Forward, causal modeling of galaxy photometry

Joint self-calibration of SEDs and broadband photometry

Boris Leistedt — @ixkael, www.ixkael.com
NASA Einstein Fellow, New York University
precision (more data) -> accuracy (better methods)
Spectroscopic vs. photometric surveys

Flux vs. Wavelength [Å]

redshift
statistics-limited $\rightarrow$ systematics-limited

SDSS
$10^7$ galaxies
$10^6$ quasars

DES/KIDS
$volume \times 10$
$10^8$ galaxies
$10^6-7$ quasars

LSST
$volume \times 1000$
$10^9$ galaxies
$10^7$ quasars

SDSS
$10^7$ galaxies
$10^6$ quasars

DES/KIDS
$volume \times 10$
$10^8$ galaxies
$10^6-7$ quasars

LSST
$volume \times 1000$
$10^9$ 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

using knowledge of observed or synthetic SEDs, bandpasses, etc
<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>template fitting</strong></td>
<td>Fitting SEDs to photometry using <strong>likelihood function</strong>.</td>
</tr>
<tr>
<td></td>
<td><strong>Requires calibrated SEDs/priors &amp; unbiased data.</strong></td>
</tr>
<tr>
<td><strong>machine learning</strong></td>
<td>Construct <strong>flexible model</strong> from spectroscopic training.</td>
</tr>
<tr>
<td></td>
<td><strong>No likelihood, built-in prior, needs representative data</strong>.</td>
</tr>
<tr>
<td><strong>clustering redshifts</strong></td>
<td>Constrain N(z) using spatial <strong>cross-correlations</strong>.</td>
</tr>
<tr>
<td></td>
<td><strong>Requires overlapping samples, bias model</strong>.</td>
</tr>
</tbody>
</table>
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

![Redshift distributions](image1)

![Magnitude distribution](image2)

![Color distribution](image3)
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

<table>
<thead>
<tr>
<th>Statistical</th>
<th>Systematic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td><strong>Model</strong></td>
</tr>
<tr>
<td>Aleatoric uncertainties</td>
<td>Data biases</td>
</tr>
<tr>
<td><em>true data noise,</em></td>
<td><em>misestimated fluxes,</em></td>
</tr>
<tr>
<td><em>flux variances,</em></td>
<td><em>zeropoints,</em></td>
</tr>
<tr>
<td><em>etc</em></td>
<td><em>variance,</em></td>
</tr>
<tr>
<td>Epistemic uncertainties</td>
<td>Model biases</td>
</tr>
<tr>
<td><em>unmodeled SED effects,</em></td>
<td><em>miscalibrated SEDs or</em></td>
</tr>
<tr>
<td><em>variability,</em></td>
<td><em>priors ( p(z, t, \ell, \text{etc}) )</em></td>
</tr>
<tr>
<td><em>variance,</em></td>
<td></td>
</tr>
</tbody>
</table>
Full hierarchical model

Prior distributions $p(z, t, m, \ldots)$

Object properties: $z, t, m, \ldots$

Model photometry

Observed photometry

Model SEDs (corrected)

Base SED templates

Linear SED corrections

Photometric zeropoint/error biases

Number of objects $N_{\text{obj}}$

Number of templates $N_{\text{templates}}$

$\beta$

$\alpha$

$\gamma$
Full hierarchical model

Prior distributions $p(z, t, m, \ldots)$

Object properties: $z, t, m, \ldots$

Model photometry

Model SEDs (corrected)

Base SED templates

Linear SED corrections

Photometric zeropoint/error biases

$N_{\text{templates}}$

$N_{\text{obj}}$

Machine learning (GMMs)

Machine learning (NNs)

Standard SED fitting
Hierarchical model: SEDs + corrections

- **Base SEDs**: CWW library (8) + interpolated SEDs

- **Linear corrections**: NMF/PCA of CWW and PEGASE SEDs + Gaussian corrections

\[
    f_{\text{t corrected}}(\lambda) = f_{\text{t base}}(\lambda) + \sum_i \alpha_{it} f_i^{\text{correction}}(\lambda)
\]

- **SED variance constructed from corrections**

\[
    \text{Var}_t(\lambda) = \left( \sum_i \beta_{it} f_i^{\text{correction}}(\lambda) \right)^2
\]
Hierarchical model: priors

- **Factorization:** $p(z, m, t) = p(z|m, t) \, p(t|m) \, p(m)$

  - **Redshifts**
  - **Types**
  - **Magnitudes**

- **Magnitude prior:** $p(ell \ or \ m)$ uniform (in reference band)

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

  = Dirichlet prior on the simplex, with 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: \( \hat{F}_b \rightarrow \hat{F}_b \times w_b(m) \)

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

• Minimum magnitude error per band:

  Quadratic in reference magnitude: \( \sigma_{\hat{m}_b}^2 \rightarrow \max[\sigma_{\hat{m}_b}^2, w'_b(m)] \)

  General (neural network): \( \sigma_{\hat{m}_b}^2 \rightarrow \max[\sigma_{\hat{m}_b}^2, w'_b(m_1, \cdots, m_B)] \)
Hierarchical model: posterior

\[ p(\tilde{\alpha}, \tilde{\beta}, \tilde{H}|\{\hat{F}_i\}) \propto p(\tilde{\alpha}, \tilde{\beta}, \tilde{H}) \prod_{i=1}^{N_{\text{obj}}} \sum_{t=1}^{N_{\text{types}}} Q_{it}(\tilde{\alpha}, \tilde{\beta}, \tilde{H}) \]

- **Alpha**: parameters of the SEDs / flux model
- **Beta**: parameters of the data error recalibration
- **H**: parameters of the prior \( p(z, t, l) \)
- **Q_{it}**: 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!*
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

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

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

Extra $g$ mag error

Extra $r$ mag error

Extra $i$ mag error

Extra $z$ mag error

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

Extra $g$ mag error

Extra $r$ mag error

Extra $i$ mag error

Extra $z$ mag error
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)
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