Emulating SPS and how to use it

1. **Neural emulation of SPS models (“Speculator”)**
2. **Hierarchical inference of n(z) under SPS models**

Justin Alsing, Oskar Klein Centre, Stockholm

Hiranya Peiris, Daniel Mortlock, Joel Leja, Boris Leistedt, George Efstathiou, ChangHoon Hahn and others

Why emulate SPS?

- Want to be able to analyse photometry under SPS models (for photo-z, and also galaxy evolution studies)

- SPS is too slow to scale up for LSST => need emulation

arXiv:1911.11778
github.com/justinalsing/speculator

Emulating SPS spectra

1. Generate training set of SPS parameters and spectra
2. PCA decompose the spectra
3. Train neural network to learn PCA coefficients, as a function of the SPS parameters

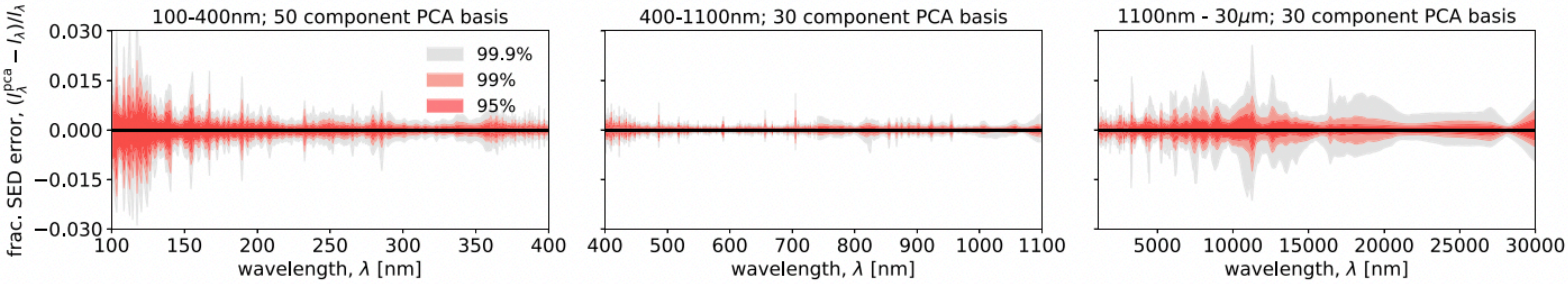
State-of-the-art 15-parameter Prospector-alpha model
Training set of a few million spectra (~ couple CPU days)

Parameter	Description	Prior
M	Total stellar mass formed	Log uniform $[10^7, 10^{12.5}]M_\odot$
$r^1_{\rm SFH},\ldots,r^6_{\rm SFH}$	Ratio of log SFR between adjacent bins	Clipped student's-t: $\sigma = 0.3$, $\nu = 2$, $\lvert r^i_{\rm SFH}\rvert \leqslant 5$
$t_{\rm age}$	Age of universe at the lookback time of galaxy	Uniform [2.6, 13.7] Gyr, $(0 < z < 2.5)$
τ_2	Diffuse dust optical depth	Normal $\mu = 0.3$, $\sigma = 1$, min $= 0$, max $= 4$
τ_1	Birth cloud optical depth	Truncated normal in τ_1/τ_2 $\mu = 1$, $\sigma = 0.3$, min $= 0$, max $= 2$
n	Index of Calzetti et al. (2000) dust attn. curve	Uniform $[-1, 0.4]$
$\ln(Z_{\rm gas}/Z_\odot)$	Gas phase metallicity	Uniform $[-2, 0.5]$
$f_{\rm AGN}$	Fraction of bolometric luminosity from AGN	Log uniform $[10^{-5}, 3]$
$\tau_{\rm AGN}$	Optical depth of AGN torus	Log uniform [5, 150]
$\ln(Z/Z_\odot)$	Stellar metallicity	Truncated normal with μ and σ from Gallazzi et al. (2005) mass–metallicity relation (see Section 4), limits min $= -1.98$, max $= 0.19$
z	Redshift	Uniform [0.5, 2.5]

arXiv:1911.11778
github.com/justinalsing/speculator

Emulating SPS spectra

1. Generate training set of SPS parameters and spectra
2. PCA decompose the spectra
3. Train neural network to learn PCA coefficients, as a function of the SPS parameters

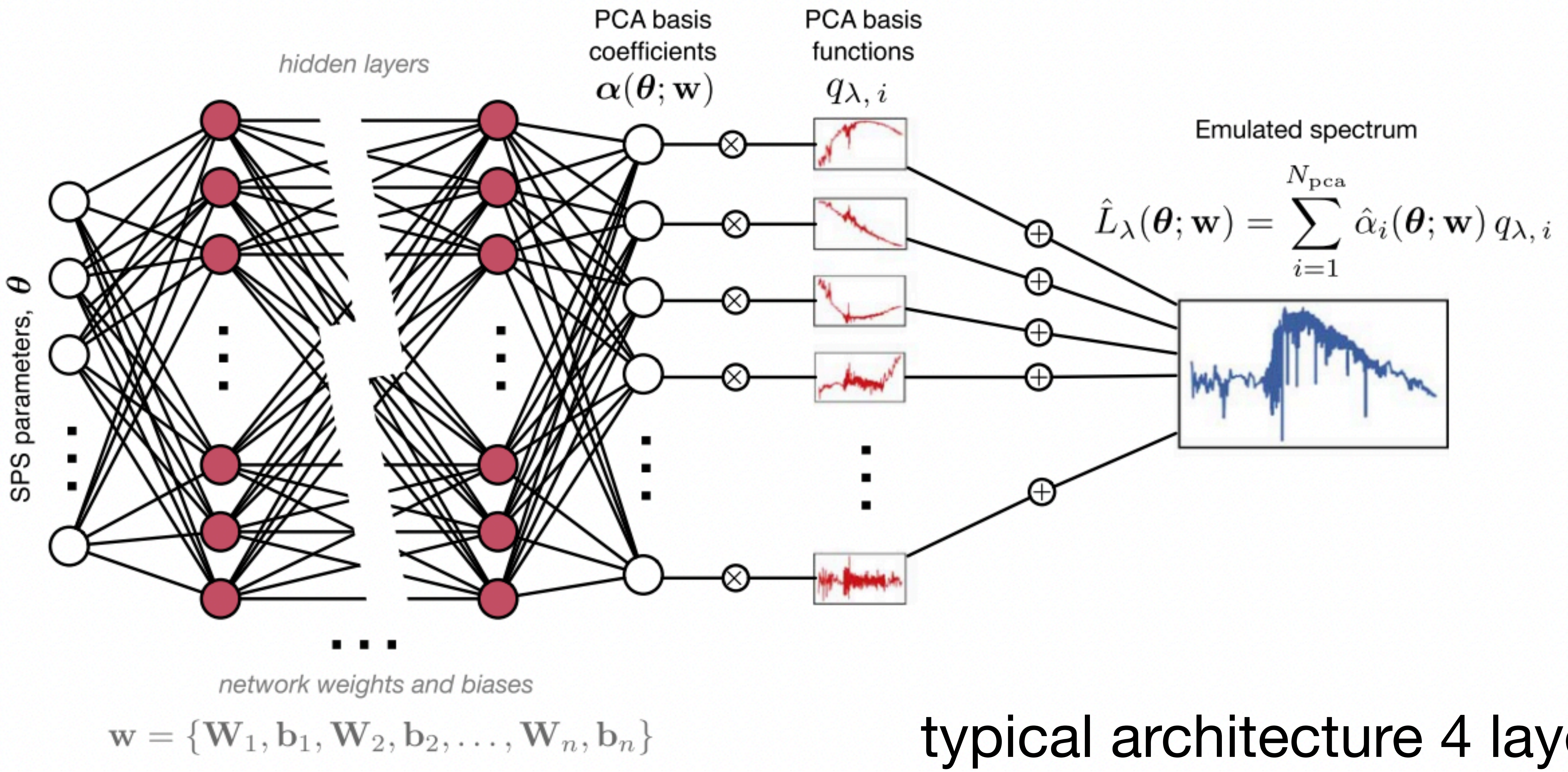


accuracy increases with more components if needed

arXiv:1911.11778
github.com/justinalsing/speculator

Emulating SPS spectra

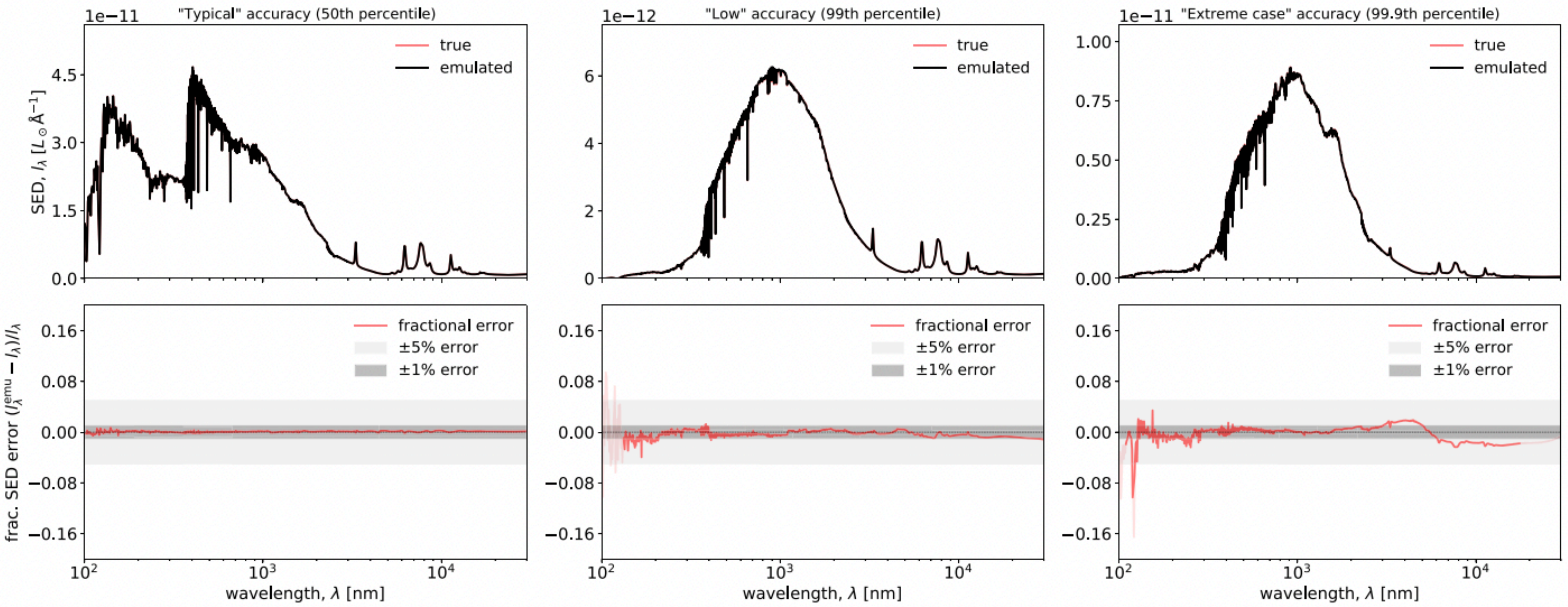
1. Generate training set of SPS parameters and spectra
2. PCA decompose the spectra
3. Train neural network to learn PCA coefficients



typical architecture 4 layers x 256 units

arXiv:1911.11778
github.com/justinalsing/speculator

Emulating SPS spectra

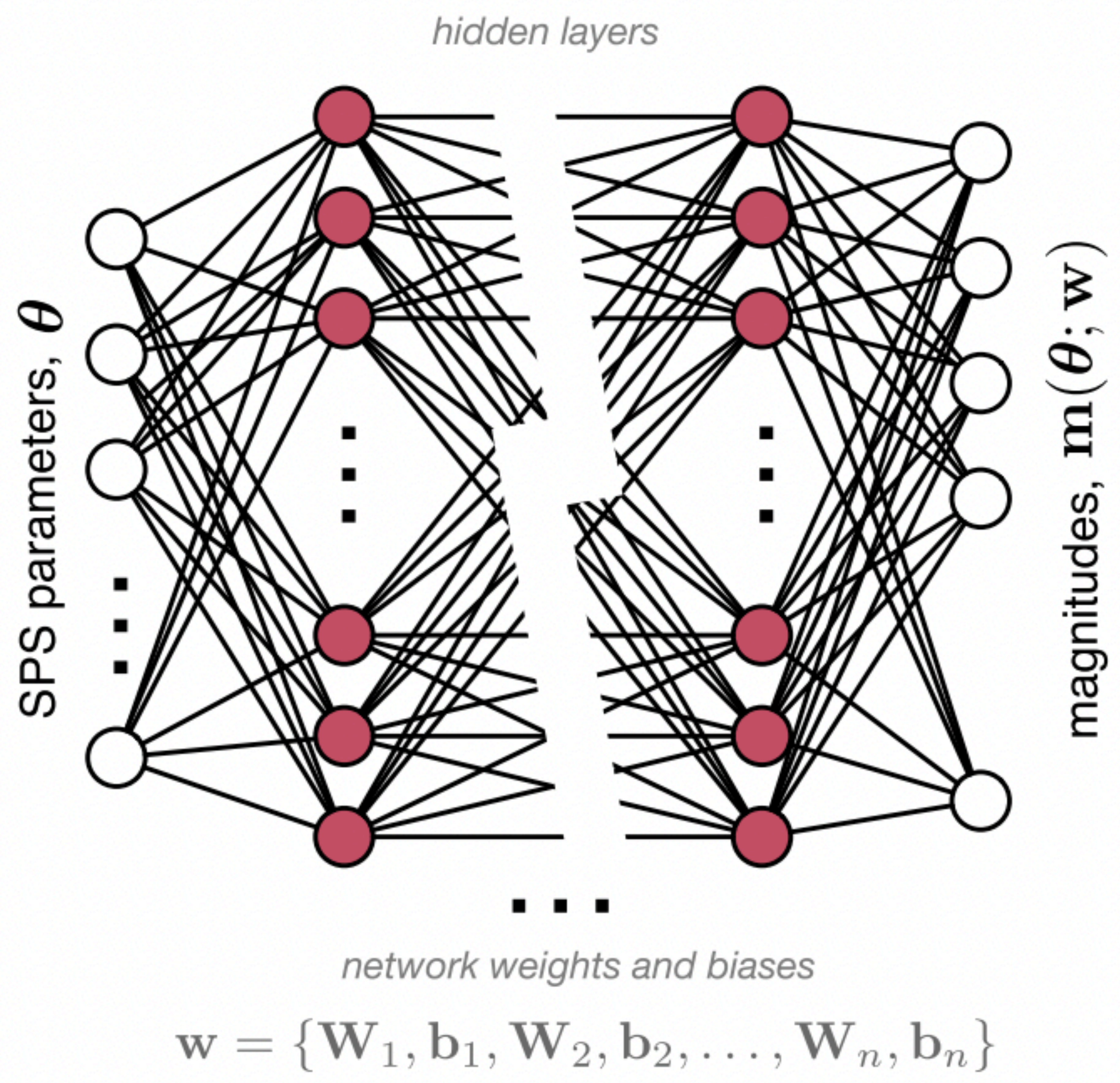


10^4 x speed-up, %-level accuracy, differentiable

arXiv:1911.11778
github.com/justinalsing/speculator

Emulating SPS photometry

1. Generate training set of SPS parameters and photometry
2. Train neural network to learn magnitudes directly



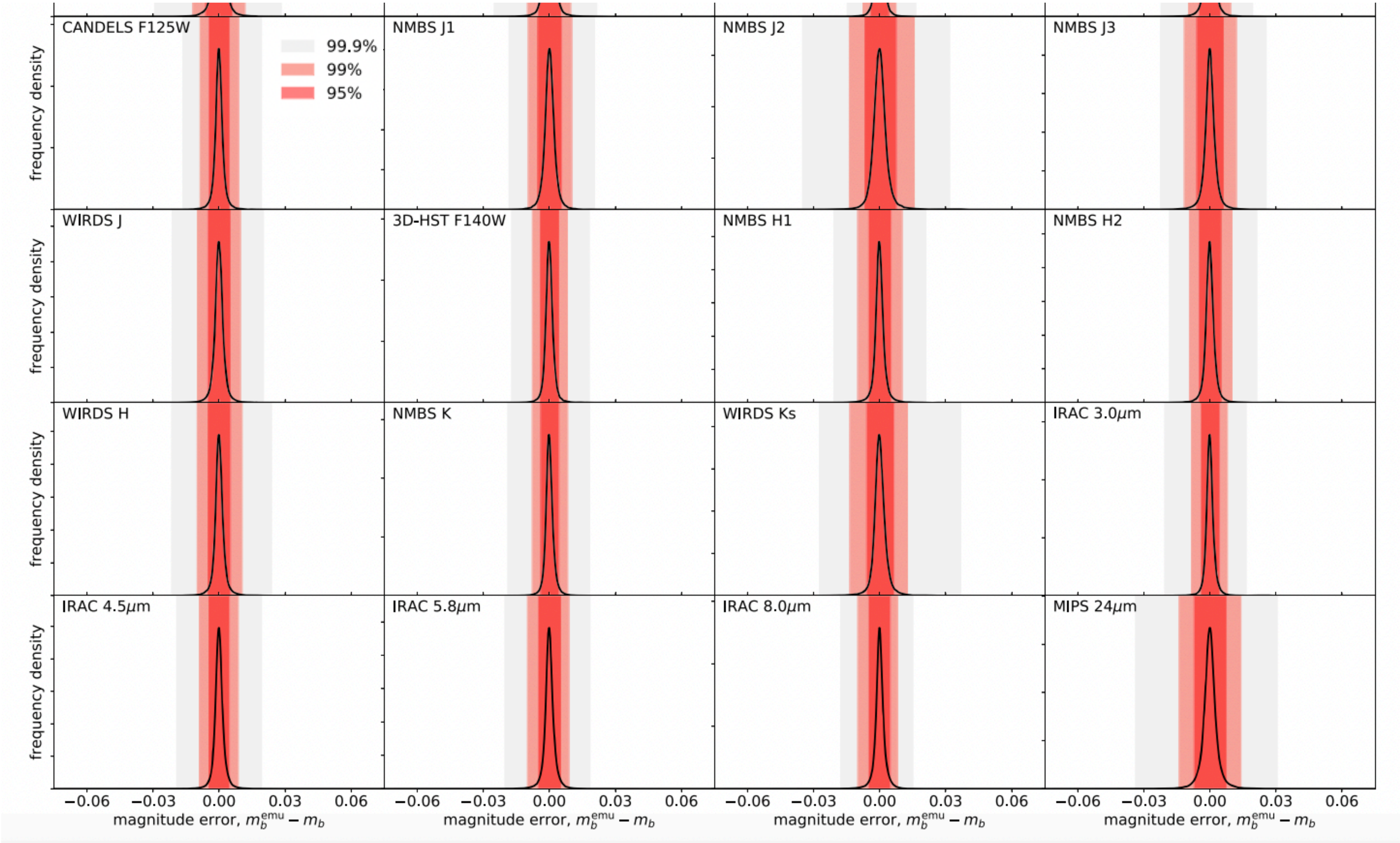
arXiv:1911.11778
github.com/justinalsing/speculator

Emulating SPS photometry

10^4 x speed-up

< %-level accuracy

Differentiable



What to do with SPS emulators

1. Fast SPS-parameter and redshift posteriors (<10 sec per galaxy)
2. Photometric likelihoods for embedding into LSS hierarchical models (eg BORG)
3. Forward modelling and hierarchical inference of cosmological redshift distributions n(z)

Forward modeling for n(z)

1. ***Population model:*** Galaxies properties drawn from population
2. ***SPS model:*** Generate photometry given physical properties
3. ***Data model:*** Add noise, do calibration
4. ***Selection model:*** Apply selection cuts

prior

population model parameters

data model parameters

$$P(\theta_{1:N}, z_{1:N}, \psi, \eta | \hat{\mathbf{f}}_{1:N}) \propto P(\psi, \eta) \times \prod_{i=1}^{N_{\mathrm{galaxies}}} P(\theta_i, z_i | \psi) \times P(\hat{\mathbf{f}}_i | \mathbf{f}(\theta_i, z_i), \eta) \times [P(S_i | \psi, \eta)]^{-1}$$

joint posterior

observed fluxes

selection integral

SPS parameters, redshift

model fluxes

Forward modeling for n(z)

Where does n(z) fit into all this?

chain rule, expand, etc

$$n(z) \equiv P(z|S)$$

$$= \frac{1}{P(S)} \int \left[\iint P(S|\hat{\mathbf{f}}, \theta, z) P(\hat{\mathbf{f}}|\theta, z, \sigma) P(\sigma) d\hat{\mathbf{f}} d\sigma \right] P(\theta, z) d\theta$$

n(z) *is* just an integral over selection x data model x population model

prior

population model parameters

data model parameters

$$P(\theta_{1:N}, z_{1:N}, \psi, \eta | \hat{\mathbf{f}}_{1:N}) \propto P(\psi, \eta) \times \prod_{i=1}^{N_{\mathrm{galaxies}}} P(\theta_i, z_i|\psi) \times P(\hat{\mathbf{f}}_i | \mathbf{f}(\theta_i, z_i), \eta) \times [P(S_i|\psi, \eta)]^{-1}$$

joint posterior

observed fluxes

selection integral

SPS parameters, redshift

model fluxes

Forward modeling for n(z)

Advantages

- *Does not rely on spec-z calibration*
- *Auxiliary data (spec-z, extra surveys) can be included seamlessly (extended data vector or extra priors for objects with extra information)*
- *"Turns photo-z back into a physics problem", synergies with galaxy-evo community*

prior

population model parameters

data model parameters

$$P(\theta_{1:N}, z_{1:N}, \psi, \eta|\hat{\mathbf{f}}_{1:N}) \propto P(\psi, \eta) \times \prod_{i=1}^{N_{\mathrm{galaxies}}} P(\theta_i, z_i|\psi) \times P(\hat{\mathbf{f}}_i|\mathbf{f}(\theta_i, z_i), \eta) \times [P(S_i|\psi, \eta)]^{-1}$$

joint posterior

observed fluxes

selection integral

SPS parameters, redshift

model fluxes

Forward modeling for n(z)

Research questions

1. *Can we forward model well enough to infer n(z) with high-fidelity?*
2. *If so, can we actually “do” the inference under this model?*

prior

population model parameters

data model parameters

$$P(\theta_{1:N}, z_{1:N}, \psi, \eta | \hat{\mathbf{f}}_{1:N}) \propto P(\psi, \eta) \times \prod_{i=1}^{N_{\mathrm{galaxies}}} P(\theta_i, z_i | \psi) \times P(\hat{\mathbf{f}}_i | \mathbf{f}(\theta_i, z_i), \eta) \times [P(S_i | \psi, \eta)]^{-1}$$

joint posterior

observed fluxes

selection integral

SPS parameters, redshift

model fluxes

Forward model

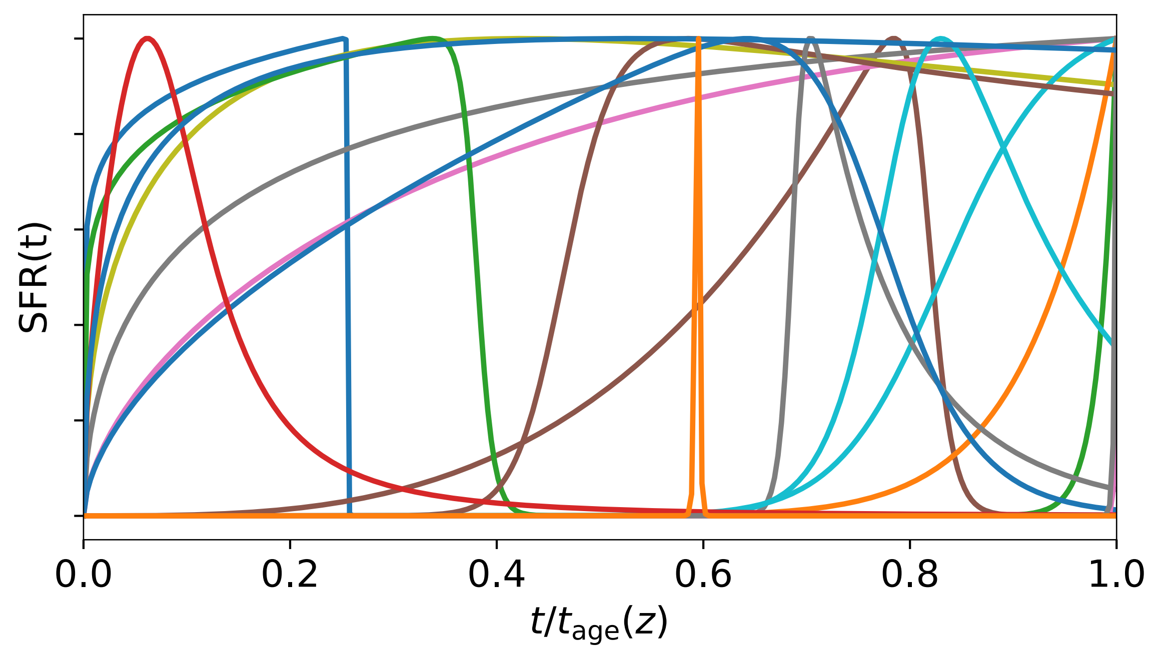
9-parameter SPS model

Mass, redshift

Gas phase metallicity (stellar metallicity history follows mass buildup)

$$\theta, z = (M, z, \log Z, \boldsymbol{\theta}_{\mathrm{SFH}}, \tau_1, \tau_2, n_{\mathrm{dust}})$$

Double power-law SFH



Dust (ISM and birth cloud screens), attenuation law index

Forward model

Population model

Double Schechter function (Leja & Speagle 2021)
11 parameters, tight priors

Star-forming sequence (Leja & Speagle 2021)
Normalizing flow, 34 parameters, modest priors

$$
\begin{aligned}
P(\theta, z) = & P(M, z) \\
& P(\mathrm{SFR}(\boldsymbol{\theta}_{\mathrm{SFH}}) | M, z) \\
& P(\log Z | \mathrm{SFR}(\boldsymbol{\theta}_{\mathrm{SFH}}), M) \\
& P(\tau_1 | \mathrm{SFR}(\boldsymbol{\theta}_{\mathrm{SFH}}), M) \\
& P(\tau_2 | \tau_1) \\
& P(n_{\mathrm{dust}} | \tau_1)
\end{aligned}
$$

Fundamental metallicity relation (Curti+ 2020)
5 parameters, modest priors

Birth cloud ~ diffuse dust (Leja+ 2020)
fixed, broad

Mass-dependent dust-SFR relation (similar to Tanaka 2015)
bilinear+scatter, 5 parameters, broad priors

Dust index ~ optical depth
Quadratic relation + scatter: 4 parameters, modest priors

Forward model

Data model

SPS modeling error
1 parameter per band

Emulator error
fixed (measured from emulator validation)

$$\hat{f} = \alpha_{\rm ZP}\, f(\theta, z) + n_{\rm noise} + n_{\rm SPS} + n_{\rm emission-line} + n_{\rm emulator}$$

Zero points
1 parameter per band

Emission-line modelling error
2 parameters (mean and variance of error) per emission-line

Measurement errors
Students-t, no free parameters

Forward model

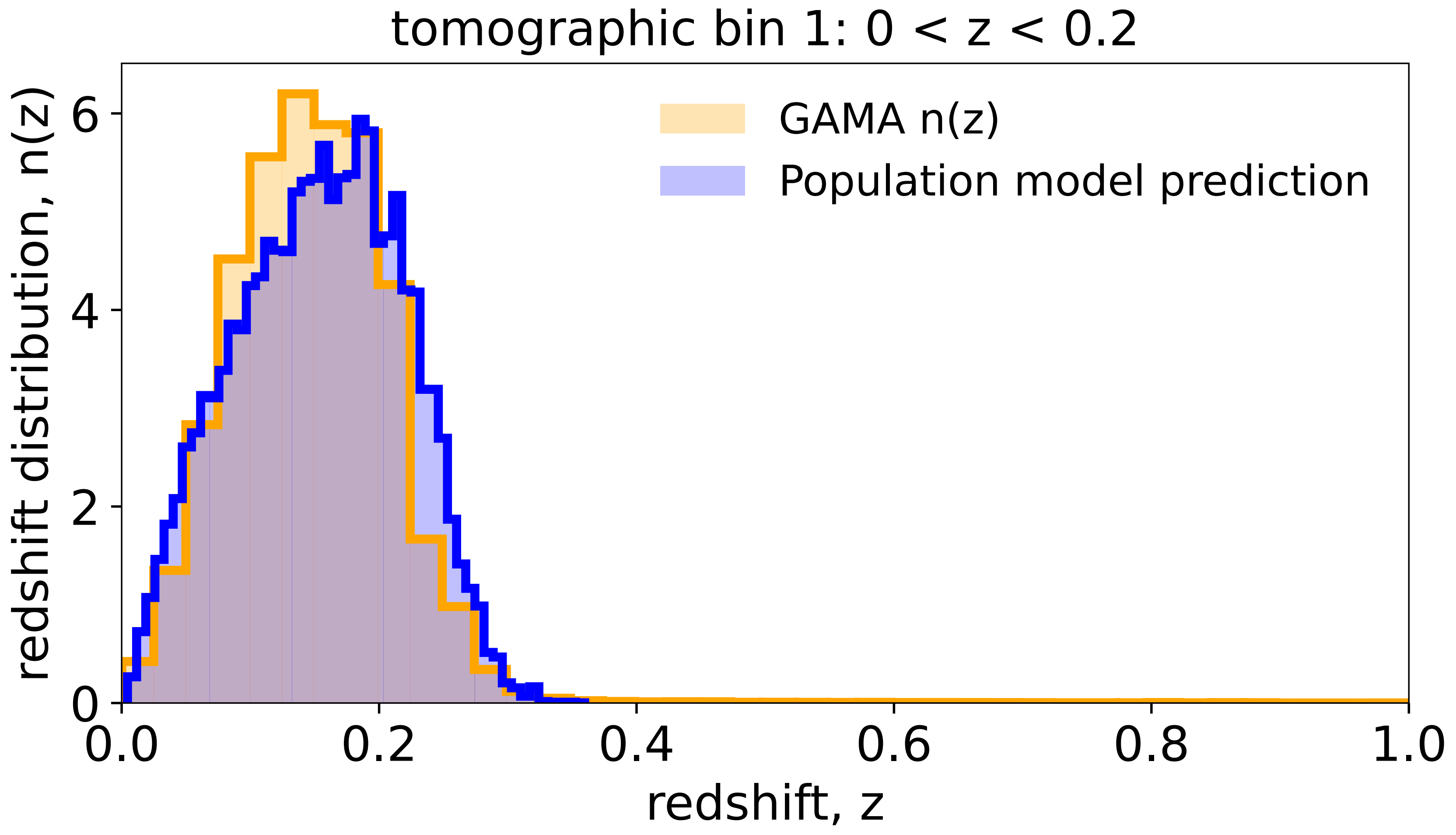
Selection model

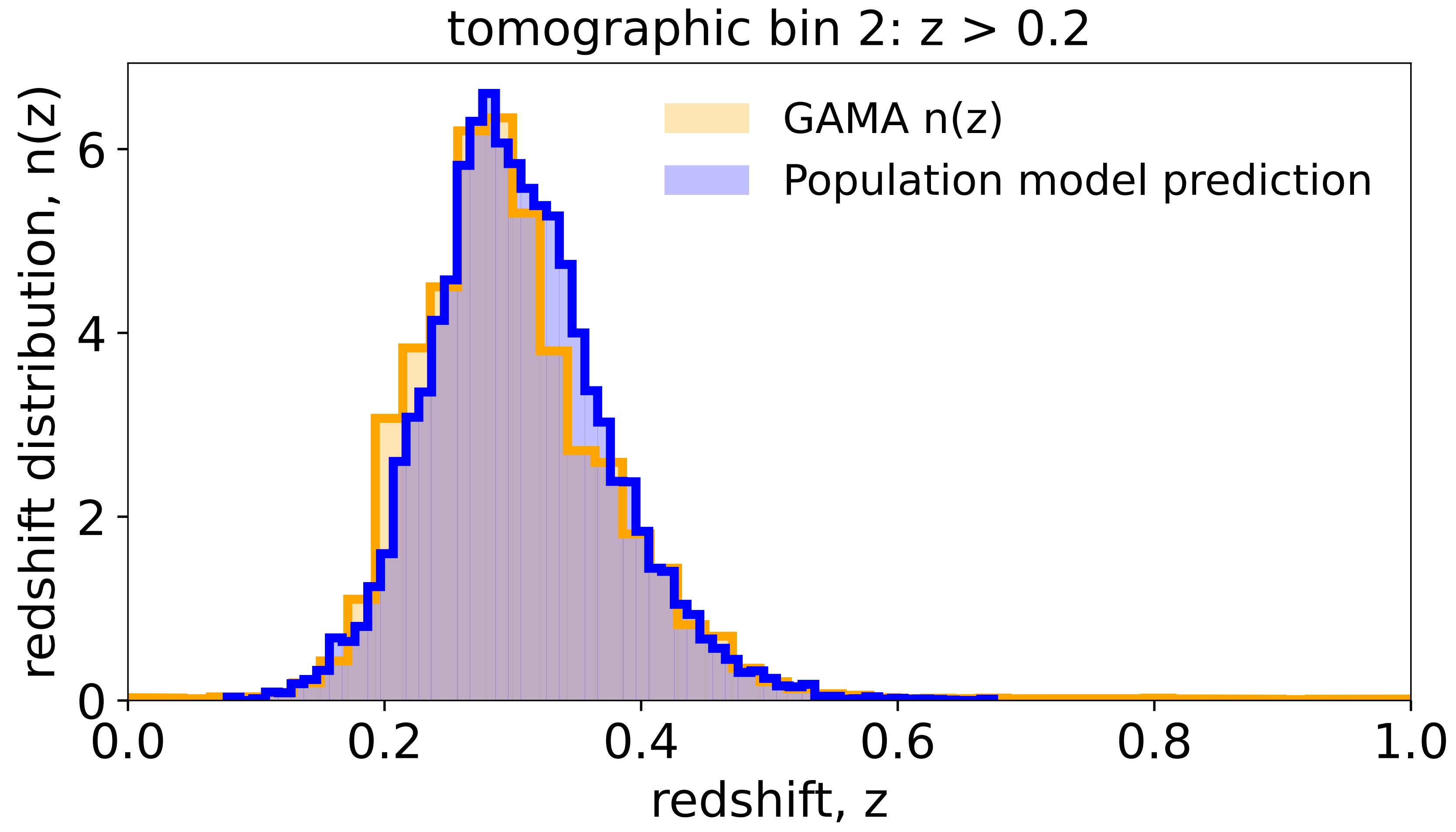
Two spec-z surveys with straightforward selection cuts (great for validation)

1. GAMA (ugriZYJHKs): $r < 19.65$, *(J-Ks)* > 0.025
2. VVDS (UBVRI): $I < 22.5$, *SG separation done at level of spectra*

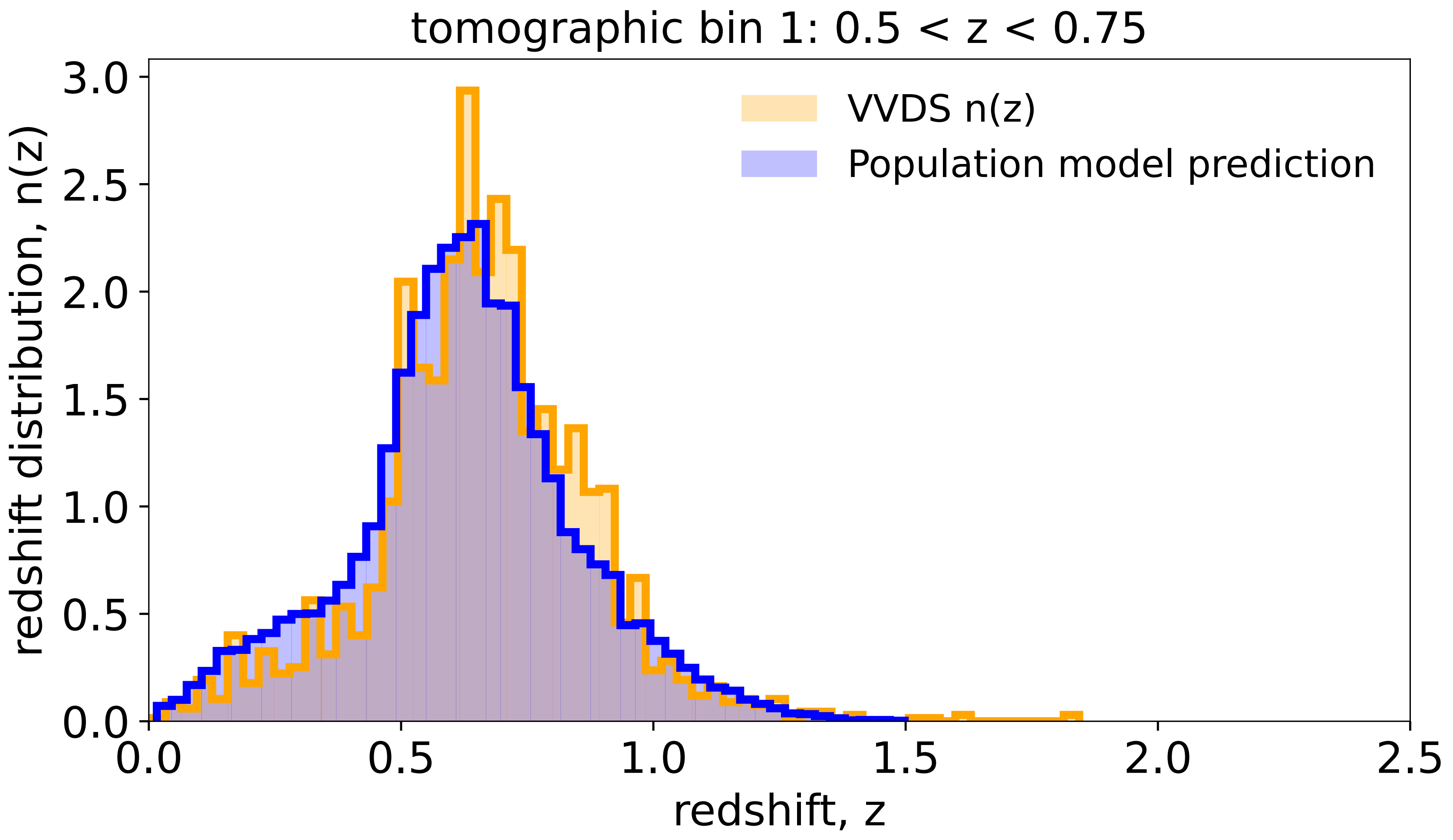
How good is the baseline model?

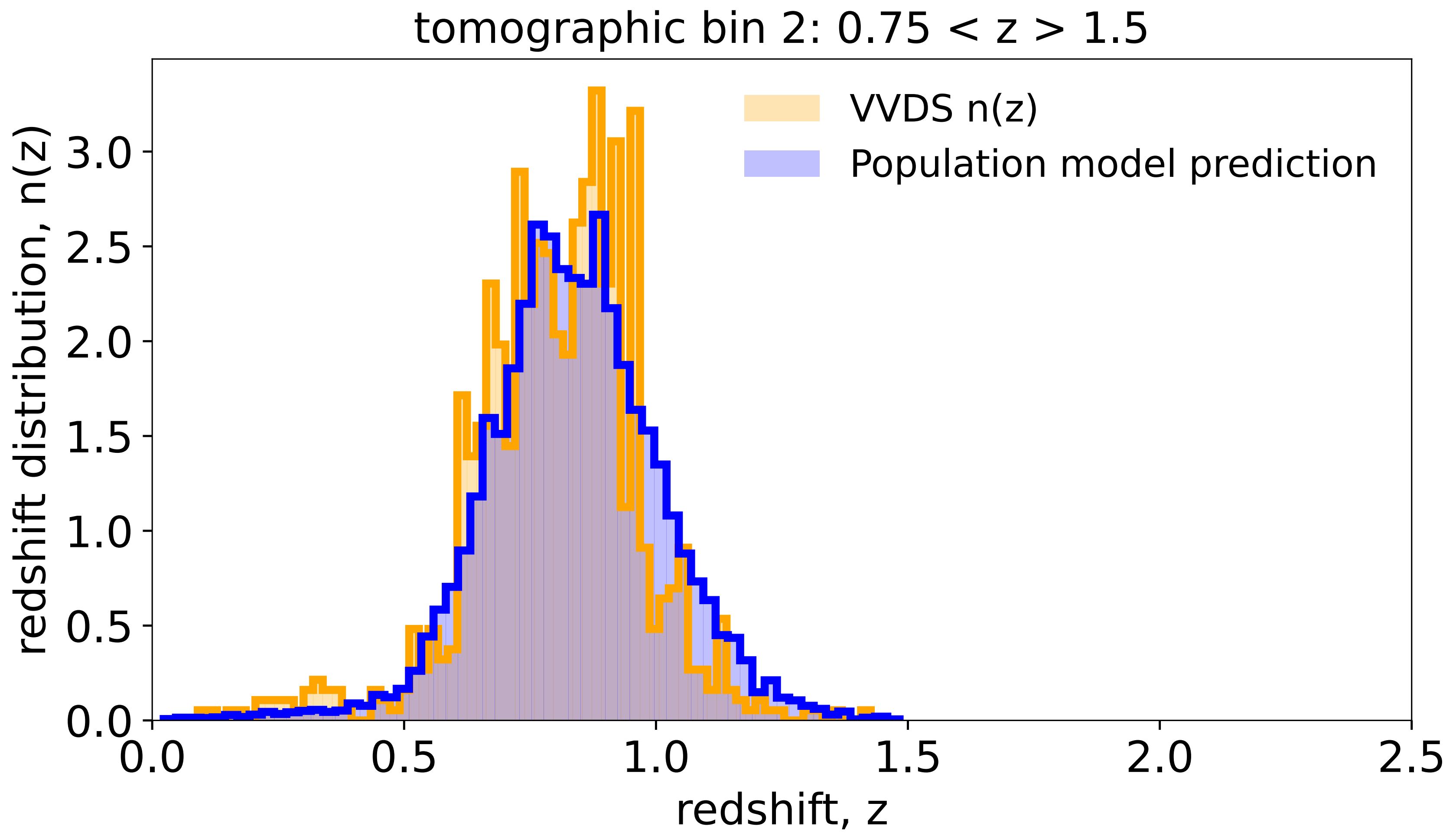
GAMA n(z)
tomography
population model X
data model X
selection





VVDS n(z)
tomography
population model X
data model X
selection





Baseline model bias < 0.03 on n(z) *<u>before parameter inference</u>* *(no data!)*

Doing the inference

Method 1: Hierarchical sampling

$$P(\theta_{1:N}, z_{1:N}, \psi, \eta | \hat{\mathbf{f}}_{1:N}) \propto P(\psi, \eta) \times \prod_{i=1}^{N_{\mathrm{galaxies}}} P(\theta_i, z_i | \psi) \times P(\hat{\mathbf{f}}_i | \mathbf{f}(\theta_i, z_i), \eta) \times [P(S_i | \psi, \eta)]^{-1}$$

Emulating the selection term:

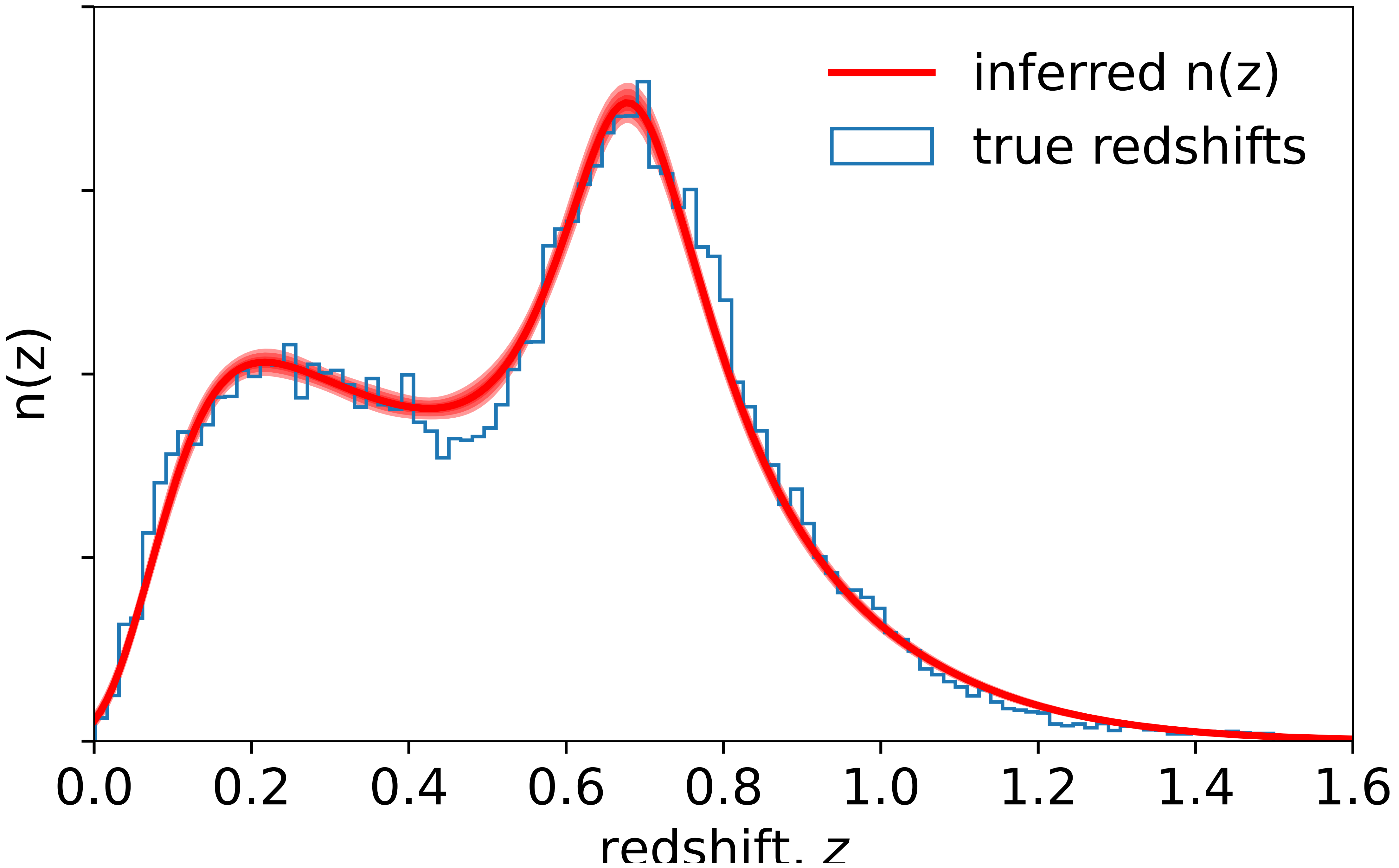
1. Generate mock galaxies (drawing hyper parameters from prior) to give training set {S, hyper-parameters}
2. Train a neural classifier to learn P(S | hyper-parameters)
3. Embed into hierarchical model and sample

NB: For selection effect connoisseurs, the Poisson term has already been integrated out here

Doing the inference

Method 1: Hierarchical sampling

This works: validation test 200k galaxies with complex selection, small sub-set of hyper-parameters (5 hyper-parameters)



Scaling up to full hyper-parameter set may be challenging

Doing the inference

Method 2: Likelihood-free inference

Forward simulate mock catalogues (including selection)
Compress to handful of summary statistics
Use those summaries as basis for density-estimation LFI (Delfi) or similar

Advantages

- Avoids explicitly calculating the selection integrals
- Bypasses latent-parameter sampling (can do hyper-parameters only)
- Easy to increase model complexity (eg could include blending explicitly)

work-in-progress…