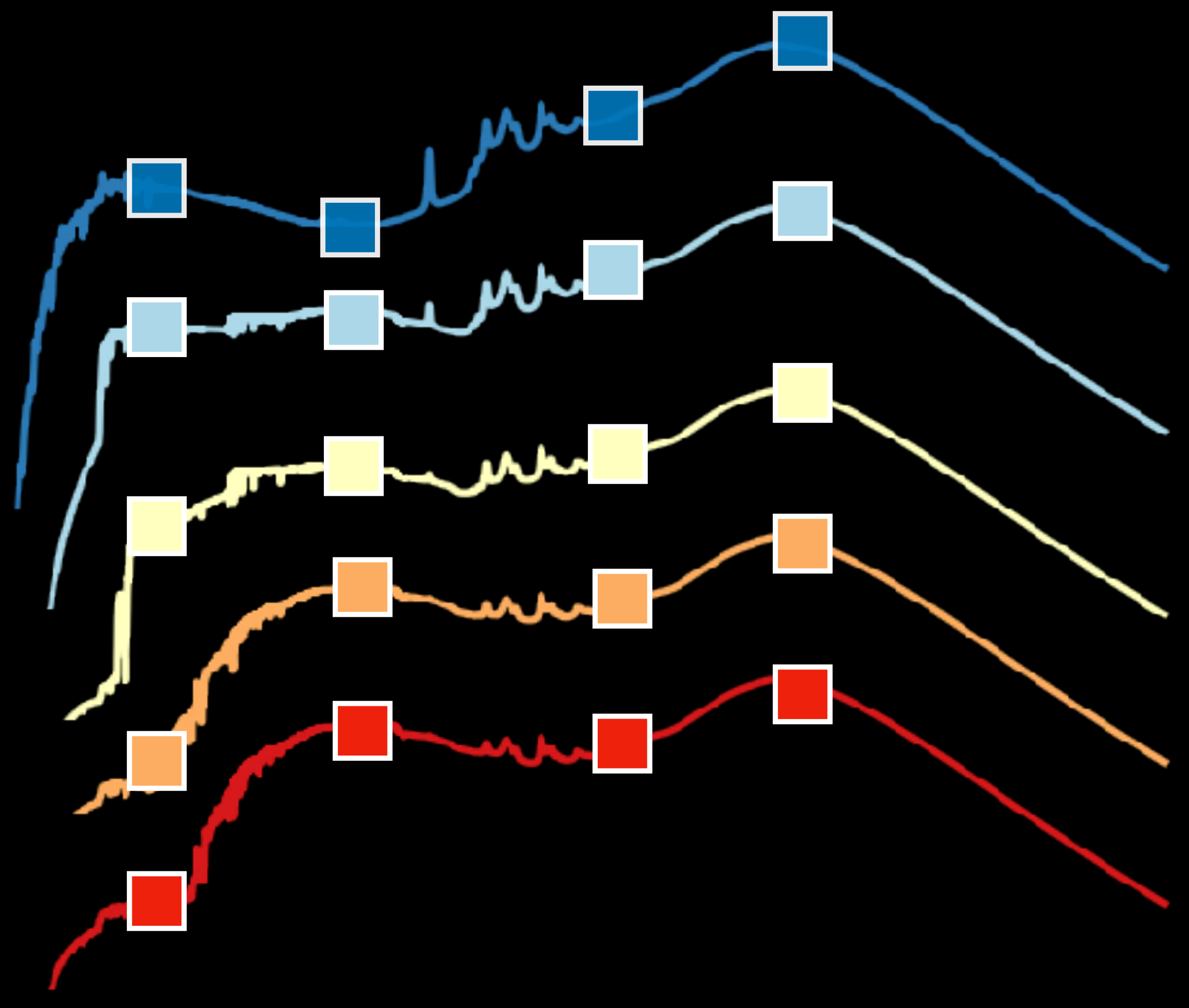
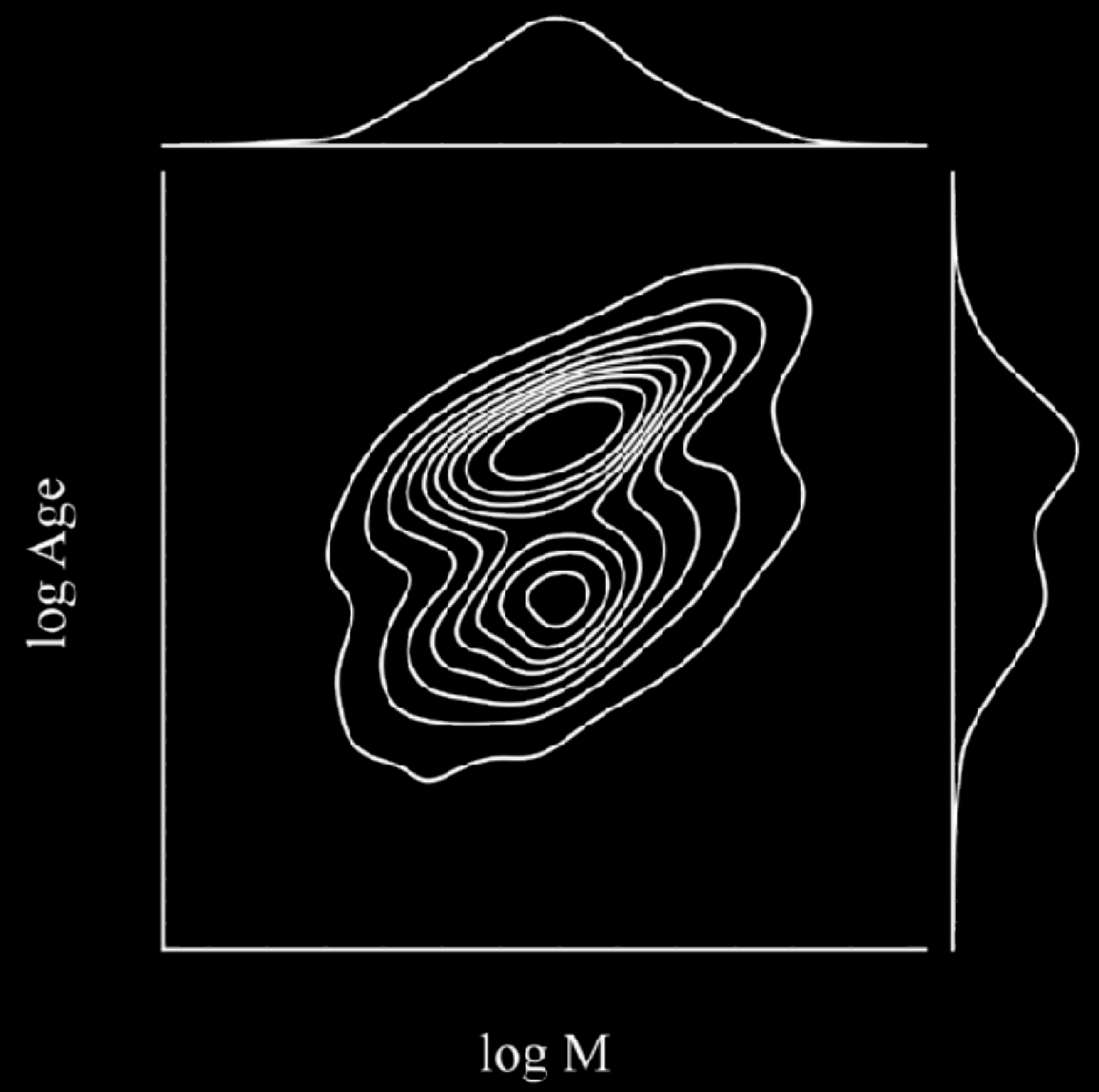
Population-Level Inference for Galaxy Properties from Spectral Energy Distribution



SEDs of a galaxy population



Population distribution of physical parameters

Jiaxuan Li 李嘉轩

Oct 29, 2024

Astro x Data Science Seminar, Yale

arxiv:2309.16958

with Peter Melchior, ChangHoon Hahn, Song Huang

How do we learn about galaxy properties?

Observable

Galaxy Spectral Energy Distribution (SED)

SED fitting

Stellar Population Synthesis

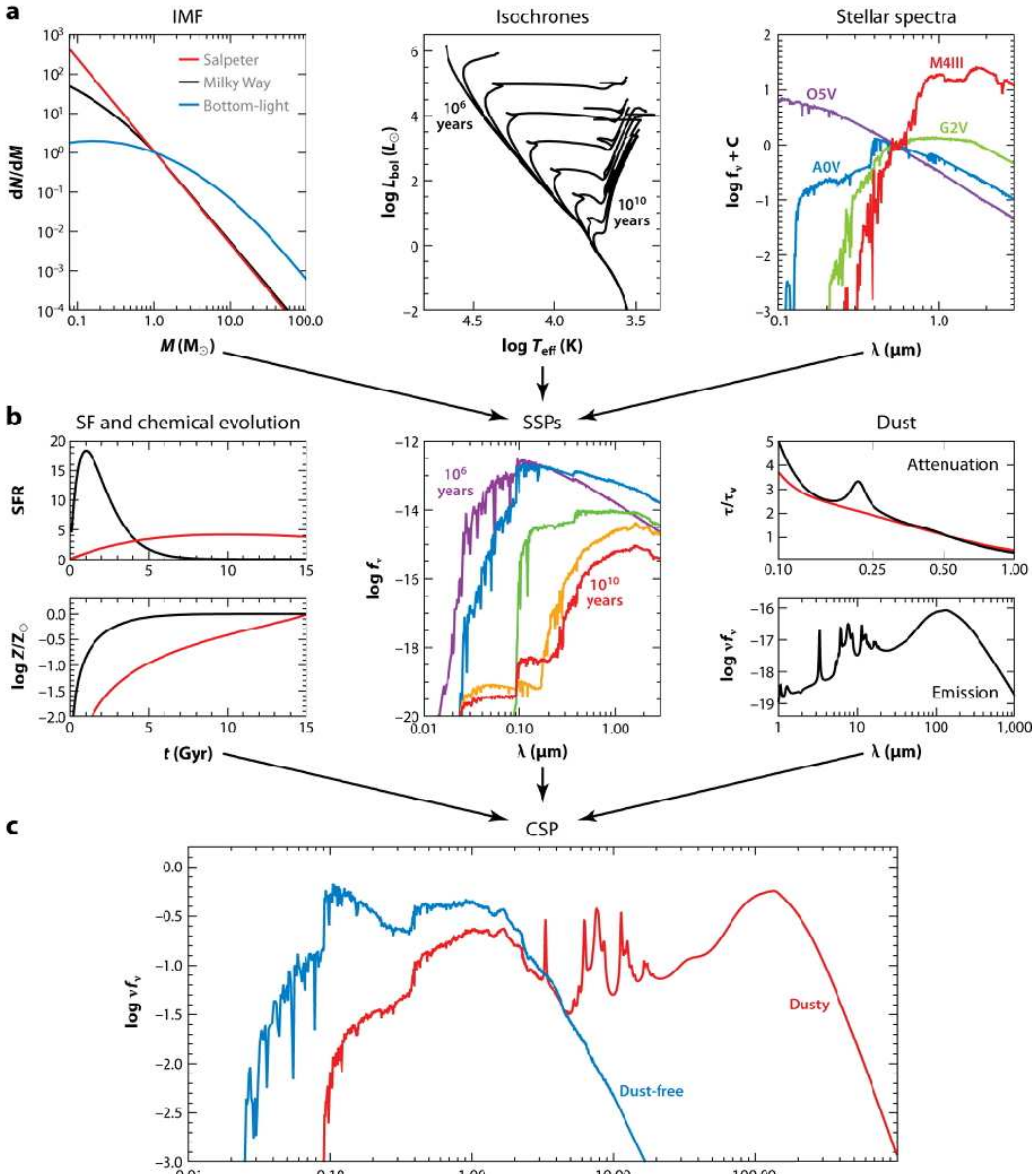
Thanks to Tinsley!

Ingredients

Initial Mass Function (IMF)

Star formation history (SFH)
Chemical enrichment history (ZH)

Dust attenuation and emission



Conroy (2013)

How do we understand galaxy properties?

Observable

Galaxy Spectral Energy Distribution (SED)

SED fitting ↓ ↑ **Stellar Population Synthesis**

Ingredients

Initial Mass Function (IMF)

Star formation history (SFH)
Chemical enrichment history (ZH)

Dust attenuation and emission

How to solve this problem?

Galaxy physical properties: θ

Observed data: X_o

Forward SPS model: $\hat{X} = F(\theta)$

Likelihood: $P(X|\theta) \sim \exp \sum (\hat{X} - X_o)^2/2\sigma^2$

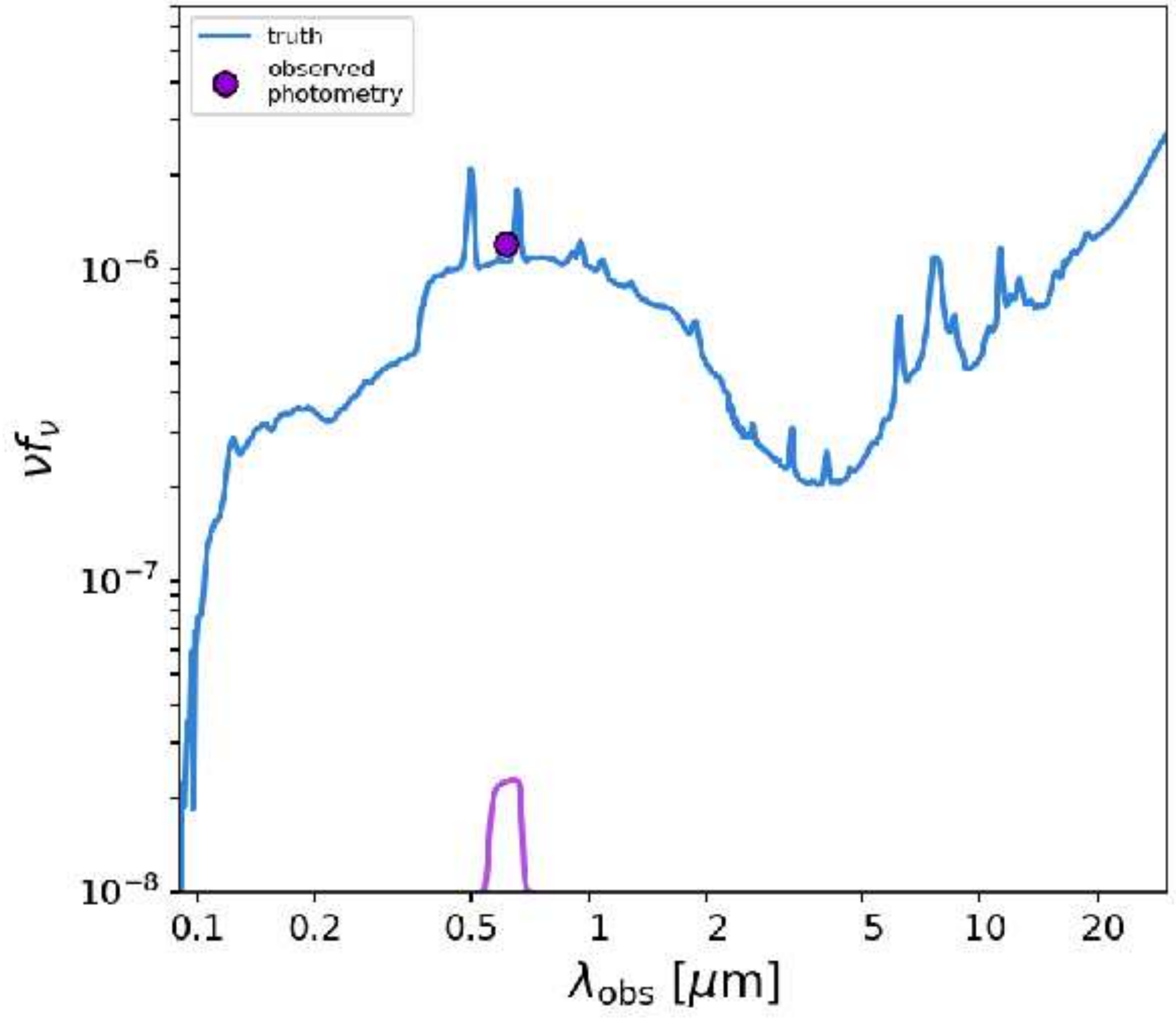
Posterior: $P(\theta|X) \propto P(X|\theta)P(\theta)$

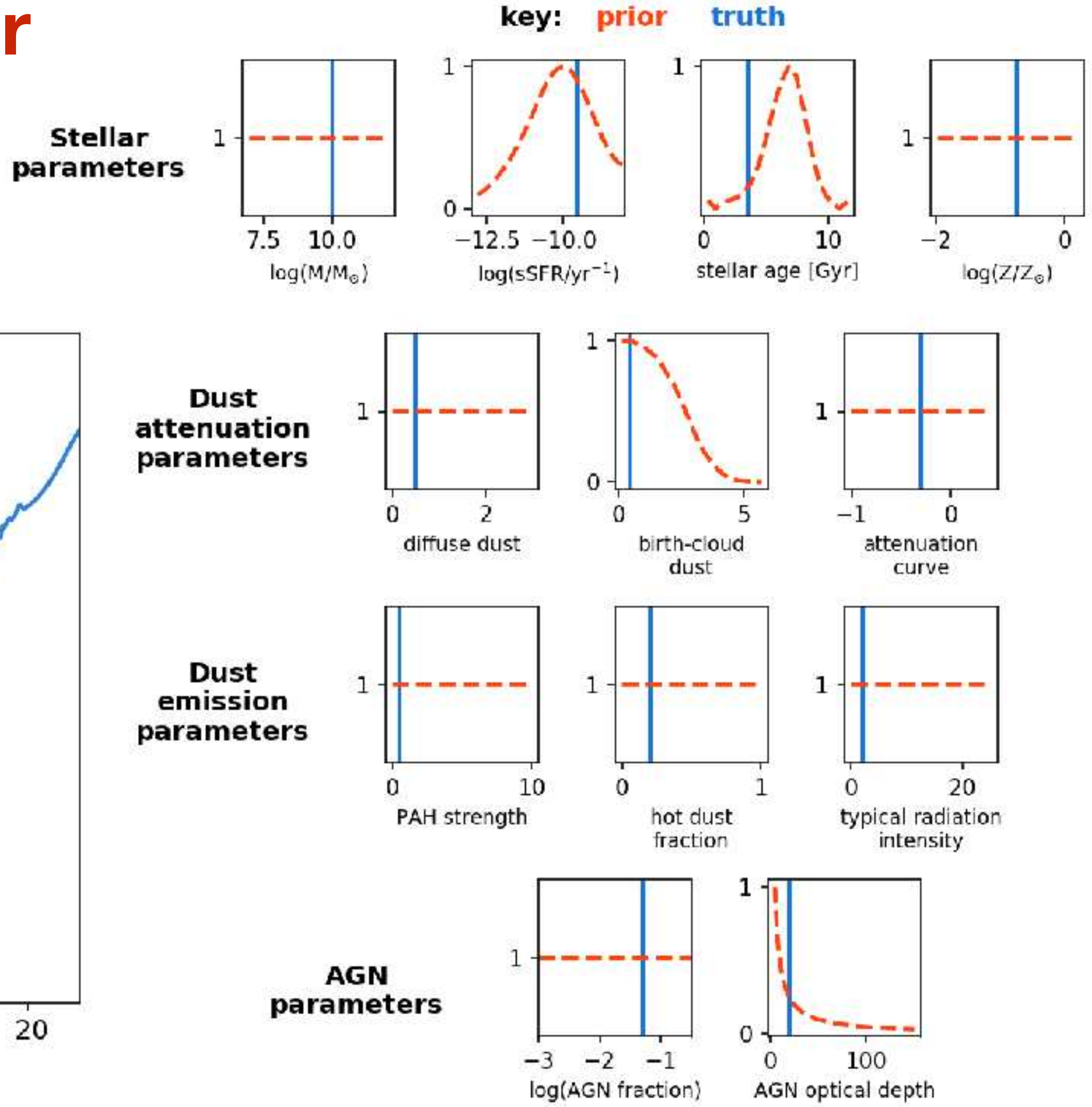
Conroy (2013)

State-of-the-art code

optical: SDSS *r*

Prospector





- High dimensionality: $n > 10$
- Slow model evaluation (FSPS)
- ~ 20 CPU-hours per galaxy (Leja et al. 2019)

Carnall et al. (2018)
Leja et al. (2021)
Johnson et al. (2021)

LSST Era

100 Million Galaxies × 20 CPU-hours

=

2 Billion CPU-hours

=

$ 80 Million*

Amazon AWS: $0.04 per CPU-hour

Solutions

Accelerate SPS

Neural emulator for FSPS (Alsing+20)

Differentiable SPS using JAX (Hearin+21)

Accelerate inference

Hamiltonian MC

Simulation-based Inference (Hahn+22, Wang+23)

Now you have the posteriors…

Wait, what astronomers do is —

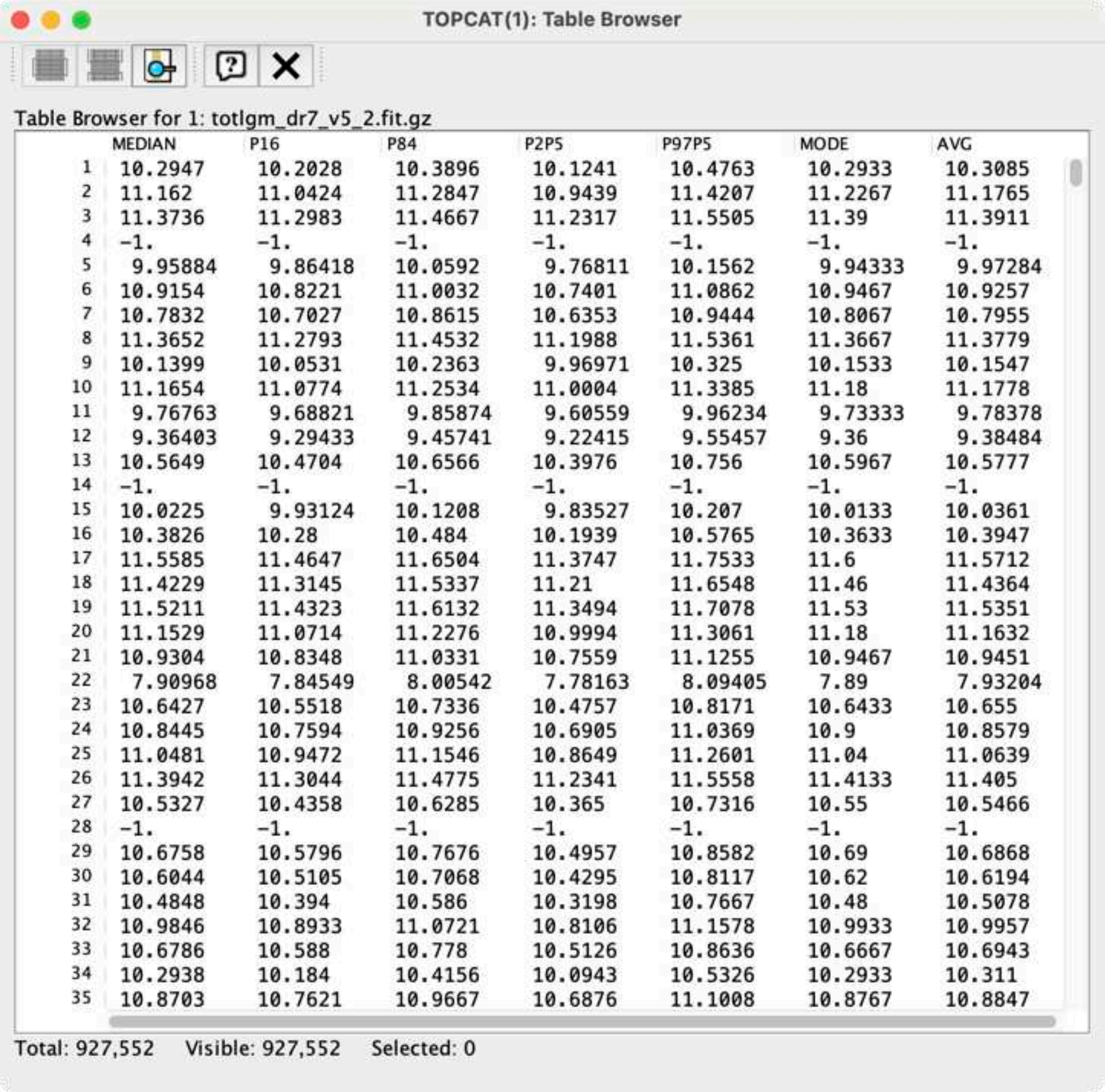
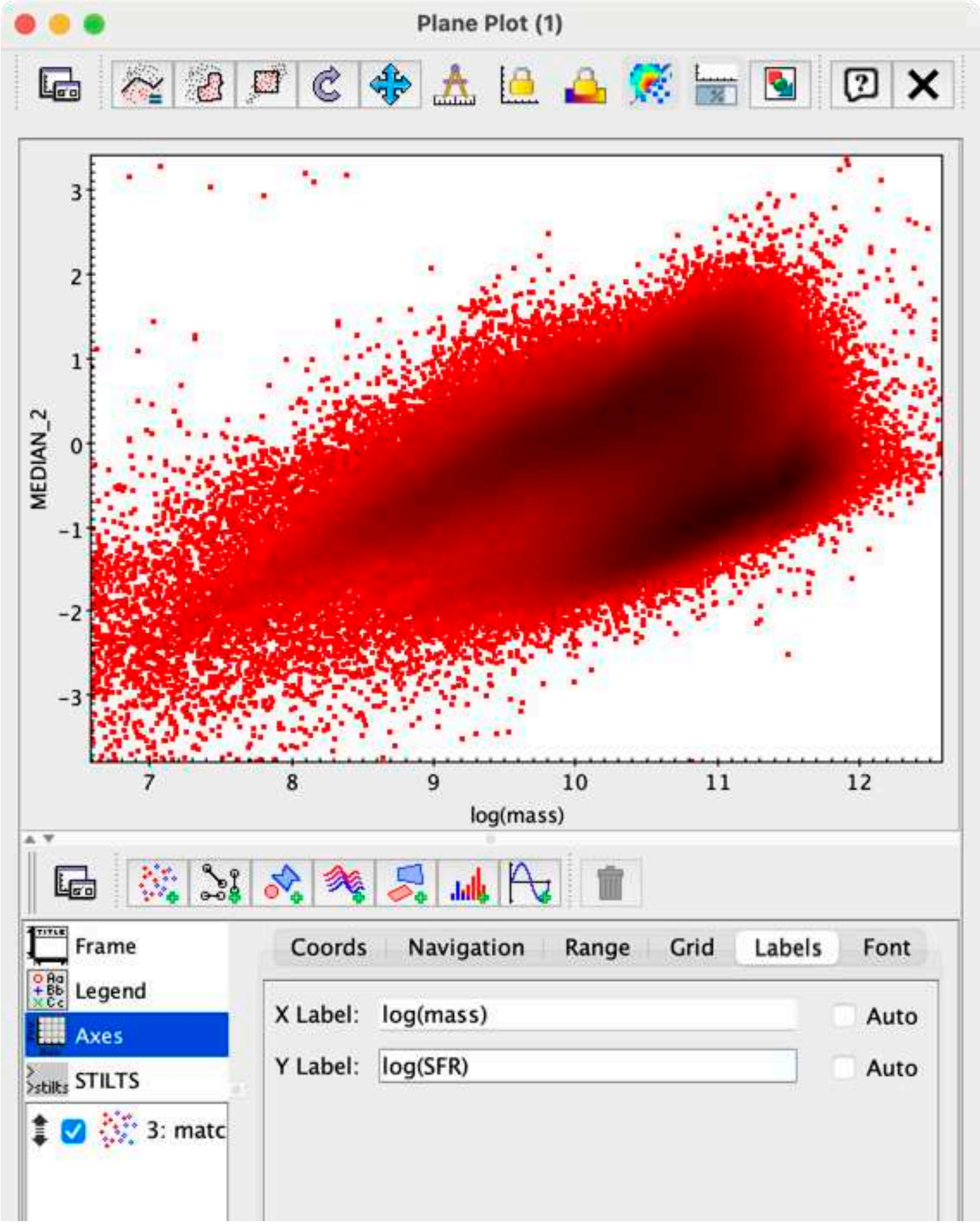

TOPCAT(1): Table Browser

Table Browser for 1: totlgm_dr7_v5_2.fit.gz

	MEDIAN	P16	P84	P2P5	P97P5	MODE	AVG
1	10.2947	10.2028	10.3896	10.1241	10.4763	10.2933	10.3085
2	11.162	11.0424	11.2847	10.9439	11.4207	11.2267	11.1765
3	11.3736	11.2983	11.4667	11.2317	11.5505	11.39	11.3911
4	-1.	-1.	-1.	-1.	-1.	-1.	-1.
5	9.95884	9.86418	10.0592	9.76811	10.1562	9.94333	9.97284
6	10.9154	10.8221	11.0032	10.7401	11.0862	10.9467	10.9257
7	10.7832	10.7027	10.8615	10.6353	10.9444	10.8067	10.7955
8	11.3652	11.2793	11.4532	11.1988	11.5361	11.3667	11.3779
9	10.1399	10.0531	10.2363	9.96971	10.325	10.1533	10.1547
10	11.1654	11.0774	11.2534	11.0004	11.3385	11.18	11.1778
11	9.76763	9.68821	9.85874	9.60559	9.96234	9.73333	9.78378
12	9.36403	9.29433	9.45741	9.22415	9.55457	9.36	9.38484
13	10.5649	10.4704	10.6566	10.3976	10.756	10.5967	10.5777
14	-1.	-1.	-1.	-1.	-1.	-1.	-1.
15	10.0225	9.93124	10.1208	9.83527	10.207	10.0133	10.0361
16	10.3826	10.28	10.484	10.1939	10.5765	10.3633	10.3947
17	11.5585	11.4647	11.6504	11.3747	11.7533	11.6	11.5712
18	11.4229	11.3145	11.5337	11.21	11.6548	11.46	11.4364
19	11.5211	11.4323	11.6132	11.3494	11.7078	11.53	11.5351
20	11.1529	11.0714	11.2276	10.9994	11.3061	11.18	11.1632
21	10.9304	10.8348	11.0331	10.7559	11.1255	10.9467	10.9451
22	7.90968	7.84549	8.00542	7.78163	8.09405	7.89	7.93204
23	10.6427	10.5518	10.7336	10.4757	10.8171	10.6433	10.655
24	10.8445	10.7594	10.9256	10.6905	11.0369	10.9	10.8579
25	11.0481	10.9472	11.1546	10.8649	11.2601	11.04	11.0639
26	11.3942	11.3044	11.4775	11.2341	11.5558	11.4133	11.405
27	10.5327	10.4358	10.6285	10.365	10.7316	10.55	10.5466
28	-1.	-1.	-1.	-1.	-1.	-1.	-1.
29	10.6758	10.5796	10.7676	10.4957	10.8582	10.69	10.6868
30	10.6044	10.5105	10.7068	10.4295	10.8117	10.62	10.6194
31	10.4848	10.394	10.586	10.3198	10.7667	10.48	10.5078
32	10.9846	10.8933	11.0721	10.8106	11.1578	10.9933	10.9957
33	10.6786	10.588	10.778	10.5126	10.8636	10.6667	10.6943
34	10.2938	10.184	10.4156	10.0943	10.5326	10.2933	10.311
35	10.8703	10.7621	10.9667	10.6876	11.1008	10.8767	10.8847

Total: 927,552 Visible: 927,552 Selected: 0

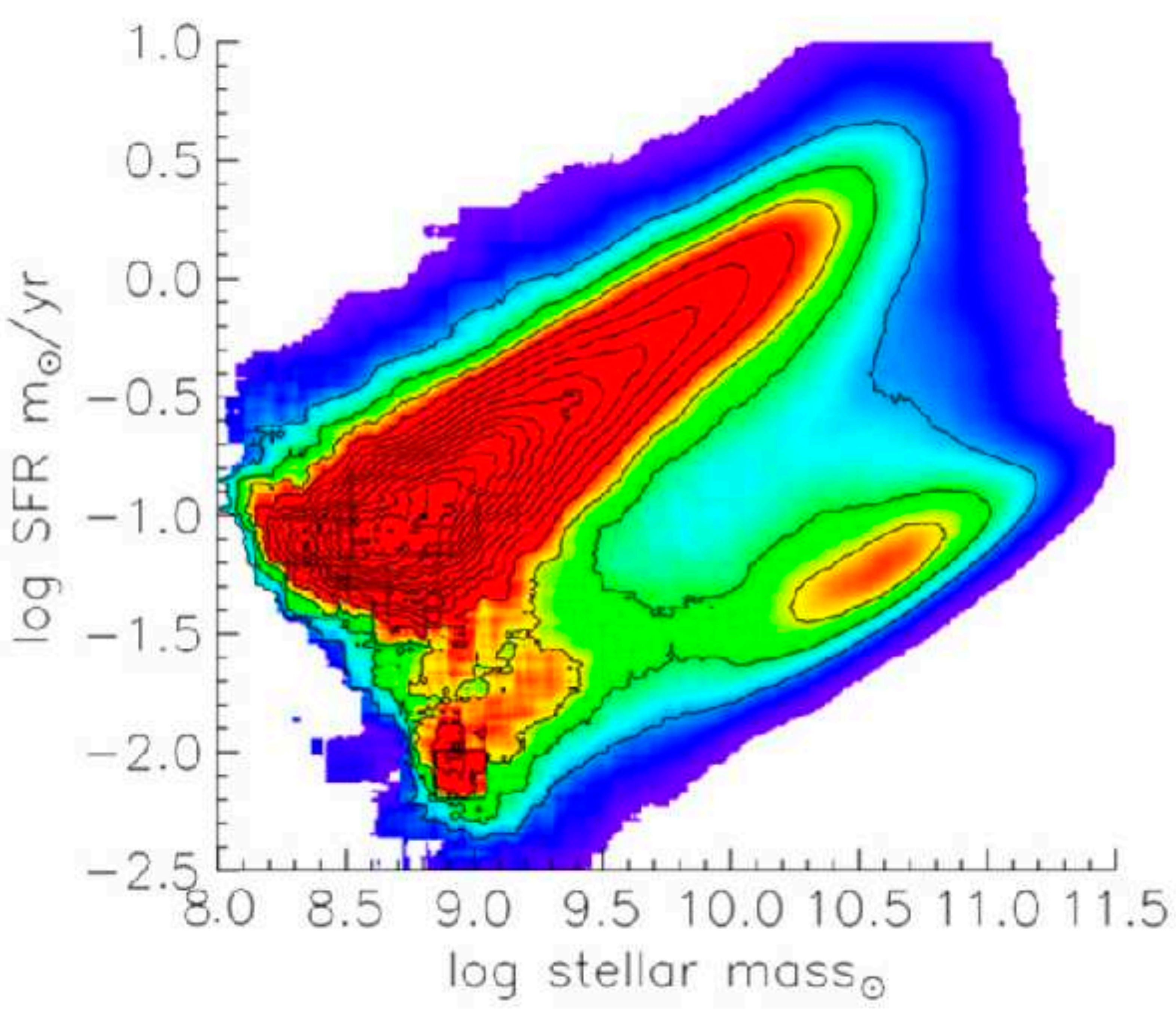
Download catalog



Plot the median

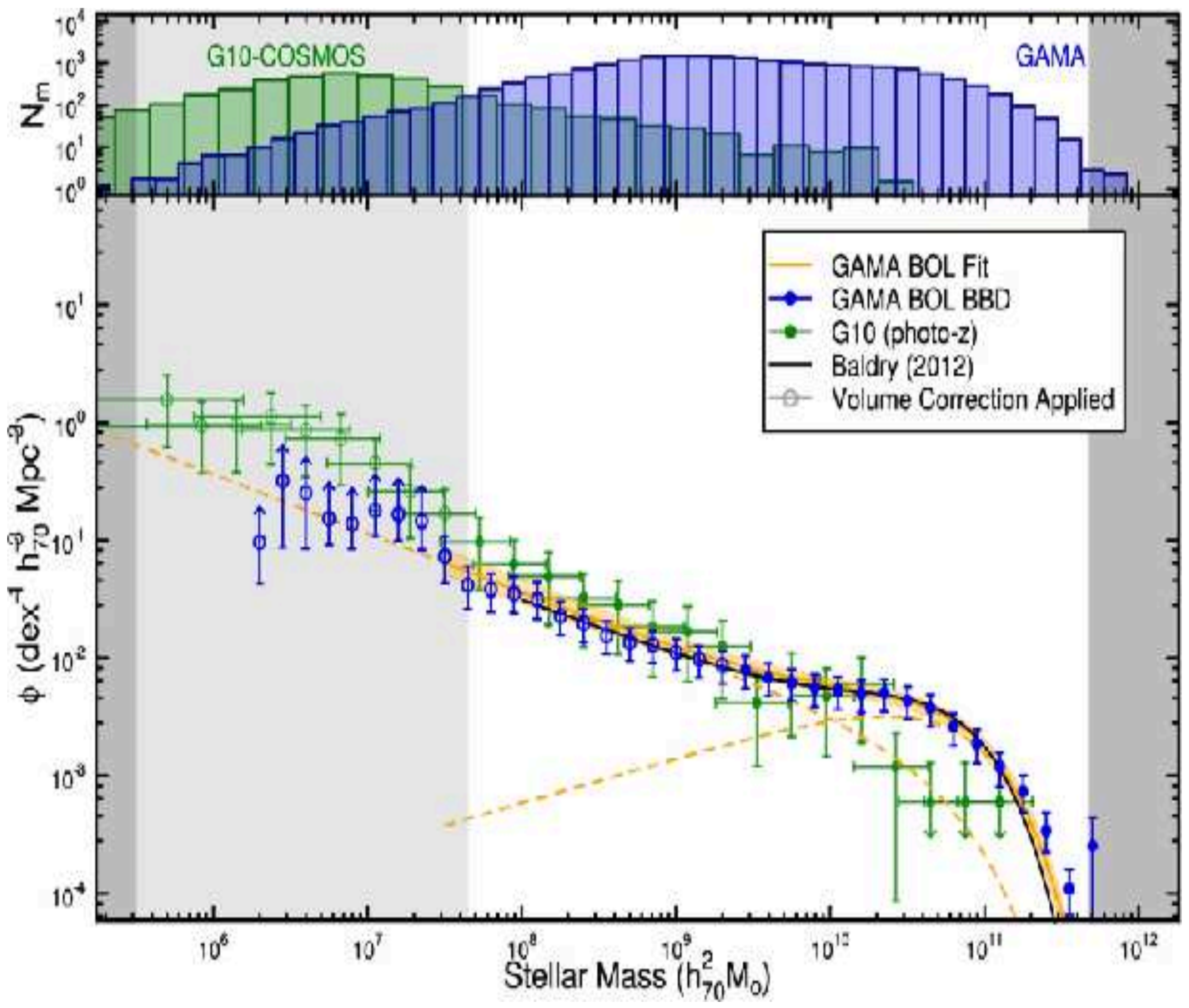
What astronomers care is —
scaling relations

Star-Forming Main Sequence



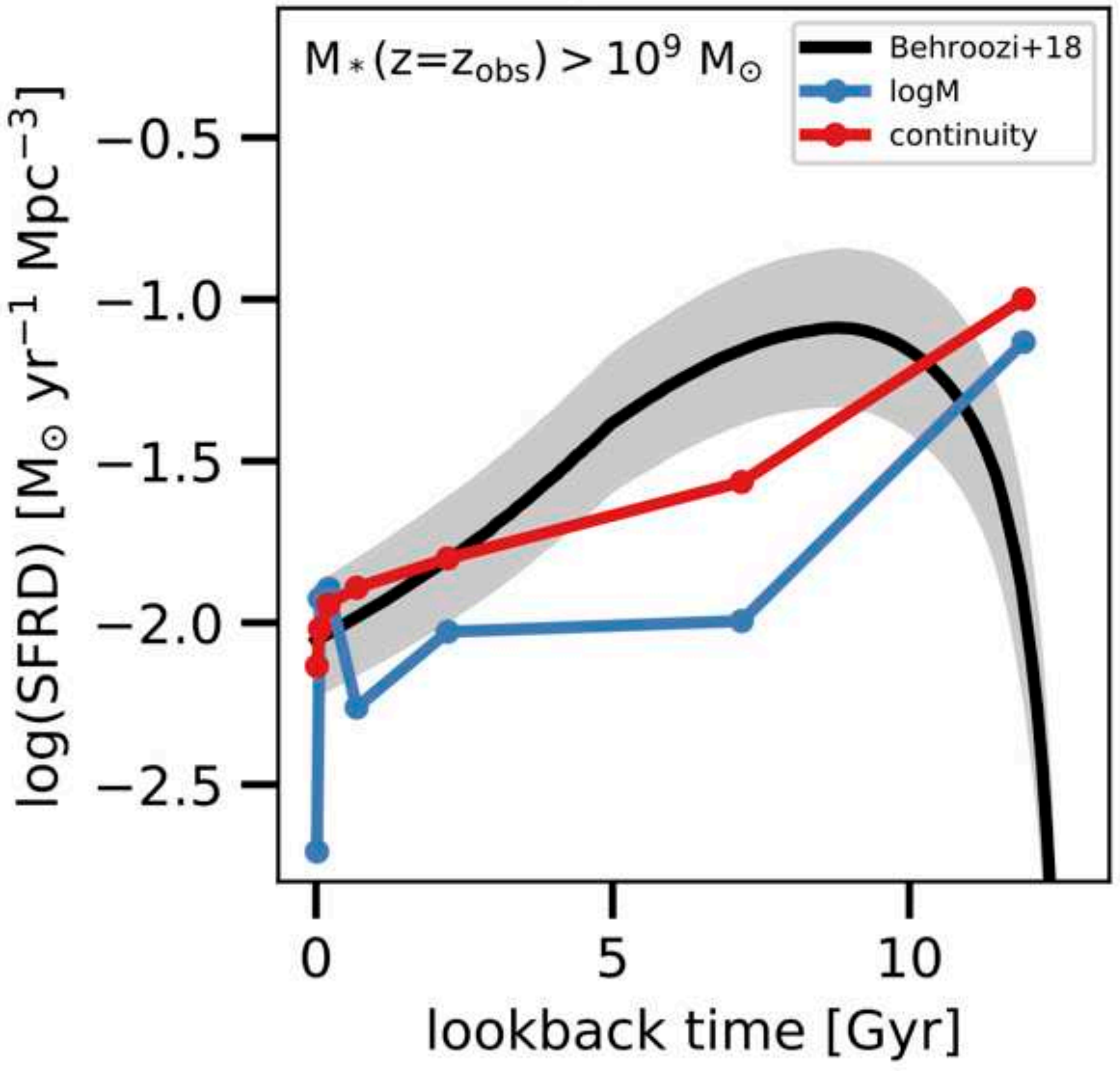
Renzini & Peng (2015)

Stellar-Mass Function



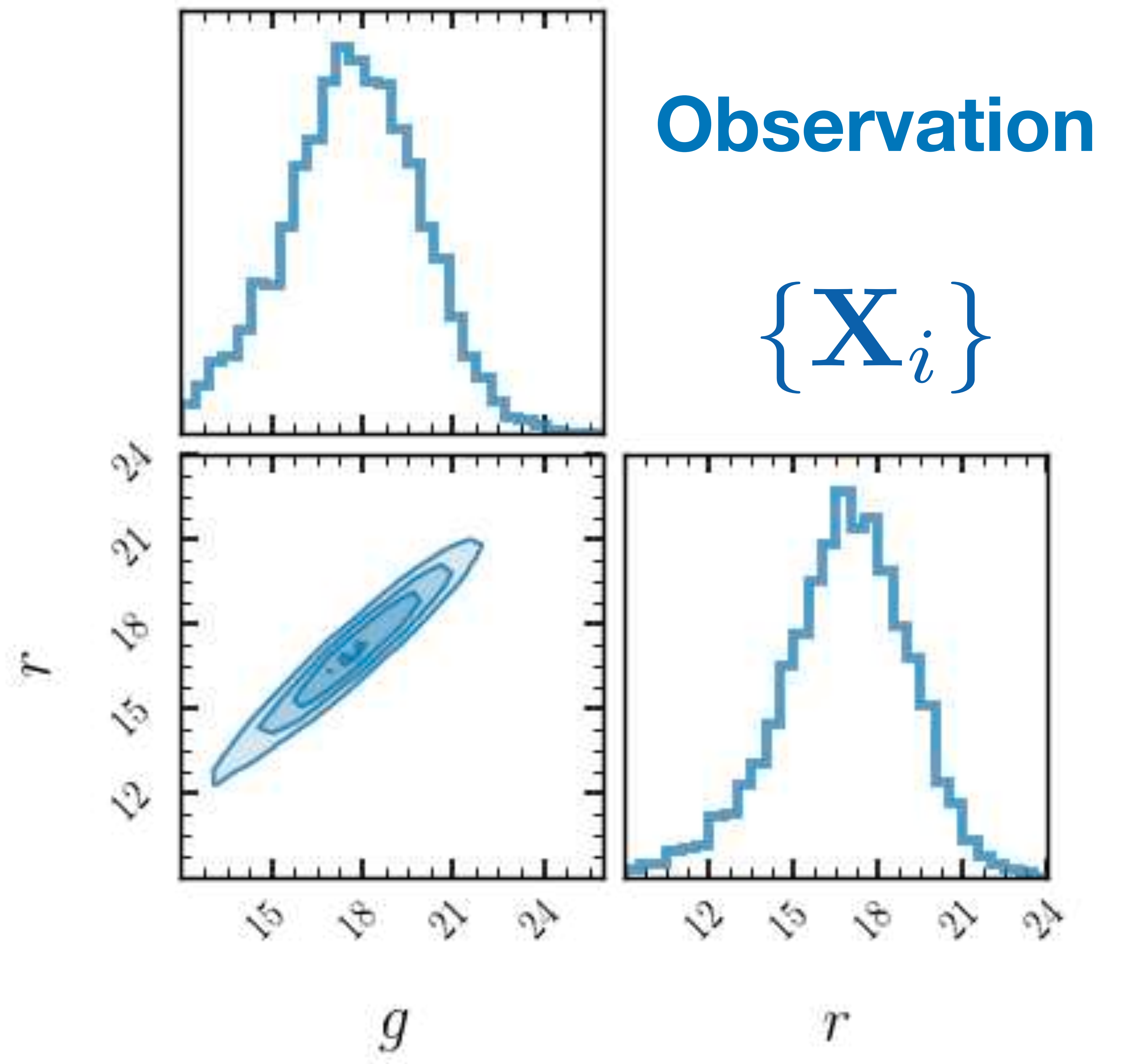
Wright et al. (2017)

Cosmic SF history



Leja et al. (2018)

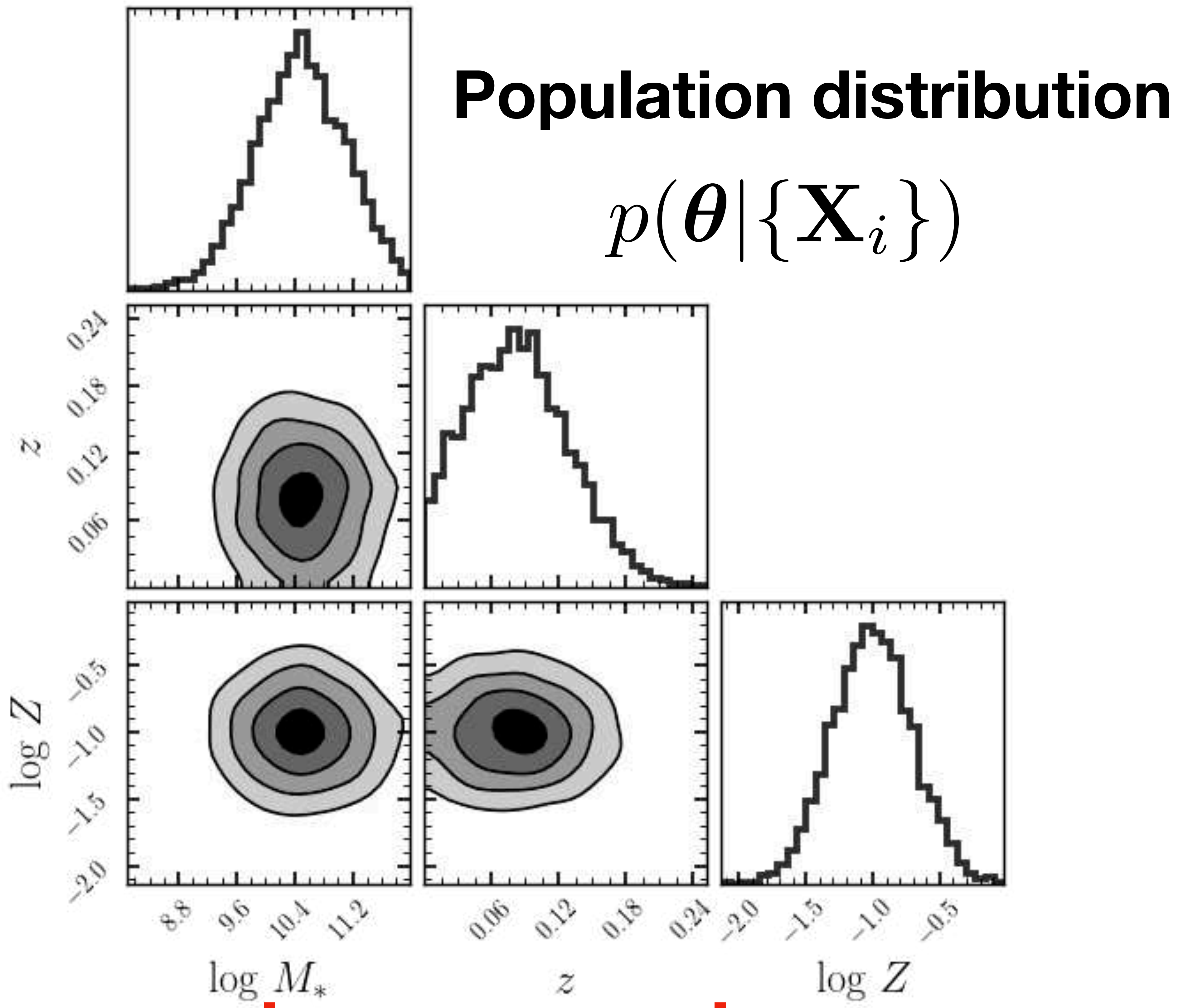
A faster way to directly get population-level information?



Photometry Dataset

1 million galaxies

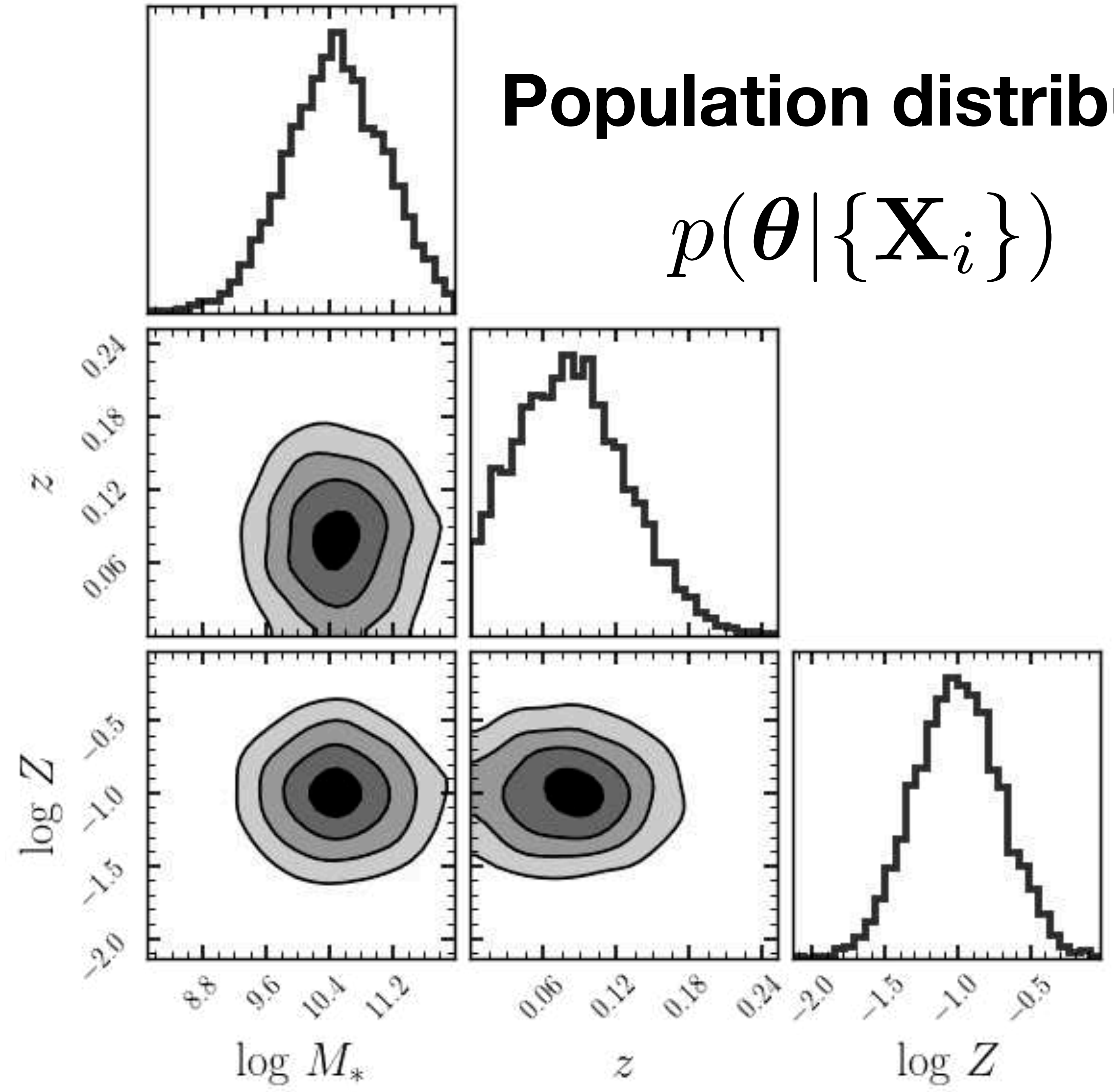
POPSED

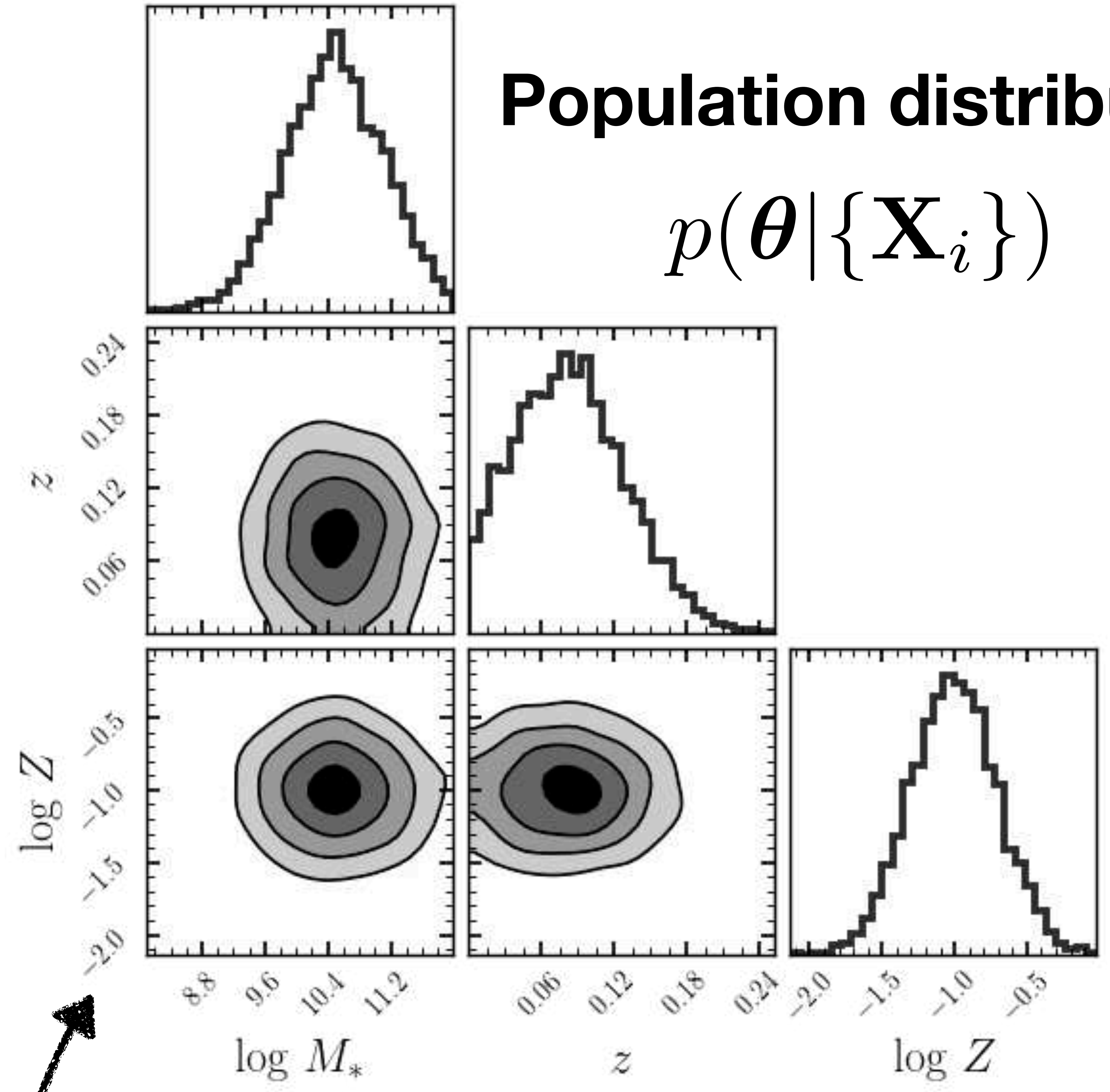


Galaxy Population & scaling relations

Population distribution

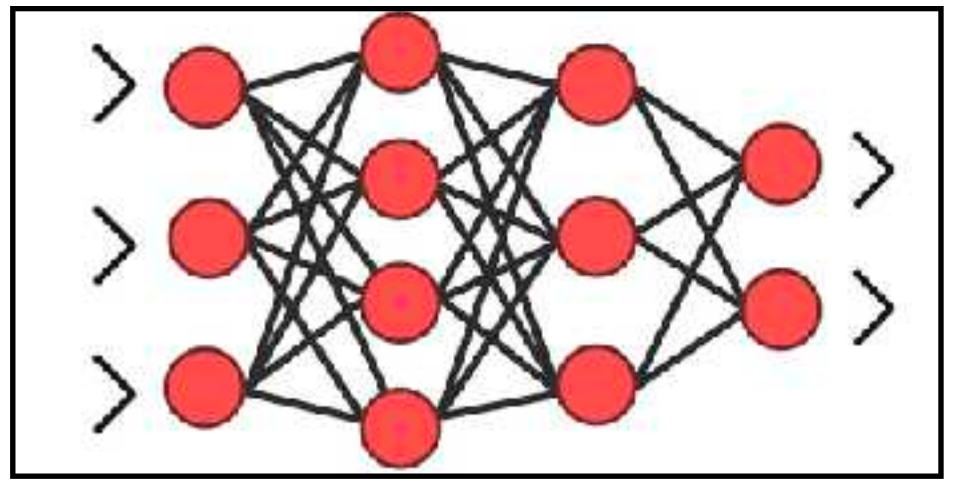
$$p(\boldsymbol{\theta}|\{\mathbf{X}_i\})$$





Population distribution

$$p(\boldsymbol{\theta}|\{\mathbf{X}_i\})$$



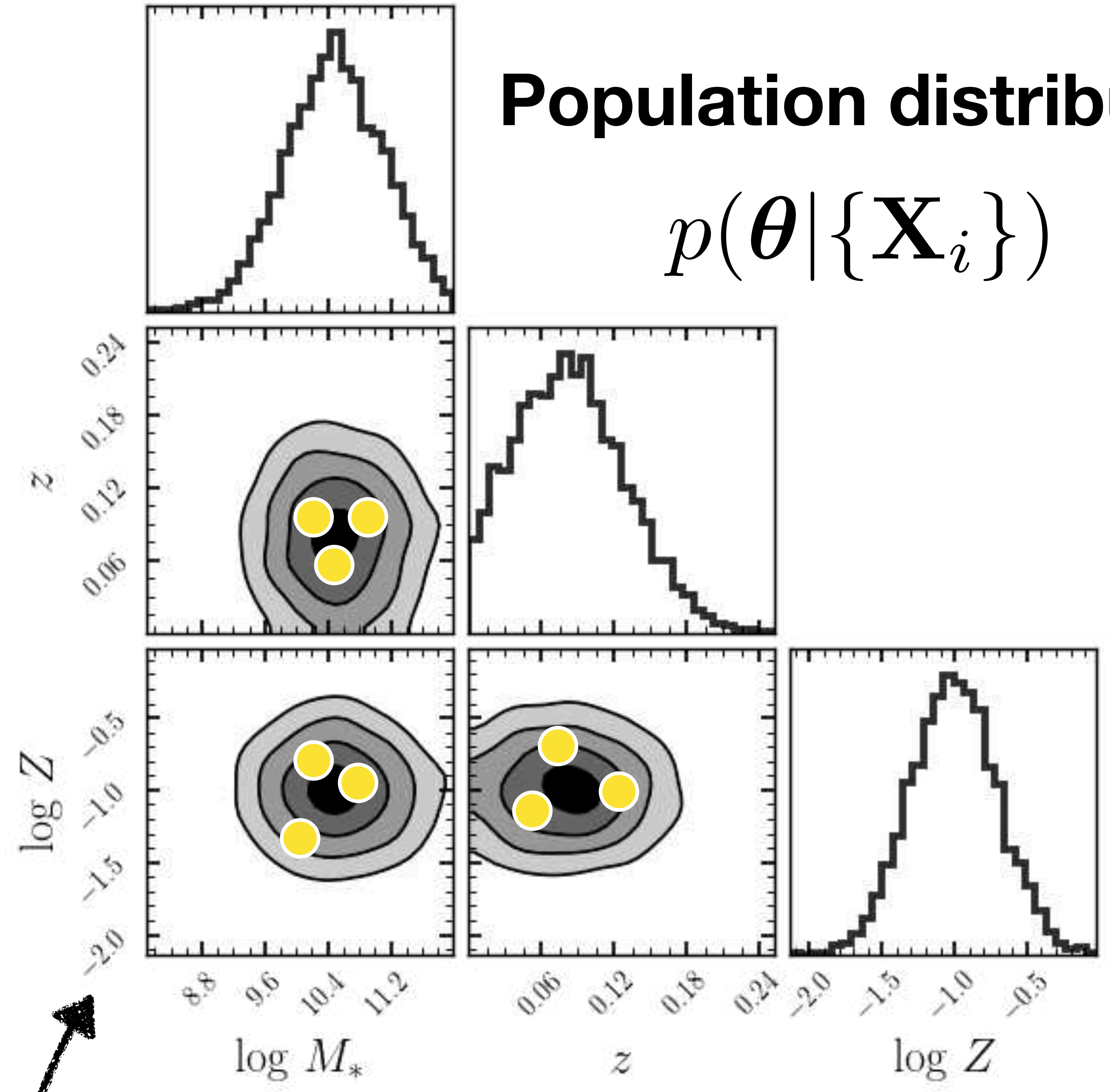
Probability Density Estimator $q_\phi(\boldsymbol{\theta})$

e.g., Gaussian mixture, normalizing flows, diffusion model

As long as you can draw samples from it

Population distribution

$$p(\boldsymbol{\theta}|\{\mathbf{X}_i\})$$



Sample from normalizing flow

$$\boldsymbol{\theta}_j^{\phi} \sim q_{\phi}(\boldsymbol{\theta})$$

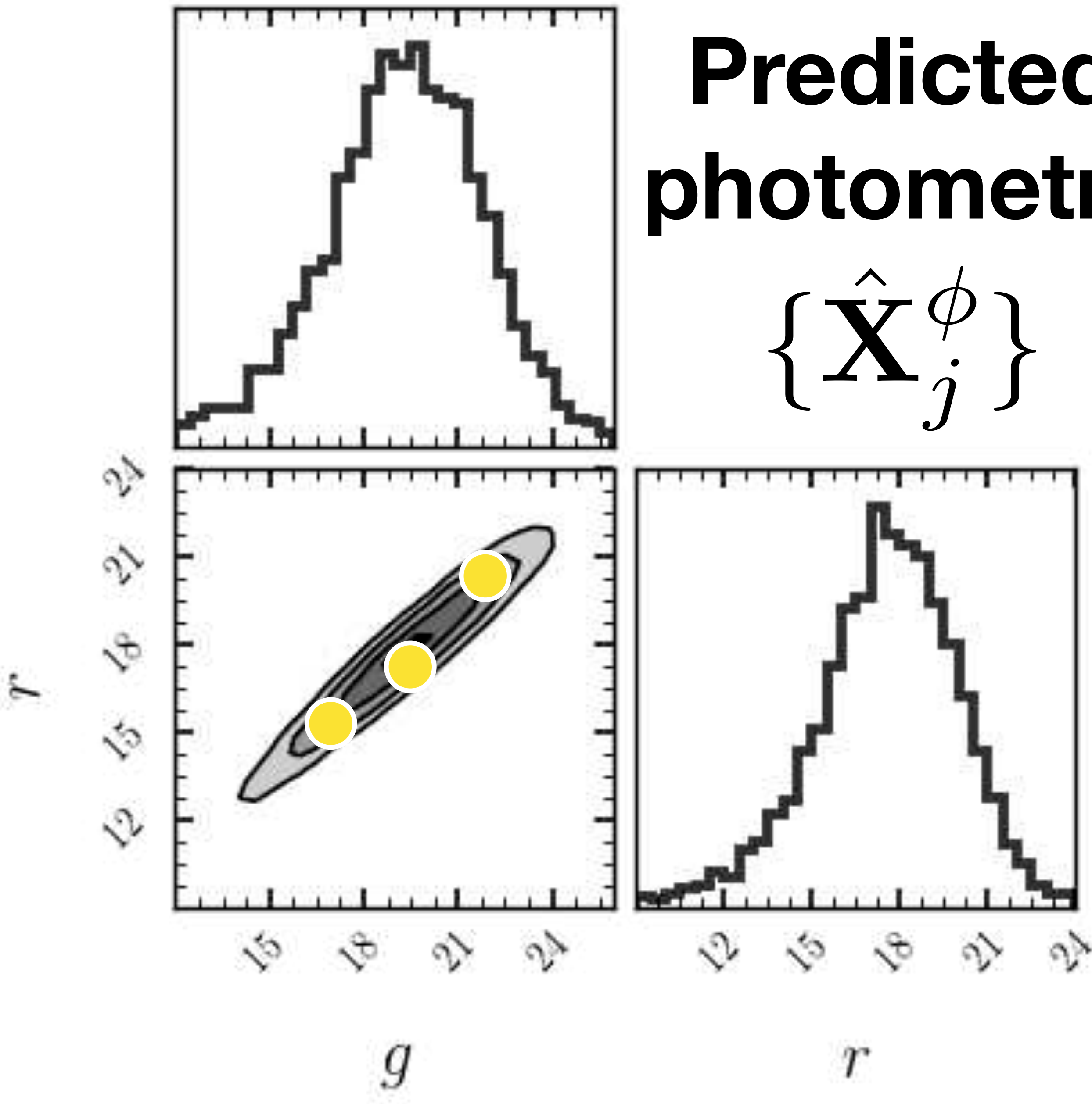
Stellar population synthesis

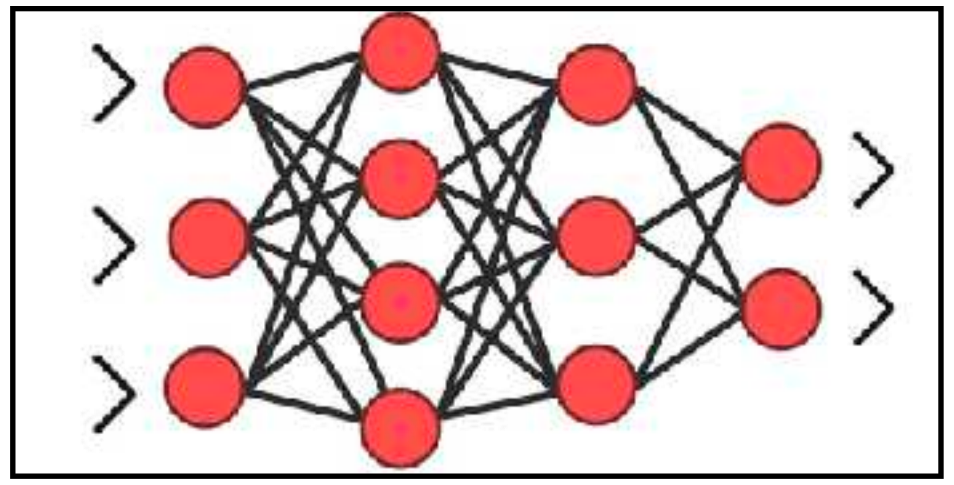
$$\hat{\mathbf{X}}_j^{\phi} = F(\boldsymbol{\theta}_j^{\phi})$$

+ obs uncertainty + selection

Predicted photometry

$$\{\hat{\mathbf{X}}_j^{\phi}\}$$

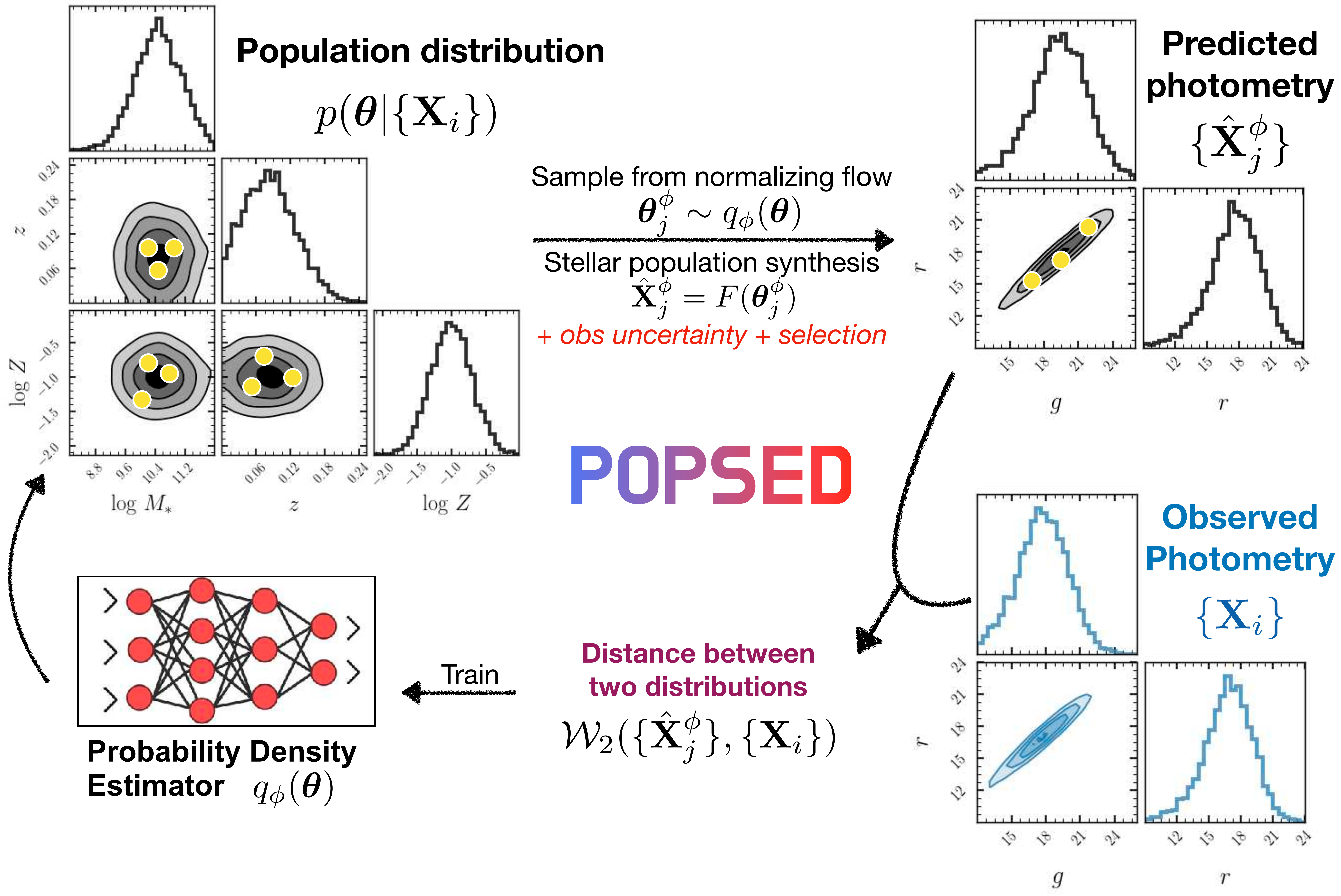




Probability Density Estimator $q_{\phi}(\boldsymbol{\theta})$

e.g., Gaussian mixture, normalizing flows, diffusion model

As long as you can draw samples from it



Population distribution
$p(\boldsymbol{\theta}|\{\mathbf{X}_i\})$
Sample from normalizing flow
$\boldsymbol{\theta}_j^{\phi} \sim q_{\phi}(\boldsymbol{\theta})$
Stellar population synthesis
$\hat{\mathbf{X}}_j^{\phi} = F(\boldsymbol{\theta}_j^{\phi})$
+ obs uncertainty + selection
Predicted photometry
$\{\hat{\mathbf{X}}_j^{\phi}\}$
POPSED
Observed Photometry
$\{\mathbf{X}_i\}$
Distance between two distributions
$\mathcal{W}_2(\{\hat{\mathbf{X}}_j^{\phi}\}, \{\mathbf{X}_i\})$
Train
Probability Density Estimator $q_{\phi}(\boldsymbol{\theta})$
log M_*
z
log Z
g
r

Stellar Population Synthesis

$$\hat{\mathbf{X}}_j^{\phi} = F(\boldsymbol{\theta}_j^{\phi})$$ F is the SPS model

(or your favorite SPS model)

- IMF: fixed Chabrier (2003)
- SFH: linear combination of bases + burst
- ZH: constant metallicity
- Dust: optical depths + Calzetti attenuation
- Mass, redshift

Parameter	Description
z	Redshift
$\log(M_\star/M_\odot)$	log10 stellar mass
$\beta_1, \beta_2, \beta_3, \beta_4$	Coefficients of SFH bases (Eq. 1)
$t_{\rm burst}$ [Gyr]	The lookback time when star formation burst happens (Eq. 1)
$f_{\rm burst}$	The fraction of total stellar mass formed in the star formation burst (Eq. 1)
$\log(Z_\star/Z_\odot)$	stellar metallicity ($Z_\odot = 0.019$)
$n_{\rm dust}$	The power-law index of the Calzetti et al. (2000) attenuation curve
τ_1	Birth-cloud dust optical depth
τ_2	Diffuse dust optical depth

But do we wanna use FSPS?

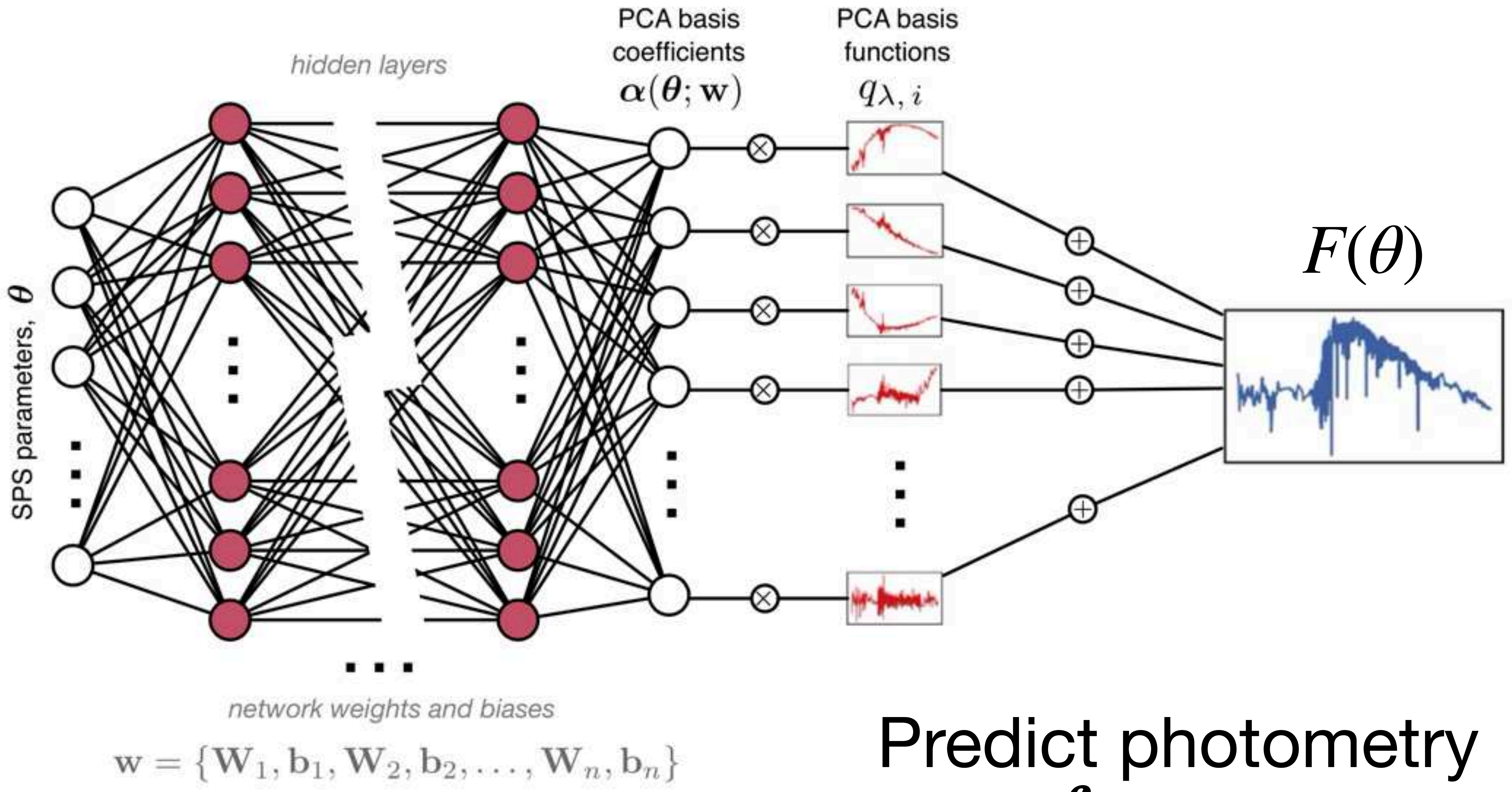
Hahn et al. (2023)

Emulator: accelerate model evaluation

$$L_\lambda = F_{\rm SPS}(\theta) \approx f_{\rm NN}(\theta)$$

Spectrum emulator based on neural network

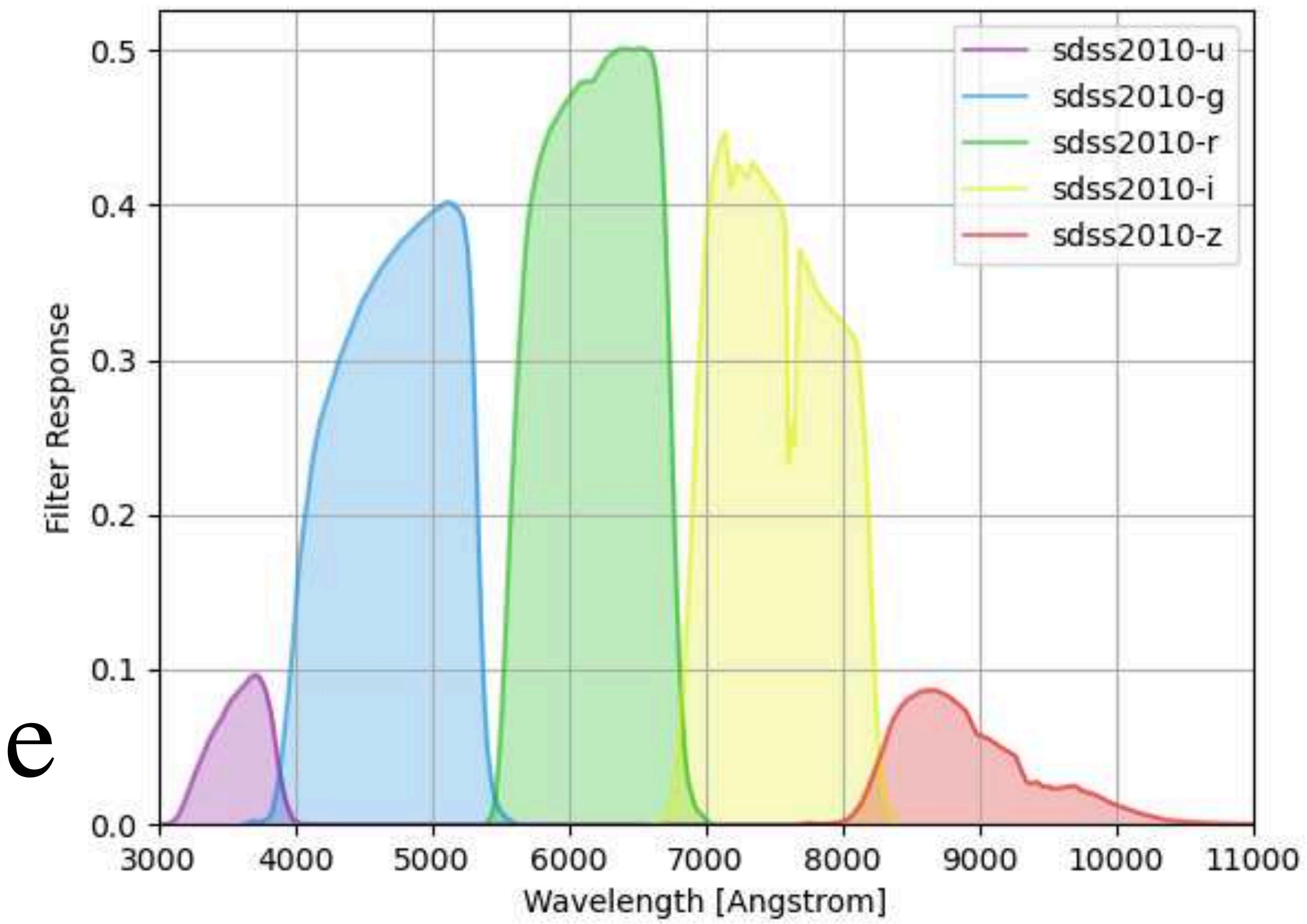
Emulator is $\sim 10^{3-4}$ faster than direct SPS computation

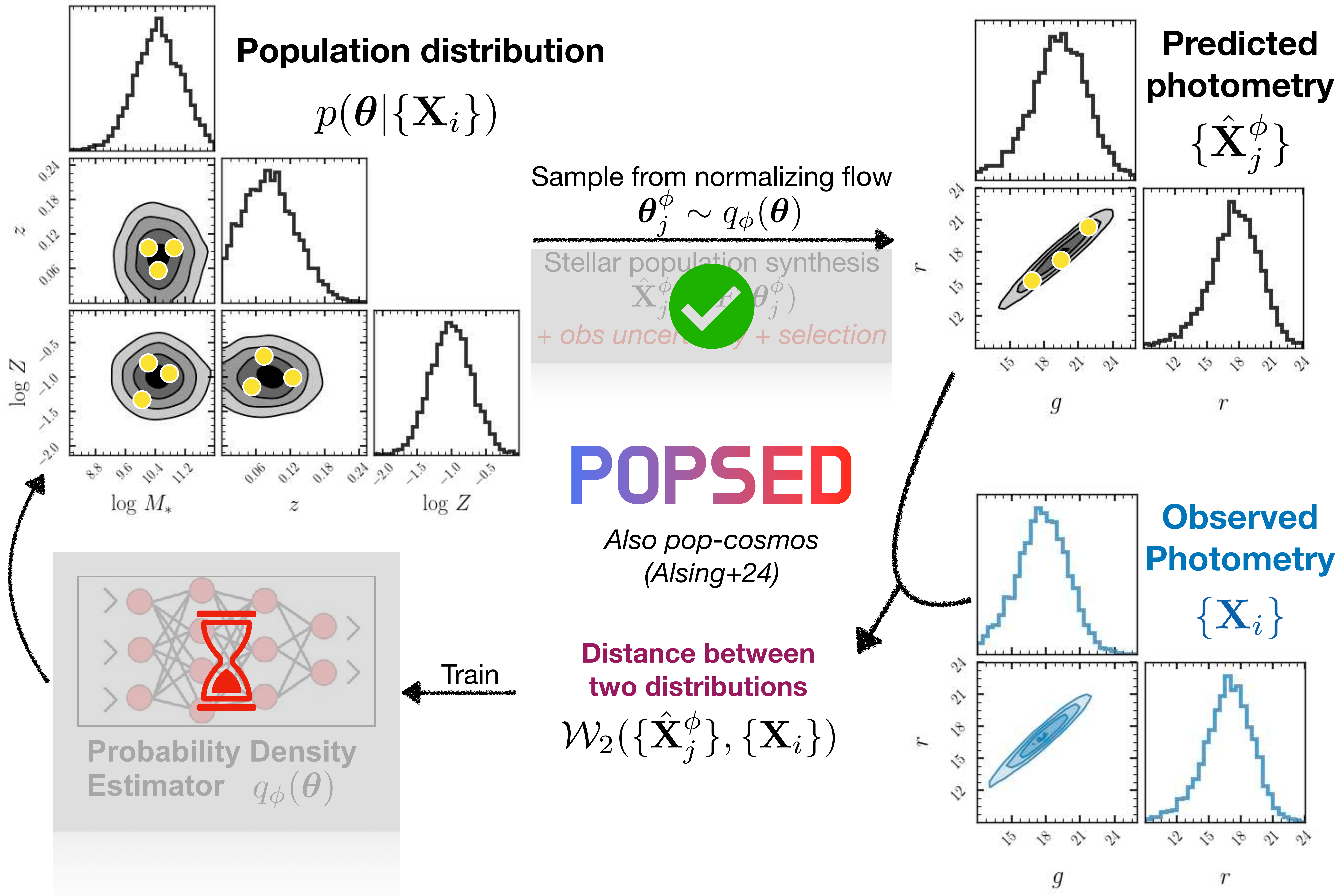


Speculator (Alsing et al. 2020)

Predict photometry

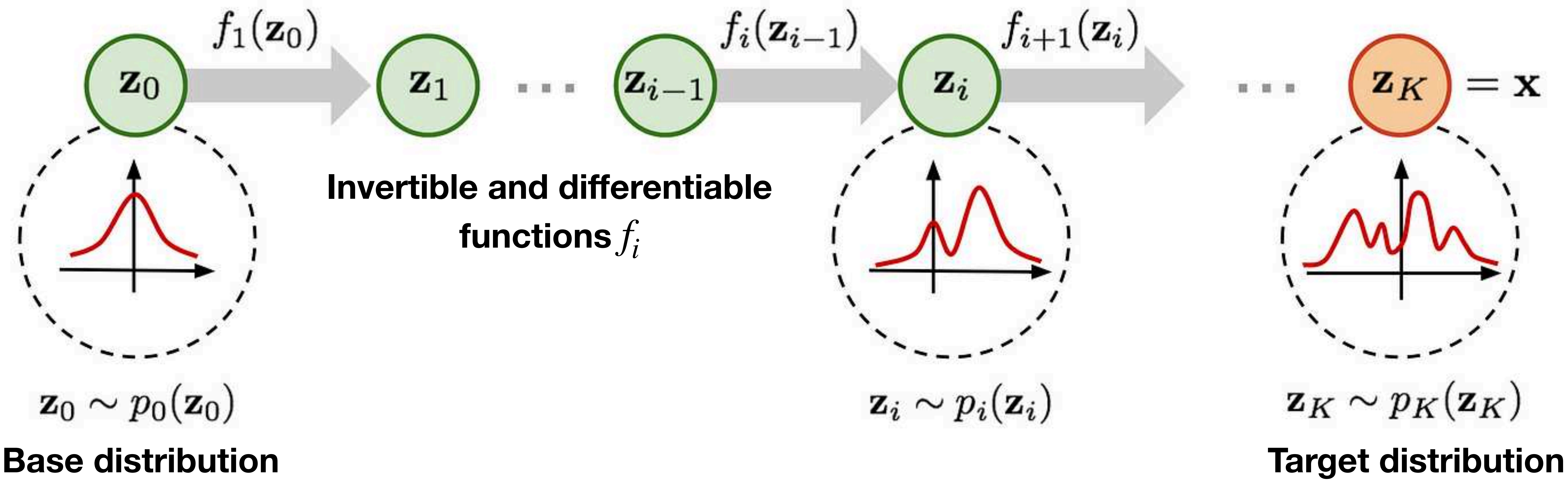
$$X_j = \int L_\lambda R_j(\lambda) \mathrm{d}\lambda + \mathrm{Noise}$$





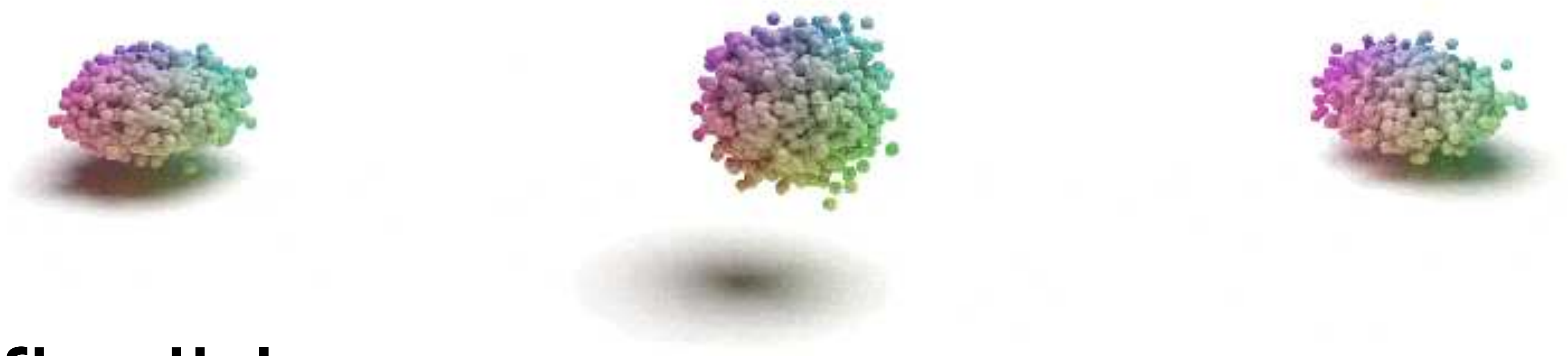
Population distribution
$p(\boldsymbol{\theta}|\{\mathbf{X}_i\})$
Sample from normalizing flow
$\boldsymbol{\theta}_j^{\phi} \sim q_{\phi}(\boldsymbol{\theta})$
Stellar population synthesis
+ selection
Predicted photometry
$\{\hat{\mathbf{X}}_j^{\phi}\}$
POPSED
Also pop-cosmos
(Alsing+24)
Observed Photometry
$\{\mathbf{X}_i\}$
Distance between
two distributions
$\mathcal{W}_2(\{\hat{\mathbf{X}}_j^{\phi}\}, \{\mathbf{X}_i\})$
Train
Probability Density
Estimator $q_{\phi}(\boldsymbol{\theta})$
$\log M_*$
z
$\log Z$
g
r

Normalizing Flow



- Transforms a base distribution (standard Gaussian) to any arbitrary distribution

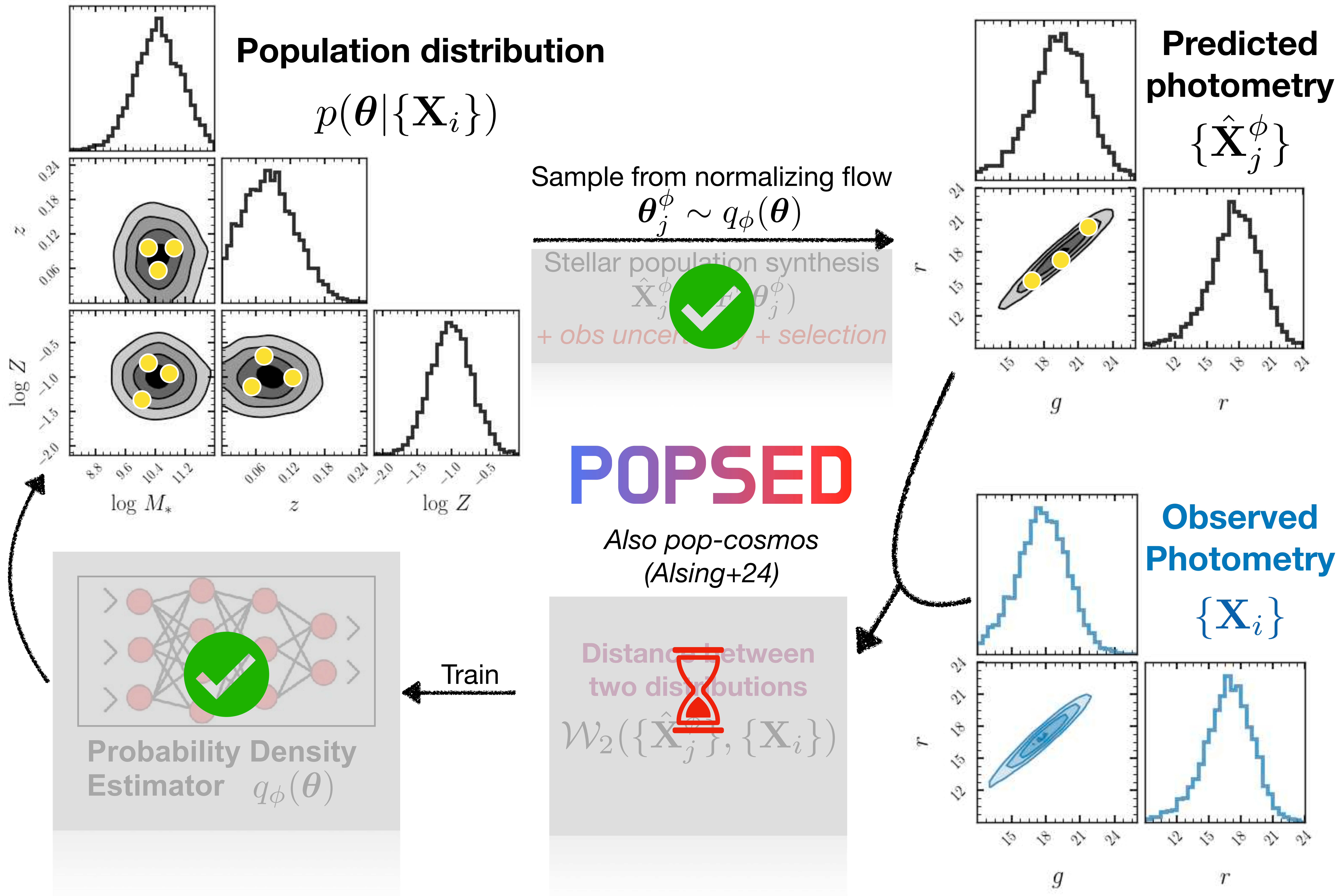
- Transformation functions are parameterized by neural nets



- Very flexible

- Easy to sample from

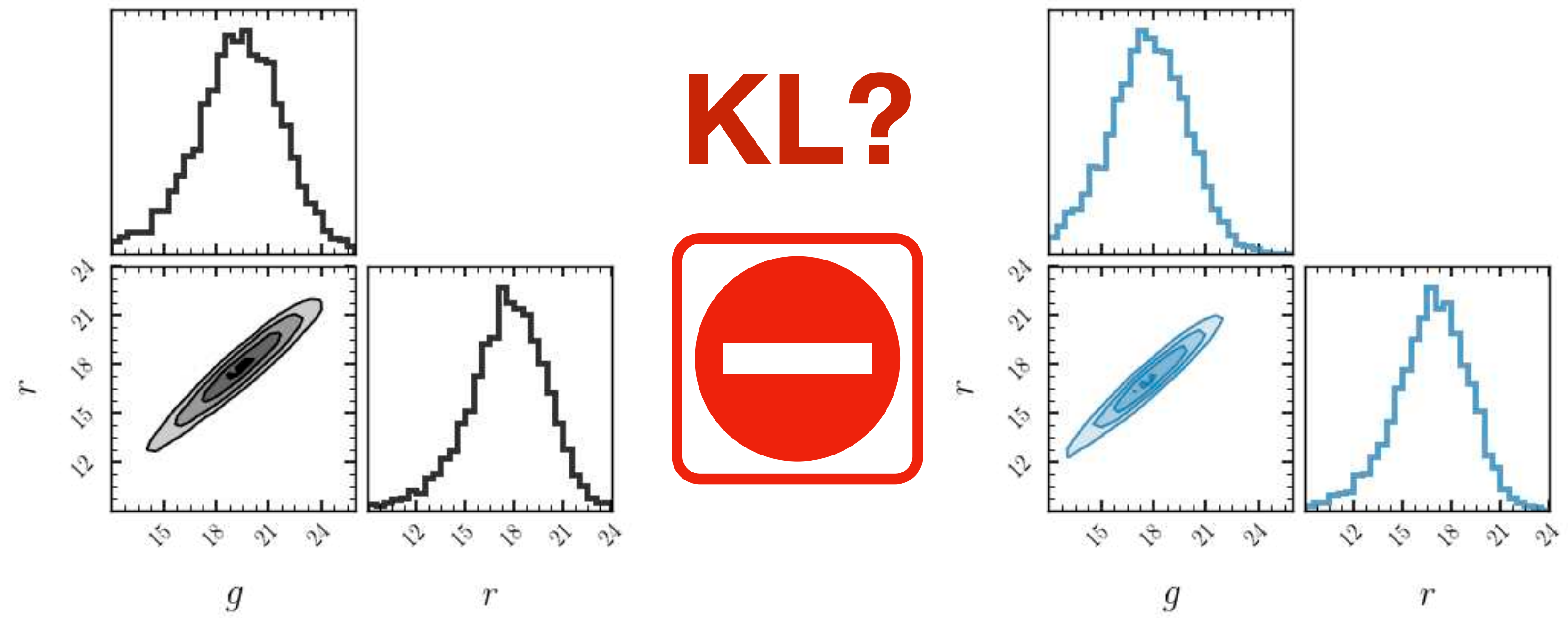
- We use “Neural Spline Flows”


Population distribution
$p(\boldsymbol{\theta}|\{\mathbf{X}_i\})$
Sample from normalizing flow
$\boldsymbol{\theta}_j^{\phi} \sim q_{\phi}(\boldsymbol{\theta})$
Stellar population synthesis
+ selection
Predicted photometry
$\{\hat{\mathbf{X}}_j^{\phi}\}$
POPSED
Also pop-cosmos (Alsing+24)
Observed Photometry
$\{\mathbf{X}_i\}$
Train
$\mathcal{W}_2(\{\hat{\mathbf{X}}_j^{\phi}\}, \{\mathbf{X}_i\})$
Probability Density Estimator $q_{\phi}(\theta)$
log M_*
z
log Z
g
r

Wasserstein Distance

Predicted photometry

Real observation

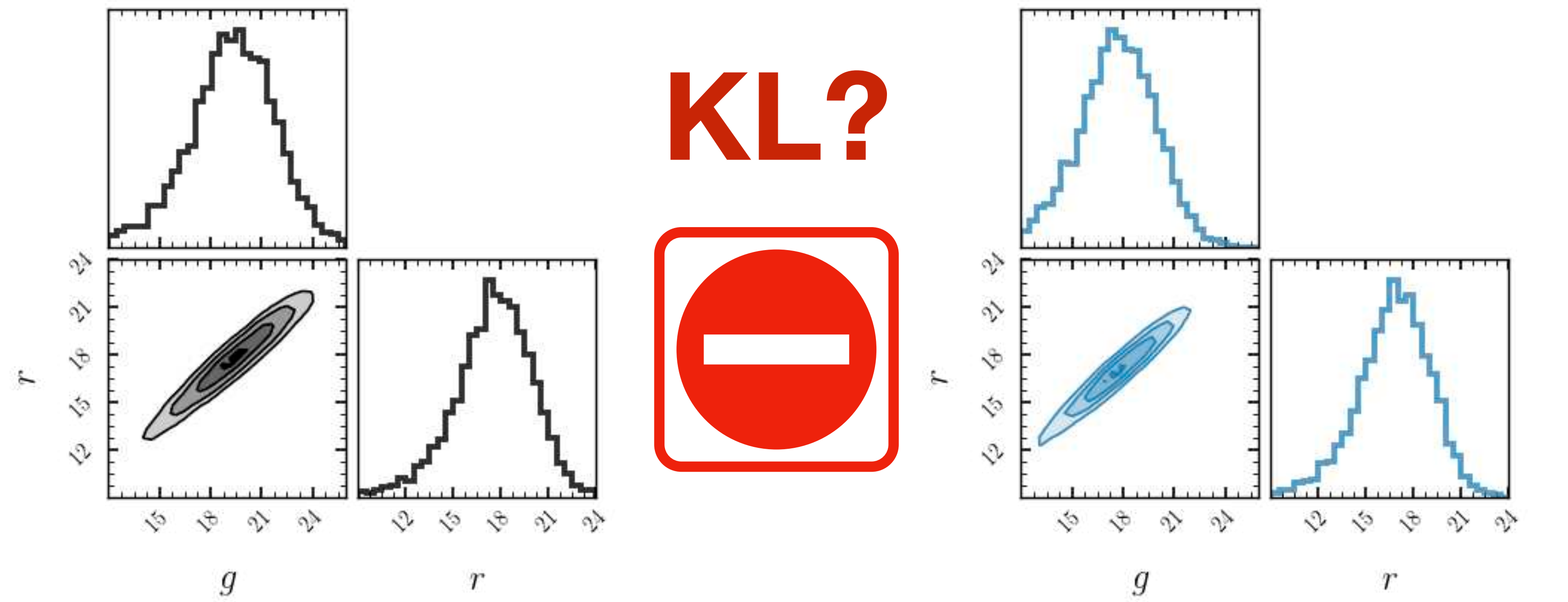


Earth mover's distance (optimal transport)

Wasserstein Distance

Predicted photometry

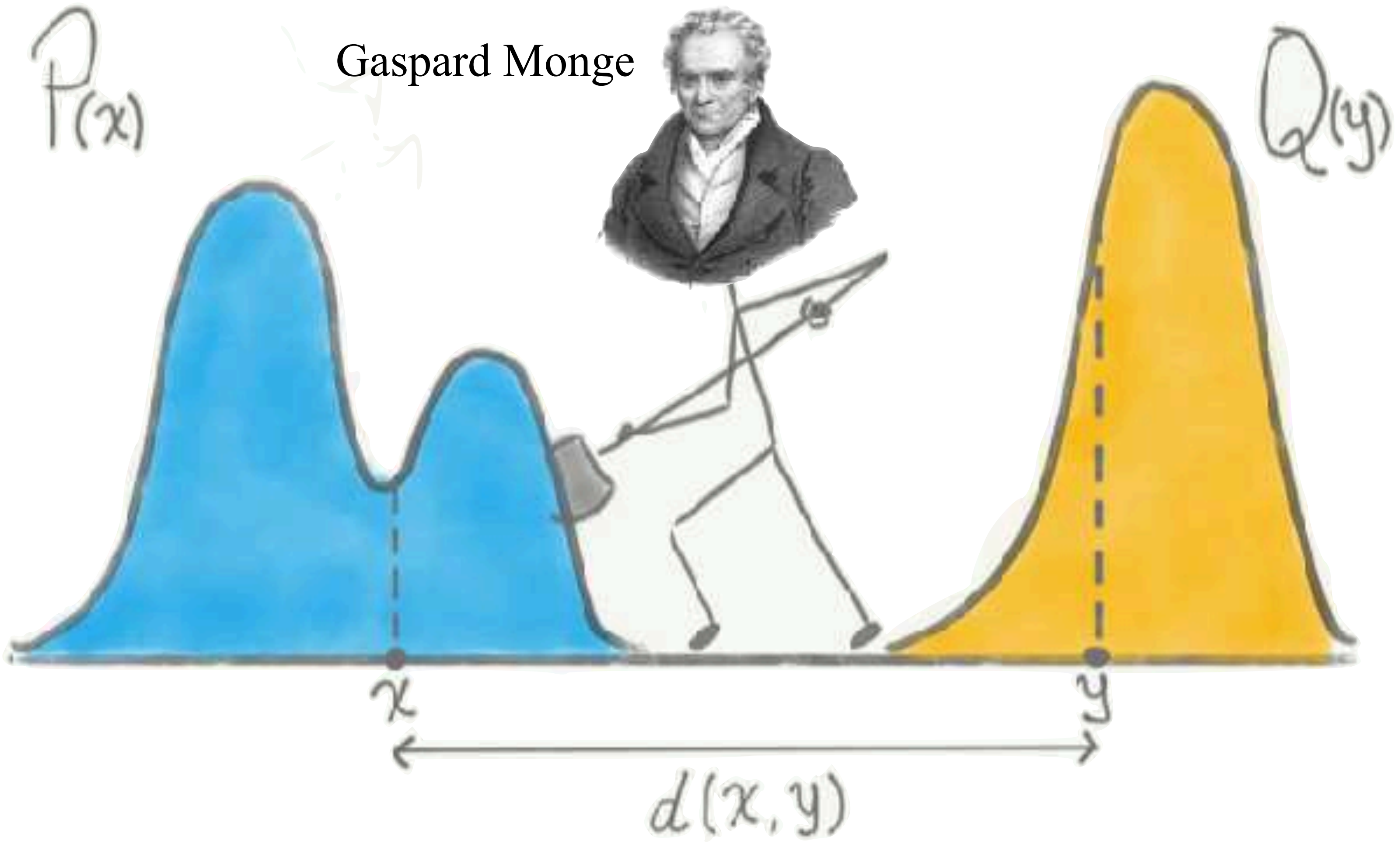
Real observation



Two point masses: $\mu_1 = \delta_{a_1}, \mu_2 = \delta_{a_2}$

Wasserstein distance: $W_p(a_1, a_2) \propto \left(|a_1 - a_2|^p \right)^{1/p}$

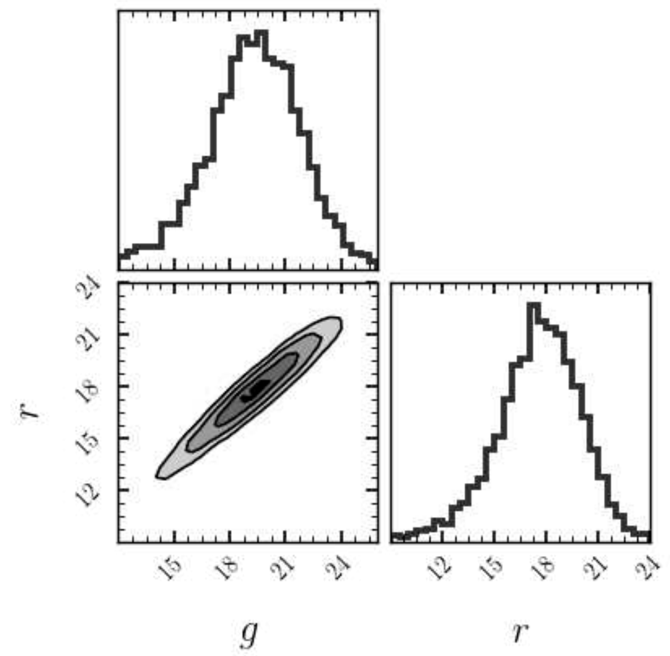
1-D case: $W_p = \left(\frac{1}{n} \sum_{i=1}^{n} \|X_{(i)} - Y_{(i)}\|^p \right)^{1/p}$



Earth mover's distance (optimal transport)

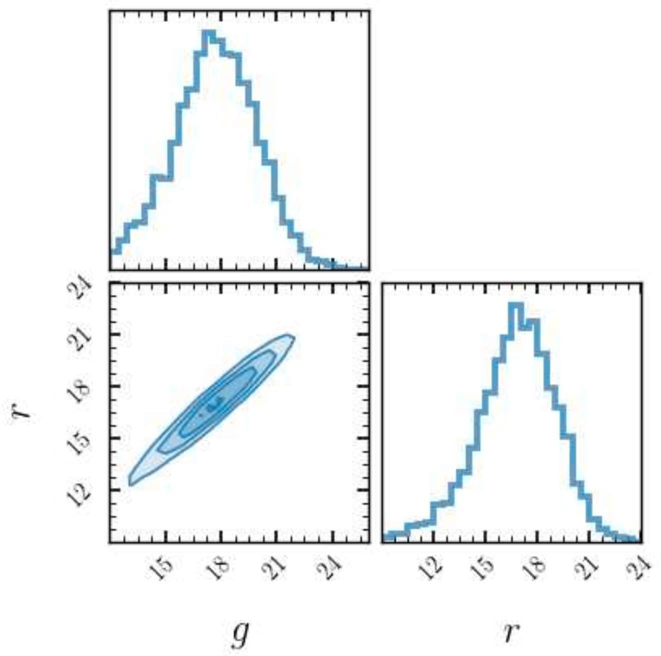
Wasserstein Distance

Predicted photometry



KL?

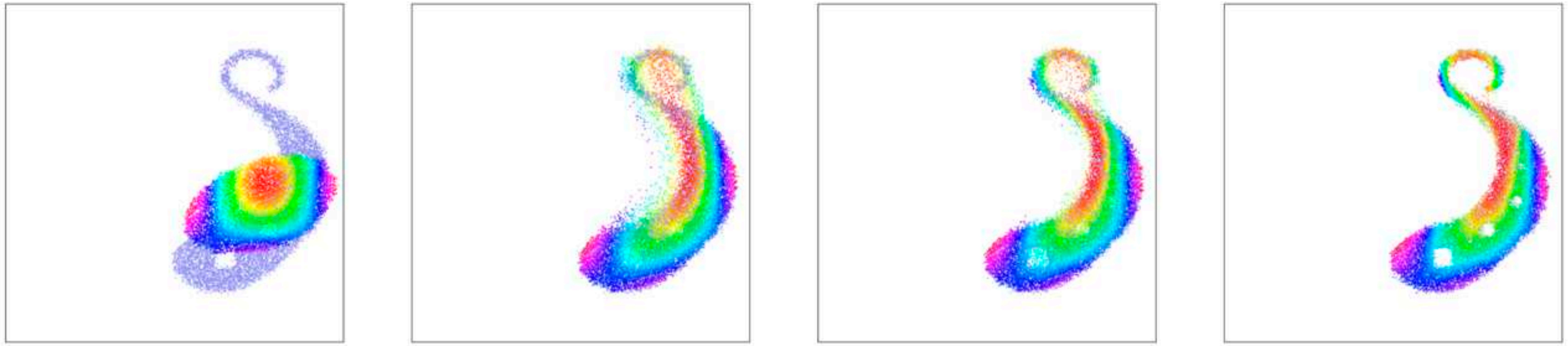
Real observation



High dimension: Sinkhorn iteration

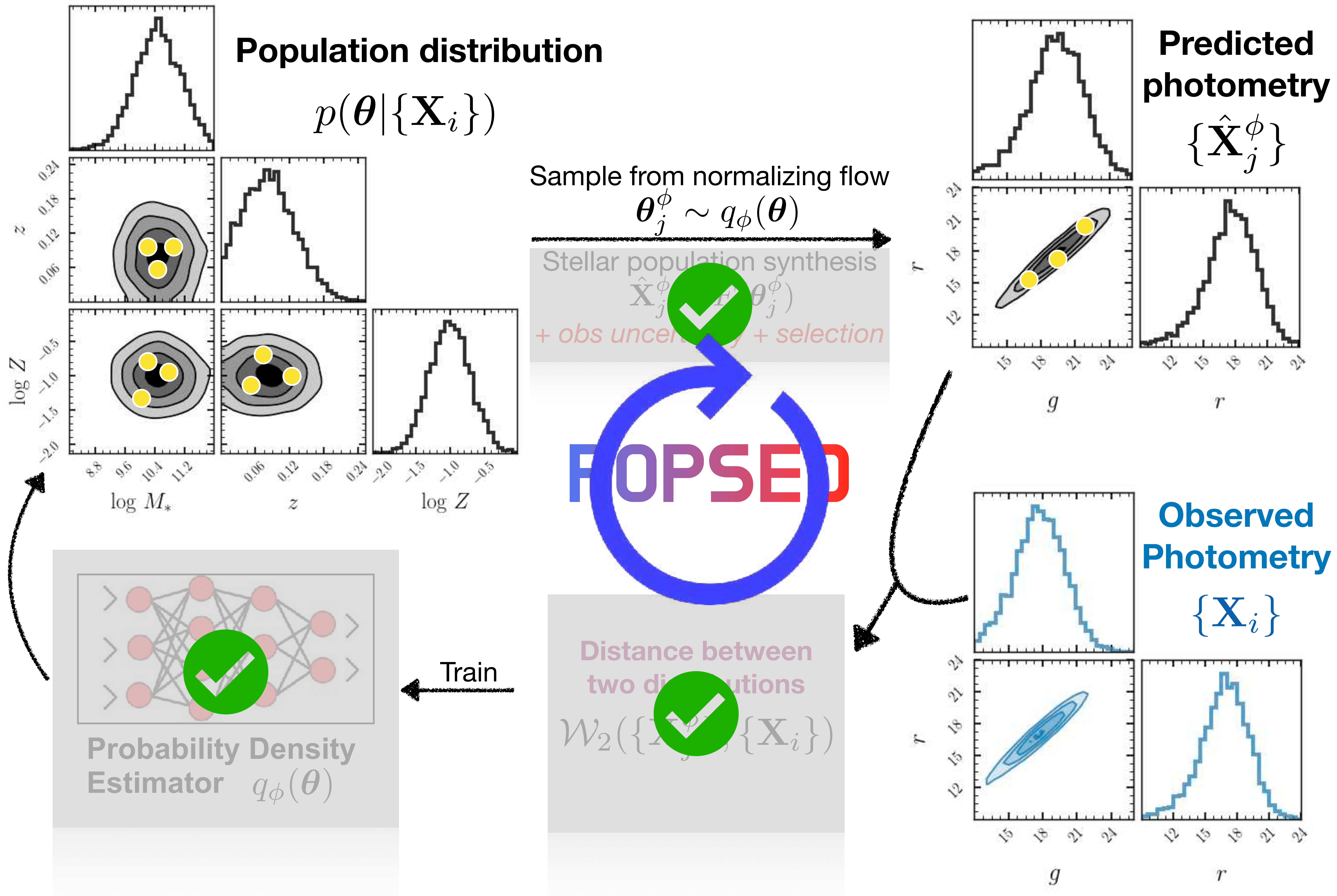
- Symmetric
- Differentiable
- Relatively fast O(N^2)

- Well-defined on high-dimensional discrete distributions
- Can tune the **blurring** scale



(a) $\sqrt{\varepsilon} = 1.00$ (b) $\sqrt{\varepsilon} = .10$ (c) $\sqrt{\varepsilon} = .05$ (d) $\sqrt{\varepsilon} = .01$

geomloss (Feydy et al., 2019)


Population distribution
$p(\boldsymbol{\theta}|\{\mathbf{X}_i\})$
Sample from normalizing flow
$\boldsymbol{\theta}_j^{\phi} \sim q_{\phi}(\boldsymbol{\theta})$
Predicted photometry
$\{\hat{\mathbf{X}}_j^{\phi}\}$
Stellar population synthesis
+ selection
POPSED
Observed Photometry
$\{\mathbf{X}_i\}$
Train
Distance between
Probability Density Estimator $q_{\phi}(\theta)$
log M_*
z
log Z
g
r

🏋️ Training

Noise Anti-Annealing

- Add realistic noise in the beginning confuses the flow
- Gradually add noise
- Stabilize training

Blurring Annealing

- Start with large blurring, to capture global landscape of the target distribution
- Reduce blurring later on to match subtle details between distributions

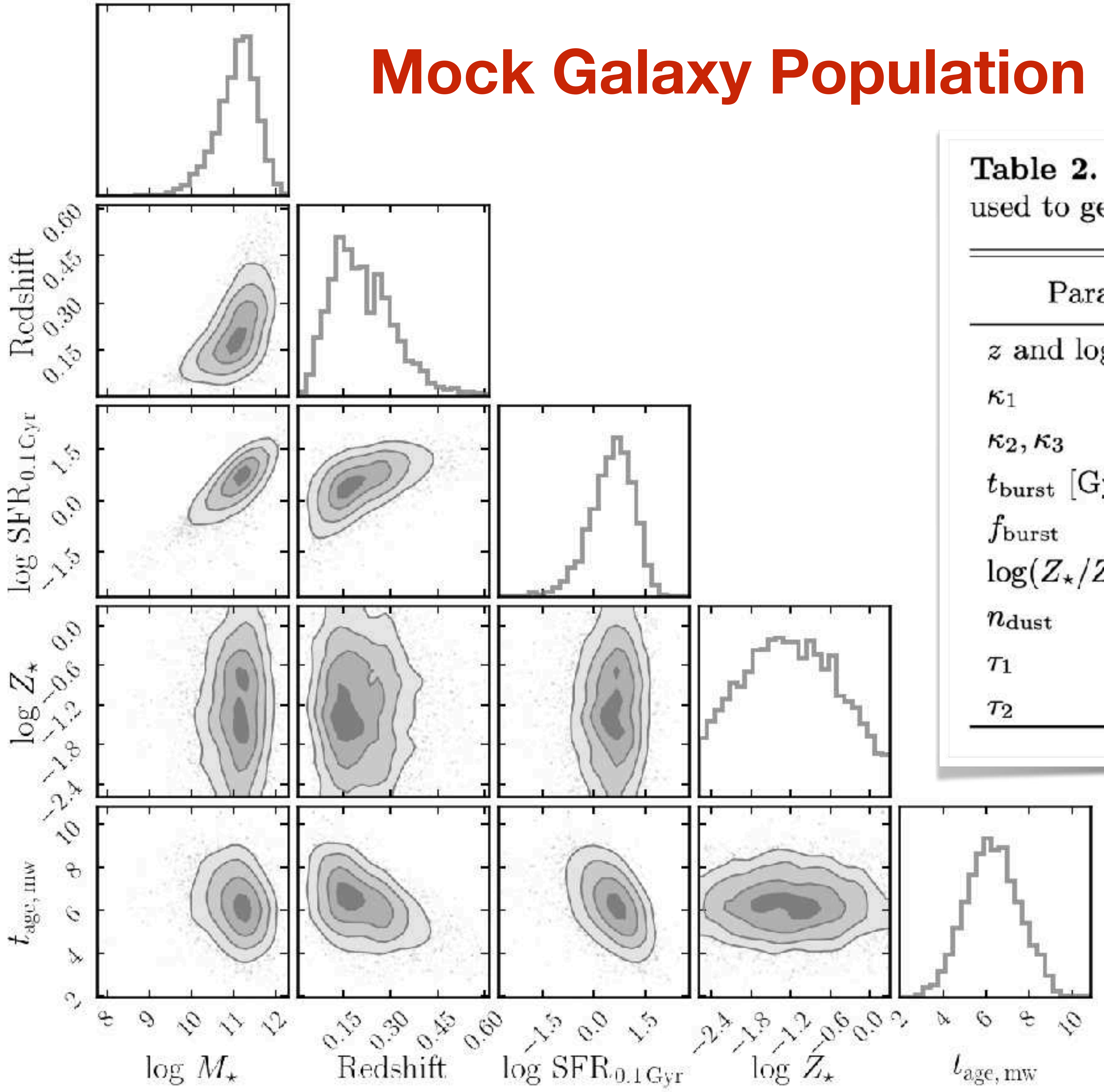
Ensemble Learning

- One flow is one sample of the population posterior.
- A number of flows will approximate the population posterior.

https://cims.nyu.edu/~andrewgw/deepensembles/

Mock Test

Mock Galaxy Population

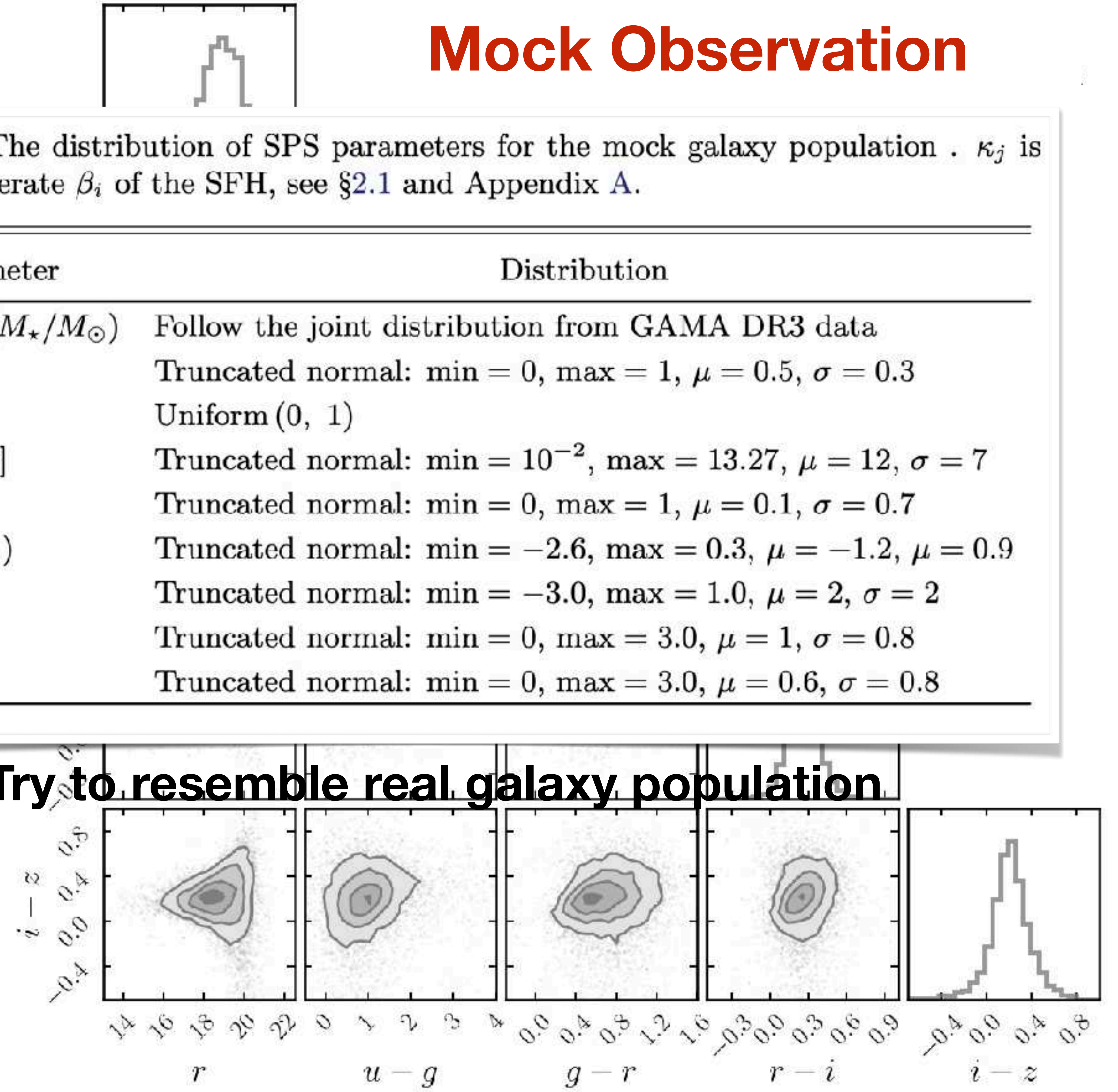


Mock Observation

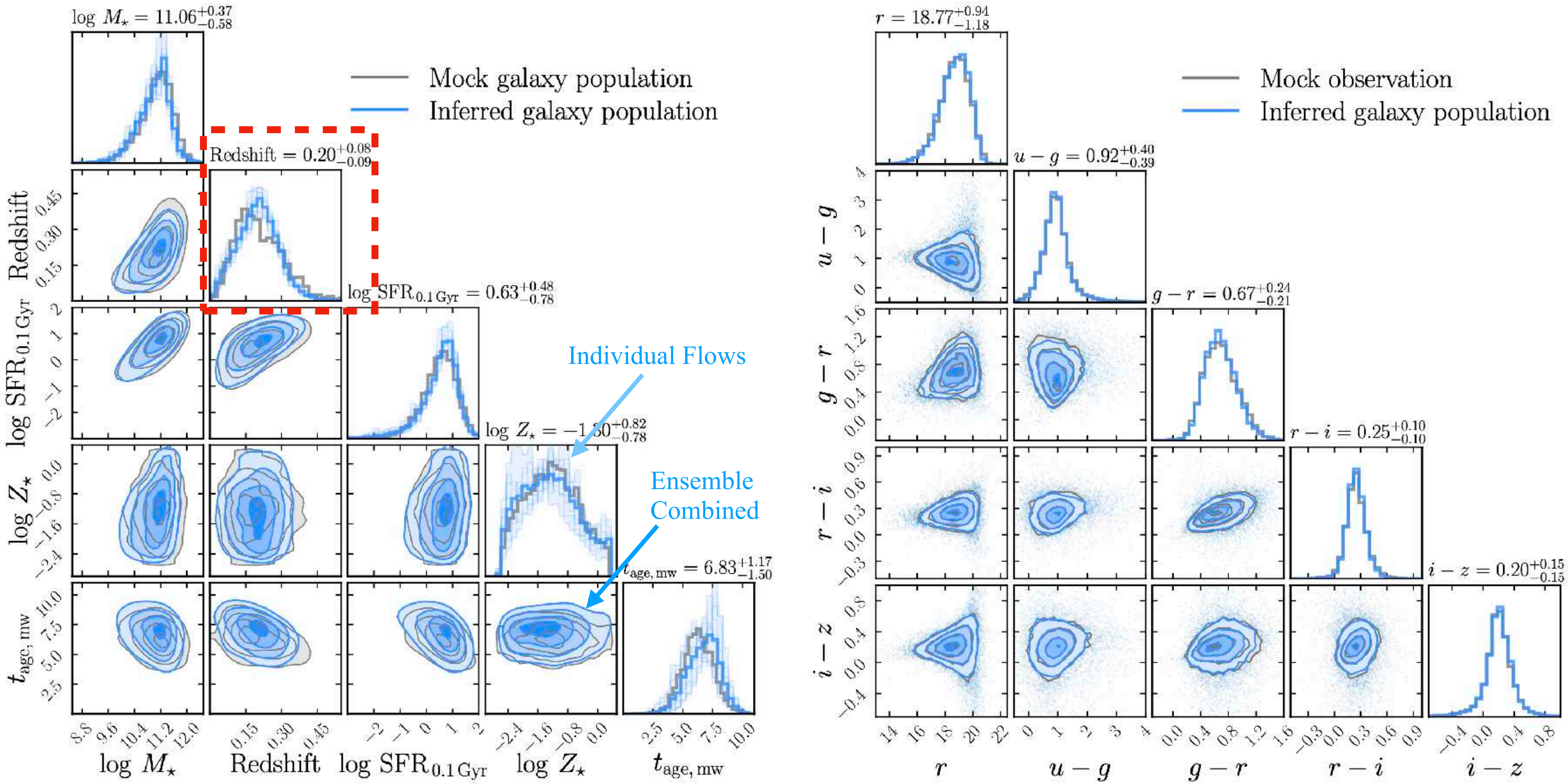
Table 2. The distribution of SPS parameters for the mock galaxy population . κ_j is used to generate β_i of the SFH, see §2.1 and Appendix A.

Parameter	Distribution
z and $\log(M_\star/M_\odot)$	Follow the joint distribution from GAMA DR3 data
κ_1	Truncated normal: min $=0$, max $=1$, $\mu=0.5$, $\sigma=0.3$
κ_2, κ_3	Uniform $(0,\ 1)$
t_{burst} [Gyr]	Truncated normal: min $=10^{-2}$, max $=13.27$, $\mu=12$, $\sigma=7$
f_{burst}	Truncated normal: min $=0$, max $=1$, $\mu=0.1$, $\sigma=0.7$
$\log(Z_\star/Z_\odot)$	Truncated normal: min $=-2.6$, max $=0.3$, $\mu=-1.2$, $\mu=0.9$
n_{dust}	Truncated normal: min $=-3.0$, max $=1.0$, $\mu=2$, $\sigma=2$
τ_1	Truncated normal: min $=0$, max $=3.0$, $\mu=1$, $\sigma=0.8$
τ_2	Truncated normal: min $=0$, max $=3.0$, $\mu=0.6$, $\sigma=0.8$

Try to resemble real galaxy population



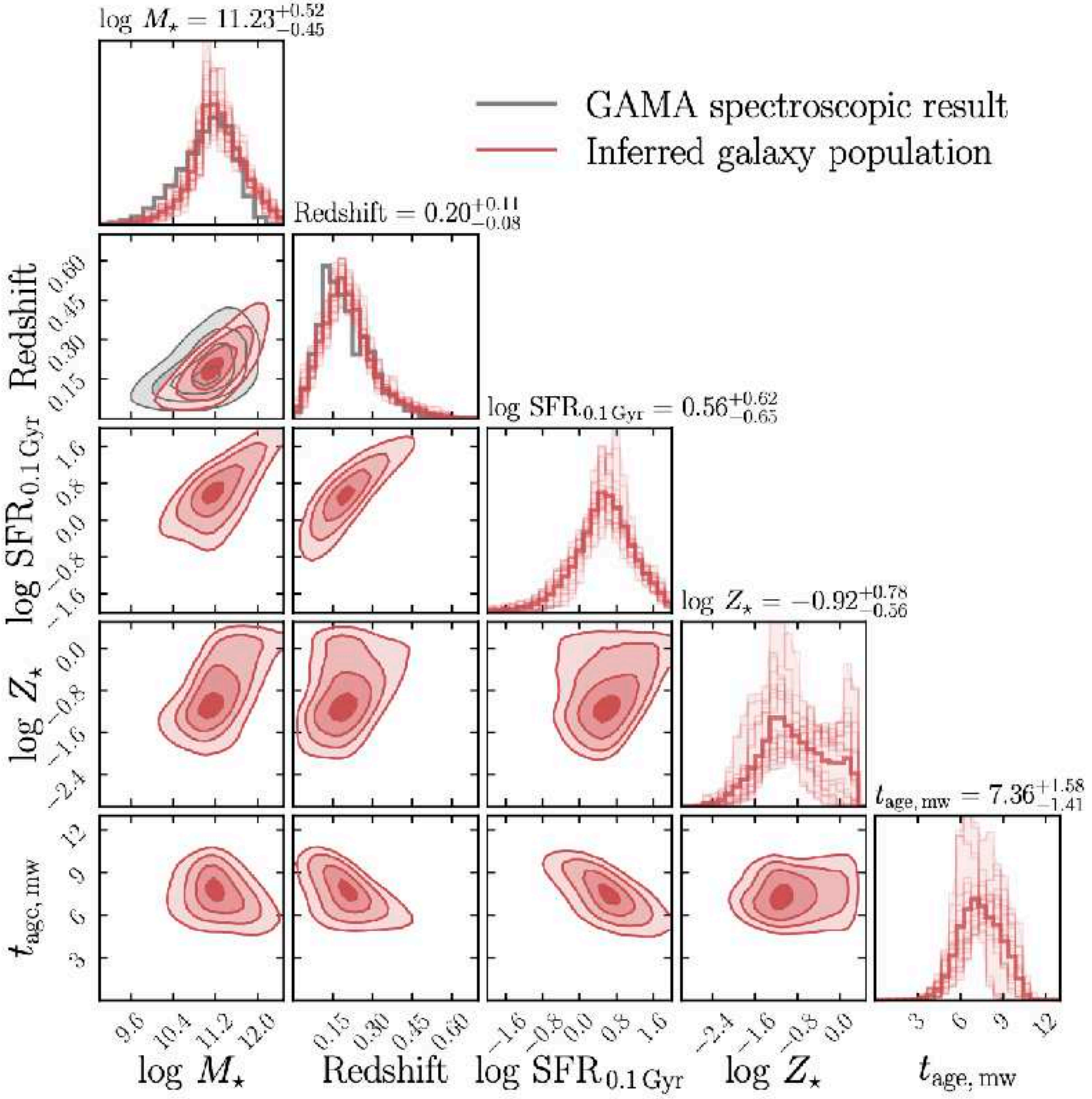
10^5 galaxies, 1 GPU hour for one flow

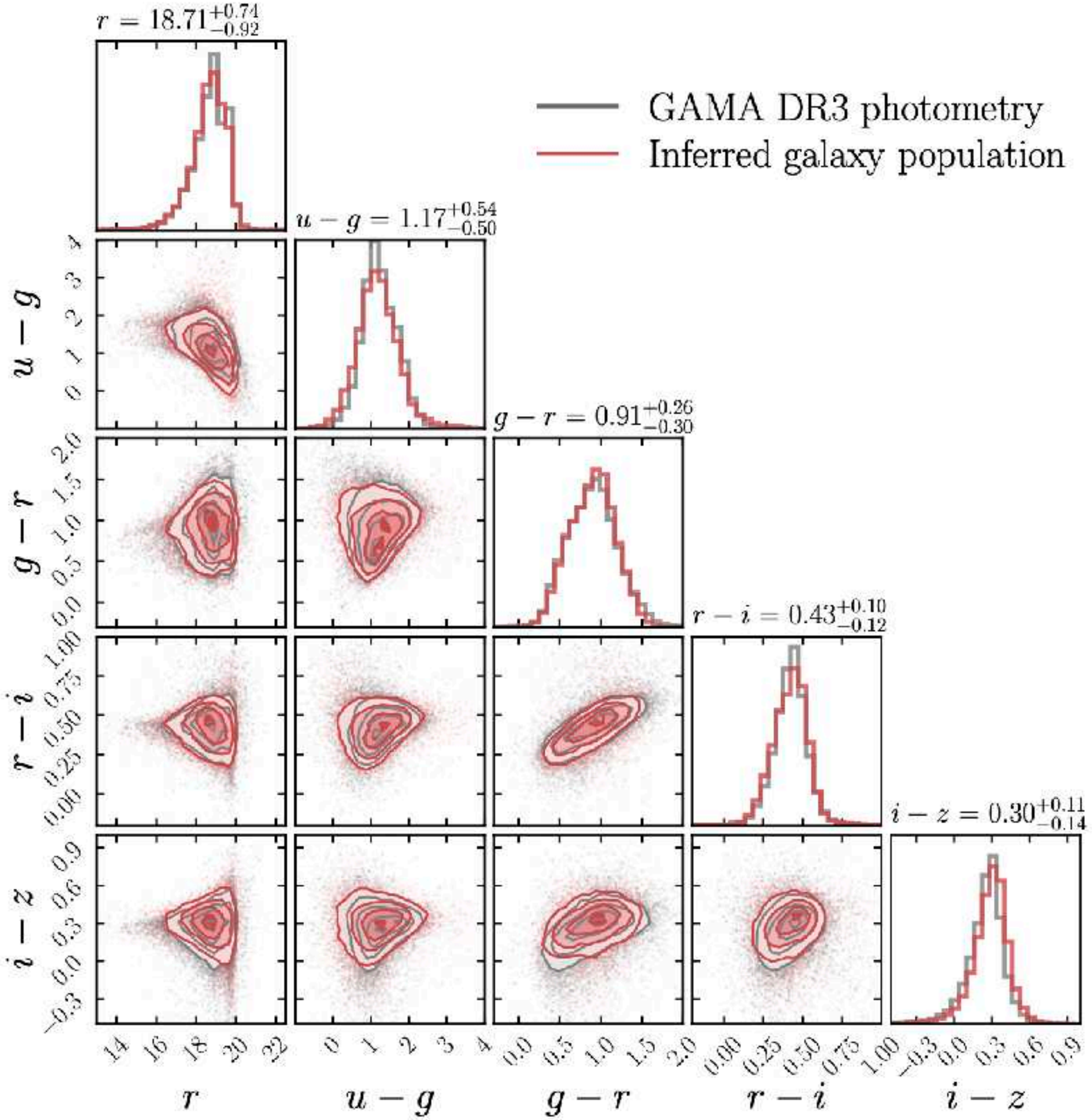


GAMA DR3 data

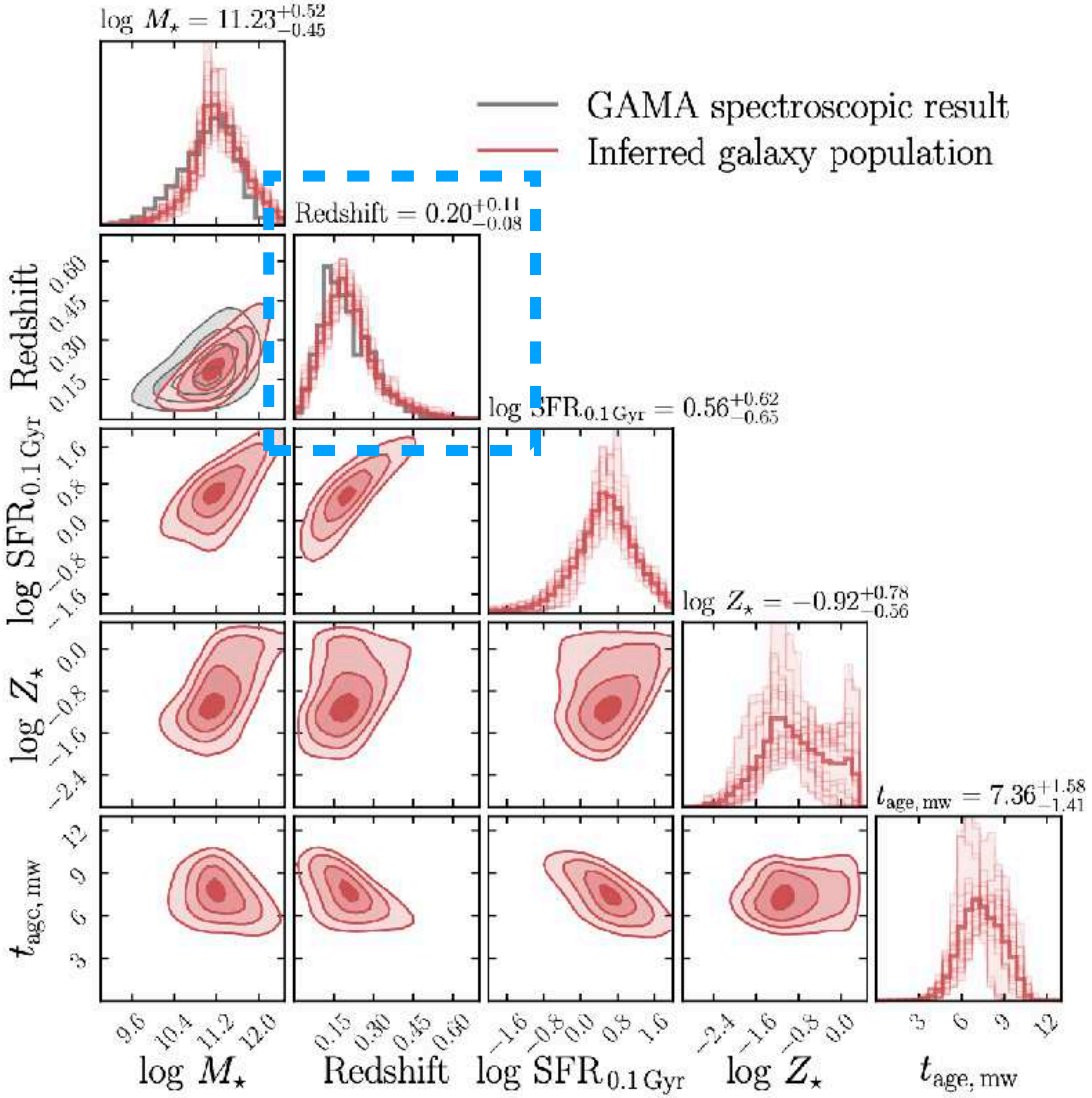
83692 galaxies
~20 GPU hours

SDSS photometry + spectroscopic follow-ups for r < 19.0 mag





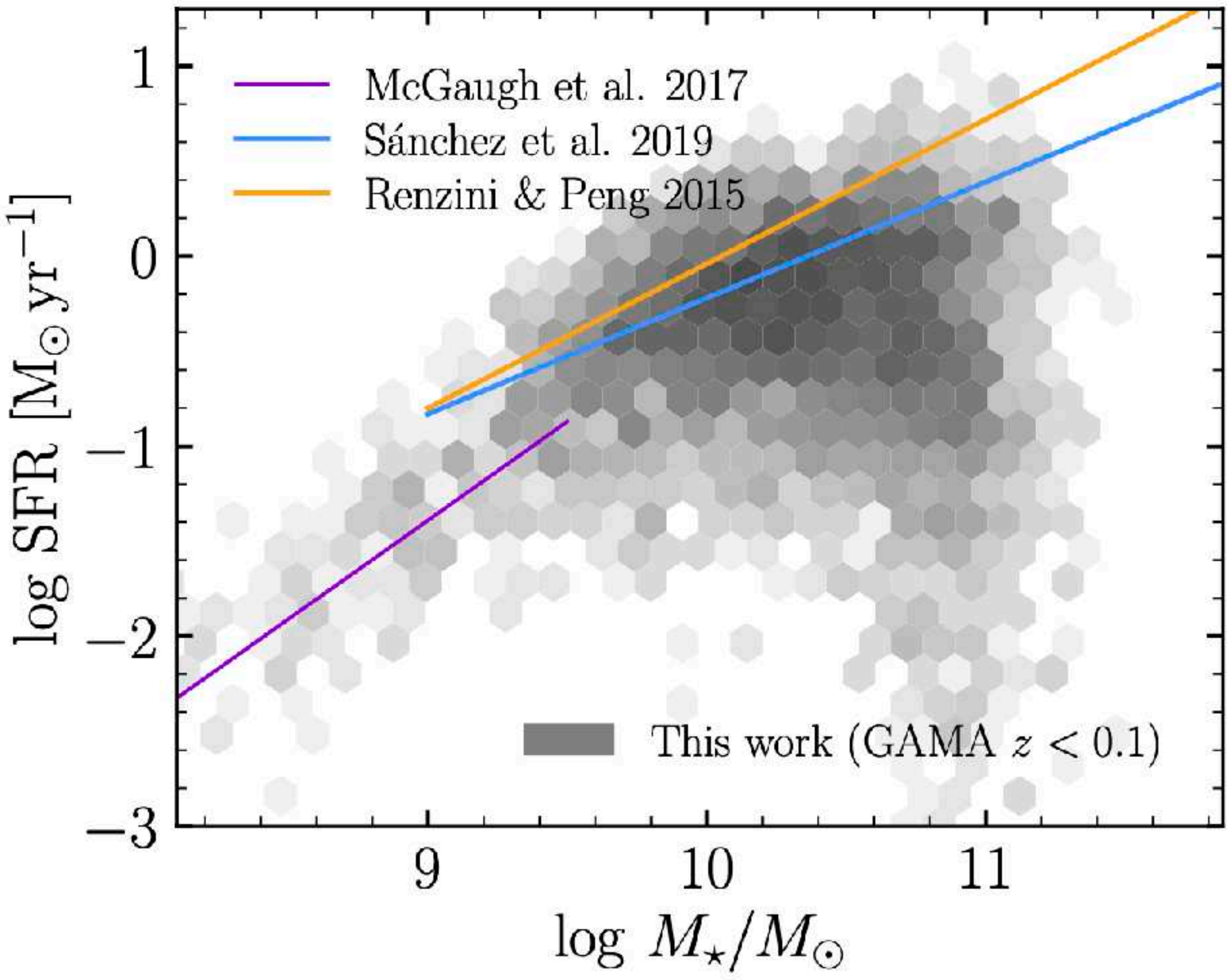
GAMA DR3 data



- Photo-z distribution agrees with spec-z distribution quite well!
- $M_\star$ distribution has a longer tail.
- Poor constraints on metallicity!

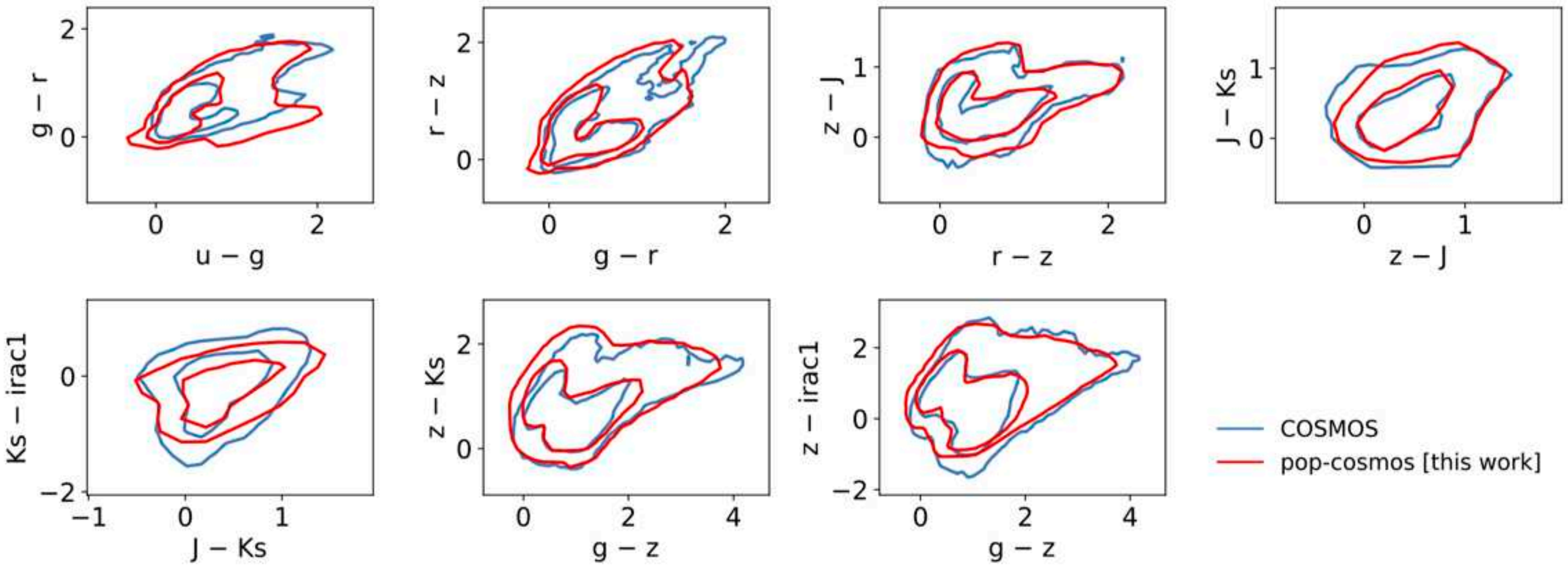
GAMA DR3 data

Star-Forming Main Sequence

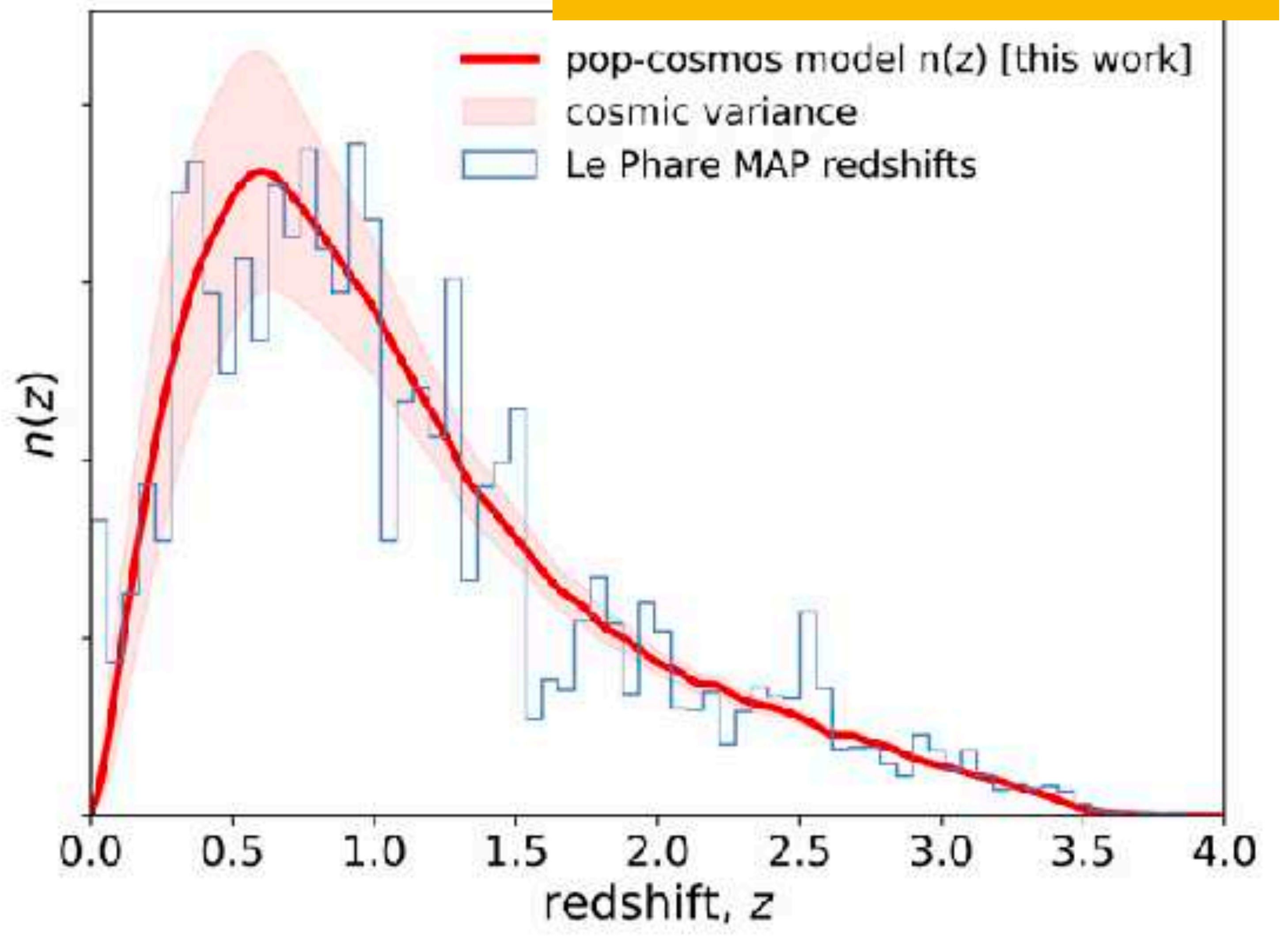


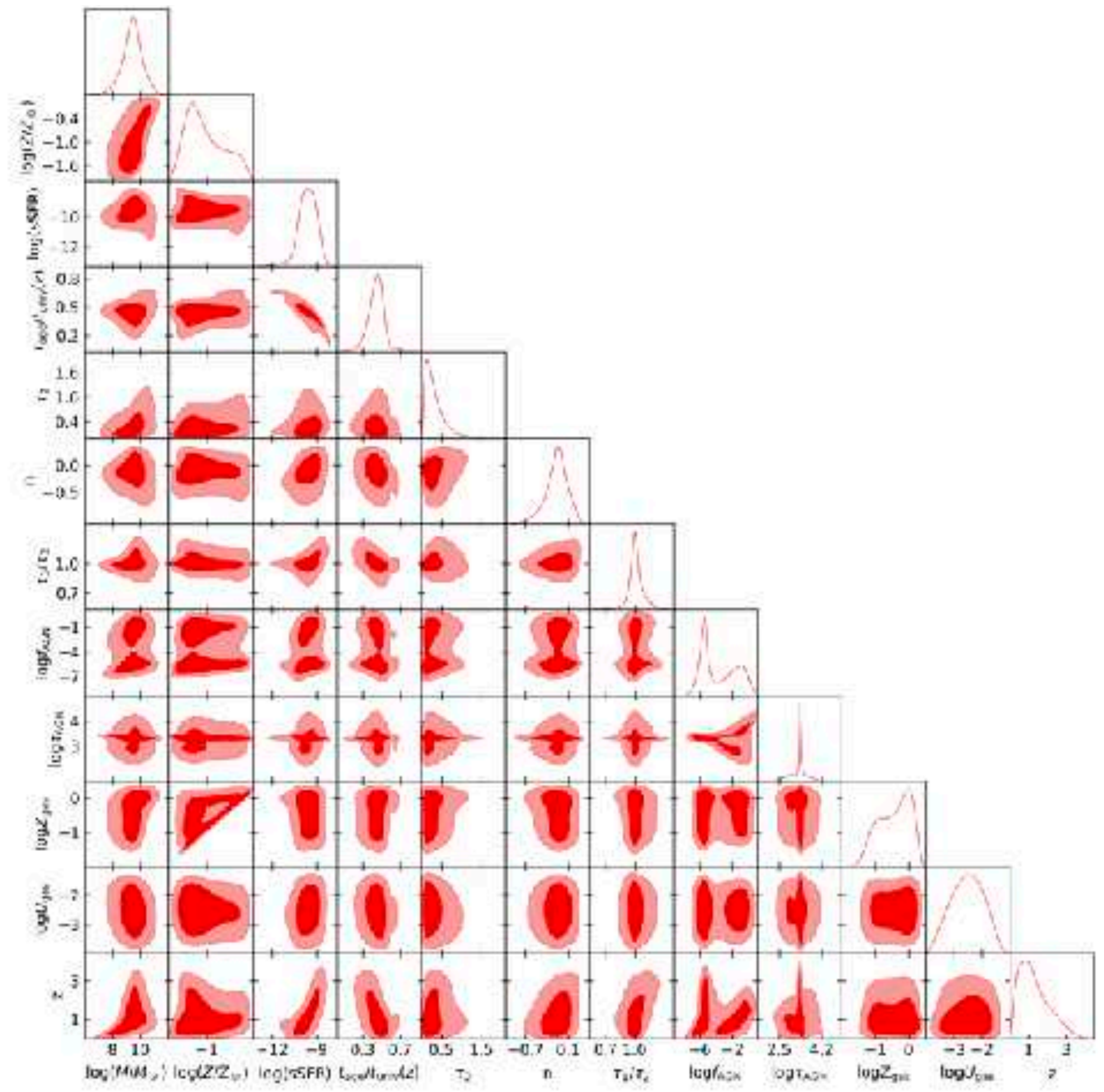
- We already have population distribution $p(\boldsymbol{\theta}|\{\mathbf{X}_i\})$, we can slice it and study scaling relations.
- We select $z < 0.1$ and plot the star-forming main sequence.
- PopSED captures the SFMS, quiescent galaxies, and their transition.

pop-cosmos (Alsing+24)



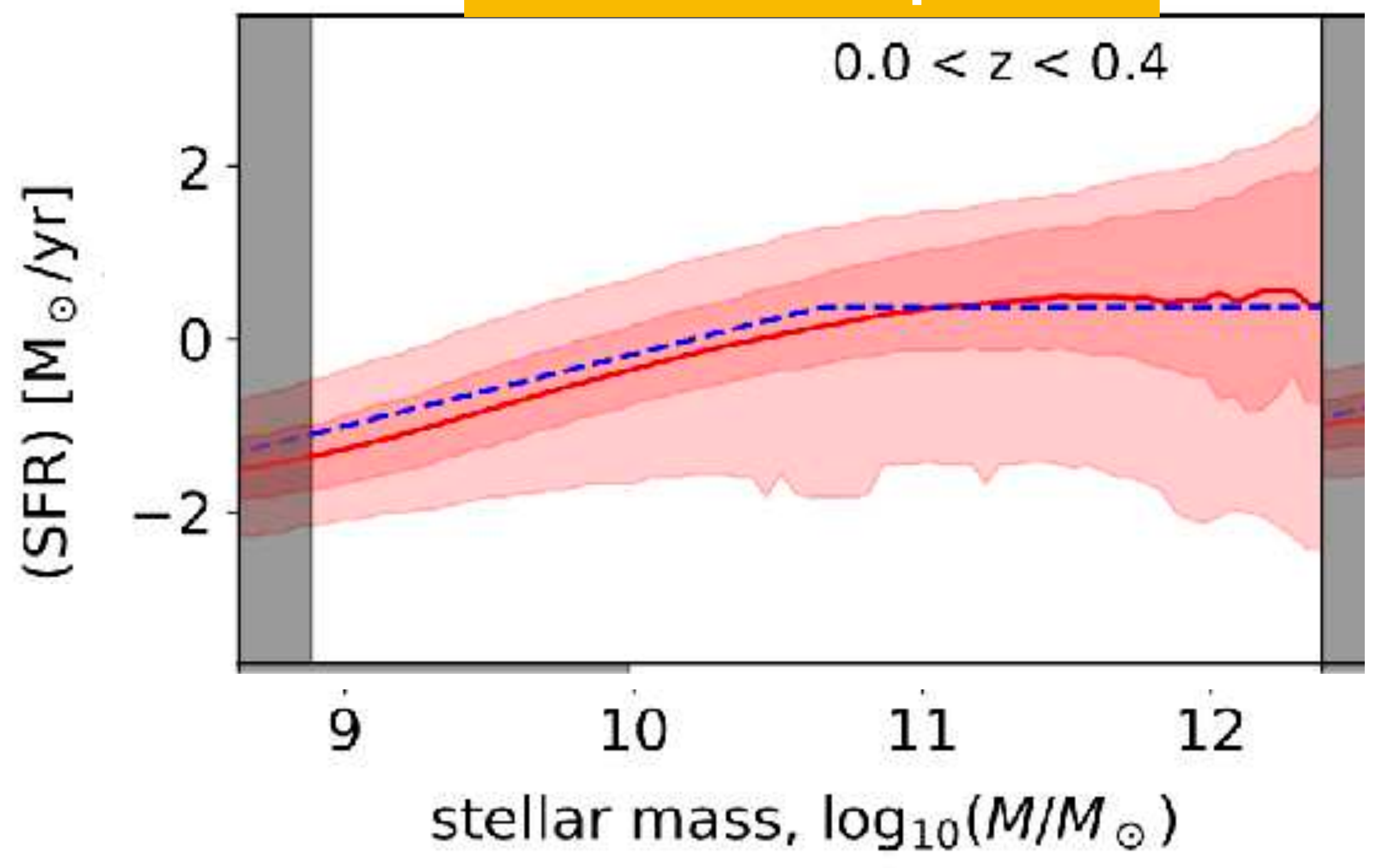
Redshift distribution





- COSMOS 2020 data (Weaver+22)
- Added emission lines in SPS
- Added selection function in forward model
- Used diffusion model

SF main sequence



Advantages of PopSED

PopSED is faster

Simulation-based inference

~100x slower

JAX-SPS + HMC

~100x slower

Hearin et al. (2021)
Hahn et al. (2022)

Don't need to combine individual posteriors

- Can't just multiply posteriors of each galaxy!
- Need to save samples of individual posteriors.
- Additional modeling to derive population-level distribution.

Can be applied to

- Pop-level distribution is a good prior for individual object (e.g., Thorp+24)
- Weak lensing photo-z
- Photometric survey design/ target selection
- Outlier detection?
- PopSpec? PopSTAR?

Possible improvements

Even faster!

Current Wasserstein distance is ~$O(N^2)$

“Sliced” Wasserstein distance is $O(N \log N)$

Better SPS models

Differential SPS (DSPS) is both fast and not an emulator

Add emission lines, AGN contributions, etc.

Selection effects

Careful noise model

Survey selection and completeness

Better deal with non-detections