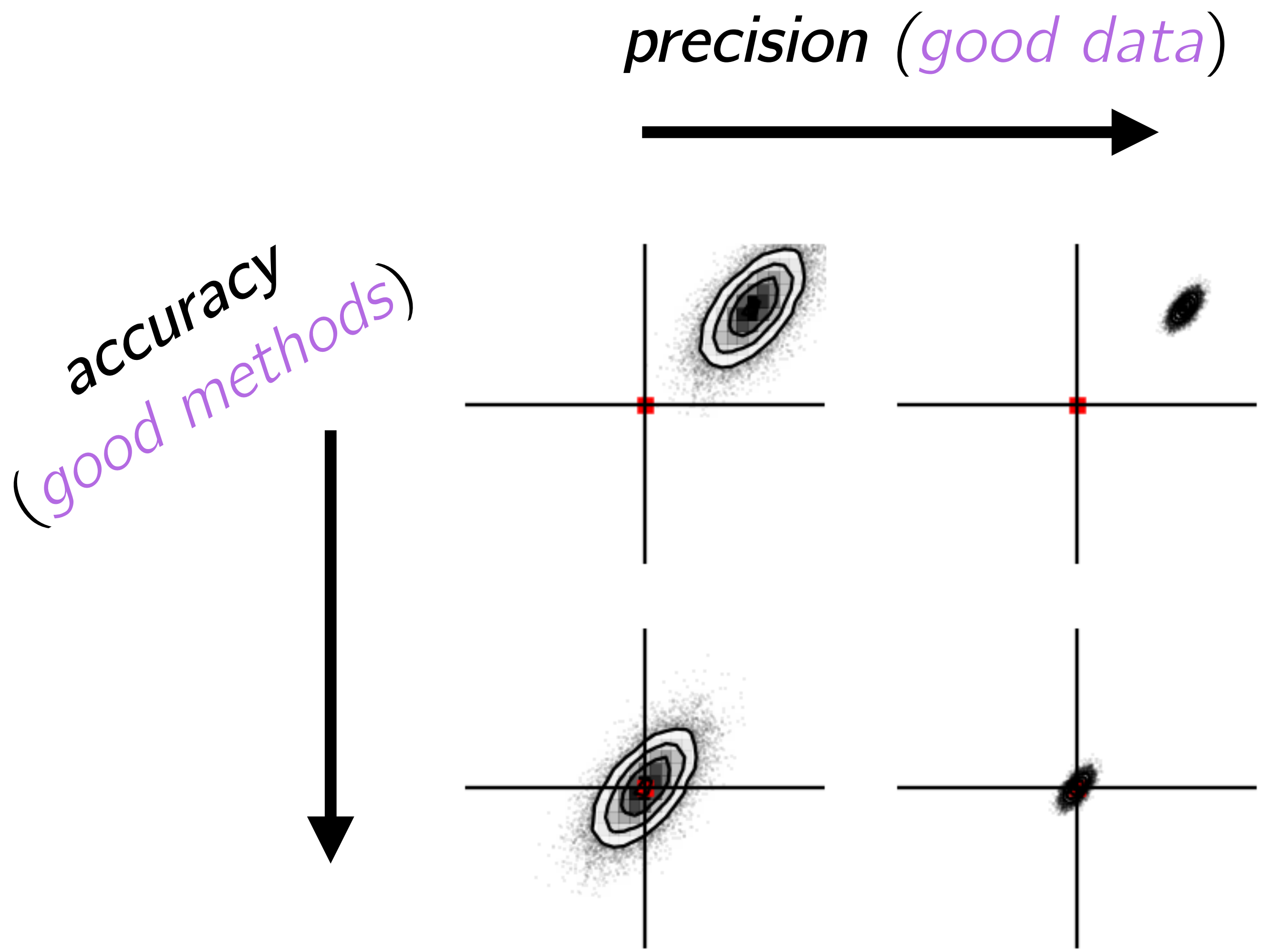
Hierarchical modeling and statistical calibration for photometric redshifts

Context and motivation
Standard methods and challenges
New method applied to DES SV data

NYU

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

work with D. Hogg (NYU), R. Wechsler, J. De Rose (Stanford)

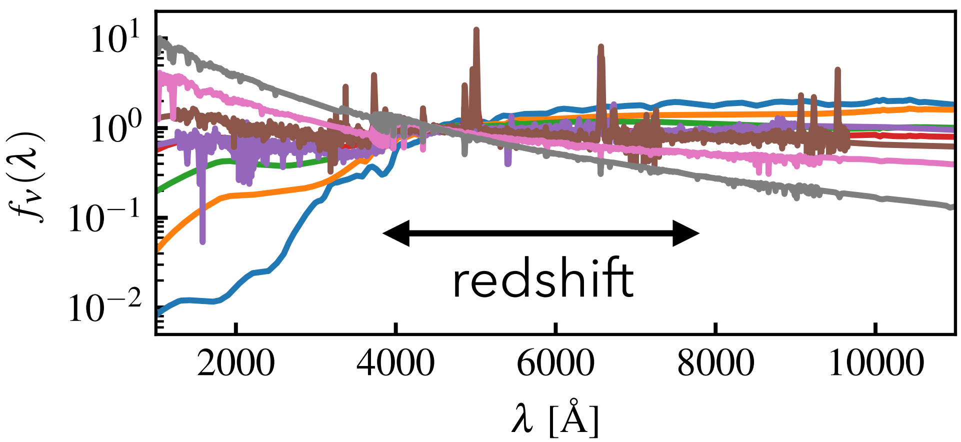


<u>This work</u>: hierarchical SED modeling
for both precise & accurate photometric redshifts
(applied to public DES SV data)

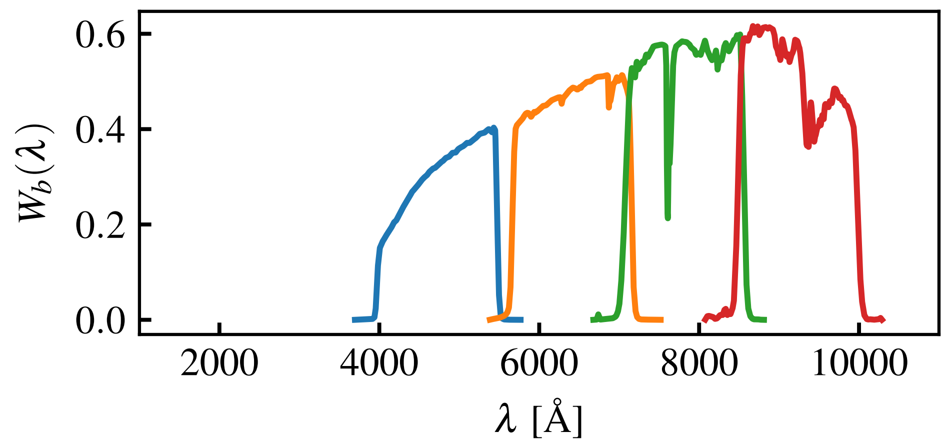
Photometric redshift

= estimating redshift from noisy broadband photometry

Galaxy spectra

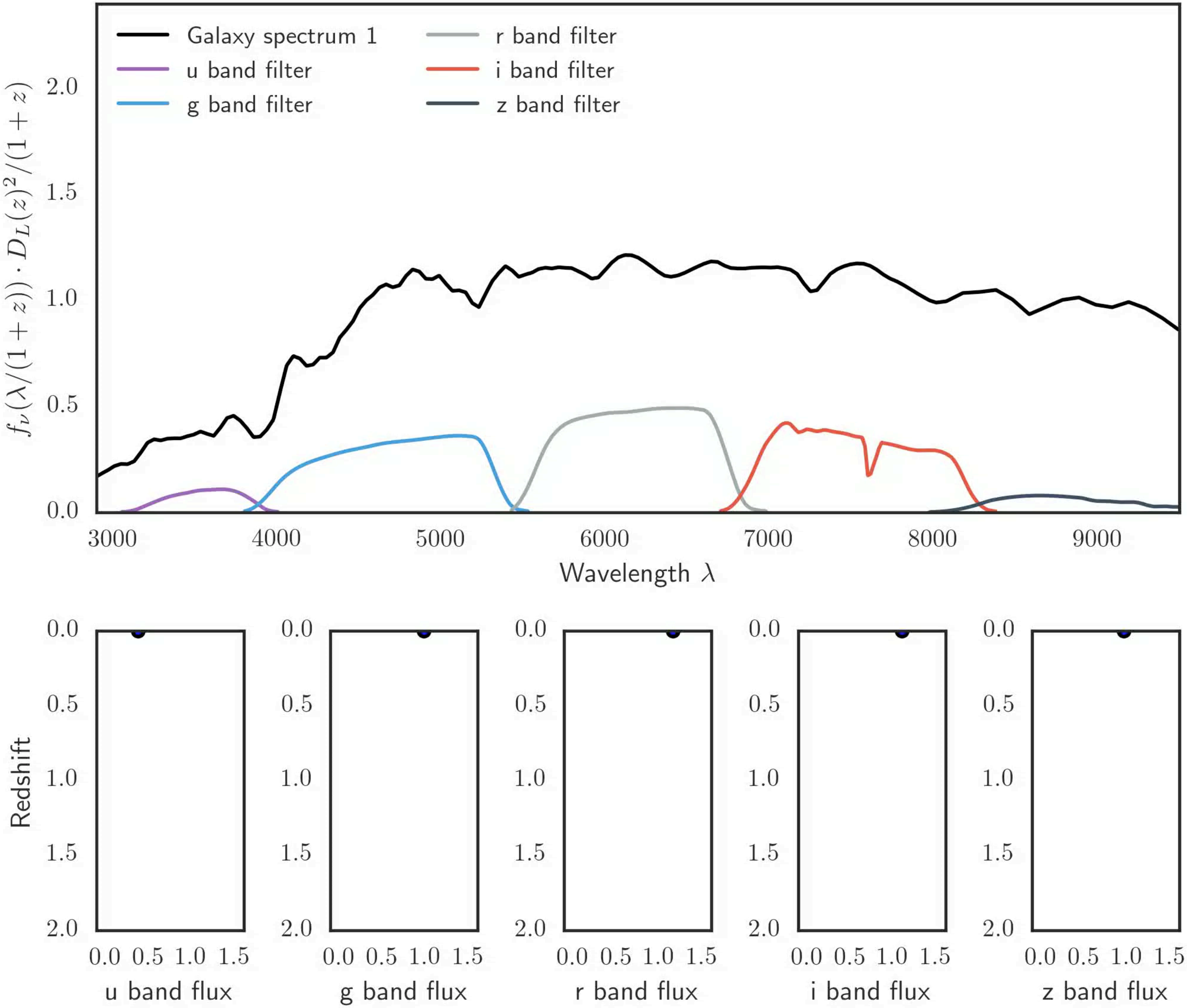


DECAM photometric band passes



using knowledge of galaxy SEDs, redshift, bandpasses, etc
OR seing it as redshift=f(fluxes) regression problem.

This is an animation - sorry if you're watching this in PDF



Context and motivation

- DES and LSST see 10^{7-9} galaxies to z=1-2
- **Cosmological requirements**: exquisite precision needed on N(z) for galaxy clustering + cosmic shear constraints
- **Complicated data:** 5 band photometry, selection effects, biases. Manageable impact on single galaxies. Huge impact on N(z)'s. *Example: poorly estimated but significant probability at high-z.*
- We will ***never*** have representative spectroscopic training data.
- **Accuracy** requirements not met. Validation is difficult.

The Right Way To Do It™

Simultaneously optimize/infer a photo-z model on all of our data

Data (possibly split in training/validation):

- Target 4-5-band photometric survey (e.g., DES, KIDS)
- Some spectroscopy, some extra photometric bands

Model:

- <u>Observations</u>: bandpasses, noise levels, detection criteria
- <u>Physics</u>: galaxy SEDs, relative abundances, redshift distributions possibly from stellar population models, luminosity function, etc

Objective function: full posterior distribution

Never done but now possible with new methods/technology

Three classes of methods

template fitting

Fitting SEDs to photometry using **likelihood function**
p(fluxes | noise, SED(z, t, l) model)

Requires calibrated SEDs/priors & unbiased data

machine learning

Construct **flexible model** for p(z|data) or fluxes(z) from spectroscopic training data

No likelihood, built-in prior, needs representative data

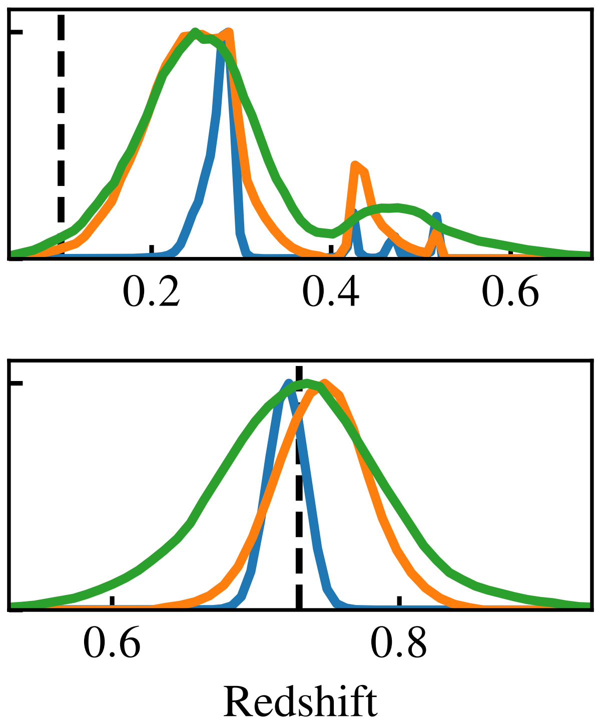
clustering redshifts

Constrain N(z) using spatial **cross-correlations** with spectroscopic or photometric samples

Requires overlapping samples, bias model

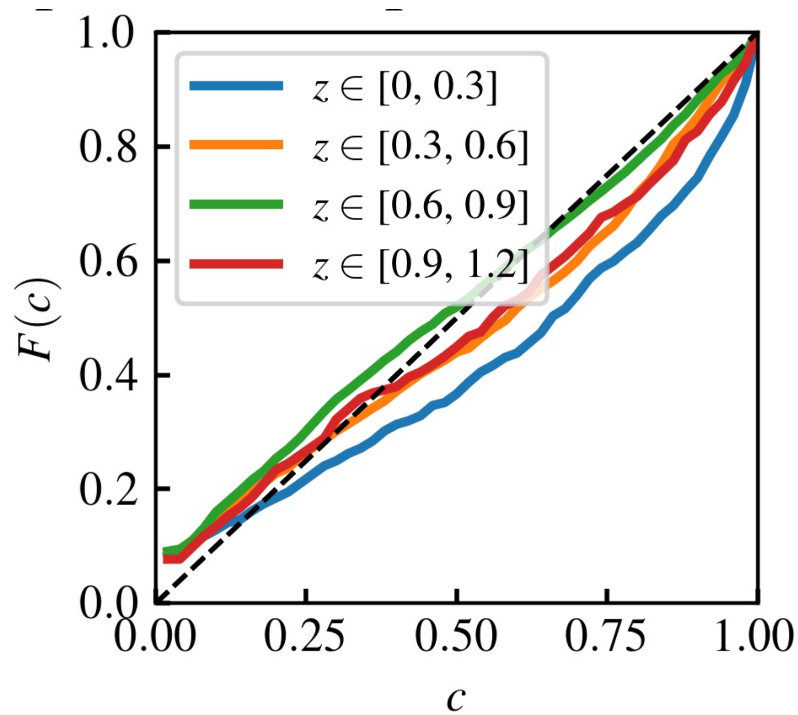
Precision

= compact redshift PDFs



Accuracy

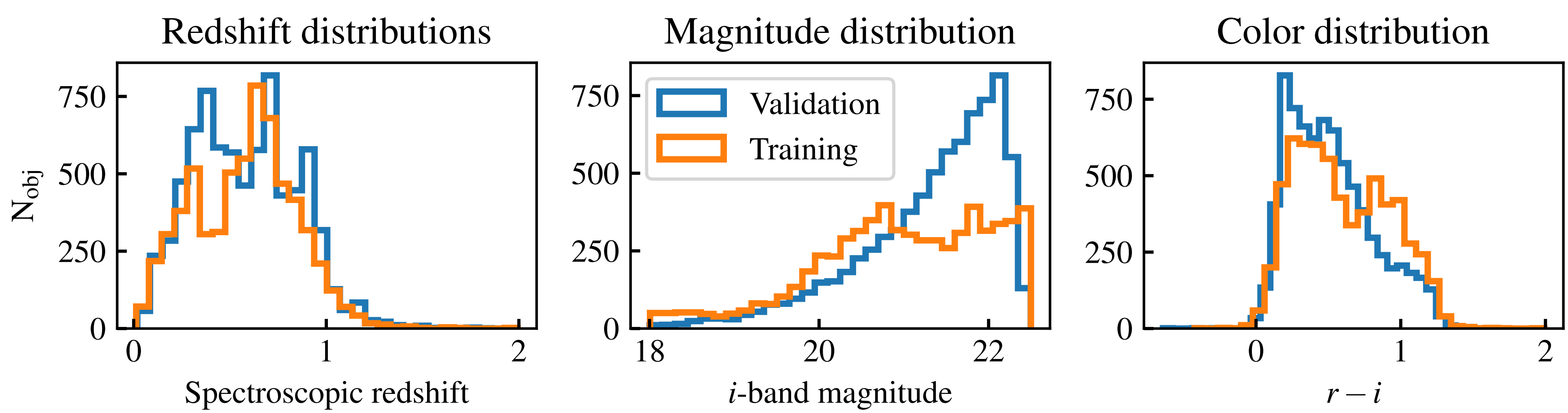
= diagonal QQ plots



= validating fraction of galaxies in redshift PDF confidence intervals

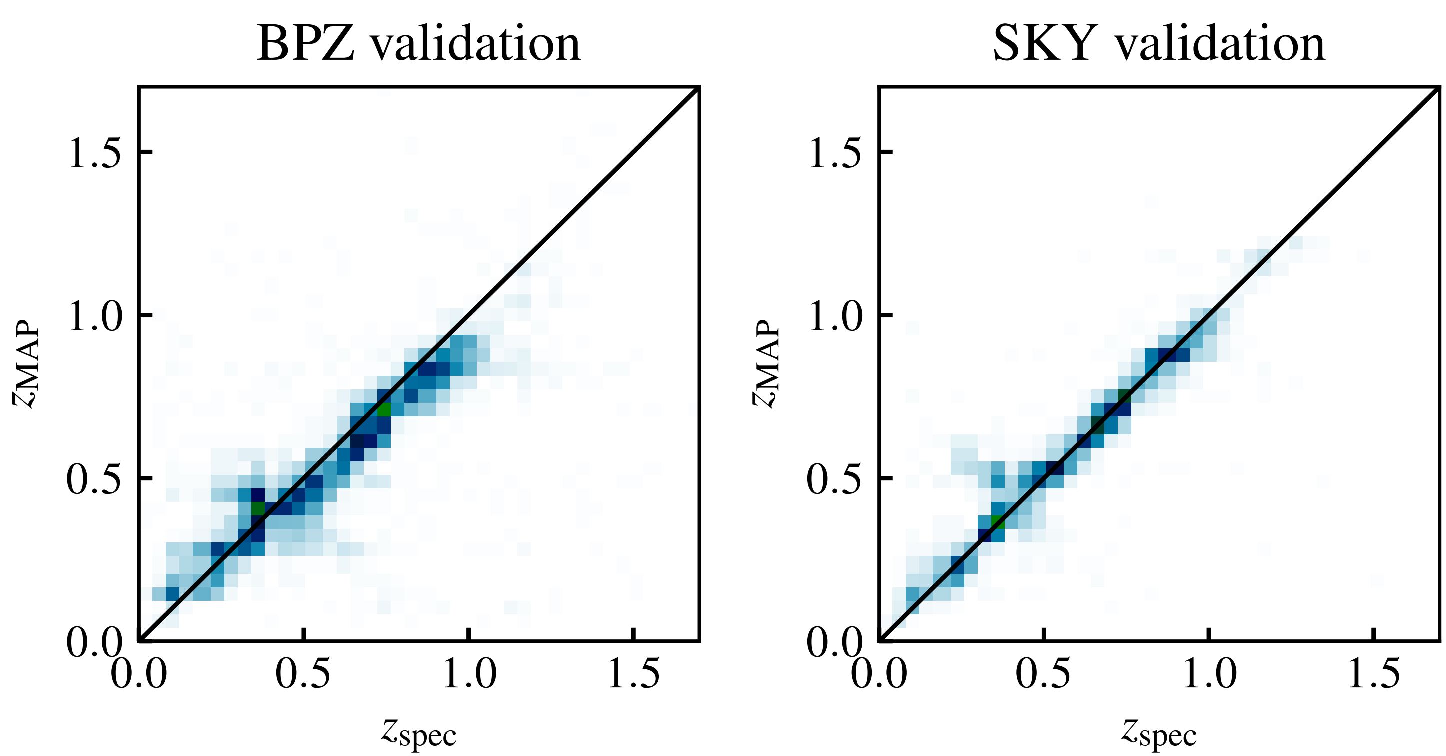
Data set: DES SV

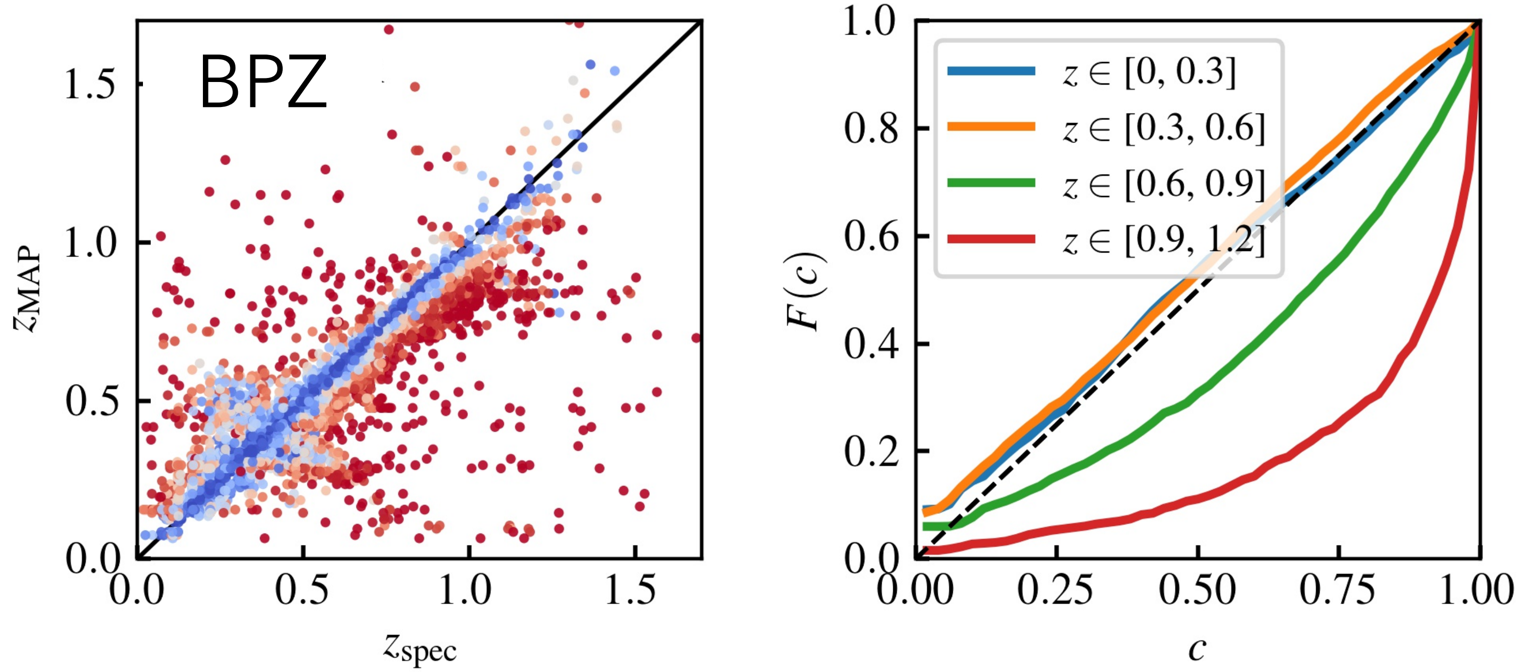
- Full SV data: 20+ million objects
- Gold sample: 18 < i magnitude < 22.5
- <u>Training</u>: VVDS, VIPERS, OzDES, ACES, 8k objects
- <u>Validation</u>: zCOSMOS, 8k objects



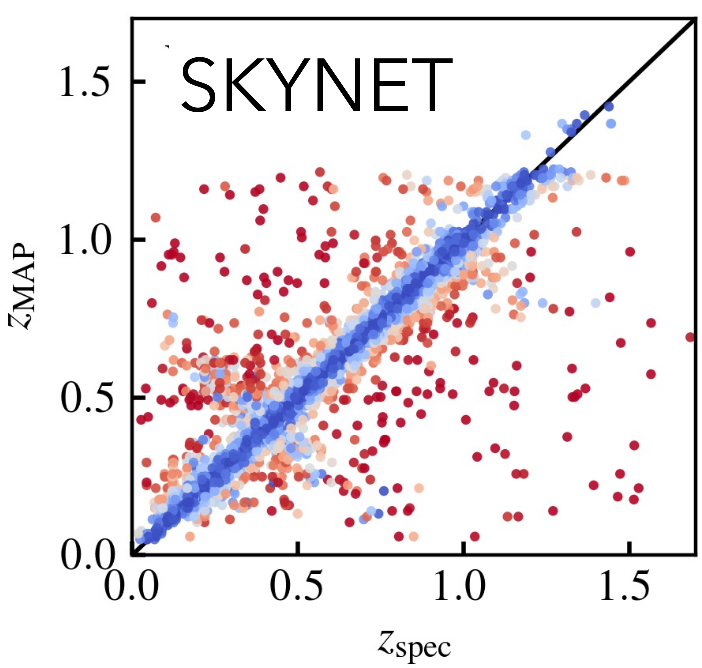
DES SV photo-z's (Bonnett + 2015)

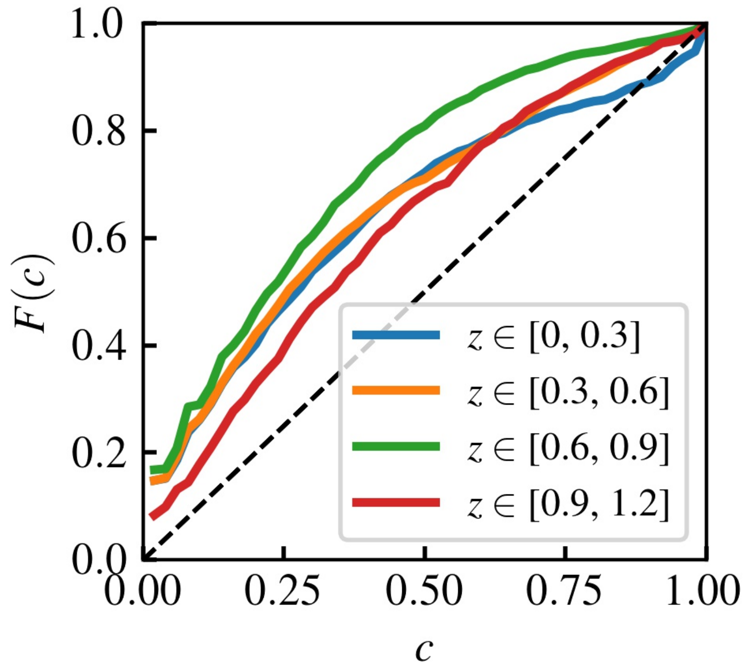
BPZ: template fitting, 8+interpolated SEDs, calibrated priors
SKYNET: machine learning (Mixture Density Networks)





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





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)*

Wish list

Precise and accurate redshift PDFs

4 sources of uncertainties modeled & propagated explicitly

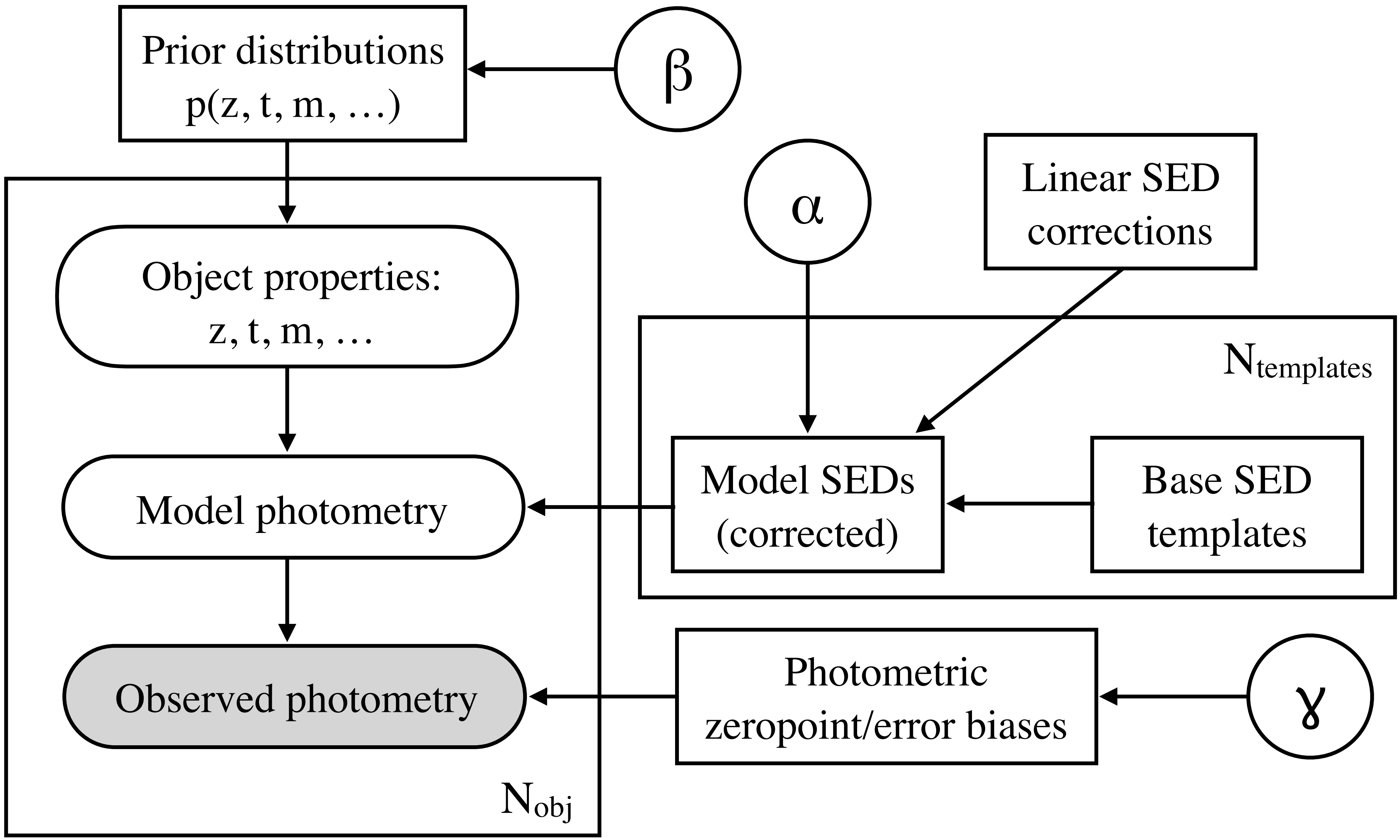
Interpretable PDFs and flux(z) model/priors

Combine SED models with machine learning and clustering

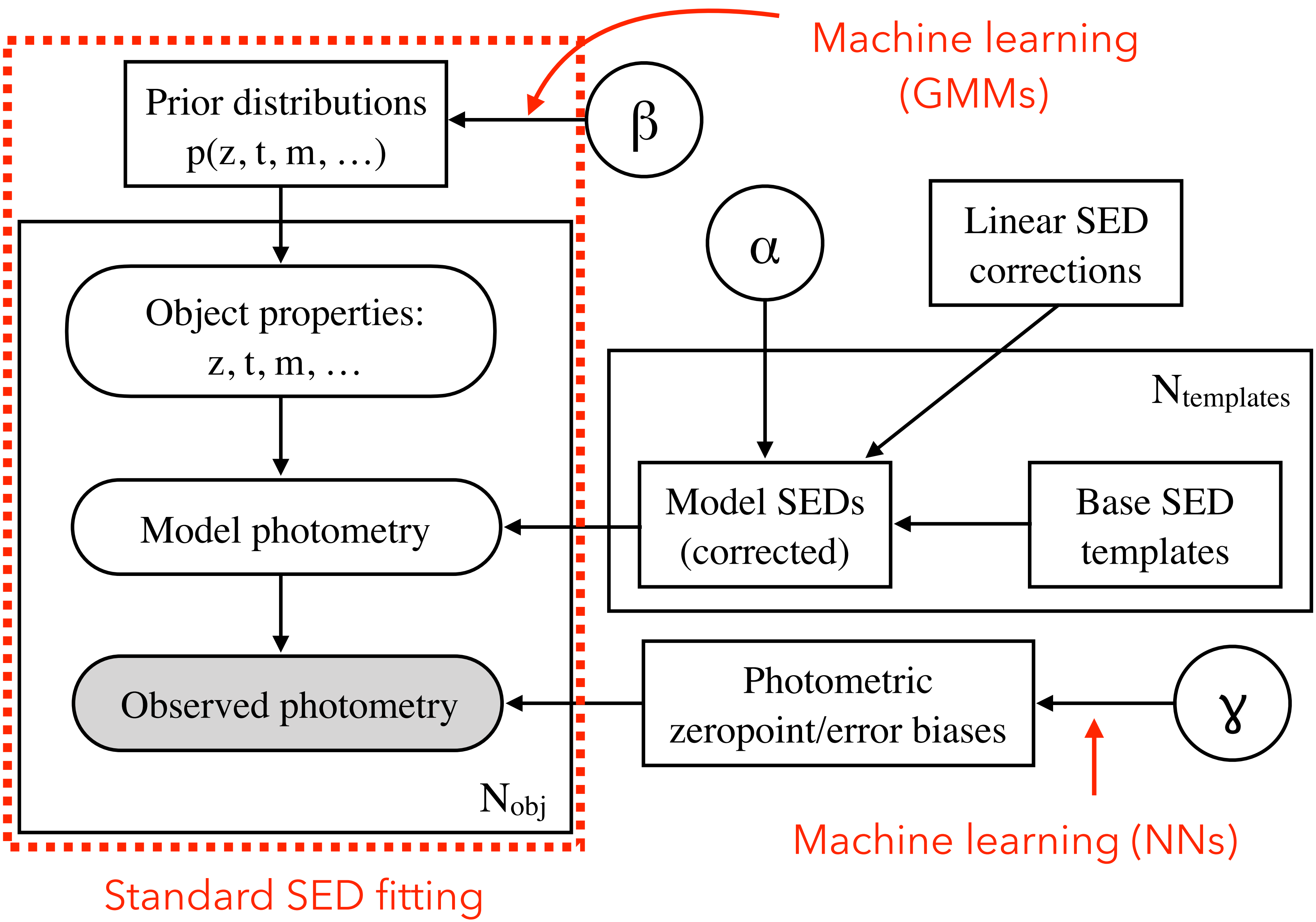
Likelihood function for hierarchical N(z) inference

*Need the **flexibility** of machine learning*
*and the **interpretability/generalization** of template fitting*
Solution: hierarchical probabilistic modeling

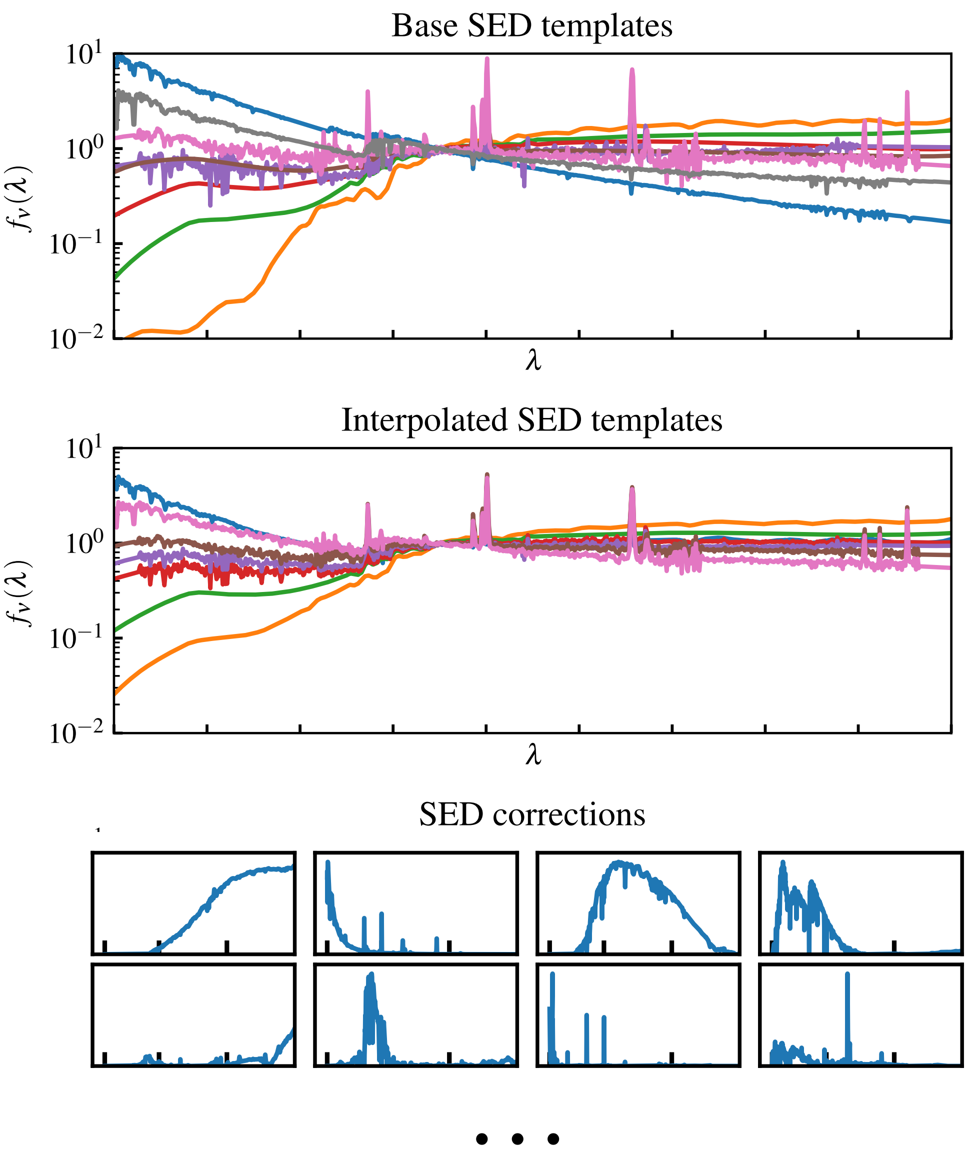
Full hierarchical model



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

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

- SED variance constructed from corrections

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

Hierarchical model: priors

- <u>**Factorization**</u>: $p(z,m,t) = \boxed{p(z|m,t)} \boxed{p(t|m)} \boxed{p(m)}$

  redshifts | types | magnitudes

- <u>**Magnitude prior**</u>: *p(ell or m)* uniform (in reference band)

- <u>**Type prior**</u>: $p(\text{type}=t|m) = v_t(m)$ with $\sum_t v_t(m) = 1\ \forall m$

  = Dirichlet prior on the simplex, with $v_t(m)$ quadratic in m

- <u>**Redshift prior**</u>: (all parameters quadratic in m)

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

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

Hierarchical model: flux/noise corrections

- **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	$\mathrm{N_{par}}$	$\log[Q]/\mathrm{N_{obj}}$ (training)	$\log[Q]/\mathrm{N_{obj}}$ (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

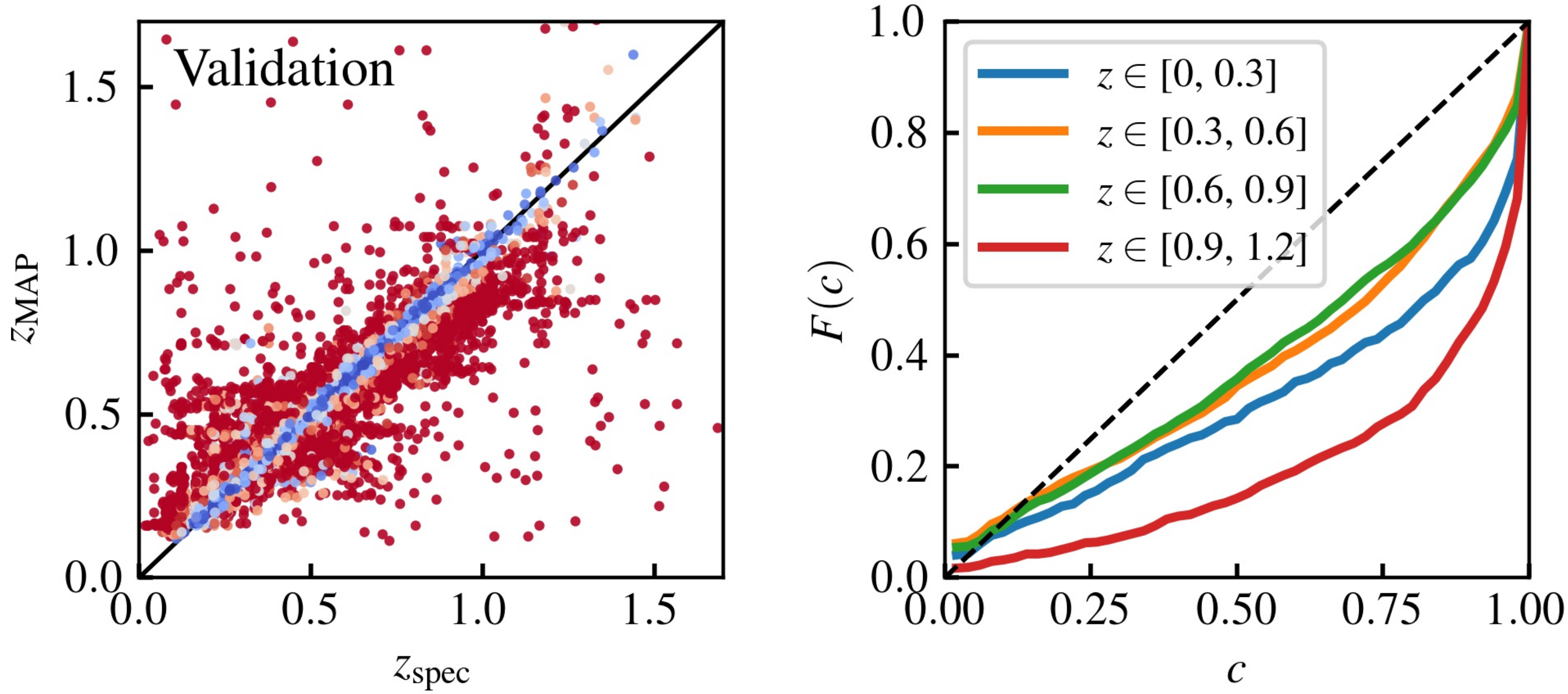
1. Cannot eliminate bias without SED corrections or variance (simultaneously optimized with SED priors)

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

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

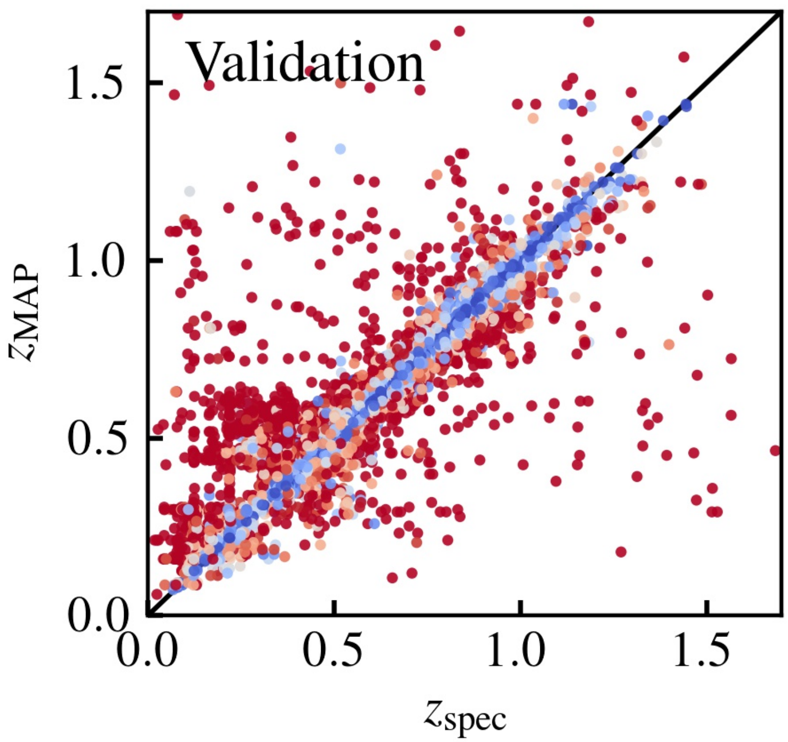
4. Redshift PDFs are more compact/precise

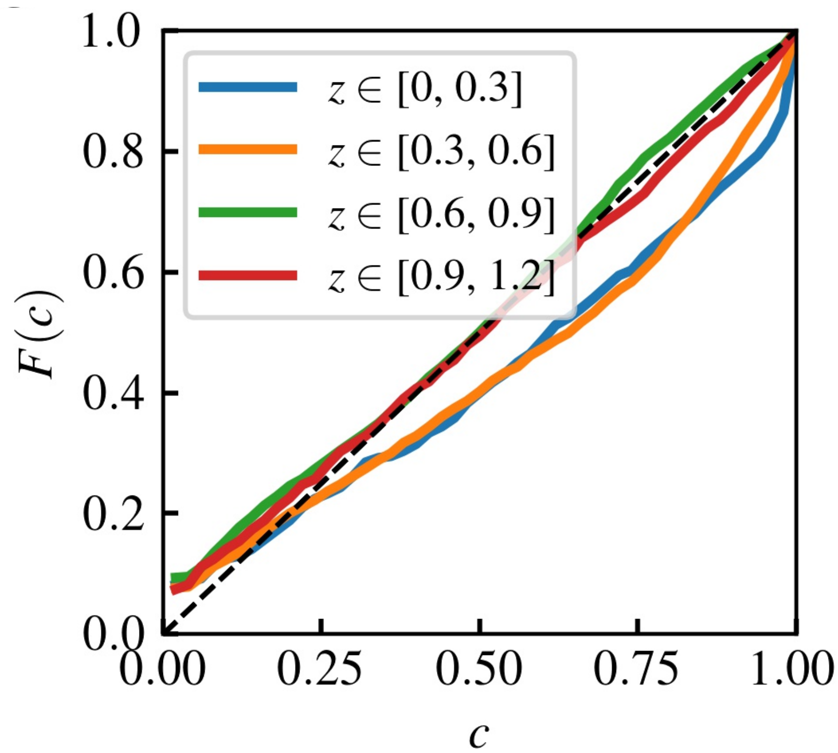
1. Cannot eliminate bias without SED corrections or variance



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

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

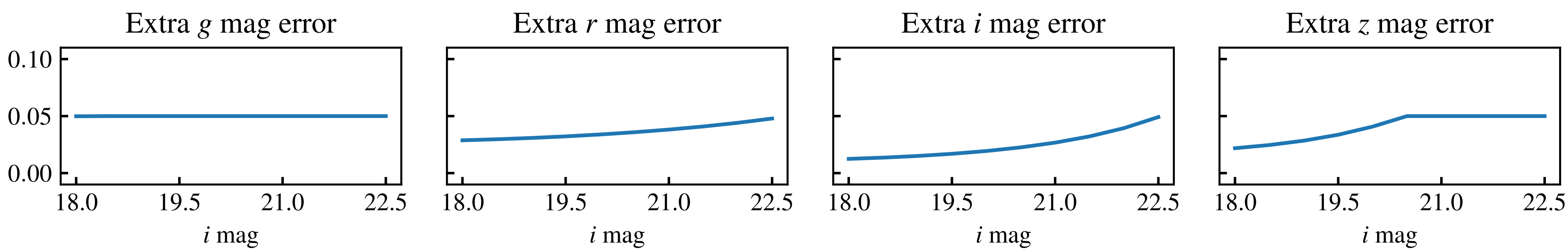




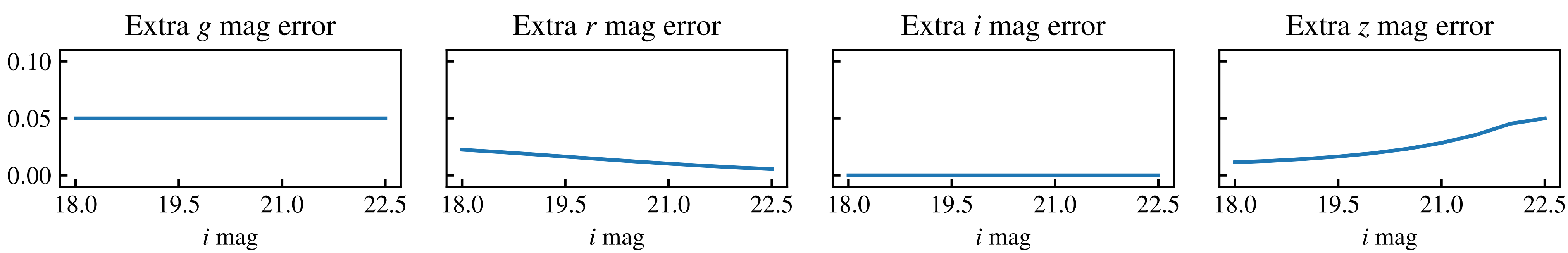
HM: 2 interpolated SEDs with SED variance & extra noise

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

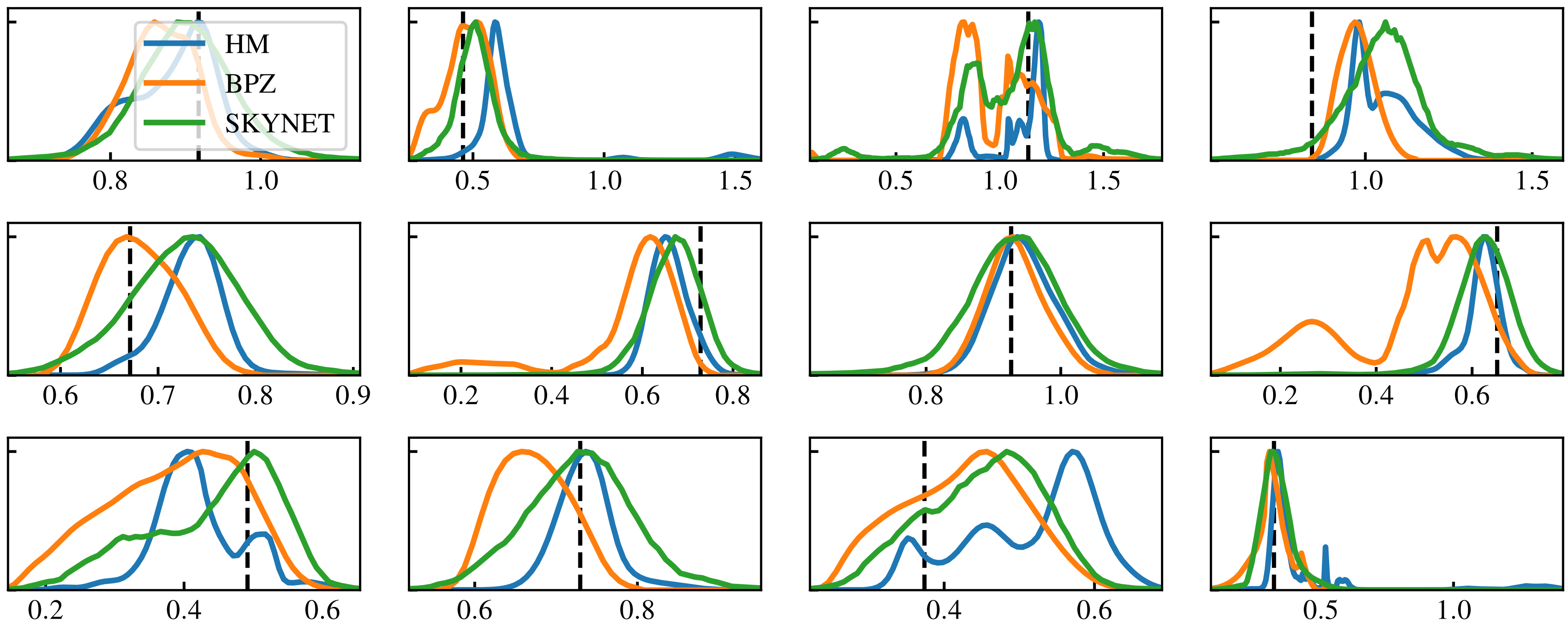
HM: simple prior, 2 interpolated SEDs, with SED corrections, magerr corrections



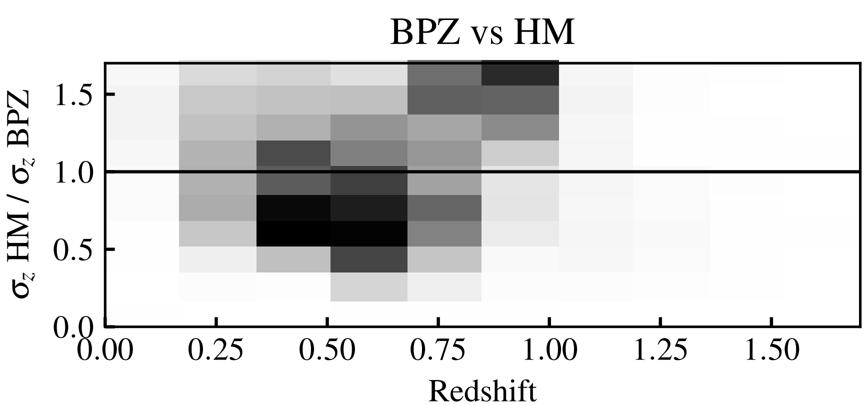
HM: simple prior, 2 interpolated SEDs, with SED corrections, variance, magerr corrections

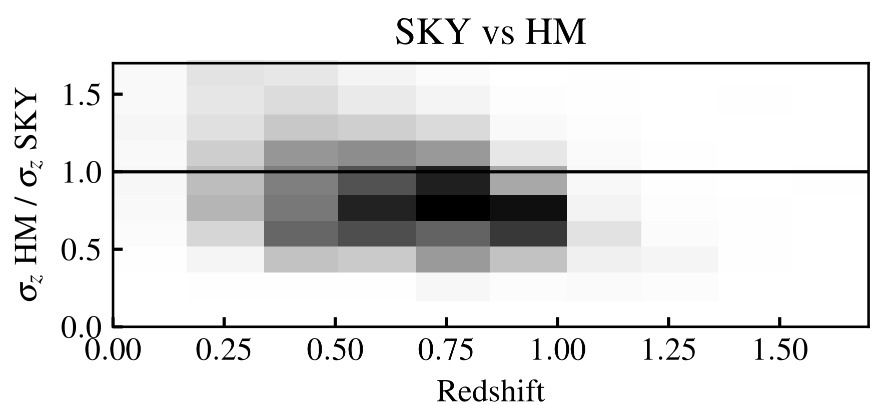


4. Redshift PDFs are more compact/precise



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



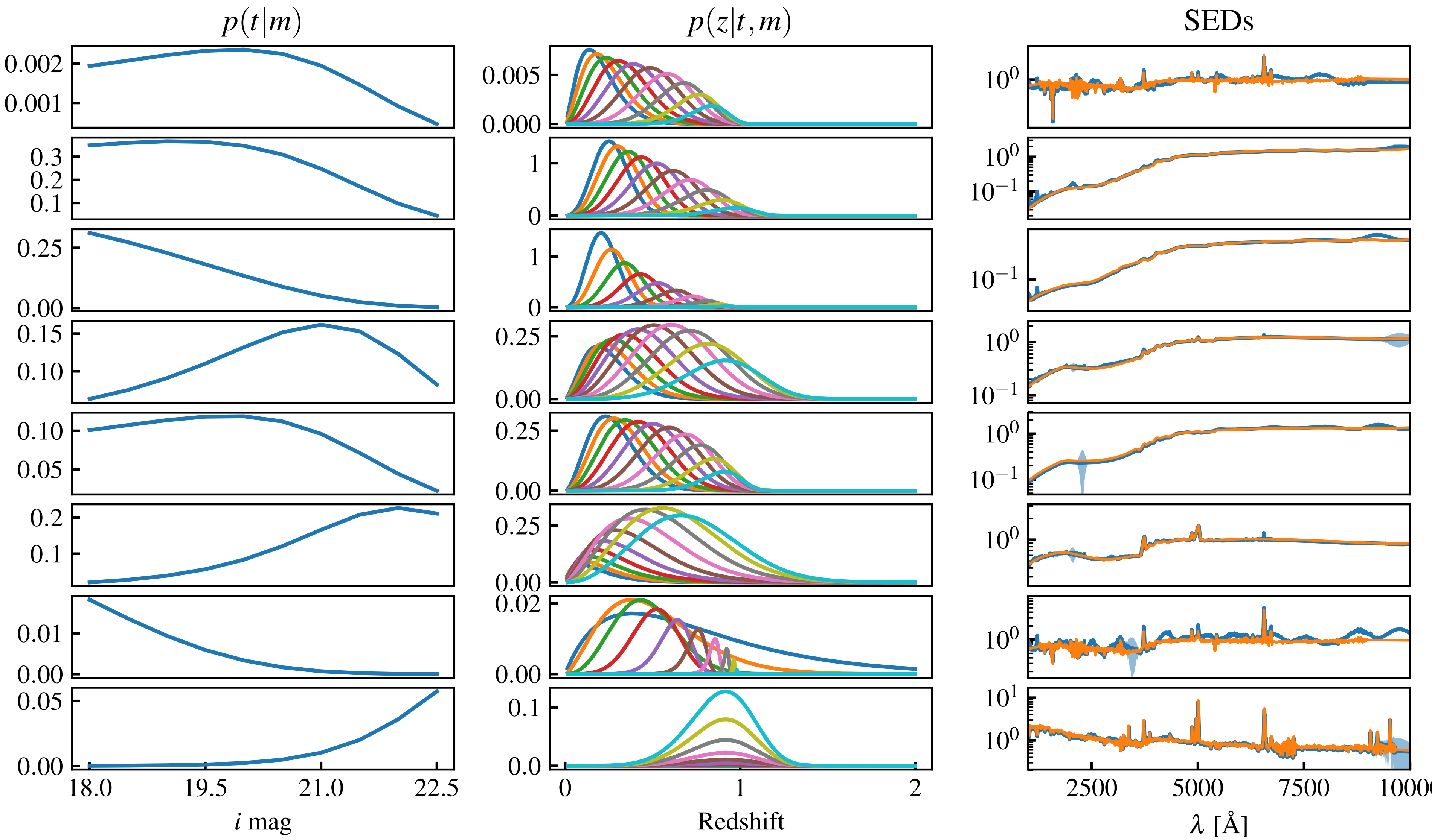


Findings (continued)

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 helps

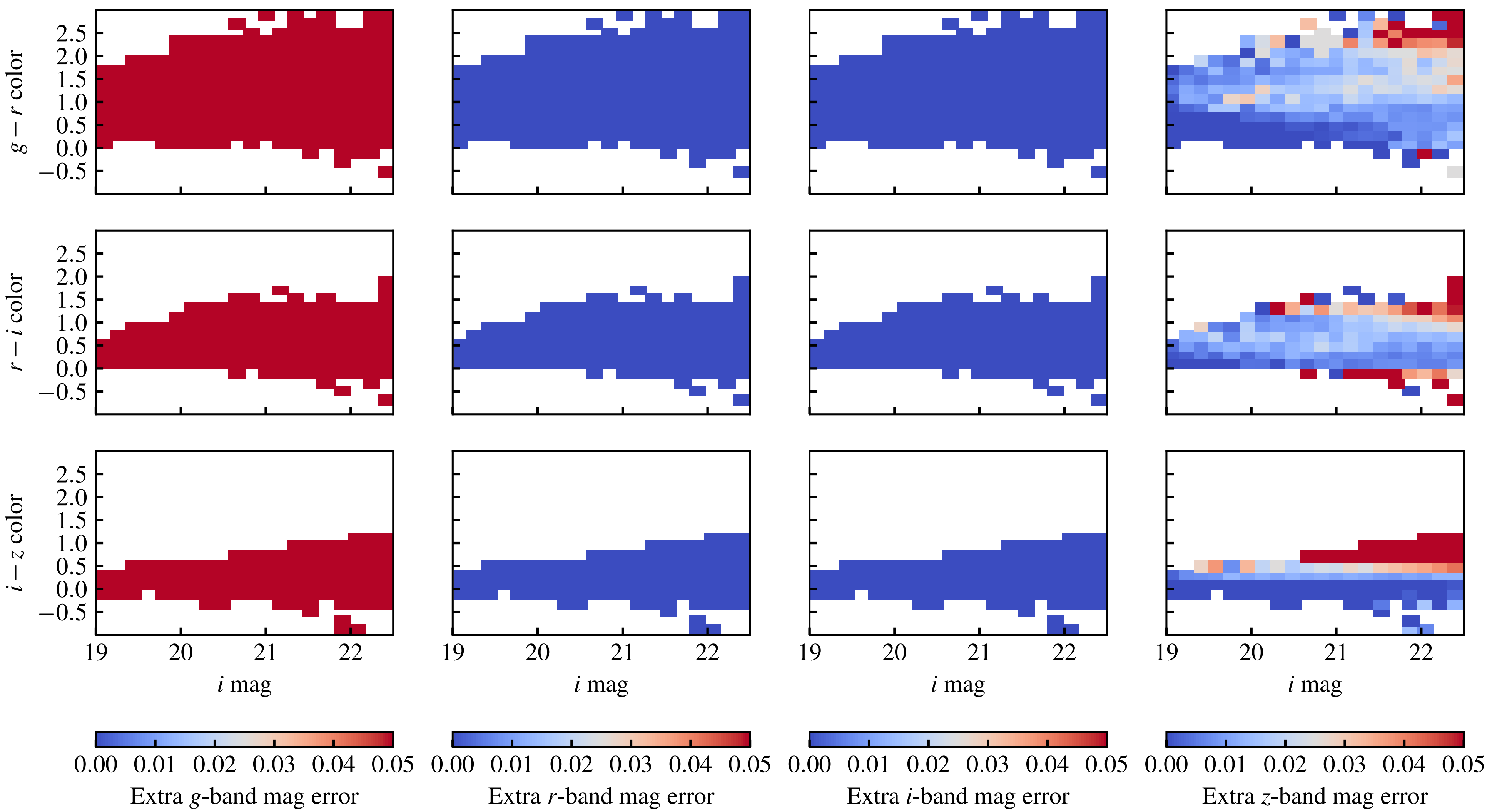
Example of SEDs and priors (top 8)

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



Example of NN noise corrections

HM: simple prior, 4 interpolated SEDs, with SED corrections, variance, magerr corrections (NNs)



Conclusions

Imaging surveys require exquisite photo-z's

Standard methods lack ***flexibility*** *or* ***interpretability***
We have the technology+power to do it right

Hierarchical SED models *deliver*
accurate + accurate + interpretable redshifts probabilities,
and interpretable physical model + data recalibration

FUTURE*: improved models + semi-supervised learning*
+ combine with N(z) inference