Forward-modelling galaxy surveys for precise cosmological measurements

Luca Tortorelli

Collaborators: D. Gruen, S. Fischbacher, A. Refregier, J. McCullough, T. Kacprzak, S. Bellstedt, A. Robotham and others

LSST DESC Dark Energy Science Collaboration

WAVES Wide Area Vista Extragalactic Survey

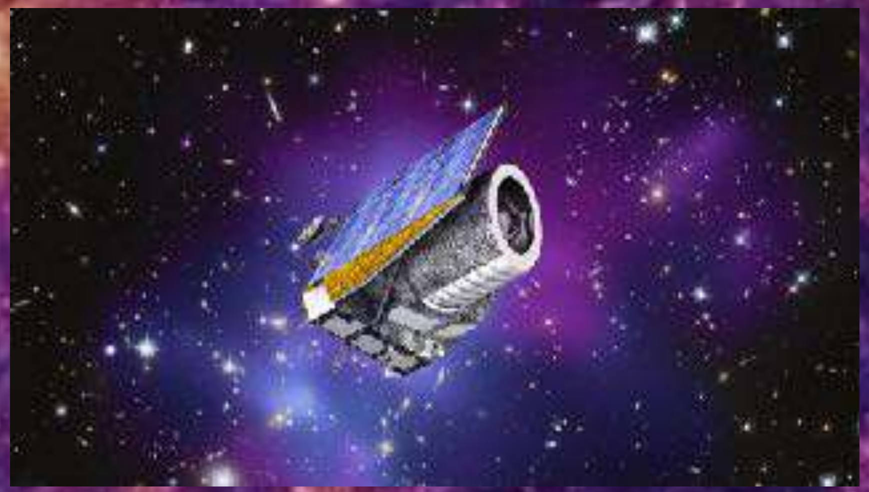
PAU

The golden era of Cosmology with wide-field galaxy surveys

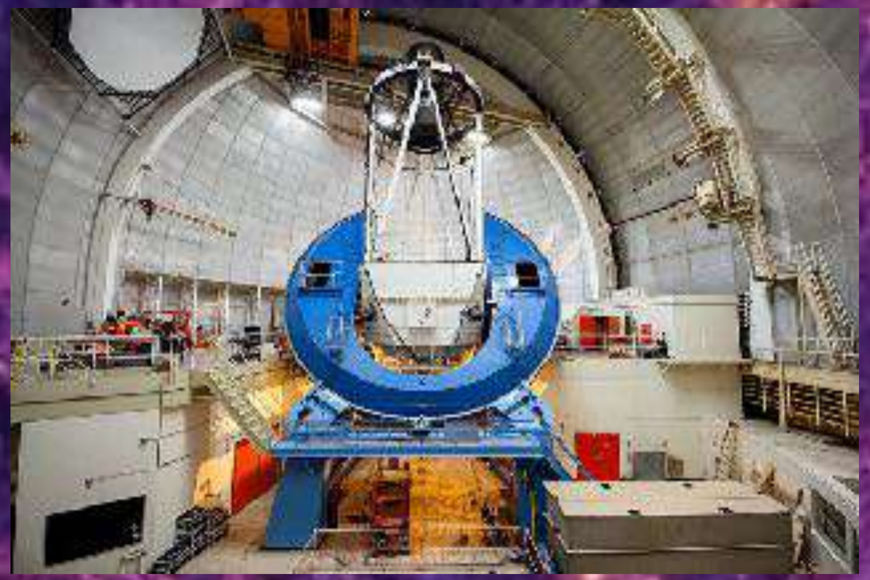
4MOST

Rubin/LSST

HSC



EUCLID

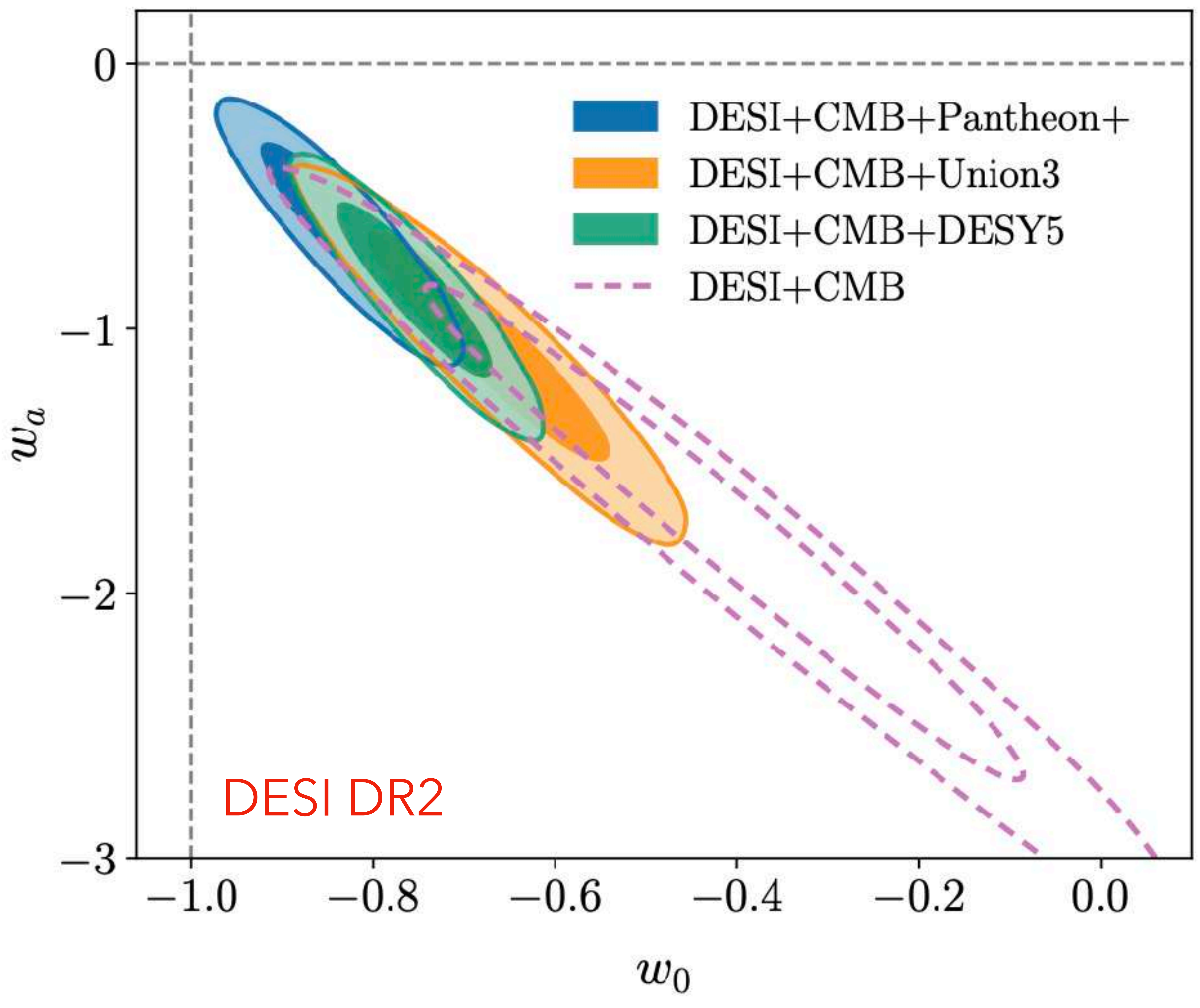


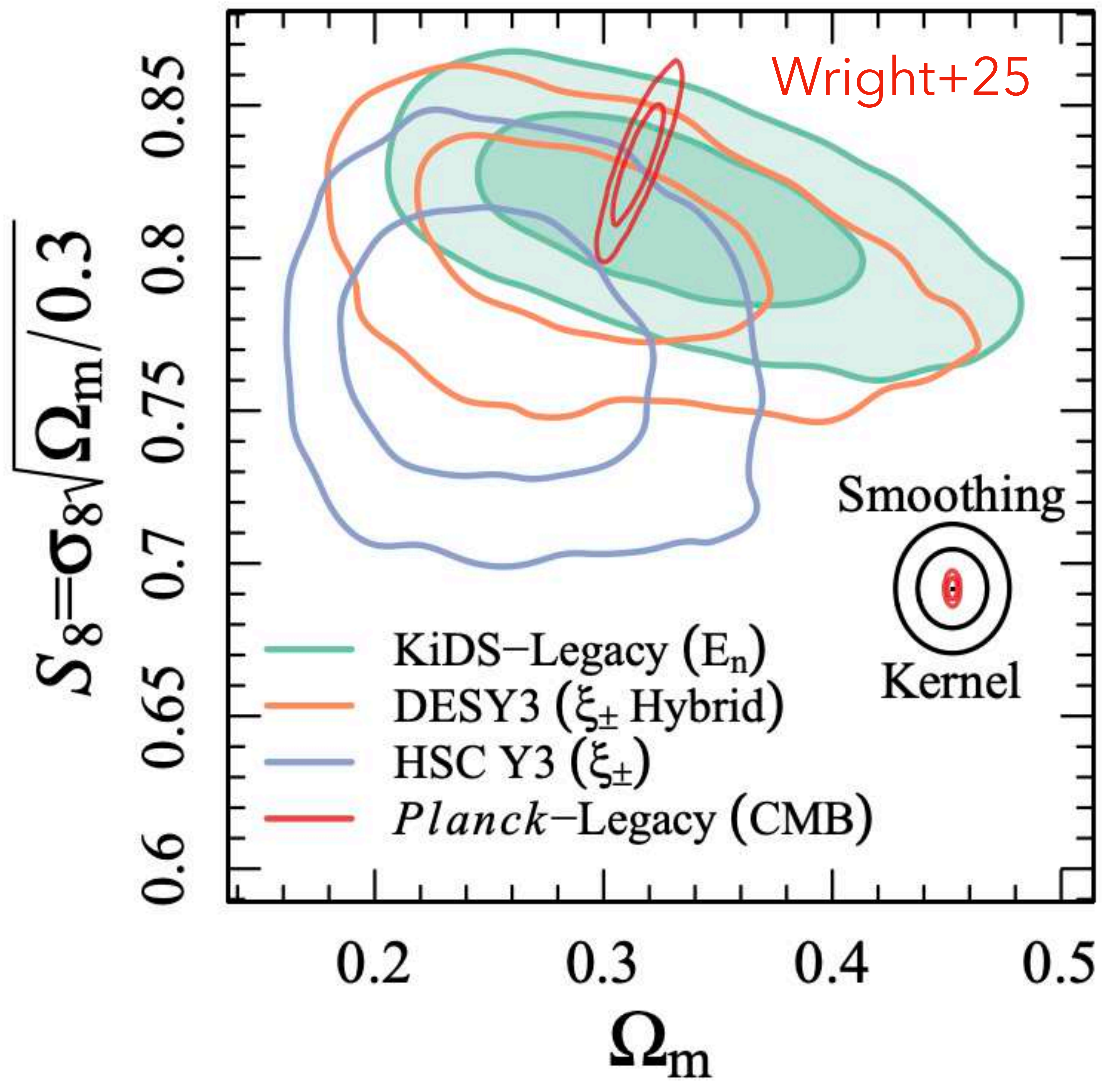
DESI

KiDS

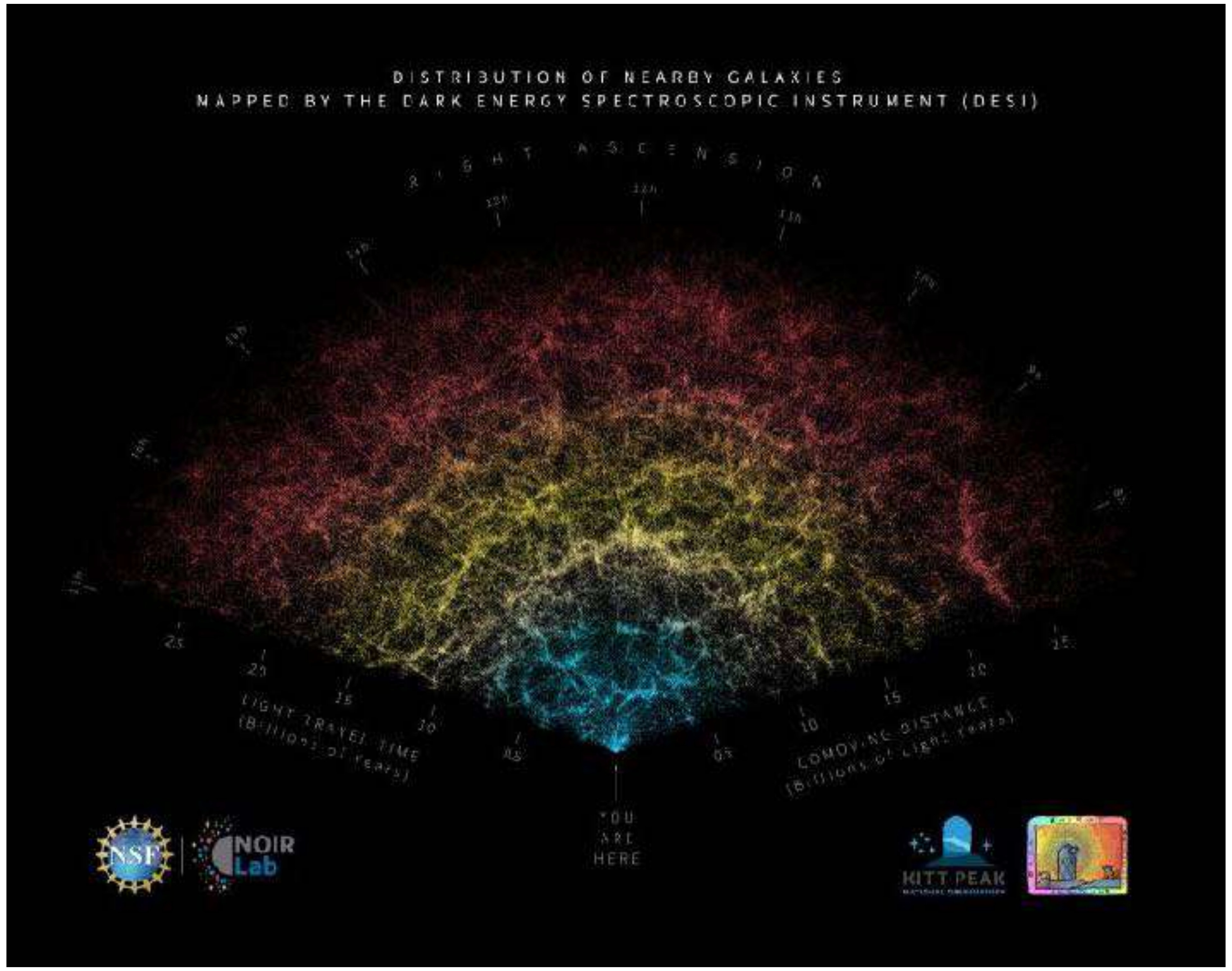
DES

Exciting new results in Cosmology



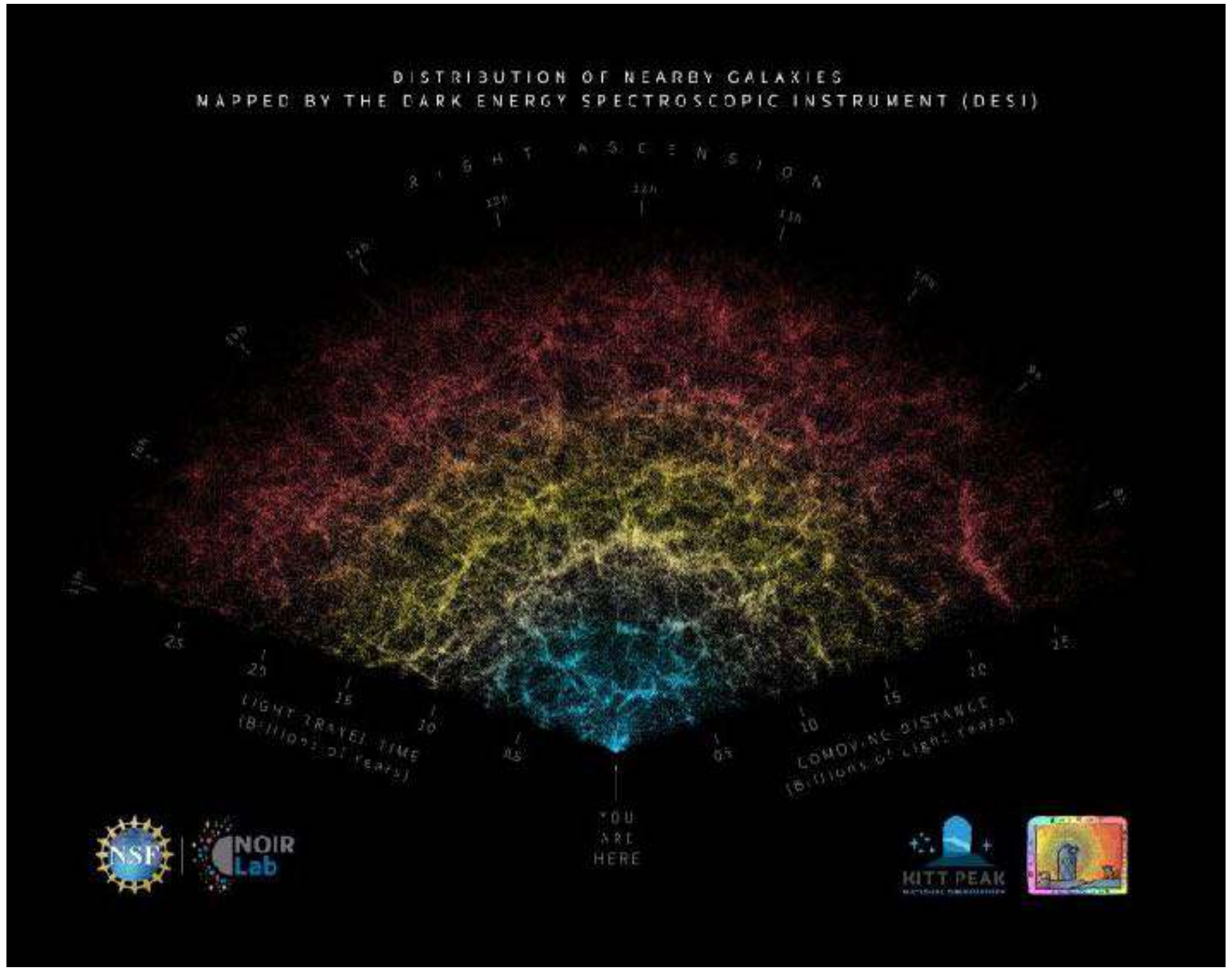


More are coming with LSS probes in Stage IV surveys



Galaxy clustering and **weak lensing** as two of the main probes to measure cosmological information with LSS.

With great (statistical) power comes great (systematics) responsibility



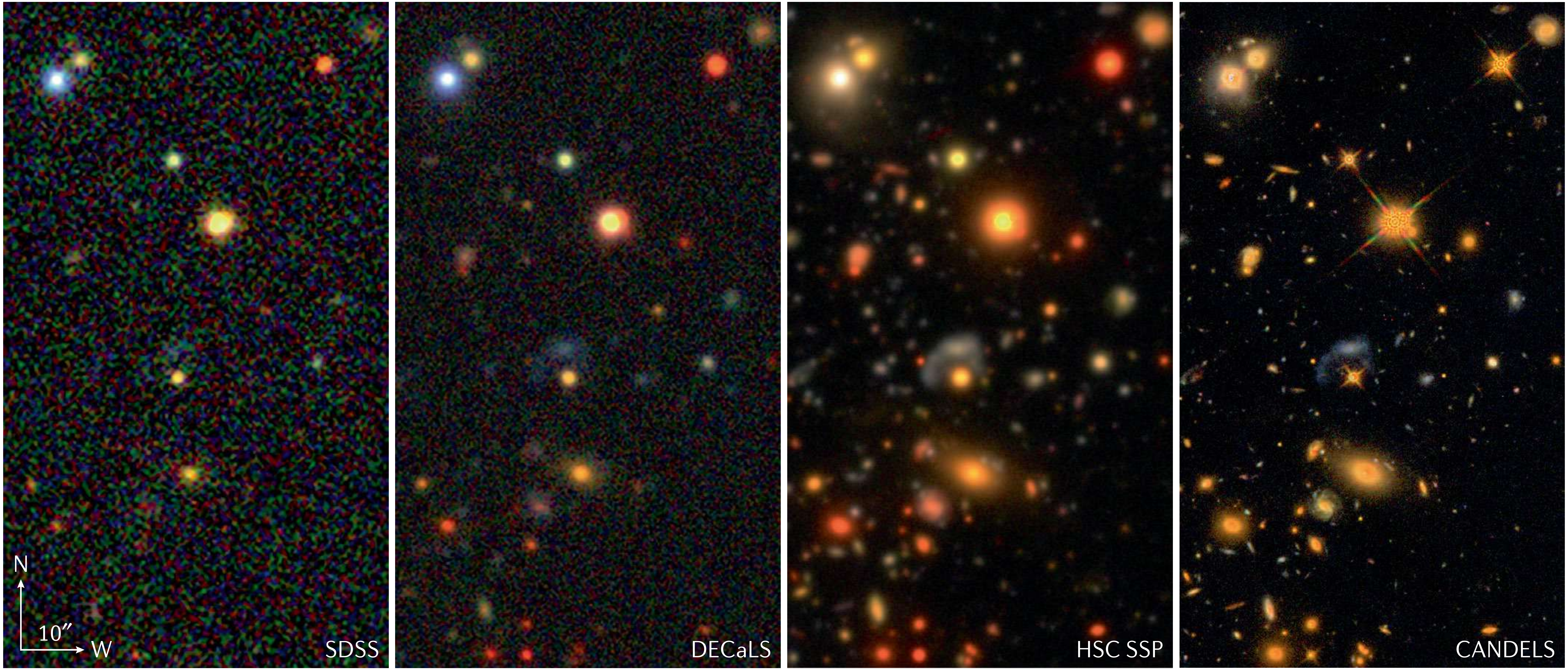
Accurate assessment of survey incompletenesses, selection effects on galaxies and biases on measurements of their properties to fully exploit these data.

Systematics in LSS measurements

Both probes are affected by a number of systematics, e.g. for weak lensing:

- Noise bias in shape estimation (e.g. Kacprzak+12,14).
- Modelling colour-dependent PSF (e.g. Paulin-Henriksson+08).
- Spatially varying survey properties (e.g. Rodriguez-Monroy+22).
- **Object blending (e.g. Sanchez+21).**
- Accurate treatment of intrinsic alignment (e.g. Fischbacher+23, McCullough+24).
- **Robust estimates of galaxy redshift distributions** (e.g. Crocce+16, Myles+21).
- Modelling of the non-linear power spectrum.

Blending in Stage IV surveys



Melchior+21

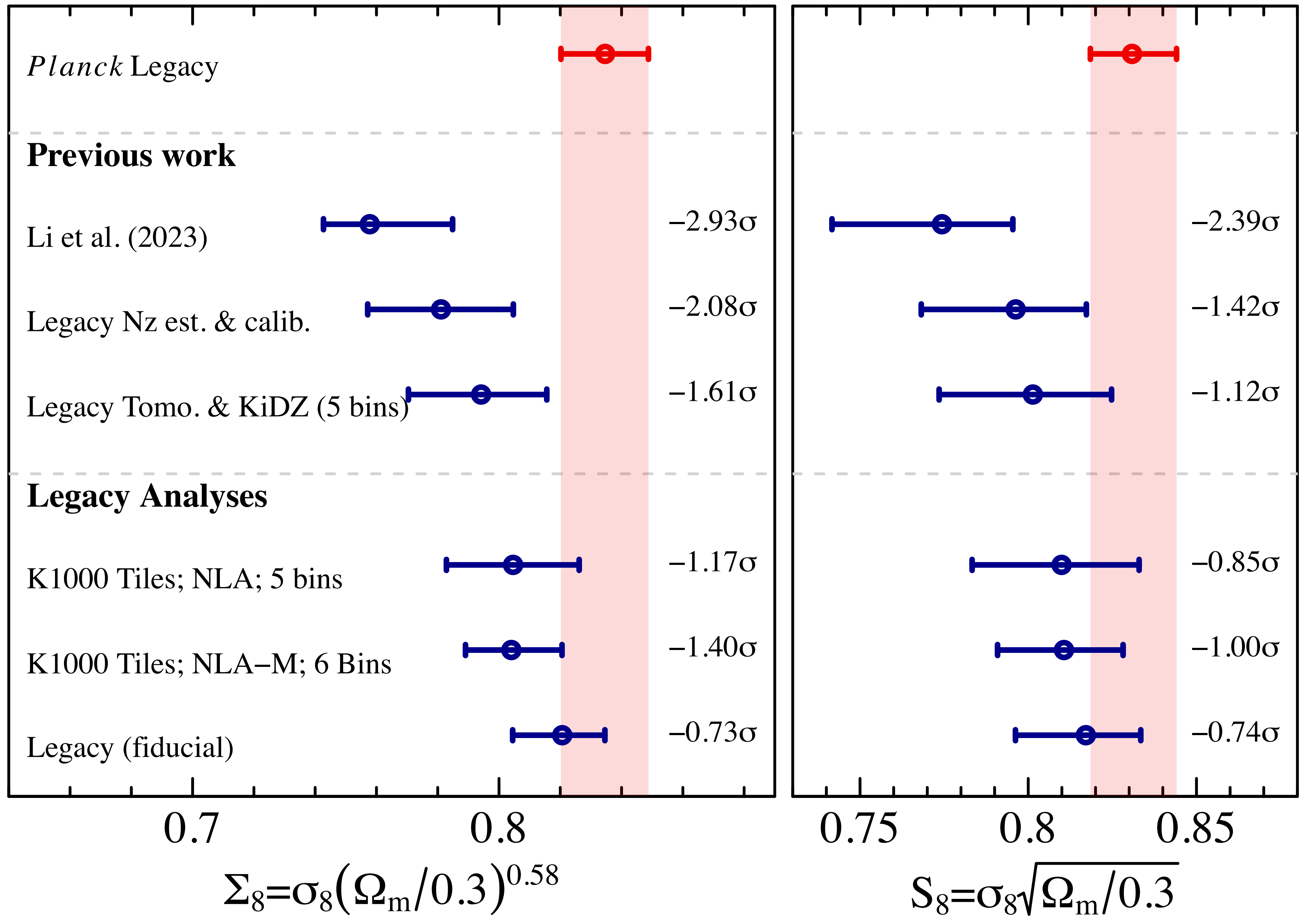
Blending in Stage IV surveys

Melchior+21

Requirements in tomographic n(z) uncertainties

Ω_M, σ_8 are highly sensitive to uncertainties in the n(z) of source galaxies, especially the mean redshift (e.g. Amon+22, van den Busch+22, Li+23, Dalal+23, Zhang+25).

Wright+25



shape measurement + conservative cuts

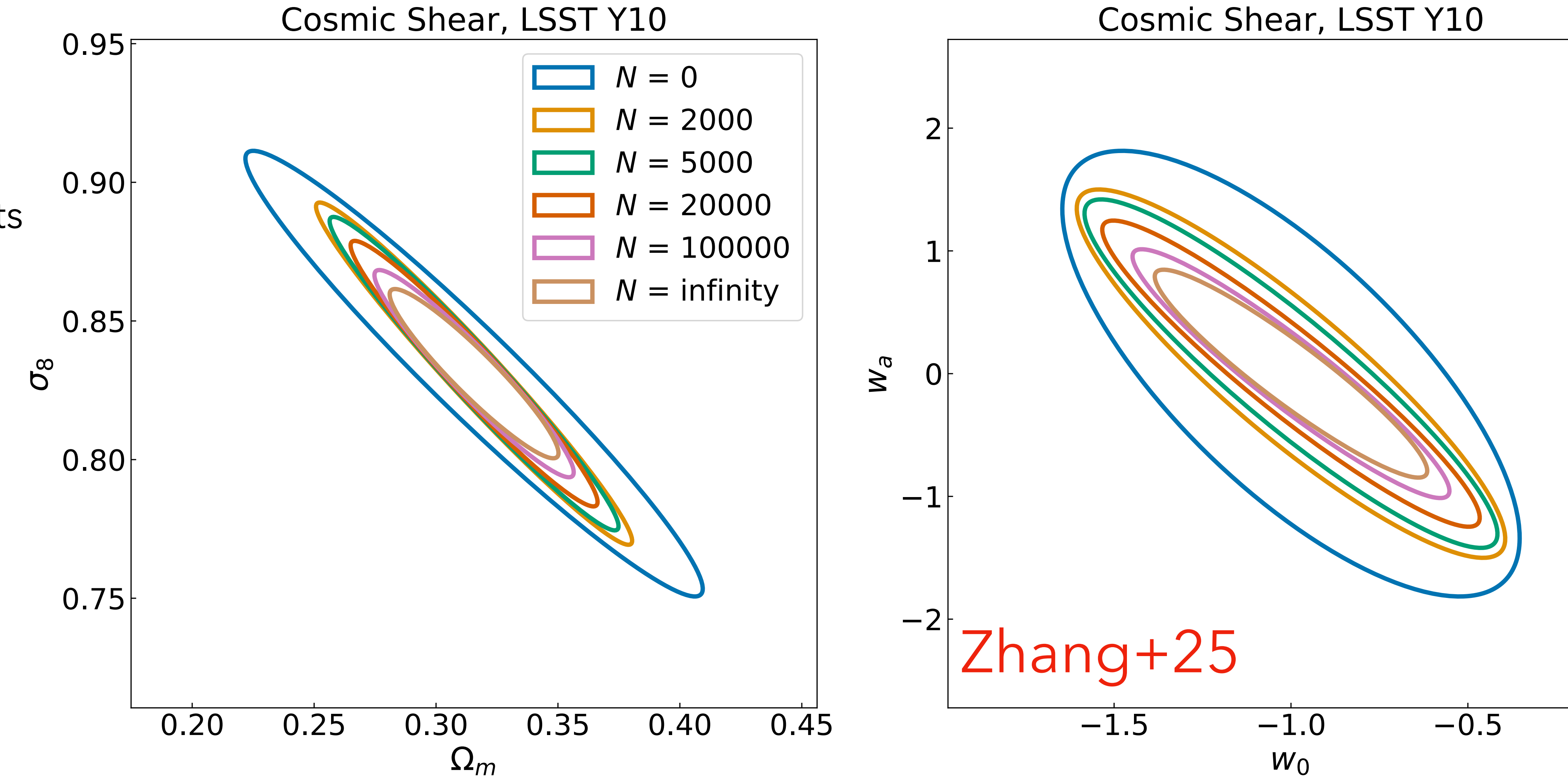
new redshift calibration method

new redshift calibration sample

new images and photo-z

Additional z bin, new IA

Additional area



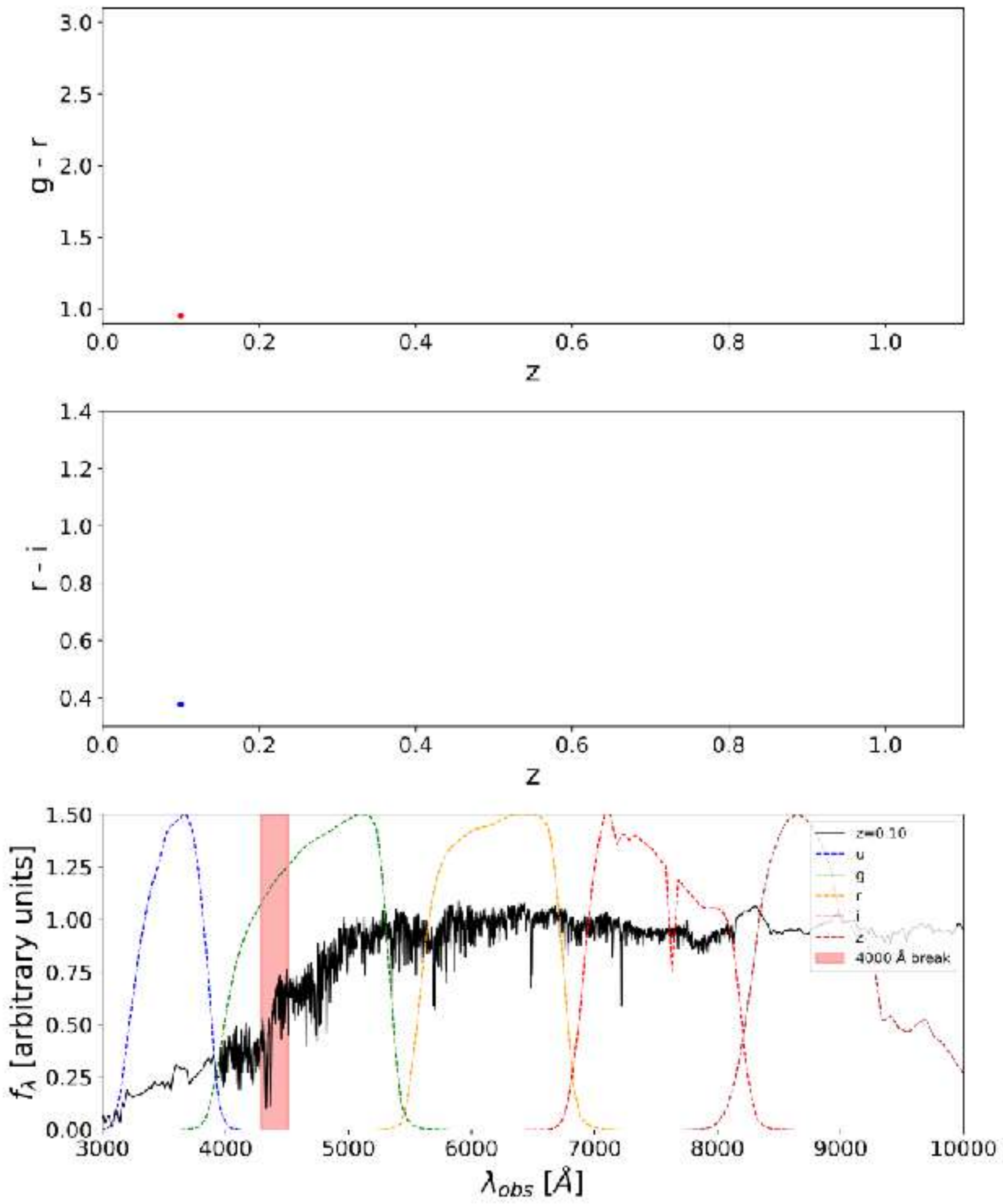
Rubin-LSST and Euclid set stringent requirements on tomographic redshift distribution systematic uncertainties:

$$\Delta\bar{z} = 0.001 \times (1+z) \text{ WL Y10}$$
$$\Delta\sigma_z = 0.003 \times (1+z) \text{ WL Y10}$$
$$\Delta\bar{z} = 0.003 \times (1+z) \text{ GCL Y10}$$
$$\Delta\sigma_z = 0.03 \times (1+z) \text{ GCL Y10}$$

LSST SRD 18

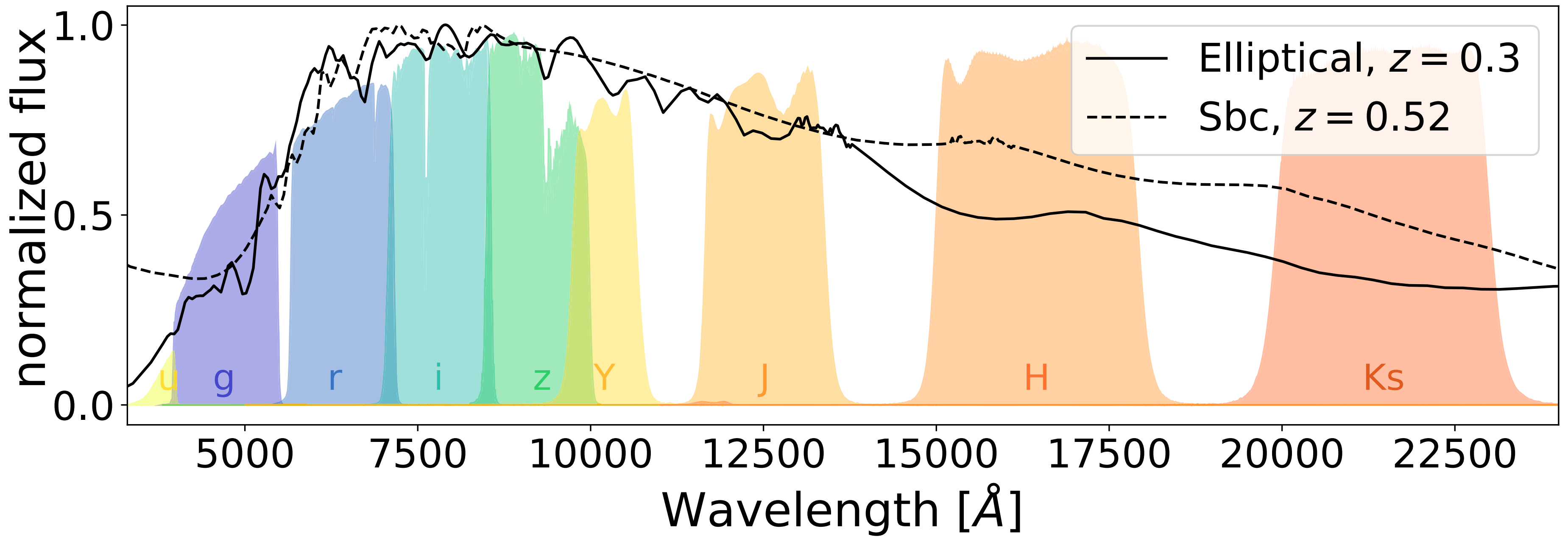
What prevents us to have well measured n(z)?

Gains in performance: regular photo-z methods (template fitting, machine learning) are not precise enough (Newman&Gruen22 review).

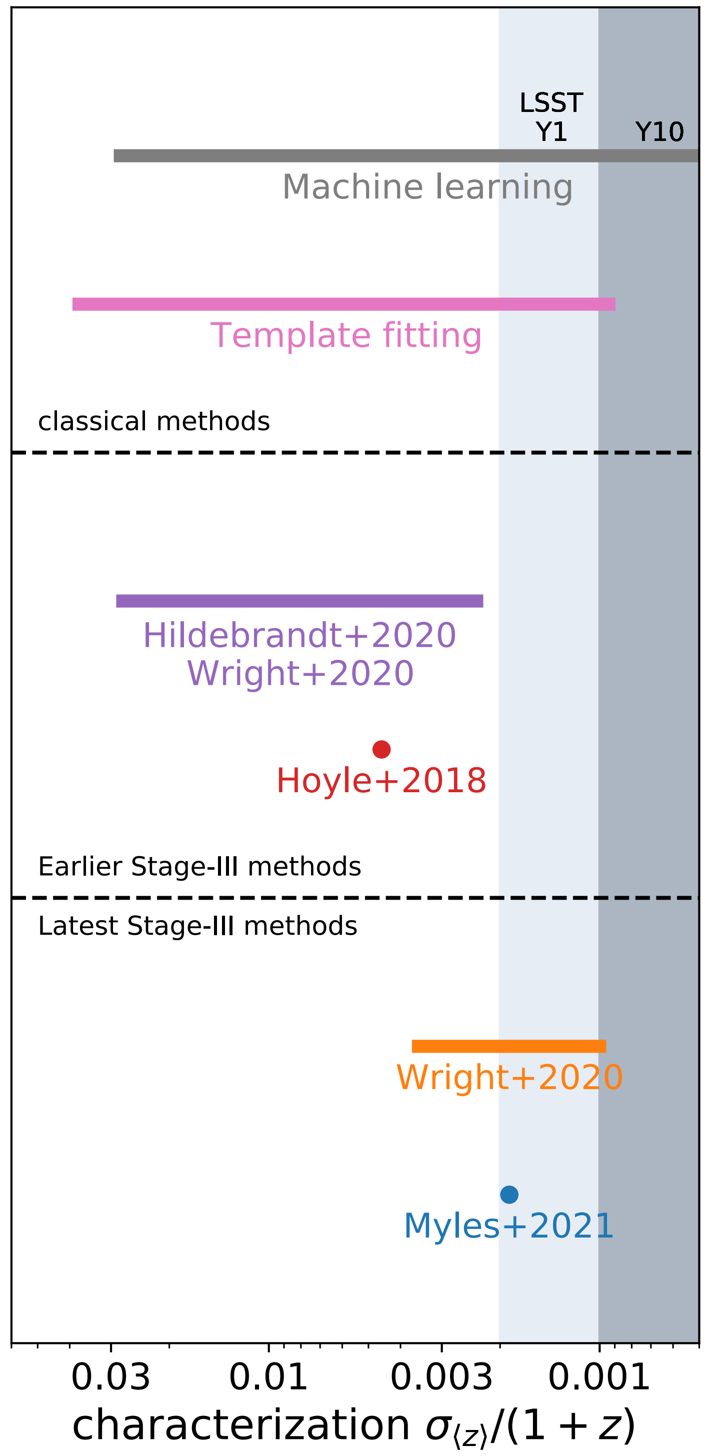


Calibration uncertainties on the mean redshift are $|\Delta z| \sim 0.01 \times (1+z)$ (Schmidt+2020), too high for Stage IV.

$\sigma_z \sim 0.01 \times (1+z)$

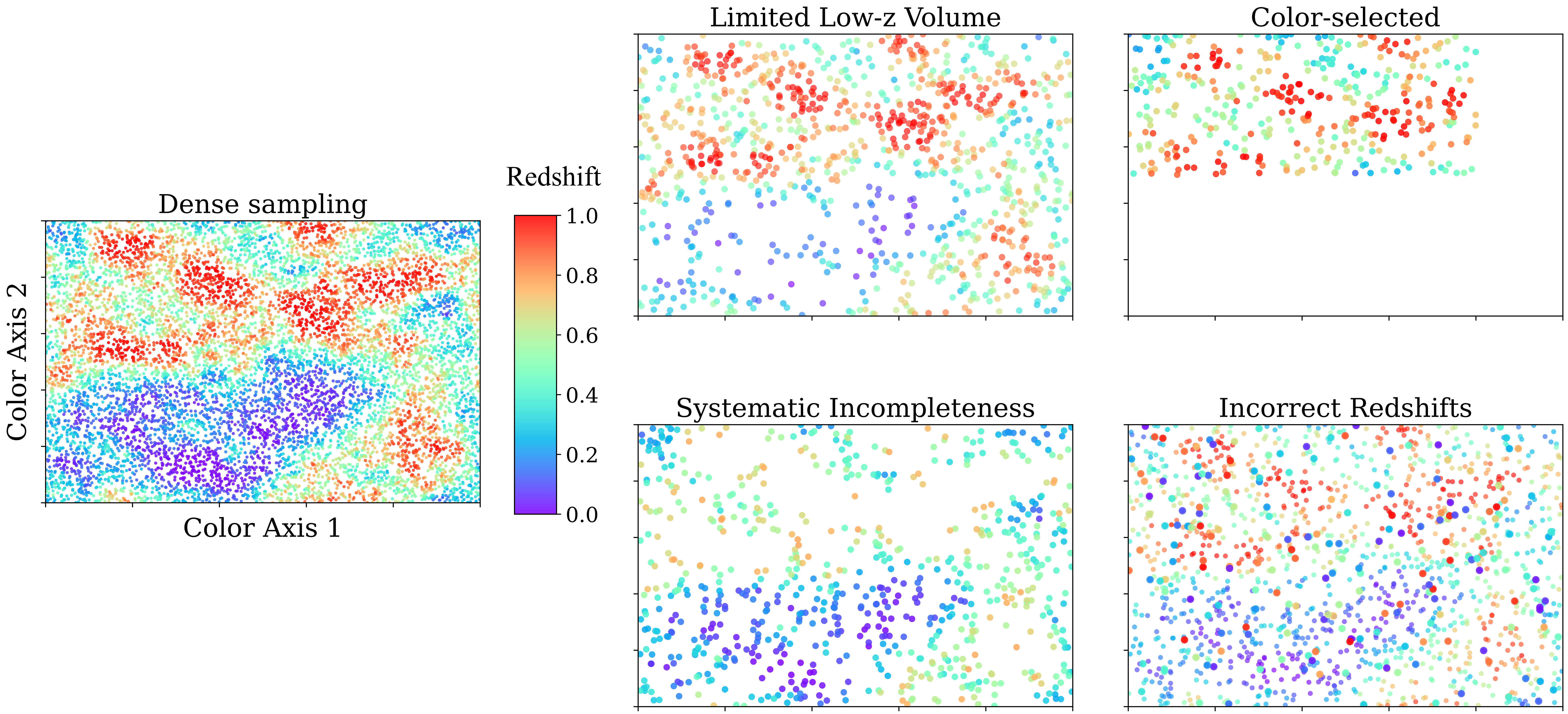


Buchs+2019



What prevents us to have well measured n(z)?

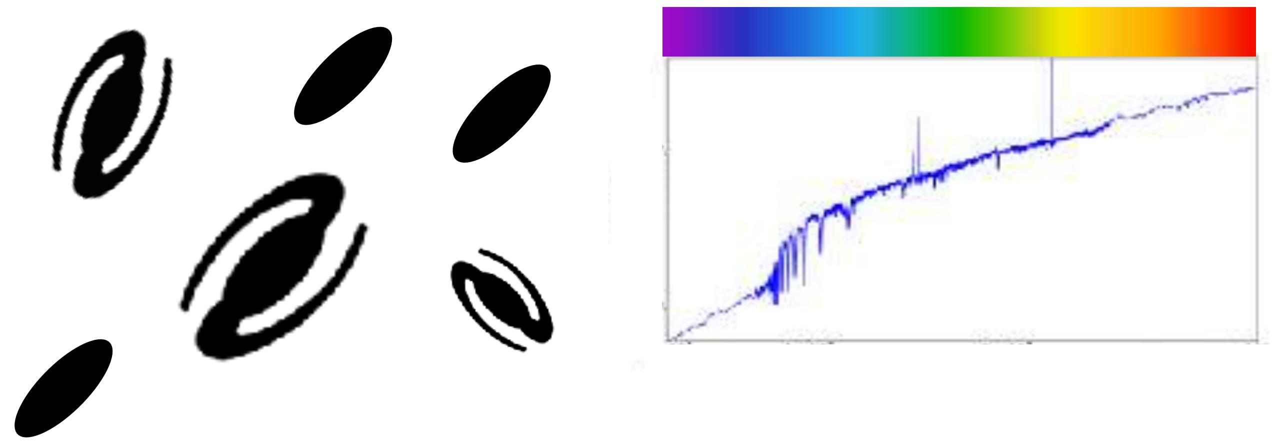
Incomplete knowledge of the galaxy population: sets informative prior, existing samples of spec-z may systematically miss some population of objects leading to biased colour-z relations (Newman&Gruen22 review).



You need model for the galaxy population!

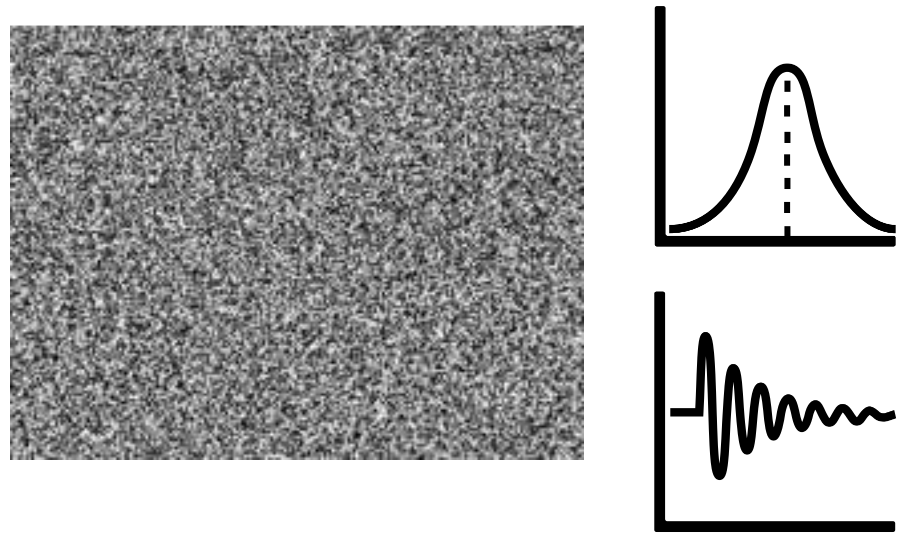
Pixel-level Forward-modelling of galaxy surveys

Galaxy Population Model



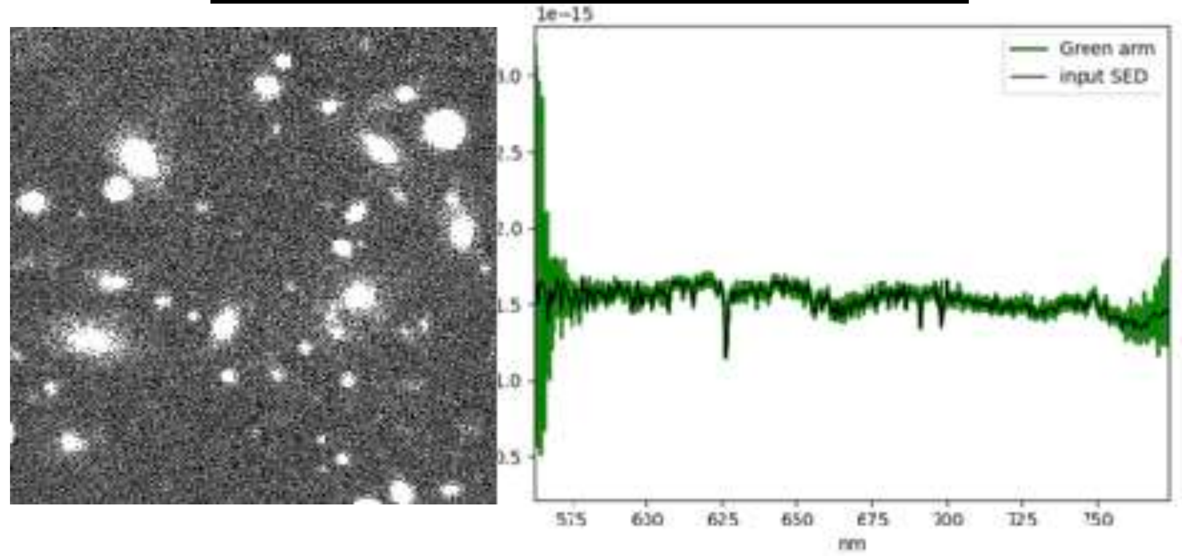
$$P(\psi) \times \prod_{i=1}^{N_{\text{galaxies}}} P(f_{\text{true}}(\theta_i, z_i) \,|\, \psi)$$

Instrumental effects



$$P(\sigma_{\text{obs}}), P(\sigma_{\text{instr}})$$

Simulated data



$$P(f_{\text{obs},i} \,|\, f_{\text{true}}(\theta_i, z_i), \sigma_{\text{instr,obs}})$$

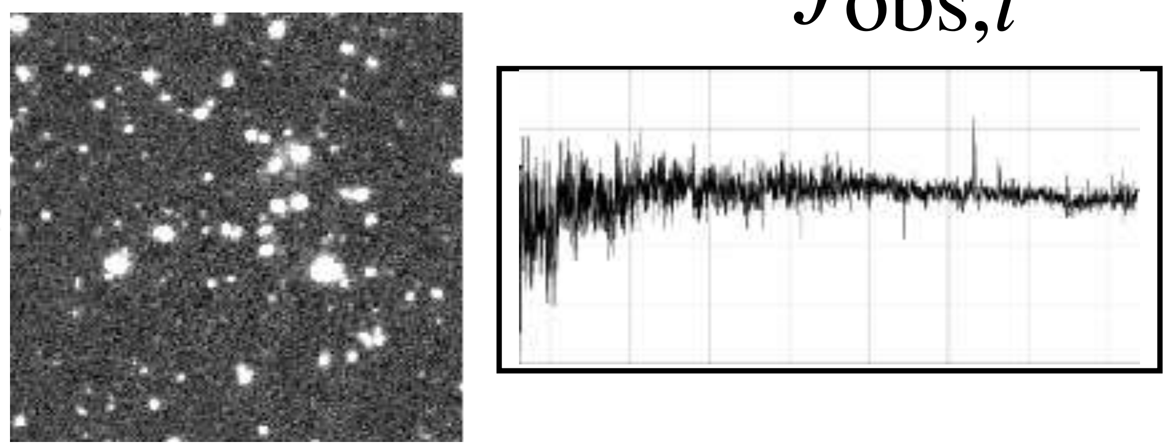
$$[P(S_i \,|\, f_{\text{obs},i})]$$

	Mag	Size	z
Obj 1	…	…	…
Obj 2	…	…	…
Obj 3	…	…	…
…	…	…	…

Observed data $f_{\text{obs},i}$

DESC

WAVES Wide Area Vista Extragalactic Survey



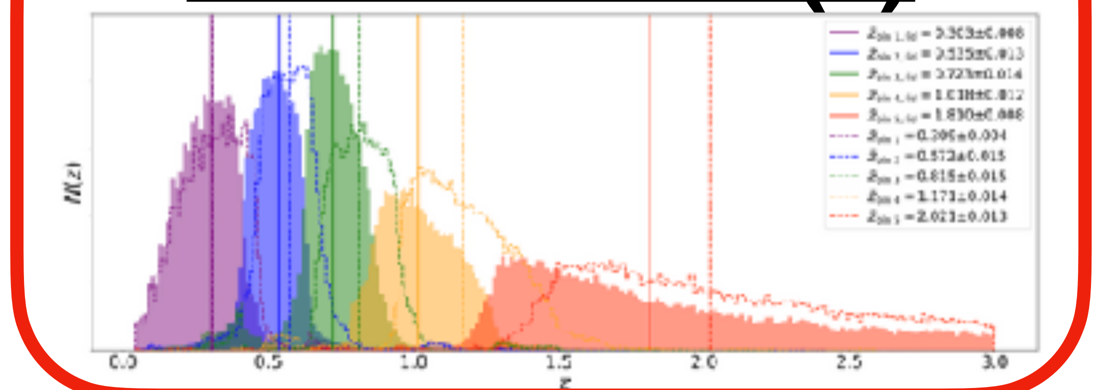
$$[P(S_i \,|\, f_{\text{obs},i})]$$

	Mag	Size	z
Obj 1	…	…	…
Obj 2	…	…	…
Obj 3	…	…	…
…	…	…	…

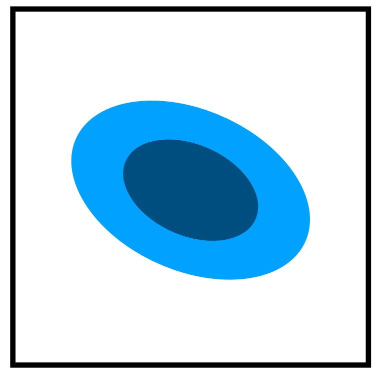
Model constraint via SBI

$$P(\theta_{1:N}, z_{1:N}, \psi \,|\, f_{\text{obs},1:N})$$

Simulated n(z)

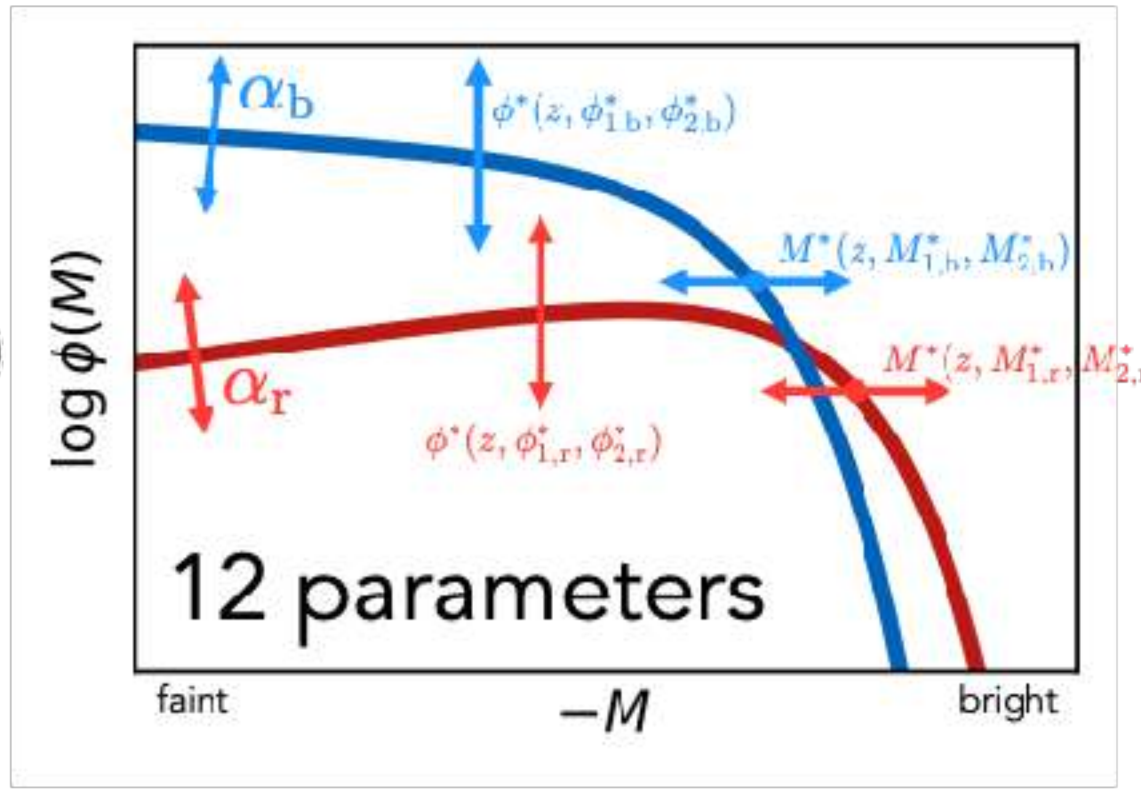


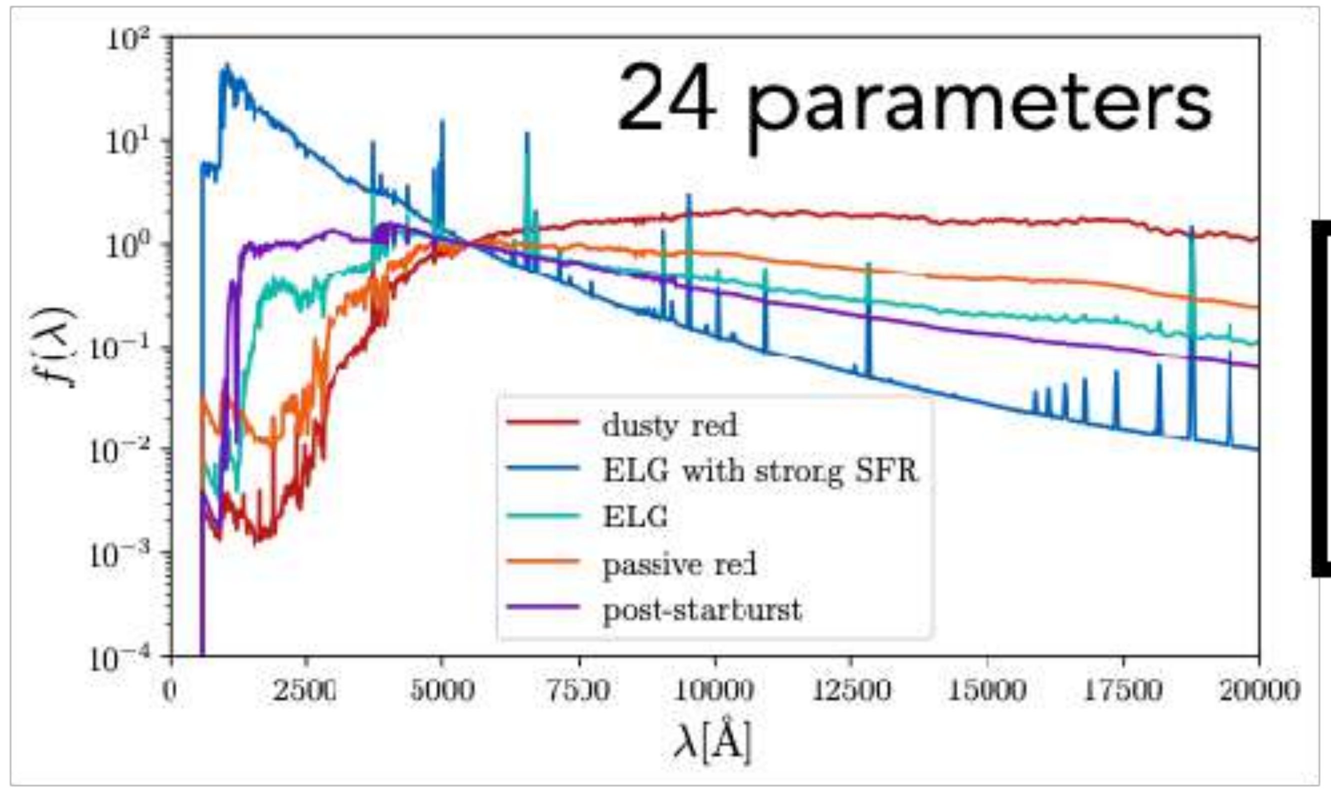
Cosmological constraints



GalSBI: phenomenological galaxy population model

Blue luminosity function → *z, M* → Linear combination of templates → *SED*





Red luminosity function → *z, M* → Linear combination of templates → *SED*

Filter throughputs
Galactic extinction

17 parameters

apparent magnitude → Morphology models → size, ellipticity, Sérsic

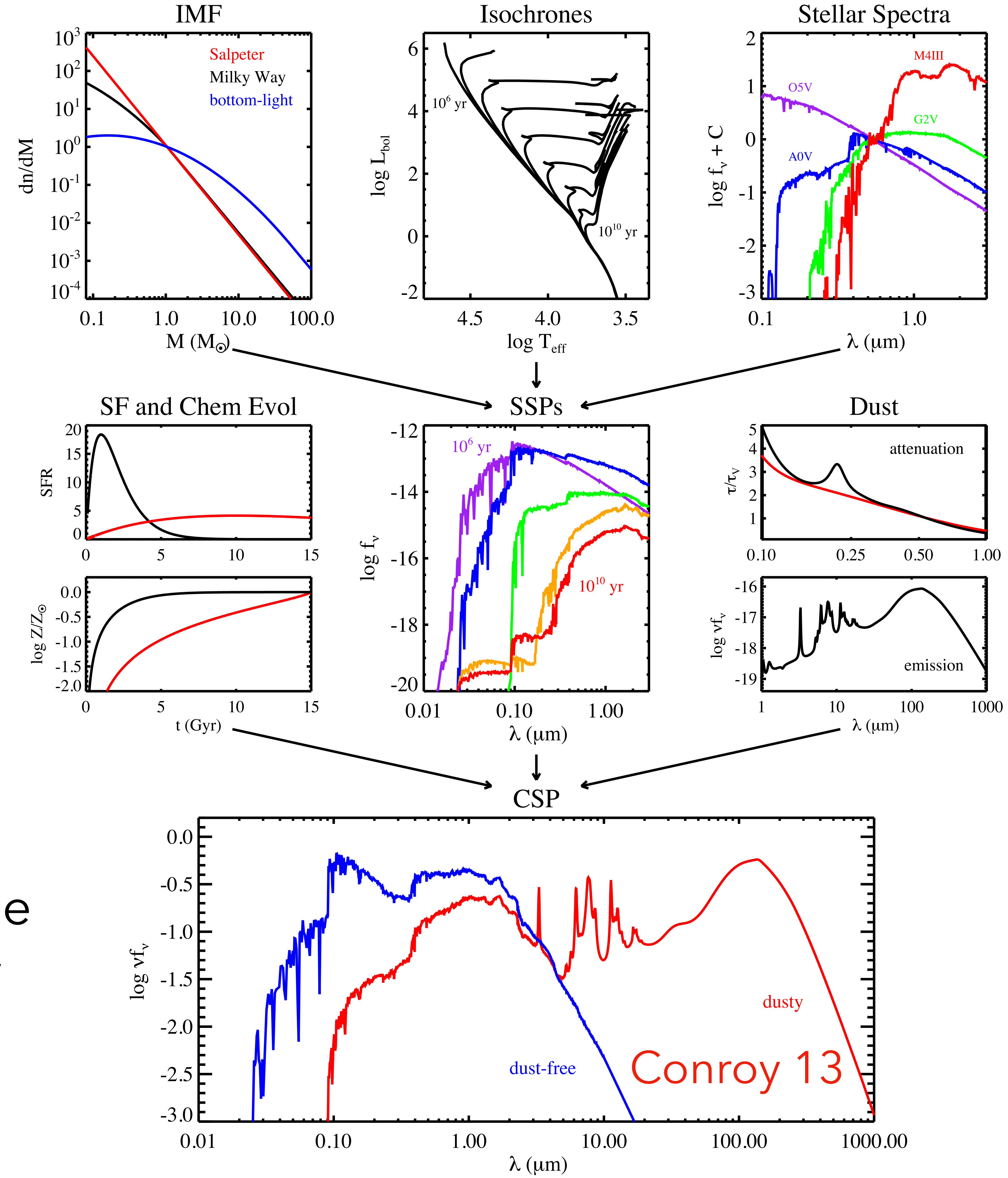
interactive luminosity function plot with live updated image simulation

https://cosmo-docs.phys.ethz.ch/galsbi/galpop_pheno.html

Publicly released in Fischbacher, Kacprzak, LT+24, based on past works by Herbel+17, Tortorelli+18,20,21, Moser+24.

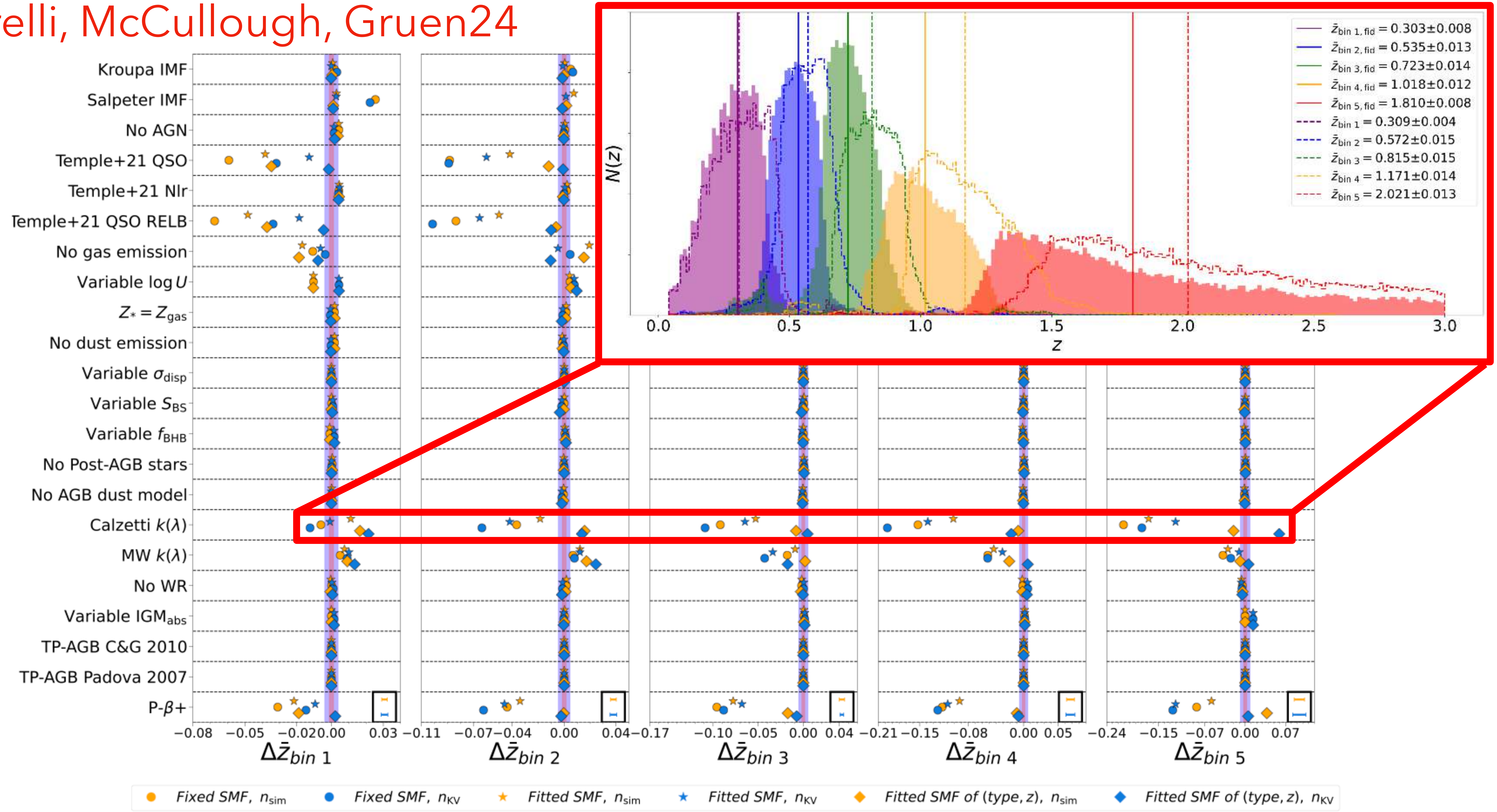
Advantages of a stellar population synthesis (SPS)-based model

- Can we extrapolate beyond template coverage?
- Physics-based model coupled SPS leverages information from more constraining dataset and transfer to fainter mag-limited samples.
- High-z analogs of low-z galaxies retaining physical consistency.
- Integration of SPS-based forward modelling with image and spectra simulators allows for characterising survey selection functions, biases and incompletenesses.



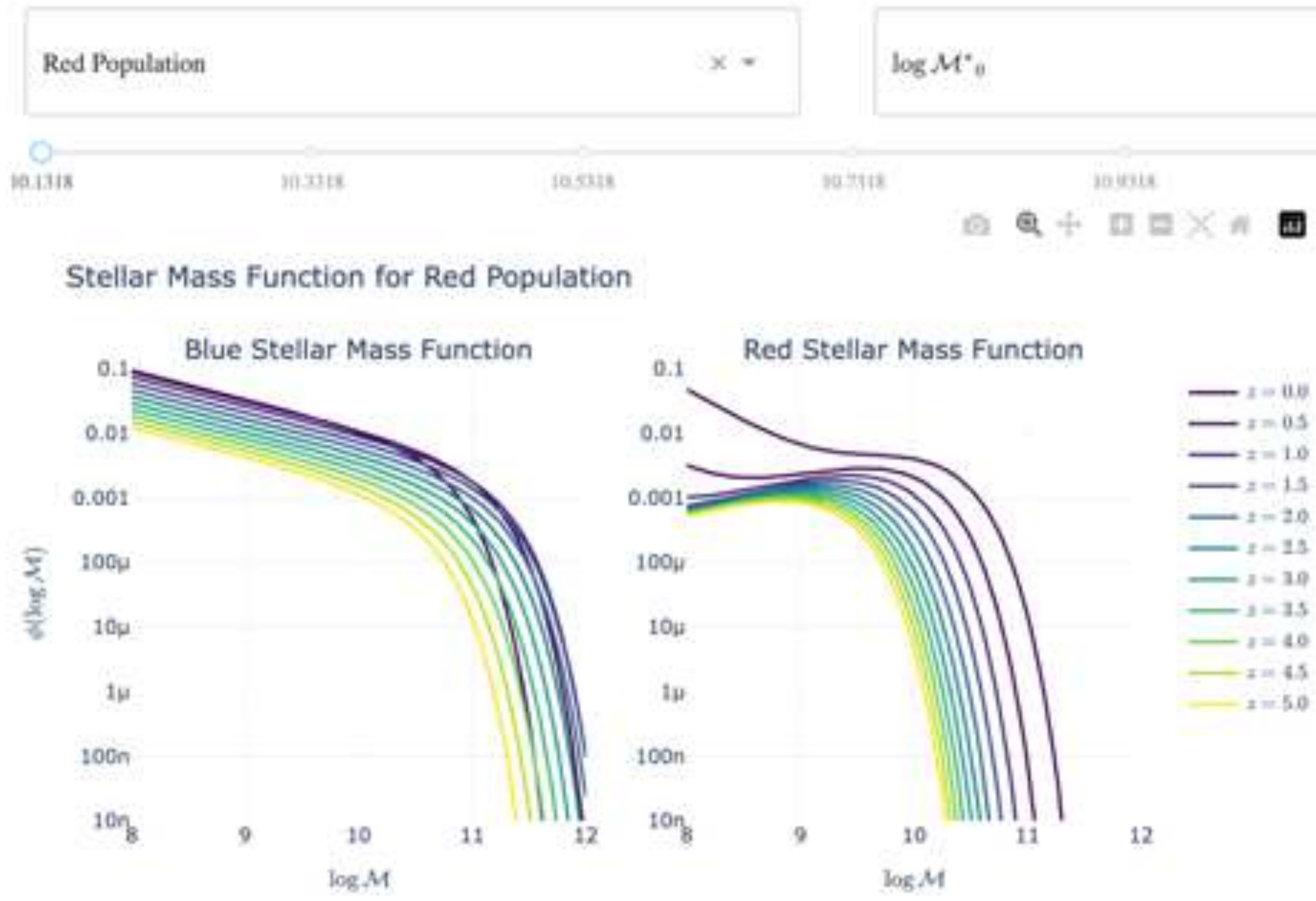
SED modelling choices impact forward modelling-based n(z)

Tortorelli, McCullough, Gruen24



GalSBI-SPS: physics-based galaxy population model

GalSBI-SPS: stellar mass function



Tortorelli+25

$$\log \mathscr{M}^*(z) = \log \mathscr{M}^*{}_0 + \log \mathscr{M}^*{}_1 \times \log(1+z) + \log \mathscr{M}^*{}_2 \times \log(1+z)^2$$

$$\phi^*_{\mathrm{l,h}}(z) = \phi^*_{\mathrm{amp,l,h}} \times (1+z)^{\phi^*_{\mathrm{exp,l,h}}}$$

$$\Phi(\log \mathscr{M}, z) = = \ln(10) e^{-10^{\log \mathscr{M} - \log \mathscr{M}^*(z)}} \times \left[\phi^*_l(z)\left(10^{\log \mathscr{M} - \log \mathscr{M}^*(z)}\right)^{\alpha_l+1}\right] \times \left[\phi^*_h(z)\left(10^{\log \mathscr{M} - \log \mathscr{M}^*(z)}\right)^{\alpha_h+1}\right]$$

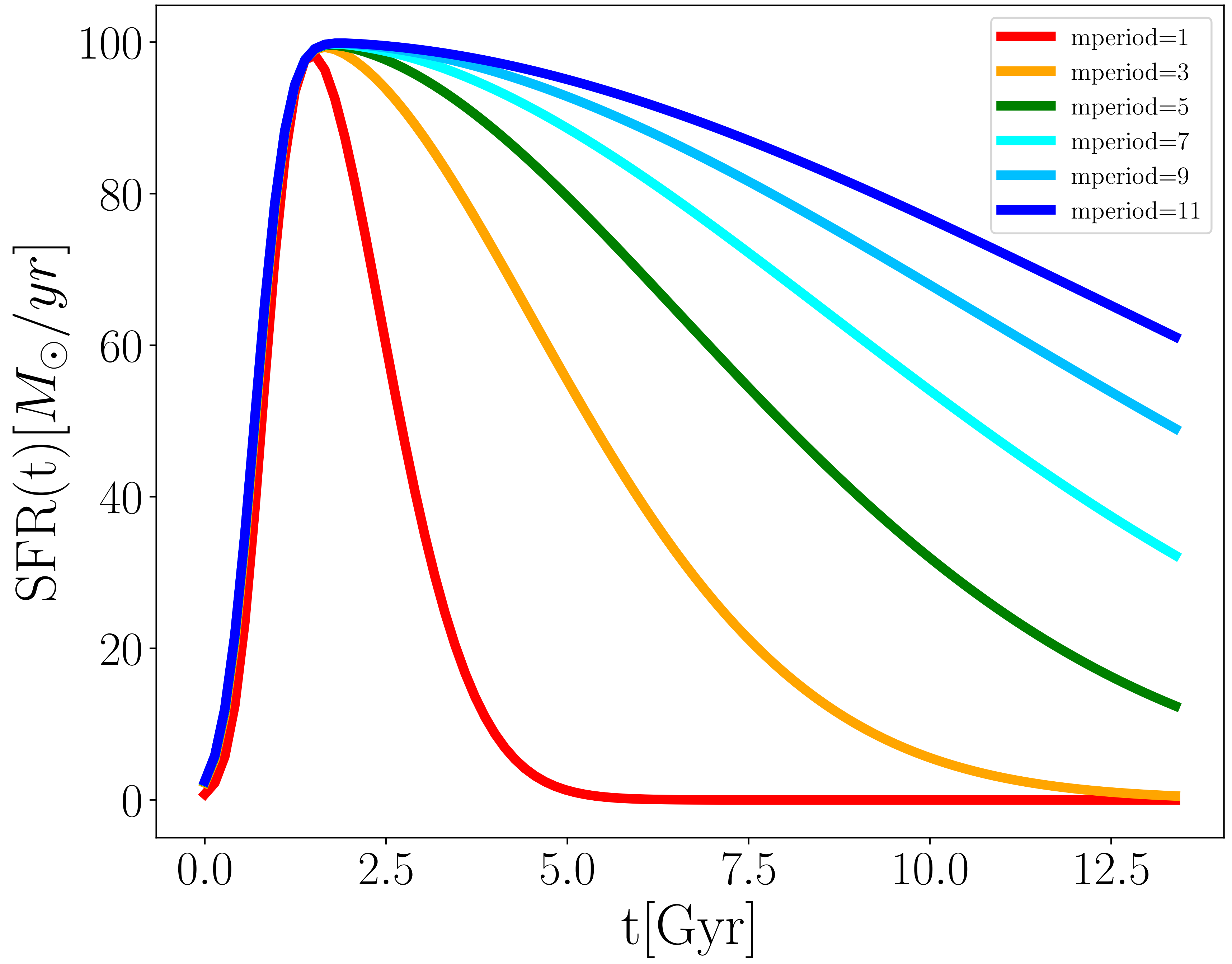
GalSBI-SPS: star formation history

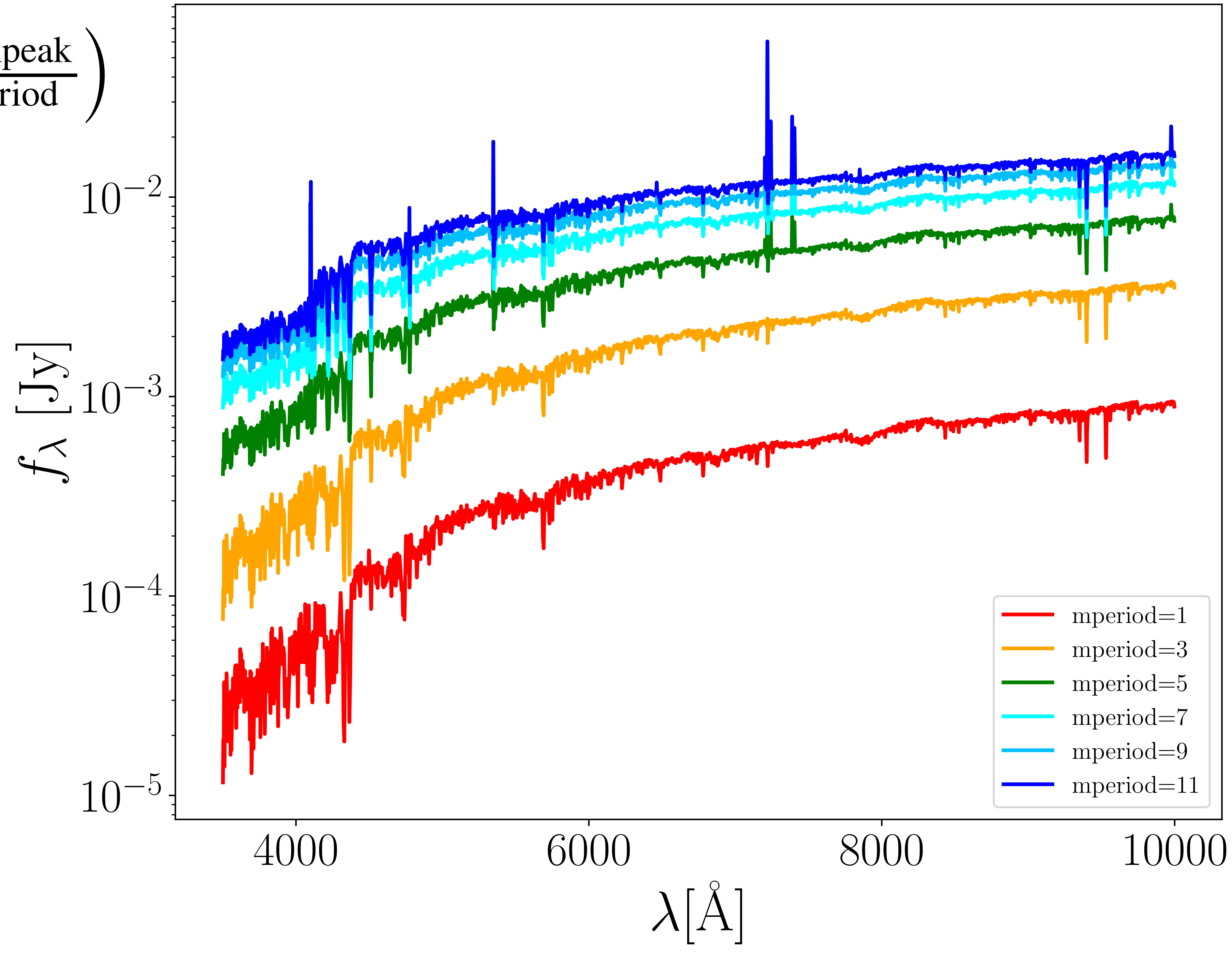
$$SFR(t) = SFR(t)_{\mathrm{norm}} \times \left[1 - \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{t-\mu}{\sigma\sqrt{2}}\right)\right]\right]$$

$$SFR(t)_{\mathrm{norm}} = \mathrm{mSFR} \times e^{\frac{X(t)^2}{2}}$$

$$X(t) = \left(\frac{t - \mathrm{mpeak}}{\mathrm{mperiod}}\right)\left(e^{\mathrm{mskew}}\right)^{\mathrm{asinh}\left(\frac{t-\mathrm{mpeak}}{\mathrm{mperiod}}\right)}$$

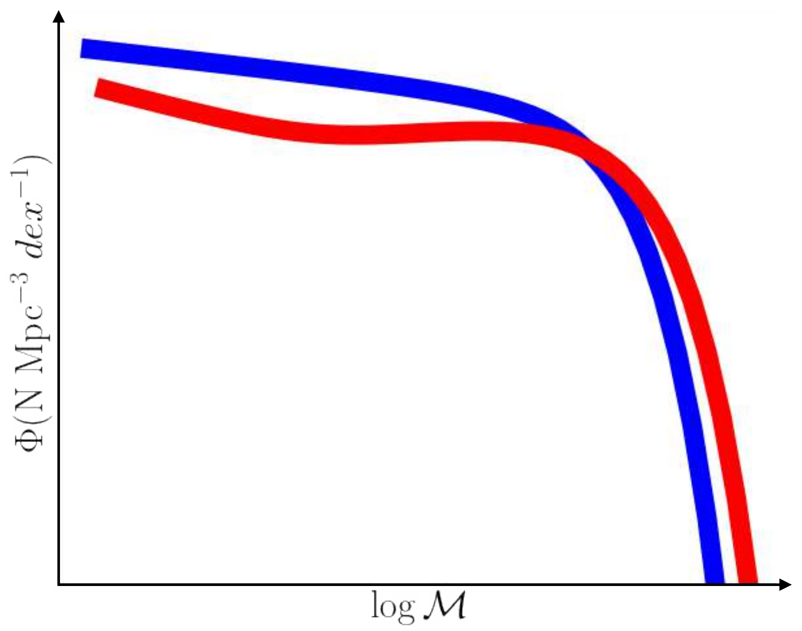
Tortorelli+25





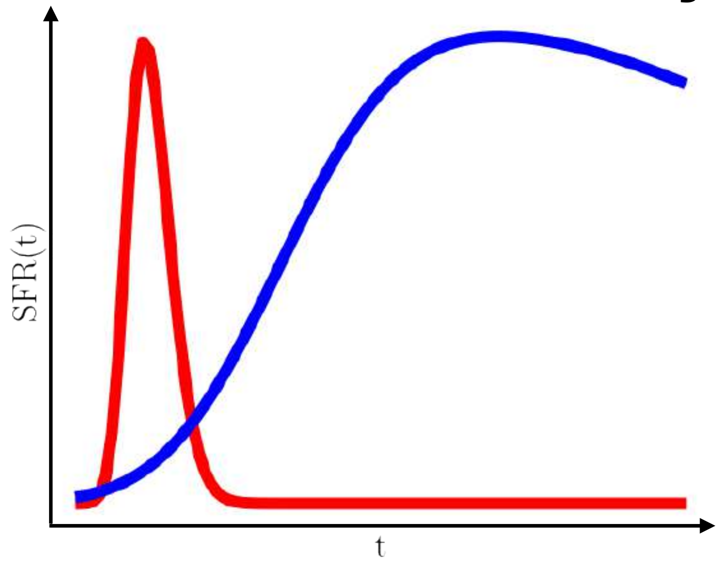
GalSBI-SPS: physics-based galaxy population model

Stellar Mass function



$P(\log \mathscr{M}, z)$

Star formation history



$P(\mathrm{SFR(t)} \,|\, \log \mathscr{M}, z)$

Metallicity history

$$Z_{\mathrm{gas}}(t) = Z_{\mathrm{gas,init}} + (Z_{\mathrm{gas,final}} - Z_{\mathrm{gas,init}}) \times \frac{1}{\mathscr{M}} \int_0^{t'} SFR(t)\mathrm{d}t$$

$$[12 + \log(O/H)] = \alpha(t_{\mathrm{lb}}) \log\left(\frac{SFR}{M_\odot yr^{-1}}\right) + \beta(t_{\mathrm{lb}}) \log\left(\frac{\mathscr{M}}{M_\odot}\right) + \gamma(t_{\mathrm{lb}})$$

Dust extinction

$$A(\lambda) = e^{-\tau(\lambda/\lambda_0)^{\nu}}$$

$$\log(\tau_{\mathrm{ISM}}) = \tau_{\mathrm{ISM},0} + \tau_{\mathrm{ISM},1} \log(1+z) + \tau_{\mathrm{ISM},2} \log\left(\frac{\mathscr{M}}{M_\odot}\right) + \tau_{\mathrm{ISM},3} \log(sSFR)$$

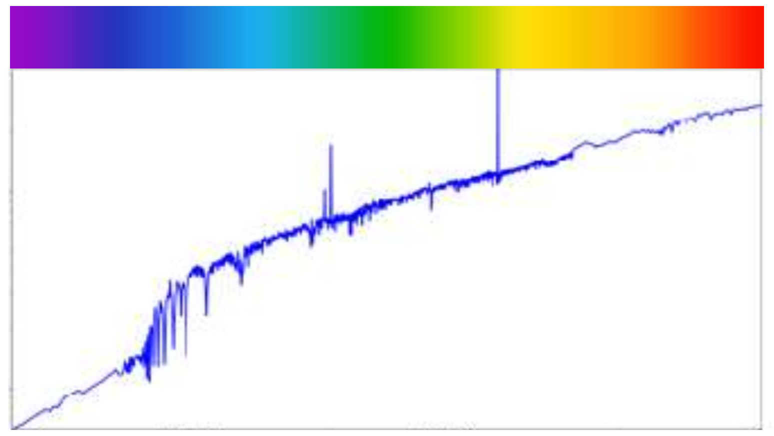
Tortorelli+25

Velocity dispersion

$$\sigma_{\mathrm{disp}}(\mathscr{M}) = \sigma_b \left(\frac{\mathscr{M}}{M_b}\right)^{s_1} \quad \text{for } \mathscr{M} \le M_b$$

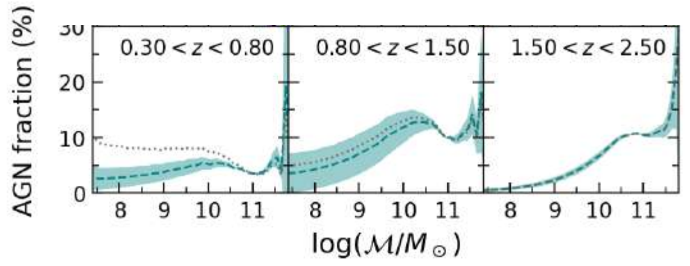
$$\sigma_{\mathrm{disp}}(\mathscr{M}) = \sigma_b \left(\frac{\mathscr{M}}{M_b}\right)^{s_2} \quad \text{for } \mathscr{M} > M_b$$

ProSpect



$$P(f_\lambda \,|\, z, SFR(t), Z_{\mathrm{gas}}(t), \log U, \log \tau, \sigma_{\mathrm{disp}}, \log f_{\mathrm{AGN}})$$

AGN



Morphology

$$R_e = \zeta(z)(\mathscr{M})^{\eta(z)} \left(1 + \frac{\mathscr{M}}{10^{\delta(z)}}\right)^{\theta(z)-\eta(z)}$$

$$n = n_0 + n_1 \times \left(\frac{\log \mathscr{M}}{\log(10^{10.5} M_\odot)}\right)^{n_2}$$

Input catalogue

Gas ionisation

$$\log U = \upsilon_0 + \upsilon_1(\log(O/H) + 4) + \upsilon_2 \log n_e + \upsilon_3(\log(sSFR) + 9)$$

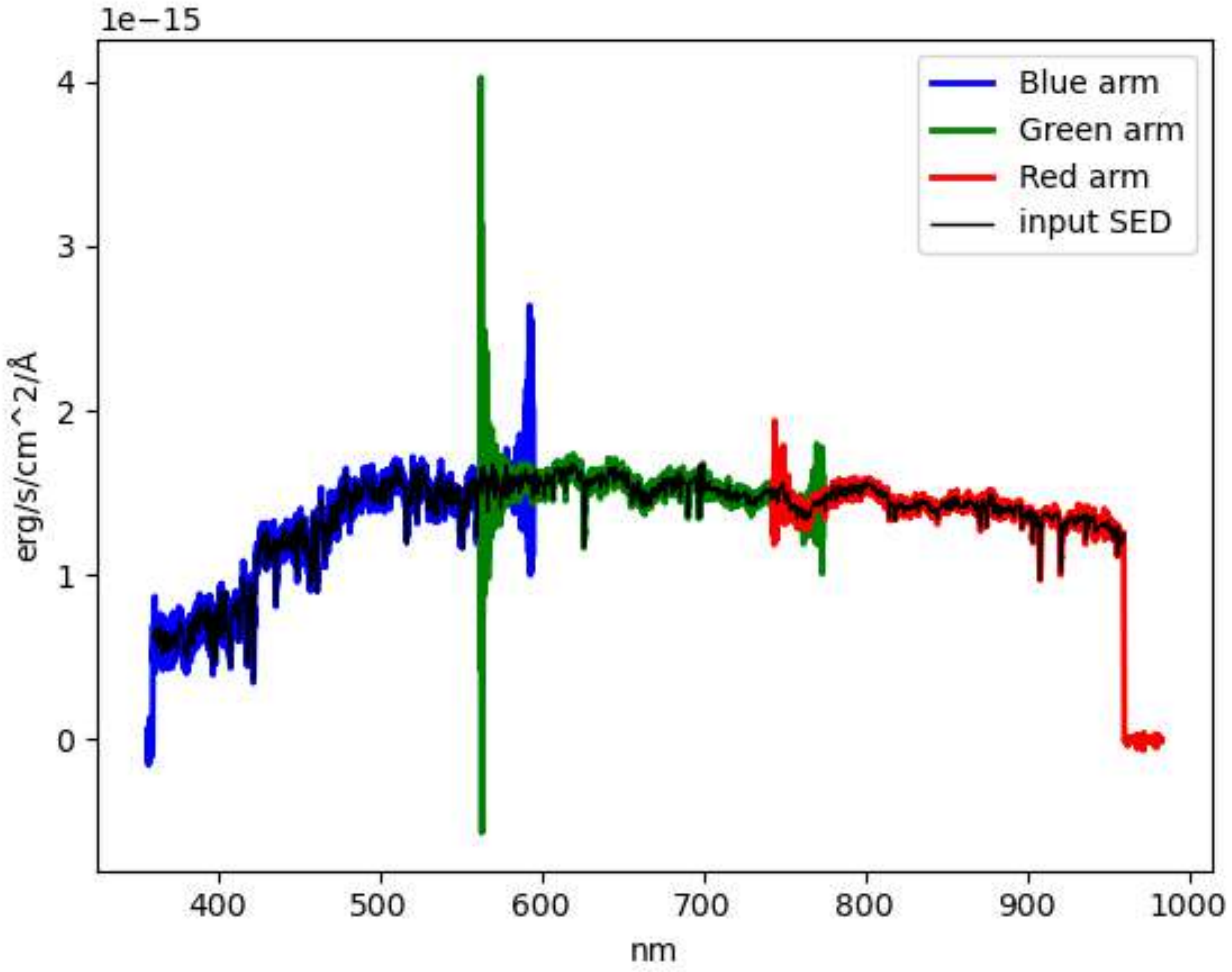
- Independent sampling
- $(\log \mathscr{M}, z)$ conditioned sampling
- $(\log \mathscr{M}, z, SFR)$ conditioned sampling
- SED generation

Pixel-level image and spectra simulations

Ultra Fast image generator (UFig)

Publicly released in Fischbacher+24 (incl. LT)

Ultra fast Spectra simulator (USpec)



Soon to be released in Tortorelli+25 in prep.

GalSBI framework is publicly available

The (Great) GalSBI

License MIT | python 3.9 | 3.10 | 3.11 | 3.12 | pypi package 0.1.0 | pipeline passed | coverage 97.00% | coverage report

arXiv 2412.08701 | arXiv 2412.08722 | Documentation | Source Code

Create realistic galaxy catalogs and astronomical images based on the GalSBI model. The GalSBI phenomenological model is described in Fischbacher et al. (2024) and the \texttt{galsbi}package is described in Fischbacher et al. (2024). The GalSBI stellar population synthesis-based model is described in Tortorelli et al. (2025) and the package is described in Tortorelli et al. (2025).

Installation

The package can be installed via pip:

```
pip install galsbi
```

Usage

To generate a catalog of galaxies with their intrinsic properties from the phenomenological model, you can use the following code snippet:

```
from galsbi import GalSBI

model = GalSBI("Fischbacher+24")
model()
cats = model.load_catalogs()
```

To generate a catalog of galaxies with their intrinsic properties from the stellar population synthesis-based model, you can use the following code snippet:

```
from galsbi import GalSBI

model = GalSBI("Tortorelli+25")
model()
cats = model.load_catalogs()
```

More examples and detailed documentation can be found in the documentation.

Citation

If you use GalSBI in your work, please cite the science papers Fischbacher et al. (2024), Tortorelli et al. (2025) for using the GalSBI model and the code release papers Fischbacher et al. (2024), Tortorelli et al. (2025) for using the package. If you are using specific models or parametrizations, please also cite the corresponding papers. If you are not sure which papers to cite, use the following with your model instance:

```
model.cite()
```

Credits

This package was developed by the Cosmology group at ETH Zurich and updated by the ACAI group at LMU Munich. The package is currently maintained by Silvan Fischbacher and Luca Tortorelli.

Contributions

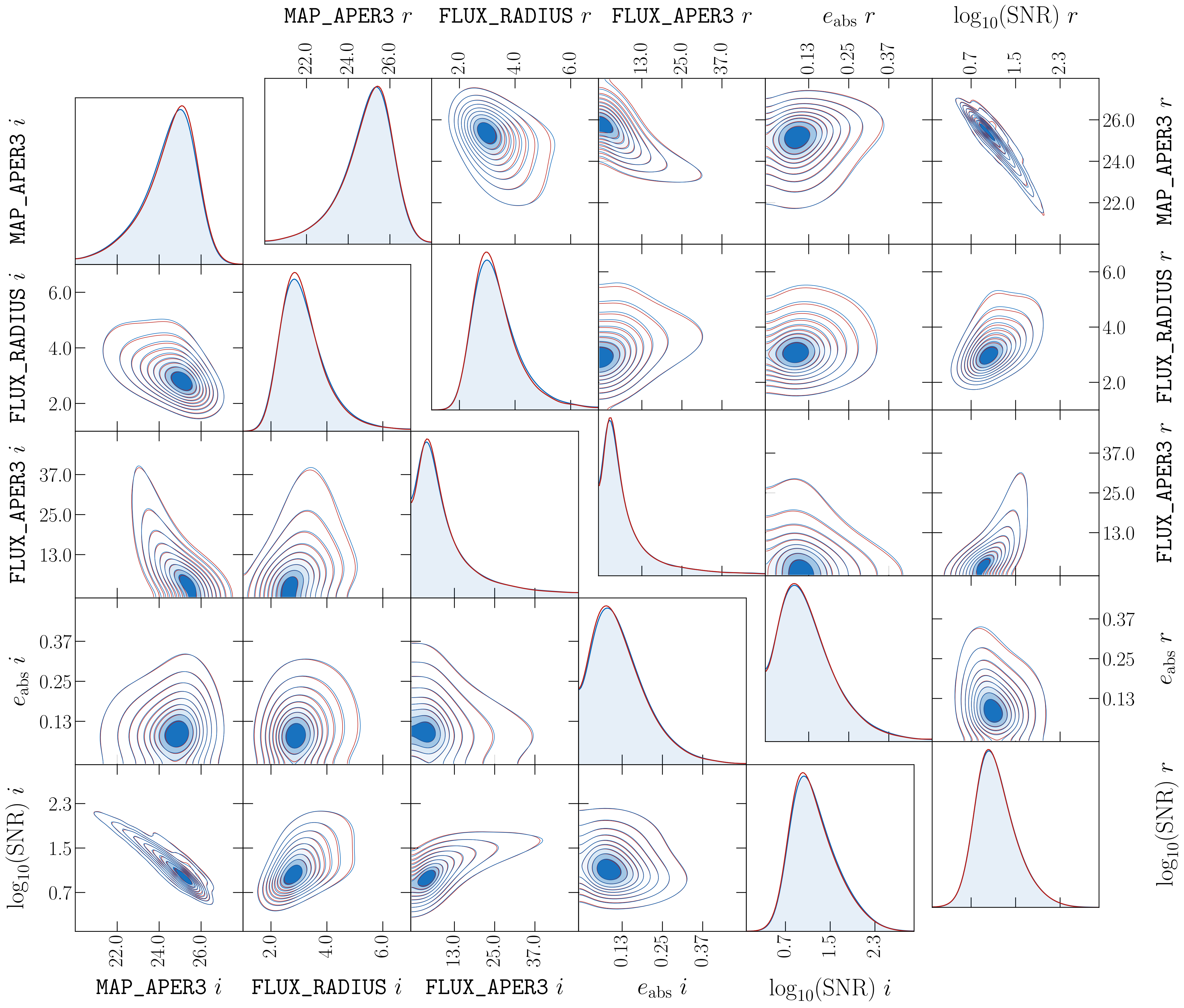
Contributions are welcome, and they are greatly appreciated! Every little bit helps, and credit will always be given.

galsbi on gitlab

https://cosmo-docs.phys.ethz.ch/galsbi/index.html

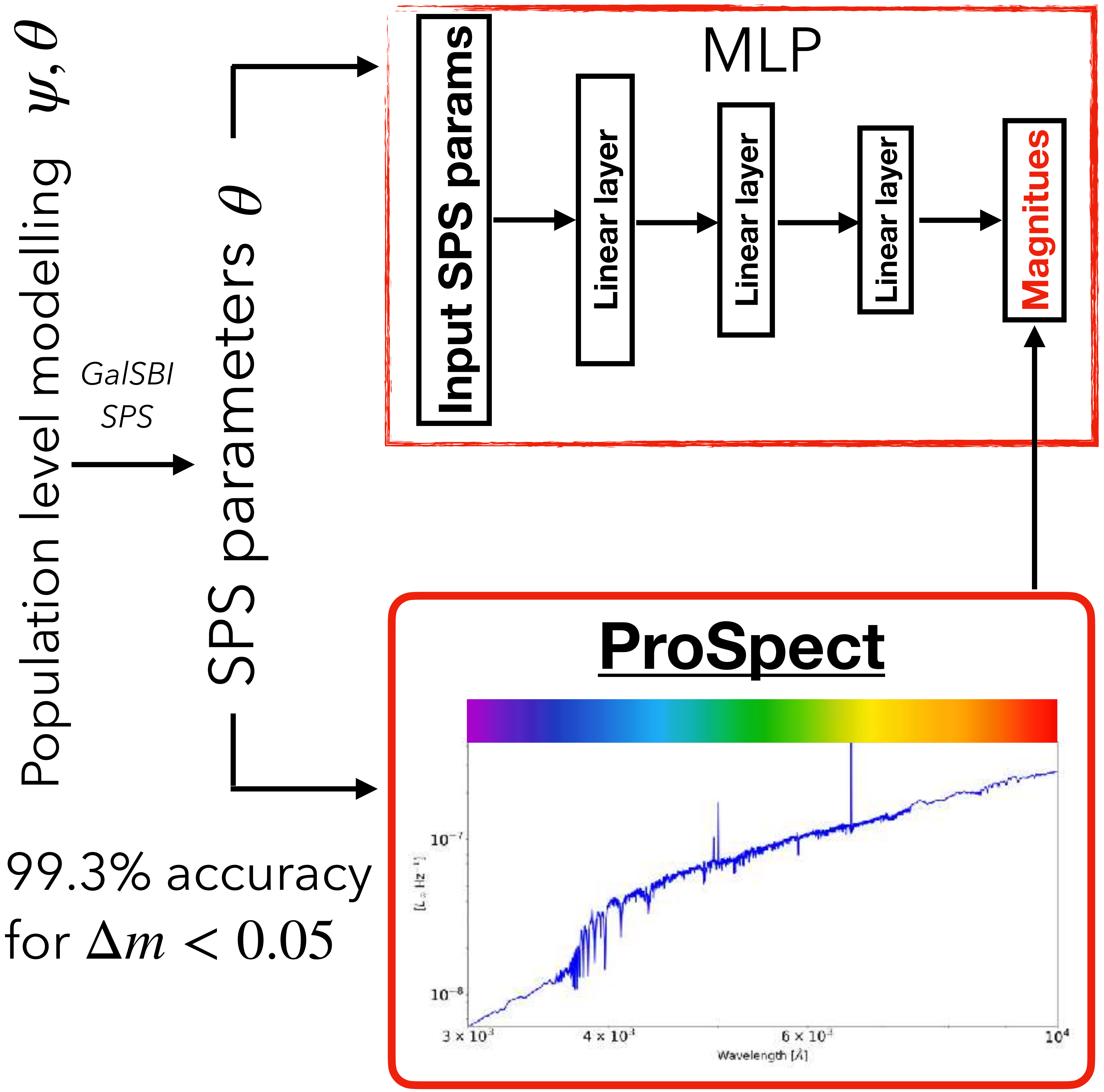
Tutorial here

Accelerating forward-modelling with emulators



Input parameters (different for each band)	Input parameters (the same for all bands)	Output parameters (different for each band)
- apparent magnitude m	- size r50	- MAG_APER
- PSF FWHM	- absolute ellipticity	- FLUX_RADIUS
- BKG noise ampl. and std	- object type	- FLUX_APER
- number of photons	- Sérsic index	- absolute ellipticity
- $n_{gal,m<i}$ with $i = 22, 23, 24$	- redshift	- SNR

Fischbacher, Kacprzak, LT+24



99.3% accuracy for $\Delta m < 0.05$

Tortorelli+25,in prep.

Positions in GalSBI

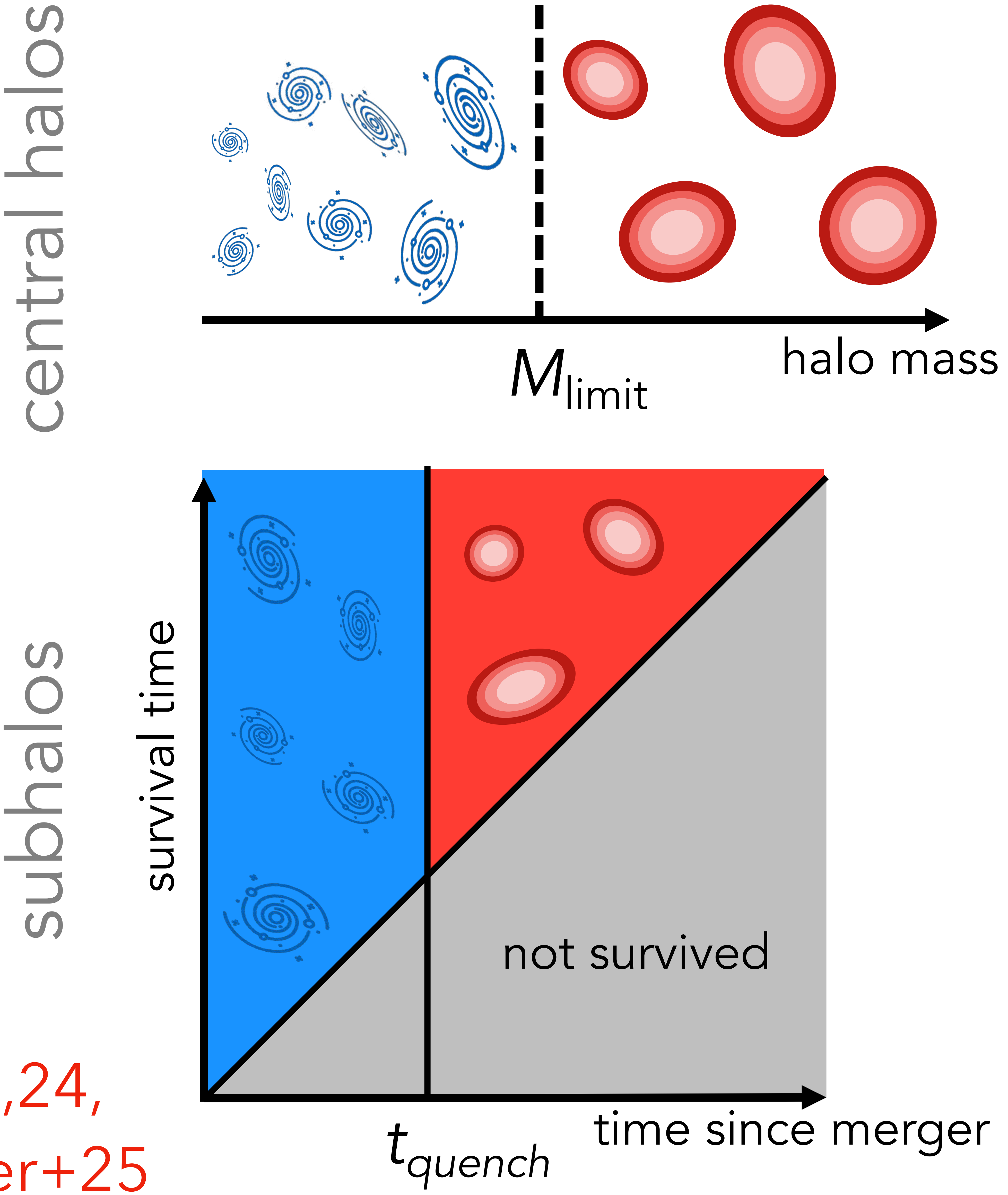
- Current status: Random sampling from a uniform distribution
- Work in progress: Using Subhalo Abundance Matching (SHAM)

SHAM overview:

- Full-sky DM-only simulation (PINOCCHIO)

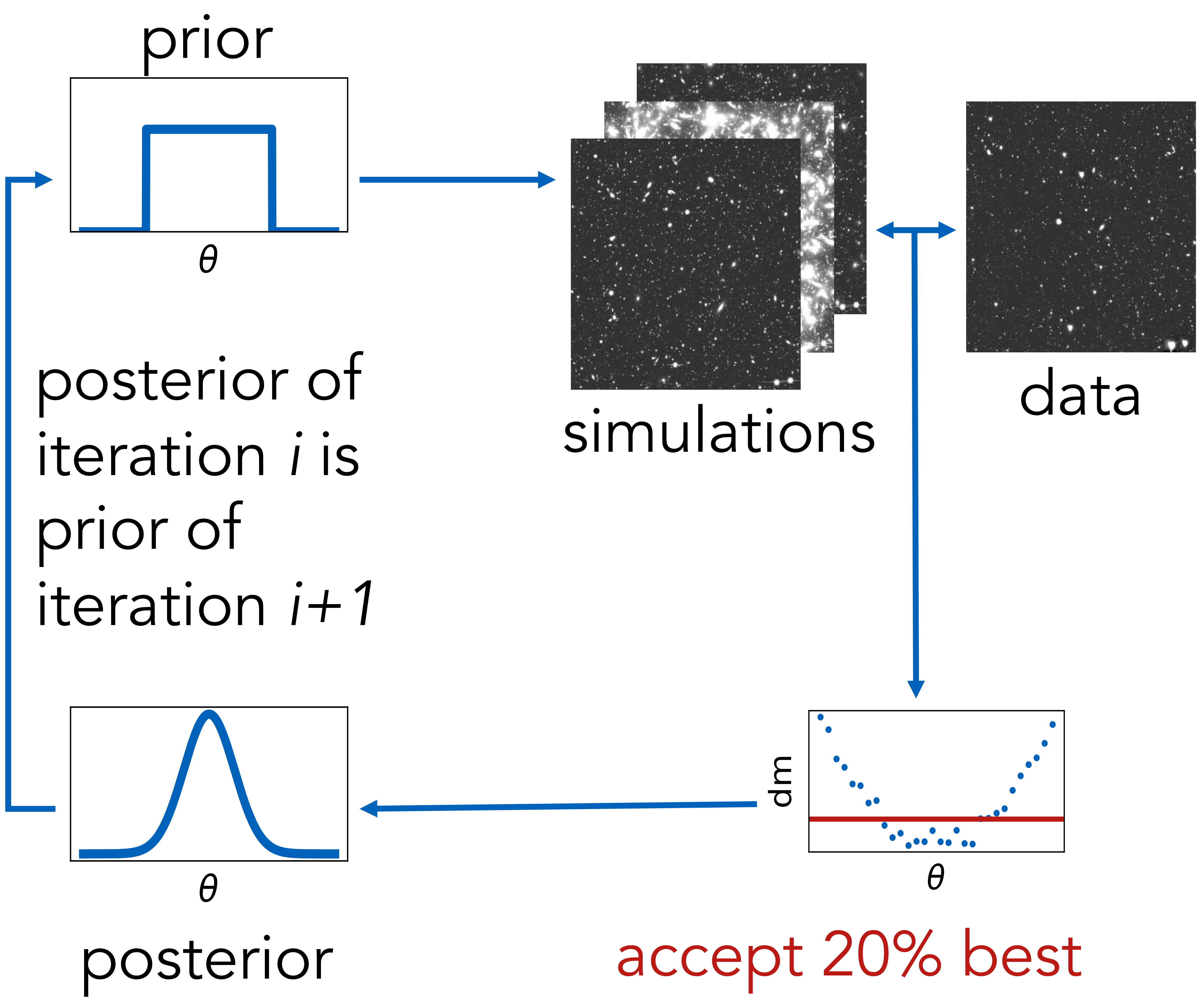
➢ Halo catalog + merger trees

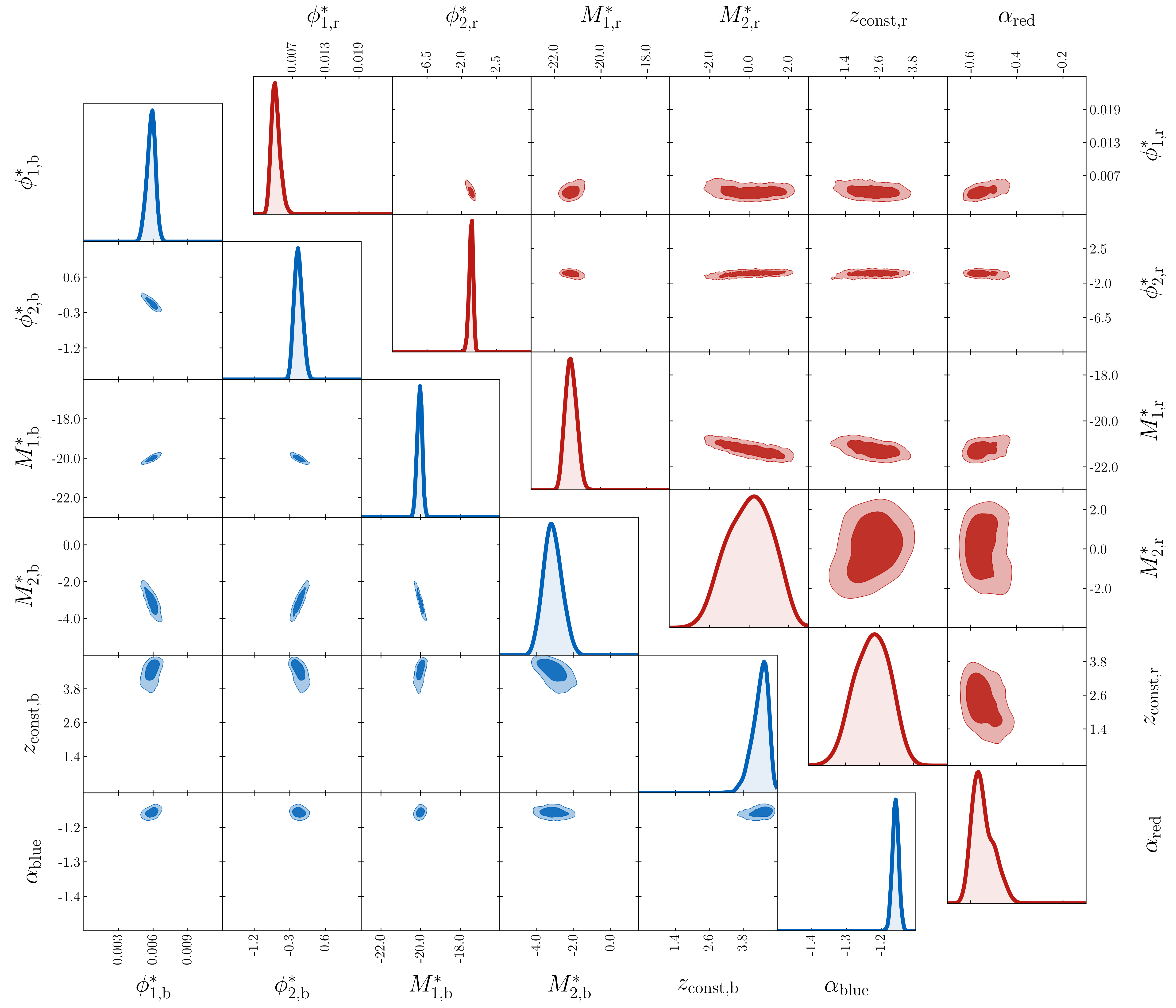
- Extract subhalos
- Classify (sub)halos as host for red/blue galaxies
- Match magnitudes to (sub)halo mass
  - ➢ assign positions



Berner+22,24, Fischbacher+25

SBI for model parameters constraints

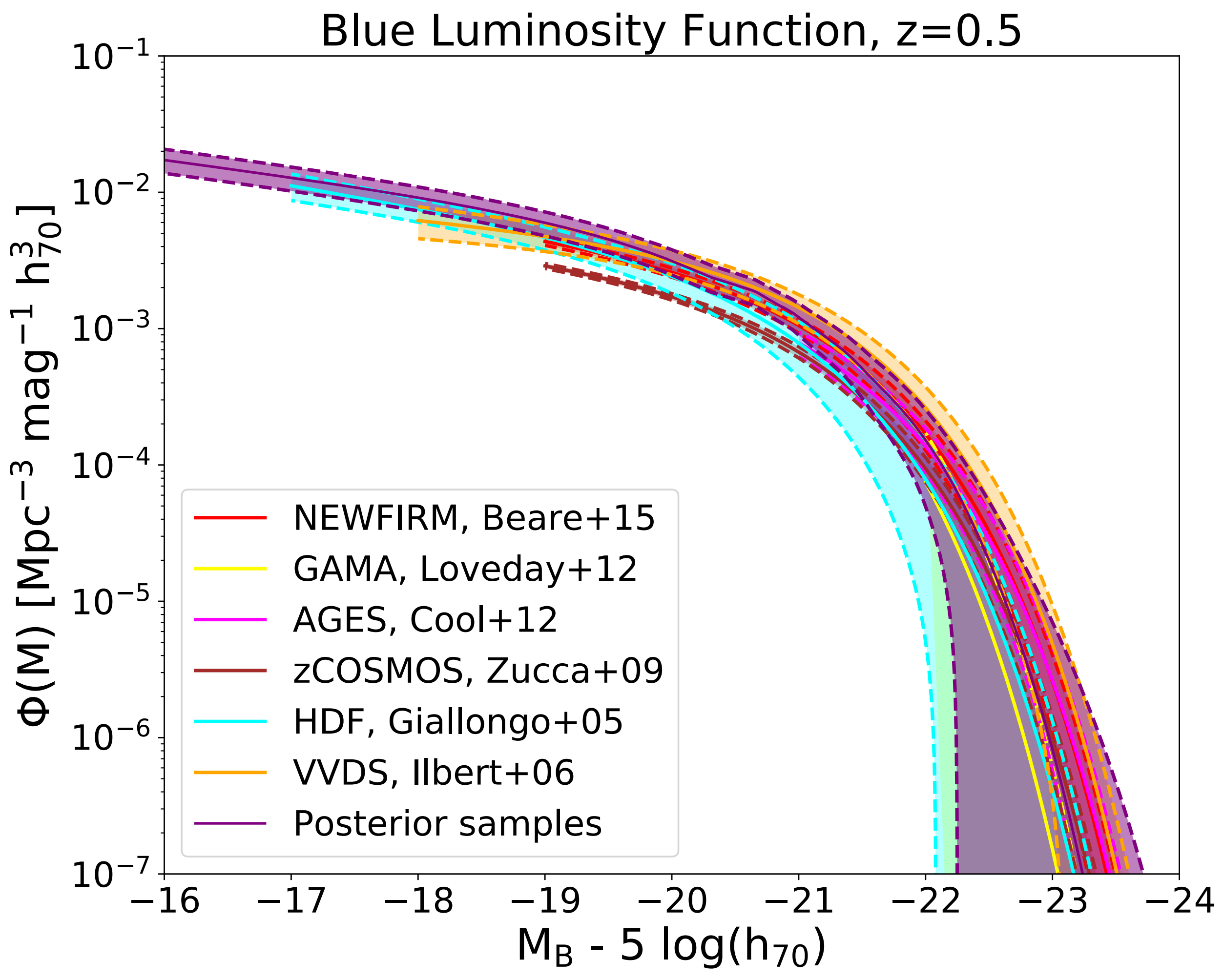


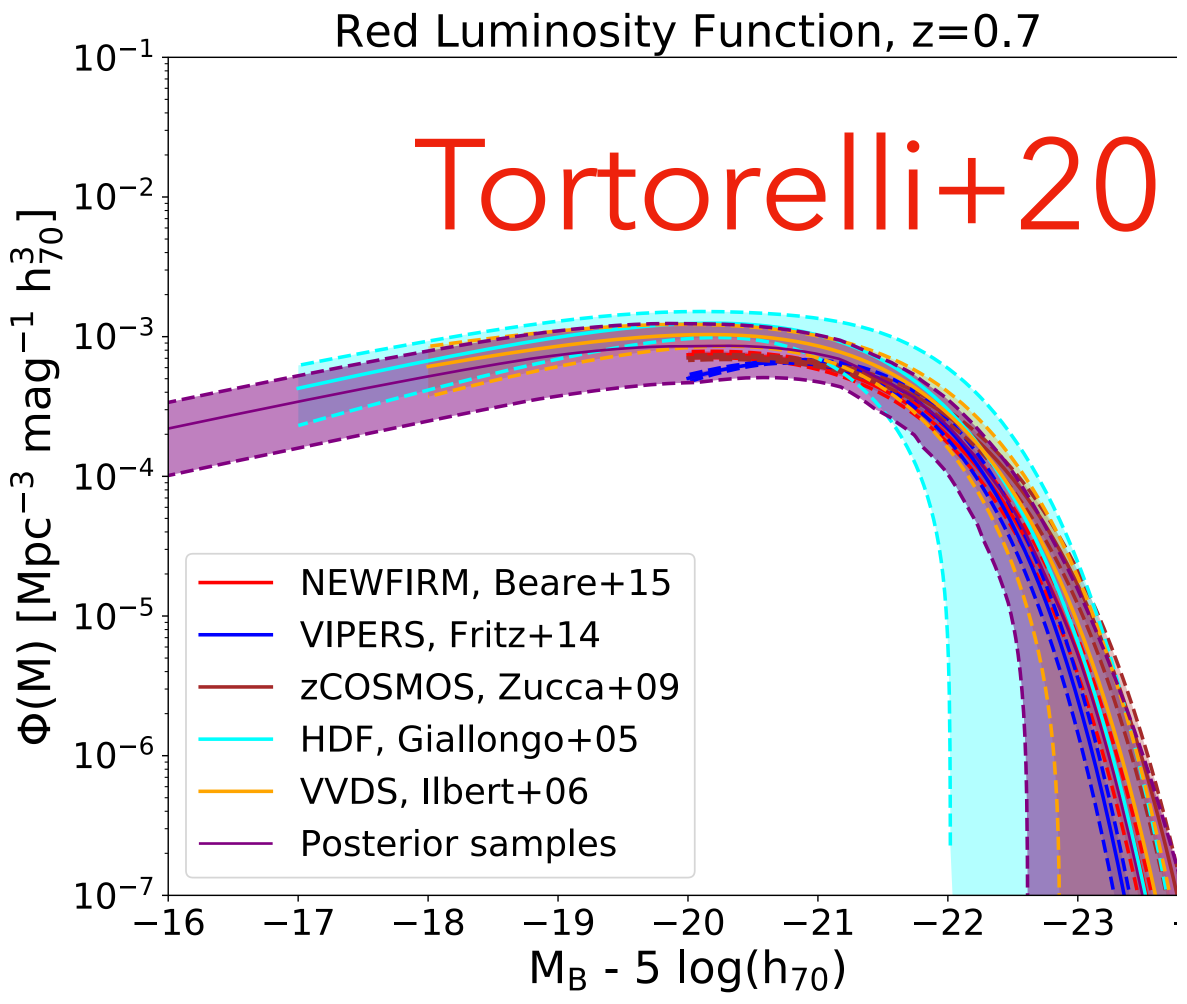


$$p\left(\theta|y\right) \simeq p\left(\theta|\rho\left(x,y\right) \leq \epsilon\right)$$

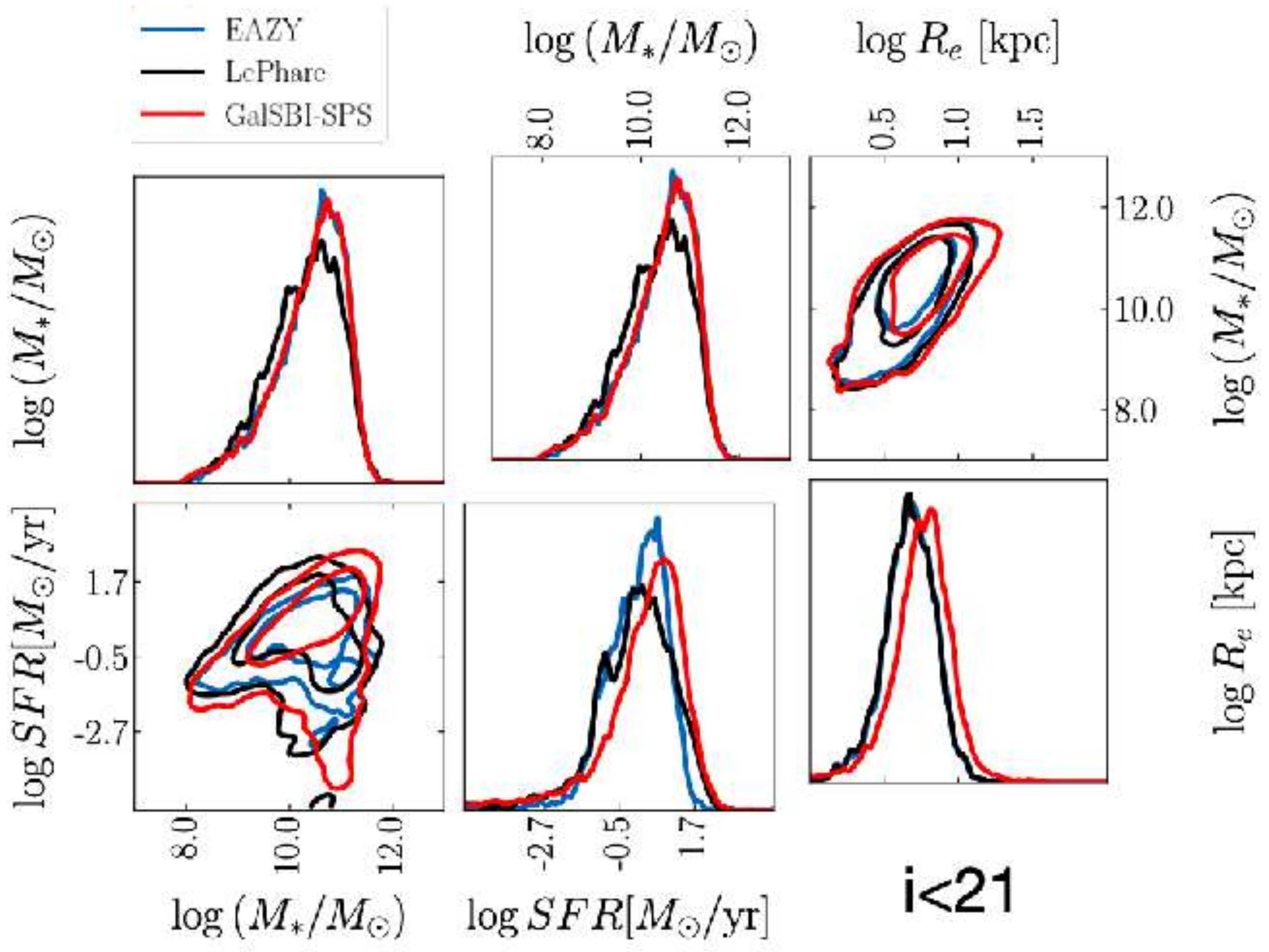
Fischbacher, Kacprzak, LT+24

Posterior predictions for galaxy evolution

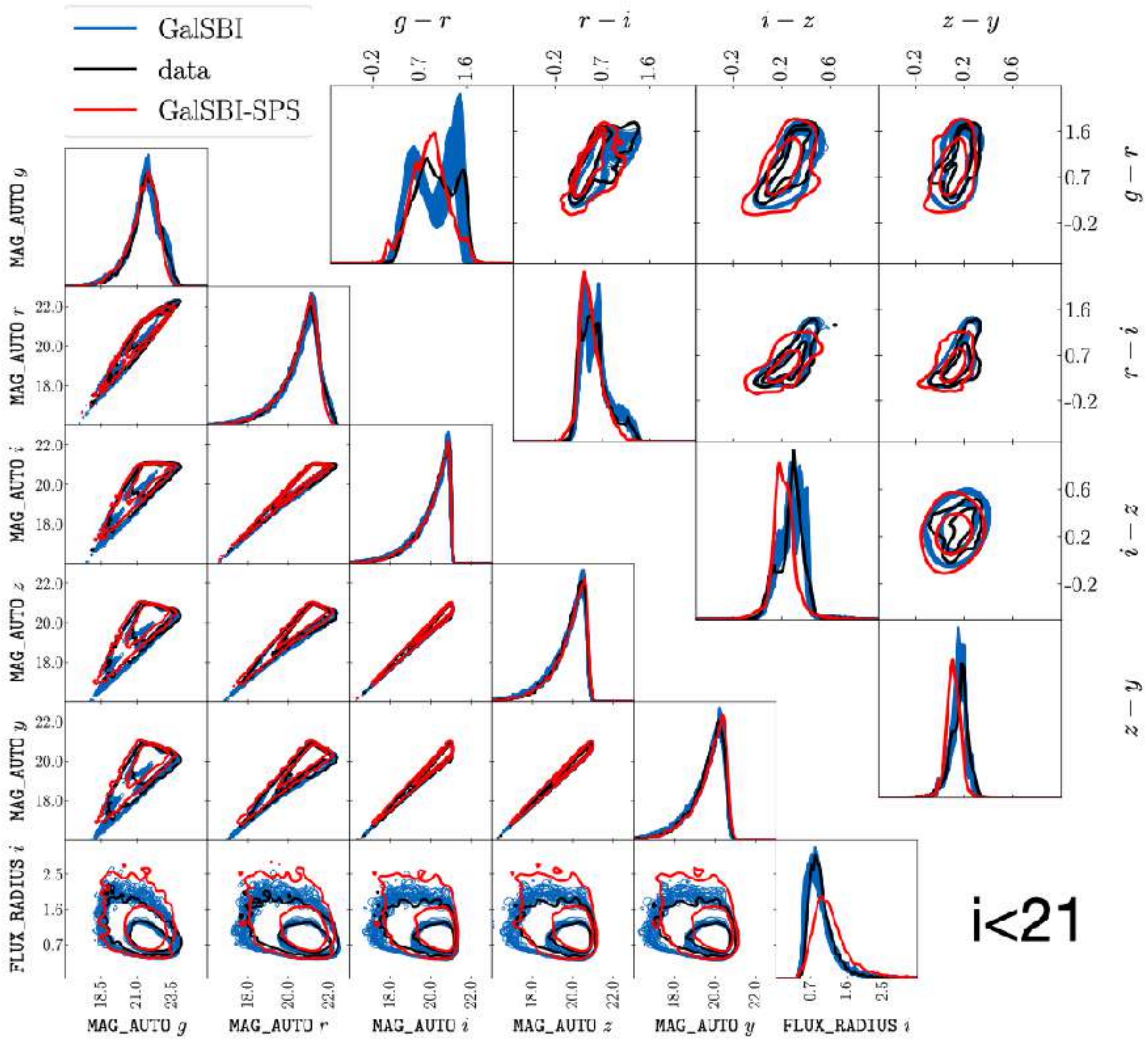




Tortorelli+20



i<21



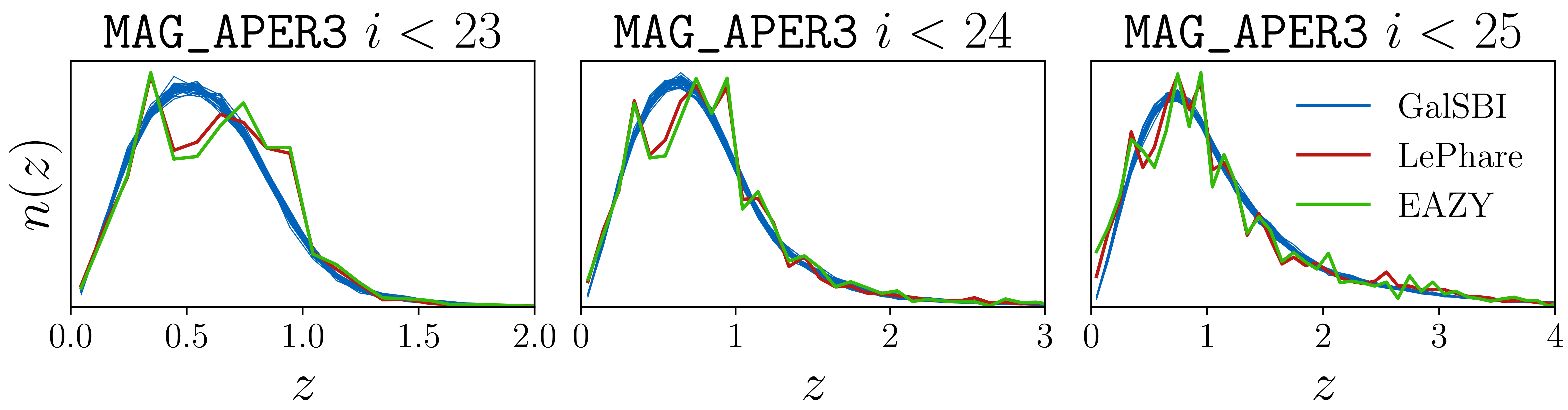
i<21

Tortorelli+25

Forward-modelling the 4C3R2 spectroscopic survey

$$P(\text{redshift success} \,|\, M_*, SFR, m, \sigma_{\text{obs.}}, \sigma_{\text{instr.}})$$

Posterior predictions for cosmology



	$m<23$	$m<24$	$m<25$
$\bar{z}_{\text{GalSBI}}$	0.596 ± 0.006	0.817 ± 0.009	1.143 ± 0.013
$\bar{z}_{\text{LP}}$	0.618	0.824	1.110
$\bar{z}_{\text{LP, mocks}}$	0.604 ± 0.012	0.823 ± 0.019	1.109 ± 0.019
$\bar{z}_{\text{E}}$	0.632	0.842	1.110
$\bar{z}_{\text{E, mock}}$	0.617 ± 0.012	0.841 ± 0.019	1.108 ± 0.018

		Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
S-III	$\bar{z}_{\text{GalSBI}}$	0.364 ± 0.006	0.547 ± 0.005	0.780 ± 0.003	1.225 ± 0.017	-
	$\bar{z}_{\text{LePhare}}$	0.362	0.554	0.827	1.211	-
	$\bar{z}_{\text{EAZY}}$	0.375	0.570	0.844	1.236	-
S-IV	$\bar{z}_{\text{GalSBI}}$	0.515 ± 0.018	0.722 ± 0.012	0.993 ± 0.009	1.34 ± 0.006	2.20 ± 0.05
	$\bar{z}_{\text{LePhare}}$	0.554	0.754	0.998	1.33	1.90
	$\bar{z}_{\text{EAZY}}$	0.558	0.766	1.011	1.34	1.87
	$\tilde{z}_{\text{GalSBI}}$	0.408 ± 0.004	0.677 ± 0.006	0.934 ± 0.005	1.291 ± 0.004	2.06 ± 0.03
	$\tilde{z}_{\text{LePhare}}$	0.406	0.707	0.945	1.293	2.06
	$\tilde{z}_{\text{EAZY}}$	0.410	0.728	0.959	1.327	2.05

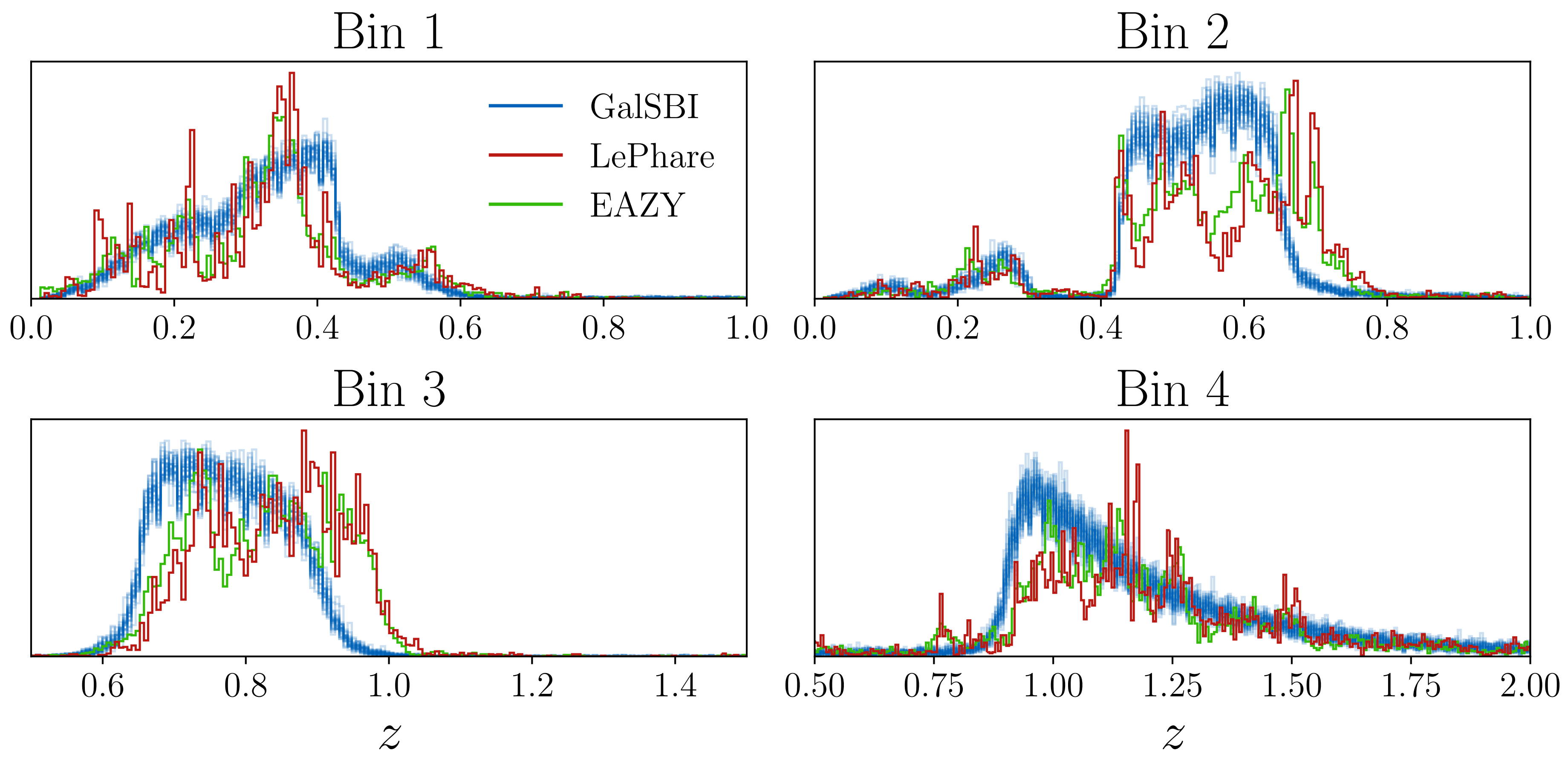


Figure 6: Tomographic bin assignment for the Stage-III survey setup.

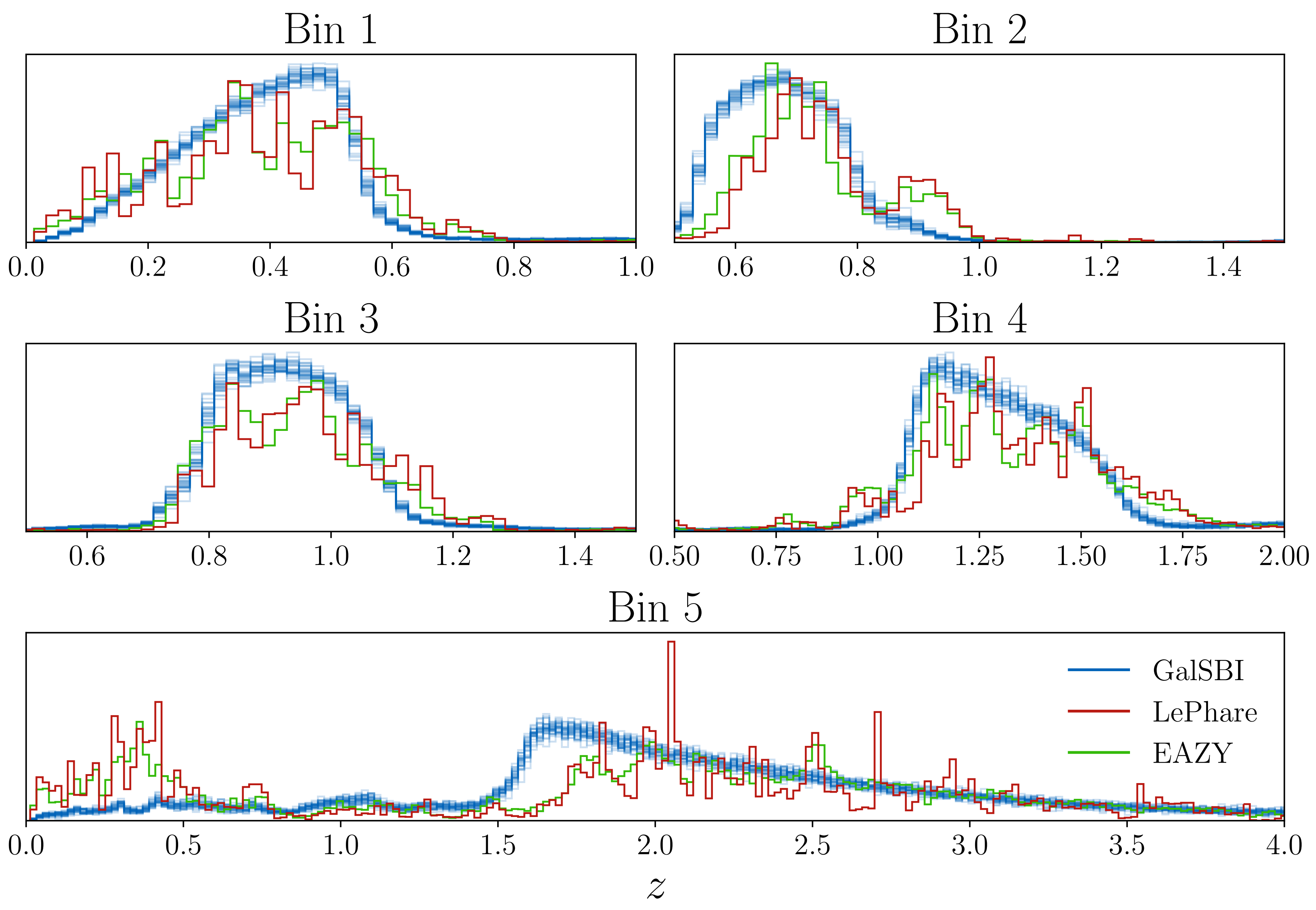


Figure 7: Tomographic bin assignment for the Stage-IV survey setup.

Fischbacher,Kacprzak,LT+24

Looking ahead… to Stage IV surveys

- GalSBI framework has the potential to obtain precise redshift distributions for a Stage-IV setup.
- However that requires constraints not only on photometric data, but also on spectroscopic data… And realistic clustering.
- Forward-modelled n(z) can already be used for a Stage IV precursor cosmological re-analysis (e.g. HSC).
- And in the future for LSST and Euclid.

galsbi on gitlab