

Astro Data Lab



---

# Outlier Detection in Spectroscopy & Photometry

---

Peter Melchior



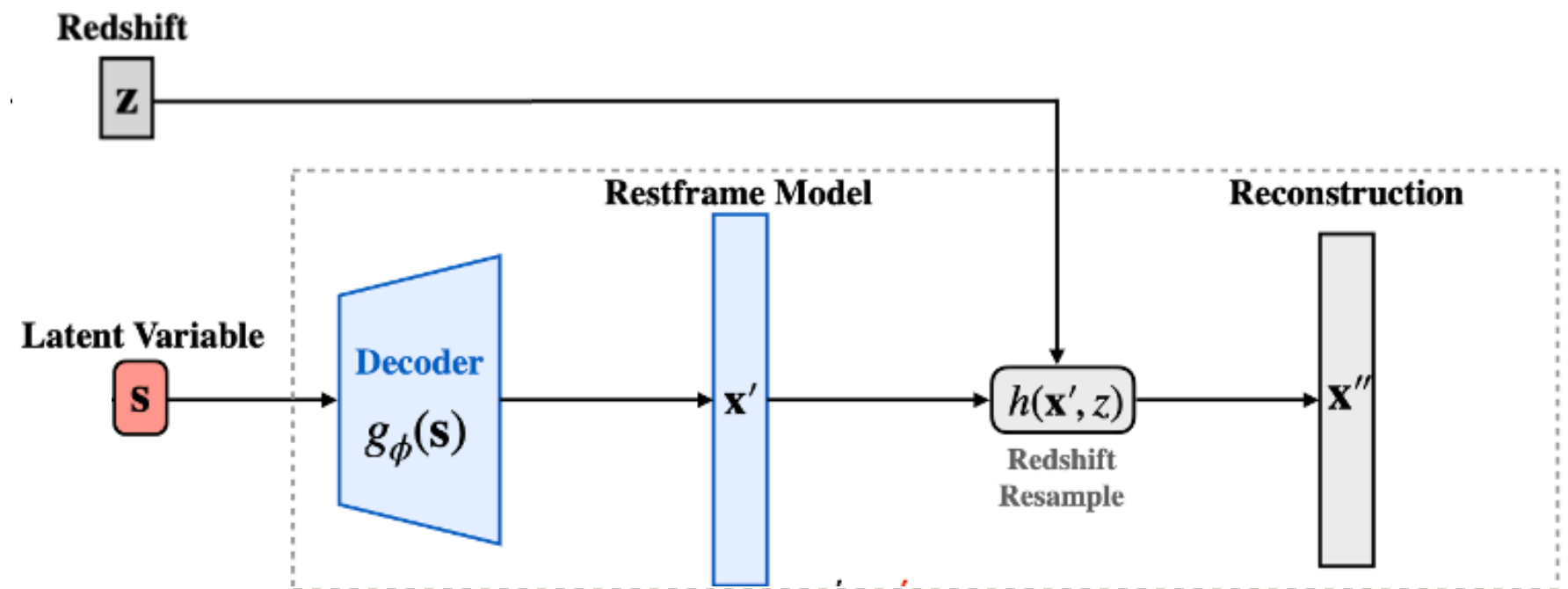
Astrophysical Sciences & CSML

---

# Spectroscopic Analysis: Spender

Code: <https://github.com/pmelchior/spender>

Melchior et al. (2023), *AJ*, 166, 74

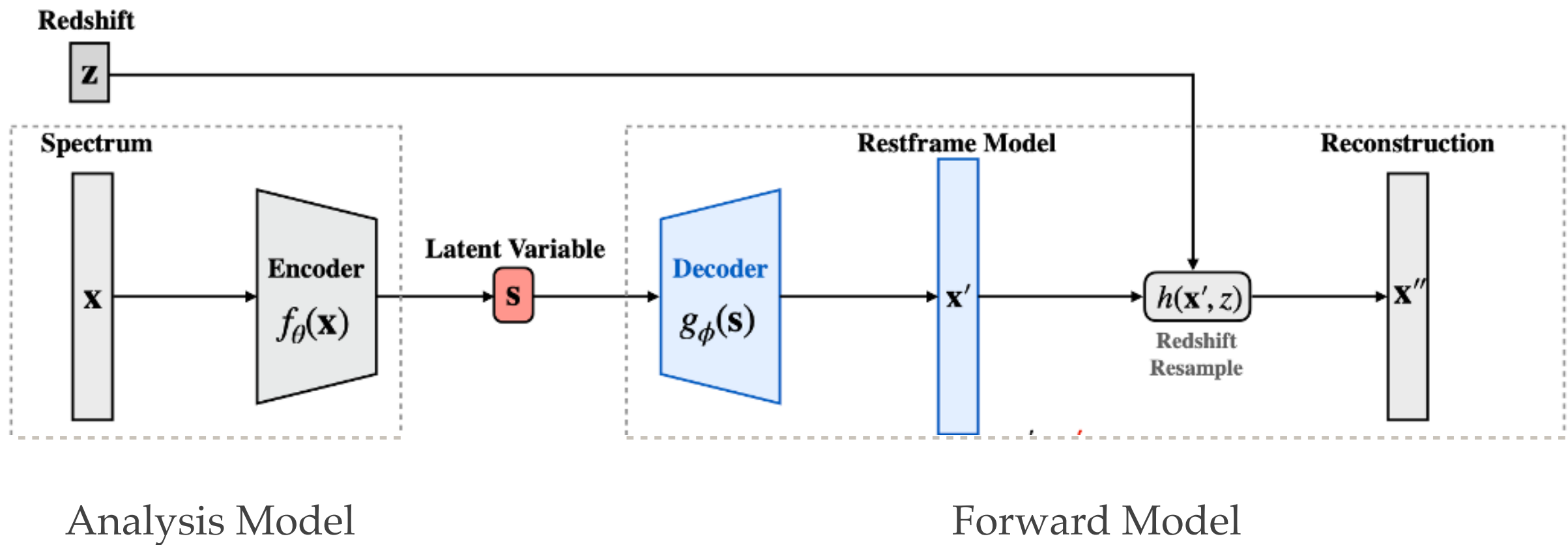


Forward Model

# Spectroscopic Analysis: Spender

Code: <https://github.com/pmelchior/spender>

Melchior et al. (2023), AJ, 166, 74



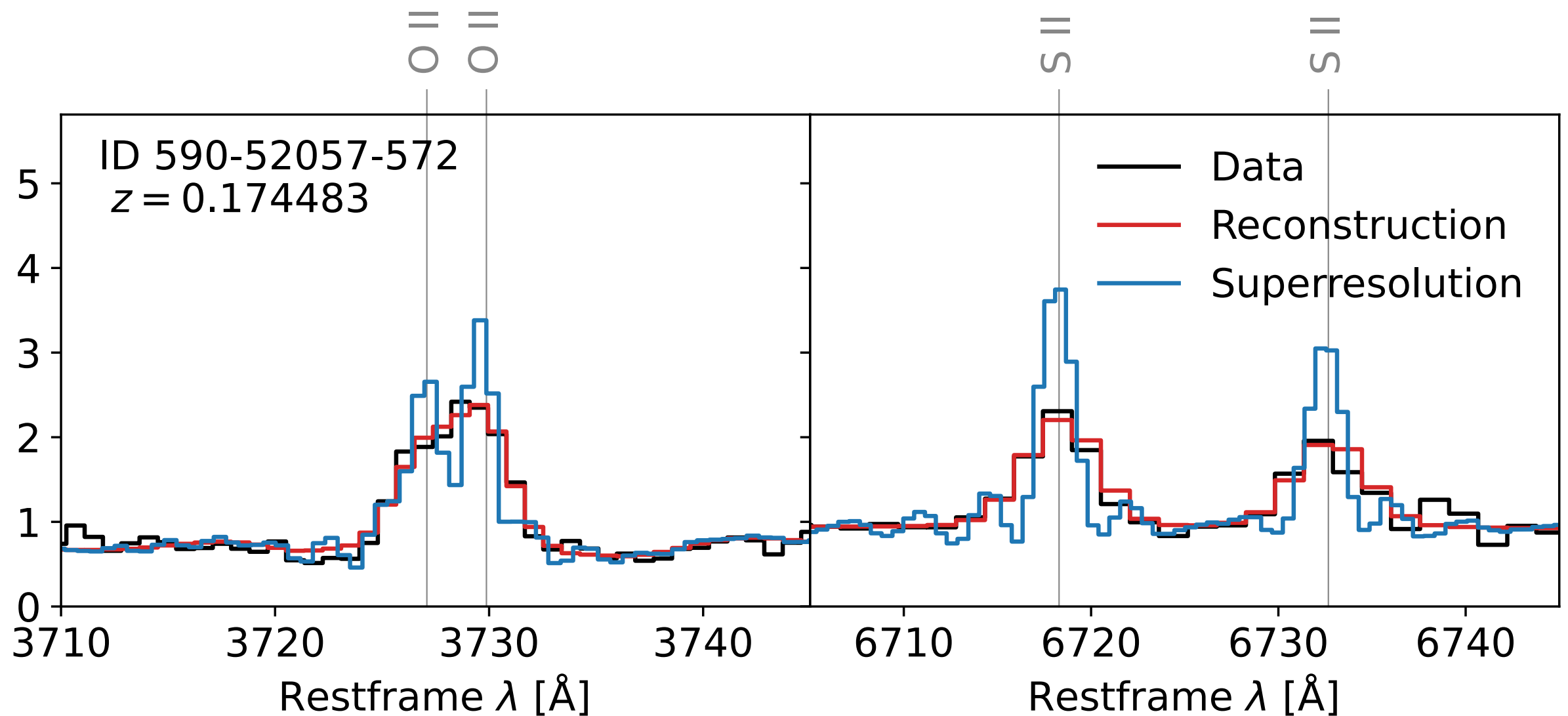




# Super-Resolution Model

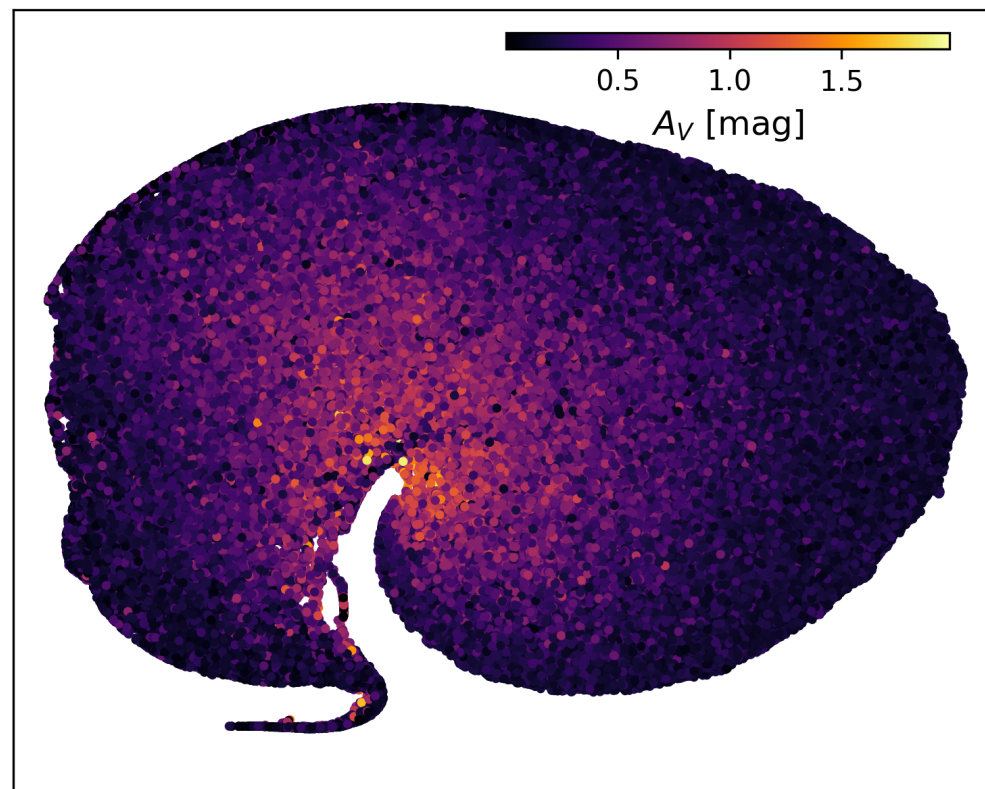
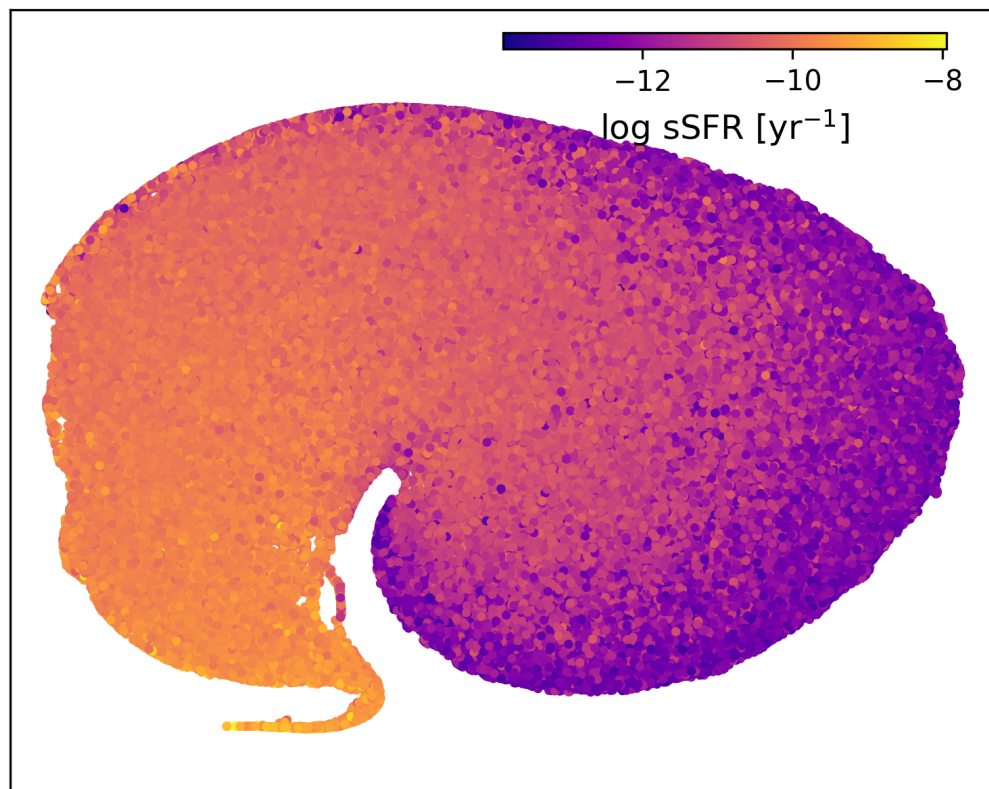
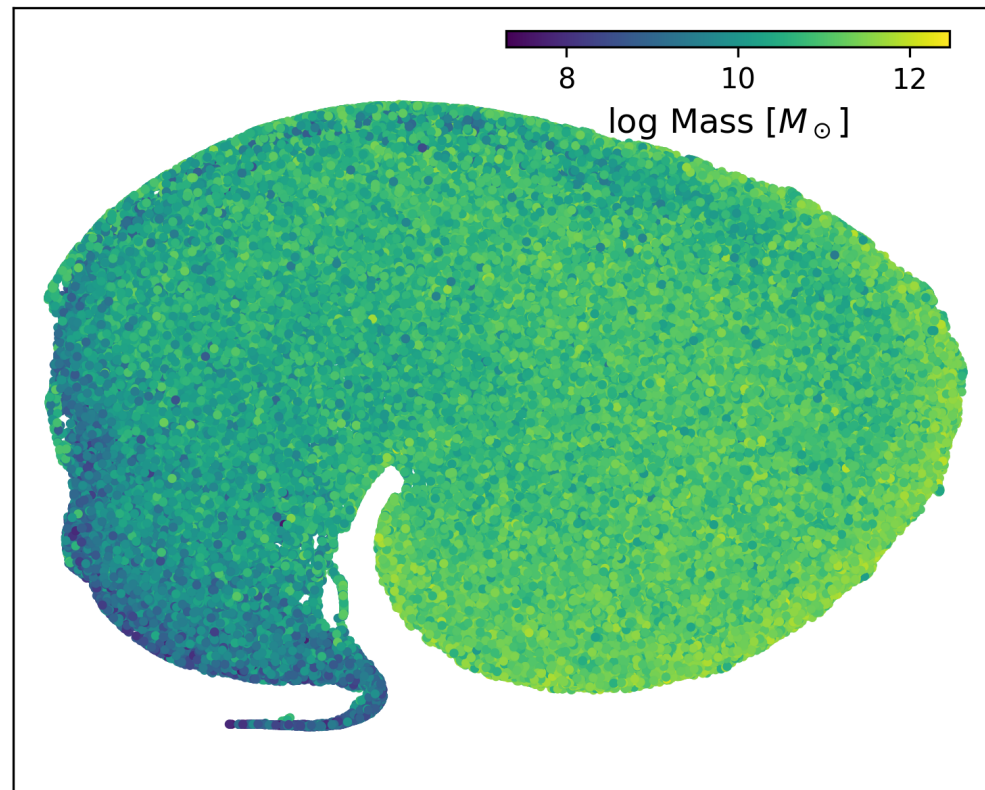
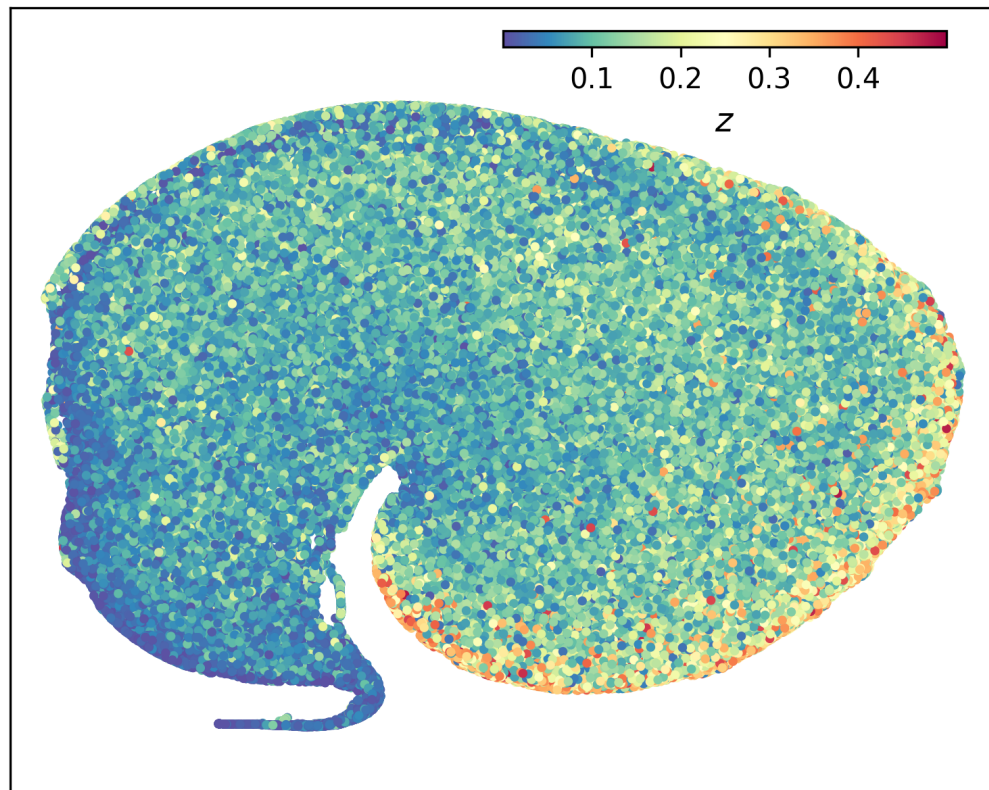
Melchior et al. (2023), AJ, 166, 74

- ❖ Internally  $\times 2$  oversampling
- ❖ Deconvolve from the (unknown) SDSS-I LSF



# The Latent Space

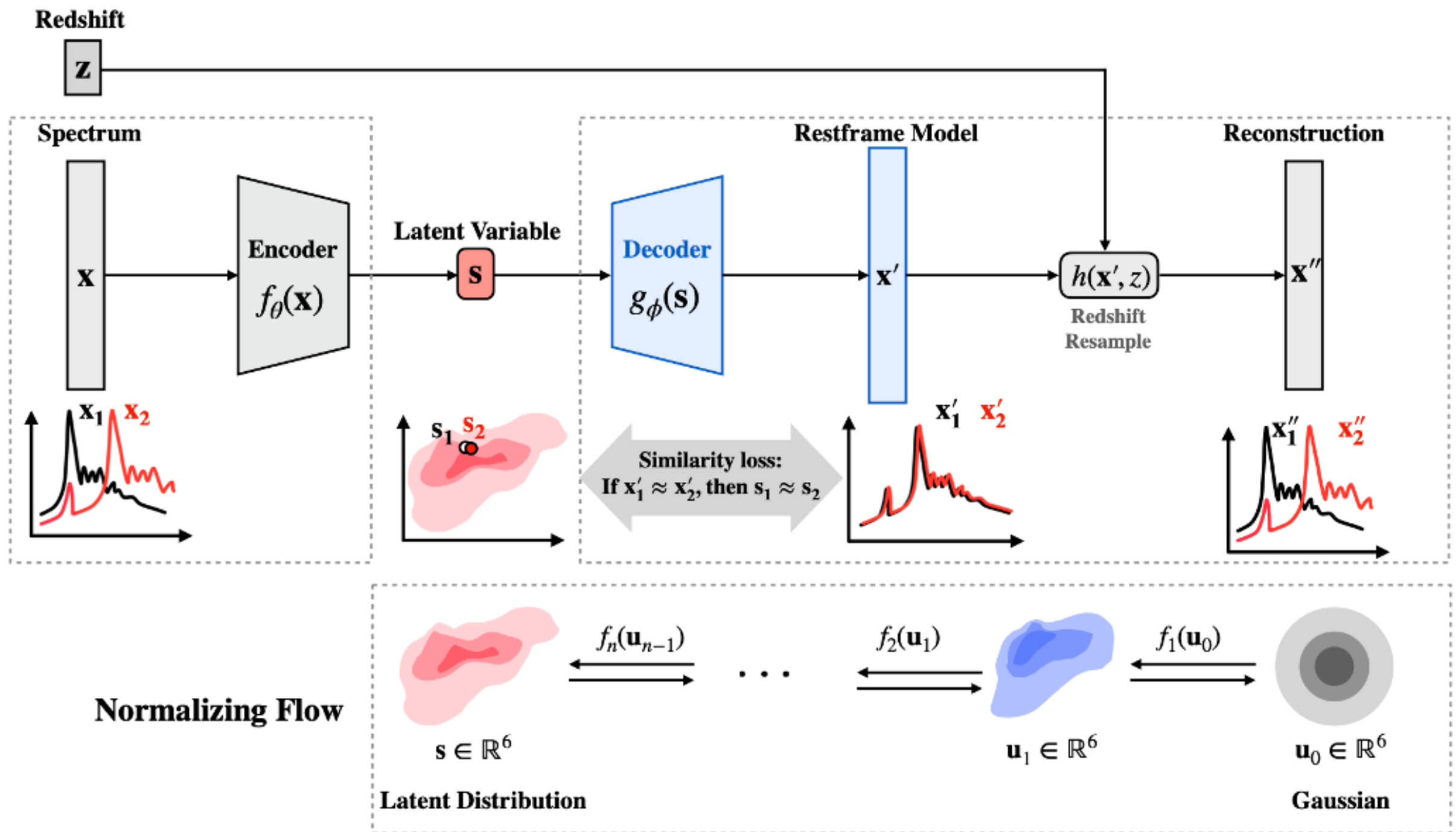
Liang et al. (2023)  
AJ, 166, 75



Yan Liang

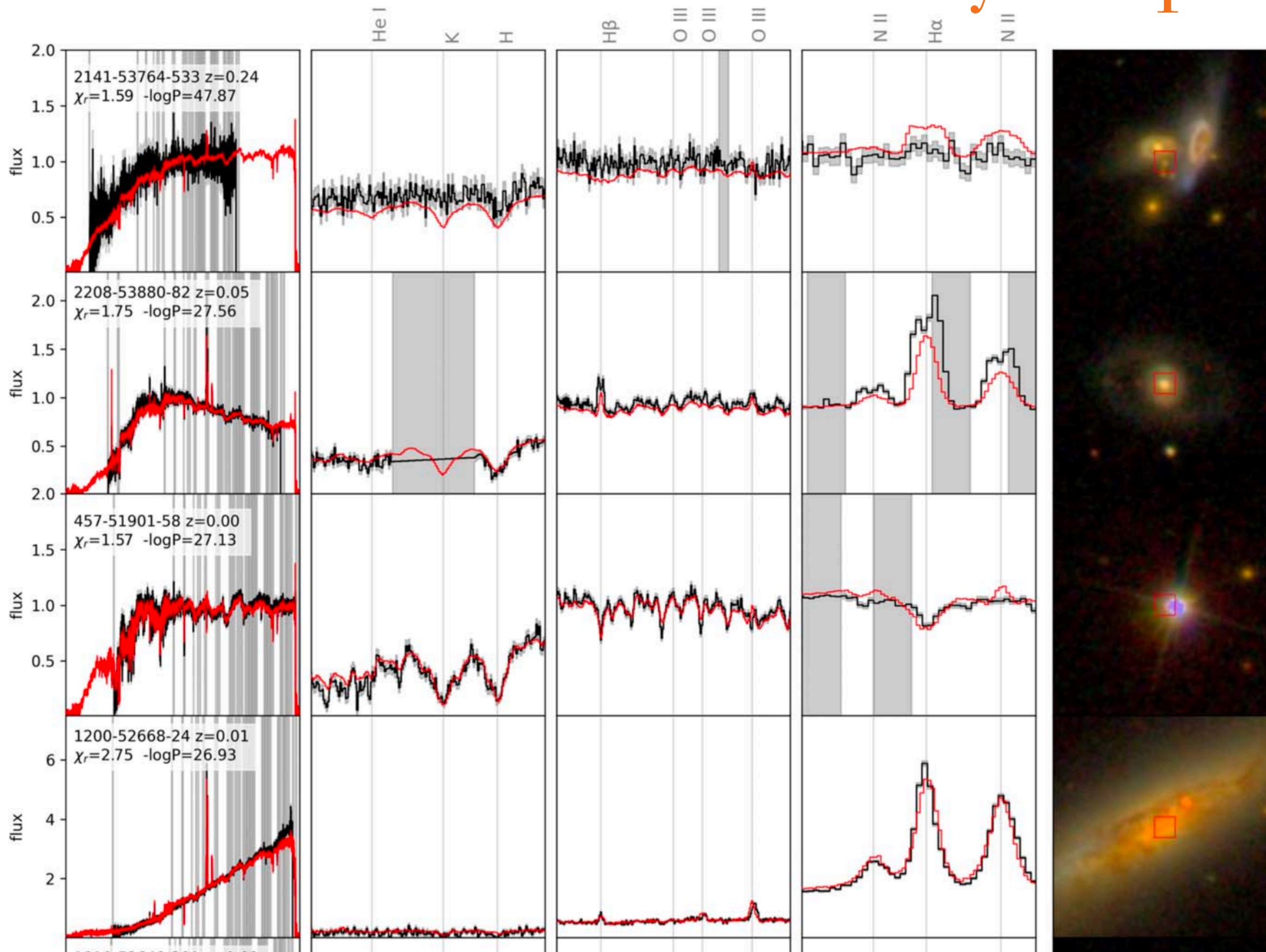
# Latent Space Distribution

Liang et al. (2023)  
AJ, 166, 75



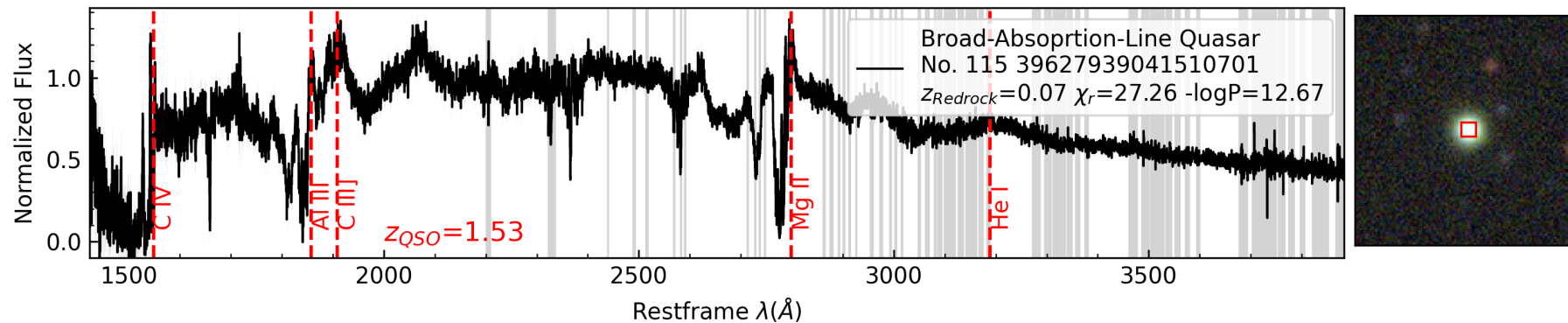
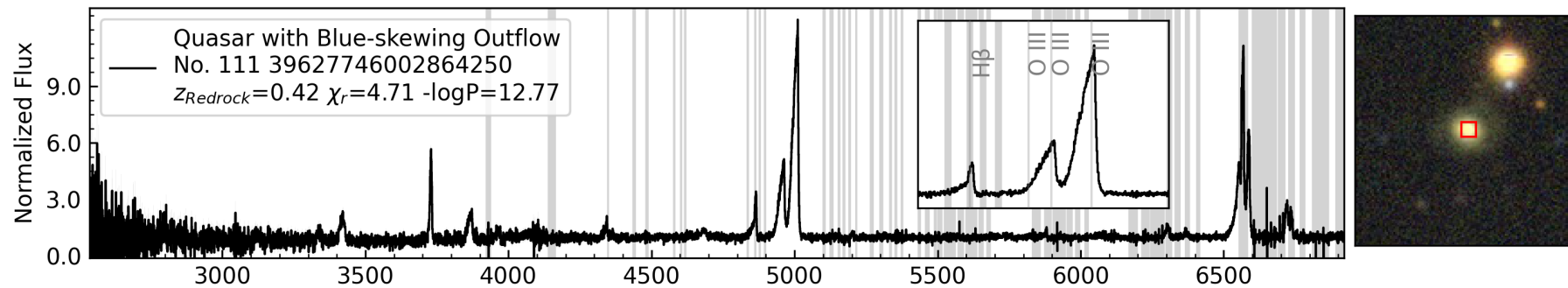
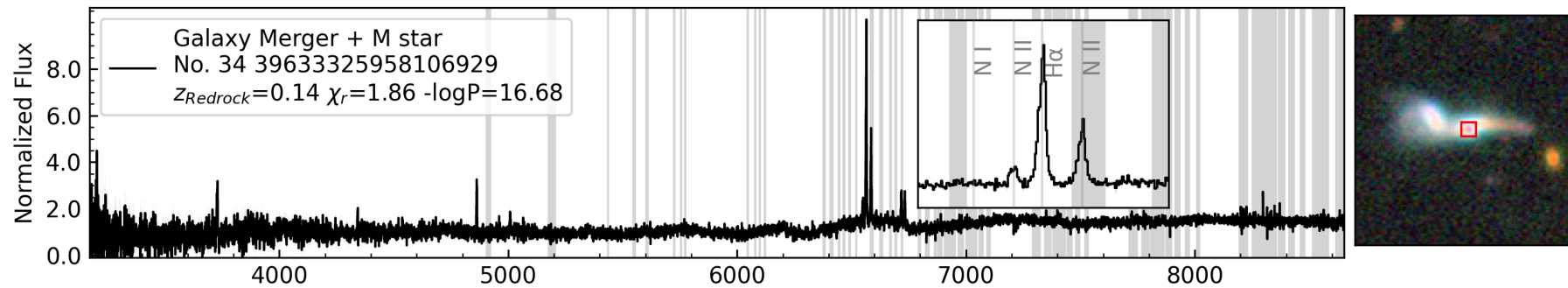
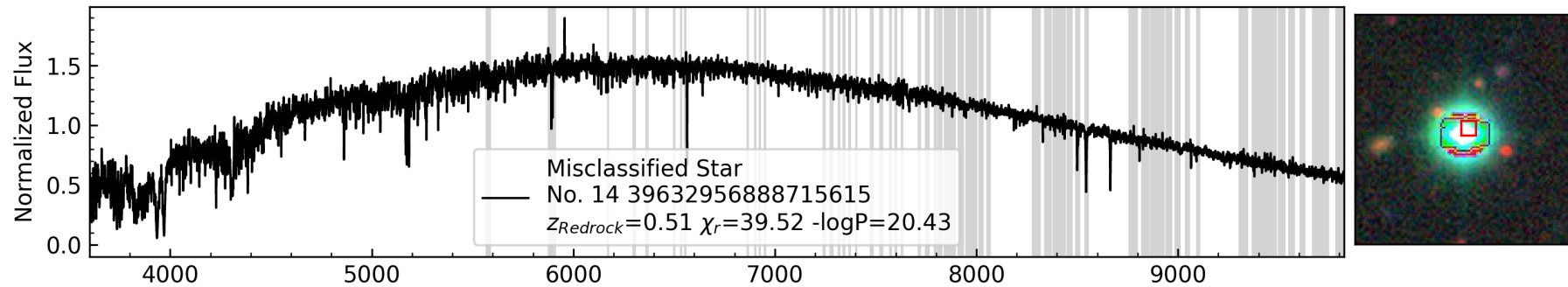
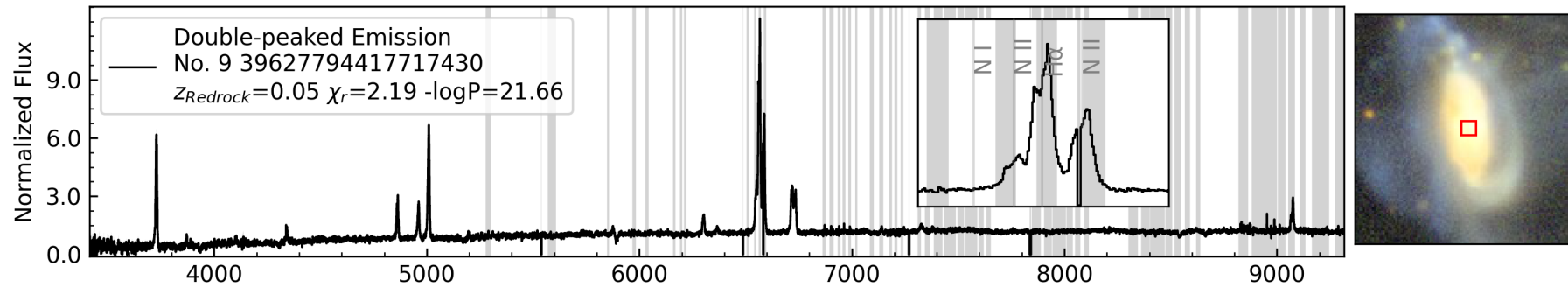
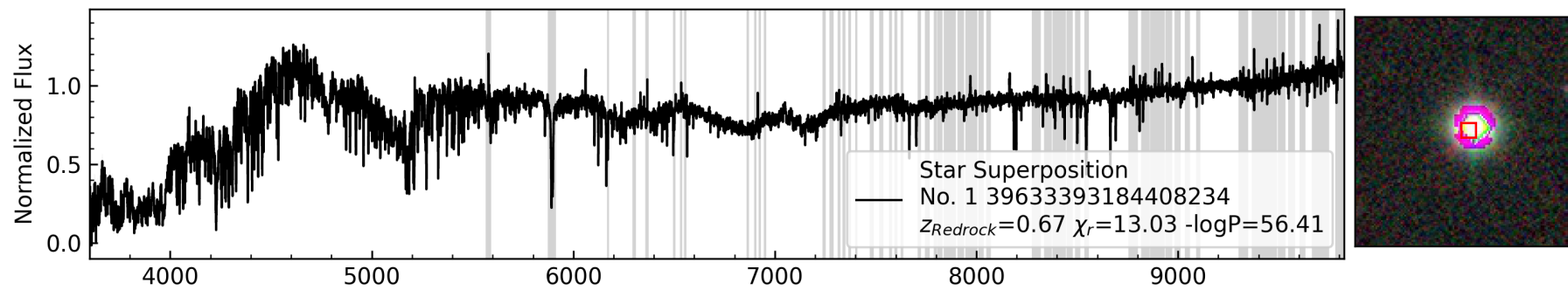


# Outliers in SDSS-I Main Galaxy Sample



# in DESI

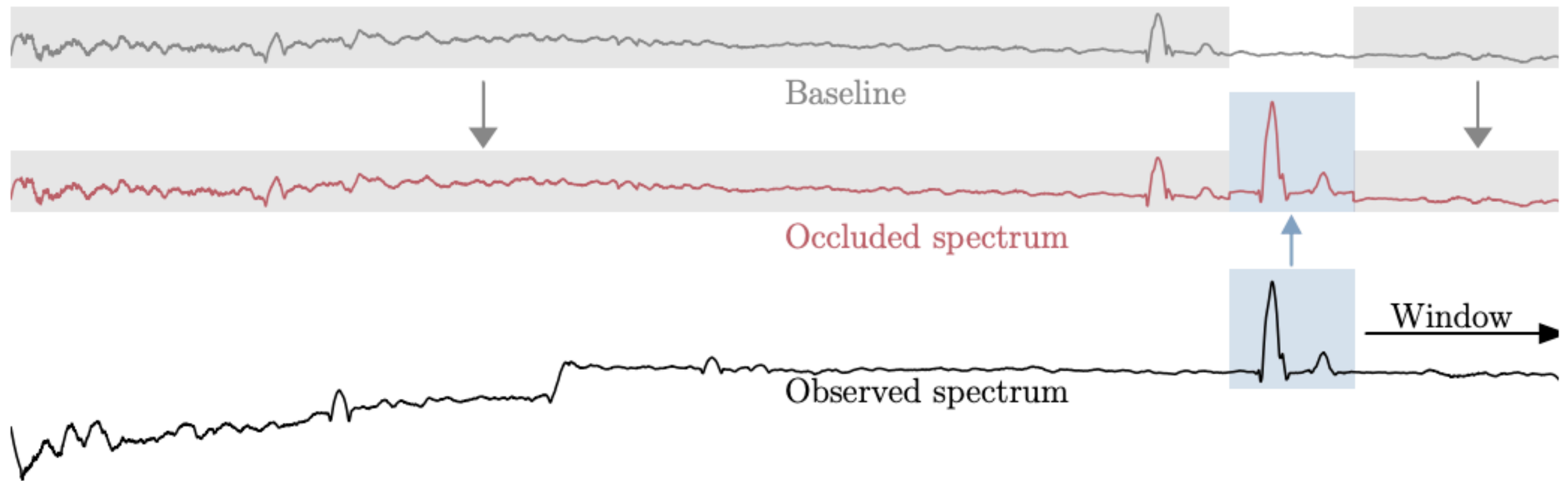
1. Stars
2. Blends
3. Velocity structure
4. weird QSOs



Liang et al. (2023)  
ApJ, 956, L6

# Feature Attribution for Outliers

Shen & Melchior (2023)  
arXiv:2310.20012

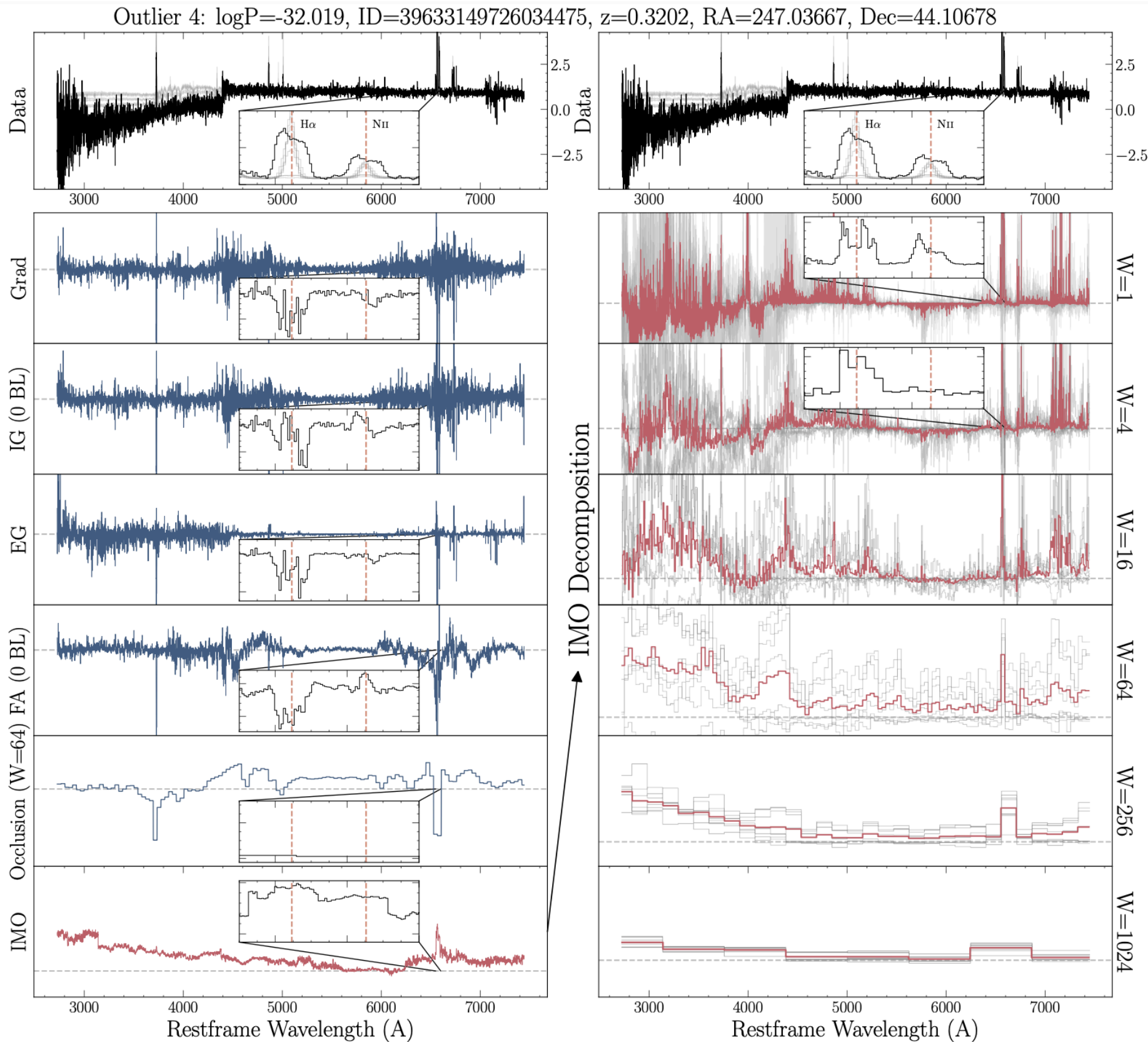


Jeff Shen



# Feature Attribution for Outliers

Shen & Melchior (2023)  
arXiv:2310.20012



Jeff Shen

# Galaxy Photometry: PROVABGS



ChangHoon Hahn

Code: <https://github.com/changhoonhahn/provabgs>

## ❖ Stellar population synthesis

$$f_{\text{csp}}(\lambda, t) = \int_{t'=0}^{t'=t} f_{\text{ssp}}(t', Z(t')) \text{SFR}(t') e^{-\tau_{\text{dust}}(t')}$$

## ❖ Expensive evaluation

## ❖ Traditionally O(h) / galaxy

## ❖ Not feasible for O(10<sup>9</sup>) galaxies!

## ❖ Normalizing flow for $p(\theta | \mathcal{D})$

**Table 1**  
Parameters of the PROVABGS SPS Model and Their Priors Used for Joint SED Modeling of DESI Photometry and Spectroscopy

Name	Description	Prior
$\log M_*$	log galaxy stellar mass	uniform over [7, 12.5]
$\beta_1, \beta_2, \beta_3, \beta_4$	NMF basis coefficients for SFH	Dirichlet prior
$f_{\text{burst}}$	fraction of total stellar mass formed in starburst event	uniform over [0, 1]
$t_{\text{burst}}$	time of starburst event	uniform over [10 Myr, 13.2 Gyr]
$\gamma_1, \gamma_2$	NMF basis coefficients for ZH	log uniform over [ $4.5 \times 10^{-5}$ , $1.5 \times 10^{-2}$ ]
$\tau_{\text{BC}}$	Birth cloud optical depth	uniform over [0, 3]
$\tau_{\text{ISM}}$	diffuse-dust optical depth	uniform over [0, 3]
$n_{\text{dust}}$	Calzetti (2001) dust index	uniform over [-2, 1]
$f_{\text{fiber}}$	spectrum fiber-aperture effect normalization	Gaussian $\mathcal{N}(\hat{f}_r^{\text{fiber}}, \frac{f_r^{\text{fiber}}}{f_r} \sigma_r)$



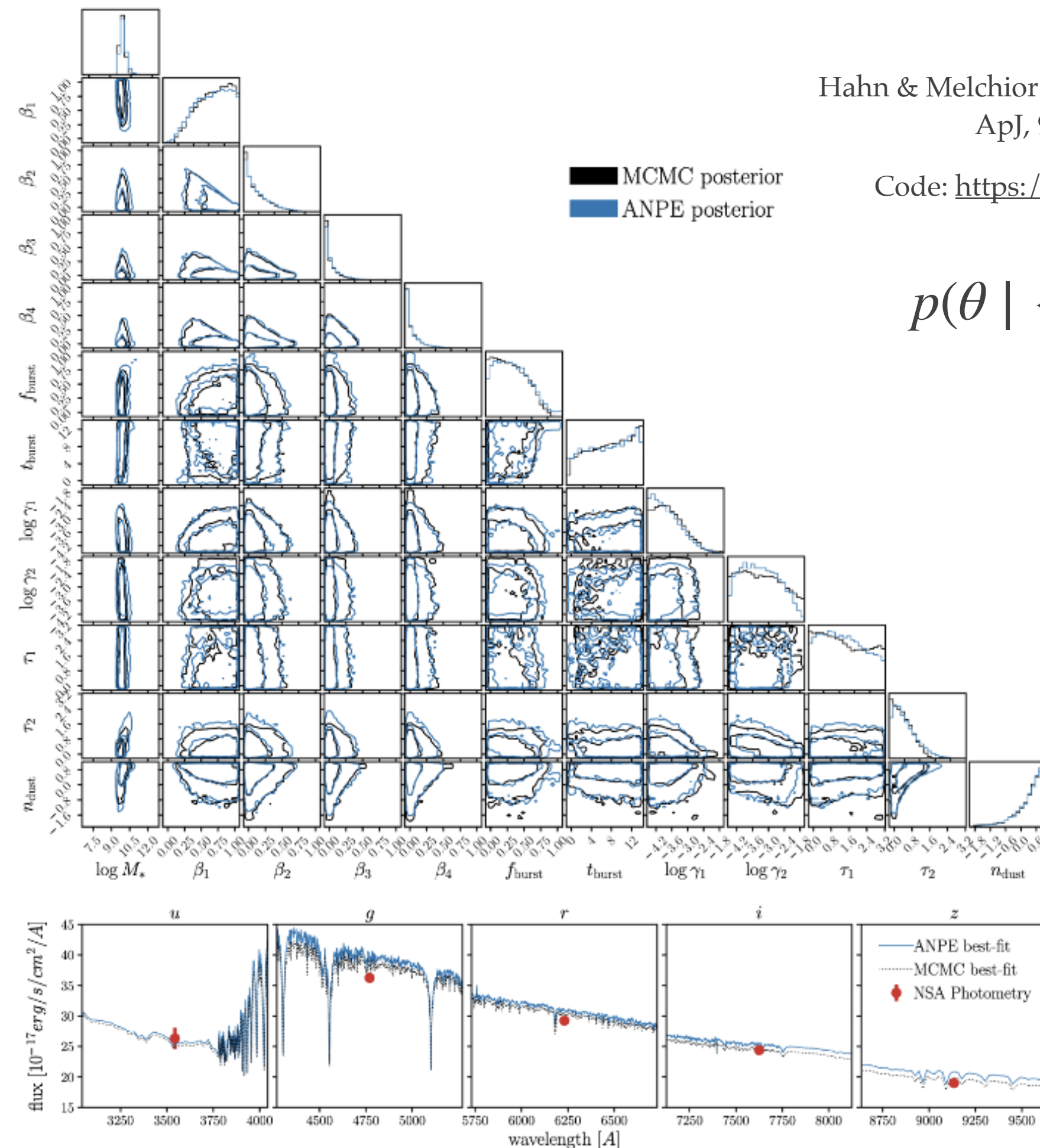
Hahn & Melchior (2022)  
ApJ, 938, 11

# SEDFlow

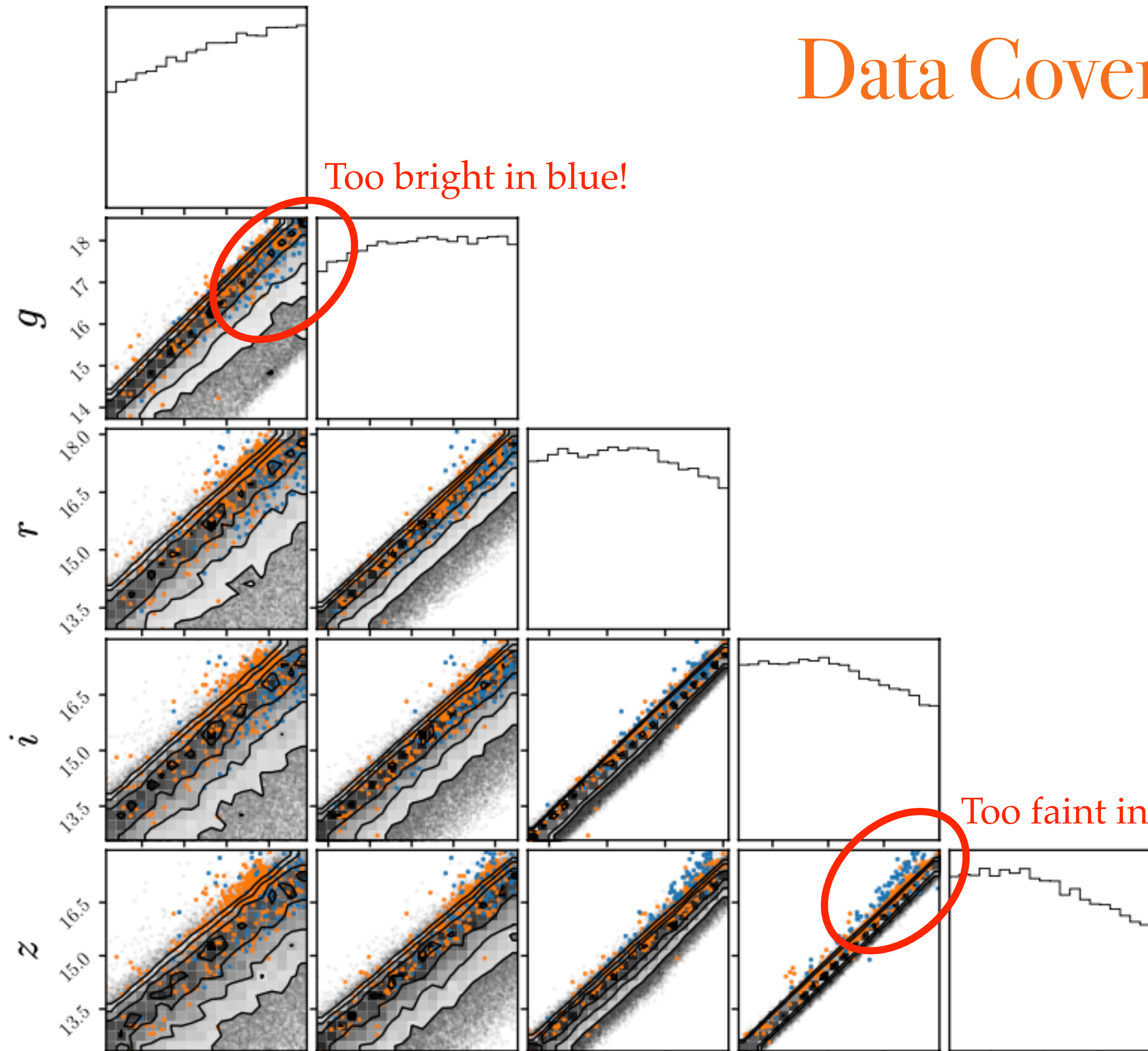
Code: <https://github.com/changhoonhahn/SEDflow>

$$p(\theta | \{g, r, i, z\}, \{\sigma_g, \sigma_r, \sigma_i, \sigma_z\})$$

- ❖ O(1s) instead of O(100h)!
- ❖ Done with broadband photometry
- ❖ Application to spectra coming
- ❖ Built explicit model  $p(\{\sigma_x\} | \{\text{mag}_x\})$

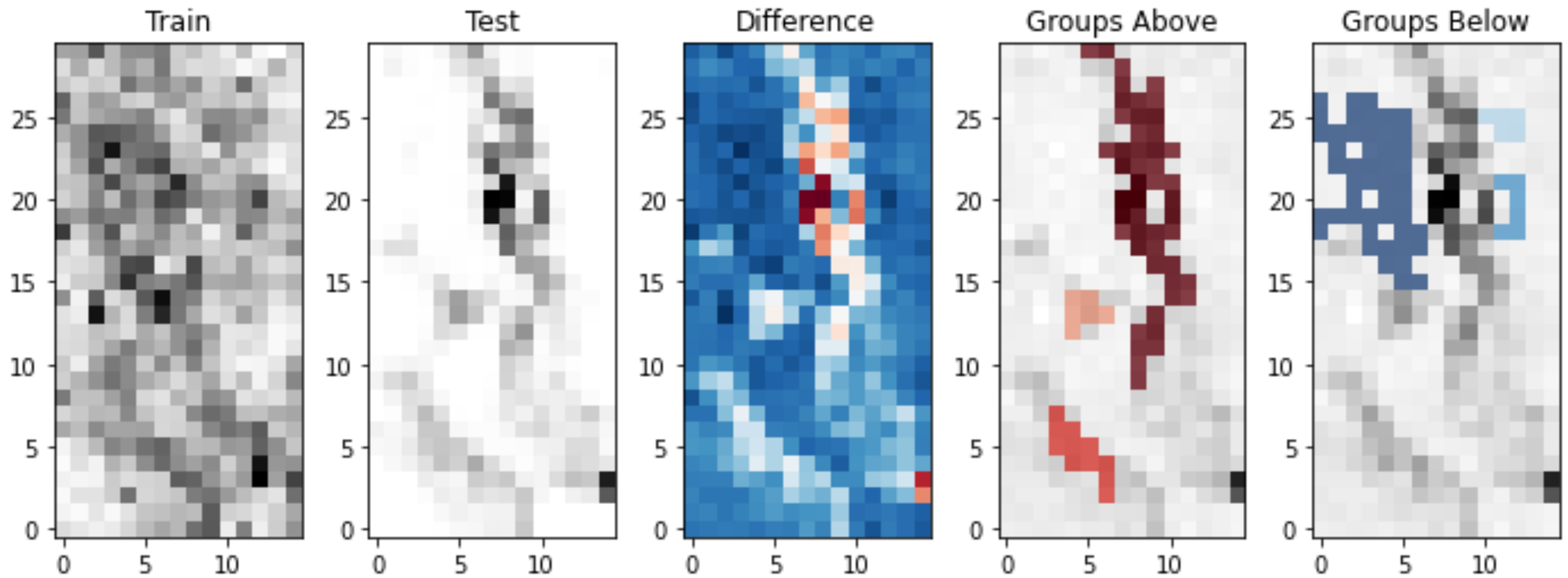


# Data Coverage



# Outliers in Photometry: Shroom

Code: <https://github.com/astro-data-lab/shroom>



# What Outliers are Not

---

- ❖ **A Class:** Many ways to be weird, no point in enforcing class

# What Outliers are Not

---

- ❖ **A Class:** Many ways to be weird, no point in enforcing class
- ❖ **Quantifiable:**
  1. Examples from tails of distribution, properties hard to learn
  2. Unexpected, completeness is ill-defined

# What Outliers are Not

---

- ❖ **A Class:** Many ways to be weird, no point in enforcing class
- ❖ **Quantifiable:**
  1. Examples from tails of distribution, properties hard to learn
  2. Unexpected, completeness is ill-defined
- ❖ **Self-evident:** Detection hard, interpretation harder

# What Outliers are Not

---

- ❖ **A Class:** Many ways to be weird, no point in enforcing class
- ❖ **Quantifiable:**
  1. Examples from tails of distribution, properties hard to learn
  2. Unexpected, completeness is ill-defined
- ❖ **Self-evident:** Detection hard, interpretation harder
- ❖ **Boring:** (ideally)

Success of outlier detection measured by our surprise





Source: Florida Center for  
Instructional Technology



# Princeton Astro Data Lab

## Funding



ChangHoon Hahn



Charlotte Ward



Benjamin Remy



Christian Jespersen



Yan Liang



Jiaxuan Li



Matt Sampson



Jeff Shen



Jared Siegel

