ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Probabilistic photometric redshifts in the era of Petascale Astronomy

Matías Carrasco Kind

PhD Committee:
Robert J. Brunner (Chair)
Athol J. Kemball
Paul M. Ricker
Jon J. Thaler

PhD Thesis defense August 19$^{\text{th}}$, 2014

Publications arisen from this thesis

Refereed publications

- **Carrasco Kind, M.**, & Brunner, R. J., 2013, MNRAS, 432, 1483
- **Carrasco Kind, M.**, & Brunner, R. J., 2014, MNRAS, 438, 3409
- **Carrasco Kind, M.**, & Brunner, R. J., 2014, MNRAS, 441, 3550
- **Carrasco Kind, M.**, & Brunner, R. J., 2014, MNRAS, 442, 3380
- Sanchéz, C., **Carrasco Kind, M.**, Lin, H., Miquel, R., et al. 2014, MNRAS submitted., arXiv: 1406.4407
- Banerji, M., Jouvel, S., Lin, H., McMahon, R.G., Lahav, O., Castander, F., Abdalla, F., Bertin, E., Bosman, S., Carnero, A., **Carrasco Kind, M.**, et al. 2014, MNRAS submitted., arXiv: 1407.380

Other publications

- **Carrasco Kind, M.** & Brunner, R. J. 2013, ASPC, 475, 69C
- Newman, J., et al. Snowmass 2013 white paper, arXiv: 1309.5384
- Abate, A., et al. LSST-DESC white paper, 2012, arXiv: 1211.0310

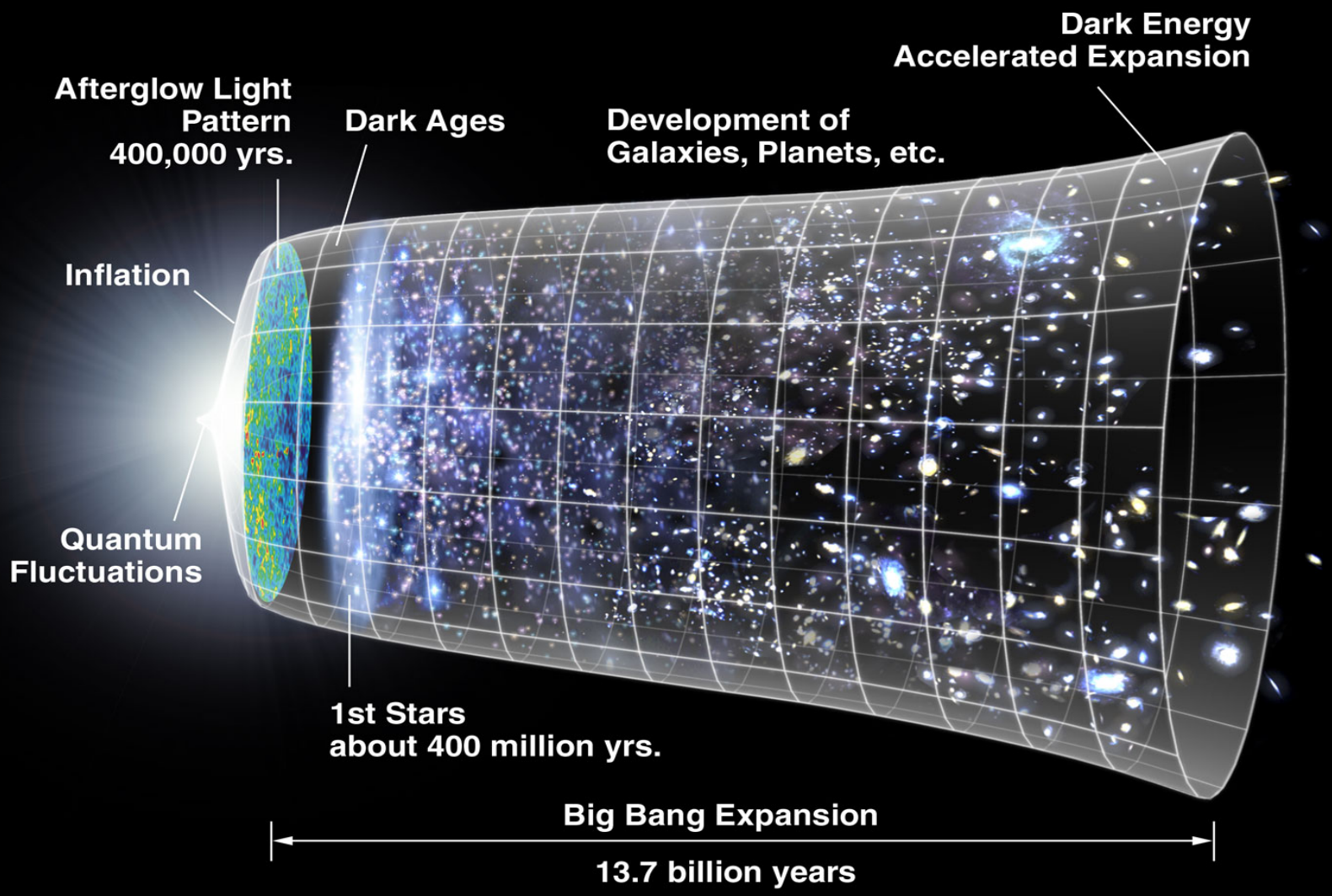
Outline

- INTRODUCTION
- HOW TO COMPUTE PHOTO-Z PDFs
- HOW TO COMBINE PHOTO-Z PDFs METHODS
- HOW TO REPRESENT PHOTO-Z PDF
- APPLICATIONS
- CONCLUSIONS AND FUTURE WORK

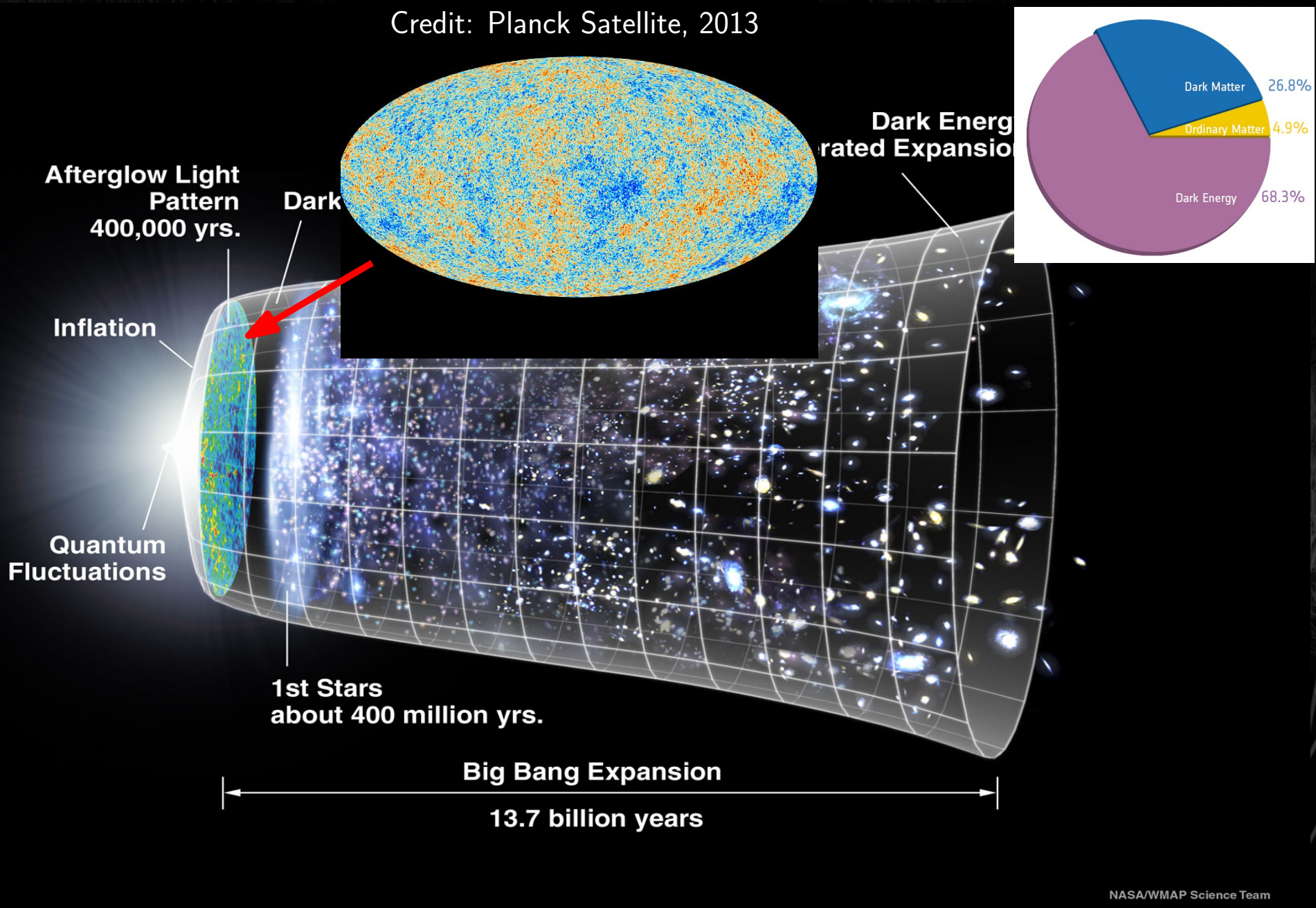
Outline

- INTRODUCTION
- HOW TO COMPUTE PHOTO-Z PDFs
- HOW TO COMBINE PHOTO-Z PDFs METHODS
- HOW TO REPRESENT PHOTO-Z PDF
- APPLICATIONS
- CONCLUSIONS AND FUTURE WORK

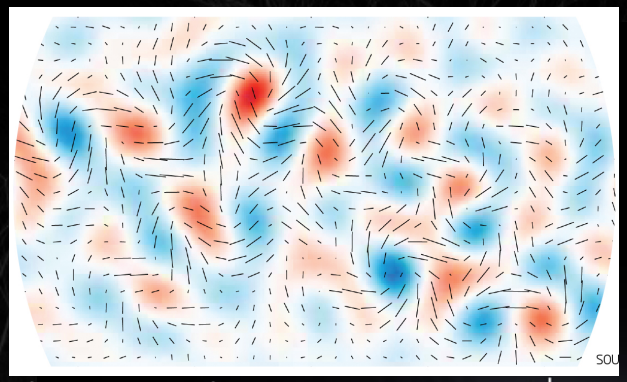
Current picture of the Universe



Current picture of the Universe

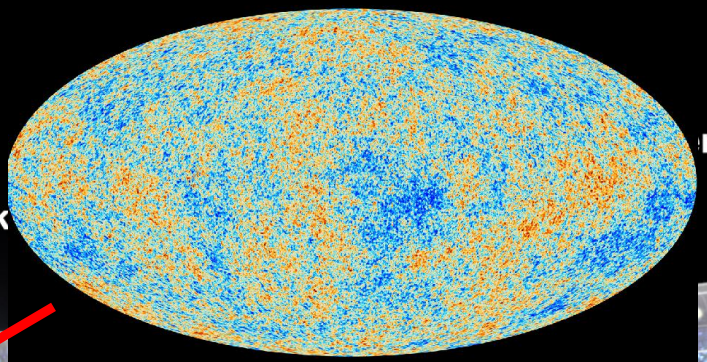


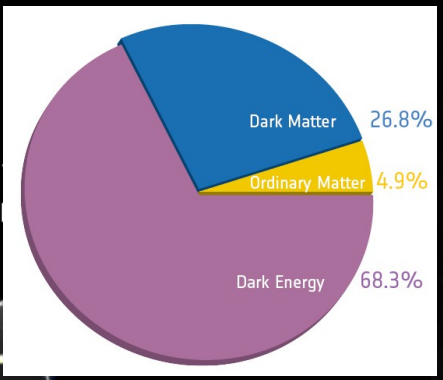
Current picture of the Universe



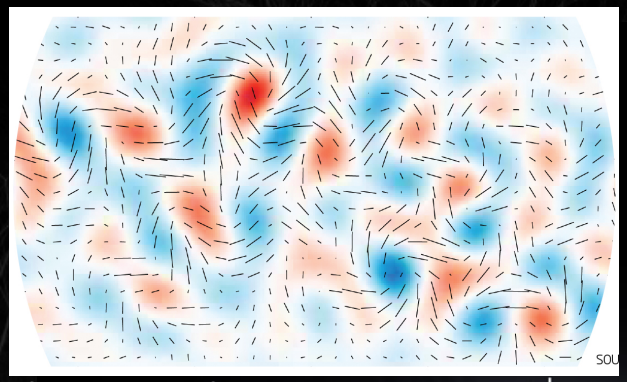
Credit: BICEP2, 2014

Credit: Planck Satellite, 2013



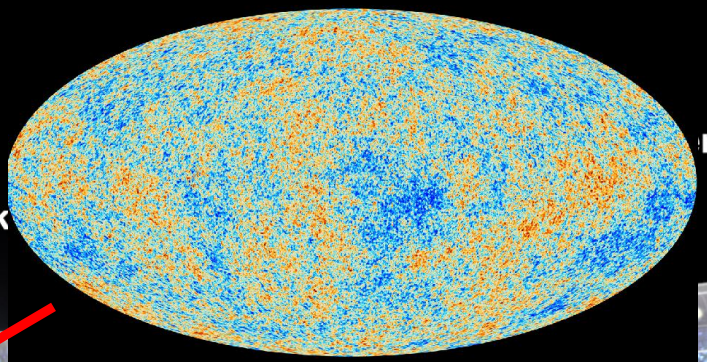


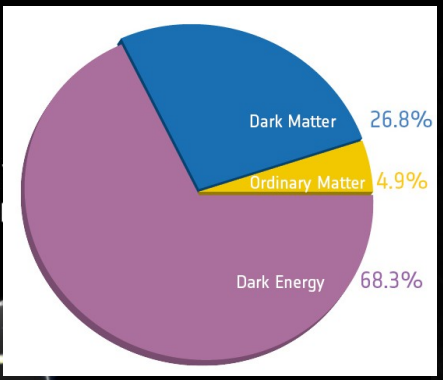
Current picture of the Universe



Credit: BICEP2, 2014

Credit: Planck Satellite, 2013





Dark Energ
rated Expansio

Dark

Inflation

Quantum Fluctuations

1st Stars
about 400 million yrs.

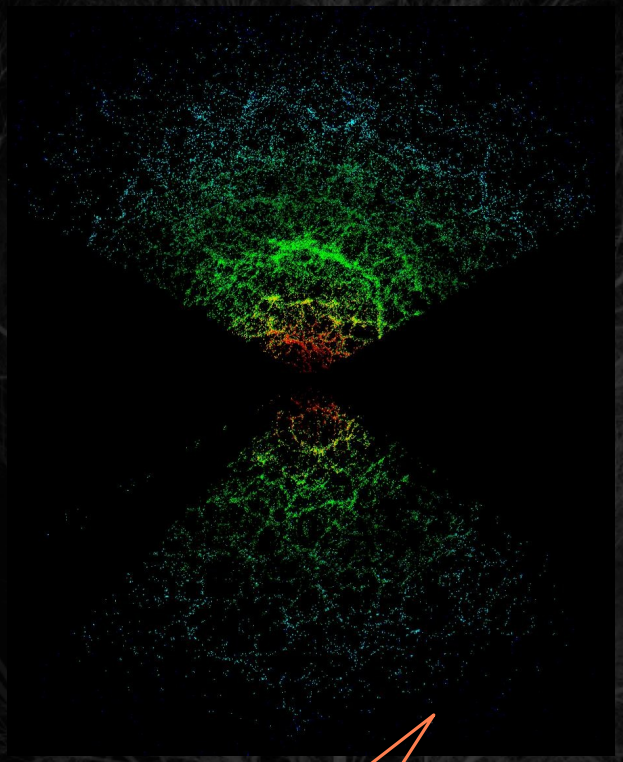
Big Bang Expansion

13.7 billion years

observables

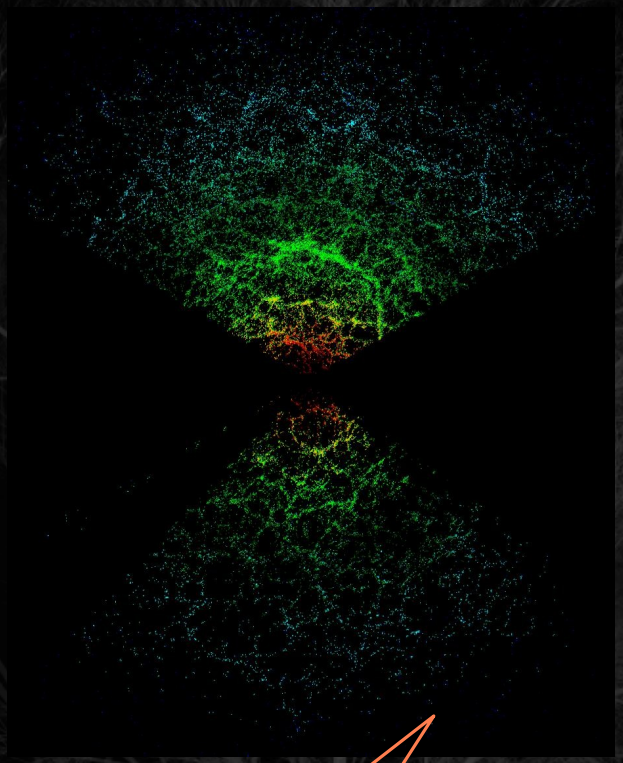
NASA/WMAP Science Team

Large-scale structure



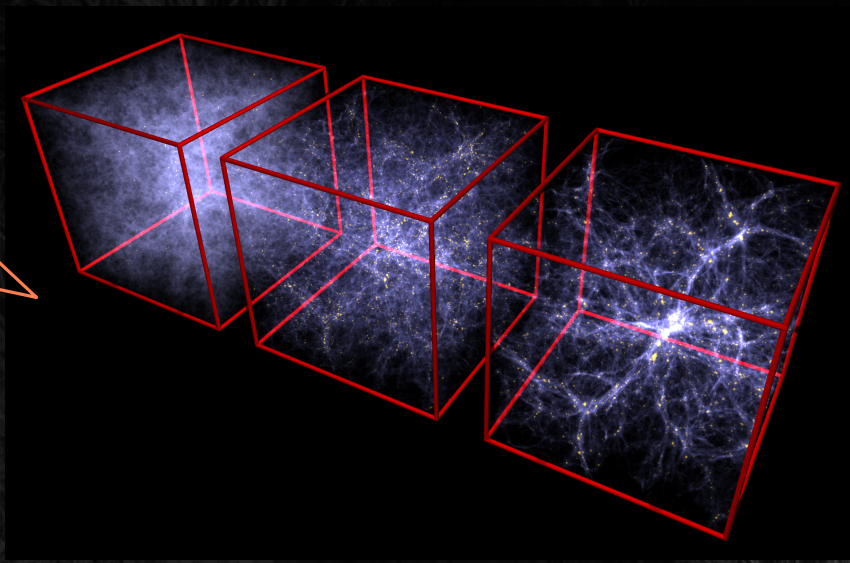
Distribution of galaxies shows clustering and cosmic web

Large-scale structure



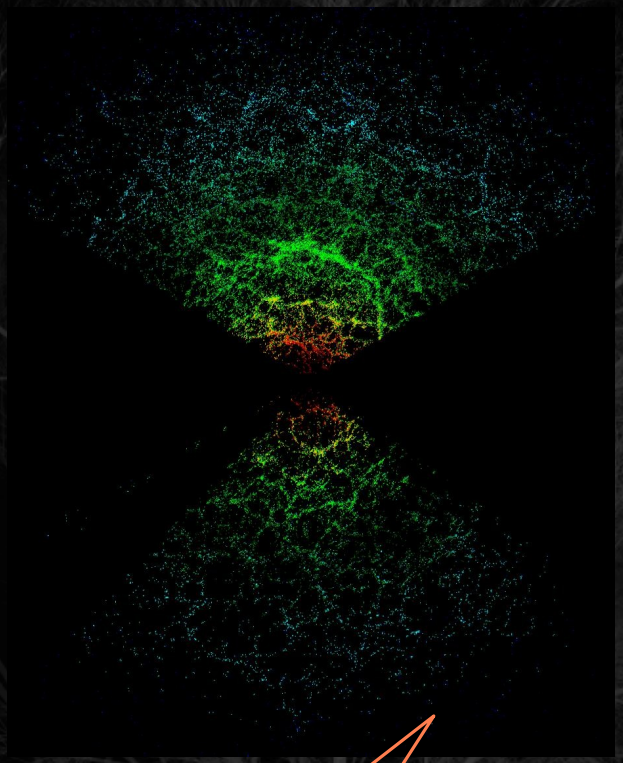
Distribution of galaxies shows clustering and cosmic web

Clustering evolution with cosmic time



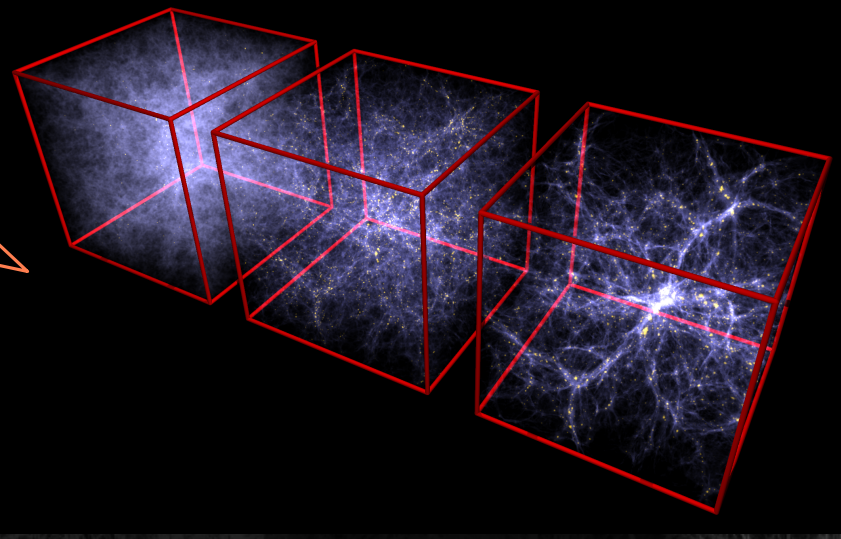
Cosmological Observables

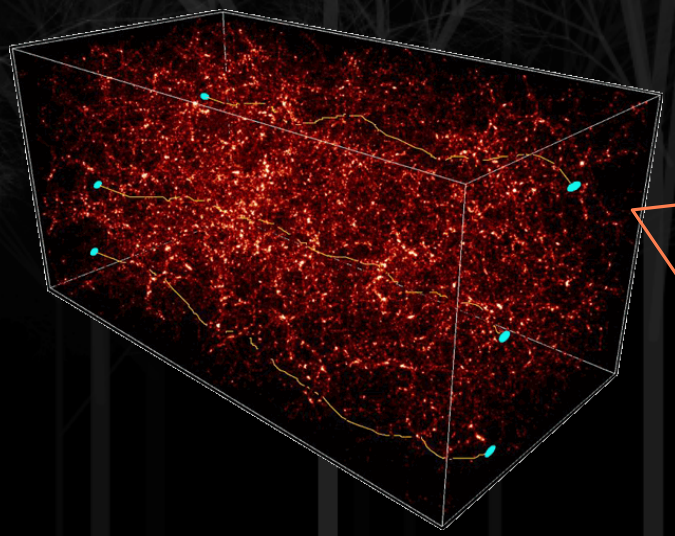
Large-scale structure



Distribution of galaxies shows clustering and cosmic web

Clustering evolution with cosmic time



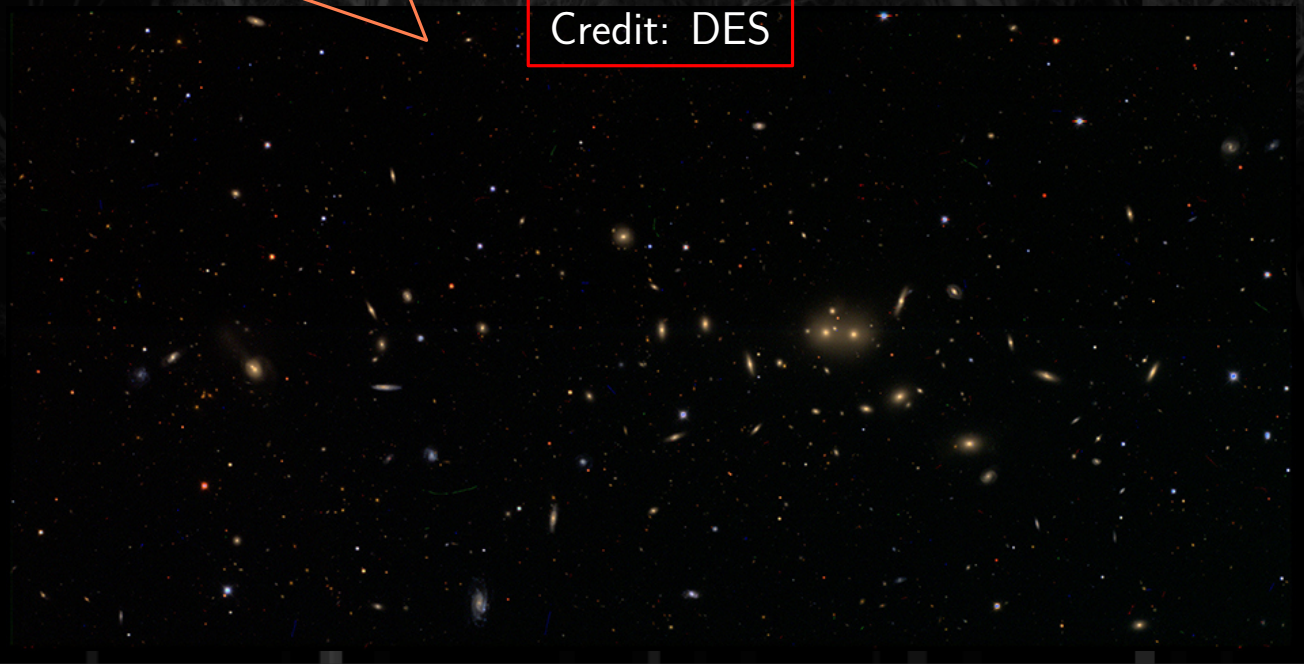


Gravitational effects deflect light given a mass distribution

Cluster-scale structure

Cluster of galaxies, gravitationally bounded structures

Cluster members, alignments, mass, etc...



Credit: DES

Cluster-scale structure

Cluster of galaxies, gravitationally bounded structures

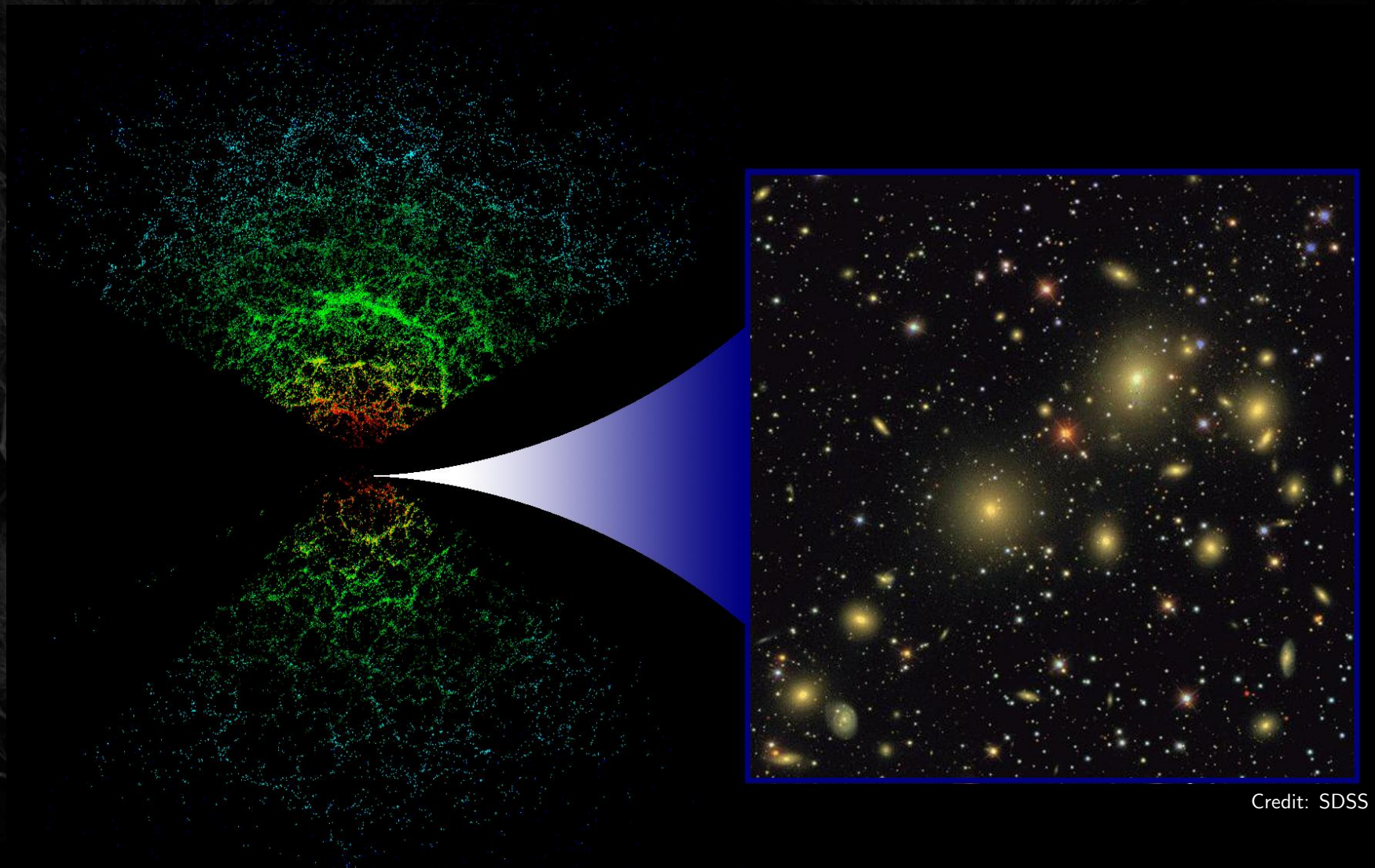
Cluster members, alignments, mass, etc...

Credit: HST

Credit: DES

Strong lensing from cluster of galaxies. GR effects, total mass of lens clusters, very far lensed galaxies

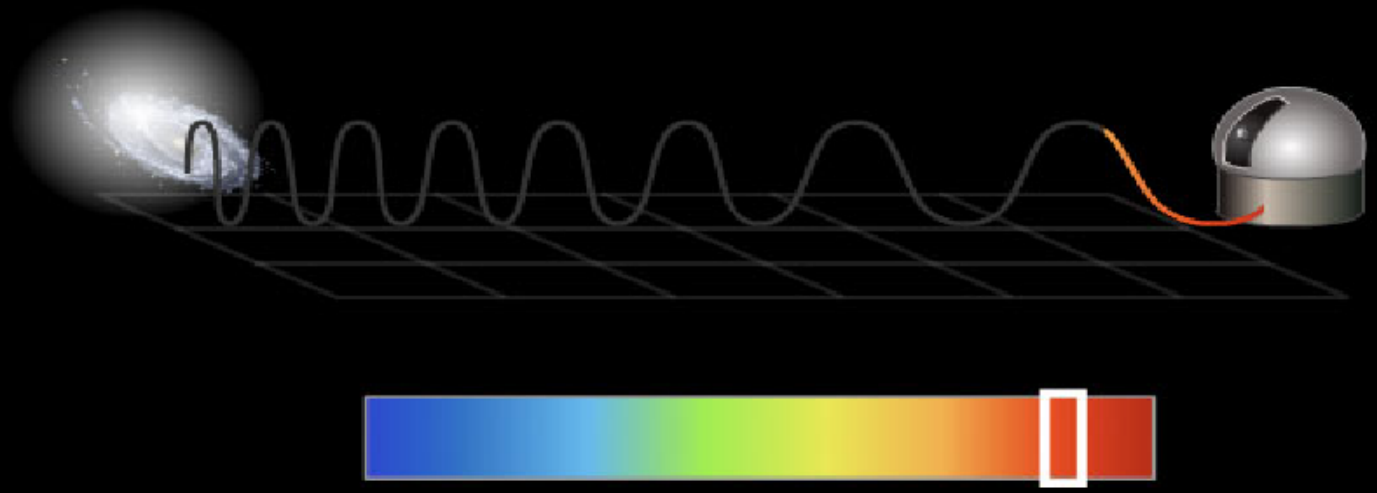
The need of distances in cosmology



Credit: SDSS Collaboration

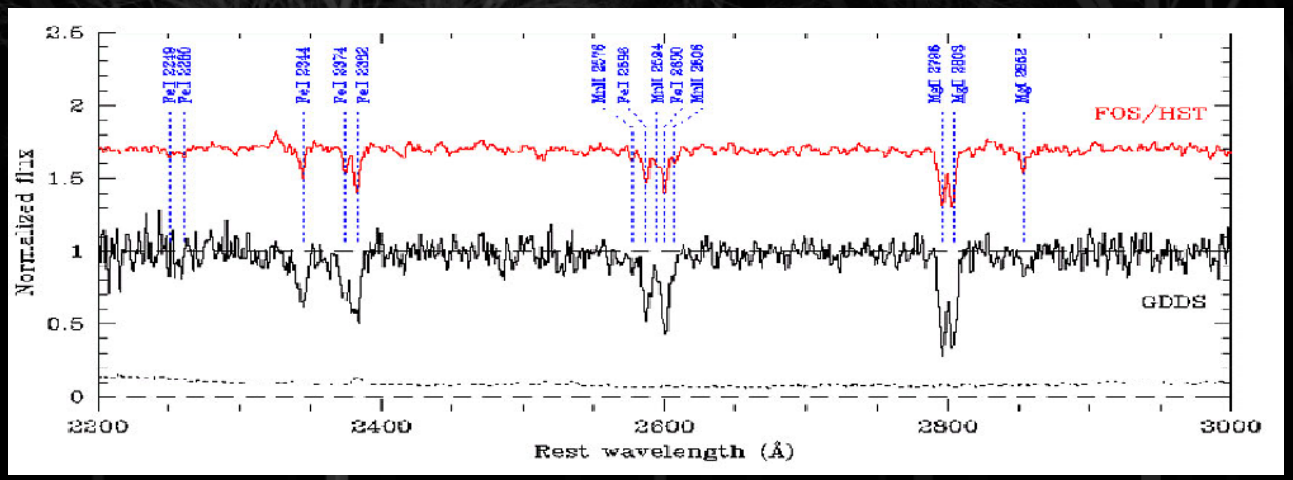
3D Clustering of galaxies as a probe in cosmology, e.g., 2 point correlation function, power spectrum of the galaxy distribution, etc.

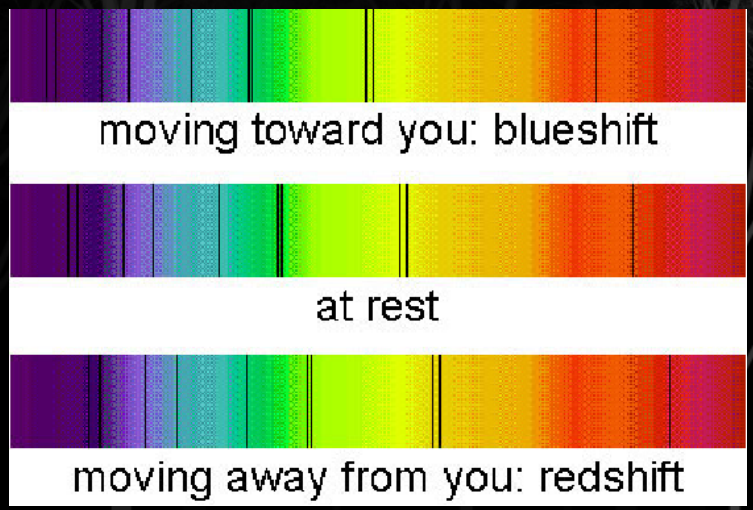
Galaxy redshifts



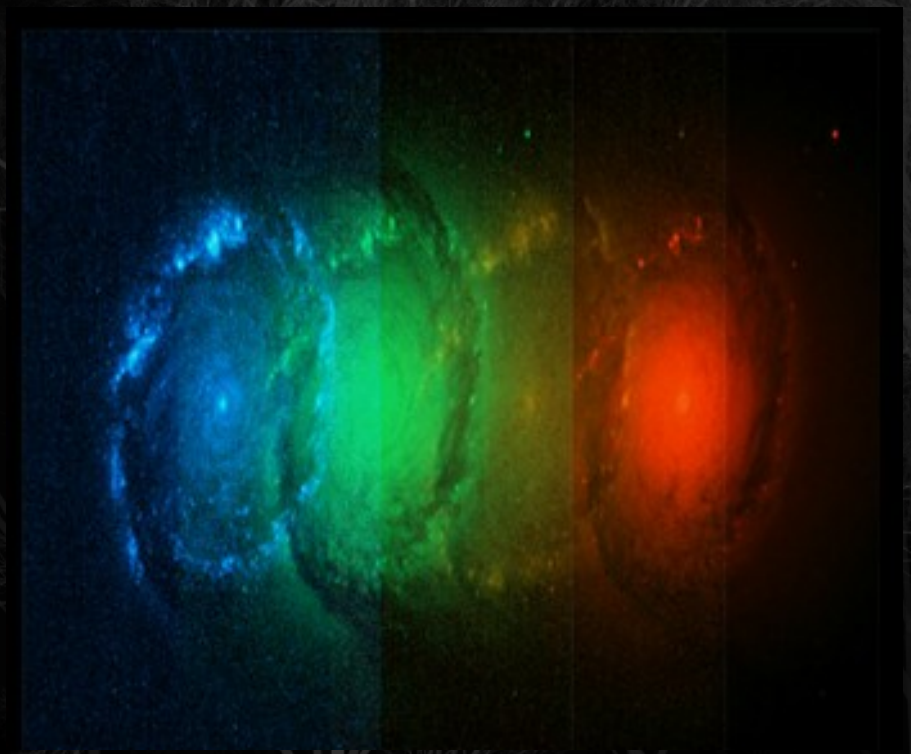
Galaxies are moving away from us, we can measure their velocities and distances by looking at the shifted lines in the spectrum

$$z = \frac{\lambda_{obs} - \lambda_{emit}}{\lambda_{emit}}$$

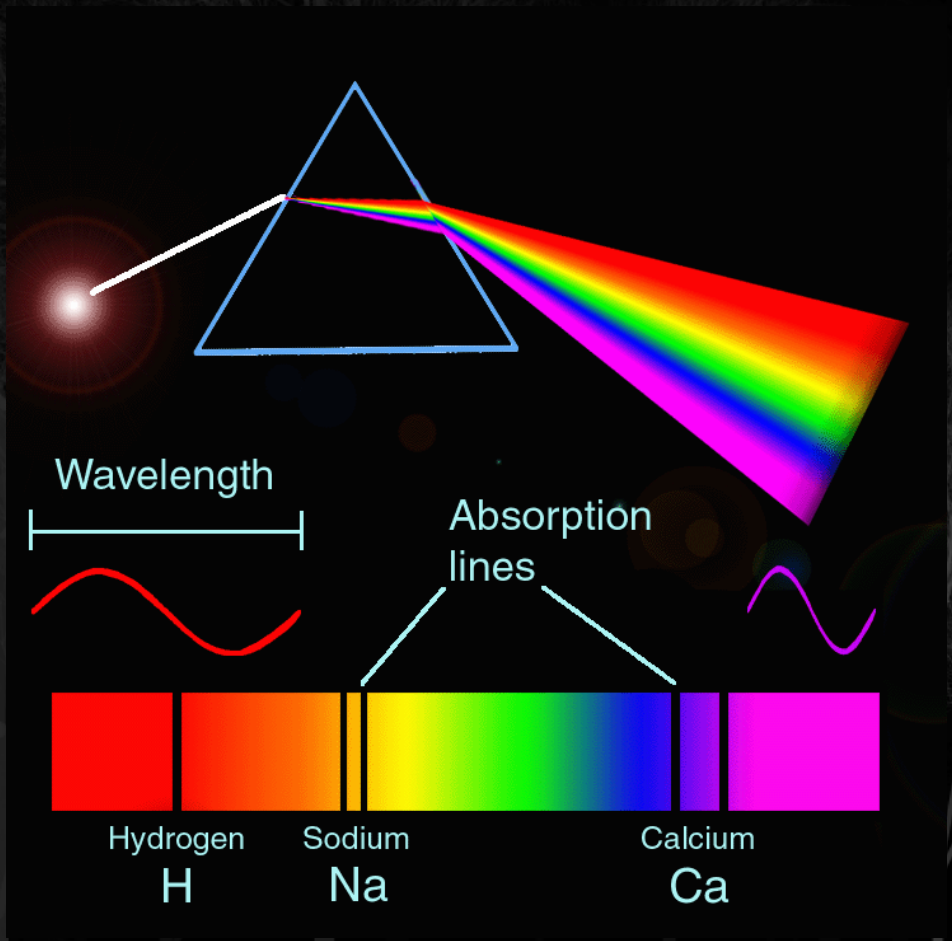




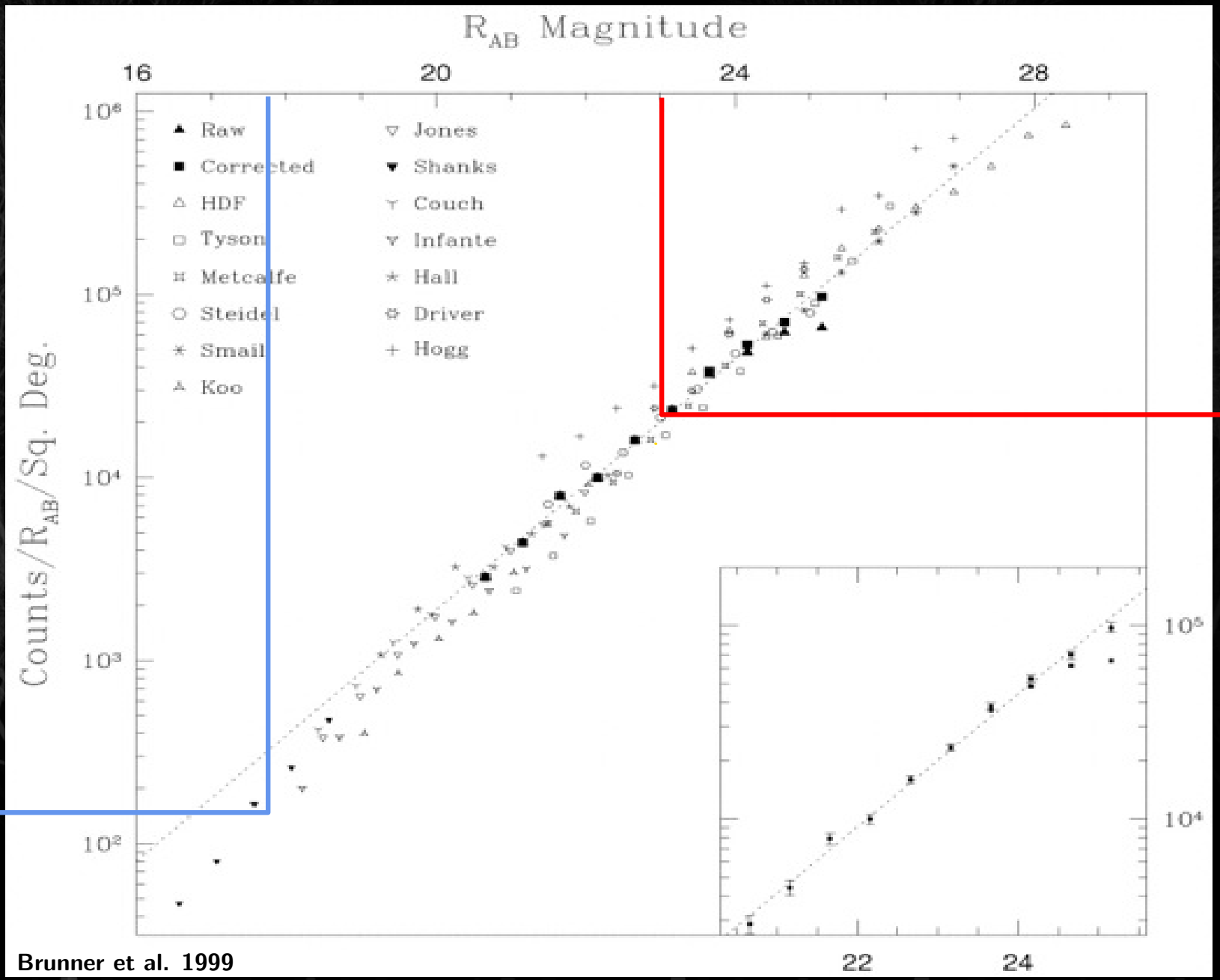
Photometry vs. Spectroscopy



VS.



Photometric surveys

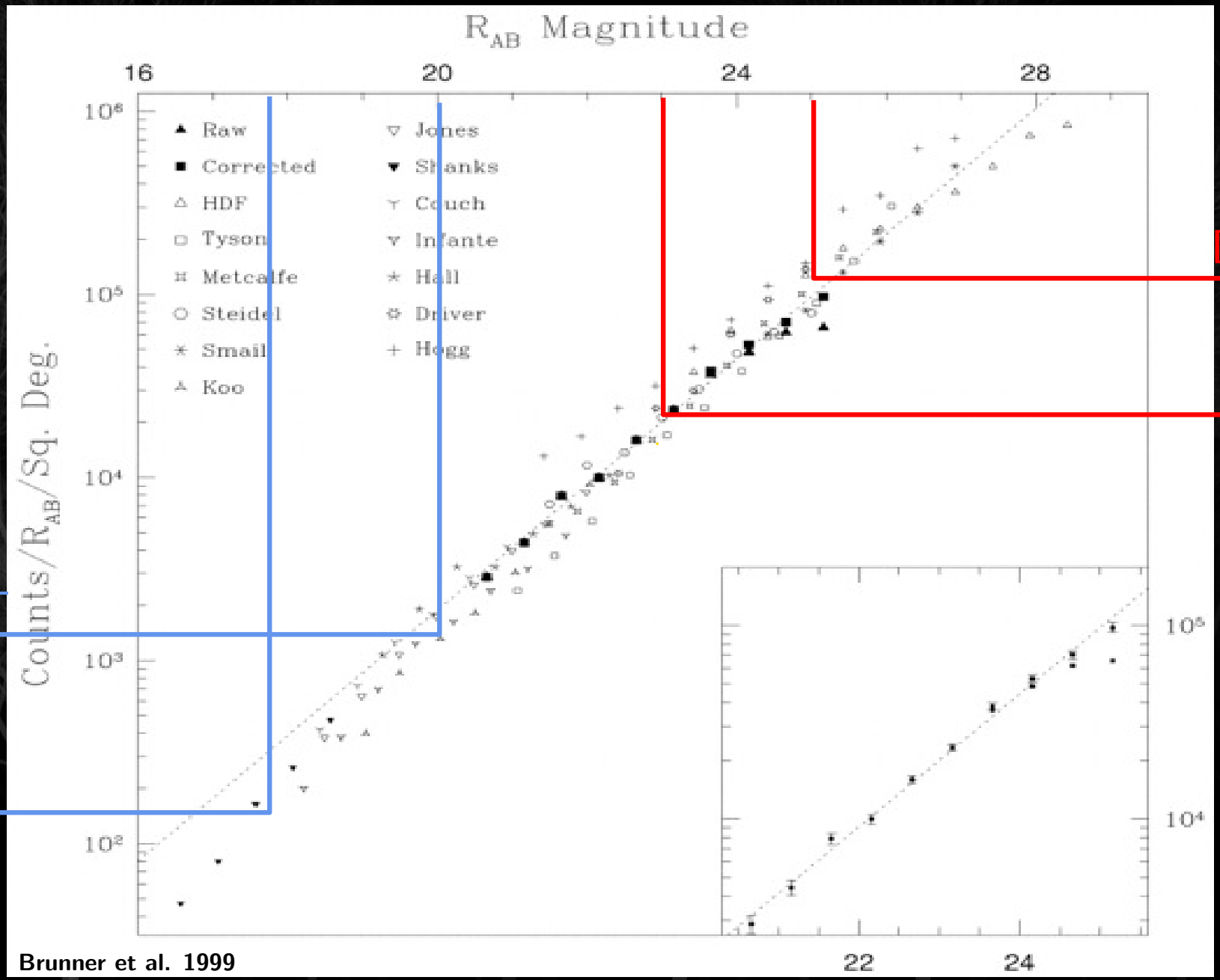


SDSS

SDSS

Brunner et al. 1999

Photometric surveys



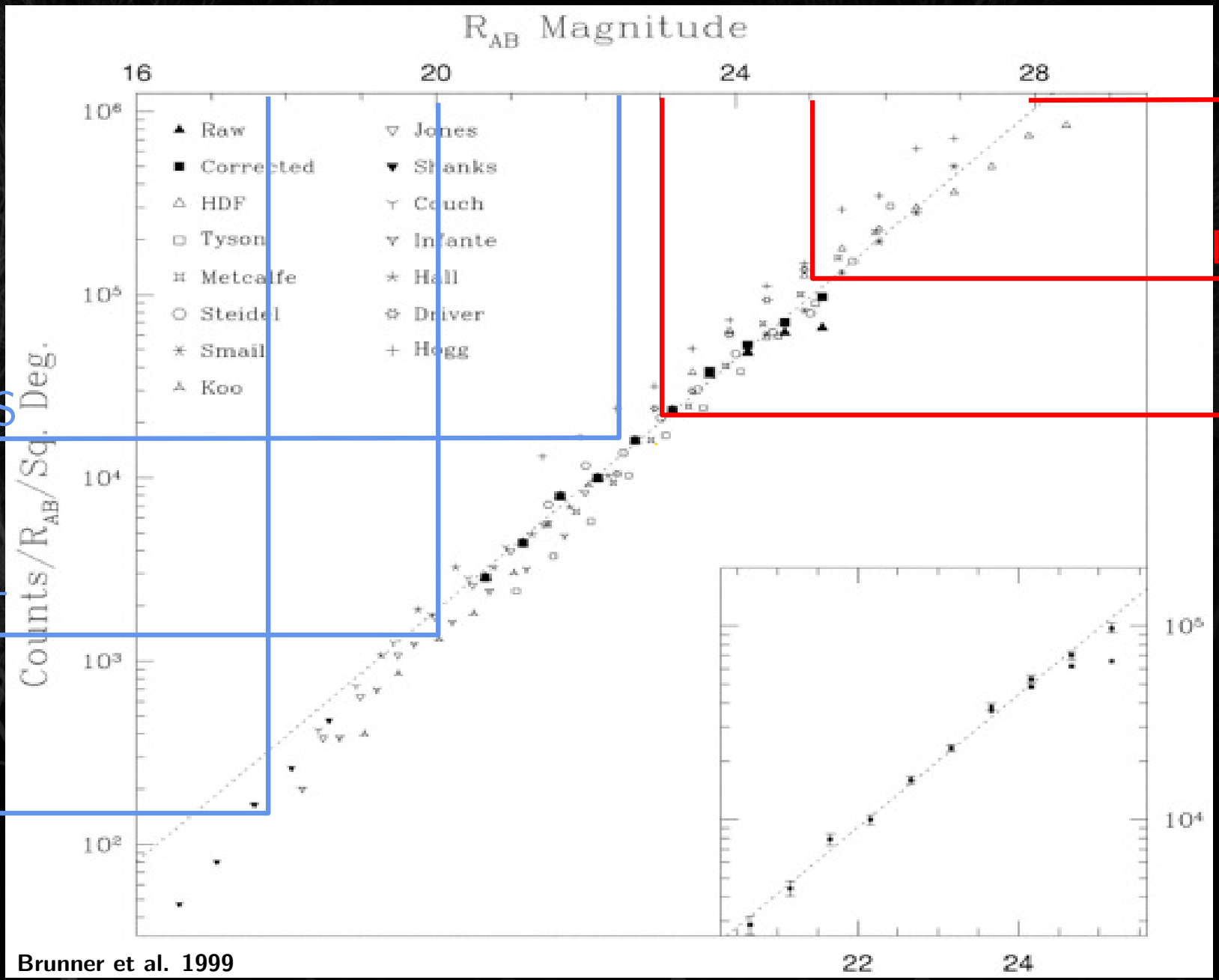
LAMOST

SDSS

DES final

SDSS

Photometric surveys



LSST

DES final

SDSS

Big BOSS

LAMOST

SDSS

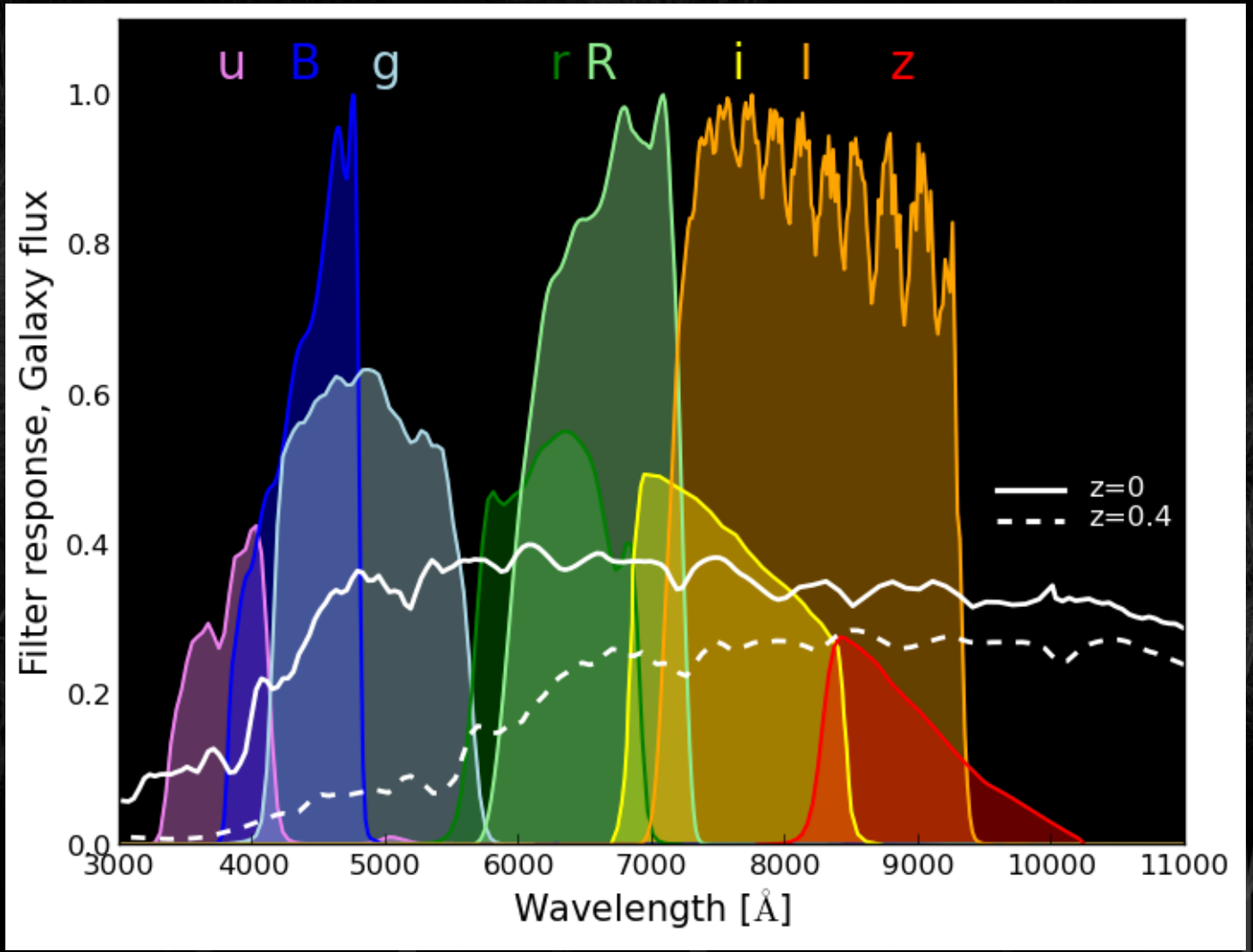
Photometric redshift (photo-z)

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



Motivation

- Photo-z Probability Density Functions needed
- Several methods/codes to compute photo-z
- Need for a meta-algorithm that combines multiple techniques
- PDFs are good **but** for large datasets, storage and I/O will be an issue
- Machine Learning and statistical tools

Motivation

- Photo-z Probability Density Functions needed
- Several methods/codes to compute photo-z
- Need for a meta-algorithm that combines multiple techniques
- PDFs are good **but** for large datasets, storage and I/O will be an issue
- Machine Learning and statistical tools

Motivation

- Photo-z Probability Density Functions needed
- Several methods/codes to compute photo-z
- Need for a meta-algorithm that combines multiple techniques
- PDFs are good **but** for large datasets, storage and I/O will be an issue
- Machine Learning and statistical tools

Motivation

- Photo-z Probability Density Functions needed
- 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 Probability Density Functions needed
- 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

Data from Surveys in this Thesis

SDSS III

CFHTLenS

LSST
Large Synoptic Survey Telescope

DEEP2
(Davis et al. 2003, Newman et al. 2013)

THE DARK ENERGY SURVEY

PHAT
(Hildebrandt et al. 2012)

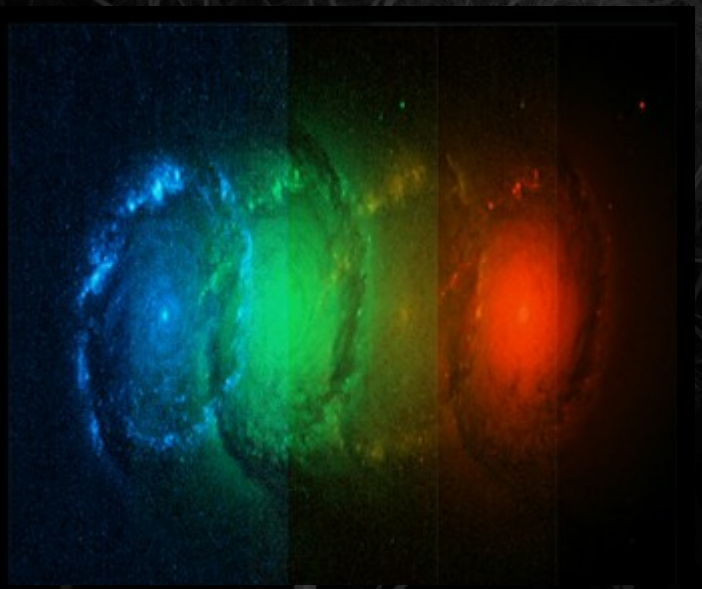
Outline

- INTRODUCTION
- HOW TO COMPUTE PHOTO-Z PDFs
- HOW TO COMBINE PHOTO-Z PDFs METHODS
- HOW TO REPRESENT PHOTO-Z PDF
- APPLICATIONS
- CONCLUSIONS AND FUTURE WORK

Outline

- INTRODUCTION
- HOW TO COMPUTE PHOTO-Z PDFs
- HOW TO COMBINE PHOTO-Z PDFs METHODS
- HOW TO REPRESENT PHOTO-Z PDF
- APPLICATIONS
- CONCLUSIONS AND FUTURE WORK

Photo-z PDF estimation



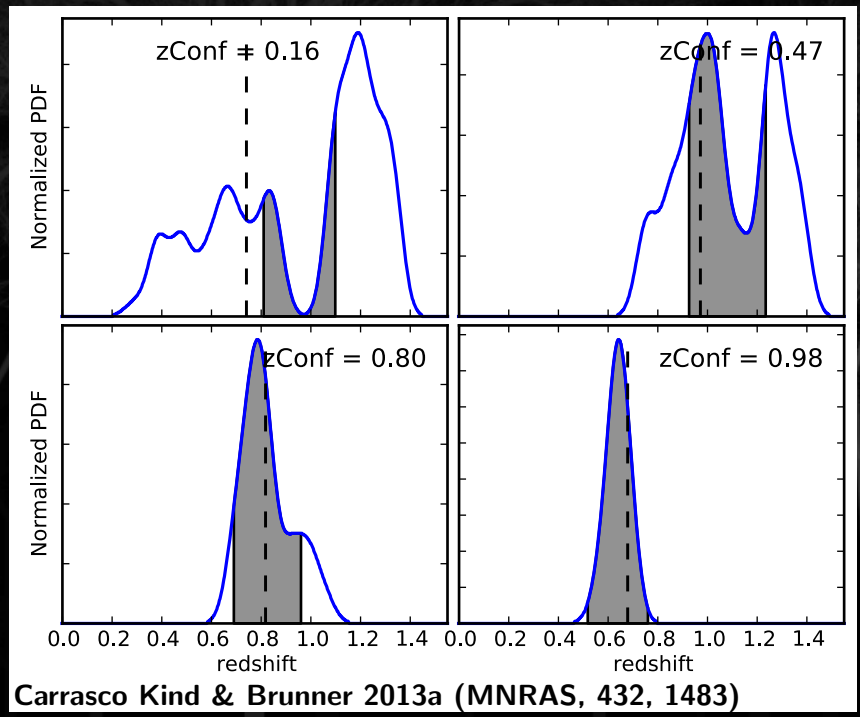
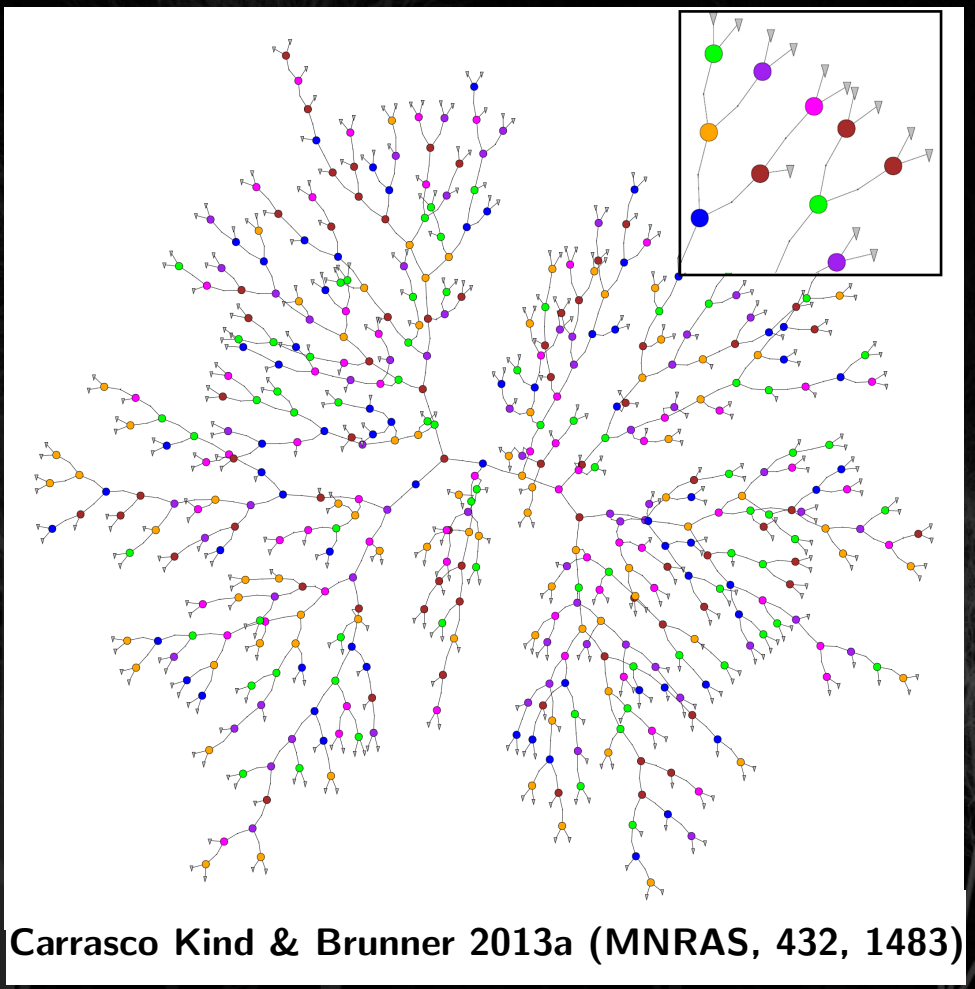


Photo-z PDF estimation: TPZ

- 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)

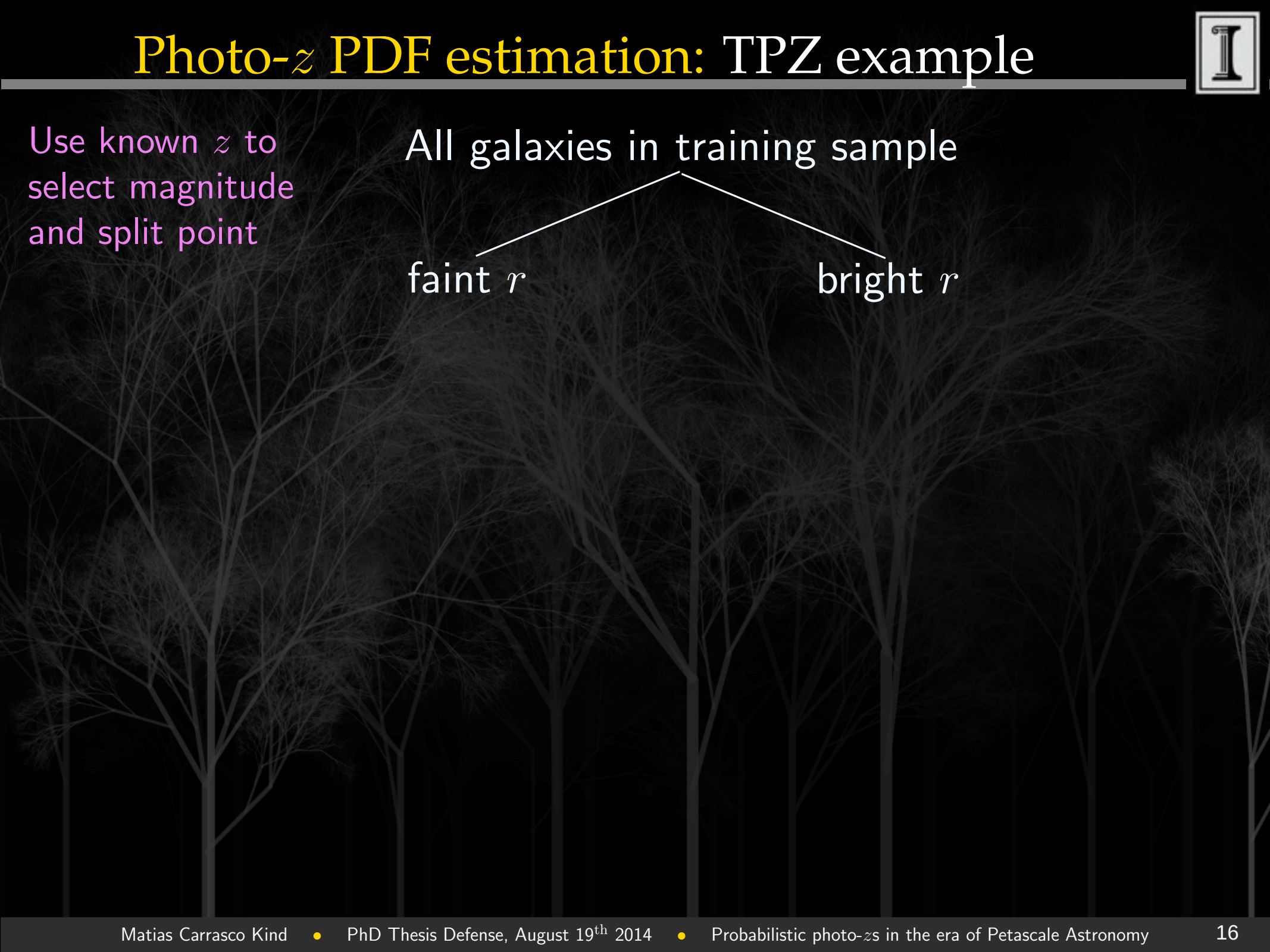
http://lcdm.astro.illinois.edu/code/mlz.html

Photo-z PDF estimation: TPZ example

Use known z to select magnitude and split point

All galaxies in training sample

Photo-z PDF estimation: TPZ example



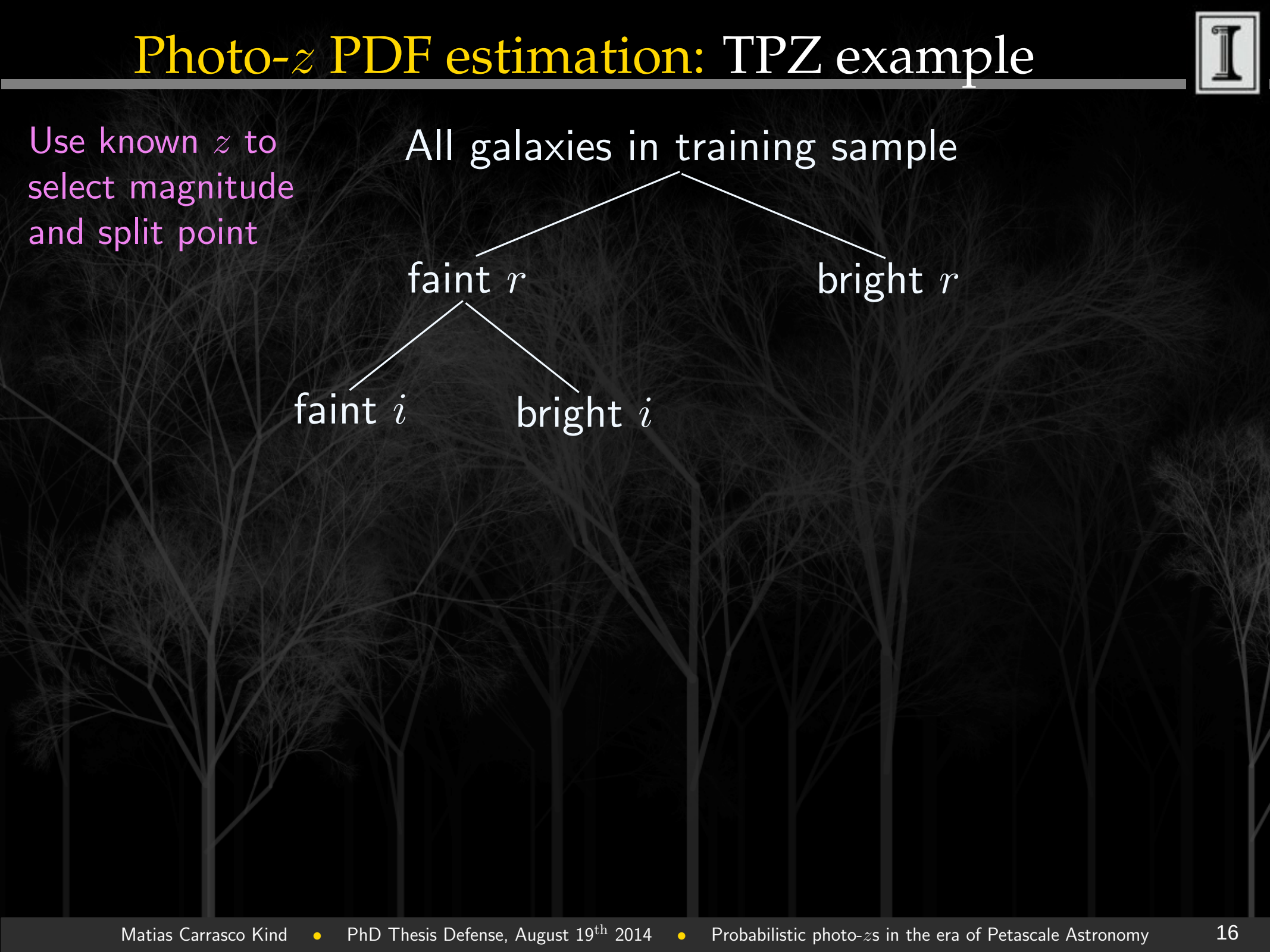
Use known z to select magnitude and split point

All galaxies in training sample

faint r

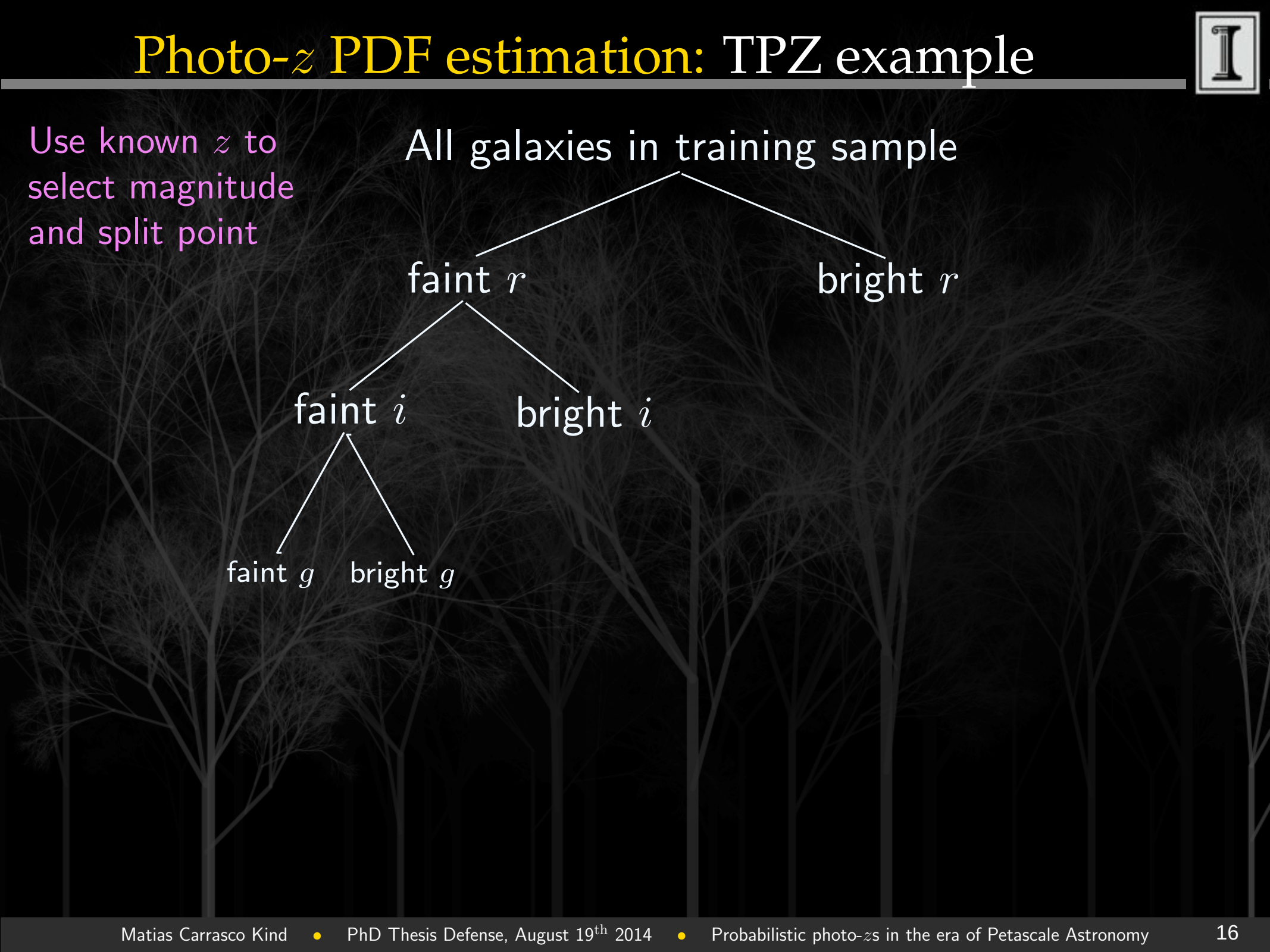
bright r

Photo-z PDF estimation: TPZ example



Use known z to select magnitude and split point

Photo-z PDF estimation: TPZ example



Use known z to select magnitude and split point

Photo-z PDF estimation: TPZ example

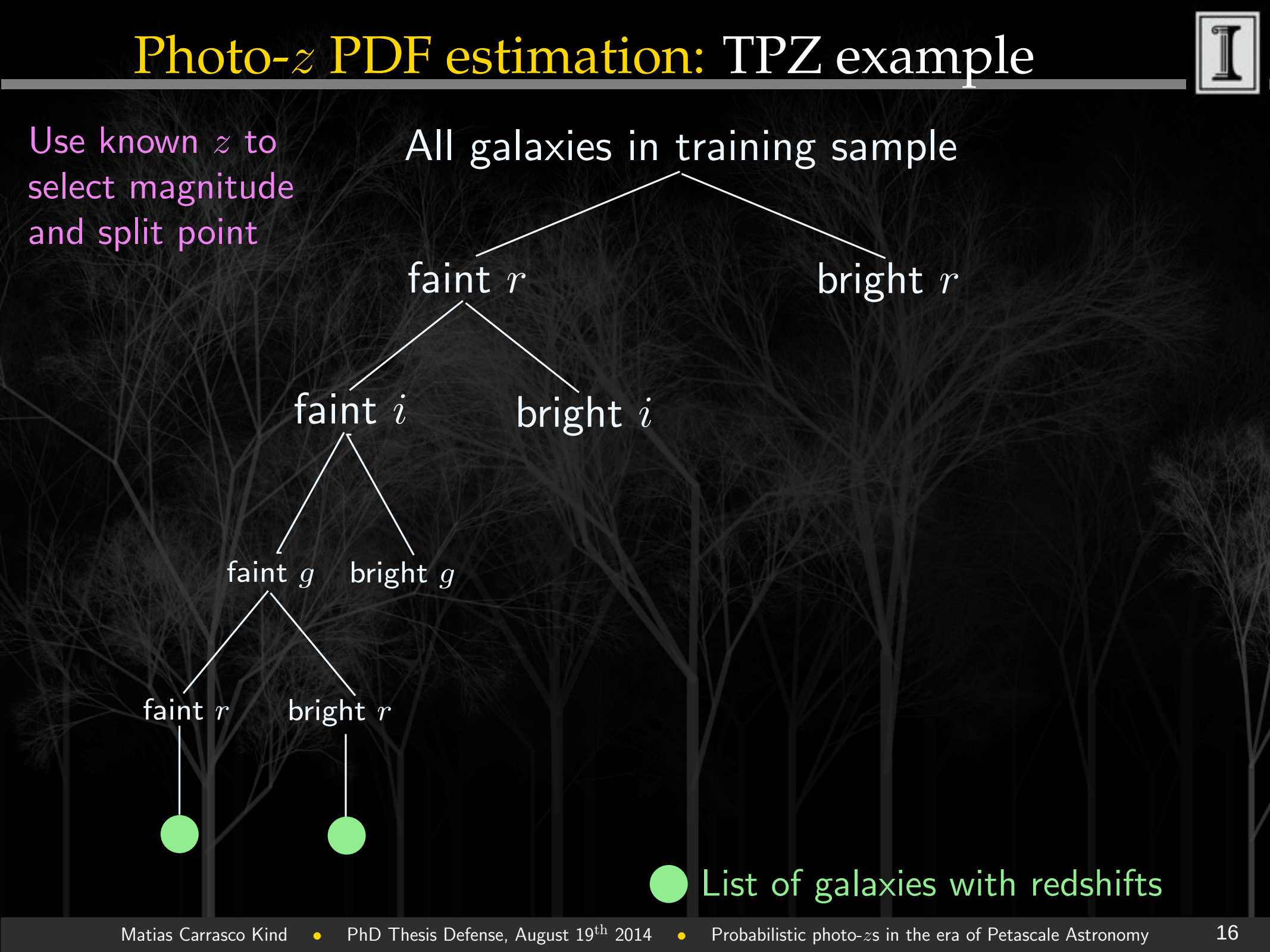


Photo-z PDF estimation: TPZ example

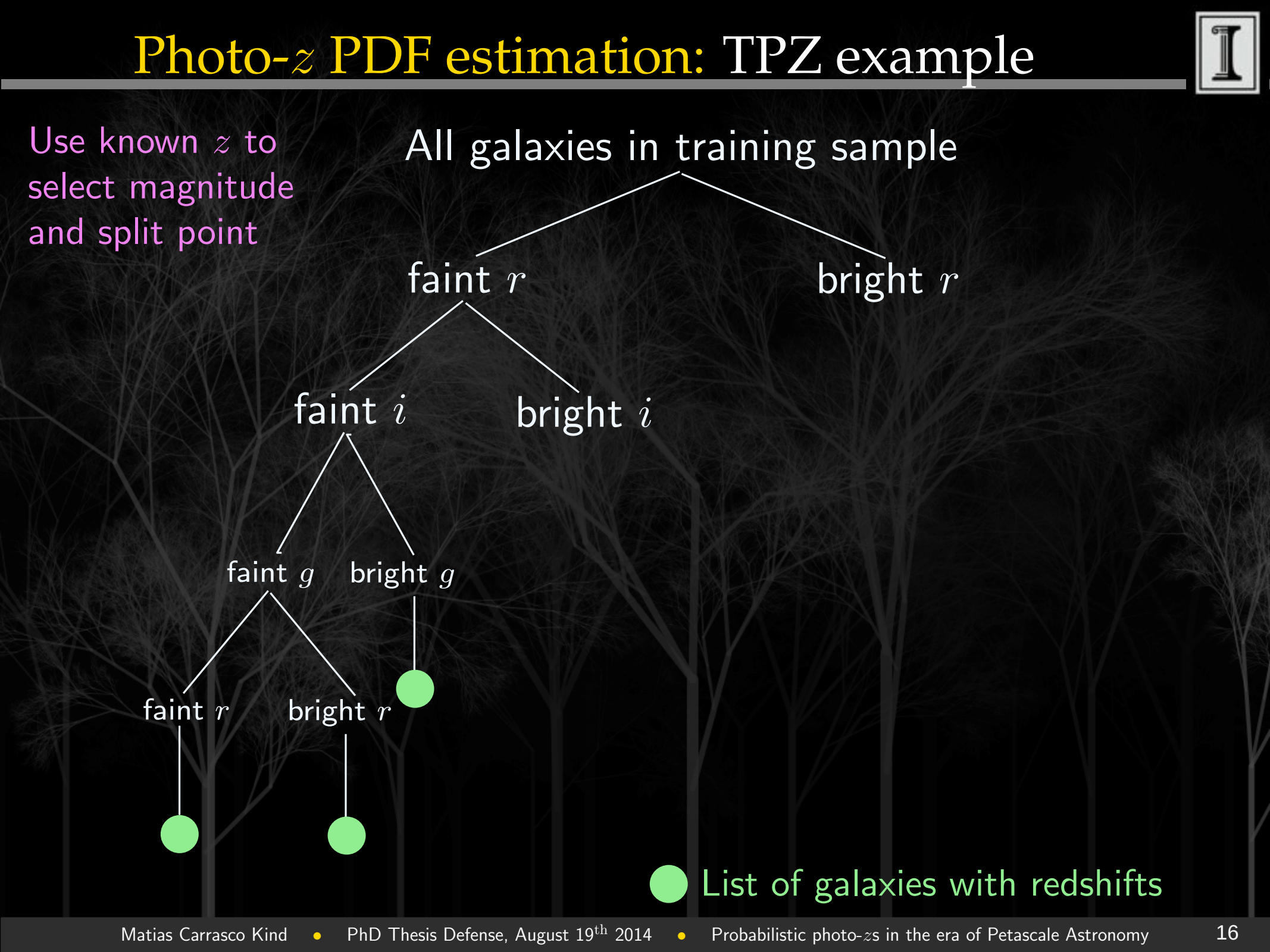


Photo-z PDF estimation: TPZ example

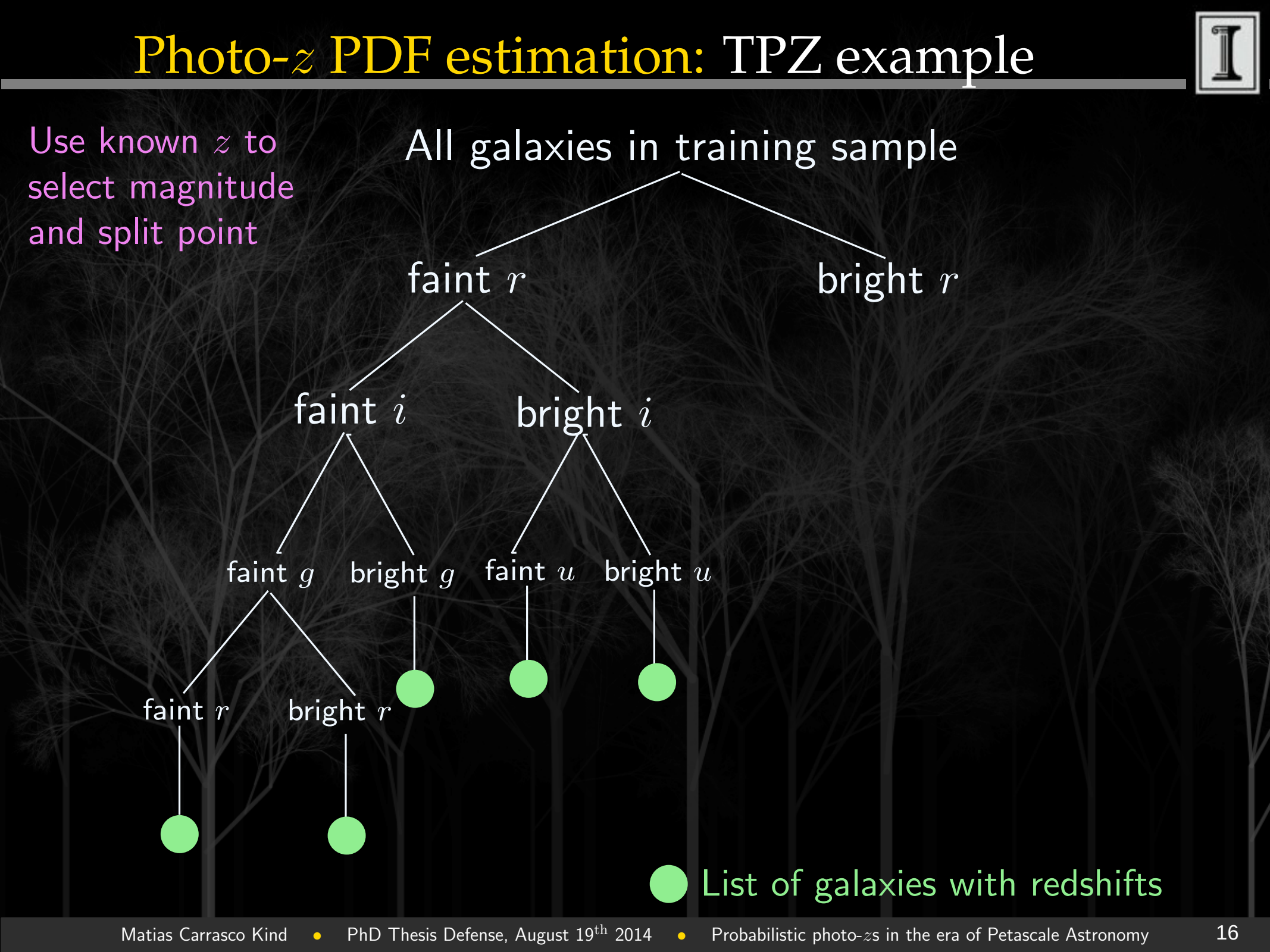


Photo-z PDF estimation: TPZ example

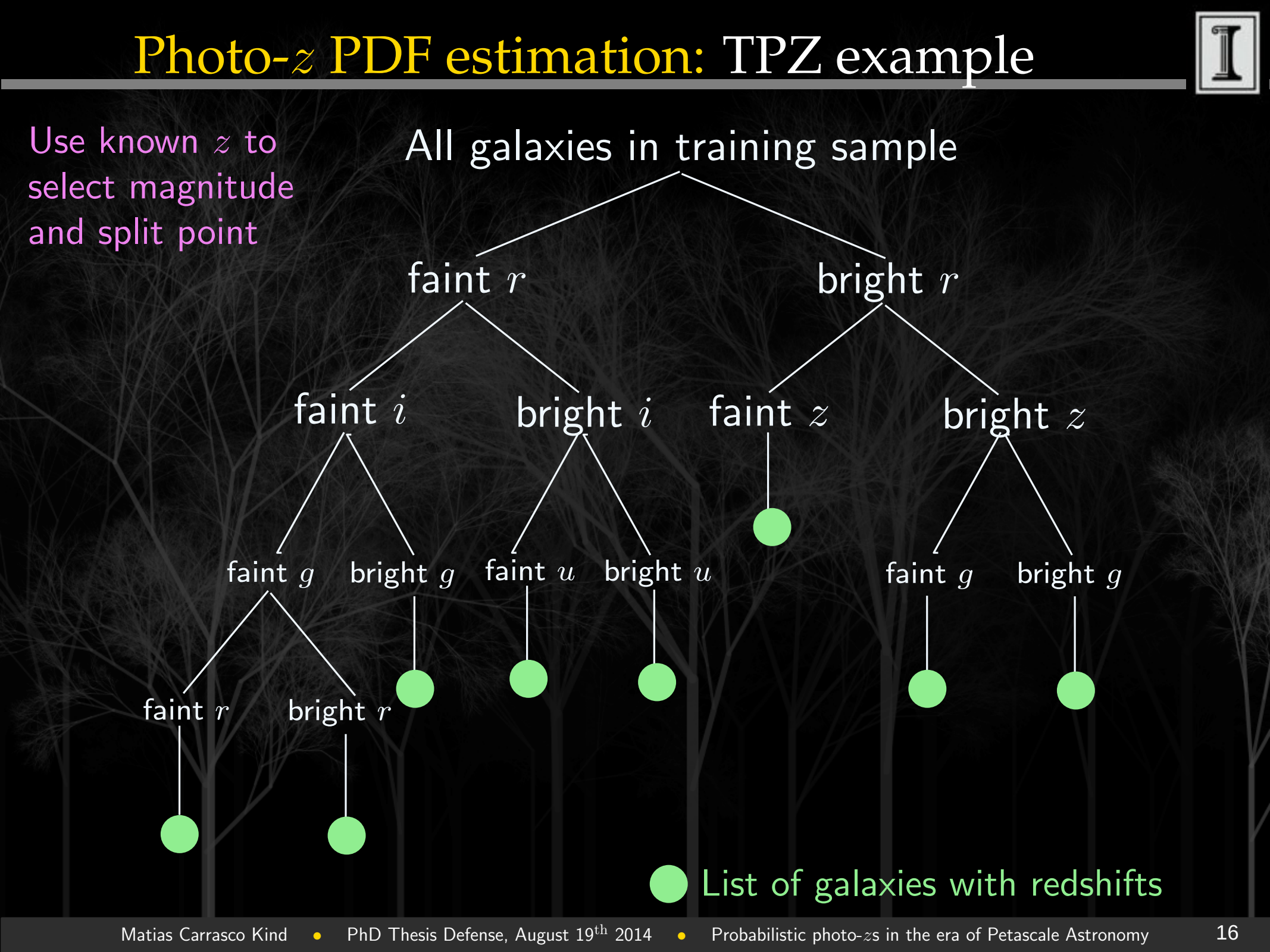


Photo-z PDF estimation: TPZ example

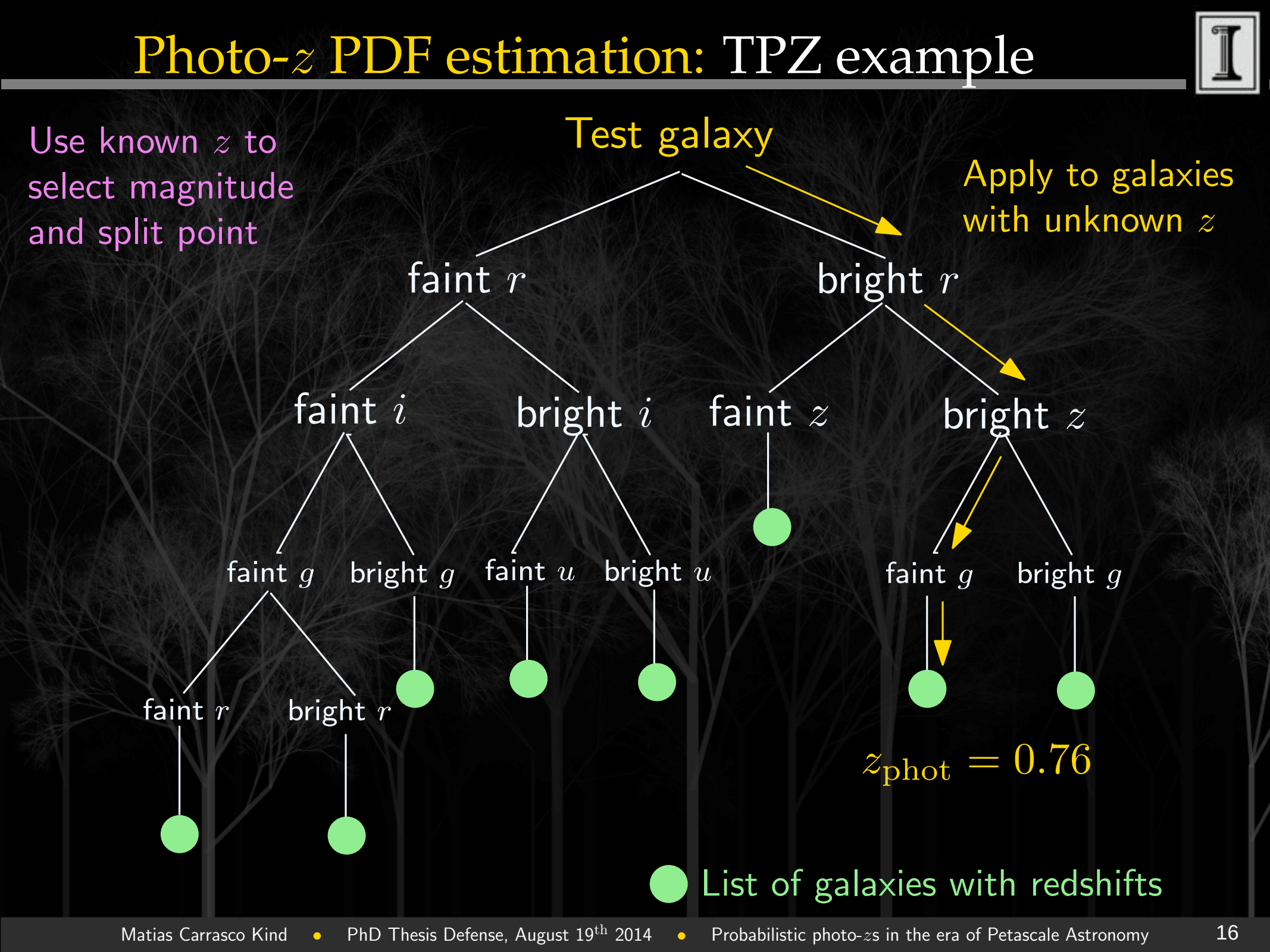
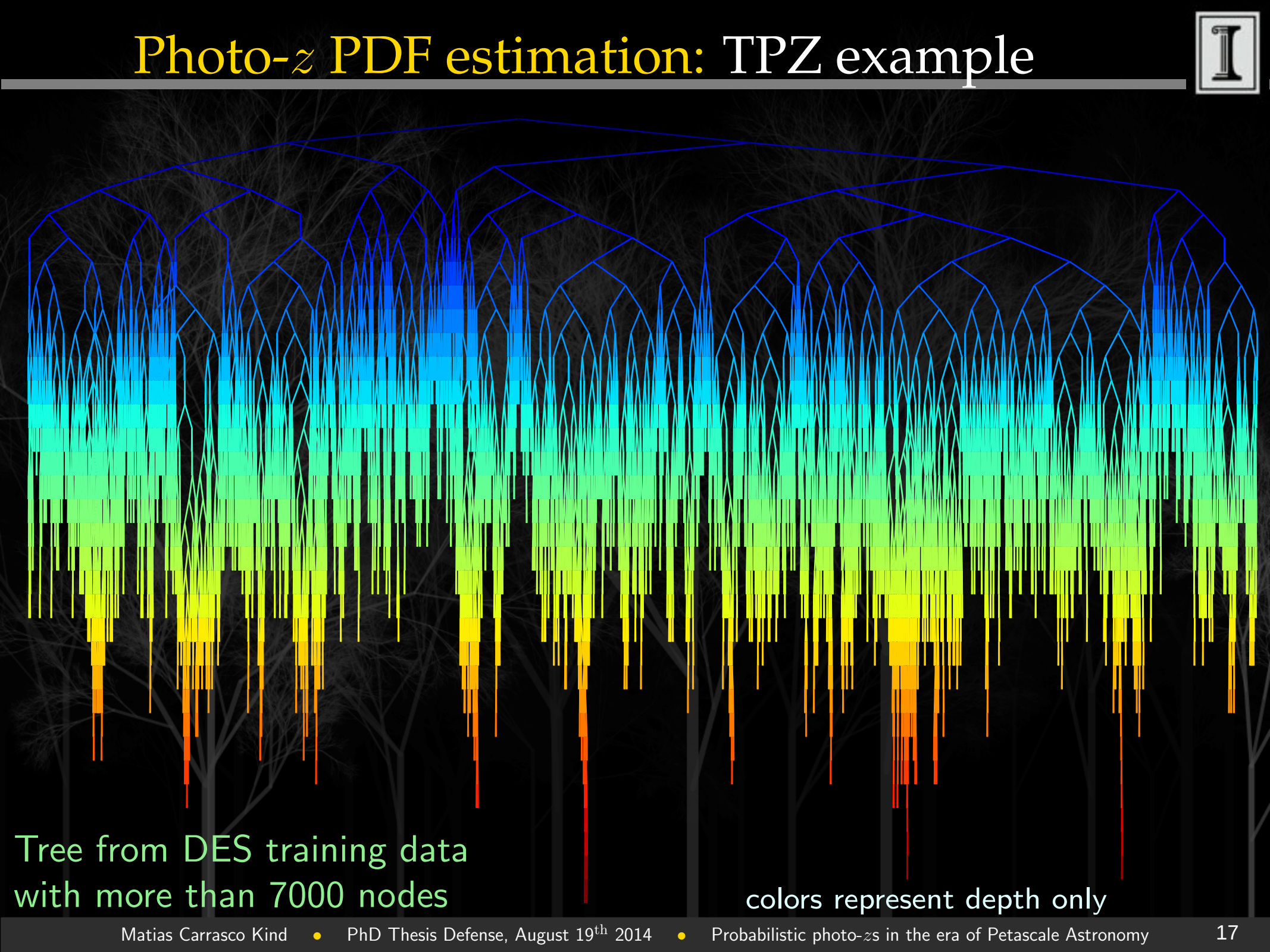


Photo-z PDF estimation: TPZ example



Tree from DES training data
with more than 7000 nodes

colors represent depth only

Photo-z PDF estimation: TPZ Random forest

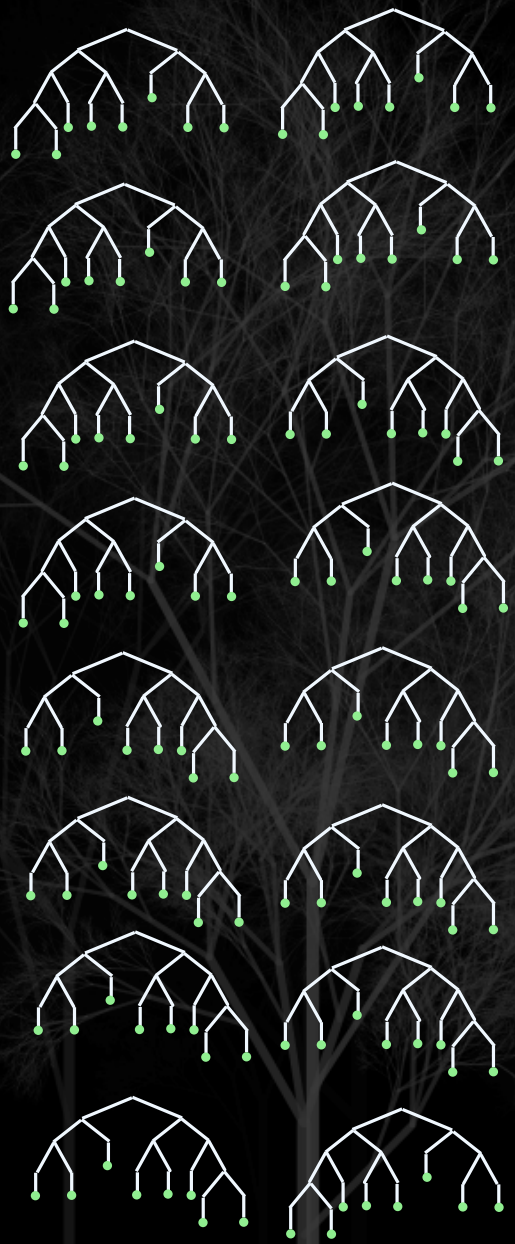
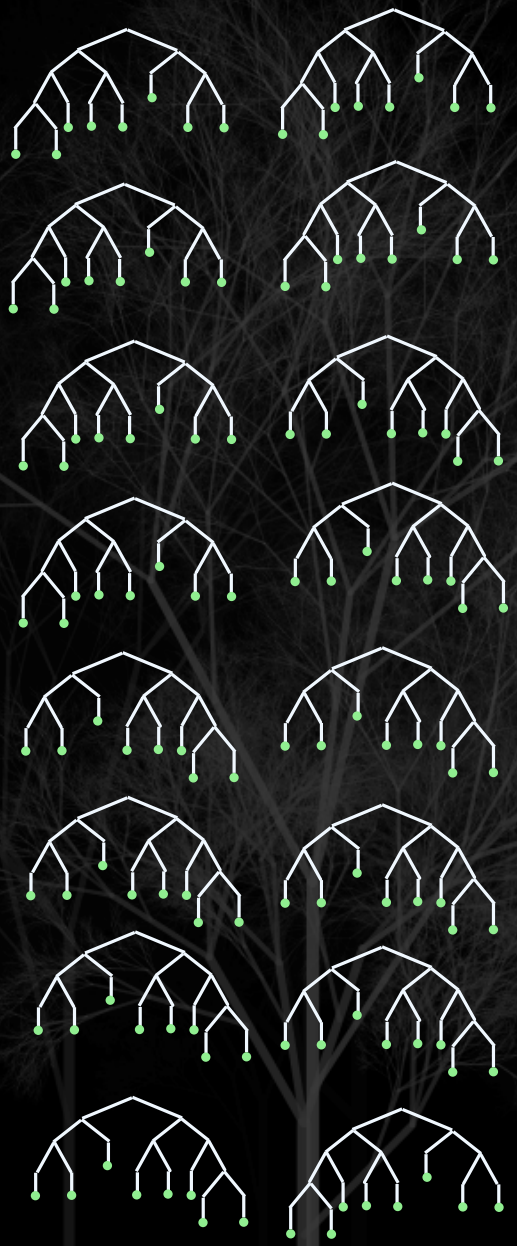
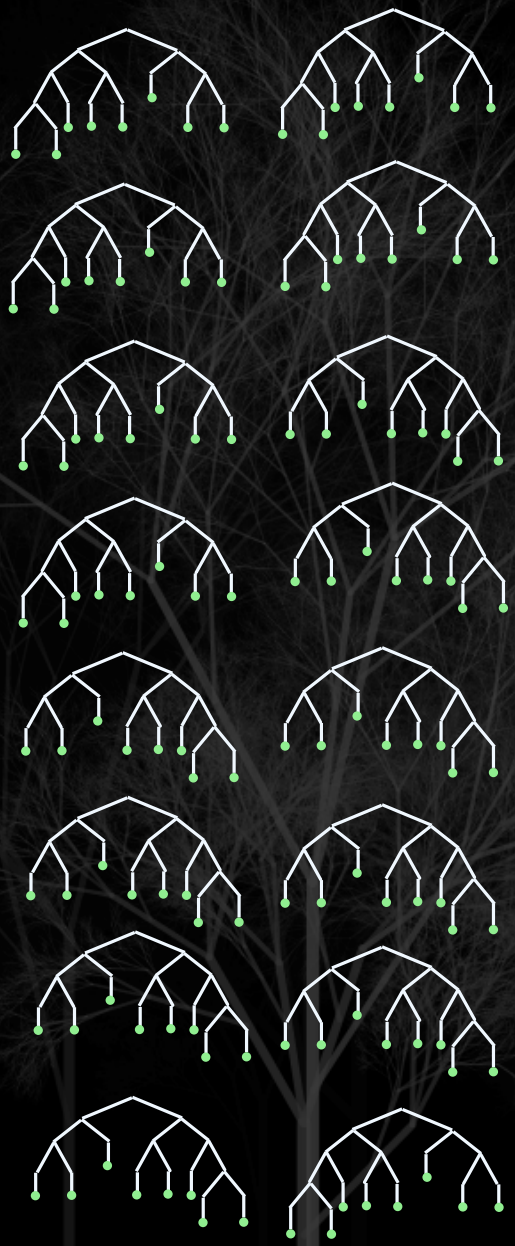


Photo-z PDF estimation: TPZ Random forest



Combine predictions from trees

Photo-z PDF estimation: TPZ Random forest



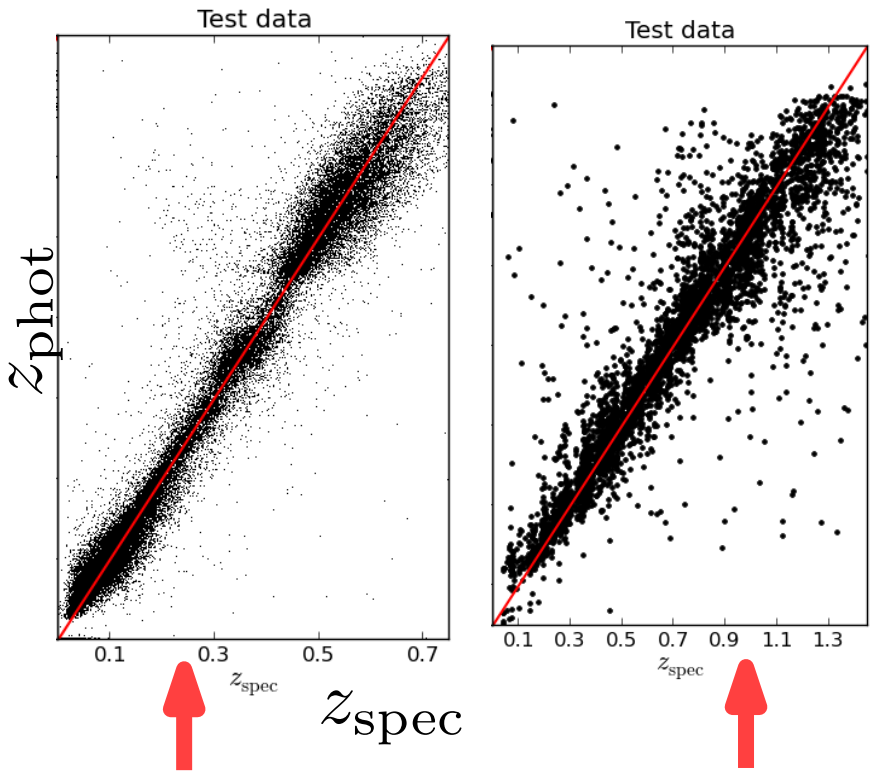
Combine predictions from trees

Trees are ideally uncorrelated and strong

Bootstrapping and error sampling

Random features at each node

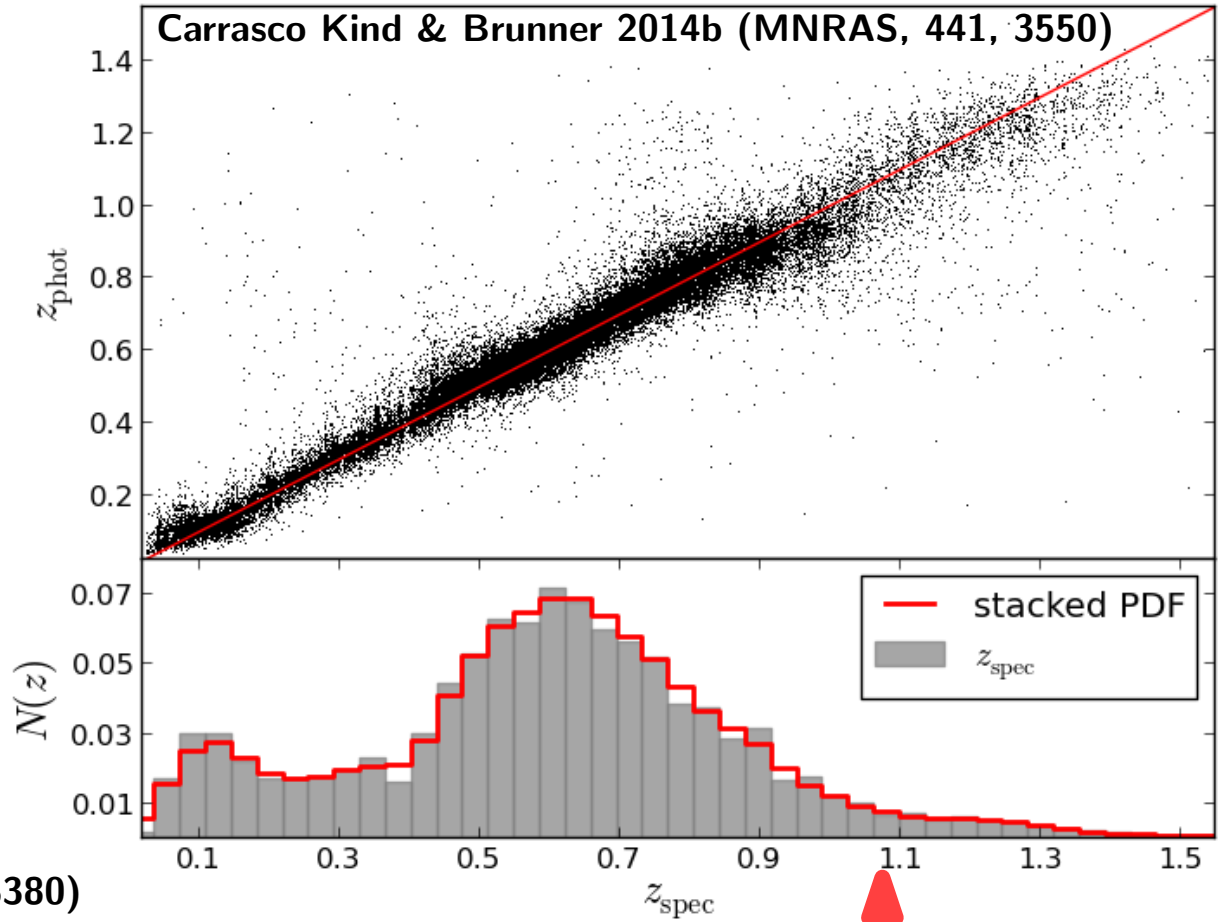
Photo-z PDF estimation: TPZ applications



SDSS DR10

DEEP2

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

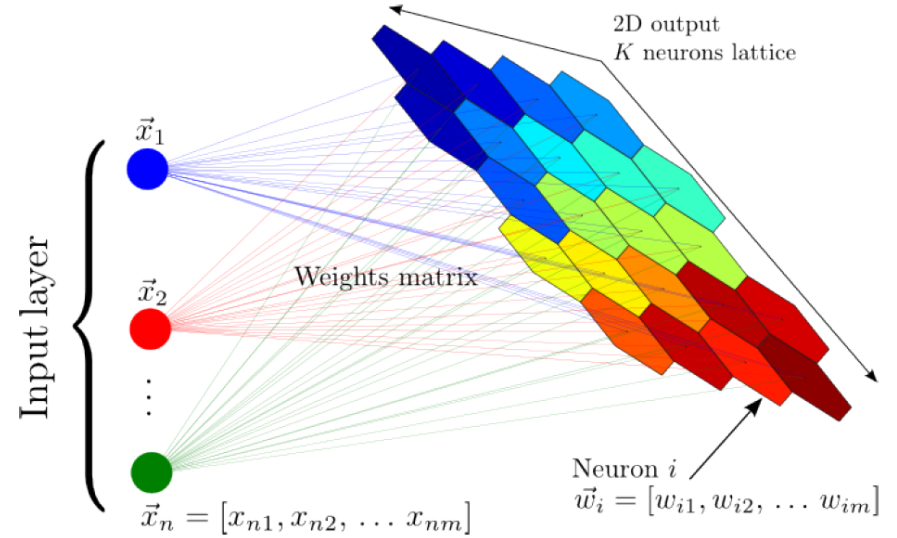


CFHTLenS

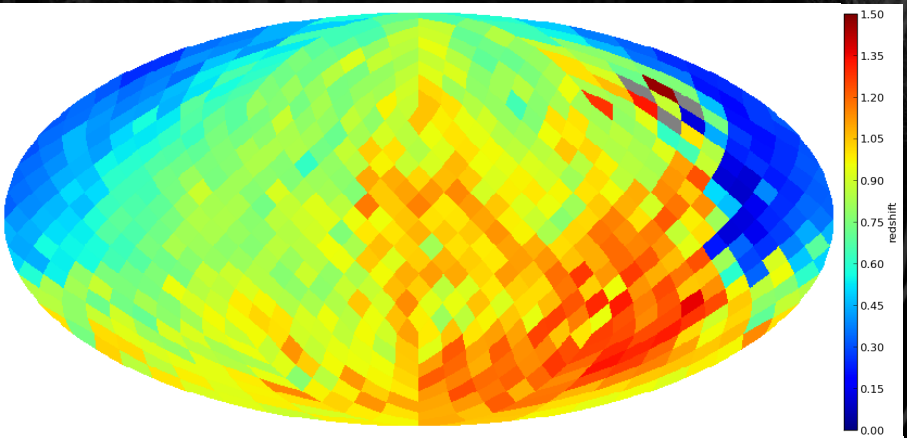
TPZ has been tested in several databases with remarkable results

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



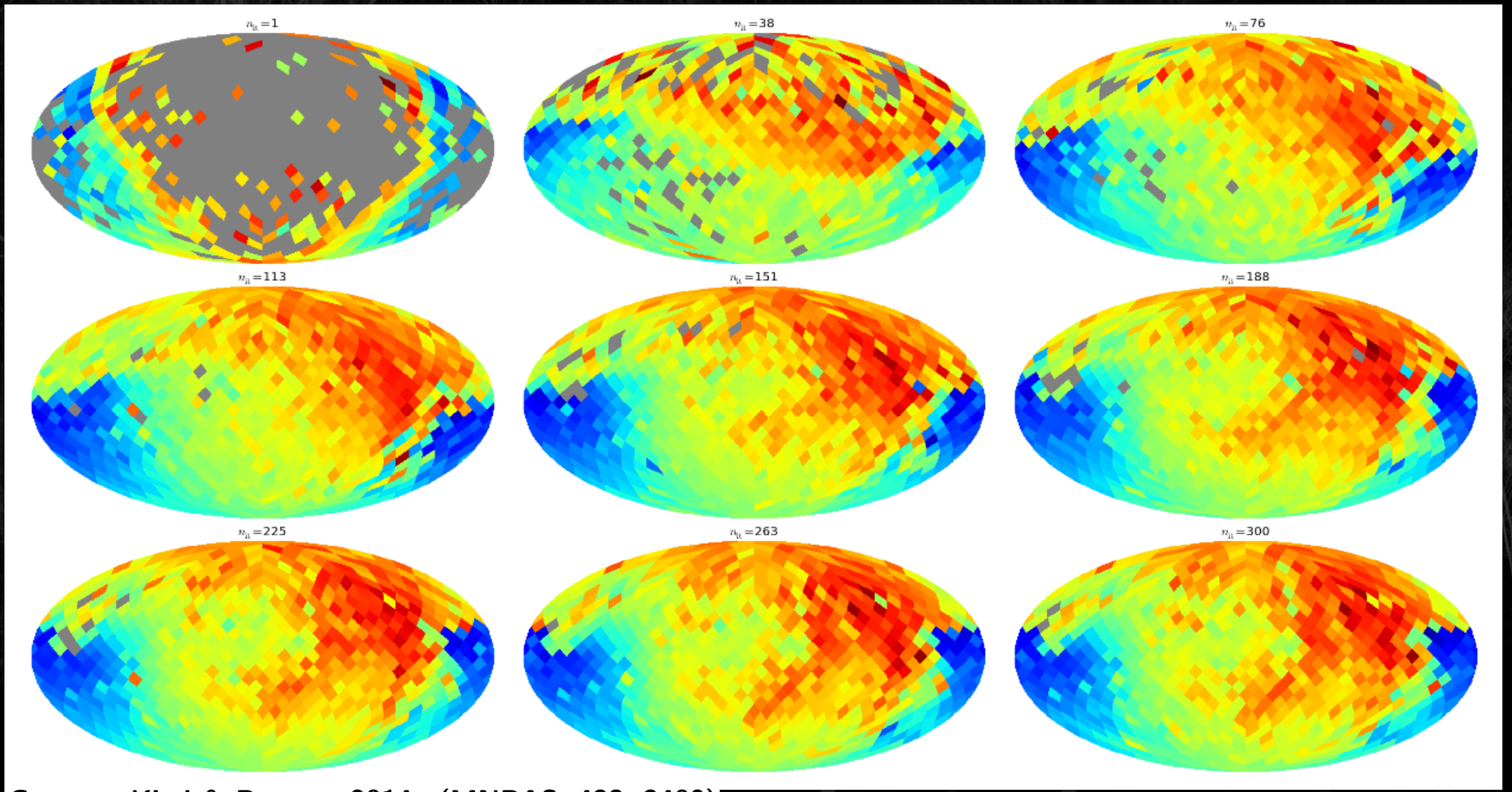
Carrasco Kind & Brunner 2014a (MNRAS, 438, 3409)



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

Photo-z PDF estimation: SOM

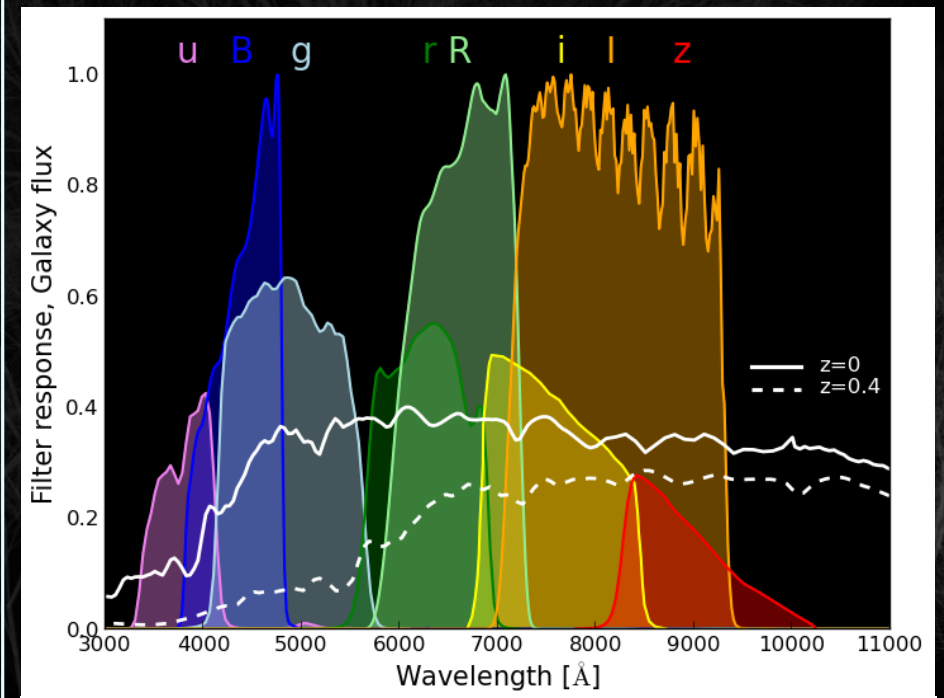
Self organized map construction, colors indicate median redshift of each cell



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



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

Photo-z PDF estimation: BPZ Templates

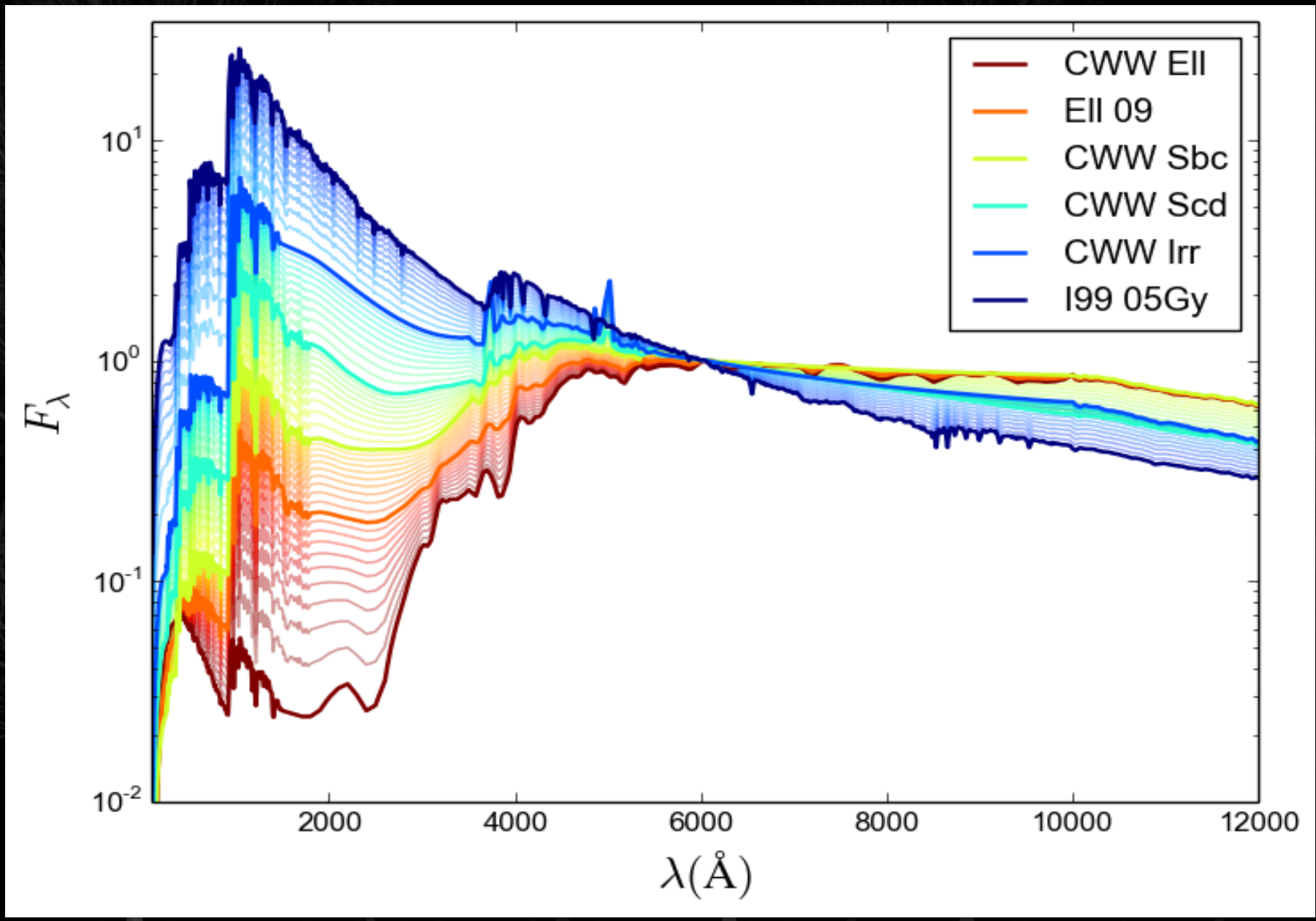


Photo-z PDF estimation: BPZ

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

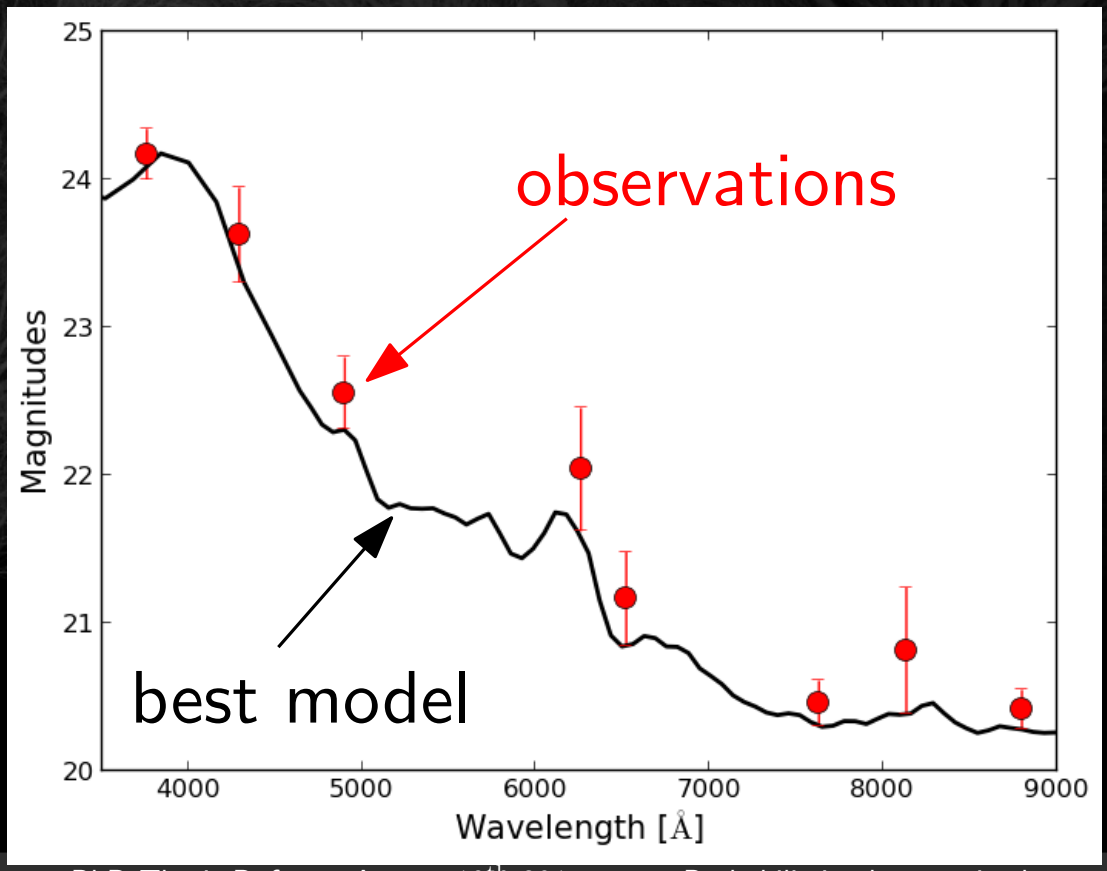
Suppose a set of templates T and n magnitudes $m_1, m_2, ..., m_n$, the probability is:

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

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

Prior

Likelihood



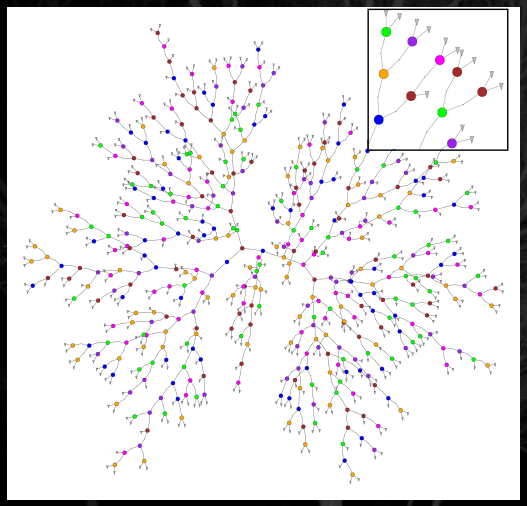
Outline

- INTRODUCTION
- HOW TO COMPUTE PHOTO-Z PDFs
- HOW TO COMBINE PHOTO-Z PDFs METHODS
- HOW TO REPRESENT PHOTO-Z PDF
- APPLICATIONS
- CONCLUSIONS AND FUTURE WORK

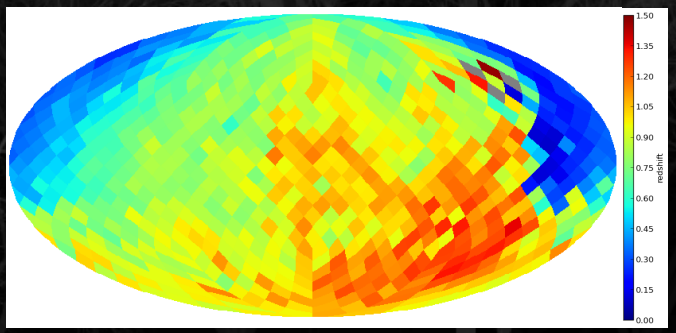
Outline

- INTRODUCTION
- HOW TO COMPUTE PHOTO-Z PDFs
- HOW TO COMBINE PHOTO-Z PDFs METHODS
- HOW TO REPRESENT PHOTO-Z PDF
- APPLICATIONS
- CONCLUSIONS AND FUTURE WORK

Photo-z PDF combination



+



+

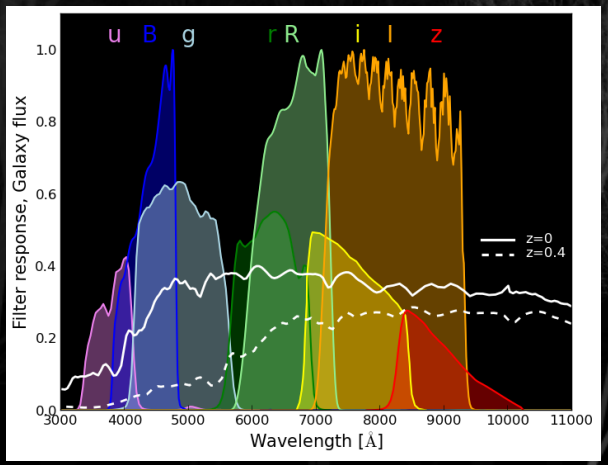


Photo-z PDF combination: Bayesian framework

Motivation

- Different methods have different strengths and weaknesses
- Not previous work on combined techniques
- Extract all possible information from data
- Bayesian framework for combination models
- Easy to incorporate other techniques

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:

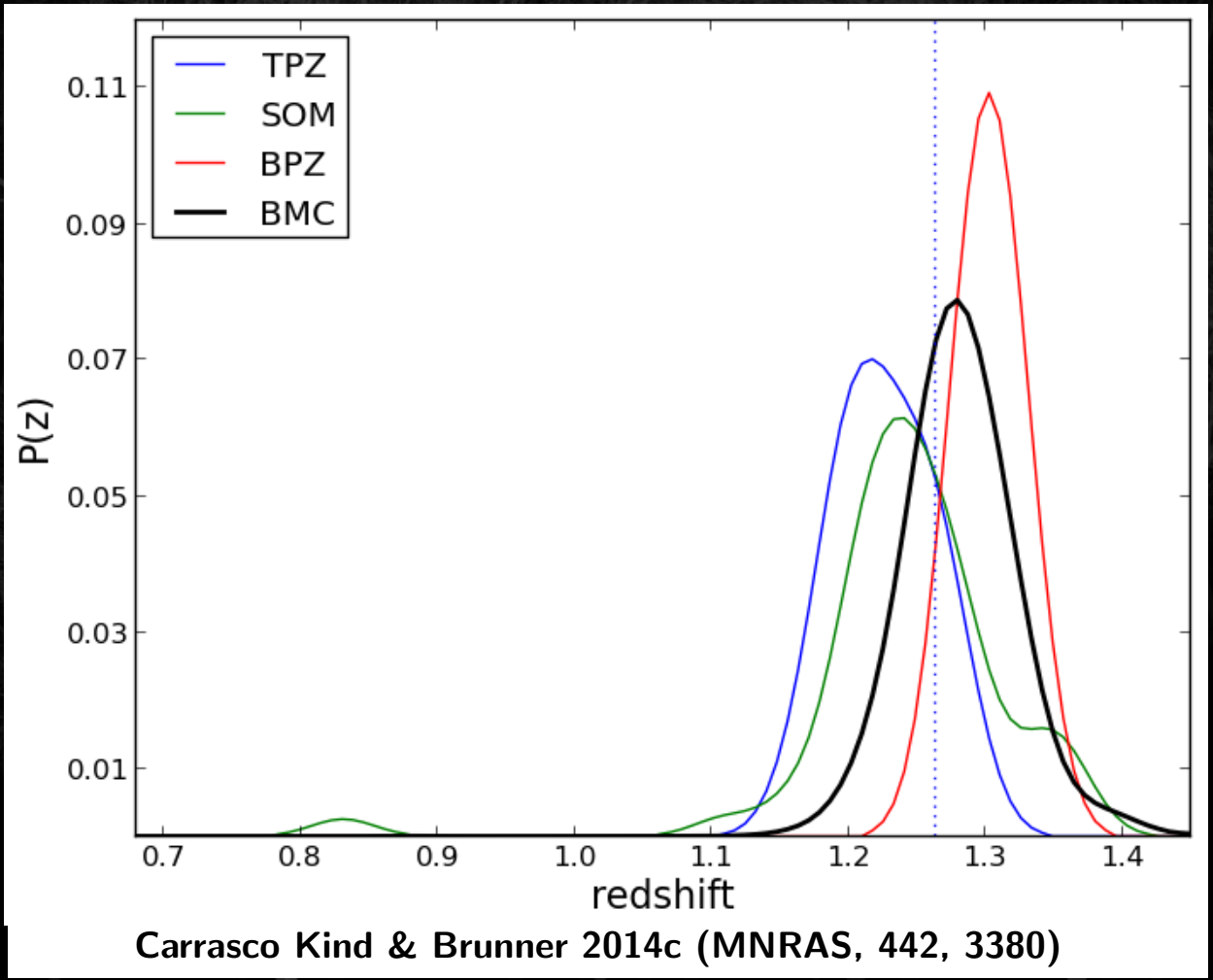
$$P(z \mid \mathbf{x}, \mathbf{D}, \mathbf{M}) \propto \sum_k P(z \mid \mathbf{x}, M_k) P(M_k) \times (1-\epsilon_k)^{N_d - N_k^{(b)}} (\epsilon_k)^{N_k^{(b)}}$$

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)

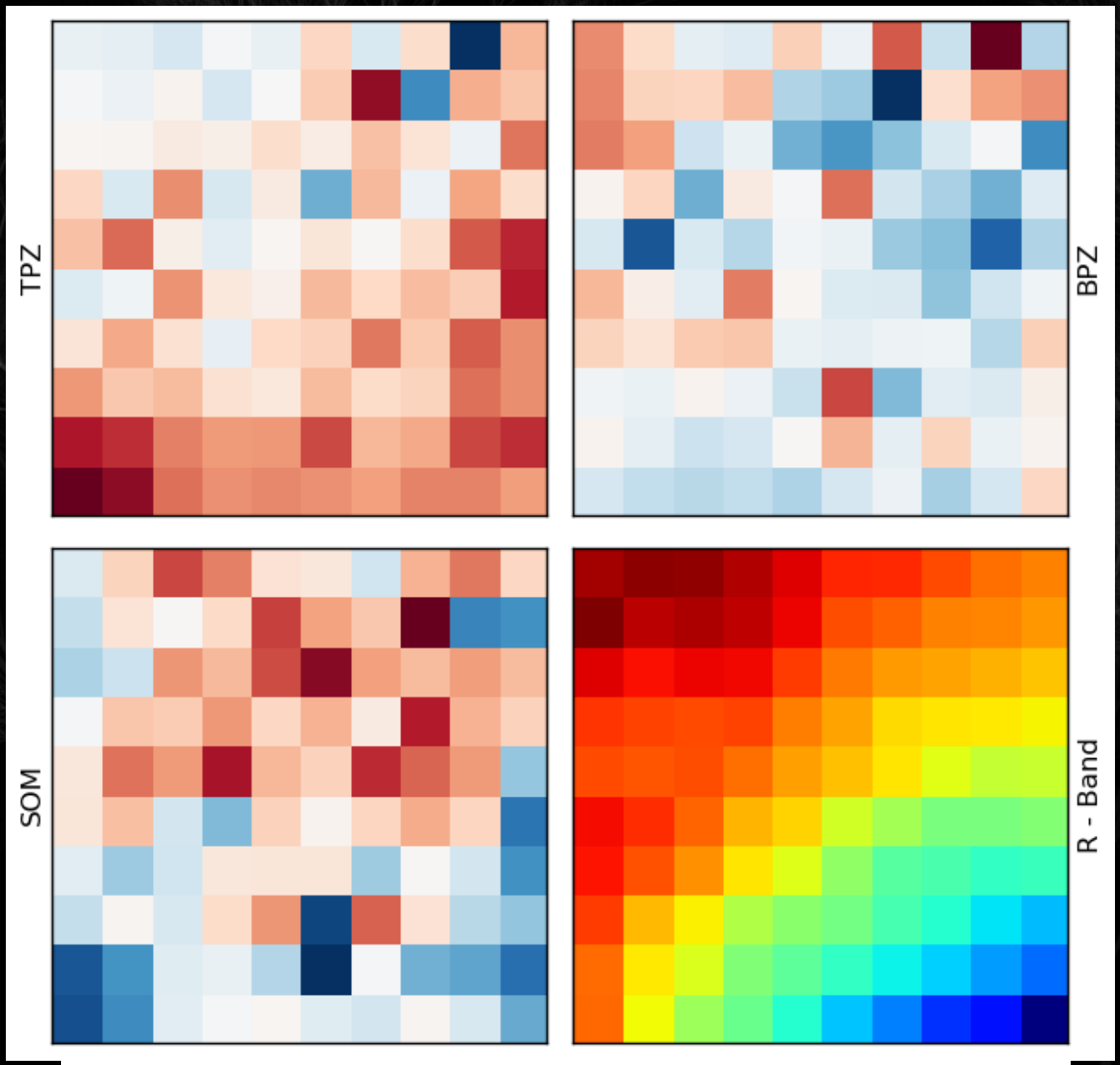
Combined PDF is, on average, better

BMA: marginalize over error in model

BMC: marginalize over error in combined models

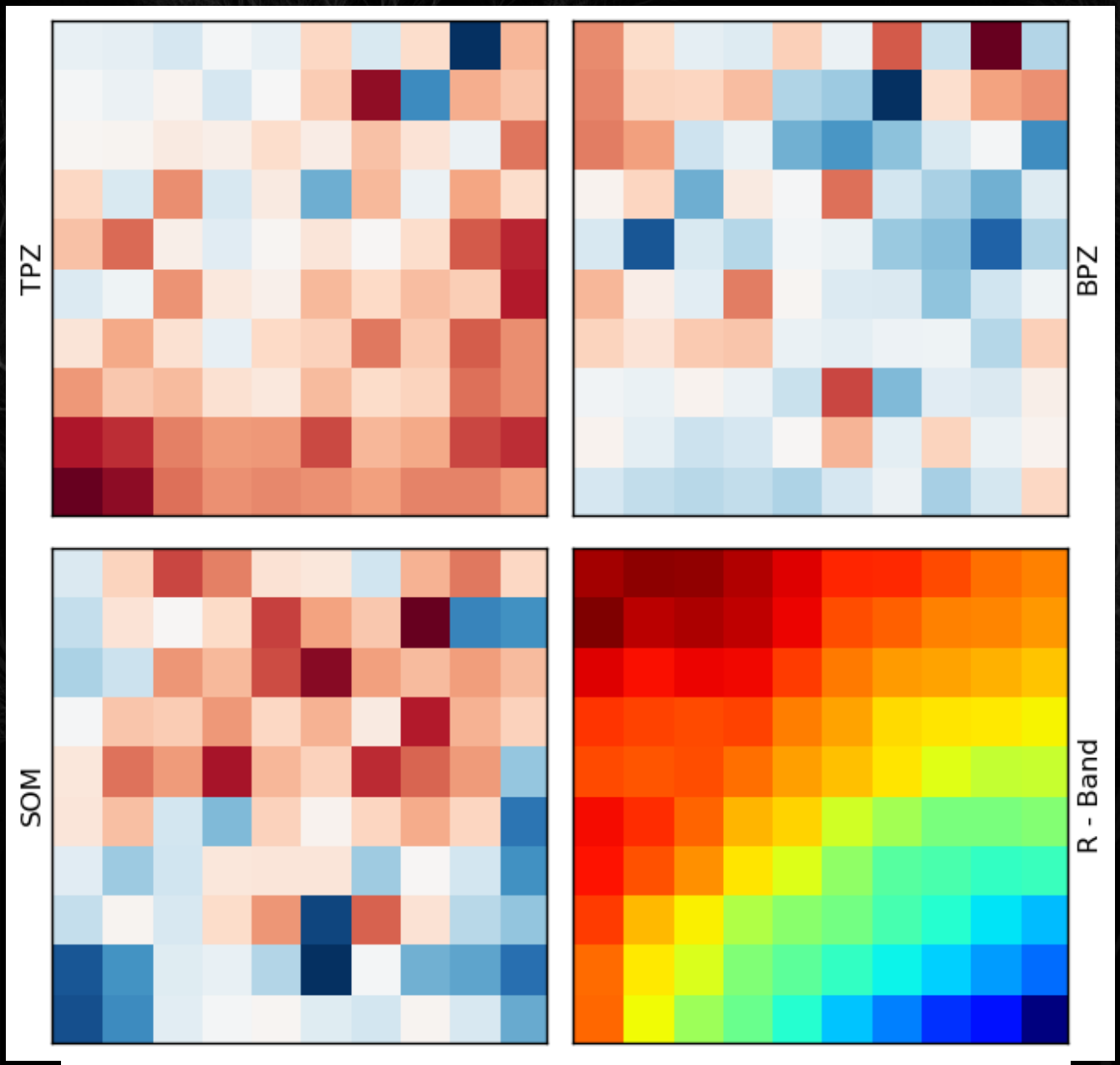
Combination depends on galaxy colors

Photo-z PDF combination: Bayesian framework



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

Photo-z PDF combination: Bayesian framework



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

Our approach

Supervised method

+

Unsupervised method

+

Template fitting

+

Weighting scheme

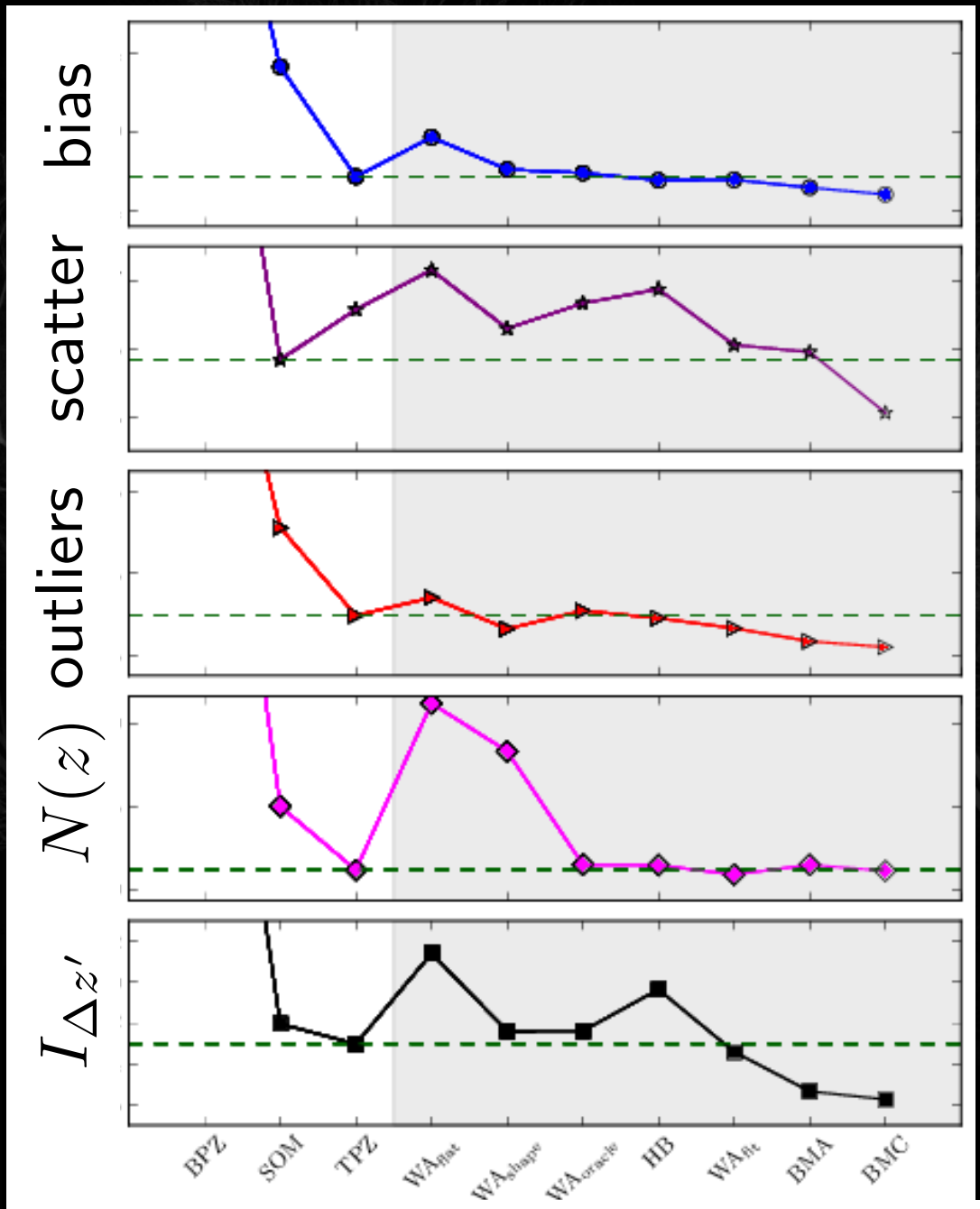
⇓

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: 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:

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

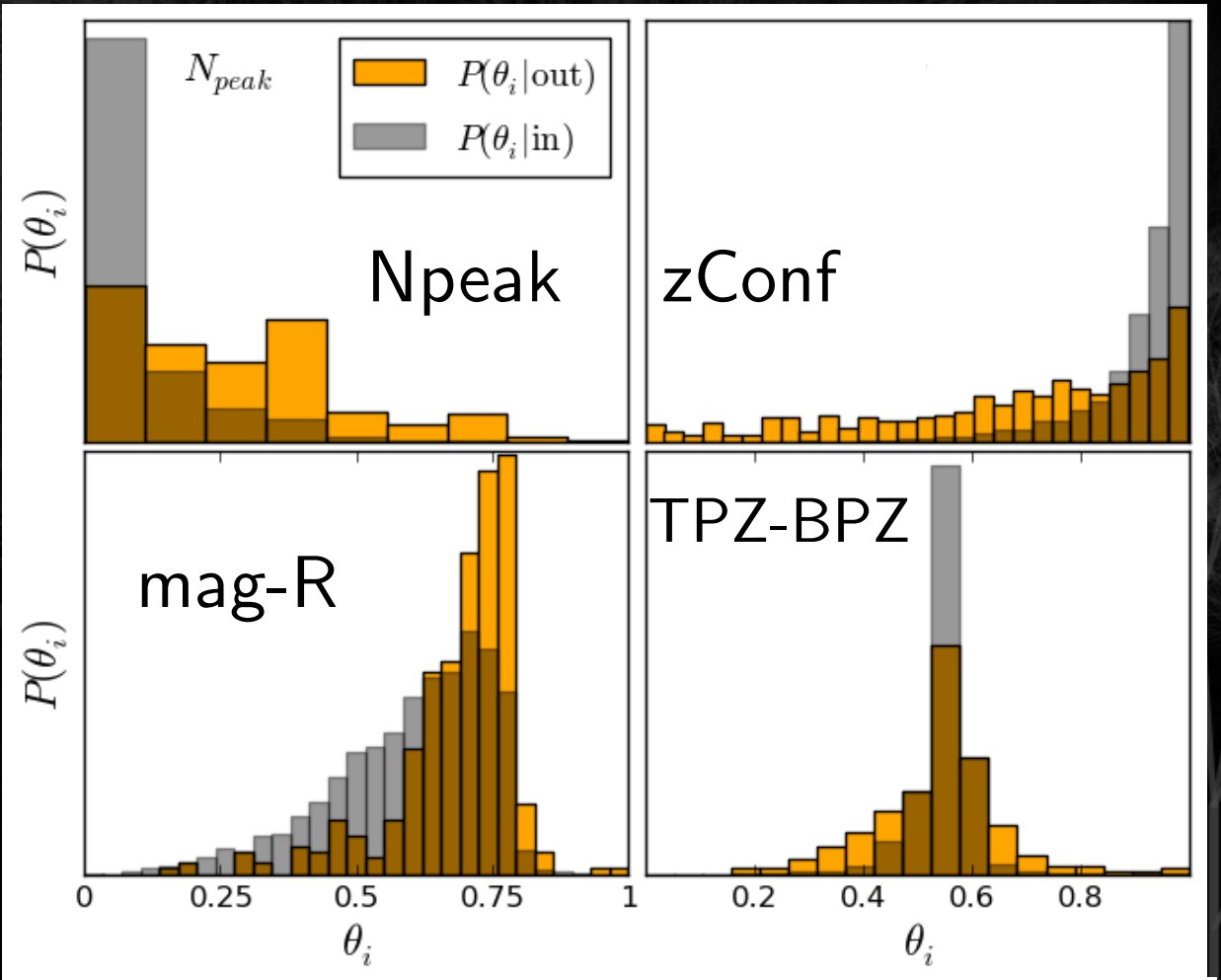
$\boldsymbol{\theta}$ includes: number of peaks, magnitudes, shape of PDF, differences, etc...

Photo-z PDF combination: Outliers

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

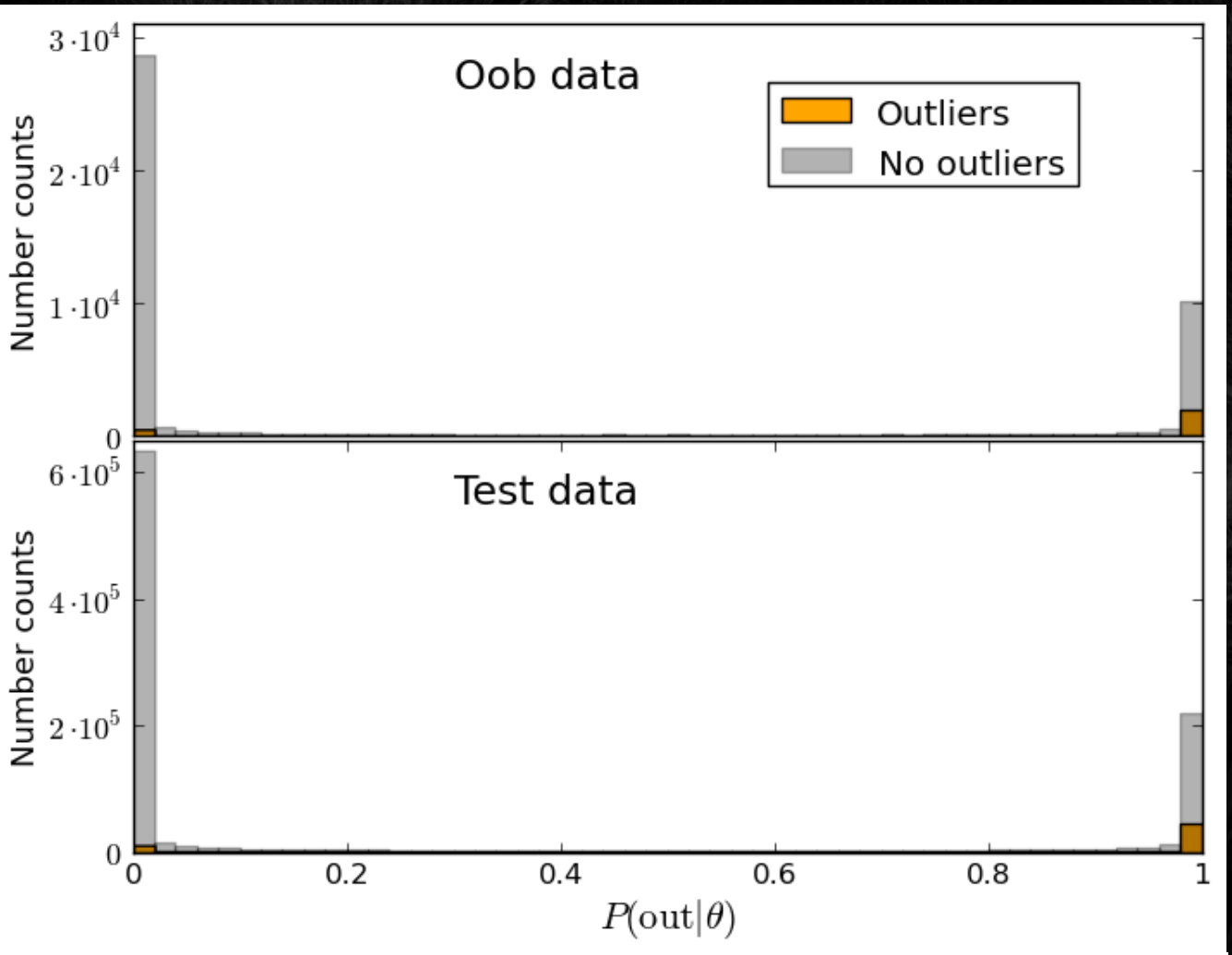
Each feature provides information about these two classes, and can be combined to make a stronger classifier

From plot, outliers have:
larger number of peaks
lower zConf values
fainter magnitudes
larger differences in photo-z



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

Photo-z PDF combination: Outliers



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

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

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

Outline

- INTRODUCTION
- HOW TO COMPUTE PHOTO-Z PDFs
- HOW TO COMBINE PHOTO-Z PDFs METHODS
- HOW TO REPRESENT PHOTO-Z PDF
- APPLICATIONS
- CONCLUSIONS AND FUTURE WORK

Outline

- INTRODUCTION
- HOW TO COMPUTE PHOTO-Z PDFs
- HOW TO COMBINE PHOTO-Z PDFs METHODS
- HOW TO REPRESENT PHOTO-Z PDF
- APPLICATIONS
- CONCLUSIONS AND FUTURE WORK

Photo-z PDF representation and storage

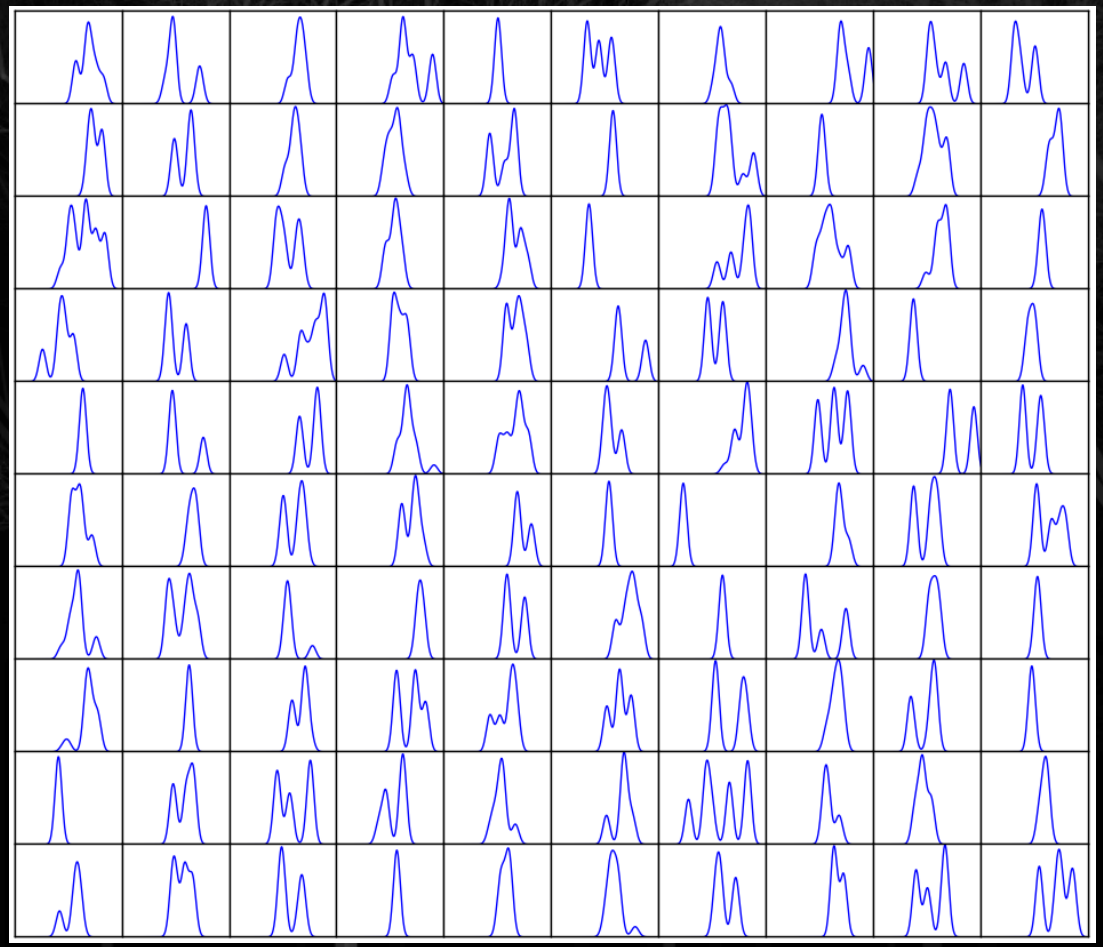


Photo-z PDF storage: Strategies

Single Gaussian fit

Multi-Gaussan fit

Monte Carlo sampling

Sparse representation techniques

Reduce number of points while increasing accuracy

Photo-z PDF storage: Strategies

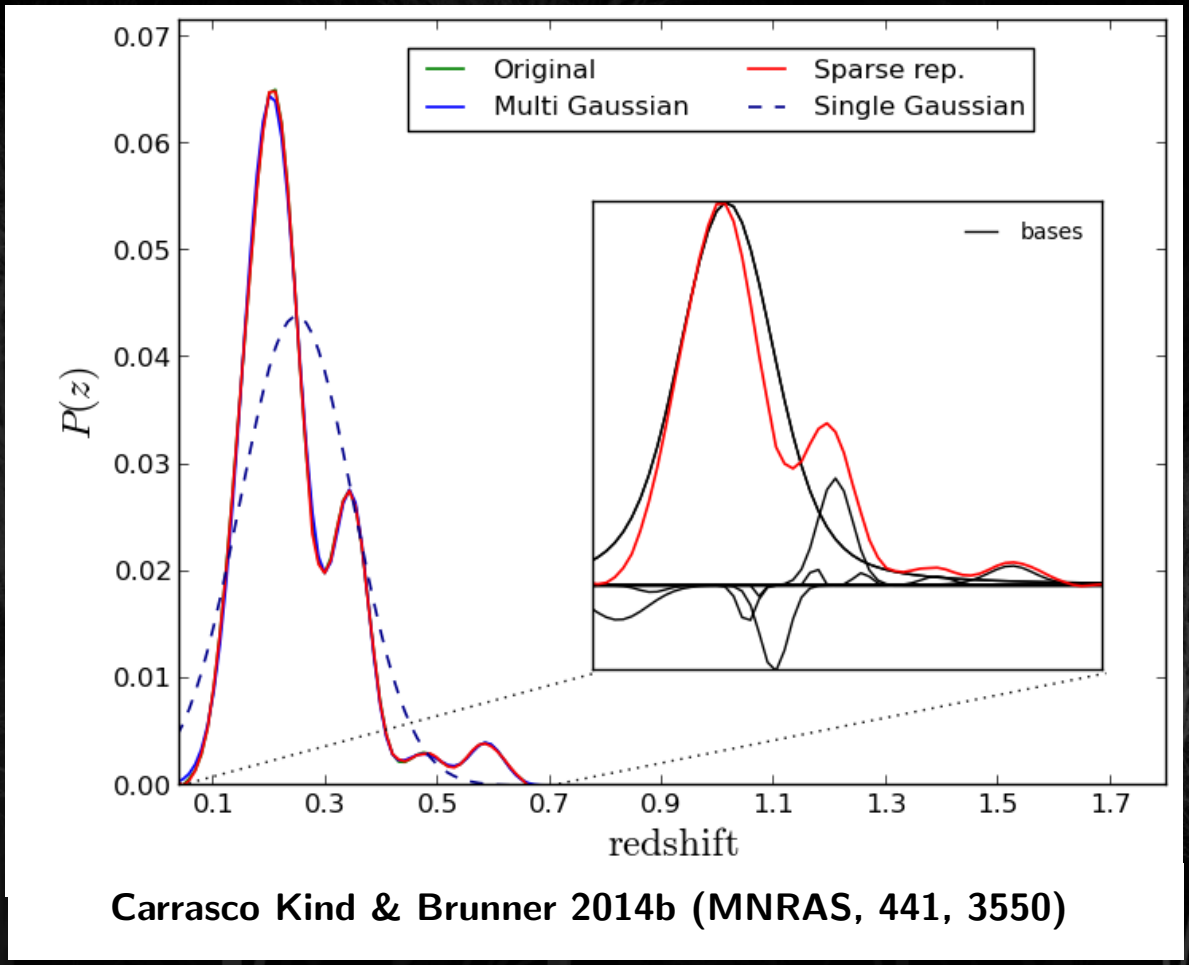
Single Gaussian fit

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

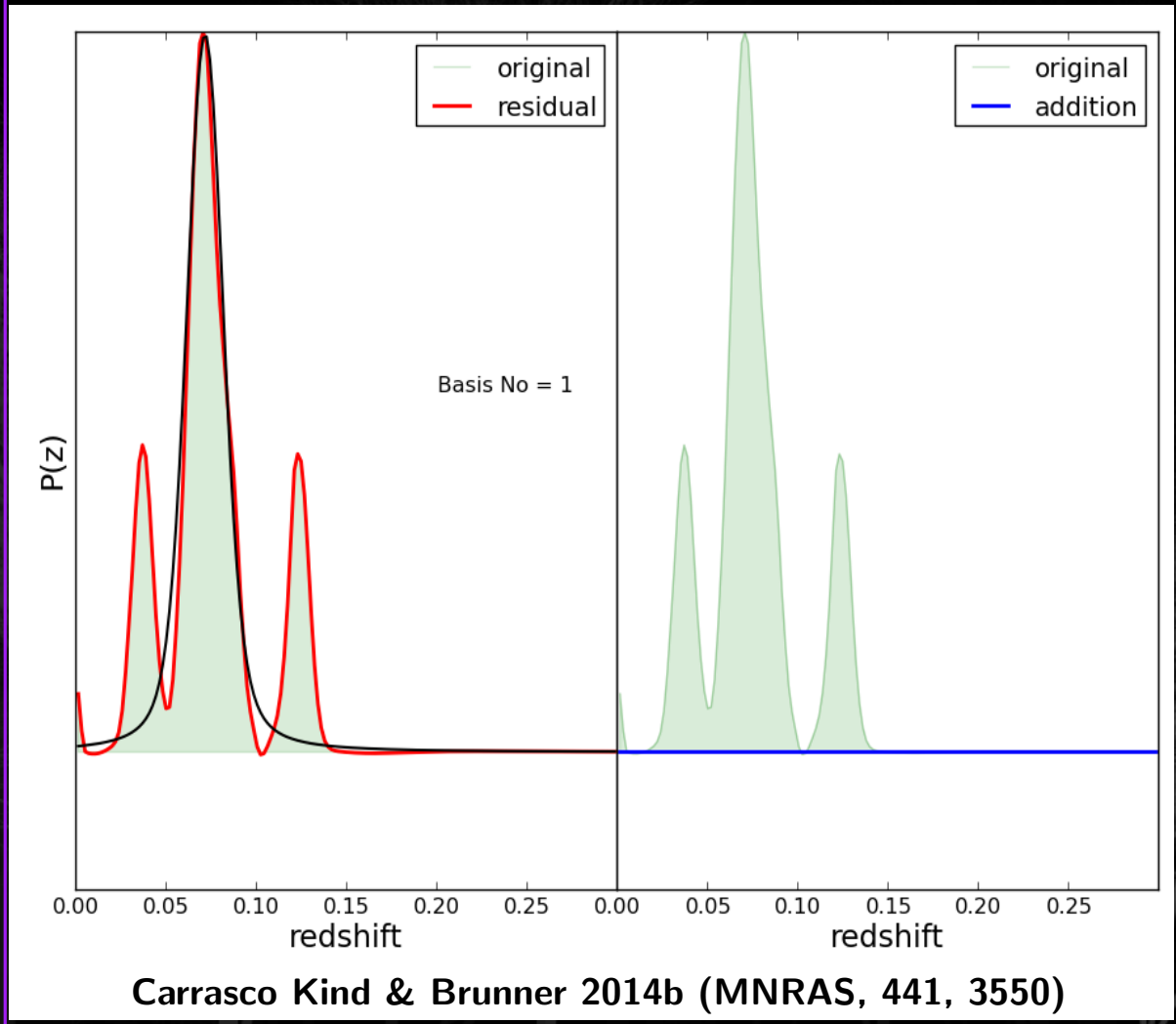
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude 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, 441, 3550)

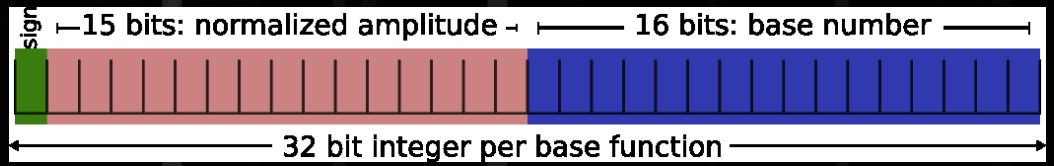


Photo-z PDF storage: Sparse representation

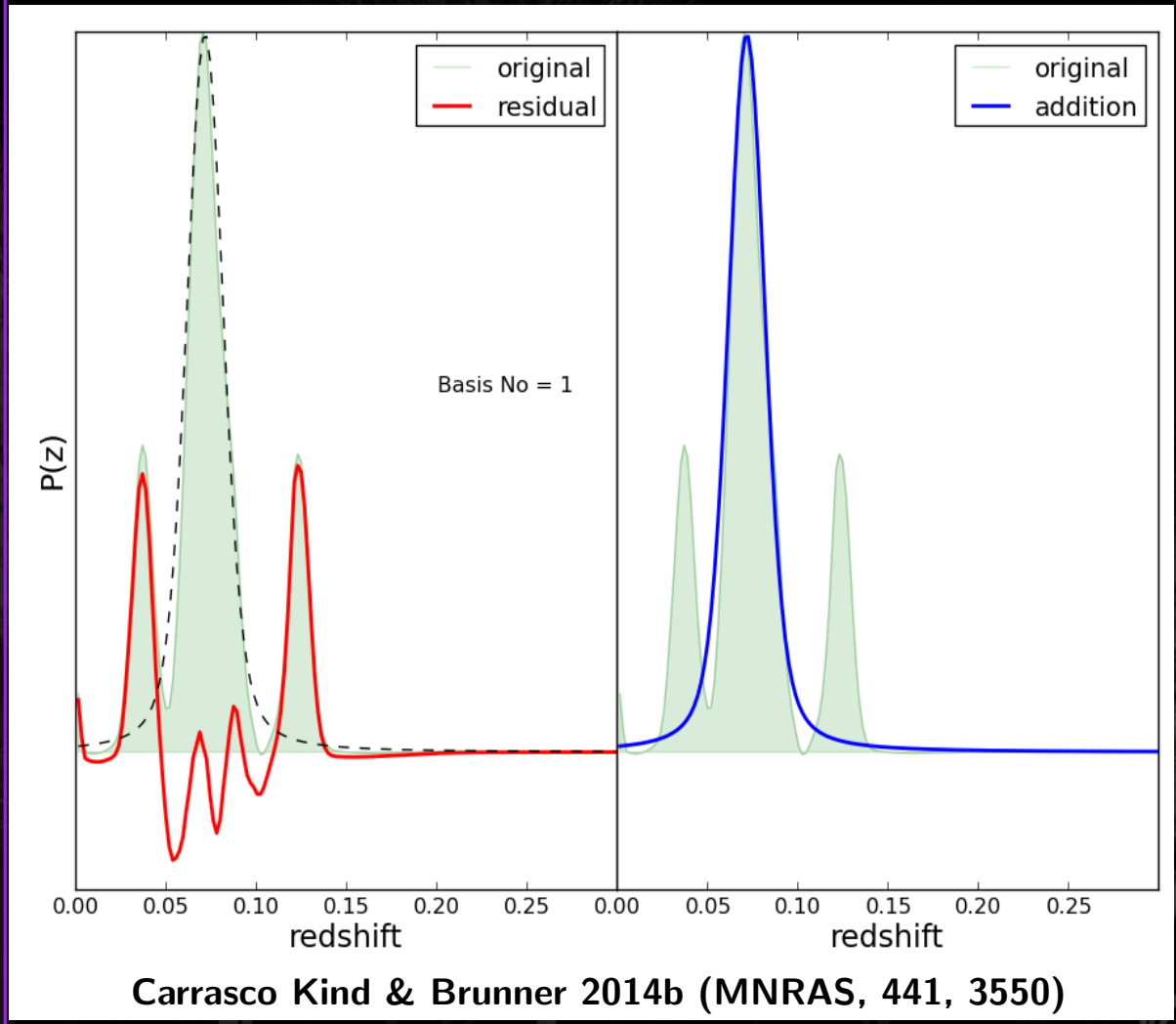
Use Gaussian and Voigt profiles as bases, need N^2_{original} bases

Find basis and amplitude 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, 441, 3550)

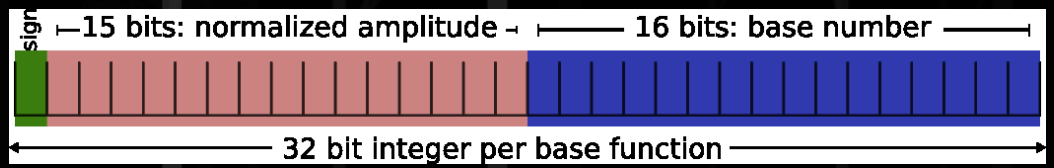


Photo-z PDF storage: Sparse representation

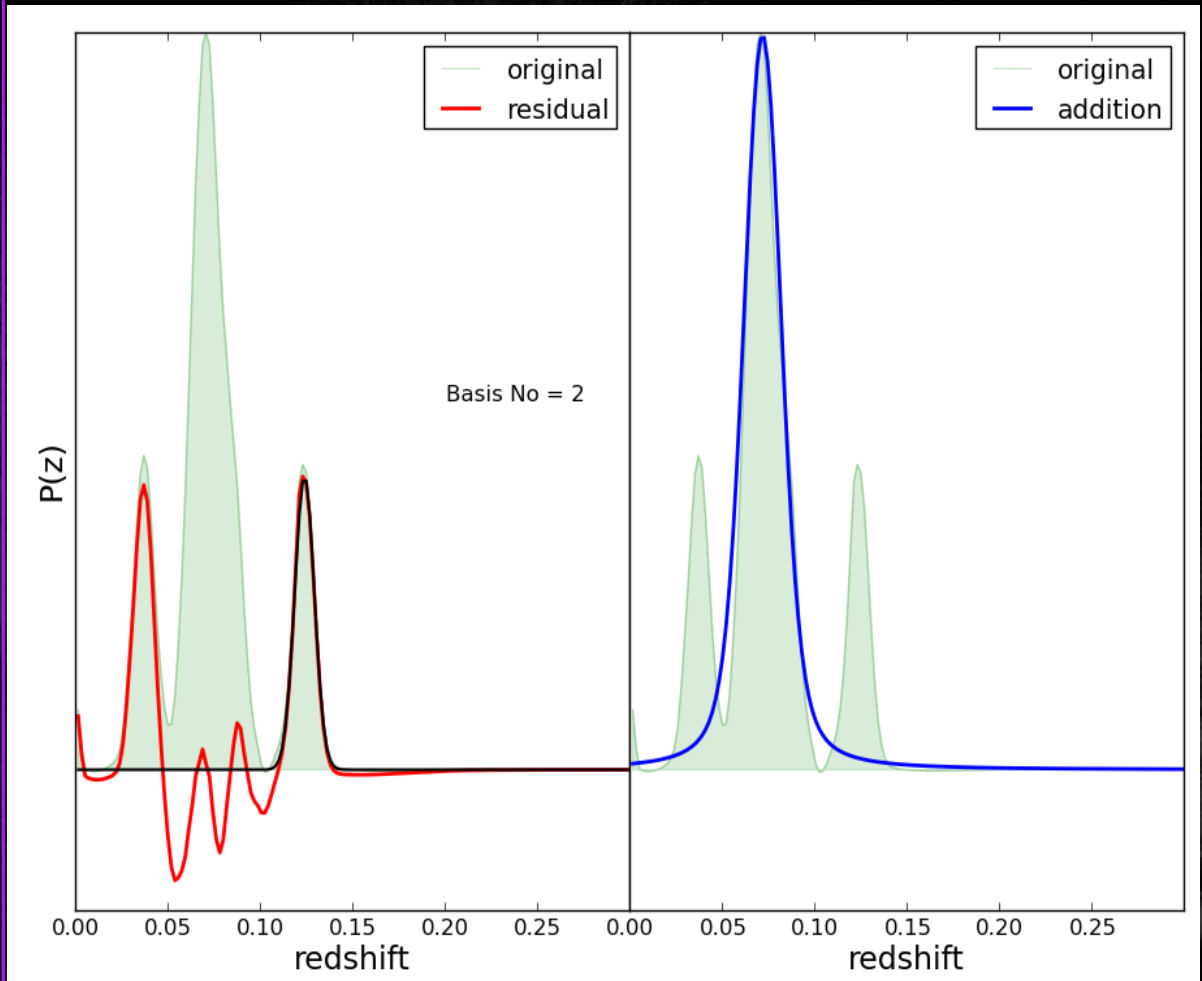
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude 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, 441, 3550)

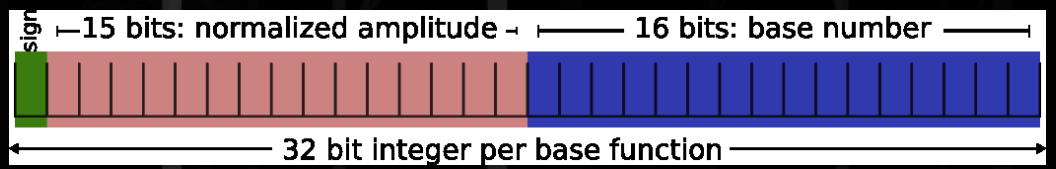


Photo-z PDF storage: Sparse representation

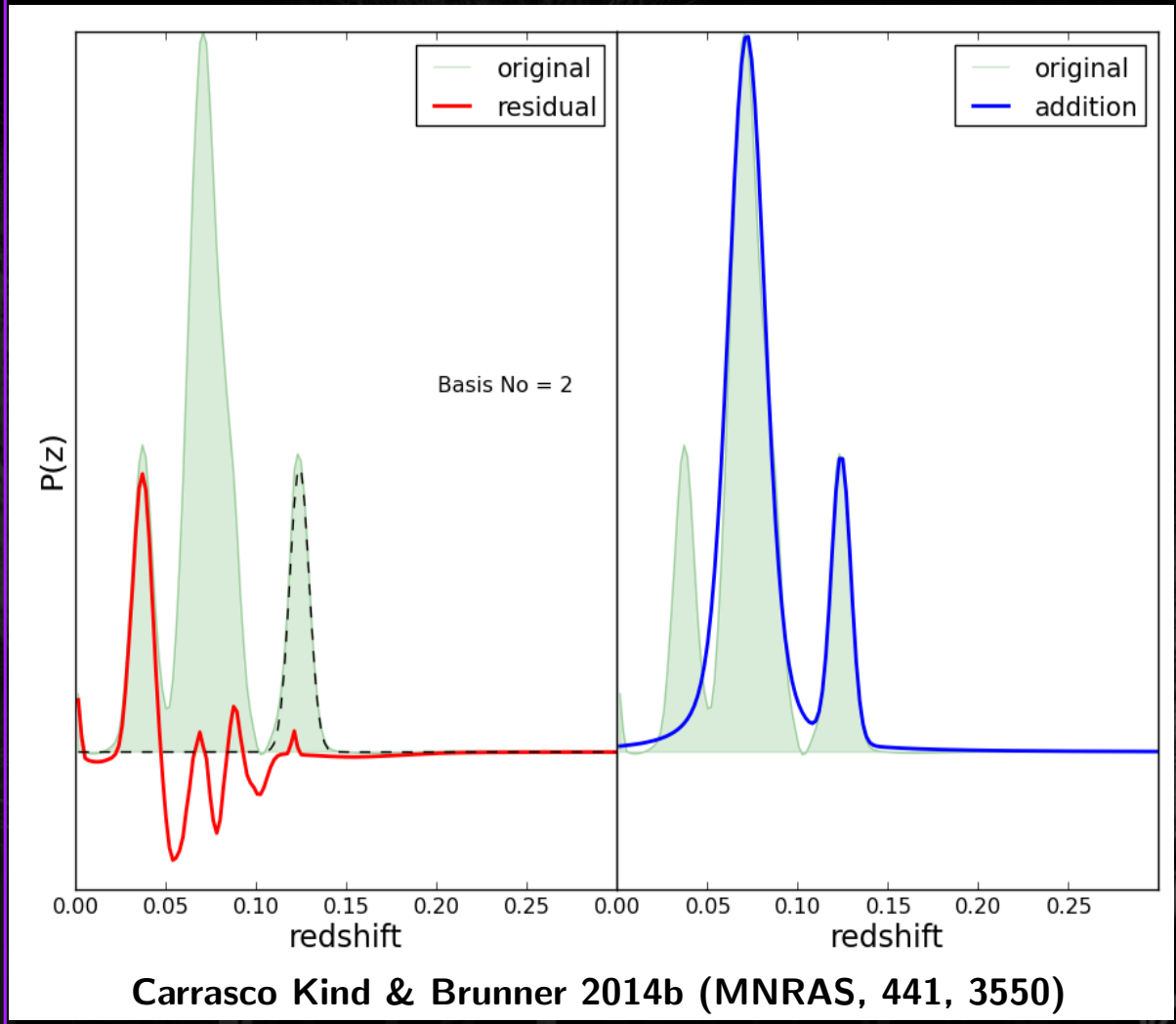
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude 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, 441, 3550)

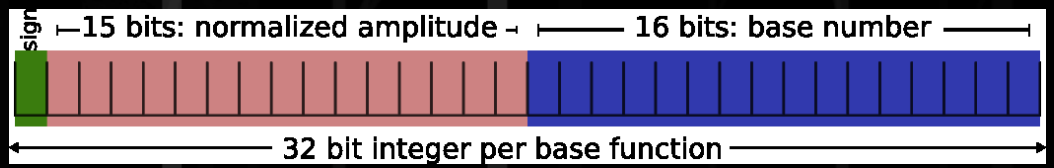


Photo-z PDF storage: Sparse representation

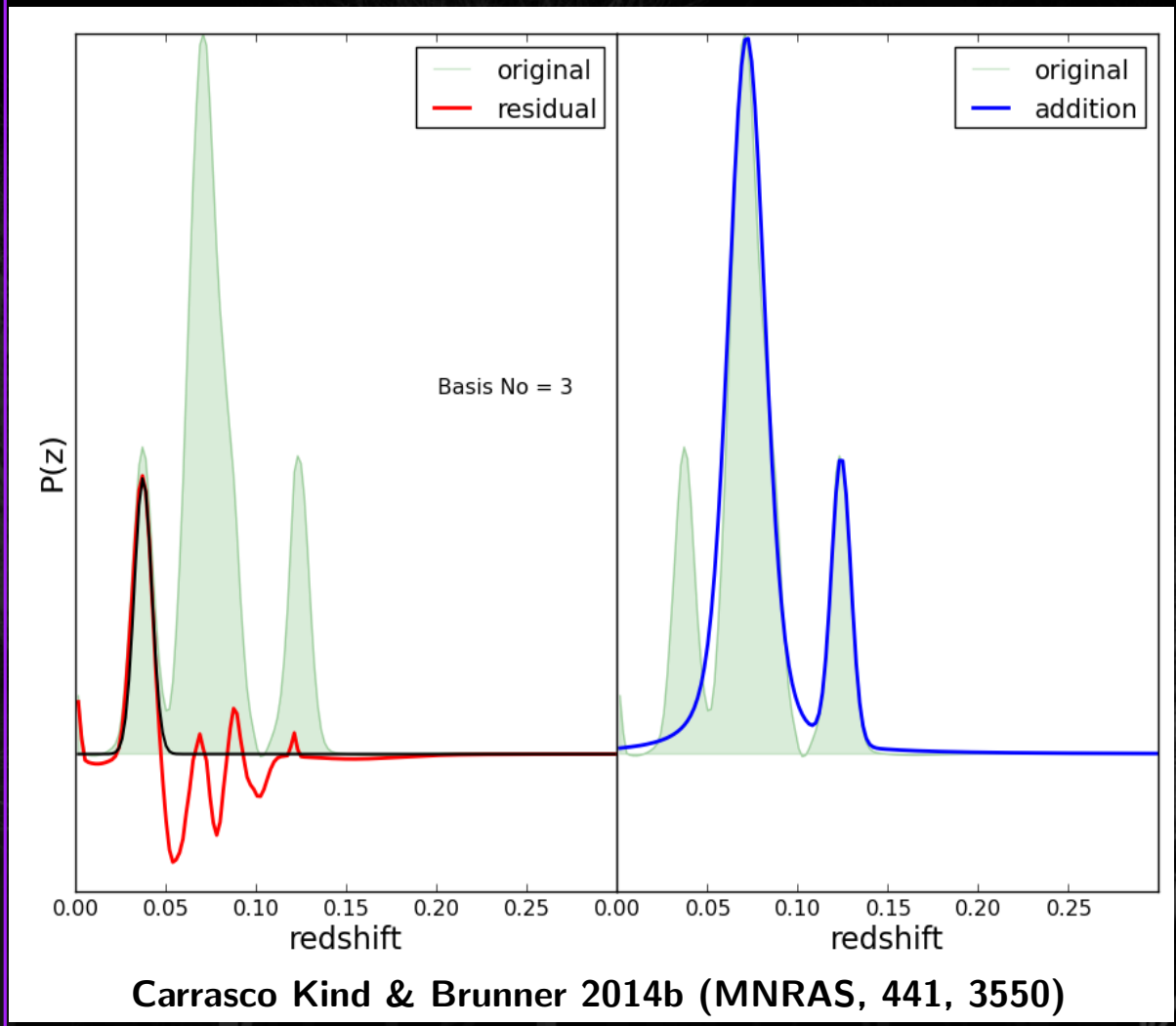
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude 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, 441, 3550)

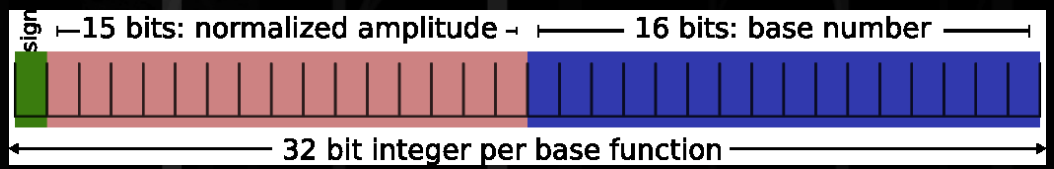


Photo-z PDF storage: Sparse representation

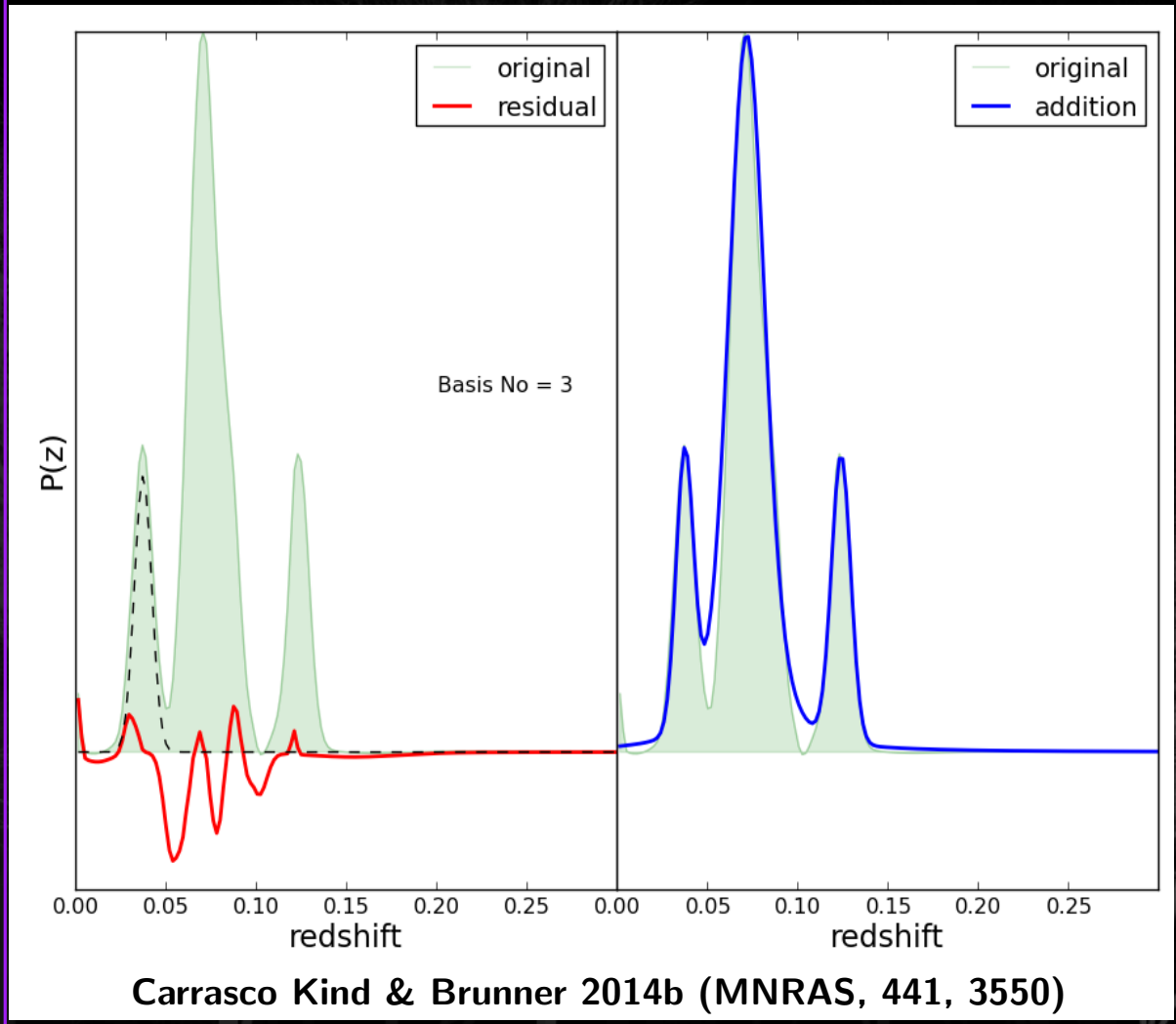
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude 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, 441, 3550)

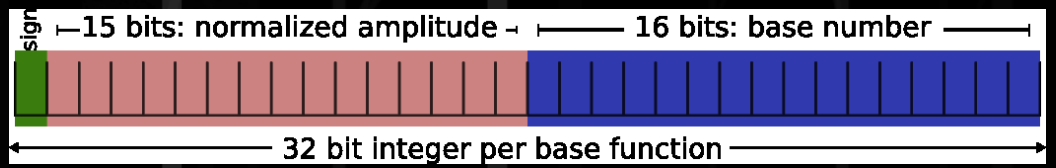


Photo-z PDF storage: Sparse representation

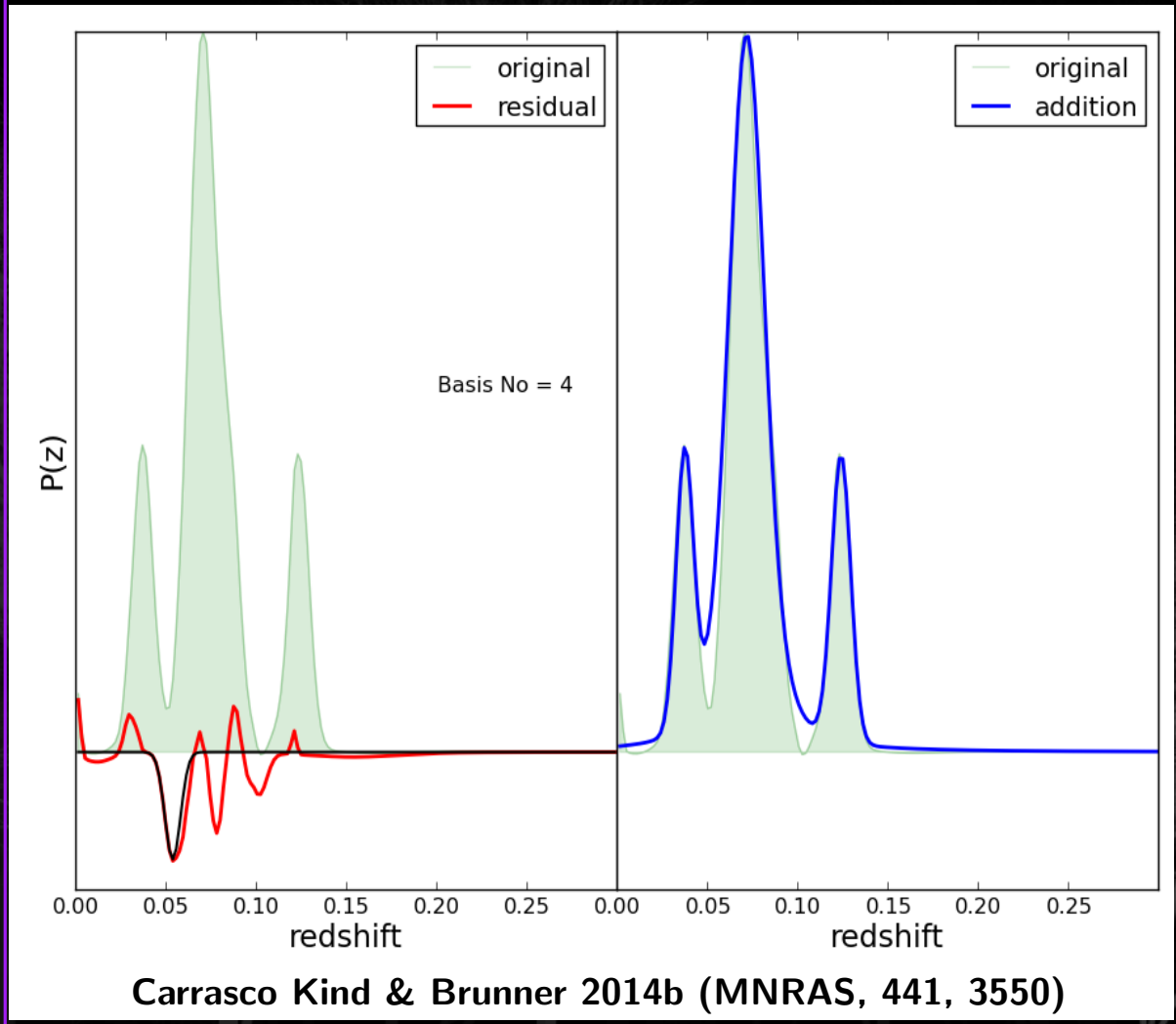
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude 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, 441, 3550)

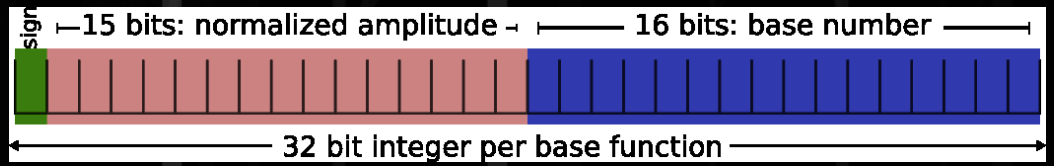


Photo-z PDF storage: Sparse representation

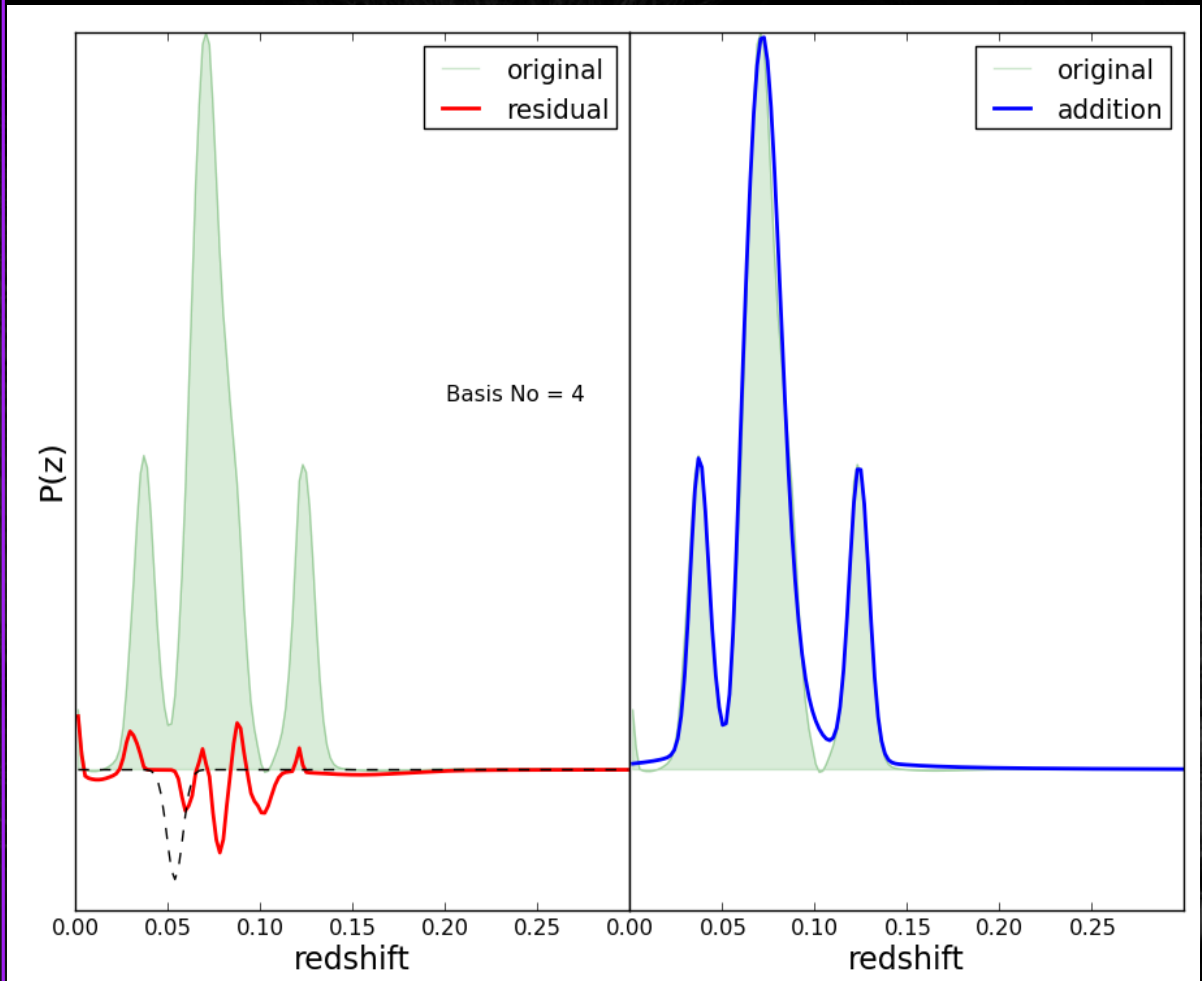
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude 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, 441, 3550)

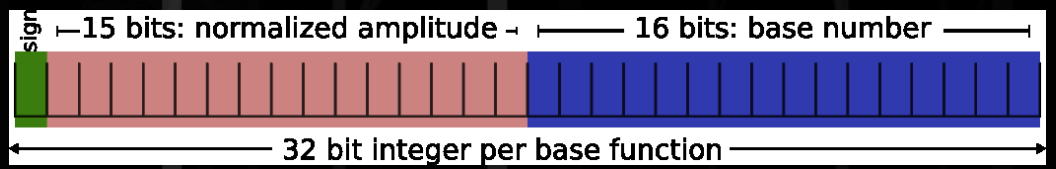


Photo-z PDF storage: Sparse representation

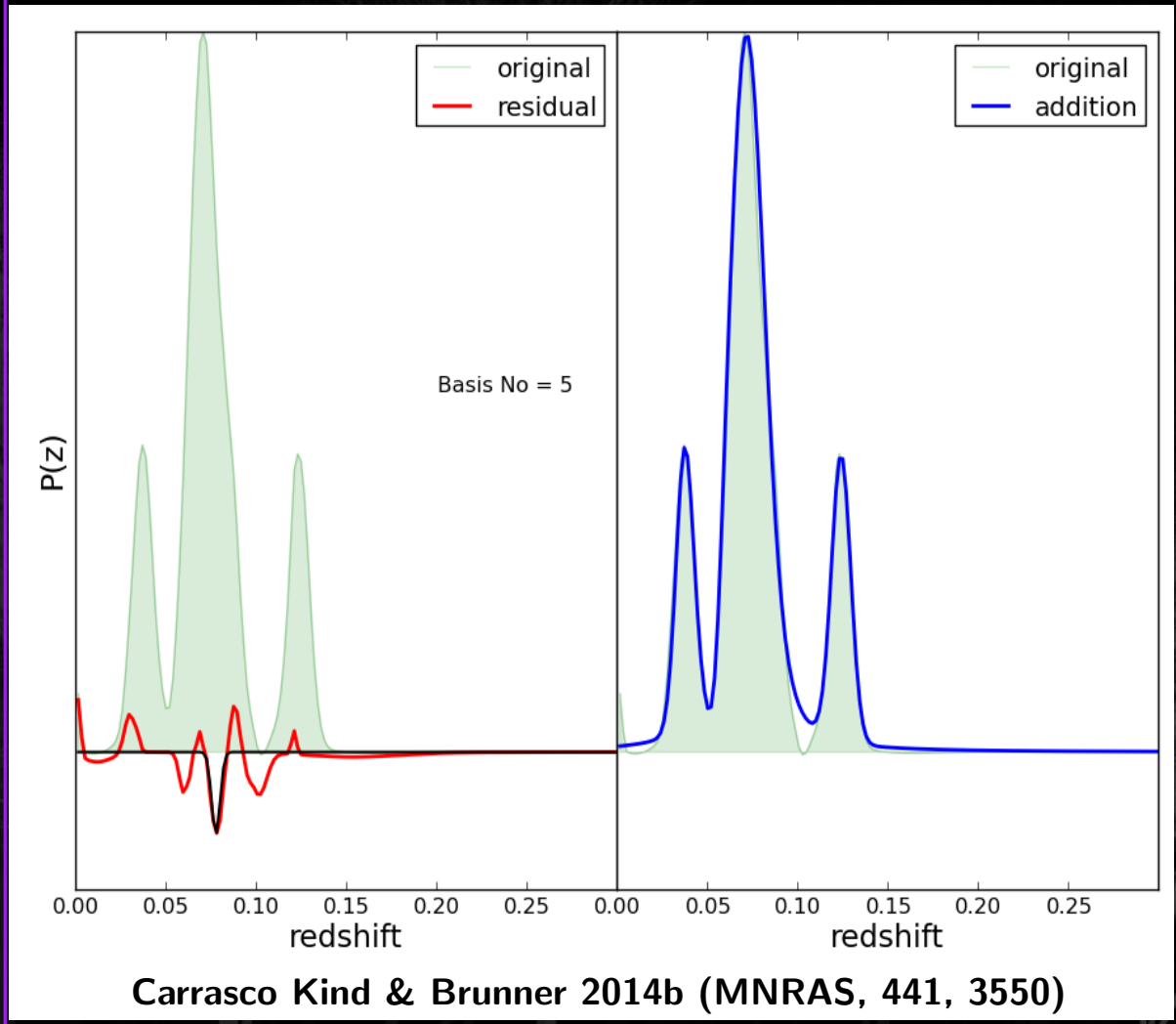
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude 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, 441, 3550)

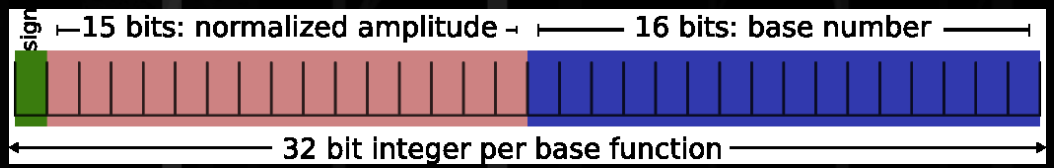


Photo-z PDF storage: Sparse representation

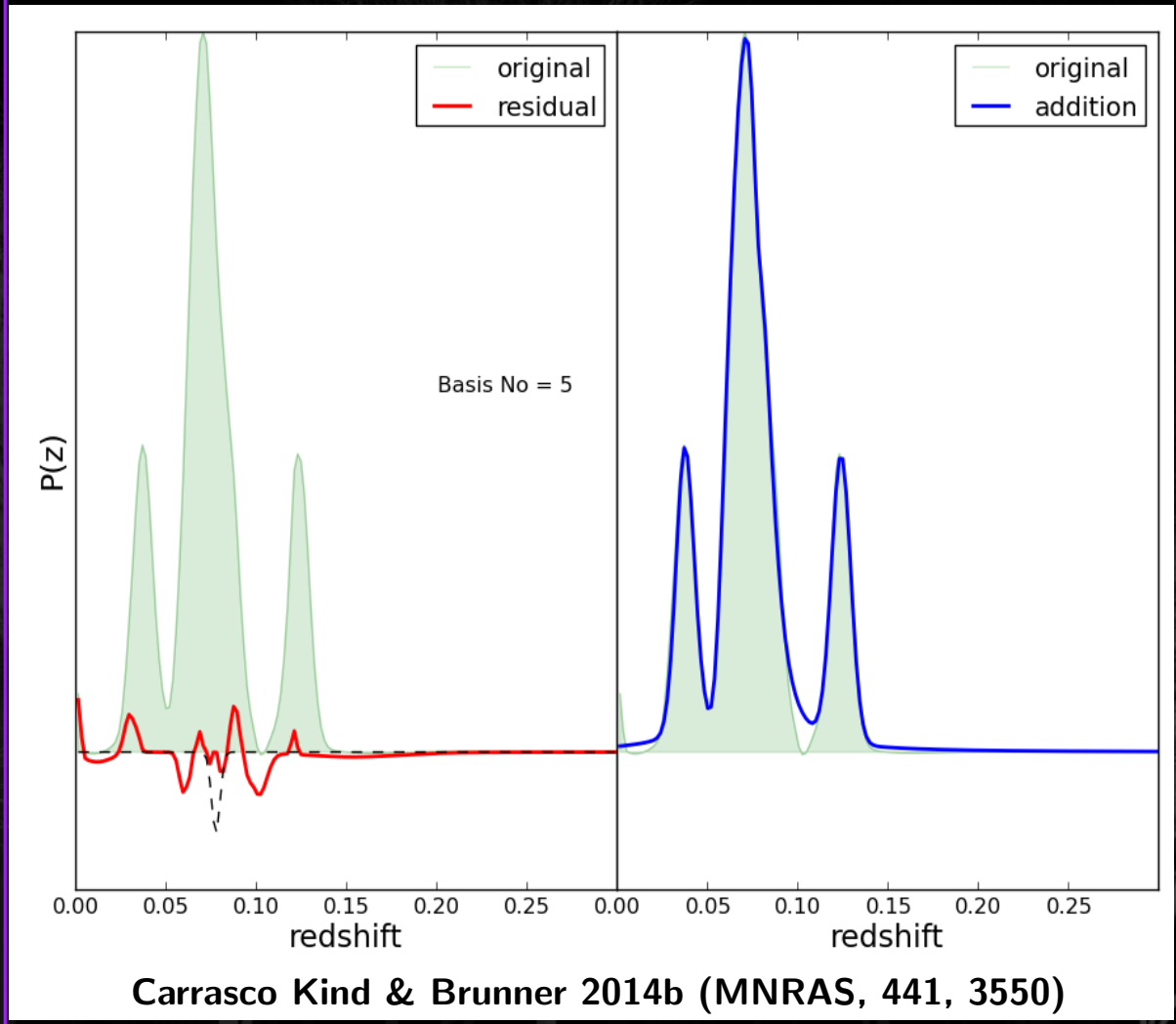
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude 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, 441, 3550)

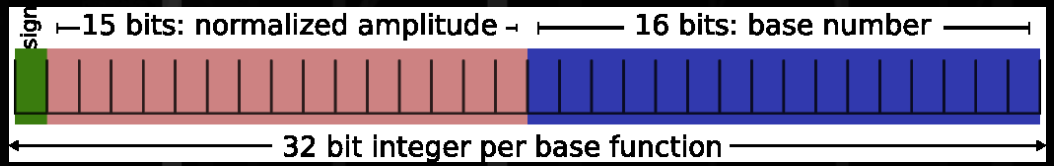


Photo-z PDF storage: Sparse representation

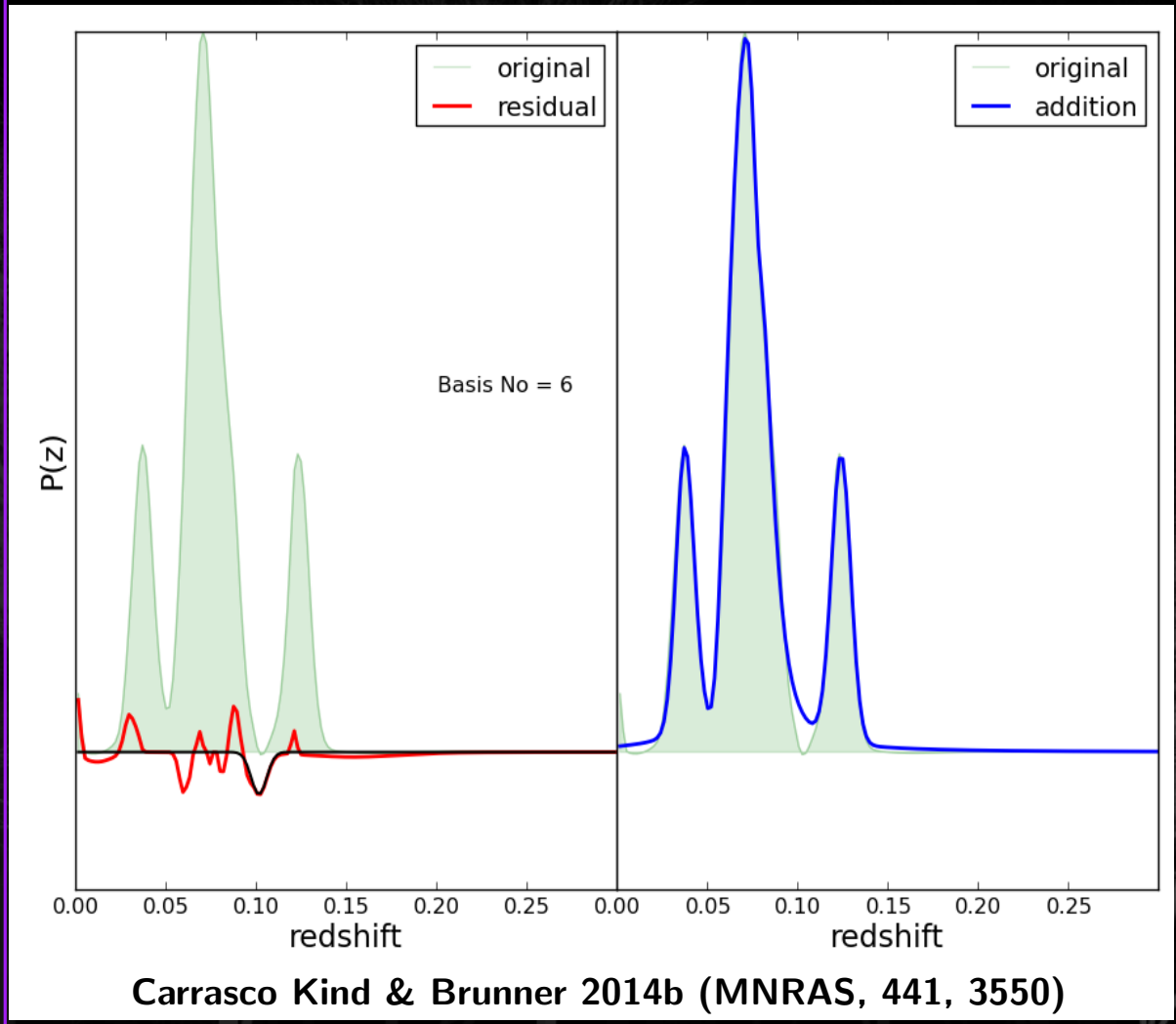
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude 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, 441, 3550)

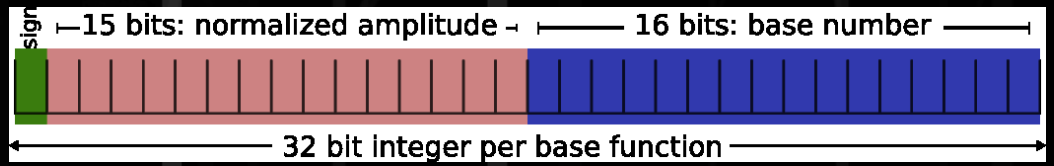


Photo-z PDF storage: Sparse representation

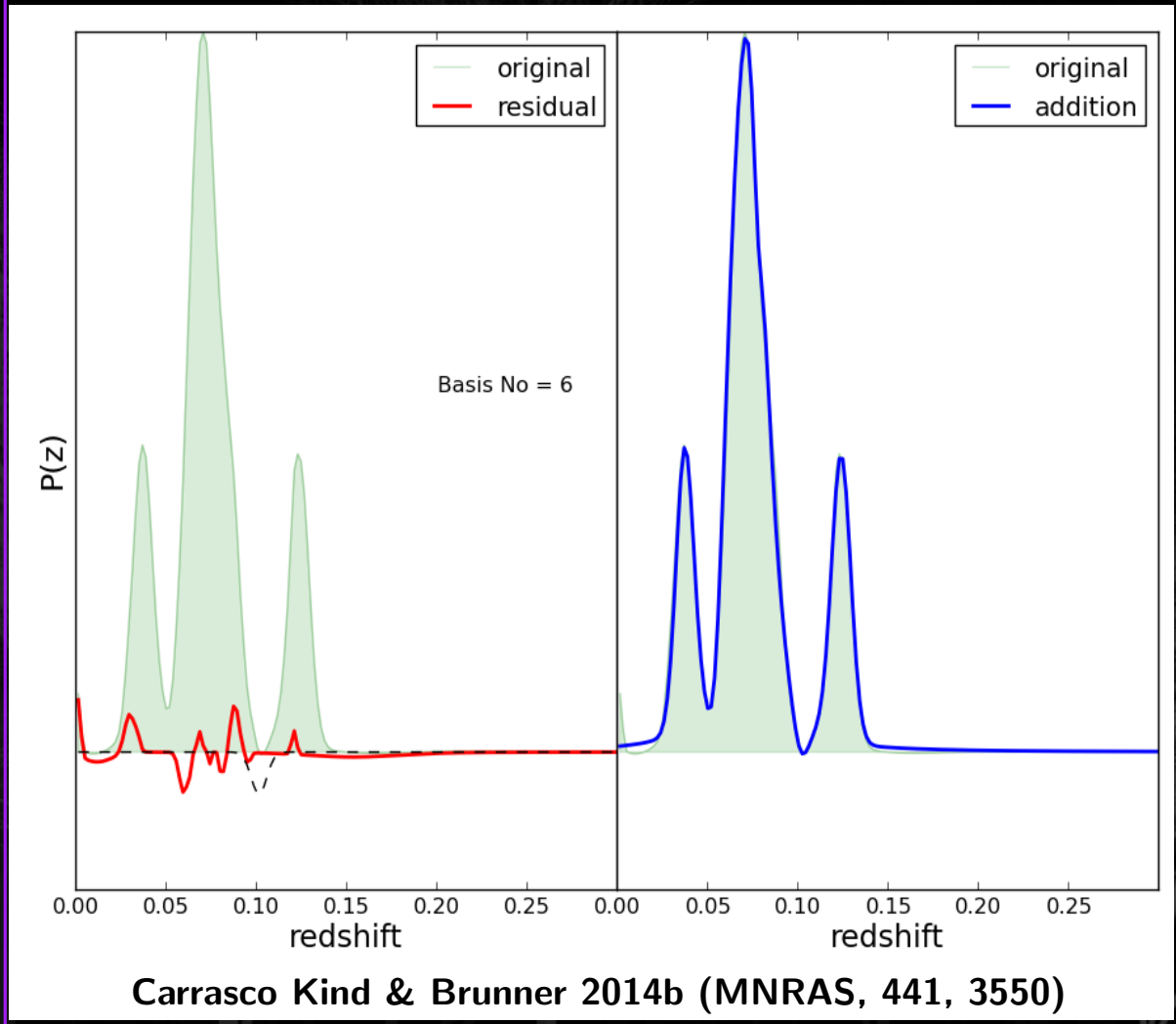
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude 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, 441, 3550)

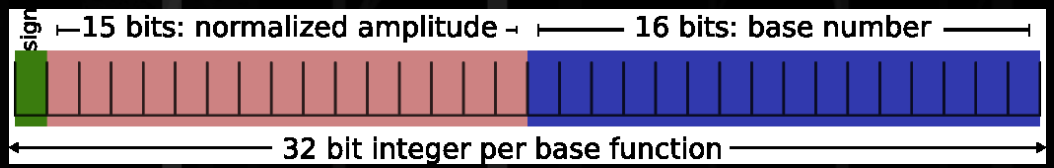


Photo-z PDF storage: Sparse representation

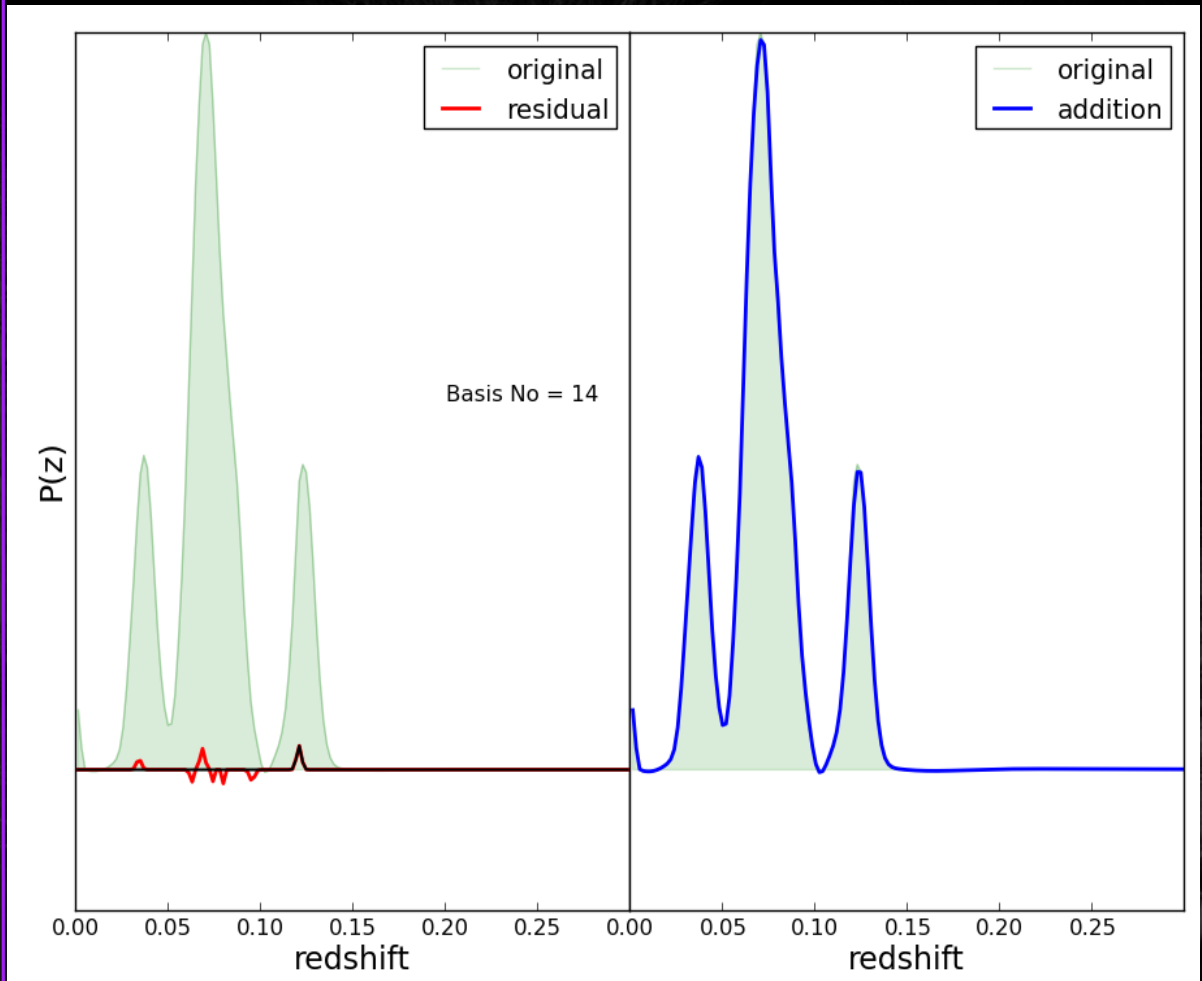
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude 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, 441, 3550)

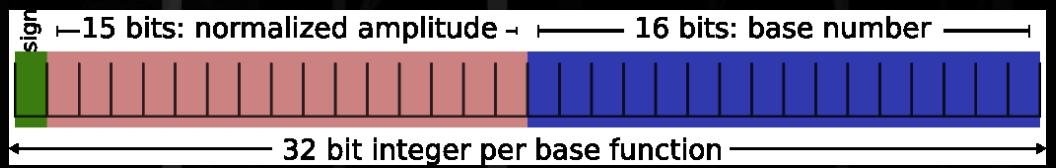


Photo-z PDF storage: Sparse representation

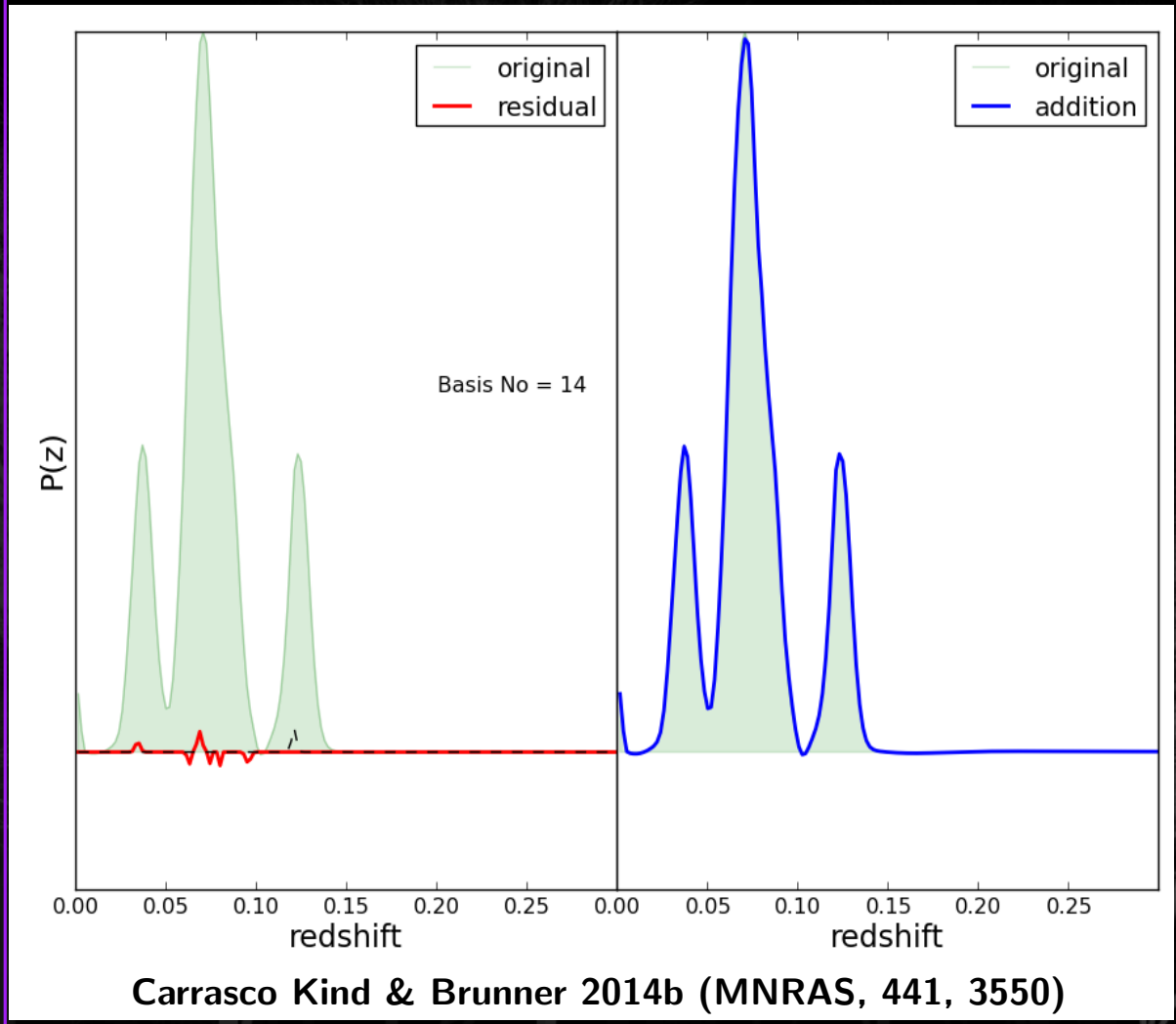
Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

Find basis and amplitude to reduce residual on each step

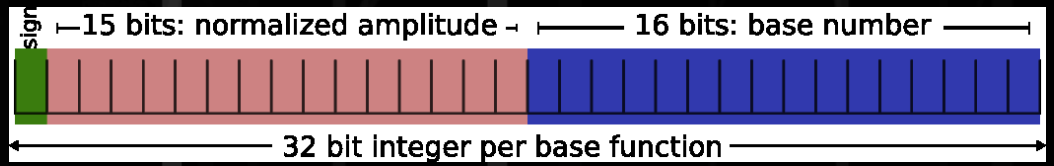
With only 10-20 bases achieve 99.9 % accuracy

Use 32-bits integer per basis, compression

Store Multiple PDFs



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



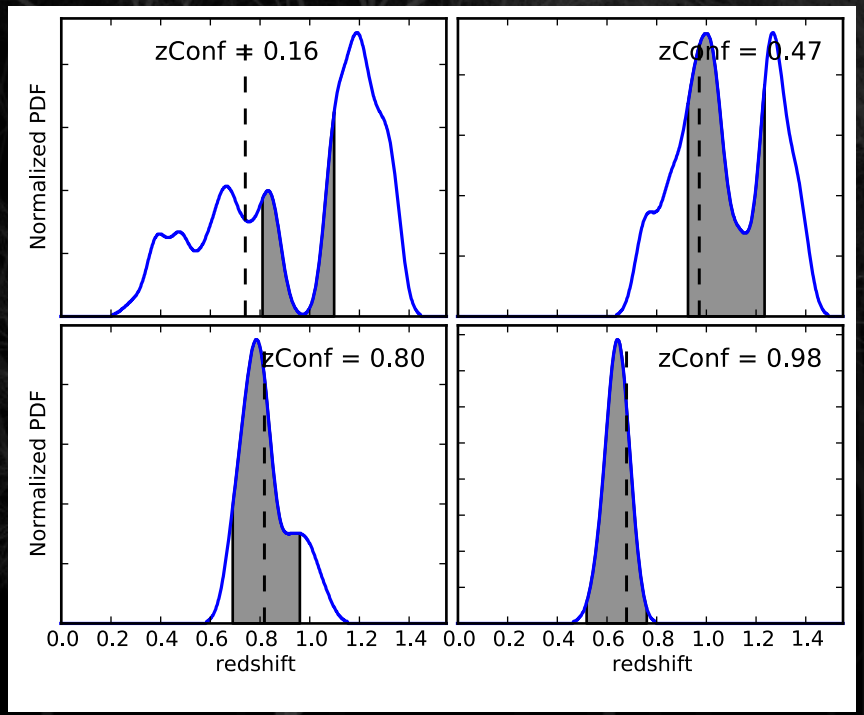
Outline

- INTRODUCTION
- HOW TO COMPUTE PHOTO-Z PDFs
- HOW TO COMBINE PHOTO-Z PDFs METHODS
- HOW TO REPRESENT PHOTO-Z PDF
- APPLICATIONS
- CONCLUSIONS AND FUTURE WORK

Outline

- INTRODUCTION
- HOW TO COMPUTE PHOTO-Z PDFs
- HOW TO COMBINE PHOTO-Z PDFs METHODS
- HOW TO REPRESENT PHOTO-Z PDF
- APPLICATIONS
- CONCLUSIONS AND FUTURE WORK

Photo-z PDF applications



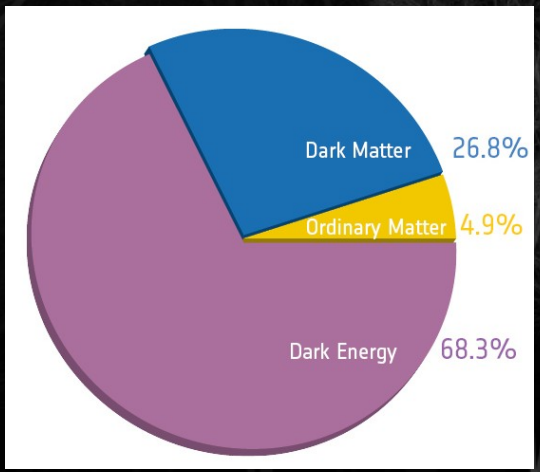


Photo-z PDF application: $N(z)$

$N(z)$ 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

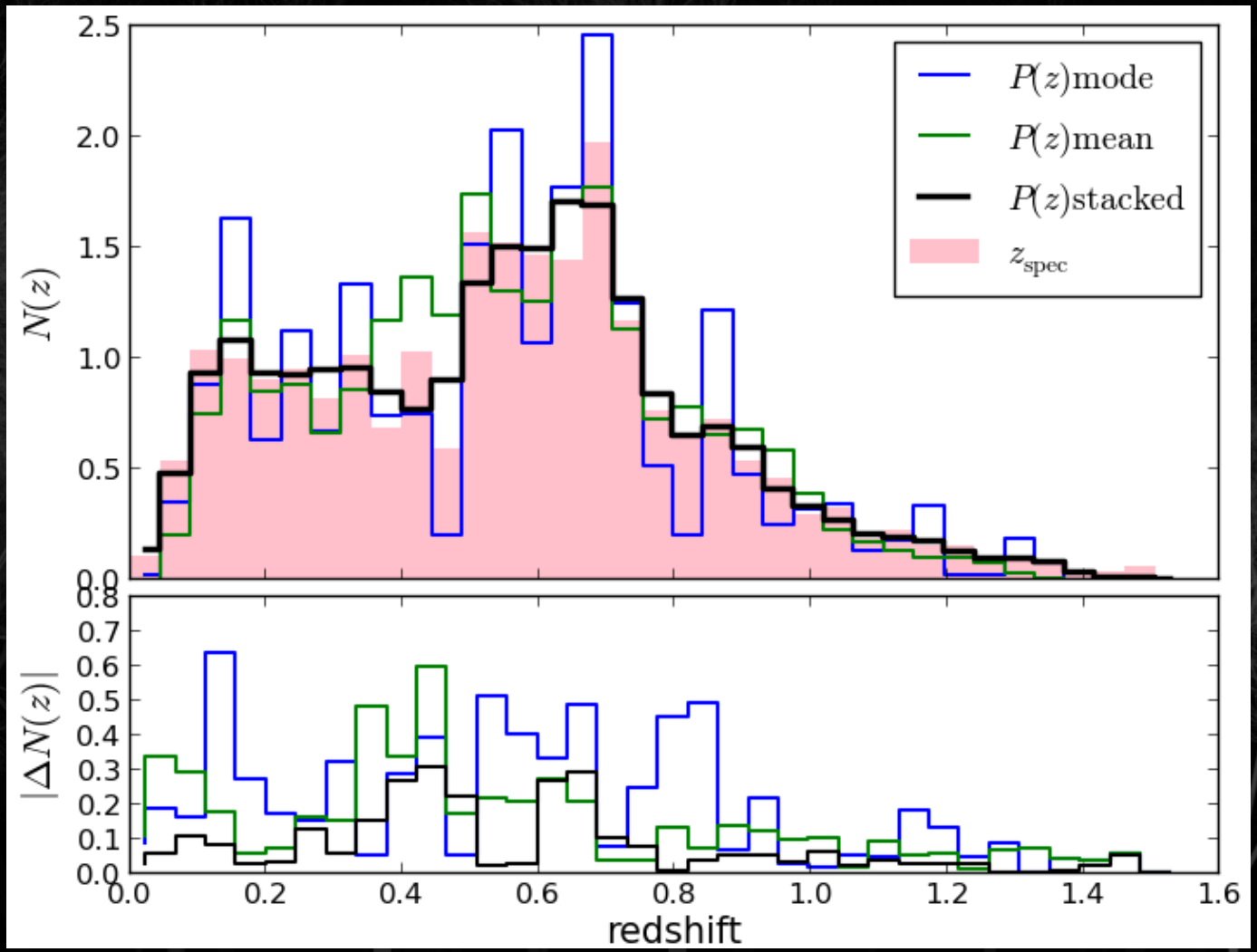


Photo-z PDF application: $N(z)$ and sparse rep.

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: $N(z)$ and sparse rep.

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: $N(z)$ and sparse rep.

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: $N(z)$ and sparse rep.

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: $N(z)$ and sparse rep.

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

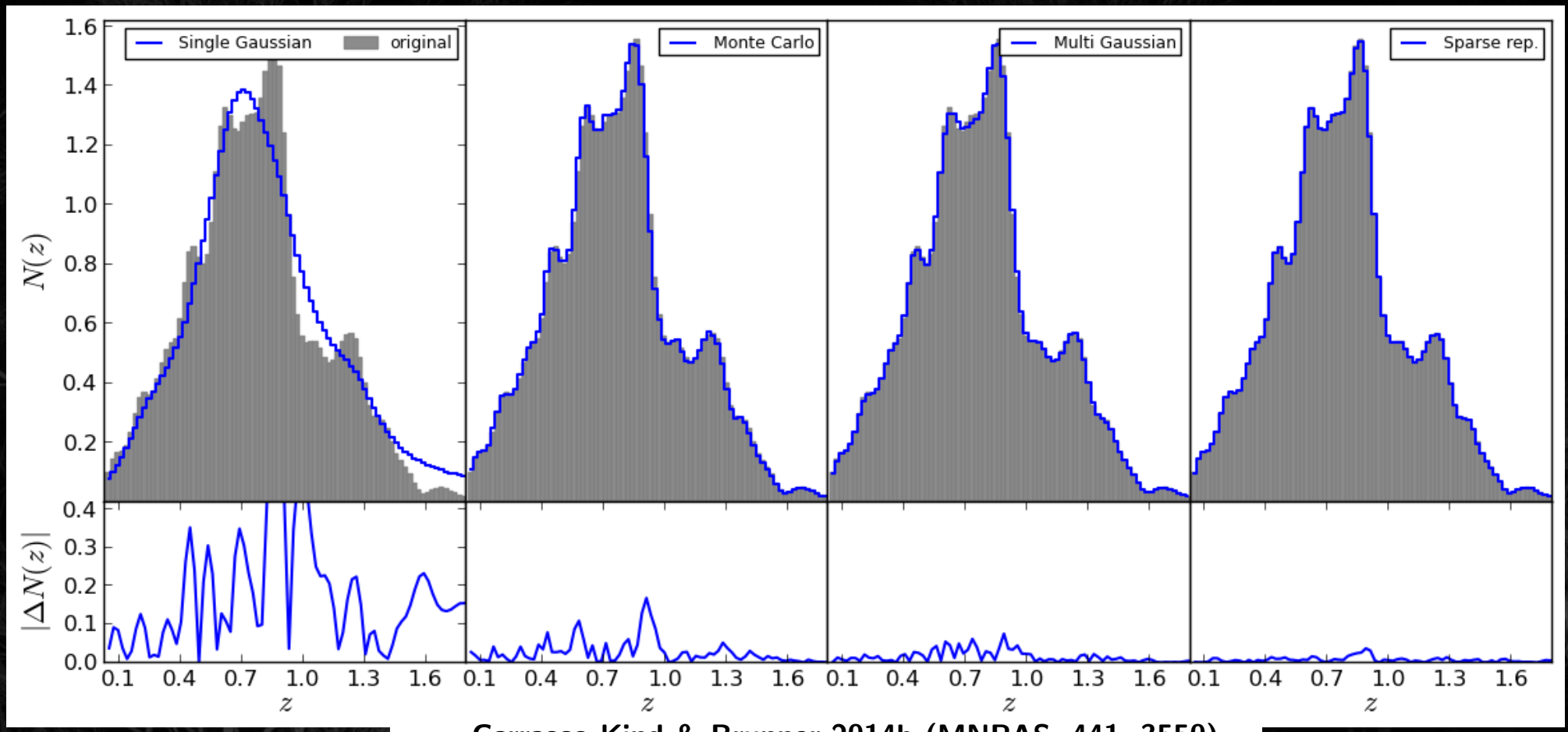
by precomputing:

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

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

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

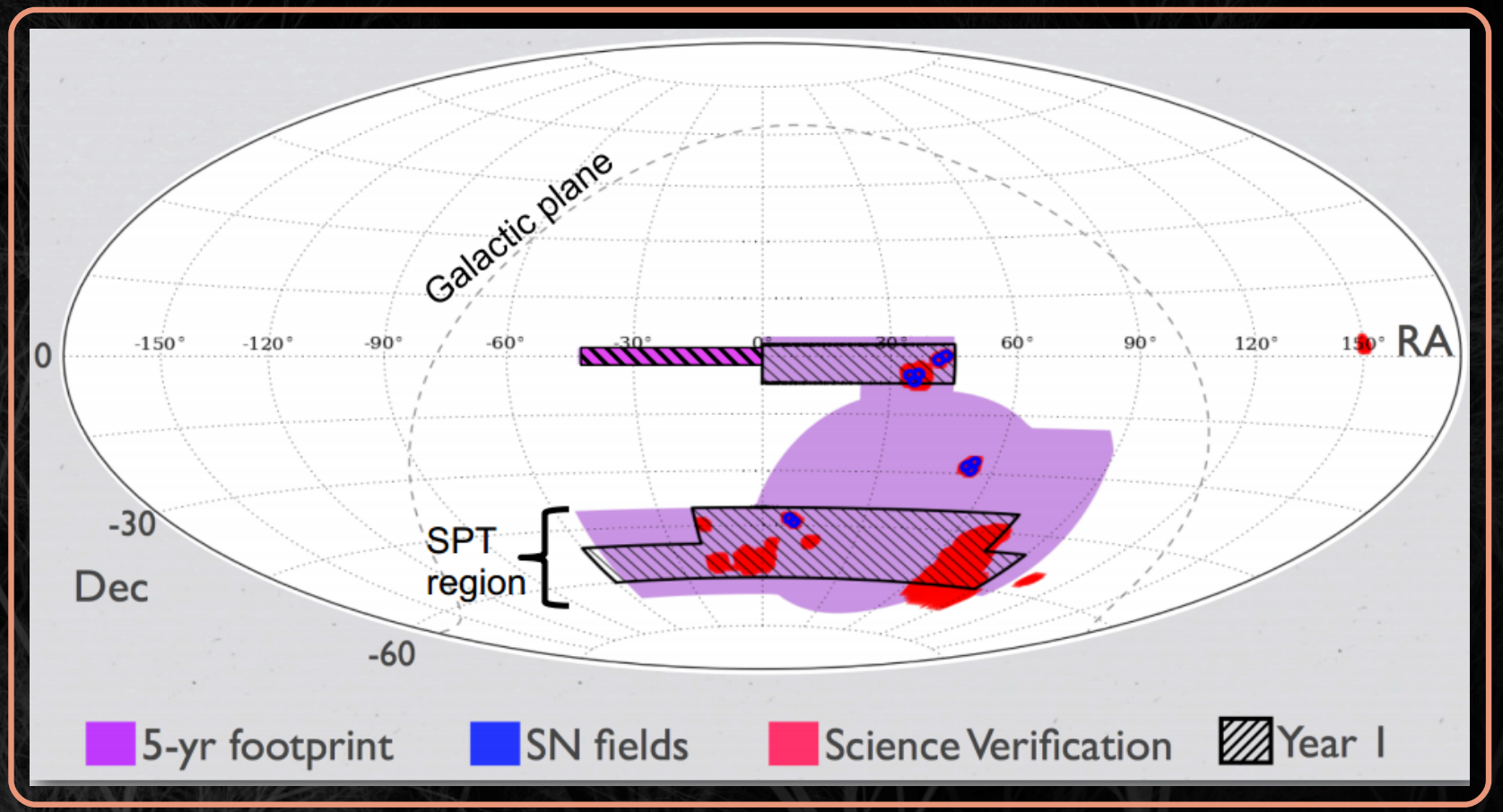
Photo-z PDF application: $N(z)$ and sparse rep.



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

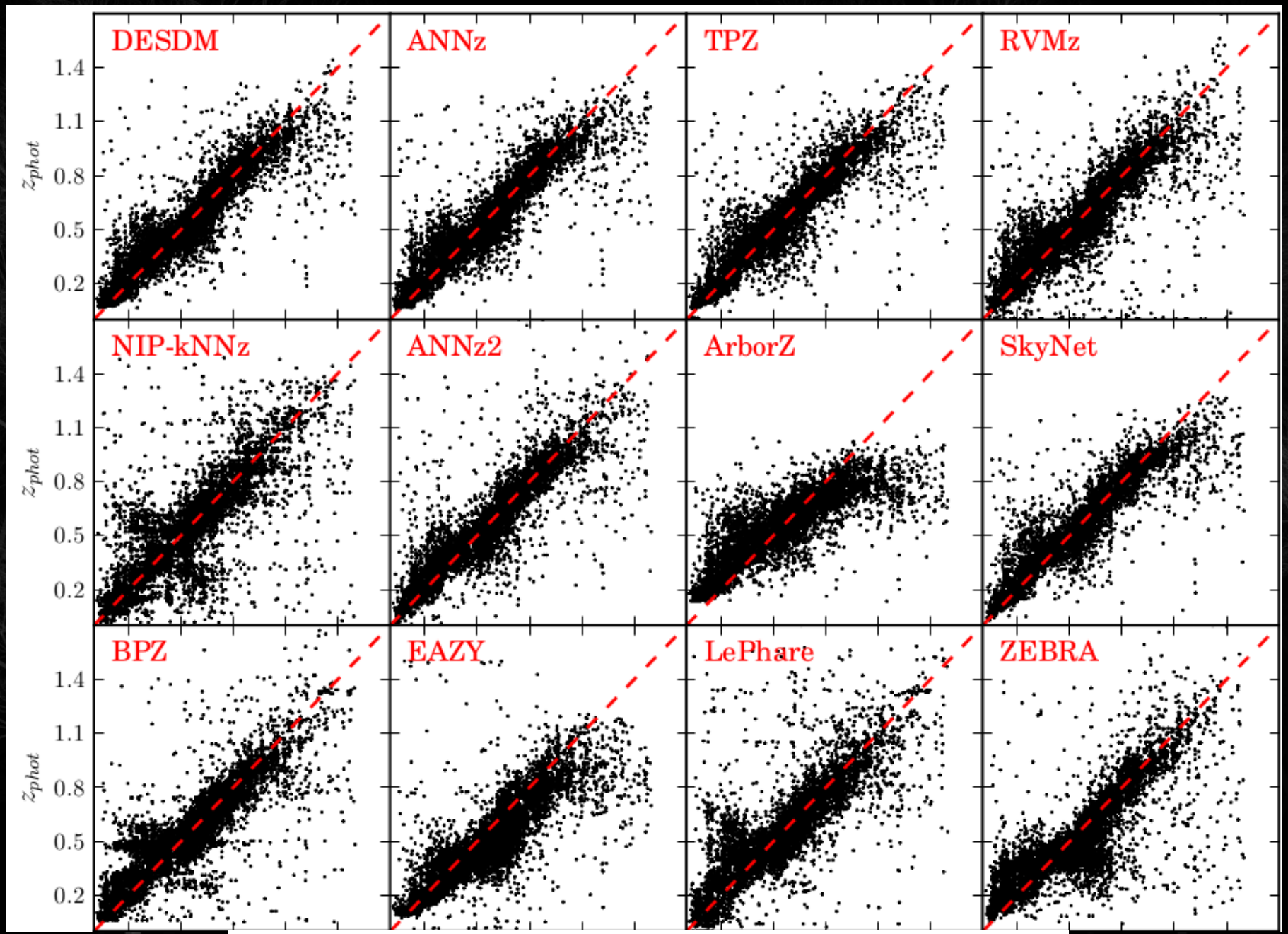
$N(z)$ original (gray) compared to 4 PDF representation methods, Single Gaussian, Monte Carlo, Multi Gaussian, Sparse rep.

Photo-z PDF application: DES SV data



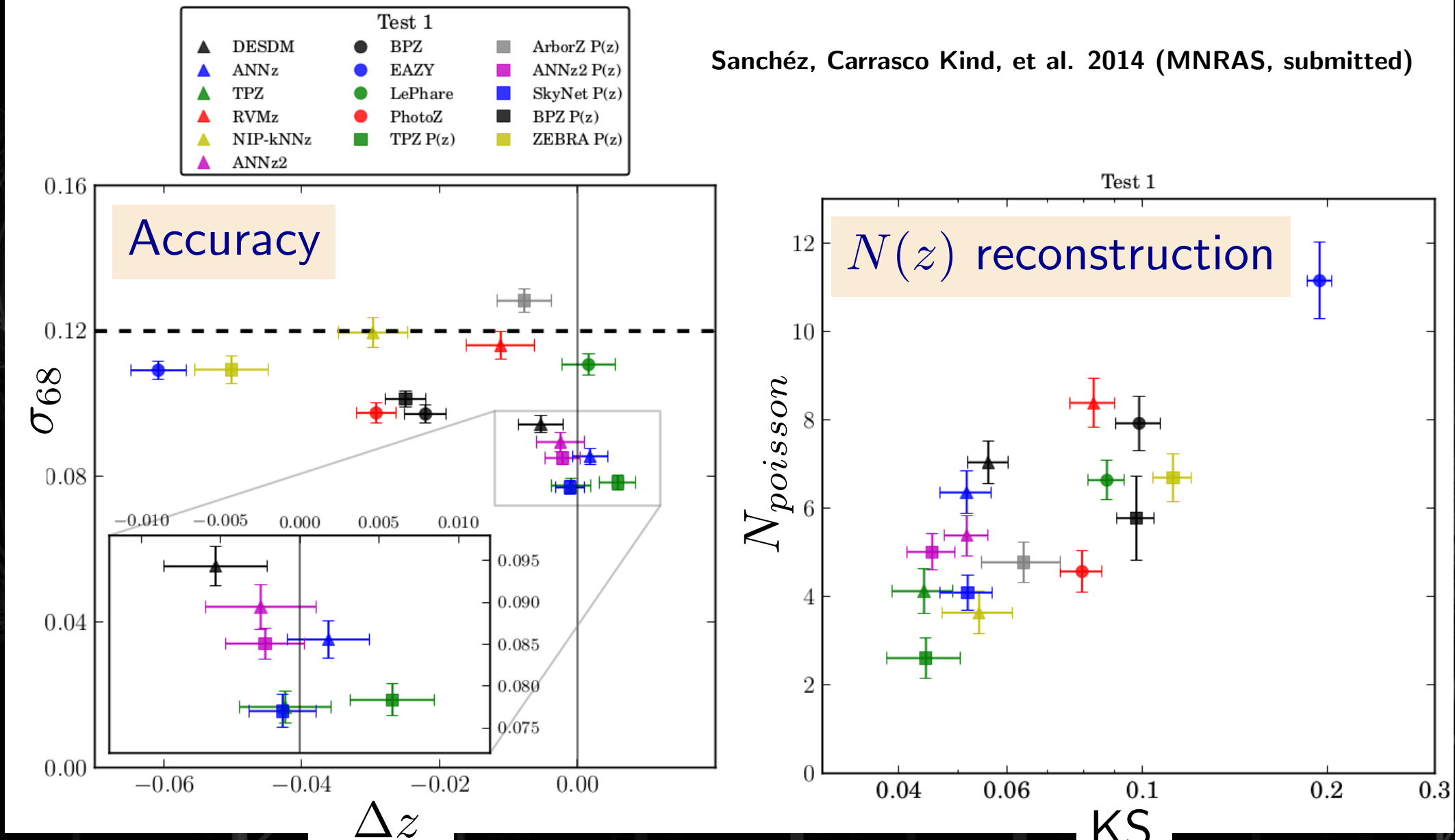
- DES Science Verification data
- Photo-z code comparison and analysis
- TPZ widely used in DES collaboration

Photo-z PDF application: DES SV data



Sanchéz, Carrasco Kind, et al. 2014 (MNRAS, submitted)

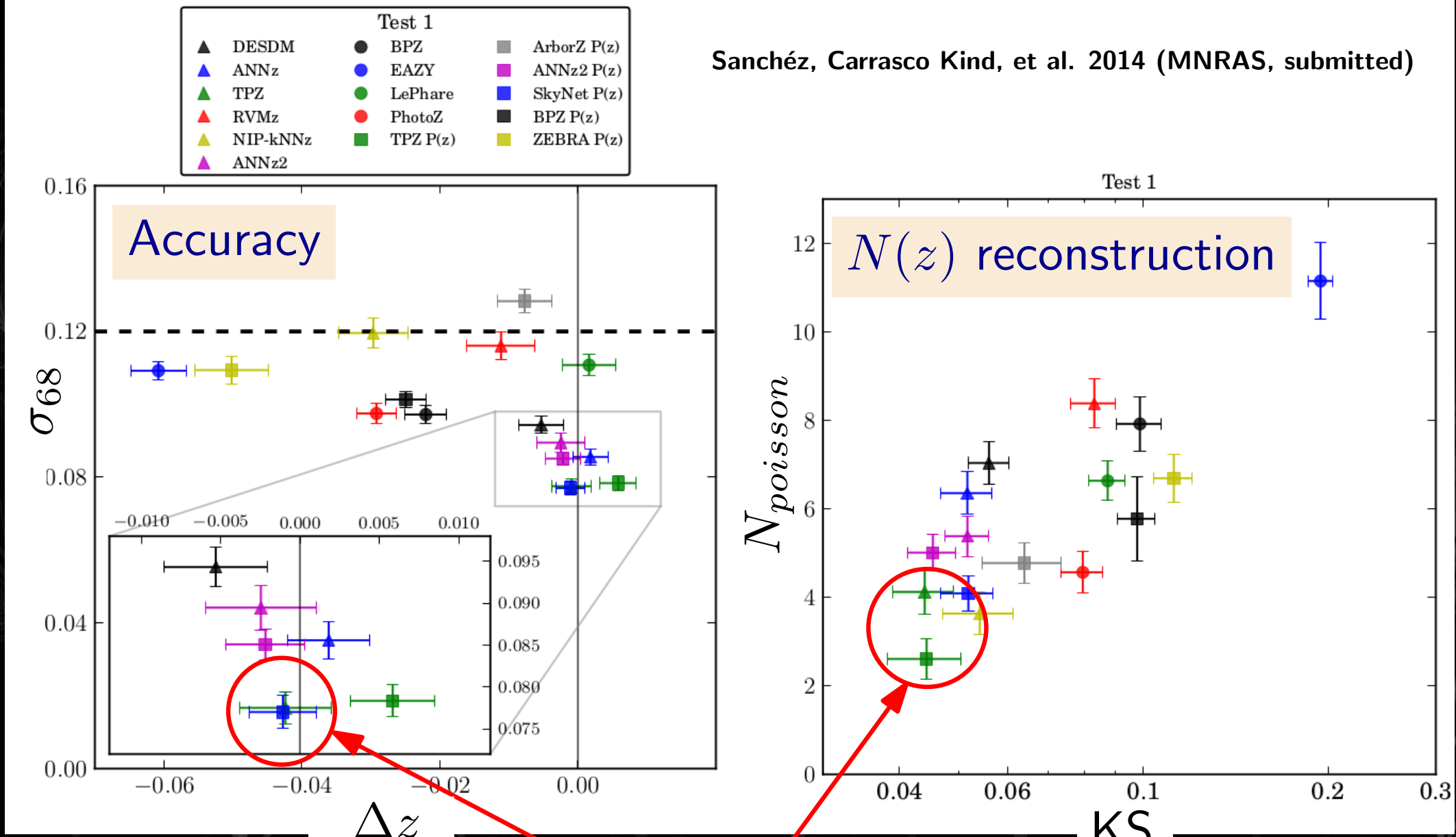
Photo-z PDF application: DES SV data



Sanchéz, Carrasco Kind, et al. 2014 (MNRAS, submitted)

13 photo-z codes comparison

Photo-z PDF application: DES SV data



Sanchéz, Carrasco Kind, et al. 2014 (MNRAS, submitted)

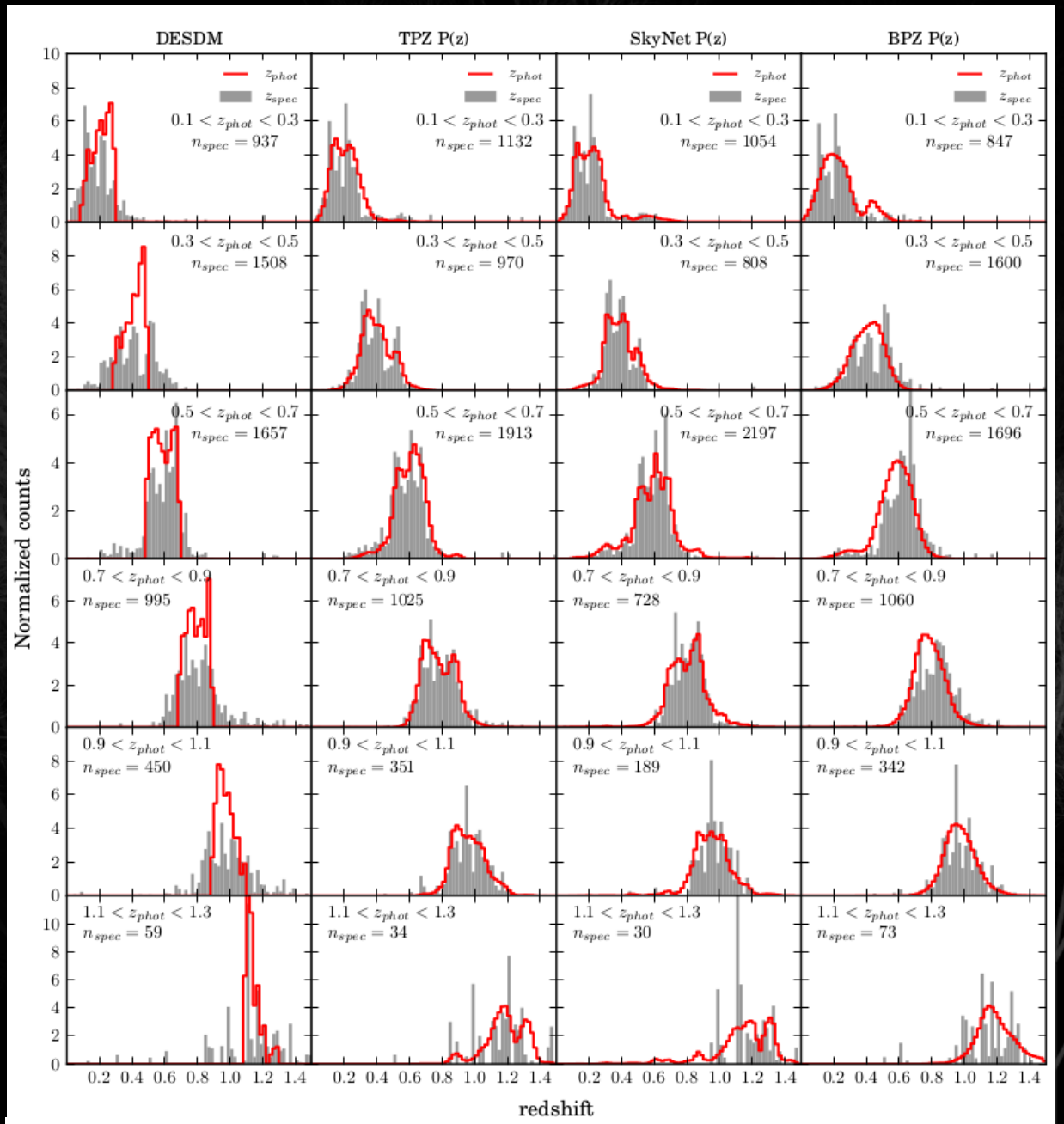
Photo-z PDF application: DES SV data

4 codes recommendation

Default, 2 training and 1 template

PDF methods are better for $N(z)$

Combination methods! (not used here)



Sanchéz, Carrasco Kind, et al. 2014 (MNRAS, submitted)

Photo-z PDF application: Simulated data

We use TPZ to generate photo-z for all $\sim$43 M galaxies.

100,00 for training

5 magnitudes only

Store 43 million PDFs for analysis

No outlier removal

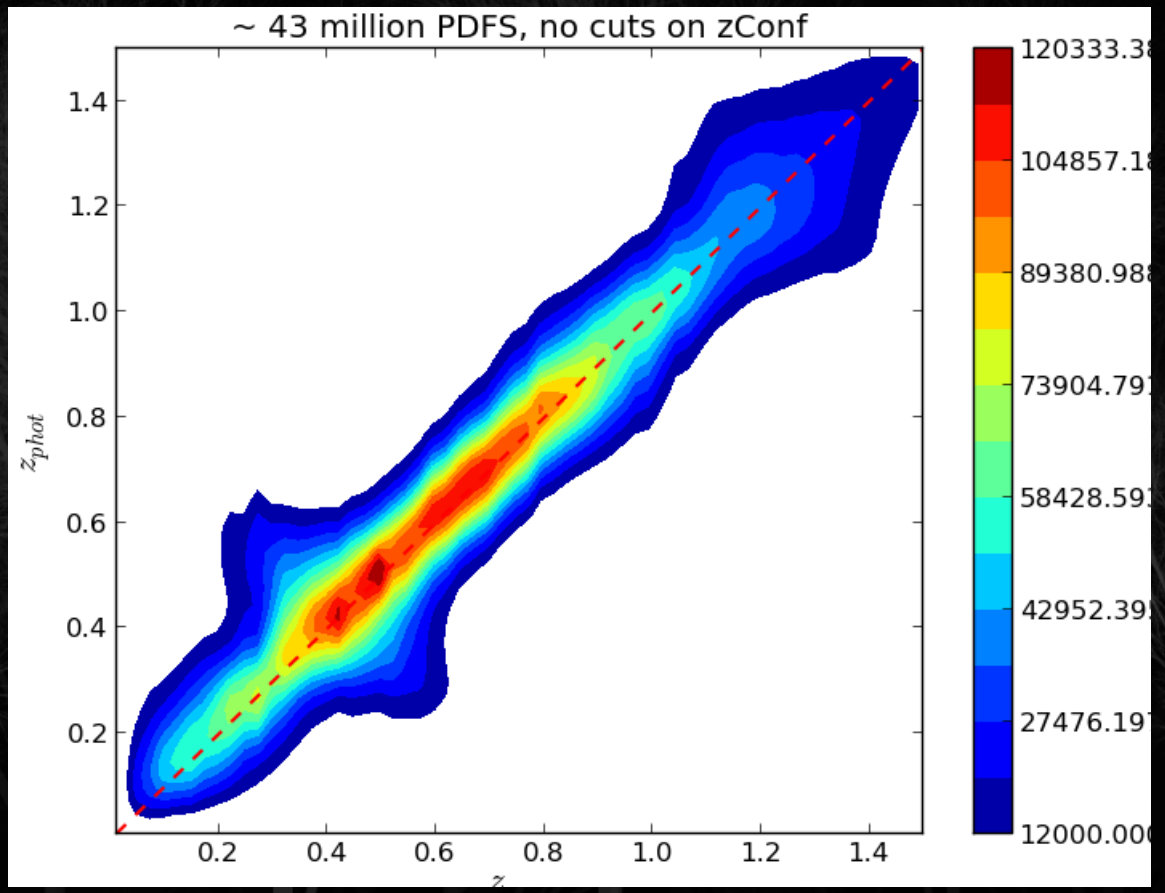


Photo-z PDF application: Simulated data $N(z)$

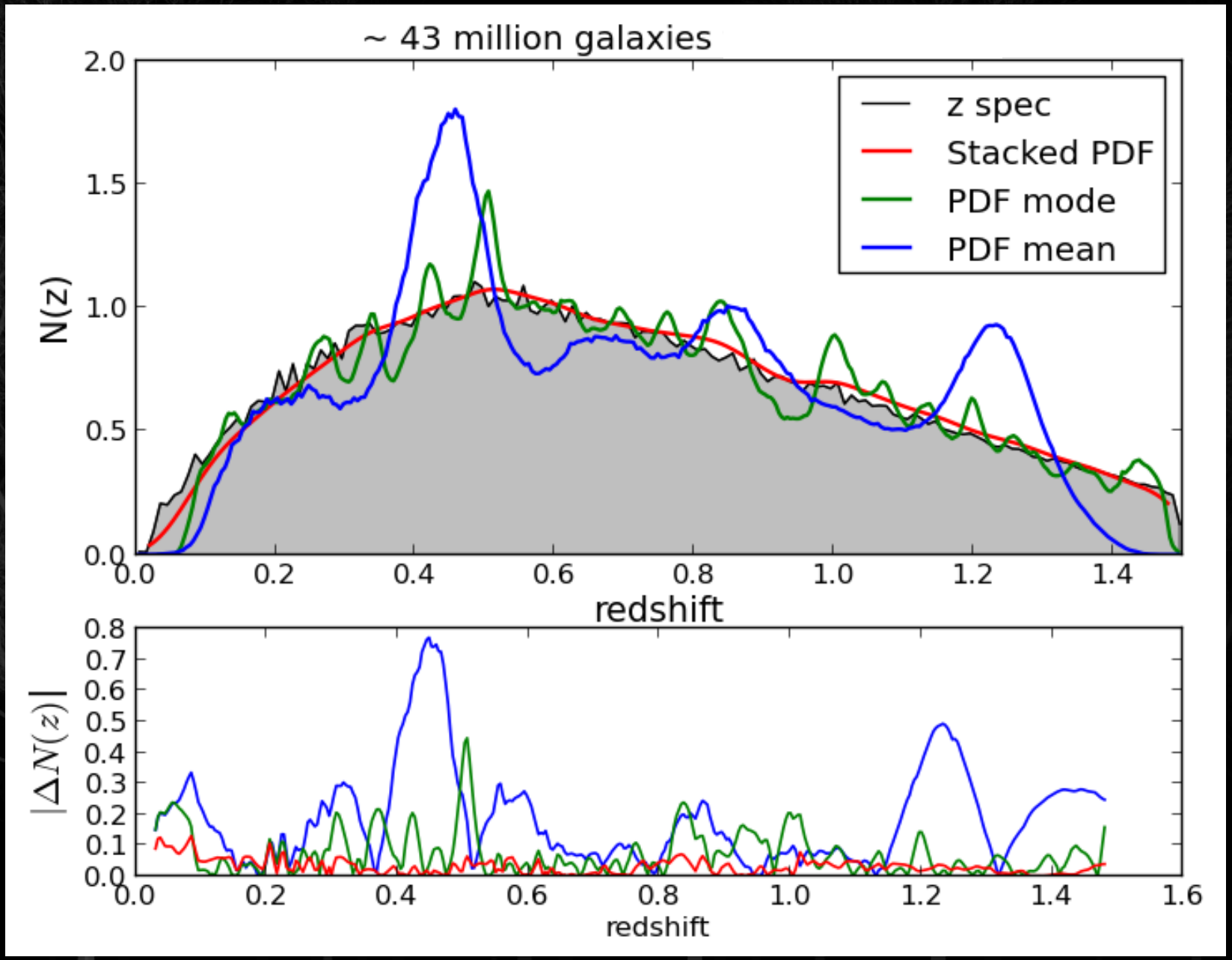


Photo-z PDF application: APS

- The angular power spectrum (APS) C_ℓ 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

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

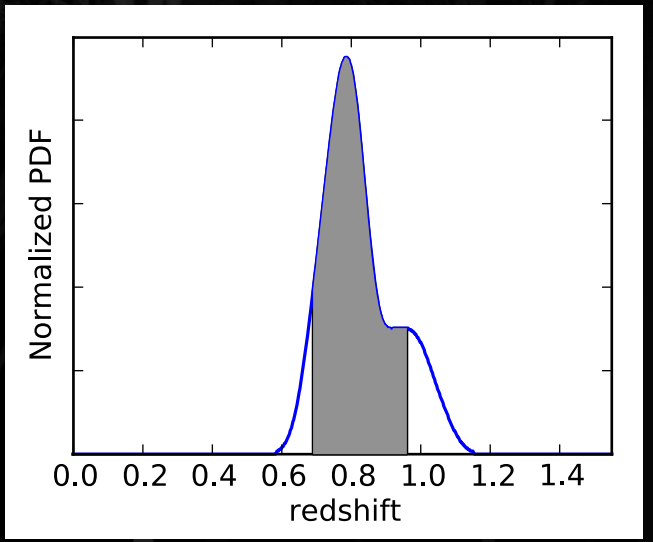


Photo-z PDF application: C_ℓ

The projected 3D $P(k)$ in 2D is given by:

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

$P(k,z), H(z), r(z)$ and b are computed from a cosmological model

Photo-z PDF application: C_ℓ

The projected 3D $P(k)$ in 2D is given by:

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

$P(k,z), H(z), r(z)$ and b are computed from a cosmological model

$\phi(z)$ is the galaxy distribution $N(z)$ and comes from the data

Photo-z PDF application: C_ℓ

The projected 3D $P(k)$ in 2D is given by:

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

$P(k,z), H(z), r(z)$ and b are computed from a cosmological model

$\phi(z)$ is the galaxy distribution $N(z)$ and comes from the data

Fitting using Monte Carlo Markov Chain methods

$$\chi^2(a_p) = \sum_{bb'} (\ln C_b - \ln C_b^T) C_b F_{bb'} C_{b'} (\ln C_{b'} - \ln C_{b'}^T)$$

Outline

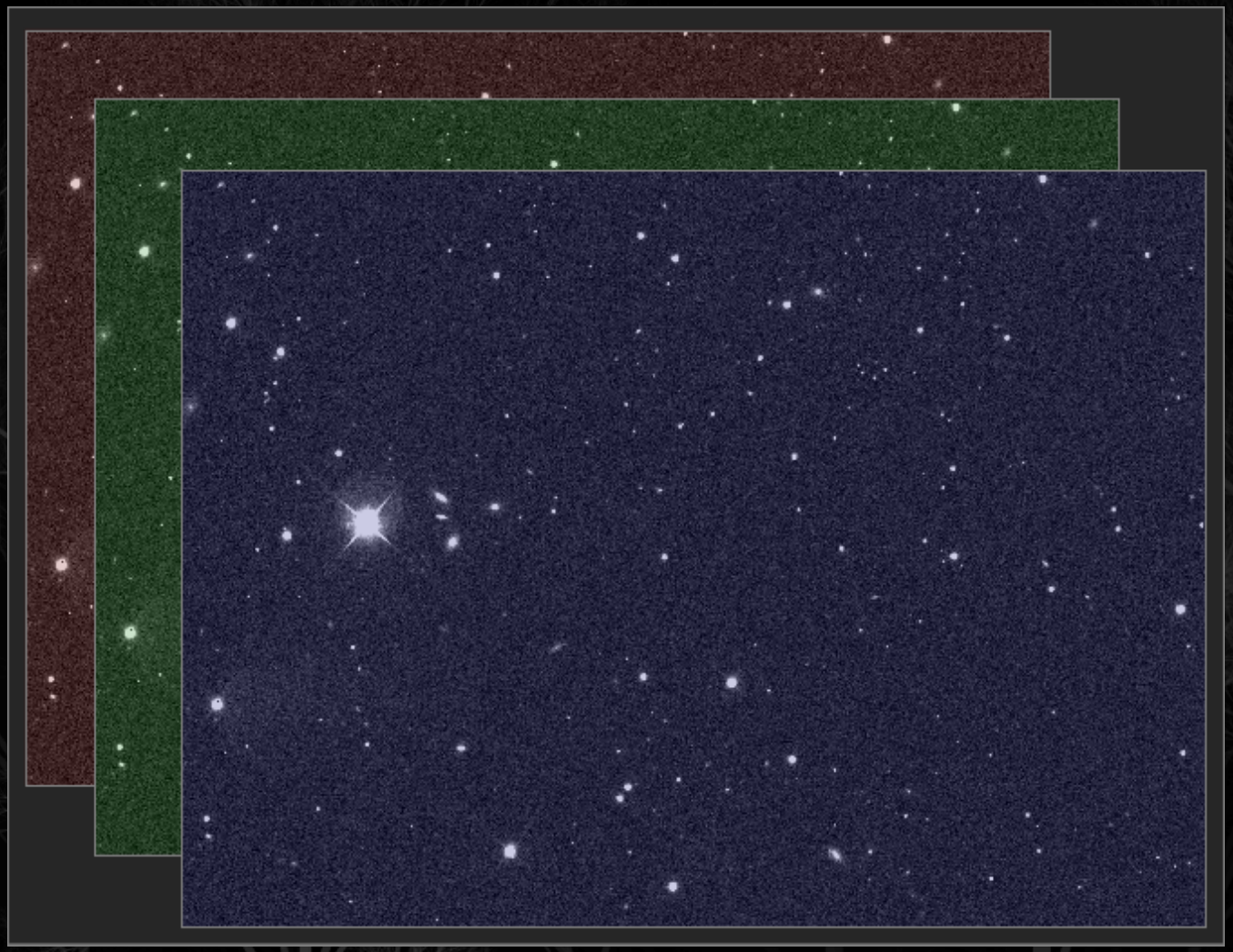
- INTRODUCTION
- HOW TO COMPUTE PHOTO-Z PDFs
- HOW TO COMBINE PHOTO-Z PDFs METHODS
- HOW TO REPRESENT PHOTO-Z PDF
- APPLICATIONS
- CONCLUSIONS AND FUTURE WORK

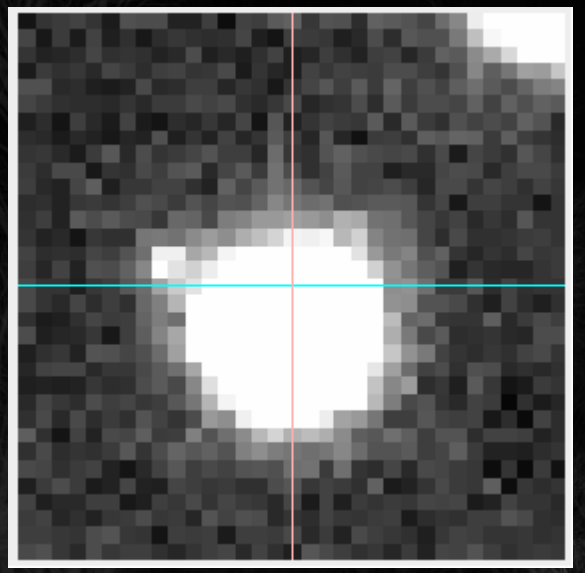
Outline

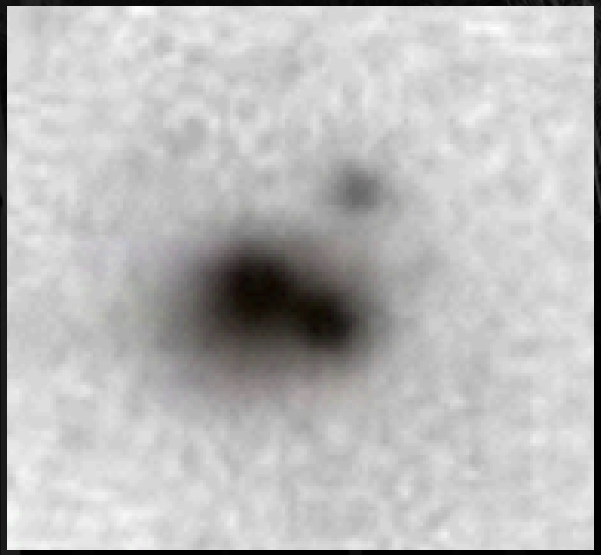
- INTRODUCTION
- HOW TO COMPUTE PHOTO-Z PDFs
- HOW TO COMBINE PHOTO-Z PDFs METHODS
- HOW TO REPRESENT PHOTO-Z PDF
- APPLICATIONS
- CONCLUSIONS AND FUTURE WORK

Current and future work

Future work: Photo-z at pixel level

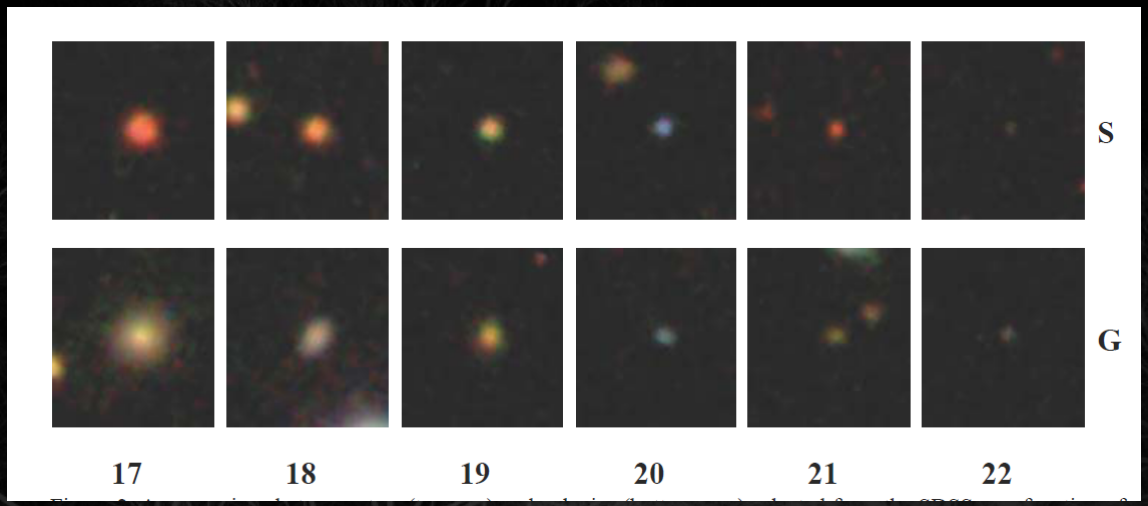






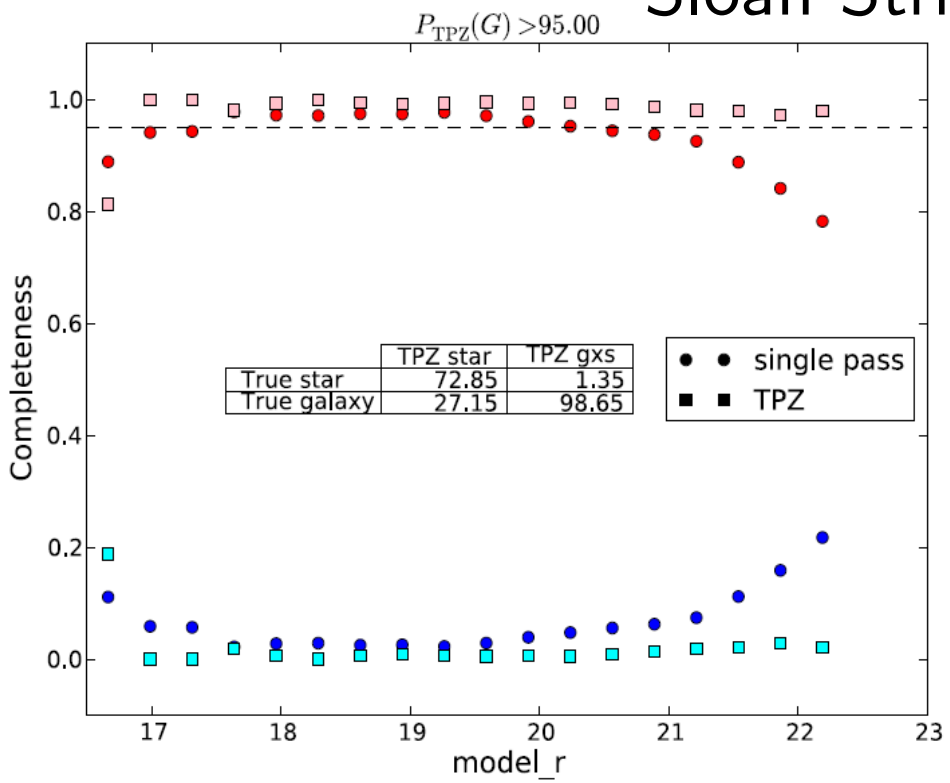
Use information from pixels to compute photo-z, promising for faint and/or blended objects

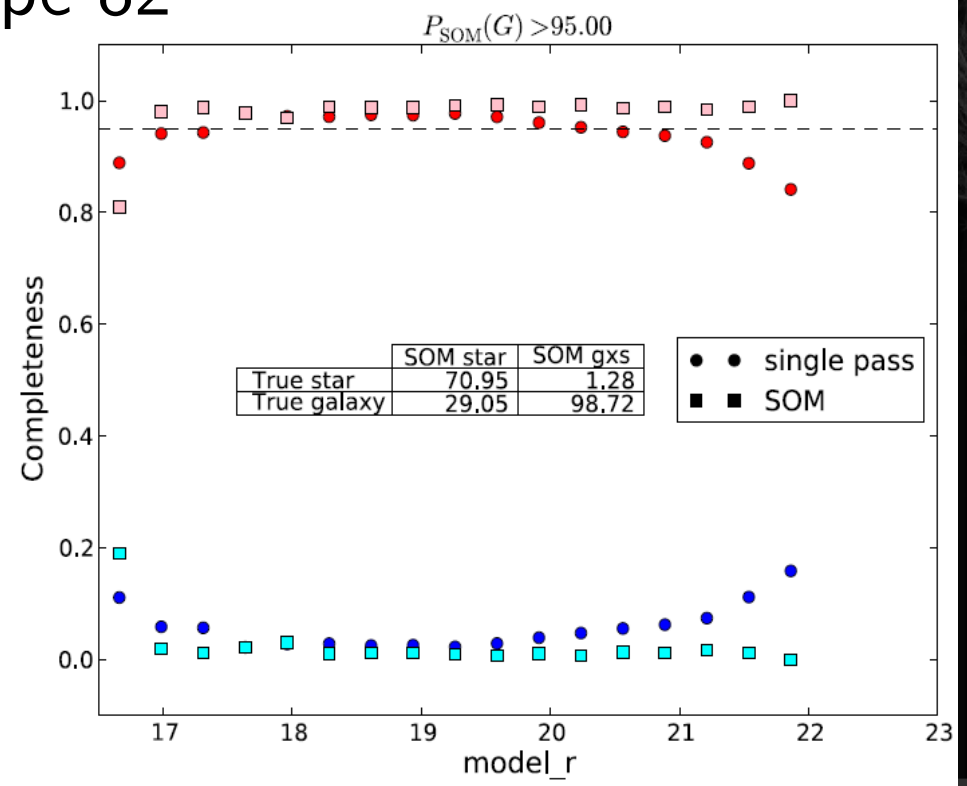
Future work: S/G classification



Machine Learning applications to S/G separation. Challenging for faint objects

Sloan Stripe 82





Future work: Sparse rep. applications

Basis representation for photo-z PDFs

$$\mathbf{pz}_k \approx \mathbf{D} \cdot \boldsymbol{\delta}_k$$

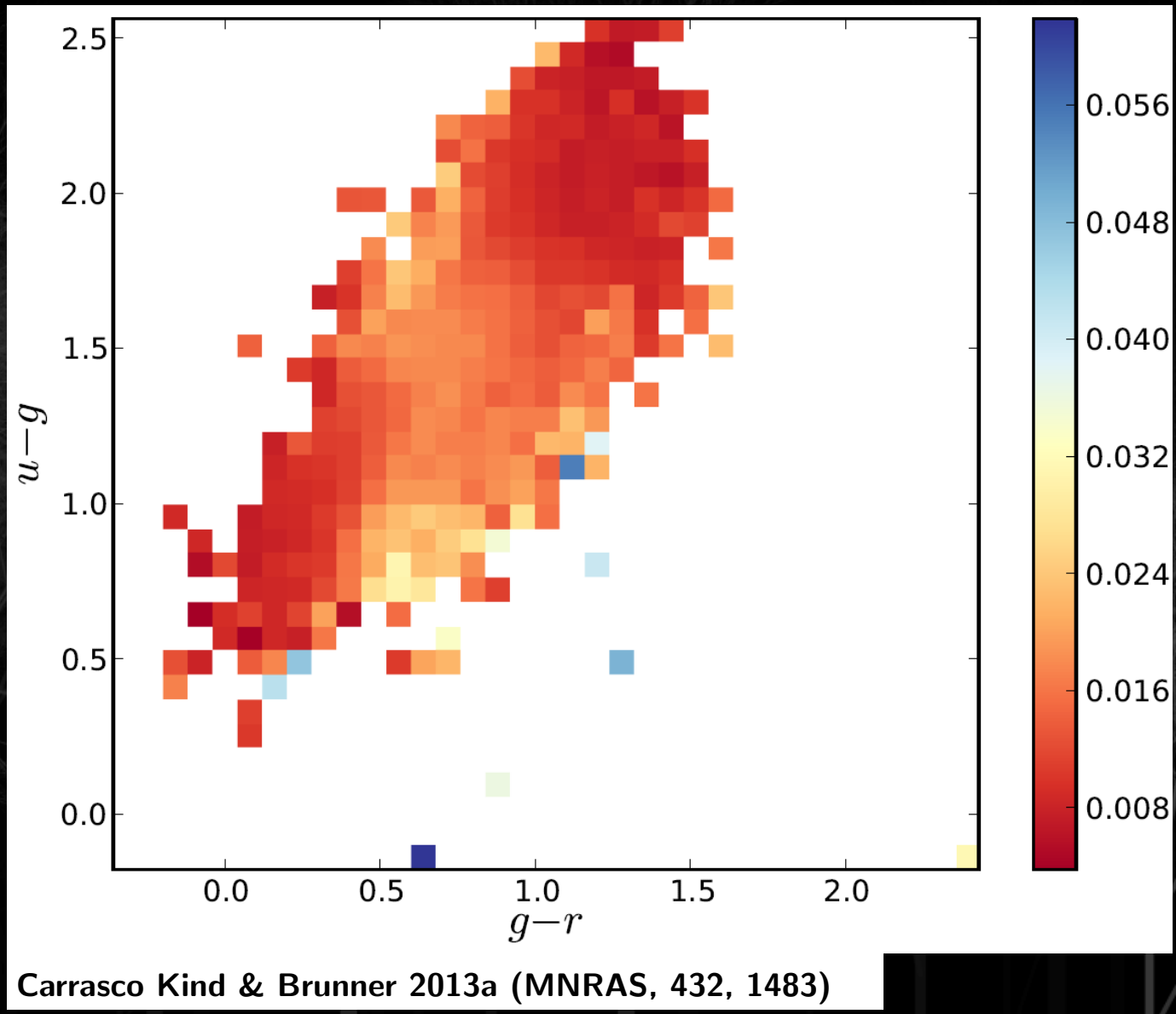
$\mathbf{D}$ is the dictionary, $\boldsymbol{\delta}_k$ is the sparse vector

Extend the sparse representation to other cosmological applications:

- Cross-correlation methods to recover $N(z)$
- Include $N(z)$ errors in cosmological fitting
- Others...

Future work: Spectroscopic calibration

- Map of performance using two most important colors
- The redder the better
- Bimodality of SDSS galaxies
- Target follow up observations



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

How to maximize telescope time?

Conclusions

✓ We have developed and end-to-end framework for probabilistic photo-z

✓ Two state-of-the-art ML approaches. Most accurate photo-z

✓ Advanced Bayesian model combination to exhaust information

✓ Highly compressed PDF representation

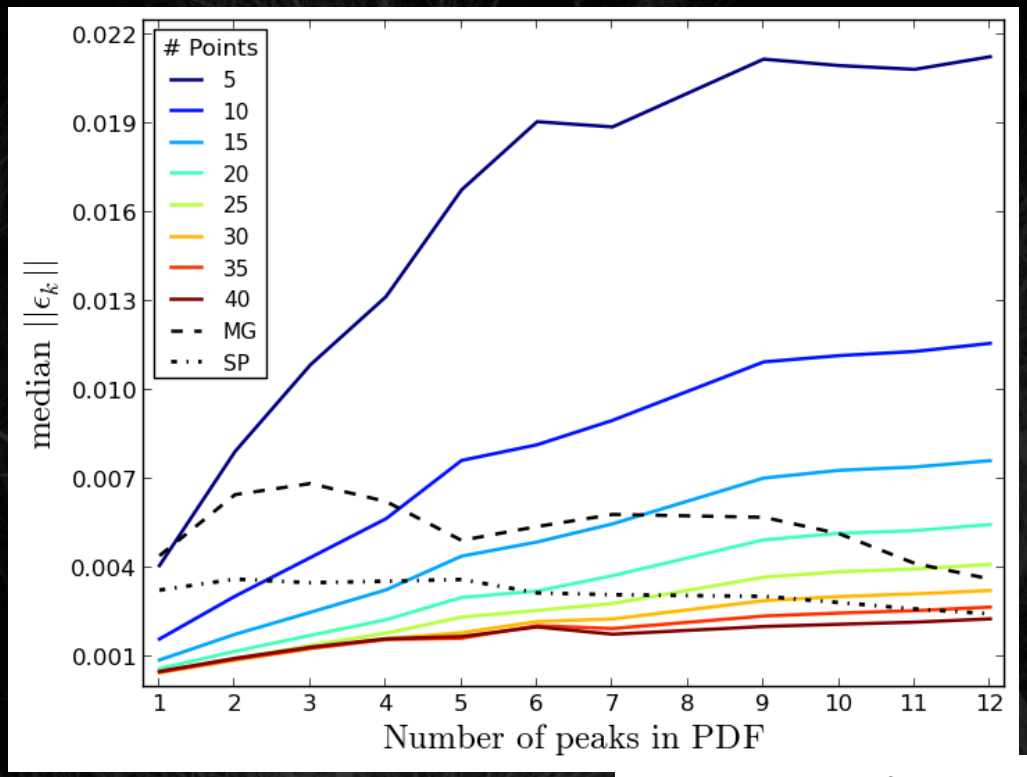
✓ Valuable contribution to the field with many applications

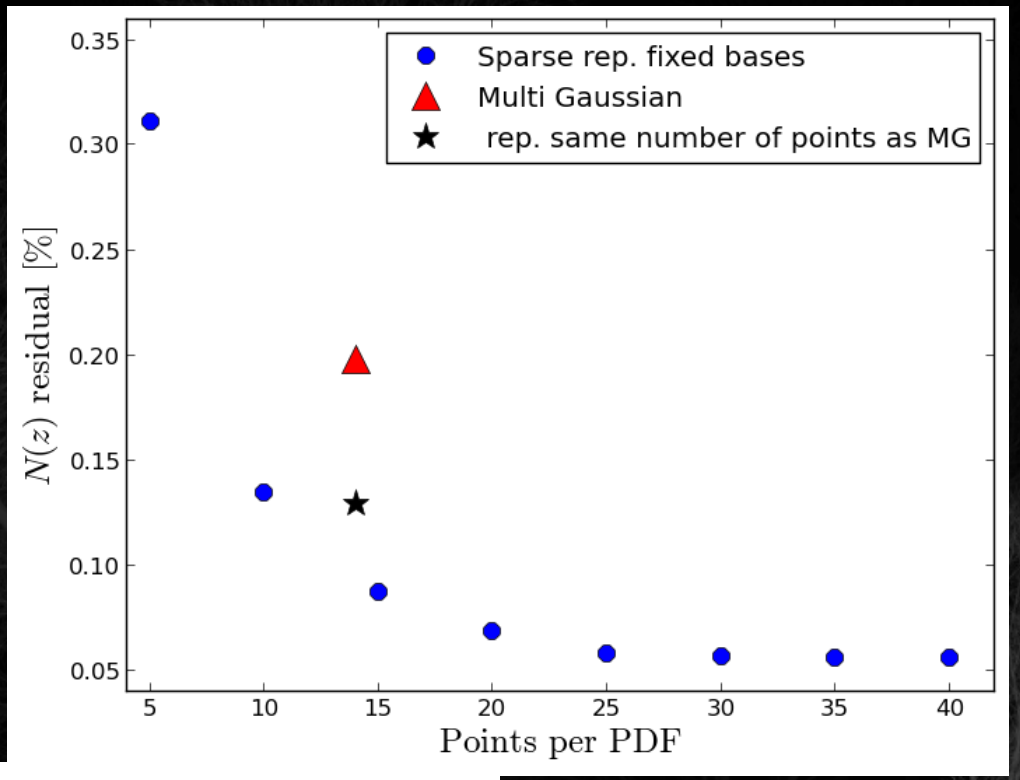
✓ Better constrains on models of galaxy formation and evolution


Thanks!

EXTRA SLIDES

Photo-z PDF storage: Results



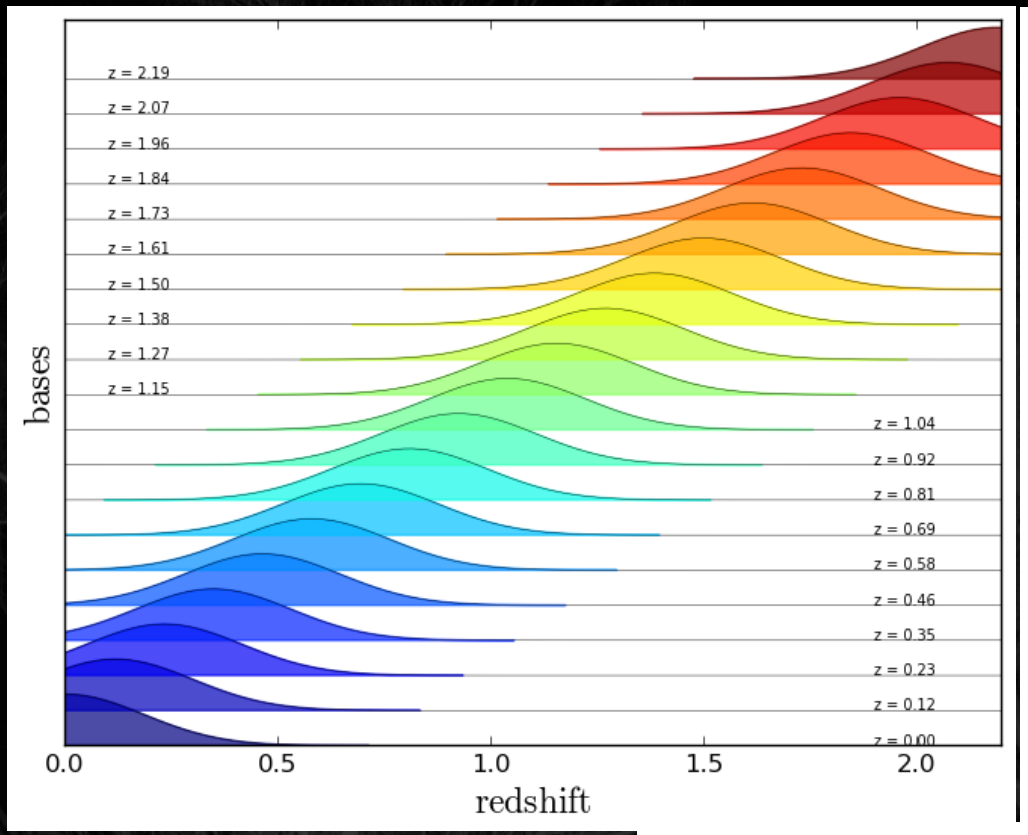


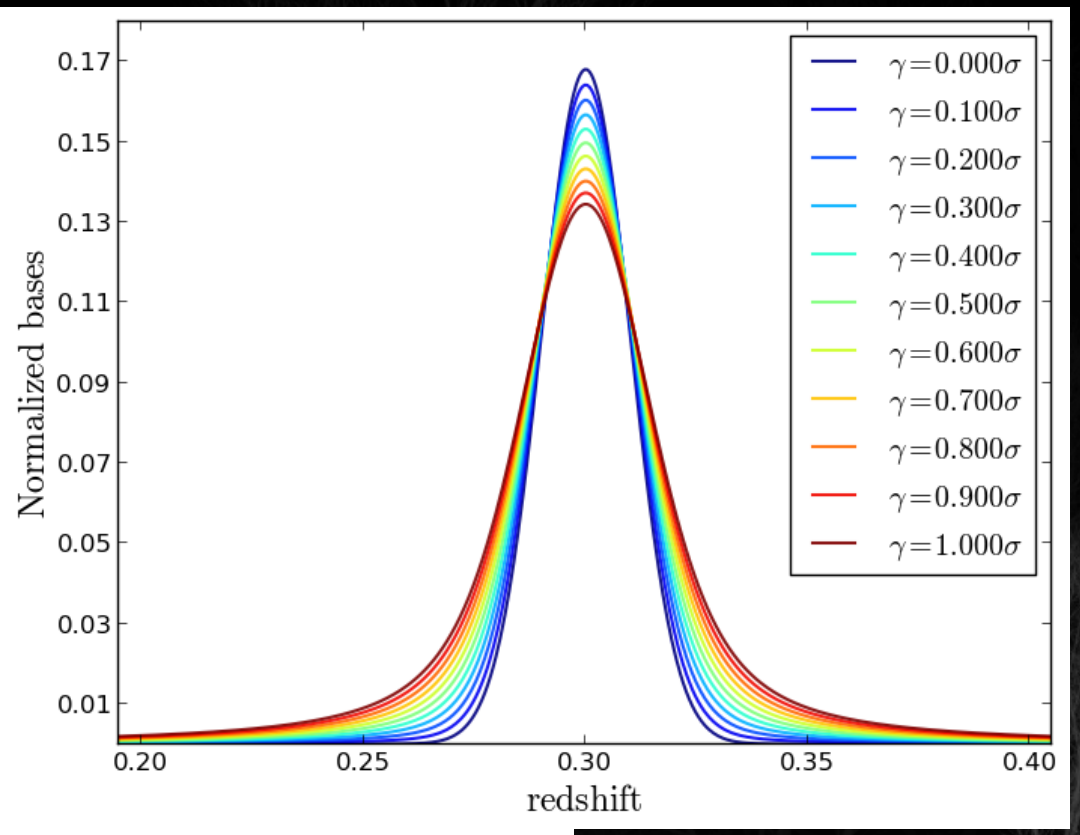
Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)

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)

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 estimation: Error and validation

Out of Bag (cross-validation) data used to validate trees/maps

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

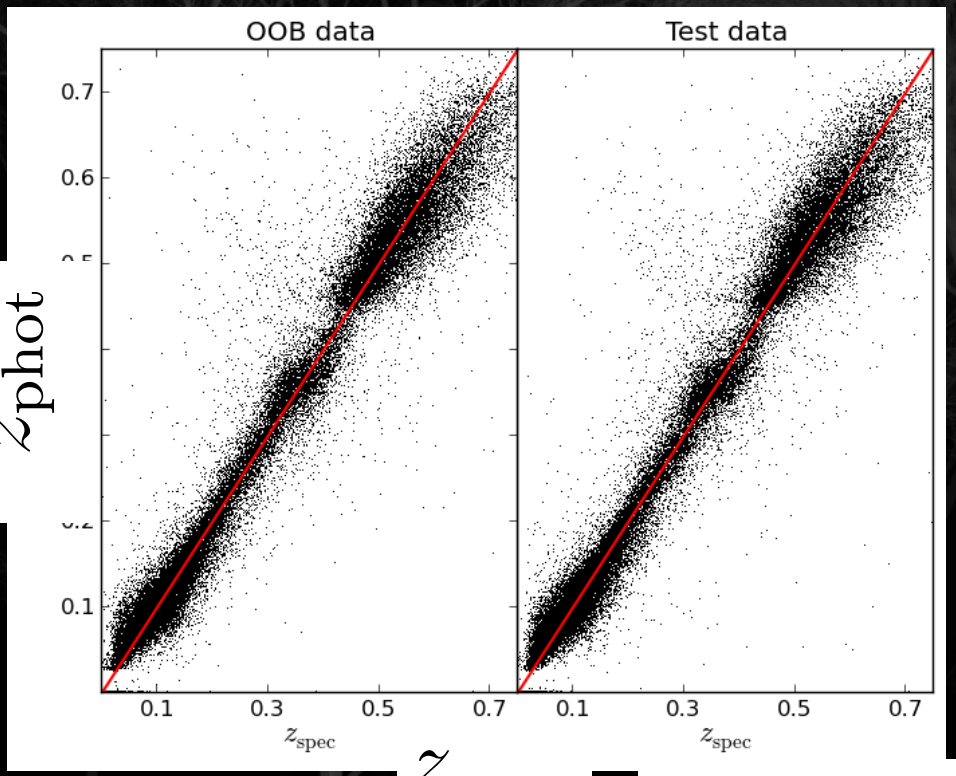
We can learn from the cross-validation data!

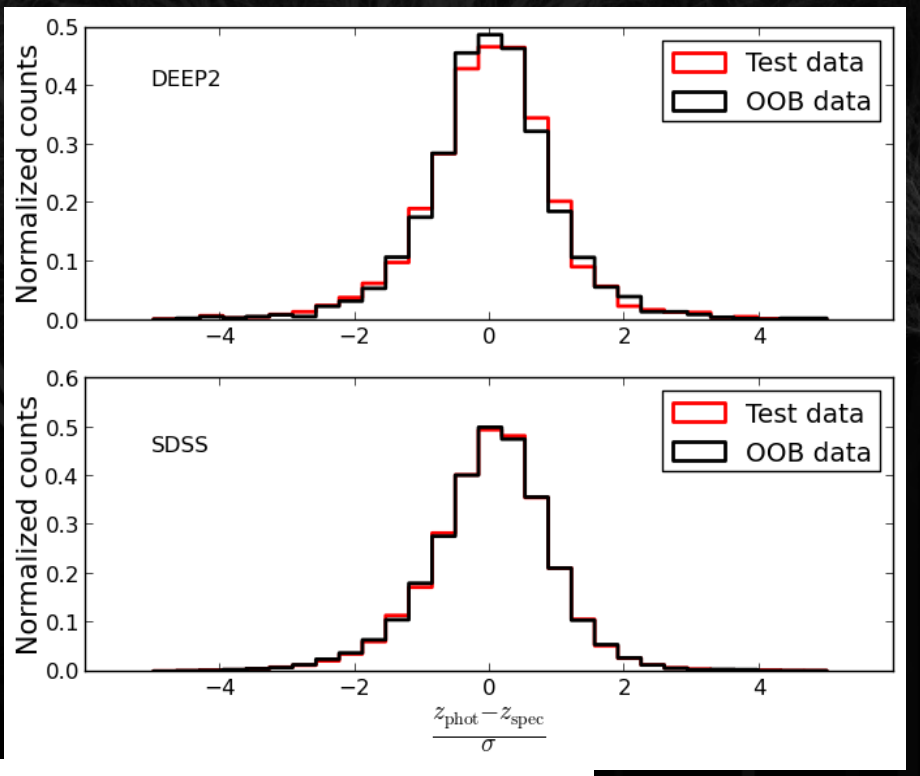
Photo-z PDF estimation: Error and validation

Out of Bag (cross-validation) 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, 442, 3380)

Photo-z PDF estimation: SOM 2D toy example

Suppose 2D data distributed in a given space

De-project the data in a 2D map

Each cell will contain objects with similar properties

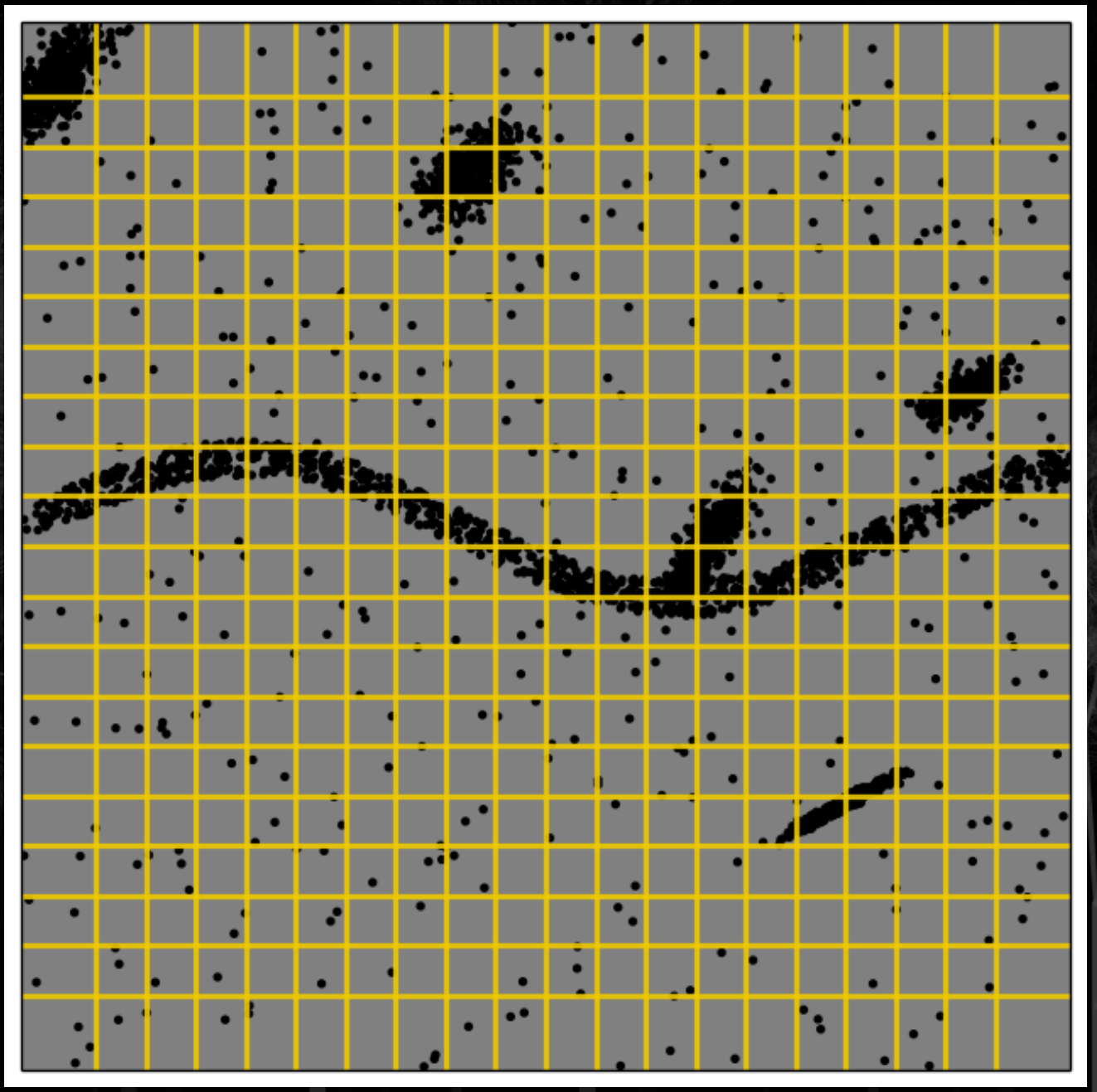
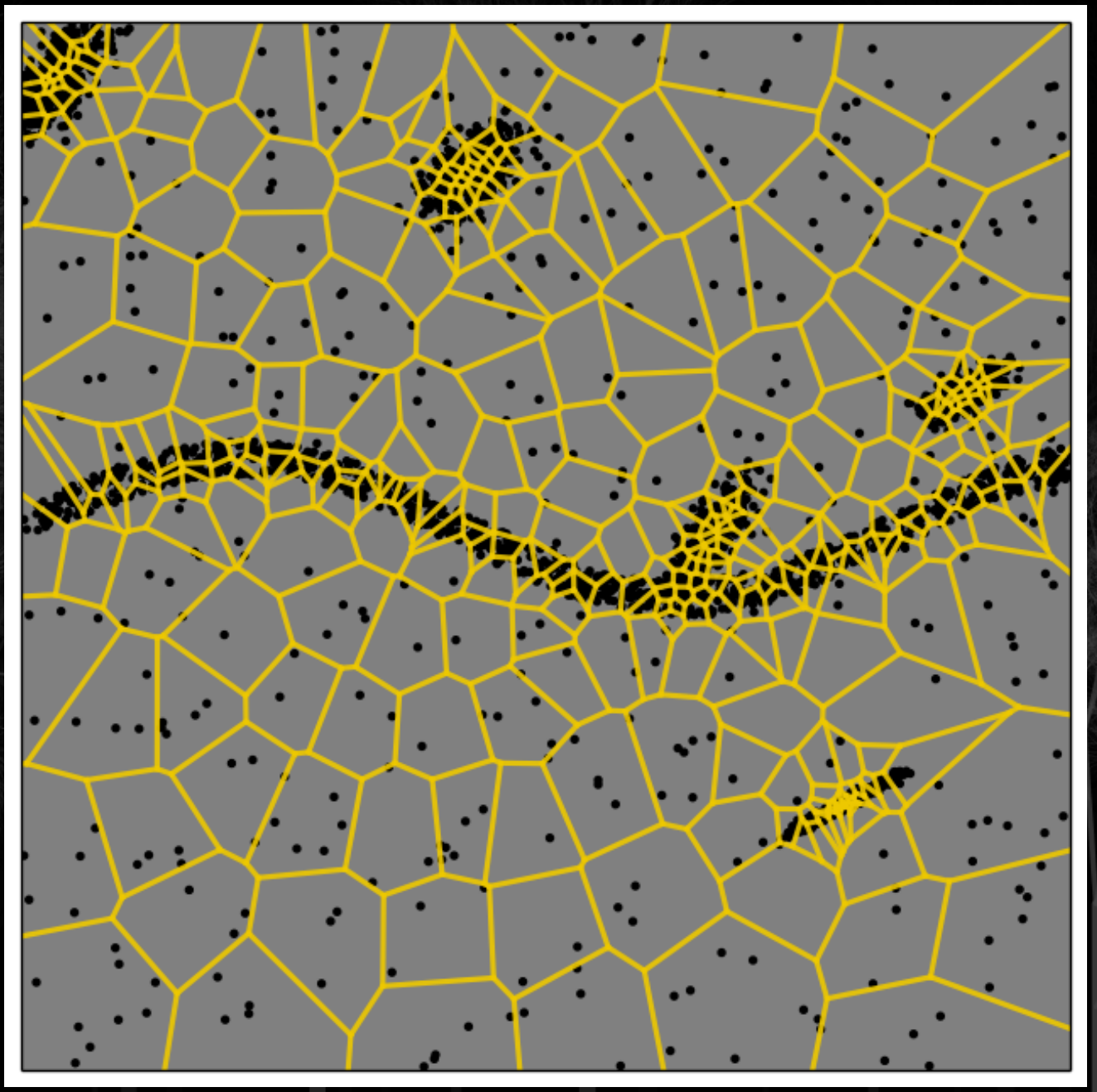


Photo-z PDF estimation: SOM 2D toy example

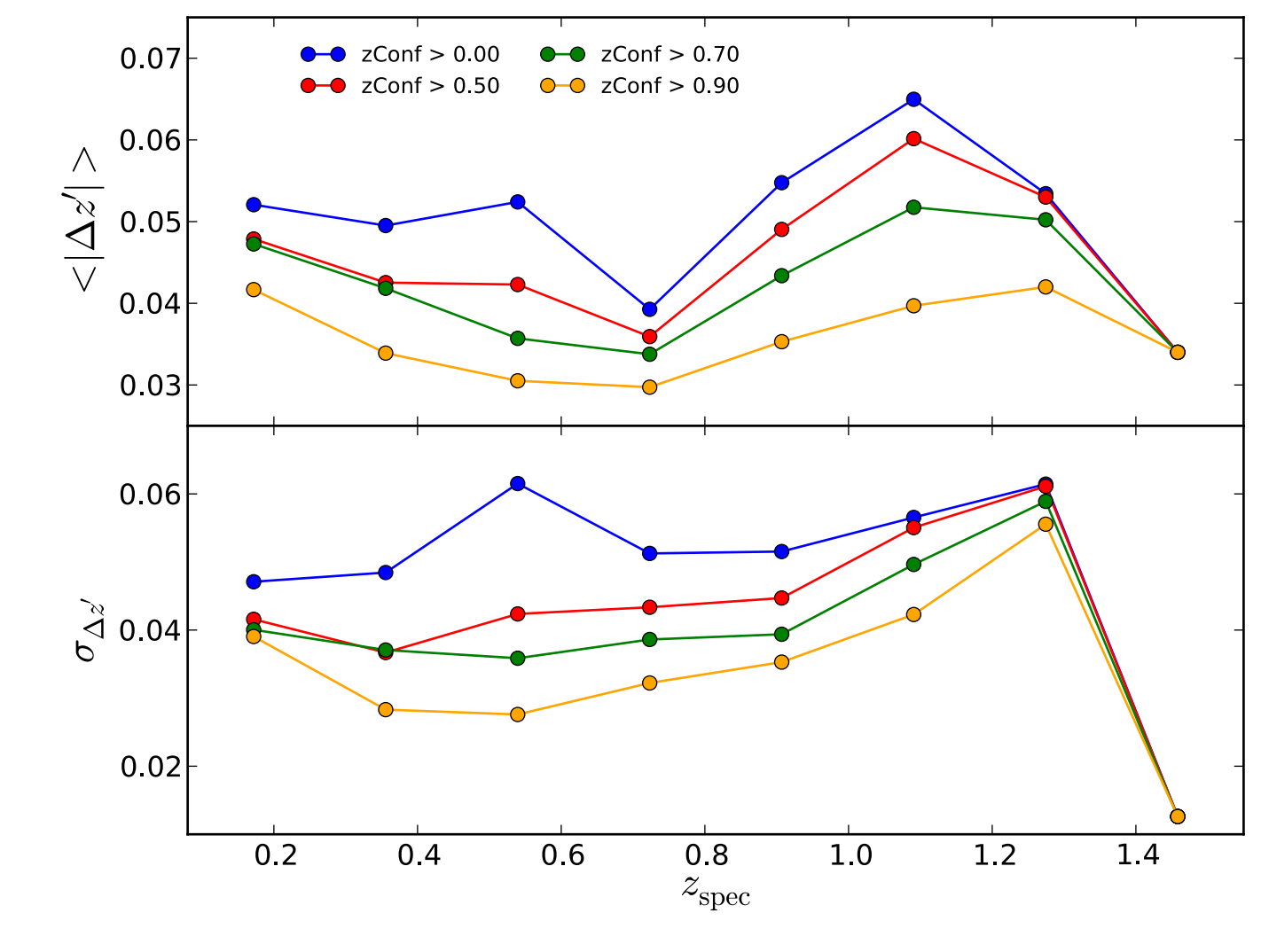


Suppose 2D data distributed in a given space

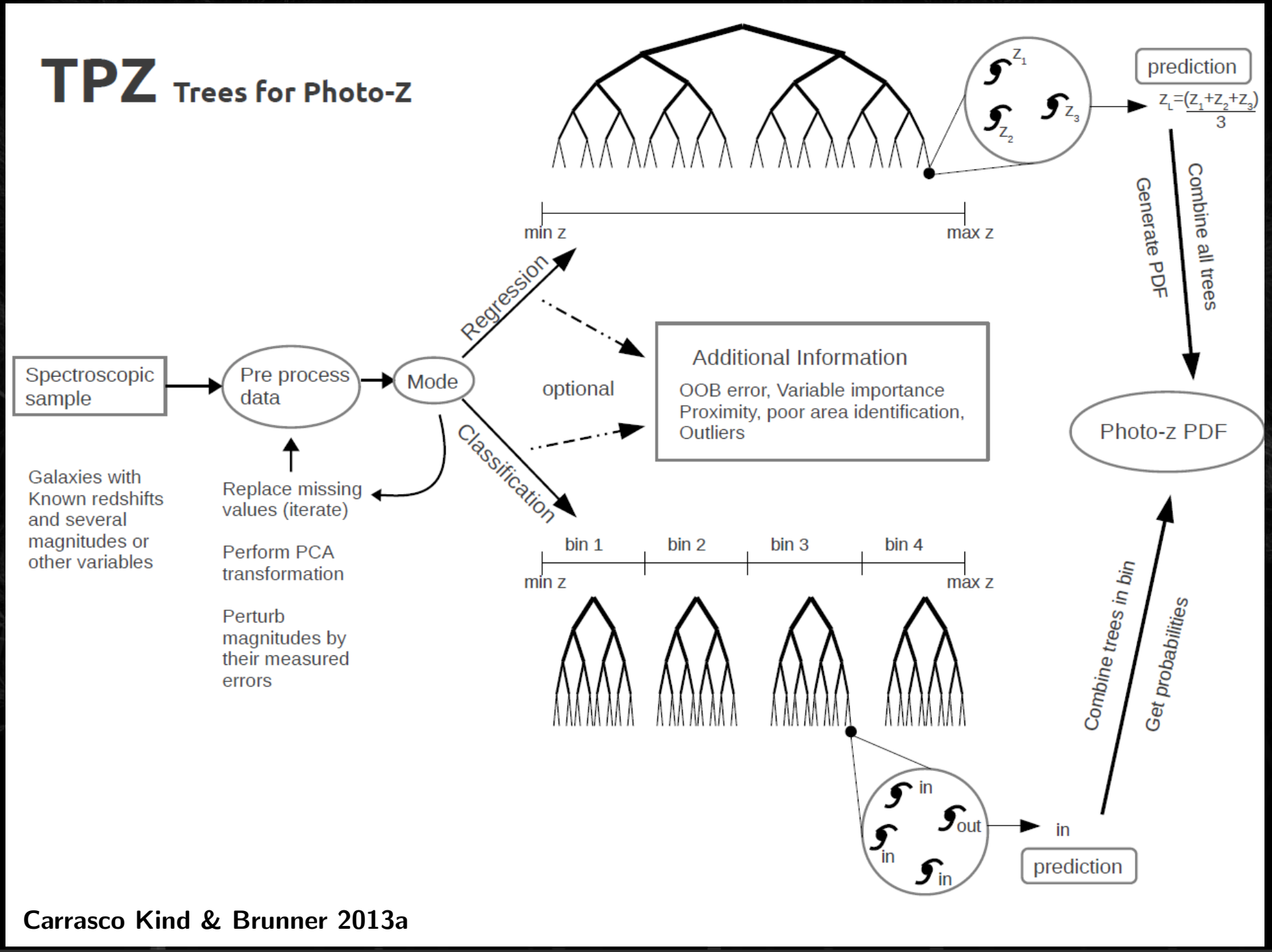
De-project the data in a 2D map

Each cell will contain objects with similar properties

Metrics vs. zConf

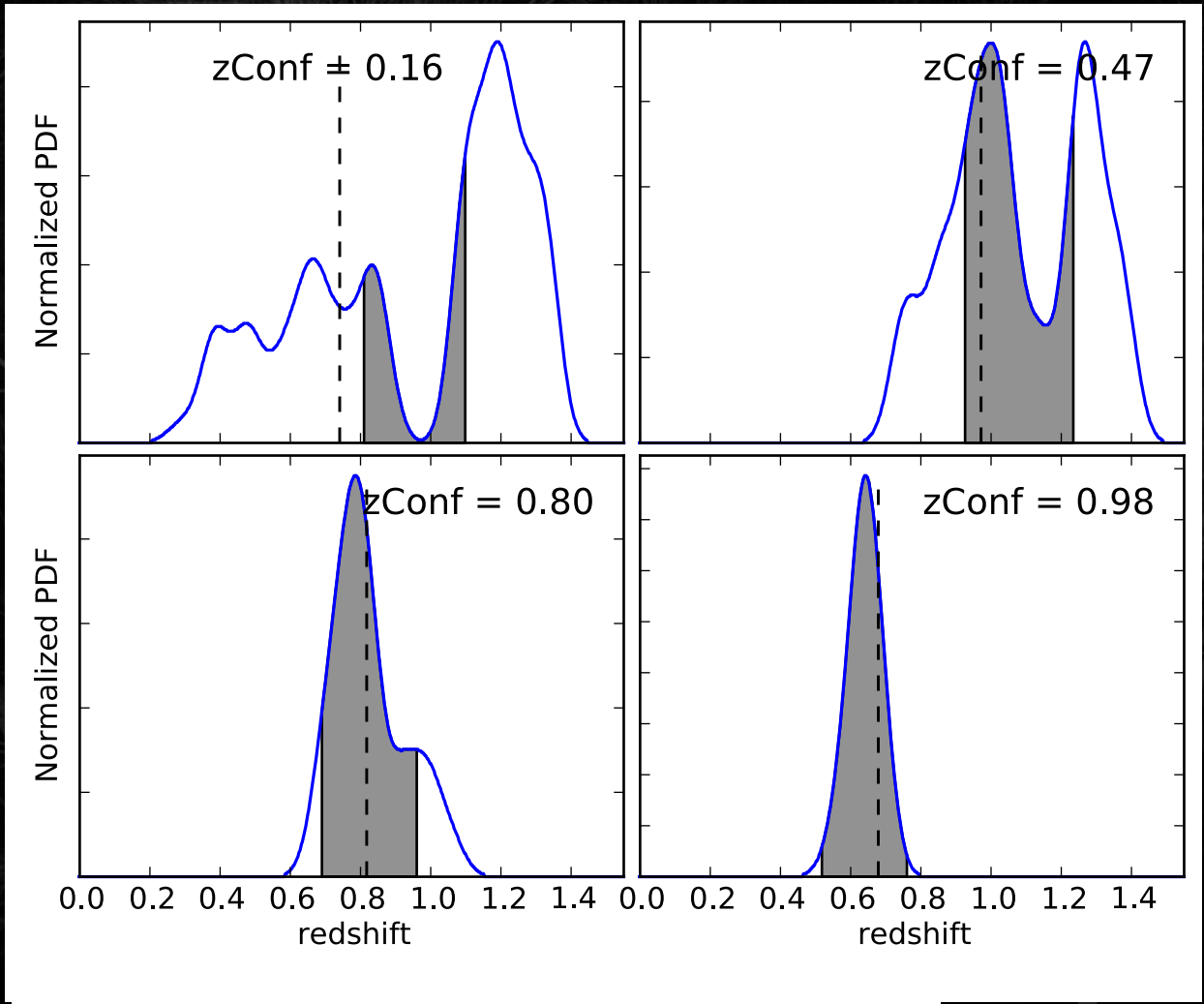


TPZ : Scheme



Carrasco Kind & Brunner 2013a

Photo-z PDF estimation: TPZ PDFs



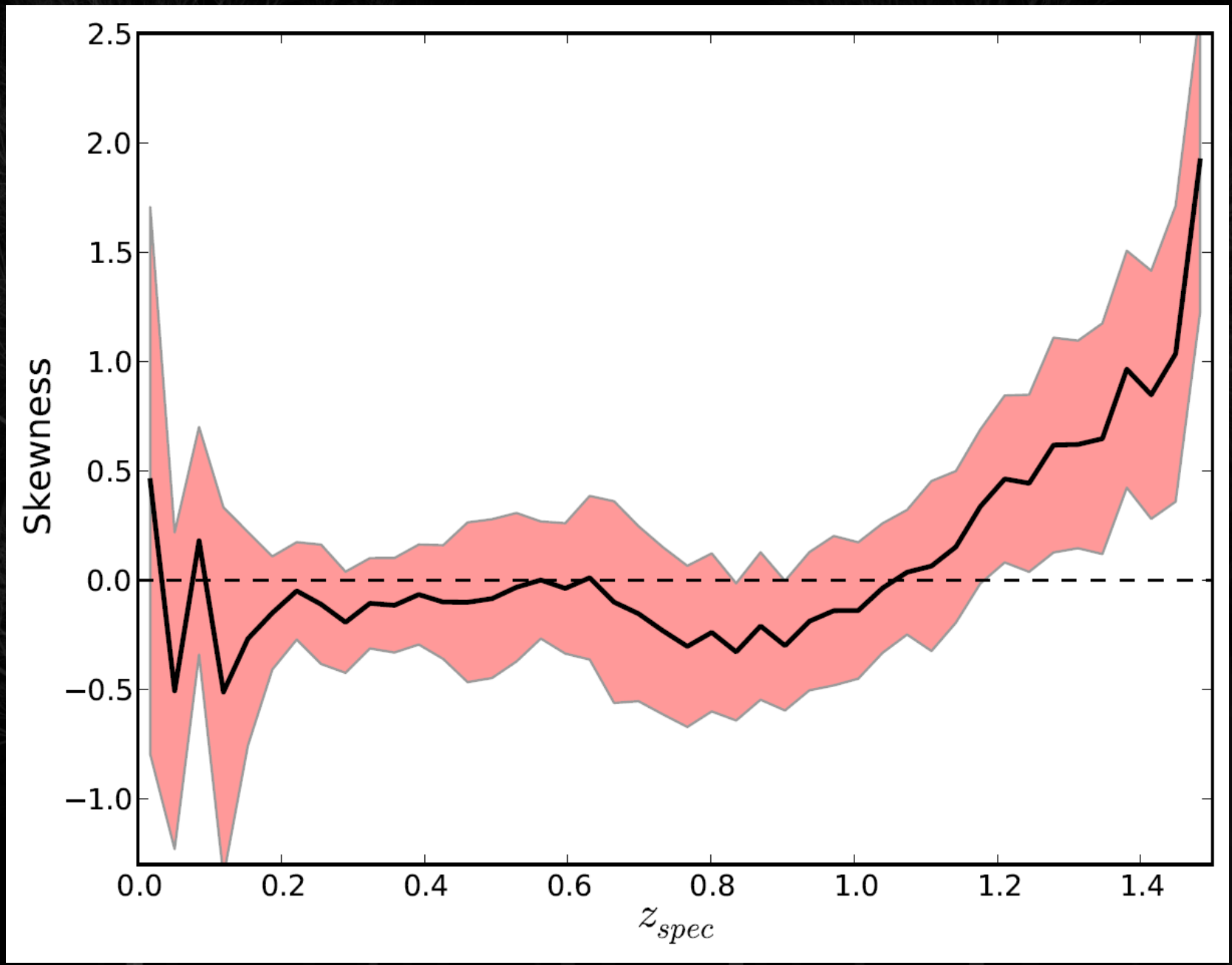
Carrasco Kind & Brunner 2013a (MNRAS, 432, 1483)

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

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

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

Skewness of DEEP2 PDFs



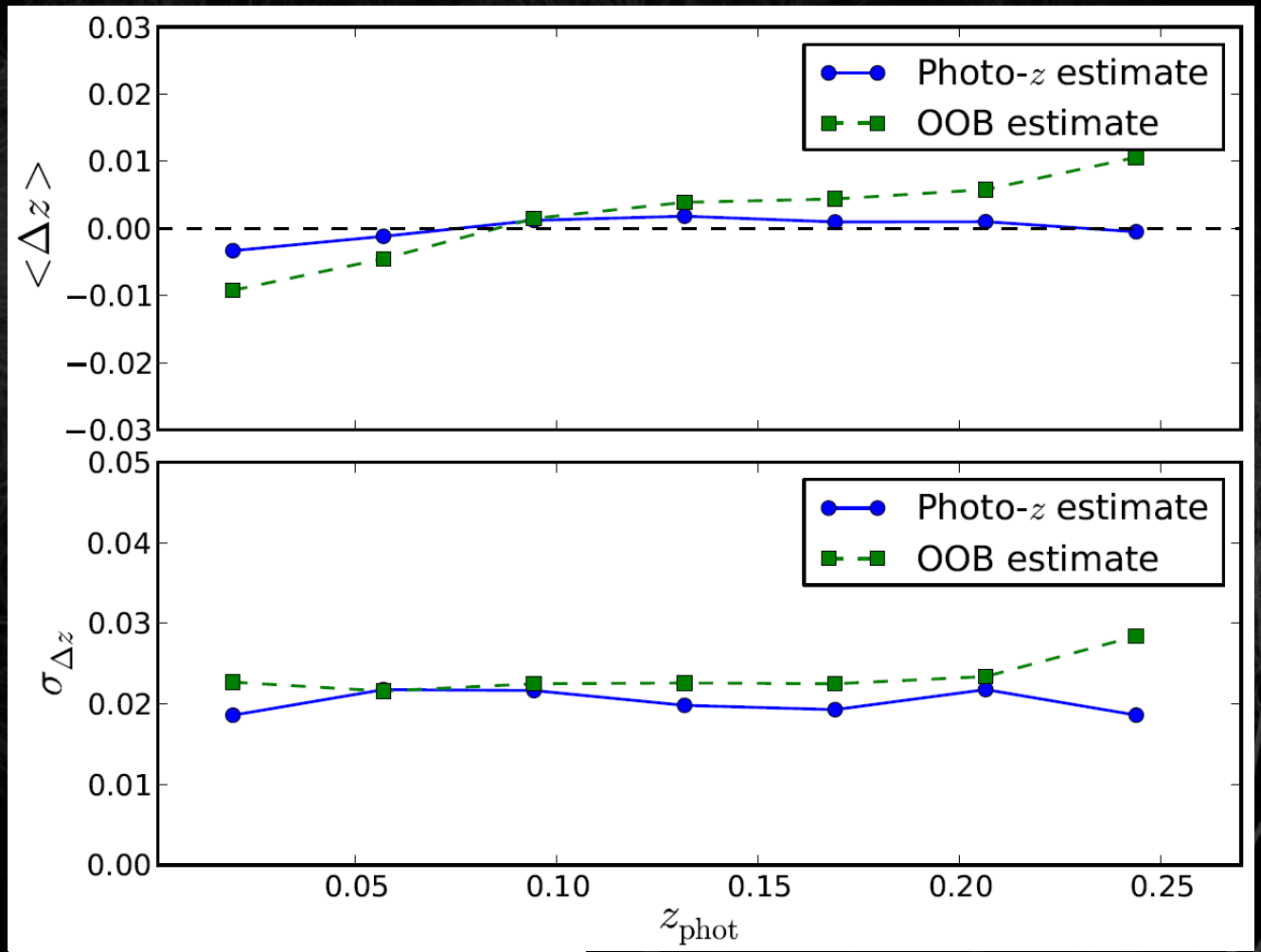
TPZ: Ancillary information - *prior error* -

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!



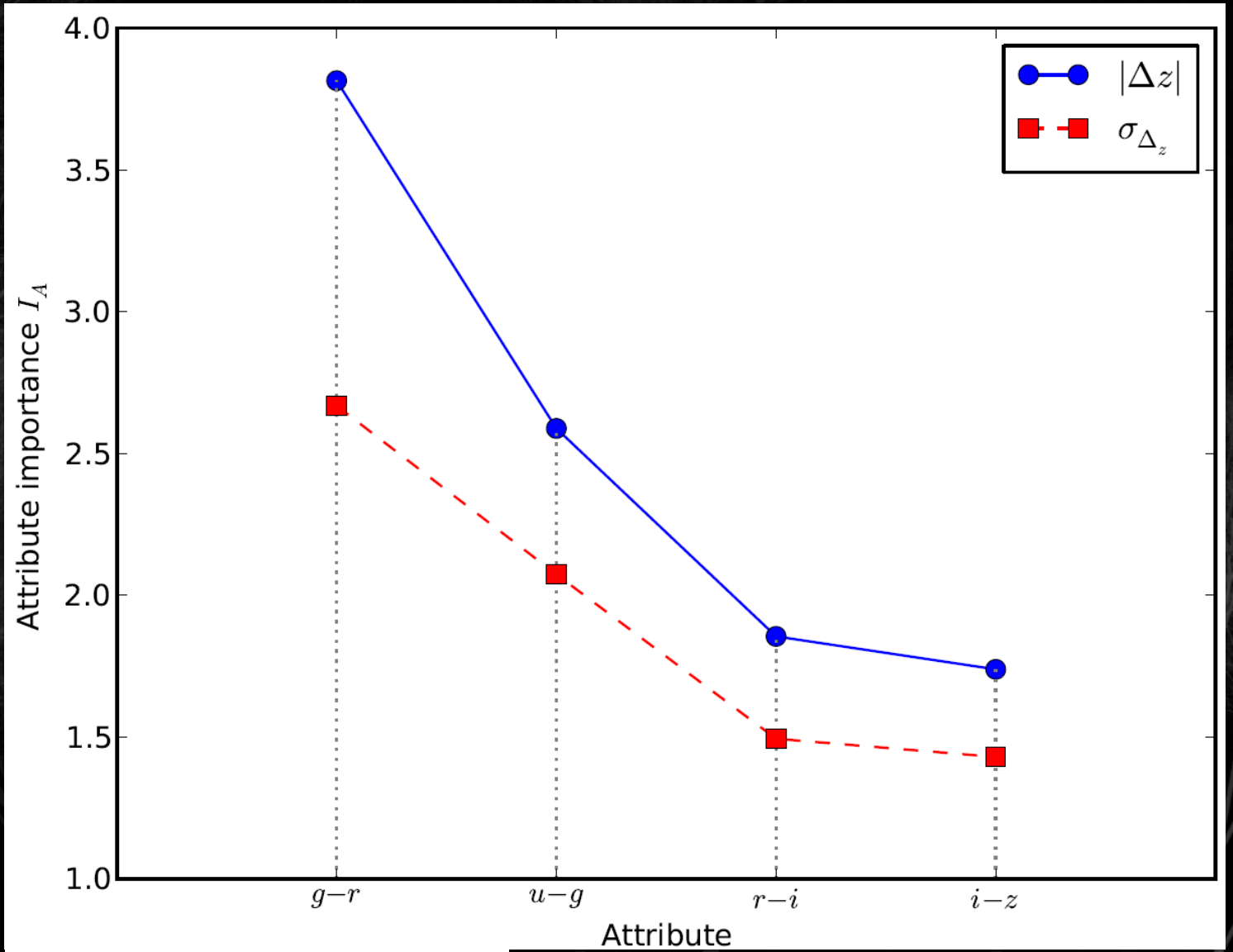
Carrasco Kind & Brunner 2013a

TPZ: Ancillary information - *Attribute importance* -

Ranking statistical only

Useful for removing unimportant variables reducing the noise

Most important attributes to construct impotance map

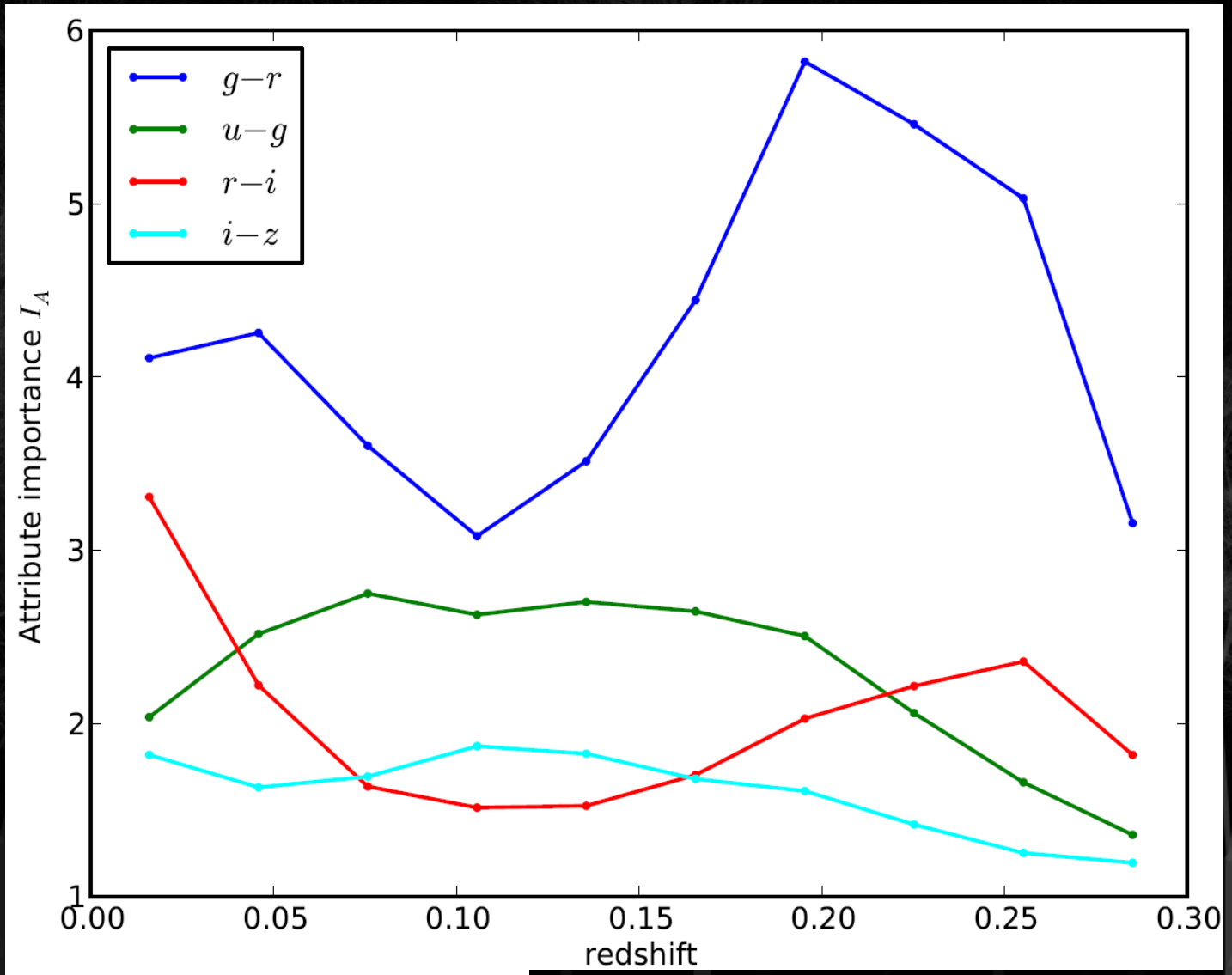


Carrasco Kind & Brunner 2013a

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

Dark energy model

From Einstein's field equations we can derive Friedmann's-Lemaître's equations:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}$$

and from the conservation of energy to:

$$\dot{\rho} = -3\frac{\dot{a}}{a}(p + \rho)$$

Dark energy model

From Einstein's field equations we can derive Friedmann's-Lemaître's equations:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}$$

and from the conservation of energy to:

$$\dot{\rho} = -3\frac{\dot{a}}{a}(p + \rho)$$

Equation of state : $\omega = \frac{p}{\rho}$

Dark energy model

From Einstein's field equations we can derive Friedmann's-Lemaître's equations:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}$$

geometry

and from the conservation of energy to:

$$\dot{\rho} = -3\frac{\dot{a}}{a}(p + \rho)$$

growth of structure

Equation of state : $\omega = \frac{p}{\rho}$

Dark energy model

From Einstein's field equations we can derive Friedmann's-Lemaître's equations:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}$$ geometry

and from the conservation of energy to:

$$\dot{\rho} = -3\frac{\dot{a}}{a}(p + \rho)$$ growth of structure

Equation of state : $\omega = \frac{p}{\rho}$

- $\omega = \frac{1}{3}$: radiation dominated
- $\omega = 0$: matter dominated
- $\omega < -\frac{1}{3}$: dark energy
- $\omega <= -1$: vacuum energy

Dark energy model

From Einstein's field equations we can derive Friedmann's-Lemaître's equations:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}$$ geometry

and from the conservation of energy to:

$$\dot{\rho} = -3\frac{\dot{a}}{a}(p + \rho)$$ growth of structure

Equation of state : $\omega = \frac{p}{\rho}$

DE models
$\omega(a) = \omega_0 + \omega_a(1 - a)$

- $\omega = \frac{1}{3}$: radiation dominated
- $\omega = 0$: matter dominated
- $\omega < -\frac{1}{3}$: dark energy
- $\omega <= -1$: vacuum energy

Dark energy model

From Einstein's field equations we can derive Friedmann's-Lemaître's equations:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}$$

geometry

and from the conservation of energy to:

$$\dot{\rho} = -3\frac{\dot{a}}{a}(p + \rho)$$

growth of structure

Equation of state : $\omega = \frac{p}{\rho}$

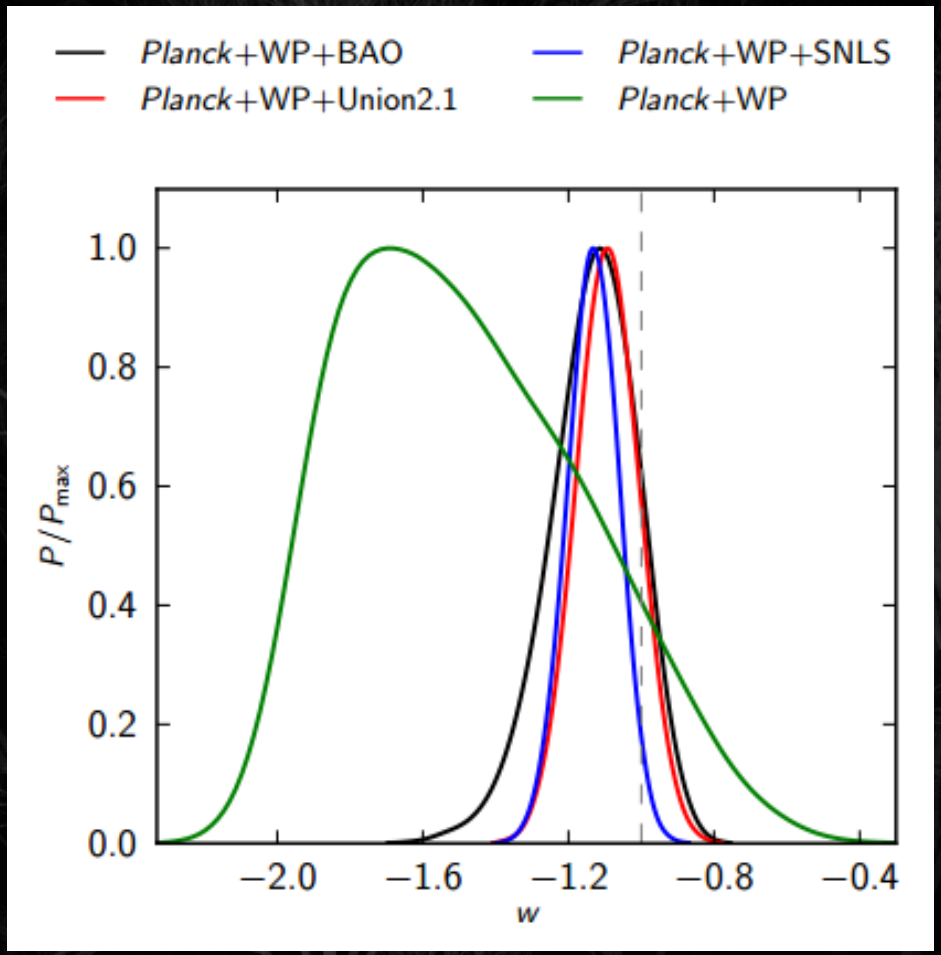
DE models

$$\omega(a) = \omega_0 + \omega_a(1 - a)$$

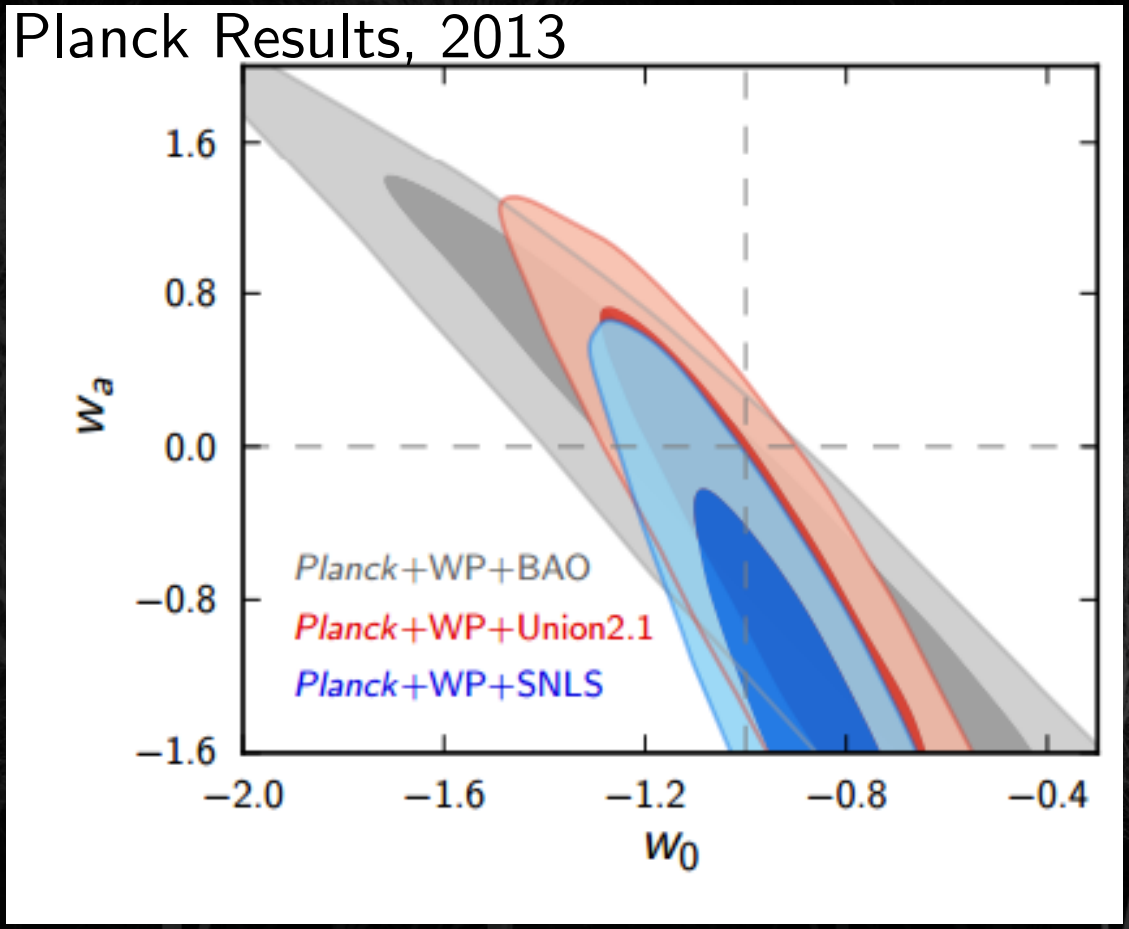
There are alternative models than can be tested/ruled out

Equation of state

$$\omega(a) = \omega_0 + \omega_a(1-a)$$

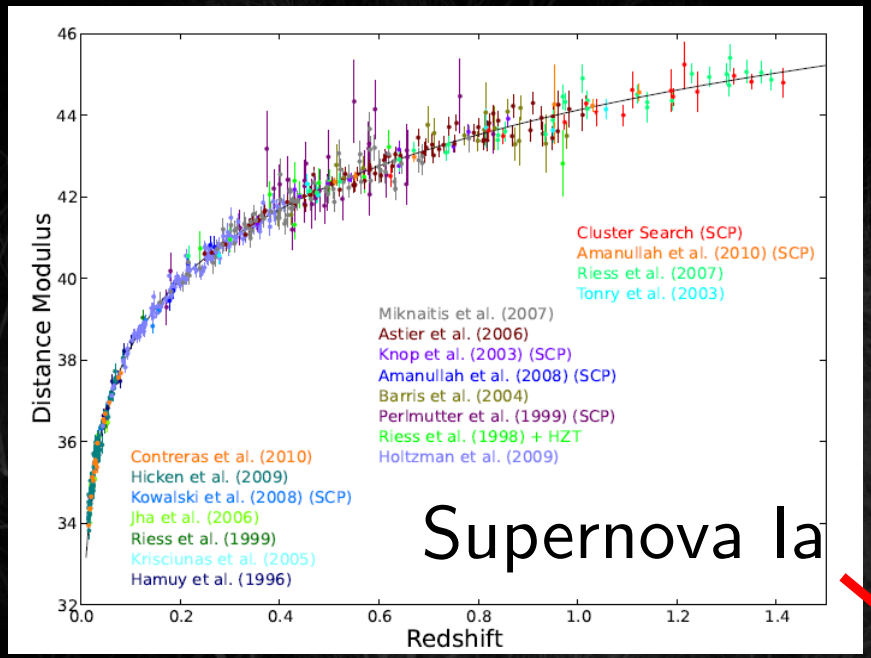


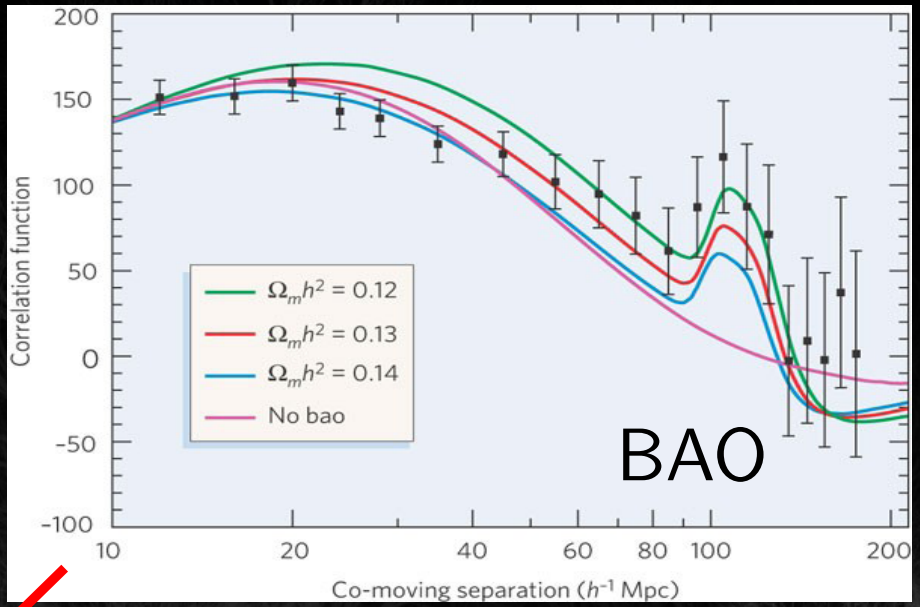
Planck Results, 2013

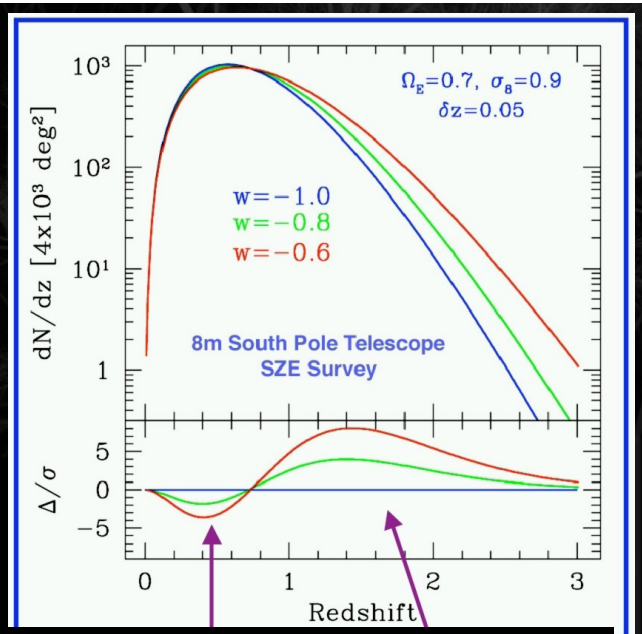


Planck does not give very tight contrains on ω and ω_a, need complementary approach. A big survey!

Observational Probes of Dark Energy

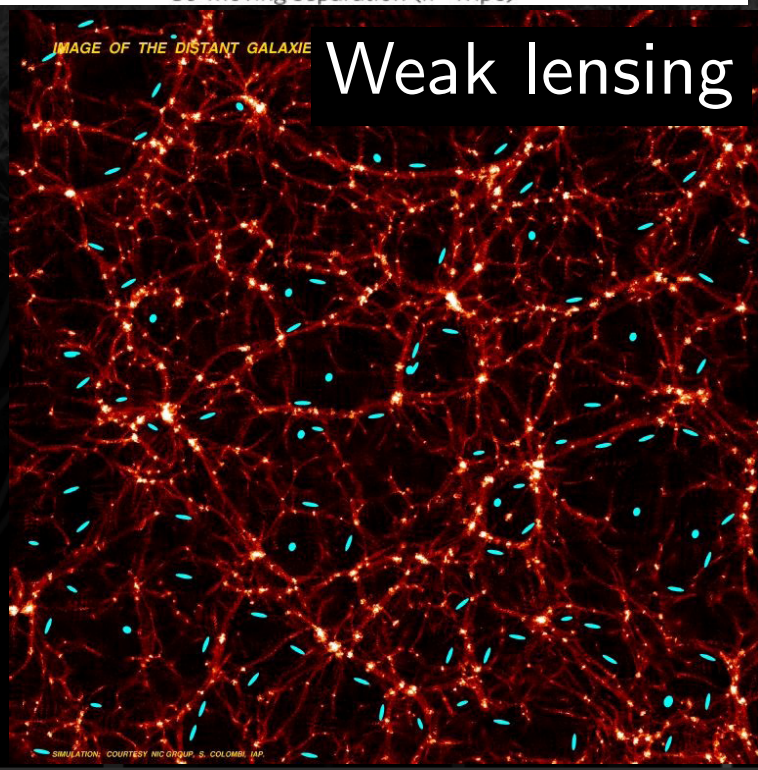






Geometry

Growth of structure



Probes of Dark Energy : Combined

$$\omega(a) = \omega_0 + \omega_a(1-a)$$

- DES first to combine 4 probes with one dataset
- Combined probes provide tighter constrains
- Cross-check for systematics

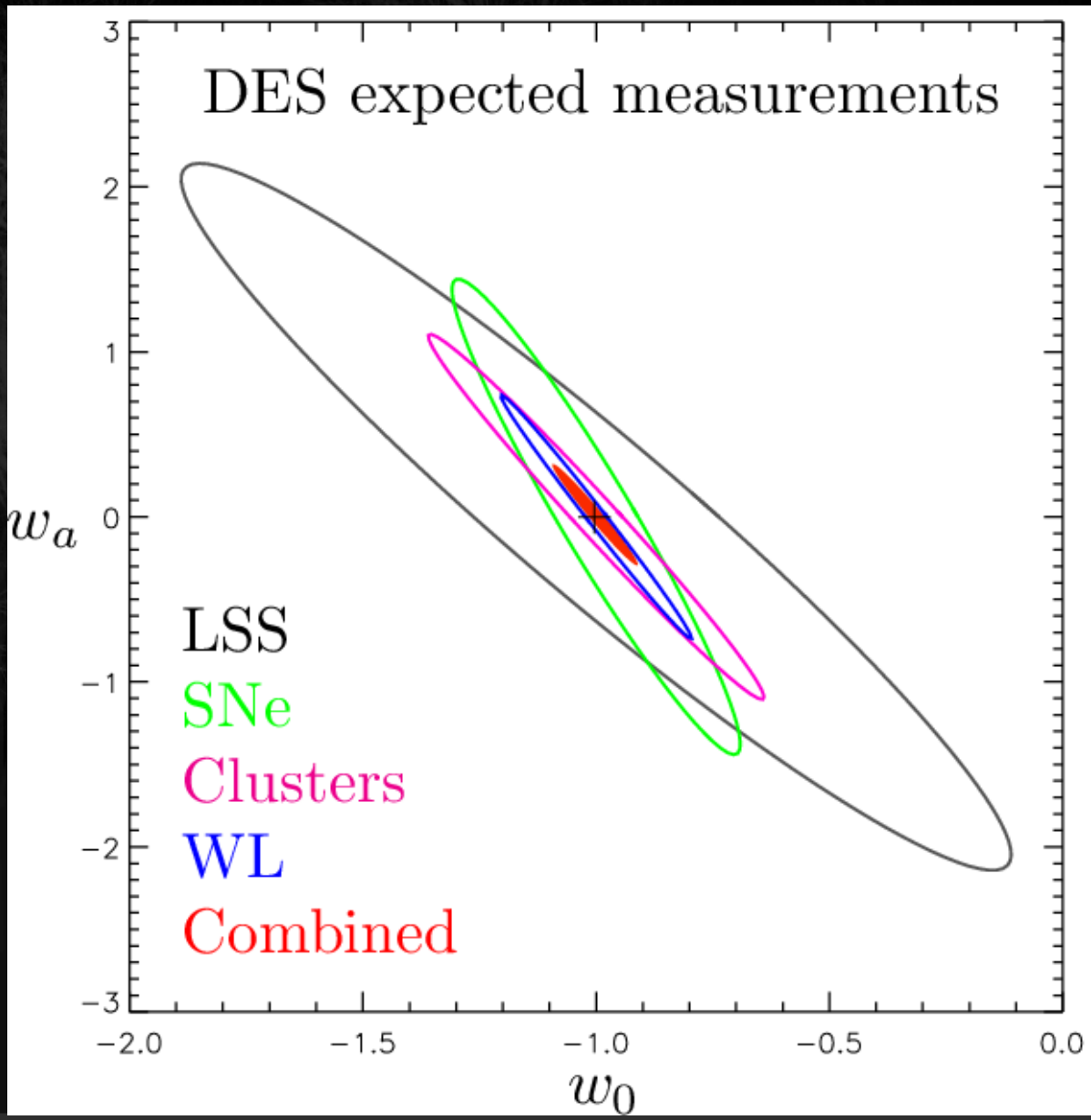


Photo-z PDF combination: BMC

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: HB

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

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

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

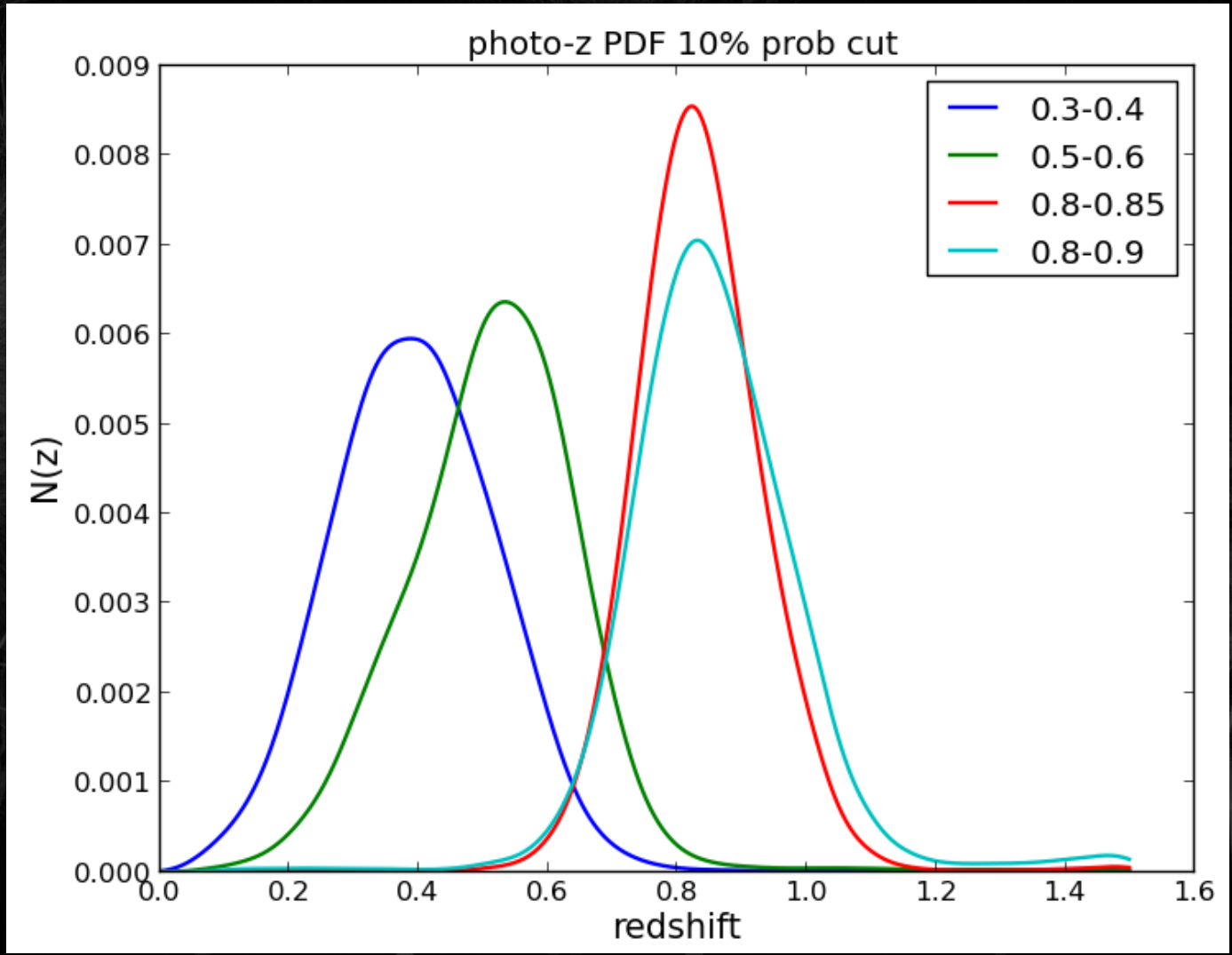
$$P(z \mid \mathbf{x}, \mathbf{D}, \theta) = \prod_k P(z \mid \mathbf{x}, \mathbf{D}, M_k, \theta_k)^{1/\beta}$$

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

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

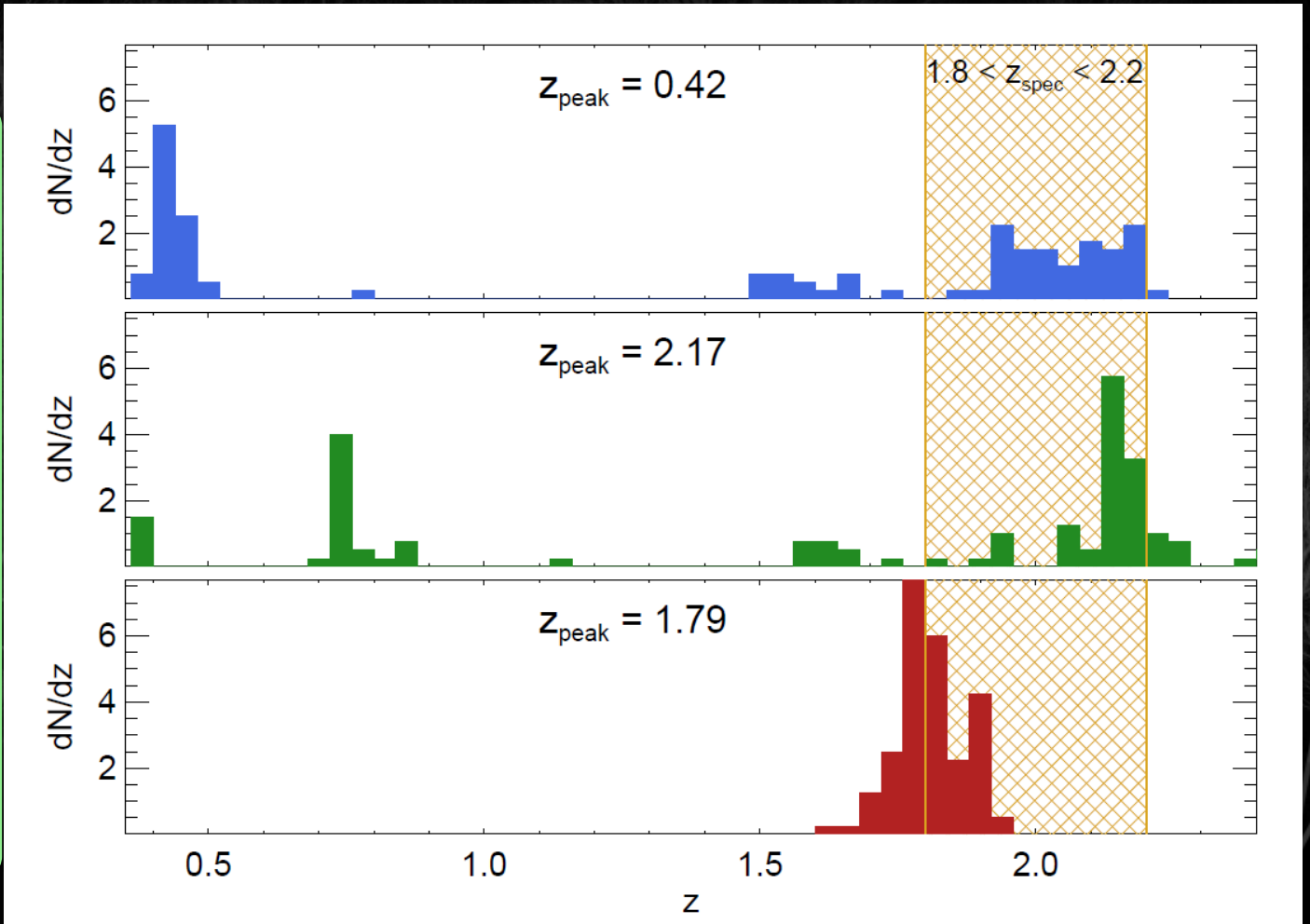


Example application of photo-z PDF

Incorporating PDF on clustering measurements

Problems of using mode of photo-z PDF

Extend to other measurements



Myers, White & Ball 2009

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

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

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

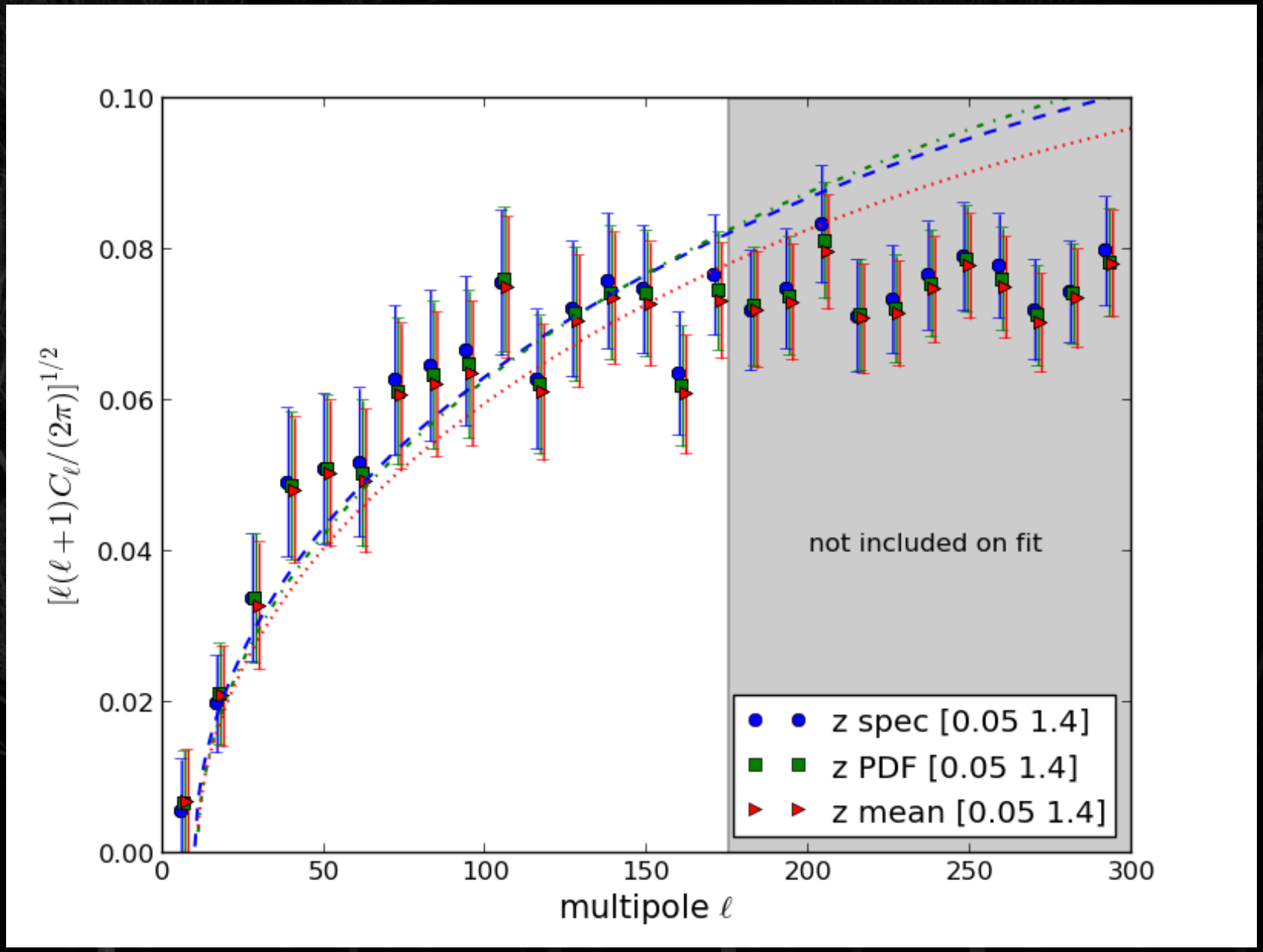
CAMB and HALOFIT for non linear $P(k,z)$

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

Fitting using Monte Carlo Markov Chain methods

$$\chi^2(a_p) = \sum_{bb'} (\ln C_b - \ln C_b^T) C_b F_{bb'} C_{b'} (\ln C_{b'} - \ln C_{b'}^T)$$

Preliminary results on APS



Photometric redshift PDFs using TPZ

Metrics

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

$< \Delta z >= 0.0088$

$< |\Delta z| >= 0.089$

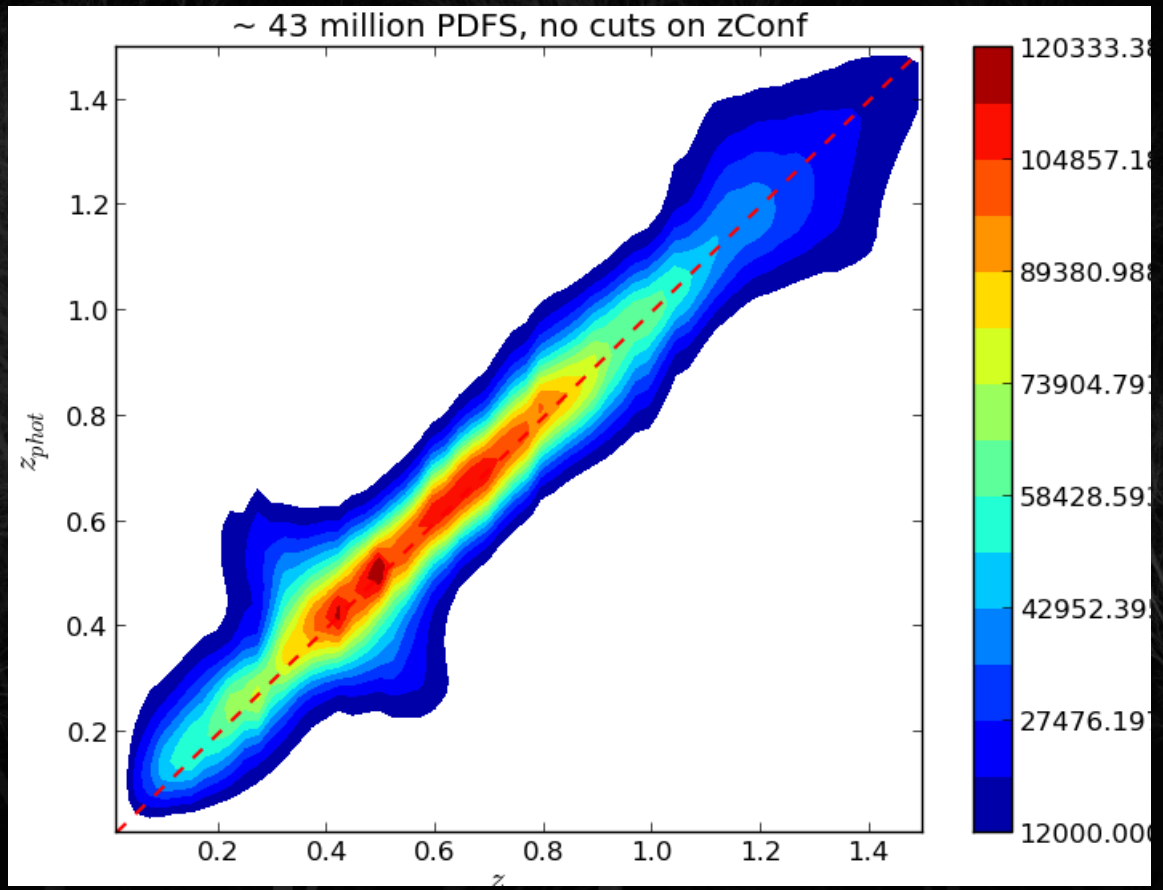
$\sigma_{\Delta z} = 0.1421$

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

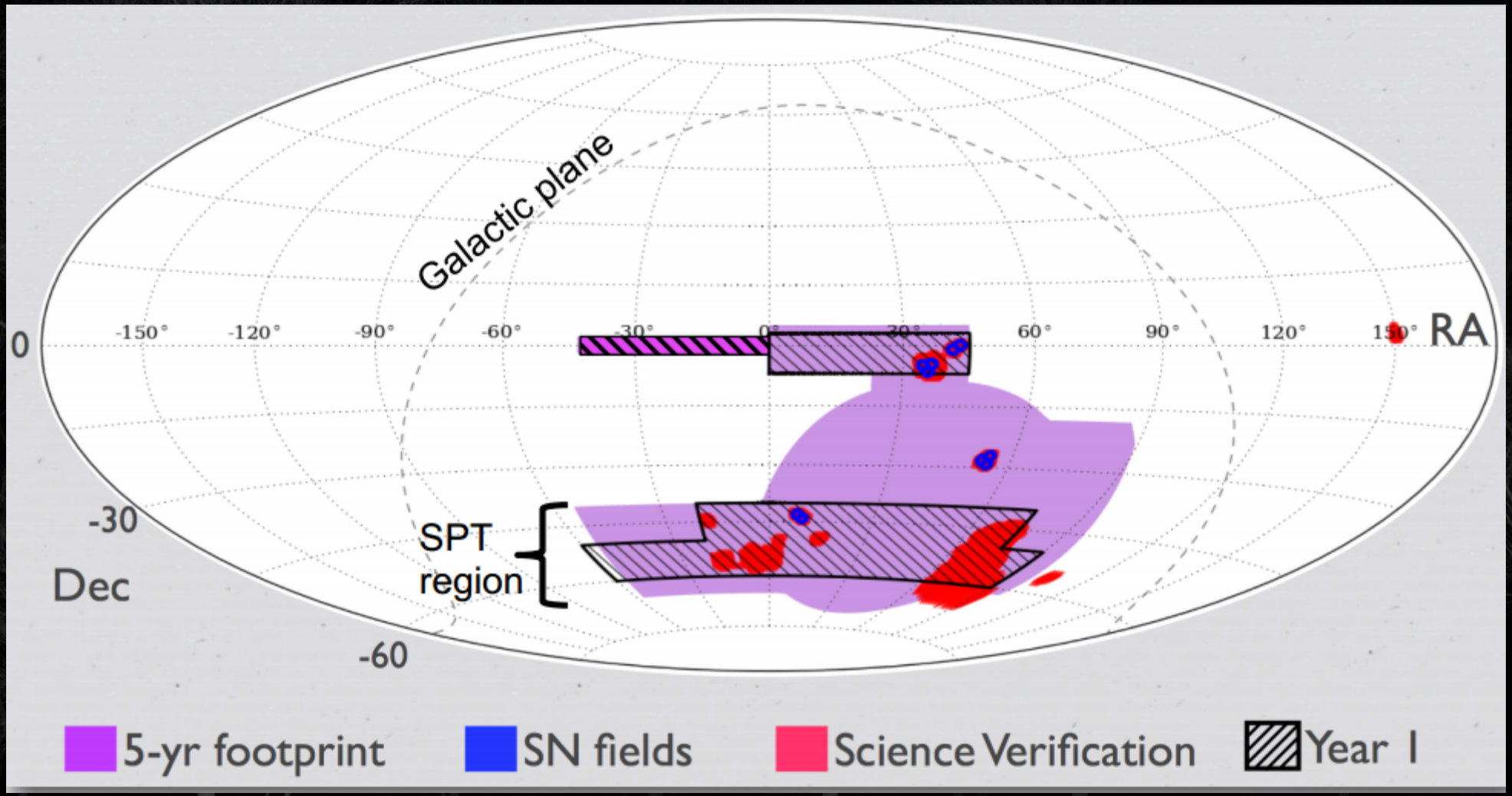
$\sigma_{68} = 0.0885$

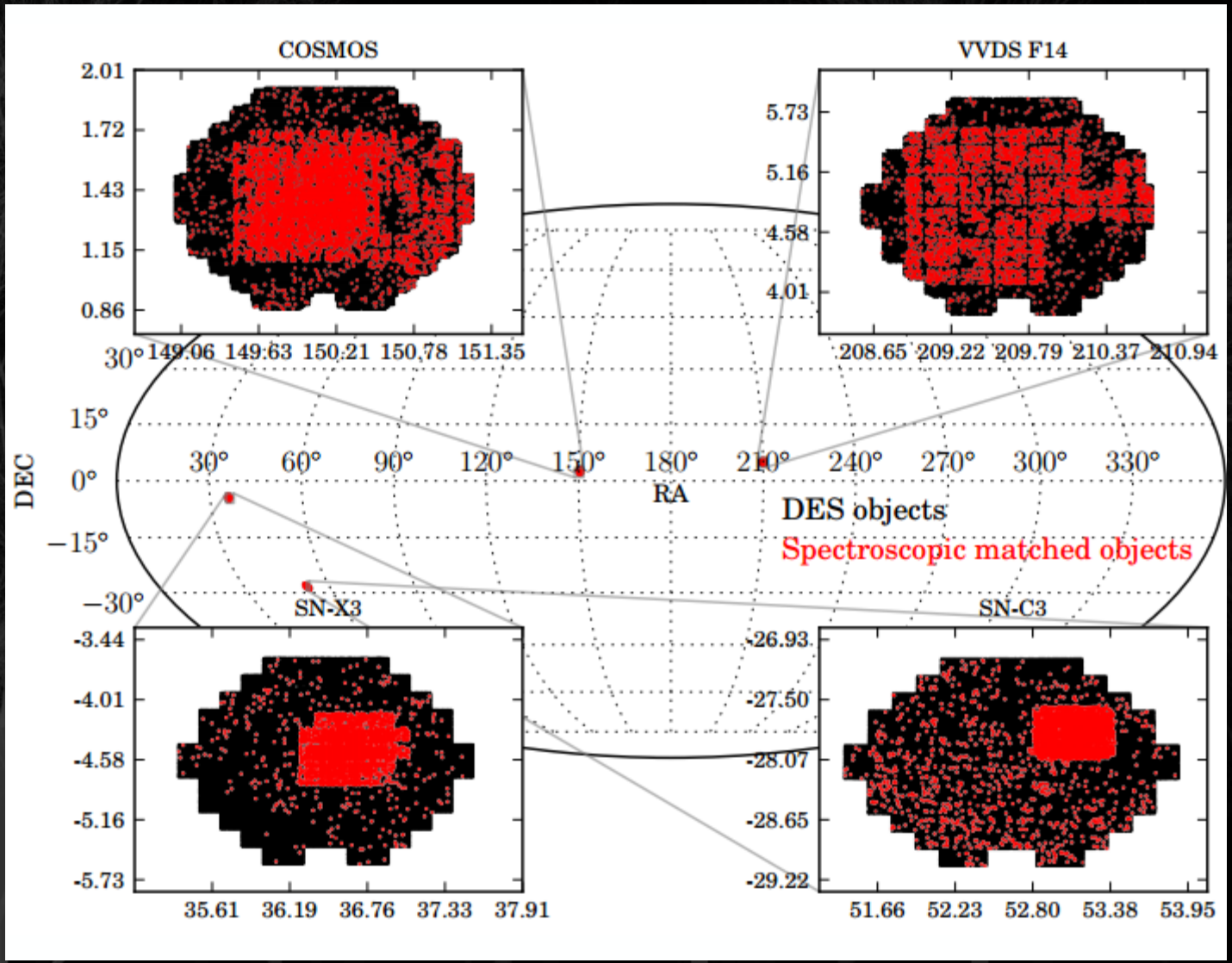
$frac > 2\sigma = 0.0531$

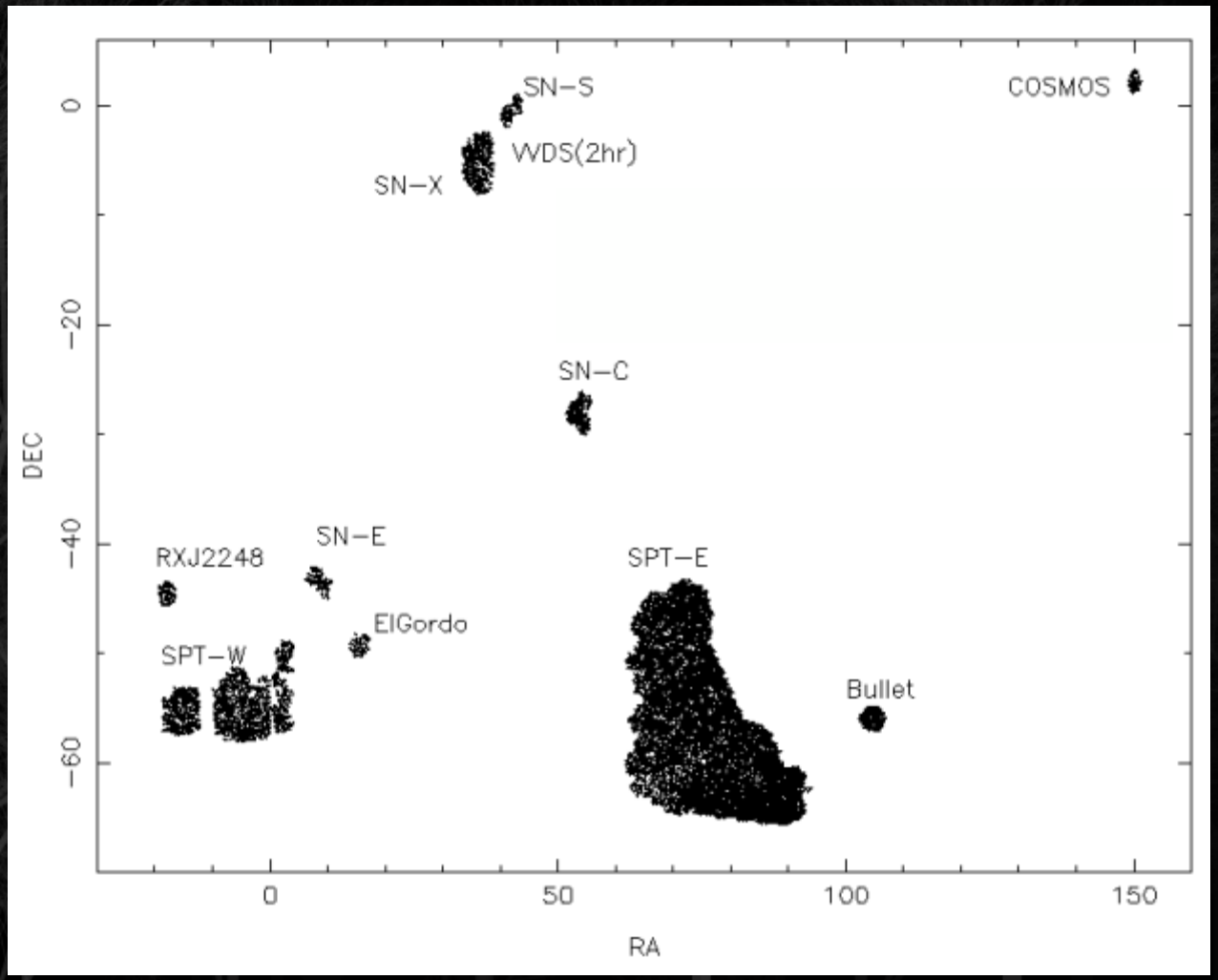
$frac > 3\sigma = 0.0207$



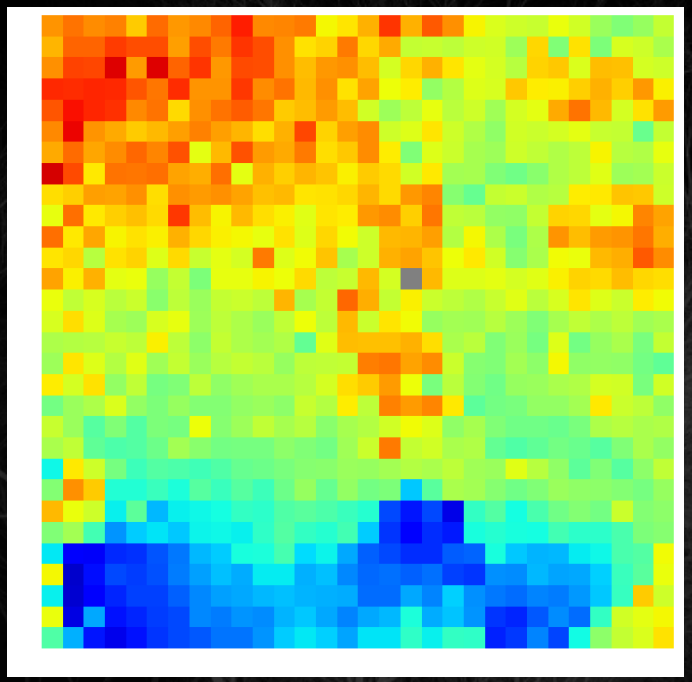
DES footprint

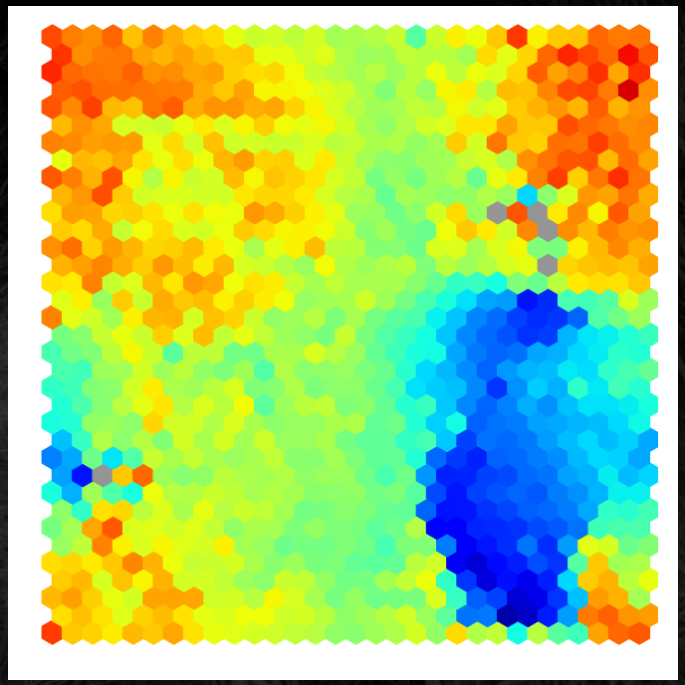


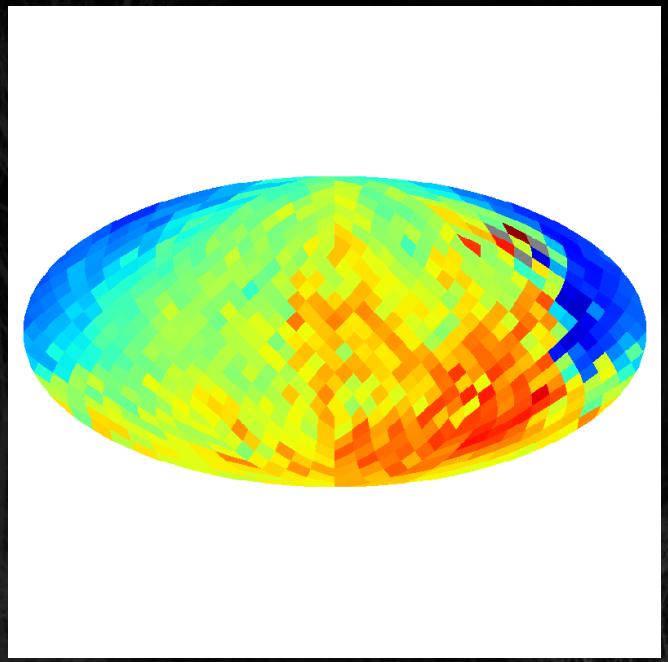

COSMOS
VVDS F14
SN-X3
SN-C3
DES objects
Spectroscopic matched objects
RA
DEC
30°
60°
90°
120°
150°
180°
210°
240°
270°
300°
330°
15°
0°
−15°
−30°
2.01
1.72
1.43
1.15
0.86
149.06
149.63
150.21
150.78
151.35
5.73
5.16
4.58
4.01
208.65
209.22
209.79
210.37
210.94
-3.44
-4.01
-4.58
-5.16
-5.73
35.61
36.19
36.76
37.33
37.91
-26.93
-27.50
-28.07
-28.65
-29.22
51.66
52.23
52.80
53.38
53.95

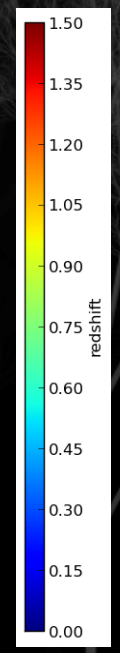

SN−S
COSMOS
WDS(2hr)
SN−X
SN−C
SN−E
RXJ2248
SPT−E
ElGordo
SPT−W
Bullet
DEC
RA
0
−20
−40
−60
0
50
100
150

SOM topologies



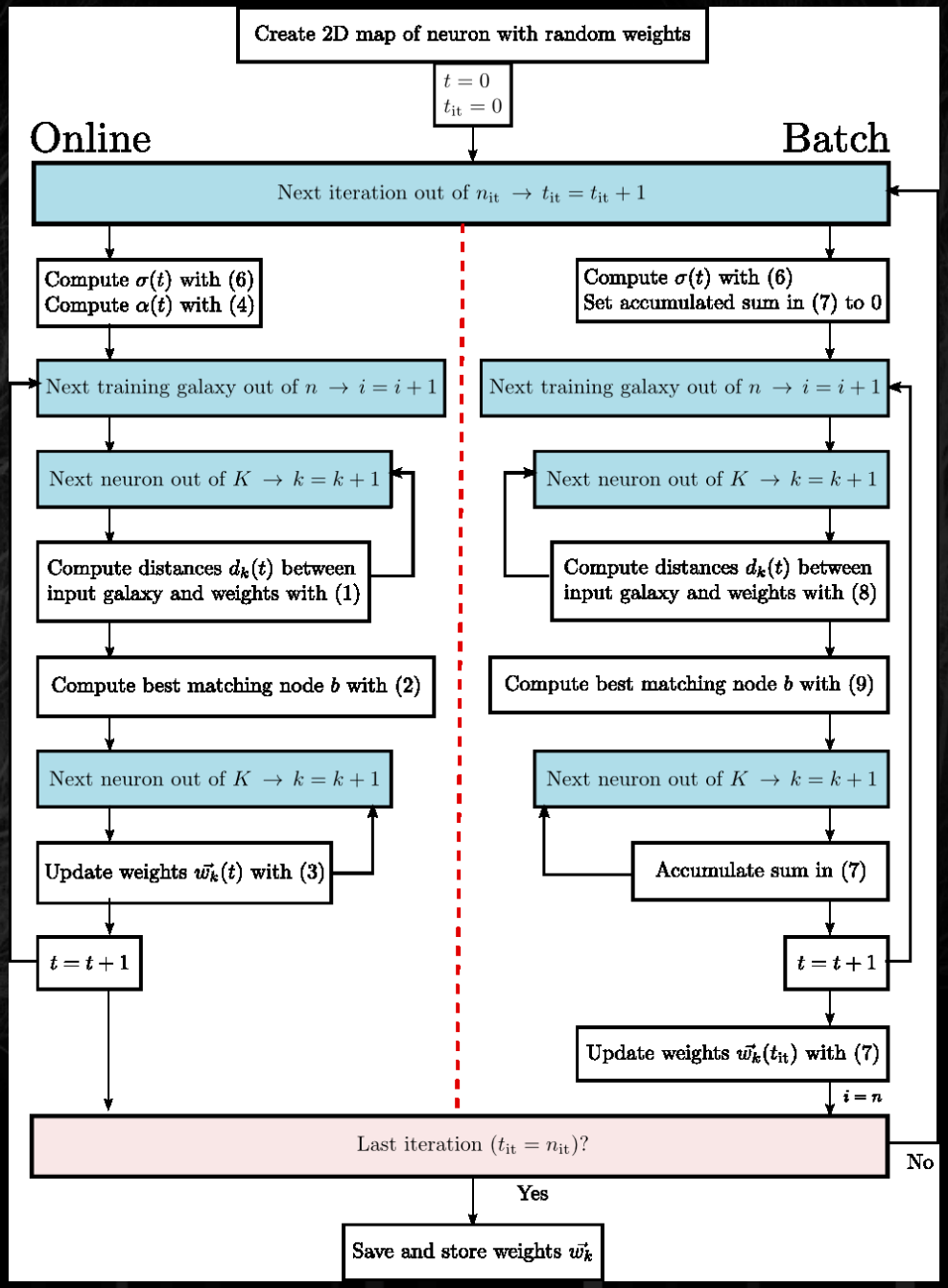




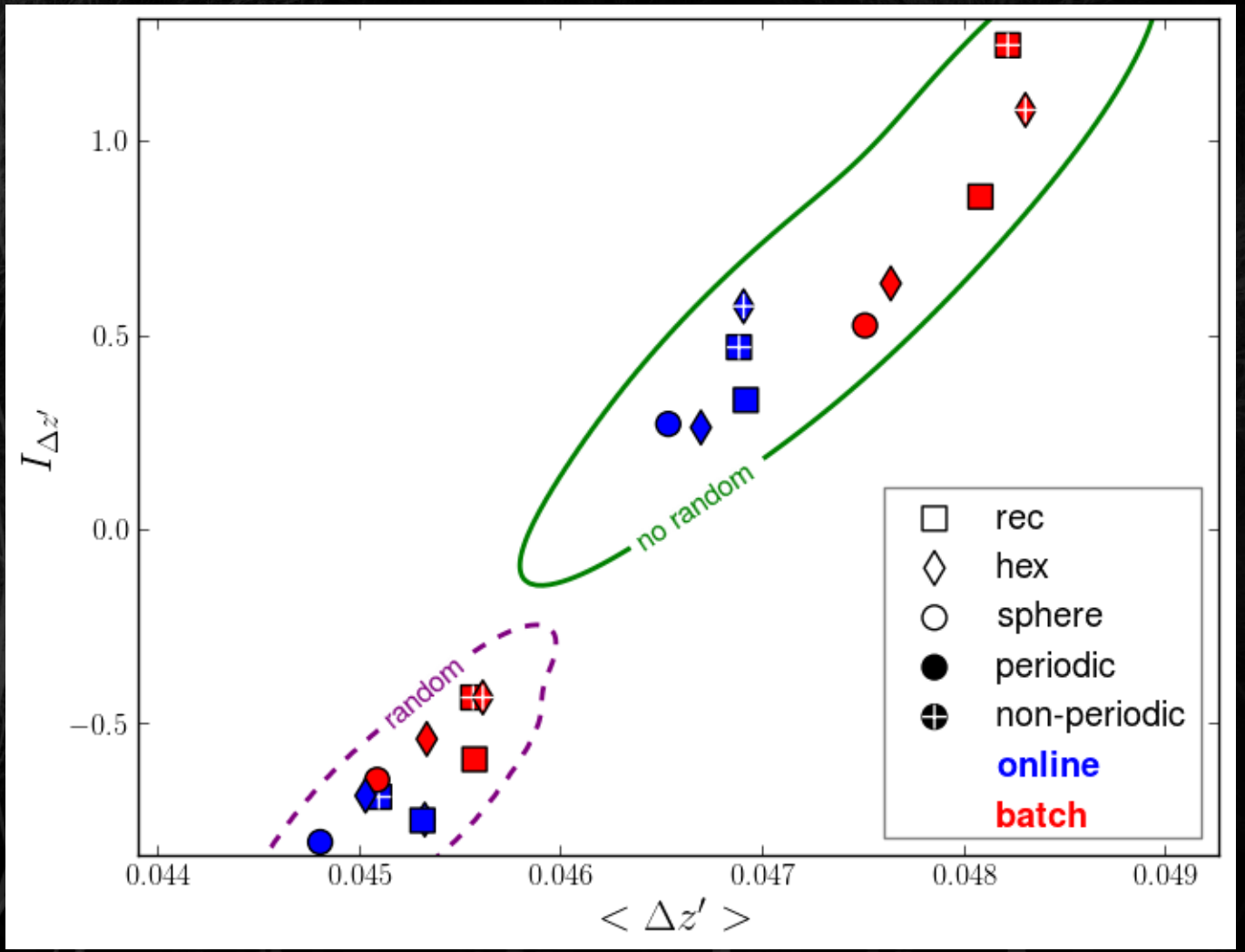


Different topologies can be used with or without periodic boundary conditions

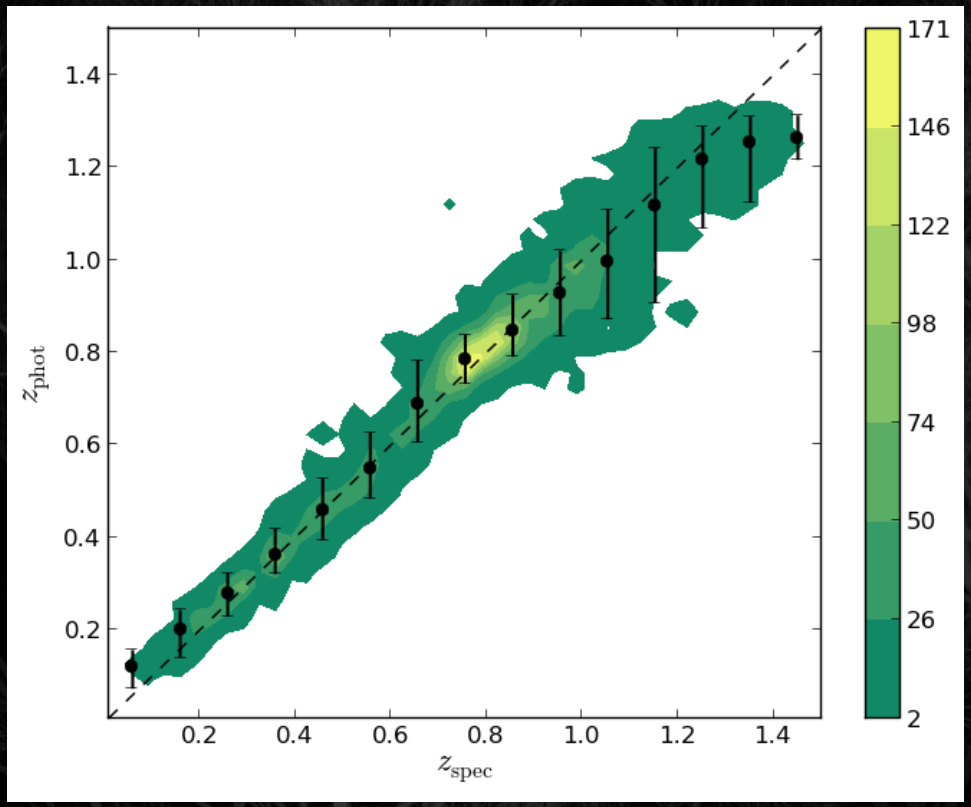
SOM scheme

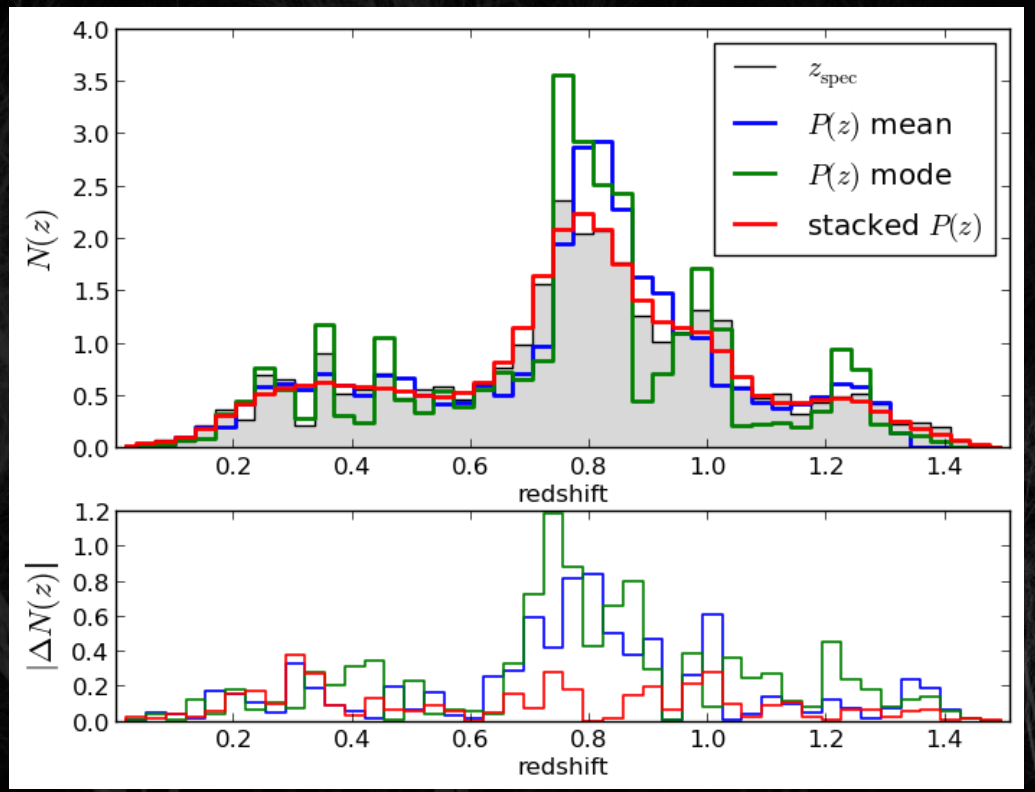

Create 2D map of neuron with random weights
$t = 0$
$t_{\rm it} = 0$
Online
Batch
Next iteration out of $n_{\rm it}$ $\rightarrow$ $t_{\rm it} = t_{\rm it} + 1$
Compute $\sigma(t)$ with (6)
Compute $\alpha(t)$ with (4)
Compute $\sigma(t)$ with (6)
Set accumulated sum in (7) to 0
Next training galaxy out of n $\rightarrow$ $i = i + 1$
Next training galaxy out of n $\rightarrow$ $i = i + 1$
Next neuron out of K $\rightarrow$ $k = k + 1$
Next neuron out of K $\rightarrow$ $k = k + 1$
Compute distances $d_k(t)$ between input galaxy and weights with (1)
Compute distances $d_k(t)$ between input galaxy and weights with (8)
Compute best matching node b with (2)
Compute best matching node b with (9)
Next neuron out of K $\rightarrow$ $k = k + 1$
Next neuron out of K $\rightarrow$ $k = k + 1$
Update weights $\vec{w}_k(t)$ with (3)
Accumulate sum in (7)
$t = t + 1$
$t = t + 1$
Update weights $\vec{w}_k(t_{\rm it})$ with (7)
$i = n$
Last iteration ($t_{\rm it} = n_{\rm it}$)?
No
Yes
Save and store weights $\vec{w}_k$

SOMz results



SOMz results





Big data problem

It's happening! ☺

$\sim$ 300 millions galaxies up to z =1.5

5,000 squares degrees (1/8 sky)

Data managment at NCSA

DES specially designed to probe the origin of dark energy

S/G class and photo-z needed

1 TB of data per day

1 year done, 4 to go

Big data problem

It's happening! ☺

$\sim$ 300 millions galaxies up to z =1.5

5,000 squares degrees (1/8 sky)

Data managment at NCSA

DES specially designed to probe the origin of dark energy

S/G class and photo-z needed

1 TB of data per day

1 year done, 4 to go

Large Synoptic Survey Telescope

$\sim$ 2020 first light

Half of sky

30 TB of data nightly for 10 years!!

NCSA involved in data analysis

Big Data Challenge, $\sim$1B galaxies