ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

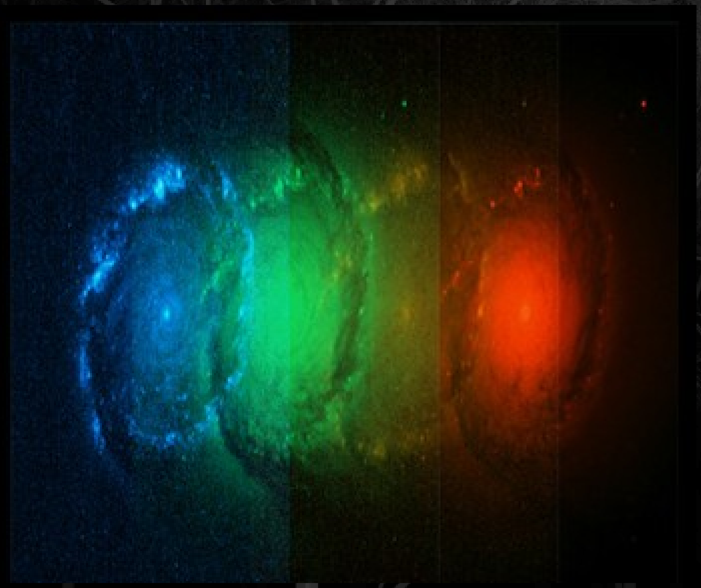
Notes on combination and storage of P(z)

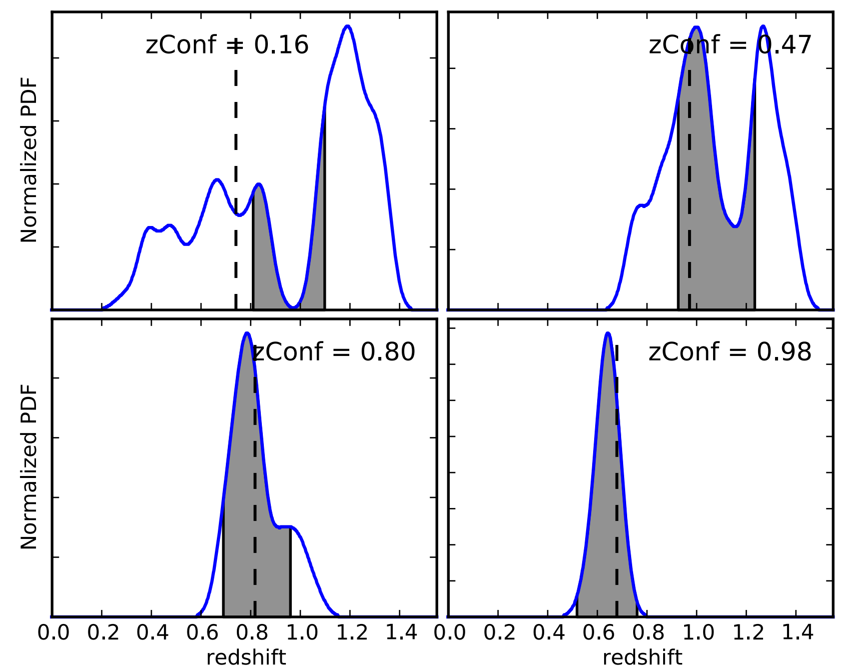
Matías Carrasco Kind

NCSA/Department of Astronomy
University of Illinois at Urbana-Champaign

LSST Photo-z Workshop @ Pittsburgh
April 5$^{\mathrm{th}}$ - 7$^{\mathrm{th}}$, 2016

Photo-z PDF estimation





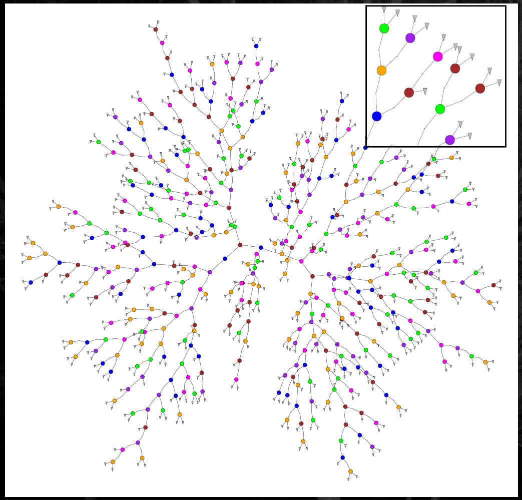
Carrasco Kind & Brunner 2013a (MNRAS, 432, 1483)

Photo-z PDF combination: Bayesian framework

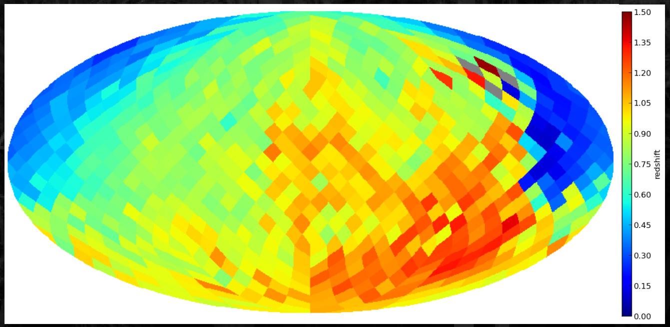
Motivation

- Different methods (> 20) have different strengths and weaknesses
- ... and these depend on color and redshift
- Extract all possible information from data
- Better identify outliers
- Easy to incorporate other techniques

Photo-z PDF combination



+



+

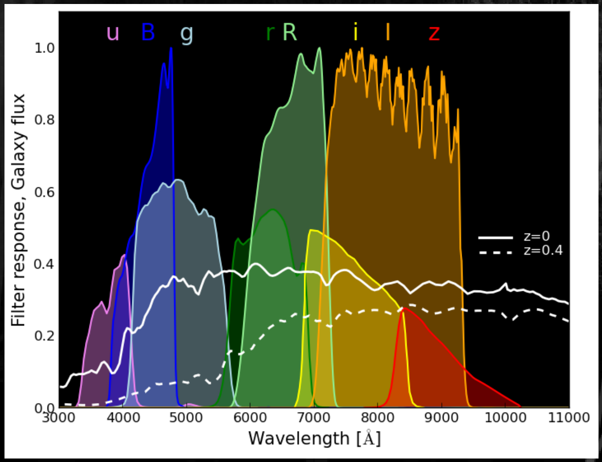


Photo-z PDF combination: Bayesian Model Averaging

$P(z)$ given by:

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

Photo-z PDF combination: Bayesian Model Averaging

$P(z)$ 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

Photo-z PDF combination: Bayesian Model Averaging

$P(z)$ 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:

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

Photo-z PDF combination: Bayesian Model Averaging

$P(z)$ 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:

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

Photo-z PDF combination: Bayesian Model Averaging

$P(z)$ 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:

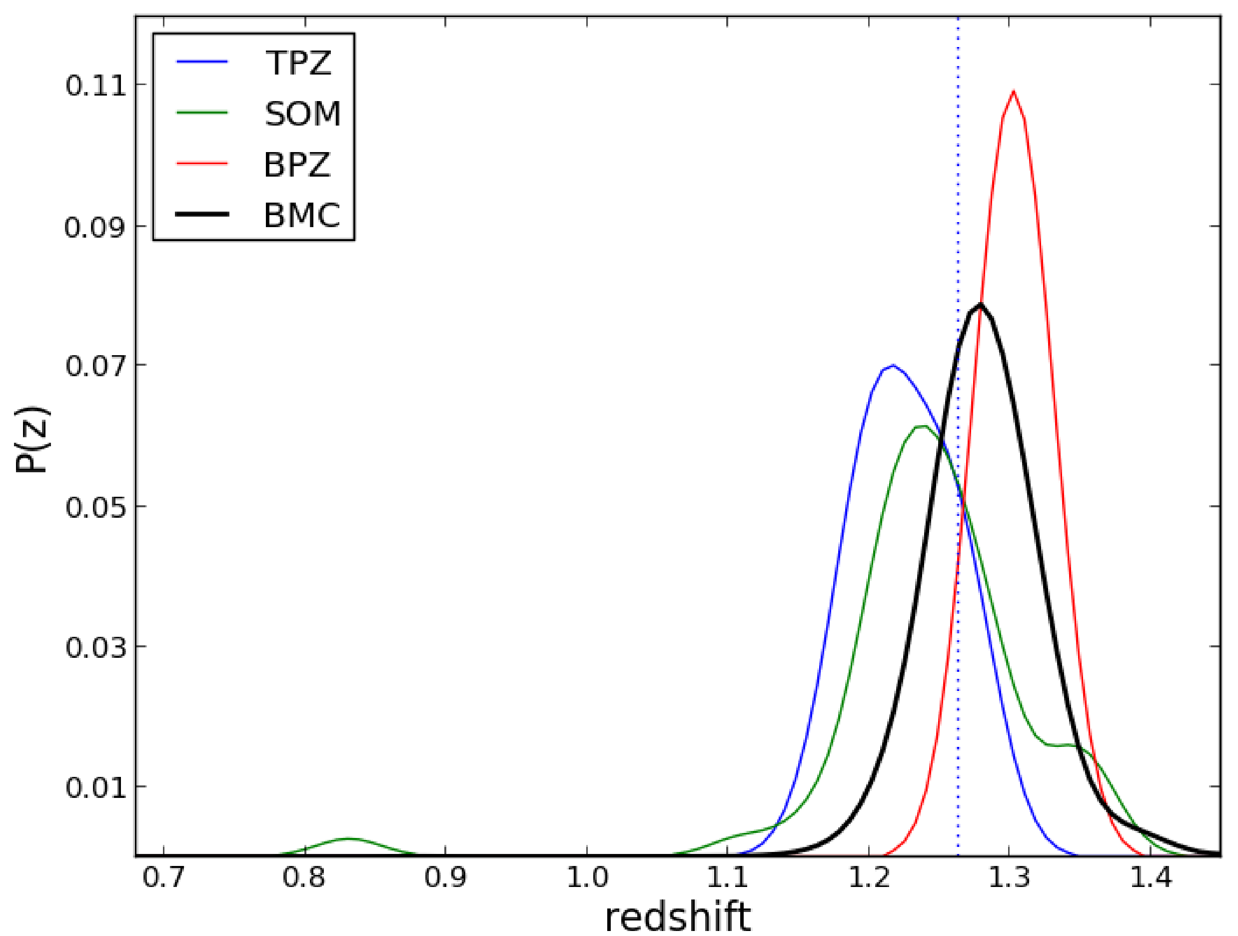
$$\boxed{P(z \mid \mathbf{x}, \mathbf{D}, \mathbf{M}) \propto \sum_k P(z \mid \mathbf{x}, M_k) P(M_k) \times (1 - \epsilon_k)^{N_d - N_k^{(b)}} (\epsilon_k)^{N_k^{(b)}}}$$

Photo-z PDF combination: BMC

Similarly to BMA, instead of selecting from models, we select from combined models (>100), we have $P(e \mid \mathbf{D})$ instead of $P(M_k \mid \mathbf{D})$. These models are generated by a Dirichlet process

Photo-z PDF combination: BMC

Similarly to BMA, instead of selecting from models, we select from combined models (>100), we have $P(e \mid \mathbf{D})$ instead of $P(M_k \mid \mathbf{D})$. These models are generated by a Dirichlet process



Carrasco Kind & Brunner 2014c (MNRAS, 442, 3380)

Combined PDF is, on average, better

BMA: marginalize over error in model

BMC: marginalize over error in combined models

Combination depends on galaxy colors

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 $P(M_k \mid \mathbf{D})$ and models are generated by a Dirichlet process

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

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

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

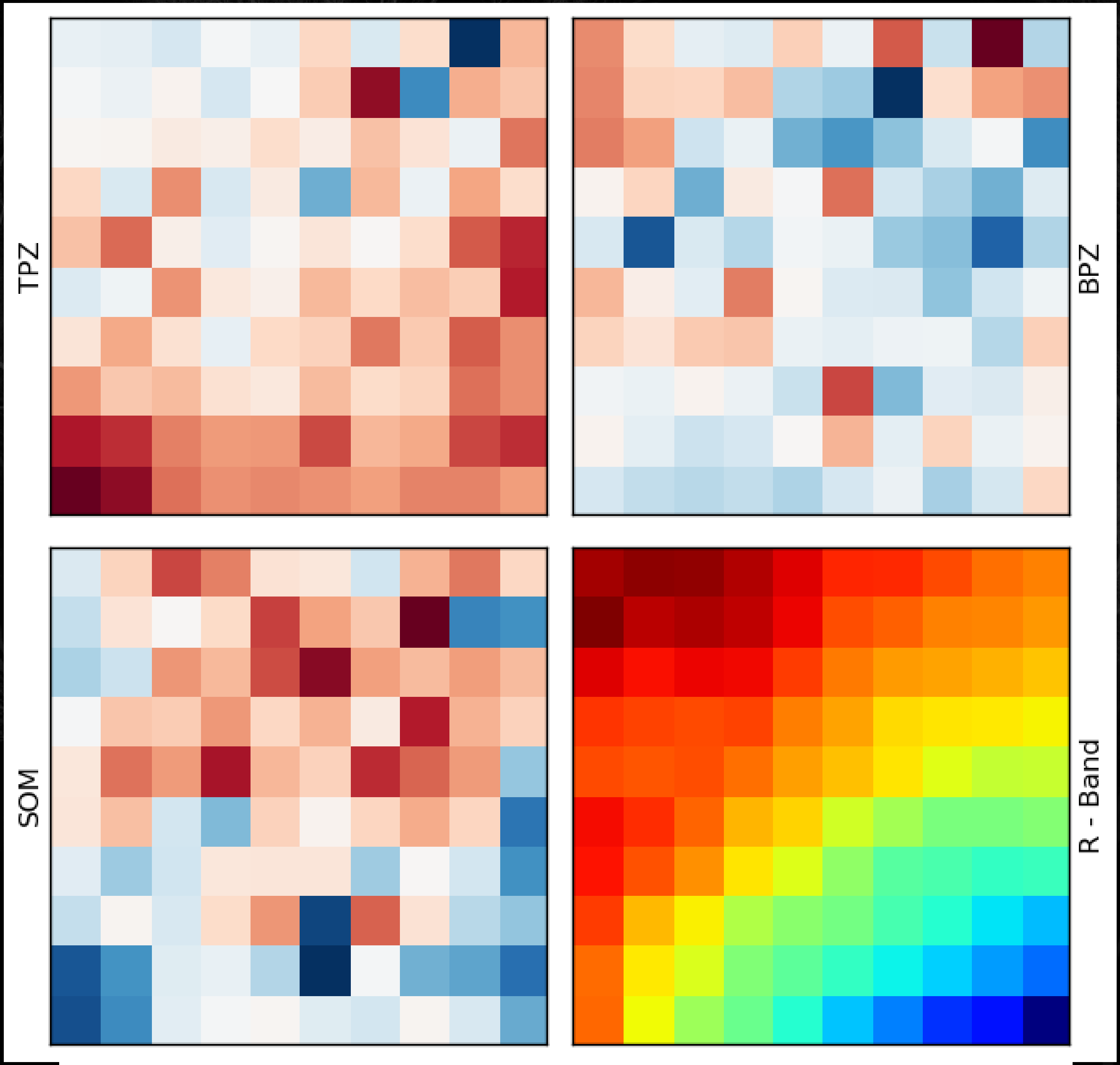
$$P(\mathbf{w}) \sim \mathcal{D}\mathrm{ir}(\boldsymbol{\alpha}) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k w_k^{\alpha_k - 1}$$

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: Bayesian framework



Carrasco Kind & Brunner 2014c (MNRAS 442, 3380)

This approach

Supervised method

+

Unsupervised method

+

Template fitting

+

Weigthing scheme

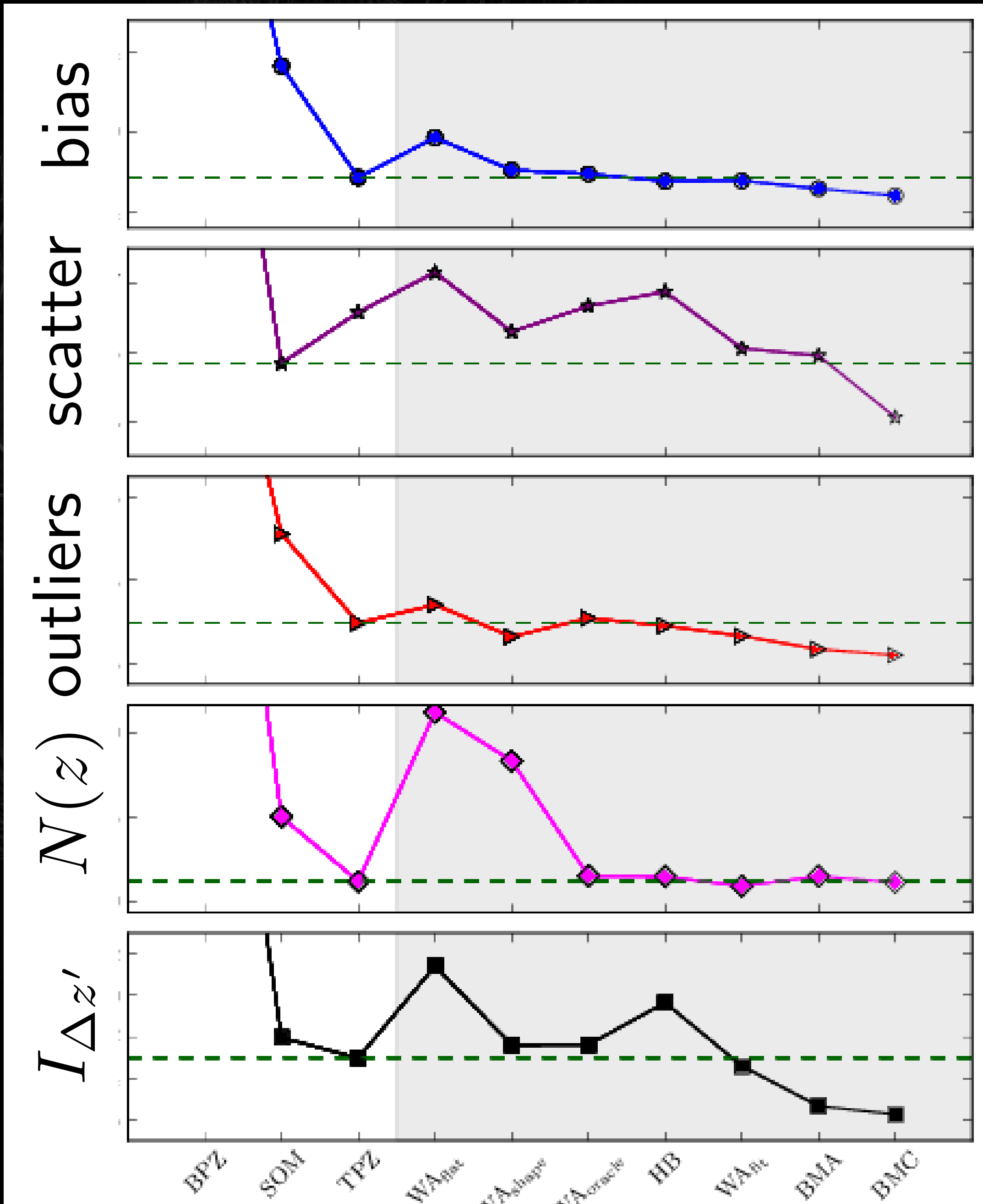
$\Downarrow$

photo-z PDF

+

Outliers

Photo-z PDF combination: Results



Carrasco Kind & Brunner 2014c (MNRAS 442, 3380)

- 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: Correlation

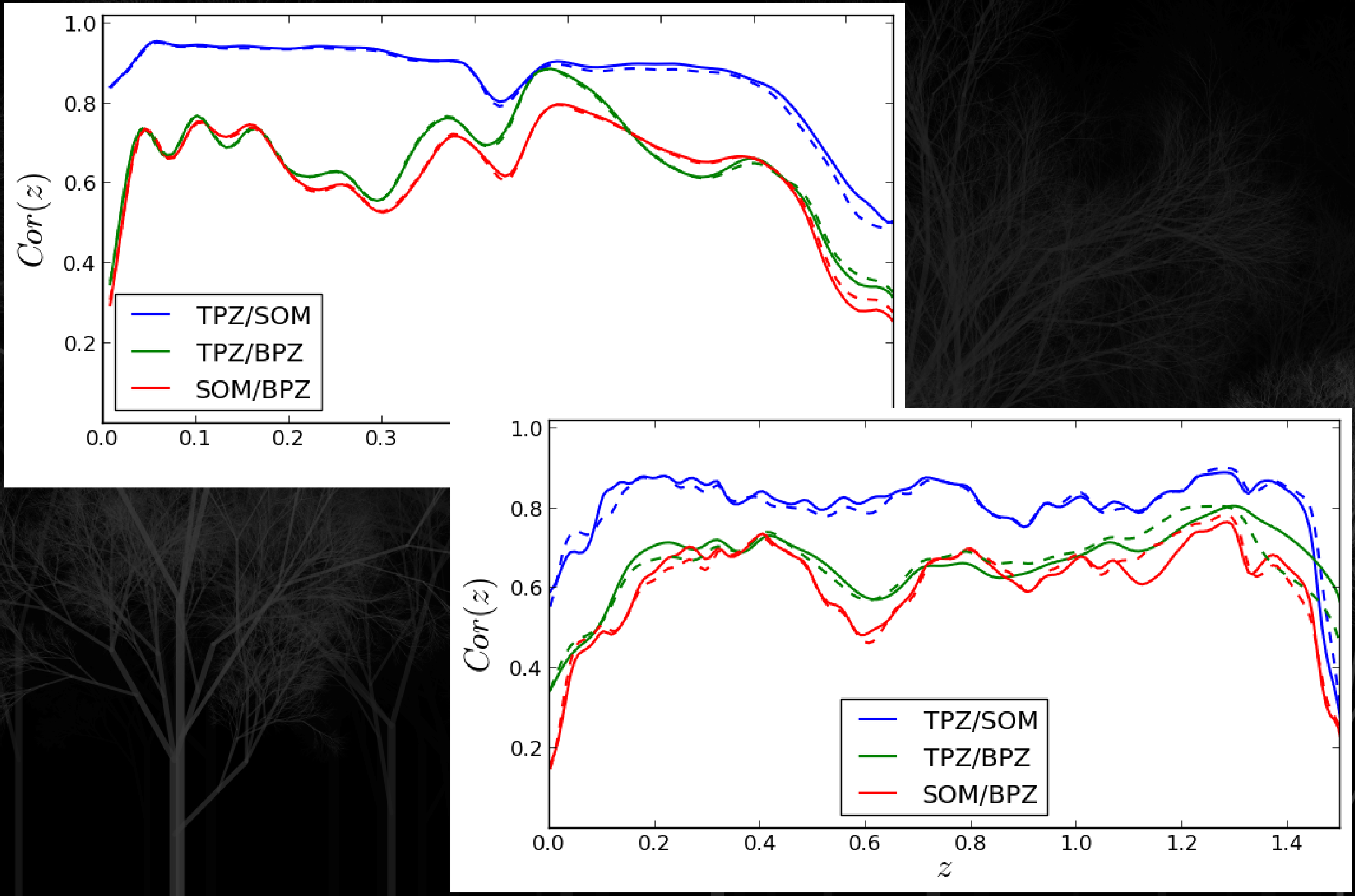


Photo-z PDF combination: Outliers

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

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:

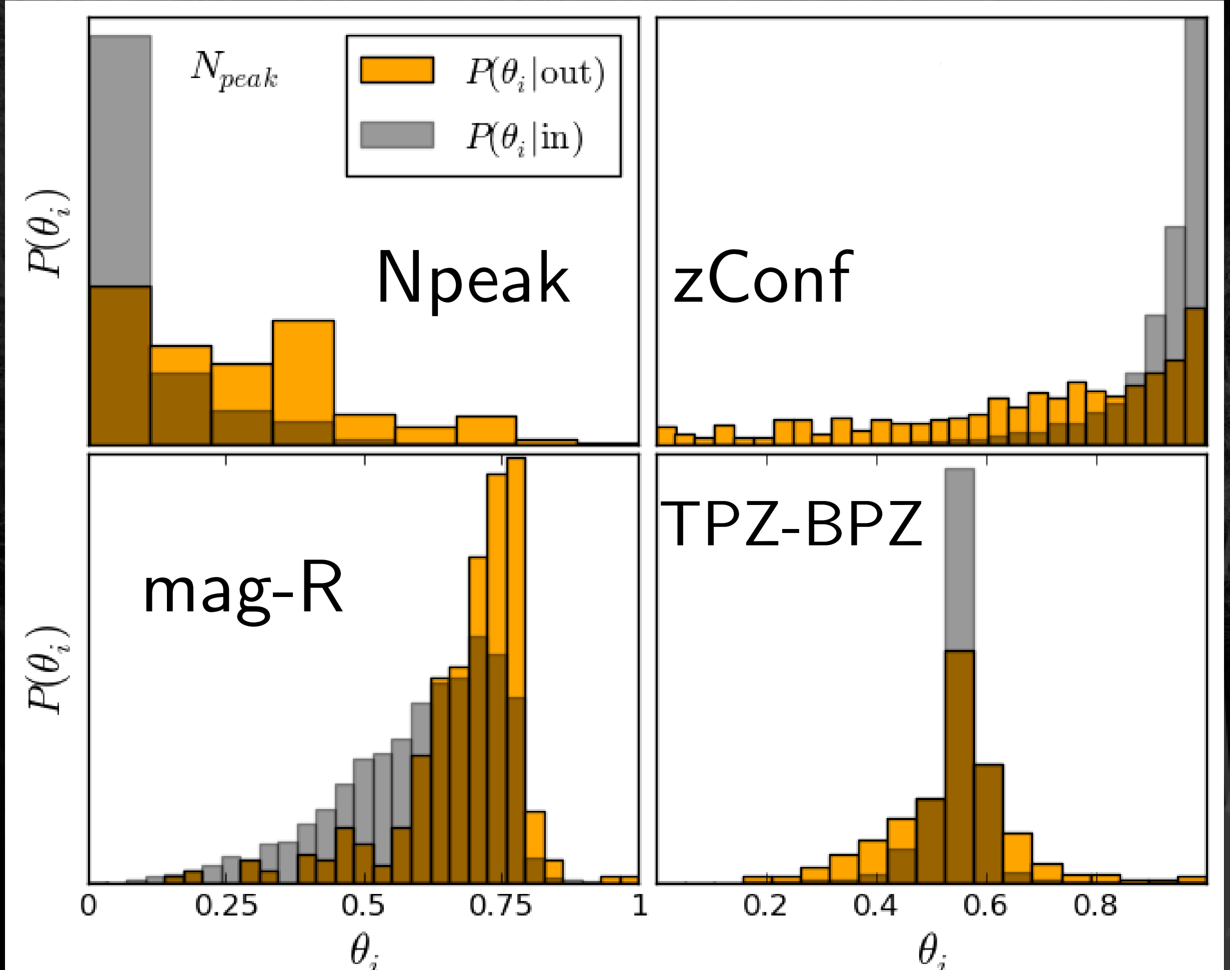
$$P(\text{out} \mid \boldsymbol{\theta}) = \frac{P(\text{out}) \prod P(\theta_i \mid \text{out})}{\prod P(\theta_i \mid \text{out}) + \prod P(\theta_i \mid \text{in})}$$

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

Photo-z PDF combination: Outliers

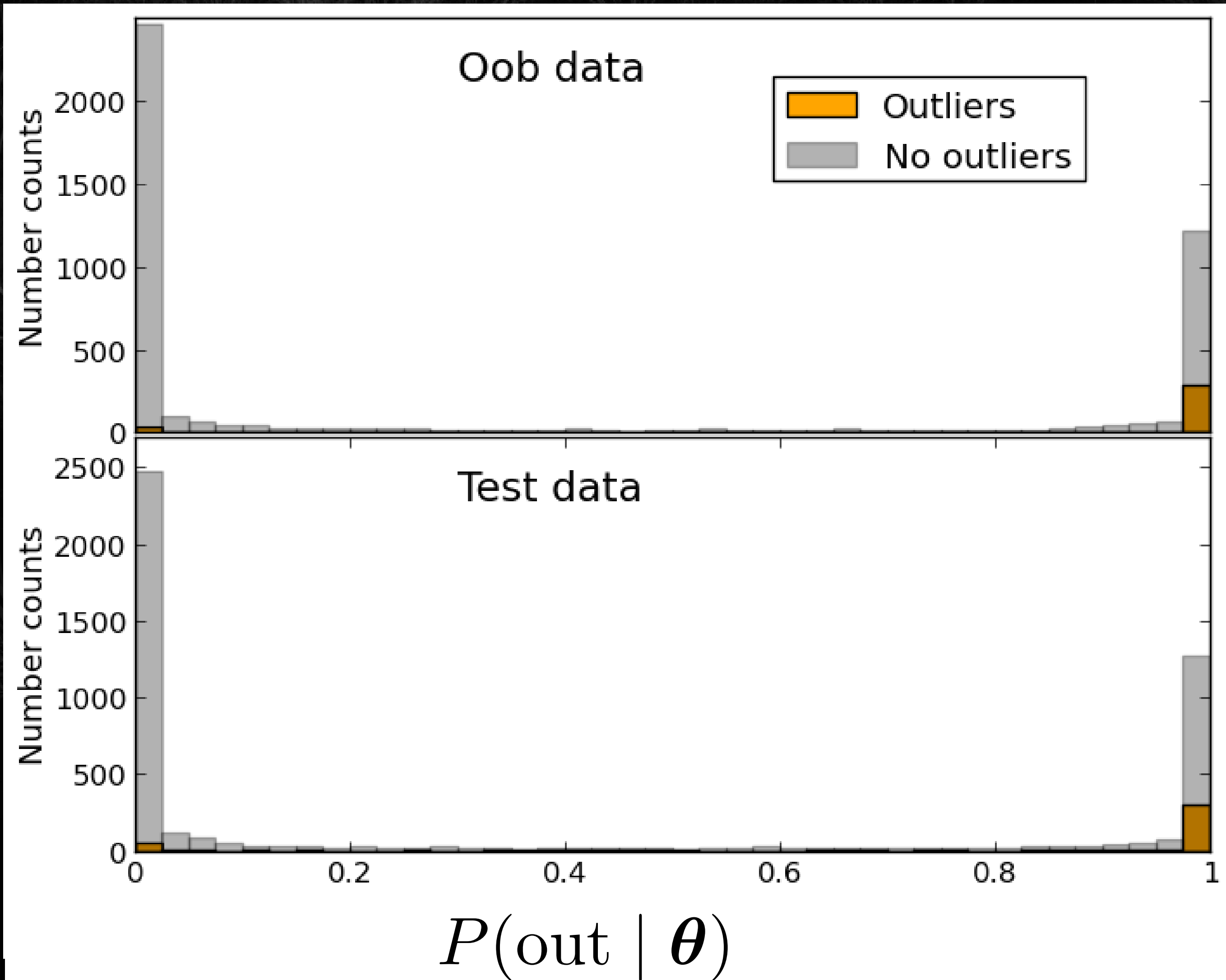
Naïve Bayes Classifier (same used for spam emails) to identify "spam" galaxies using information from multiple techniques

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



Carrasco Kind & Brunner 2014c (MNRAS 442, 3380)

Photo-z PDF combination: Outliers



Carrasco Kind & Brunner 2014c (MNRAS 442, 3380)

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

Photo-z PDF representation and storage

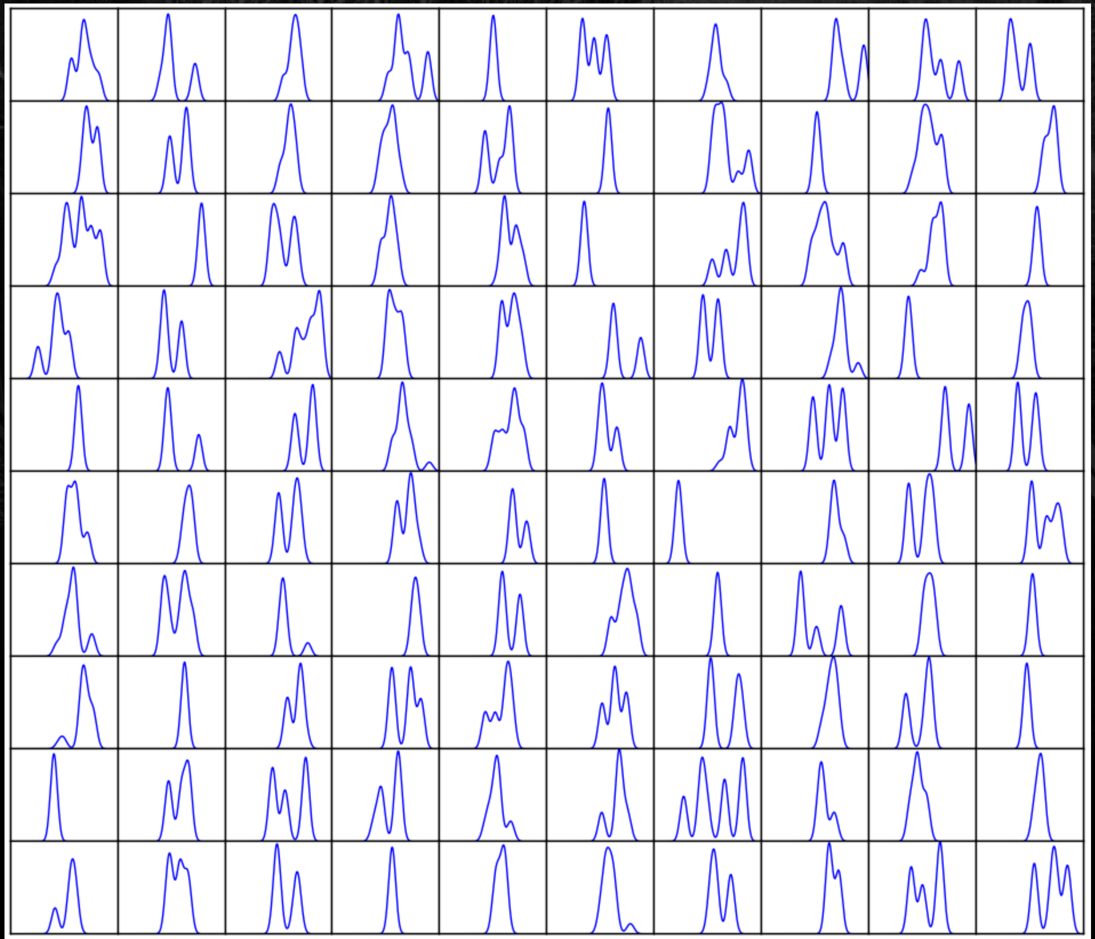


Photo-z PDF storage: Strategies

Single Gaussian fit

Multi-Gaussan fit

Monte Carlo sampling

Sparse representation techniques

Reduce number of points while increasing accuracy

Photo-z PDF storage: Strategies

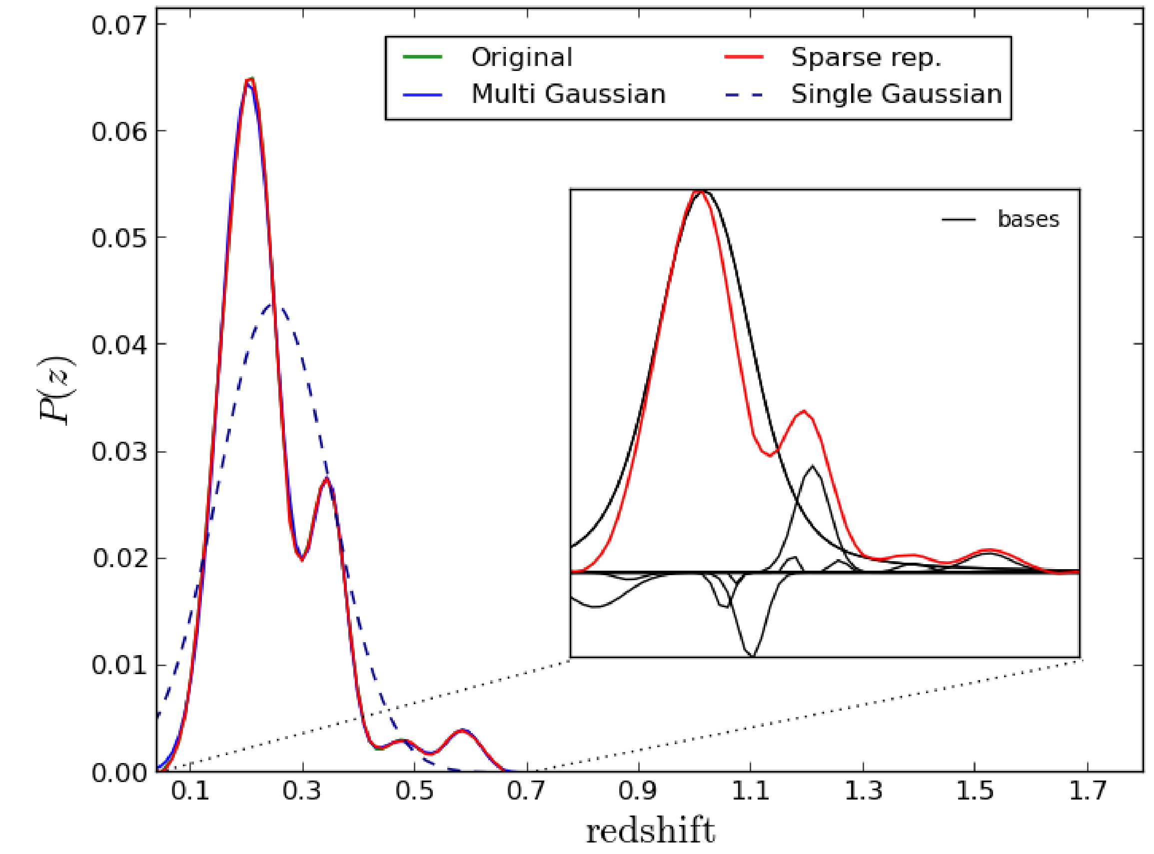
Single Gaussian fit

Multi-Gaussan fit

Monte Carlo sampling

Sparse representation techniques

Reduce number of points while increasing accuracy



Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)

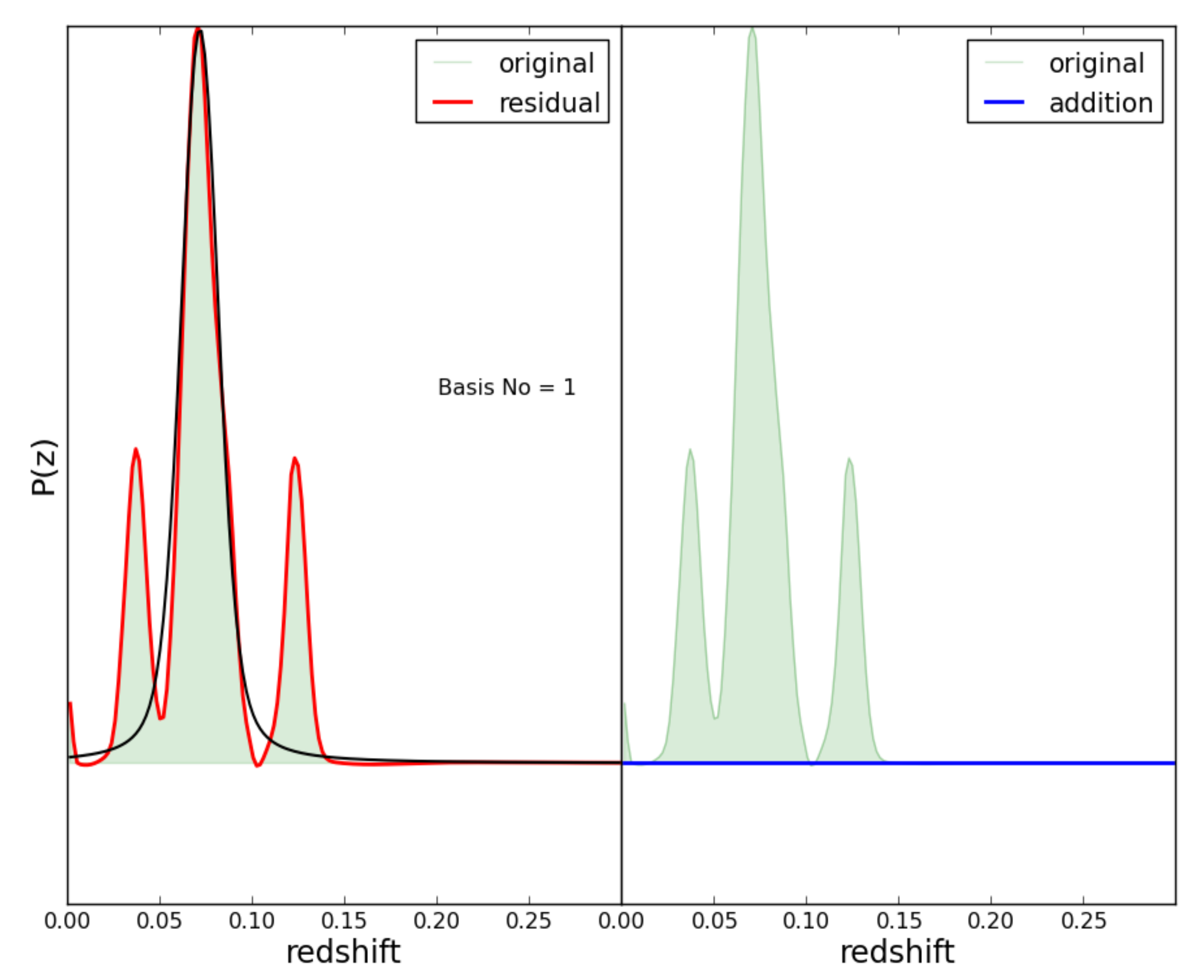
Photo-z PDF storage: Sparse representation

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

With only 10-20 bases achieve 99.9 % accuracy

Use 32-bits integer per basis, compression

Store Multiple PDFs



Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)

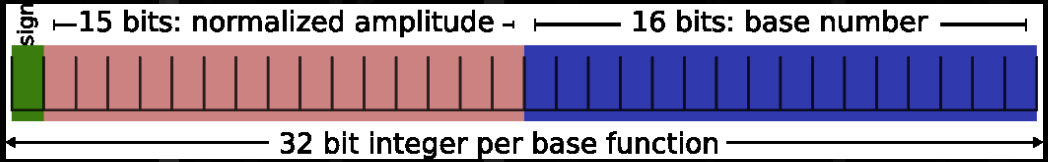


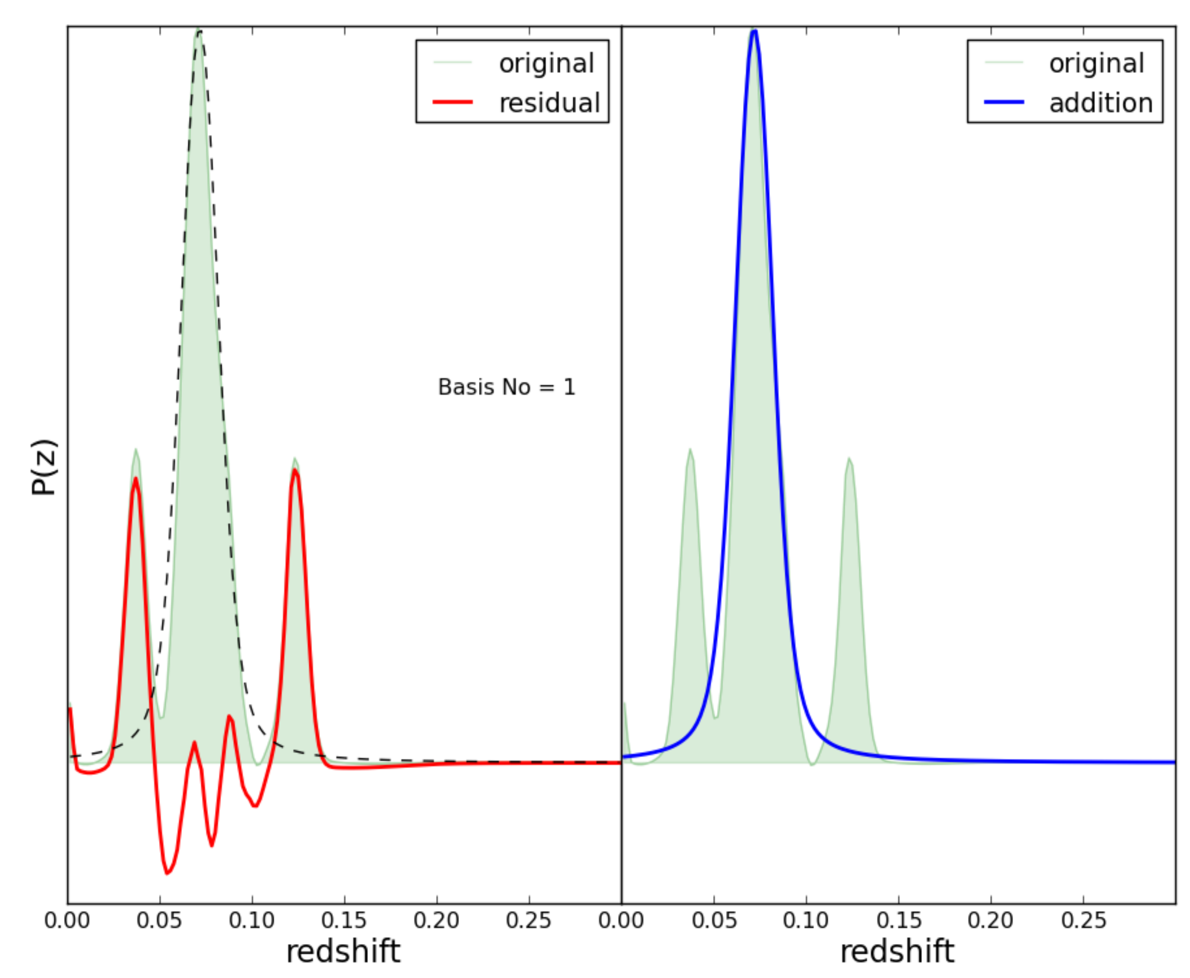
Photo-z PDF storage: Sparse representation

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

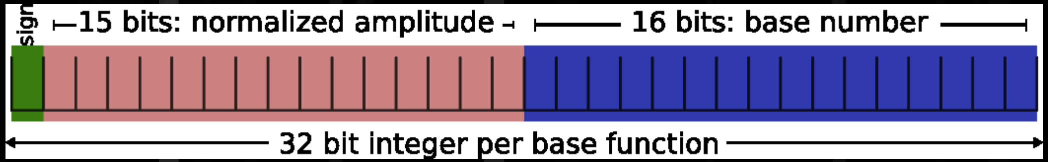
With only 10-20 bases achieve 99.9 % accuracy

Use 32-bits integer per basis, compression

Store Multiple PDFs



Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)

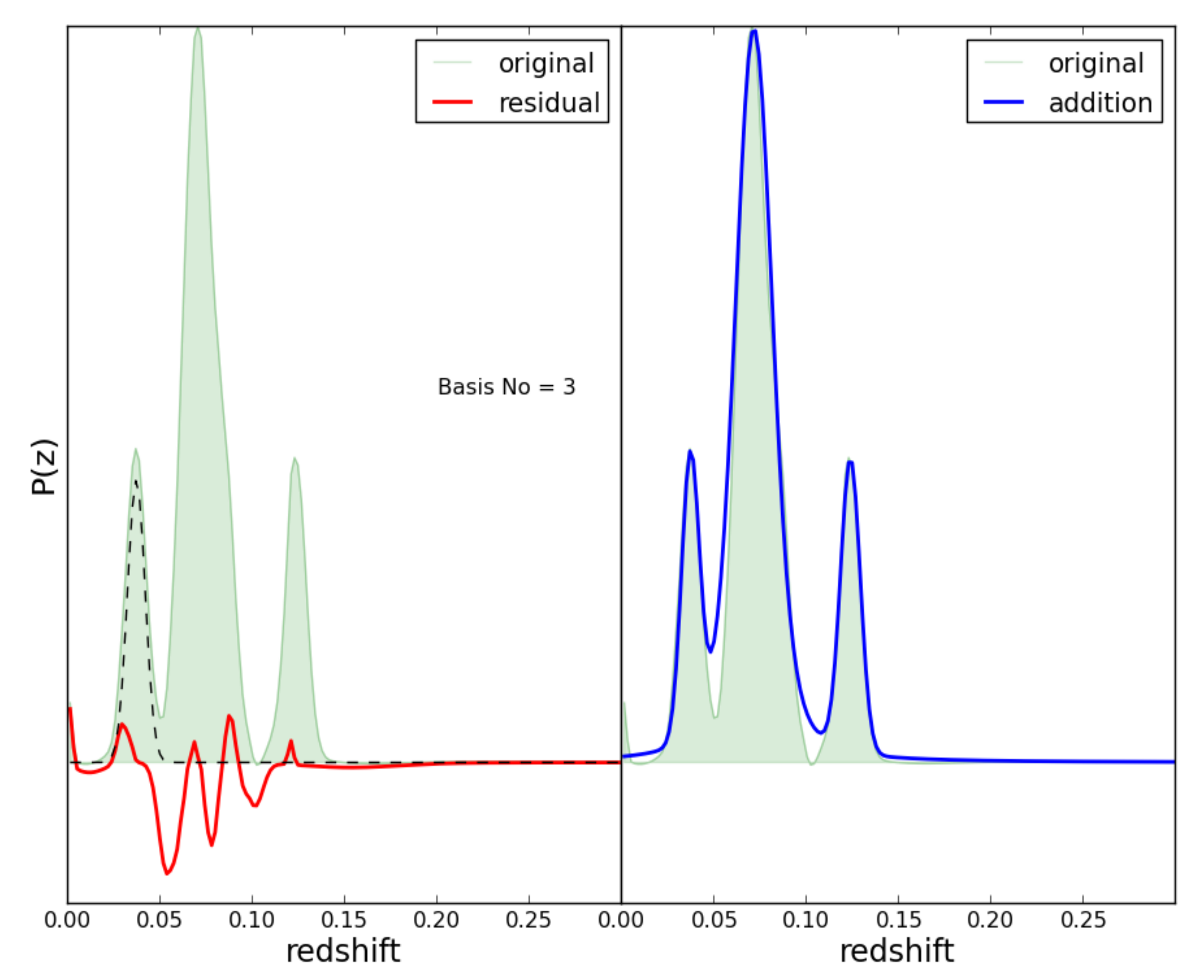


Use Gaussian and Voigt profiles as bases, need N^2_{original} bases

With only 10-20 bases achieve 99.9 % accuracy

Use 32-bits integer per basis, compression

Store Multiple PDFs



Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)

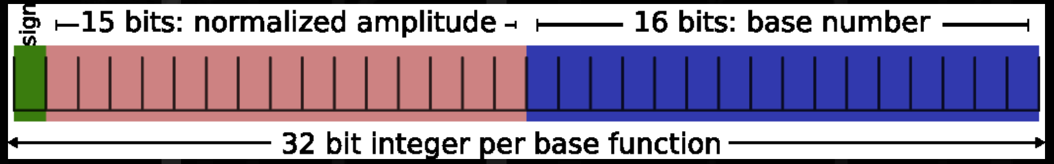


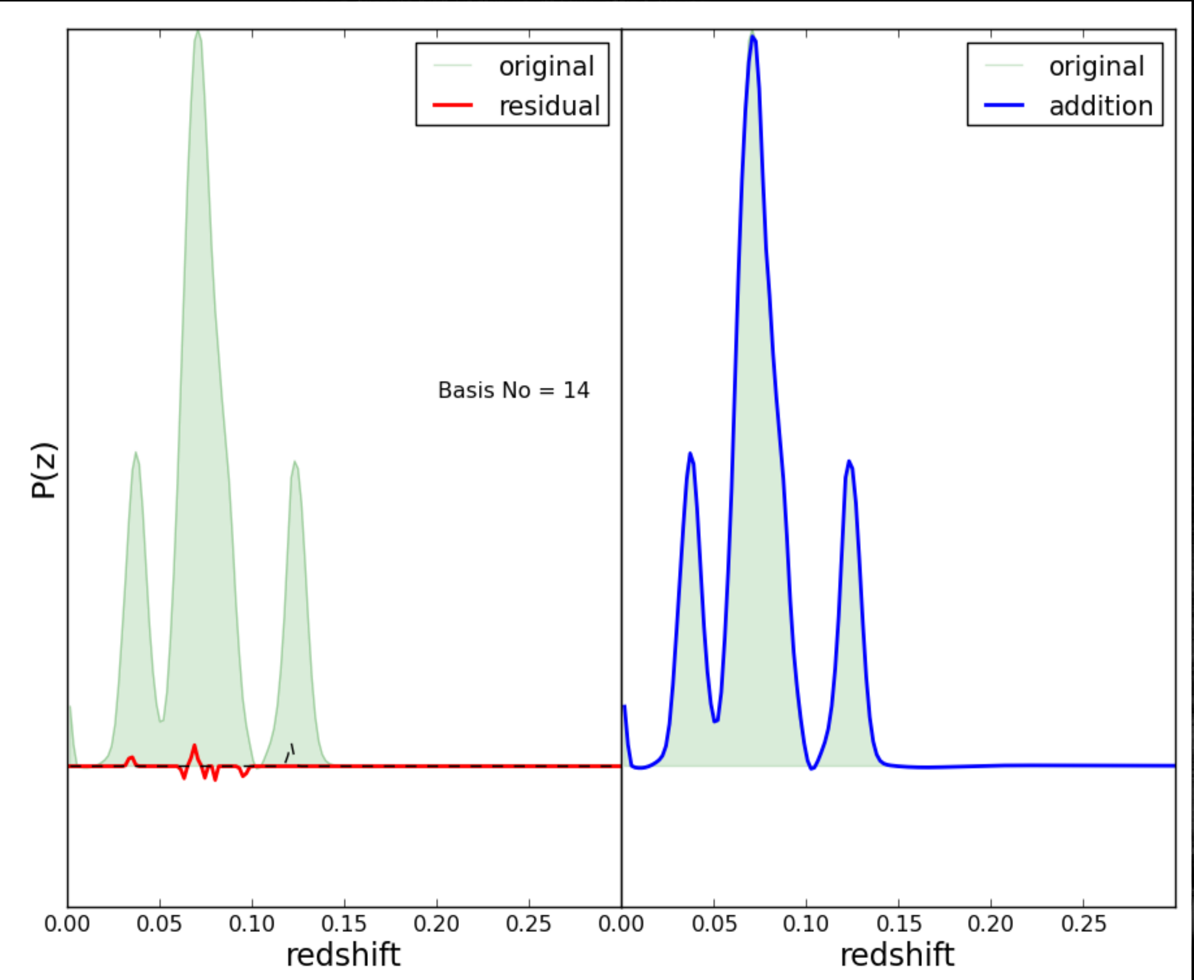
Photo-z PDF storage: Sparse representation

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

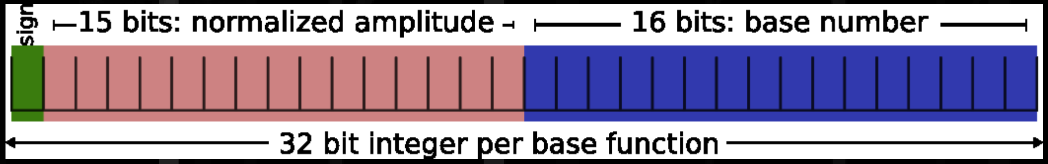
With only 10-20 bases achieve 99.9 % accuracy

Use 32-bits integer per basis, compression

Store Multiple PDFs



Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)



Conclusions

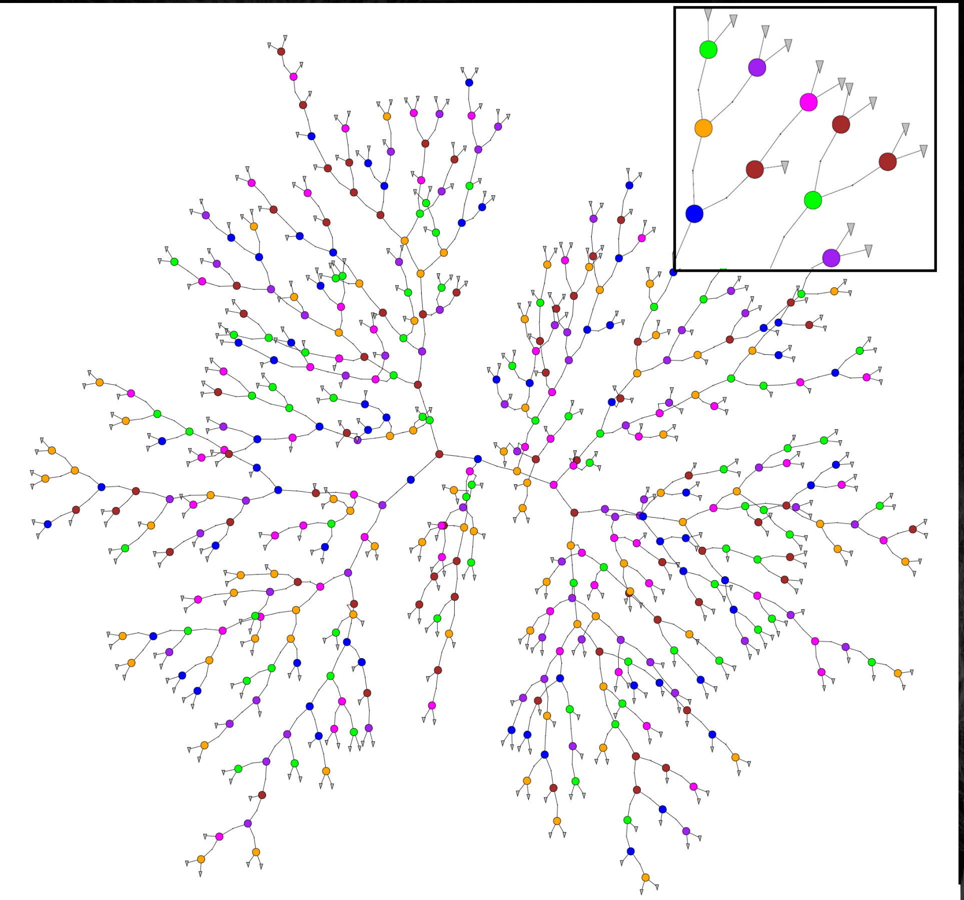
- Several photo-z methods out there
- Advantages of combining results
- Outliers identification
- Efficient storage of PDF

Matias Carrasco Kind
NCSA/UIUC
mcarras2@ncsa.illinois.edu
http://matias-ck.com/

EXTRA SLIDES

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

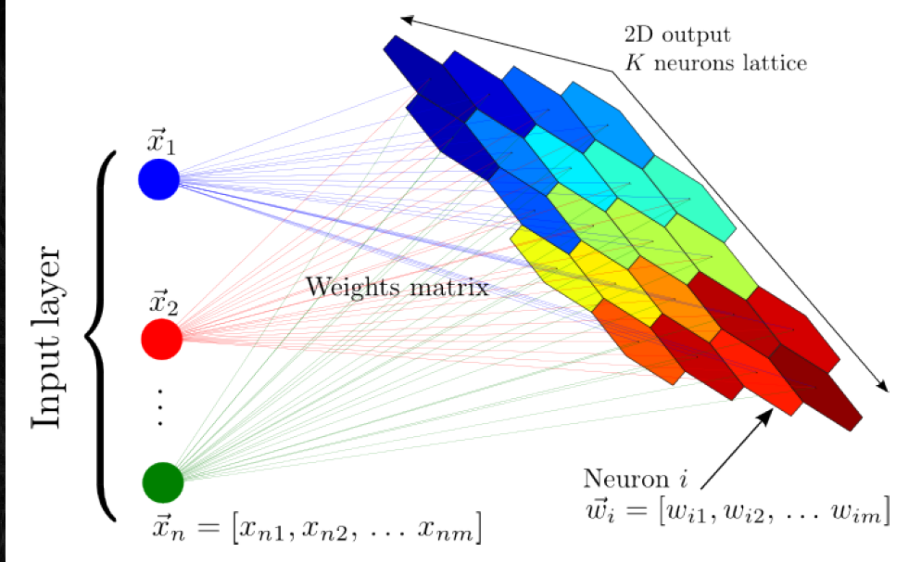


Carrasco Kind & Brunner 2013a (MNRAS, 432, 1483)

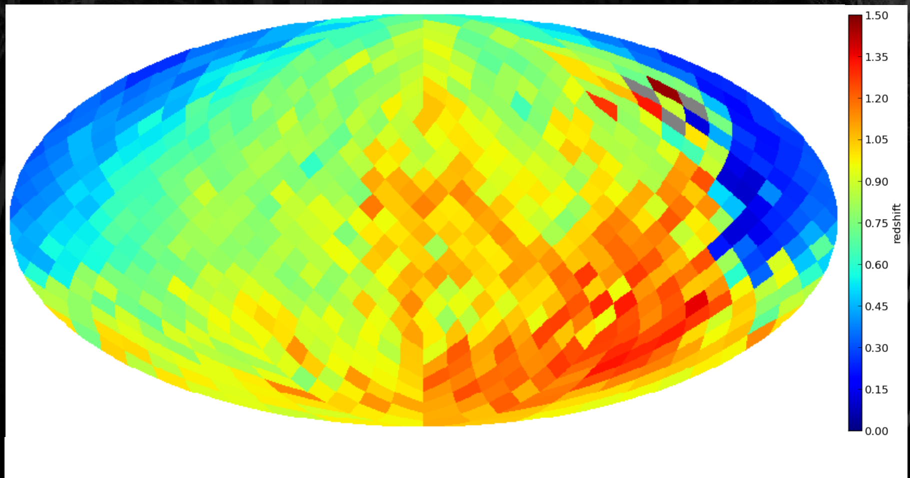
http://matias-ck.com/mlz

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: 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

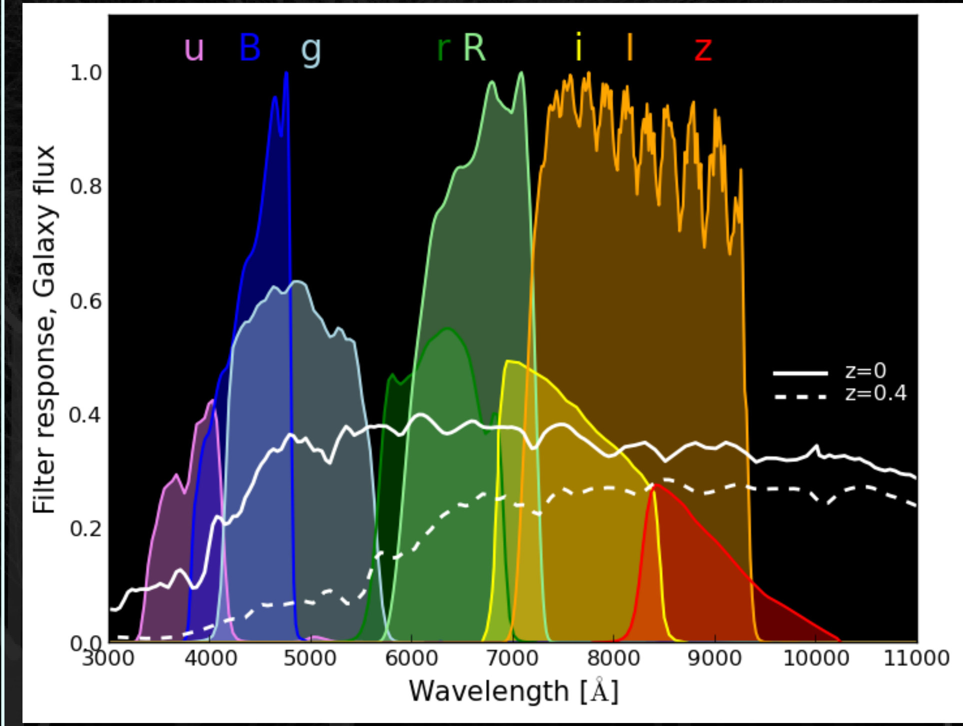
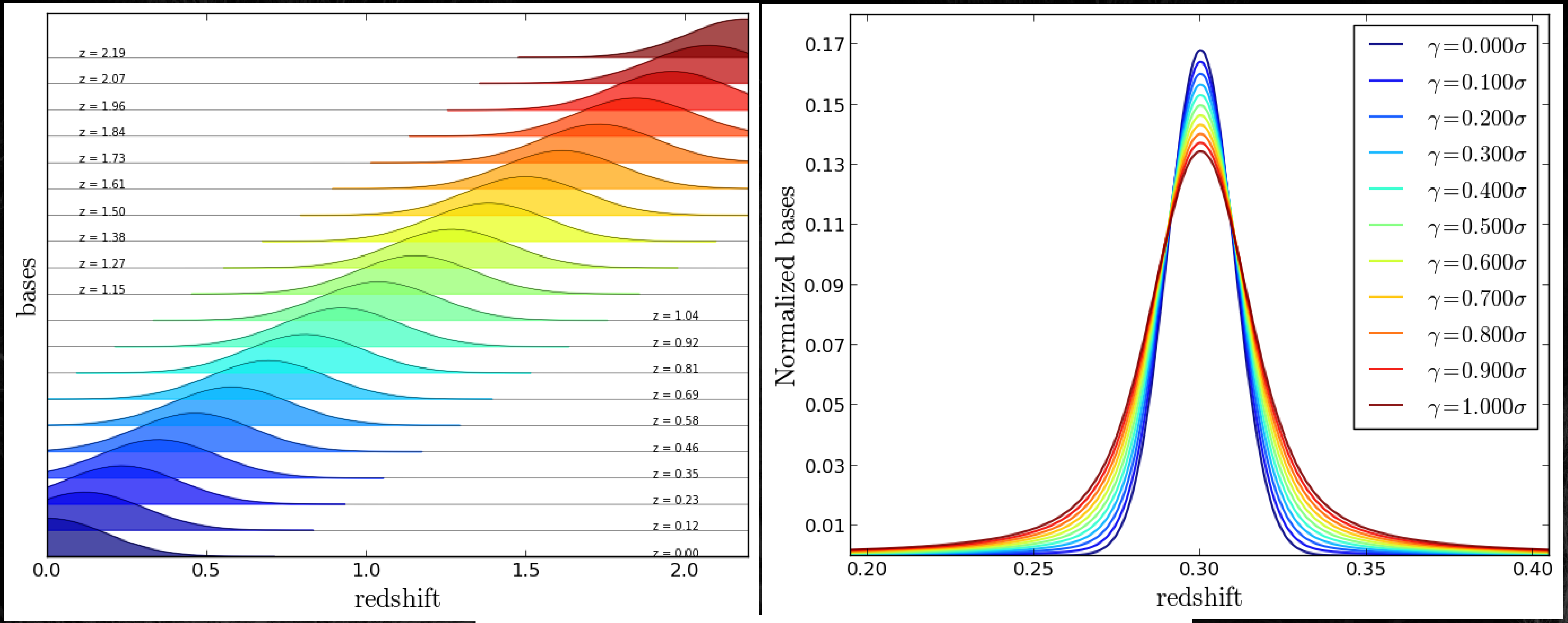


Photo-z PDF storage: Dictionary

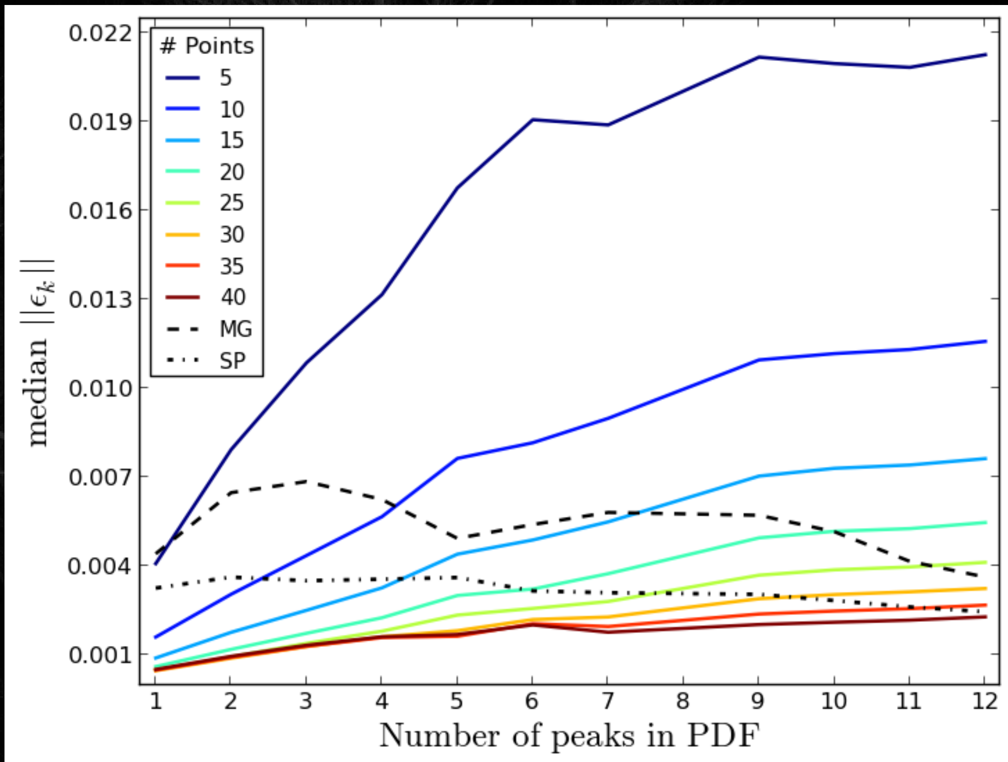


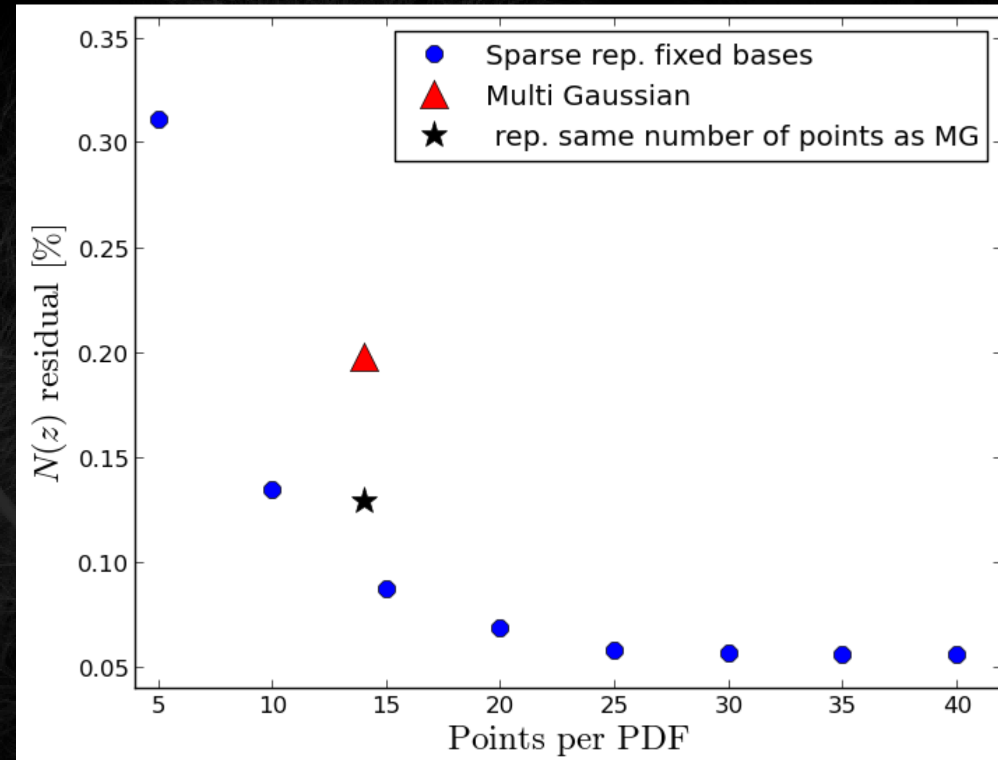
Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)

Combination of Gaussian and Voigt profiles

Covering the whole redshift space, at each location we have several bases

Photo-z PDF storage: Results





Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)

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)