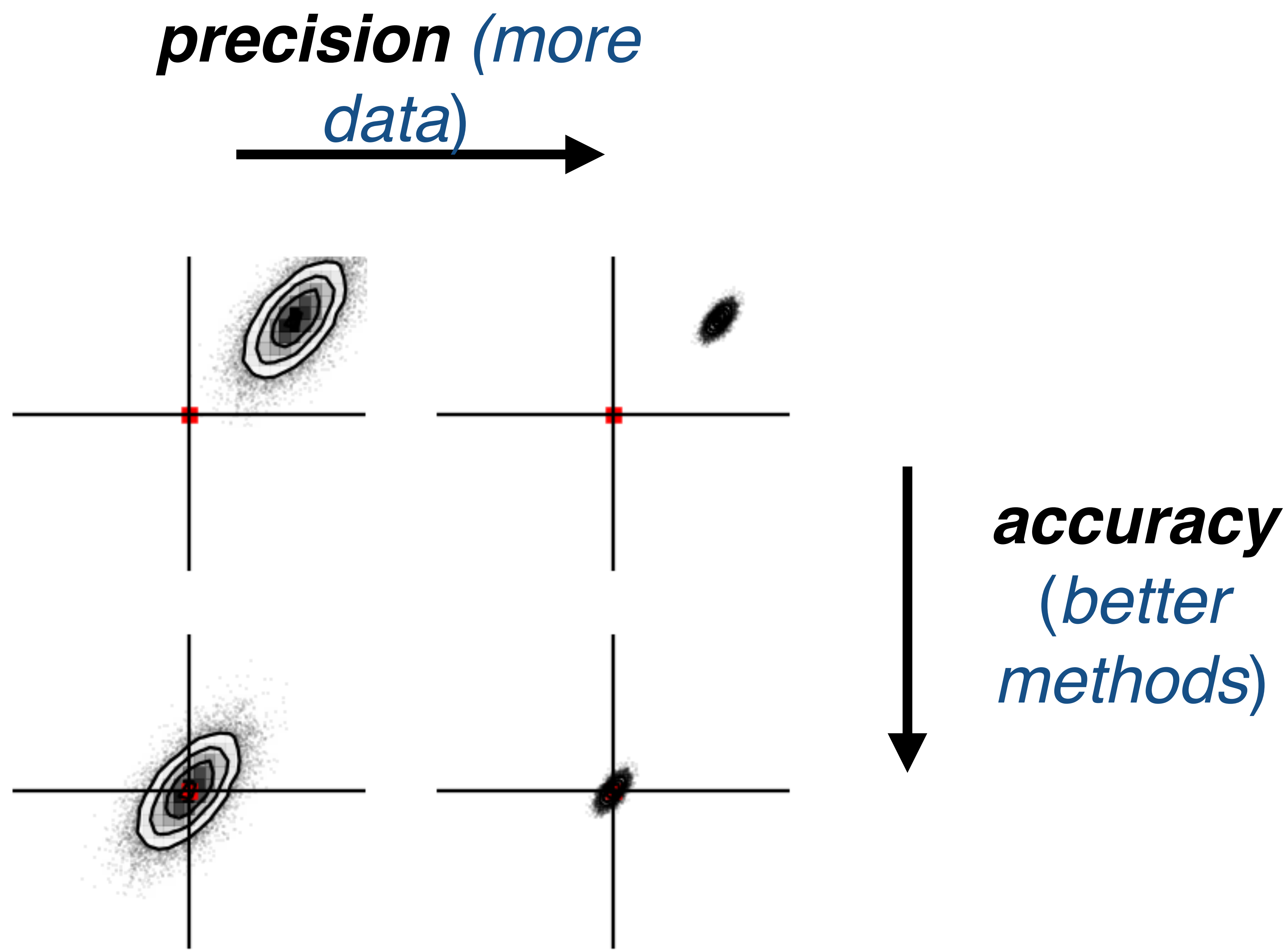
Forward, causal modeling of galaxy photometry

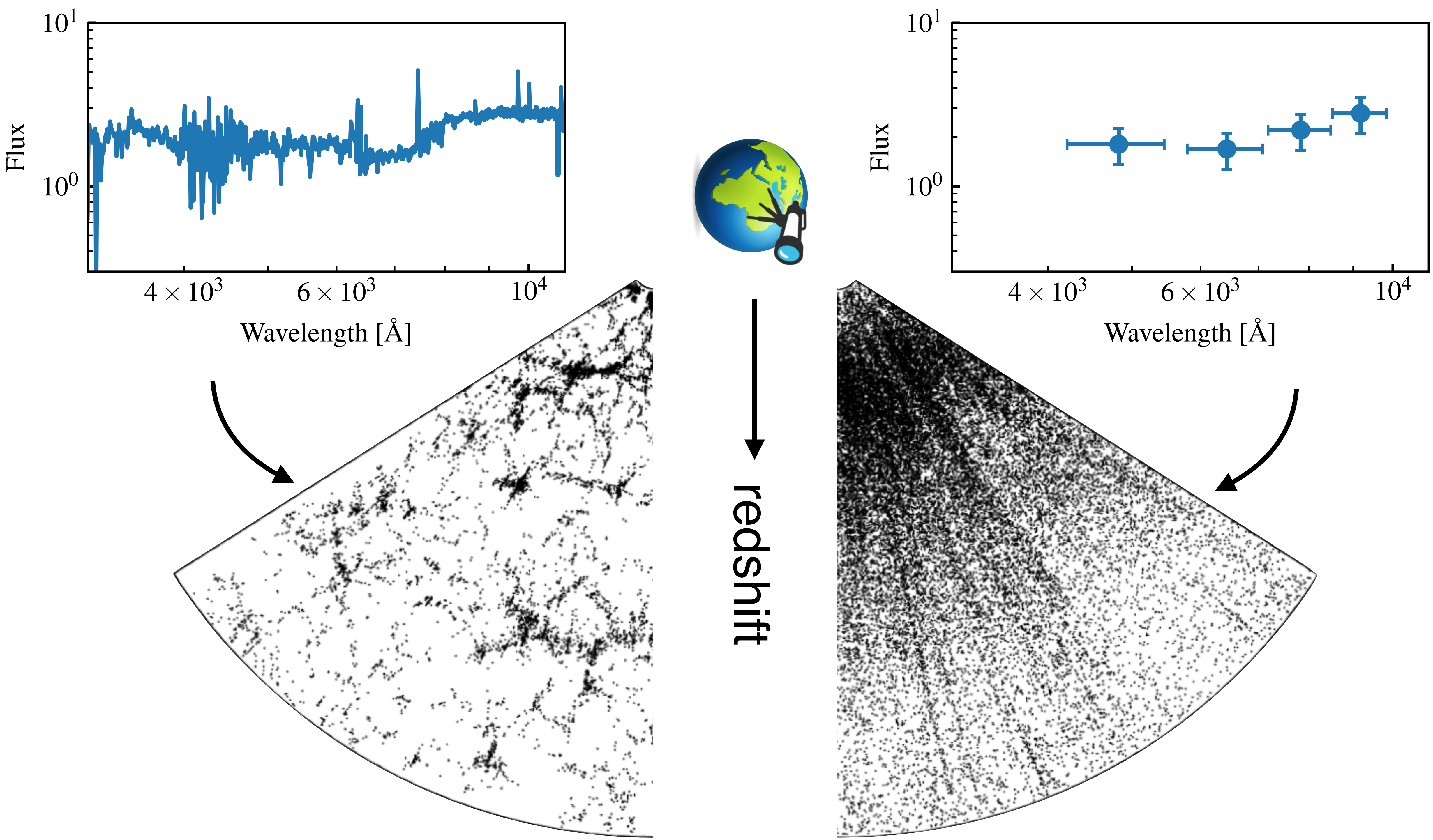
Joint self-calibration of SEDs and broadband photometry

NYU

Boris Leistedt — @ixkael, www.ixkael.com
NASA Einstein Fellow, New York University

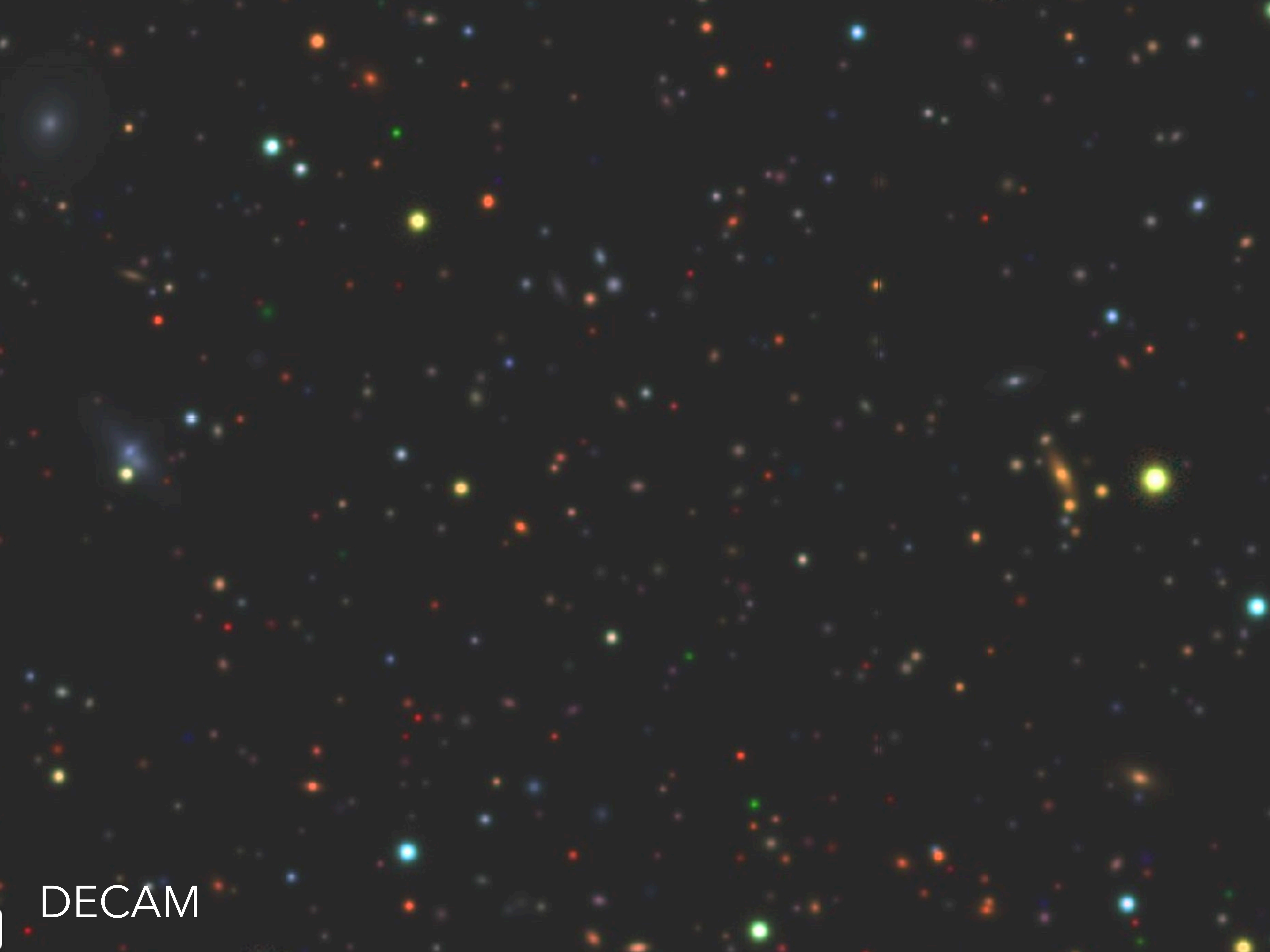

precision (more data)
accuracy (better methods)

Spectroscopic vs. photometric surveys

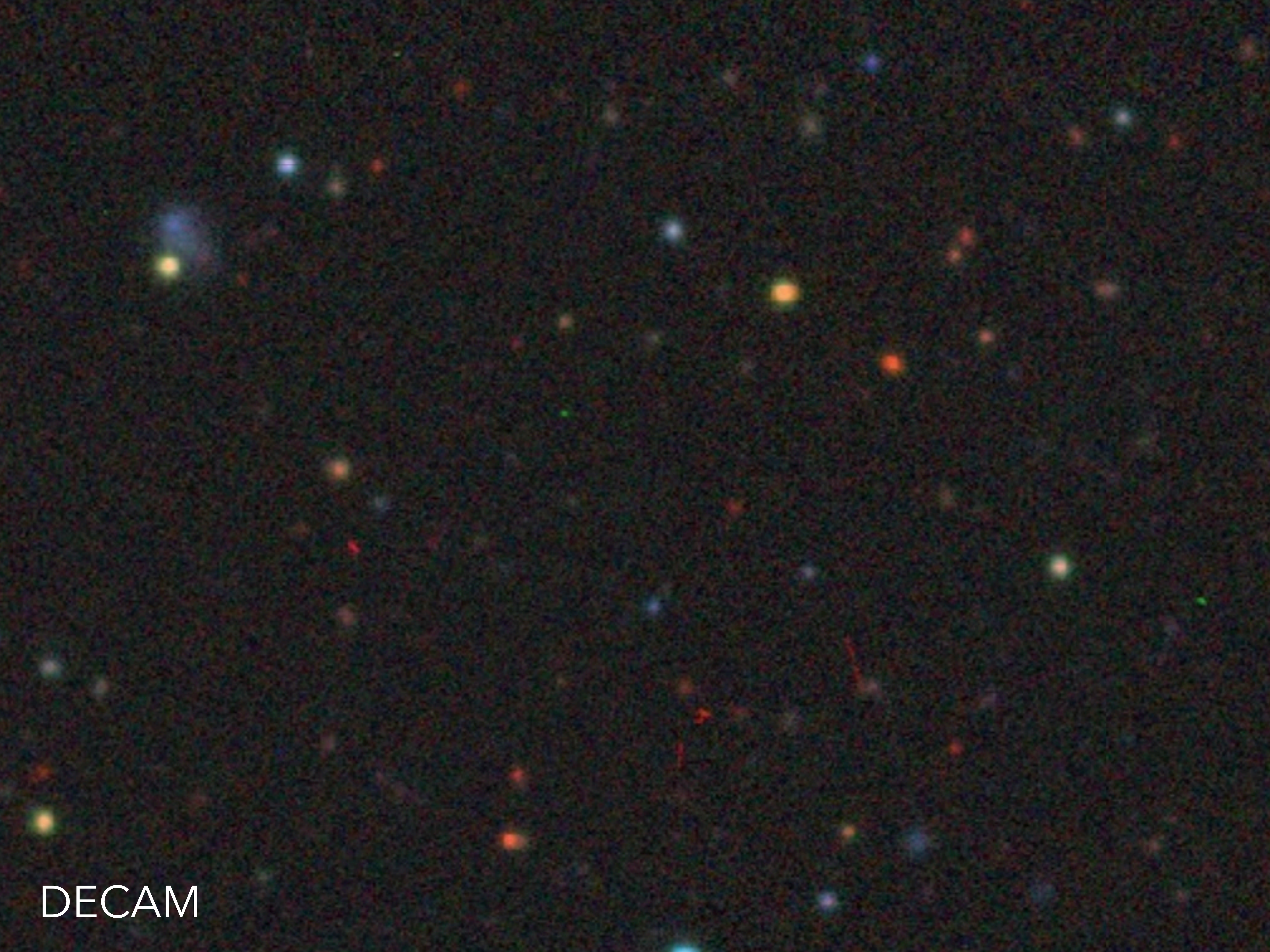



SDSS

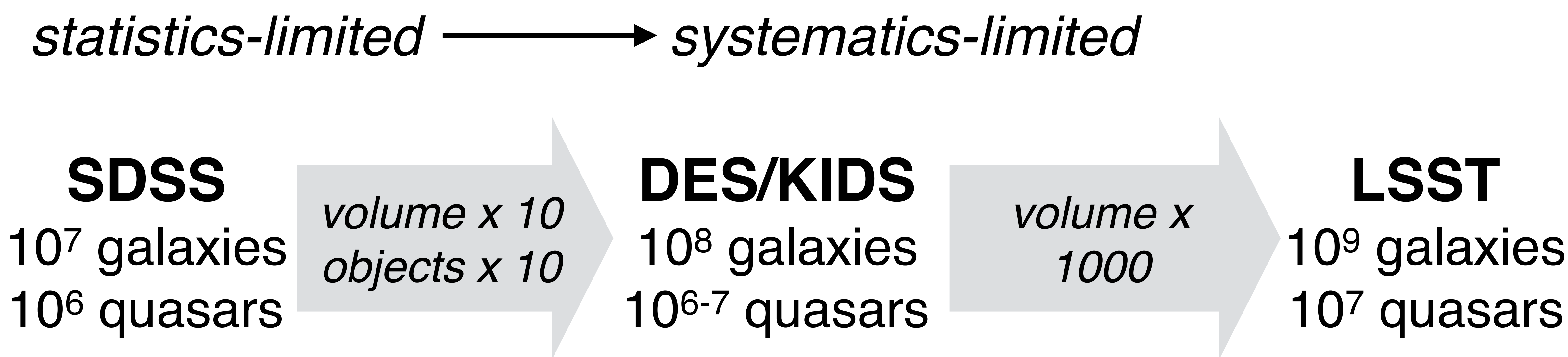

DECAM


DECAM


SDSS


DECAM


DECAM



statistics-limited
systematics-limited
SDSS
10^7 galaxies
10^6 quasars
volume x 10
objects x 10
DES/KIDS
10^8 galaxies
10^6-7 quasars
volume x
1000
LSST
10^9 galaxies
10^7 quasars

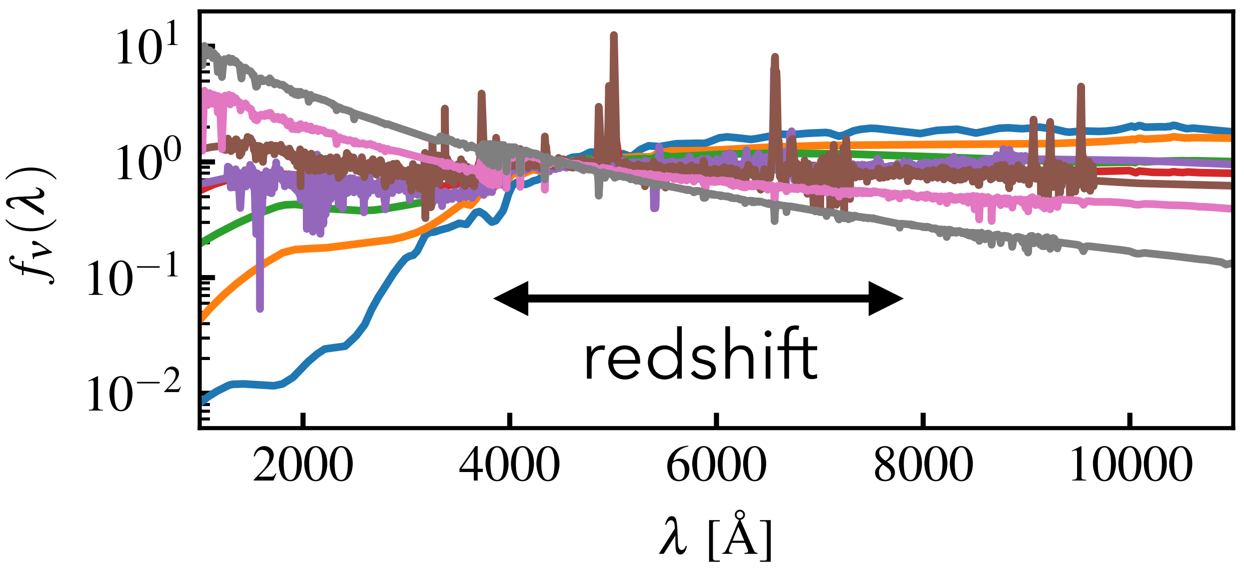
Challenges of photometric surveys

- Flux and shear measurements
- Modeling intrinsic alignments
- Modeling small scales baryonic physics
- Form of covariances, likelihoods
- ***Photometric redshifts, redshift distributions***
- ***Image artefacts, blending***
- ***Simulations of realistic galaxies and photometry***

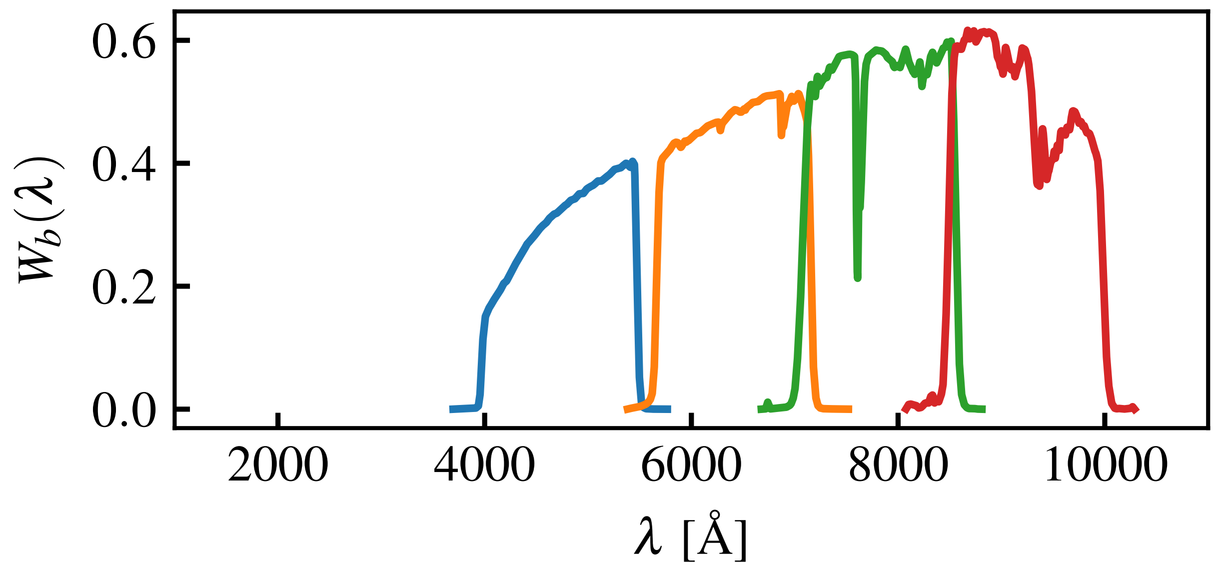
Photometric redshift

= estimating redshift from noisy broadband photometry

Galaxy spectra



DECAM photometric band passes



using knowledge of observed or synthetic SEDs, bandpasses, etc

Three classes of methods

template fitting

Fitting SEDs to photometry using **likelihood function**

Requires calibrated SEDs/priors & unbiased data

machine learning

Construct **flexible model** from spectroscopic training

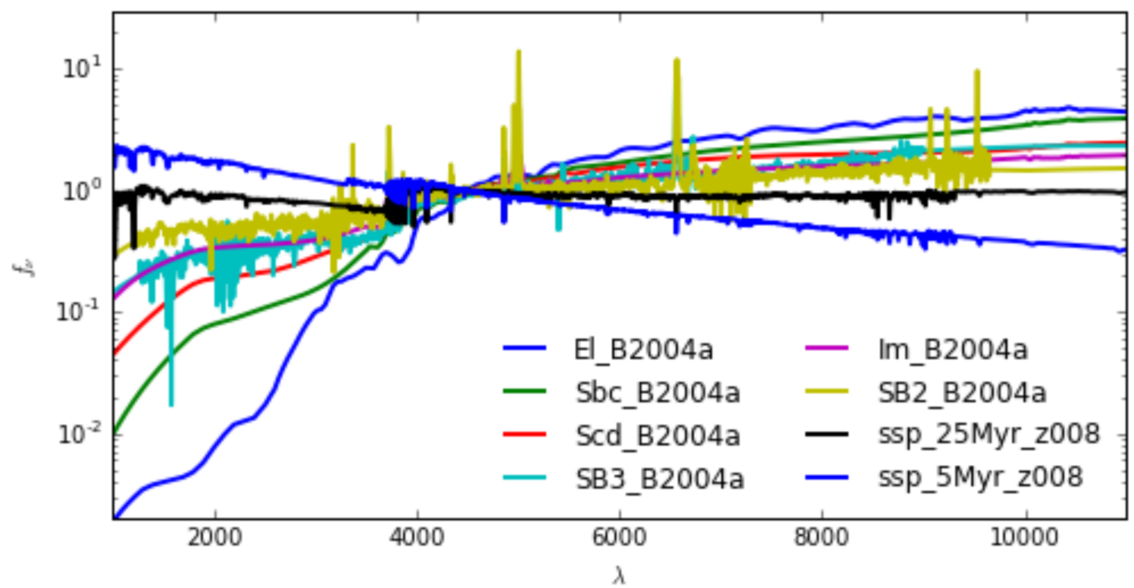
No likelihood, built-in prior, needs representative data

clustering redshifts

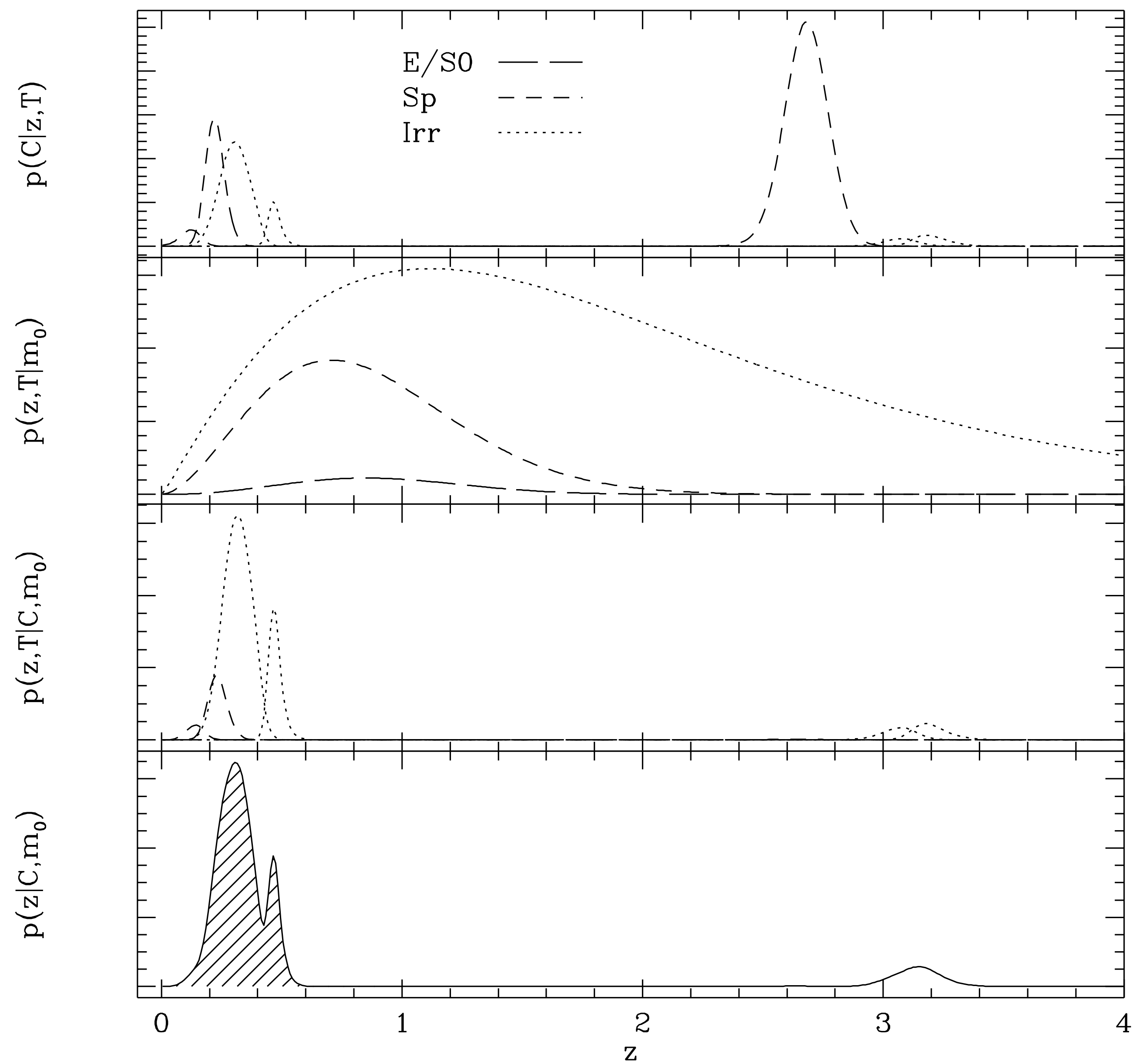
Constrain N(z) using spatial **cross-correlations**

Requires overlapping samples, bias model

Template fitting in cosmology



BPZ code applied to 5-band photometry



Benitez et al, arXiv:9811189

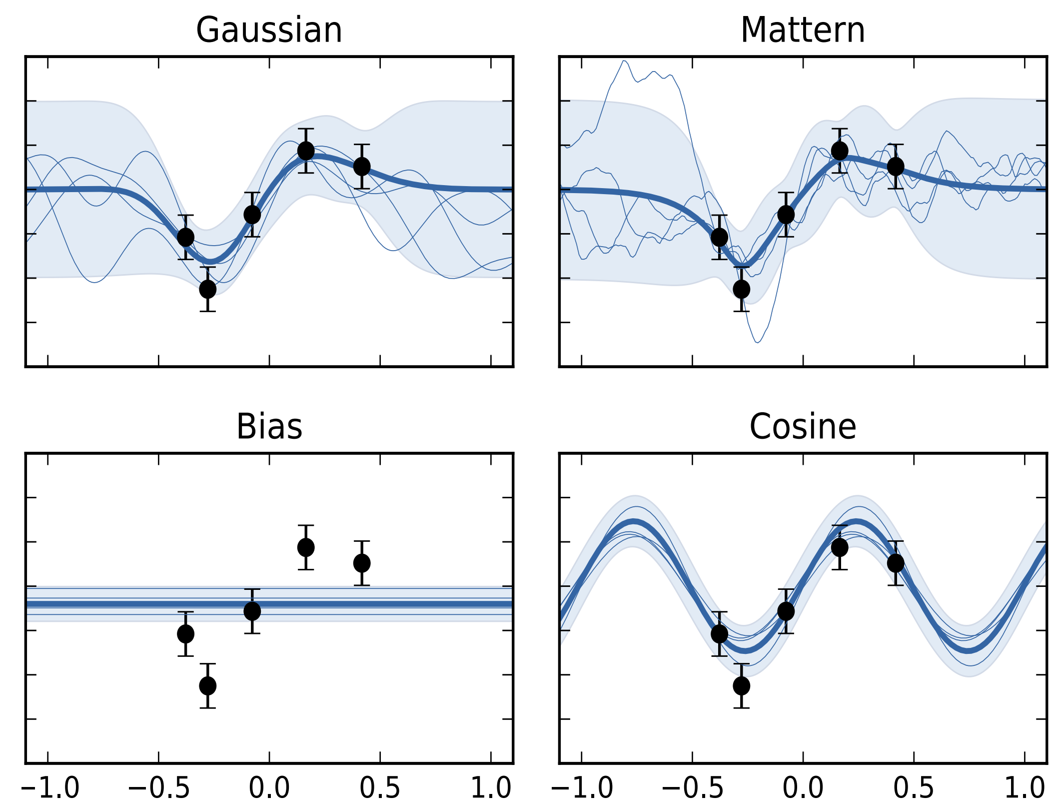
Trust photometry and recalibrate SEDs/priors?

Trust SED model and recalibrate photometry?

How many templates? Form of priors?
What about spatially-varying photometry?
Unrepresentative spectroscopic testing data?

Machine learning in physics 1

- ML absorbs data complexity
- But time/training wasted on learning known physics
- ***Encode physics in ML*** to *generalize/extrapolate* outside of training data



- ML forced to satisfy physics of redshift to improve robustness to (un)representative training (see BL & Hogg, 1703.08112)
But less robust to data complexity… => hierarchical modeling

Machine learning in physics 2

Emulation: speed up simulations or function evaluations (no representativeness issue)

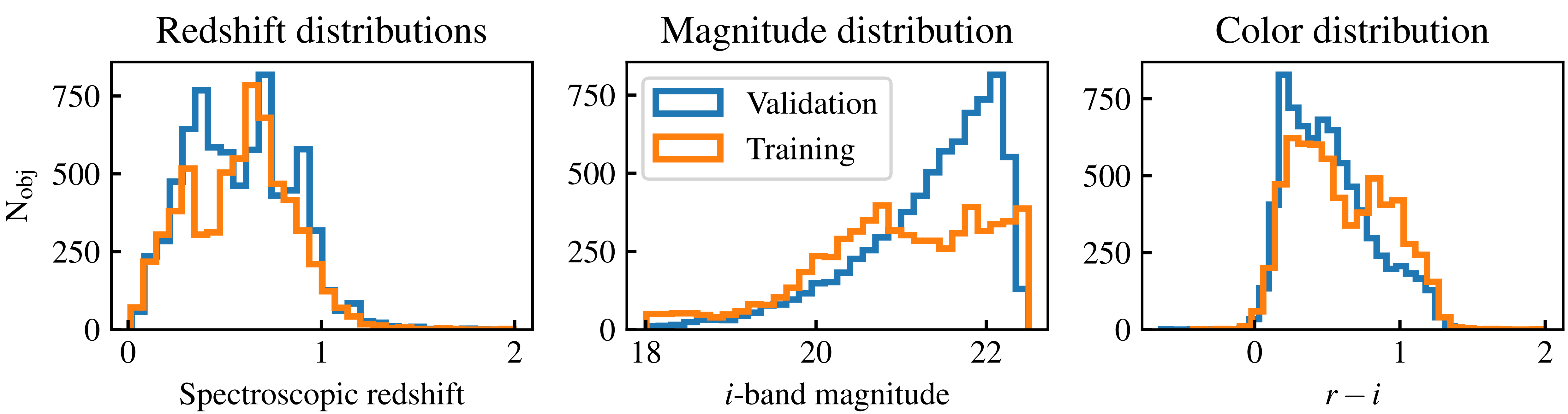
Parts of a model I don't care about or have no intuition for (unknown functional form and no need to extrapolate).

Example of hierarchical SED modelling with embedded machine learning

BL, Hogg, Wechsler, DeRose (arXiv:1807.0139)

DES SV & photo-z's (Bonnett + 2015)

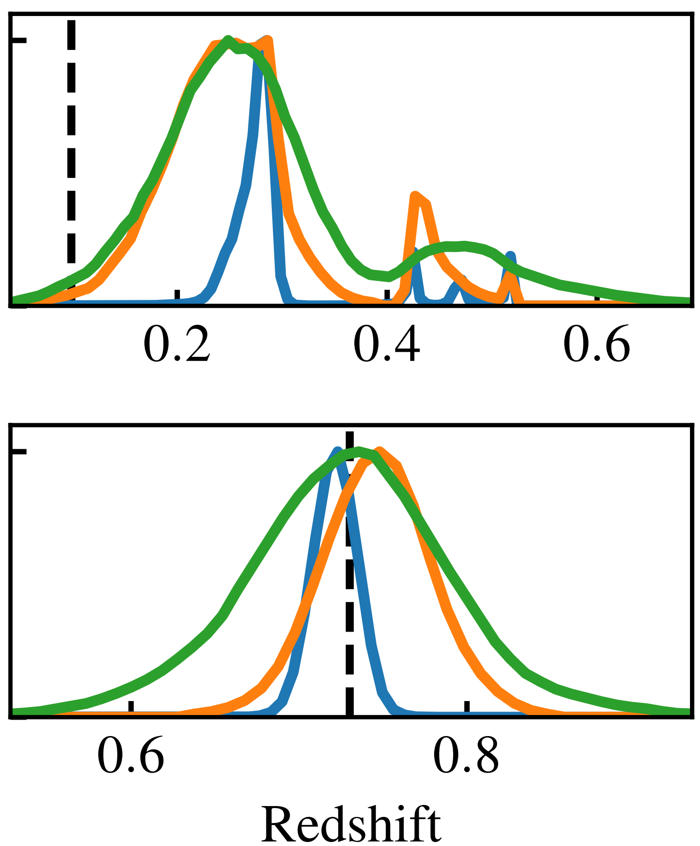
- Full SV data: 20+ million objects
- Gold sample: 18 < *i* magnitude < 22.5
- Training: VVDS, VIPERS, OzDES, ACES, 8k objects
- Validation: zCOSMOS, 8k objects



Criteria

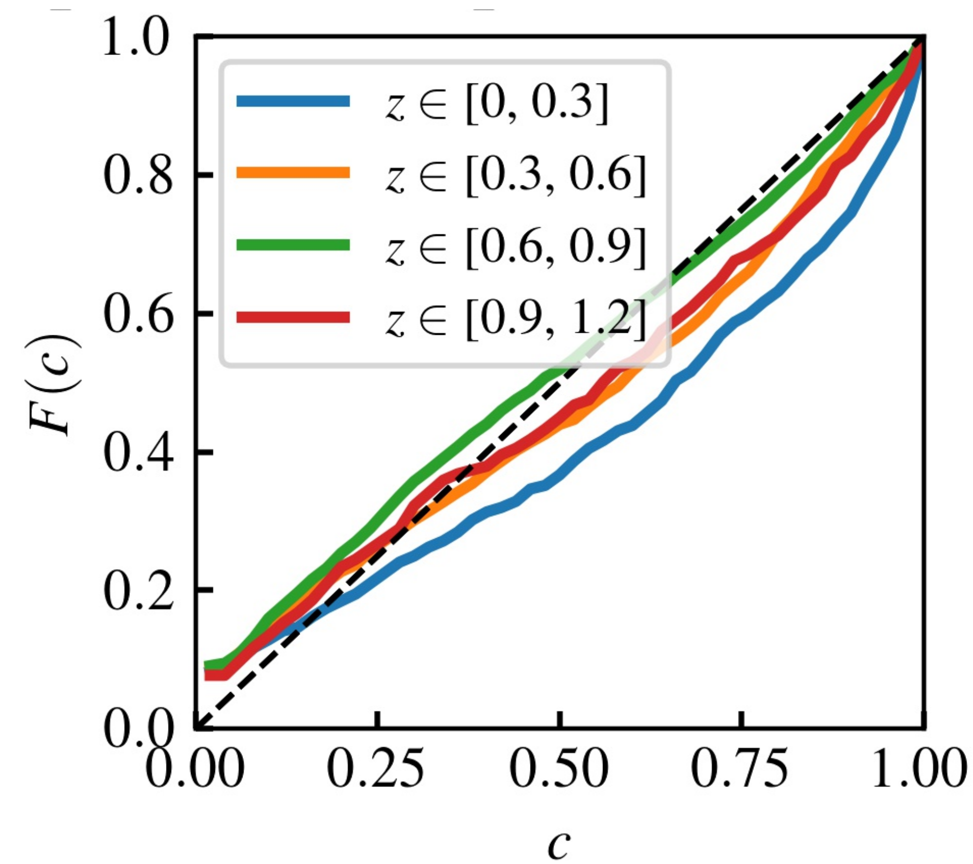
a) **Precision**

e.g. compact redshift PDFs



b) **Accuracy**

e.g. diagonal QQ plots

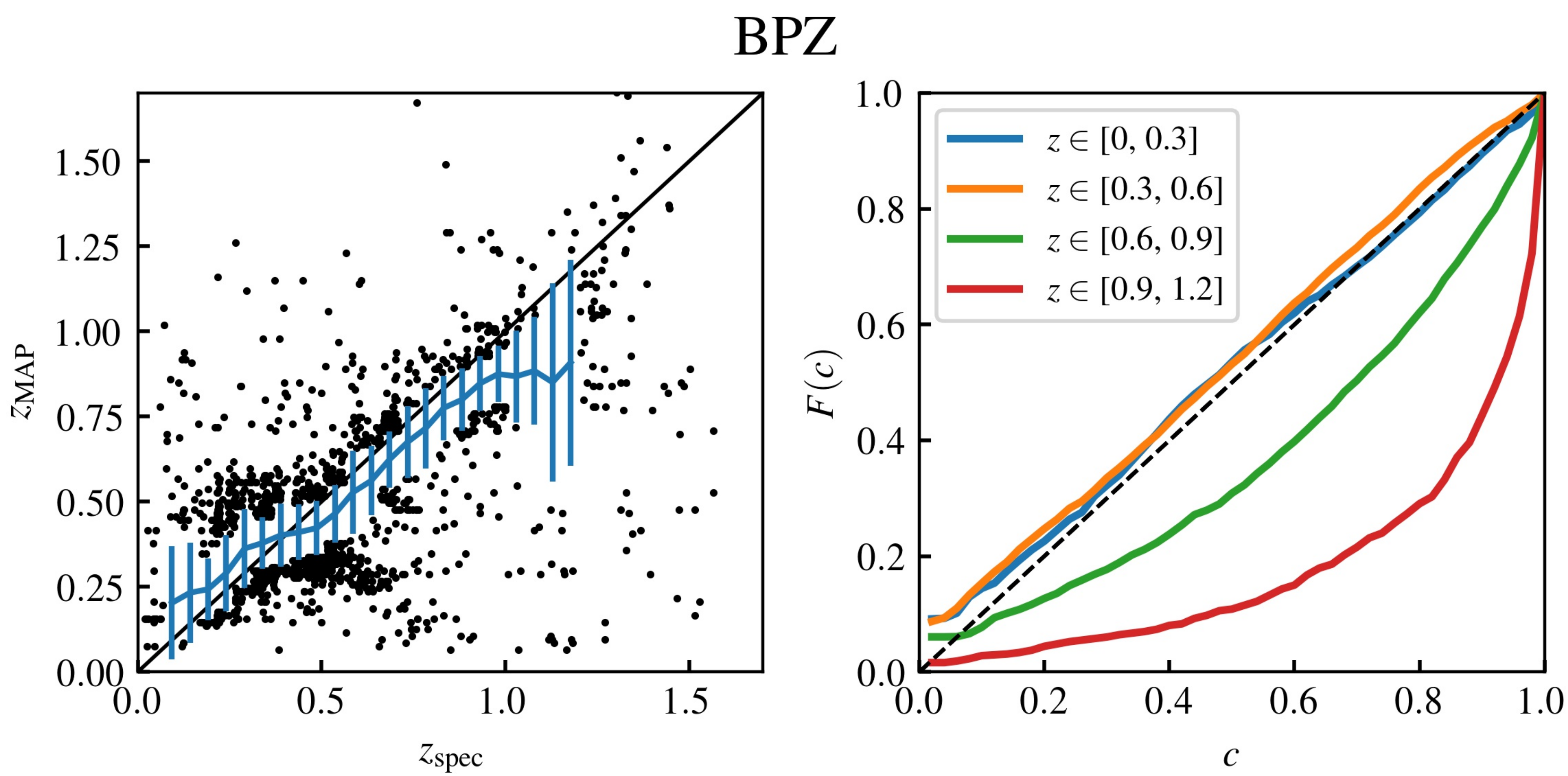


= validating fraction of galaxies in redshift PDF confidence intervals

c) **Interpretability**

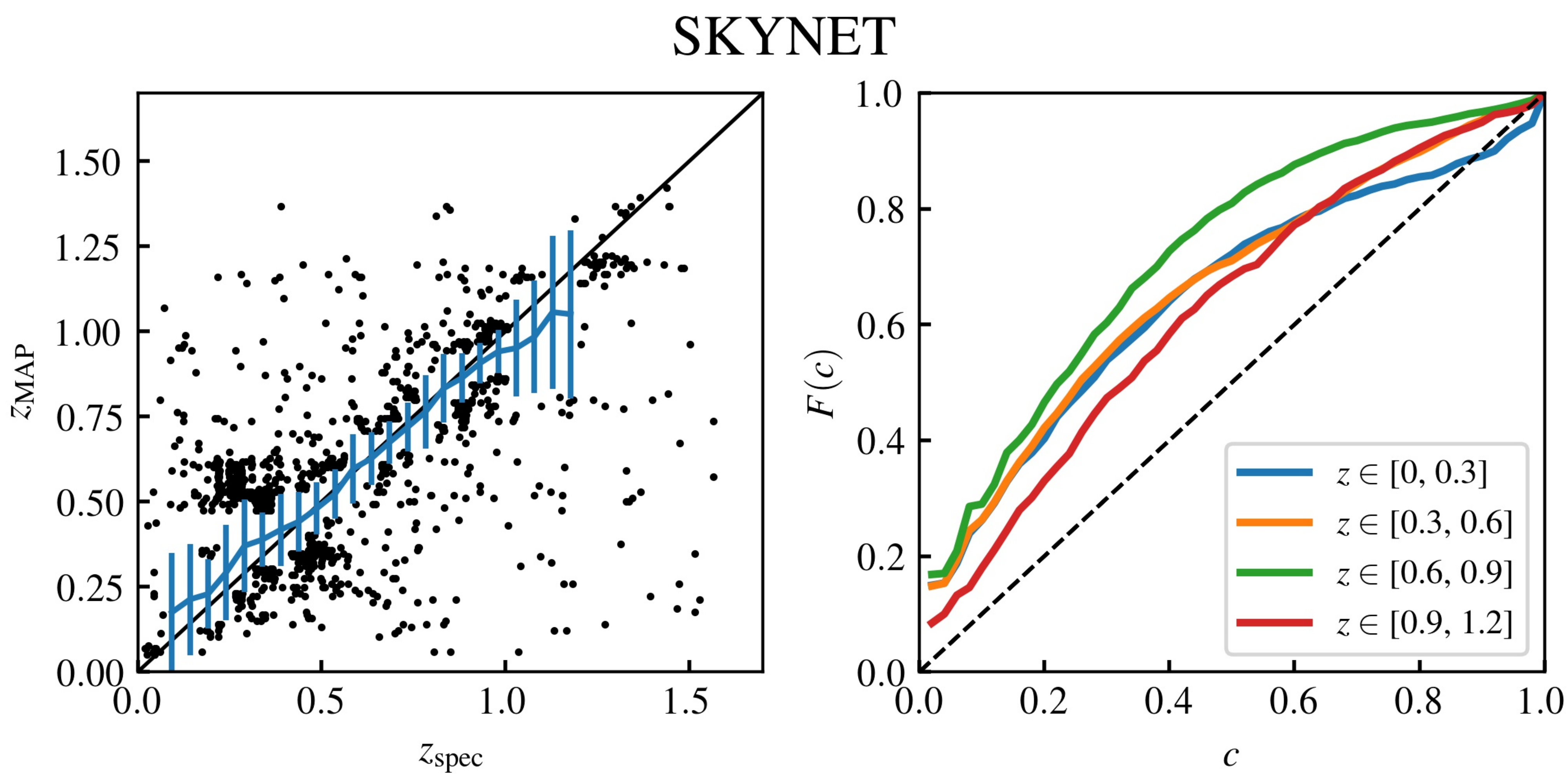
e.g. SED model

BPZ: template fitting, 8+interpolated SEDs, simple priors



interpretable model but biased photo-z’s & under-estimated errors

SKYNET: machine learning (Mixture Density Networks)

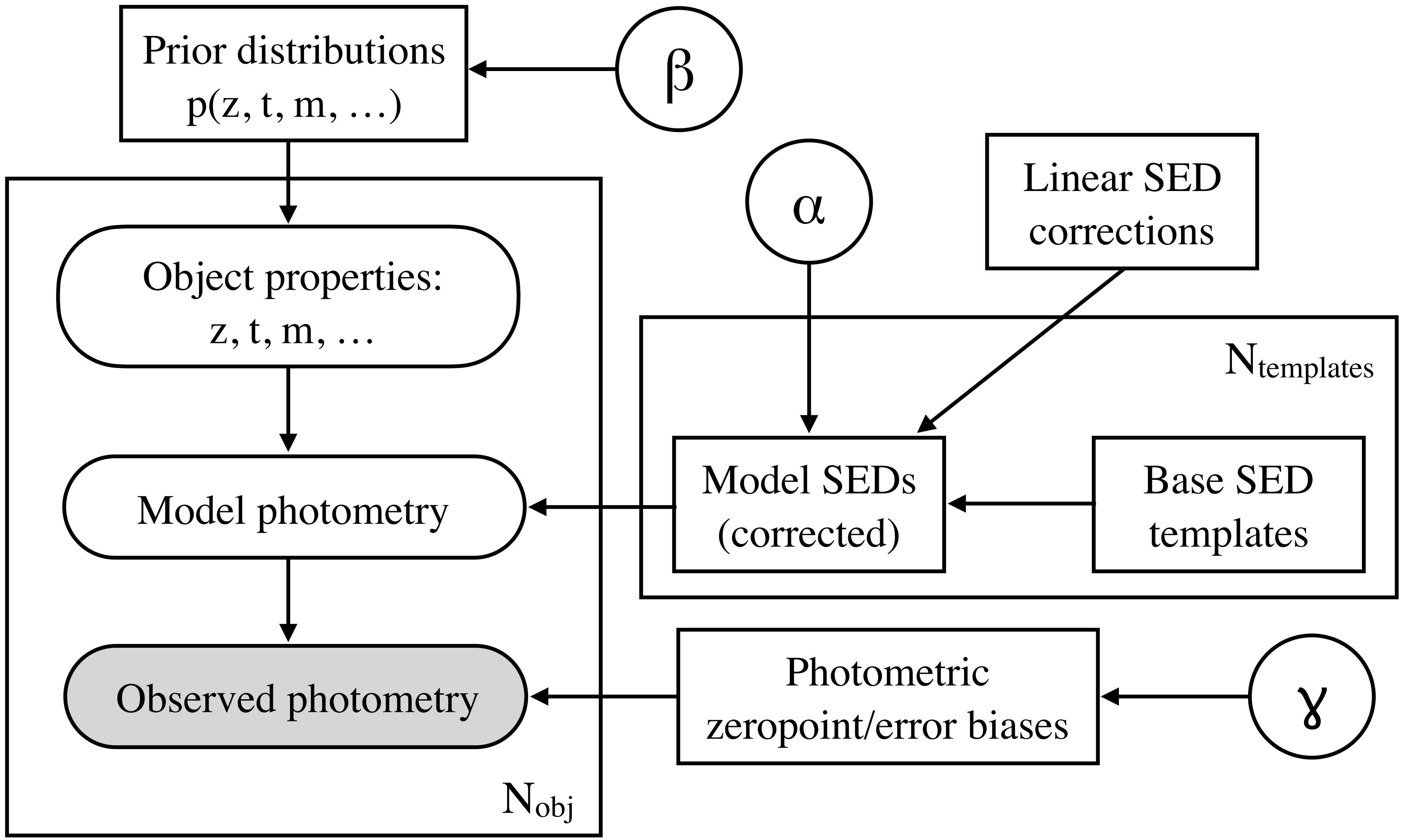


unbiased photo-z's but not interpretable & over-estimated errors

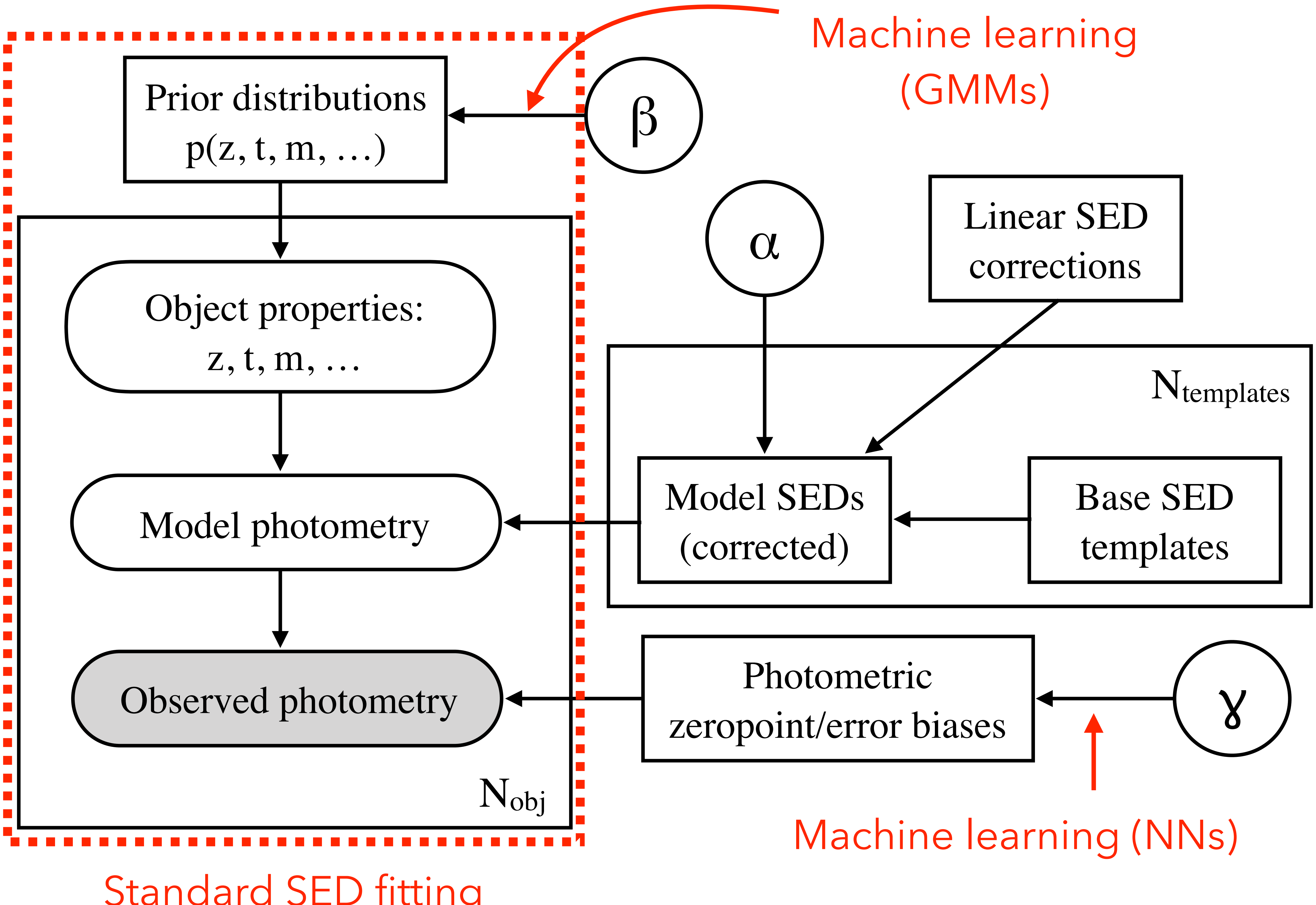
Photo-z uncertainty budget

	Statistical	Systematic
Data	**Aleatoric uncertainties** *true data noise,* *flux variances, etc*	**Data biases** *misestimated fluxes,* *zeropoints, variance, etc*
Model	**Epistemic uncertainties** *unmodeled SED effects,* *variability, variance, etc*	**Model biases** *miscalibrated SEDs or* *priors p(z, t, ell, etc)*

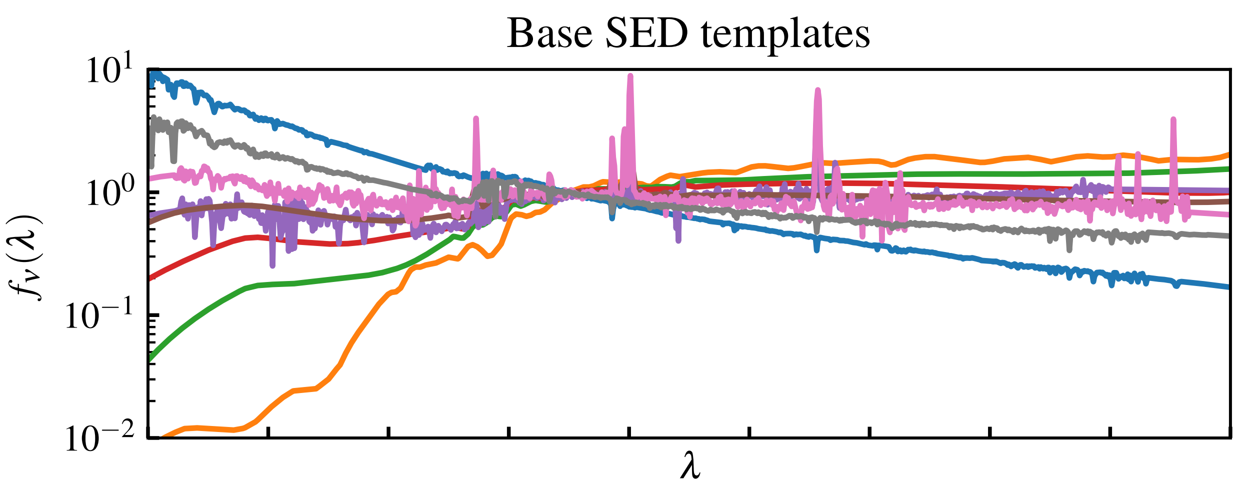
Full hierarchical model

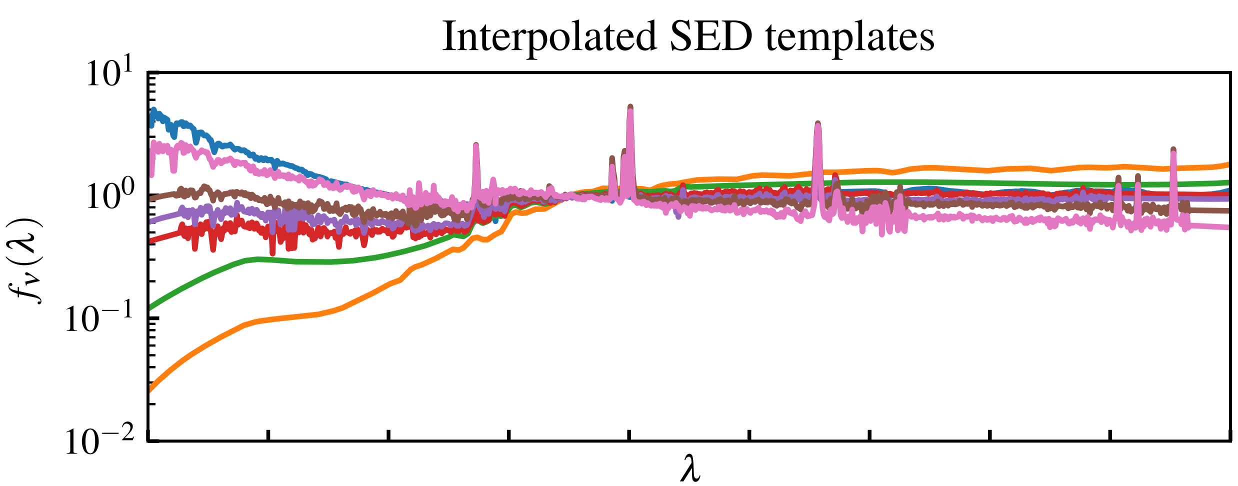


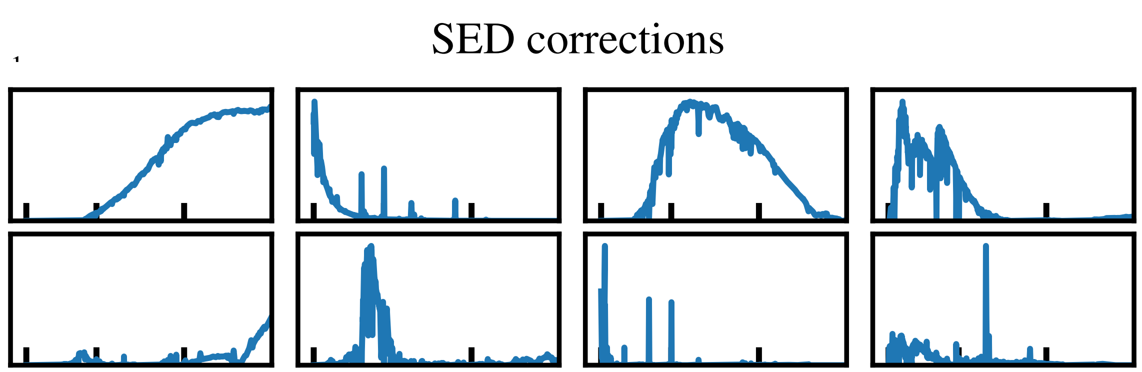
Full hierarchical model



Hierarchical model: SEDs + corrections







…

- **Base SEDs**: CWW library (8)
  + interpolated SEDs
- **Linear corrections**: NMF/PCA of CWW and PEGASE SEDs + Gaussian corrections

$$f_t^{\text{corrected}}(\lambda) = f_t^{\text{base}}(\lambda) + \sum_i \alpha_{it} f_i^{\text{correction}}(\lambda)$$

- SED variance constructed from corrections

$$\mathrm{Var}_t(\lambda) = \Big(\sum_i \beta_{it} f_i^{\text{correction}}(\lambda)\Big)^2$$

Hierarchical model: priors

- **Factorization**: $p(z, m, t) \;=\; \boxed{p(z|m,t)}\;\boxed{p(t|m)}\;\boxed{p(m)}$
  redshifts types magnitudes
- **Magnitude prior**: *p(ell or m)* uniform (in reference band)
- **Type prior**: $p(\mathrm{type} = t|m) = v_t(m)$ with $\sum_t v_t(m) = 1 \; \forall m$

  = Dirichlet prior on the simplex, with $v_t(m)$ quadratic in m
- **Redshift prior**: (all parameters quadratic in m)

  Simple N(z): $p(z|m,t) = \frac{z}{\bar{z}_t(m)} \exp\left(-\frac{z^2}{2\bar{z}_t(m)}\right)$

  Gridded Gaussian Mixture: $p(z|m,t) = \sum_i \gamma_i(m) \mathcal{N}(\mu_i - z; \Delta)$

Hierarchical model: flux/noise

- **Multiplicative zero point corrections:**

  Quadratic in reference magnitude: $\hat{F}_b \longrightarrow \hat{F}_b \times w_b(m)$

  General form (neural network!): $\hat{F}_b \longrightarrow \hat{F}_b \times w_b(\hat{F}_1, \cdots, \hat{F}_B)$

- **Minimum magnitude error per band:**

  Quadratic in reference magnitude: $\sigma^2_{\hat{m}_b} \longrightarrow \max[\sigma^2_{\hat{m}_b}, w'_b(m)]$

  General (neural network): $\sigma^2_{\hat{m}_b} \longrightarrow \max[\sigma^2_{\hat{m}_b}, w'_b(m_1, \cdots, m_B)]$

Hierarchical model: posterior

$$p\big(\vec{\alpha}, \vec{\beta}, \vec{H} | \{\hat{\vec{F}}_i\}\big) \ \propto \ p(\vec{\alpha}, \vec{\beta}, \vec{H}) \prod_{i=1}^{\mathrm{N_{obj}}} \sum_{t=1}^{\mathrm{N_{types}}} Q_{it}(\vec{\alpha}, \vec{\beta}, \vec{H})$$

- **Alpha**: parameters of the SEDs / flux model
- **Beta**: parameters of the data error recalibration
- **H**: parameters of the prior p(z, t, l)
- **Qit**: marginal evidence of the i-th object under the model
- Analytic solution for ell marginalization since additive or multiplicative scaling in Gaussian likelihood
- *Here for spectroscopic training set, but could be written for photometric data too!*

TensorFlow™

- Google's toolskit for linear algebra, covering numpy+scipy functionalities
- Build graphs of data/operations + gradients with automatic/symbolic differentiation
- Best optimizers on the market
- Interfaces with deep learning & probabilistic inference libraries
- Great for optimization and modeling. Advanced inference/sampling via external libraries such as Edward.

Models

interp SEDs	prior $p(z,t,m)$	SED mean corrections	SED variances	mag error corrections	N_{par}	$\log[Q]/N_{obj}$ (training)	$\log[Q]/N_{obj}$ (validation)
2	simple	✓			2398	-9.04	-7.49
2	simple			f(m)	210	17.49	18.34
2	simple	✓		f(m)	2410	19.43	20.00
0	simple	✓	✓		1672	18.57	19.24
2	simple	✓	✓		4598	19.87	20.43
2	GMM	✓	✓		5126	19.73	20.21
2	simple	✓	✓	f(m)	4610	19.83	20.35
2	GMM	✓	✓	f(m)	5138	19.93	20.44
2	simple	✓	✓	NN	5022	19.73	20.33
4	simple	✓	✓	NN	7948	20.43	20.84

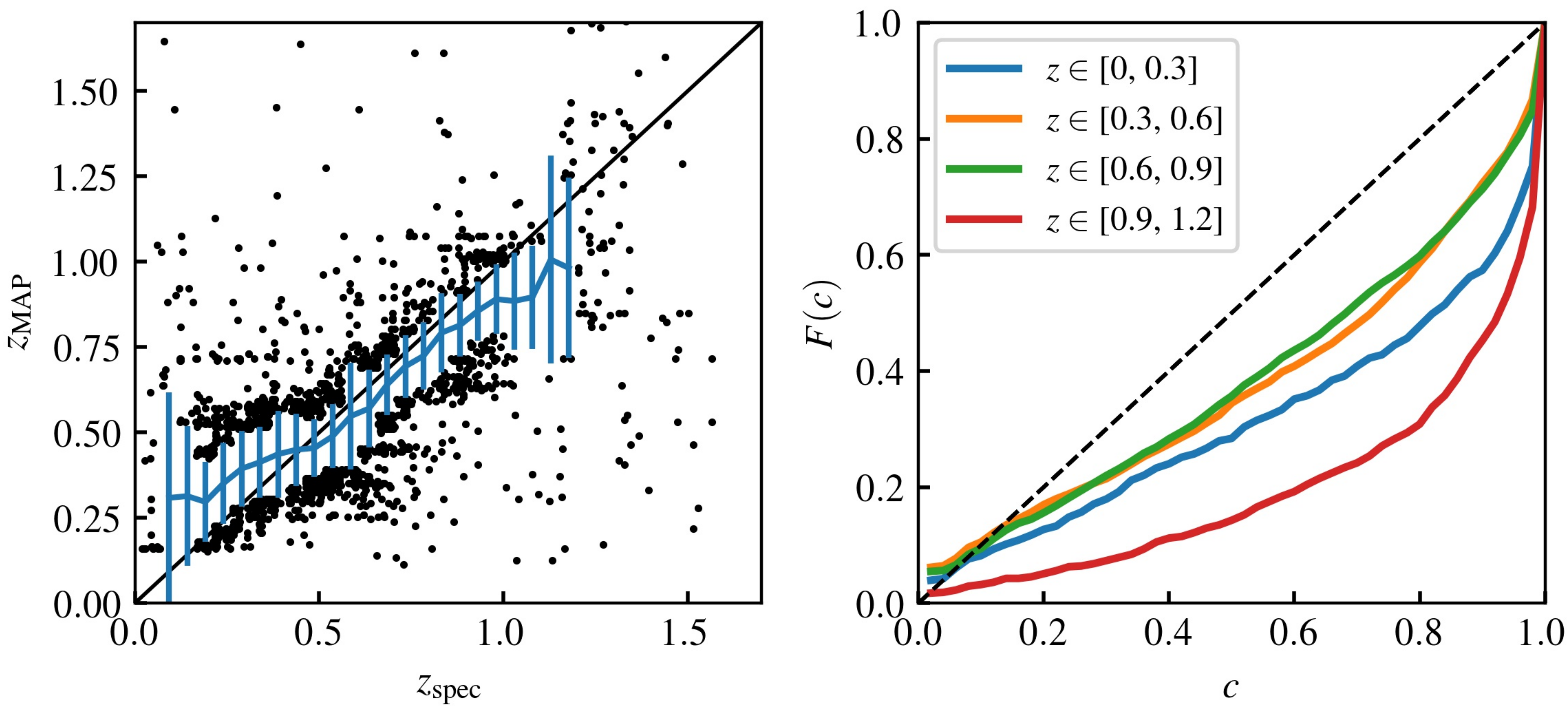
Findings

1. Cannot eliminate bias without SED corrections or variance
   (simultaneously optimized with SED priors)

2. Models with SED variance or noise have good QQ metrics

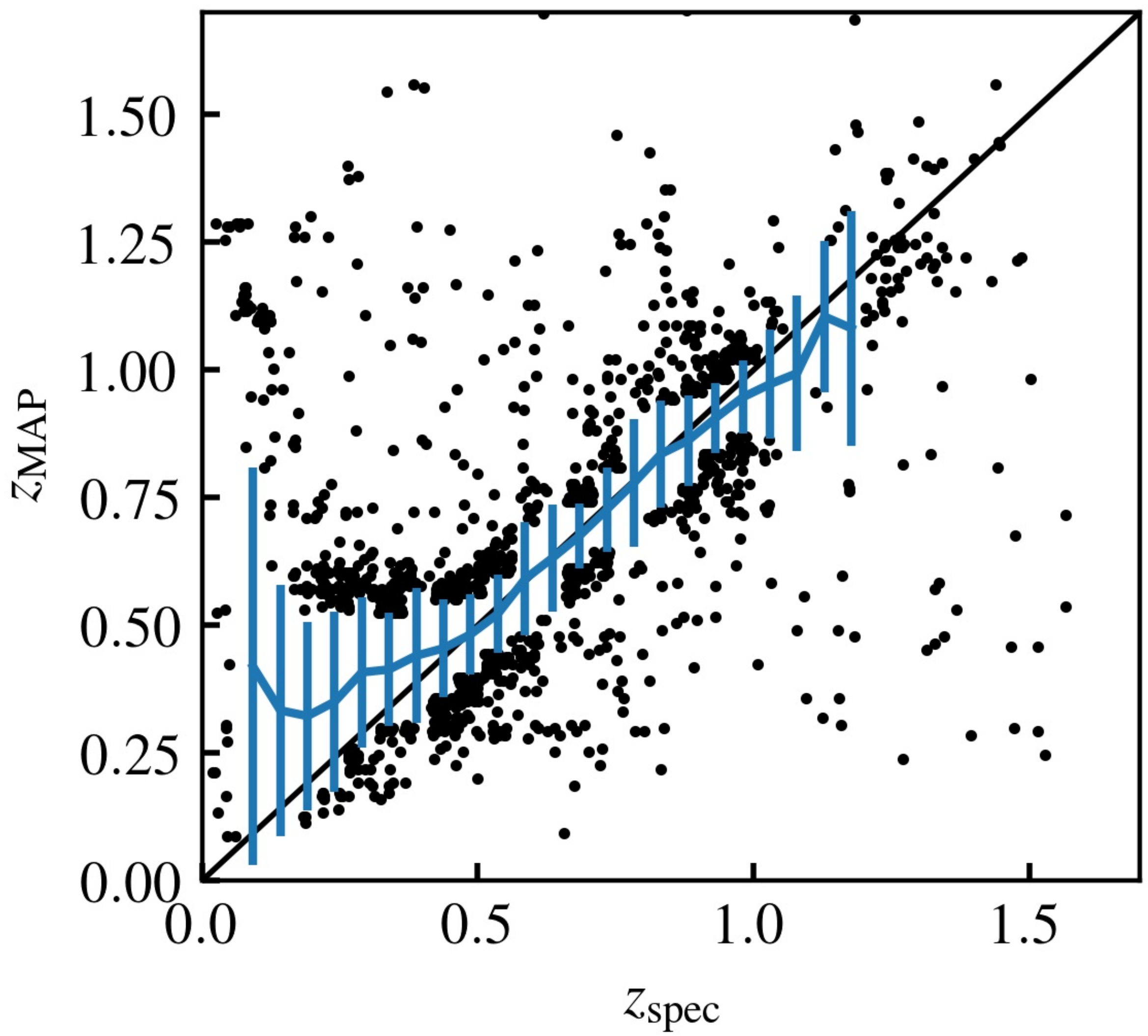
3. Even with SED variance, some extra g-band noise is

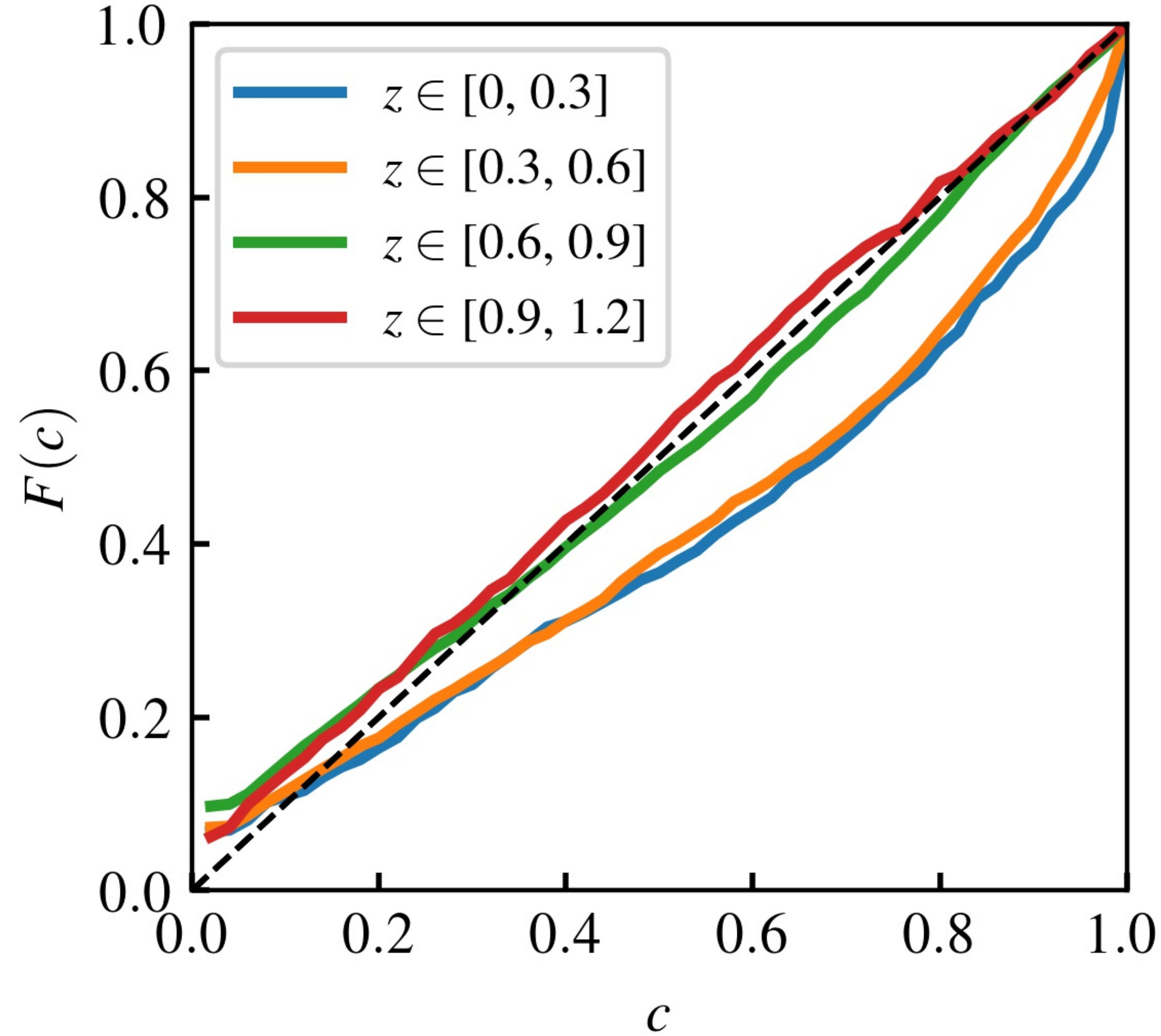
1. Cannot eliminate bias without SED corrections or variance



HM: 2 interpolated SEDs, extra photometric noise
(no SED corrections or variance)

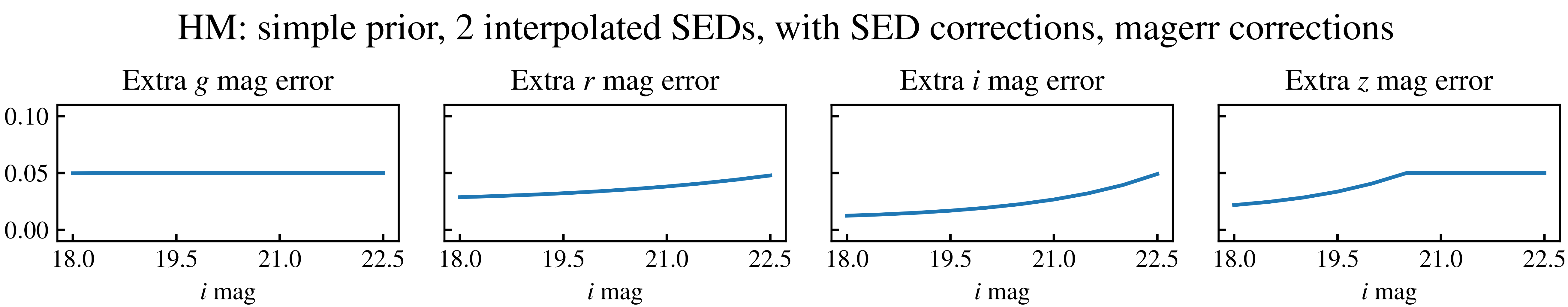
2. Models with SED variance or noise have good QQ metrics

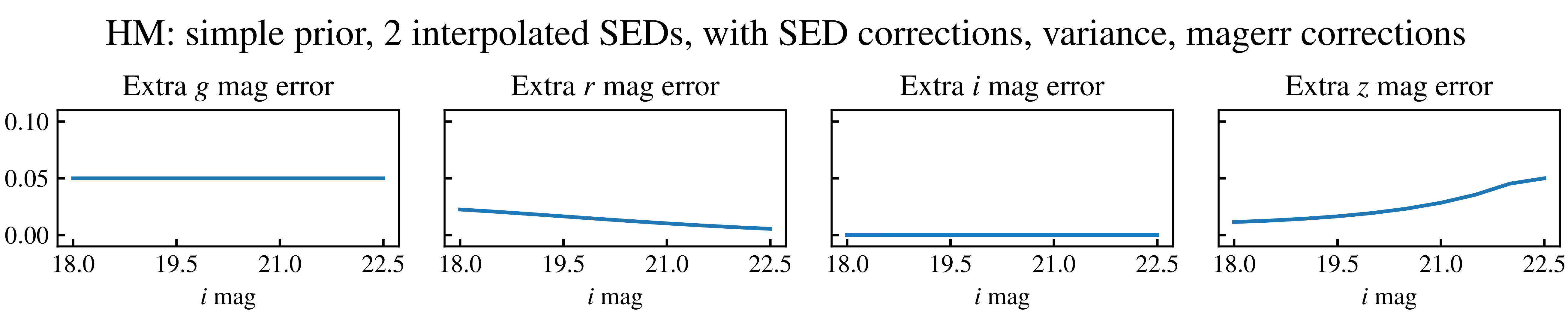




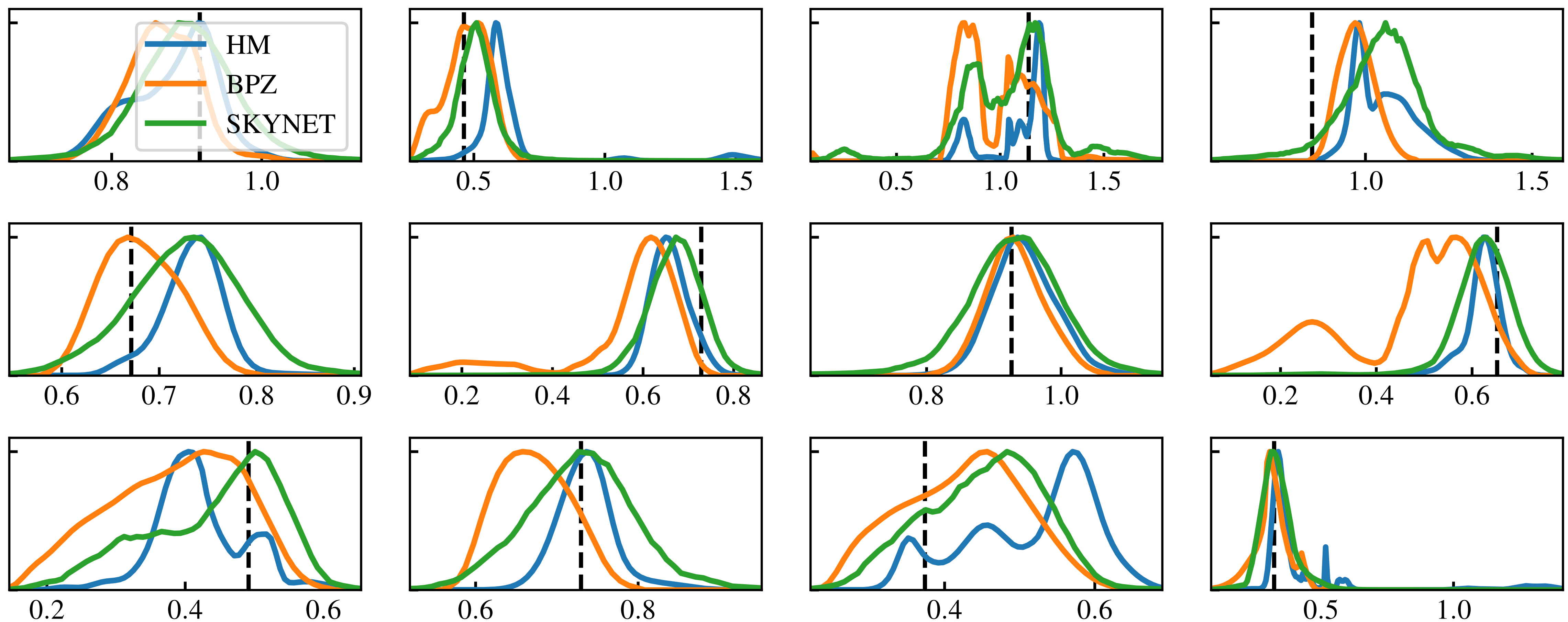
HM: 2 interpolated SEDs with SED variance & extra noise

3. Even with SED variance, some extra g-band noise is needed

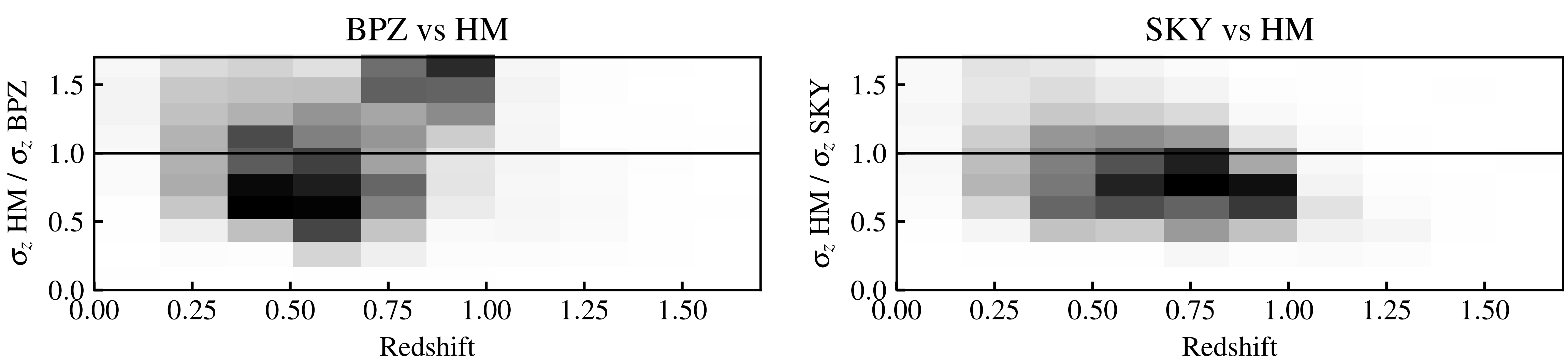




4. Redshift PDFs are more compact/precise



HM: simple prior, 2 interpolated SEDs, with SED corrections, variance, magerr corrections

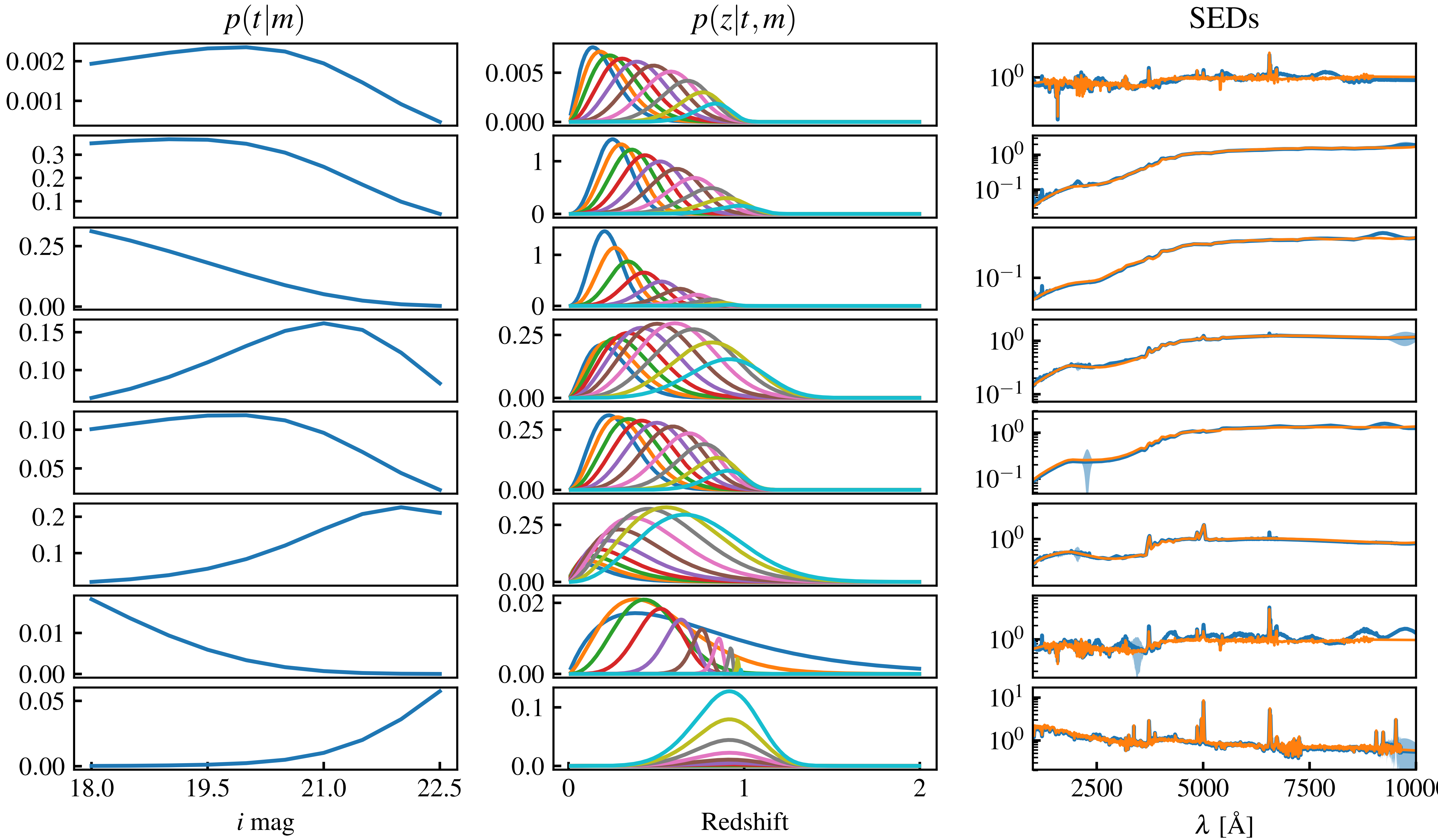


Findings (continued)

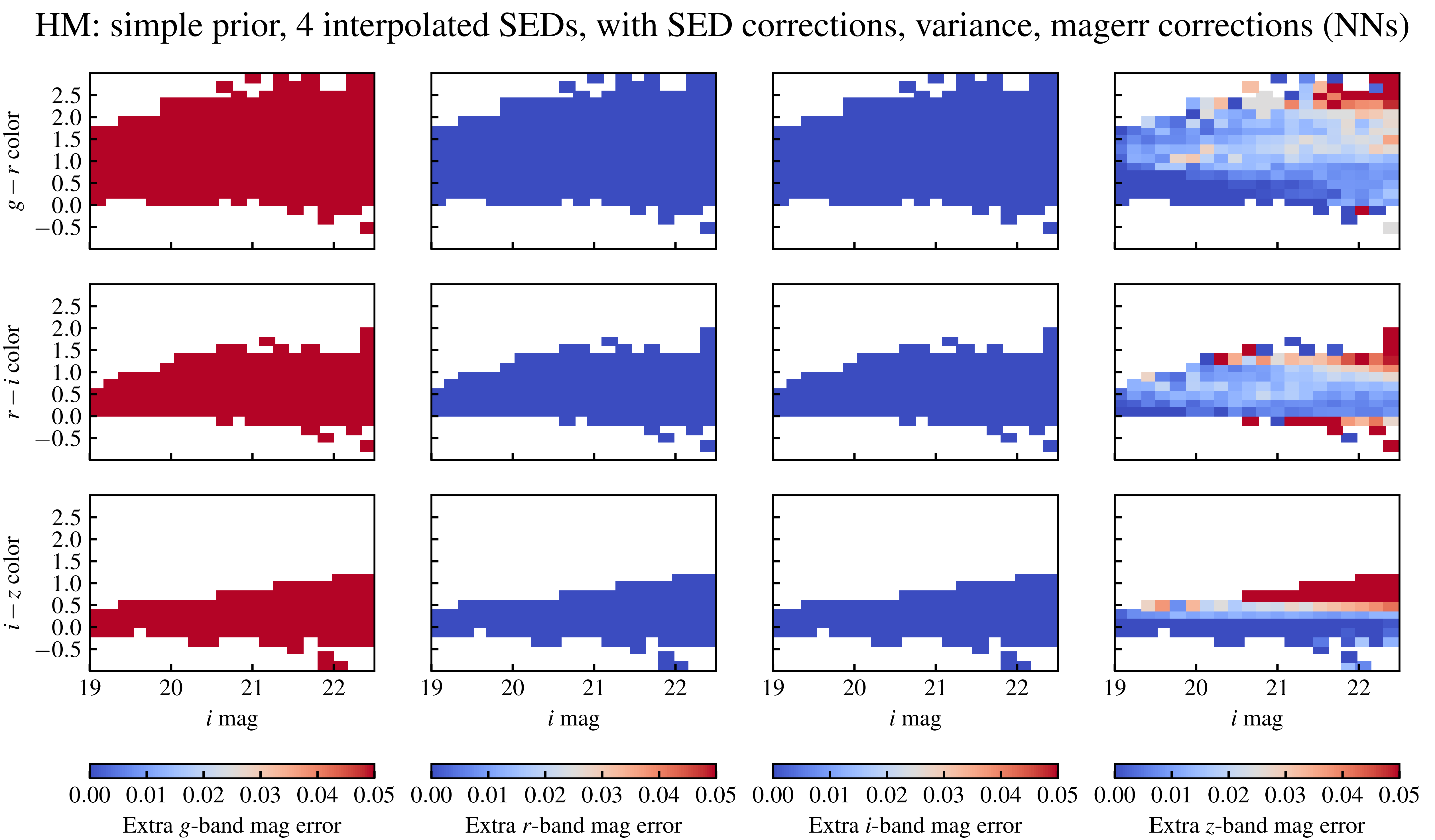
5. Outliers are consistent across models
6. SED priors and corrections are interpretable
7. More complex redshift priors marginally helps
8. Number of interpolated SEDs marginally helps
9. More complex noise corrections marginally

Example of SEDs and priors (top 8)

HM: simple prior, 2 interpolated SEDs, with SED corrections, variance, magerr corrections



Example of NN noise corrections



Summary

Hierarchical model for self-calibration of photometry & SEDs to self-consistently generate survey data at high accuracy
and derive photometric redshifts

Current: re-calibration of SED grid + priors + photometry

Soon: redshift/luminosity-dependent data-driven SEDs, AGN component, spatially-varying photometry

Future: filter responses, image artefacts