



THE UNIVERSITY OF
CHICAGO

Argonne NATIONAL LABORATORY

WEAK LENSING TOMOGRAPHIC REDSHIFT DISTRIBUTION INFERENCE FOR HSC YEAR 3

MARKUS MICHAEL RAU

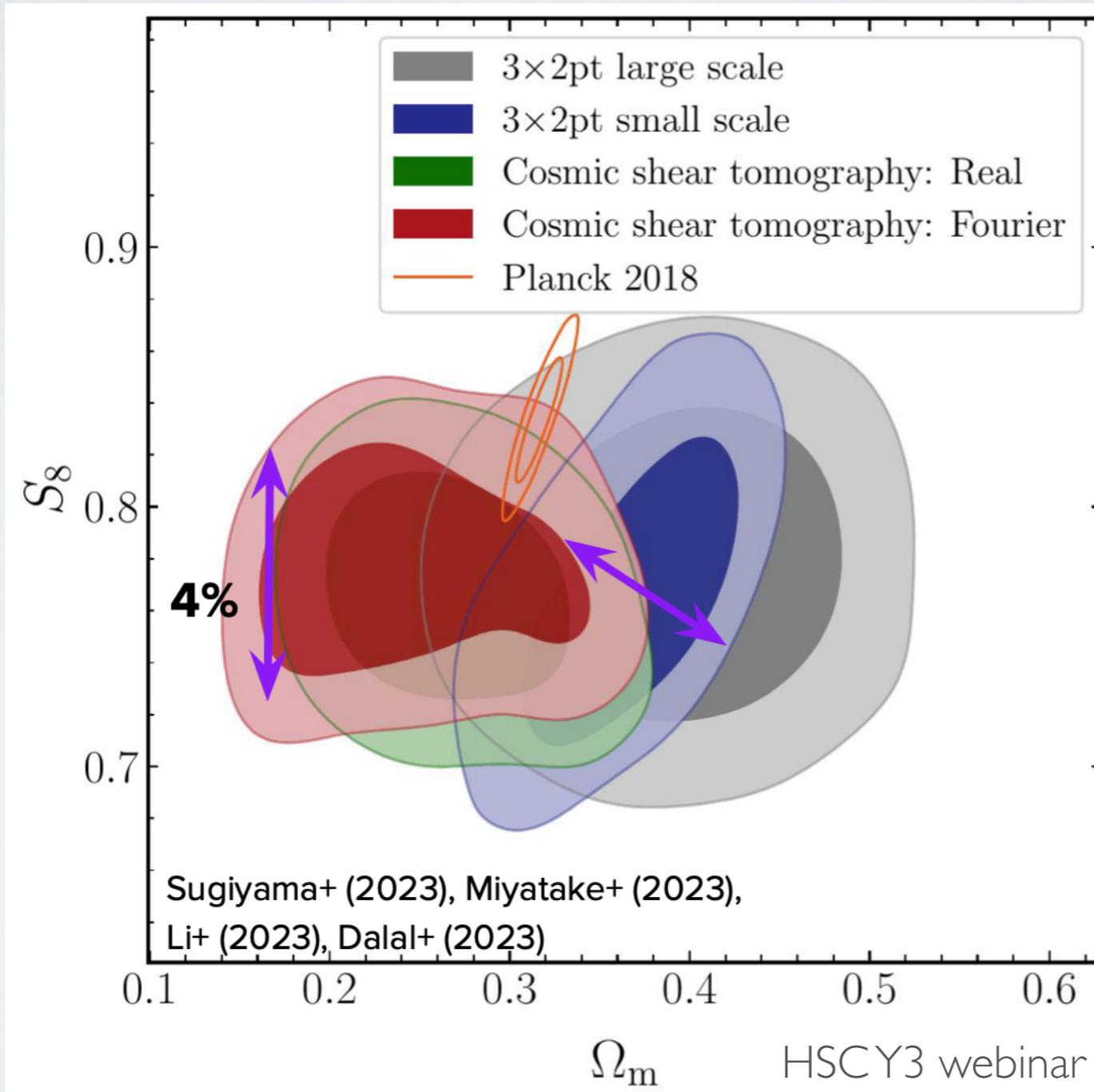


SciDAC
Scientific Discovery
through
Advanced Computing

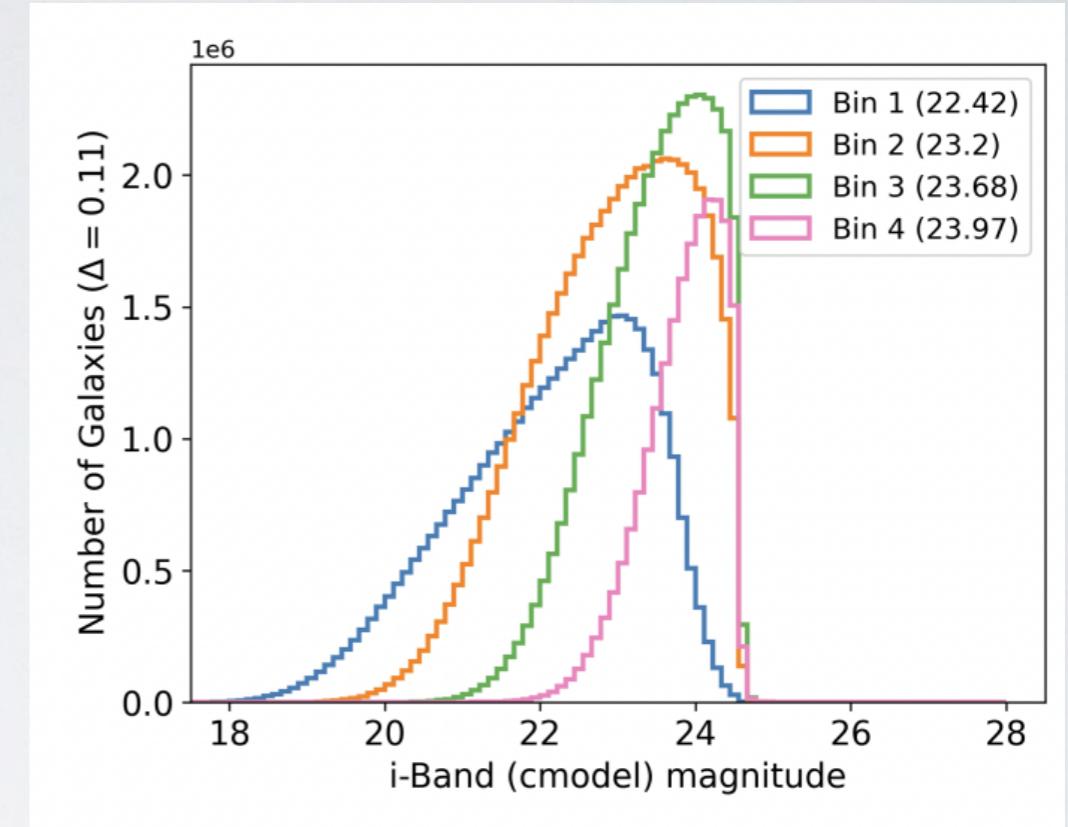
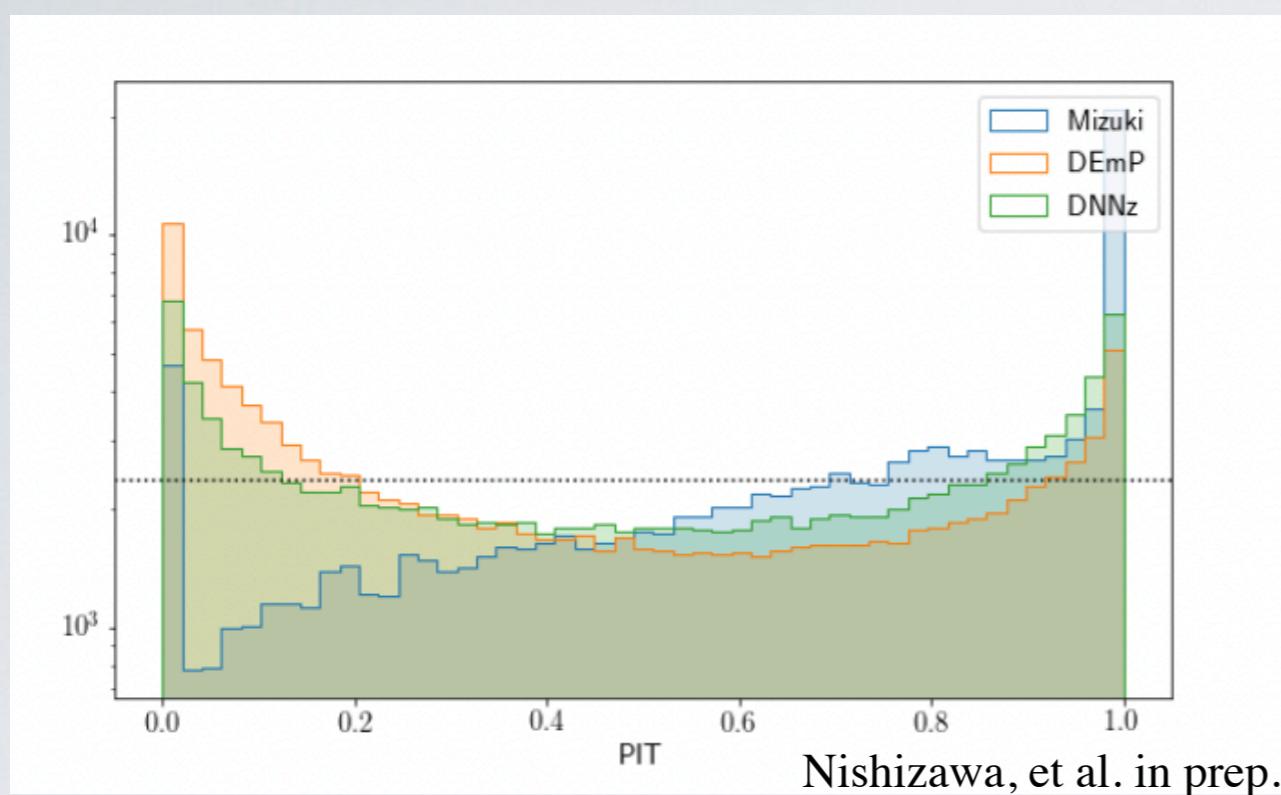
OVERVIEW

- HSC Y3 Tomographic Sample Redshift Distribution Inference
 - Tomographic Sample Redshift Inference Methodology
 - Choice of prior for weak lensing cosmological analysis
- Outlook towards LSST
 - Upcoming Challenges for Redshift Estimation
- Conclusions

HSC YEAR 3 RESULTS

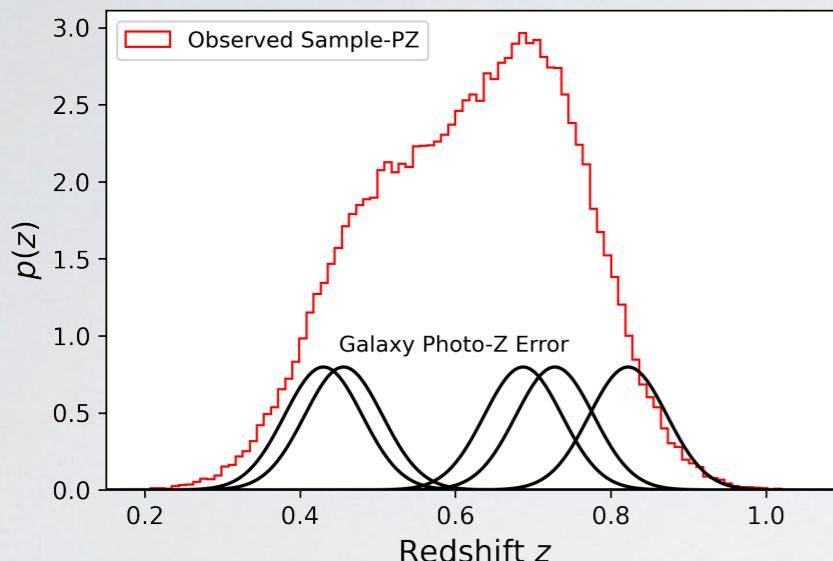


PZ FOR THE HSCY3 SHAPE CATALOG



- HSCY3 shape catalog: 417 sq. deg. Area
- 4 tomographic bins: raw (effective) galaxy number densities: 3.92 (3.77), 5.63 (5.07), 4.68 (4.00), 2.60 (2.12) arcmin $^{-2}$
- 3 photometric redshift codes run on the data: DnnZ (ML), Dempz (ML) and Mizuki (SED fit)
- Scope: Perform Sample redshift inference based on the provided codes

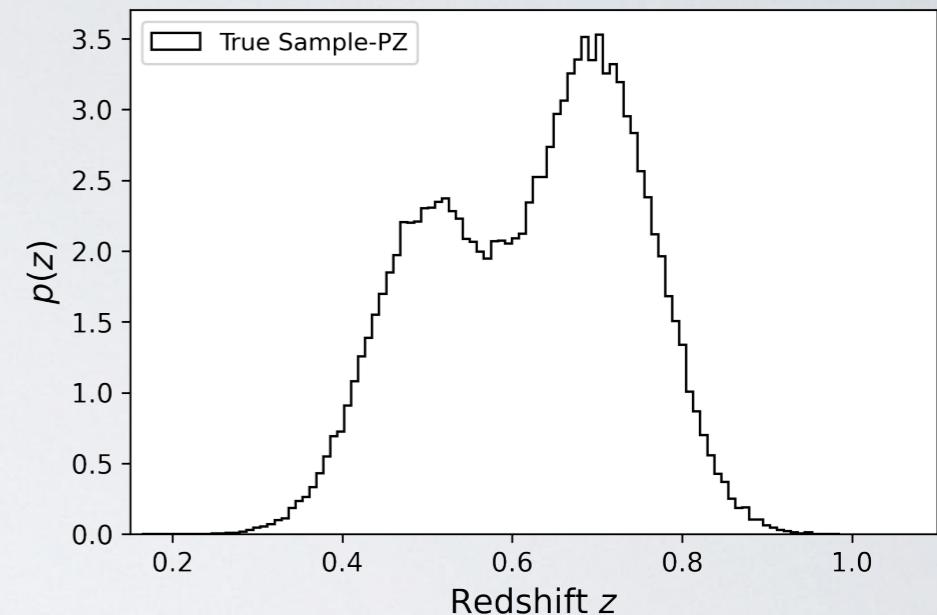
THE PHOTOMETRIC REDSHIFT PROBLEM



Deconvolution

Recover Redshift distributions using:

Spatial Distribution
Photometry



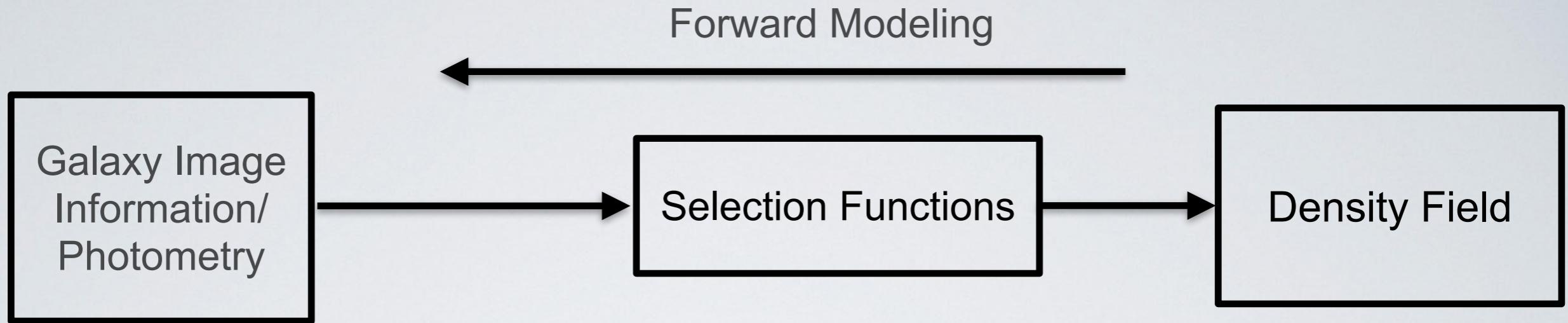
**Map a high dimensional parameter space
that describes galaxy populations to a low
dimensional data vector**

Challenges:

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

HSC PZ analysis Strategy:

- Optimize sample selection to minimize the impact of multiple solutions
- Include using spatial cross-correlations with CAMIRA-LRG
- Perform Sample Redshift inference using multiple individual galaxy redshift pz. Construct a conservative error budget



$$p(\hat{\mathbf{F}} | \nu) = \prod_{i=1}^{N_{\text{gal}}} \int dz_i \omega_i p(\mathbf{f}_i | z_i) p(z_i | \nu)$$

Image info
Of Galaxies
 Selection
Weights
 Likelihood
for each Galaxy
 Sample Redshift
Distribution

Illustration

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

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

Condition number can be large —> Regularization is Key

Assumption: Photometry-Redshift mapping representative.



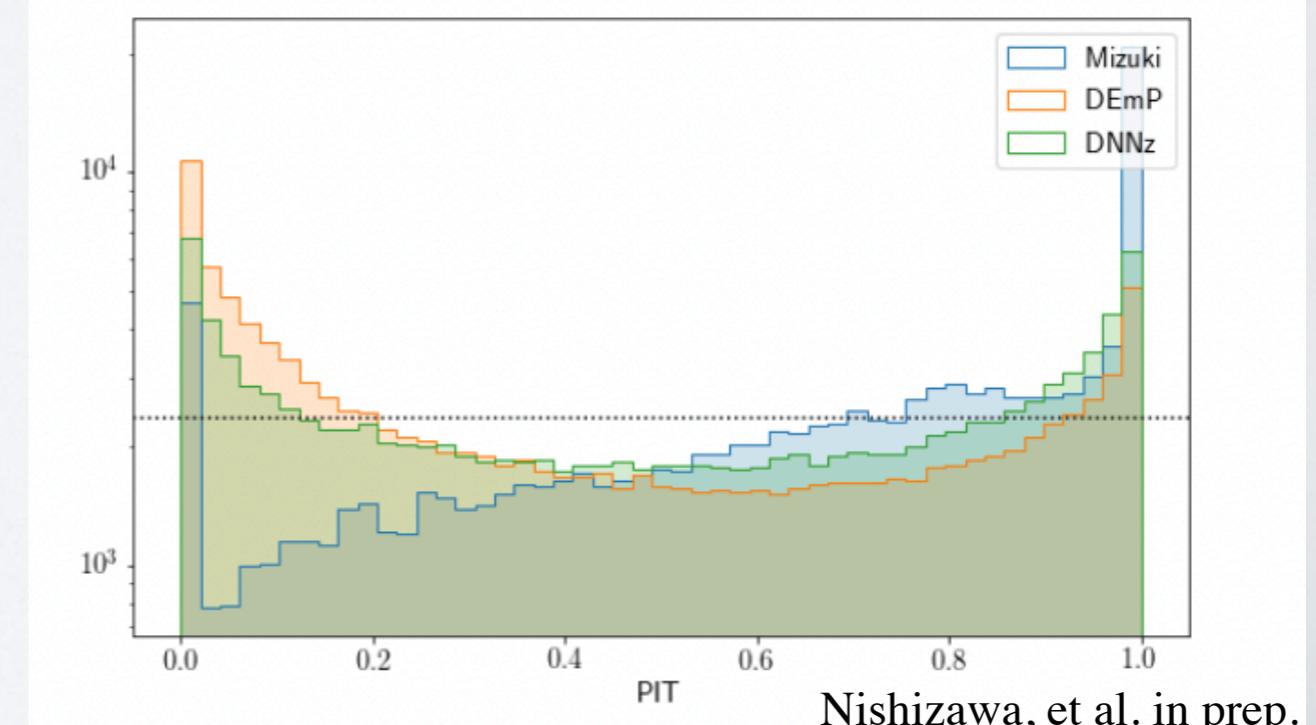
$$p_{\text{samp}}(z) = \int p(z | \mathbf{f}) p(\mathbf{f}) d\mathbf{f}$$

Density Estimate: Sample
Redshift Distribution
(Kernel, KNN, ...)

Conditional Density Estimate
trained on a calibration dataset

Density Estimate of the
observed photometry

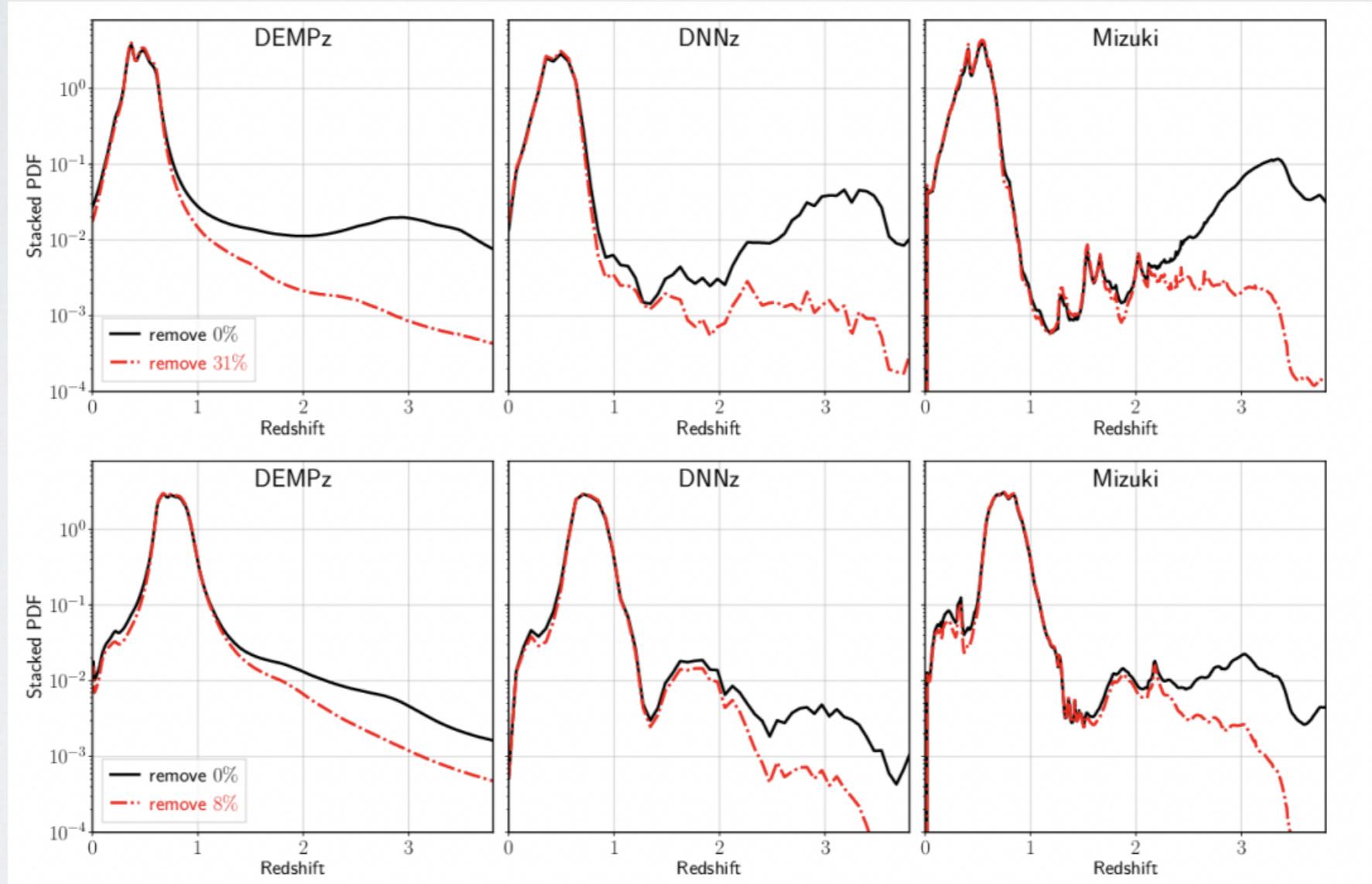
Based on the consistency with WX
and the available calibration data,
we selected the DNNz method
from the available techniques
provided by the HSC PZ working
group.



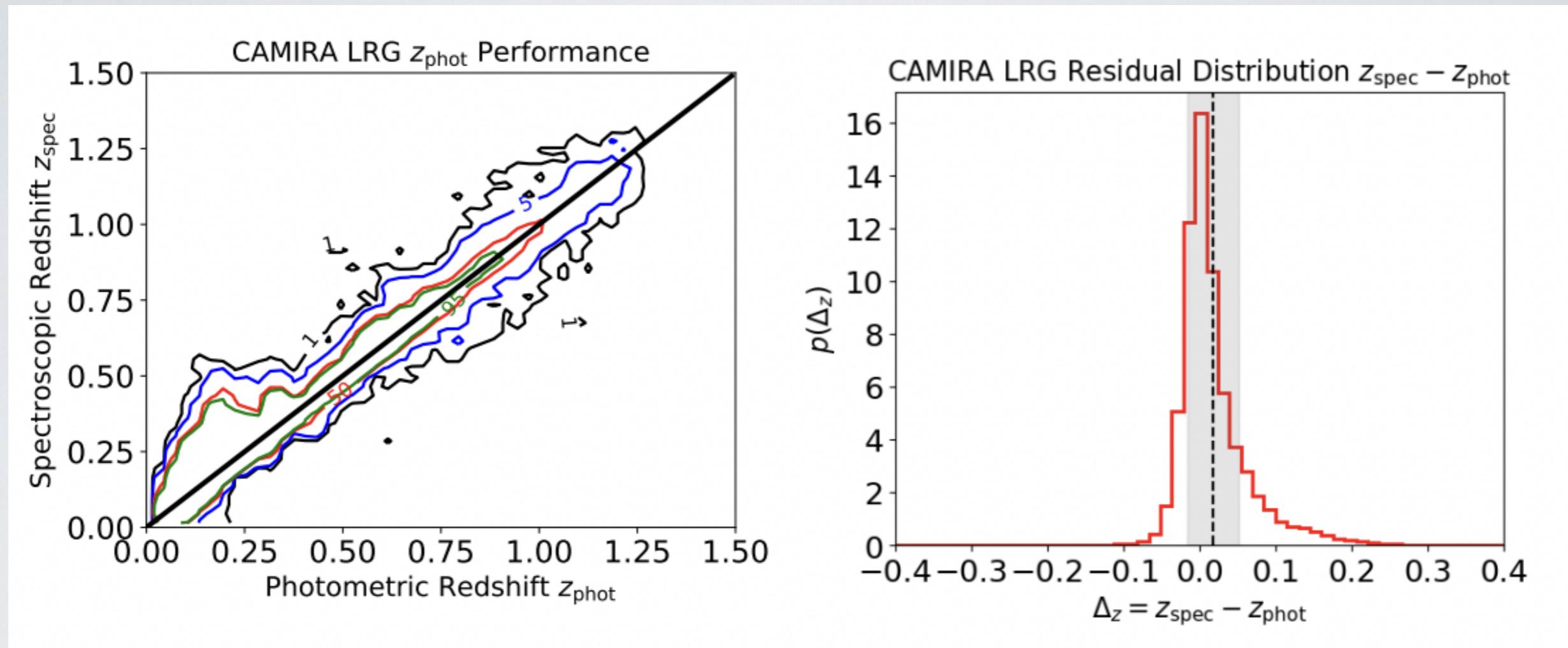
Sample Selection

Criterium:

$$\left(z_{0.975,i}^{\text{Mizuki}} - z_{0.025,i}^{\text{Mizuki}} \right) < 2.7 \quad \text{and} \quad \left(z_{0.975,i}^{\text{DNNz}} - z_{0.025,i}^{\text{DNNz}} \right) < 2.7,$$



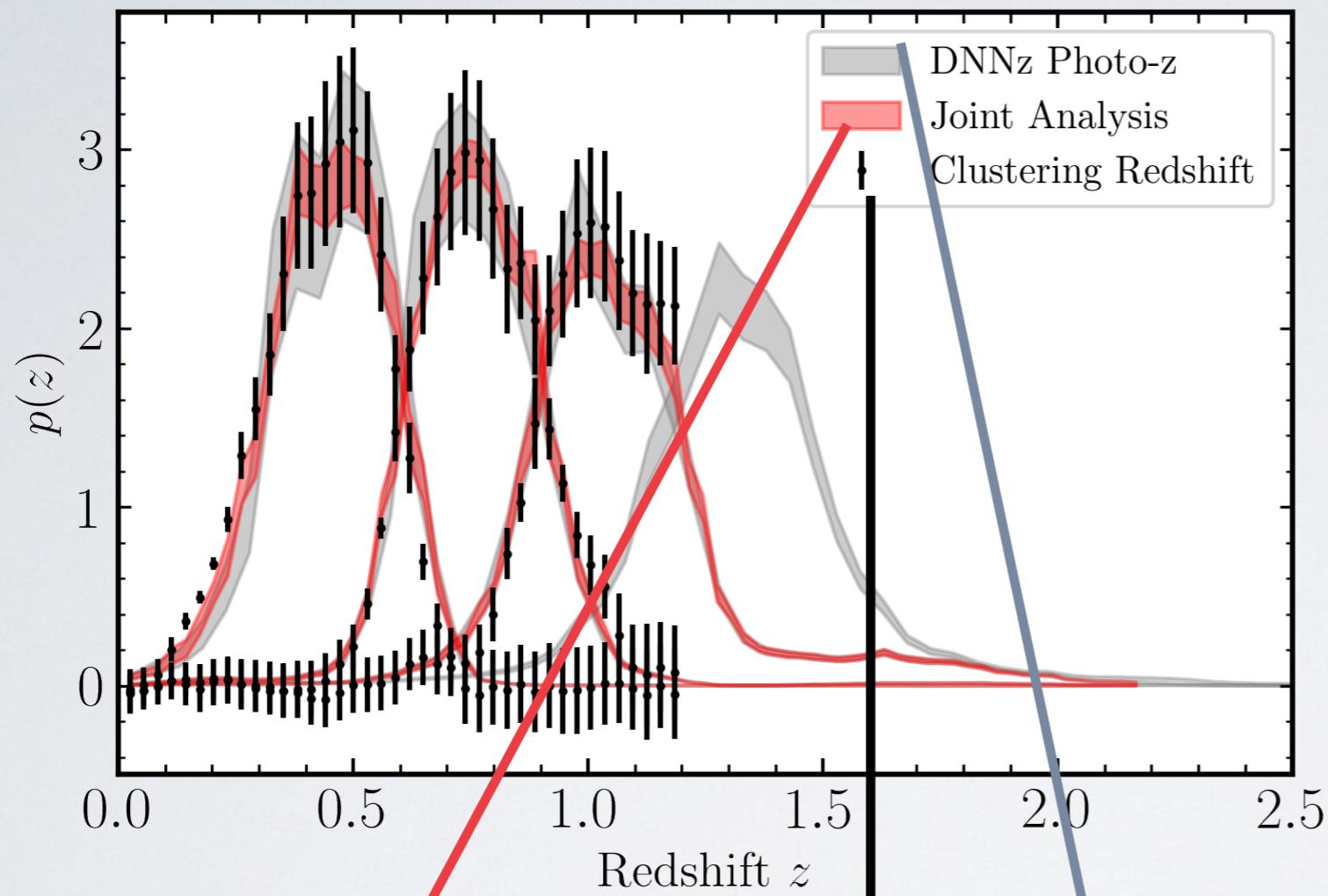
CAMIRA LRG PHOTOMETRIC REDSHIFT ERROR



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
- Marginalize over this redshift error in the analysis

Rau et al. 2023: HSC SSP Year 3 Results



- r:** Pair Counts
- F:** Set of photometry of galaxies
- f_i:** Photometry of single galaxy
- Ω:** Additional Parameters
- p(z):** Redshift probability density function
- ν:** Parameters of basis function model
- φ_i(z):** Basis Functions
- N_{Basis}:** Number of Basis Functions

Formulate a composite likelihood

$$p(\mathbf{r}, \mathbf{F} | \nu, \Omega) = p(\mathbf{r} | \nu, \Omega)p(\mathbf{F} | \nu)$$

$$p(z) = \sum_{i=1}^{N_{\text{Basis}}} \nu_i \phi_i(z)$$

Impose a logistic Gaussian process prior to $p(z)$ or ν
spatial model to quantify spatial variations in density

LIMITATIONS

- 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

Due to these limitations, the HSC Y3 analysis uses a flat prior on the posterior mean for the high redshift tomographic bins.

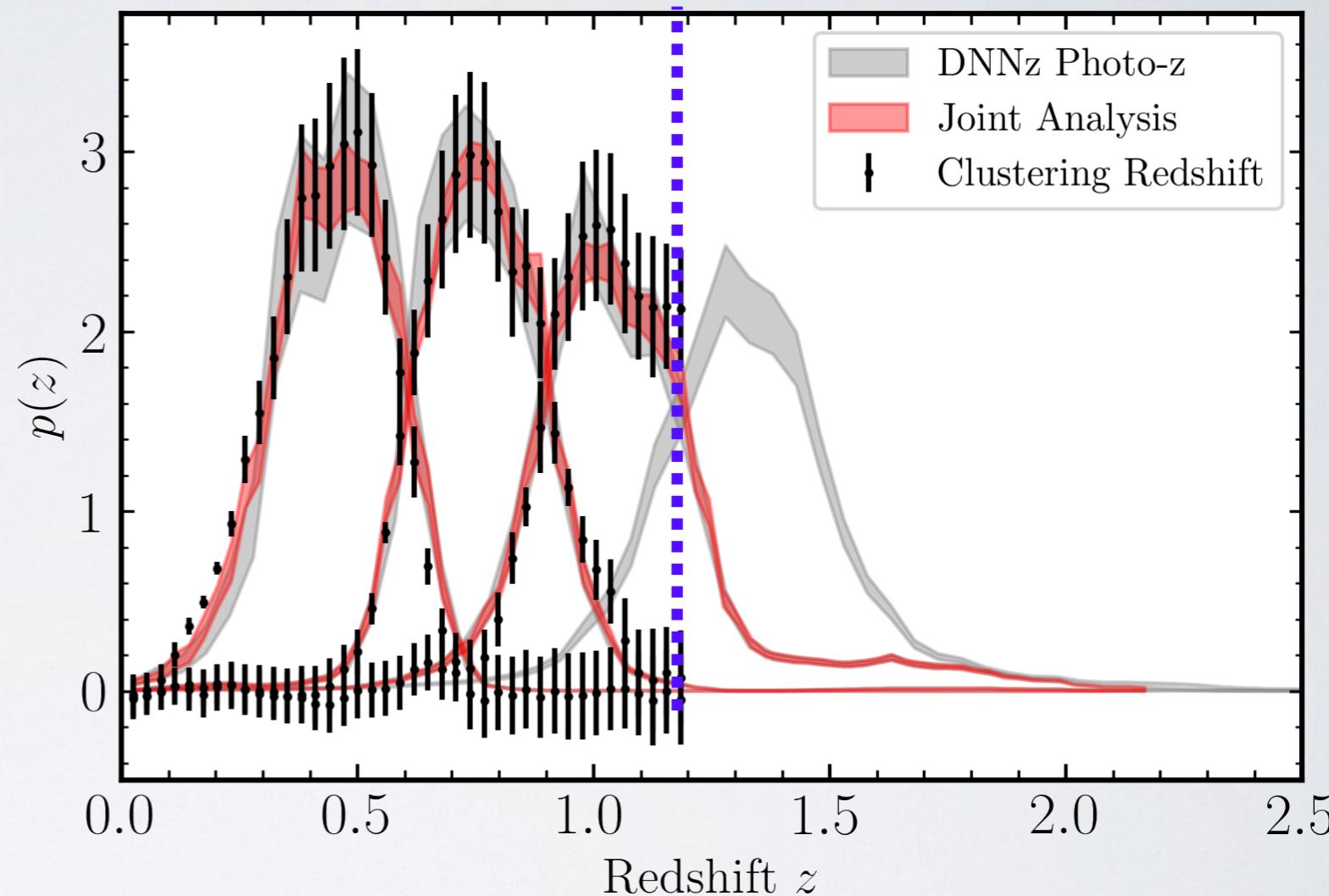
PRIOR CHOICE FOR THE COSMOLOGICAL ANALYSIS

$$\Delta z_1 = 0.452 \pm 0.024$$

$$\Delta z_2 = 0.766 \pm 0.022$$

$$\Delta z_3 = \mathcal{U}(-1, 1)$$

$$\Delta z_4 = \mathcal{U}(-1, 1)$$



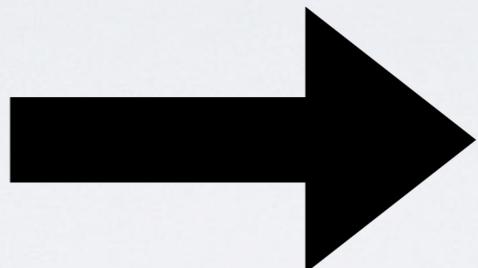
We adopt informative priors based on the presented analysis for the low-redshift bins and a flat prior at high redshift where, our calibration is incomplete

Moving from Stage III (HSC) to Stage IV (LSST)



Credit: hsc.mtk.nao.ac.jp

HSC SSP in Year 3: a “small LSST”
Area $\sim 400 \text{ deg}^2$



LSST:
Area $\sim 20000 \text{ deg}^2$

The **Hyper Supreme-Cam Subaru Strategic Program** (HSC SSP), a great precursor survey to develop LSST relevant Data Science methodology

Big Data Survey Science

Rubin Observatory Operations: Sites & Data Flows

Data and compute sizes:

Final volume of raw image data = 60 PB
Final catalog size (DR11) = 15 PB
Peak compute power in Rubin Observatory data centers
= about 2 PFLOPS

Network bandwidths:

Summit (Cerro Pachón) - Base (La Serena)
= 600 Gbps
Base (La Serena) to Archive (SLAC)
= 2 x 100 Gbps

Alert Production:

Real-time alert latency = 60 seconds
Estimated number of alerts per night
= up to about 10 million

Data Releases:

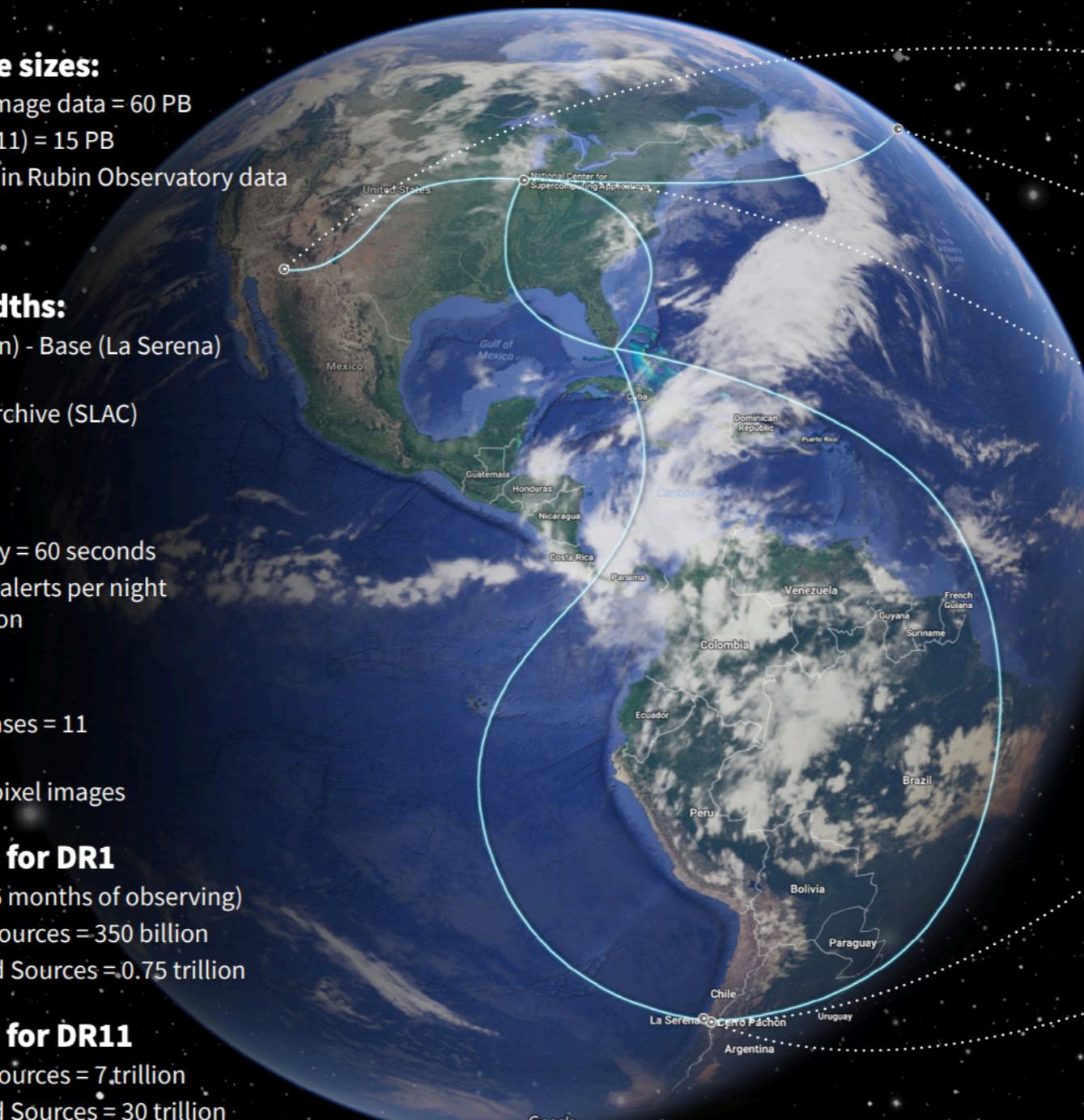
Number of Data Releases = 11
Images collected
= 5.5 million 3:2 Gigapixel images

Estimated counts for DR1

(produced from first 6 months of observing)
Objects = 18 billion; Sources = 350 billion
(single epoch); Forced Sources = 0.75 trillion

Estimated counts for DR11

Objects = 37 billion; Sources = 7 trillion
(single epoch); Forced Sources = 30 trillion



HQ Site

Tucson, AZ

Science Operations
Observatory Management
Education & Public Outreach

French Site

CC-IN2P3, Lyon, France

French Data Facility
Data Release Production
Long-term Storage (copy 3)

Rubin Observatory Data Facility

SLAC National Accelerator Laboratory (SLAC), Menlo Park, CA

Processing Center
Alert Production
Data Release Production
Calibration Products Production
EPO Infrastructure
Long-term Storage (copy 2)
Data Access Center
Data Access and User Services

Summit Site

Cerro Pachón, Chile

Telescope & Camera
Data Acquisition
Crosstalk Correction

Base Site

La Serena, Chile

Base Center
Long-term storage (copy 1)
Data Access Center
Data Access & User Services

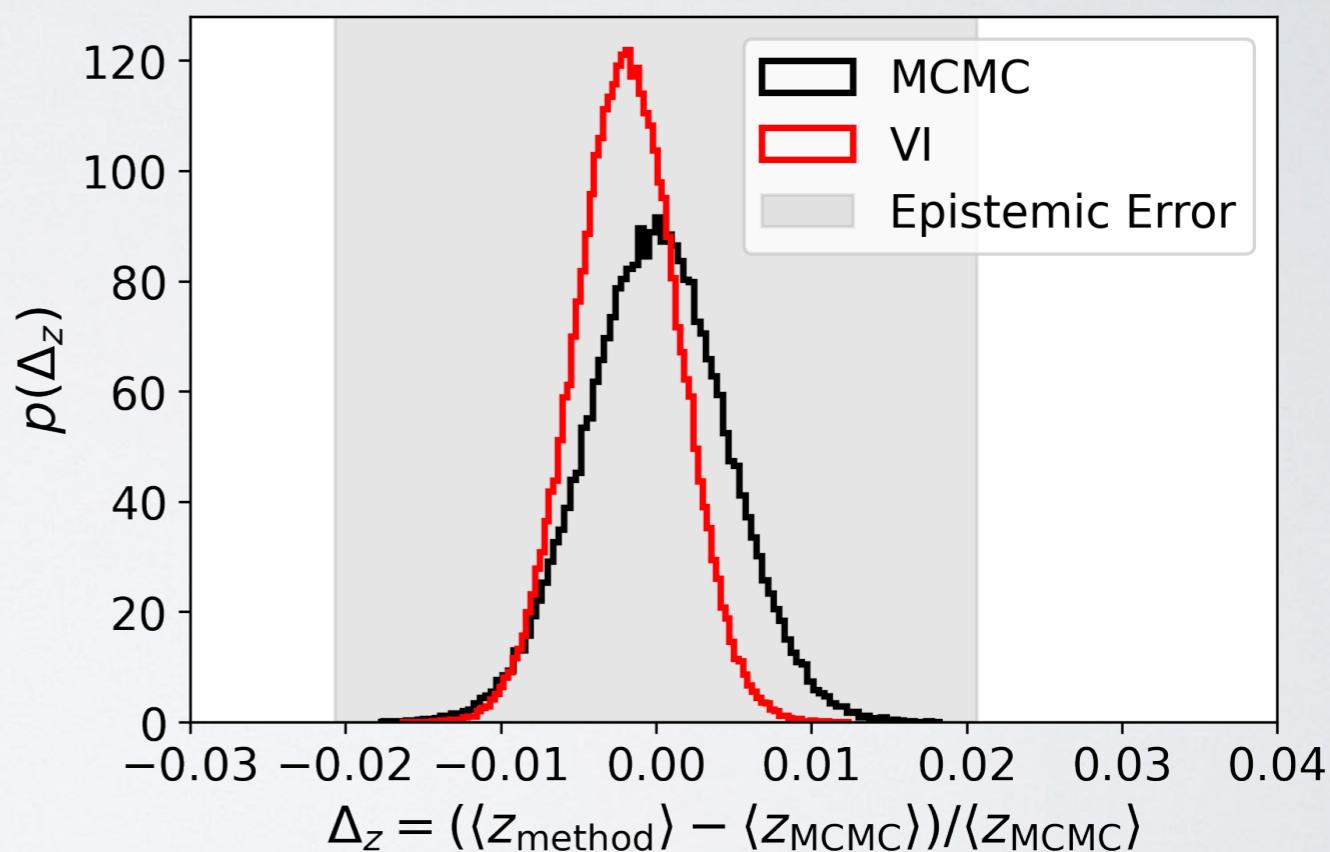
Credit: lsst.org

SCALABILITY

Large Epistemic error
in astrophysical modeling

Approximate Inference schemes
In combination with appropriate
Regularization

→ Dramatic Speedup with
acceptable additional misestimation
Error budget

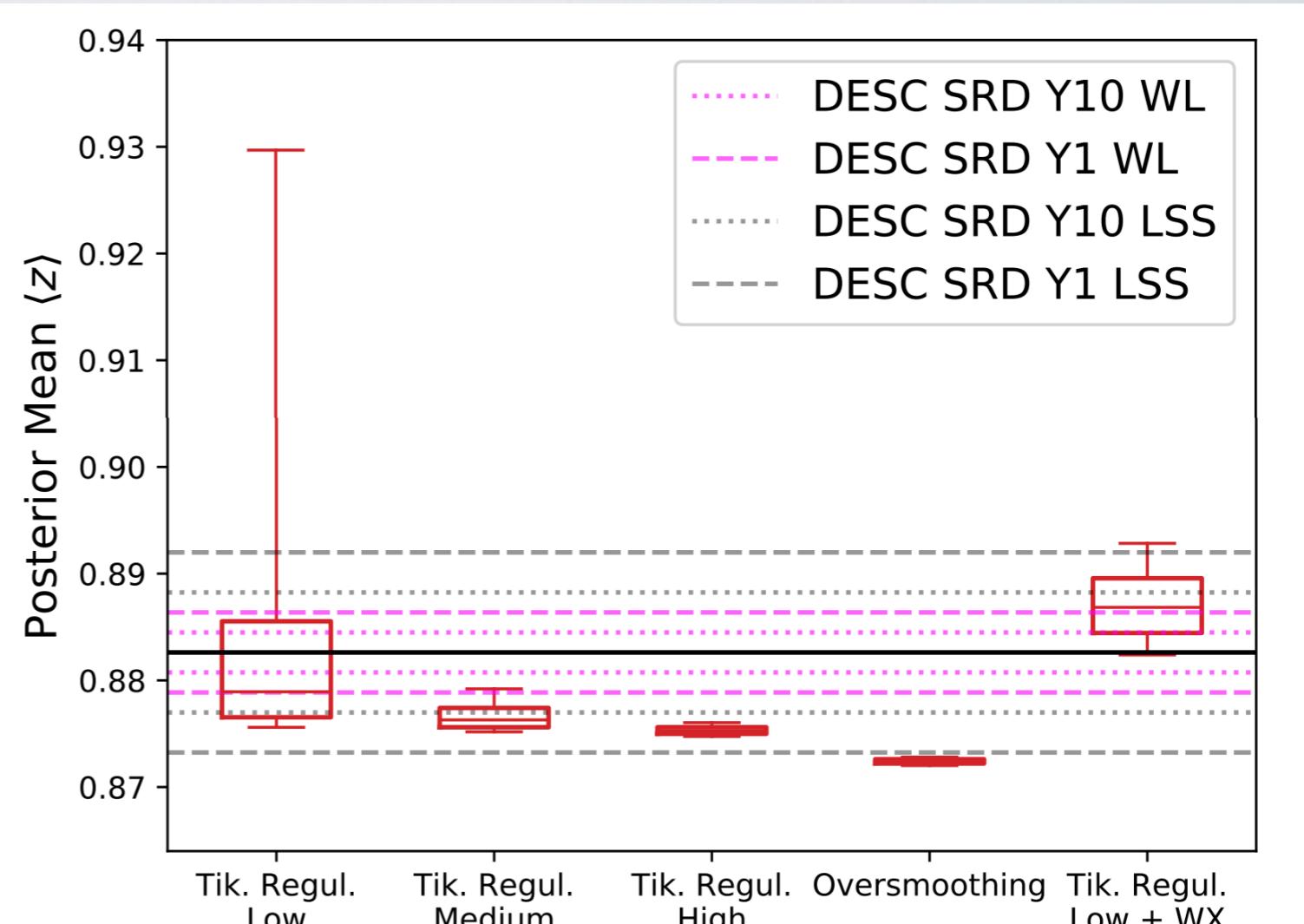


See Rau et al. 2023

REGULARIZATION



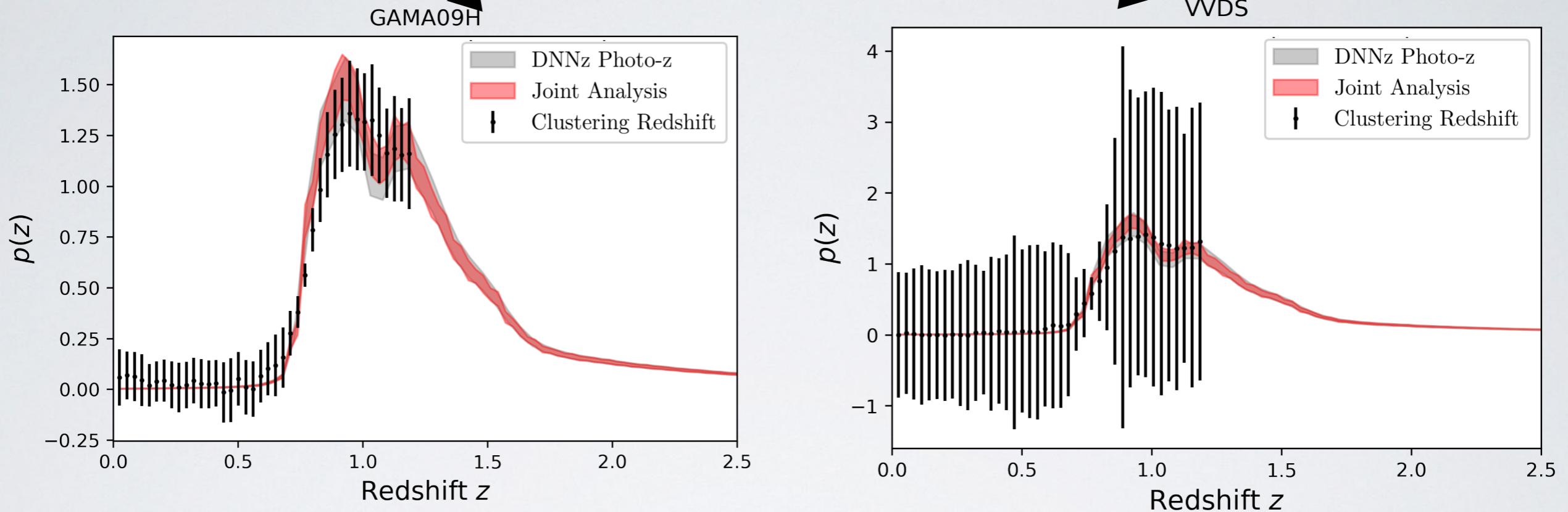
**Bins are merged
to reduce variance**



Rau et al. 2022

Oversmoothing and Tikhonov's regularization induce under-coverage.
Solution: inference in high resolution and merge bins to reduce variance
Adaptation of Burrus, Rust 1972, Kuusela 2016

Spatial variation in number density in different regions on the sky



Rau, et al. in prep.

The inference scheme developed by Rau et al. 2020, 2023 fits a log-gaussian cox process to the galaxy density and allows quantifying field-to-field variations in data. Applied here to the HSC Year 3 cluster-lensing analysis.

CONCLUSIONS

- Survey Science rapidly advances into the era of petascale cosmology and astronomy.
- At the same time the modeling of systematics becomes more complex
- Photometric Redshift estimation is challenging due to the lack of calibration sources that are robust against model misspecification
- Inclusion of complementary data vital to reach high fidelity of LSST
- The next 2-5 years will be a critical time window for Data Science development in this area