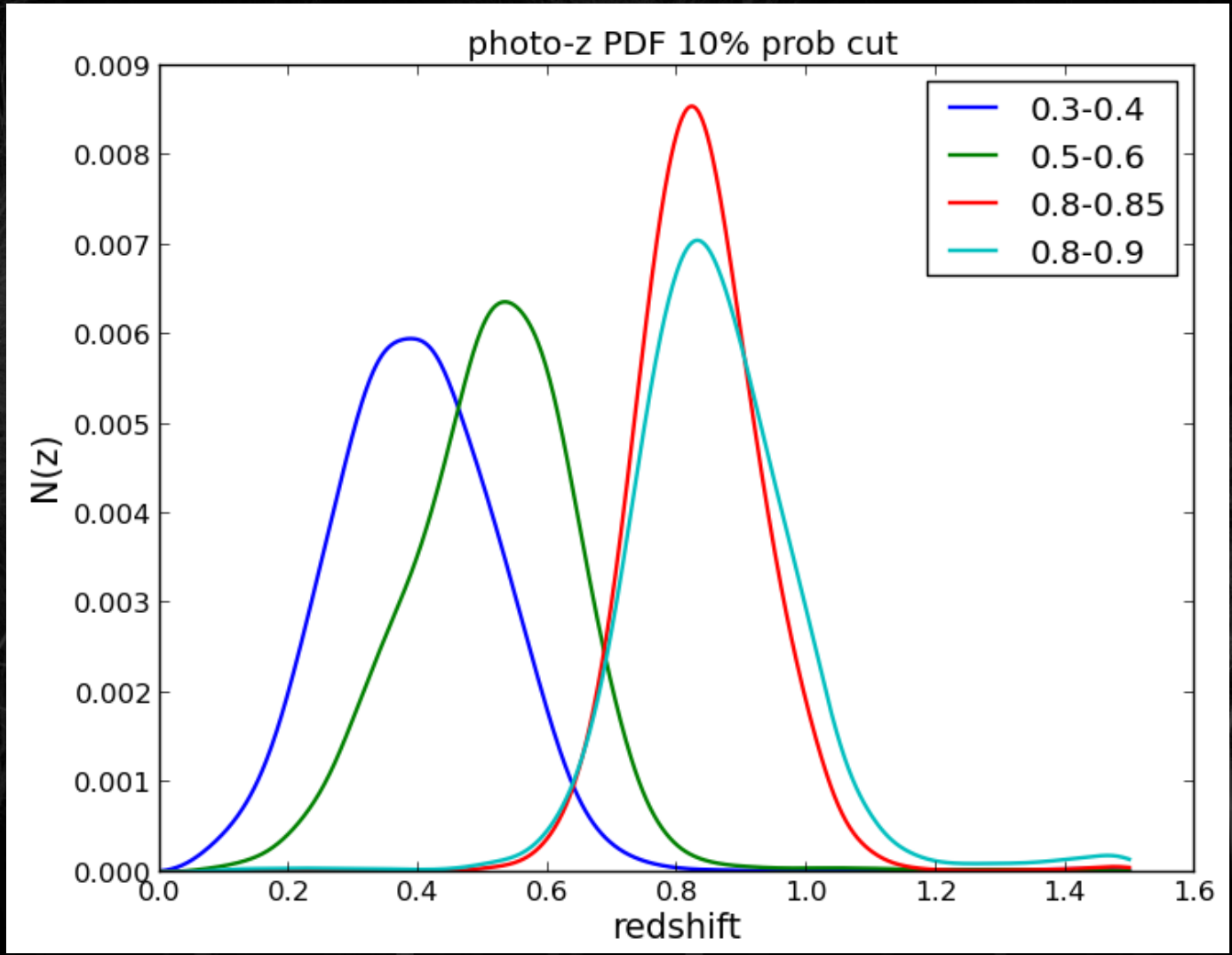
$frac > 3\sigma = 0.0207$



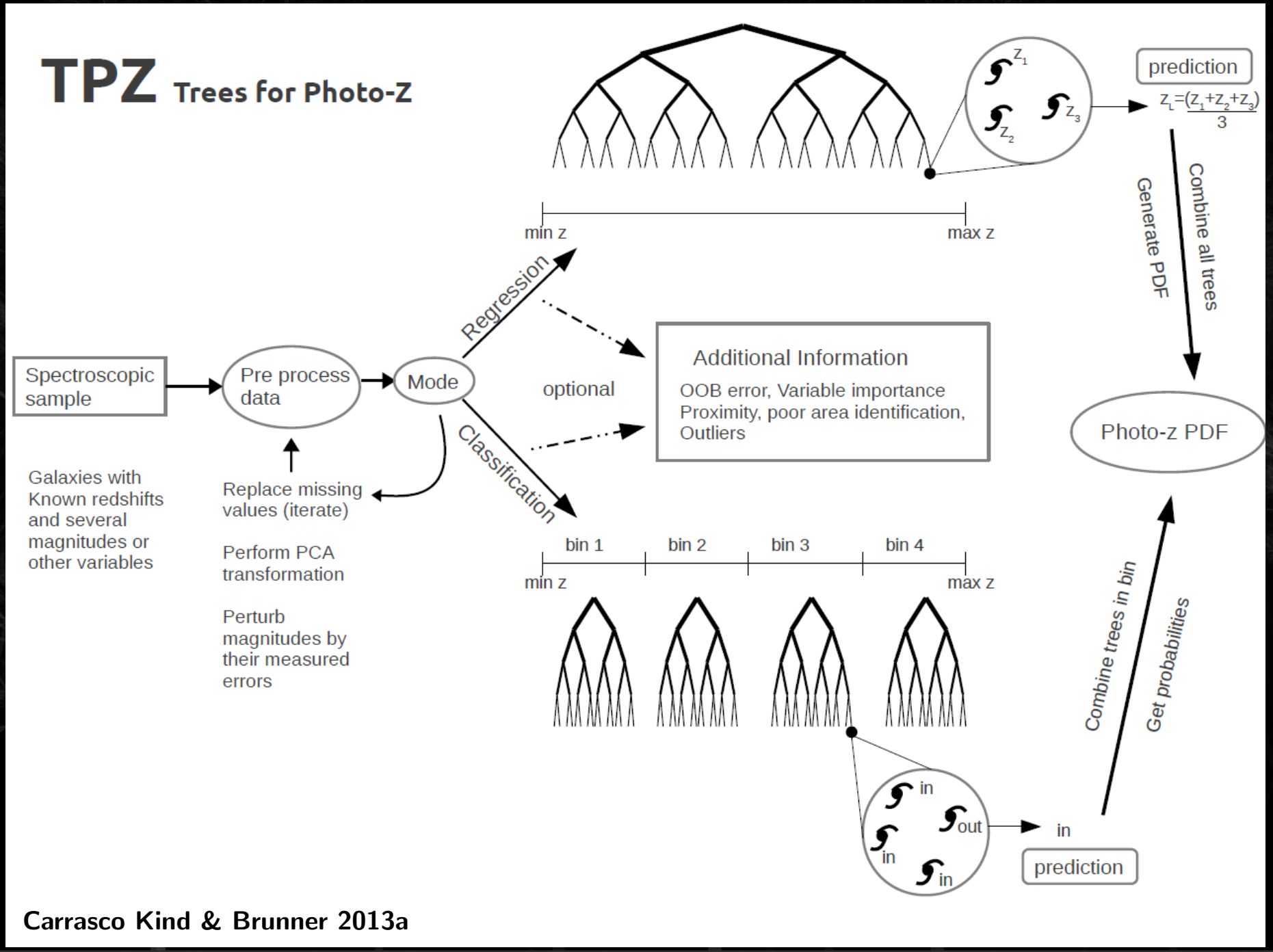
Also in redshift shells

We consider only PDF with at least 10% of its area inside redshift shell

N(z) and overdensities from stacked PDFs

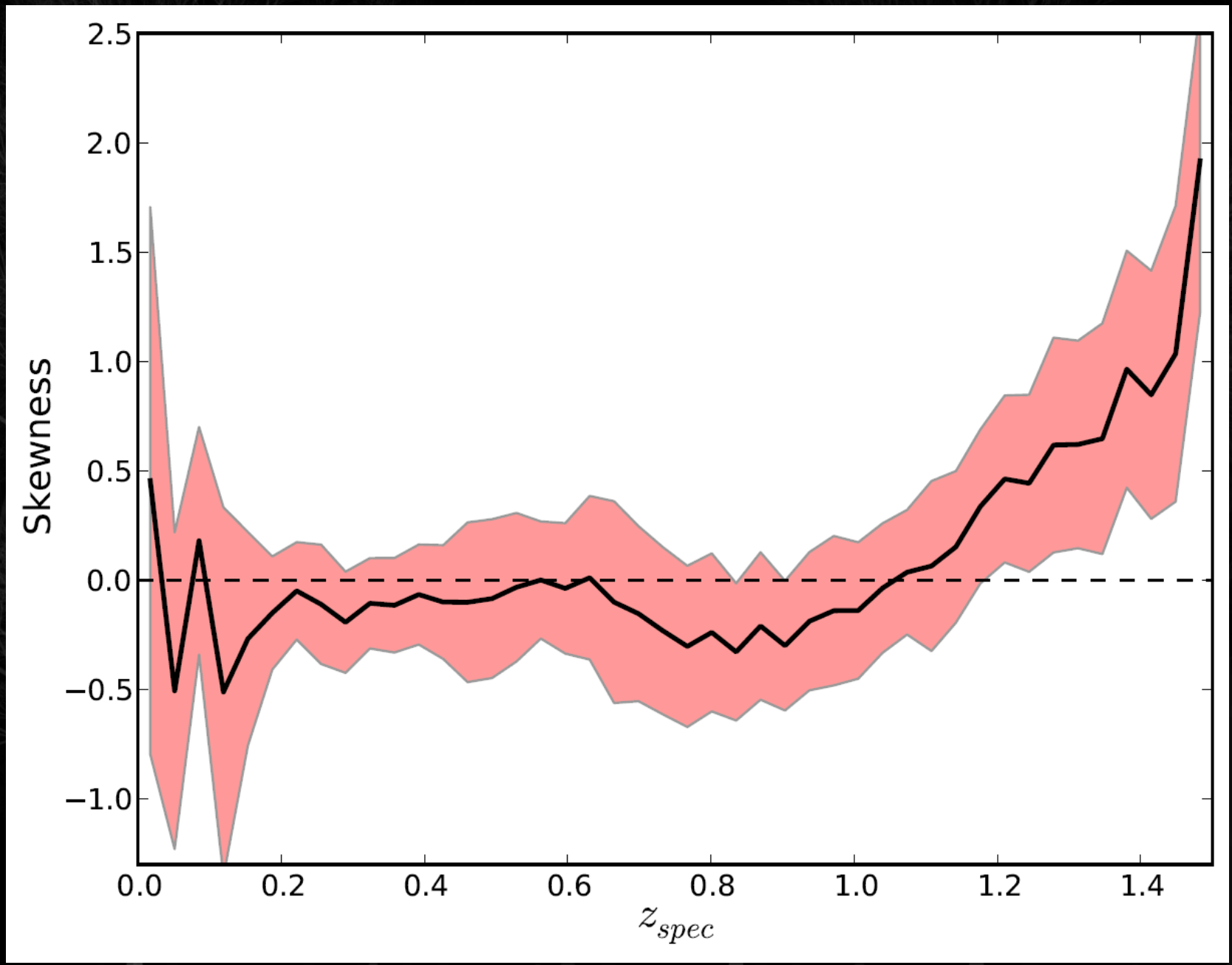


TPZ : Scheme



Carrasco Kind & Brunner 2013a

Skewness of DEEP2 PDFs



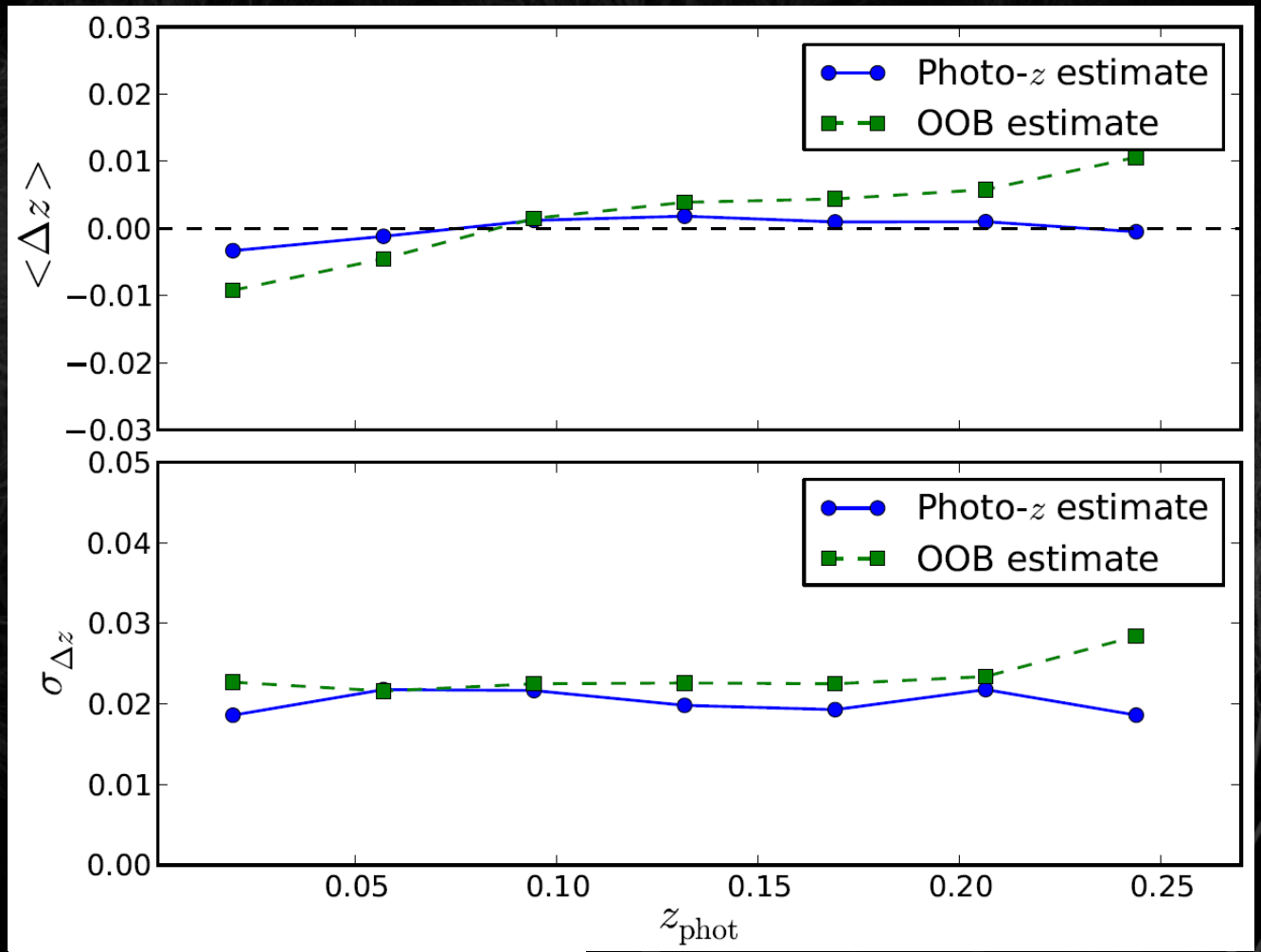
TPZ: Ancillary information - *prior error* -

Using *Out-of-Bag* data TPZ provides useful extra information

No need of a validation set, use full training set.

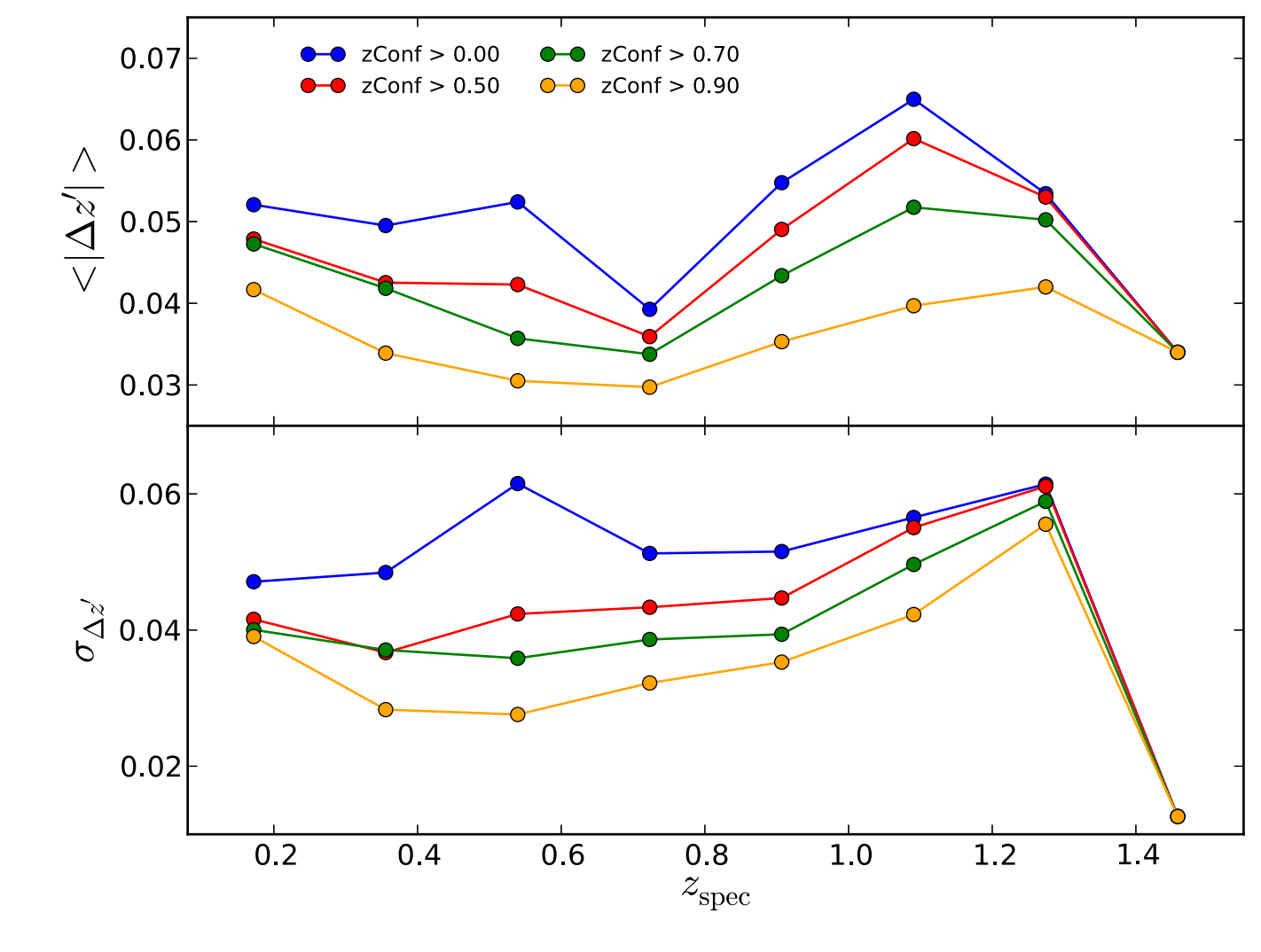
Example application on SDSS MGS, 40,000 test and 15,000 training galaxies

A prior unbiased estimations of errors!



Carrasco Kind & Brunner 2013a

Metrics vs. zConf

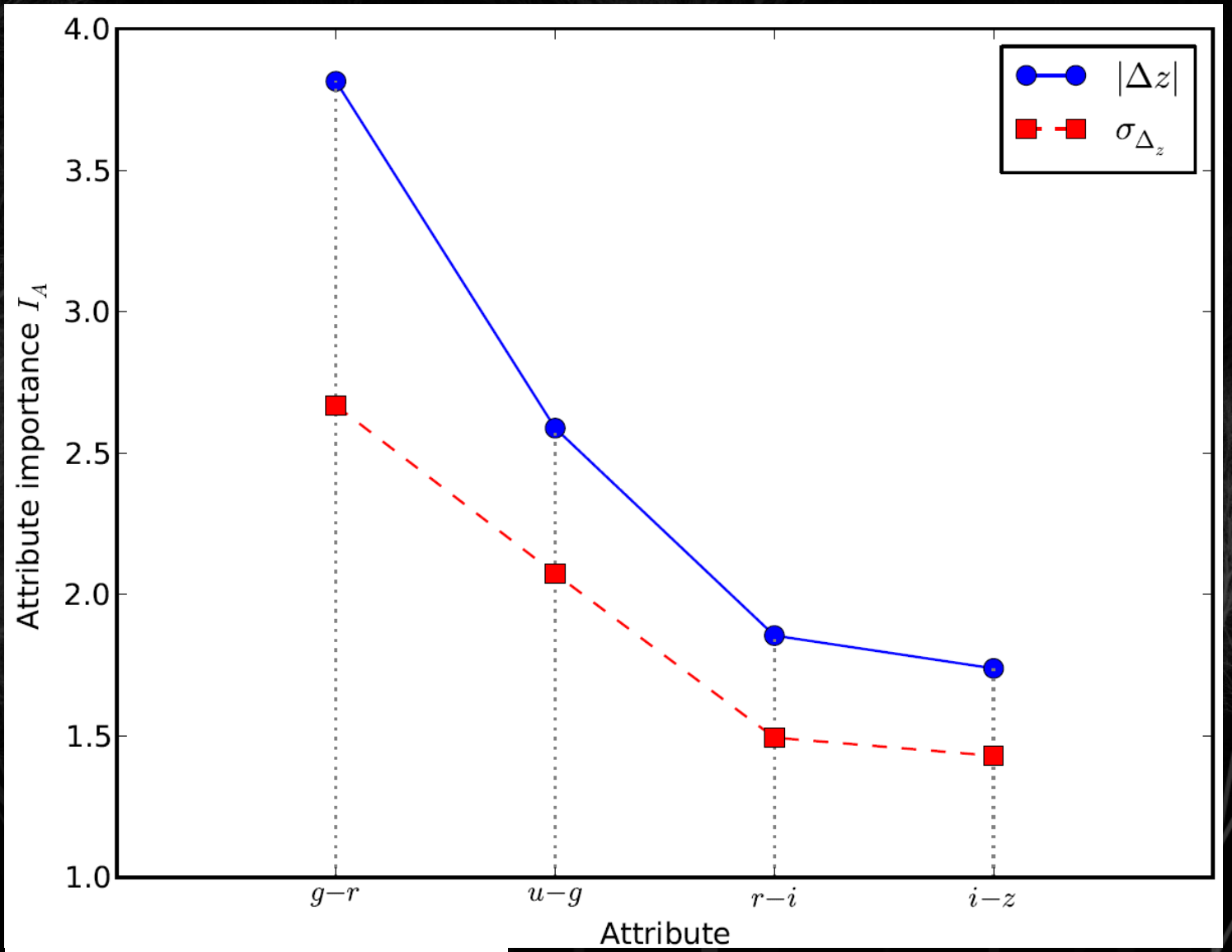


TPZ: Ancillary information - *Attribute importance* -

Ranking statistical only

Useful for removing unimportant variables reducing the noise

Most important attributes to construct impotance map

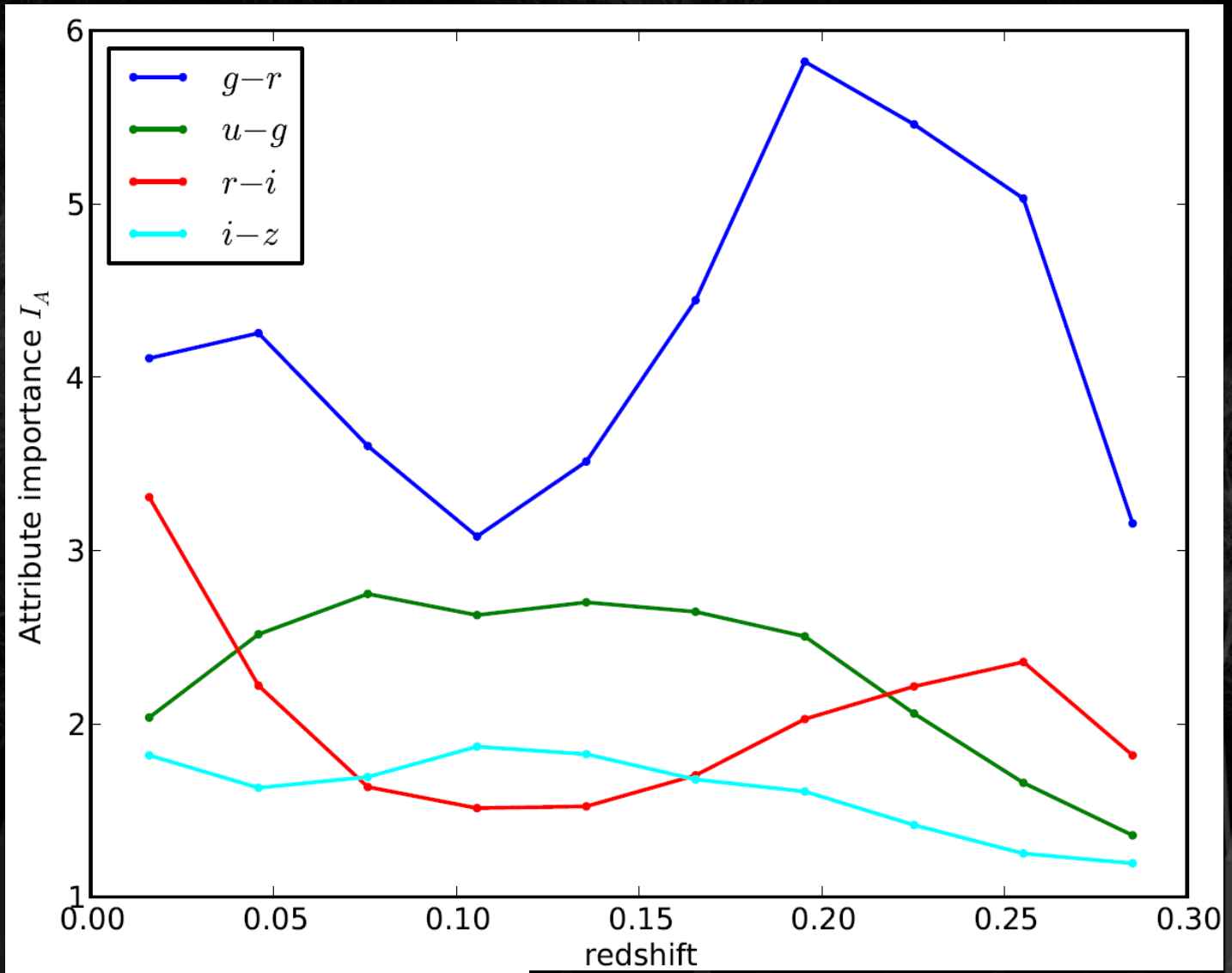


Carrasco Kind & Brunner 2013a

TPZ: Ancillary information - *Attribute importance* -

How much the metrics change as we permute the attributes one at a time

For SDSS the $g-r$ color is the most important attribute



Carrasco Kind & Brunner 2013a

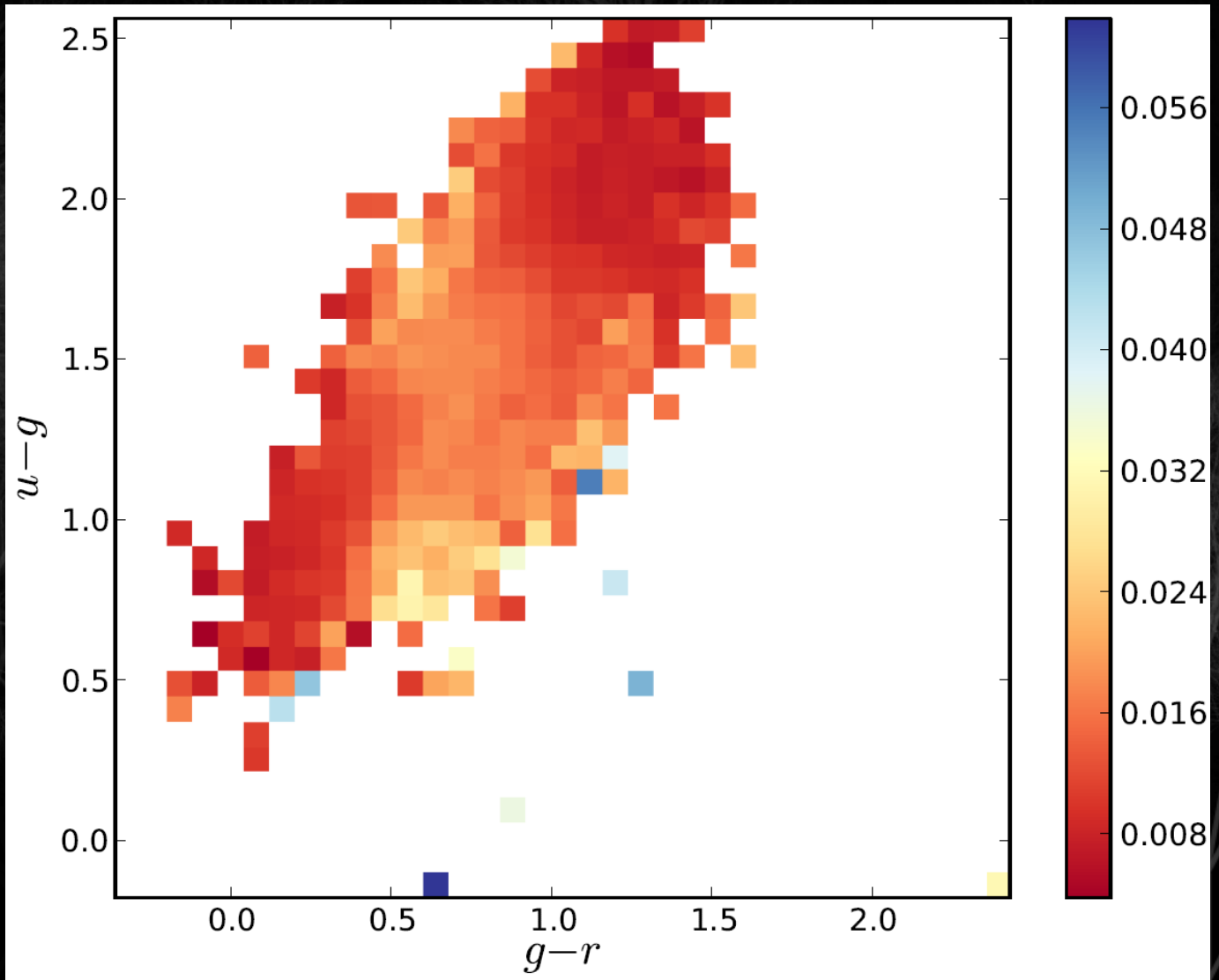
TPZ: Ancillary information - *Poor area identification* -

Map of performance using two most important colors

The redder the better

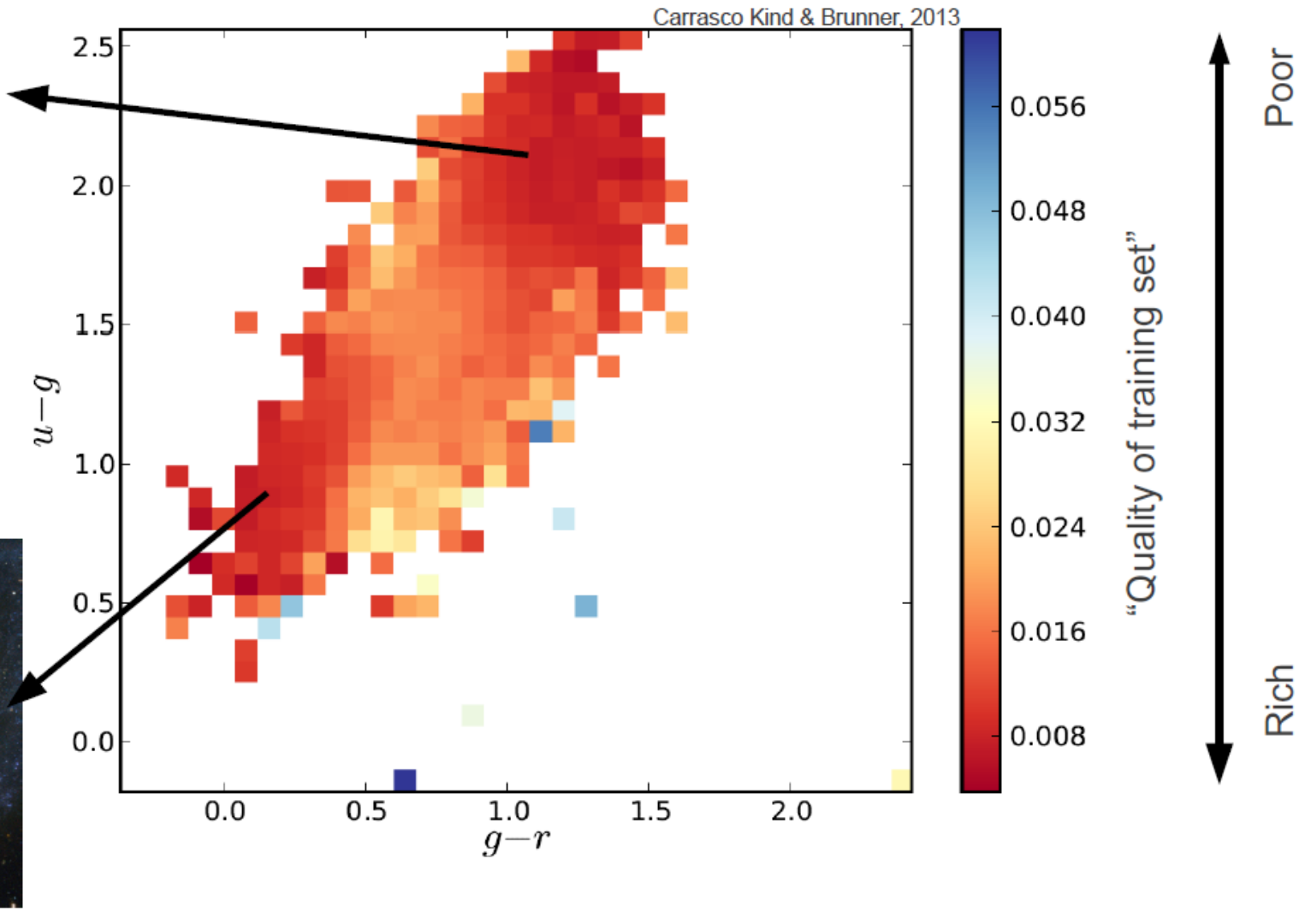
Bimodality of SDSS galaxies

Narrow follow up observations



Carrasco Kind & Brunner 2013a

TPZ: Ancillary information - *Poor area identification* -



Without any prior knowledge machine learning techniques provide physical insights to astronomical processes

Photo-z PDF combination: BMA

$P(z)$ given by: $$P(z \mid \mathbf{x}, \mathbf{D}, \mathbf{M}) = \sum_k P(z \mid \mathbf{x}, M_k) P(M_k \mid \mathbf{D})$$

"weight"

$$P(M_k \mid \mathbf{D}) = \frac{P(M_k)}{P(\mathbf{D})} P(\mathbf{D} \mid M_k) \propto P(M_k) \prod_{i=1}^{N_d} P(d_i \mid M_k)$$

d_i: training data

We define:

$$N_{k,i}^{(b)} = \begin{cases} 1 & \text{if } \int_{z_s-\delta_z}^{z_s+\delta_z} P(z \mid \mathbf{x}, d_i) dz \leq \pi_z, \\ 0 & \text{otherwise.} \end{cases}$$

then:

$$P(M_k \mid \mathbf{D}) \propto P(M_k)(1-\epsilon_k)^{N_d - N_k^{(b)}} (\epsilon_k)^{N_k^{(b)}}$$

and finally:

$$P(z \mid \mathbf{x}, \mathbf{D}, \mathbf{M}) \propto \sum_k P(z \mid \mathbf{x}, M_k) P(M_k) \times (1-\epsilon_k)^{N_d - N_k^{(b)}} (\epsilon_k)^{N_k^{(b)}}$$

Photo-z PDF combination: BMC

Similarly to BMA, instead of selecting from models, we select from combinated models (>100), we have $P(e \mid \mathbf{D})$ instead of $P(M_k \mid \mathbf{D})$

$$P(e \mid \mathbf{D}) \propto P(e) \prod_{i=1}^{N_d} P(d_i \mid e) \qquad \text{then, } P(z) :$$

$$P(z \mid \mathbf{x}, \mathbf{D}, \mathbf{M}, \mathbf{E}) = \sum_{e \in \mathbf{E}} P(z \mid \mathbf{x}, \mathbf{M}, e) P(e \mid \mathbf{D})$$

We generate models e in set $\mathbf{E}$ by a Dirichlet process:

$$P(\mathbf{w}) \sim \mathcal{D}\mathrm{ir}(\boldsymbol{\alpha}) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k w_k^{\alpha_k - 1}$$

every few steps we update $\boldsymbol{\alpha}$

$$\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^t + \max_{\mathbf{w}_e \in n_s} P(e \mid \mathbf{D})$$

We procedure as BMA to select best combinaiton

Photo-z PDF combination: HB

$$P(z \mid \mathbf{x}, \mathbf{D}, M_k, \theta_k) = \sum_j P(z \mid \mathbf{x}, \mathbf{D}, M_k, \theta_{kj}) \times P(\theta_{kj} \mid \mathbf{D}, M_k)$$

$$\sum_j P(\theta_{kj} \mid \mathbf{D}, M_k) = 1$$

$$P(z \mid \mathbf{x}, \mathbf{D}, M_k, \theta_k) = P_{def}(z \mid M_k, \theta_k)\gamma_k + P(z \mid \mathbf{x}, \mathbf{D}, M_k, \theta_k)(1 - \gamma_k)$$

$$P(z \mid \mathbf{x}, \mathbf{D}, \theta) = \prod_k P(z \mid \mathbf{x}, \mathbf{D}, M_k, \theta_k)^{1/\beta}$$

$$P(z) = \int_0^1 P(z \mid \mathbf{x}, \mathbf{D}, \theta) P(\theta) d\theta$$