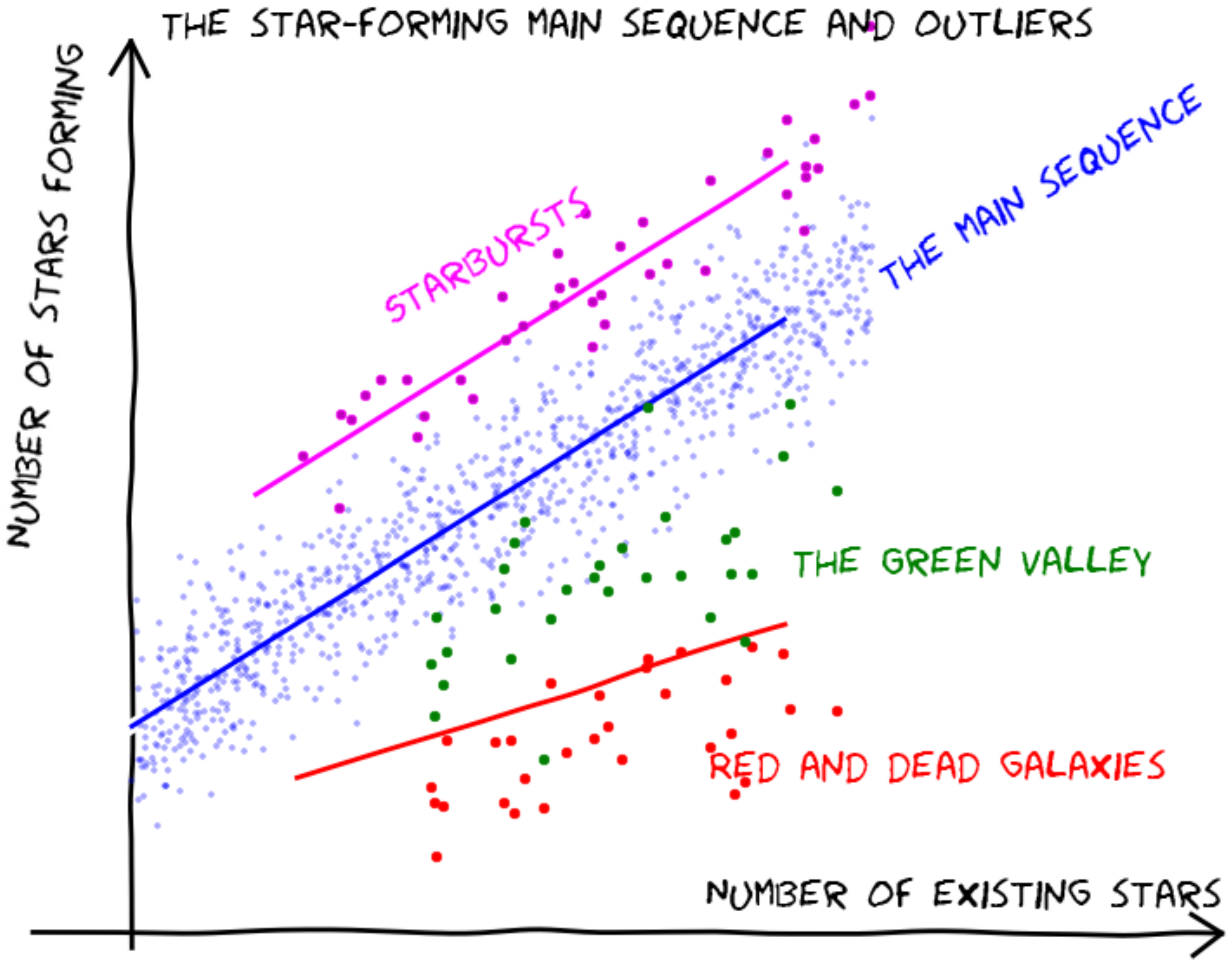
Constraining the stochasticity of star formation

Harry Bevins, Sandro Tacchella, Charlotte Simmonds

Main Sequence Galaxies

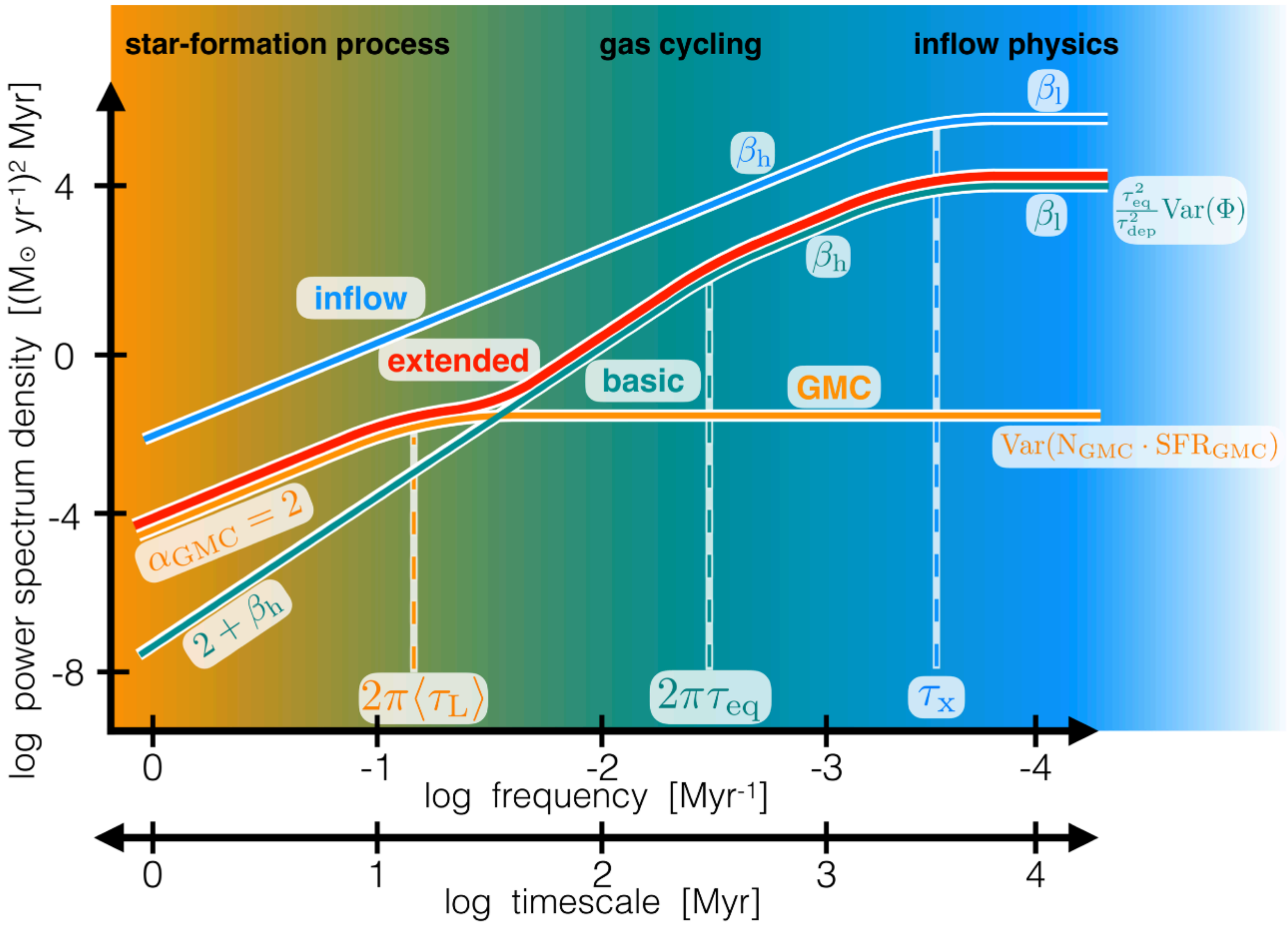
- MS galaxies lie on a relatively tight correlation between star formation rate and stellar mass $\mathrm{SFR} \propto M_*^\alpha$
- Normalisation of relationship and α vary with redshift
- Star formation driven by the same physical processes in these galaxies



CANDLES - Star formation in the Mountains

MS Power Spectral Density

- Expect formation to happen on common time scales in these galaxies
- Can define a Power Spectral Density function
- Fourier transform to get an Auto Correlation Function and define a gaussian process from which to draw SFHs
- Using a modified version of the implementation in Iyer et al. 2022 [2208.05938] (https://github.com/kartheikiyer/GP-SFH)



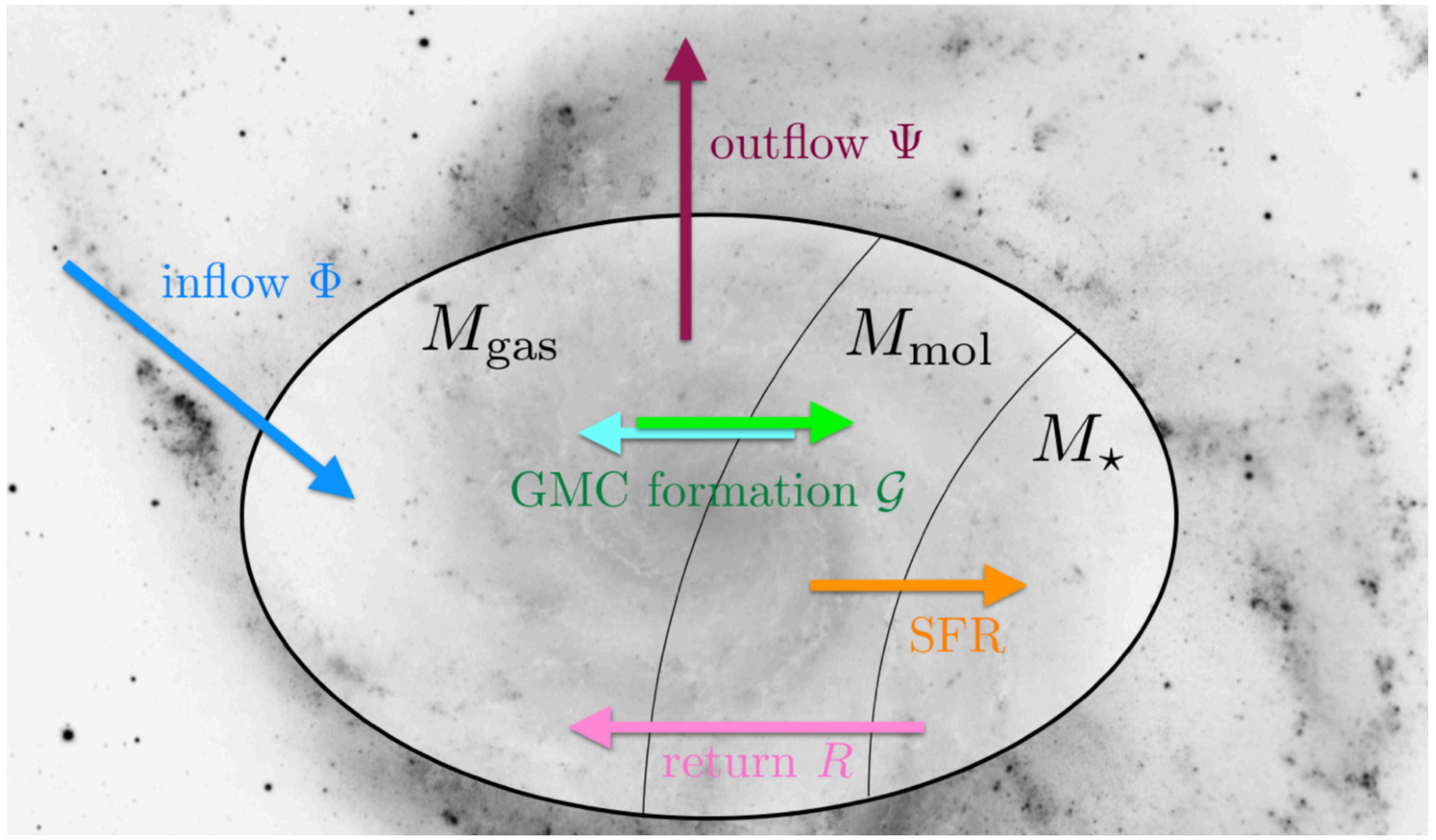
Tacchella et al 2020

Choice of PSD?

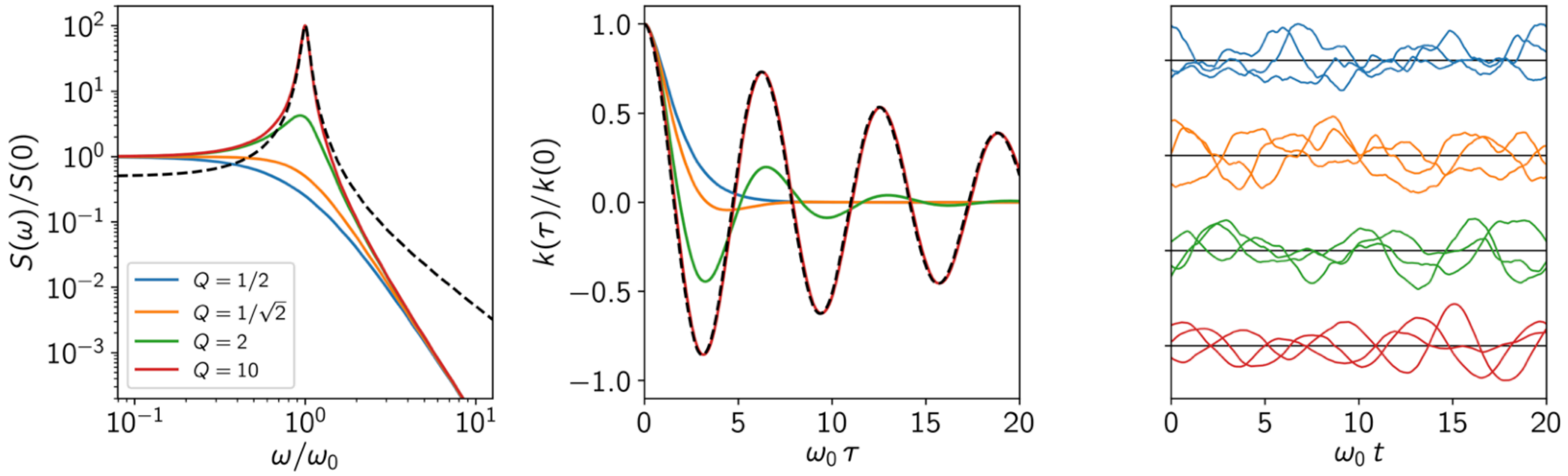
- Looked at a few different PSDs

  - Extended Regulator Model
    - Captures quasi-equilibrium between gas inflow and cycling and includes formation of Giant Molecular Clouds [Tacchella et al 2020]
    - $\sigma_{\text{reg}}, \tau_{\text{in}}, \tau_{\text{eq}}, \sigma_{\text{dyn}}, \tau_{\text{dyn}}$

  - Simple Harmonic Oscillator
    - Oscillatory SFH with preferential scale [Foreman-Mackey et al. 2017]
    - S_0, τ_0, Q

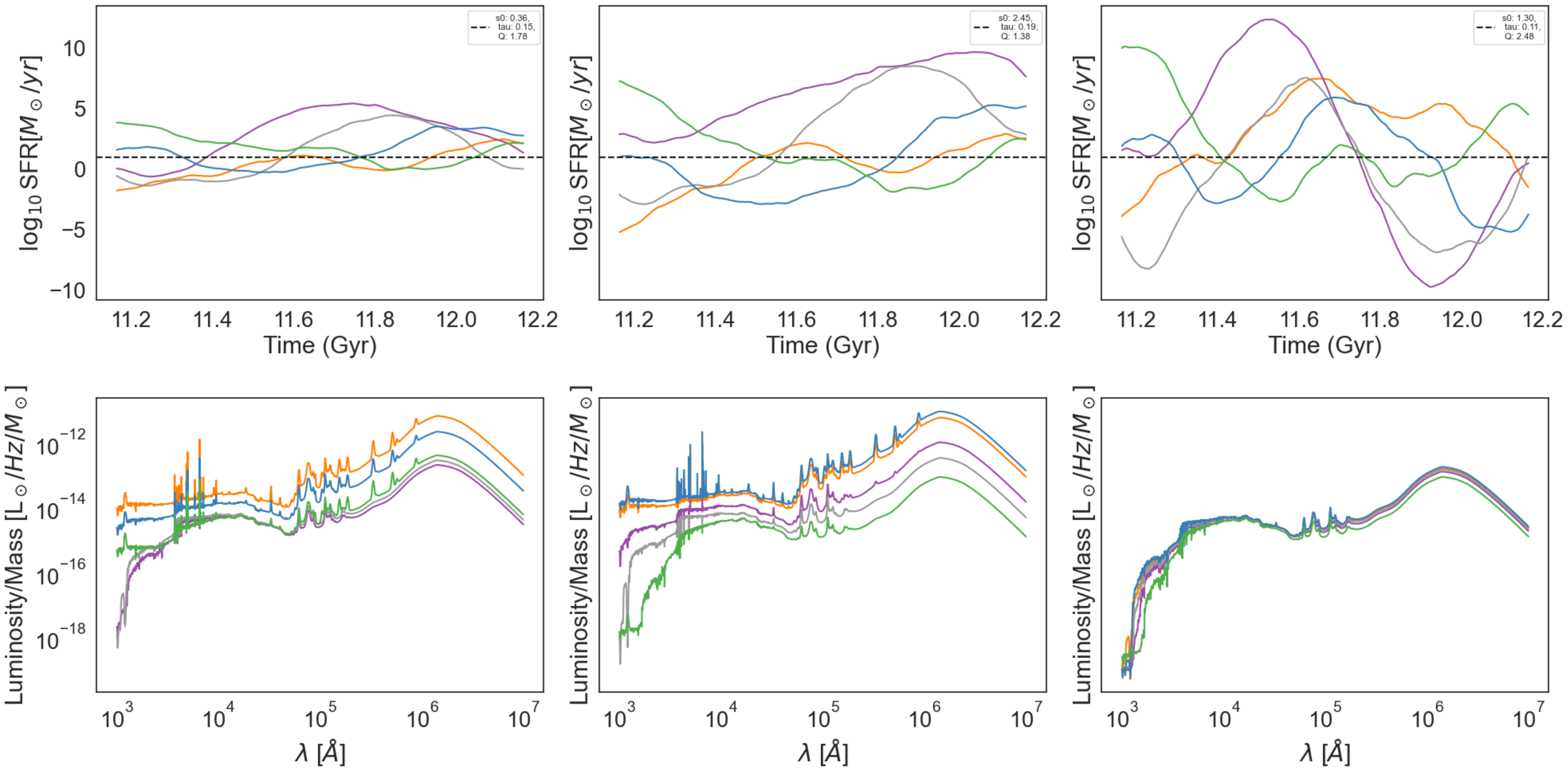


Tacchella et al. 2020



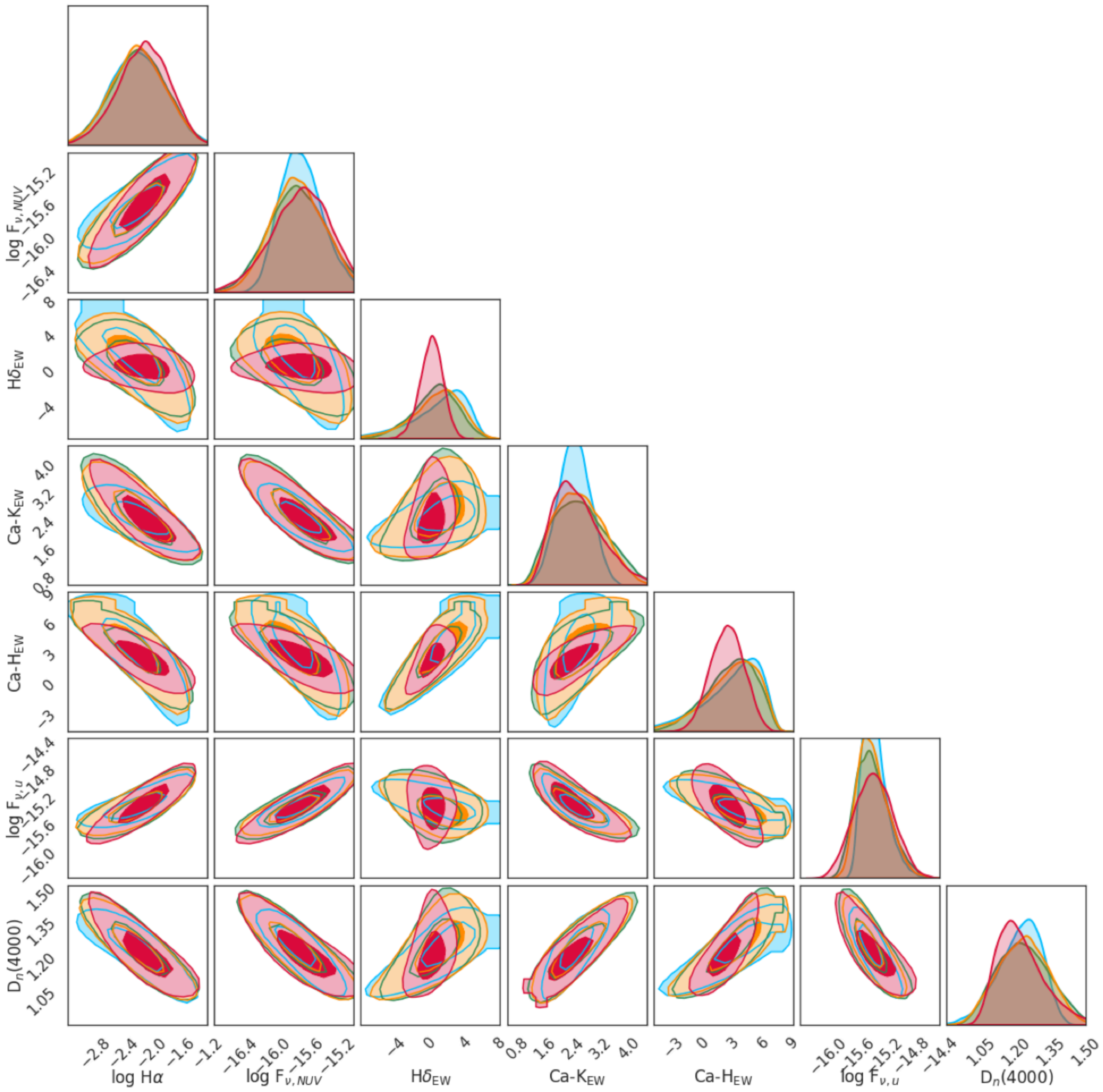
Foreman-Mackey et al. 2017

Simulating Main Sequence Galaxies



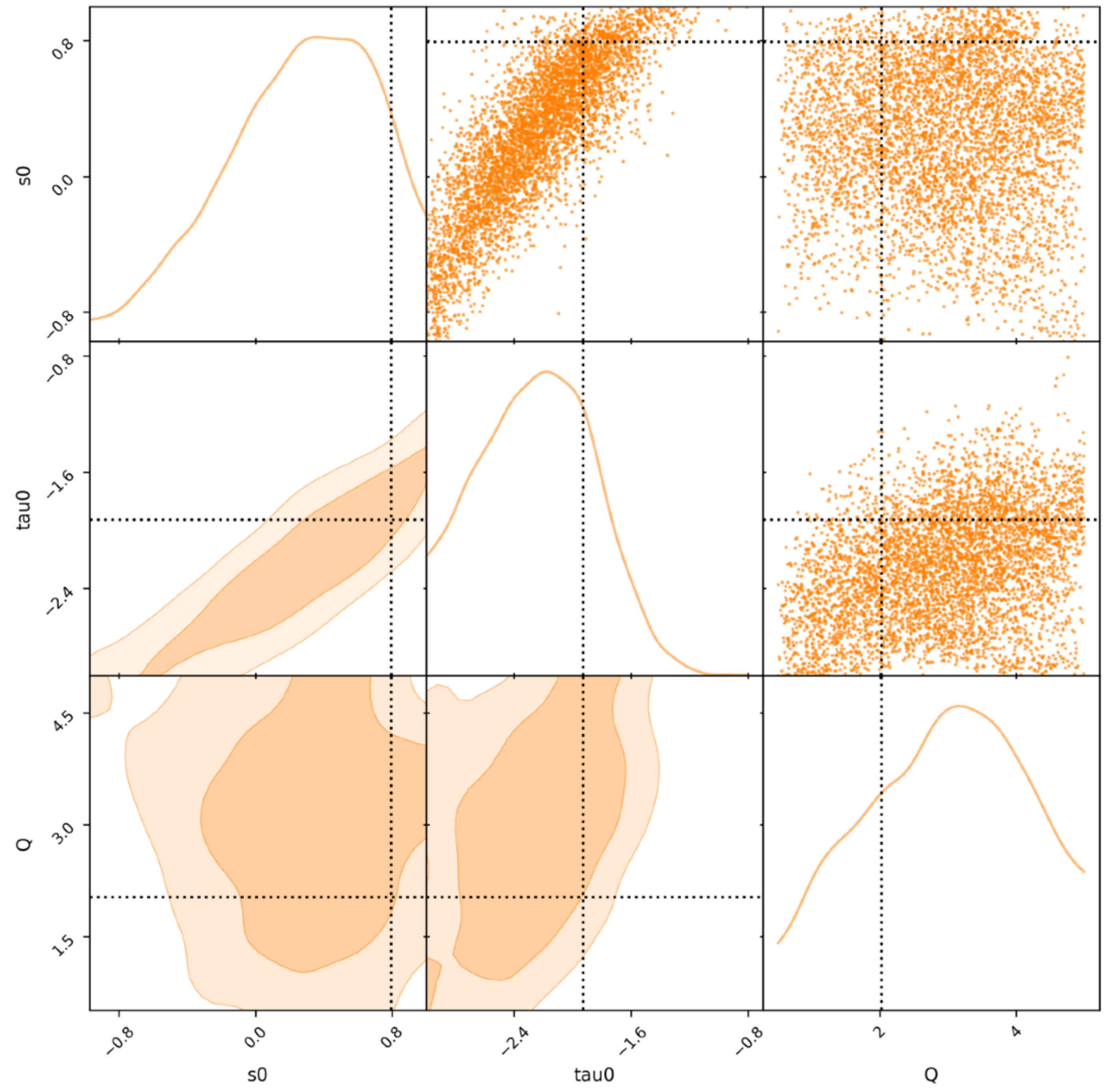
Tracing star formation

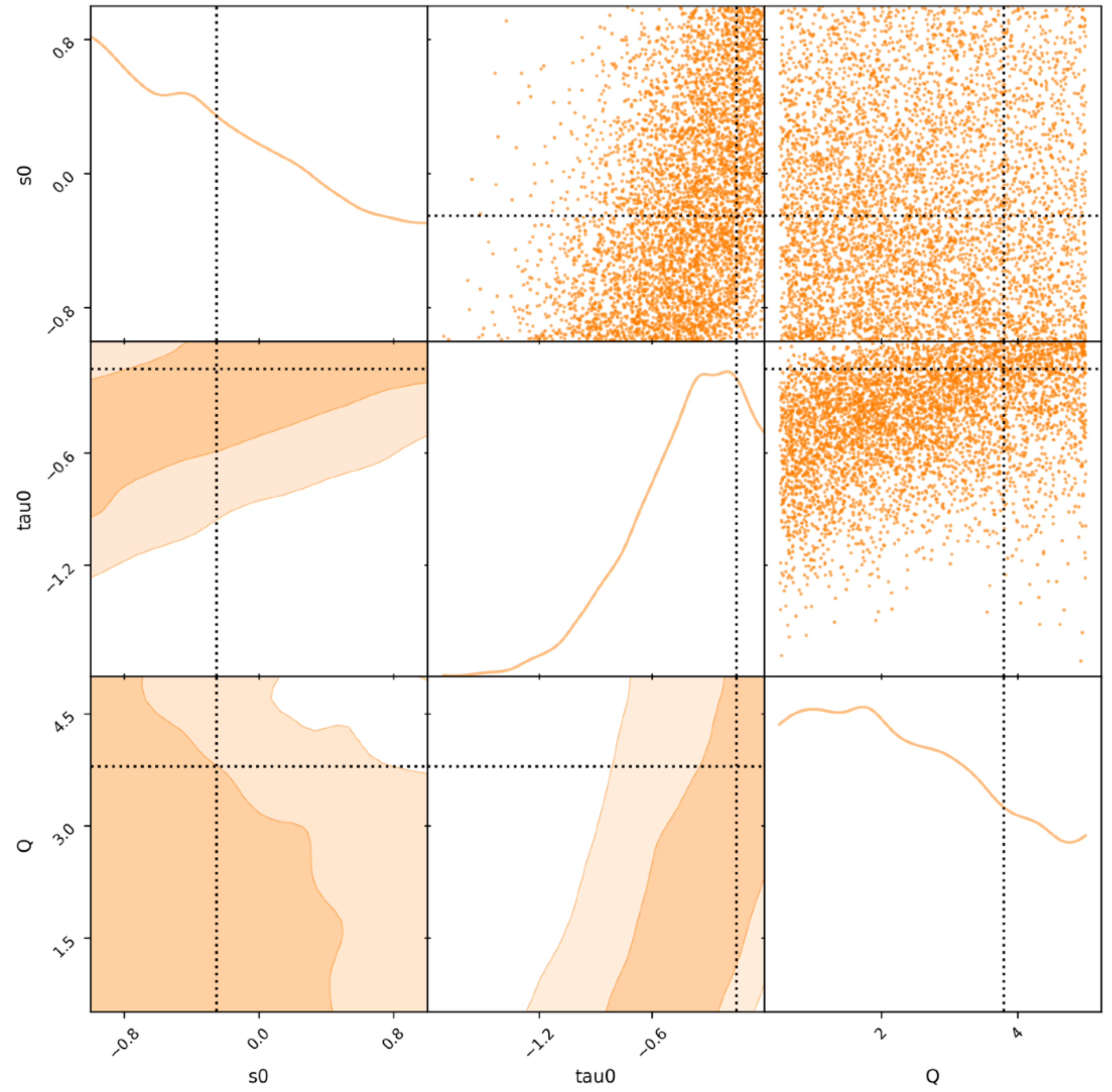
- Distribution of various star formation tracers
- $M_*, H_\alpha, D_n(4000), F_{UV}, F_{NUV}, H_\delta^{EW}$
- SDSS, WISE and 2MASS fluxes
- Iyer et al. 2022 [2208.05938]



Neural Posterior Estimation

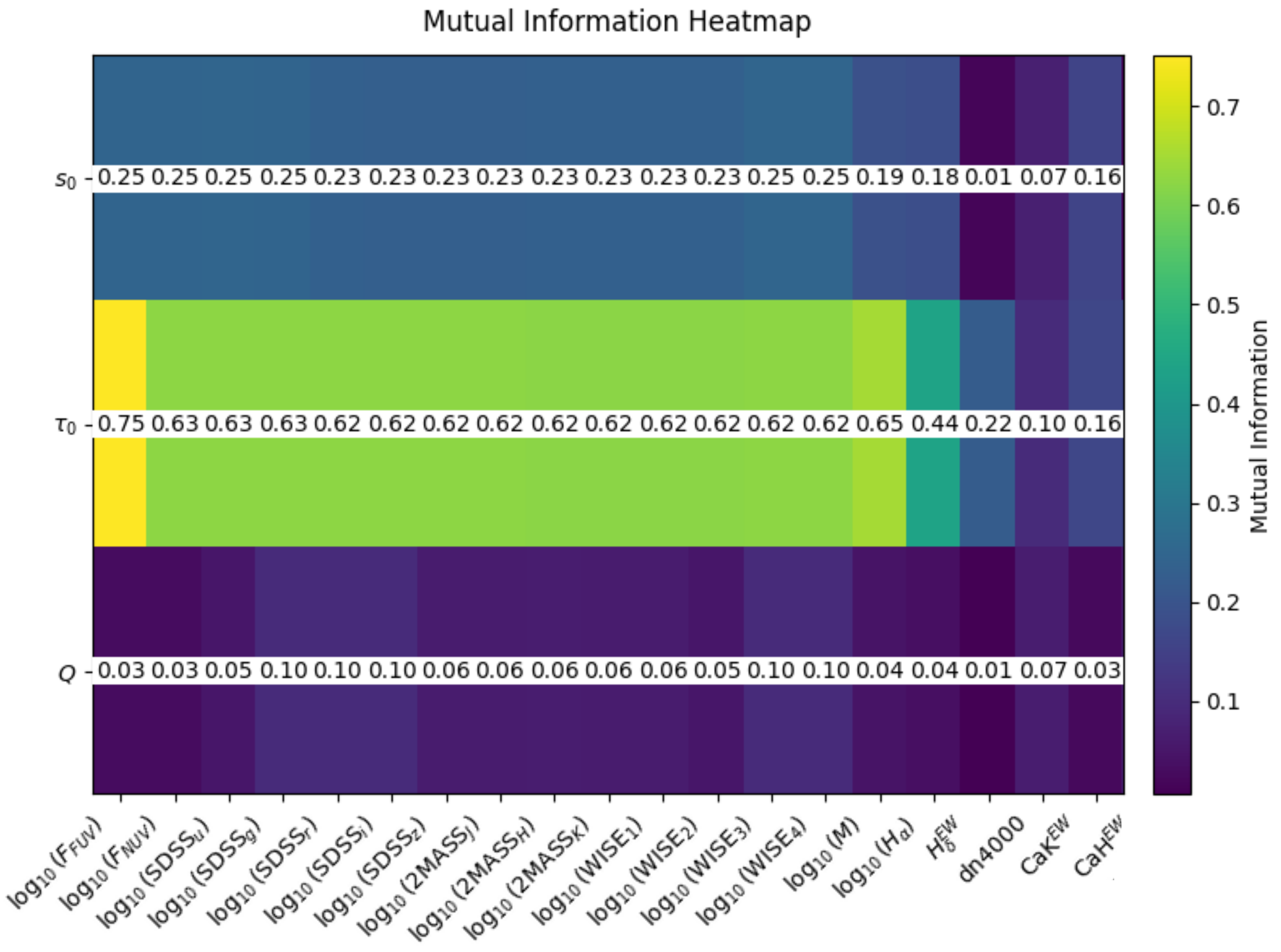
- Generate corresponding pairs of θ_{PSD} and $\{D_{\text{tracers}}\}^{N_{\text{gal}}}$
- Train a conditional normalising flow to learn $P(\theta_{PSD}\,|\,\{D_{\text{tracers}}\}^{N_{\text{gal}}})$
- Cycle through phases of training and inference contracting priors around the posterior bulk
- Terminate inference when $D_{KL}(P_i\,||\,P_{i-1}) < x$
- Validate training with C2ST





Constraining power of the lines

- Planned to apply this technique to SDSS and potentially surveys from JWST
- Very little information about the PSDs in the tracers
- Look for other probes of star formation…



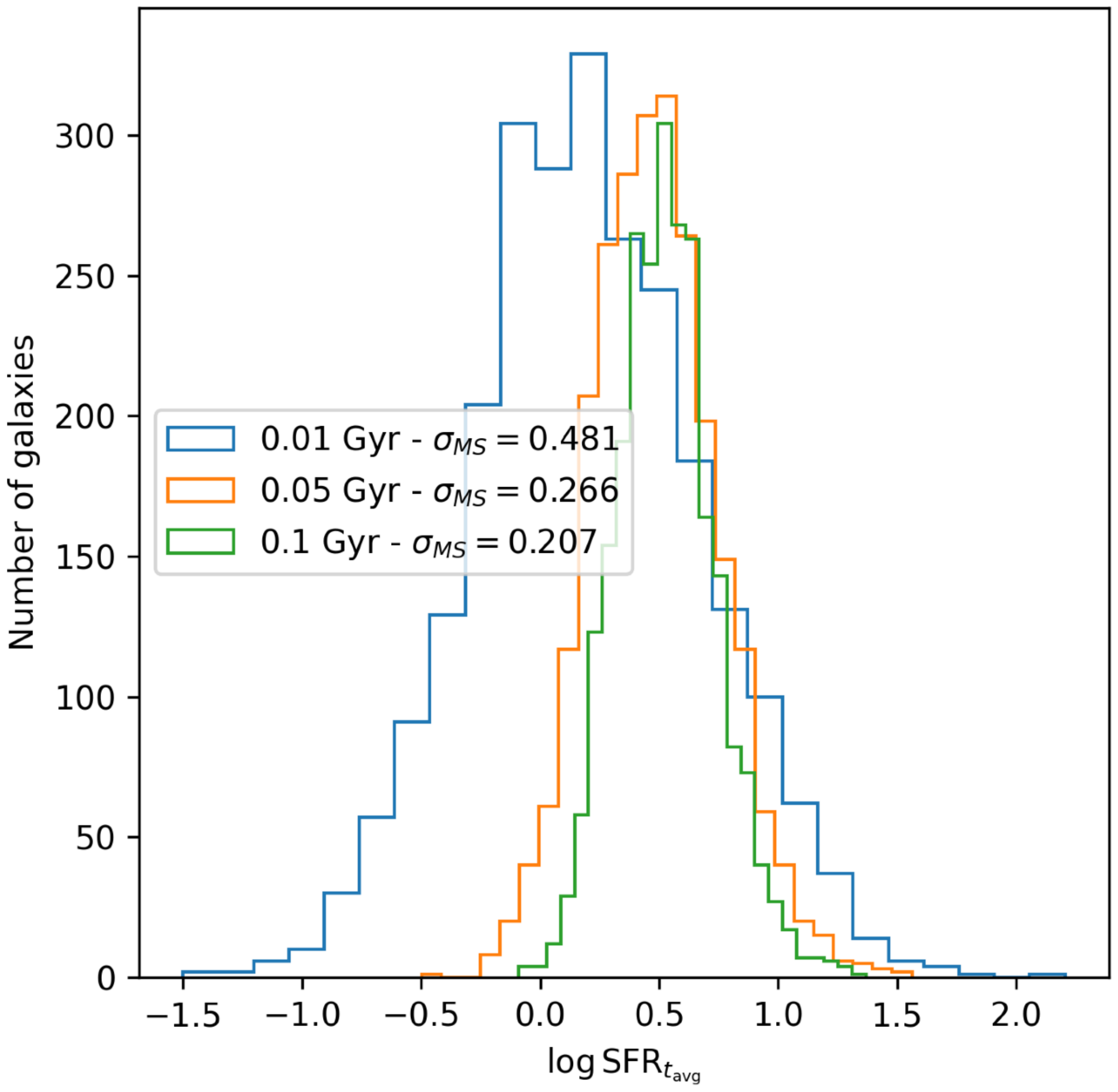
Main Sequence Scatter

- Different tracers of star formation such as H_α and $D_n(4000)$ provide different information of different time scales

- In theory we could work more directly with the SFH generated from our GP

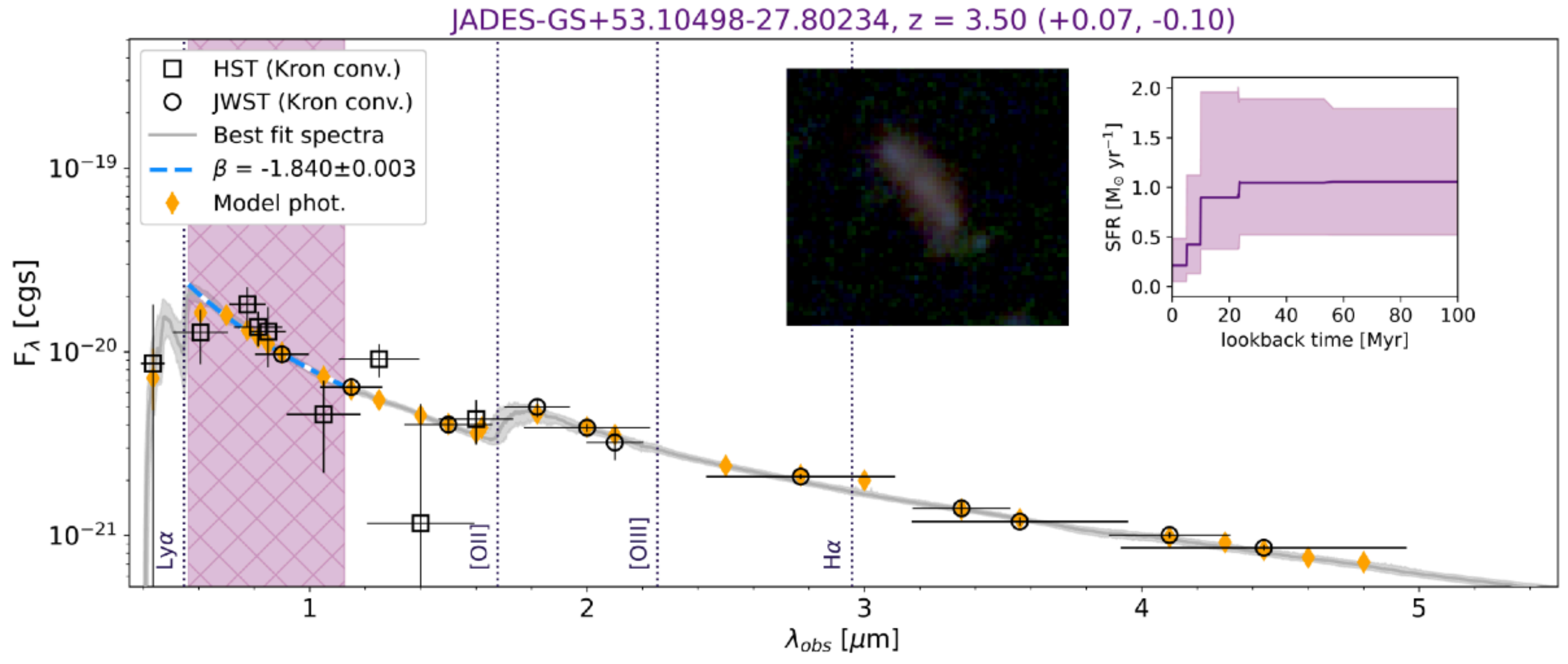
$$\mathrm{SFR}_{t_{\mathrm{avg}}} = \frac{\int_{t_{\mathrm{avg}}}^{t_0} \mathrm{SFR}(t')dt'}{t_{\mathrm{avg}}}$$

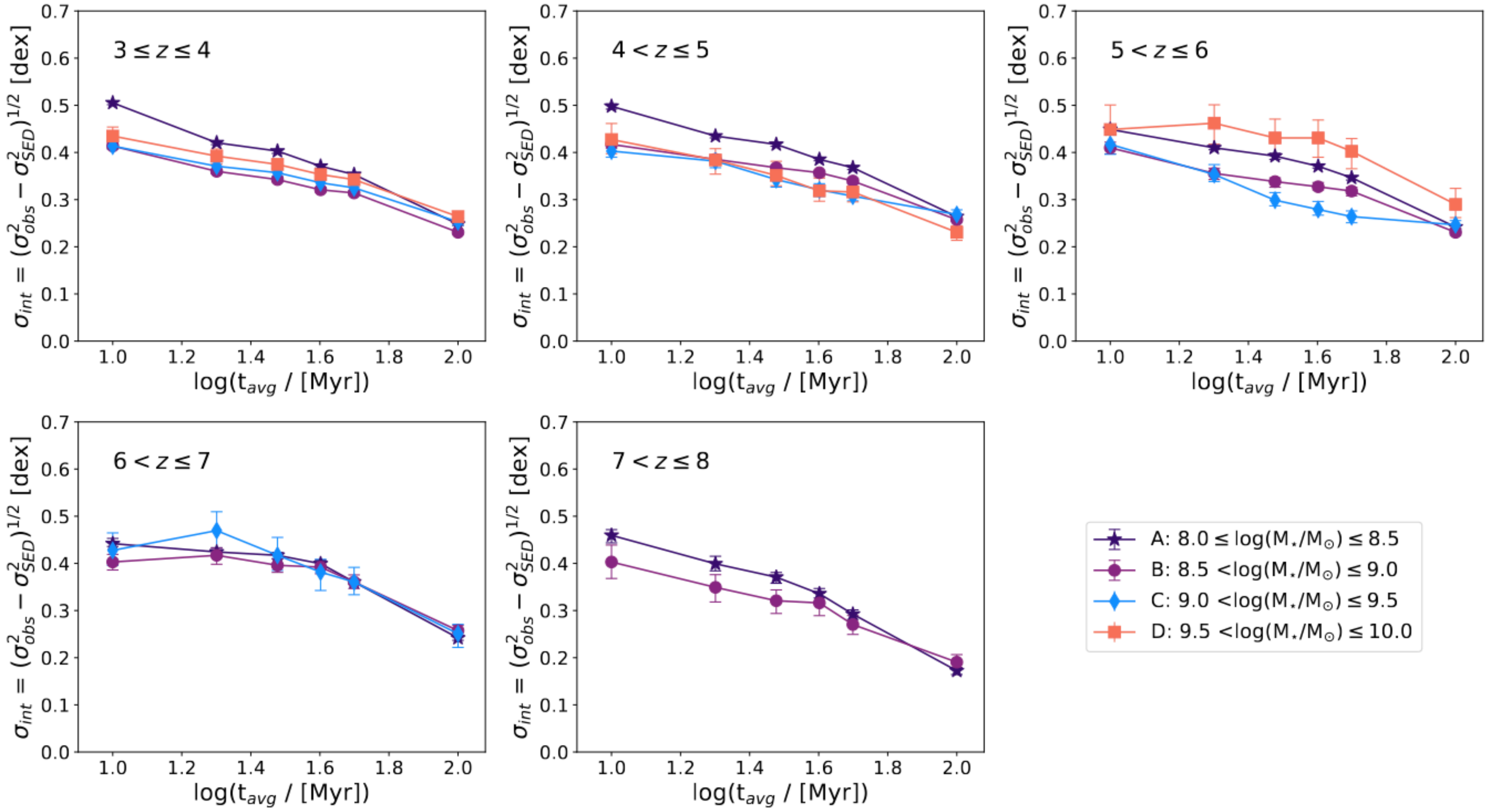
- Initially investigate constraining power of the scatter around the mean $\mathrm{SFR}_{t_{\mathrm{avg}}}$ denoted $\sigma_{MS}(t_{\mathrm{avg}})$



NIRCam Observations

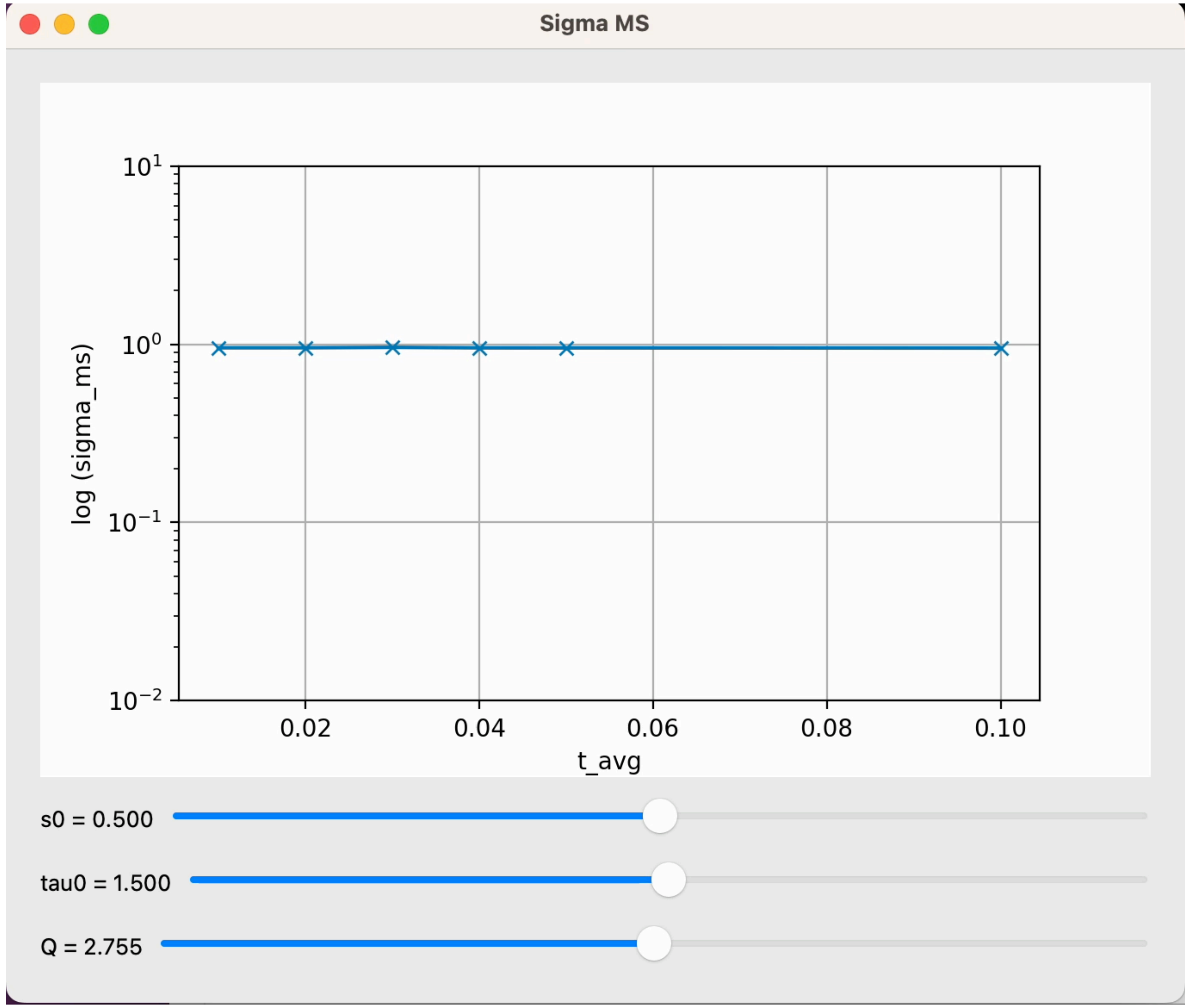
- Simmonds et al. in prep and Simmonds et al 2024 [2409.01286]
- Fit photometric catalogue ($\sim 50{,}000$ galaxies) with FSPS and piecewise SFR
- Divide catalogue into mass and redshift bins
- Integrate over SFR on a range of different timescales
- Intrinsic scatter $\sigma_{MS}^{\text{int}} = \sqrt{\sigma_{MS}^{\text{obs}\ 2} - \sigma_{\text{SED}}^{2}}$



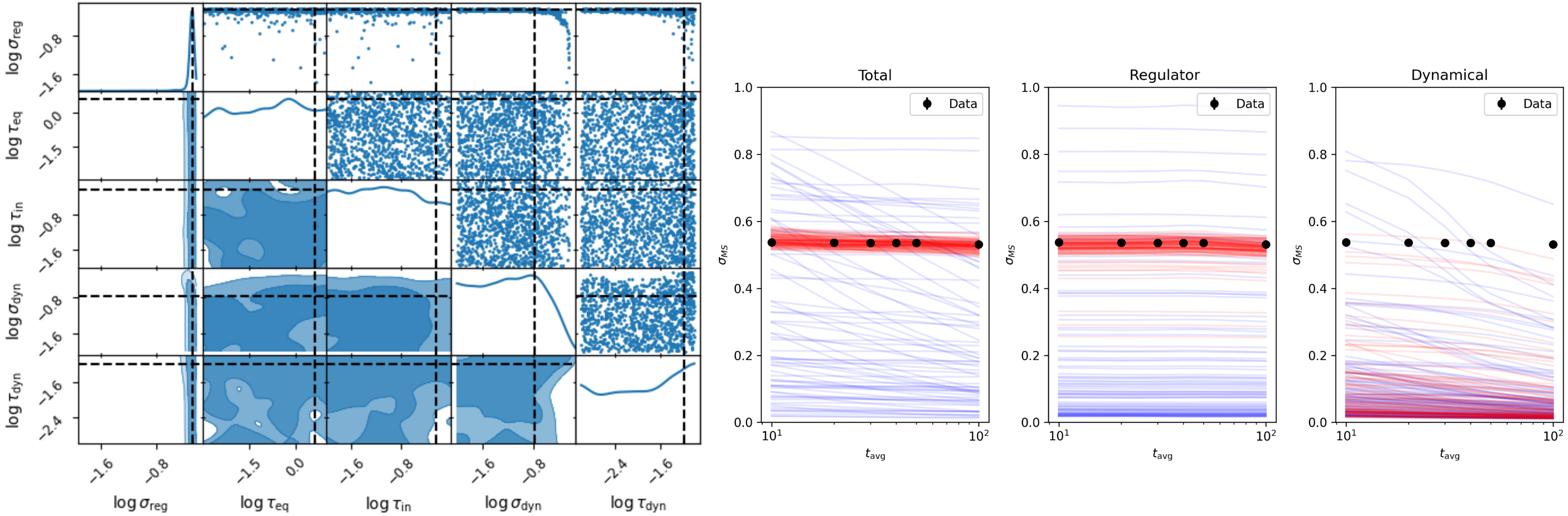


Likelihood Based Inference and Emulation

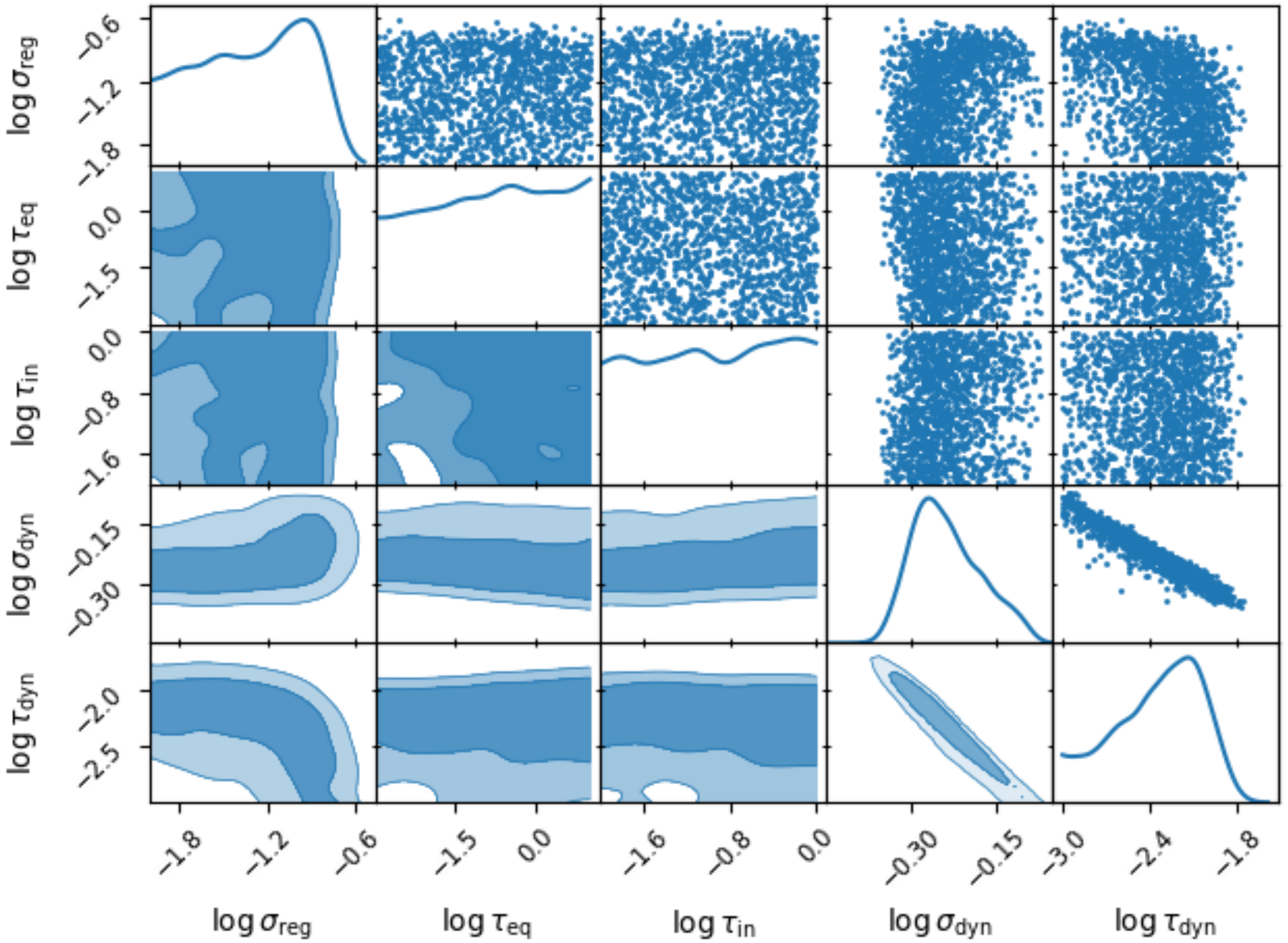
- Use the same simulation chain as before to estimate $\sigma_{\mathrm{MS}}(\theta_{PSD})$
- Generate 10,000 PSD samples and for each 2500 SFHs
- From these estimate $\sigma_{MS}(\theta_{PSD})$ and train a neural network emulator
- Assume a gaussian likelihood and use measured error on data
- Account for uncertainty in emulator



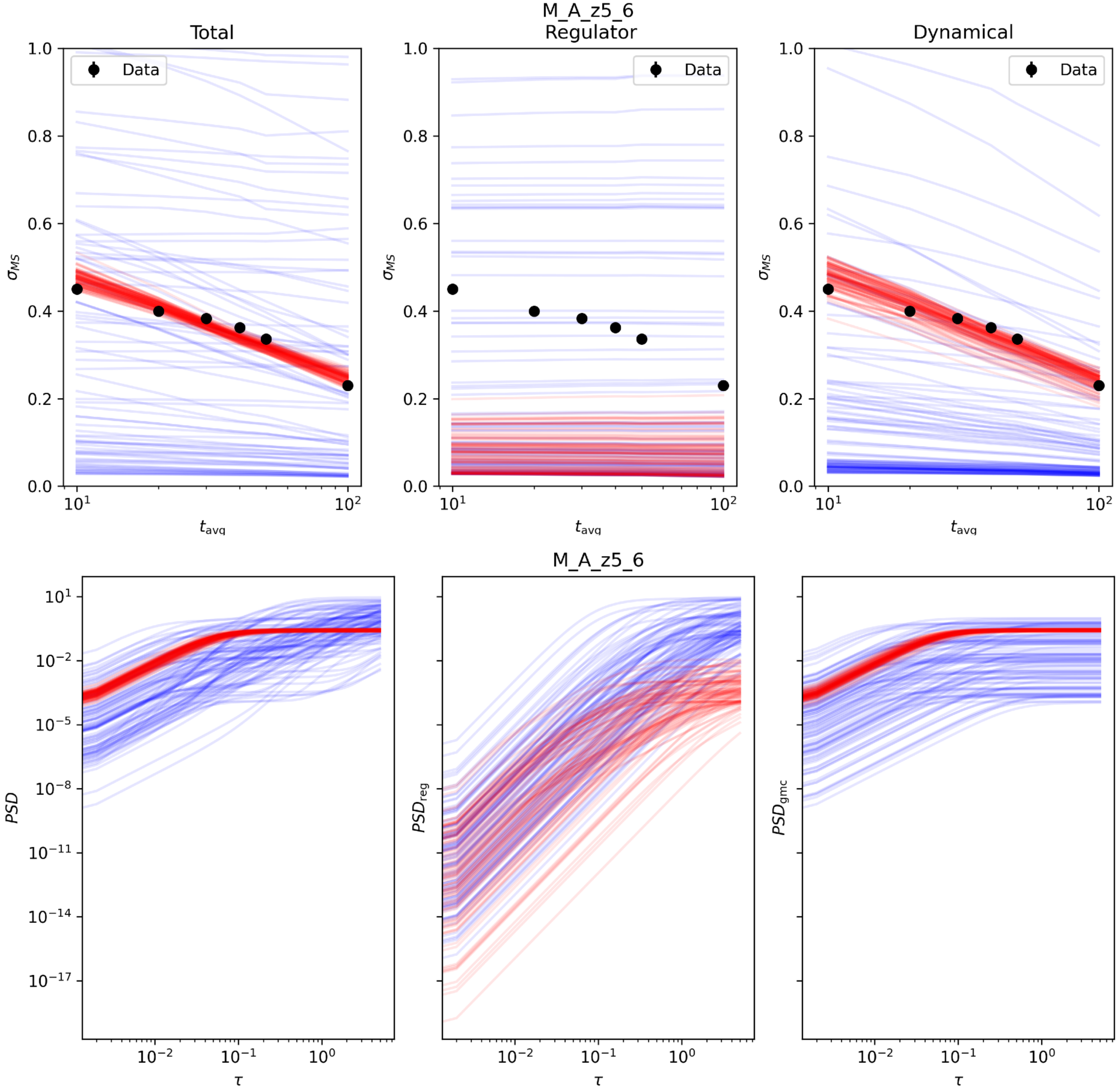
Likelihood Based Inference and Emulation



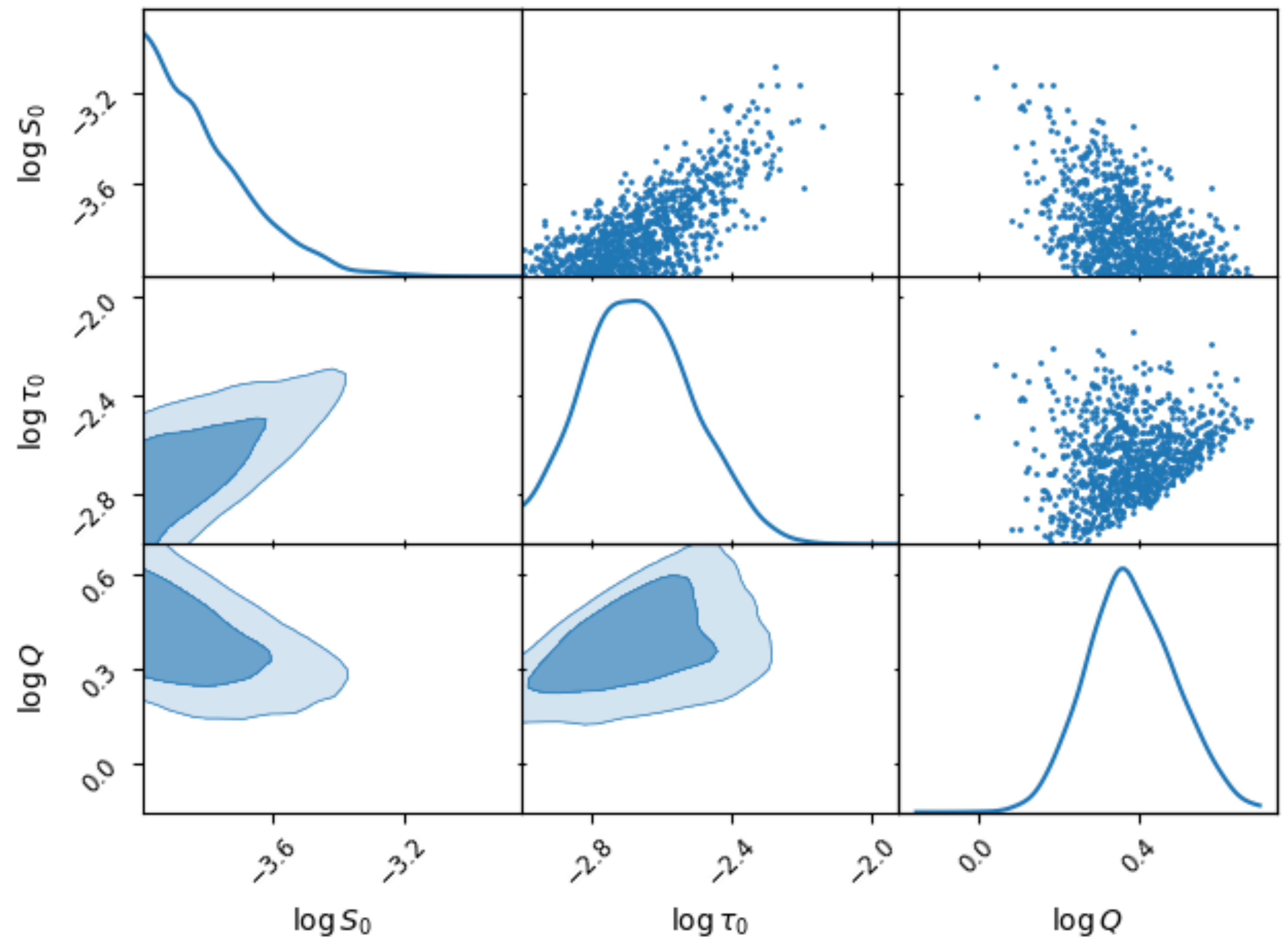
Likelihood Based Inference and Emulation



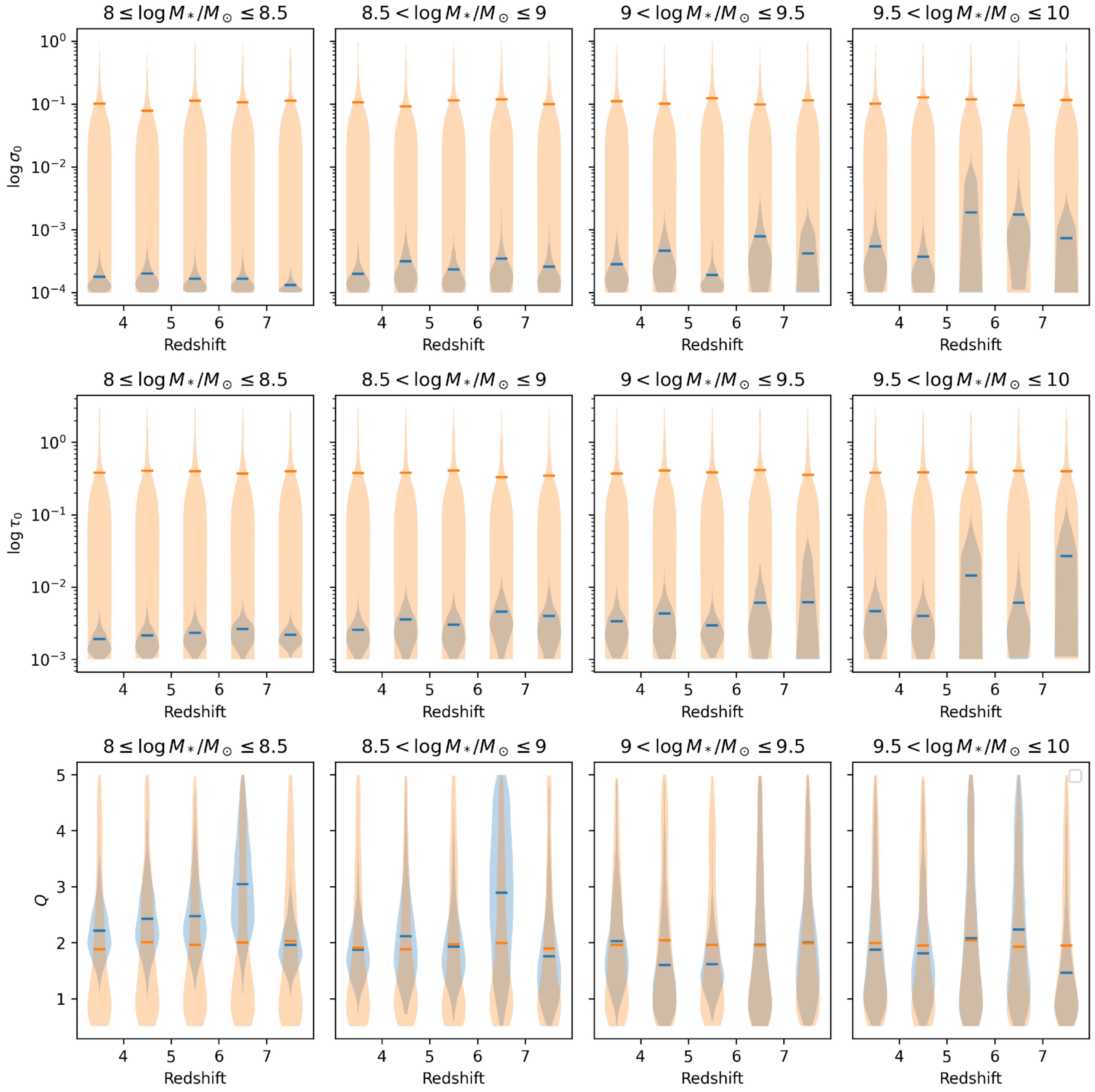
$$8 \leq \log M_*/M_\odot \leq 8.5$$
$$5 < z < 6$$



Likelihood Based Inference and Emulation



$$8 \leq \log M_*/M_\odot \leq 8.5$$
$$5 < z < 6$$



Neural Ratio Estimation

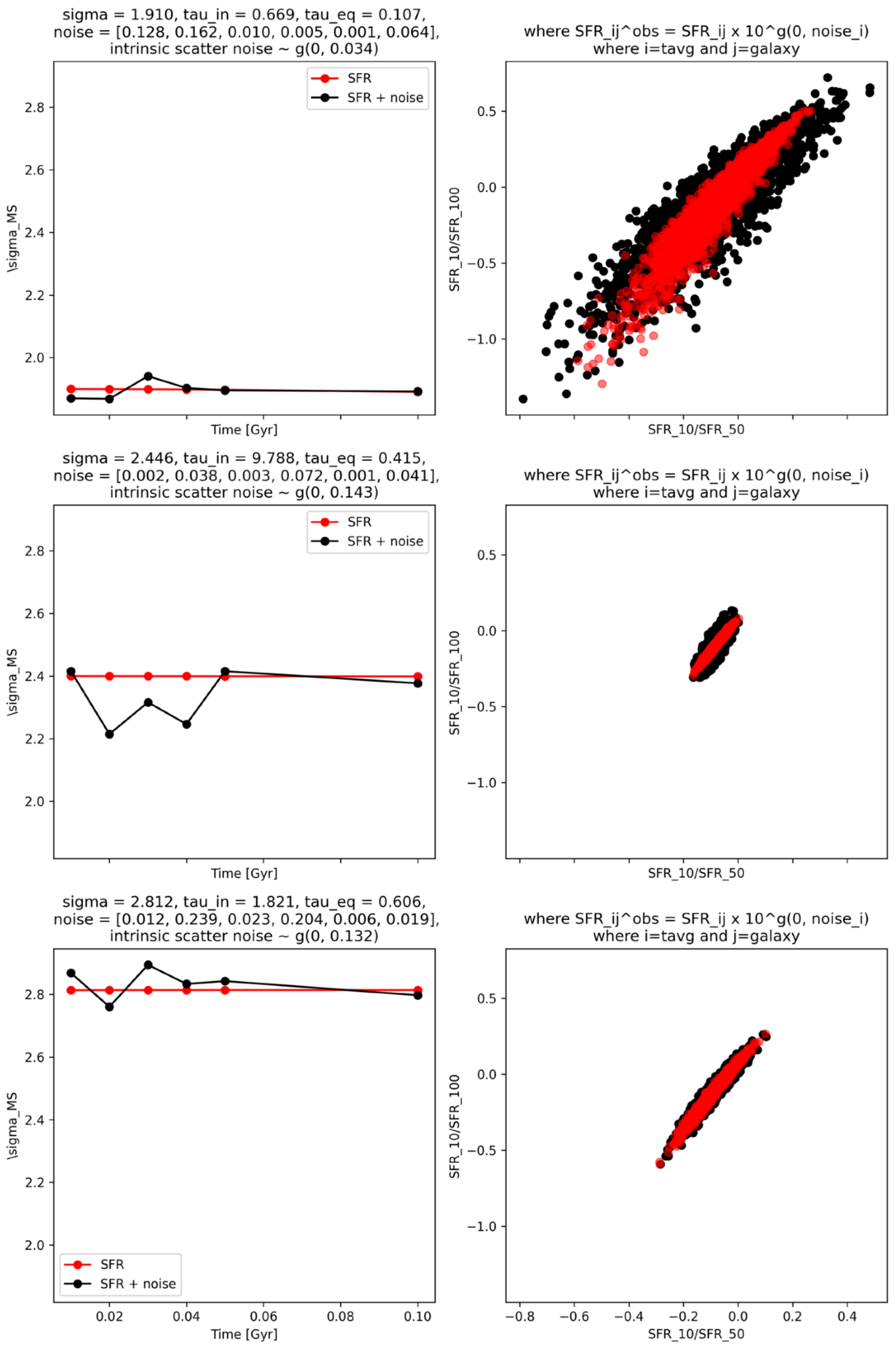
- Currently investigating applications of NRE to the estimation of $P(\theta_{\mathrm{PSD}} \,|\, \sigma_{\mathrm{MS}})$

- Goal is to better account for the uncertainty introduced by the SED modelling

- For each galaxy in our simulation we calculate

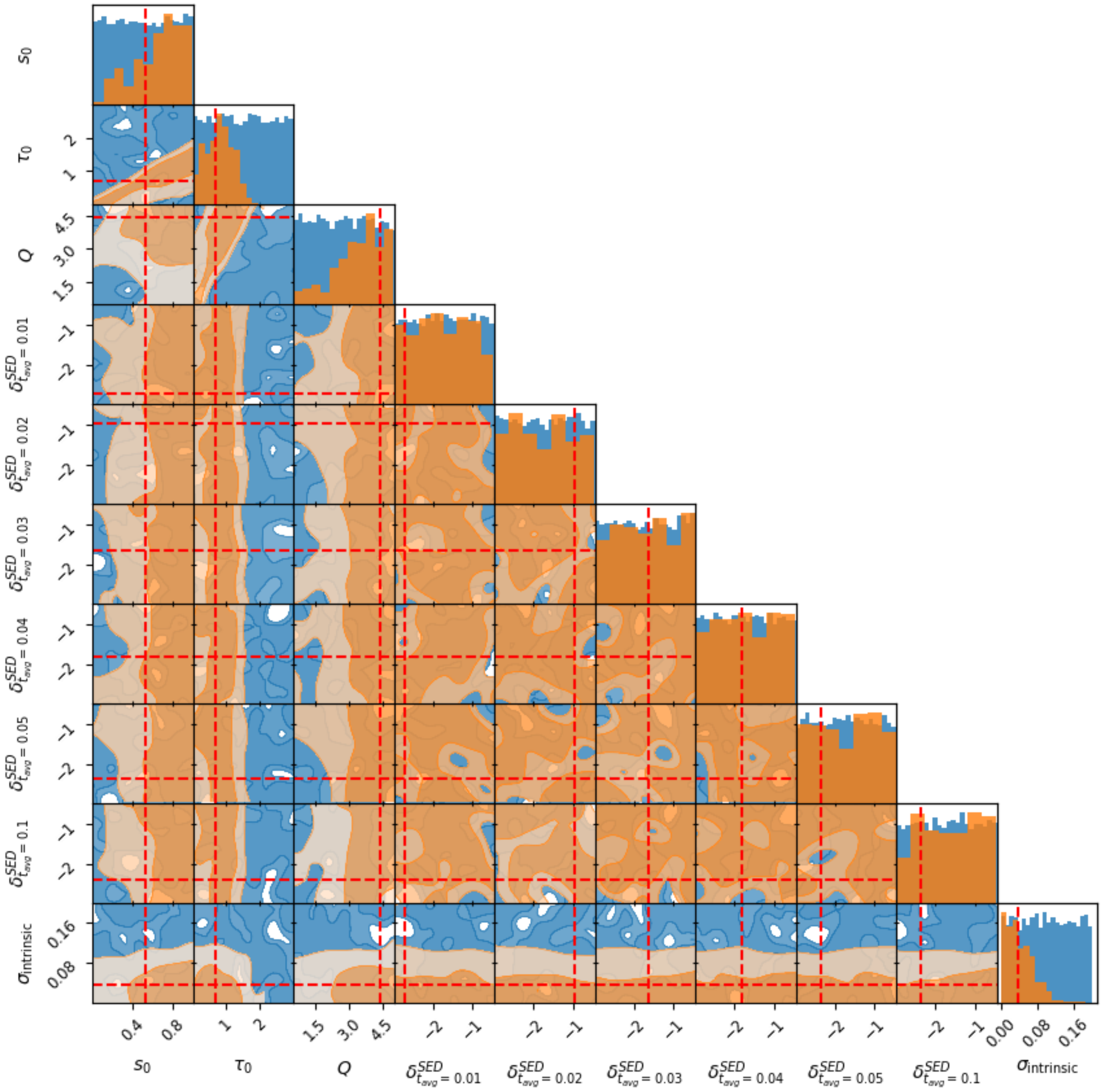
$$\mathrm{SFR}^{\mathrm{obs}}_{t_{\mathrm{avg}}} = \mathrm{SFR}^{int}_{t_{\mathrm{avg}}} \times 10^{\alpha}$$

$$\alpha \sim \mathcal{N}\left(0, \delta^{SED}_{t_{\mathrm{avg}}}\right)$$



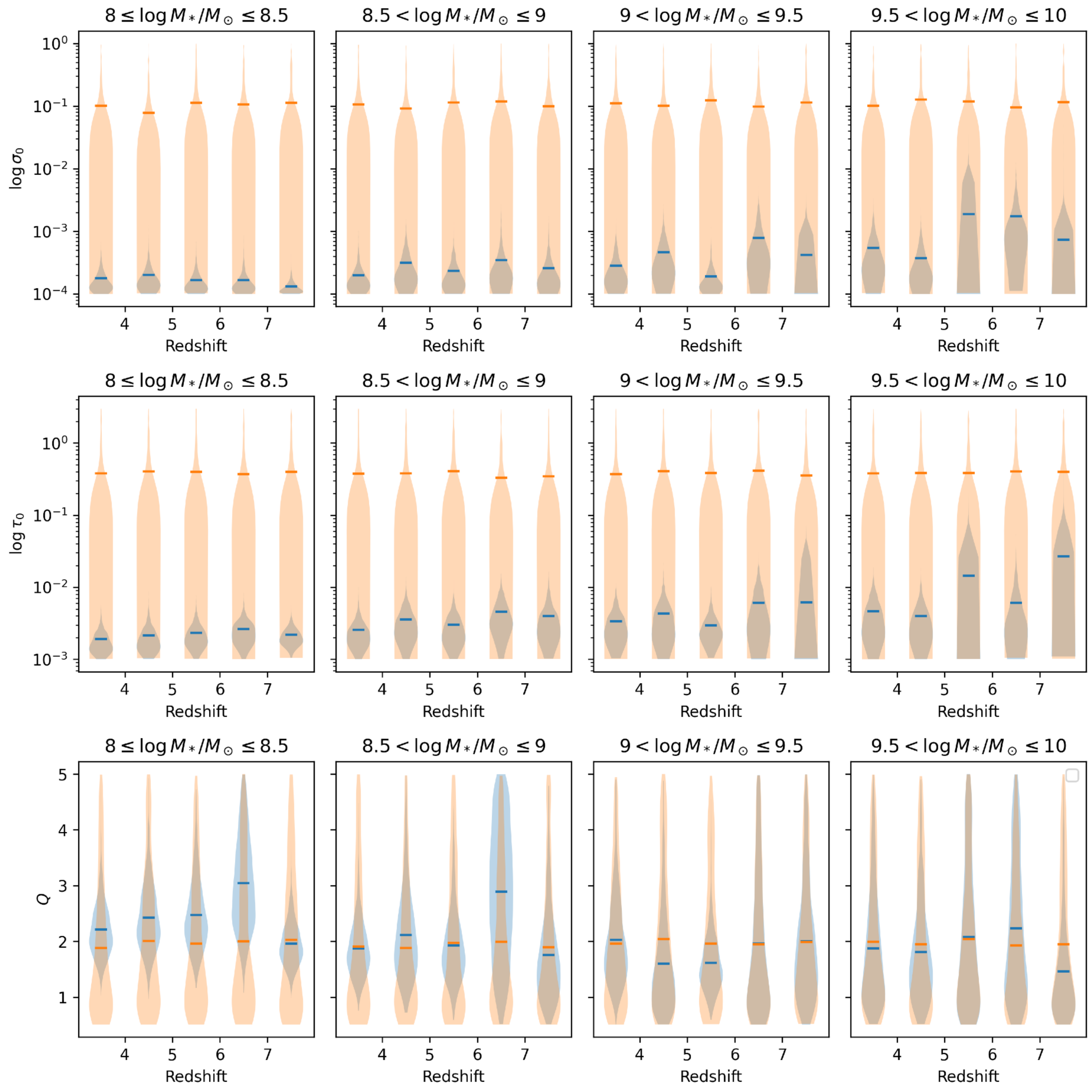
Neural Ratio Estimation

- Trained the network on 20000 PSD samples with 2500 galaxies per sample
- Focusing on simpler SHO model at the moment
- Pass the simulated data, PSD parameters and noise parameters through separate linear layers before mixing the information
- Tuned with optuna and currently testing on a large set of simulations



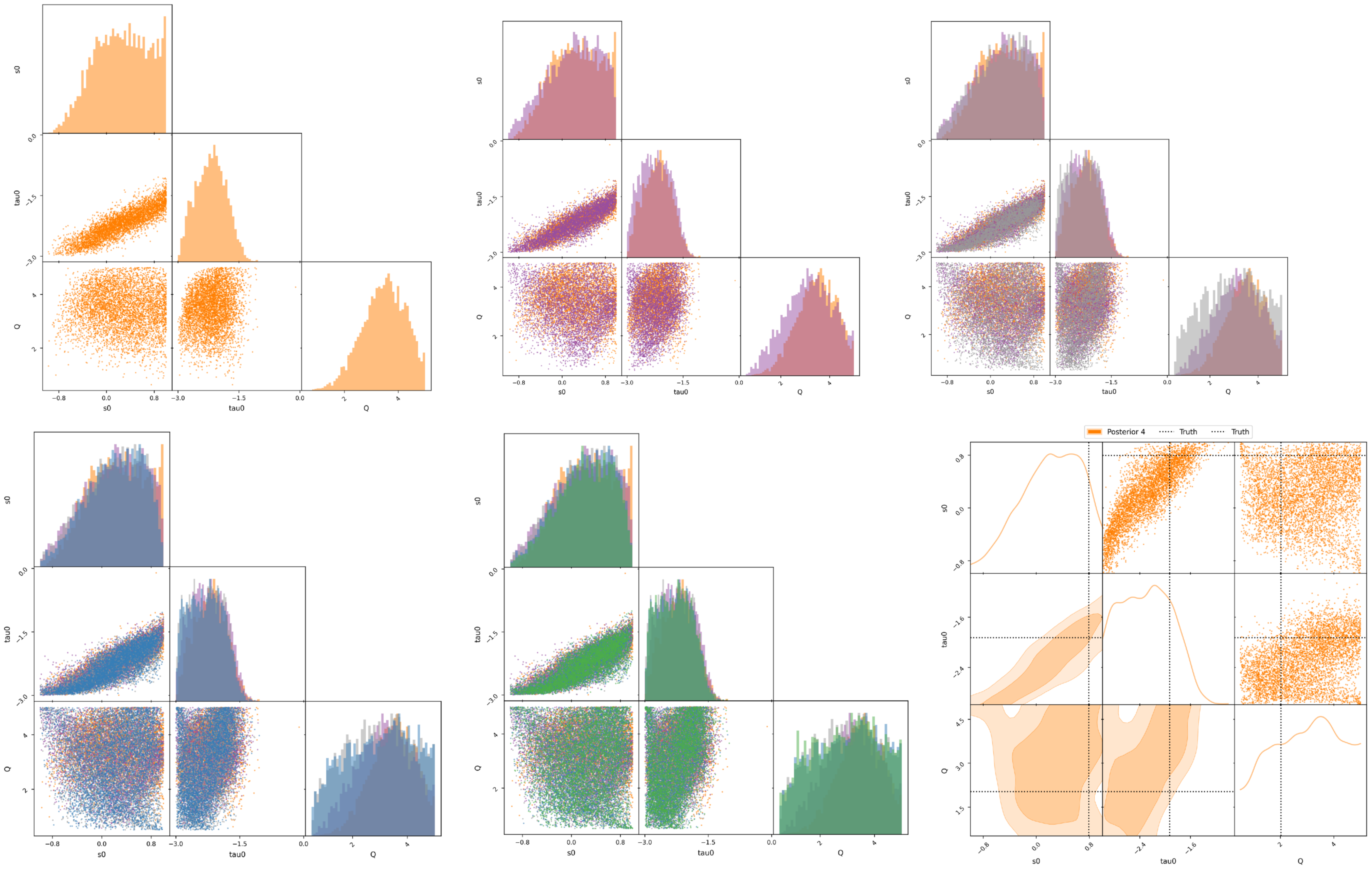
Conclusions

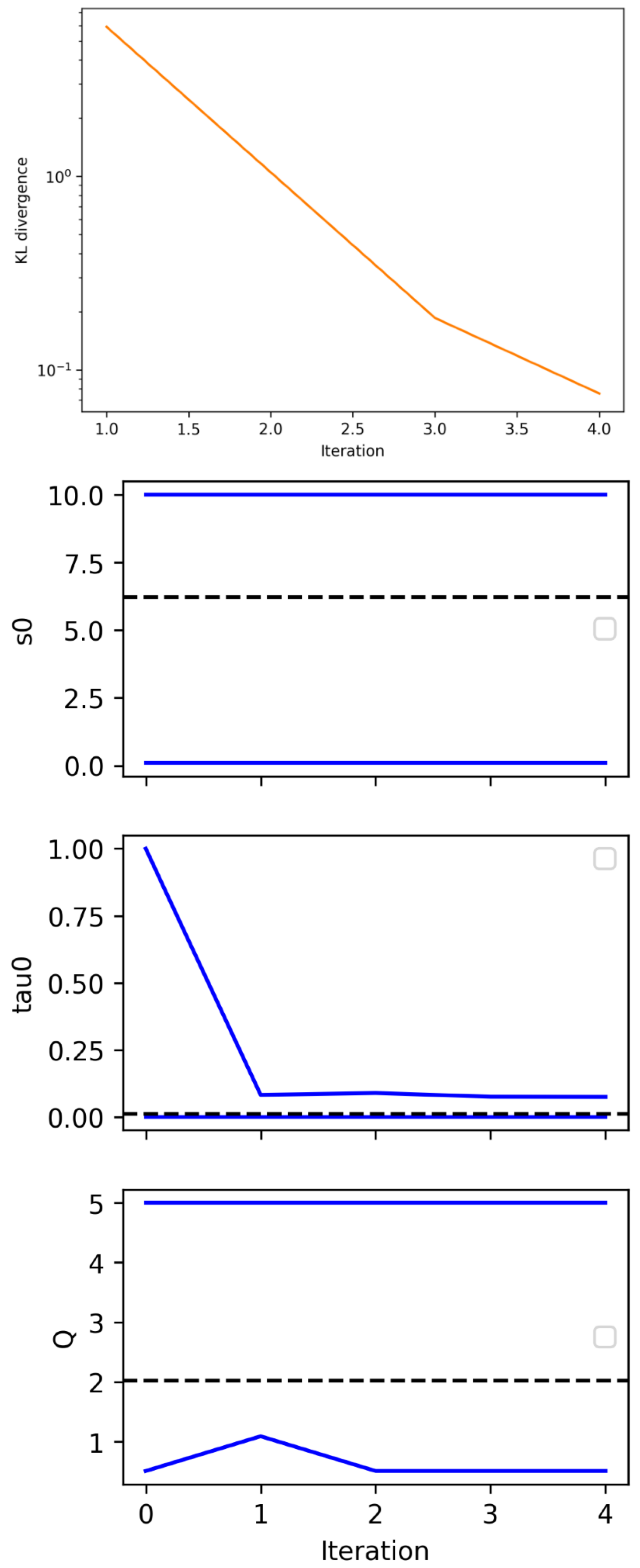
- Apply the NRE to the observations after a bit more tuning
- Planning to investigate constraints from ratios of $\mathrm{SFR}_{t_{\mathrm{avg}}}$ for various t_{avg}
- Revisit the emission line work and see if we can improve the constraints using the results from the MS scatter work to inform our tracer choices
- Write up, polish and publish some of our results



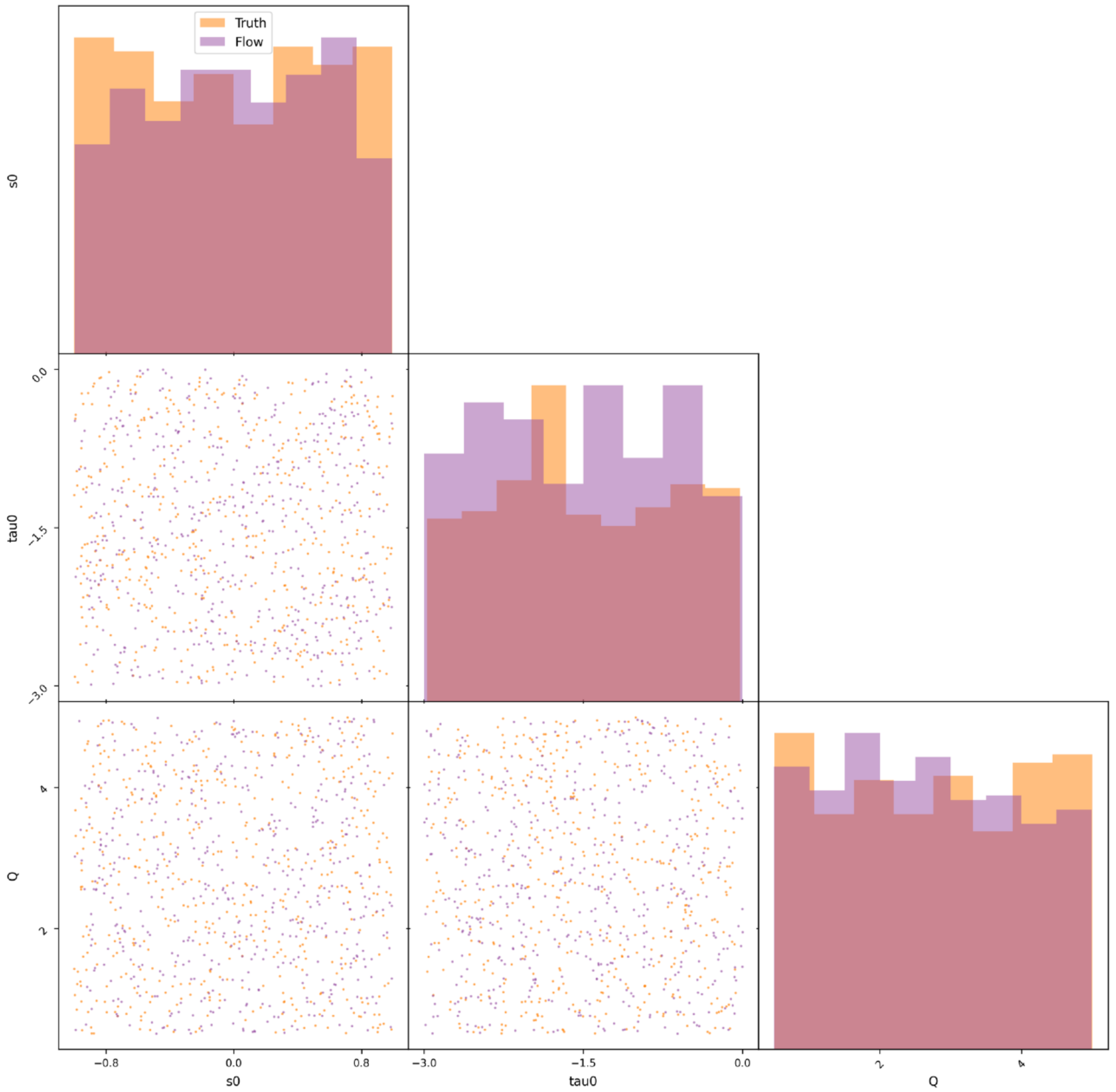
Additional Slides

NPE - Prior Contraction and Iterations

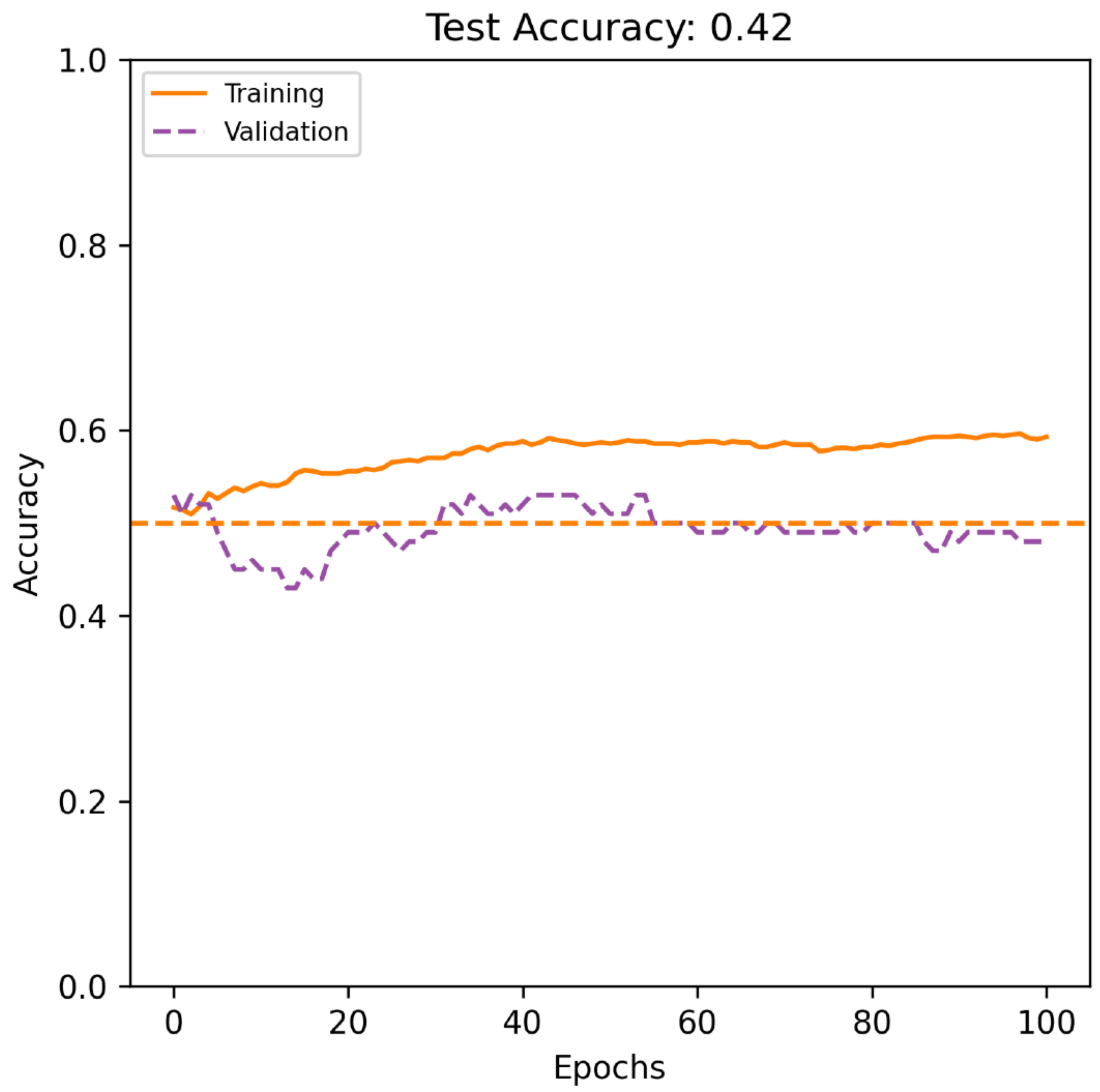




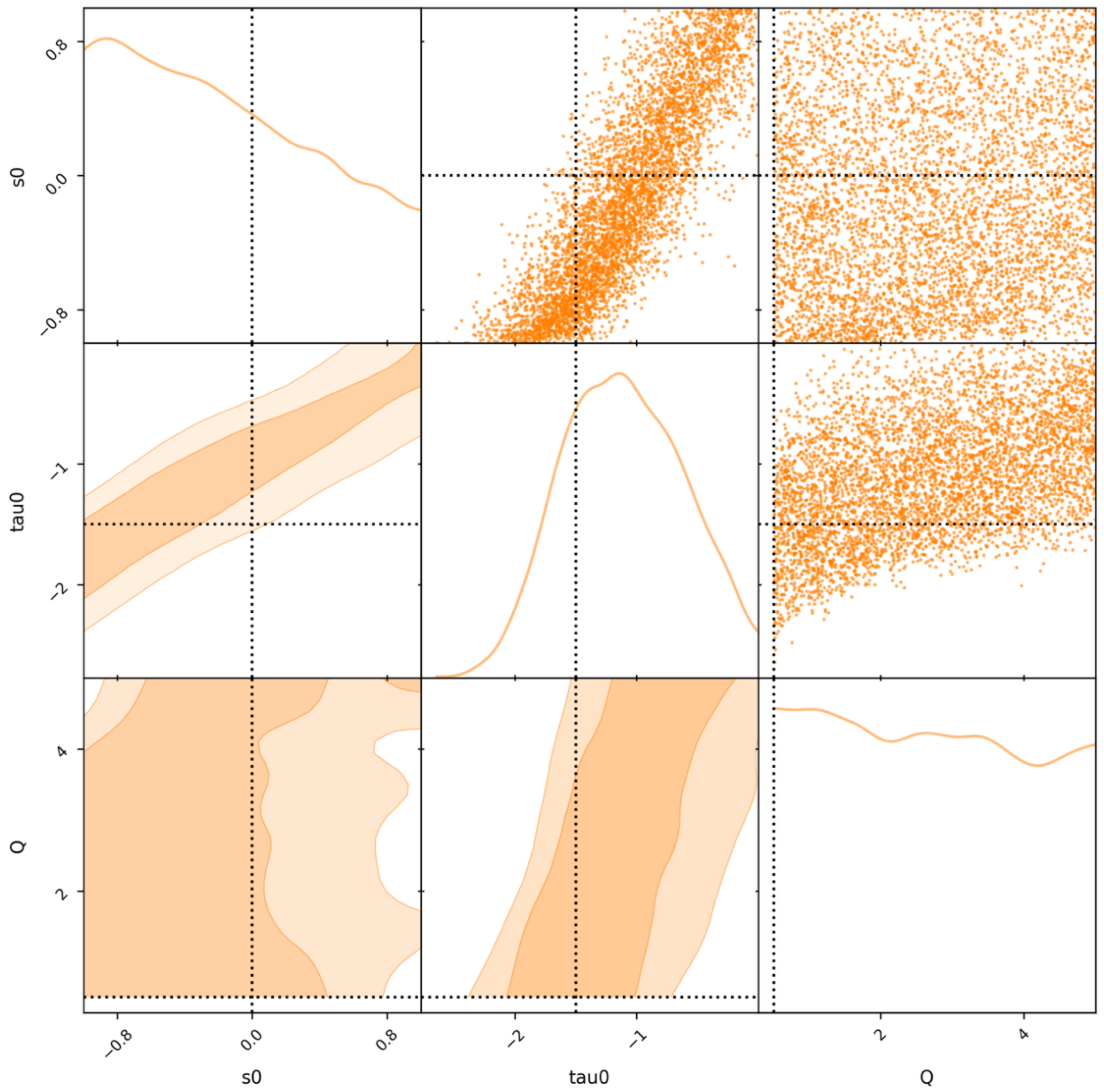
NPE - C2ST



Truth
Flow
s0
tau0
Q

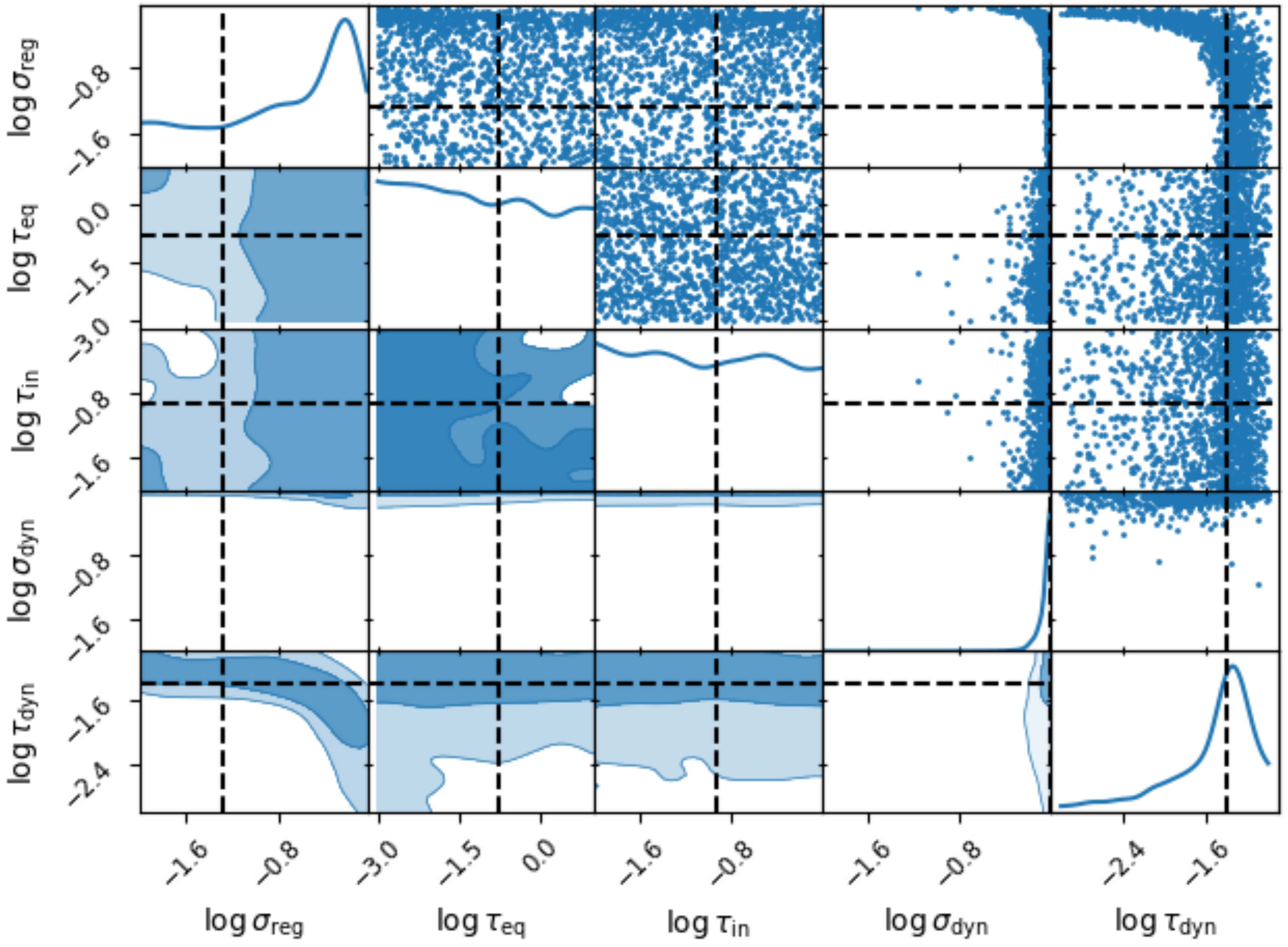


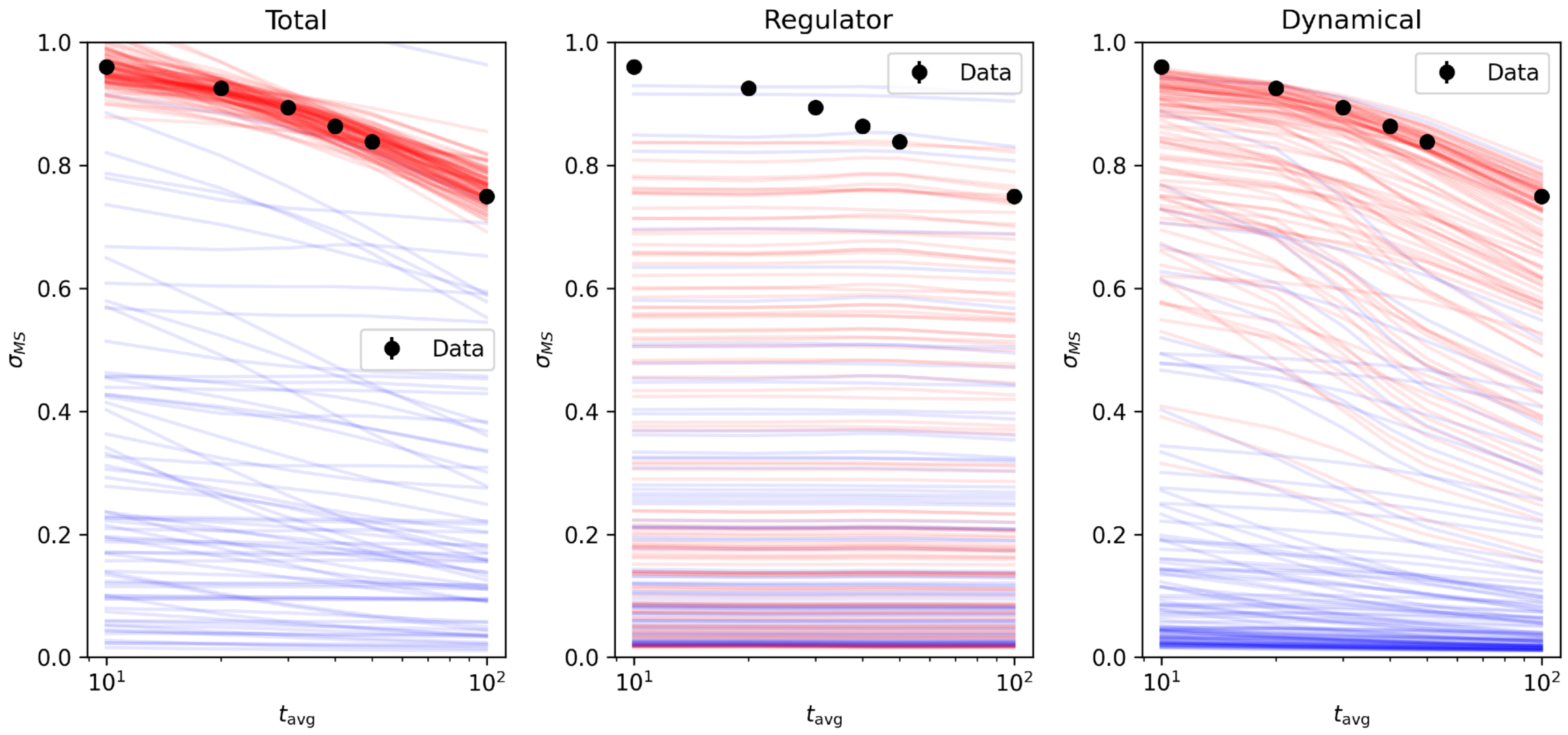
Test Accuracy: 0.42
Training
Validation
Accuracy
Epochs



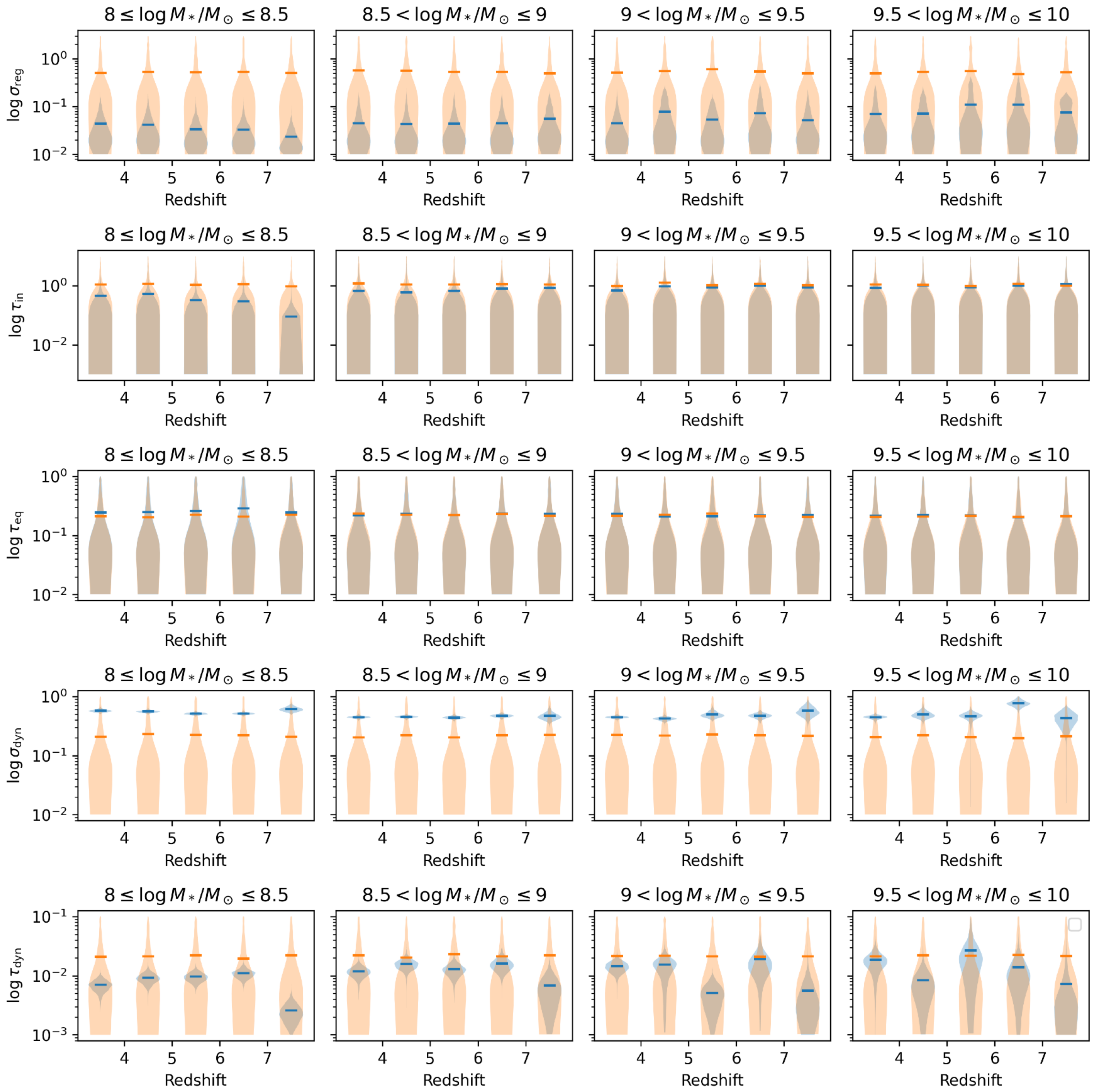
s0
tau0
Q

Likelihood Based Inference and Emulation - Example 2

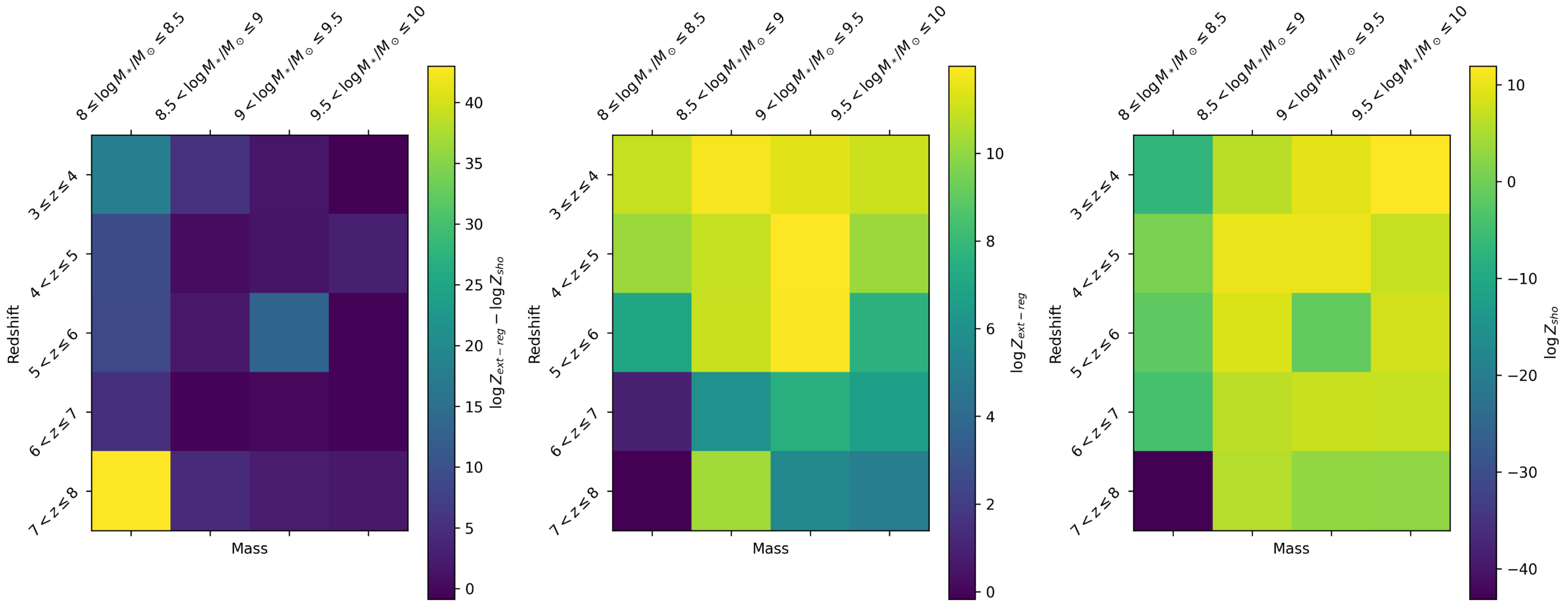




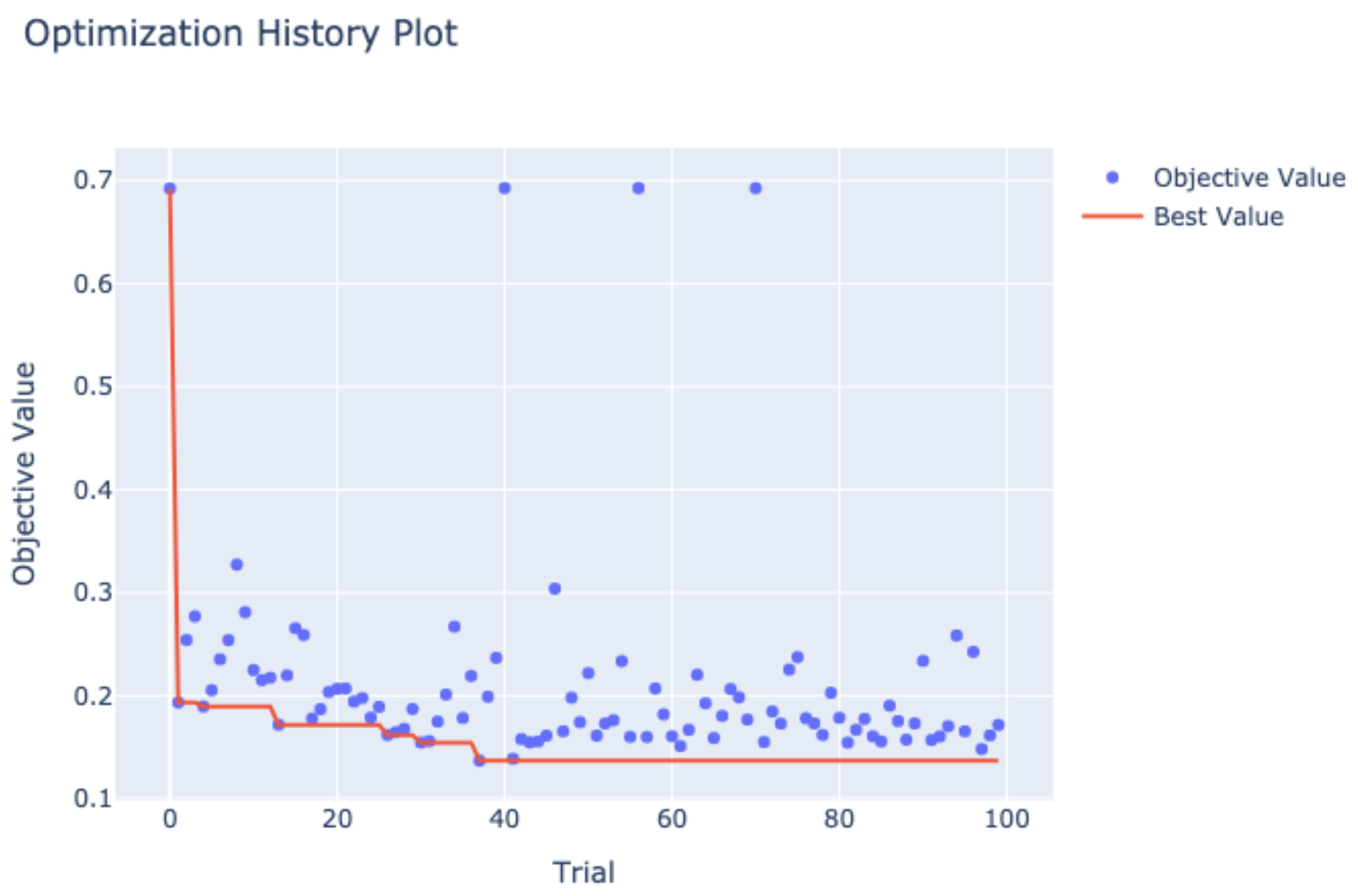
Likelihood Based Inference and Emulation

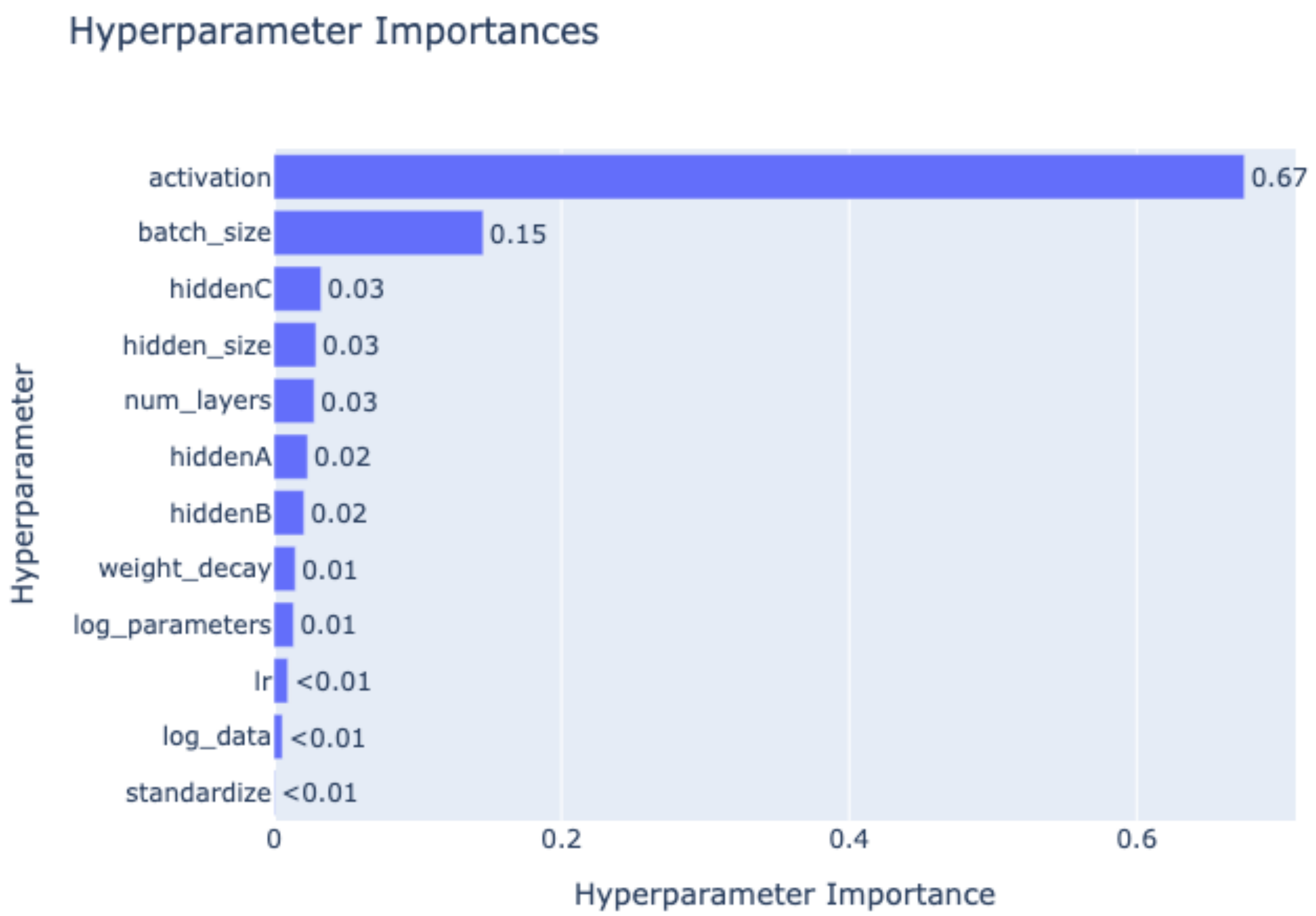


Likelihood Based Inference and Emulation

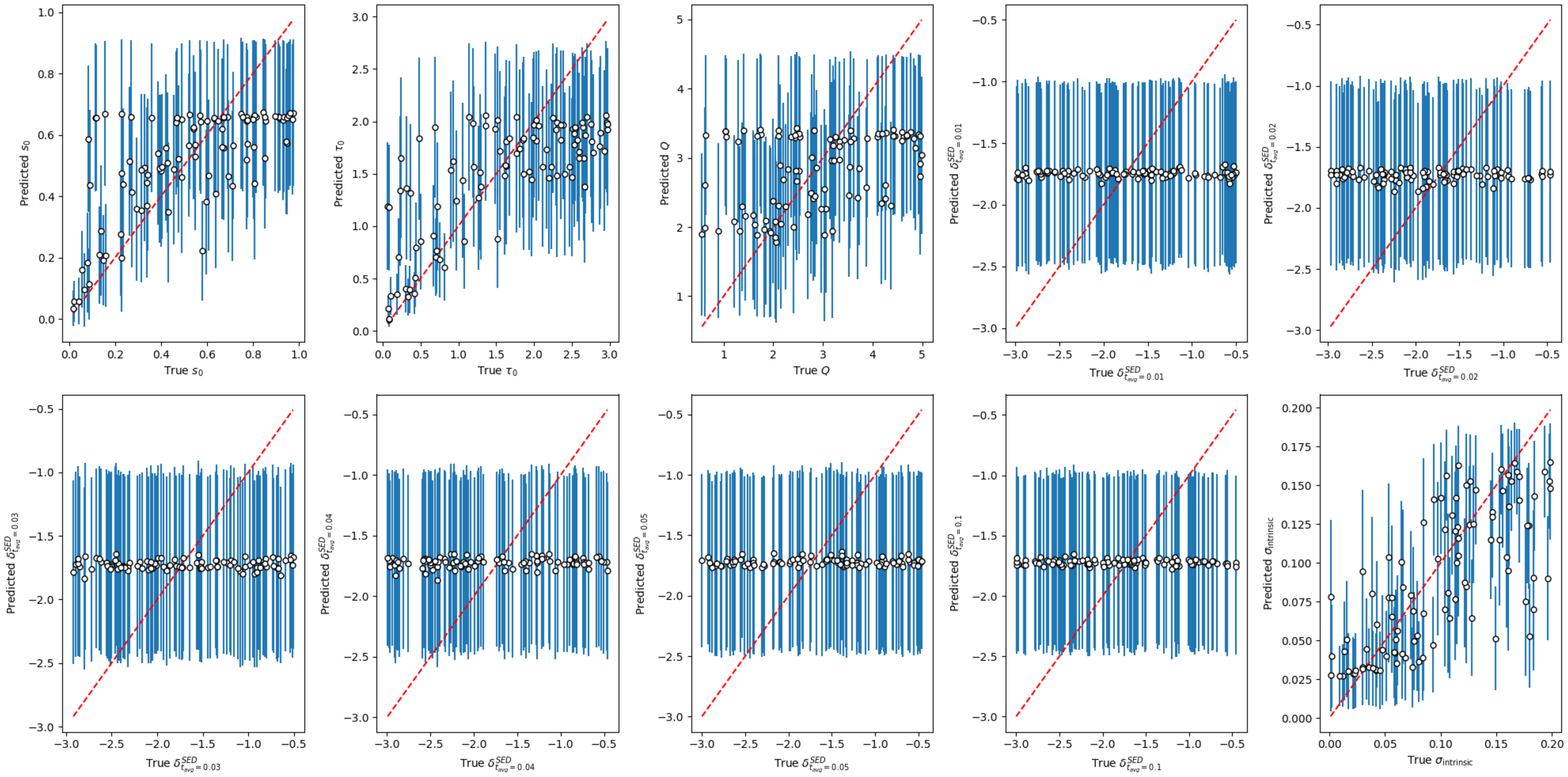


Neural Ratio Estimation - Optuna

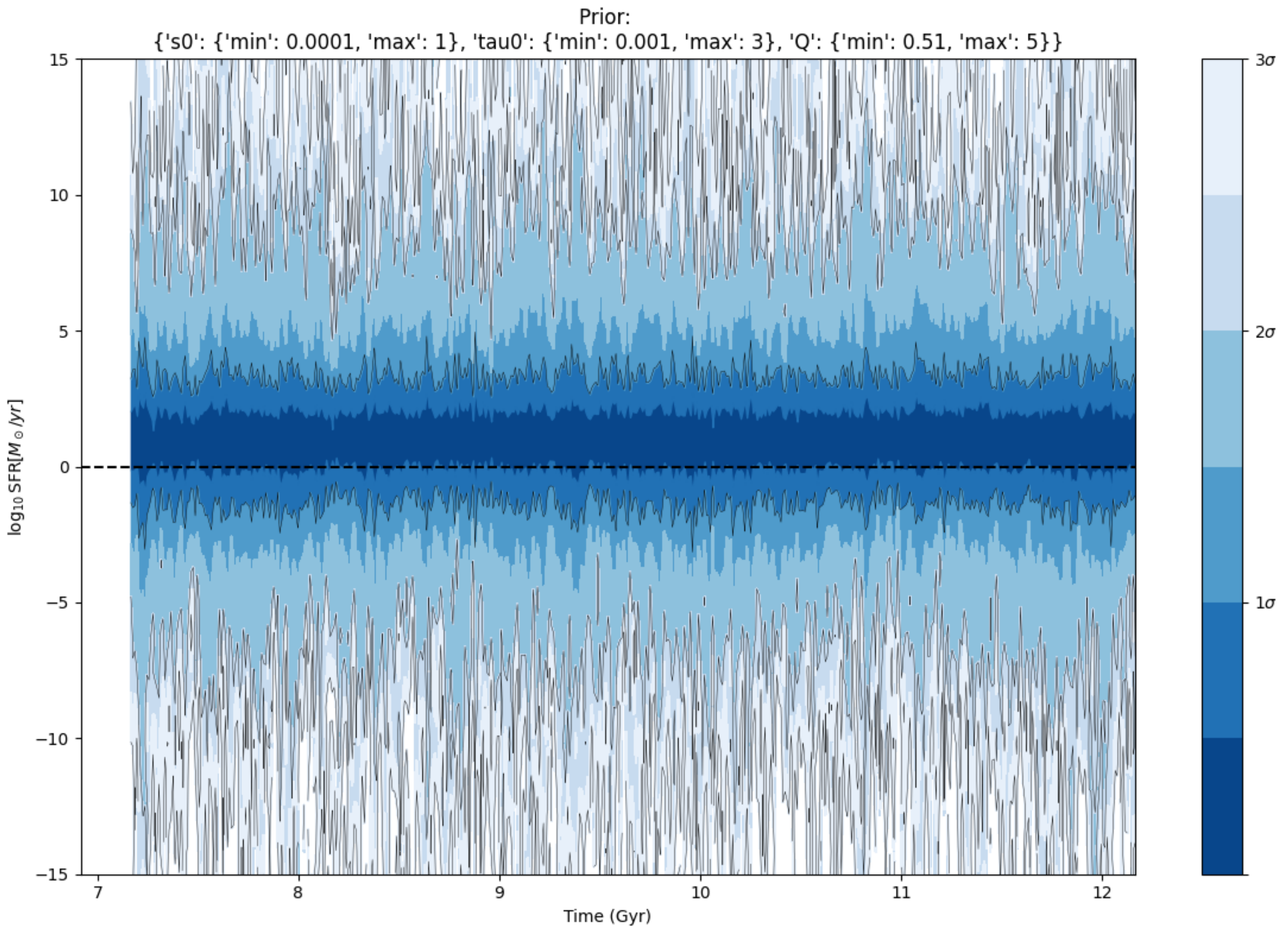




Neural Ratio Estimation



Priors


Prior:
{'s0': {'min': 0.0001, 'max': 1}, 'tau0': {'min': 0.001, 'max': 3}, 'Q': {'min': 0.51, 'max': 5}}
15
10
5
0
−5
−10
−15
$\log_{10}$ SFR$[M_\odot/yr]$
7
8
9
10
11
12
Time (Gyr)
3σ
2σ
1σ