# Accurate photometric redshift probability density estimates

Markus Michael Rau
SPV. Stella Seitz
Ben Hoyle, Kerstin Peach

# Accurate photometric redshift probability density estimation - method comparison and application

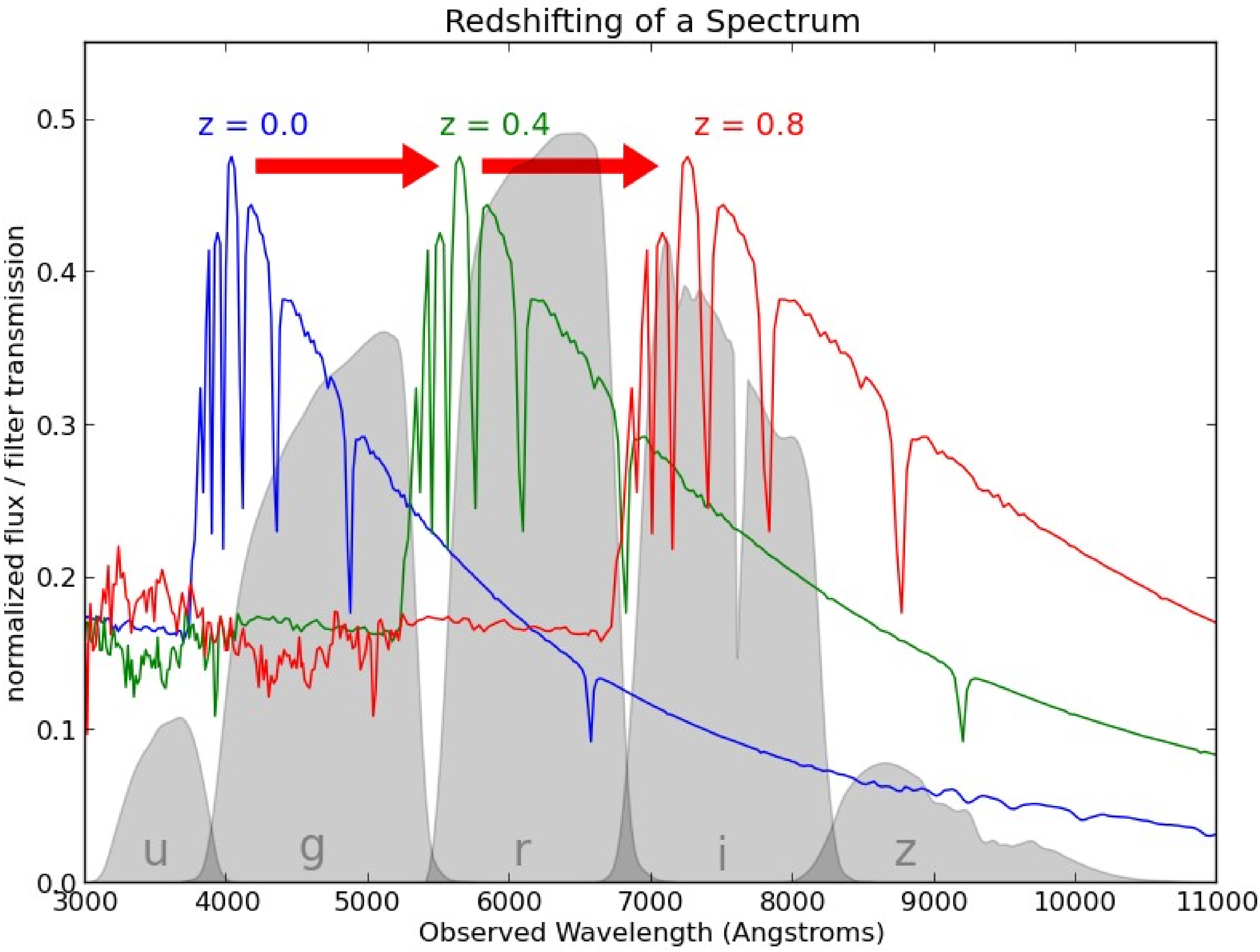
Markus Michael Rau[1,2], Stella Seitz[1,2], Fabrice Brimioulle[3], Eibe Frank[4], Oliver Friedrich[1,2], Daniel Gruen[1,2], Ben Hoyle[1,5]

MNRAS 2015 452 (4): 3710-3725

arXiv:1503.08215

- What are photometric redshifts (PhotoZs)?
- Why are PhotoZs important?
- Why should you care about modelling of PhotoZ uncertainty (PhotoZ probability density function)?
- How do we estimate PhotoZ probability density functions?
- What are the current problems?

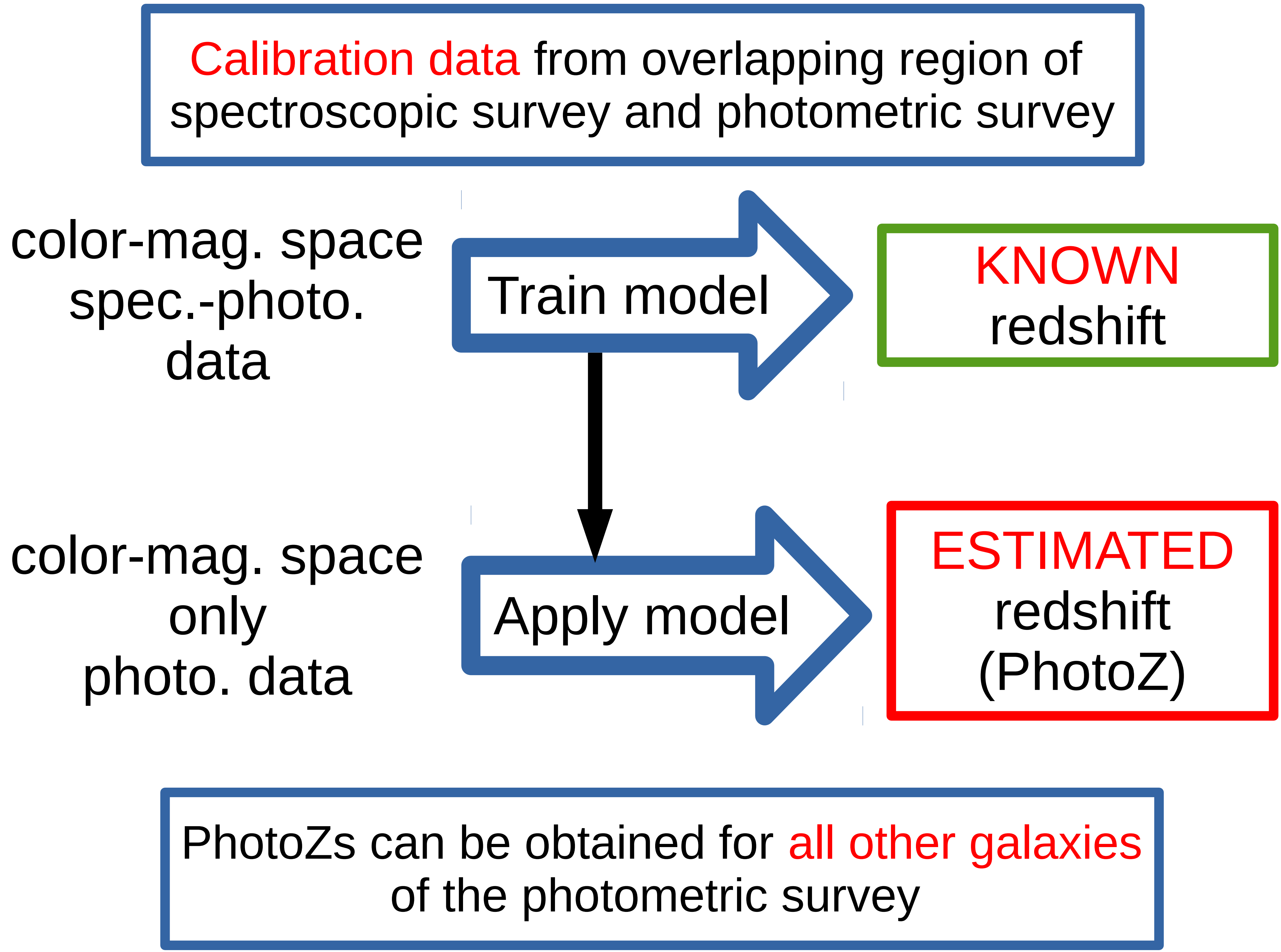
# What are photometric redshifts?



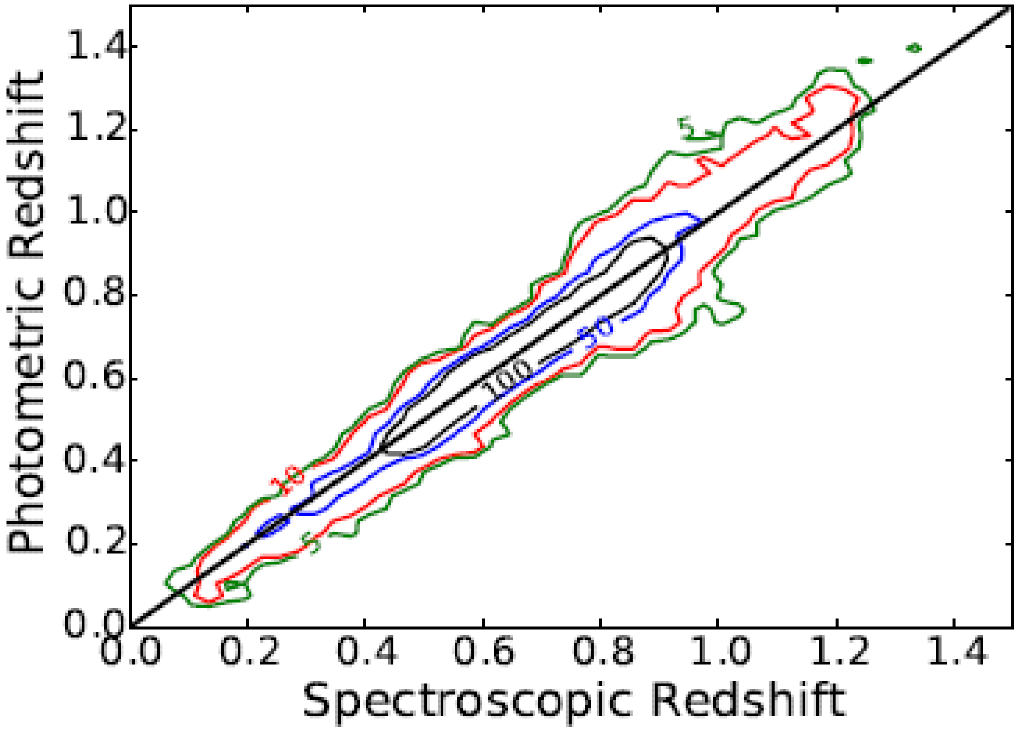
# Why are PhotoZs important?

- Spectroscopic redshifts expensive (long exposure times especially for faint objects)
- Small datasets
- Insufficient number of objects for many cosmological applications (galaxy clustering, cosmic shear, etc.)

→ Solution: Photometric surveys with spectroscopic overlap

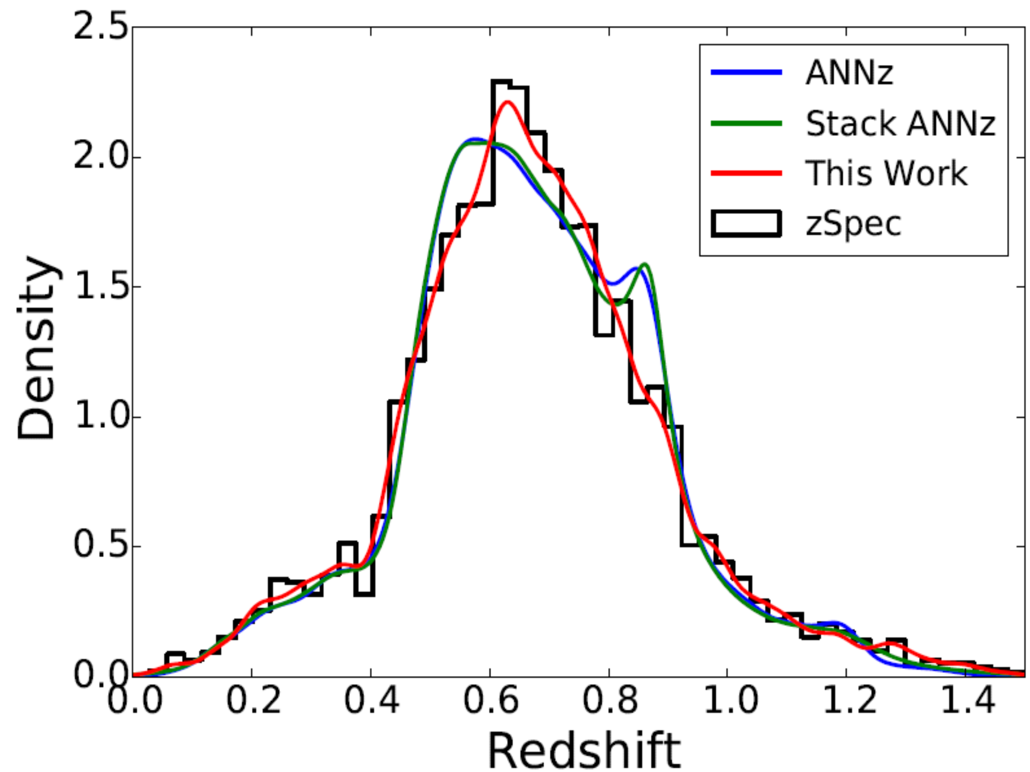

Calibration data from overlapping region of spectroscopic survey and photometric survey
color-mag. space spec.-photo. data
Train model
KNOWN redshift
color-mag. space only photo. data
Apply model
ESTIMATED redshift (PhotoZ)
PhotoZs can be obtained for all other galaxies of the photometric survey

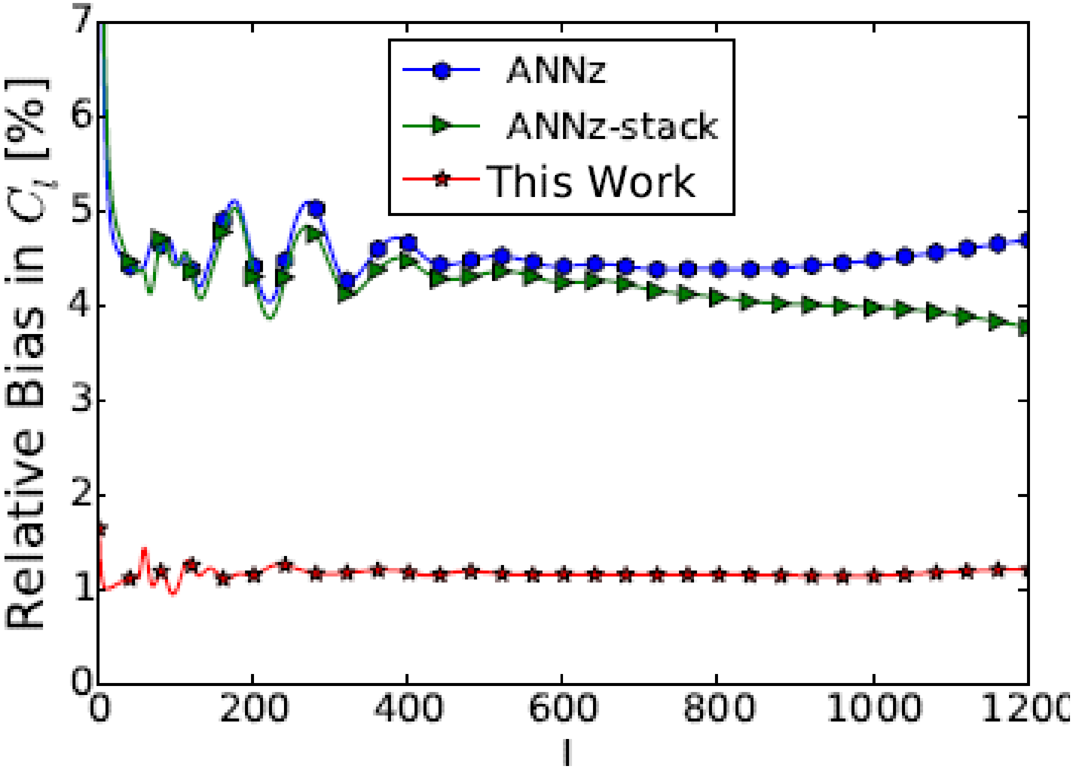
# Why should you care about the modelling of PhotoZ uncertainty?

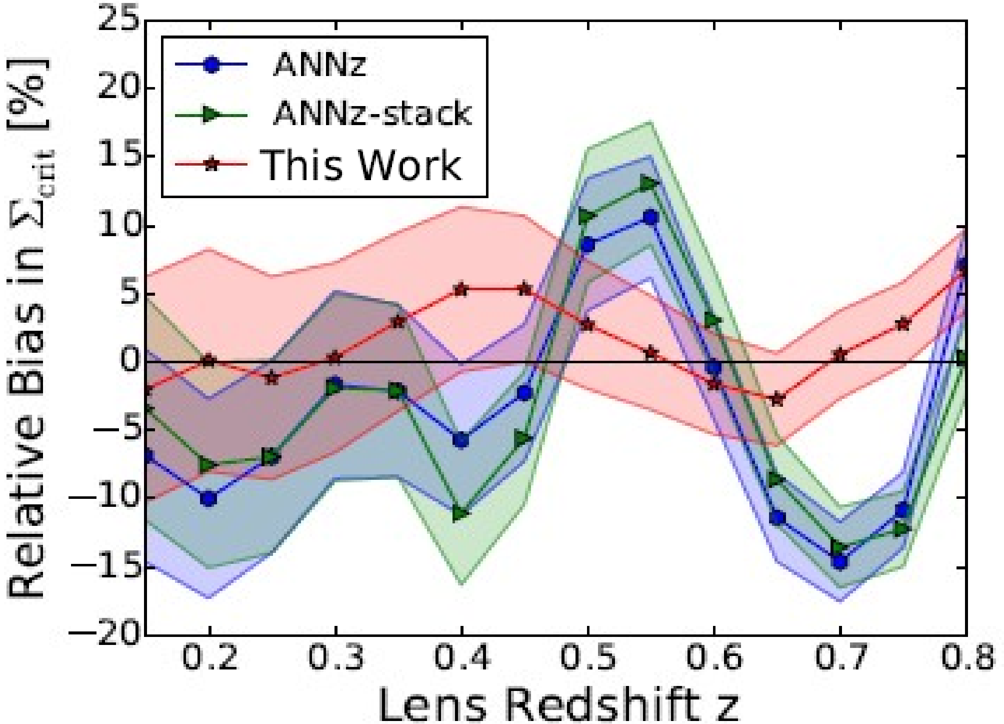


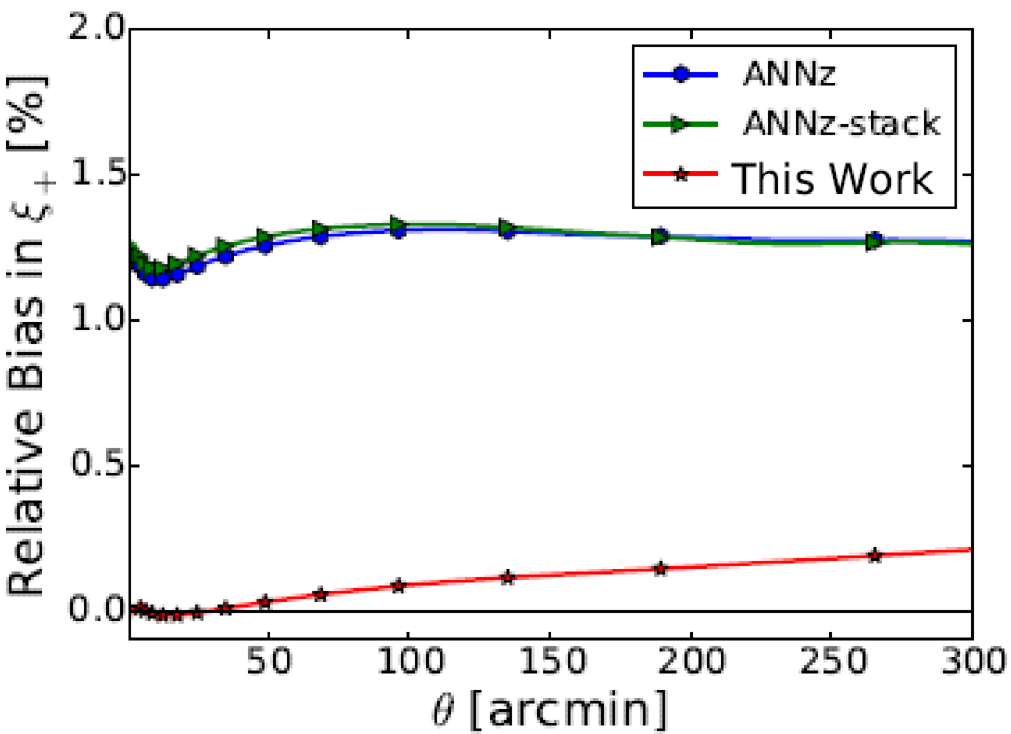

POINT PREDICTION
(ANNz code)
PhotoZ for
**CFHTLS Wide**

Inaccurate estimate of the sample redshift distribution





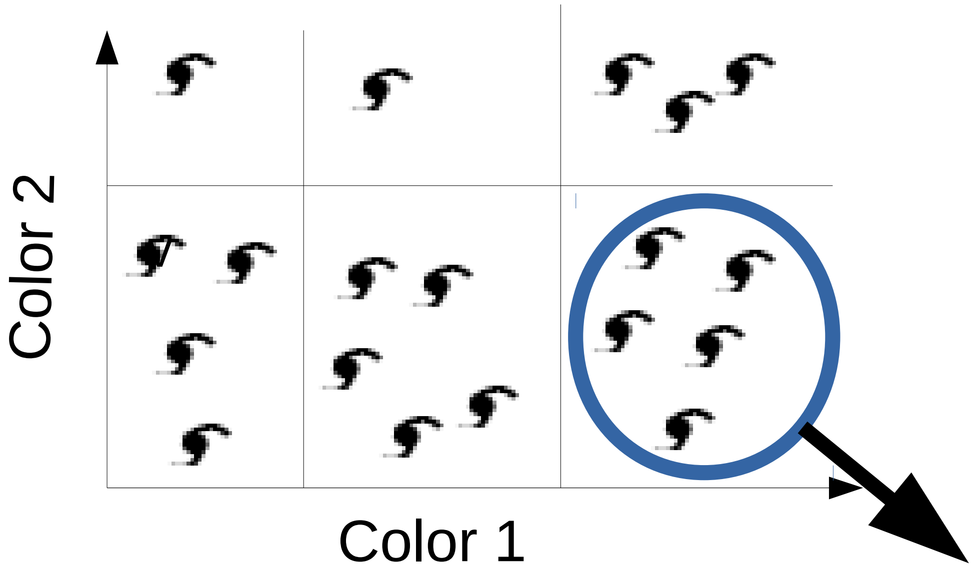




$$Relative\ Bias(X) = \frac{X_{phot} - X_{spec}}{X_{spec}}$$

Insufficient treatment of PhotoZ error distribution biases cosmological observables

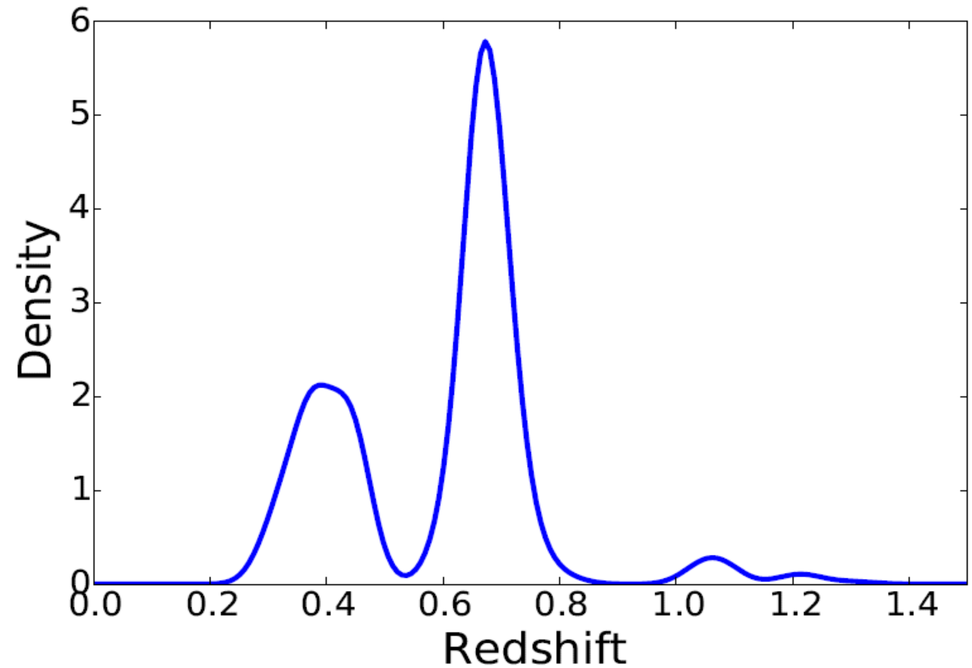
# How do we estimate PhotoZ probability density functions (PDFs)?

# How are they obtained?



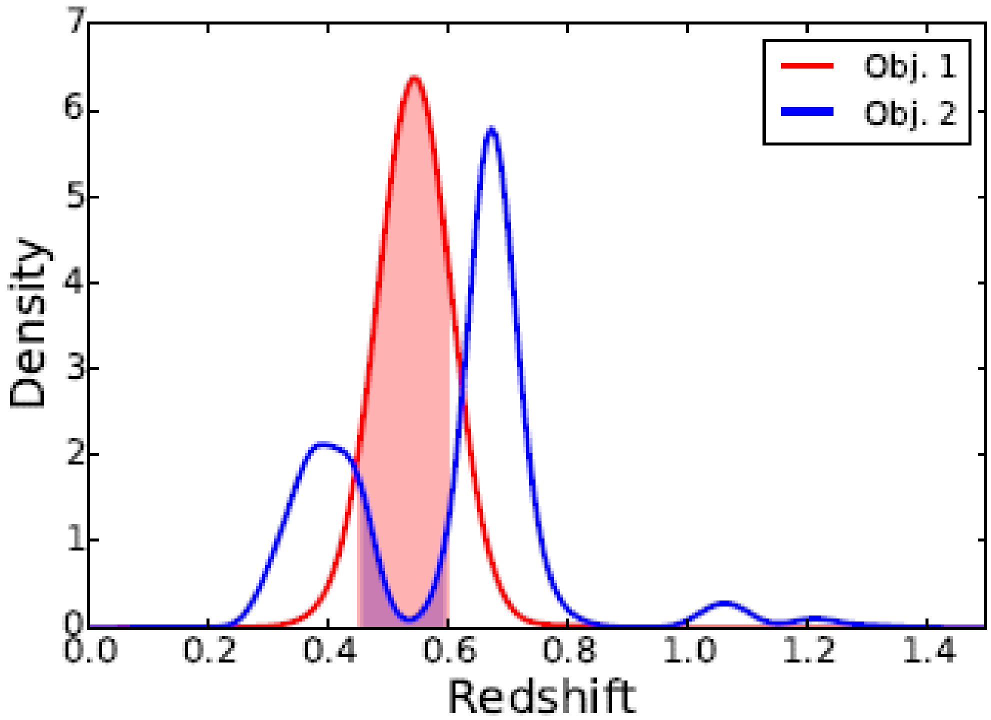

„Bin“ spectrophotometric calibration data

Put new objects in bins to get redshift distribution of calibration objects in the bin

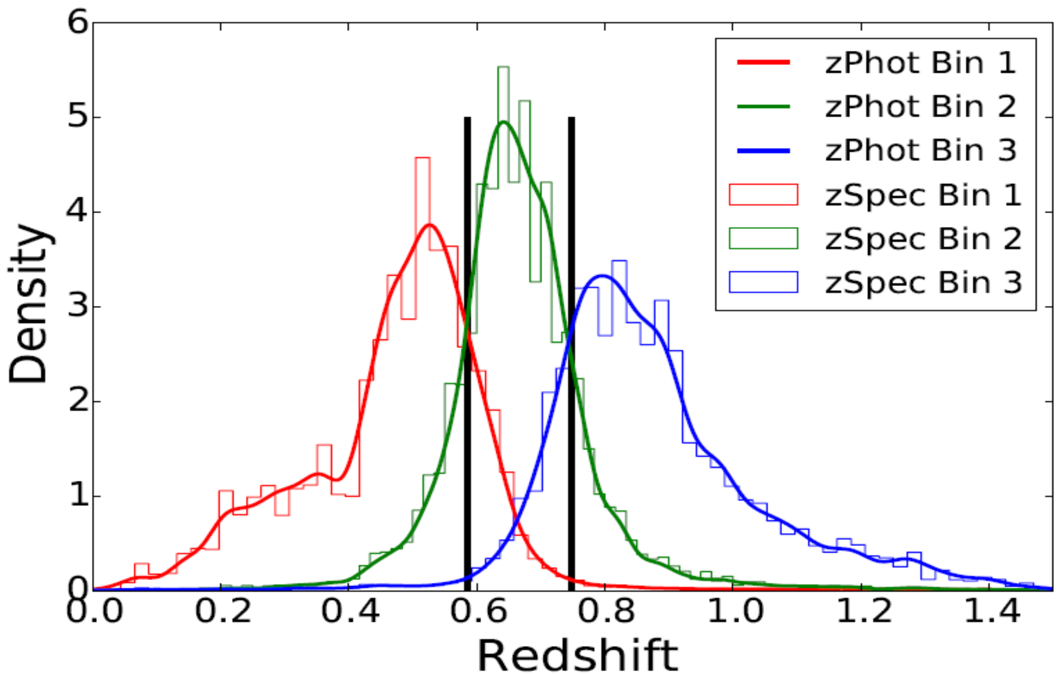


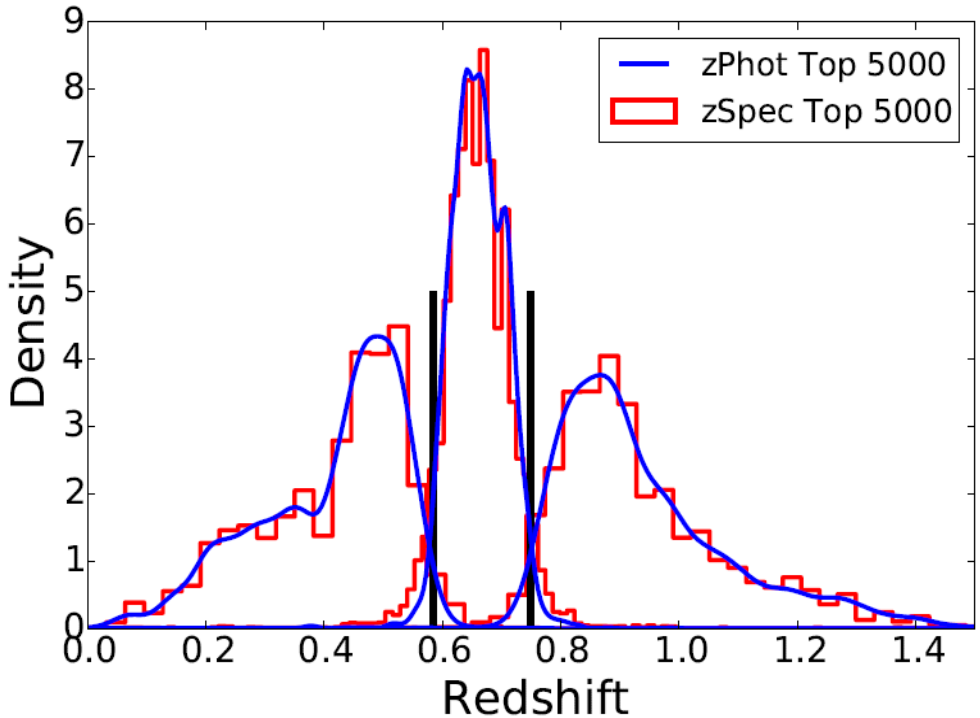
# We developed highly accurate and efficient methods

- Efficient representation of individual object redshift PDFs (5 numbers per object)
- Single number per object to estimate the sample redshift distribution
- Efficient sample selection by weighting individual object redshift PDFs by their overlap with redshift bins




- Stack individual object PDFs with weights
  → distribution centres in predefined z-range
- Applications like shear tomography



# Problem of spectroscopic incompleteness

- Problem of verification
  - Independent on the method no verification of PhotoZ quality!

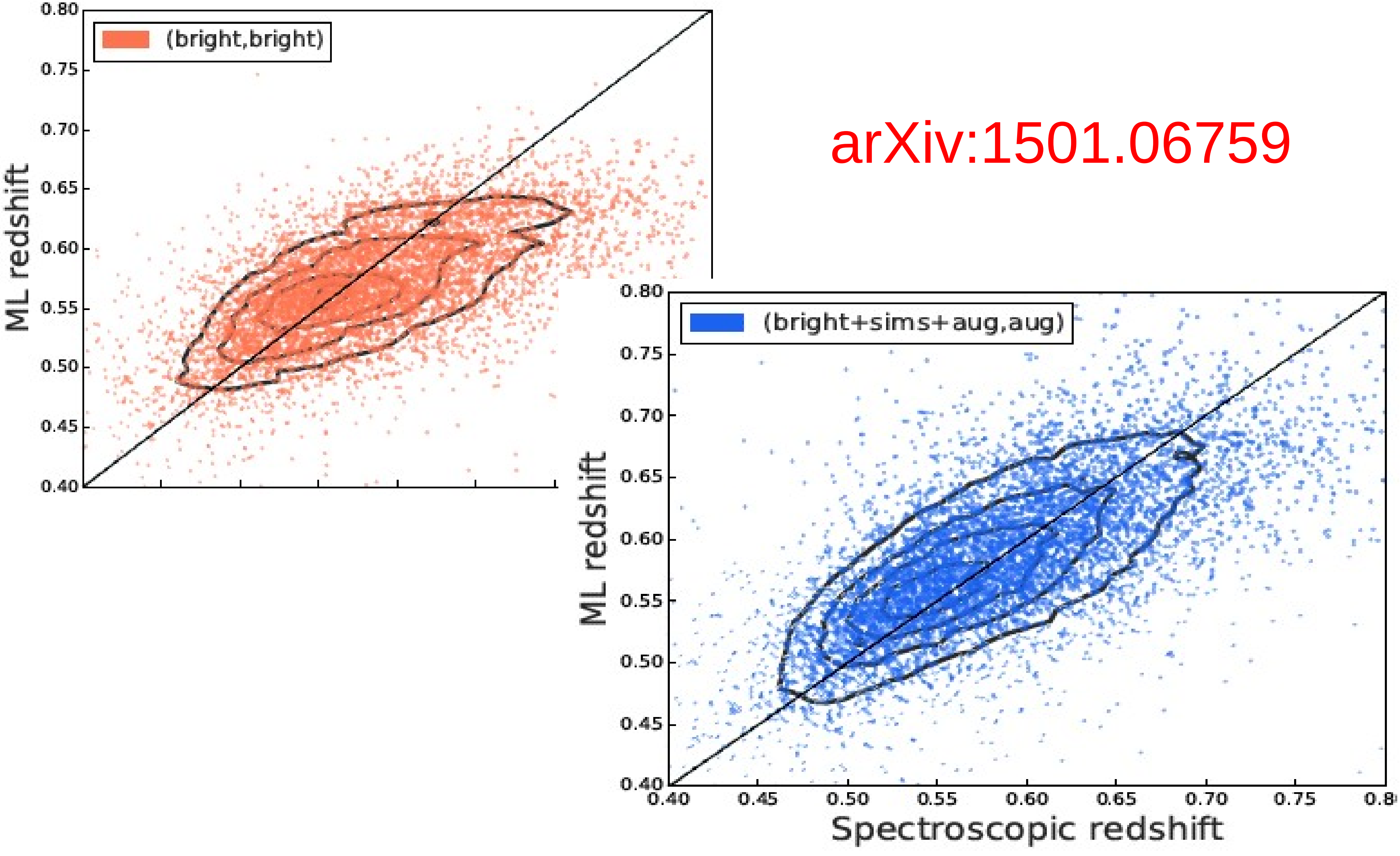
**Spectroscopic follow-up**

- Problem of accuracy
  - Too few spectroscopic data → bad estimator

**Add simulation data/template fitting**

# Data augmentation for machine learning redshifts applied to SDSS galaxies

Ben Hoyle[1,2], Markus Michael Rau[1,4], Christopher Bonnett[3], Stella Seitz[1,4]
Jochen Weller[1,2,4]




arXiv:1501.06759

# Conclusions

- Efficient and accurate modelling of the uncertainty in photometric redshifts is very important

  → Reduction of systematic biases in angular power spectrum, cosmic shear correlation functions and the critical surface density by a factor of four compared with point estimates

- Spectroscopic incompleteness is a challenge

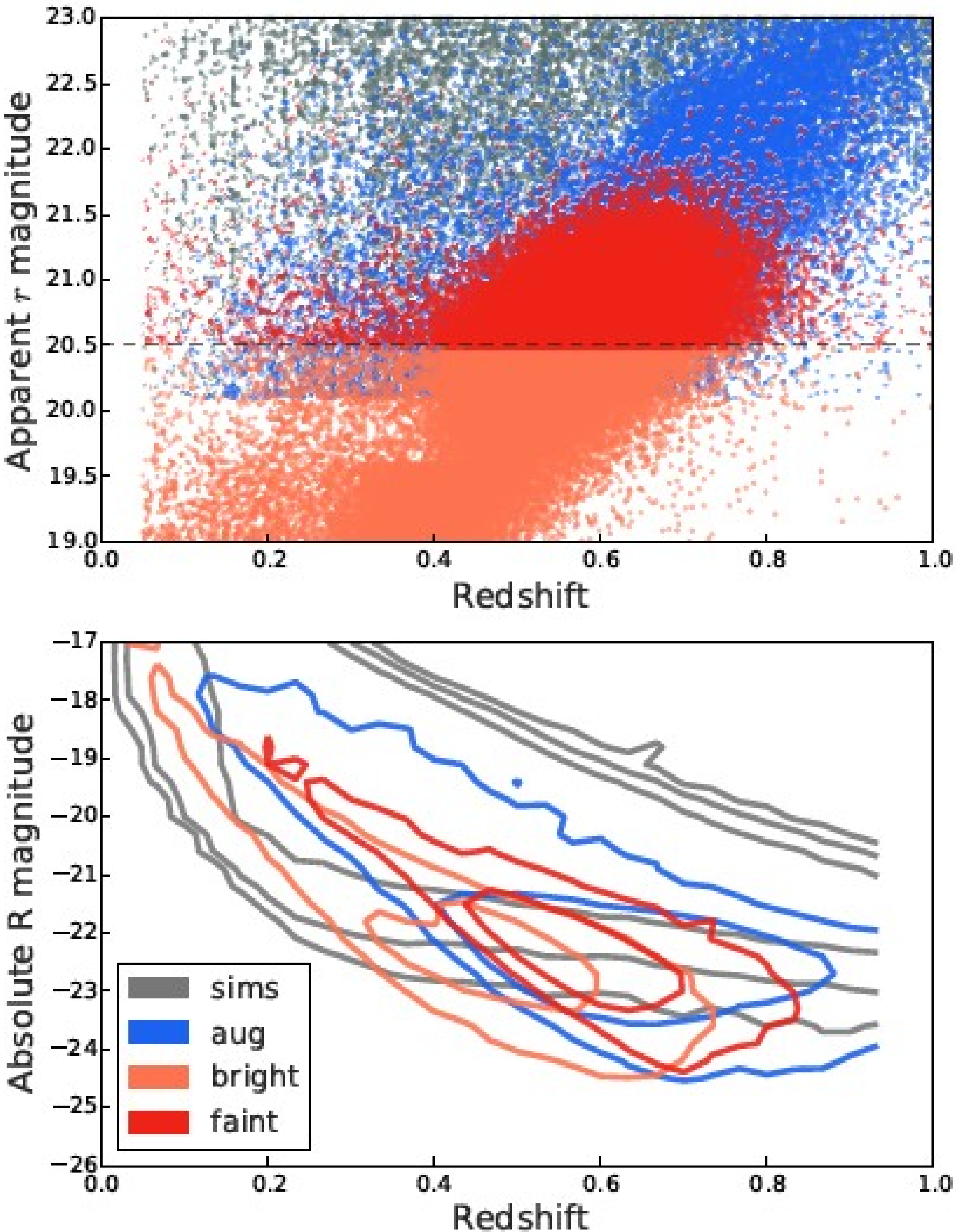
  → we are currently working on efficient ways to carry out spectroscopic follow-up

**Tuning target selection algorithms to improve galaxy redshift estimates**

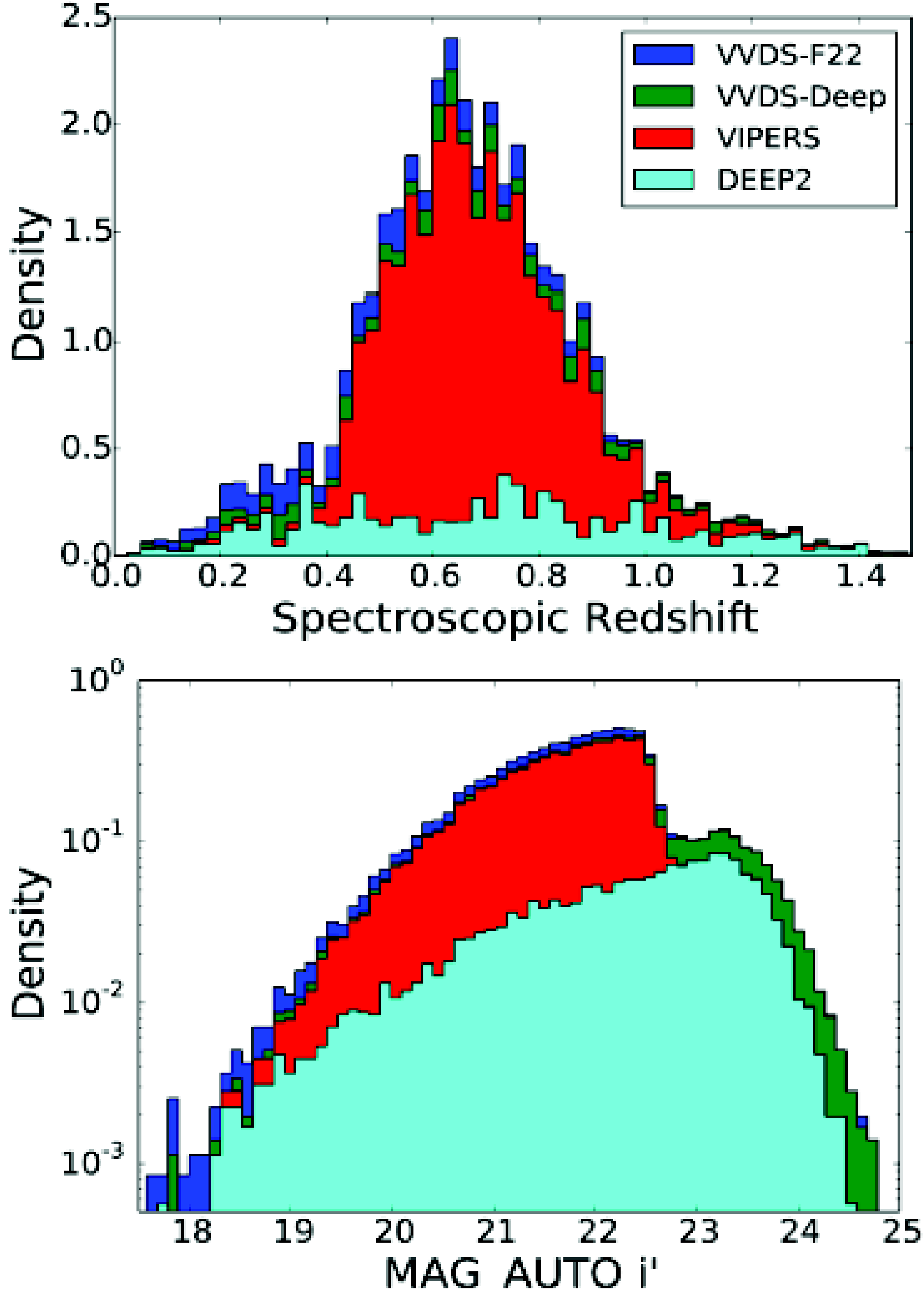
arXiv:1508.06280

Ben Hoyle[1,2], Kerstin Paech[1,2], Markus Michael Rau[1,3], Stella Seitz[1,3], Jochen Weller[1,2,3]

# Backup slides



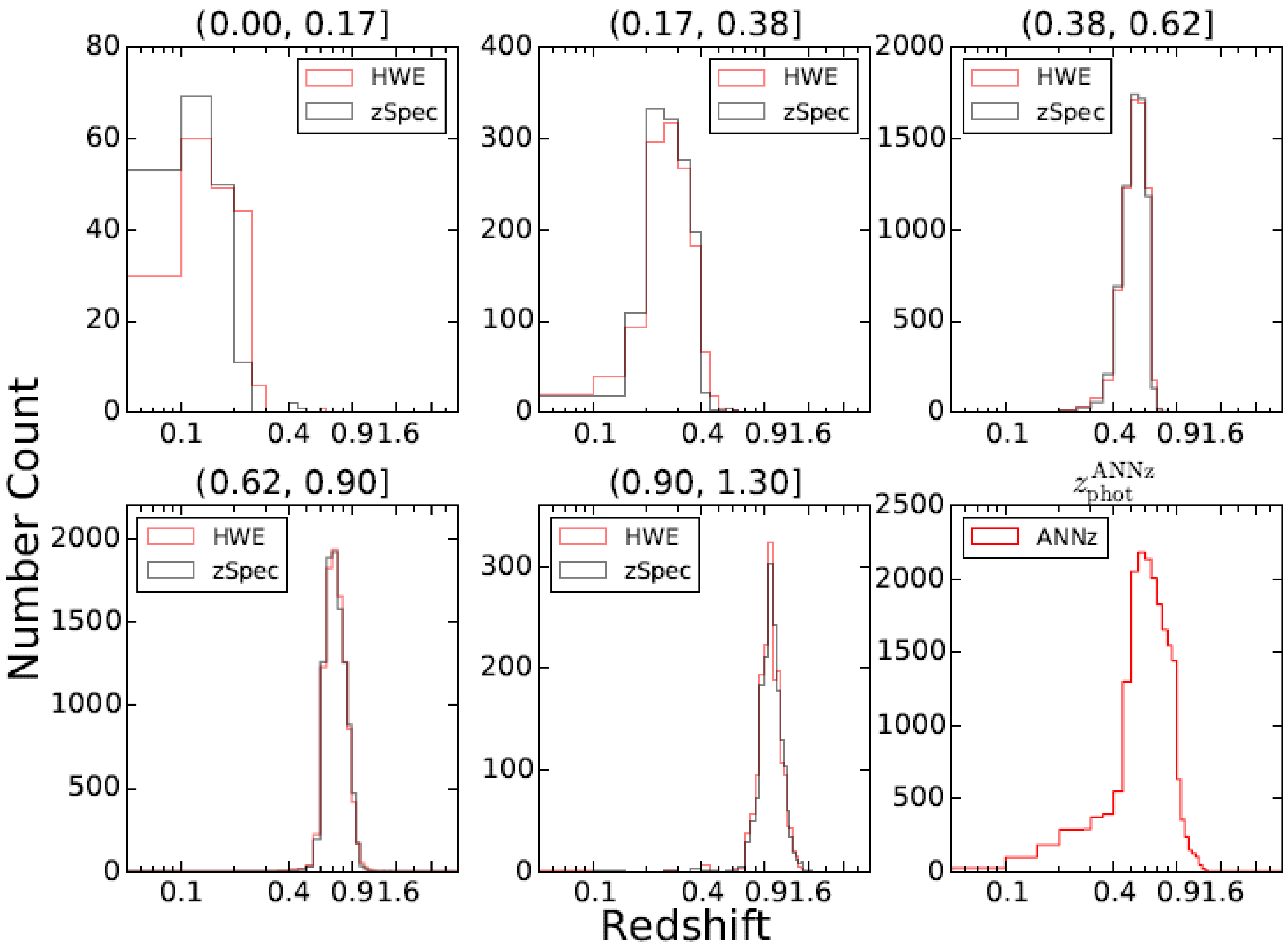
Apparent r magnitude
Redshift
Absolute R magnitude
Redshift
sims
aug
bright
faint



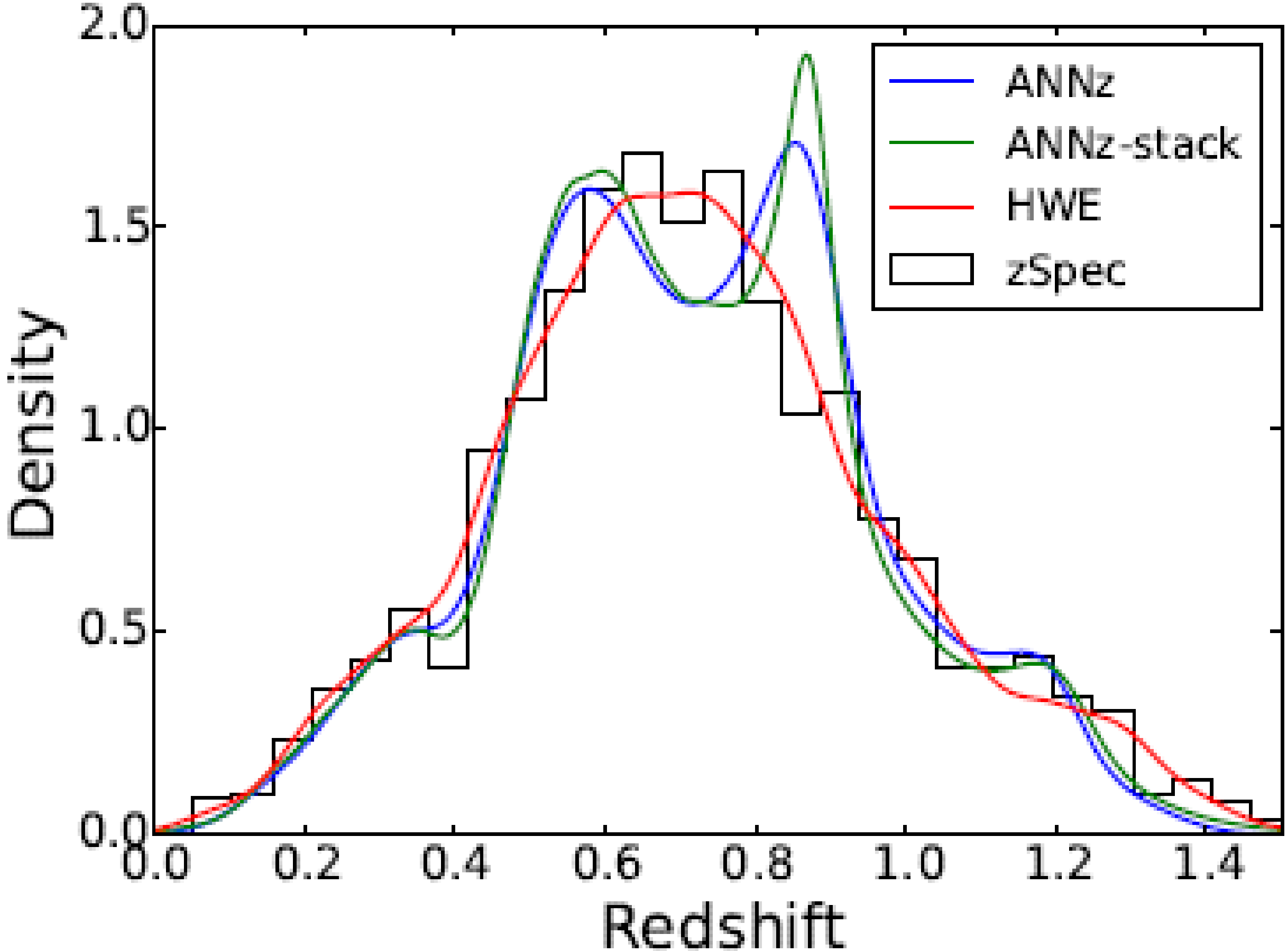
VVDS-F22
VVDS-Deep
VIPERS
DEEP2
Density
Spectroscopic Redshift
0.0
0.5
1.0
1.5
2.0
2.5
0.2
0.4
0.6
0.8
1.2
1.4
Density
MAG_AUTO i'
$10^0$
$10^{-1}$
$10^{-2}$
$10^{-3}$
18
19
20
21
22
23
24
25

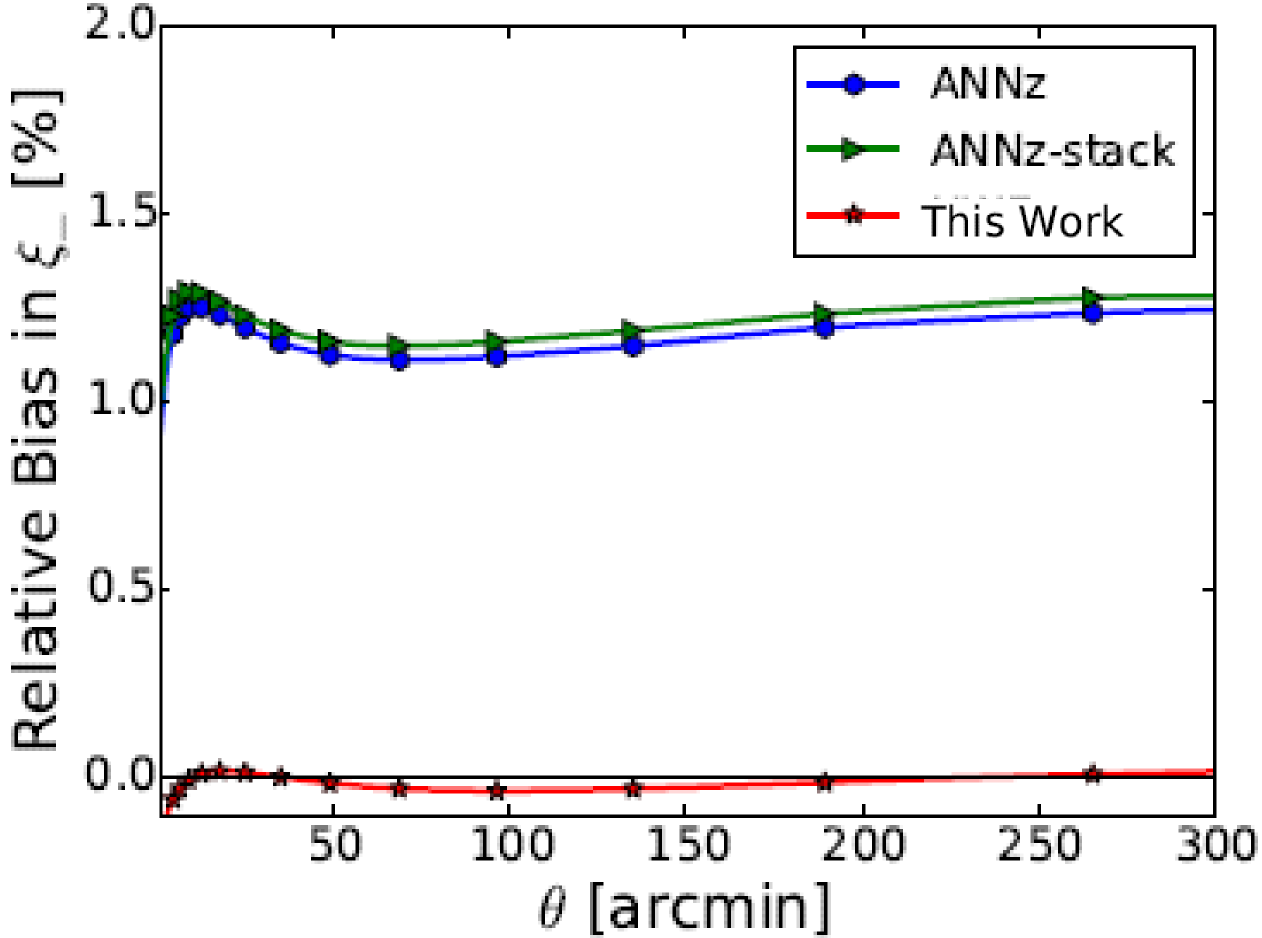
| | $\eta$ | $\sigma(\Delta z)$ | $\langle\Delta z\rangle$ | $\sigma_{68}$ |
|---|---|---|---|---|
| ANNz | 1.23% | 0.092 | -0.001 | 0.044 |
| PhotoZ | 2.27% | 0.129 | -0.008 | 0.050 |

**Table 2.** Point prediction performance of the Neural Network code ANNz and the template fitting code PhotoZ quantified by the metrics described in §5.1.

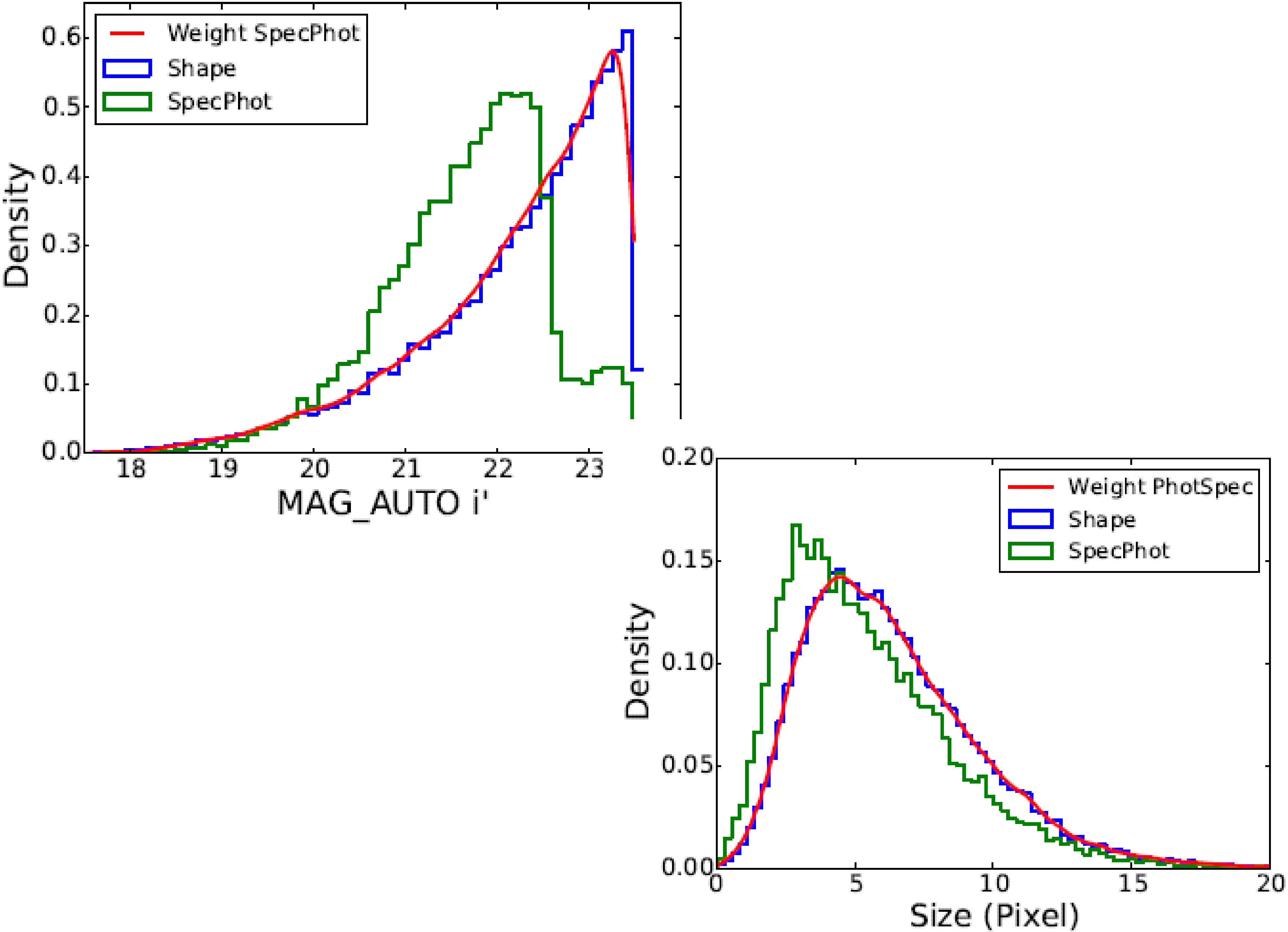


(0.00, 0.17]
(0.17, 0.38]
(0.38, 0.62]
(0.62, 0.90]
(0.90, 1.30]
$z_{\rm phot}^{\rm ANNz}$
HWE
zSpec
ANNz
Number Count
Redshift


ANNz
ANNz-stack
HWE
zSpec
Density
Redshift
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4



Relative Bias in $\xi_-$ [%]
θ [arcmin]
ANNz
ANNz-stack
This Work
2.0
1.5
1.0
0.5
0.0
50
100
150
200
250
300


Weight SpecPhot
Shape
SpecPhot
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
18
19
20
21
22
23
MAG_AUTO i'
Weight PhotSpec
Shape
SpecPhot
Density
0.00
0.05
0.10
0.15
0.20
0
5
10
15
20
Size (Pixel)