COSMOLOGICAL INFERENCE IN PHOTOMETRIC SURVEYS UNDER REDSHIFT UNCERTAINTY

COMMENTS WELCOME
ARXIV: 2211.16516

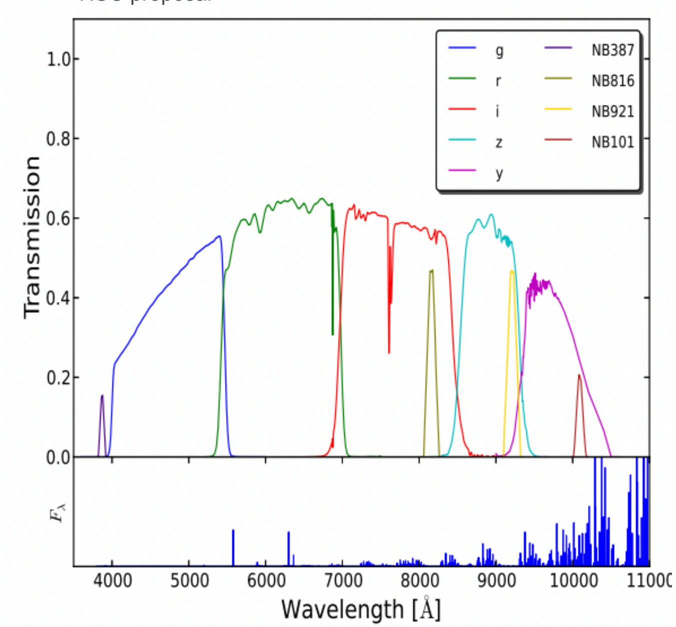
01/19/2022 Markus Michael Rau

OVERVIEW

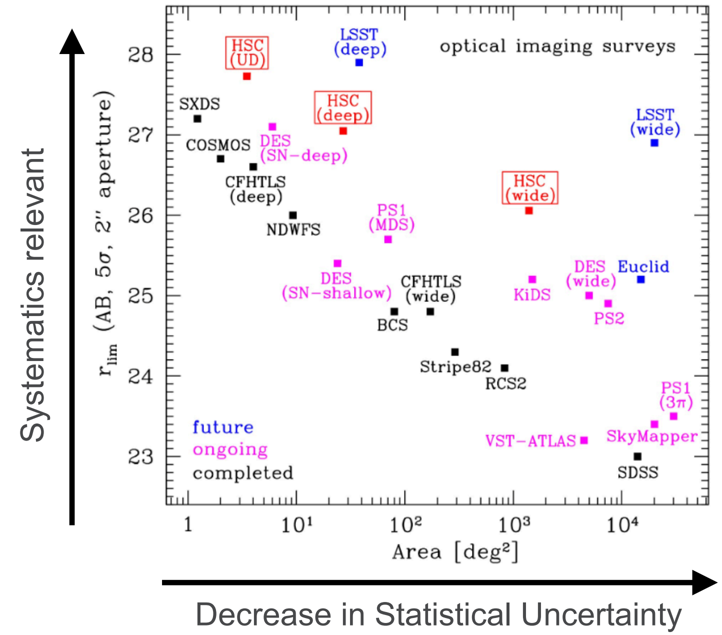
- The Hyper Suprime-Cam Subaru Strategic Program (HSC SSP)
- Motivation: Why are photometric redshifts significant for survey science
- The photometric redshift problem
- Inverse problems
- The HSC PZ analysis strategy
- HSC Year 3 results
- Conclusions

THE HSC SSP

HSC proposal



https://hsc.mtk.nao.ac.jp/ssp/survey/



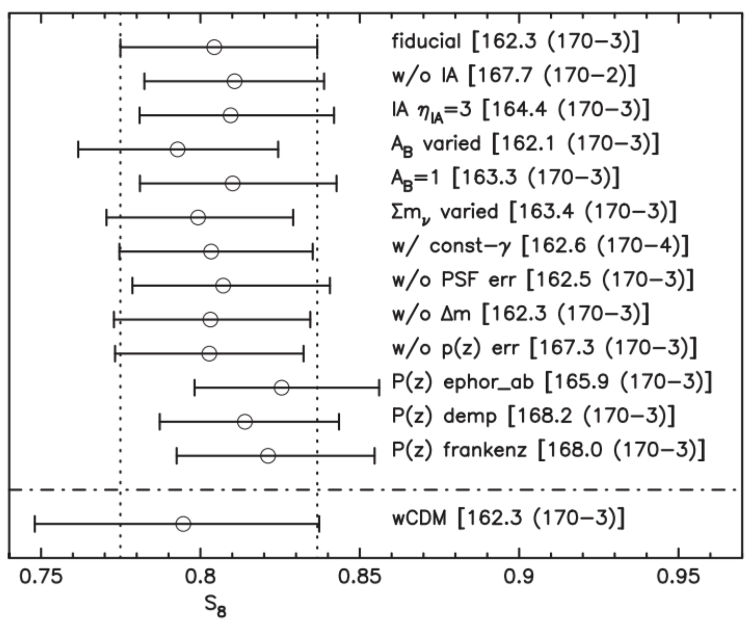
- HSC Y3 shape catalog: 417 sq. deg. Area
- HSC SSP: wide-field imaging survey with 1.77 sq. deg. field of view
- 8.2 m Subaru telescope
- 4 tomographic bins: raw (effective) galaxy number densities are 3.92 (3.77), 5.63 (5.07), 4.68 (4.00) and 2.60 (2.12) arcmin−2

MOTIVATION

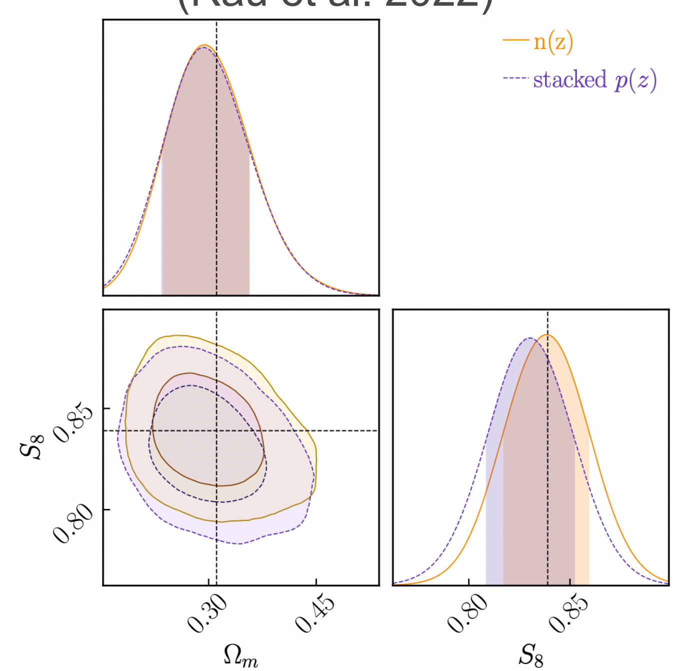
FORECAST: IMPACT OF UPDATED PZ METHODOLOGY

- S16A results indicate that photometric redshift uncertainty dominates the error budget
- Compare WL analysis of the S16A tomographic photometric redshift analysis using (S16A/novel S19A methodology) assuming an S19A covariance
- 0.5 sigma shift in the S8 towards higher values

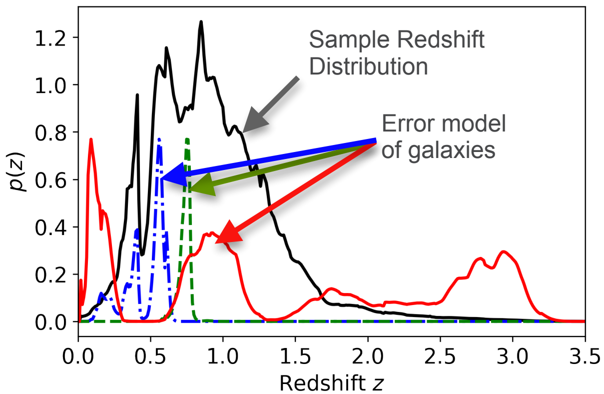
Year 1 Results
(Hamana et. al. 2020)



Forecast Year 3 Results
(Rau et al. 2022)

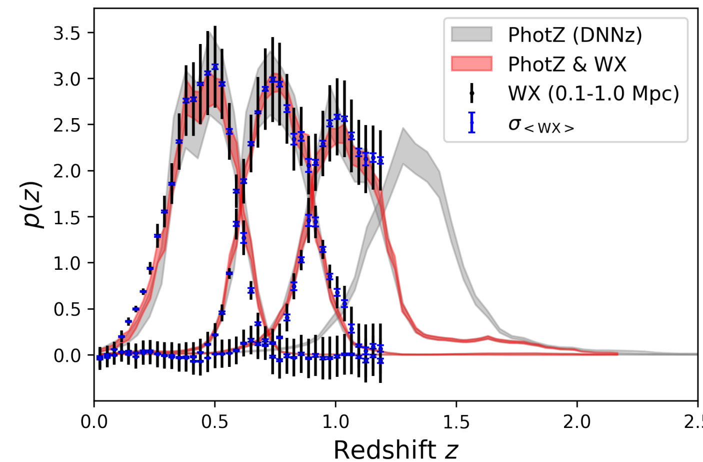


THE PHOTOMETRIC REDSHIFT PROBLEM



Deconvolution

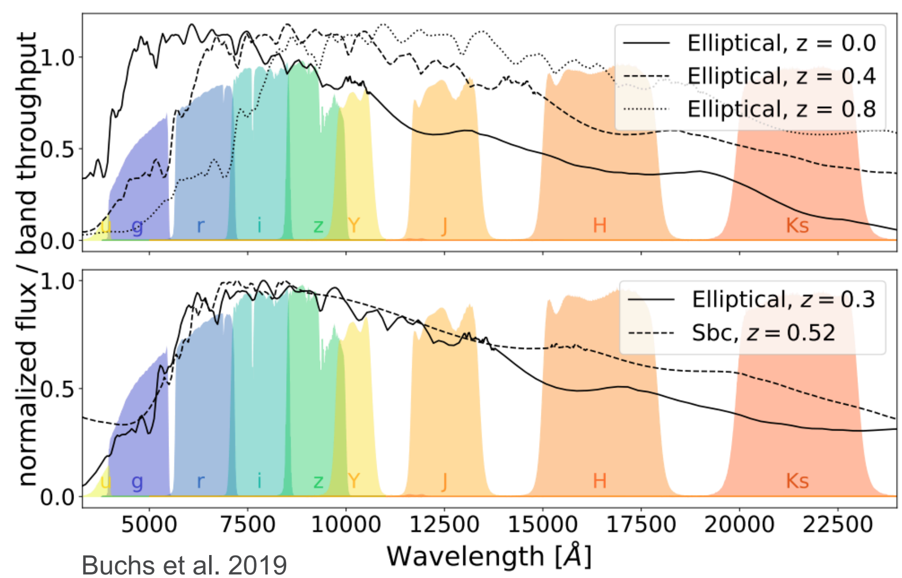
Recover using:
Spatial Distribution
Photometry



Map a high dimensional parameter space that describes galaxy populations to a low dimensional data vector

Challenges:

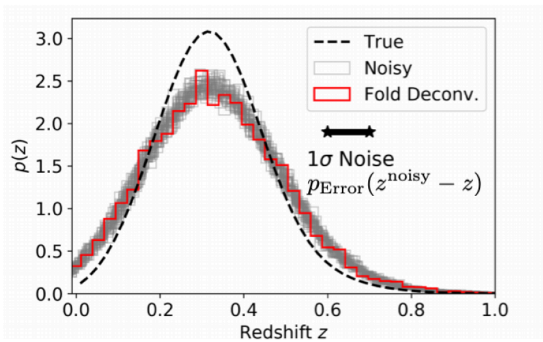
- Mapping can be ill-conditioned
- Multiple solutions reproduce similar photometry (outlier populations)
- Incomplete spectroscopic calibration for faint samples



Buchs et al. 2019

INVERSE PROBLEMS

$$p(z^{\text{noisy}})$$

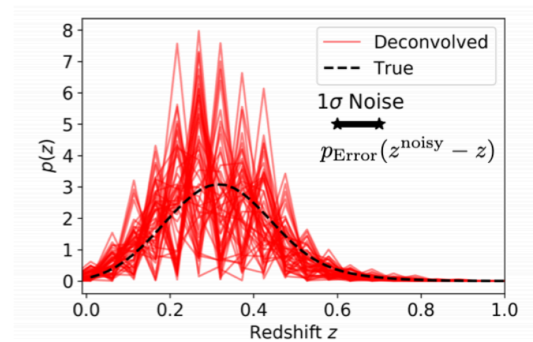


Smeared Space

Inverse Problem

Forward Problem

$$p(z)$$



True Space

Forward Model:

$$p(z^{\text{noisy}}) = \int p_{\text{Error}}(z^{\text{noisy}} - z)p(z)dz$$

$$\mathbf{p}^{\text{noisy}} = \mathbf{K}_{\text{error}} \cdot \mathbf{p}^{\text{true}}$$

$$p^{\text{true}} = \frac{\alpha_1}{\lambda_1}\phi_1 + \ldots + \frac{\alpha_n}{\lambda_n}\phi_n$$

$$\alpha_j = \phi_j^* p_{\text{noisy}}$$

Poor conditioning of Cosmological Imaging Surveys: $N_{\text{K,cond}} = 10^{10-17}$

Argonne NATIONAL LABORATORY

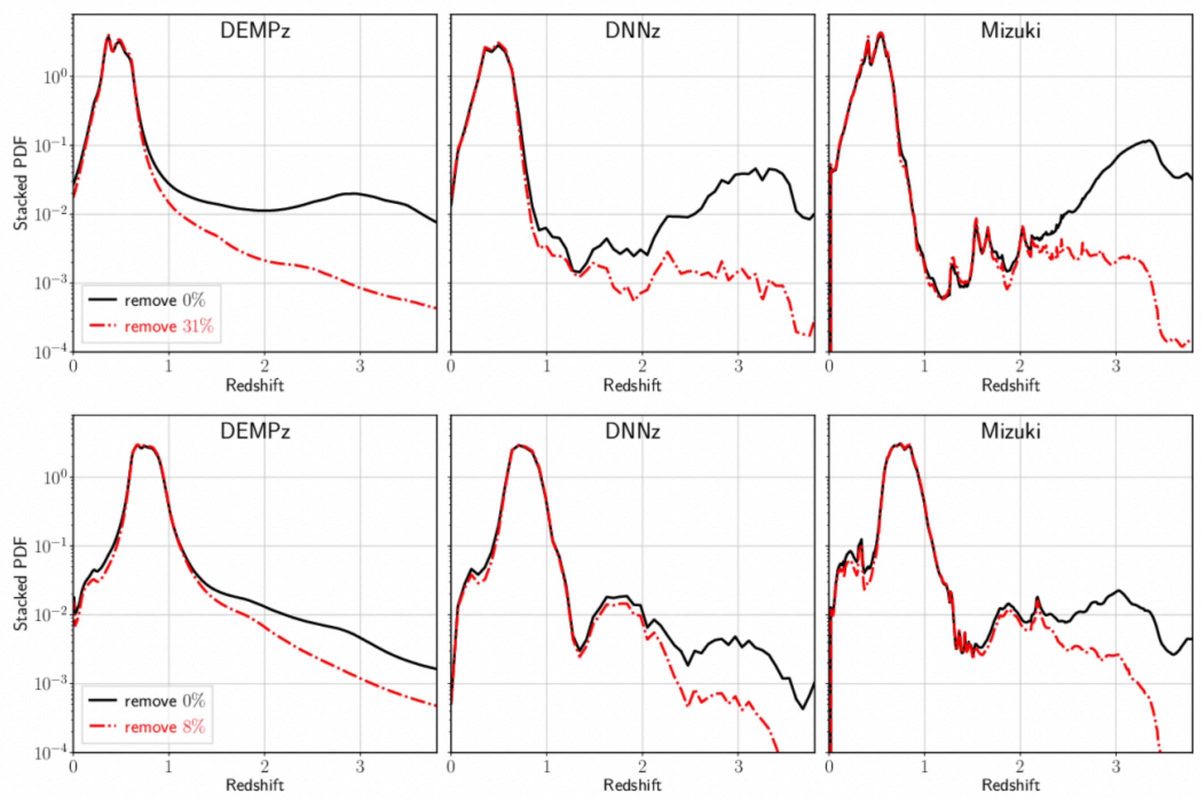
HSC PZ ANALYSIS STRATEGY

1. Optimize sample selection to avoid identifiability issues
2. Perform Sample Redshift Inference using multiple individual galaxy redshift methods.
3. Include spatial cross-correlations as an additional calibration method
4. Construct a conservative error budget for tomographic bins

SAMPLE SELECTION

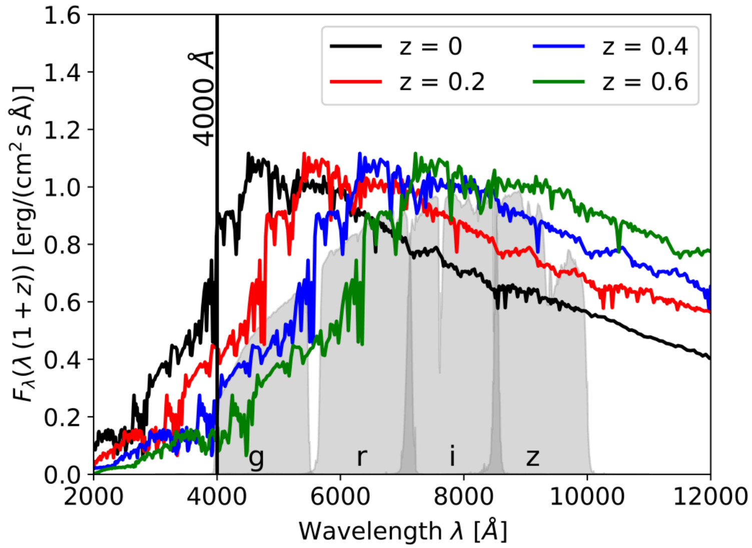
REMOVE GALAXIES WITH MULTIPLE SOLUTIONS

Criterium: $\left(z^{\rm Mizuki}_{0.975,i} - z^{\rm Mizuki}_{0.025,i}\right) < 2.7 \quad \text{and} \quad \left(z^{\rm DNNz}_{0.975,i} - z^{\rm DNNz}_{0.025,i}\right) < 2.7\,,$



Argonne NATIONAL LABORATORY

INDIVIDUAL GALAXY REDSHIFT ESTIMATION



Deconvolution/Inverse Problem →

Population Distributions (Galaxy Type, Redshift, Stellar Mass, etc.) → Selection Functions → Observed Photometry

Regression Analysis using Training Set ←

Population Distributions (Galaxy Type, Redshift, Stellar Mass, etc.) ← Selection Functions ← Observed Photometry

GALAXY REDSHIFT ESTIMATION

FORWARD MODELING BASED APPROACHES (MIZUKI TEMPLATE FITTING)

Deconvolution/Inverse Problem

Population Distributions (Galaxy Type, Redshift, Stellar Mass, etc.) → Selection Functions → Observed Photometry

$$p(\boxed{\hat{\mathbf{F}}}|\boldsymbol{\phi}_{\mathrm{nz}}, \boxed{\boldsymbol{\Omega}}) = \prod_{i=1}^{N_{\mathrm{gal}}} \int \mathrm{d}z_i \boxed{\omega_i}\, p(\boxed{\mathbf{f}_i|z_i}, \boldsymbol{\Omega})\, p(z_i|\boxed{\boldsymbol{\phi}_{\mathrm{nz}}}, \boldsymbol{\Omega}) \,.$$

Photometry all Galaxies

Quantities of Interest

Lensing Weights

Individual Galaxy Flux/Redshifts

Sample redshift histogram Heights

Likelihood of measured photometry given parameters that describe population distributions	Likelihood: Parametrization of measurement & relevant selection functions conditional on individual galaxy parameters	Prior of individual Galaxy parameters conditional on population parameters $p_{\mathrm{samp}}(z) = \sum_{i=1}^{N_{\mathrm{bins}}} \phi_{\mathrm{nz,i}} \mathbb{1}_i(z)$

GALAXY REDSHIFT ESTIMATION

CONDITIONAL DENSITY ESTIMATION APPROACHES (DEMPZ, DNNZ)

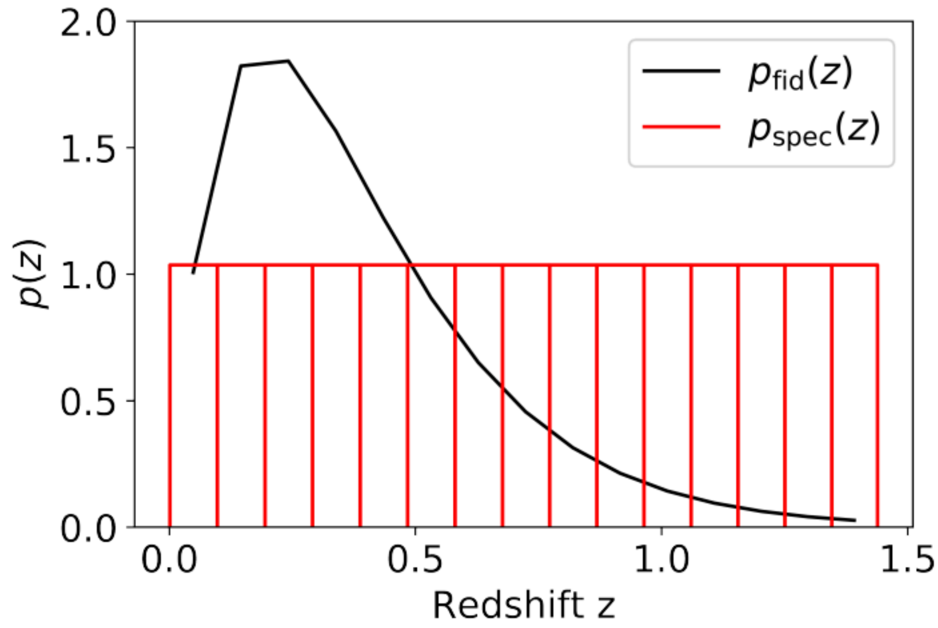
Regression Analysis

Population Distributions (Galaxy Type, Redshift, Stellar Mass, etc.) ← Selection Functions ← Observed Photometry

$$\boxed{p_{\mathrm{samp}}(z)} = \int d\mathbf{f}\, \boxed{p(z|\mathbf{f})}\, \boxed{p(\mathbf{f})} .$$

Density Estimate: Sample Redshift Distribution (Kernel, KNN, …)	Conditional Density Estimate trained on a calibration dataset	Density Estimate of the observed photometry

CROSS-CORRELATION METHOD

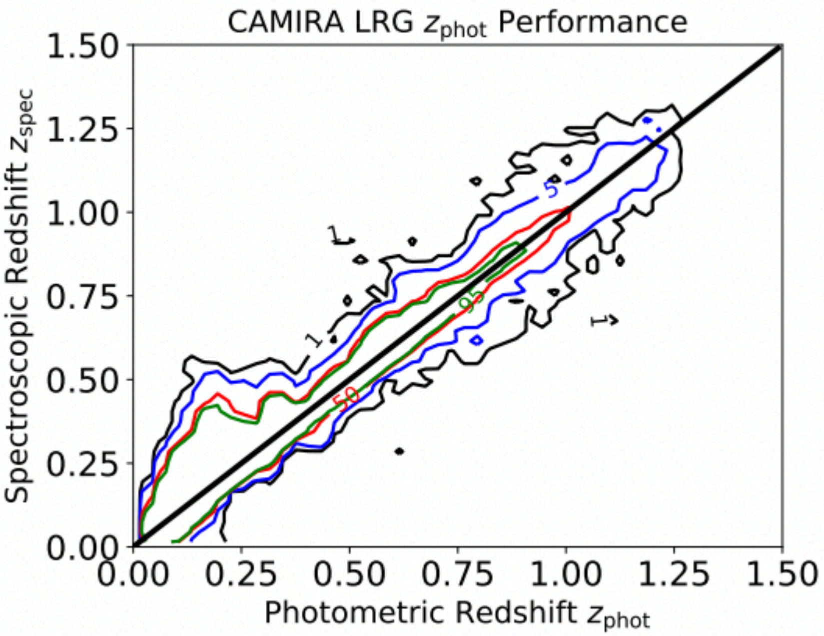


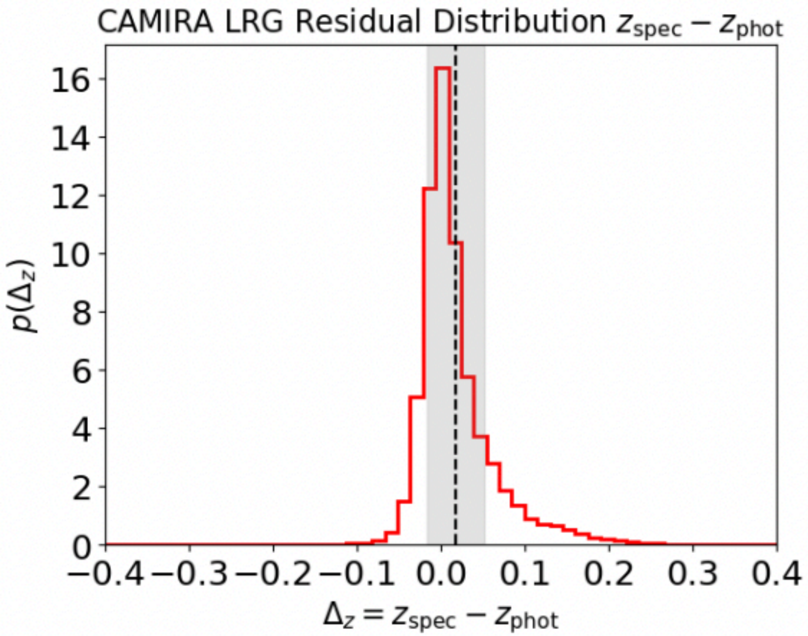
- The spatial Cross-Correlation signal between binned spectroscopic and photometric samples is proportional to the photometric redshift distribution.

$$w_{2pt} \propto p_{fid}\, p_{spec}\, b_{fid}\, b_{spec} w_{DM}$$

Argonne NATIONAL LABORATORY

UTILIZE LUMINOUS RED GALAXY SAMPLES





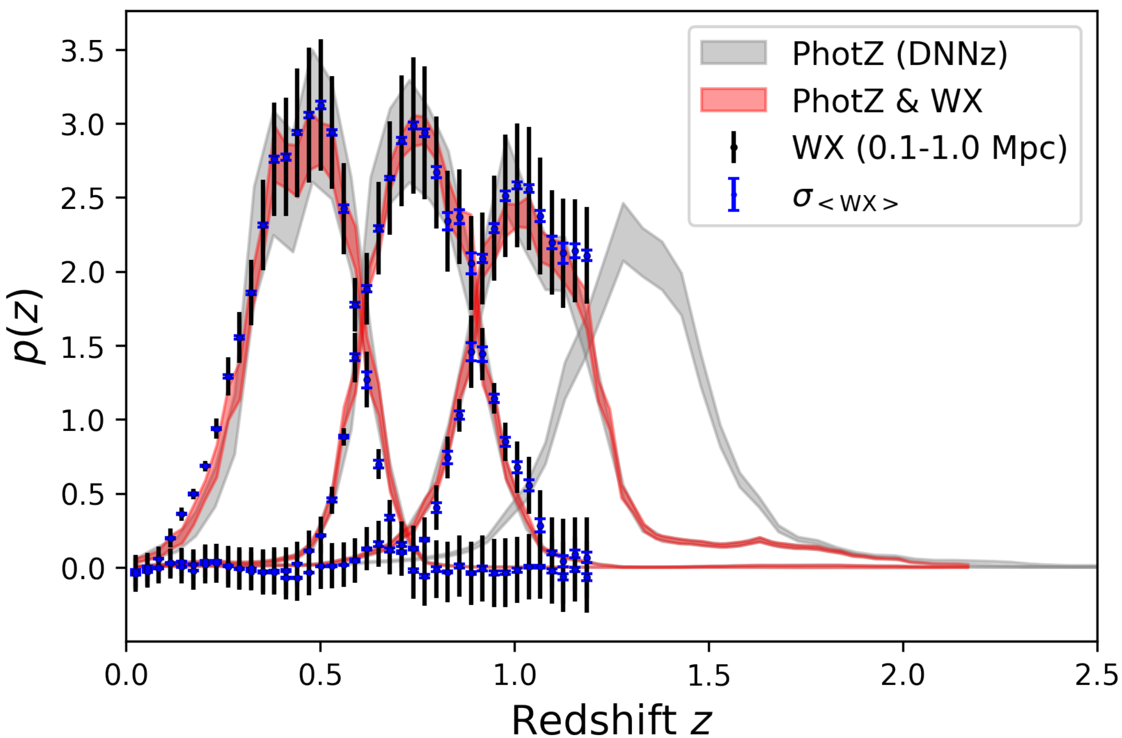
Spatial Cross-Correlations with the CAMIRA Luminous Red Galaxy Sample

- Spatial Cross-Correlations are systematically affected by PZ systematics.
- Photometric redshift error in the CAMIRA LRG sample $\sigma_{\rm z,LRG} \approx 0.03$
- Marginalize over this redshift error in the analysis

RESULTS

POSTERIOR TOMOGRAPHIC REDSHIFT DISTRIBUTION INFERENCE

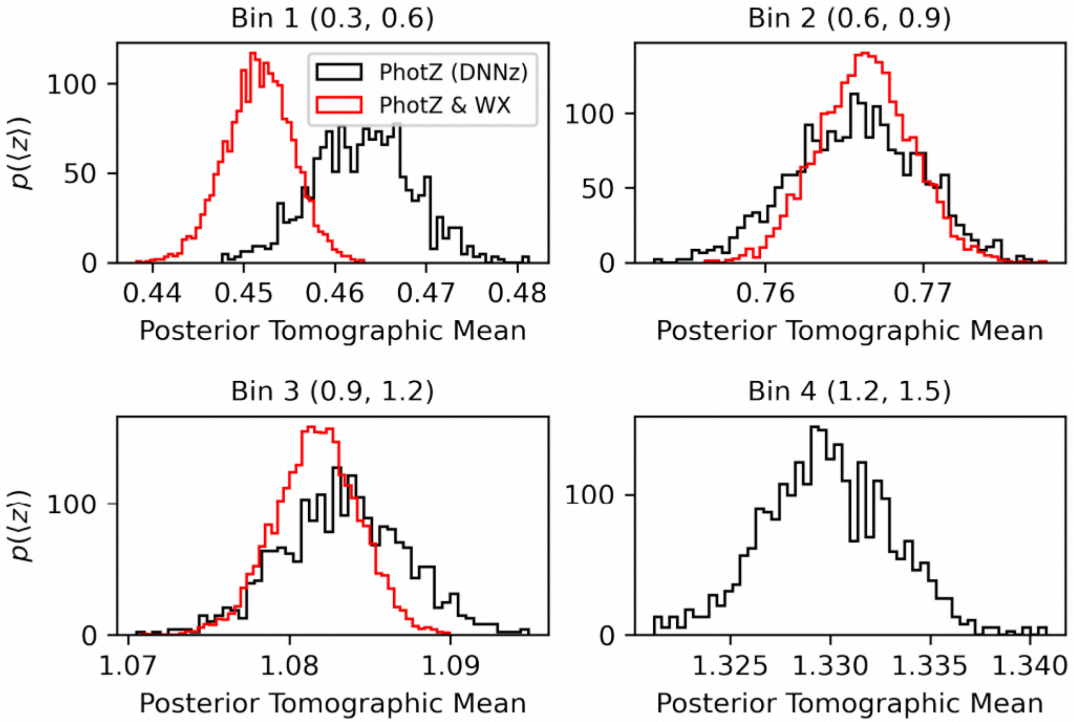
- Joint Inference between Photometry and Spatial Distribution of Galaxies (Rau, et al. 2020, 2021, 2022)
- Bayesian Hierarchical Model for the HSC Year 3 photometric Redshift Inference



Argonne NATIONAL LABORATORY

RESULTS

POSTERIOR TOMOGRAPHIC REDSHIFT MEANS



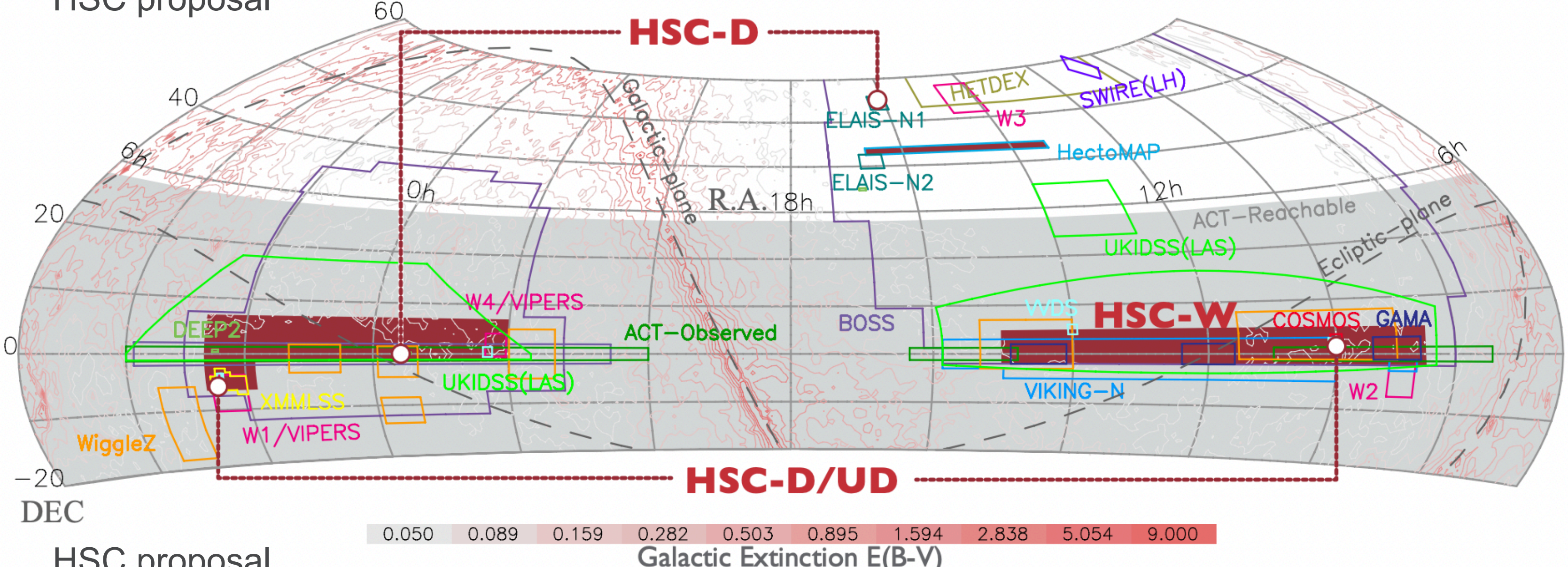
	Y1 Analysis	Y3 PhotZ (DNNz)	Y3 DEMPz	Y3 PhotZ & WX	Y3 Bayesian Surprise	Y3 Total
Bin 1	0.44 (0.0285)	0.463 (0.005)	0.463	0.452 (0.004)	3.84	0.452 (0.024)
Bin 2	0.77 (0.014)	0.766 (0.004)	0.777	0.766 (0.003)	0.10	0.766 (0.022)
Bin 3	1.05 (0.0383)	1.084 (0.004)	1.097	1.081 (0.004)	0.28	1.081 (0.031)
Bin 4	1.33 (0.0376)	1.330 (0.003)	1.350	-	-	1.330 (0.034)

SUMMARY AND CONCLUSIONS

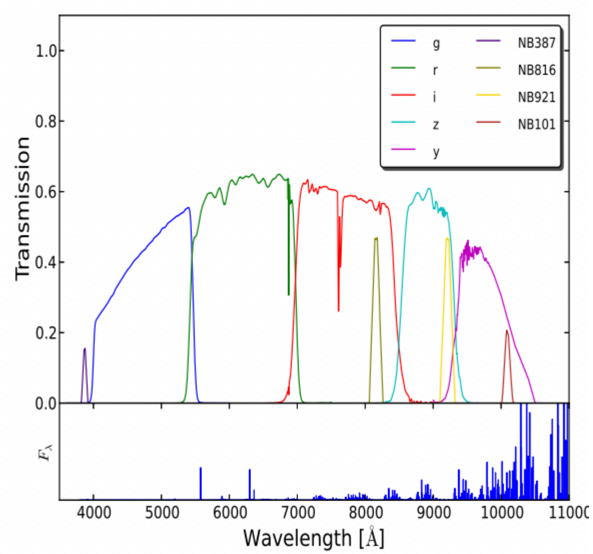
- We present a tomographic sample redshift distribution analysis of the HSC Y3 shape catalog.
- Hierarchical inference of a joint data vector informed by cross-correlations and photometry.
- We achieve good consistency between the cross-correlation and photometry-based inference.
- We present a conservative assessment of these errors and provide recommendations on prior choices
- Include multiple sources of systematic in the inference for example:
  - Cosmic Variance
  - Discrepancies between multiple models
  - Marginalization over the CAMIRA-LRG error

THE HSC SURVEY

HSC proposal



HSC proposal



- HSC Y3 shape catalog: 417 sq. deg. area
- HSC SSP: wide-field imaging survey with 1.77 sq. deg. field of view
- 8.2 m Subaru telescope
- 4 tomographic bins: raw (effective) galaxy number densities are 3.92 (3.77), 5.63 (5.07), 4.68 (4.00) and 2.60 (2.12) arcmin−2

LSST PHOTOMETRIC REDSHIFT REQUIREMENTS

Detailed requirement WL1 (Y10): Systematic uncertainty in the mean redshift of each source tomographic bin shall not exceed $0.001(1 + z)$ in the Y10 DESC WL analysis. Goal WL1 (Y1): Systematic uncertainty in the mean redshift of each source tomographic bin should not exceed $0.002(1 + z)$ in the Y1 DESC WL analysis.

Status of Stage III surveys:

KiDS-1000 and DES-Y3 currently claim to measure the mean redshifts to a precision of ~0.01. It is necessary to improve the constraints by one order of magnitude, considering we have a deeper survey.

INDIVIDUAL GALAXY REDSHIFT ESTIMATION

FORWARD MODELING BASED APPROACHES (MIZUKI TEMPLATE FITTING)

Deconvolution/Inverse Problem

Population Distributions (Galaxy Type, Redshift, Stellar Mass, etc.) → Selection Functions → Observed Photometry

$$p(\boxed{\hat{\mathbf{F}}}|\boldsymbol{\phi}_{\rm nz},\boxed{\boldsymbol{\Omega}}) = \prod_{i=1}^{N_{\rm gal}} \int \mathrm{d}z_i \boxed{\omega_i} p(\boxed{\mathbf{f}_i|z_i}, \boldsymbol{\Omega}) p(z_i|\boxed{\boldsymbol{\phi}_{\rm nz}}, \boldsymbol{\Omega}) \, .$$

Photometry all Galaxies

Quantities of Interest

Lensing Weights

Individual Galaxy Flux/Redshifts

Sample redshift histogram Heights

Likelihood of measured photometry given parameters that describe population distributions

Likelihood: Parametrization of measurement & relevant selection functions conditional on individual galaxy parameters

Prior of individual Galaxy parameters conditional on population parameters

$$p_{\rm samp}(z) = \sum_{i=1}^{N_{\rm bins}} \phi_{\rm nz,i} \mathbb{1}_i(z)$$

Argonne NATIONAL LABORATORY

INDIVIDUAL GALAXY REDSHIFT ESTIMATION

CONDITIONAL DENSITY ESTIMATION APPROACHES (DEMPZ, DNNZ)

Regression Analysis

Population Distributions (Galaxy Type, Redshift, Stellar Mass, etc.) ← Selection Functions ← Observed Photometry

$$\boxed{p_{\text{samp}}(z)} = \int d\mathbf{f}\, \boxed{p(z|\mathbf{f})}\, \boxed{p(\mathbf{f})}\,.$$

Density Estimate: Sample Redshift Distribution

(Kernel, KNN, …)

Conditional Density Estimate trained on a calibration dataset

Density Estimate of the observed photometry

LIMITATIONS AND FUTURE WORK

- Treatment of selection functions of the specXphot calibration sample
- Improvements in the quantification of model error in ML (DEMPz, DNNz), selection functions in Template Fitting (Mizuki)
- Limitations in the treatment of cosmic variance induced by redshift calibration using the specXphot calibration sample: Conditioning on color and other quantities of interest
- Quantification of photometric redshift uncertainties and systematics of CAMIRA LRG galaxies
- Astrophysical effects in modeling the cross-correlation data vector: more complex galaxy-dm bias model, magnification bias, etc.
- Improve high redshift coverage with DESI in future analysis