Overview of HSC photo-z's

Jean Coupon, Bau-Ching Hsieh, Sogo Mineo,
Jun Nakano, Atsushi Nishizawa, Josh Speagle
Masayuki Tanaka

The question

The problem

HSC photo-z paper (arXiv: 1704.05988)

Photometric Redshifts for the Hyper Suprime-Cam Subaru Strategic Program Data Release 1

Masayuki Tanaka[1], **Jean Coupon**[2], **Bau-Ching Hsieh**[3], **Sogo Mineo**[1], **Atsushi J. Nishizawa**[4], **Joshua Speagle**[5], **Hisanori Furusawa**[1], **Satoshi Miyazaki**[1,6], **Hitoshi Murayama**[7,8,9]

[1]National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
[2]Department of Astronomy, University of Geneva, ch. dÉcogia 16, 1290 Versoix, Switzerland
[3]Academia Sinica Institute of Astronomy and Astrophysics, P.O. Box 23-141, Taipei 10617, Taiwan
[4]Institute for Advanced Research, Nagoya University Furocho, Chikusa-ku, Nagoya, 464-8602 Japan
[5]Harvard University, 60 Garden St., Cambridge, MA 02138, USA
[6]SOKENDAI(The Graduate University for Advanced Studies), Mitaka, Tokyo, 181-8588, Japan
[7]Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), University of Tokyo, Chiba 277-8582, Japan
[8]Department of Physics and Center for Japanese Studies, University of California, Berkeley, CA 94720, USA
[9]Theoretical Physics Group, Lawrence Berkeley National Laboratory, MS 50A-5104, Berkeley, CA 94720

*E-mail: masayuki.tanaka@nao.ac.jp

Received ; Accepted

Abstract

Photometric redshifts are a key component of many science objectives in the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP). In this paper, we describe and compare the codes used to compute photometric redshifts for HSC-SSP, how we calibrate them, and the typical accuracy we achieve with HSC five-band photometry ($grizy$). We introduce a new point estimator based on an improved loss function and demonstrate that it works better than other commonly used estimators. We find that our photo-z's are most accurate from $0.3 \lesssim z \lesssim 1.5$, where we can straddle the 4000Å break. We achieve $\sigma(\Delta z/(1+z)) \sim 0.04$ and an outlier rate of about 10% for galaxies down to $i = 25$ within this redshift range, which should enable many science cases for HSC-SSP. We also characterize the accuracy of our redshift probability distribution function (PDF) and discover that some codes over/under-estimate the redshift uncertainties, which have implications for $N(z)$ reconstruction. Our photo-z's for the Deep and UltraDeep layers are available in the public data release, while those for the Wide layer will soon be made available. Both our catalog products (such as point estimates) and full PDFs are available from the data release site, `https://hsc-release.mtk.nao.ac.jp/`.

Key words: surveys, galaxies: distances and redshifts, galaxies: general, cosmology: observations

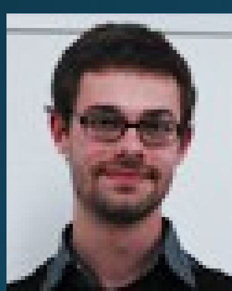
Our photo-z's for Deep and UltraDeep are publicly available.

Photo-z's for the Wide layer will be made available later this month.

The HSC photo-z team

Masayuki Tanaka
Template fitting code
(Tanaka 2015)

Atsushi Nishizawa
MLZ (Carrasco Kind+ 14)



Jean Coupon
NNPZ (no paper yet)

Sogo Mineo
EPHOR (no paper yet)

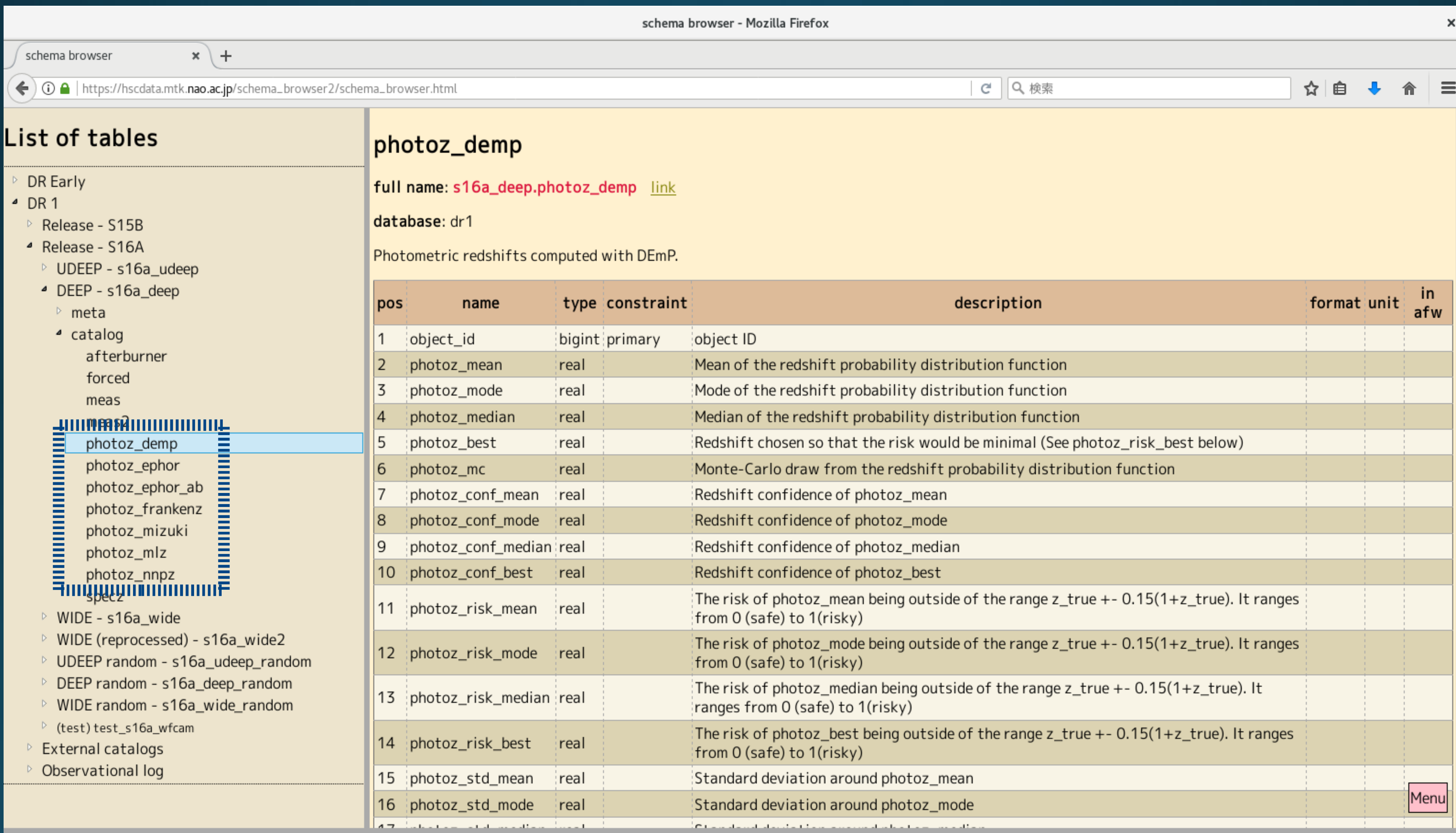
Bau-Ching Hsieh
DEmP
(Hsieh+ 2014)

Jun Nakano
Deep-learning (no paper yet)

Josh Speagle
Franken-Z (paper in prep)

Database tables

We have a database table for each code. We let science users choose the code.



photoz_demp

full name: s16a_deep.photoz_demp link

database: dr1

Photometric redshifts computed with DEmP.

pos	name	type	constraint	description	format	unit	in afw
1	object_id	bigint	primary	object ID			
2	photoz_mean	real		Mean of the redshift probability distribution function			
3	photoz_mode	real		Mode of the redshift probability distribution function			
4	photoz_median	real		Median of the redshift probability distribution function			
5	photoz_best	real		Redshift chosen so that the risk would be minimal (See photoz_risk_best below)			
6	photoz_mc	real		Monte-Carlo draw from the redshift probability distribution function			
7	photoz_conf_mean	real		Redshift confidence of photoz_mean			
8	photoz_conf_mode	real		Redshift confidence of photoz_mode			
9	photoz_conf_median	real		Redshift confidence of photoz_median			
10	photoz_conf_best	real		Redshift confidence of photoz_best			
11	photoz_risk_mean	real		The risk of photoz_mean being outside of the range z_true +- 0.15(1+z_true). It ranges from 0 (safe) to 1(risky)			
12	photoz_risk_mode	real		The risk of photoz_mode being outside of the range z_true +- 0.15(1+z_true). It ranges from 0 (safe) to 1(risky)			
13	photoz_risk_median	real		The risk of photoz_median being outside of the range z_true +- 0.15(1+z_true). It ranges from 0 (safe) to 1(risky)			
14	photoz_risk_best	real		The risk of photoz_best being outside of the range z_true +- 0.15(1+z_true). It ranges from 0 (safe) to 1(risky)			
15	photoz_std_mean	real		Standard deviation around photoz_mean			
16	photoz_std_mode	real		Standard deviation around photoz_mode			
17	photoz_std_median	real		Standard deviation around photoz_median			

Photo-z production runs

After a data release, the photo-z team prepares

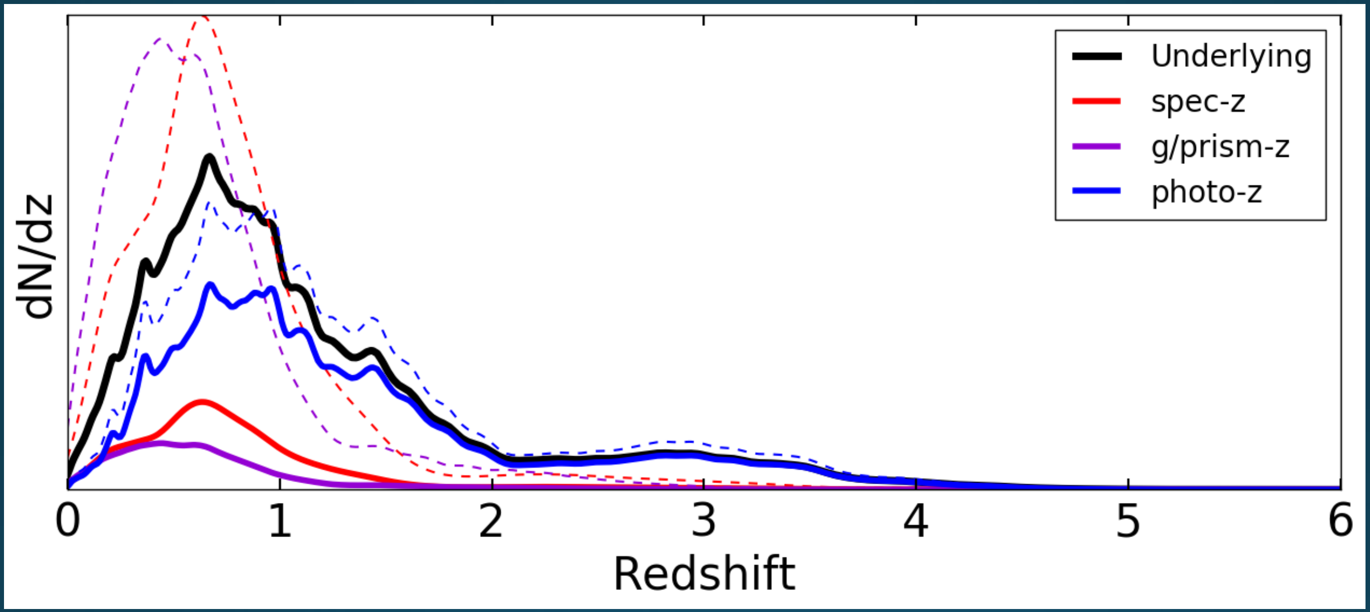
(a) training sample
(b) target sample

I let people train their codes and do the production in any way they want, but I ask them to

(1) not to apply stringent cuts on the target sample
(2) generate P(z) in a common format
(3) write a release note
(4) submit photo-z products by a deadline

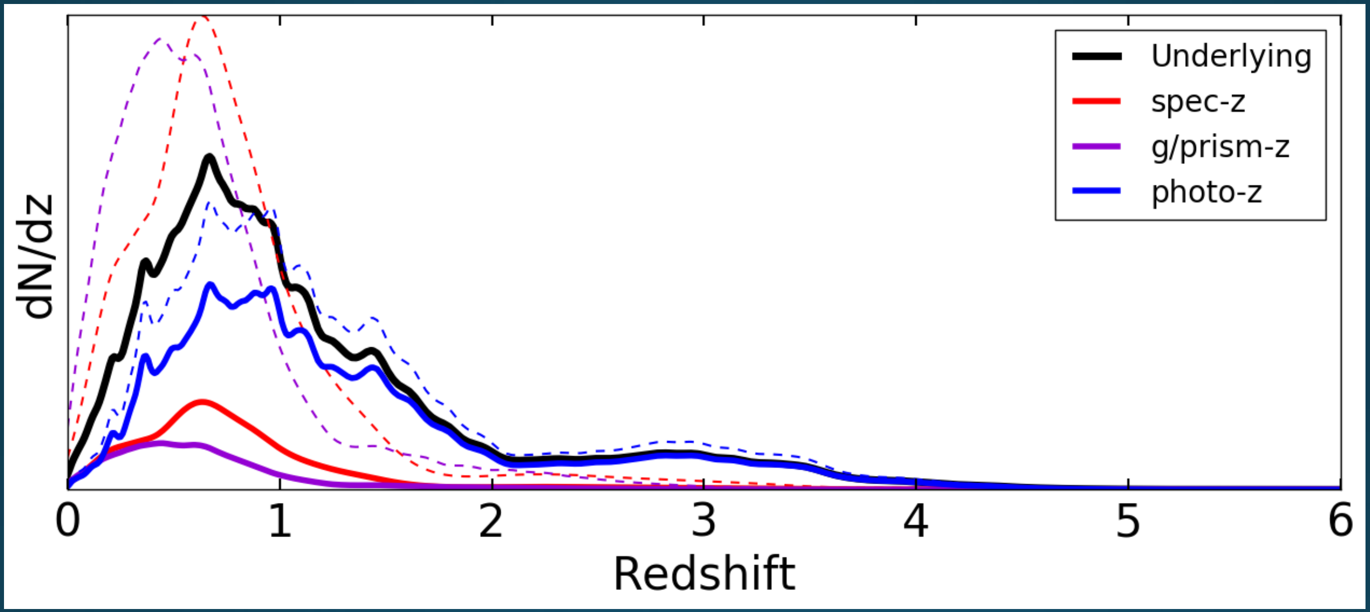
This works reasonably well, but...

Training sample construction



- The Wide layer goes down to i~26.0.
- No spectroscopic sample goes down to that faint mags…
- We rely on many-band photo-z's from COSMOS.

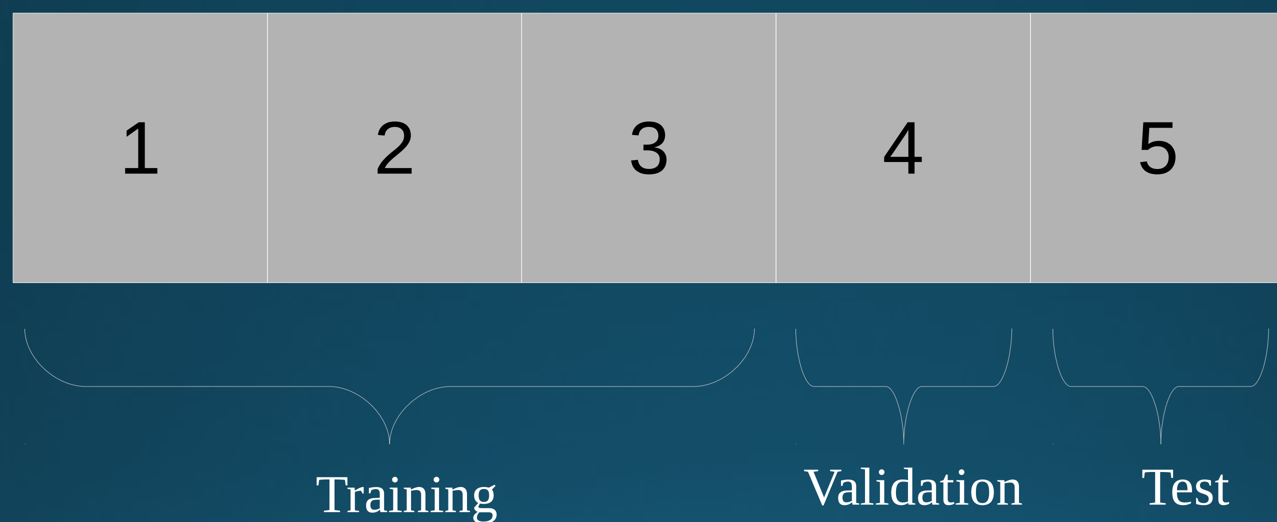
Training sample construction



- The training sample is combination of spec-z, g/prism-z and high-accuracy photo-z from the entire survey footprint.
- We compute a weight for each object in order to reproduce the Wide galaxy distribution in the multi-color space.
- We suffer from the wiggles in dn/dz in COSMOS.

Code training

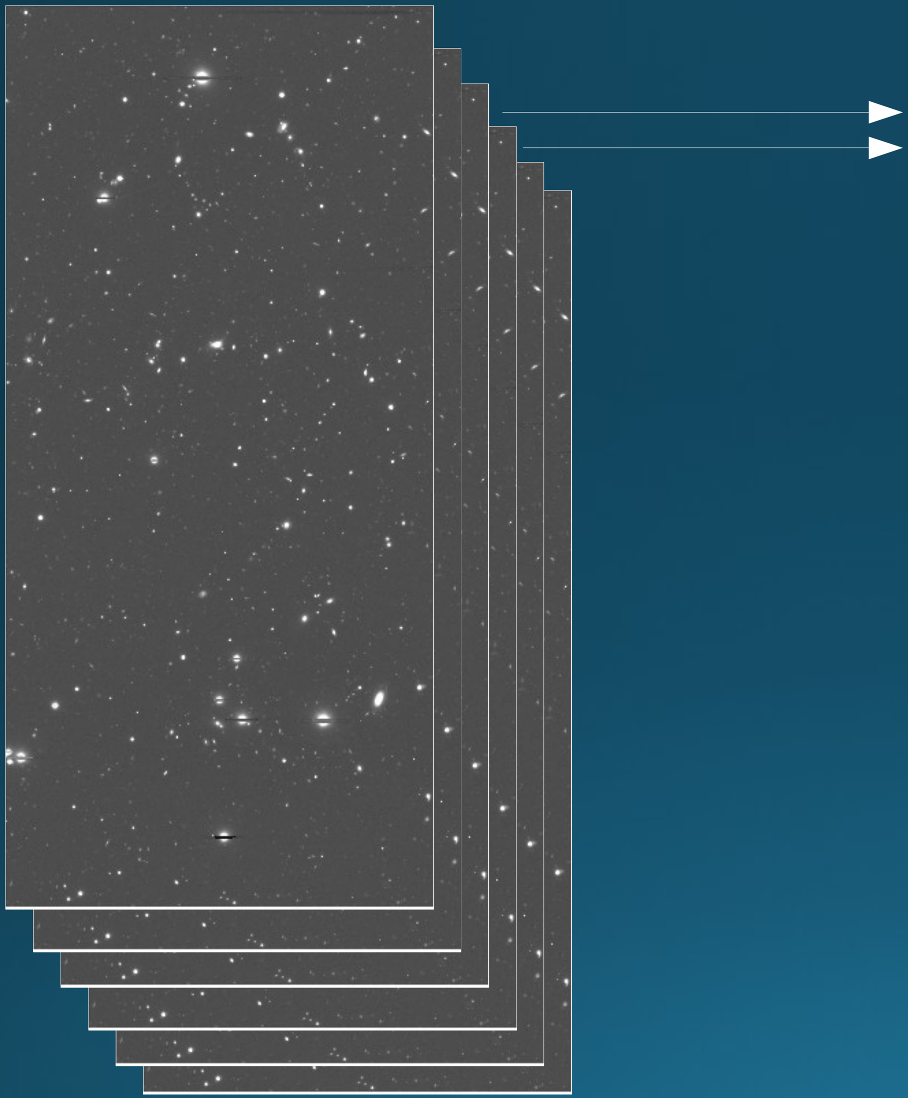
We employed the classical hold-out validation:



Each fold contains about 90k objects. One of us did cross-validation to get a sense of the performance and then used essentially all the training sample.

Lesson: ask someone outside of your photo-z group to keep the truth table for the test fold.

COSMOS wide-depth stacks



COSMOS visits taken under various seeing conditions

Stack to the Wide depth:

- 0.5 arcsec (best)
- 0.7 arcsec (median)
- 1.0 arcsec (worst)

Photo-z risk and the best point estimates

We take a minimum risk approach to make a point photo-z estimate and its photo-z *risk*. See Mineo-sensei's talk for details. We define the loss function as an inverted Lorentzian:

$$L(\Delta z) = 1 - \frac{1}{1 + \left(\frac{\Delta z}{\gamma}\right)^2}$$

$$\Delta z = (z_{\rm phot} - z_{\rm ref})/(1 + z_{\rm ref})$$

We then define a *risk* around a point-estimate:

$$R(z_{\rm phot}) = \int {\rm d}z P(z) L\left(\frac{z_{\rm phot} - z}{1 + z}\right)$$

The *best* point estimate is where the *risk* is the smallest:

$$z_{\rm best} = {\rm argmin}(R(z_{\rm phot}))$$

Some of our codes are optimized to minimize loss, but we could minimize delta_P(z).

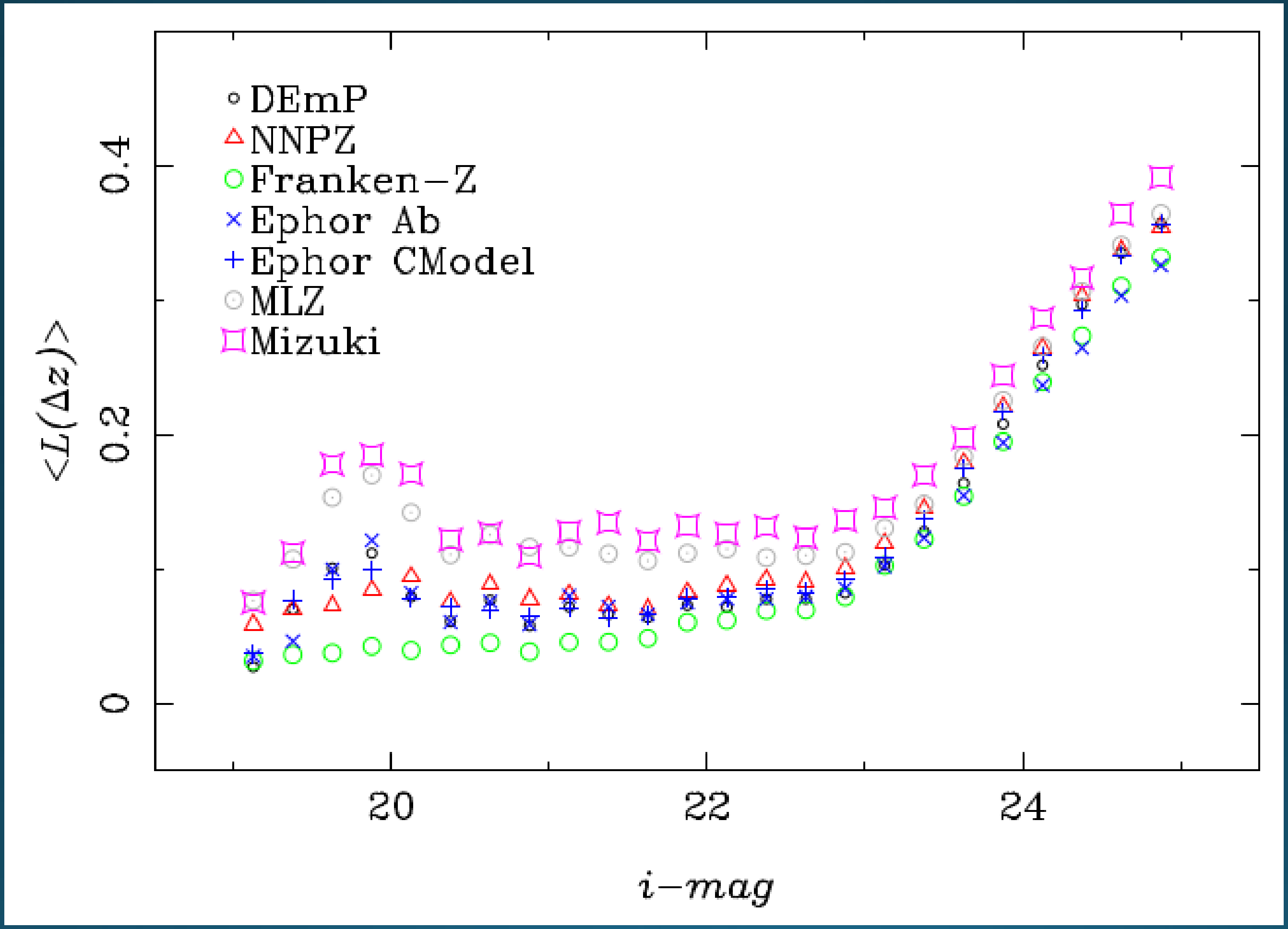
And, it works!

Point Estimator	bias	σ_{conv}	$f_{outlier,conv}$	$loss$
mean	-0.003	0.075	0.227	0.260
mode	-0.002	0.067	0.213	0.244
median	-0.001	0.066	0.199	0.236
best	-0.003	0.064	0.197	0.233

z_best gives the smallest scatter and outlier rate!

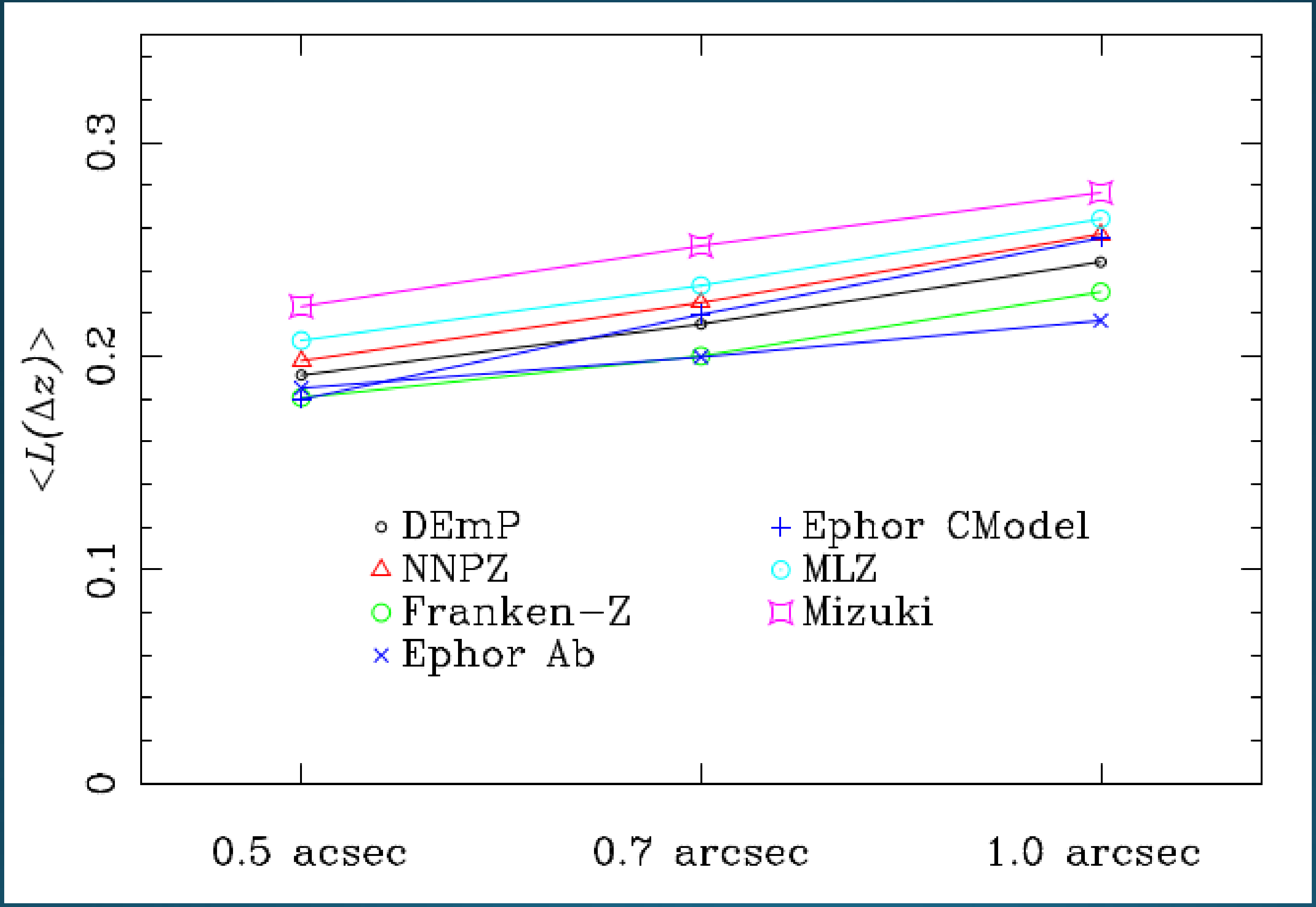
z_risk works better than z_conf (see Mineo-sensei's talk)

Code performance : point estimates



- All galaxies down to i=25 (no clipping)
- Be careful not to interpret the absolute numbers at faint mags

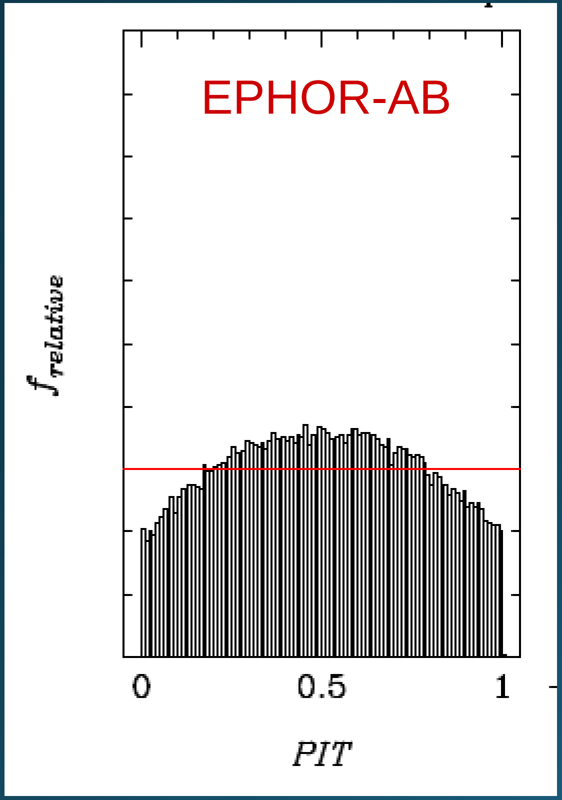
Code performance : point estimates

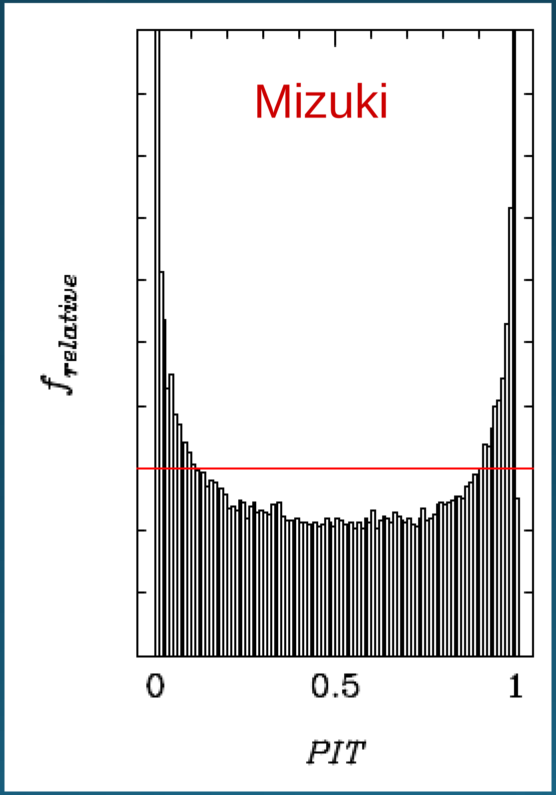


Peak-peak variation of our photo-z accuracy. Not too bad.

Photo-z performance : PDF

$$PIT(z_{ref}) = \int_0^{z_{ref}} P(z)dz$$



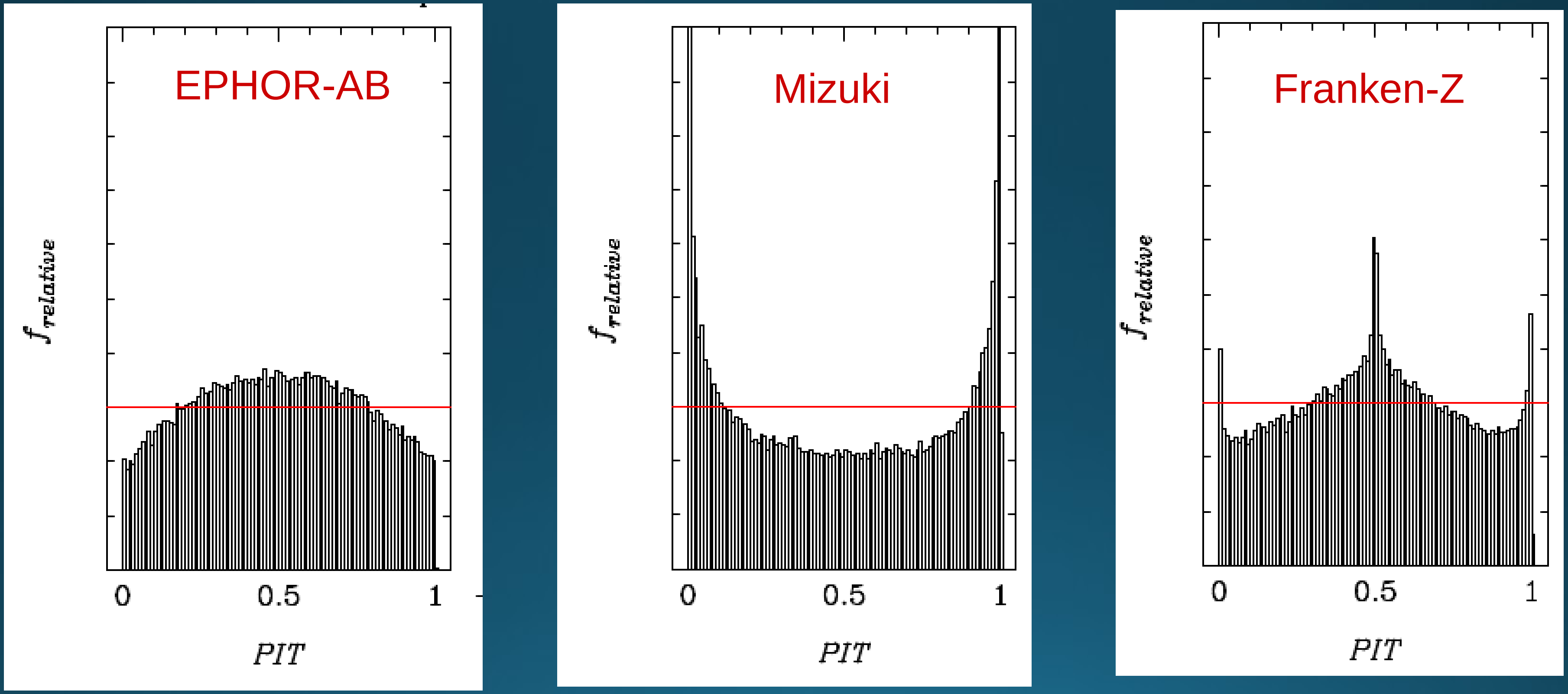


PDF is over-dispersed

PDF is under-dispersed

Photo-z performance : PDF

$$PIT(z_{ref}) = \int_0^{z_{ref}} P(z)dz$$



PDF is over-dispersed

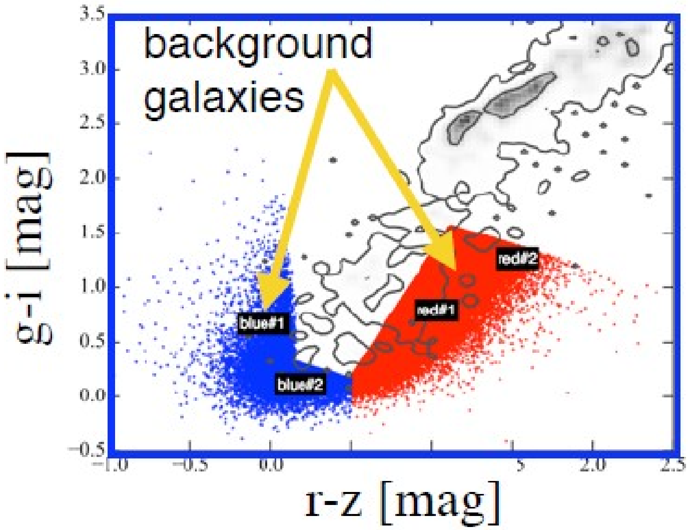
PDF is under-dispersed

PDF is too accurate

The Boldoroi corrections will be applied in our future photo-z's.

Weak-lensing tests

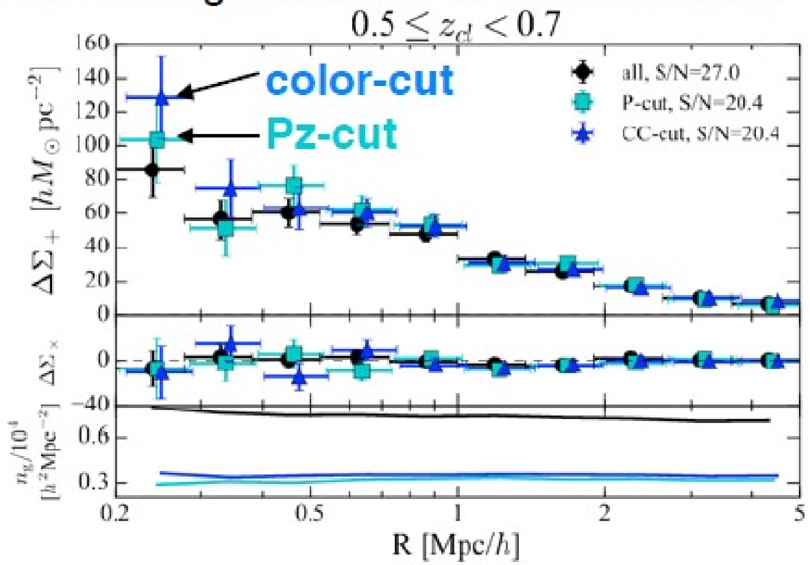
photo-z's for source selection



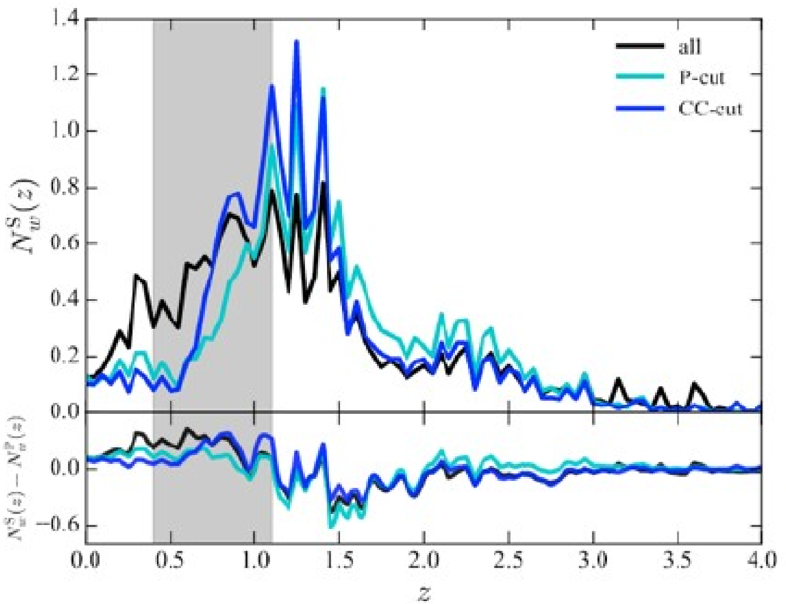
color selection vs p(z) cut

$$\int_{z_{\rm lens}+0.2}^{1.3} \mathrm{d}P(z_p) z_p > 0.98$$

Stacked WL signal around CAMIRA clusters



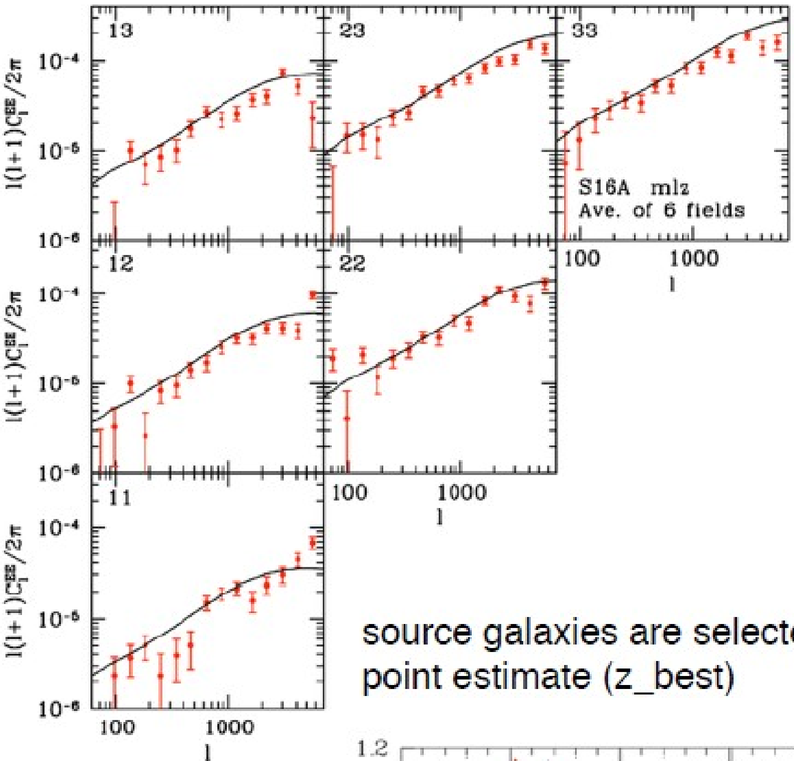
dNdz w/ re-weighting -> foreground contamination : ~ 11% (22% for no-cut)



Medezinski+ 2017 submitted

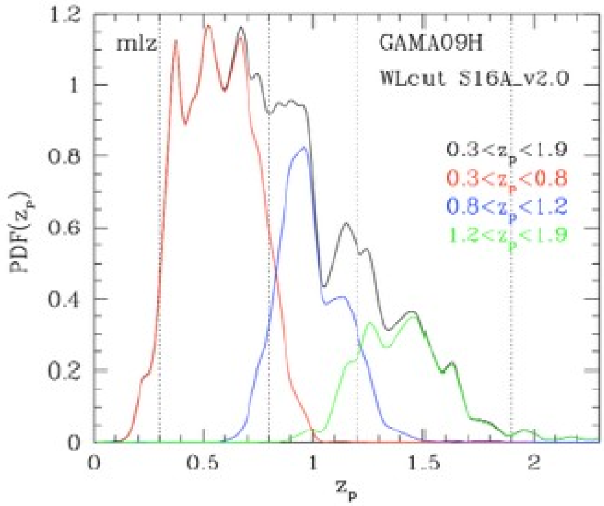
Weak-lensing tests

cosmic shear power spectra (Hikage+ 2017 in prep.)

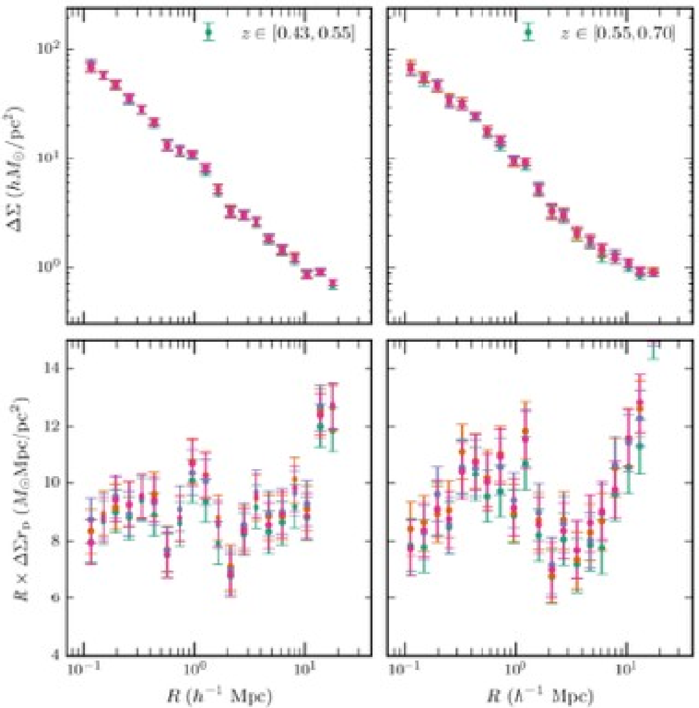


consistent with LCDM but we are still working on blind analysis

source galaxies are selected with point estimate (z_best)



CMASS galaxy lensing
(More+ 2017 in prep.)



- different colors : different photo-z codes
- No significant difference between photo-z codes

PDF files are getting very massive...

- The total file size of all the P(z)’s from all the codes amounts to ~1Tbyes (xz compressed).
- Need to compress further! Any good algorithms?

HSC public data release 1: https://hsc-release.mtk.nao.ac.jp/

Hyper Suprime-Cam Subaru Strategic Program

Data Release 1

Home | Survey | Processing | Release Data | Database | Data Access | FAQ

We peer deep into the Universe to unveil the nature of dark matter and dark energy.

Public Data Release 1

Welcome to the Hyper Suprime-Cam Subaru Strategic Program Data Release Site!
The first public release of HSC-SSP occurred on 28 February 2017. The release includes over 100 square degrees of deep multi-color data served through dedicated databases and user interfaces. The figures below shows the area covered in this release and the table gives an overview of the data in the three survey layers. Refer to our survey website for details of the survey design.