

Machine Learning methods in Cosmology

ML4ASTRO3 Plenary
Uni. of Malta, Malta

Angus H Wright, 03.09.2026

RUHR
UNIVERSITÄT
BOCHUM

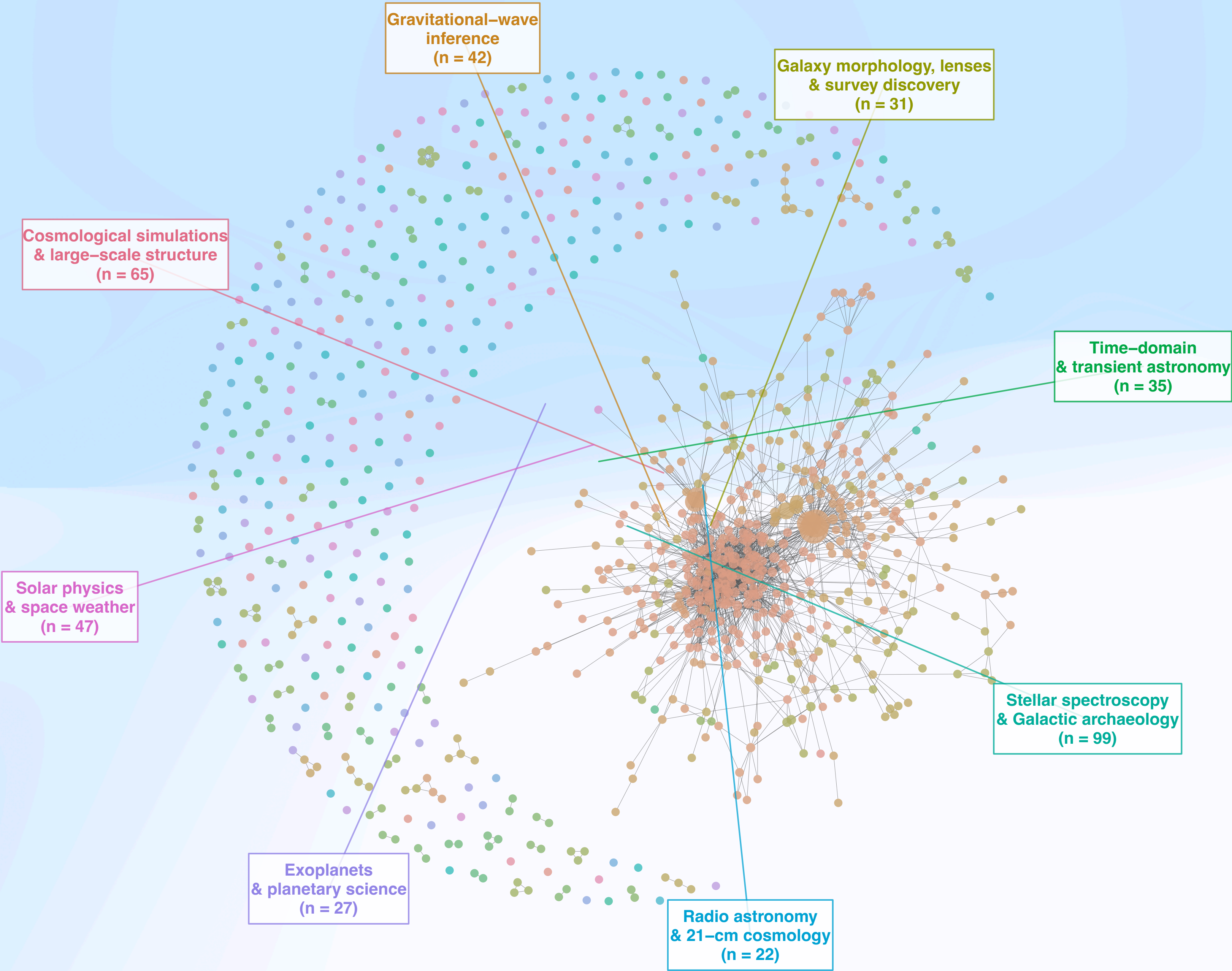
RUB

GCCL
GERMAN CENTRE FOR COSMOLOGICAL LENSING

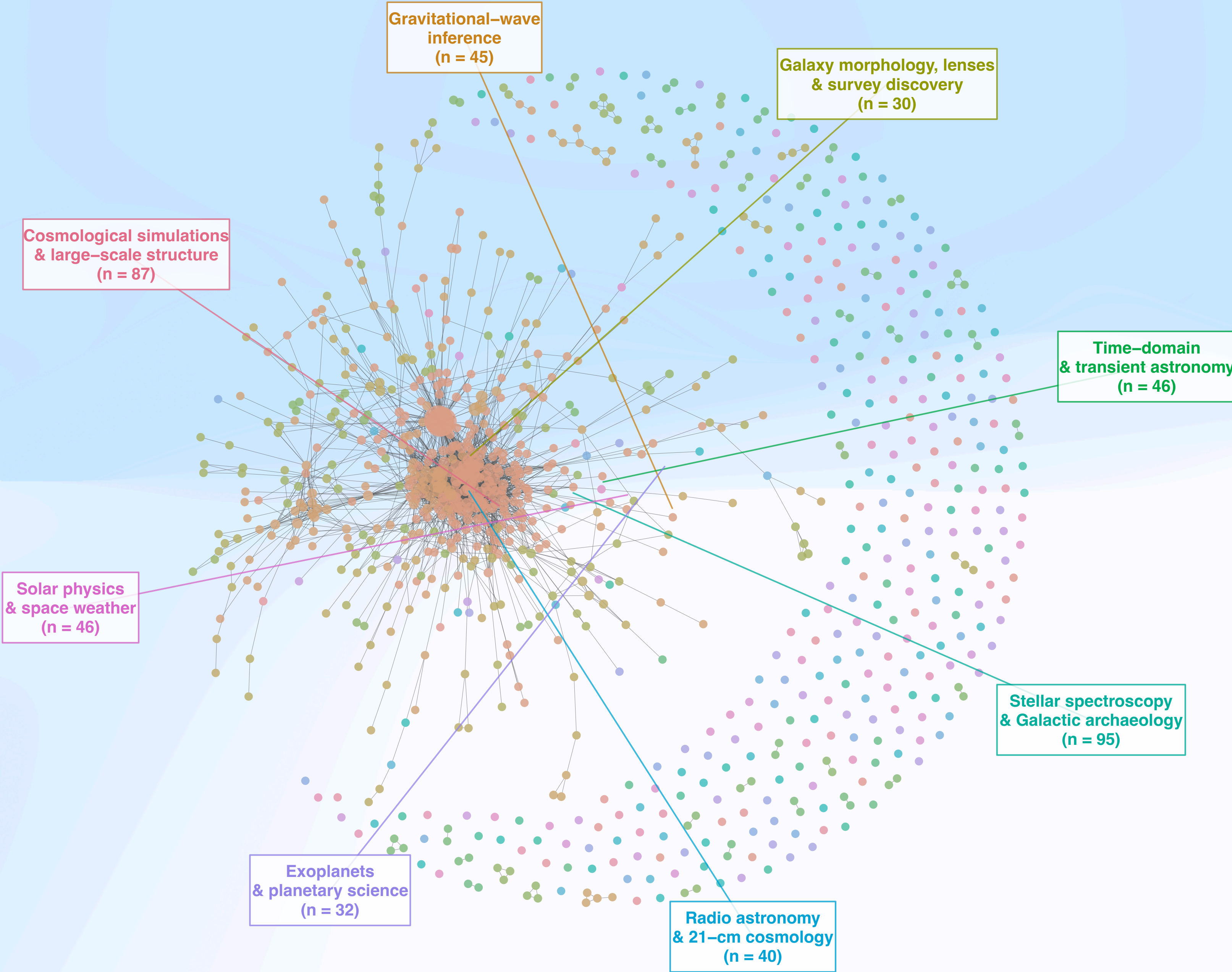


euclid
consortium

Publication Author Network (2024)



Publication Author Network (2025)



For which tasks are ML methods in cosmology applied?

ML adoption in Cosmology

What are the dominant tasks to which the community applies ML?

$$\hat{z} = f_{\theta}(\mathbf{X})$$

***Physical Parameter
Estimation***

- Photometric redshift estimation, stellar mass estimation, etc.
- Necessary for sub-sample definitions (e.g. tomography)
- Machine learning is motivated over conventional methods (e.g. template fitting) predominantly due to speed

ML adoption in Cosmology

What are the dominant tasks to which the community applies ML?

$$\hat{z} = f_{\theta}(\mathbf{x})$$

*Physical Parameter
Estimation*

$$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$$

***Classification &
Systematics Correction***

- Crucial for redshift distribution, $N(z)$, estimation.
- Machine learning is ideal due to unsupervised classification

ML adoption in Cosmology

What are the dominant tasks to which the community applies ML?

$$\hat{z} = f_{\theta}(\mathbf{x})$$

*Physical Parameter
Estimation*

$$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$$

*Classification &
Systematics Correction*

$$f_{\theta}(\phi) \approx \mathcal{M}(\phi)$$

Model Emulation

- Used primarily in sampling, to increase runtime efficiency
- Machine learning grants speed without sacrificing accuracy

ML adoption in Cosmology

What are the dominant tasks to which the community applies ML?

$$\hat{z} = f_{\theta}(\mathbf{x})$$

*Physical Parameter
Estimation*

$$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$$

*Classification &
Systematics Correction*

$$f_{\theta}(\phi) \approx \mathcal{M}(\phi)$$

Model Emulation

$$p(\phi | \mathbf{x}) \approx q_{\hat{\theta}}(\phi | \mathbf{x})$$

***Simulation Based
Inference***

- Used when objective likelihood formulation is intractable
- Machine learning grants speed without sacrificing accuracy

ML adoption in Cosmology

What are the dominant tasks to which the community applies ML?

$$\hat{z} = f_{\theta}(\mathbf{x})$$

***Physical Parameter
Estimation***

$$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$$

***Classification &
Systematics Correction***

$$f_{\theta}(\phi) \approx \mathcal{M}(\phi)$$

Model Emulation

$$p(\phi | \mathbf{x}) \approx q_{\hat{\theta}}(\phi | \mathbf{x})$$

***Simulation Based
Inference***

ML adoption in Cosmology

What are the dominant tasks to which the community applies ML?

**Photo-z &
Tomographic Binning**

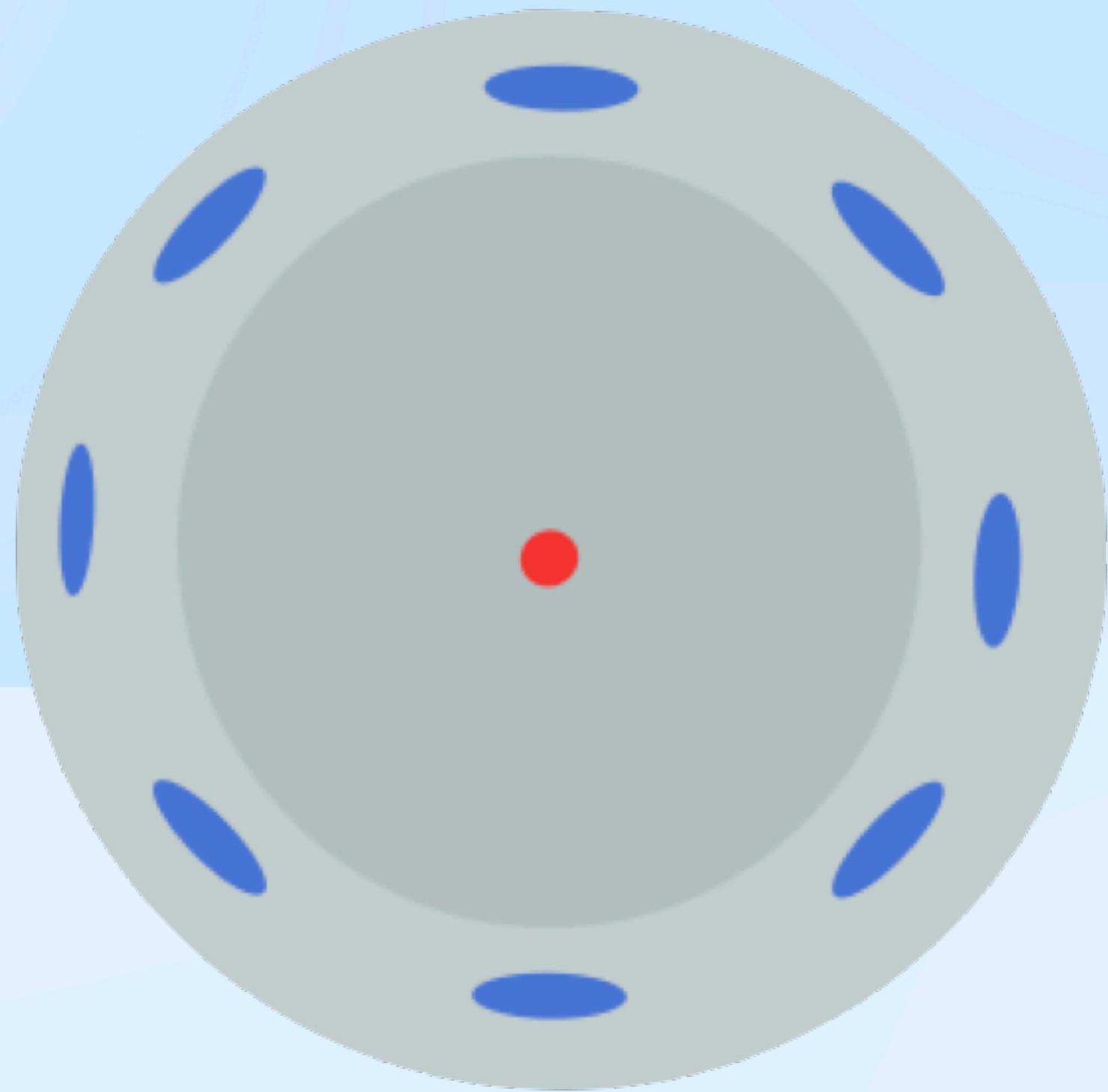
$N(z)$ estimation

**Evaluation of the
matter power
spectrum**

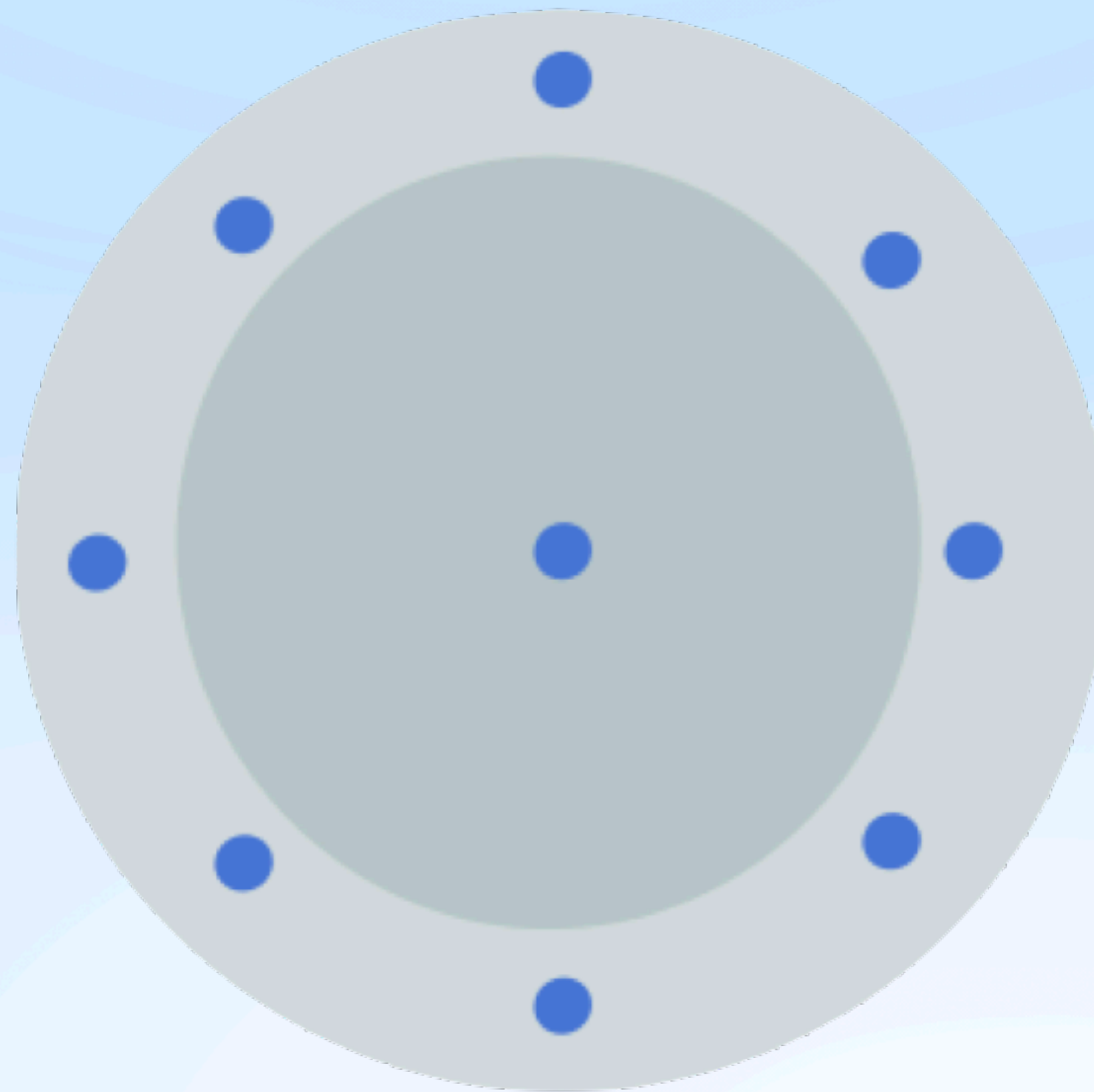
**Map-based
Inference**

A brief introduction to Cosmological analyses with images

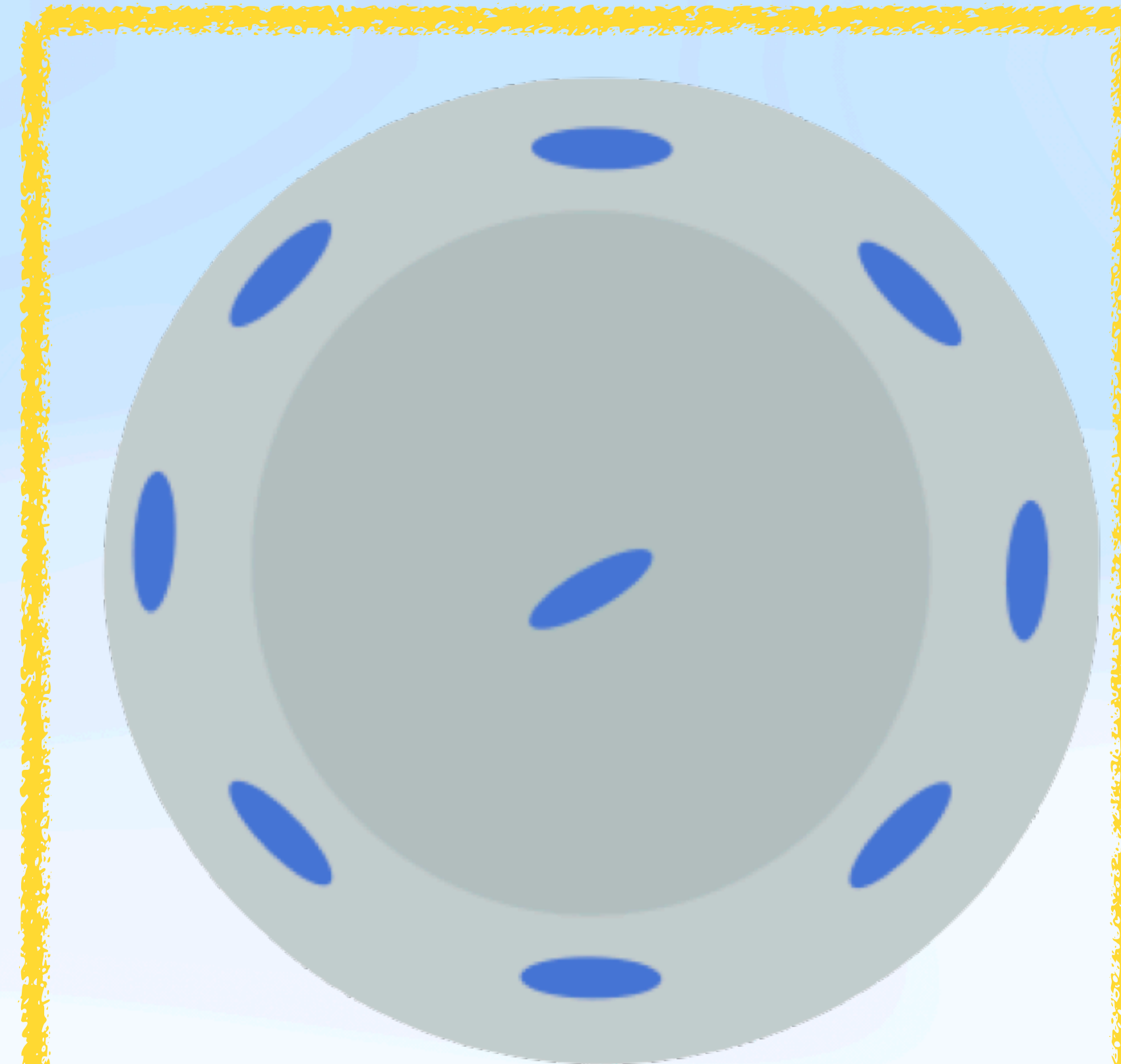
The “3x2pt” probes



Shear-position
correlation function:
galaxy-galaxy lensing



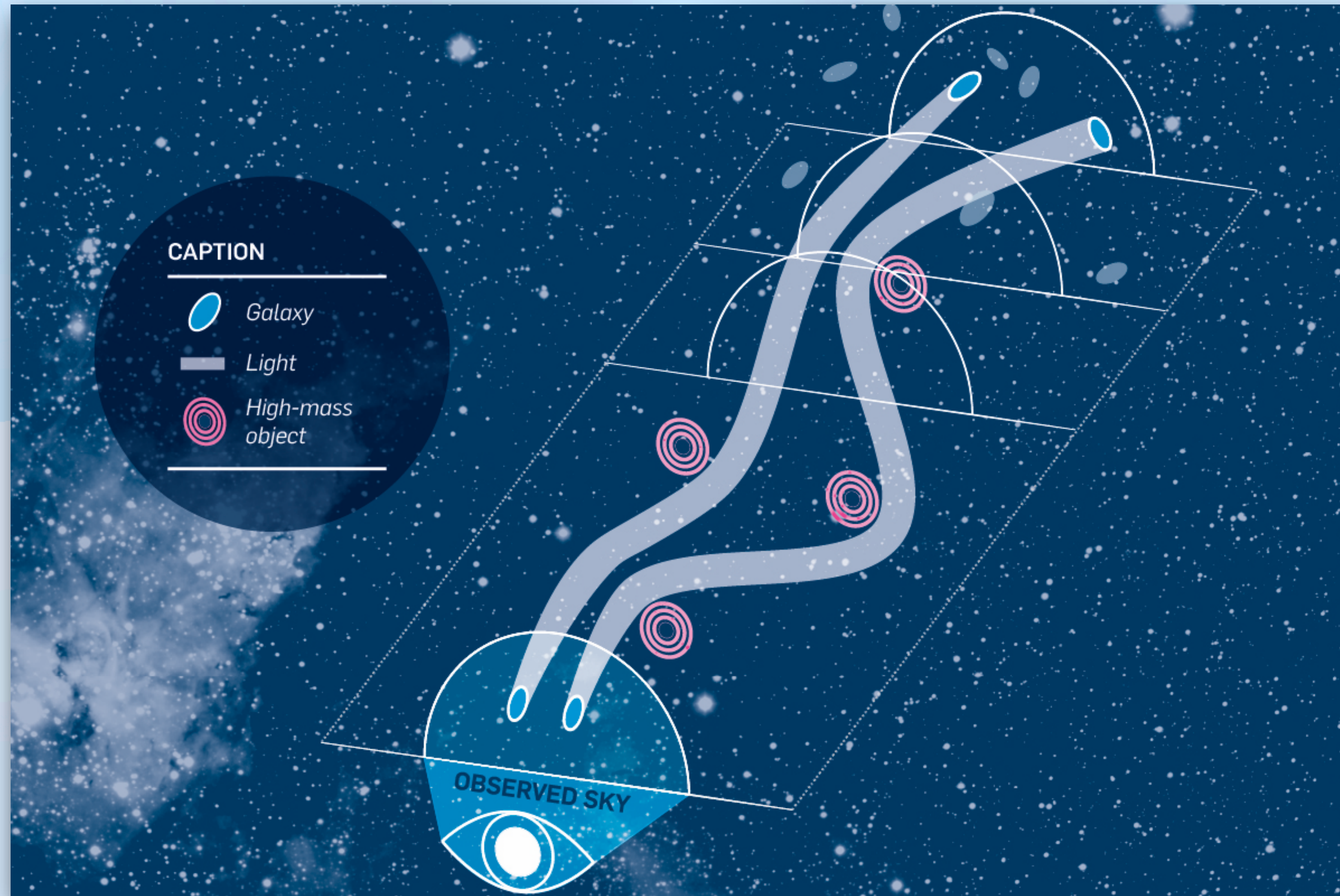
Position-position
correlation function:
photometric clustering



Shear-Shear
correlation function:
cosmic shear

Fundamentals of Cosmic Shear

Lensing by large-scale structures



Galaxies in the distant universe have (mostly) randomly distributed shapes

Light is distorted along the line-of-sight by massive structures

Propagation through similar structures imprints coherent distortions on galaxy shapes

Fundamentals of Cosmic Shear

The Shear-Shear Correlation Function

Observed & Theory

- Observed shear-shear correlation function relates observed tangential- and cross-shears
- Shear correlations directly relate to the matter power-spectrum
- Requires knowledge of the source redshift distribution

$$\xi_{\pm}(\theta) = \langle \gamma_t \gamma_t \rangle(\theta) \pm \langle \gamma_{\times} \gamma_{\times} \rangle(\theta)$$

$$\xi_{+}(\theta) = \int_0^{\infty} \frac{d\ell \ell}{2\pi} J_0(\ell\theta) P_{\kappa}(\ell)$$

$$\xi_{-}(\theta) = \int_0^{\infty} \frac{d\ell \ell}{2\pi} J_4(\ell\theta) P_{\kappa}(\ell)$$

$$P_{\kappa}(\ell) = \frac{9H_0^4 \Omega_m^2}{4c^4} \int_0^{\chi_h} d\chi \frac{g^2(\chi)}{a^2(\chi)} P_{\delta} \left(\frac{\ell}{f_K(\chi)}, \chi \right)$$

$$g(\chi) = \int_{\chi}^{\chi_h} d\chi' p_{\chi}(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')}$$

Fundamentals of Cosmic Shear

The Shear-Shear Correlation Function

Observed & Theory

- Observed shear-shear correlation function relates observed tangential- and cross-shears
- Shear correlations directly relate to the matter power-spectrum
- Requires knowledge of the source redshift distribution

$$\xi_{\pm}(\theta) = \langle \gamma_{\pm} \times \gamma_{\pm} \rangle(\theta)$$

Photo-z & Tomographic Binning

$$\xi_{+}(\theta) = \int_0^{\infty} \frac{d\ell \ell}{2\pi} J_0(\ell\theta) P_{\kappa}(\ell)$$

$$\xi_{-}(\theta) = \int_0^{\infty} \frac{d\ell \ell}{2\pi} J_4(\ell\theta) P_{\kappa}(\ell)$$

$$P_{\kappa}(\ell) = \frac{9H_0^4 \Omega_m^2}{4c^4} \int_0^{\chi_h} d\chi \frac{g^2(\chi)}{a^2(\chi)}$$

Evaluation of the matter power spectrum

$$g(\chi) = \frac{N(z) \kappa(\chi' - \chi)}{f_K(\chi')}$$

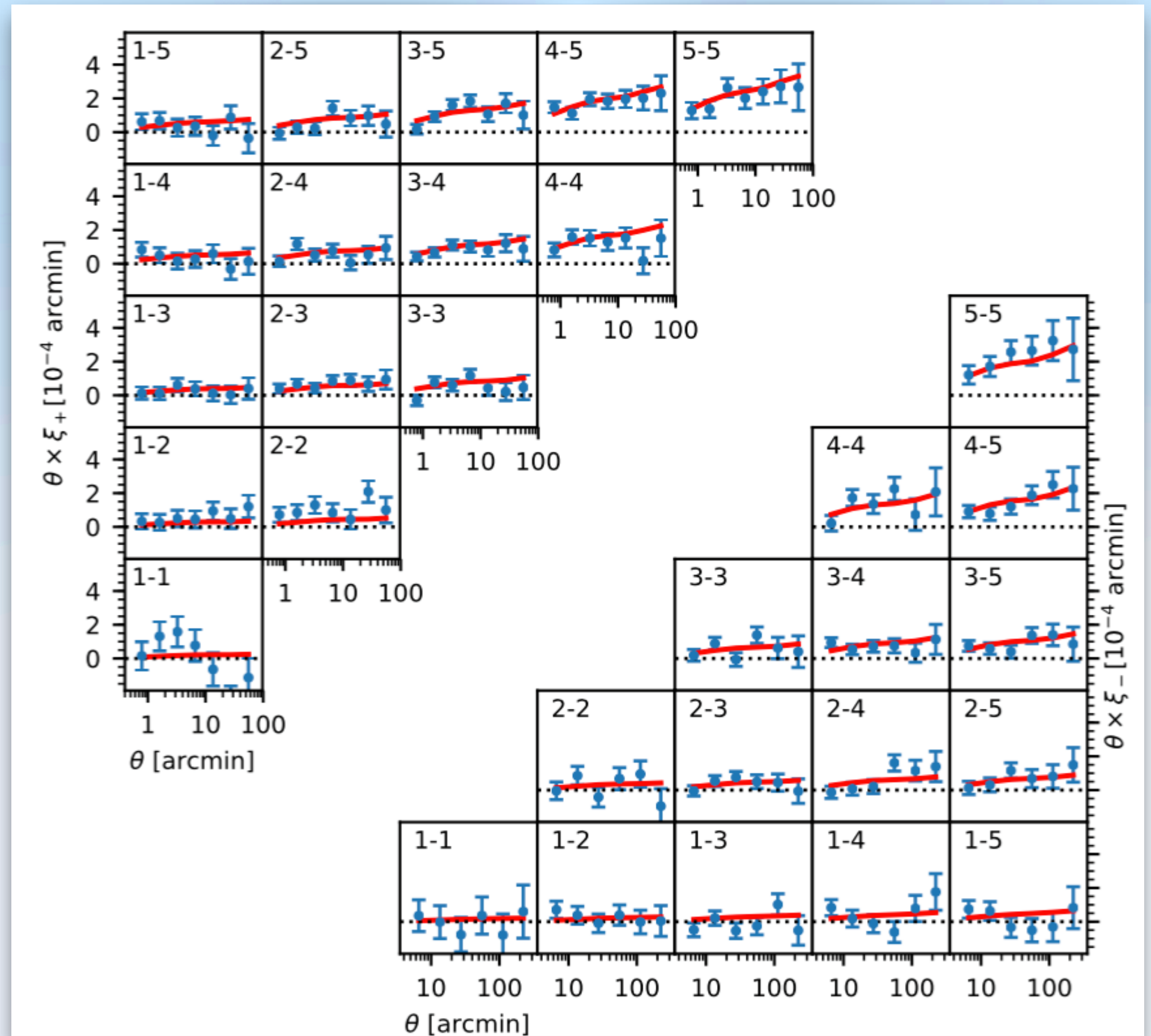
$N(z)$ estimation

Fundamentals of Cosmic Shear

The Shear-Shear Correlation Function

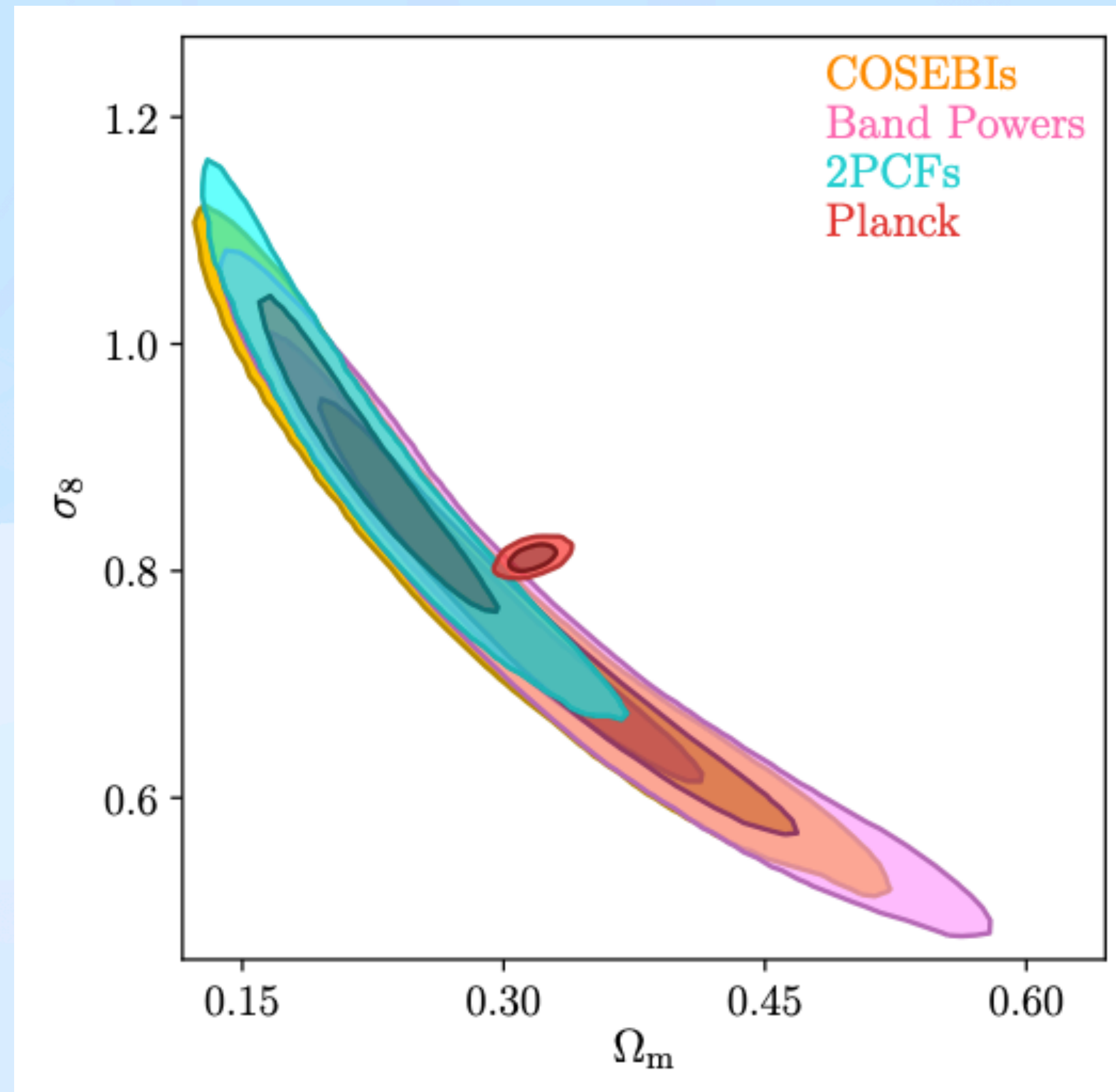
Observed & Theory

- Observed shear-shear correlation function relates observed tangential- and cross-shears
- Shear correlations directly relate to the matter power-spectrum
- Requires knowledge of the source redshift distribution



Fundamentals of Cosmic Shear

Measuring Cosmological Parameters



Degeneracy between amount of matter and strength of clustering

So we discuss cosmic shear in terms of the degeneracy:

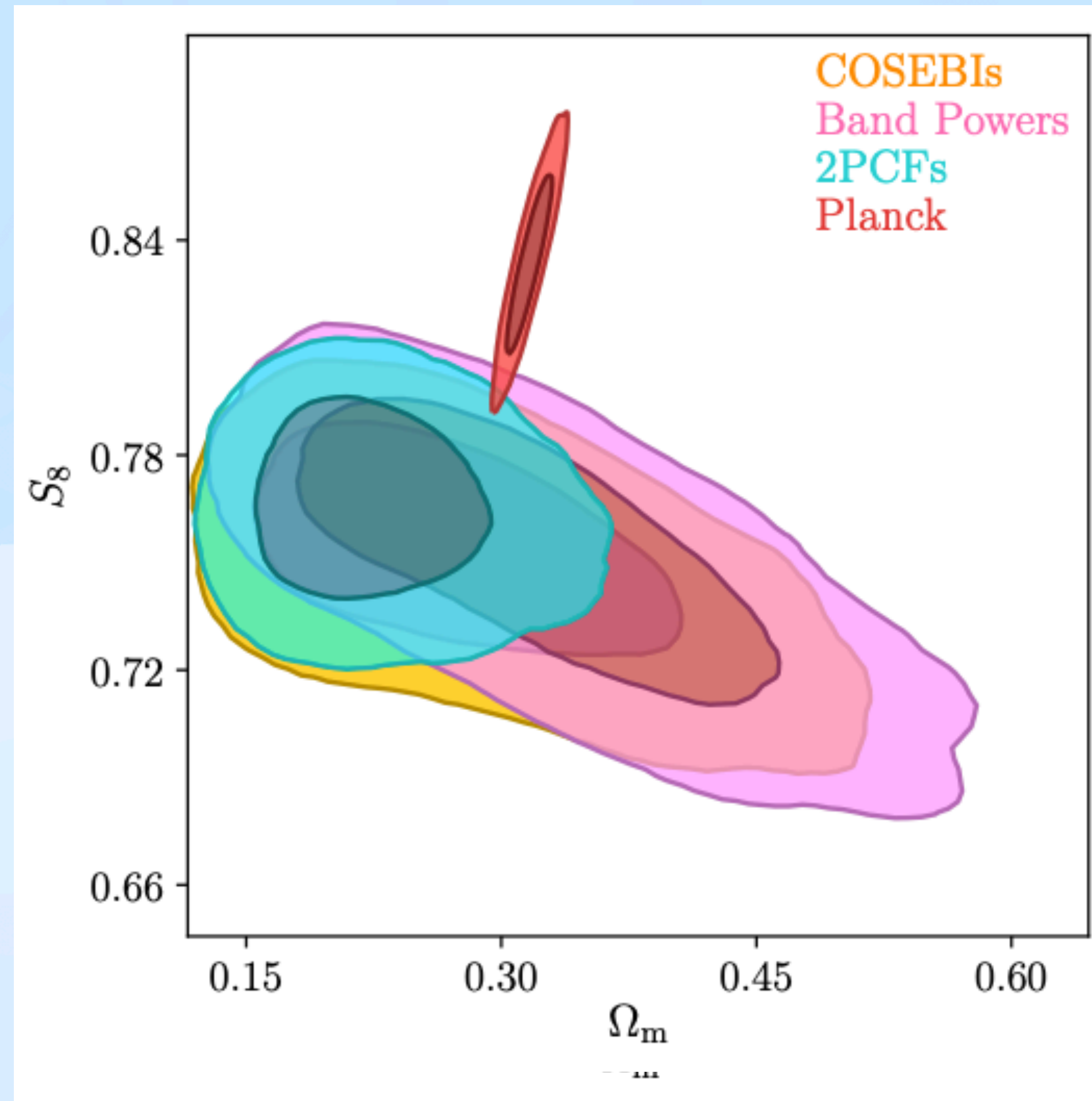
$$S8 = \sigma_8 \sqrt{\Omega_m / 0.3}$$

or more generally:

$$\Sigma_8 = \sigma_8 (\Omega_m / 0.3)^\alpha$$

Fundamentals of Cosmic Shear

Measuring Cosmological Parameters



Degeneracy between amount of matter and strength of clustering

So we discuss cosmic shear in terms of the degeneracy:

$$S_8 = \sigma_8 \sqrt{\Omega_m / 0.3}$$

or more generally:

$$\Sigma_8 = \sigma_8 (\Omega_m / 0.3)^\alpha$$

For which tasks are ML methods in cosmology applied?

$$\hat{z} = f_{\theta}(\mathbf{x})$$

Physical Parameter Estimation

Used in simulation construction, sample definition, model inference, and down stream analyses.

Particularly important: photometric redshifts (photo- z).

Recent & notable examples:

- Bilicki et al. (2021)
- Jalen et al. (2025)

Example talks this week:

- Sai Srinivas (Mon 1630)
- Nicola Napolitano (Wed 0900)
- Benjamin Csizi (Thurs 1400)

Photo-z estimation in wide-field surveys

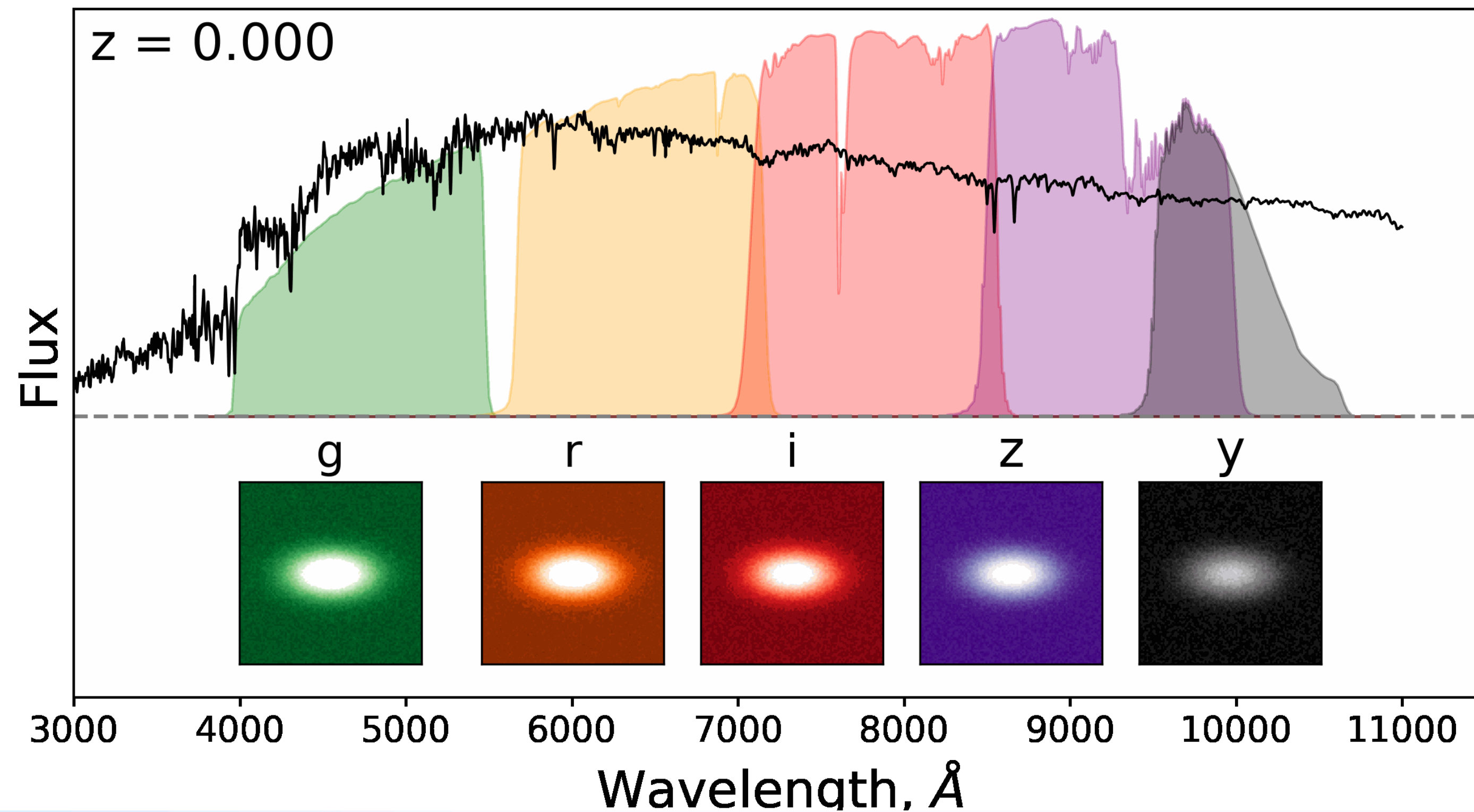
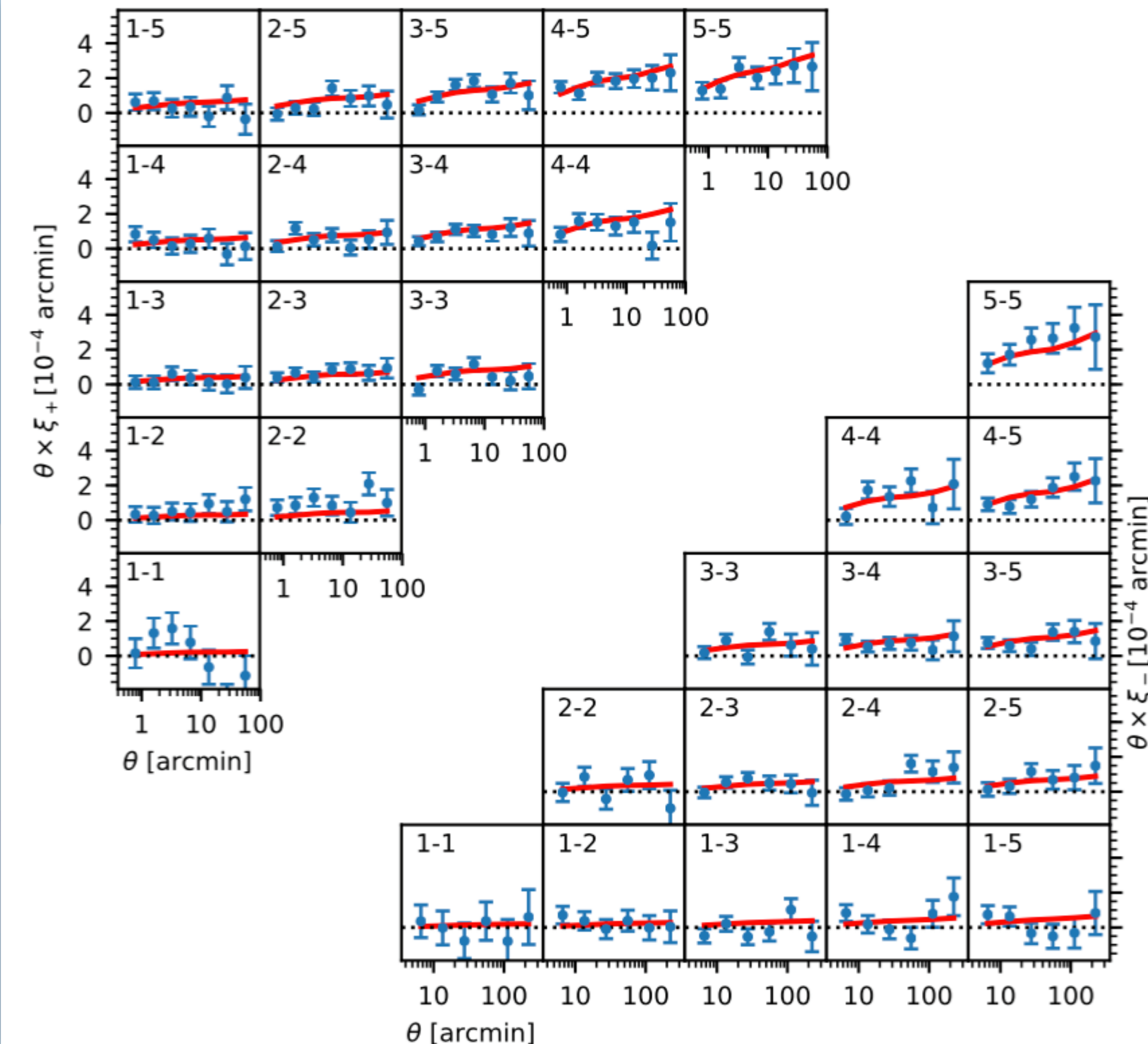


Photo- z estimation from photometry

$$\hat{z} = f_{\theta}(\mathbf{x})$$

*Physical Parameter
Estimation*

Why is Photo- z important?



- Photo- z is a convenient number that *correlates* with redshift
 - ➡ Photo- z is not redshift
- Therefore, Photo- z enables tomography
- Better photo- z
 - better line-of-sight separation
 - higher cosmic shear SNR
 - better cosmological constraints

$$\hat{z} = f_{\theta}(\mathbf{x})$$

Physical Parameter
Estimation

Template photo- z performance in KiDS

$$\hat{z} = f_{\theta}(\mathbf{x})$$

*Physical Parameter
Estimation*

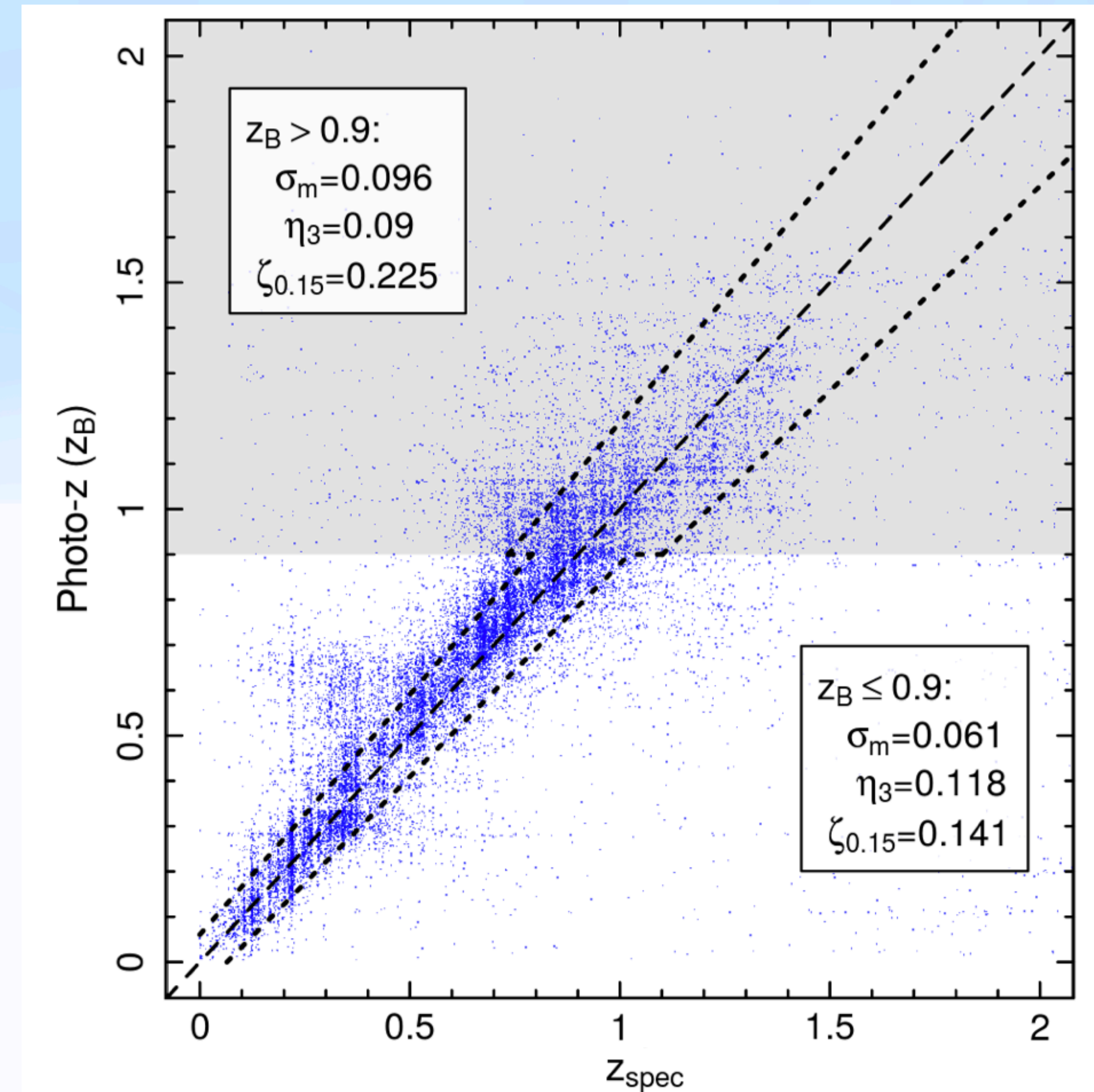
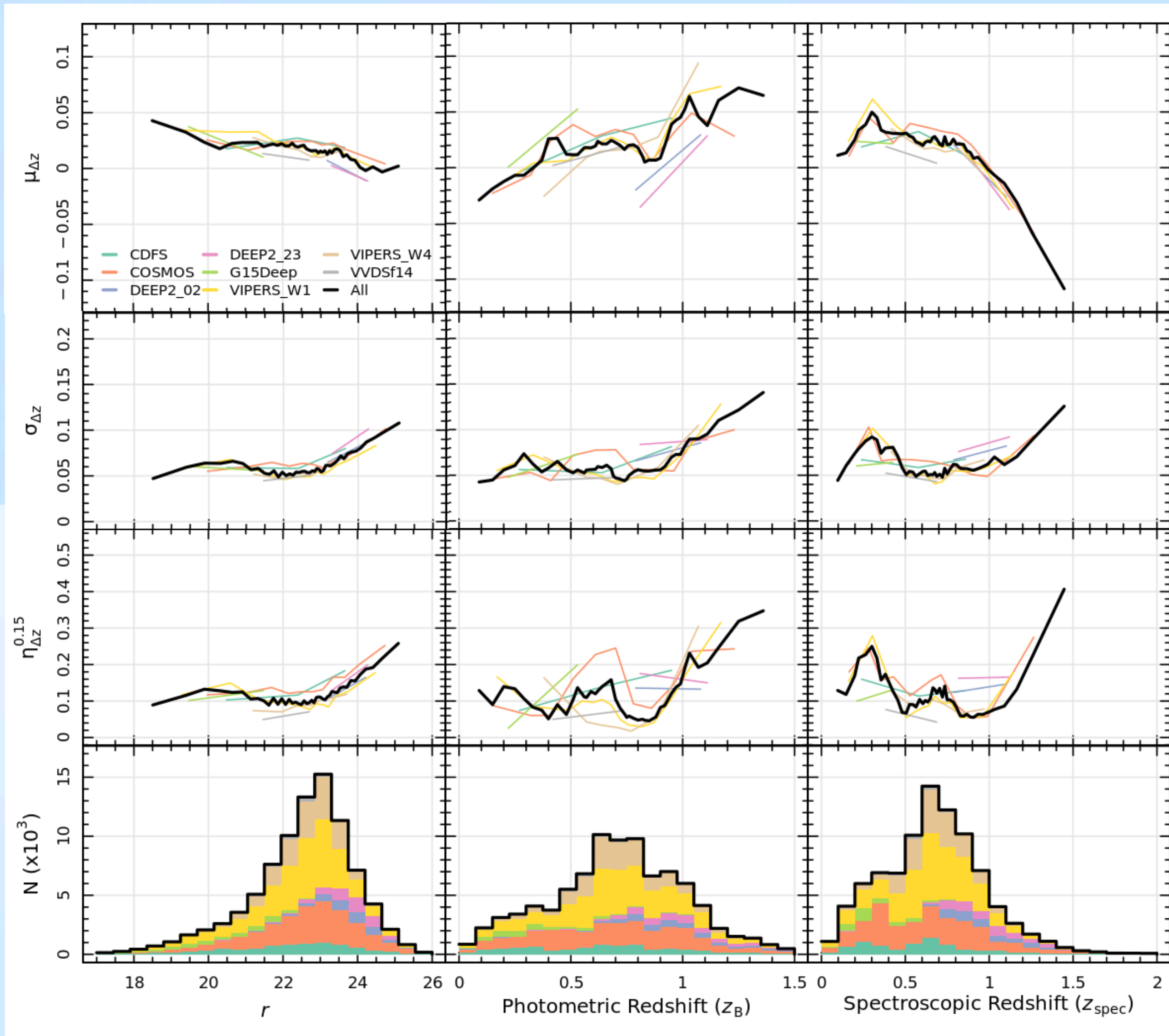
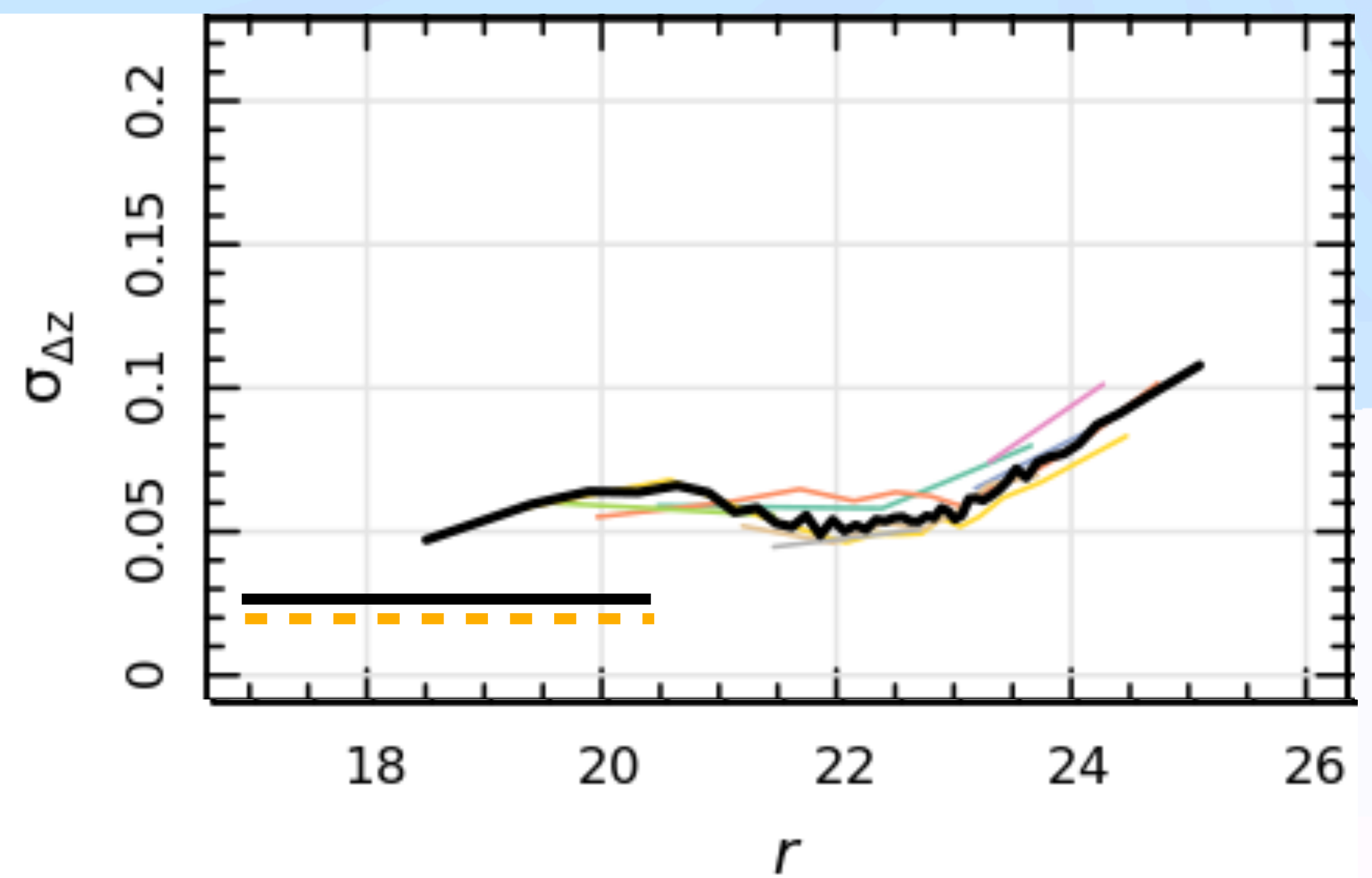
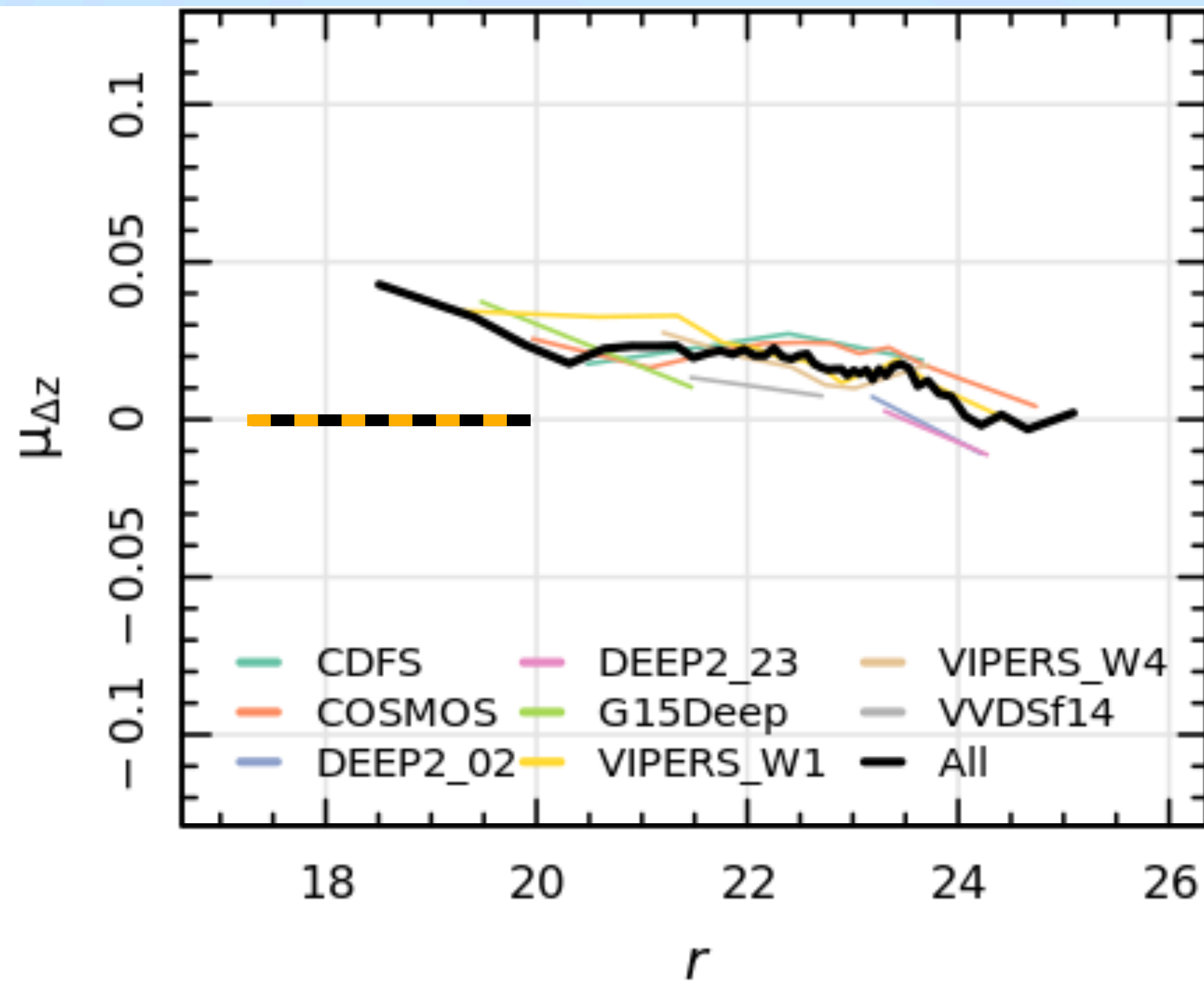


Photo-z estimation in wide-field surveys

Template fitting vs Machine Learning in KiDS



— Bilicki et al (2022): ANNz
- - Jalen et al. (2025): Hybrid-z

ML photo- z often comfortably out-performs template fitting

$$\hat{z} = f_{\theta}(\mathbf{x})$$

*Physical Parameter
Estimation*

Photo-z for stage-IV imaging surveys

Rubin LSST Tomography Optimisation Challenge (Zuntz et al. 2021)

Name	Description	Submitters
RANDOM	Random assignment (control method)	Organizers
IBANDONLY	i-band only (control method)	Organizers
RANDOMFOREST	A collection of individual decision trees	Organizers
LSTM	Long Short-Term Memory neural network	CRB, BF, GT, ESC, EJG
AUTOKERASLSTM	Automatically configured LSTM variant.	CRB, BF, GT, ESC, EJG
LGBM	Light Gradient Boosted Machine	CRB, BF, GT, ESC, EJG
TCN	Temporal Convolution Network	CRB, BF, GT, ESC, EJG
CNN	Convolutional Neural Network	CRB, BF, GT, ESC, EJG
ENSEMBLE	Combination of three other network methods	CRB, BF, GT, ESC, EJG
SIMPLESOM	Self-Organizing Map	AHW
PQNLD	Extension of SIMPLESOM combining with template-based redshift estimation	AHW
UTOPIA	Nearest-neighbours, optimised in the limit of large representative training sets	AHW
COMPLEXSOM	Extension of SIMPLESOM with an additional assignment optimization	DA, AHW, AN, EG, BA, ABr
GPzBINNING	Gaussian Process redshift estimator and binner	PH
JAXCNN/JAXRESNET	Two related CNN-based bin edge optimizers	AP
NEURALNETWORK1/2	Dense network optimizing bin assignment for two metrics	FL
PCACLUSTER	Principal Component Analysis of fluxes then optimized clustering	DG
ZOTBIN/ZOTNET	Two related neural network methods with a preprocessing normalizing flow	DK
FFNN	Feed-forward neural network	EN
FUNBINS	Random forest with various nominal bin edge selection	ABa
MLPQNA	Multi-layer perceptron	SC, MB
STACKED GENERALIZATION	Combination of Gradient Boosted Trees classifiers with standard Random Forests (50/50)	JEC

ML method

Other

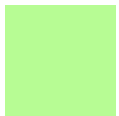
$$\hat{z} = f_{\theta}(\mathbf{x})$$

Physical Parameter Estimation

Photo-z for stage-IV imaging surveys

Euclid Photometric Redshift Challenge (Plantini et al. 2020)

Type	
Le Phare	Template-fitting
CPz	Random forest classification + template-fitting
Phosphoros	Template-fitting
EAzY	Template-fitting
METAPHOR	Machine-learning: neural network
ANNz	Machine-learning: neural network
GPz	Machine-learning: Gaussian processes
GBRT	Machine-learning: boosted decision trees
RF	Machine-learning: random forest
Adaboost	Machine-learning: boosted decision trees
DNF	Machine-learning: nearest neighbor
frankenz	Machine-learning: nearest neighbor
NNPZ	Machine-learning: nearest neighbor

 ML method

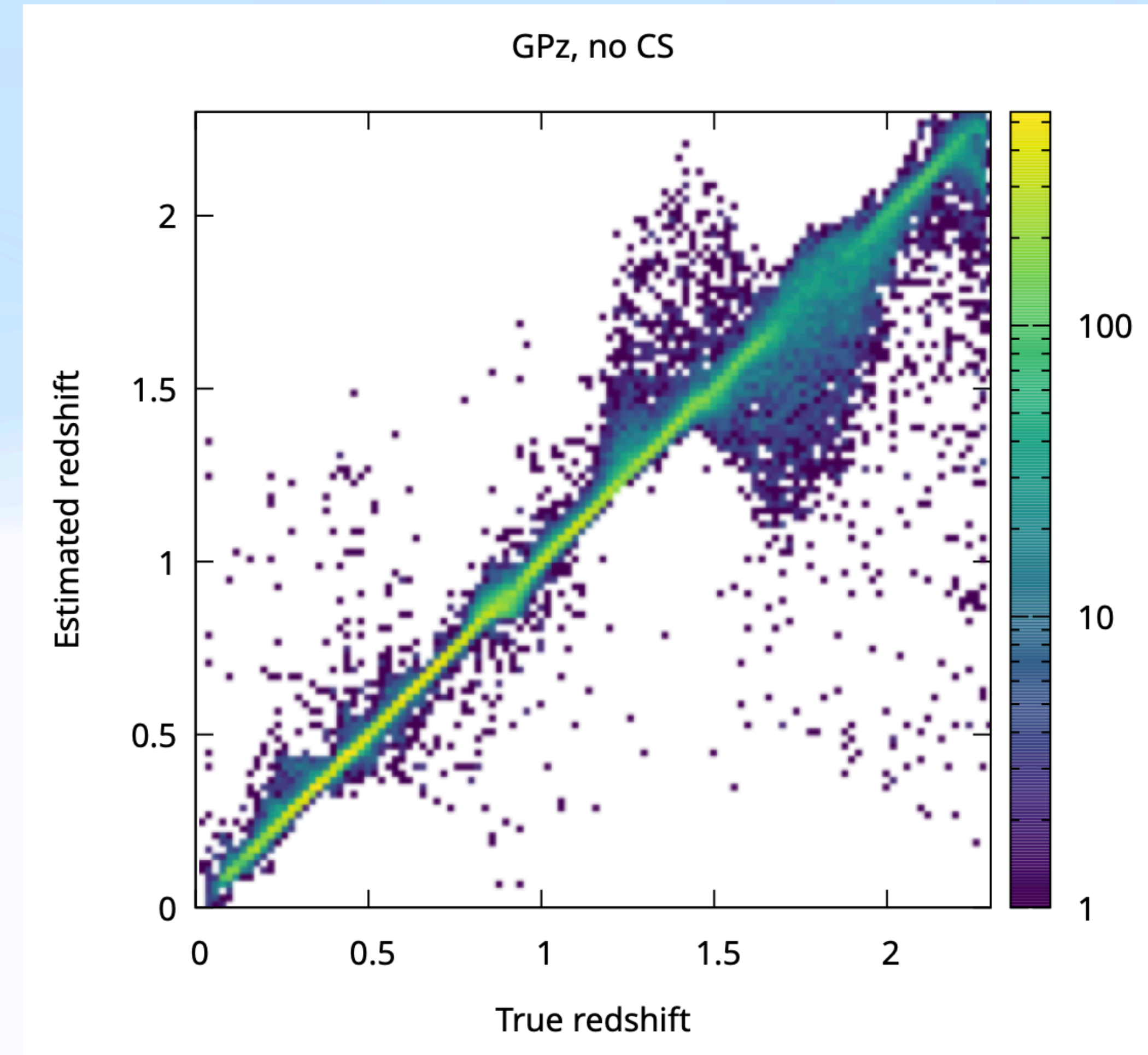
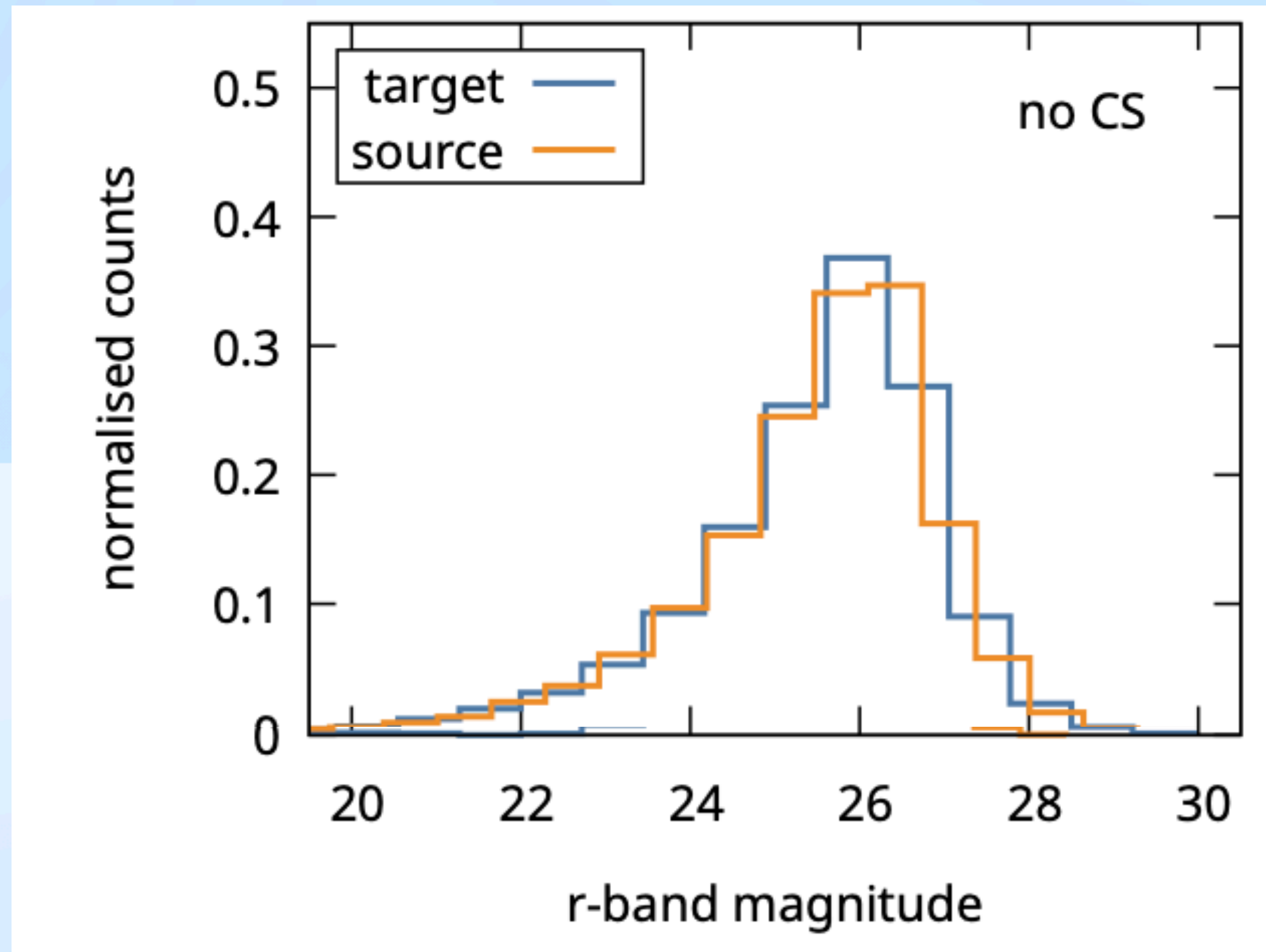
 Other

$$\hat{z} = f_{\theta}(\mathbf{x})$$

*Physical Parameter
Estimation*

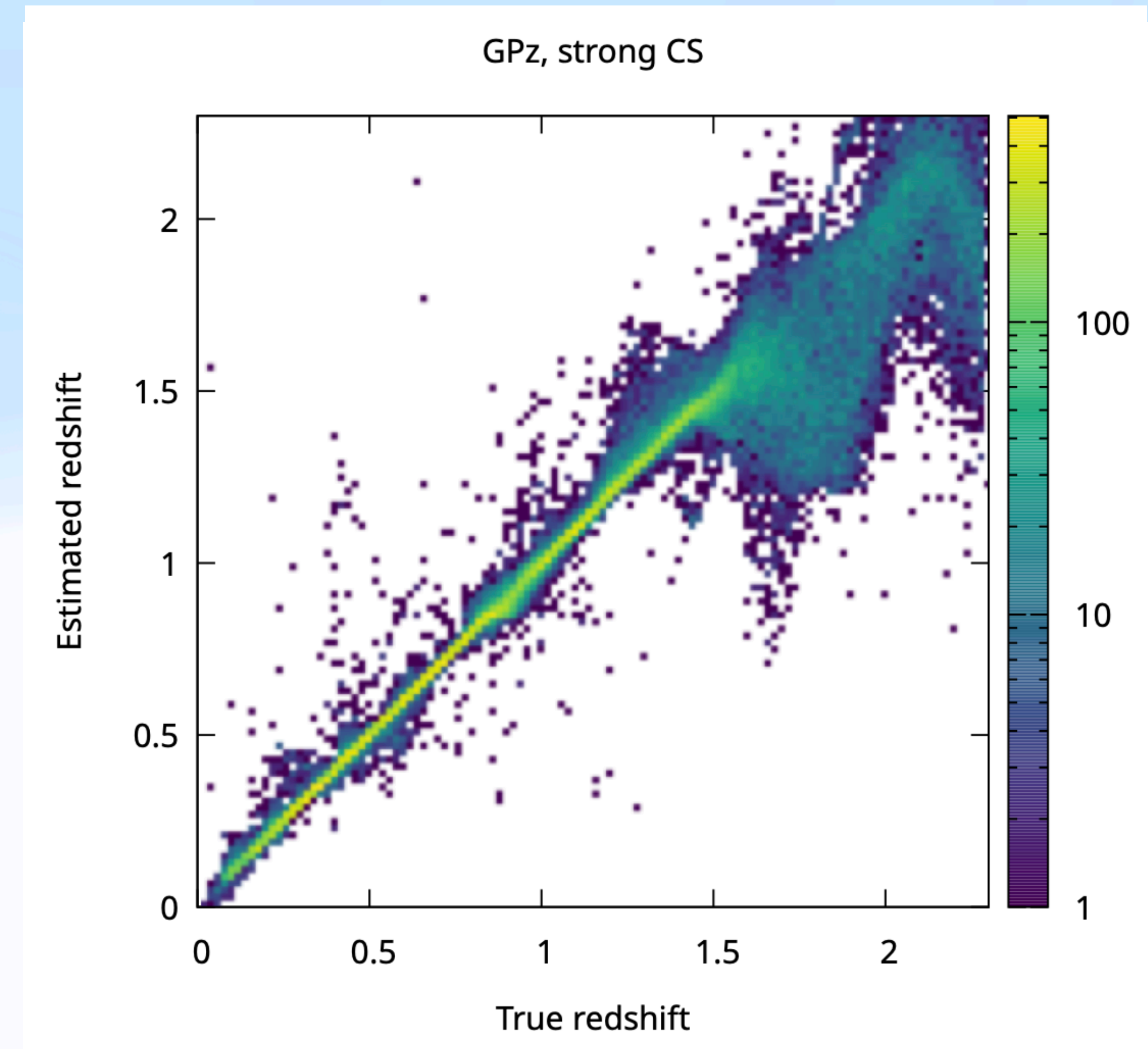
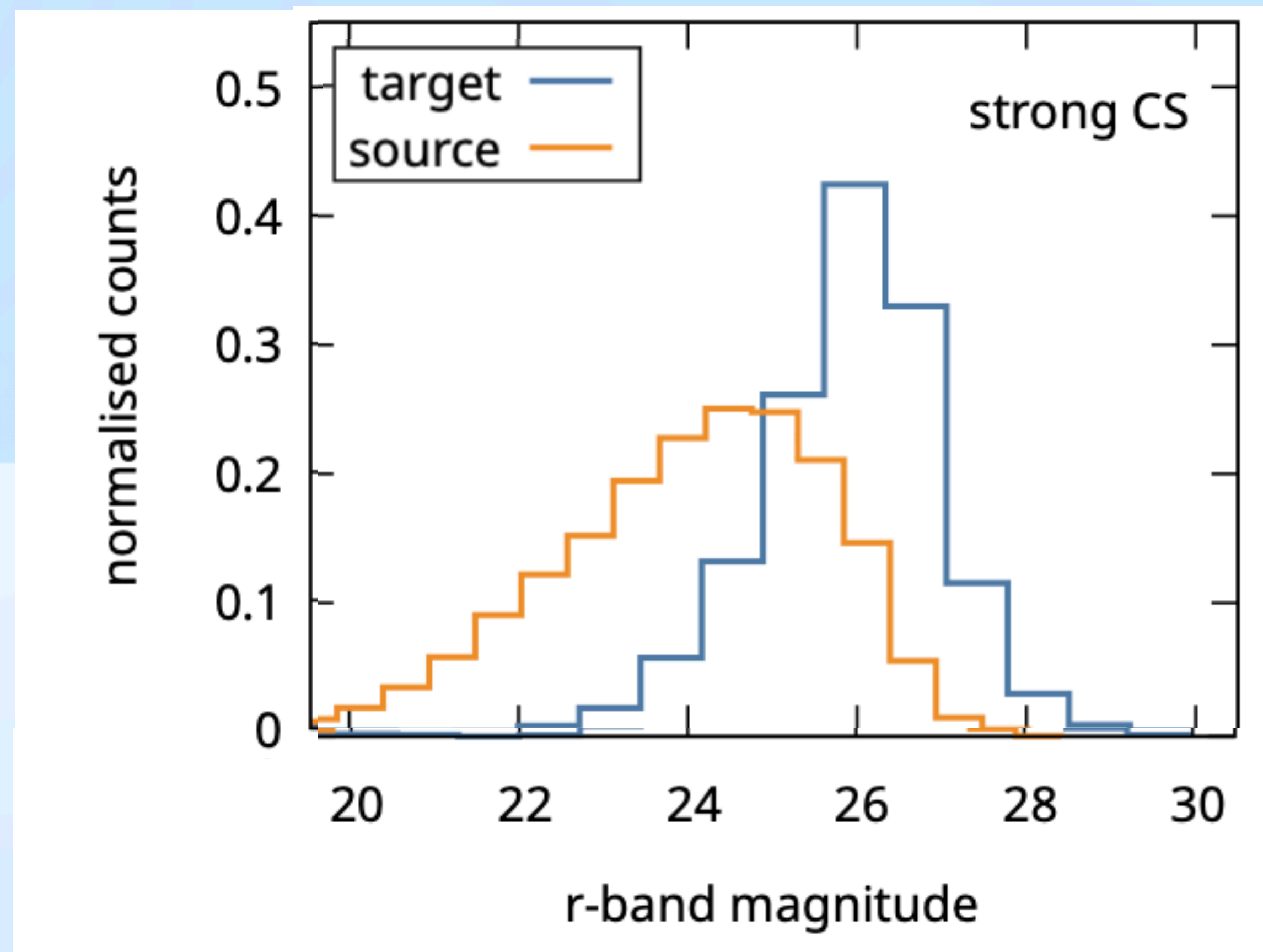
But ML methods are not perfect

Fidelity of training data, or compensating methods, is crucial



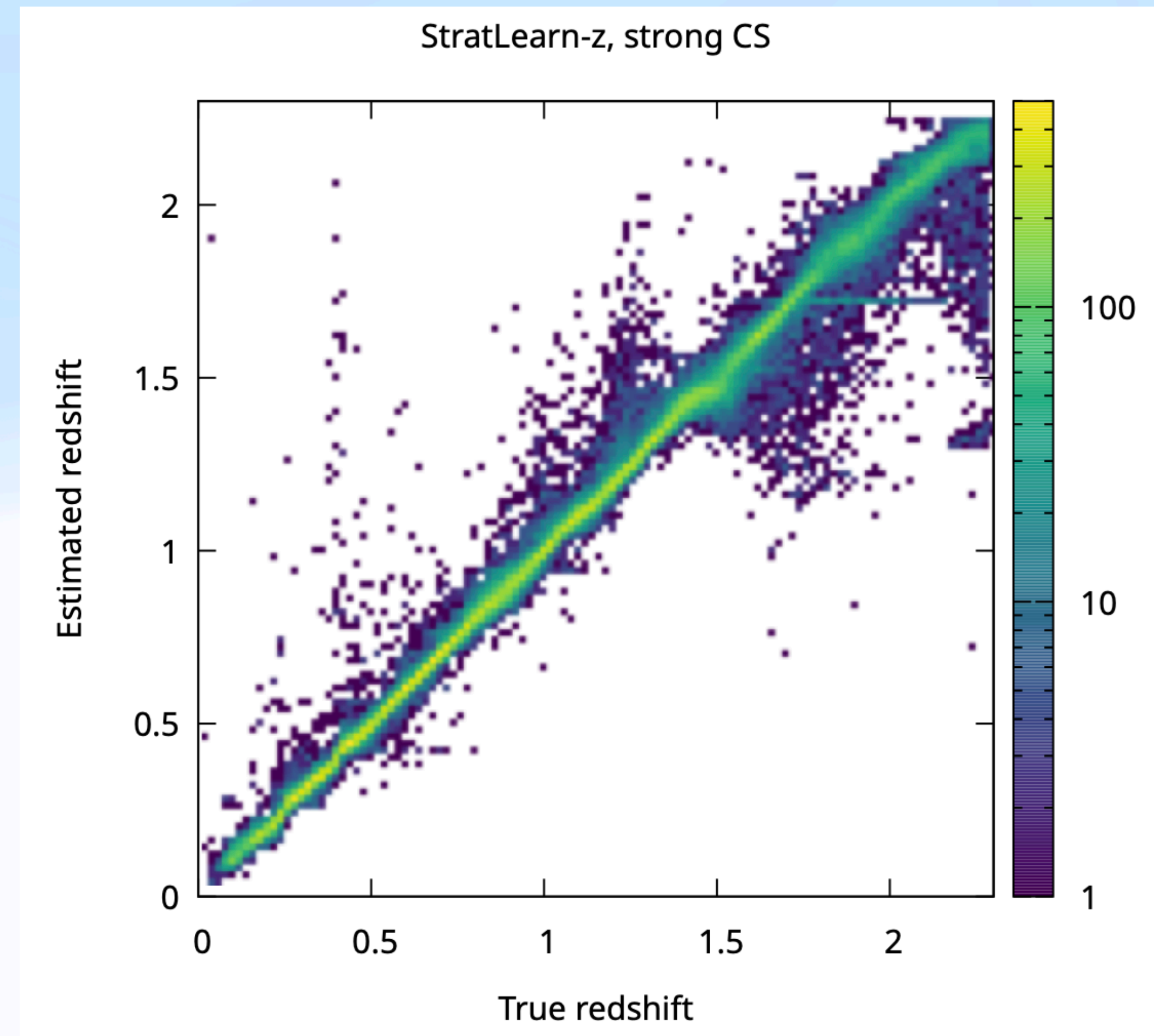
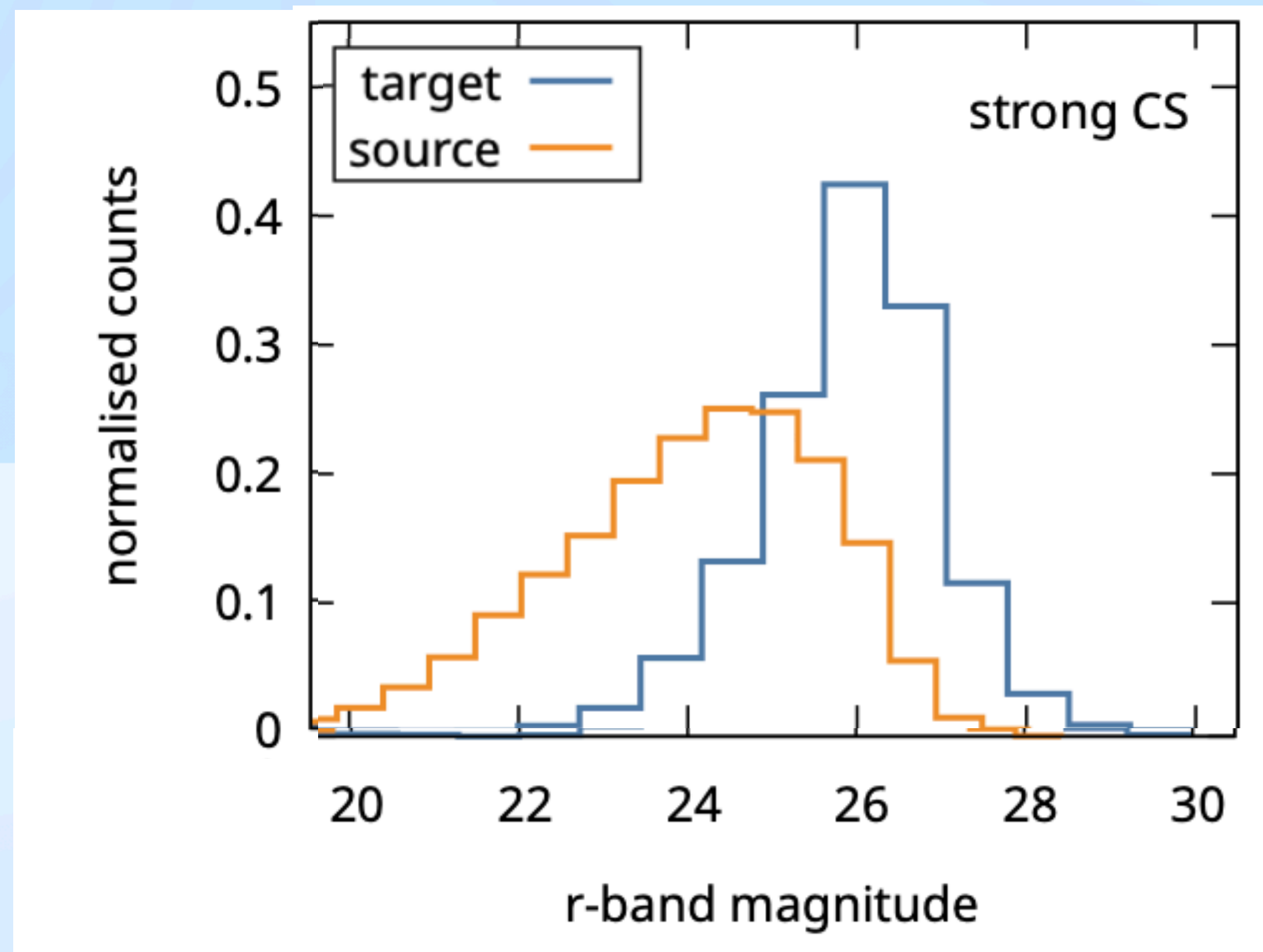
But ML methods are not perfect

Fidelity of training data, or compensating methods, is crucial



But ML methods are not perfect

Fidelity of training data, or compensating methods, is crucial



$$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$$

Classification & Systematics Correction

Extremely important for robust inference. Particularly important for the measurement of redshift distributions

Recent & notable examples:

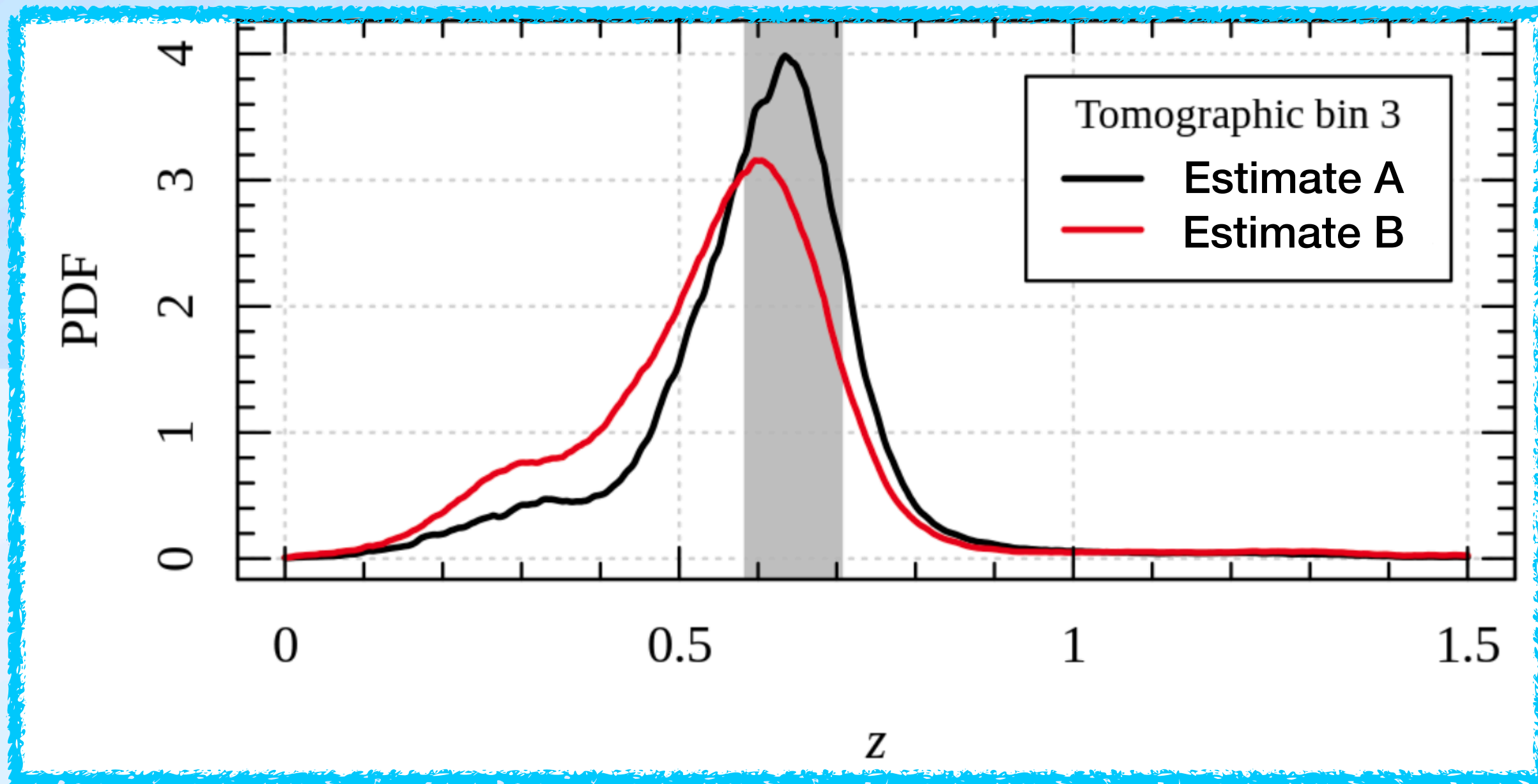
- Stolzner et al. (2021)
- Autenrieth et al. (2024)
- Ashmead et al. (2026)

Example talks this week:

- Yvonne Becherini (0900 Tues)

Sensitivity of cosmic shear to $N(z)$

Why is accuracy important?



Example Redshift Distributions

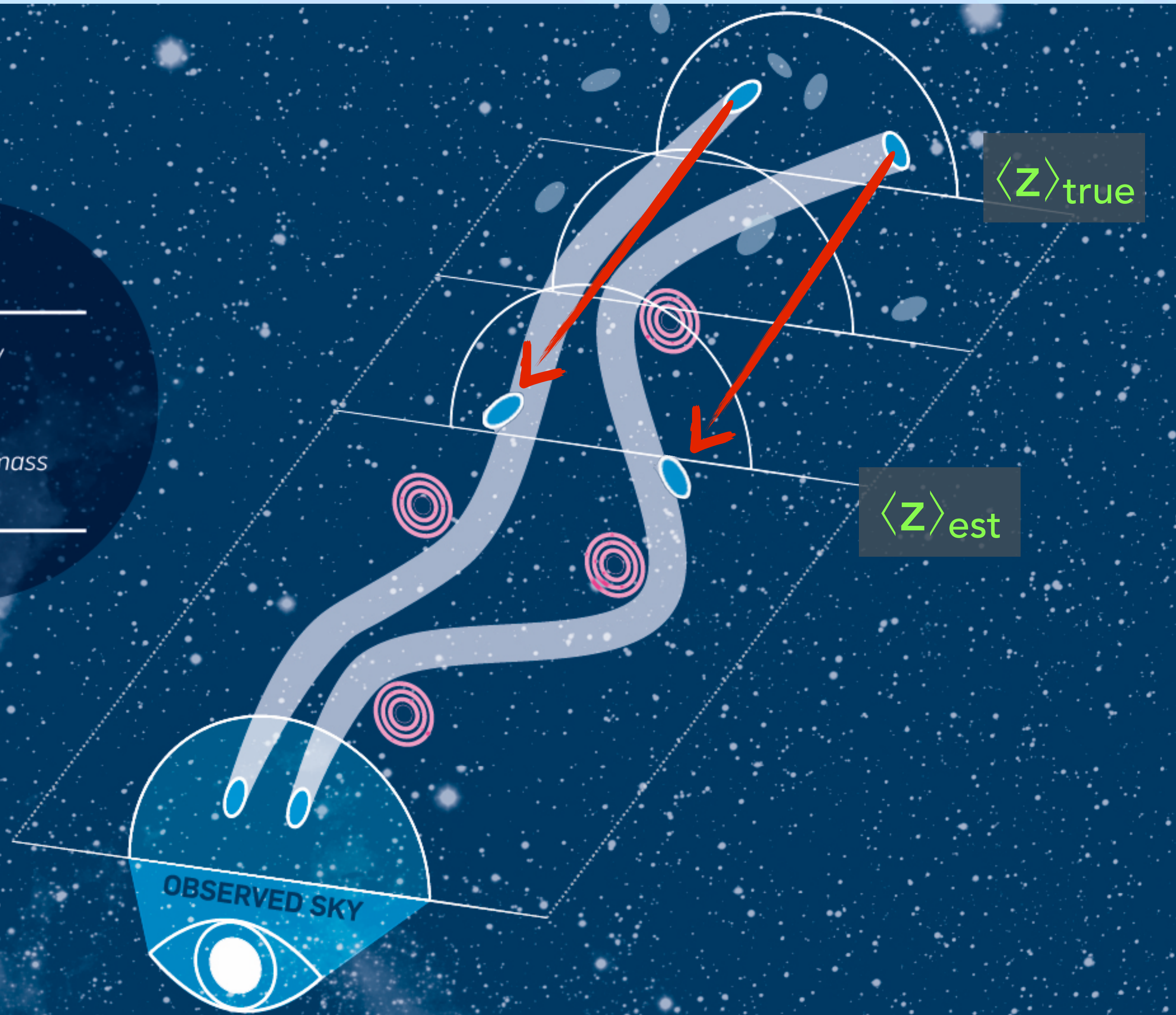
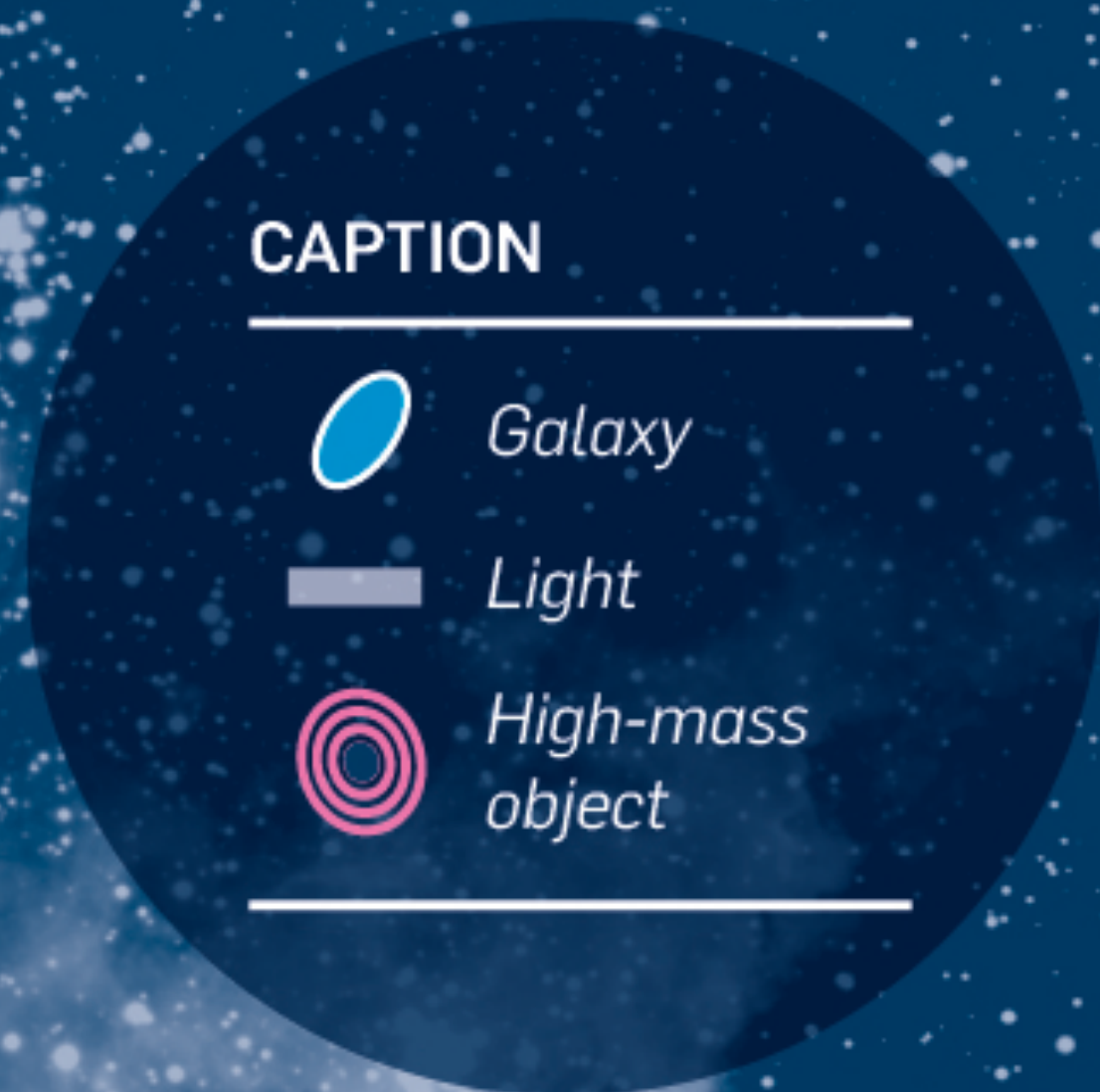
$$\xi_+(\theta) = \int_0^\infty \frac{d\ell \ell}{2\pi} J_0(\ell\theta) P_\kappa(\ell)$$

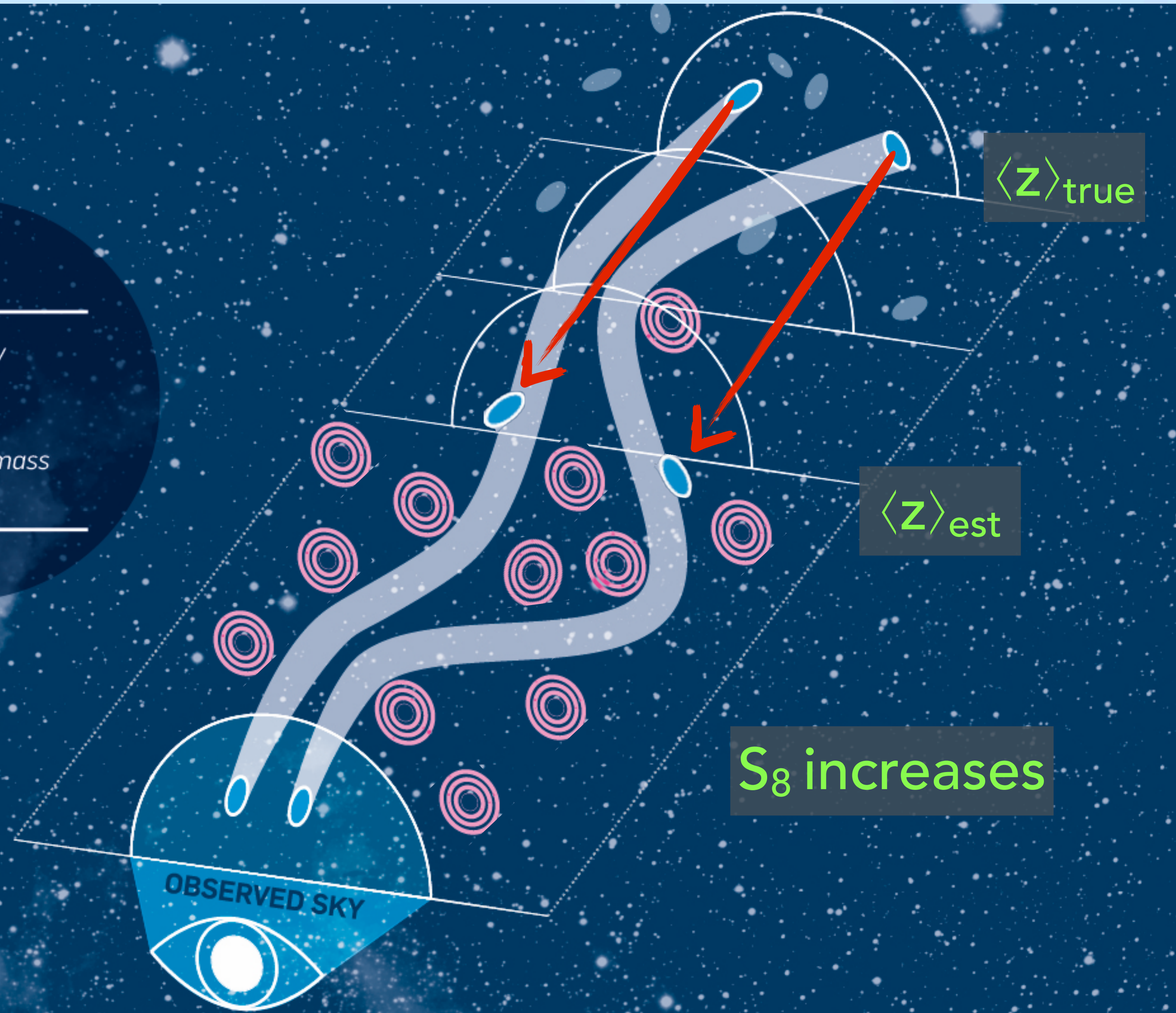
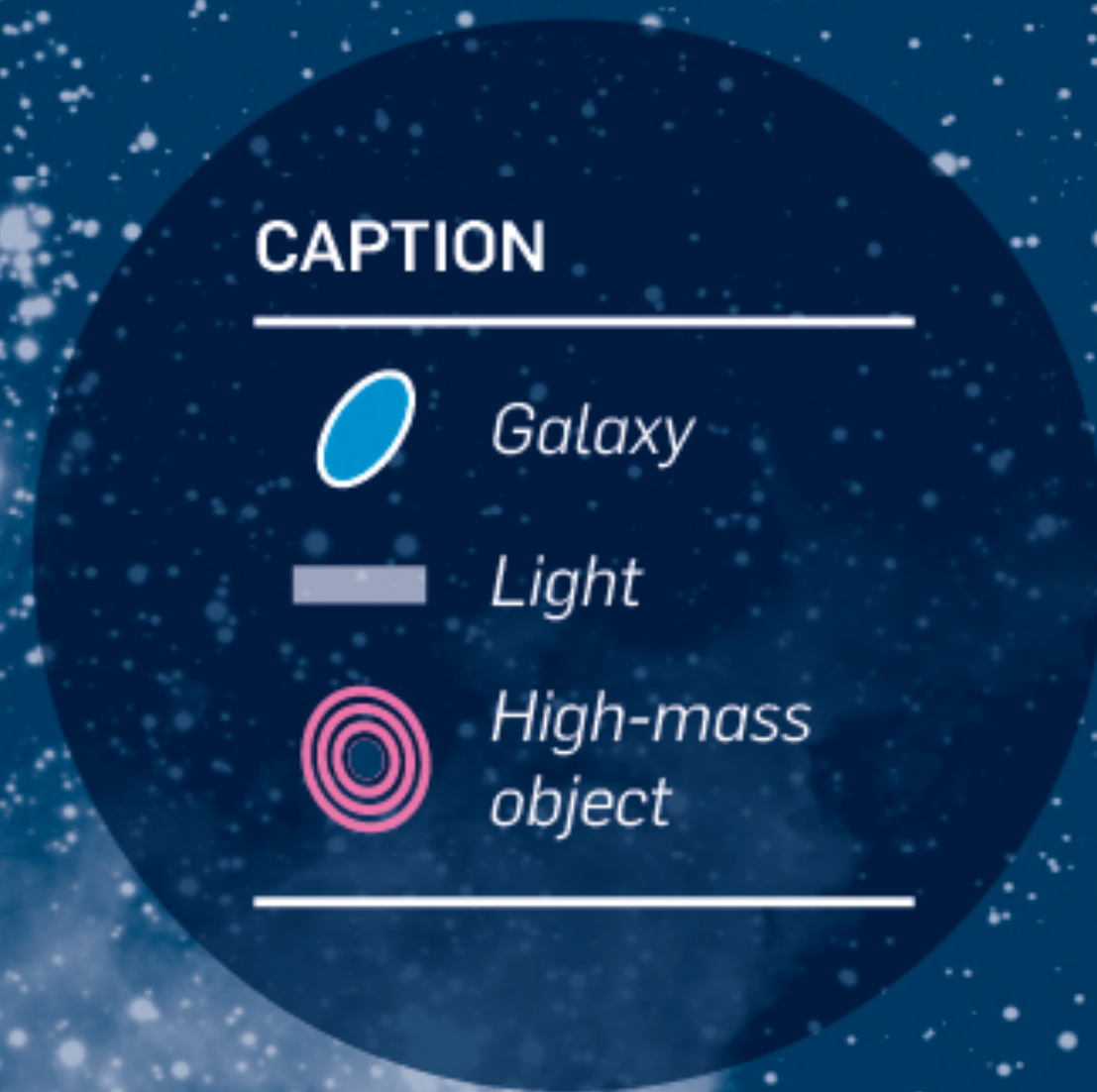
$$\xi_-(\theta) = \int_0^\infty \frac{d\ell \ell}{2\pi} J_4(\ell\theta) P_\kappa(\ell)$$

$$P_\kappa(\ell) = \frac{9H_0^4 \Omega_m^2}{4c^4} \int_0^{\chi_h} d\chi \frac{g^2(\chi)}{a^2(\chi)} P_\delta \left(\frac{\ell}{f_K(\chi)}, \chi \right)$$

$$g(\chi) = \int_\chi^{\chi_h} d\chi' p_\chi(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')}$$

Model construction

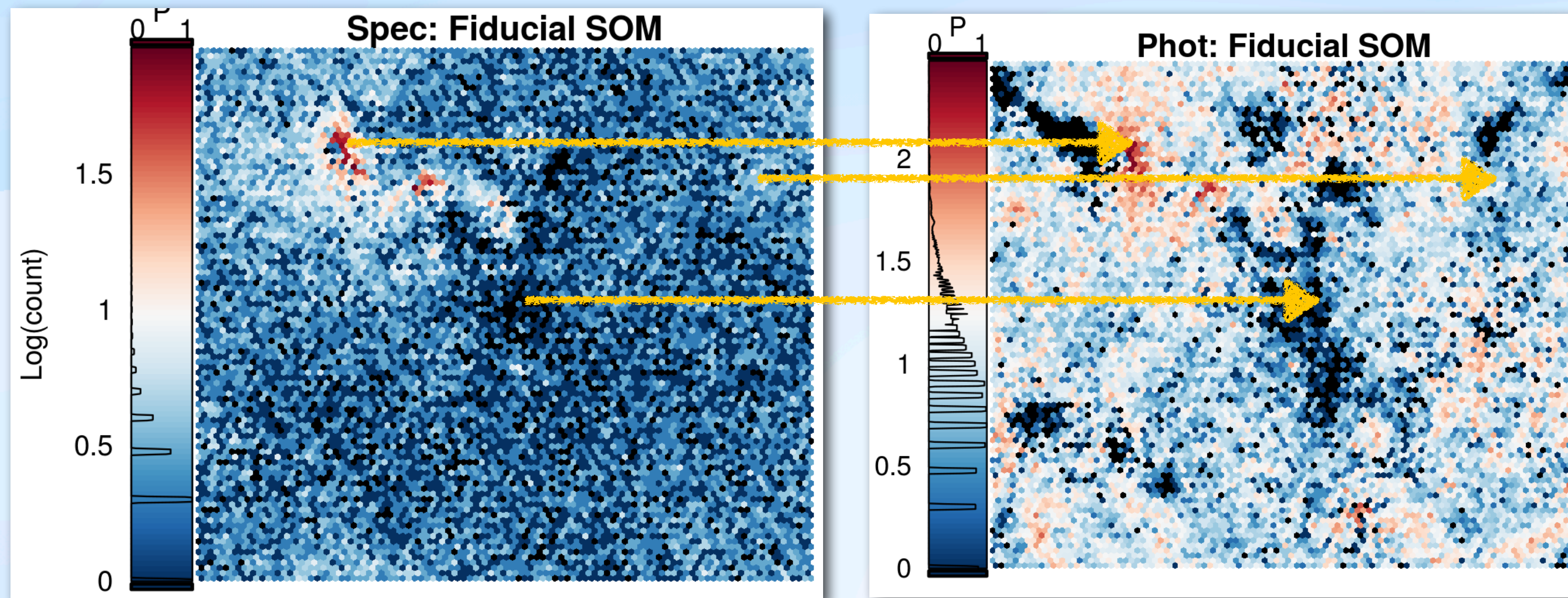




ML Redshift Distribution Calibration

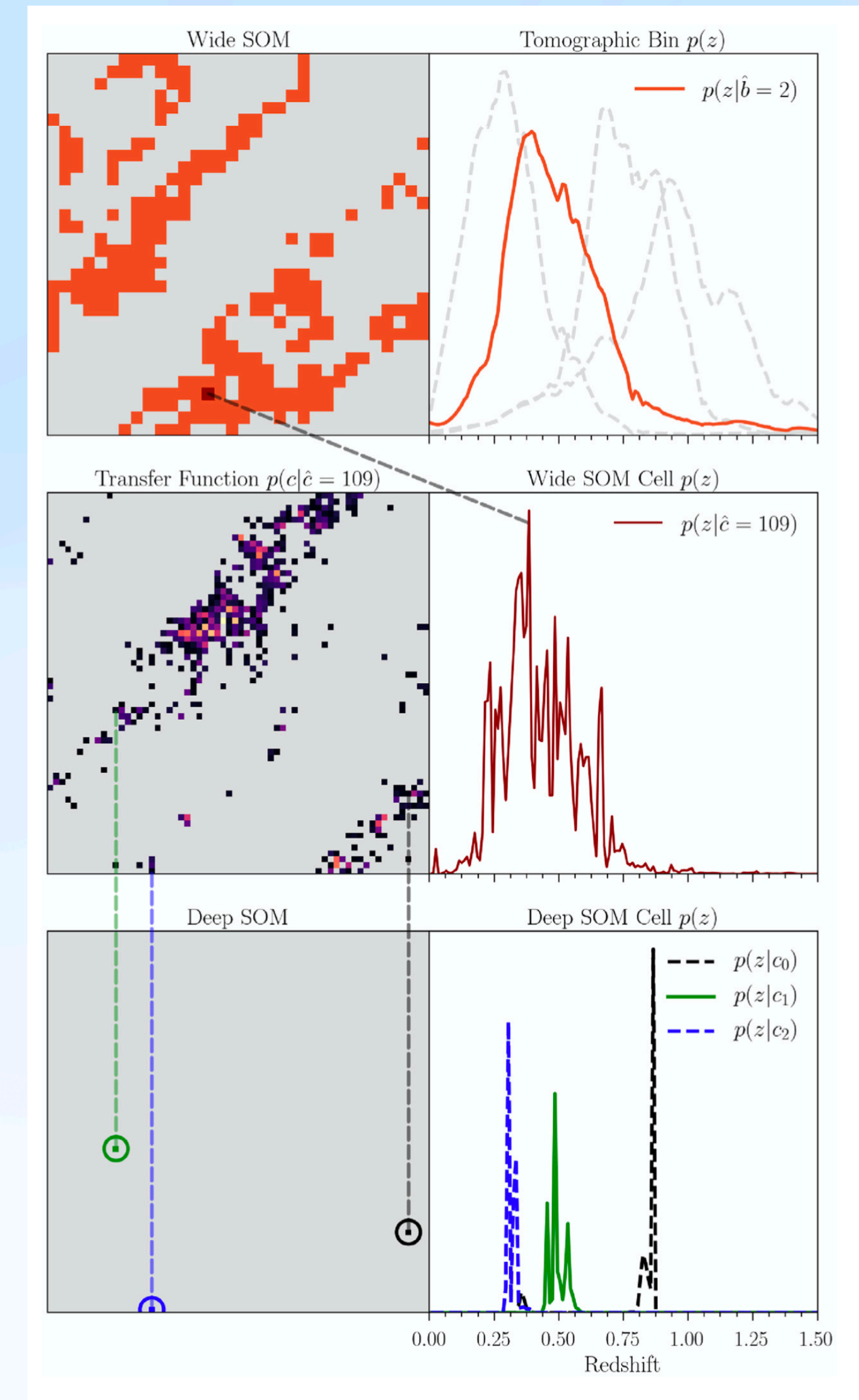
Early work used stacking photo- z , then k -nearest-neighbour weighting. But since 2020, the cosmic shear community has been caught firmly in the grip of the Self-Organising Map (SOM). This is because SOM gives us both:

- (1) Unsupervised classification &
- (2) human-interpretable representations of $\mathbb{R}^n$ -spaces



KiDS1000

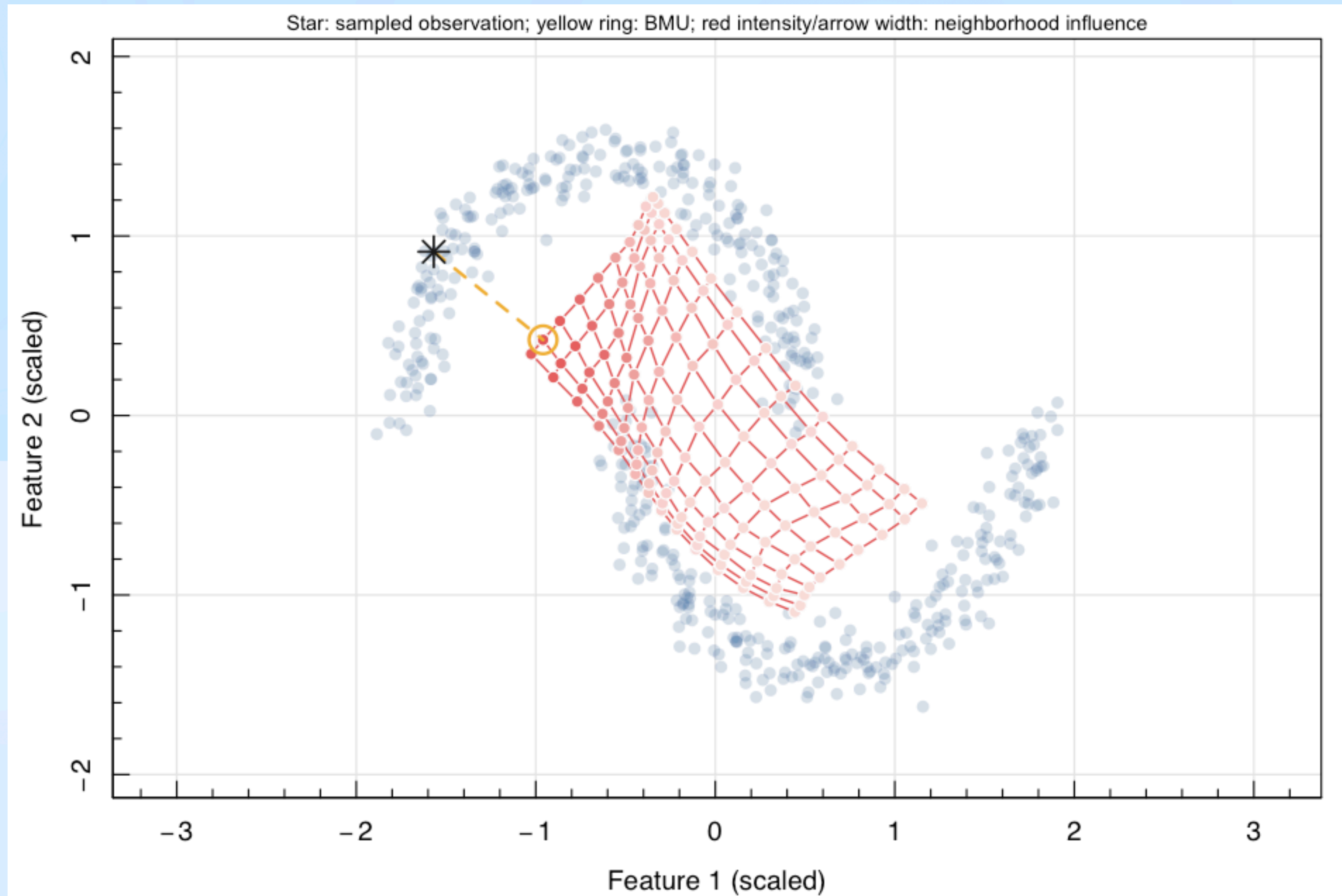
Wright et al (2020a)



DESY3

Myles et al (2021)

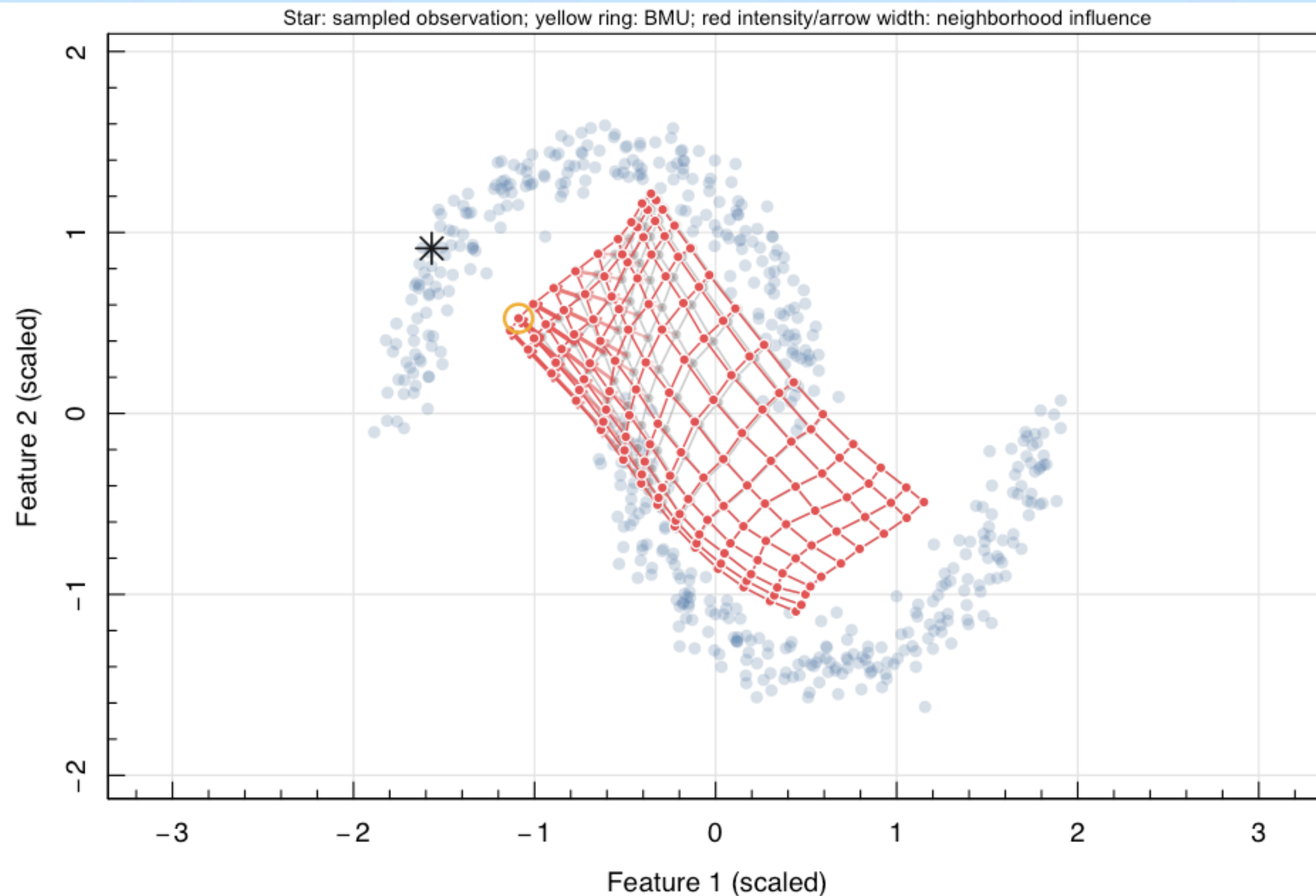
SOM mapping of complex spaces: early iteration



$$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$$

**Classification &
Systematics Correction**

SOM mapping of complex spaces: early iteration

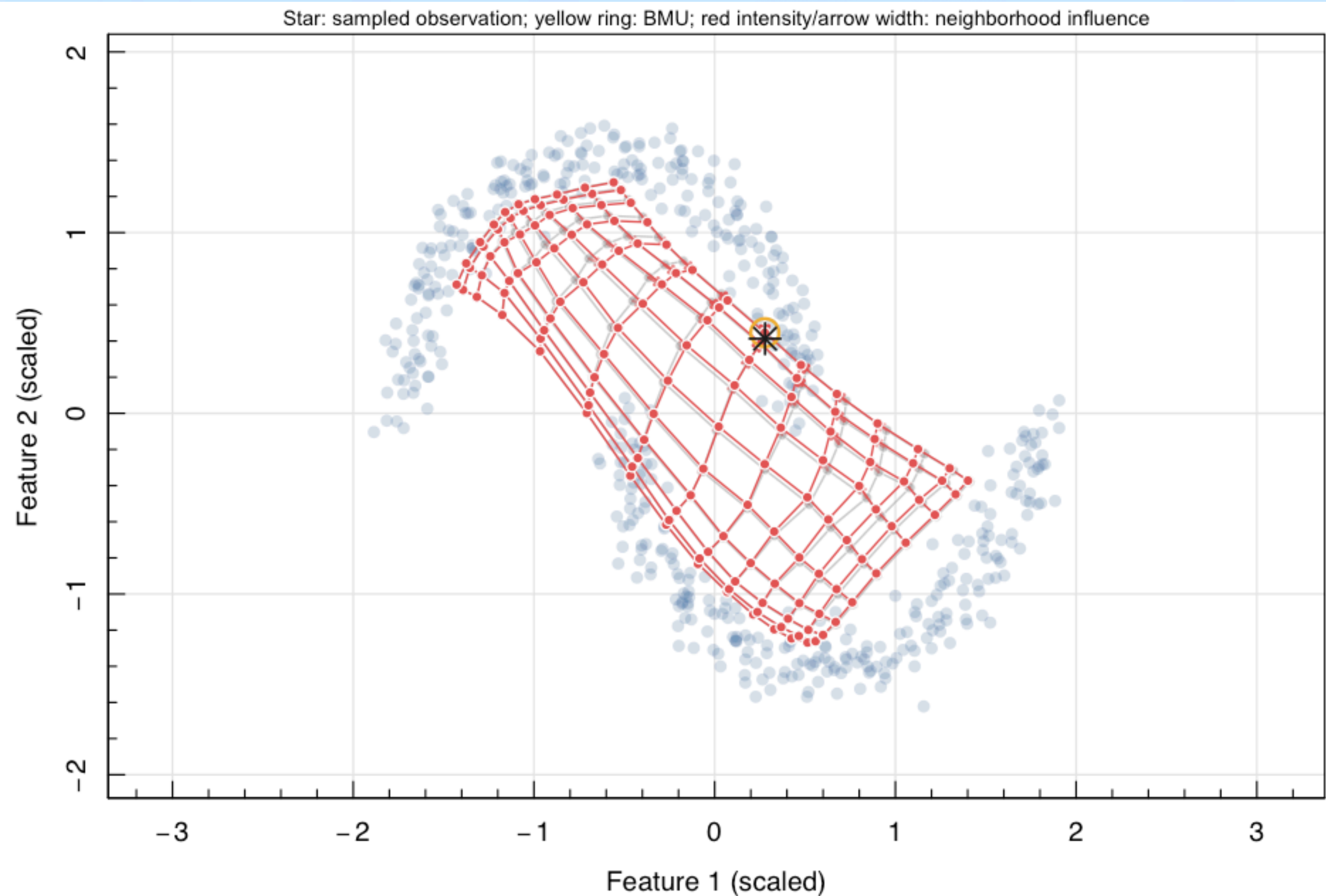


Early epochs have
a large influence
over the manifold

$$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$$

*Classification &
Systematics Correction*

SOM mapping of complex spaces: late iteration

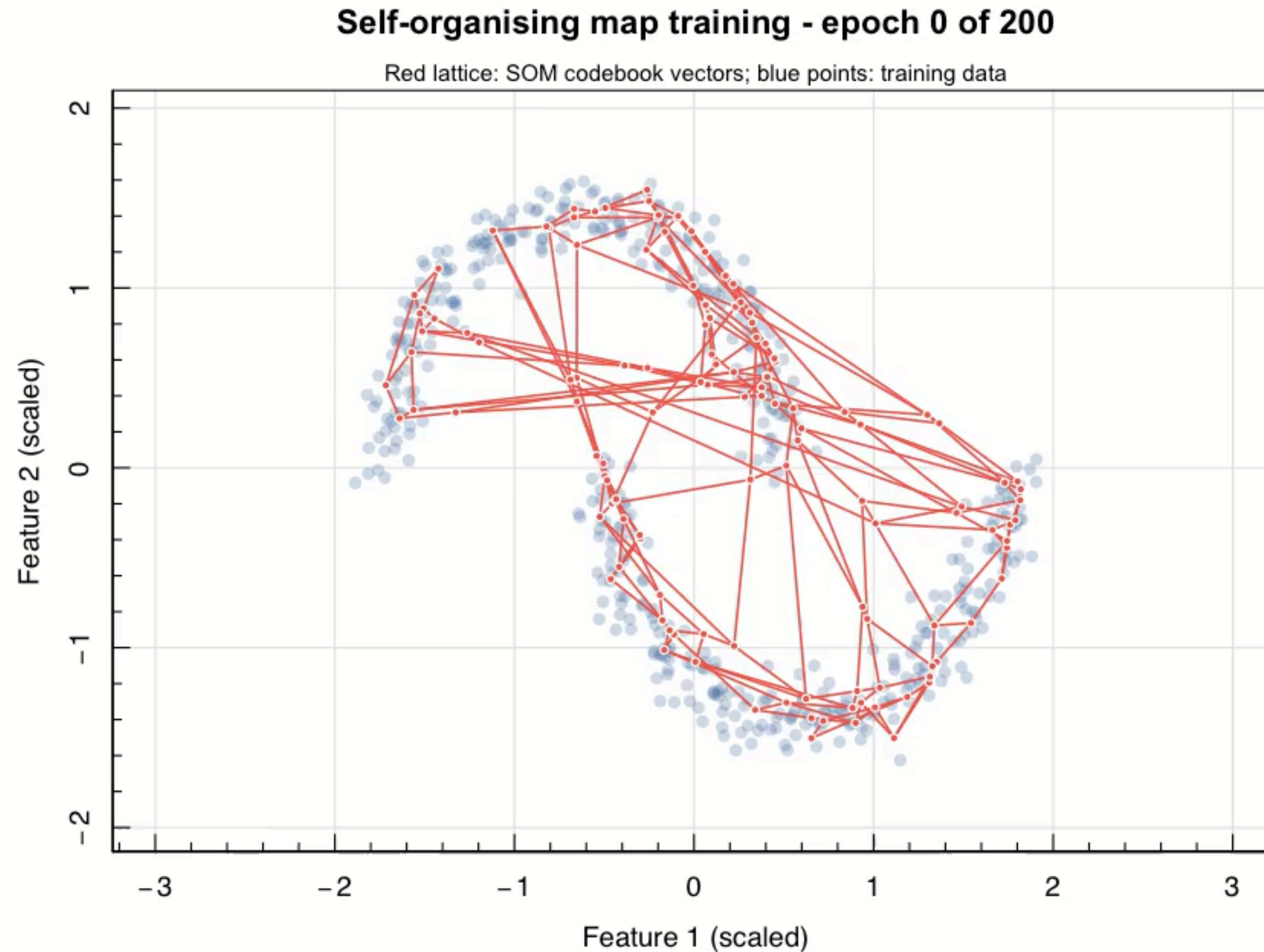


Late epochs
only influence
the BMU and its
neighbours

$$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$$

*Classification &
Systematics Correction*

SOM mapping of complex spaces: late iteration



$$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$$

**Classification &
Systematics Correction**

Ways that your ML analysis can fail

There are three classic failure modes in ML regression/classification problems:

1. Covariate shift: $p_{\text{tr}}(z | c) = p_{\text{tg}}(z | c) \ \& \ p_{\text{tr}}(c) \neq p_{\text{tg}}(c)$
2. Prior Probability shift: $p_{\text{tr}}(c | z) = p_{\text{tg}}(c | z) \ \& \ p_{\text{tr}}(z) \neq p_{\text{tg}}(z)$
3. Concept drift: $p_{\text{tr}}(z | c) \neq p_{\text{tg}}(z | c)$

These all affect redshift calibration datasets:

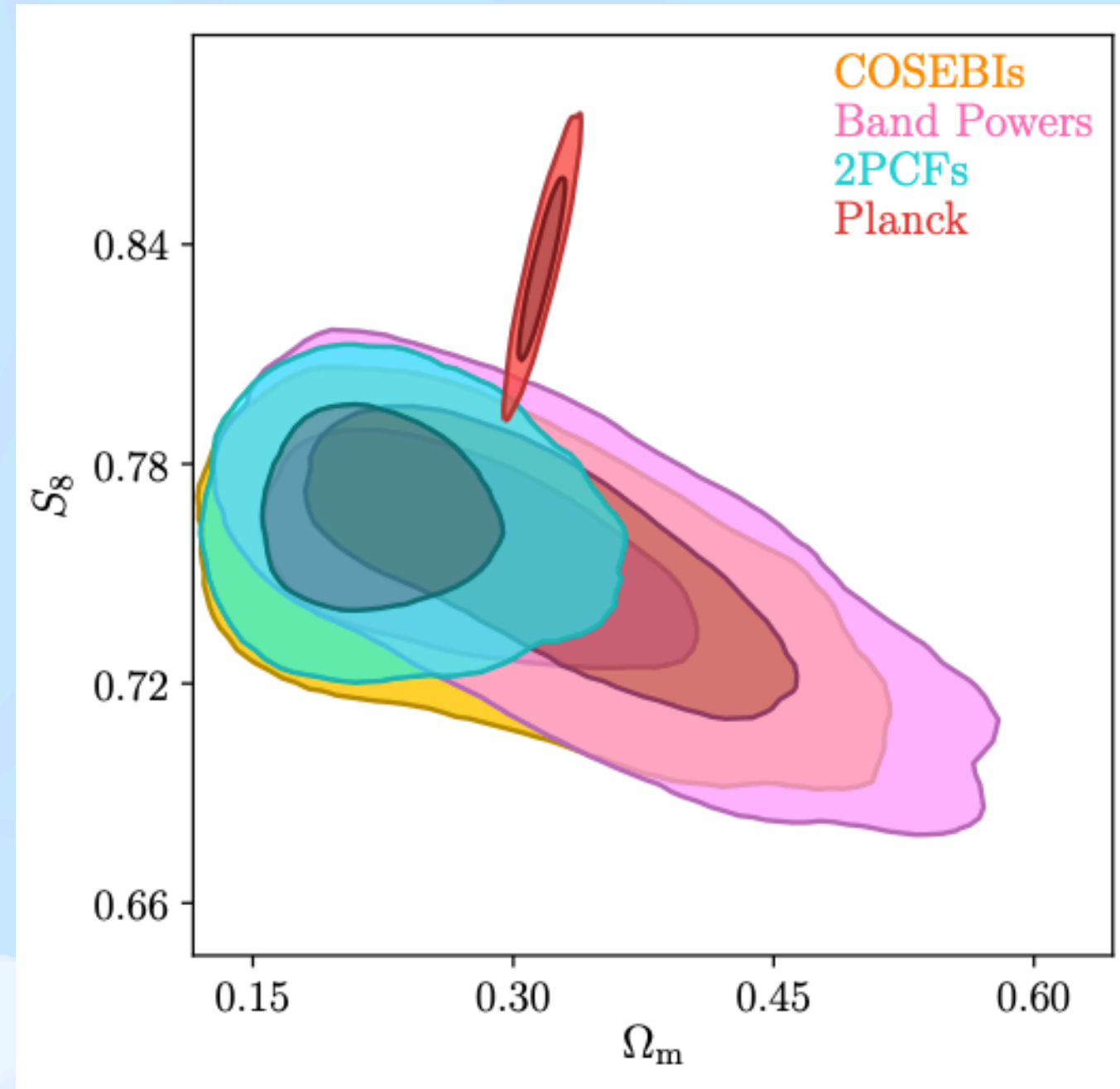
1. Targeting in spectroscopy differs from photometry
2. Redshift success and confidence is systematic
3. SOM cells have non-zero size

➡ the above effects also persist below the cell level

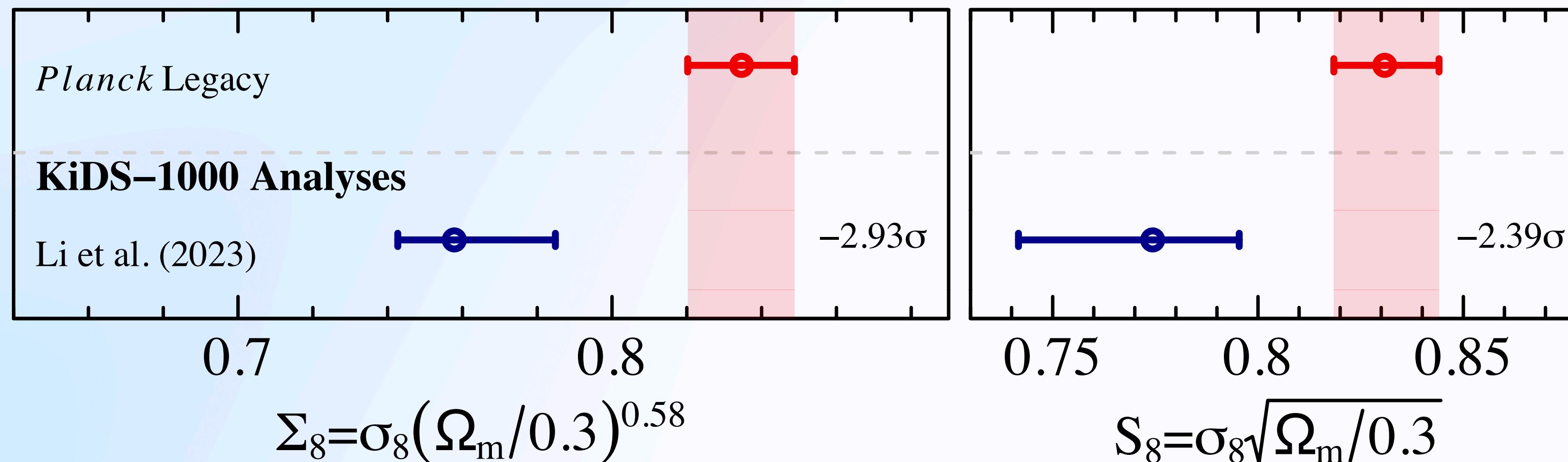
$$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$$

**Classification &
Systematics Correction**

Nz bias & covariate shift correction



- Results from KiDS-1000 (Asgari et al. 2021)
- Used SOM redshift estimation
- 3-sigma tension with Planck CMB



$$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$$

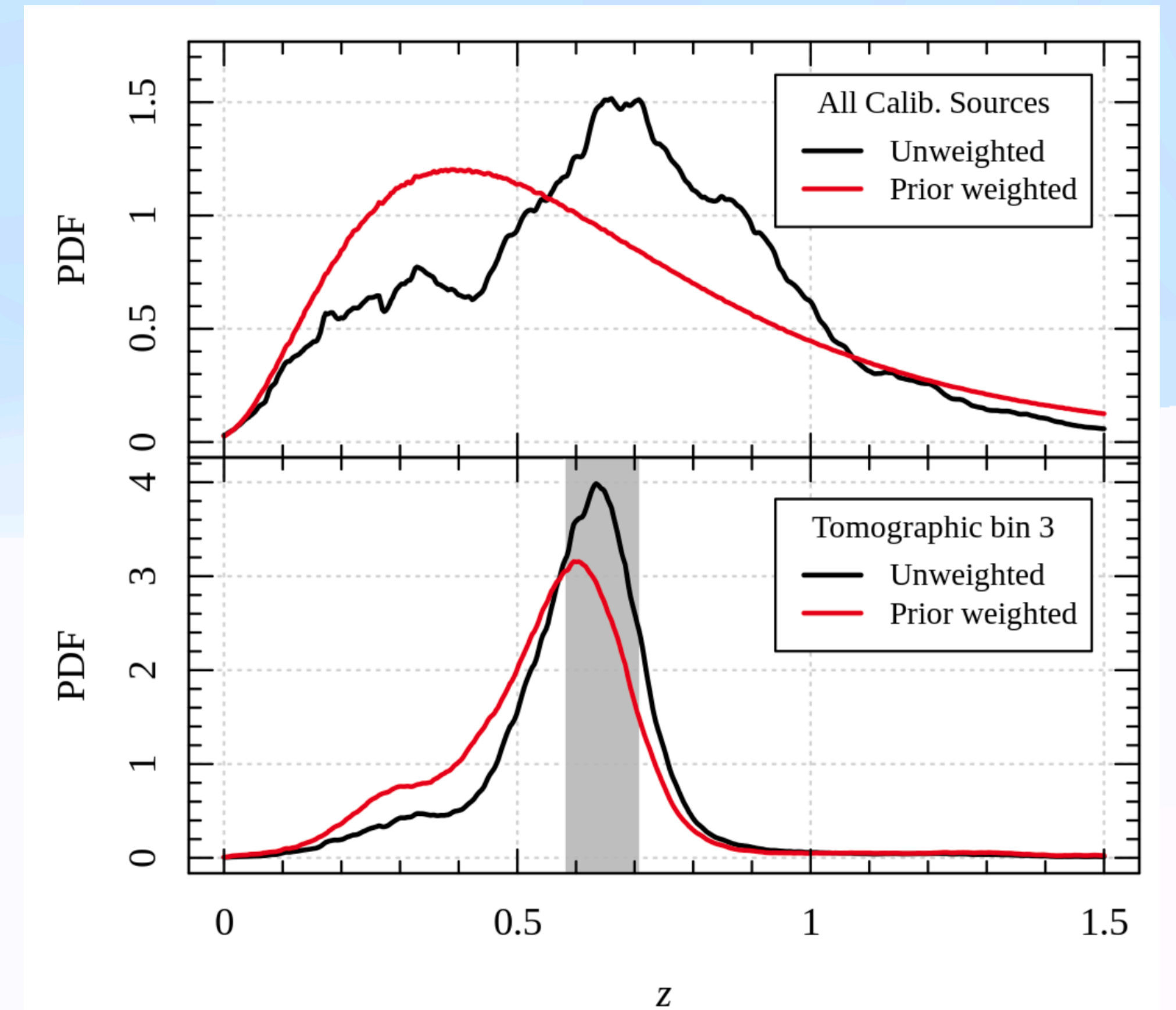
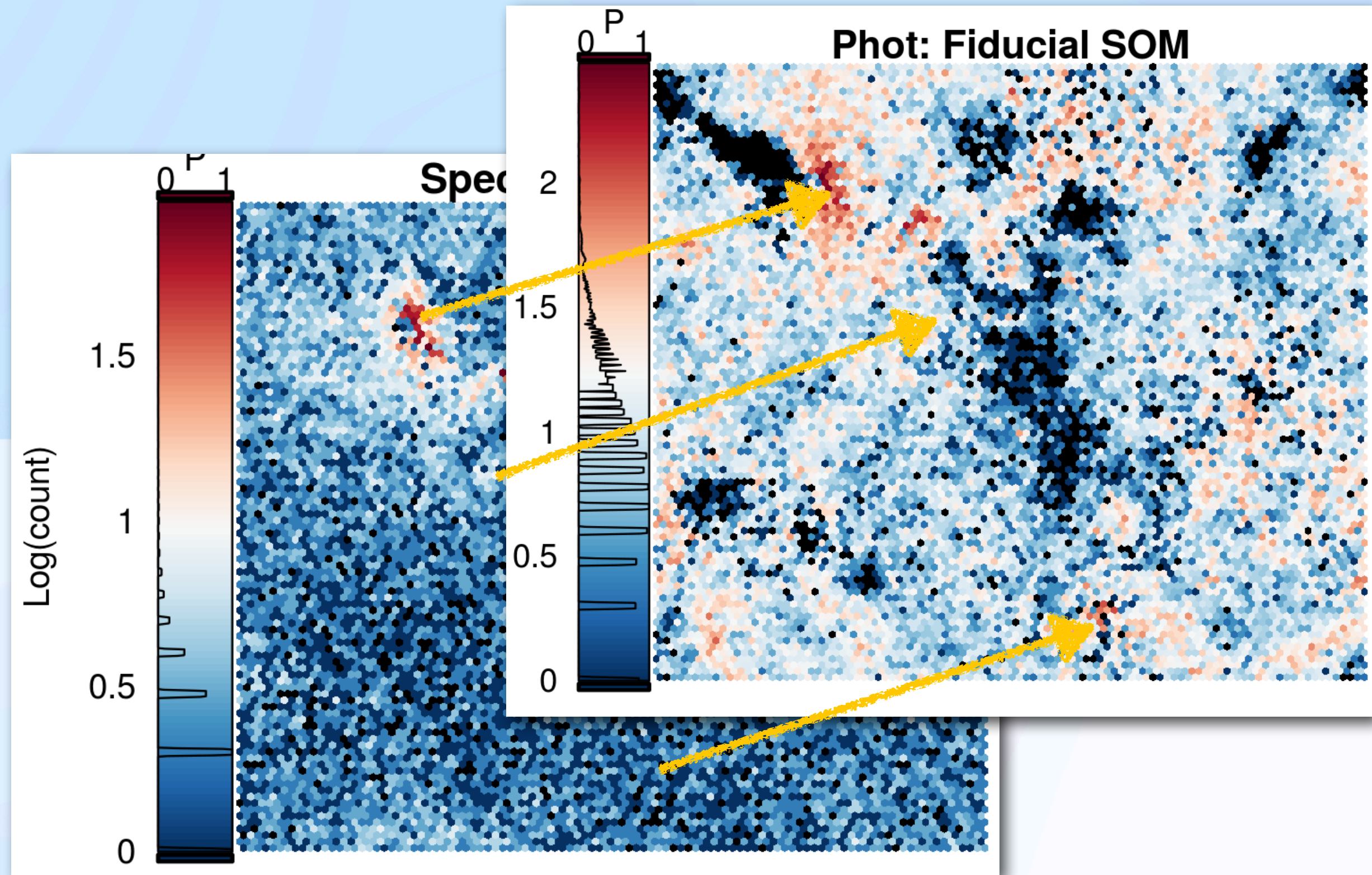
Classification & Systematics Correction

Nz bias & covariate shift correction

Correction with SOMs

$$p_t(z|\mathbf{x})p_t(\mathbf{x}) \neq p_v(z|\mathbf{x})p_v(\mathbf{x})$$

Classification & Systematics Correction



1. Covariate shift:

$$p_{\text{tr}}(z|c) = p_{\text{tg}}(z|c) \text{ \& } p_{\text{tr}}(c) \neq p_{\text{tg}}(c)$$

2. Prior Probability shift:

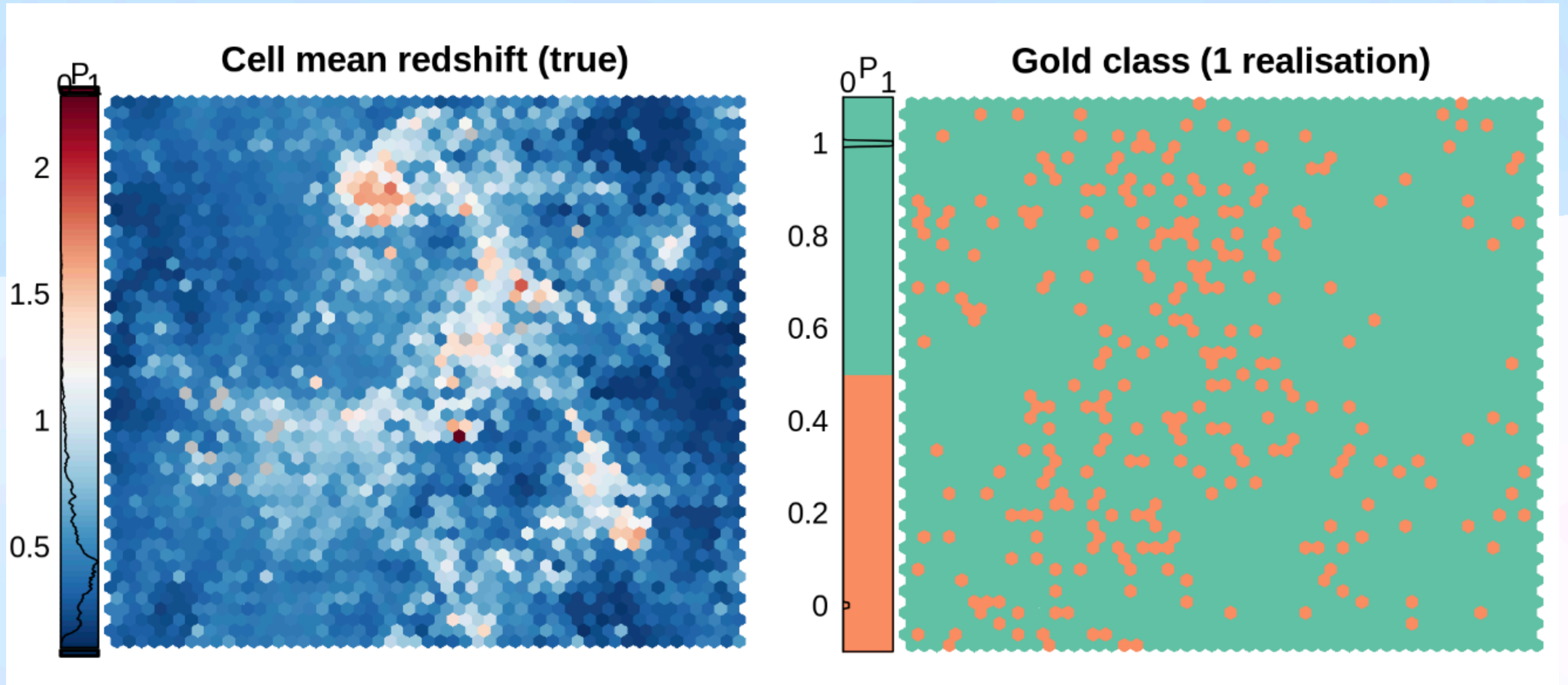
$$p_{\text{tr}}(c|z) = p_{\text{tg}}(c|z) \text{ \& } p_{\text{tr}}(z) \neq p_{\text{tg}}(z)$$

Methodological sensitivity in $N(z)$ calibration

Gold-weight vs Gold-class (Wright et al. 2025a)

$$p_t(z|\mathbf{x})p_t(\mathbf{x}) \neq p_v(z|\mathbf{x})p_v(\mathbf{x})$$

Classification & Systematics Correction

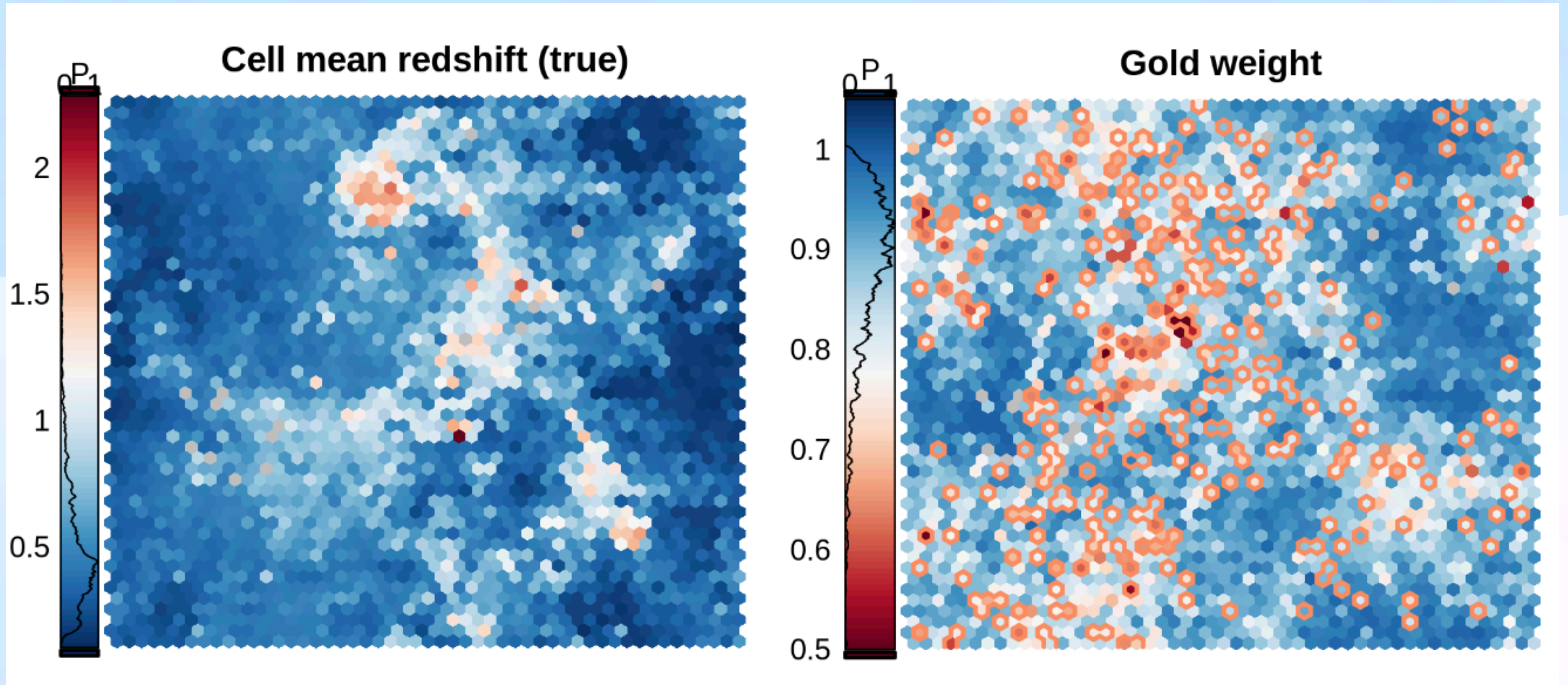


Methodological sensitivity in $N(z)$ calibration

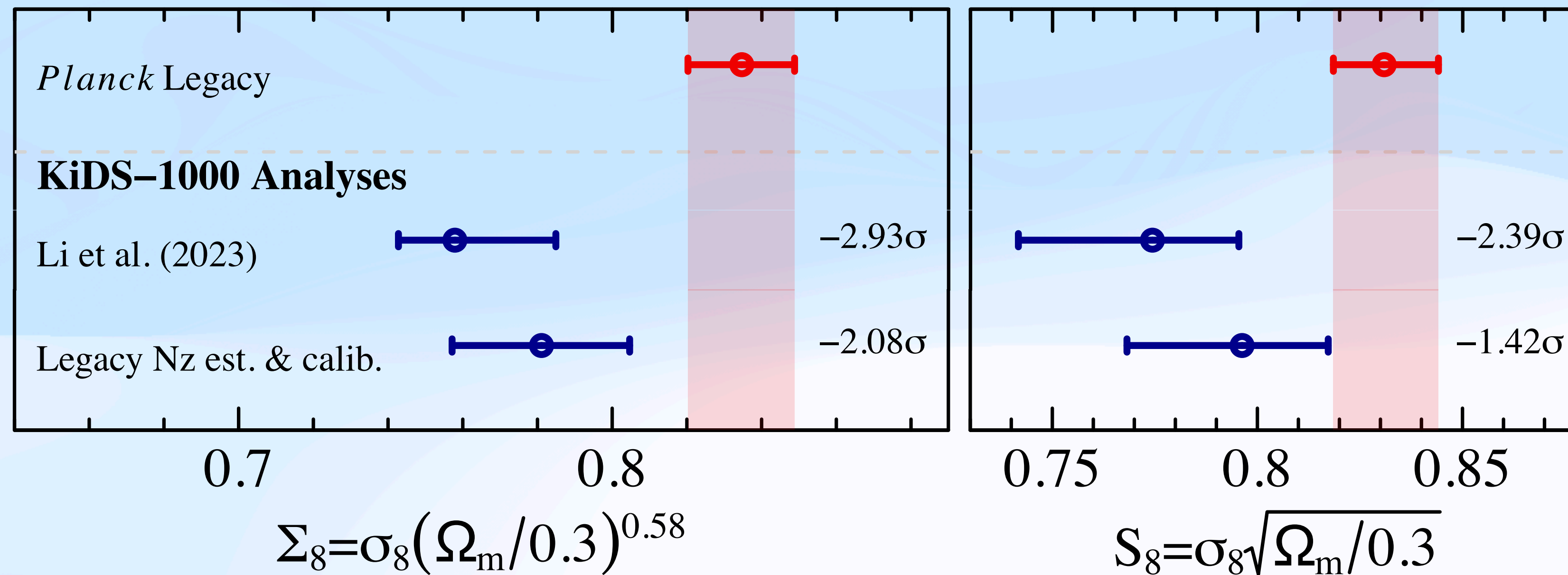
Gold-weight vs Gold-class (Wright et al. 2025a)

$$p_t(z|\mathbf{x})p_t(\mathbf{x}) \neq p_v(z|\mathbf{x})p_v(\mathbf{x})$$

Classification & Systematics Correction



Nz bias & covariate shift correction



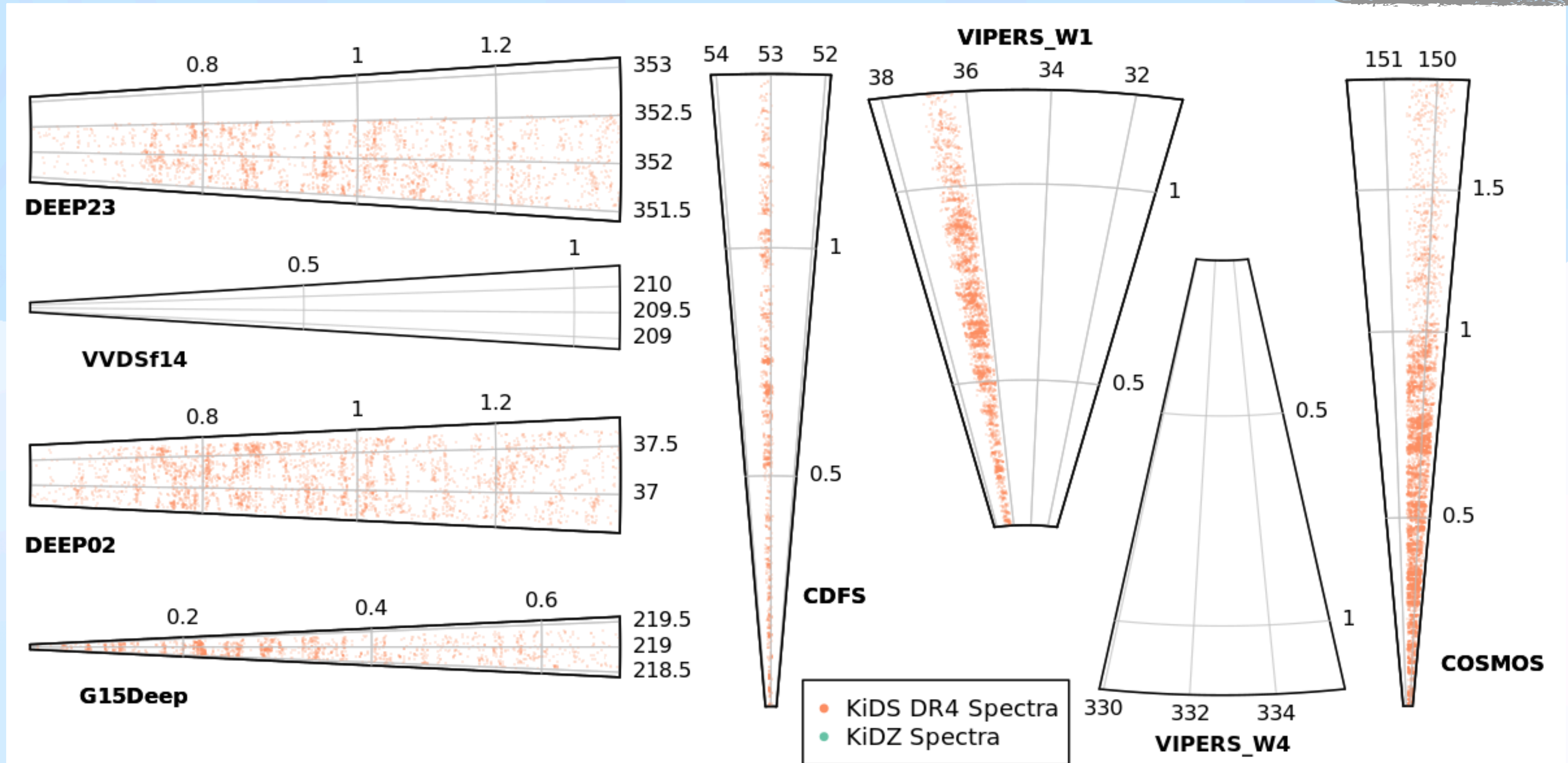
$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$

Classification & Systematics Correction

KiDZ Redshift Calibration Sample

Robust corrections with more spectra (Wright et al. 2024)

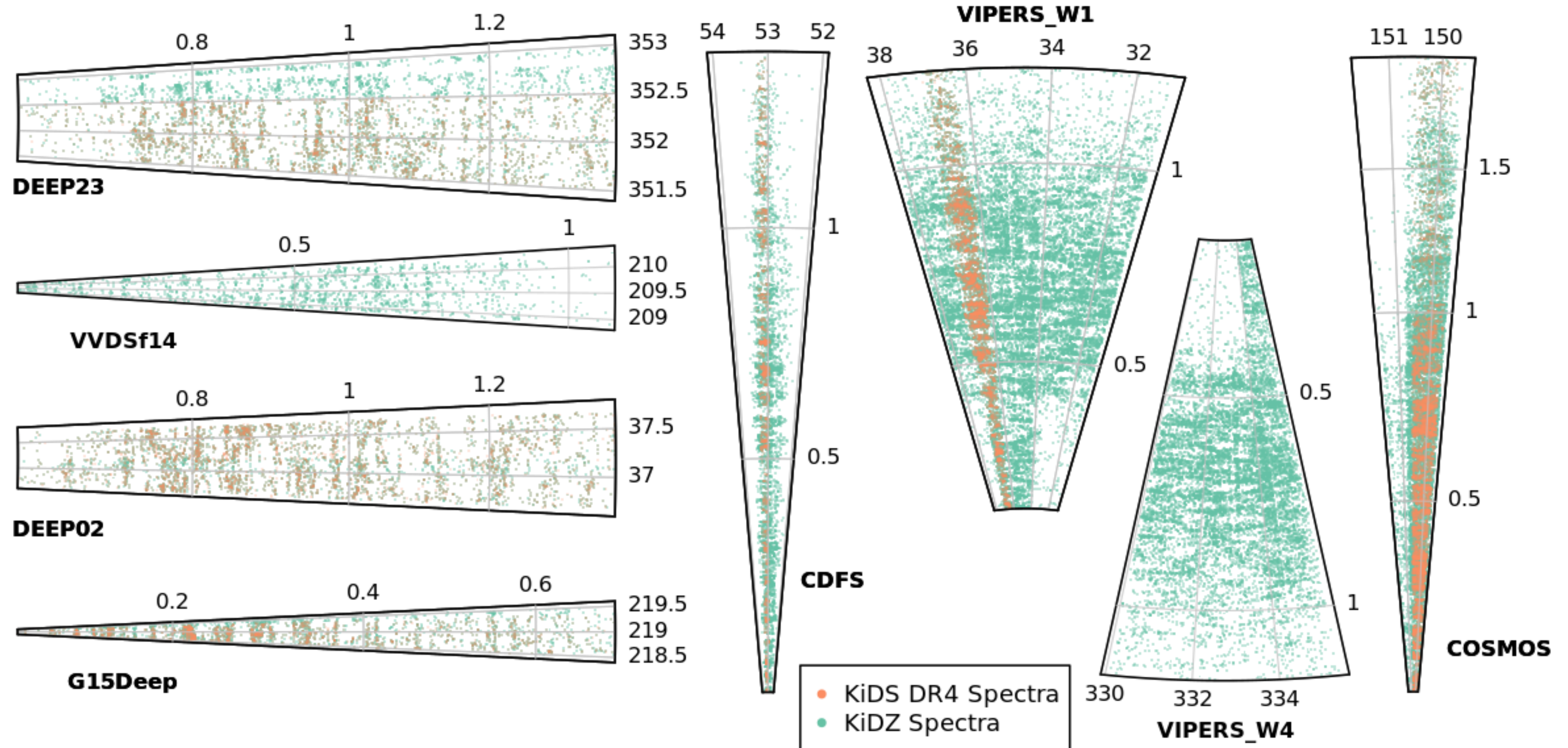
$p_t(z|\mathbf{x})p_t(\mathbf{x}) \neq p_v(z|\mathbf{x})p_v(\mathbf{x})$
Classification & Systematics Correction



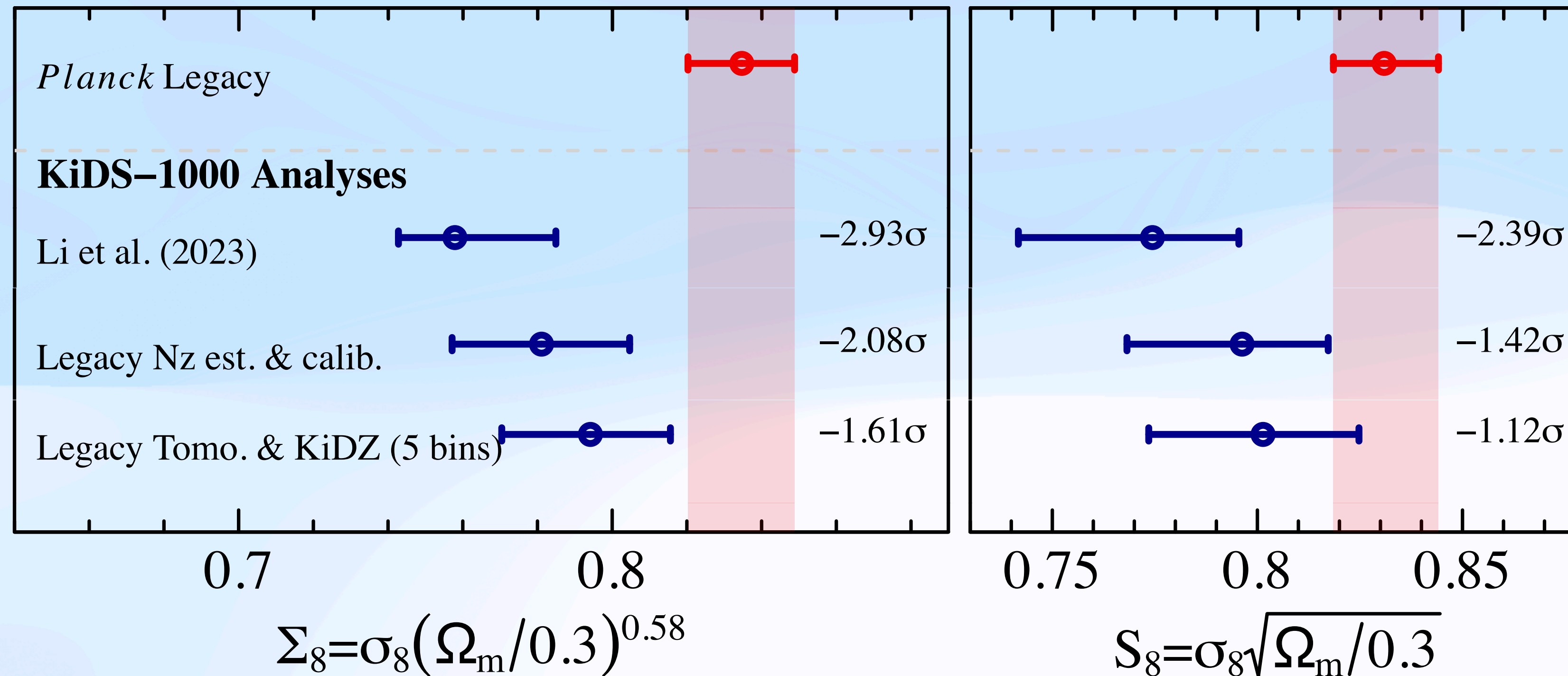
KiDZ Redshift Calibration Sample

Robust corrections with more spectra (Wright et al. 2024)

$p_t(z|\mathbf{x})p_t(\mathbf{x}) \neq p_v(z|\mathbf{x})p_v(\mathbf{x})$
Classification & Systematics Correction



Nz bias & covariate shift correction



$p_t(z | \mathbf{x}) p_t(\mathbf{x}) \neq p_v(z | \mathbf{x}) p_v(\mathbf{x})$
Classification & Systematics Correction

$$f_{\theta}(\phi) \approx \mathcal{M}(\phi)$$

Model Emulation

Used primarily in parameter inference. Symbolic regression emulators are more broad, but remove concerns of a black box.
For cosmology: emulating the matter power spectrum.

Recent & notable examples:

- Arico et al (2021,2023)
- Spurio-Mancini et al. (2022)
- Burger et al. (2025)

Example talks this week:

- Alicia Martin (Thurs 1420)
- Diane Salim (Thurs 1540)

Power spectrum emulation for MCMC

Embedded at the heart of the optimisation process

$$f_{\theta}(\phi) \approx \mathcal{M}(\phi)$$

Model Emulation

$$\xi_{+}(\theta) = \int_0^{\infty} \frac{d\ell \ell}{2\pi} J_0(\ell\theta) P_{\kappa}(\ell)$$

$$\xi_{-}(\theta) = \int_0^{\infty} \frac{d\ell \ell}{2\pi} J_4(\ell\theta) P_{\kappa}(\ell)$$

$$P_{\kappa}(\ell) = \frac{9H_0^4 \Omega_m^2}{4c^4} \int_0^{\chi_h} d\chi \frac{g^2(\chi)}{a^2(\chi)} P_{\delta} \left(\frac{\ell}{f_K(\chi)}, \chi \right)$$

$$g(\chi) = \int_{\chi}^{\chi_h} d\chi' p_{\chi}(\chi') \frac{f_K(\chi' - \chi)}{f_K(\chi')}$$

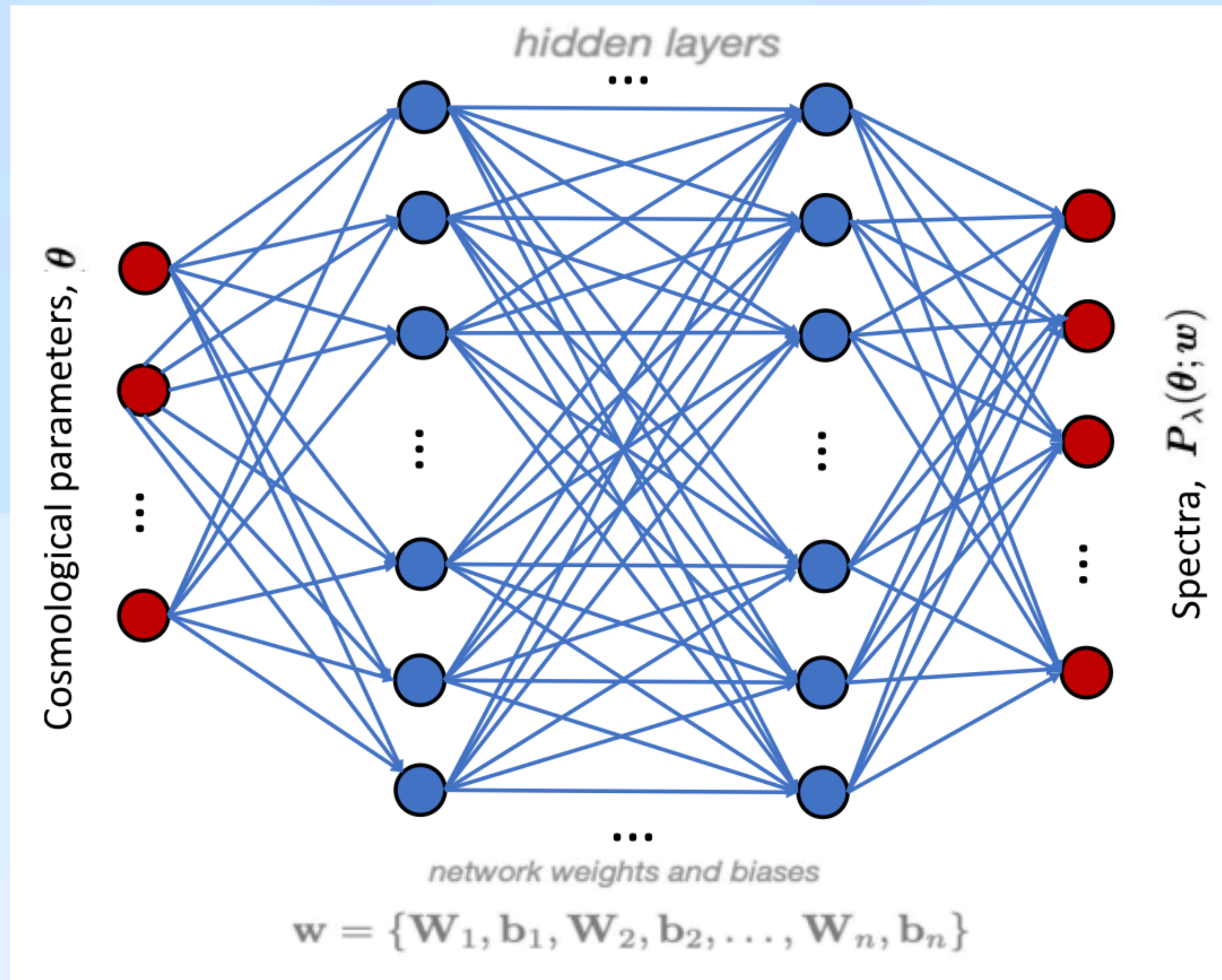
Model construction

- The problem: Boltzmann codes are slow, but crucial
- The idea: replace the Boltzmann solver with an emulator
- The concern: is it accurate enough? Does it create bias?

Power spectrum emulation for MCMC

$$f_{\theta}(\phi) \approx \mathcal{M}(\phi)$$

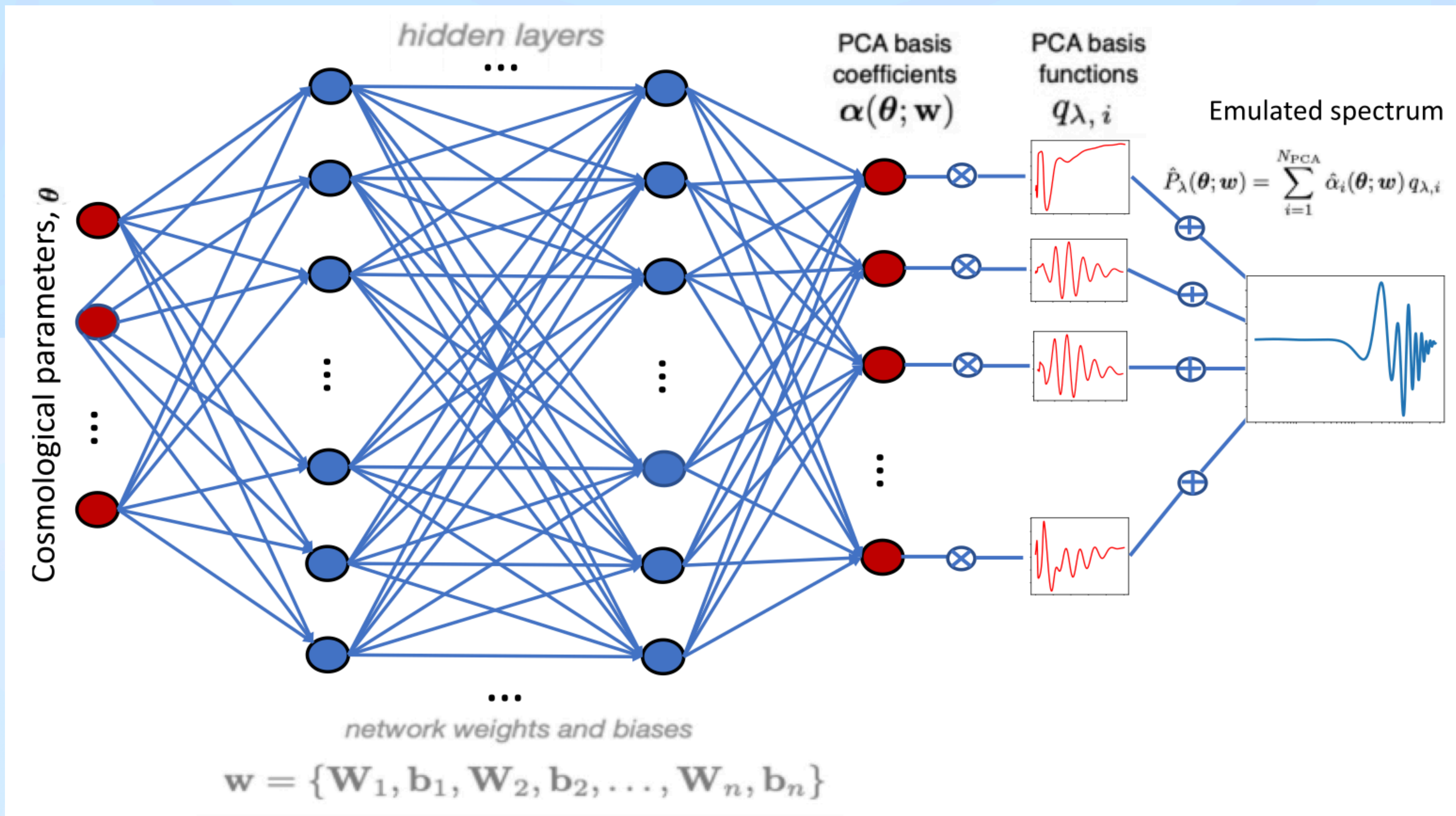
Model Emulation



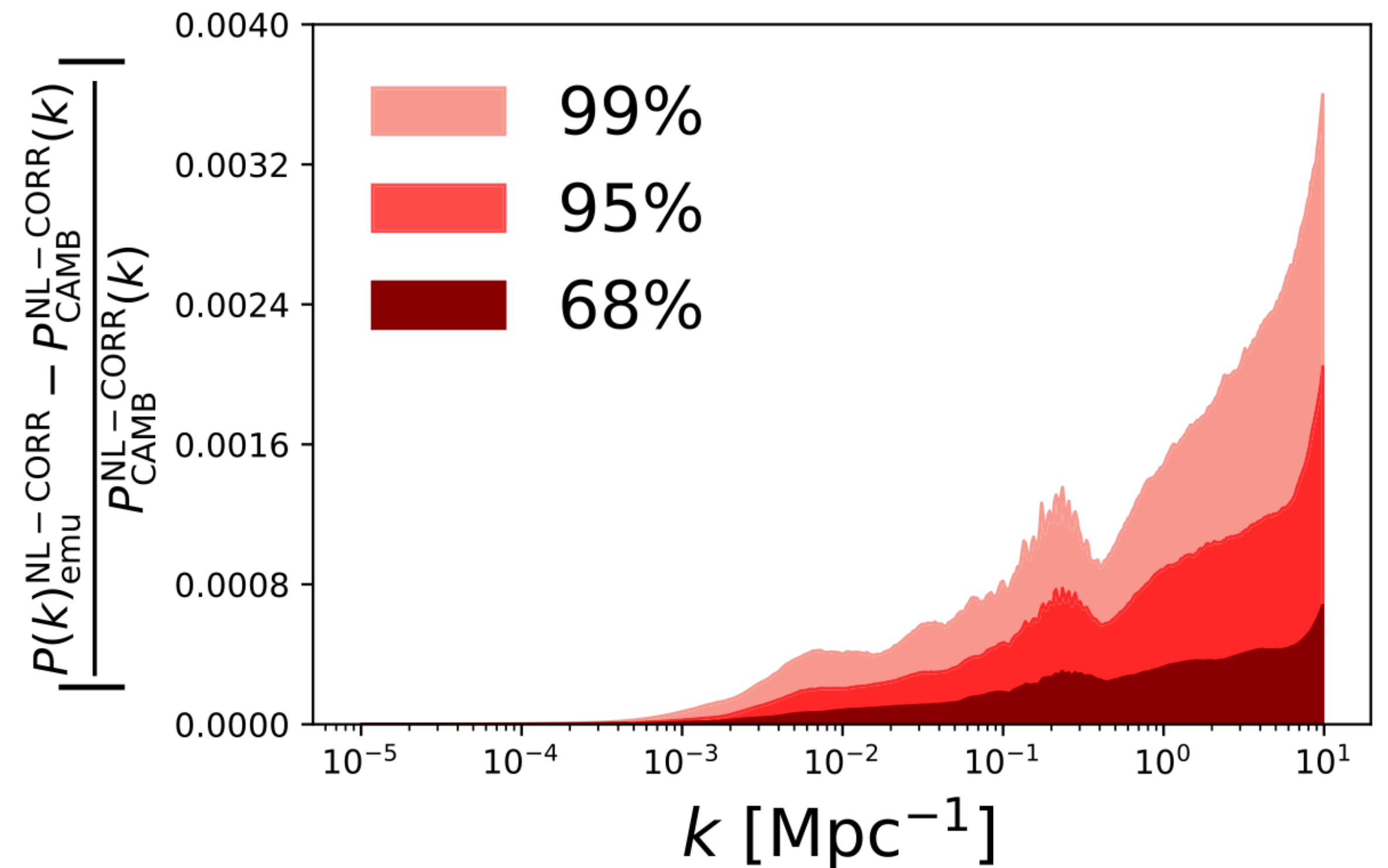
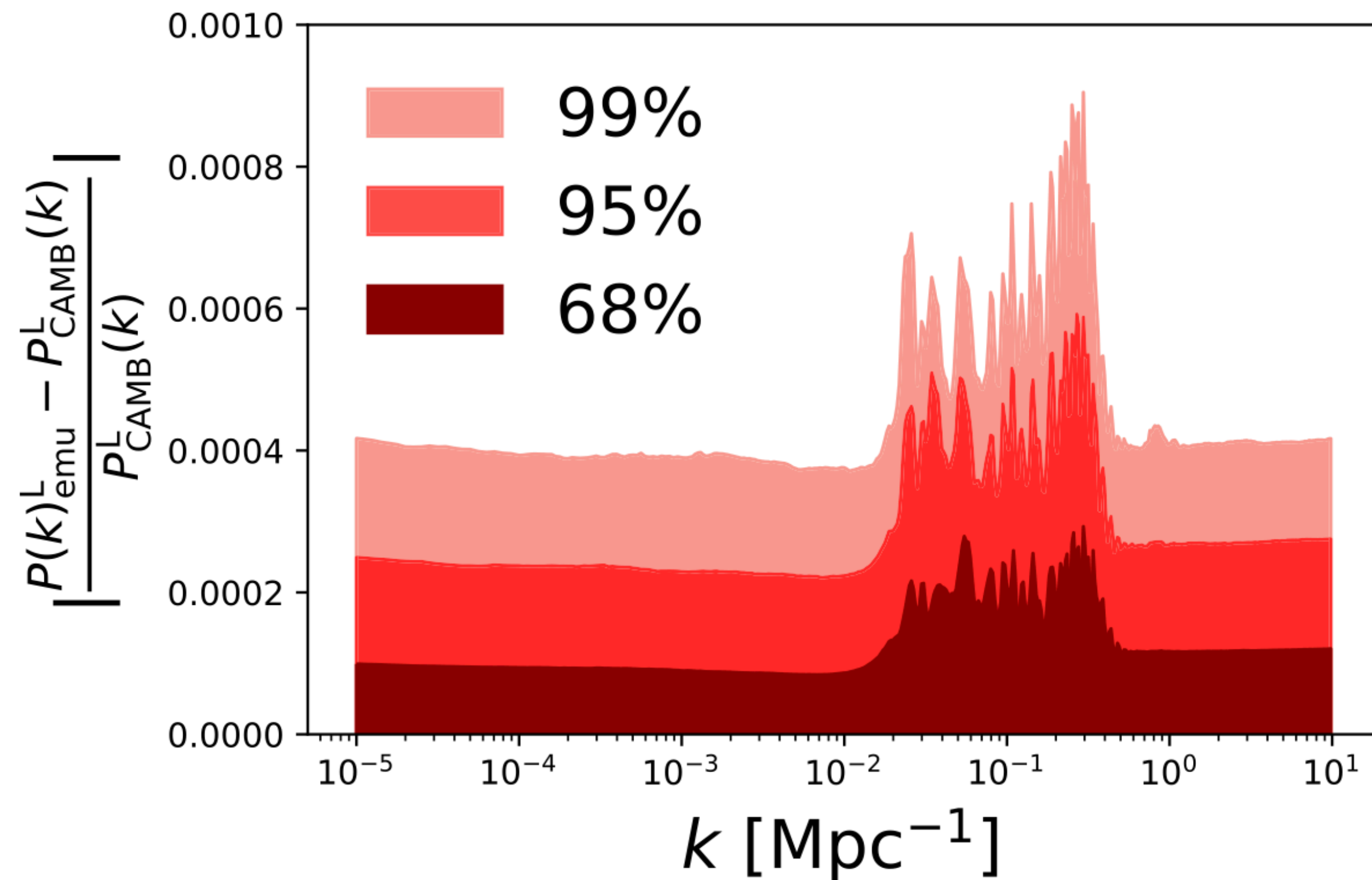
Power spectrum emulation for MCMC

$$f_{\theta}(\phi) \approx \mathcal{M}(\phi)$$

Model Emulation



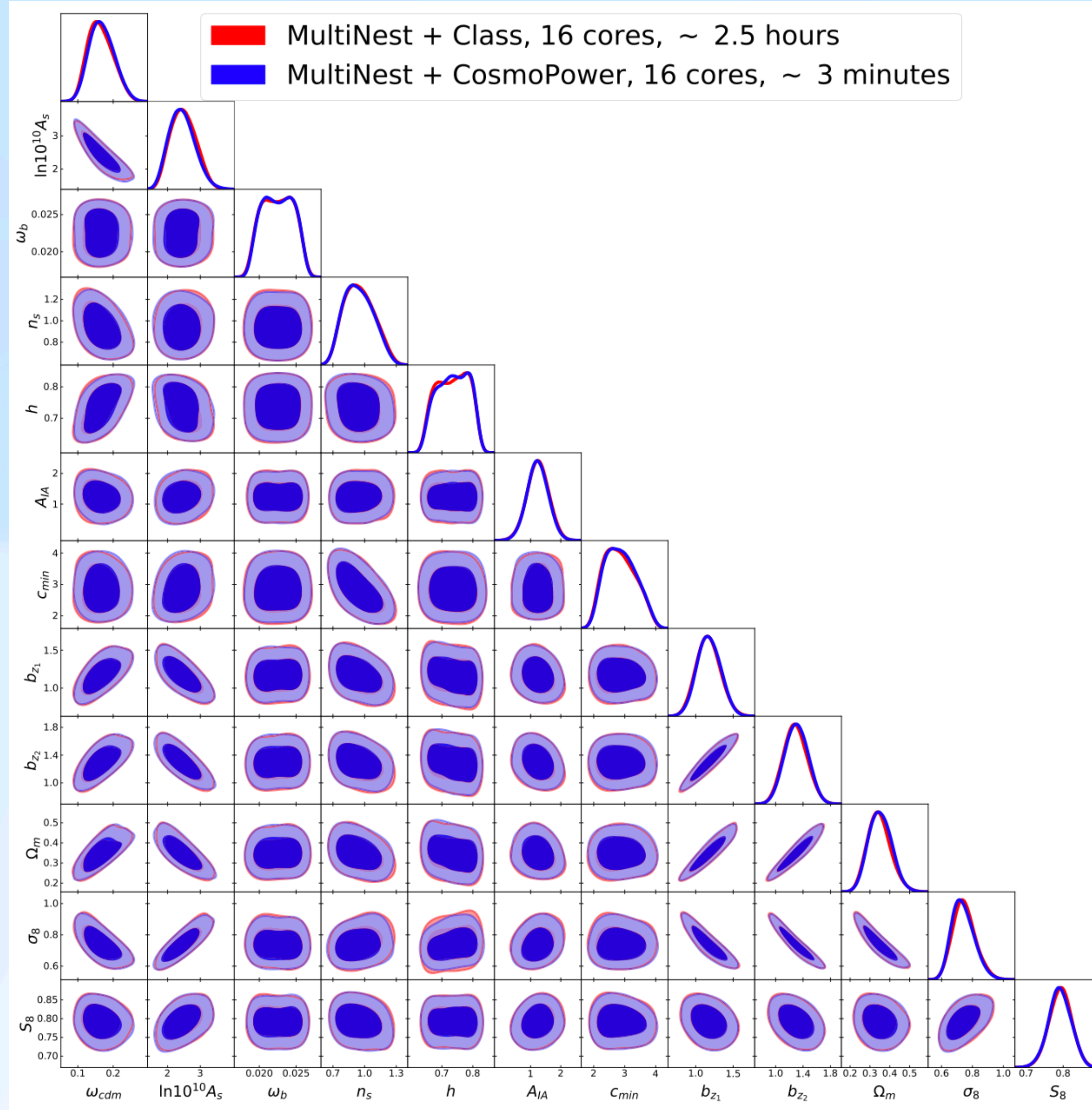
Power spectrum emulation for MCMC



$$f_{\theta}(\phi) \approx \mathcal{M}(\phi)$$

Model Emulation

Power spectrum emulation for MCMC



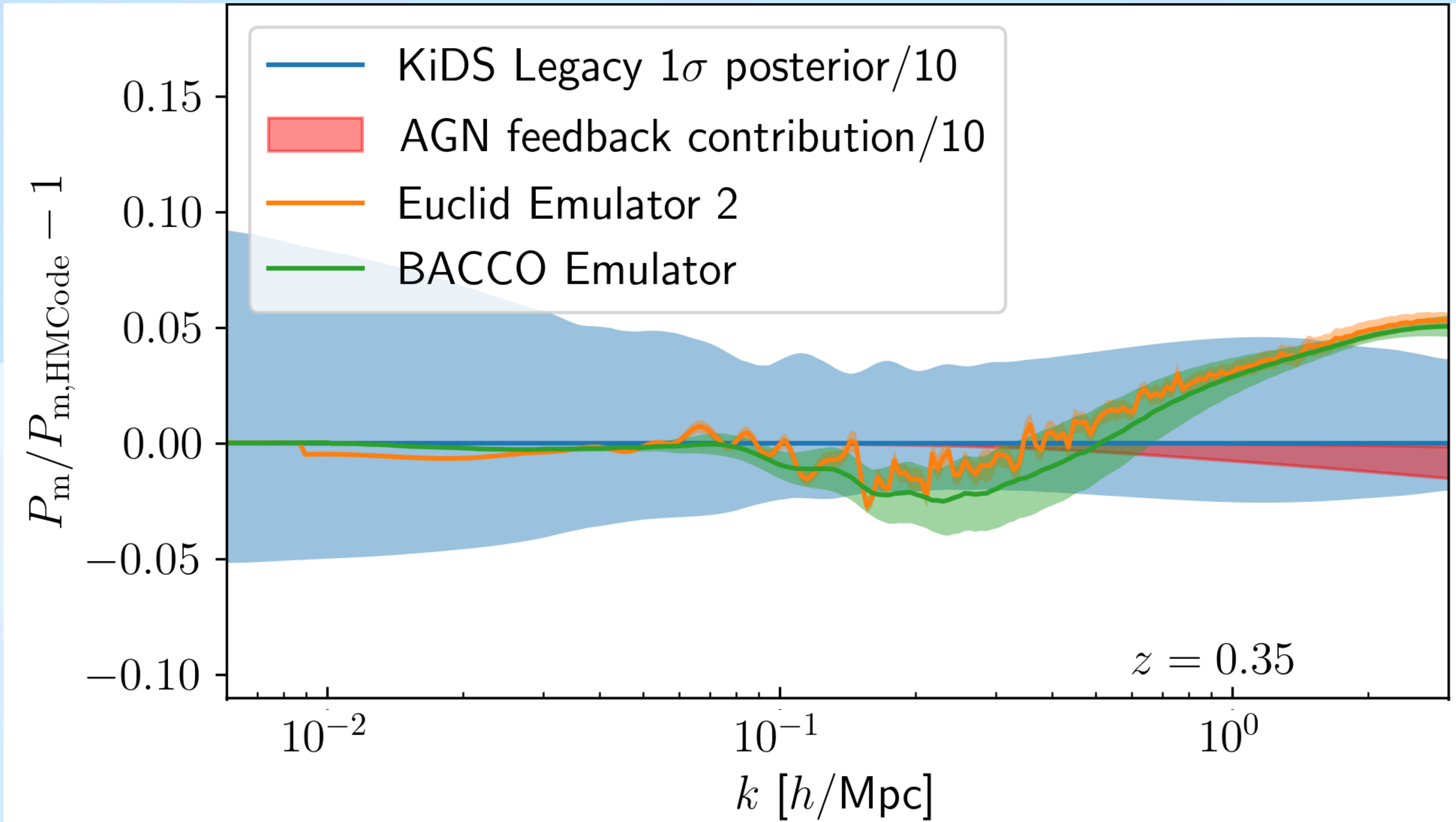
Spurio-Mancini et al. (2022)

$$f_{\theta}(\phi) \approx \mathcal{M}(\phi)$$

Model Emulation

Power spectrum emulation for MCMC

Accuracy of the emulator itself is now a sub-dominant problem



Wright et al. (2025)

$$f_{\theta}(\phi) \approx \mathcal{M}(\phi)$$

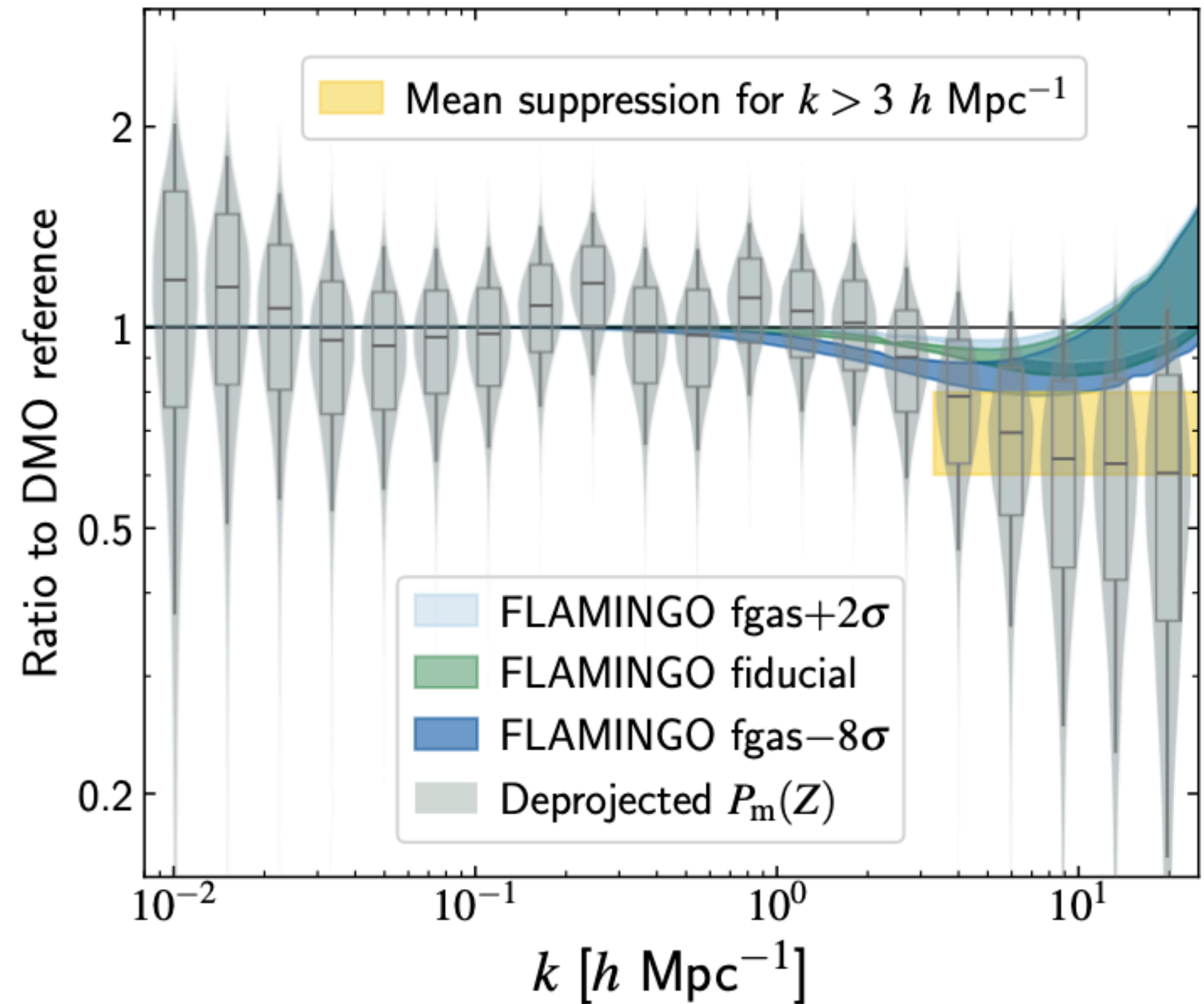
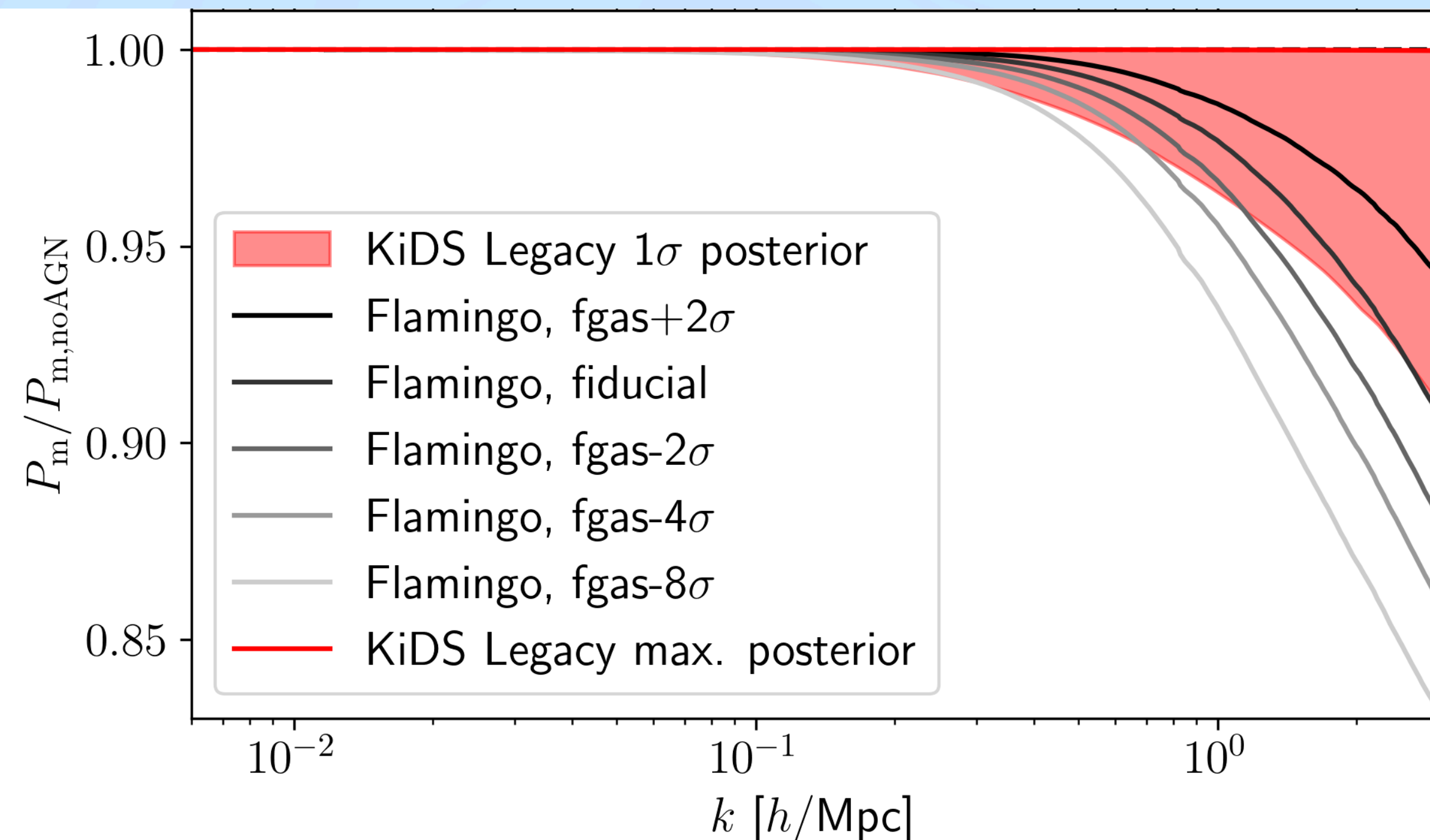
Model Emulation

Power spectrum emulation for MCMC

Accuracy of the emulator itself is a sub-dominant problem

$$f_{\theta}(\phi) \approx \mathcal{M}(\phi)$$

Model Emulation



$$p(\phi \mid \mathbf{x}) \approx q_{\hat{\theta}}(\phi \mid \mathbf{x})$$

Simulation Based Inference

Used for analysis of non-standard or complex data summaries, where likelihoods are non-analytic or intractable.

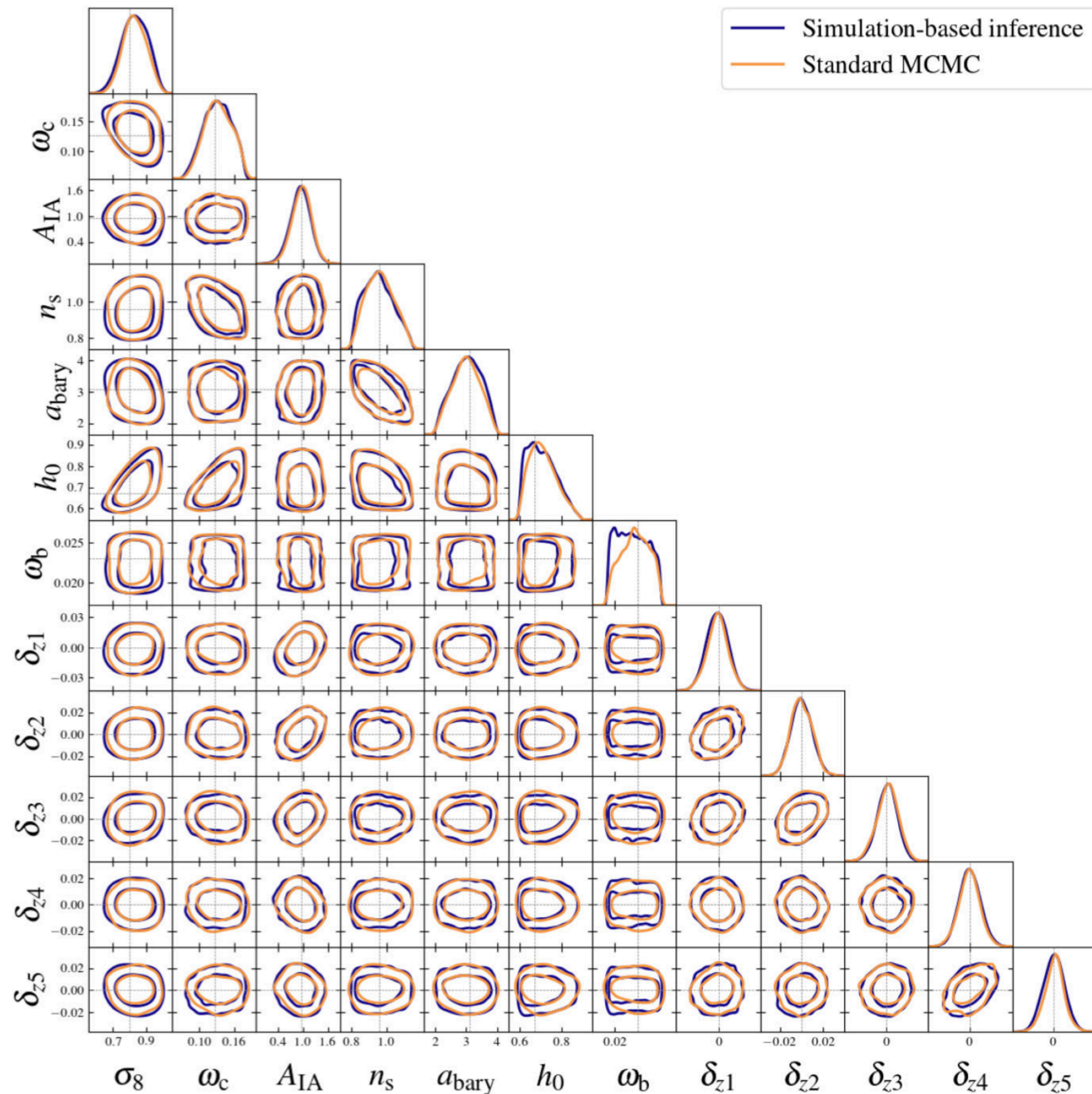
Recent & notable examples:

- Weitersheim-Kramsta et al (2024)
- Vinciguerra et al (2026)
- Thompson et al (2026)

Example talks this week:

- Gosia Suidek (Mon 0950)
- Nicola Napolitano (Wed 0900)
- Mali Land-Strykowski (Thurs 110)
- Oliver Oayda (Thurs 1150)

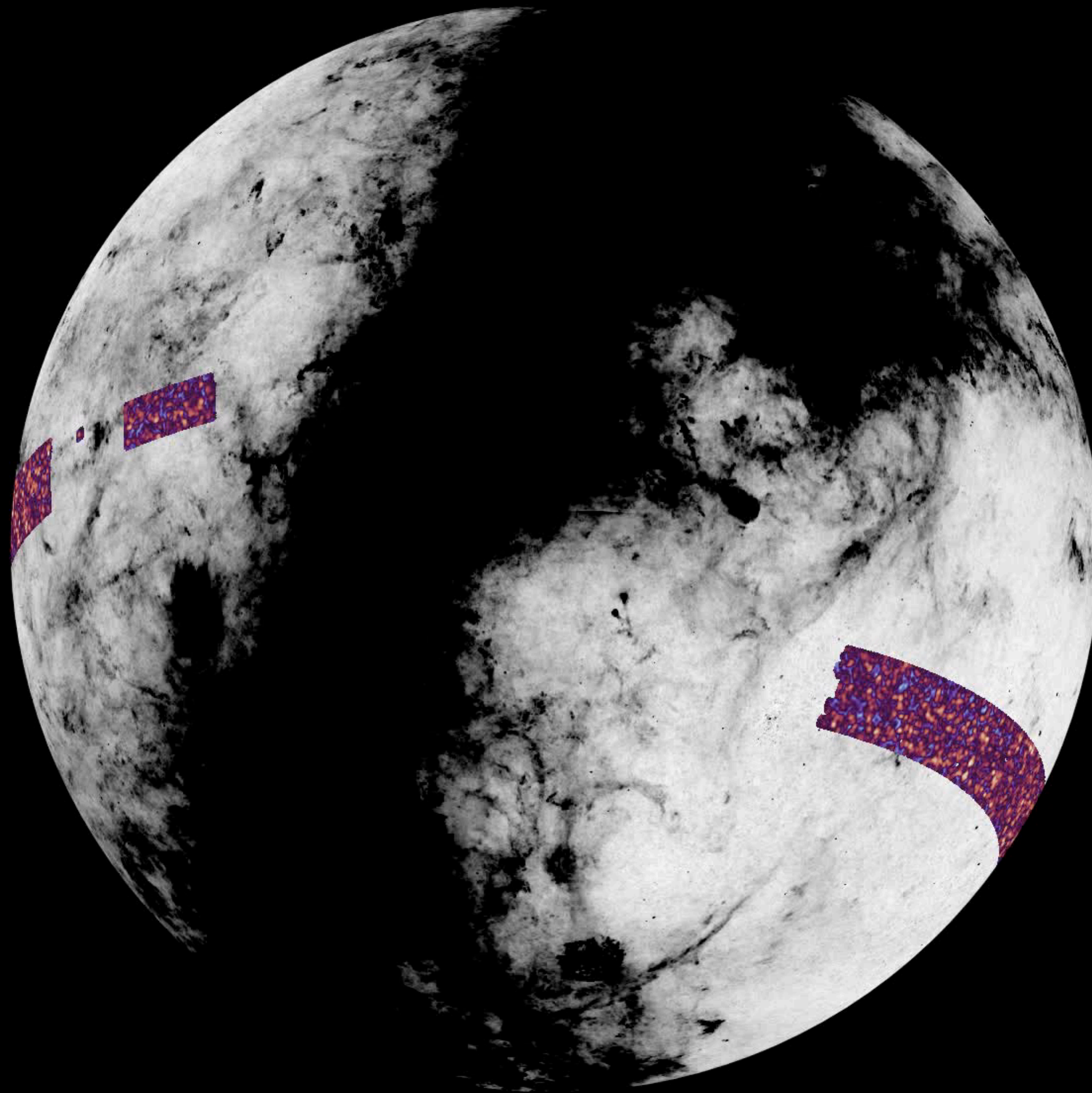
Simulation-based inference: 2pt validation



- Cosmic shear analysis of KiDS-1000 using MCMC and SBI
- Equivalent posteriors showcase consistency between methods
- Utility is in the application to non-standard summaries

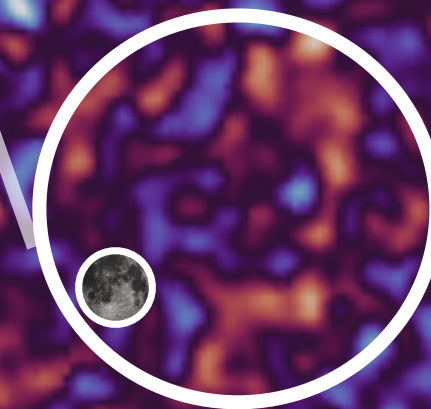
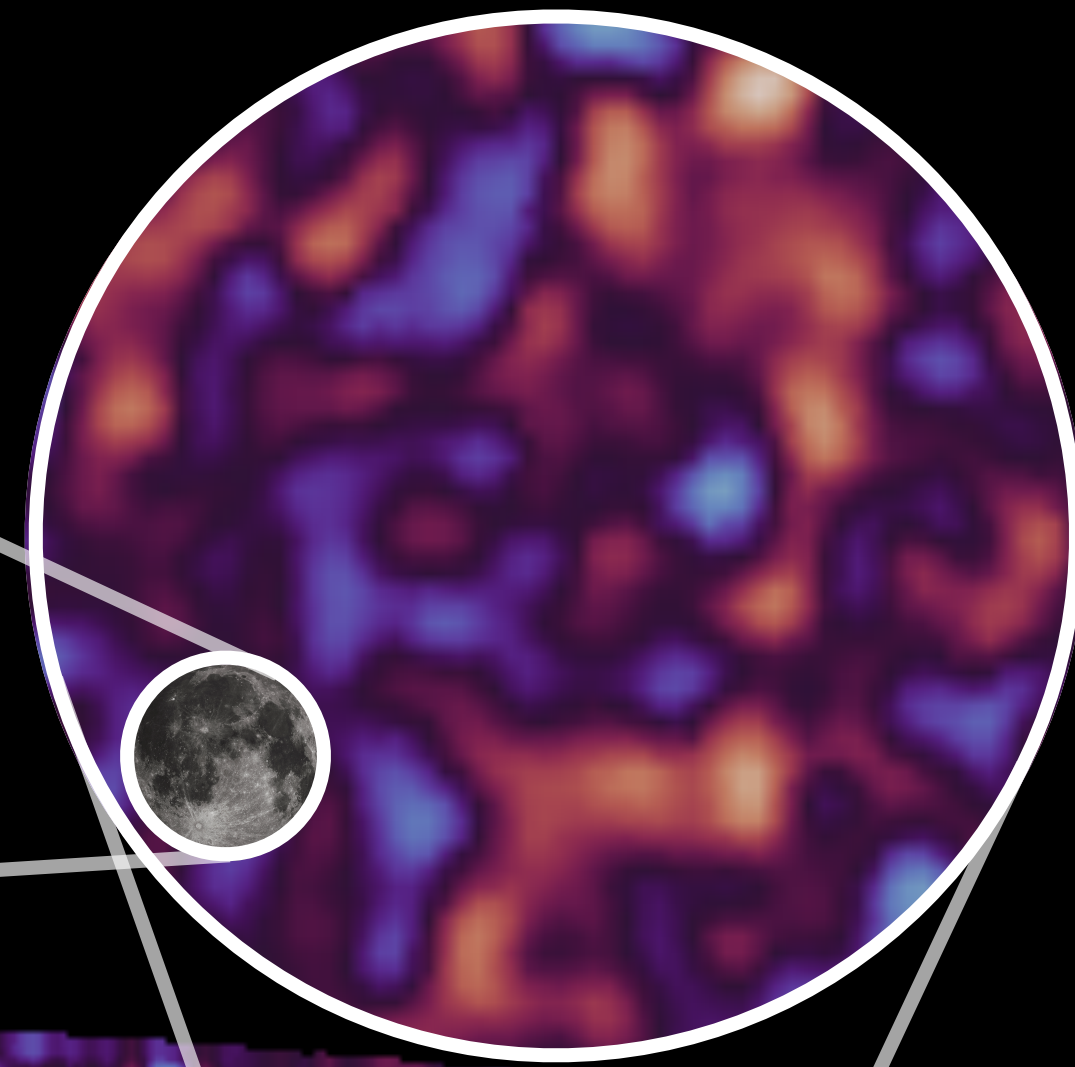
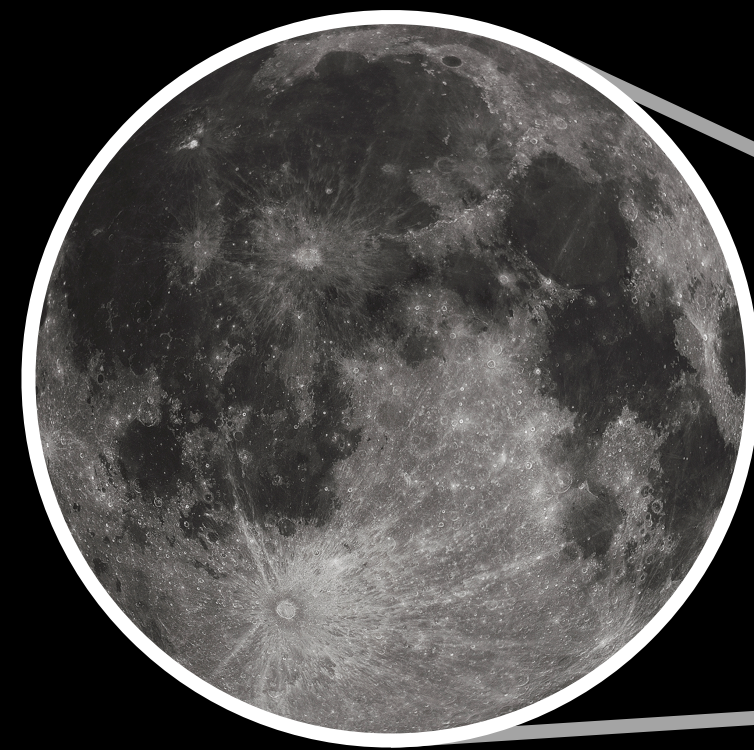
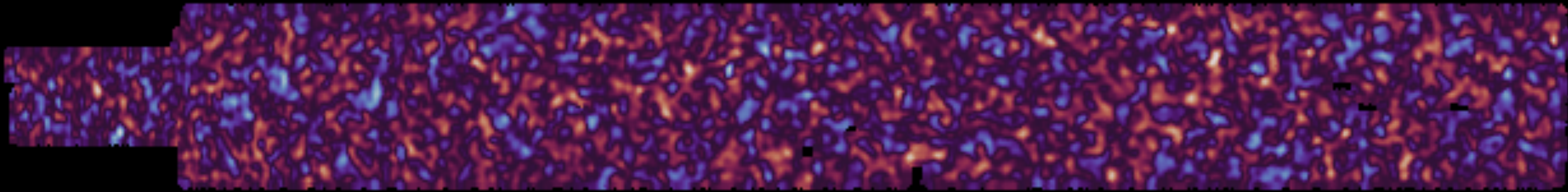
$$p(\phi \mid \mathbf{x}) \approx q_{\hat{\theta}}(\phi \mid \mathbf{x})$$

Simulation Based Inference

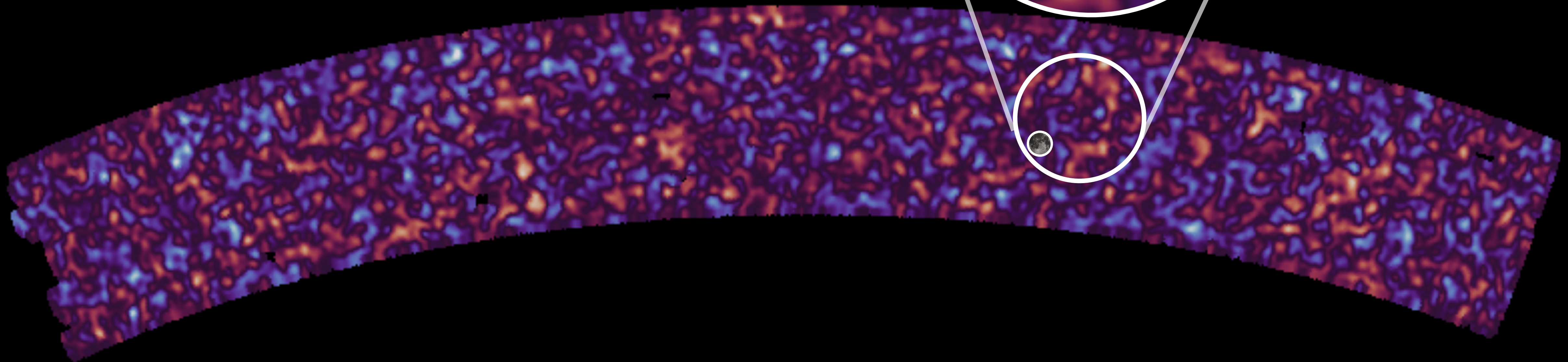


$$p(\phi \mid \mathbf{x}) \approx q_{\hat{\theta}}(\phi \mid \mathbf{x})$$

*Simulation Based
Inference*



$p(\phi \mid \mathbf{x}) \approx$
 $q_{\hat{\theta}}(\phi \mid \mathbf{x})$
*Simulation Based
Inference*



A final thought

Where is ML in cosmology heading?

Development of complete analysis workflows

📖 README 📄 GPL-3.0 license



CosmoPipe

CosmoPipe is a modular, configuration-driven pipeline for cosmic-shear data analysis and cosmological inference. It supports the path from survey catalogues through calibration, tomographic measurements, covariance modelling, and likelihood sampling with [CosmoSIS](#).

Originally developed for KiDS cosmology analyses, CosmoPipe provides reusable processing functions and reference configurations for analyses including KiDS-1000 and Euclid-style data products.

⚠ Warning

Cosmic-shear inference is sensitive to analysis choices. Review, validate, and document all configuration values, external inputs, and processing steps for your science case before interpreting results.

Where is ML in cosmology heading?

Development of complete analysis workflows

- Cosmic shear with KiDS Legacy was analysed with CosmoPipe
- CosmoPipe contains over 200 individual processing functions
- KiDS Legacy analysed more than 5000 individual MCMC chains in 2 years

The point?

- Each of the above describes a task; a single aspect of a cosmological analysis
- Modern LLM (and even local MLM!) models are capable of constructing workflows equally well as they can construct a single PF
- CosmoPipe without the Angus is within our grasp. We should act!