Big Data Inference

Combining Hierarchical Bayes and Machine Learning to Improve Photometric Redshifts

---

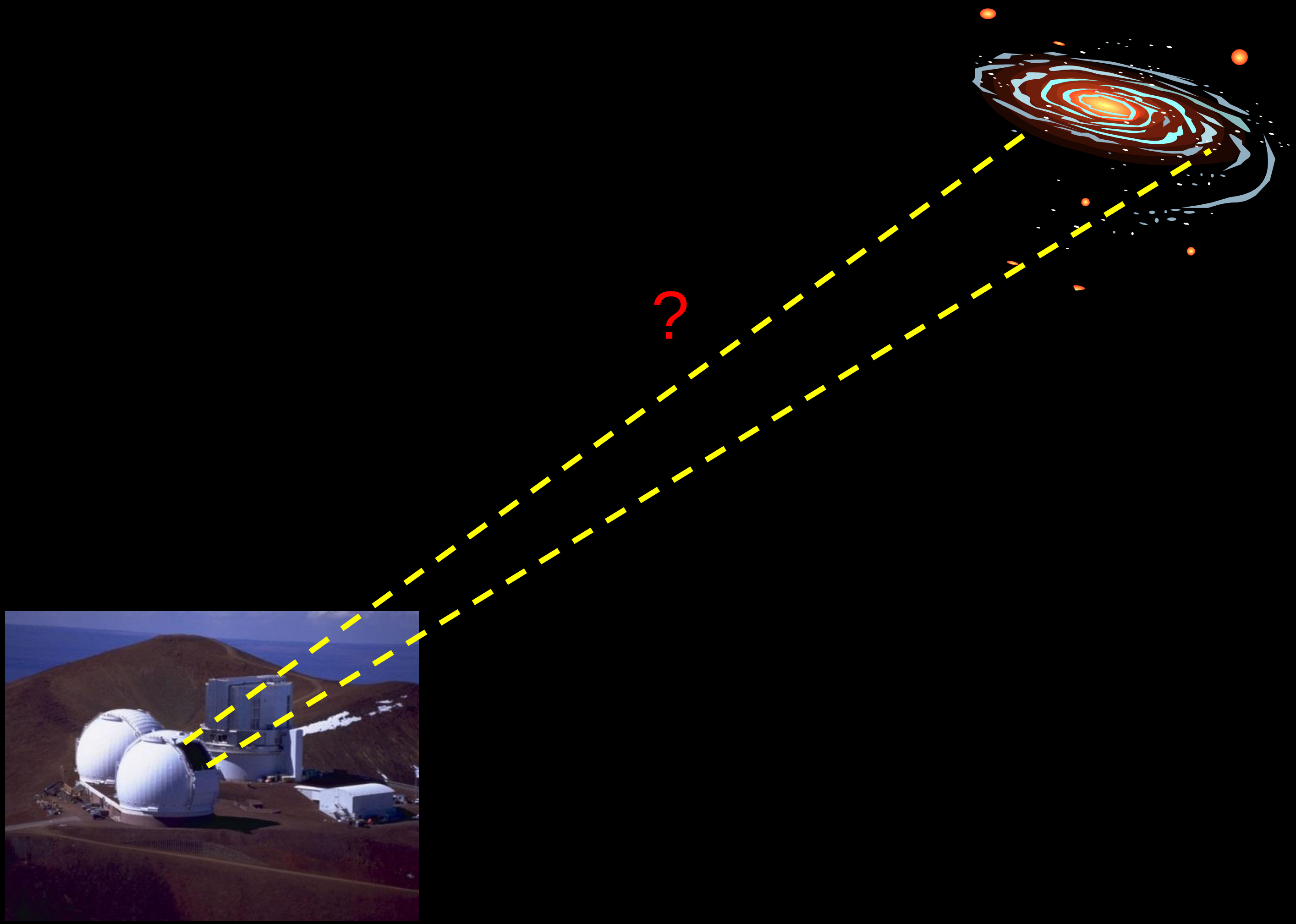
Josh Speagle[1,†]
jspeagle@cfa.harvard.edu

In collaboration with:
Boris Leistedt[2], Alexie Leauthaud[3], Daniel Eisenstein[1], Kevin Bundy[3,4],
Peter Capak[5], Ben Hoyle[6], Daniel Masters[5], Daniel Mortlock[7], and Hiranya Peiris[8,9]

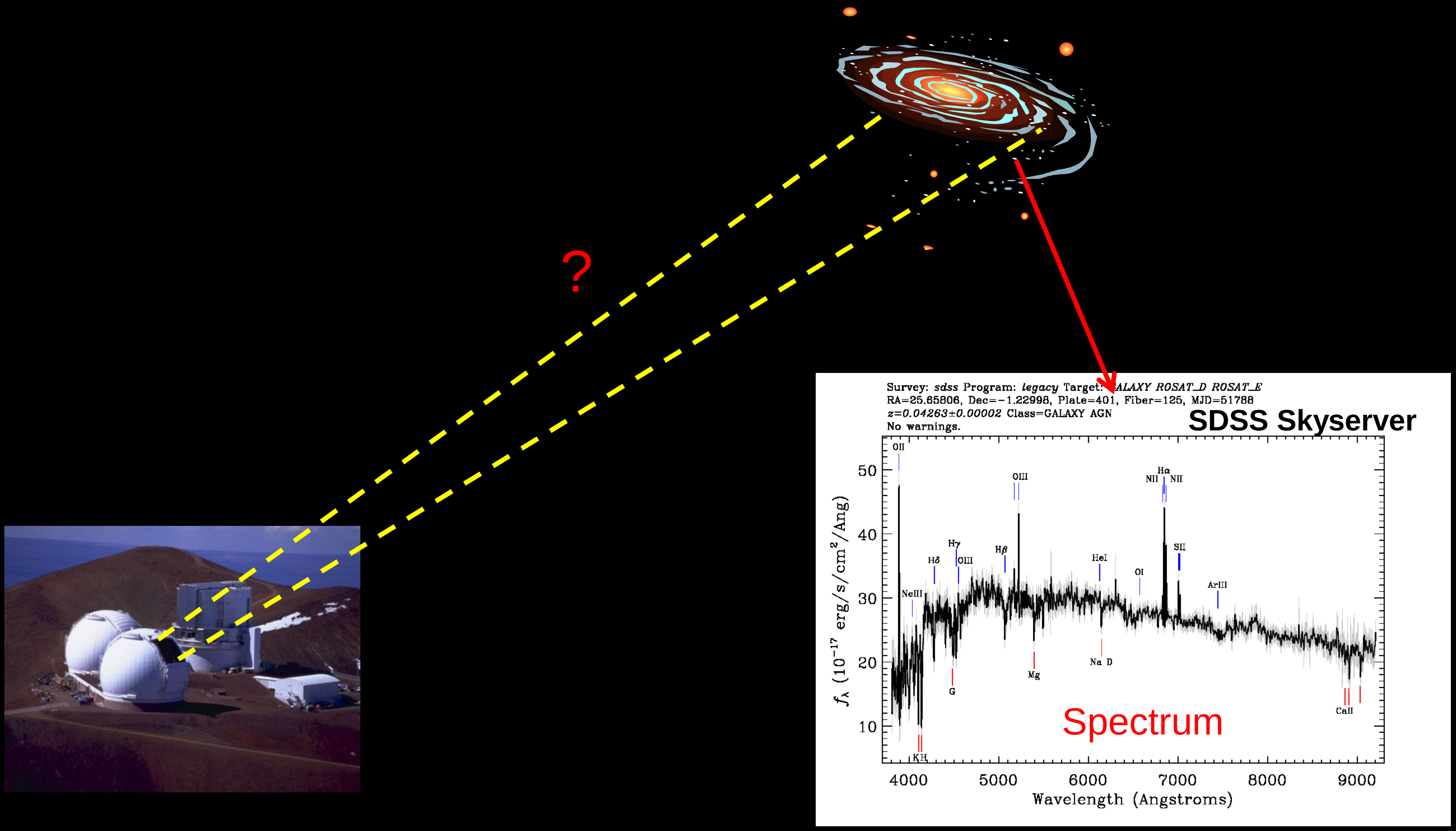
[†]NSF Graduate Research Fellow, [1]Harvard University, [2]NYU, [3]UC Santa Cruz, [4]UC Observatories, [5]IPAC, [6]LMU Munich, [7]Imperial College London, [8]University College London, [9]Oskar Klein Centre for Cosmoparticle Physics

What are Photometric Redshifts?

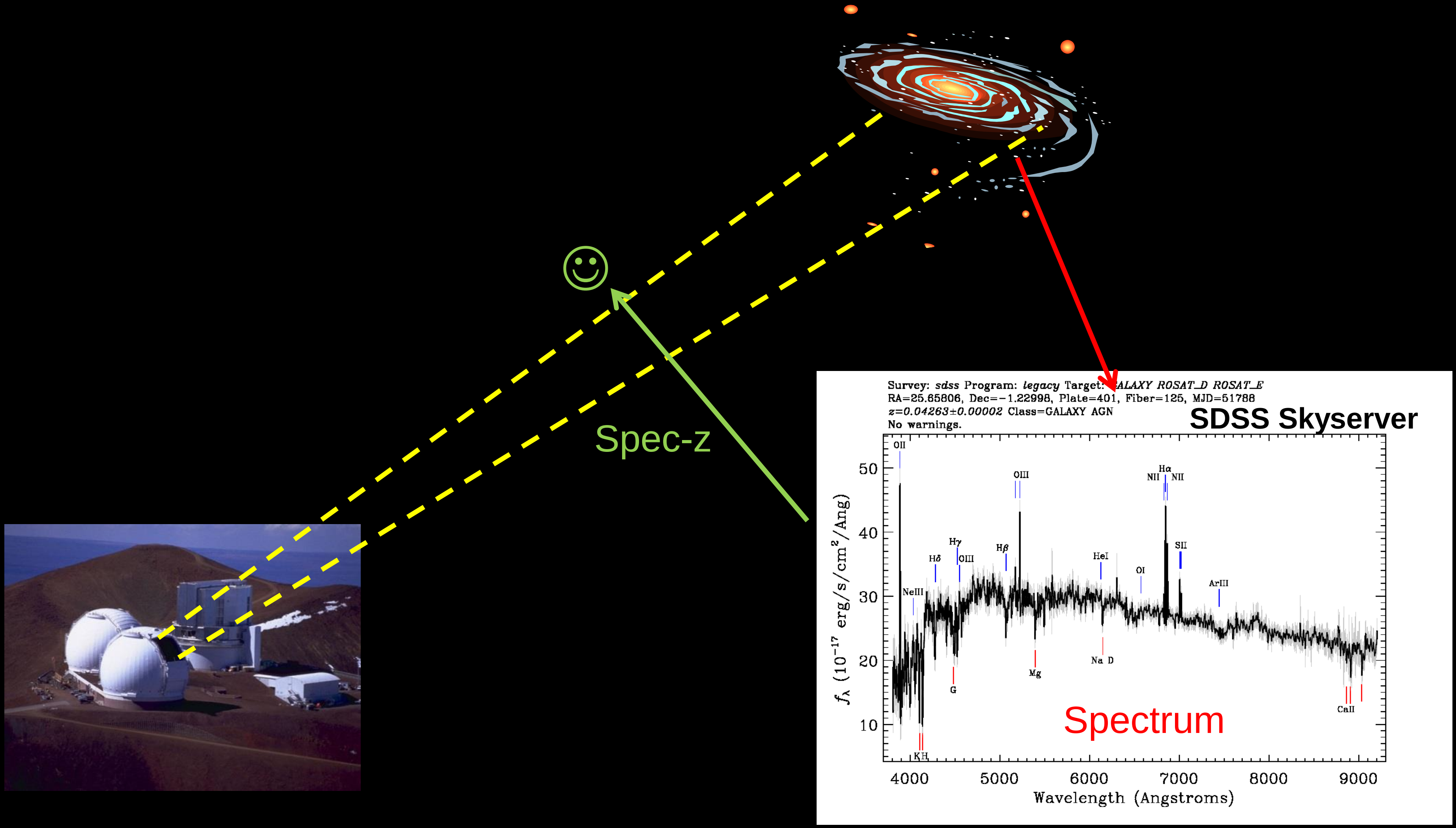
Basic Concept



Basic Concept

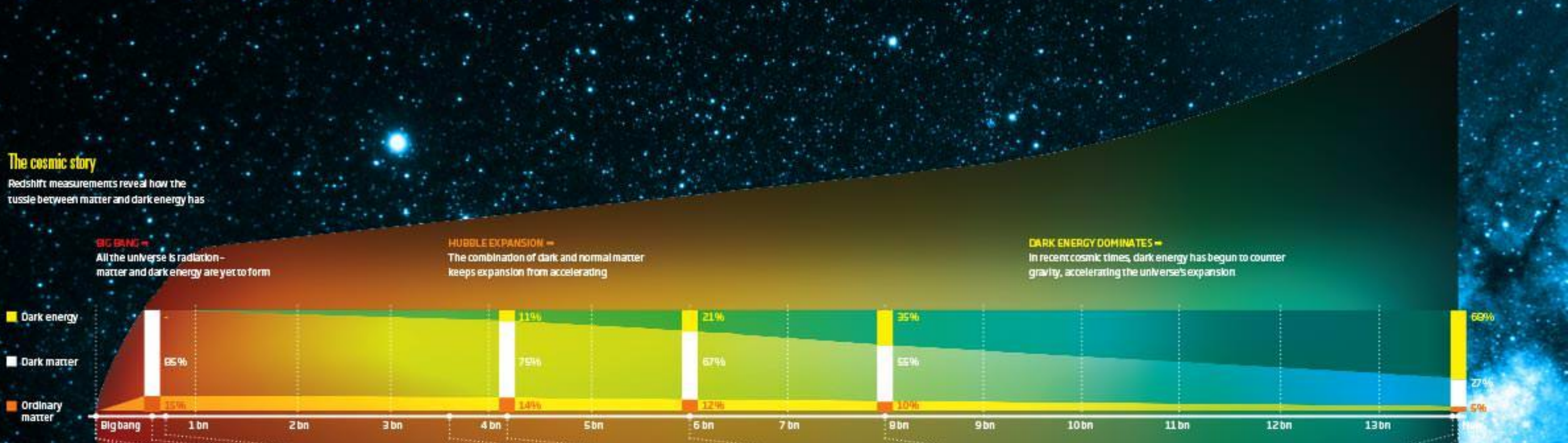


Basic Concept



The cosmic story

Redshift measurements reveal how the tussle between matter and dark energy has



Redshift infinity

BIG BANG

The equations of general relativity, which describe the universe's evolution, break down in the earliest moments of its history. No light from this time is visible to us

Redshift 1100

380,000 YEARS

The earliest light we can detect comes from this era. Its redshift is so extreme that the light now exists only as microwaves. This is the cosmic microwave background

Redshift 9

570 MILLION YEARS

The current furthest known galaxy, EGSY8p7, had formed by this time. It was found lurking in an image from the Hubble Space Telescope and confirmed by detailed observations using the Keck telescope in Hawaii in August 2015

Redshift 8

700 MILLION YEARS

Redshift 8 forms a barrier to observations using most wavelengths of light. NASA's future James Webb space telescope and the European Southern Observatory's new ALMA facility will try to stare beyond this frontier by looking for infrared radiation and radiowaves

Redshift 1.9

3.6 BILLION YEARS

The furthest supernova so far found lies at this redshift. Supernovae help us grasp cosmic evolution, but they are short-lived and therefore hard to find. The Large Synoptic Survey Telescope in Chile will search for specimens up to 10 billion light years from us

Redshift 1.6

4.2 BILLION YEARS

The universe was smaller in the past. As a result, beyond redshift 1.6, the light rays reaching us bend in a way that causes objects to look bigger, contrary to our everyday experience of distant objects appearing smaller. This complex relationship between size and redshift can be used to characterise dark energy

Redshift 1

6 BILLION YEARS

Galaxies become hard to detect at redshifts above 1 because their distance makes them appear so faint. An alternative is to make use of quasars: rare but extremely bright pinpricks of light emitted by gas getting squeezed into black holes

Redshift 0.6

8 BILLION YEARS

The complex web-like distribution of galaxies revealed by the Sloan Digital Sky Survey has been used to explore the universe's expansion at redshifts up to 0.6, corresponding to the past five or so billion years

Redshift 0.003

13.76 BILLION YEARS

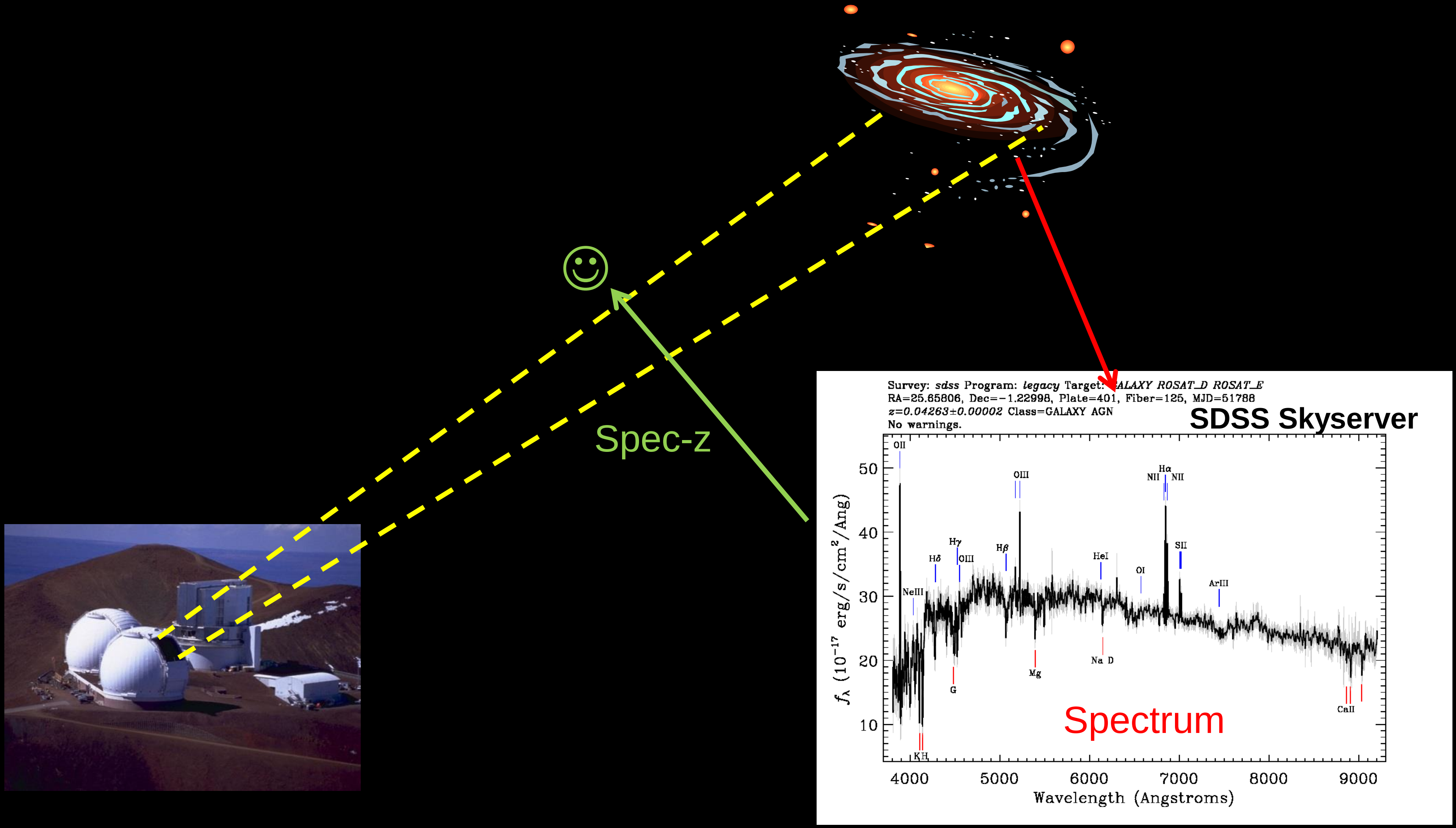
Edwin Hubble's observations in the 1920s established the crucial relationship between redshift and distance. They covered just a tiny fraction of what we can study today

Redshift zero

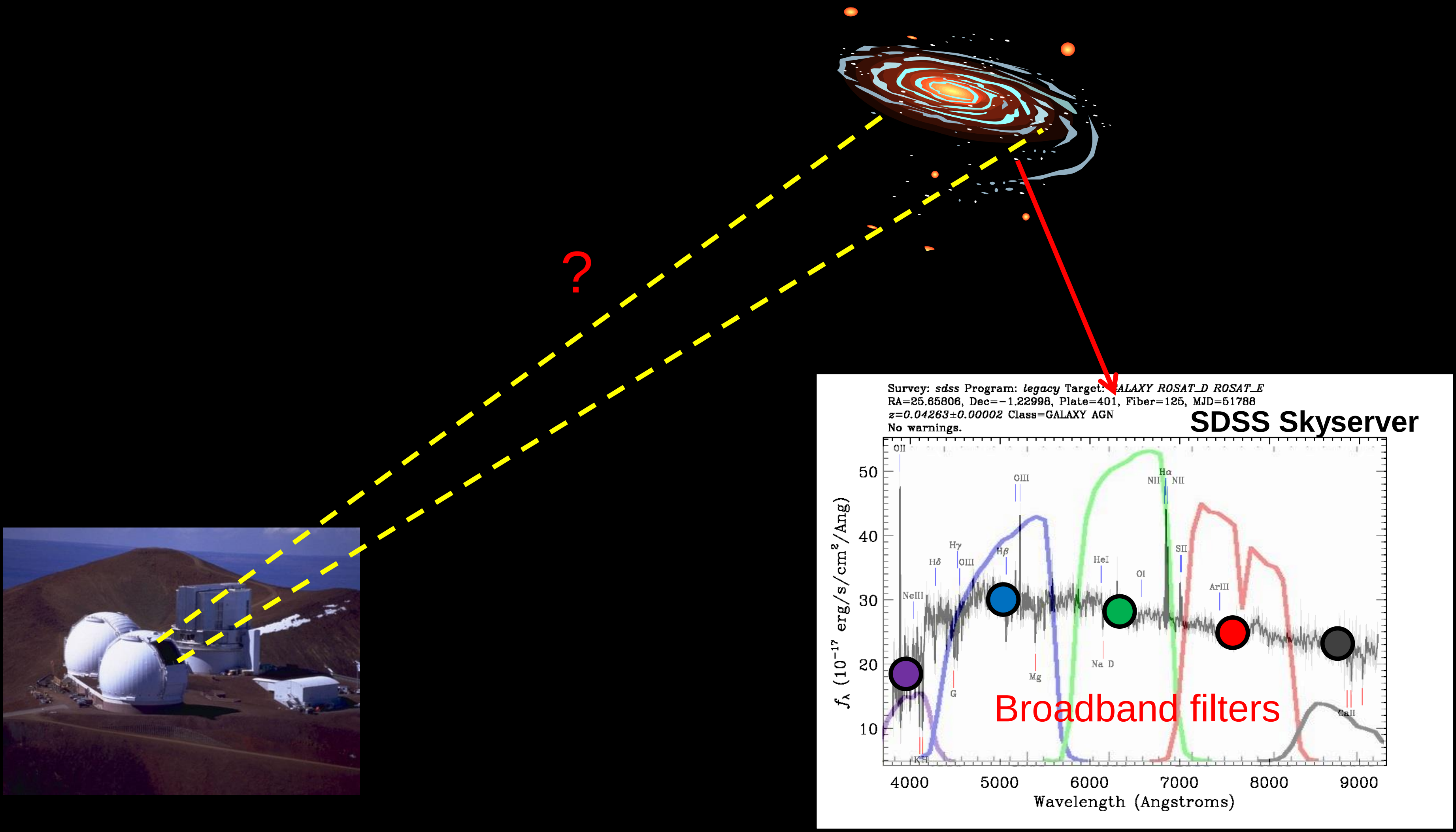
13.8 BILLION YEARS

Objects in our immediate neighbourhood – the solar system, nearby stars, and galaxies such as Andromeda – appear in their true colours, without redshift. They are the exception rather than the rule

Basic Concept

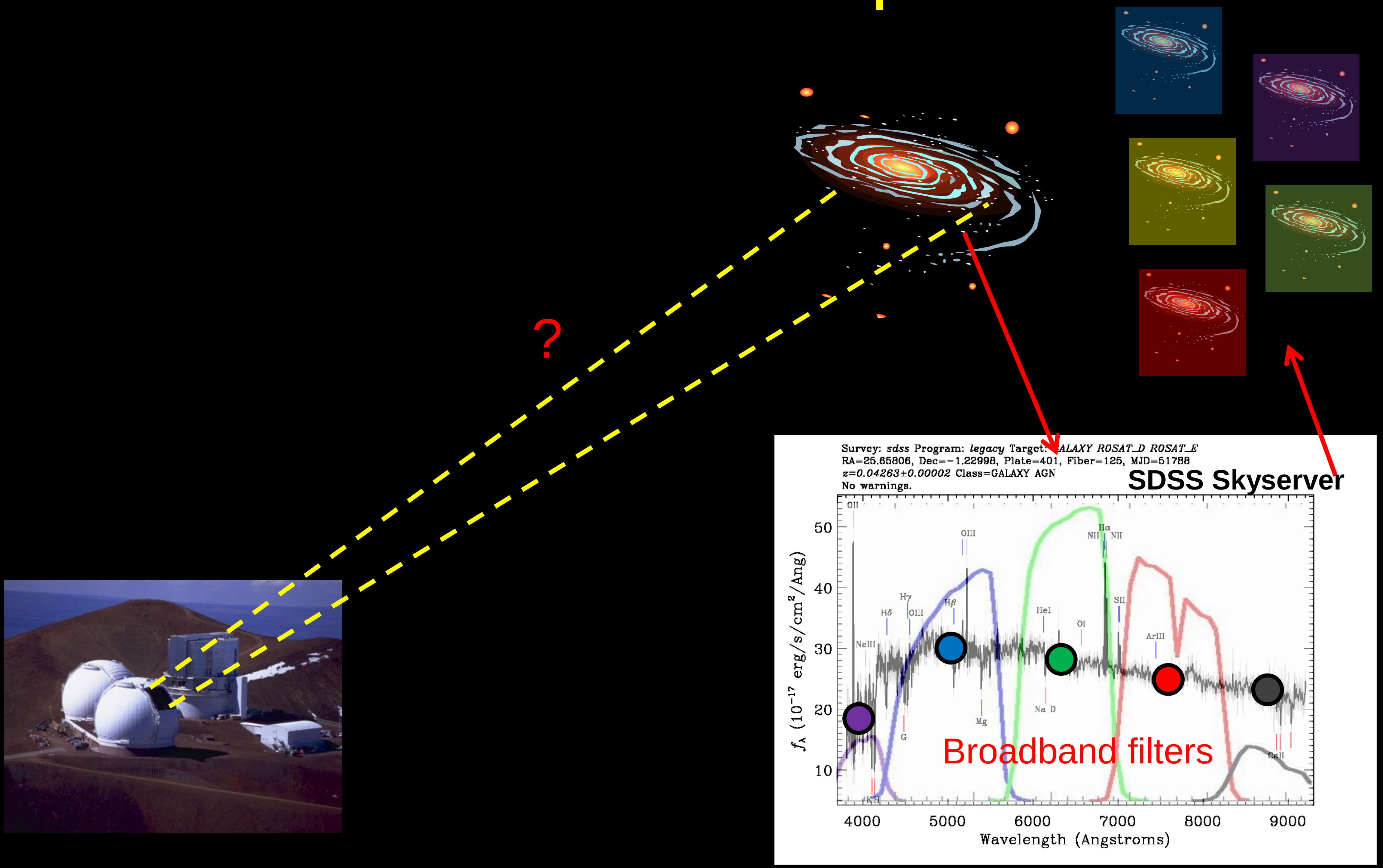


Basic Concept



Basic Concept

Spectral Energy Distribution (SED)



Basic Concept

Spectral Energy Distribution (SED)

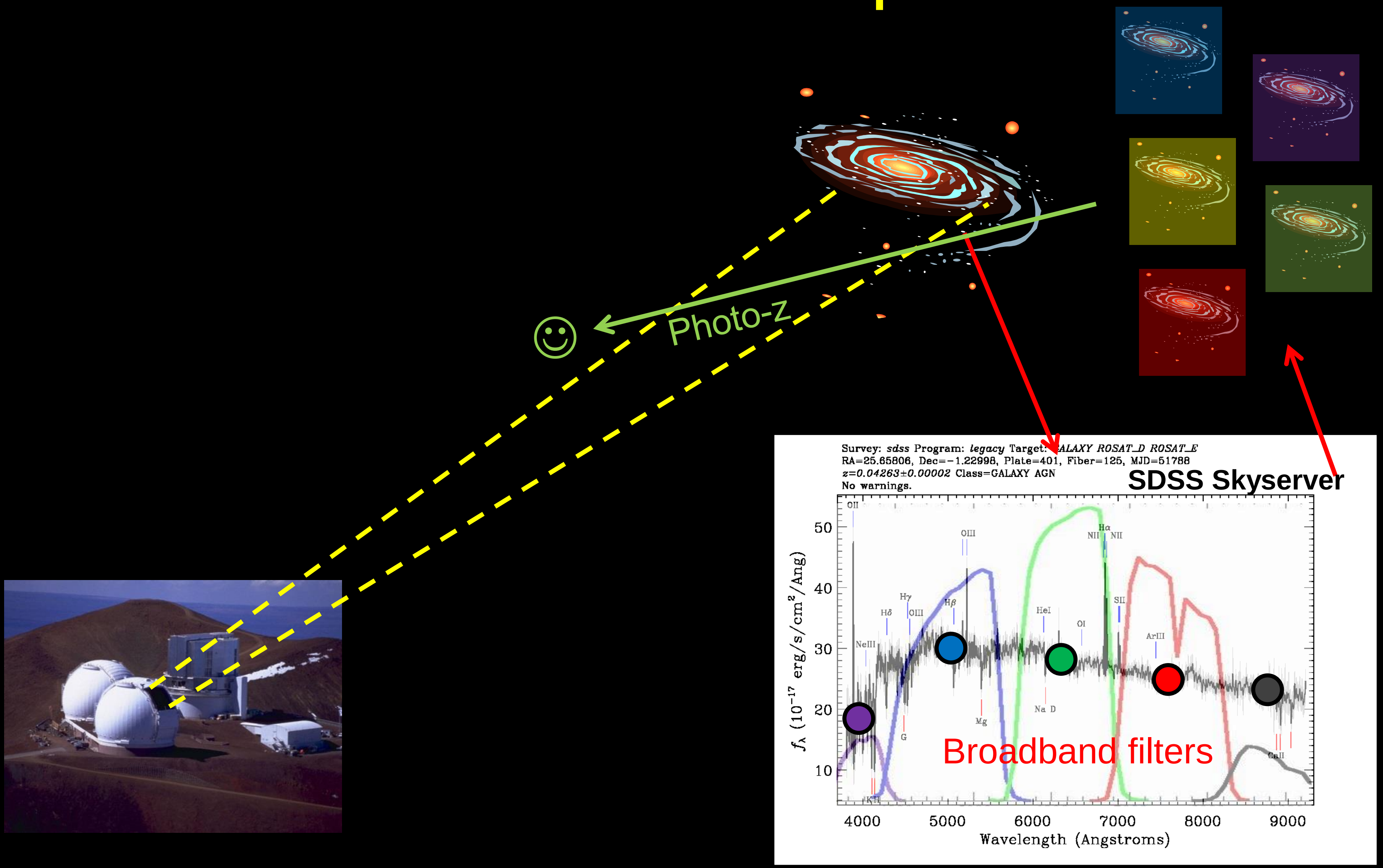


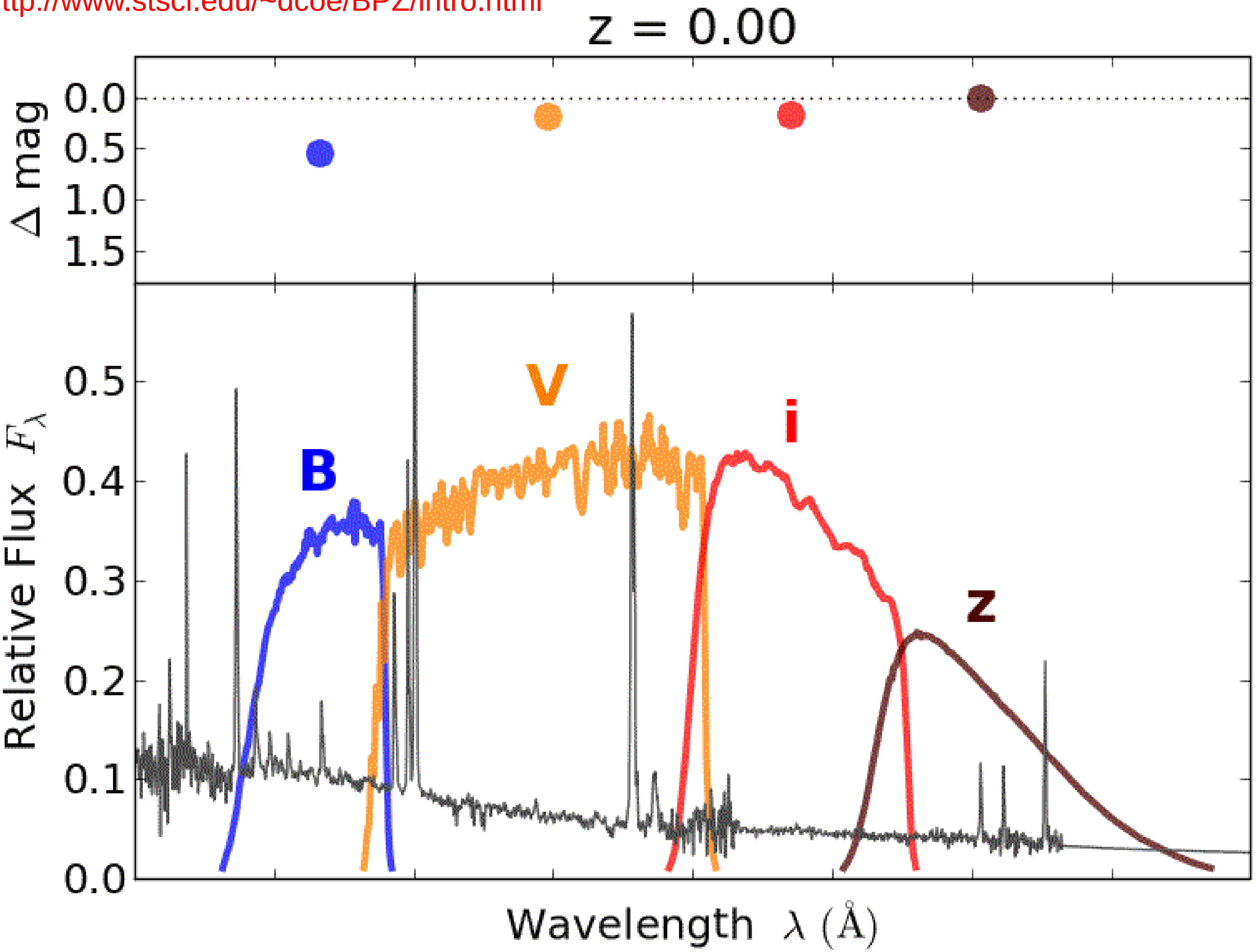
Photo-z

SDSS Skyserver

Survey: *sdss* Program: *legacy* Target: *GALAXY ROSAT_D ROSAT_E*
RA=25.65806, Dec=−1.22998, Plate=401, Fiber=125, MJD=51788
z=0.04263±0.00002 Class=GALAXY AGN
No warnings.

Broadband filters

http://www.stsci.edu/~dcoe/BPZ/intro.html



Why do we care?

Why do we care?

- **Because we have to.**

Why do we care?

- **Because we have to.**
  - Many questions now require large samples of galaxies to answer – now entering the “**Big Data**” era of astronomy.

Why do we care?

- **Because we have to.**
  - Many questions now require large samples of galaxies to answer – now entering the “**Big Data**” era of astronomy.
  - Wide-field imaging surveys **much** cheaper and faster than spectroscopic surveys. Also can see fainter objects.

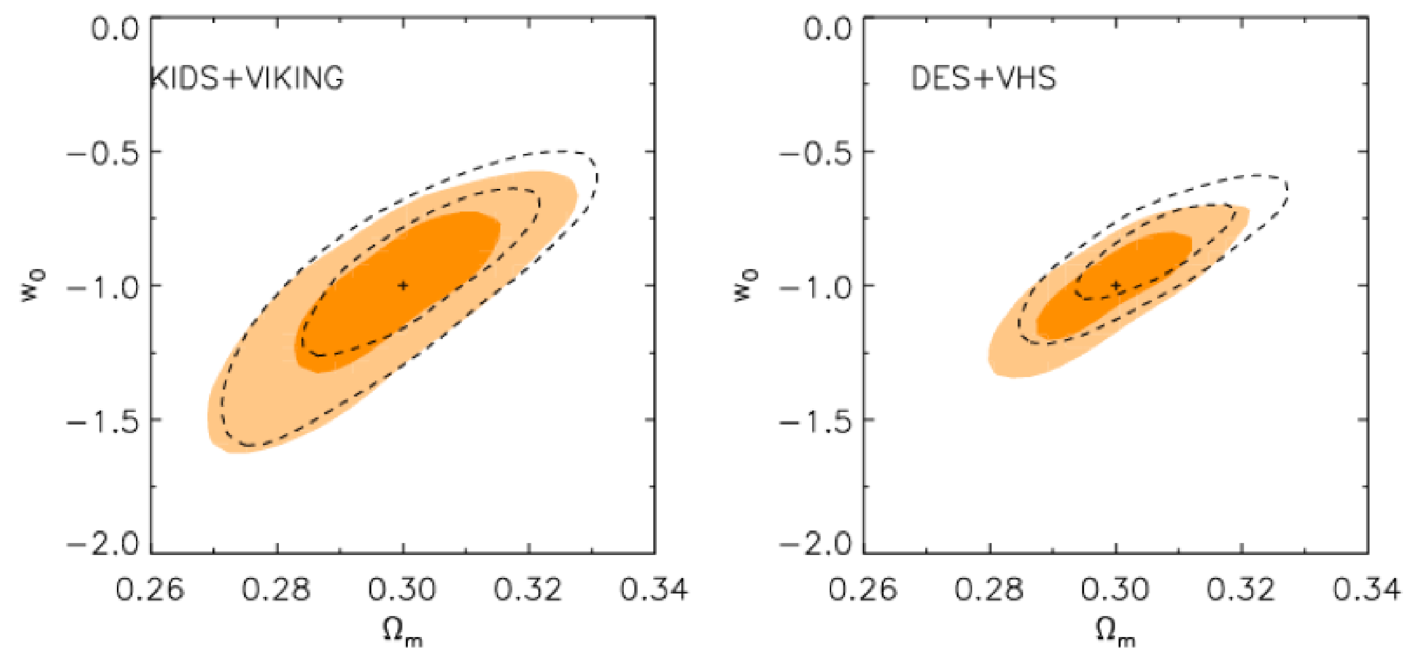
Why do we care?

- **Because we have to.**
  - Many questions now require large samples of galaxies to answer – now entering the "**Big Data**" era of astronomy.
  - Wide-field imaging surveys **much** cheaper and faster than spectroscopic surveys. Also can see fainter objects.
  - ~100x increase in sample size, diversity makes up for photo-z uncertainties. (Detailed studies can rely on ~1% spectroscopic subsample.)

Science Case

Precision cosmology

- Using large samples of galaxies to pin down the dark energy equation of state, growth of large-scale structure, etc.



Taken from http://kids.strw.leidenuniv.nl/goals.php.

Computing Photo-z’s

Photo-z’s: Statistically Speaking

Star formation rate

Star formation history

Redshift

Dust content

Stellar mass

Metallicity

AGN activity

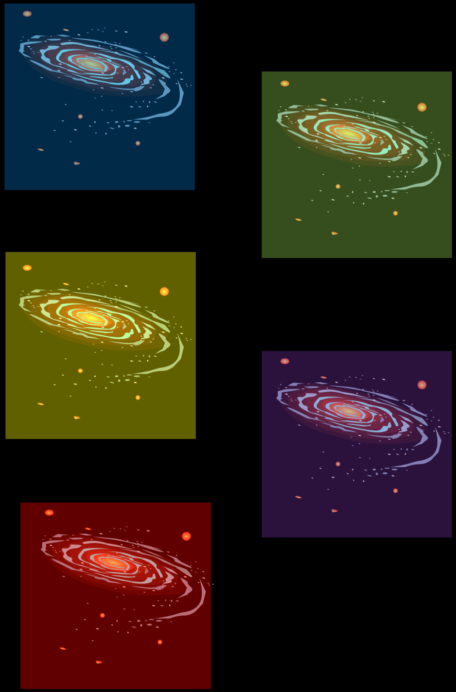


Photo-z's: Statistically Speaking

- A **forward modeling** problem: can we construct a model from parameters we care about that matches the observed SED?

Star formation rate

Star formation history

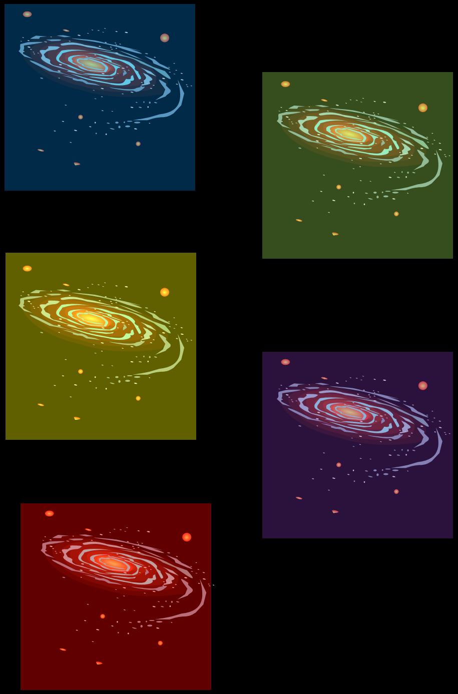
Redshift

Dust content

Stellar mass

Metallicity

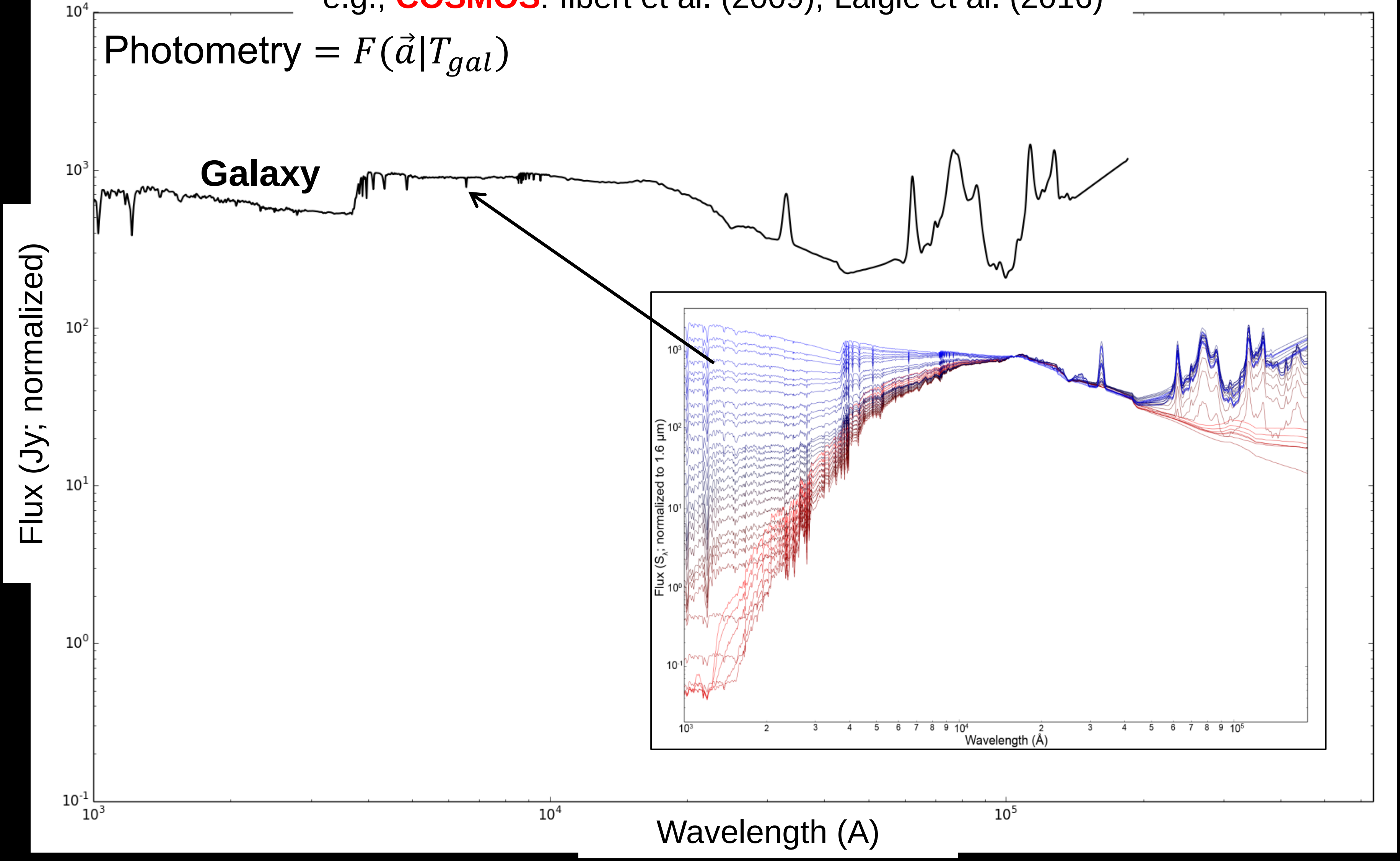
AGN activity



Traditional Approach

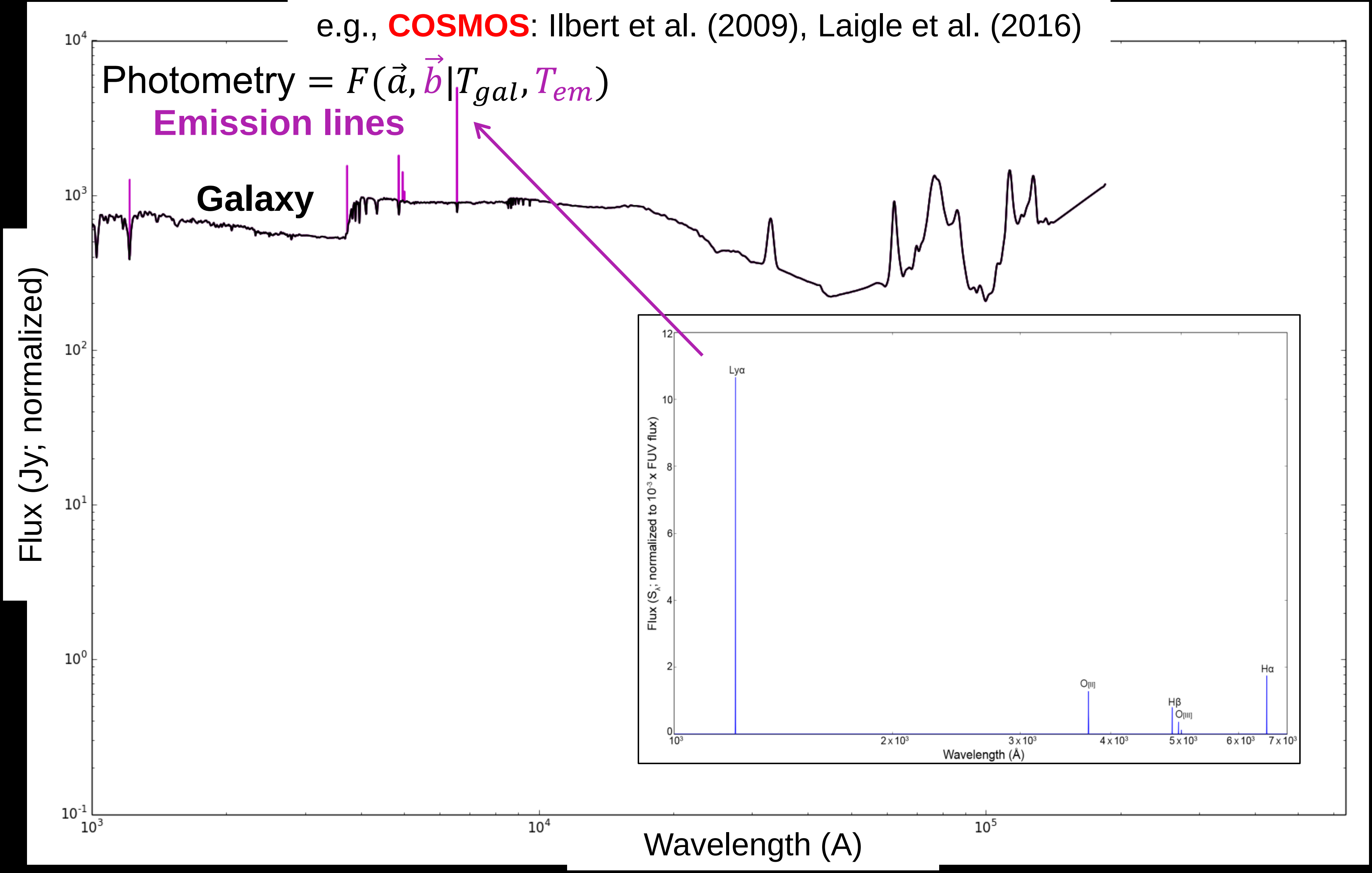
e.g., **COSMOS**: Ilbert et al. (2009), Laigle et al. (2016)

Photometry $= F(\vec{a}|T_{gal})$



Traditional Approach

e.g., **COSMOS**: Ilbert et al. (2009), Laigle et al. (2016)

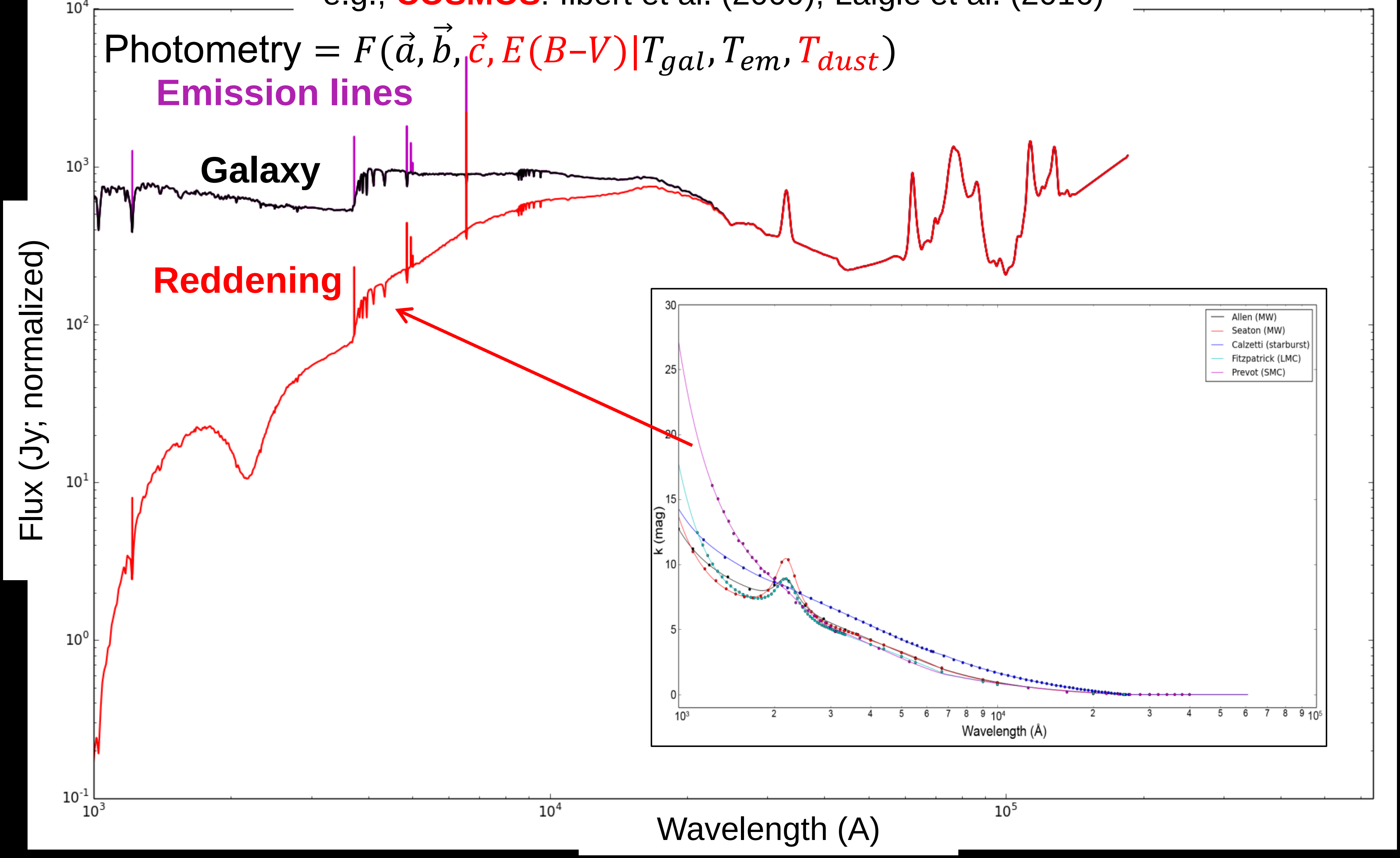


Photometry $= F(\vec{a}, \vec{b} | T_{gal}, T_{em})$

Traditional Approach

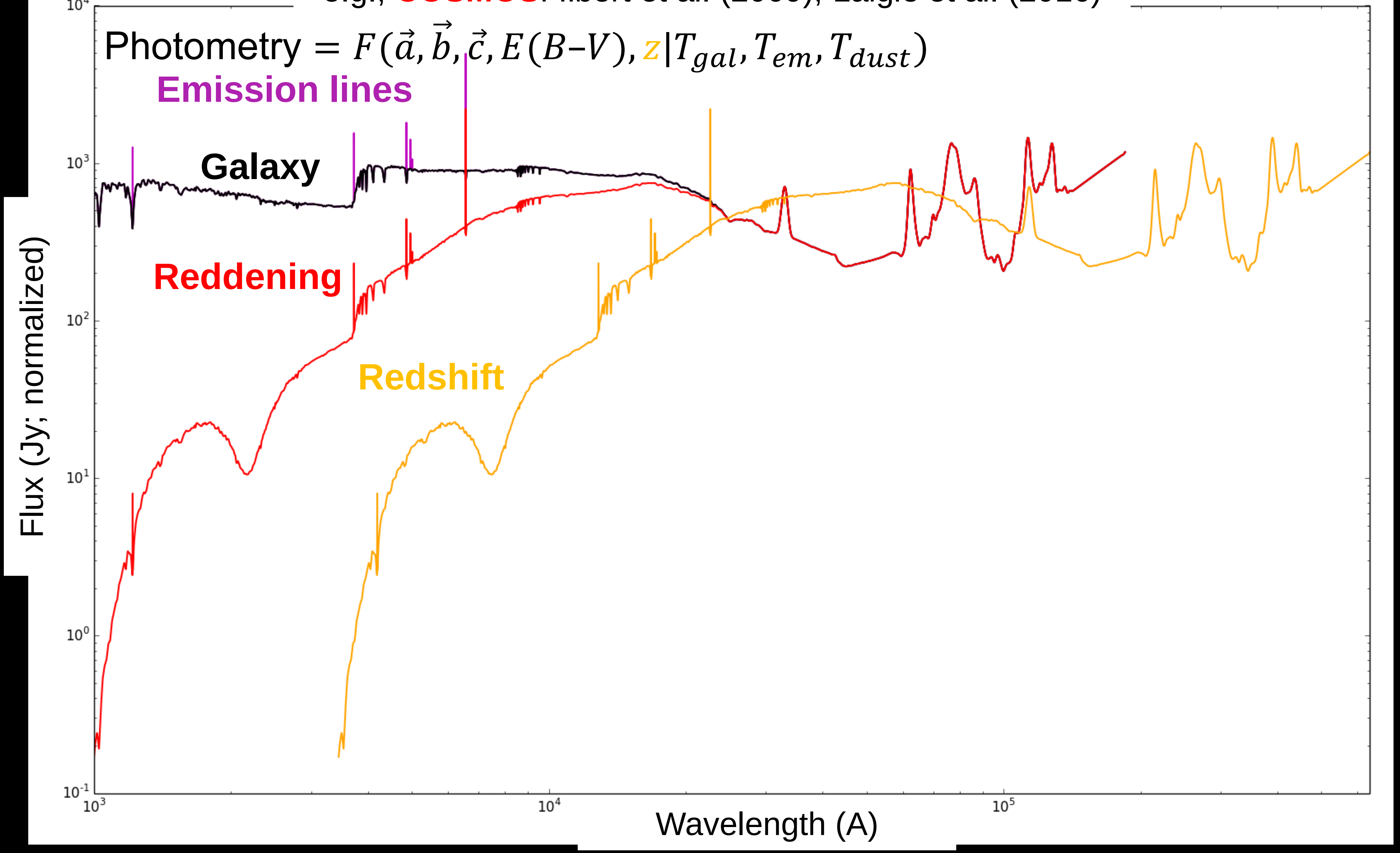
e.g., **COSMOS**: Ilbert et al. (2009), Laigle et al. (2016)

Photometry = $F(\vec{a},\vec{b},\vec{c},E(B\text{–}V)|T_{gal},T_{em},T_{dust})$



Traditional Approach

e.g., **COSMOS**: Ilbert et al. (2009), Laigle et al. (2016)

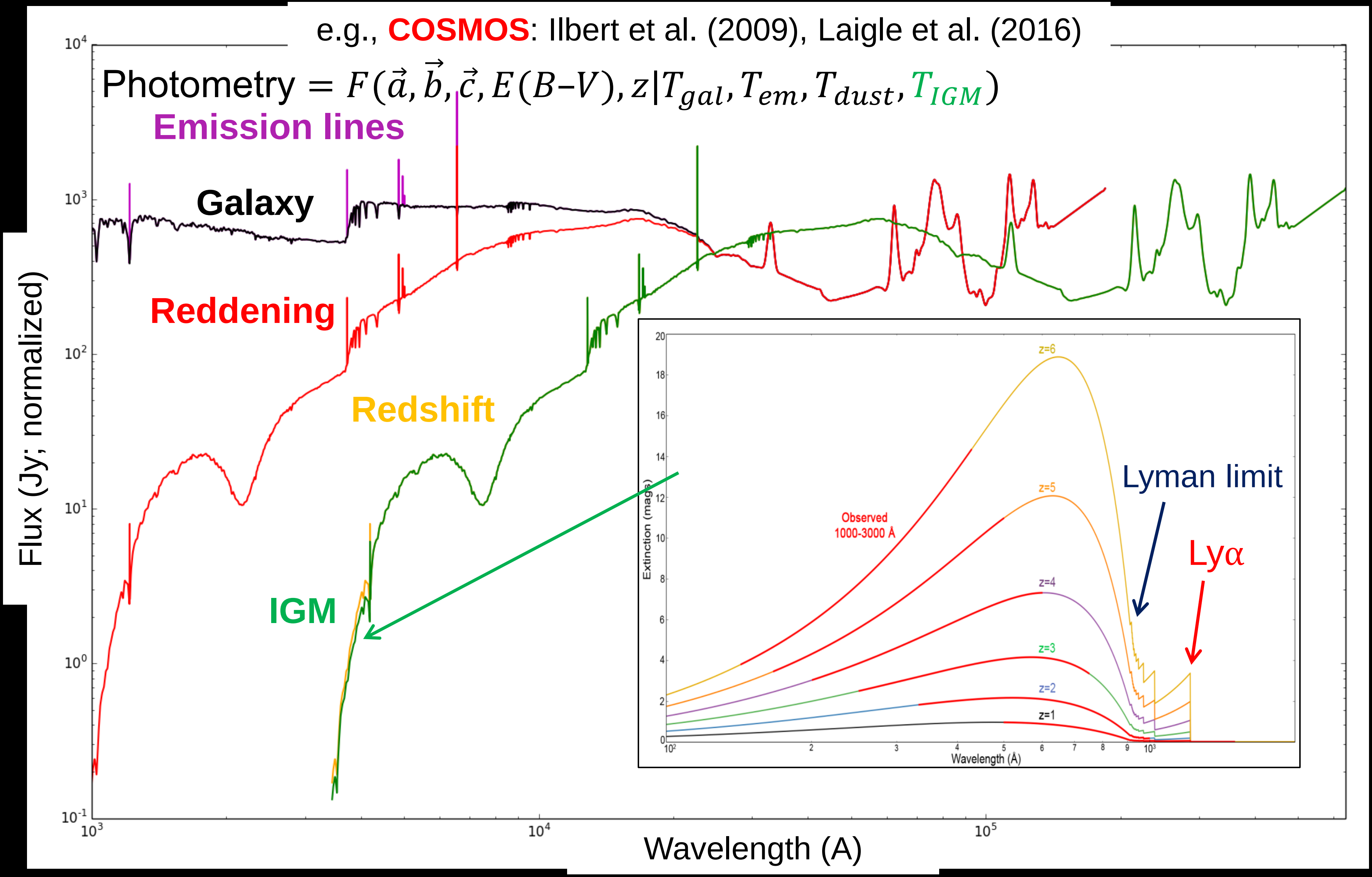


$$\text{Photometry} = F(\vec{a}, \vec{b}, \vec{c}, E(B\text{–}V), z \mid T_{gal}, T_{em}, T_{dust})$$

Traditional Approach

e.g., **COSMOS**: Ilbert et al. (2009), Laigle et al. (2016)

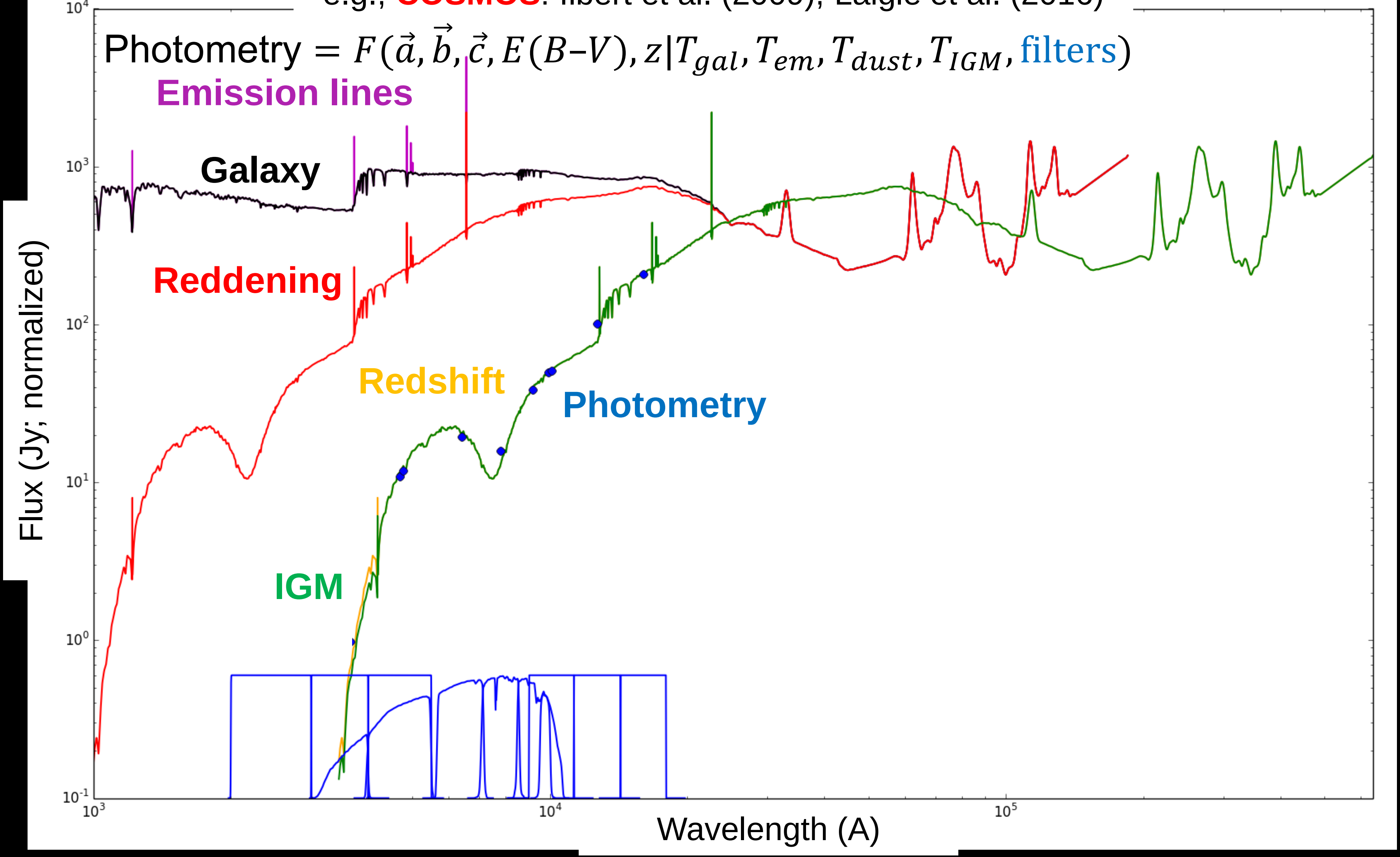
Photometry $= F(\vec{a},\vec{b},\vec{c},E(B\text{–}V),z|T_{gal},T_{em},T_{dust},T_{IGM})$



Traditional Approach

e.g., **COSMOS**: Ilbert et al. (2009), Laigle et al. (2016)

$$\text{Photometry} = F(\vec{a}, \vec{b}, \vec{c}, E(B{-}V), z \,|\, T_{gal}, T_{em}, T_{dust}, T_{IGM}, \text{filters})$$



Traditional Approach

e.g., **COSMOS**: Ilbert et al. (2009), Laigle et al. (2016)

$$\text{Photometry} = F(\vec{a}, \vec{b}, \vec{c}, E(B{-}V), z \mid T_{gal}, T_{em}, T_{dust}, T_{IGM}, \text{filters})$$

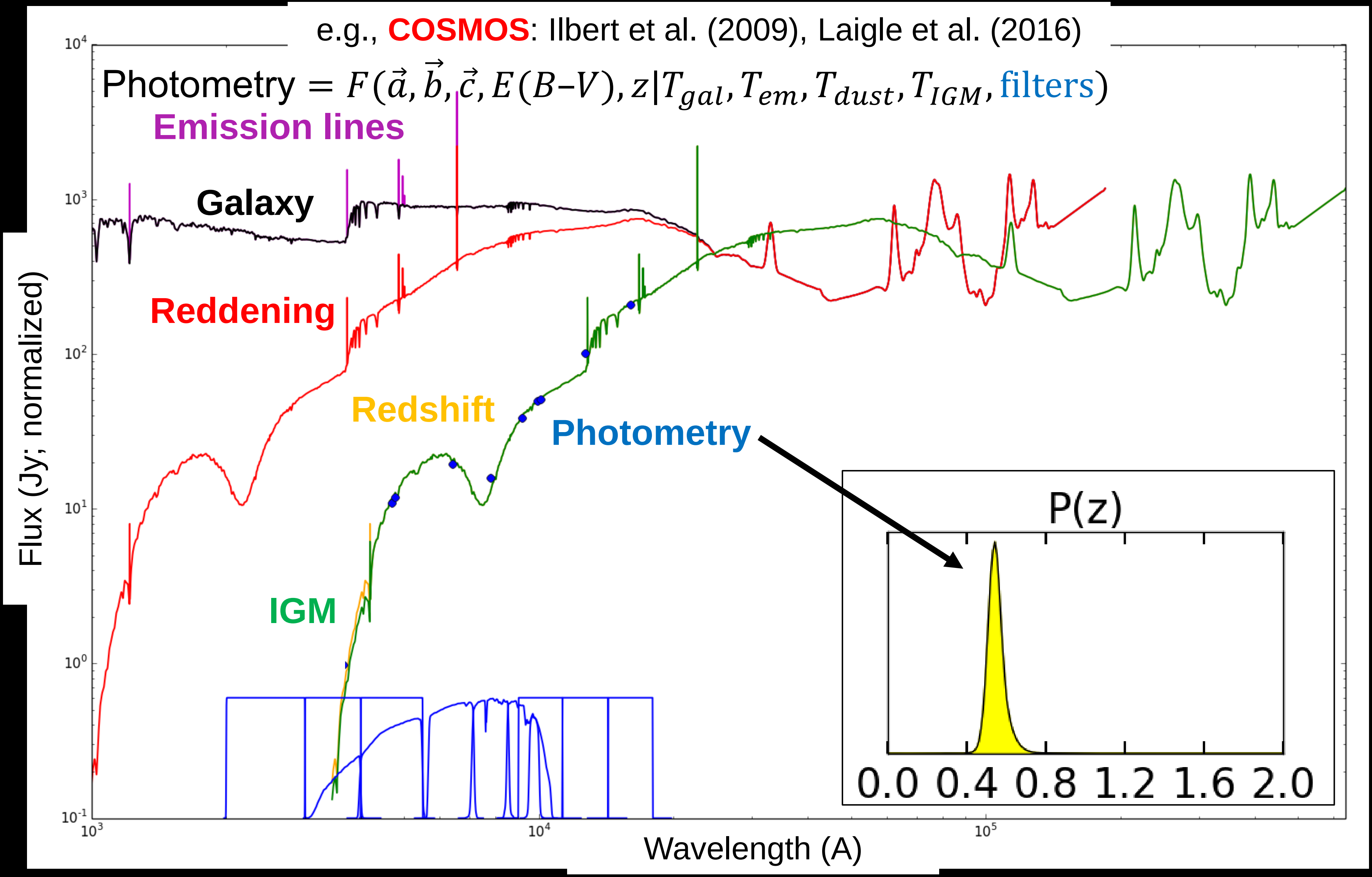
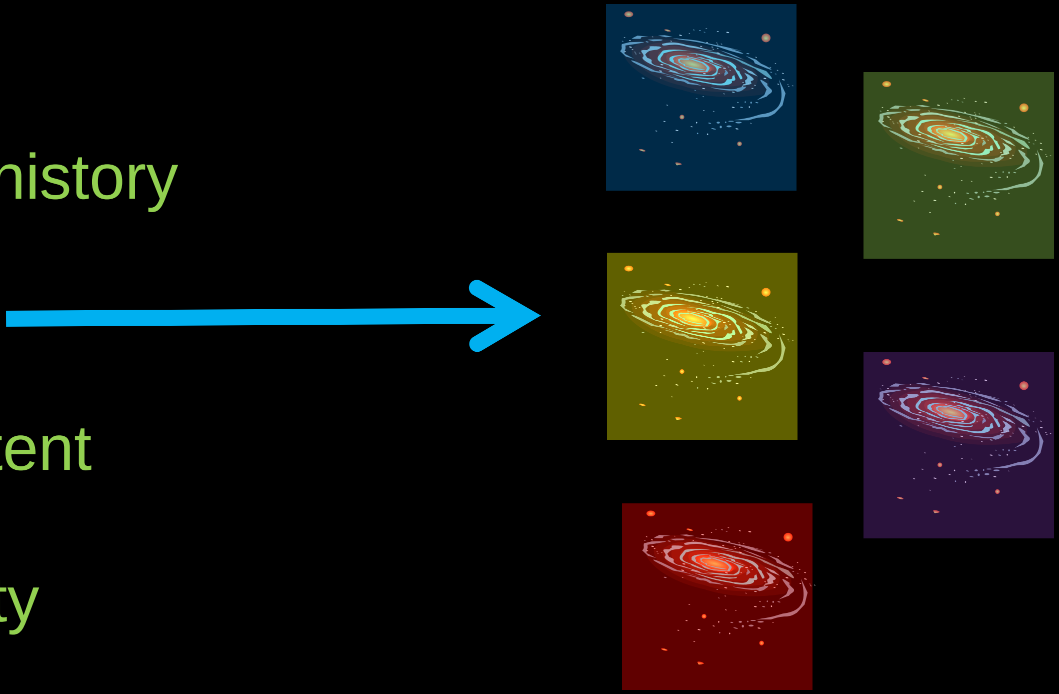


Photo-z's: Statistically Speaking

- A **forward modeling** problem: can we construct a model from parameters we care about that matches the observed SED?



Star formation rate

Star formation history

Redshift

Dust content

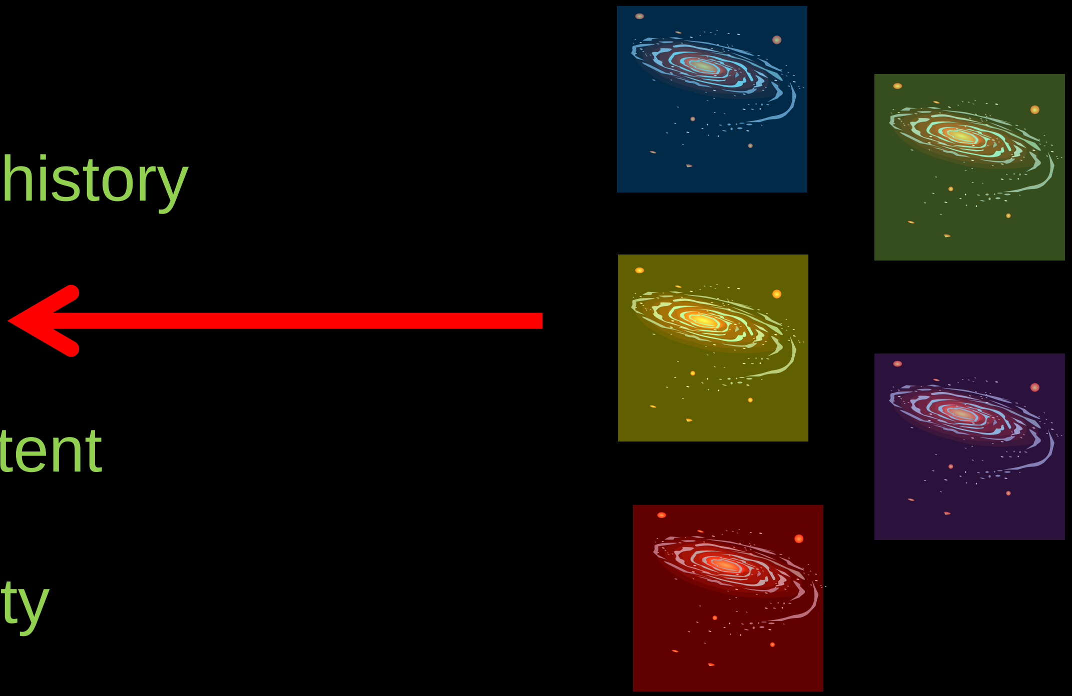
Stellar mass

Metallicity

AGN activity

Photo-z’s: Statistically Speaking

- An **inverse mapping** problem: can we use **machine learning** to construct a mapping from color to redshift?



Star formation rate

Star formation history

Redshift

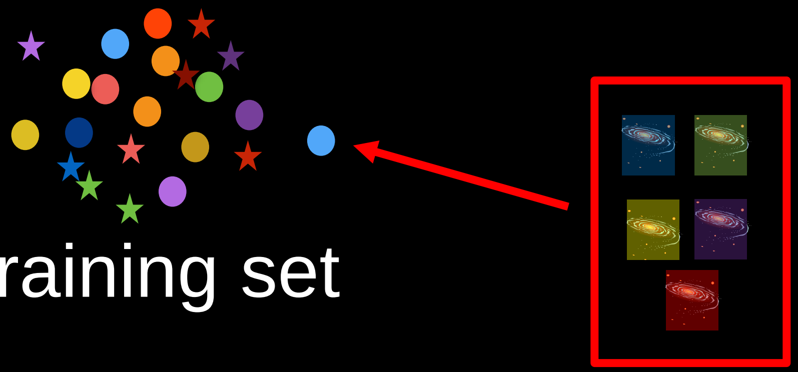
Dust content

Stellar mass

Metallicity

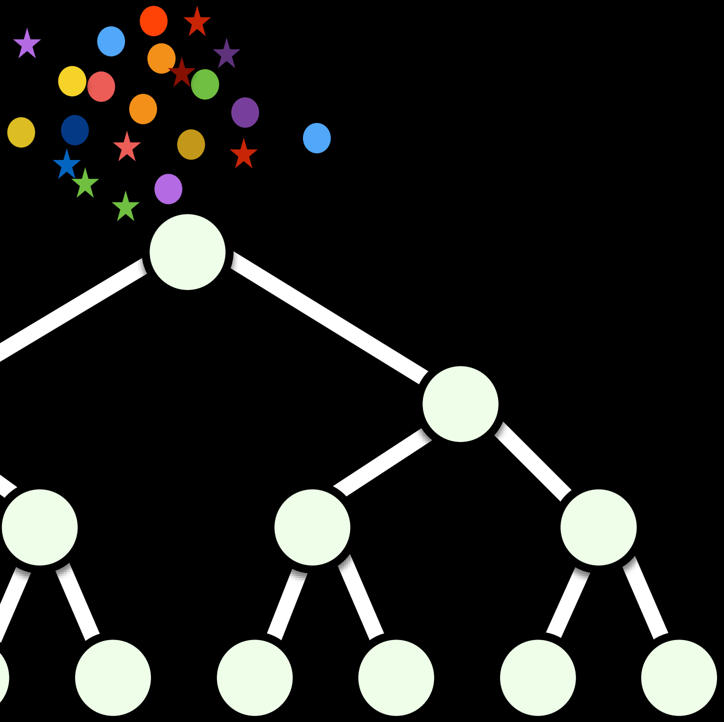
AGN activity

Machine Learning



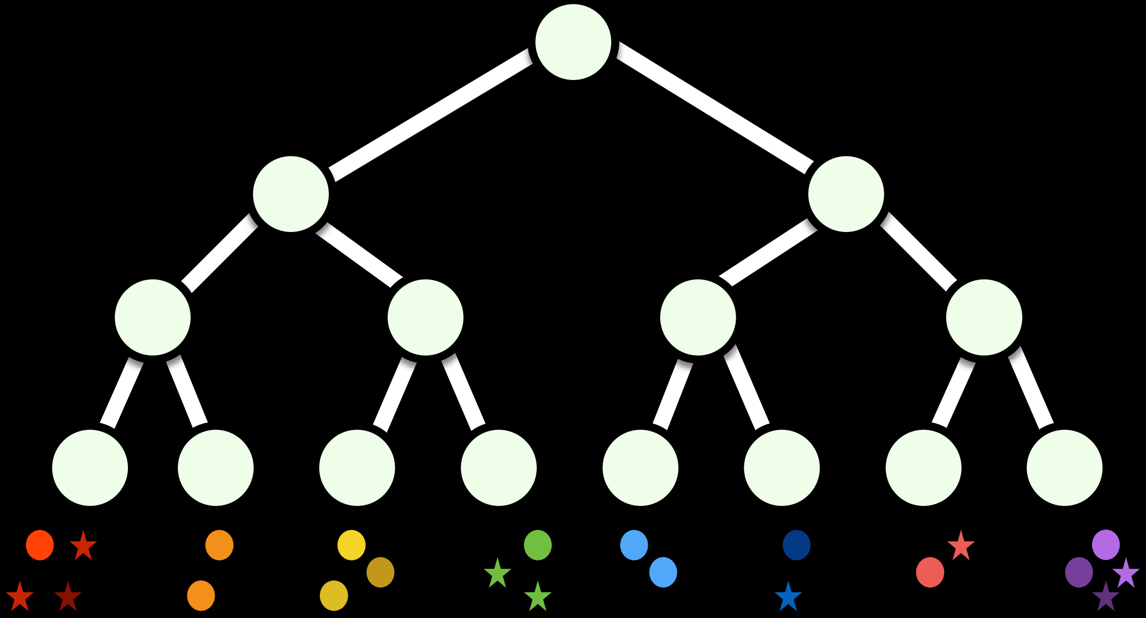
Training set

Machine Learning

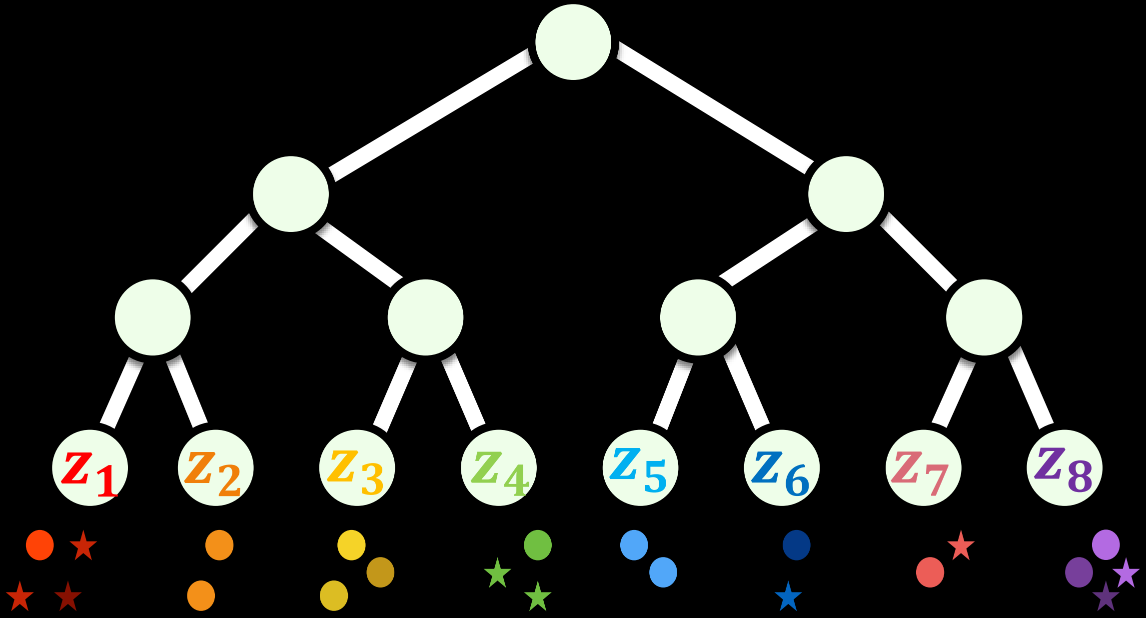


Decision
Tree

Machine Learning

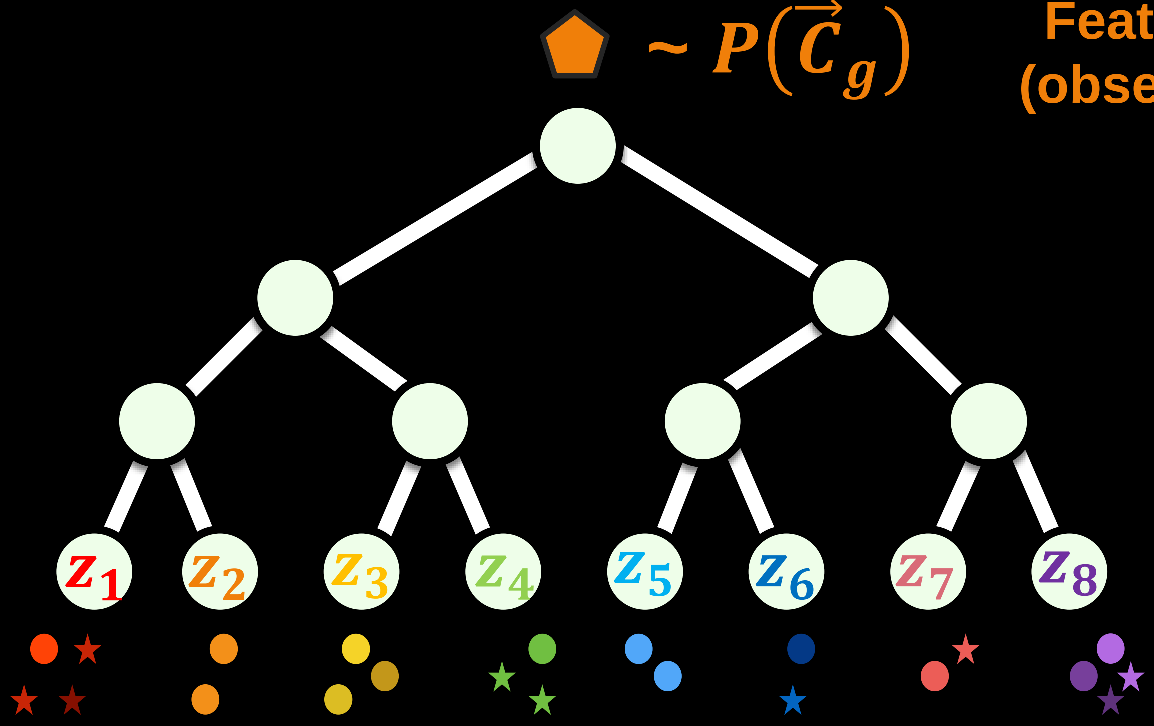


Machine Learning



Target (redshift)

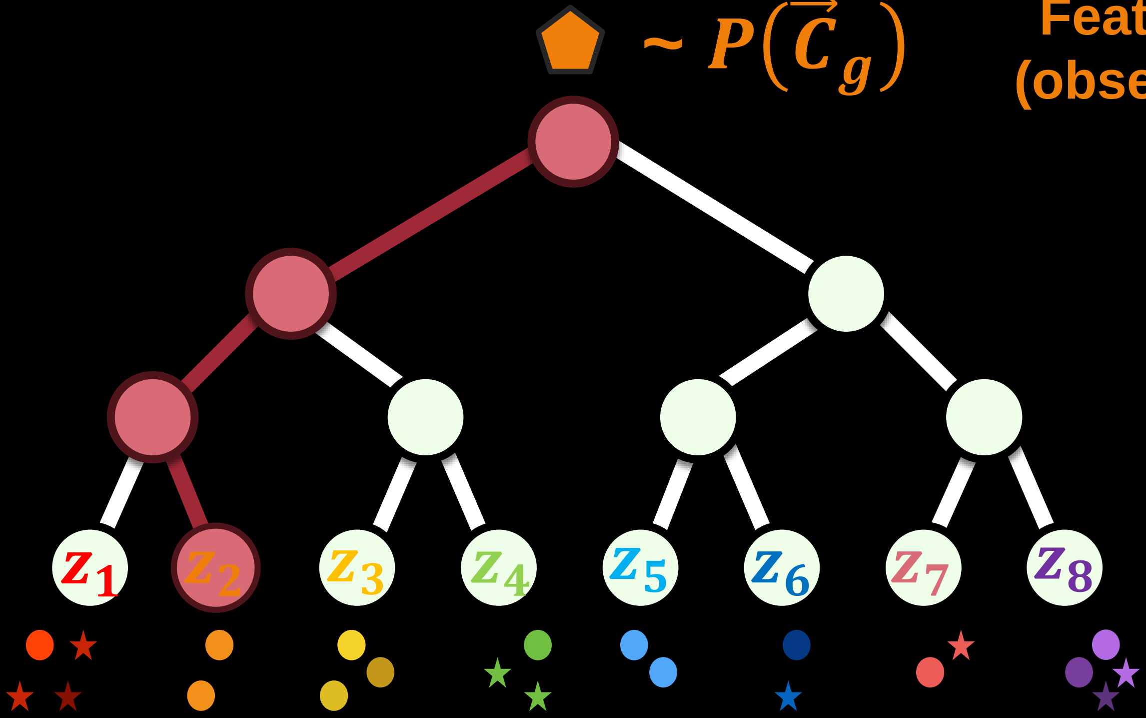
Machine Learning



Feature vector (observed color)

Target (redshift)

Machine Learning

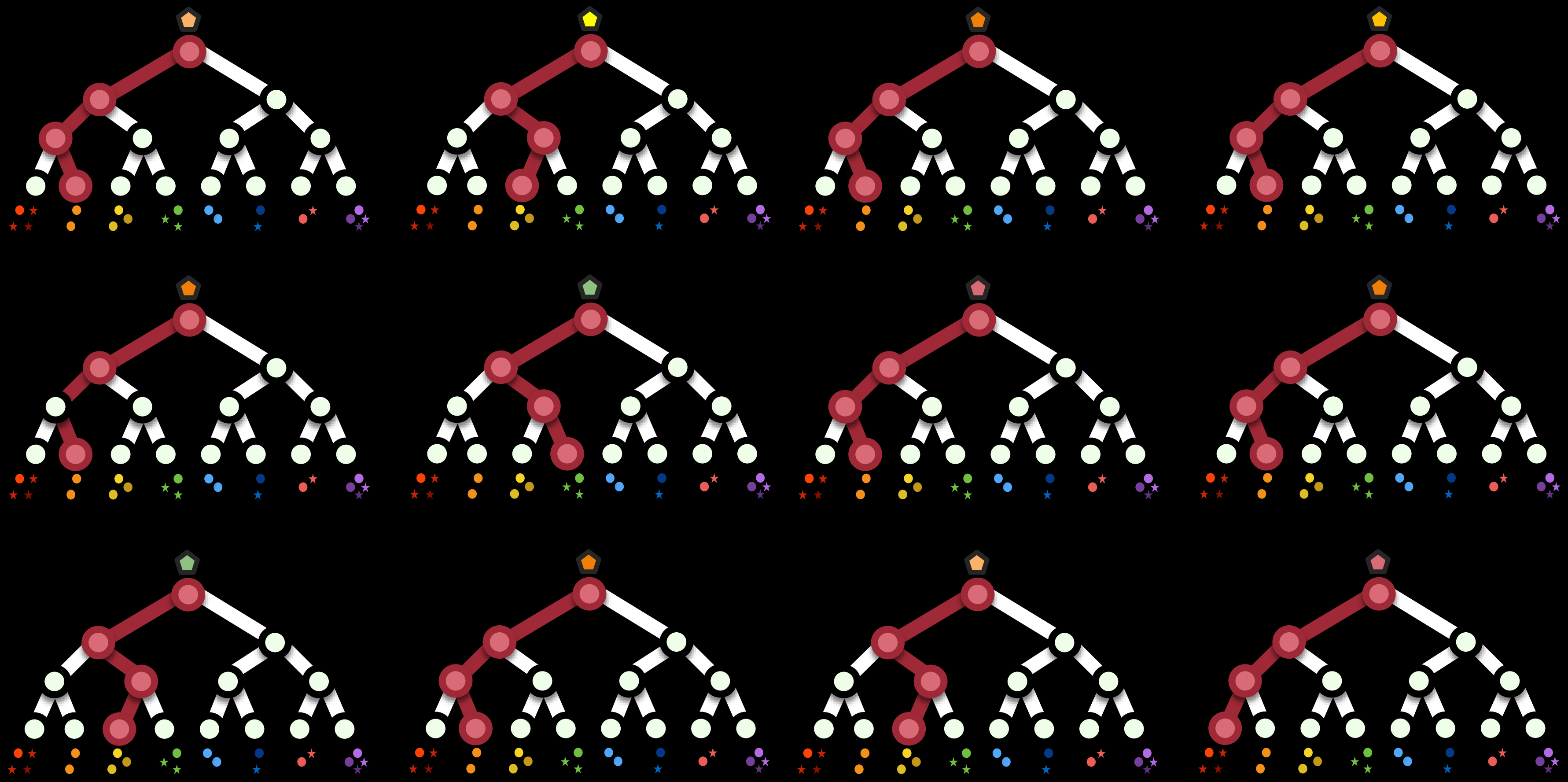


$\sim P(\vec{C}_g)$

Feature vector (observed color)

Target (redshift)

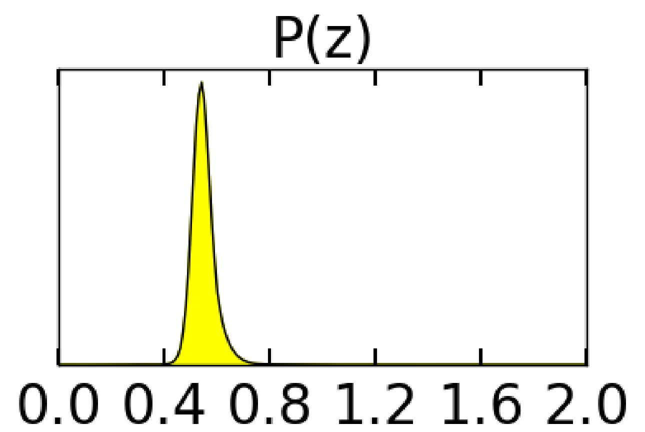
Machine Learning



Machine Learning

z'_1	z'_2	z'_3	z'_4	z'_5	z'_6	z'_7	z'_8	**Target**
$\mathbf{1}$	$\mathbf{7}$	$\mathbf{3}$	$\mathbf{1}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	**Weight**

Kernel Density Estimation



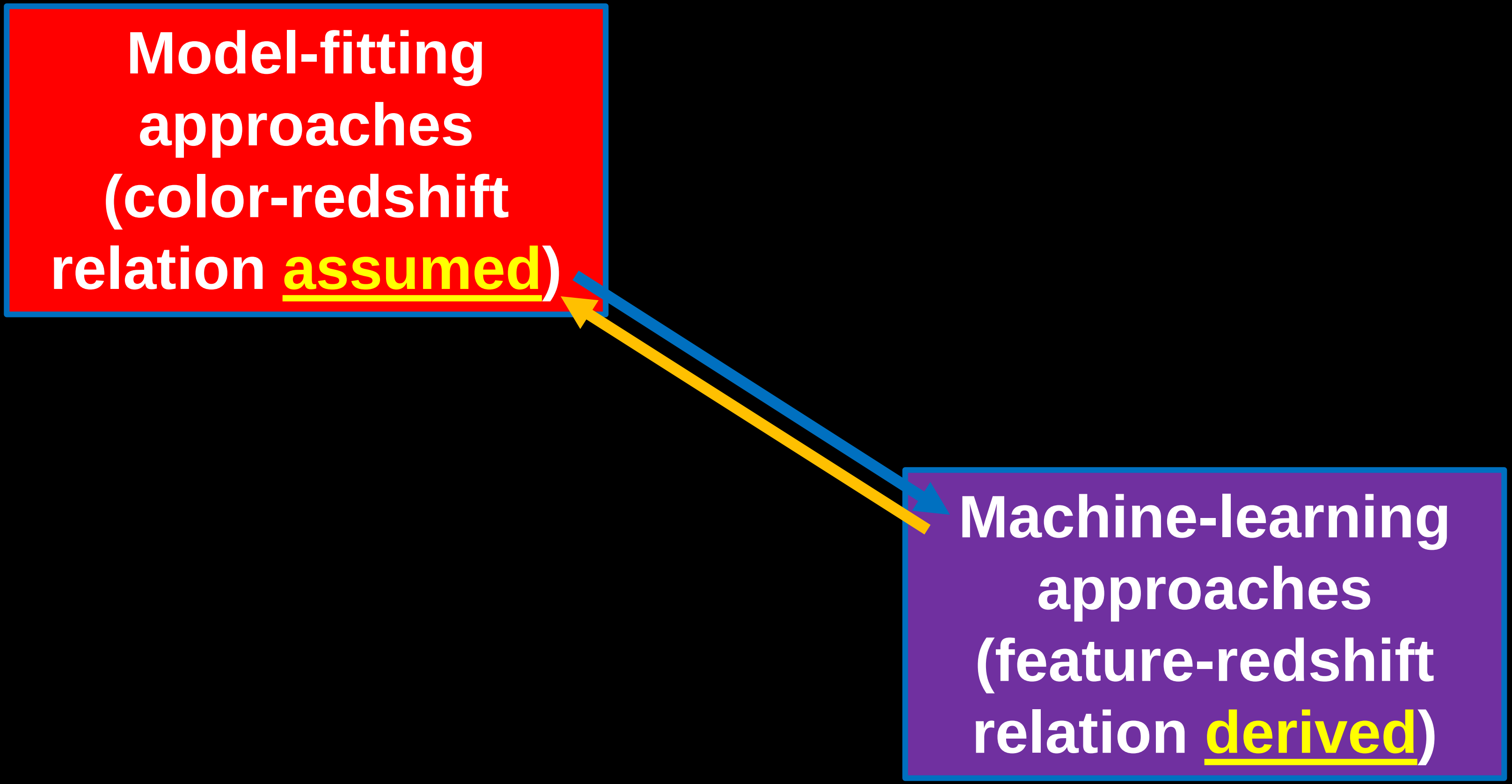
$\boldsymbol{P(z|g)}$

Two Communities

Model-fitting approaches (color-redshift relation assumed)

Machine-learning approaches (feature-redshift relation derived)

Two Communities



Model-fitting approaches (color-redshift relation assumed)

Machine-learning approaches (feature-redshift relation derived)

Two Communities

Model-fitting approaches (color-redshift relation assumed)

- Probabilistic
- Interpretable
- Sensitive to systematics
- Generally slow

Machine-learning approaches (feature-redshift relation derived)

Two Communities

Model-fitting approaches (color-redshift relation assumed)

- Probabilistic
- Interpretable
- Sensitive to systematics
- Generally slow

- Flexible, data-driven
- More robust to systematics
- Generally fast
- Difficult to interpret
- Difficult to derive PDFs

Machine-learning approaches (feature-redshift relation derived)

What are we doing?

Splitting Photo-z's into Two Parts

- Mapping from features to redshift.
  - Can be done through model fitting and/or machine learning.

$P(z|\boldsymbol{F})$ or $P(\boldsymbol{F}|z,\boldsymbol{\eta})$

Splitting Photo-z's into Two Parts

- Mapping from features to redshift.
  - Can be done through model fitting and/or machine learning.

$P(z|\boldsymbol{F})$ or $P(\boldsymbol{F}|z,\boldsymbol{\eta})$

- Propagating uncertainties in features.

$P(\boldsymbol{F}|\widehat{\boldsymbol{F}},\widehat{\boldsymbol{C}})$

Splitting Photo-z's into Two Parts

- Mapping from features to redshift.
  - Can be done through model fitting and/or machine learning.

$$\left\{P\left(z \middle| \widehat{\boldsymbol{F}}_h, \widehat{\boldsymbol{C}}_h\right)\right\}$$

$$\left\{P\left(\boldsymbol{F}_h \middle| \widehat{\boldsymbol{F}}_h, \widehat{\boldsymbol{C}}_h\right)\right\}$$

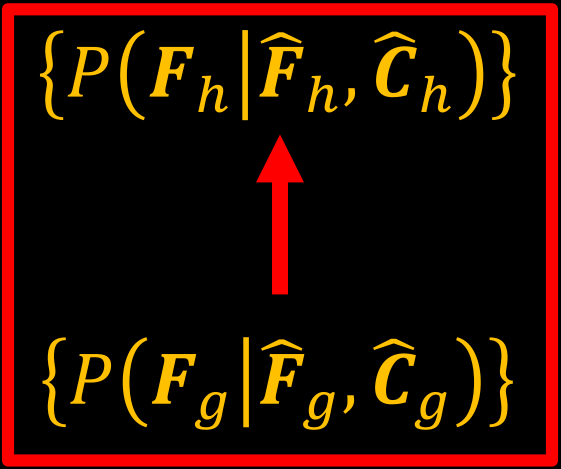
- Propagating uncertainties in features.

$$\left\{P\left(\boldsymbol{F}_g \middle| \widehat{\boldsymbol{F}}_g, \widehat{\boldsymbol{C}}_g\right)\right\}$$

Splitting Photo-z's into Two Parts

- Mapping from features to redshift.
  - Can be done through model fitting and/or machine learning.
- Propagating uncertainties in features.

$$\{P(z|\hat{\boldsymbol{F}}_h, \hat{\boldsymbol{C}}_h)\}$$



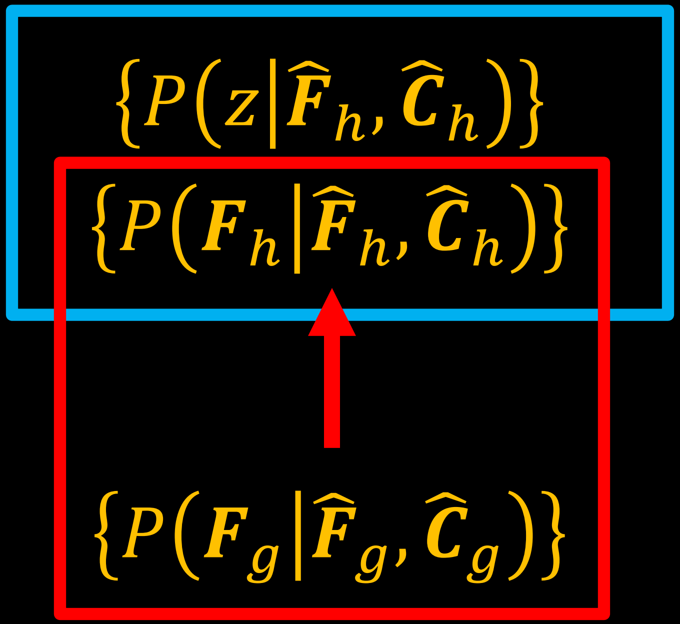
$$\{P(\boldsymbol{F}_h|\hat{\boldsymbol{F}}_h, \hat{\boldsymbol{C}}_h)\}$$

$$\{P(\boldsymbol{F}_g|\hat{\boldsymbol{F}}_g, \hat{\boldsymbol{C}}_g)\}$$

Splitting Photo-z's into Two Parts

- Mapping from features to redshift.
  - Can be done through model fitting and/or machine learning.

- Propagating uncertainties in features.



$$\{P(z|\hat{\boldsymbol{F}}_h,\hat{\boldsymbol{C}}_h)\}$$

$$\{P(\boldsymbol{F}_h|\hat{\boldsymbol{F}}_h,\hat{\boldsymbol{C}}_h)\}$$

$$\{P(\boldsymbol{F}_g|\hat{\boldsymbol{F}}_g,\hat{\boldsymbol{C}}_g)\}$$

Feature Projection

$$\begin{tikzcd}
\left\{P\left(\boldsymbol{F}_h \middle| \widehat{\boldsymbol{F}}_h, \widehat{\boldsymbol{C}}_h\right)\right\} \\
\left\{P\left(\boldsymbol{F}_g \middle| \widehat{\boldsymbol{F}}_g, \widehat{\boldsymbol{C}}_g\right)\right\} \arrow[u]
\end{tikzcd}$$

The “Big Data” Assumption

$$P(z|g) = \sum_{h} P(z|h)P(h|g)$$

The “Big Data” Assumption

redshift

redshift PDF

$$P(z|g) = \sum_{h} P(z|h)P(h|g)$$

target galaxy “g”

photometric
posterior

training galaxy “h”

The "Big Data" Assumption

$$P(z|g) = \sum_h P(z|h)P(h|g)$$

$$= \sum_h P(z|h)\frac{P(g|h)P(h)}{P(g)}$$

The “Big Data” Assumption

$$P(z|g) = \sum_h P(z|h)P(h|g)$$

$$= \sum_h P(z|h)\frac{P(g|h)P(h)}{P(g)}$$

1. What is this likelihood?

The “Big Data” Assumption

$$P(z|g) = \sum_h P(z|h)P(h|g)$$

$$= \sum_h P(z|h)\frac{P(g|h)P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?

The “Big Data” Assumption

$$P(z|g) = \sum_{h} P(z|h)P(h|g)$$

band mask

$$= \sum_{h} P(z|h)\frac{P(g|h,\boldsymbol{b})P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?
3. How do we deal with missing data?

The “Big Data” Assumption

selection effects

$$P(z|g) = \sum_h P(z|h) P\left(h \middle| g, s_g = 1, \boldsymbol{S_g}\right)$$

galaxy is
observed

$$\overset{?}{=} \sum_h P(z|h) \frac{P(g|h, \boldsymbol{b}) P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?
3. How do we deal with missing data?
4. How do we incorporate selection effects?

The “Big Data” Assumption

$$P(z|g) = \sum_h P(z|h) P\left(h \middle| g, s_g = 1, \boldsymbol{S_g}\right)$$

$$\overset{?}{=} \sum_h P(z|h) \frac{P(g|h, \boldsymbol{b}) P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?
3. How do we deal with missing data?
4. How do we incorporate selection effects?
5. What is our prior?

The “Big Data” Assumption

$$P(z|g) = \sum_{h} P(z|h) P\left(h \middle| g, s_g = 1, \boldsymbol{S_g}\right)$$

$$\overset{?}{=} \sum_{h} P(z|h) \frac{P(g|h, \boldsymbol{b}) P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?
3. How do we deal with missing data?
4. How do we incorporate selection effects?
5. What is our prior?

Likelihood / Distance Metric

- What metric(s) to use?

Training galaxy

$$P(g|h)$$

Observed galaxy

Likelihood / Distance Metric

- What metric(s) to use?

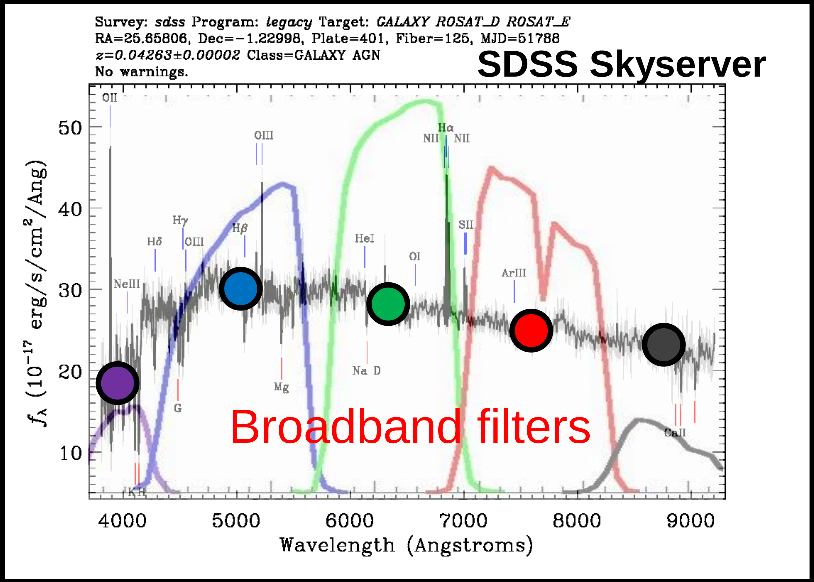
$$P(\underset{\text{Observed galaxy}}{g}|\overset{\text{Training galaxy}}{h})$$

NGC 4414

NGC 5457

Likelihood / Distance Metric

- What metric(s) to use?



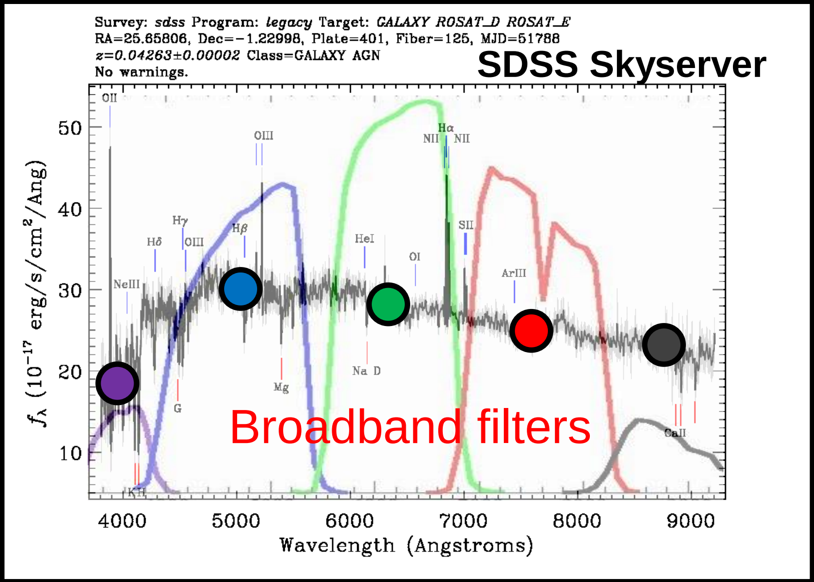
Training galaxy

$P(g|h)$

Observed galaxy

Likelihood / Distance Metric

- What metric(s) to use?



Training galaxy

$$P(g|h)$$

Observed galaxy

Assume data is normally distributed.

$$P\left(\boldsymbol{F}_g \middle| \widehat{\boldsymbol{F}}_g, \widehat{\boldsymbol{C}}_g\right) \sim N\left(\boldsymbol{F}_g \middle| \widehat{\boldsymbol{F}}_g, \widehat{\boldsymbol{C}}_g\right)$$

Likelihood / Distance Metric

- What metric(s) to use?

$P(g|h)$

Color space (traditional)

$$\chi^2 = \sum_i \frac{\left(\hat{F}_{g,i} - s\hat{F}_{h,i}\right)^2}{\hat{\sigma}_{g,i}^2 + s^2\hat{\sigma}_{h,i}^2}$$

Likelihood / Distance Metric

- What metric(s) to use?

$$P(g|h)$$

Color space (traditional)

$$\chi^2 = \sum_i \frac{\left(\hat{F}_{g,i} - s\hat{F}_{h,i}\right)^2}{\hat{\sigma}^2_{g,i} + \cancel{s^2\hat{\sigma}^2_{h,i}}}$$

Likelihood / Distance Metric

- What metric(s) to use?

$$P(g|h)$$

Color space (traditional)

$$\chi^2 = \sum_i \frac{\left(\hat{F}_{g,i} - s\hat{F}_{h,i}\right)^2}{\hat{\sigma}^2_{g,i} + \cancel{s^2\hat{\sigma}^2_{h,i}}}$$

Magnitude space (“new”)

$$\chi^2 = \sum_i \frac{\left(\hat{F}_{g,i} - \hat{F}_{h,i}\right)^2}{\hat{\sigma}^2_{g,i} + \hat{\sigma}^2_{h,i}}$$

Likelihood / Distance Metric

- What metric(s) to use?

$$P(g|h)$$

Color space (traditional)

Requires magnitude priors to account for galaxy evolution.

“Scale-free”

$$\chi^2 = \sum_i \frac{\left(\hat{F}_{g,i} - s\hat{F}_{h,i}\right)^2}{\hat{\sigma}_{g,i}^2 + \cancel{s^2\hat{\sigma}_{h,i}^2}}$$

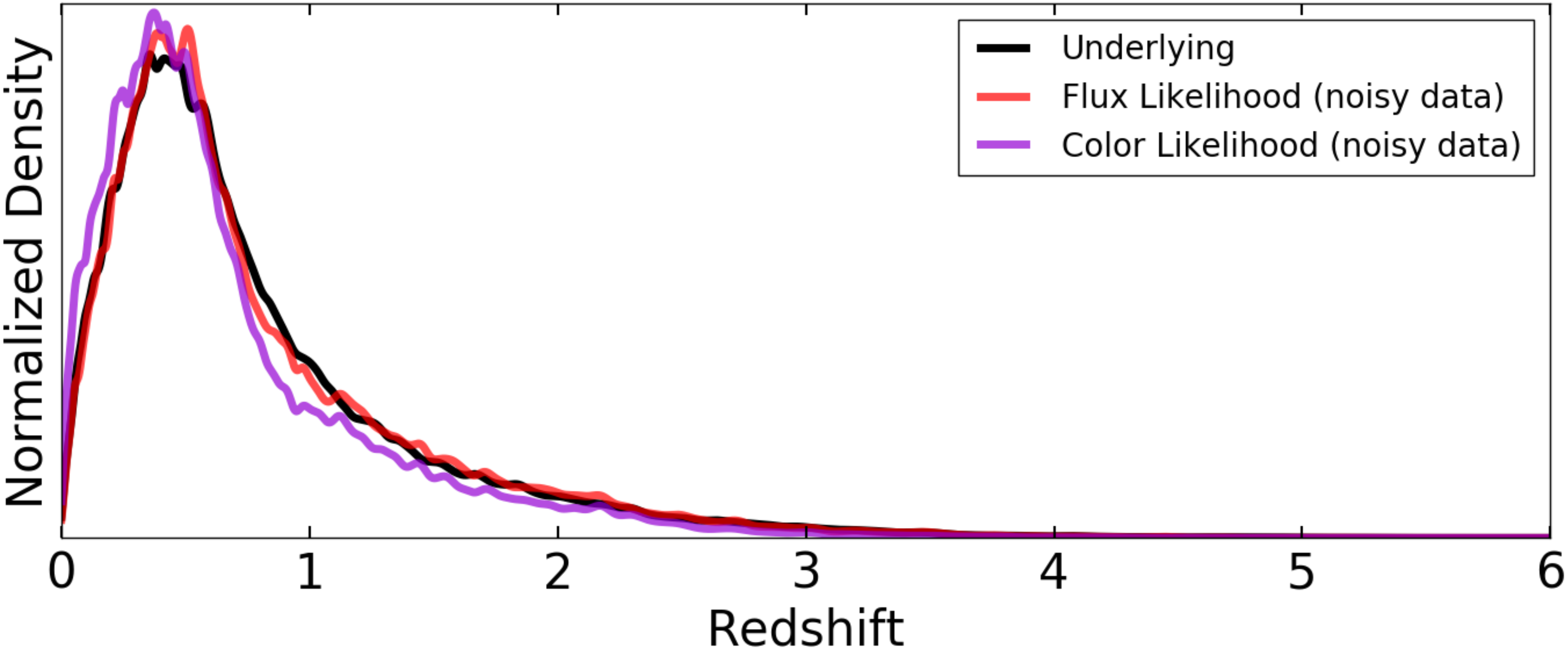
Magnitude space (“new”)

Requires good sampling in full magnitude space.

“Scale-dependent”

$$\chi^2 = \sum_i \frac{\left(\hat{F}_{g,i} - \hat{F}_{h,i}\right)^2}{\hat{\sigma}_{g,i}^2 + \hat{\sigma}_{h,i}^2}$$

Population: Mag v Color



The "Big Data" Assumption

$$P(z|g) = \sum_h P(z|h) P\left(h \middle| g, s_g = 1, \boldsymbol{S_g}\right)$$

$$\overset{?}{=} \sum_h P(z|h) \frac{P(g|h, \boldsymbol{b}) P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?
3. How do we deal with missing data?
4. How do we incorporate selection effects?
5. What is our prior?

The “Big Data” Assumption

$$P(z|g) = \sum_h P(z|h) P\left(h \middle| g, s_g = 1, \boldsymbol{S_g}\right)$$

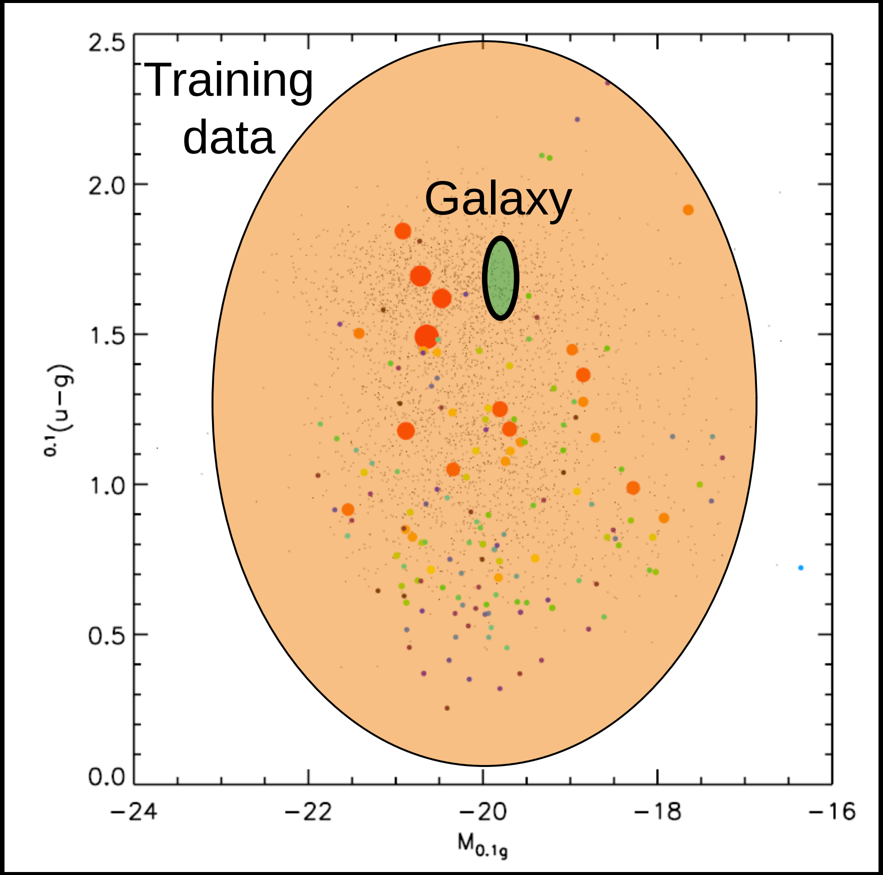
$$\stackrel{?}{=} \sum_h P(z|h) \frac{P(g|h, \boldsymbol{b}) P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?
3. How do we deal with missing data?
4. How do we incorporate selection effects?
5. What is our prior?

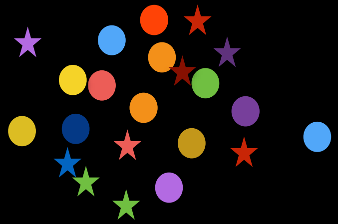
The Problem

Likelihood

$\{P(h|g)\}$

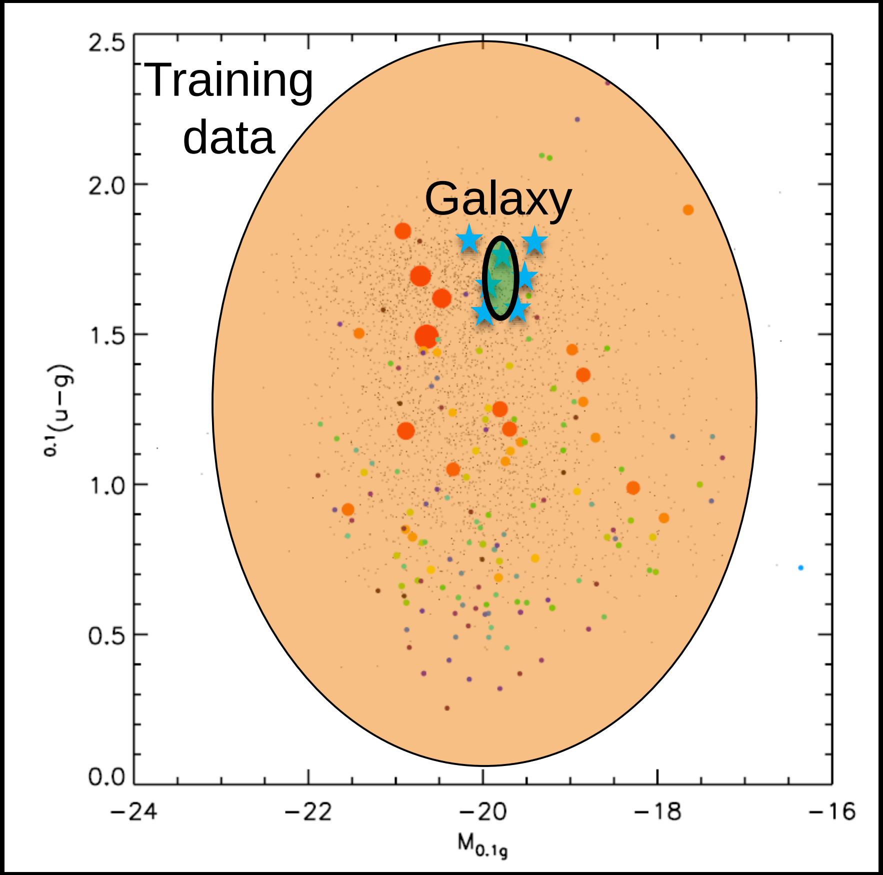


Cool et al. (2007)



Photometric training set

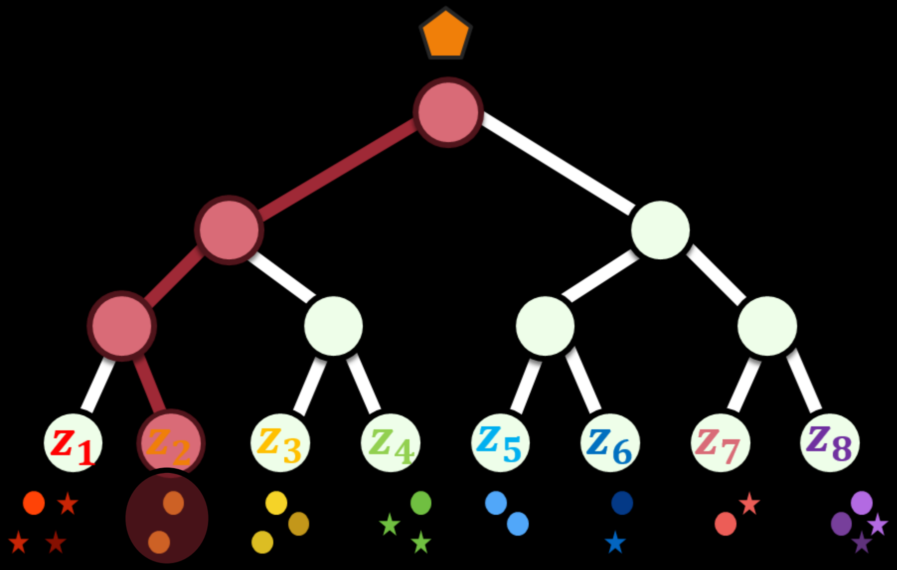
Machine Learning Approximation



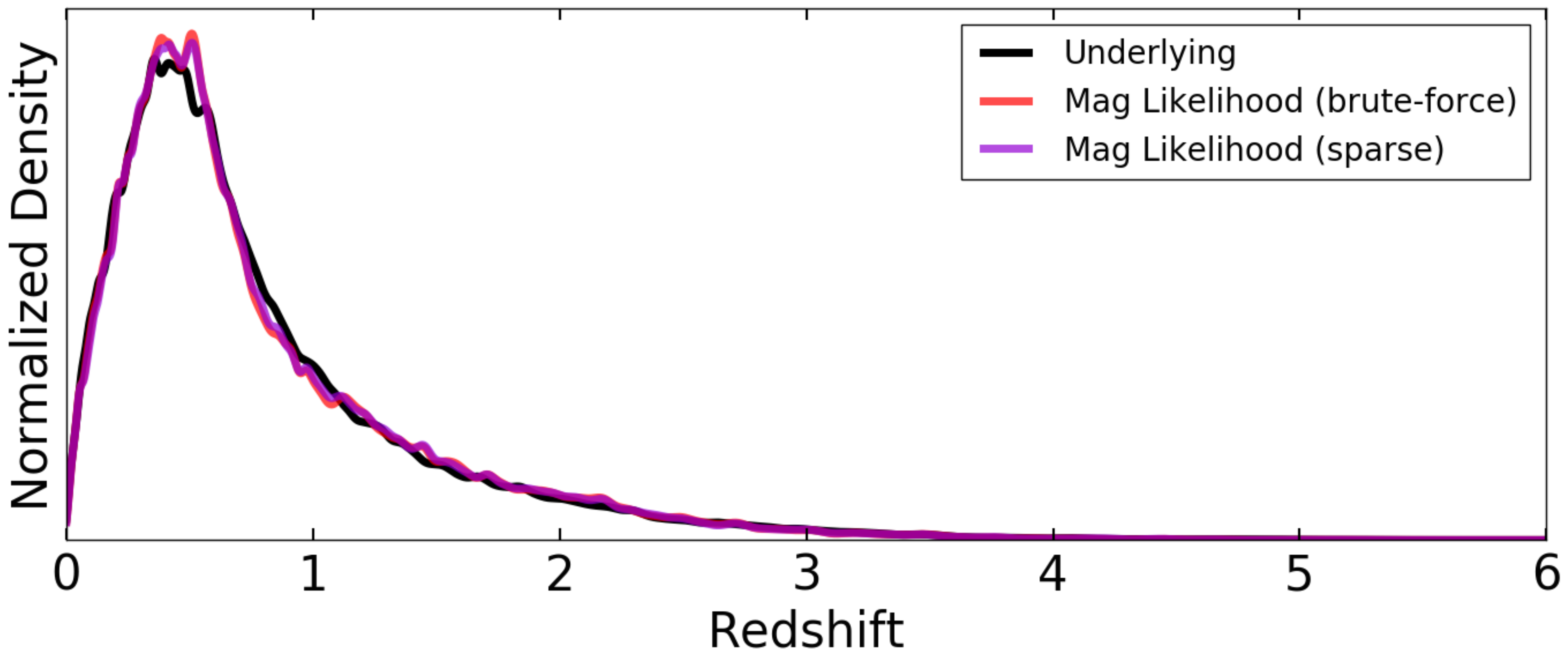
Cool et al. (2007)

Likelihood

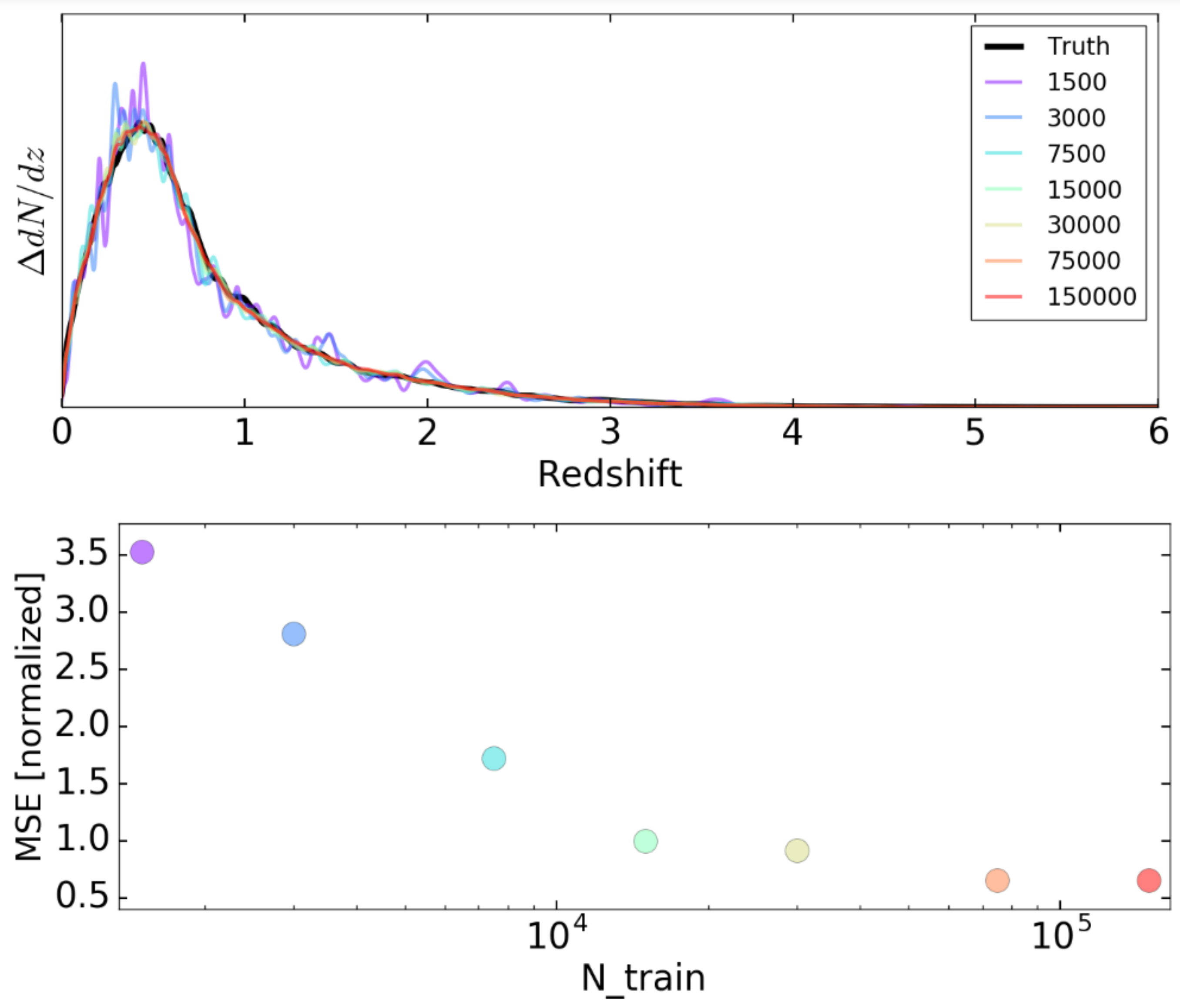
$\{\hat{P}(h|g)\}$



Population: FRANKEN-Z



How Valid is Our Approximation?



The “Big Data” Assumption

$$P(z|g) = \sum_h P(z|h) P\left(h \middle| g, s_g = 1, \boldsymbol{S_g}\right)$$

$$\overset{?}{=} \sum_h P(z|h) \frac{P(g|h, \boldsymbol{b}) P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?
3. How do we deal with missing data?
4. How do we incorporate selection effects?
5. What is our prior?

The “Big Data” Assumption

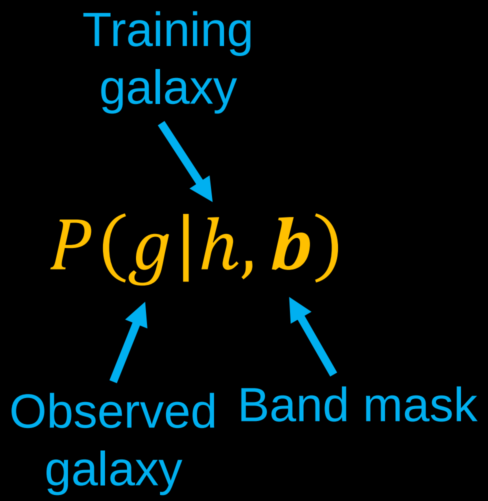
$$P(z|g) = \sum_h P(z|h) P\left(h \middle| g, s_g = 1, \boldsymbol{S_g}\right)$$

$$\overset{?}{=} \sum_h P(z|h) \frac{P(g|h, \boldsymbol{b}) P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?
3. How do we deal with missing data?
4. How do we incorporate selection effects?
5. What is our prior?

Missing Data

- How to deal with missing data?



Training galaxy

$$P(g|h,\boldsymbol{b})$$

Observed galaxy

Band mask

Now What?

- How to deal with missing data? $P(g|h, \boldsymbol{b})$

$$-2\ln L = ???$$

Naïve Likelihood: Multivariate Normal

- How to deal with missing data?

$P(g|h, \boldsymbol{b})$

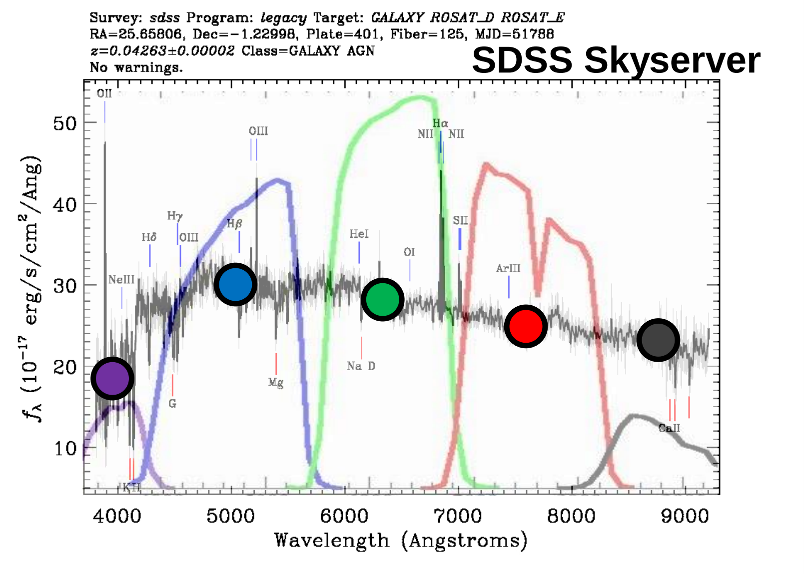
$$-2 \ln L_n \sim \chi_n^2(\delta_n)$$

$$+n \ln 2\pi + \ln\left|\hat{C}_g + \hat{C}_h\right|$$

Implications

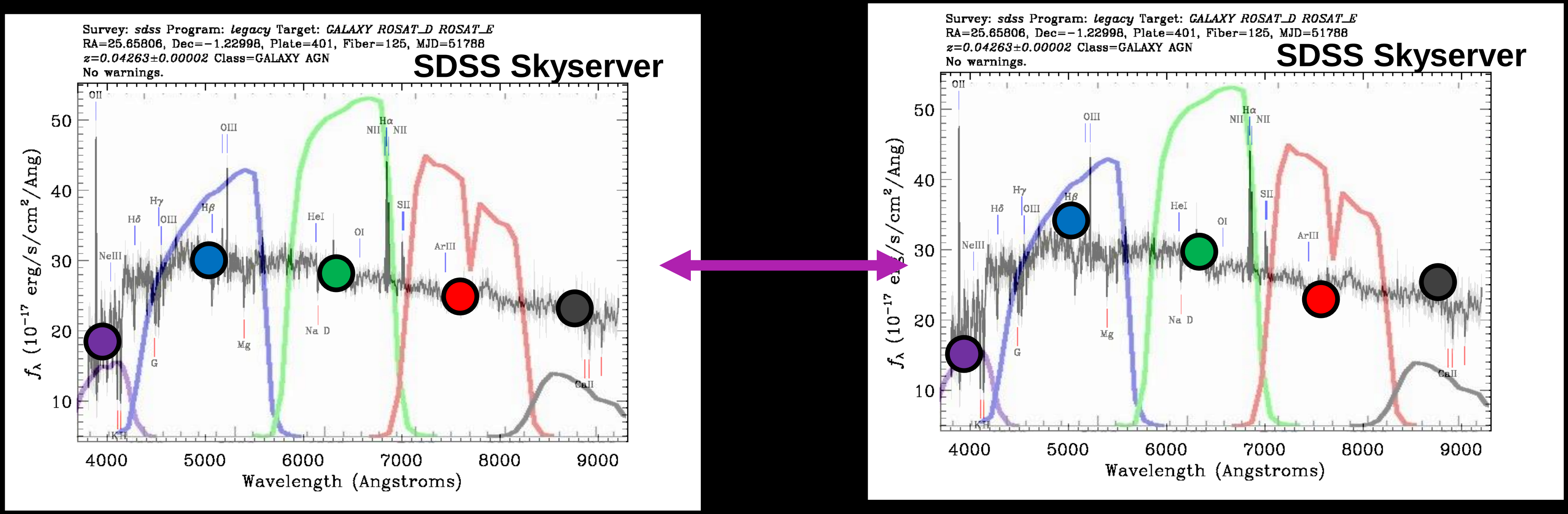
- How to deal with missing data?

$P(g|h, \boldsymbol{b})$



Implications

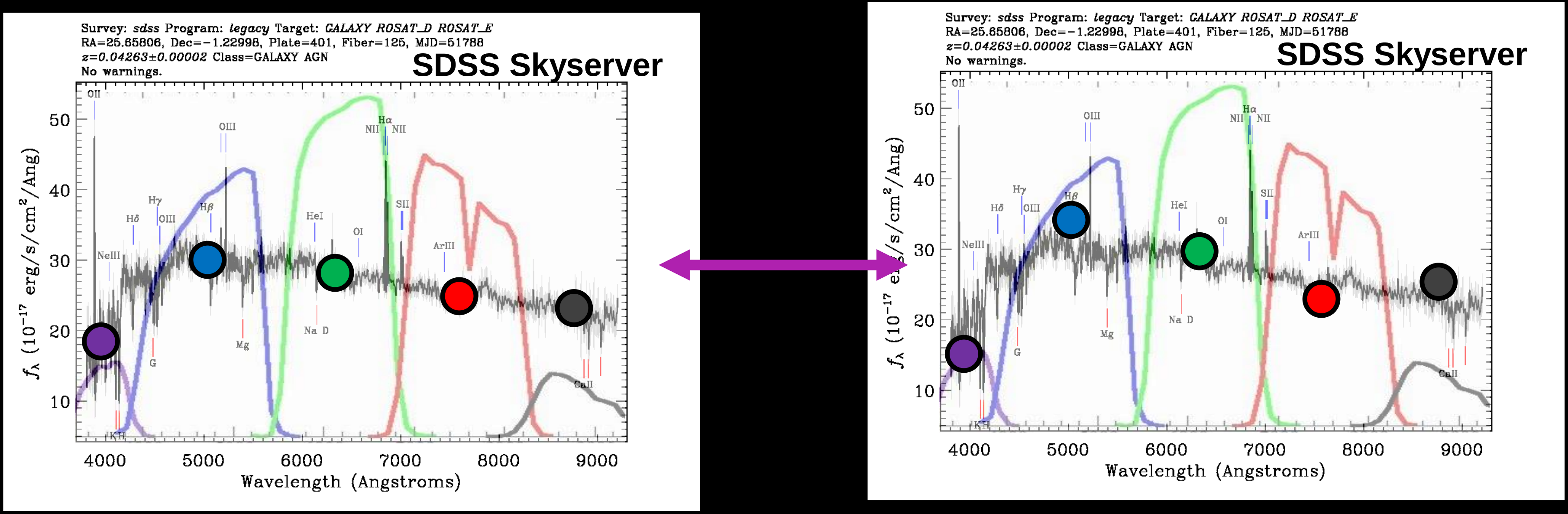
- How to deal with missing data? $P(g|h,\boldsymbol{b})$



Implications

- How to deal with missing data?

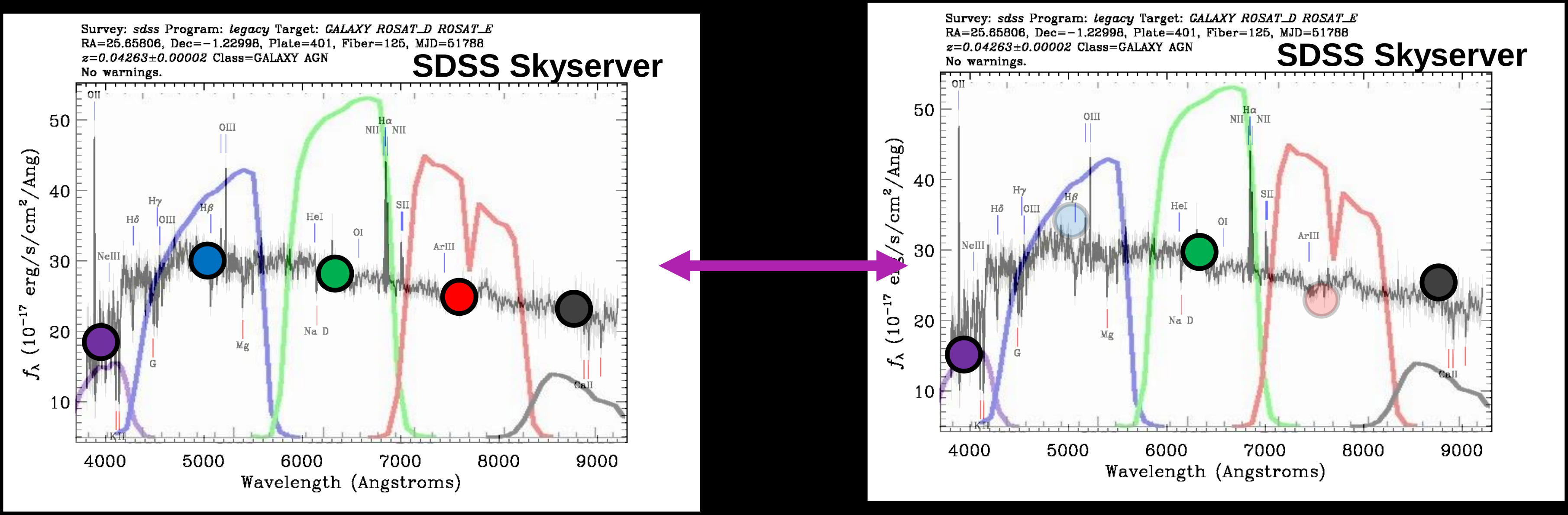
$$P(g|h,\boldsymbol{b})$$



$$-2\ln L_5 \sim \chi_5^2 + 5\ln 2\pi + \ln\left|\hat{C}_g + \hat{C}_h\right|$$

Implications

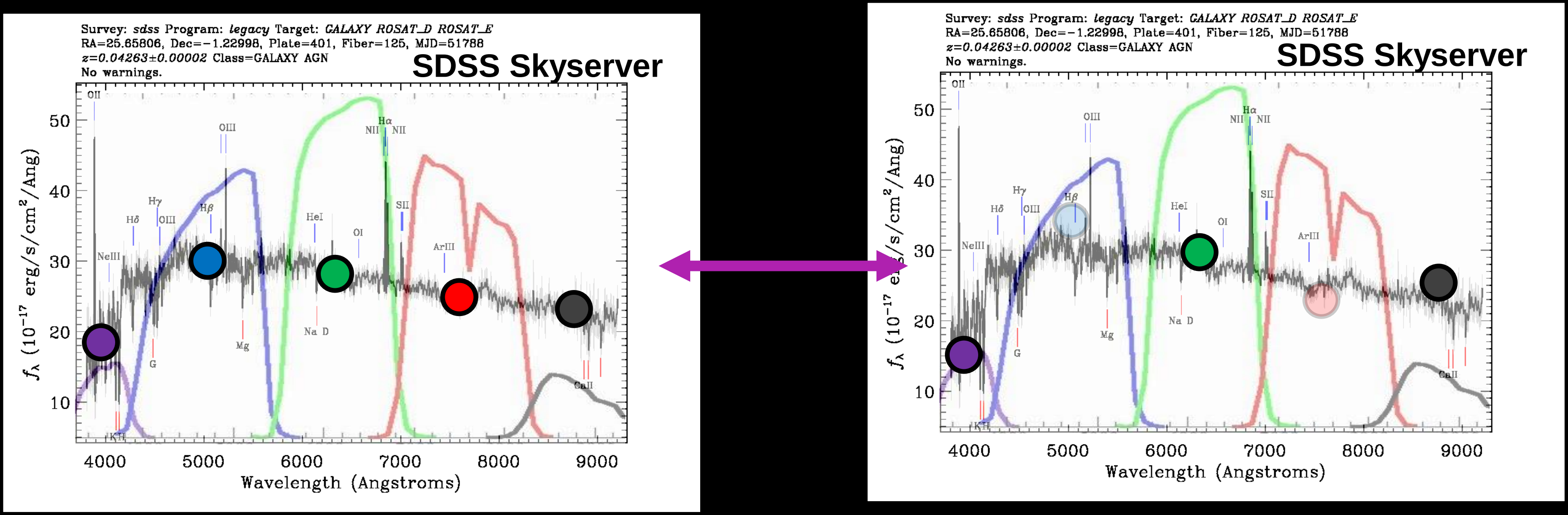
- How to deal with missing data? $P(g|h, \boldsymbol{b})$



$$-2\ln L_3 \sim \chi_3^2 + 3\ln 2\pi + \ln\left|\hat{C}_g' + \hat{C}_h'\right|$$

Solutions

- How to deal with missing data? $P(g|h,\boldsymbol{b})$



$$-2\ln L \sim X_n - X'_n, \qquad X_n, X'_n \overset{\text{i.i.d.}}{\sim} \chi^2_n(\delta_n)$$

Implementation

- How to deal with missing data? $P(g|h,\boldsymbol{b})$

$$-2\ln L = ???$$

some transformation of doubly-non-central F-distribution

Implementation

- How to deal with missing data?

$P(g|h,\boldsymbol{b})$

ignore
non-centrality

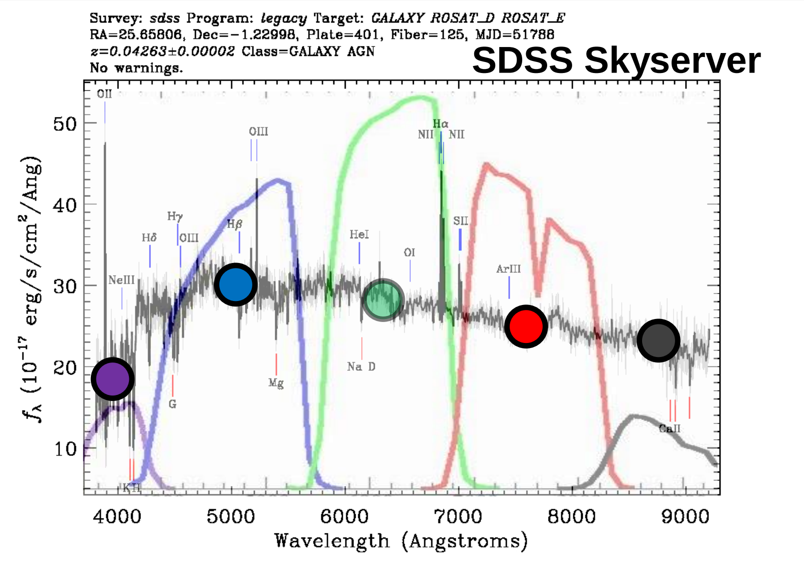
$$-2\ln L = \chi_n^2 - N_b$$

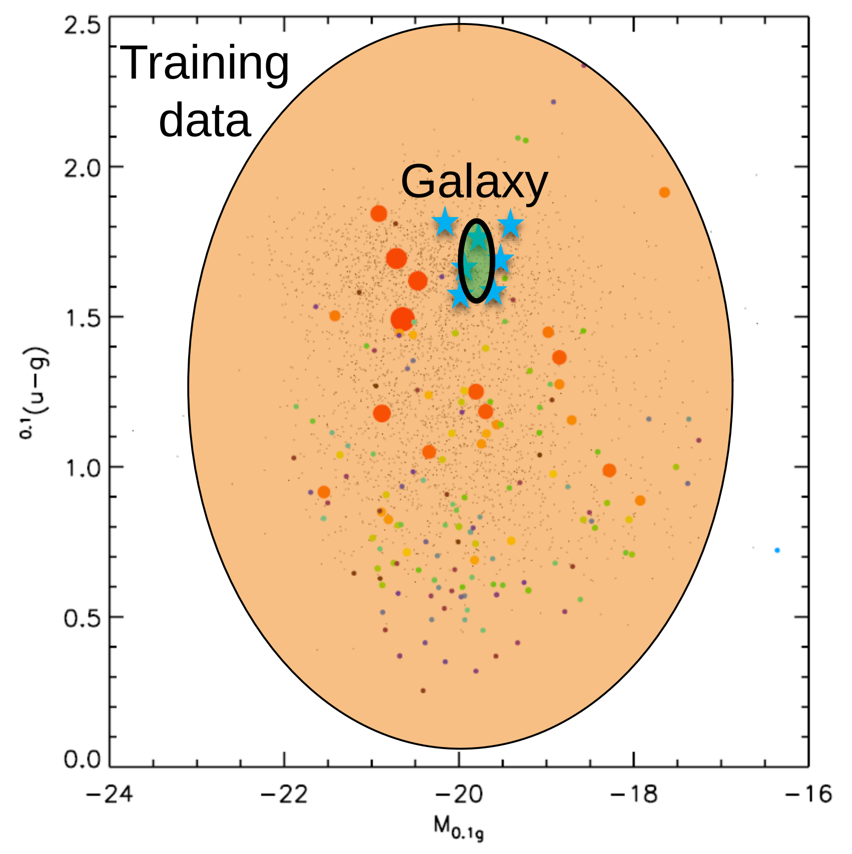
first-order
correction

$$\chi^2 = \sum_i \frac{\left(\hat{F}_{g,i} - \hat{F}_{h,i}\right)^2}{\hat{\sigma}_{g,i}^2 + \hat{\sigma}_{h,i}^2}$$

Missing Data: Searching for Neighbors

- How to deal with missing data? $\hat{P}(g|h,\boldsymbol{b})$





Cool et al. (2007)

The "Big Data" Assumption

$$P(z|g) = \sum_h P(z|h) P\left(h \middle| g, s_g = 1, \boldsymbol{S_g}\right)$$

$$\overset{?}{=} \sum_h P(z|h) \frac{P(g|h, \boldsymbol{b}) P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?
3. How do we deal with missing data?
4. How do we incorporate selection effects?
5. What is our prior?

The “Big Data” Assumption

$$P(z|g) = \sum_h P(z|h) P\left(h \middle| g, s_g = 1, \boldsymbol{S_g}\right)$$

$$\overset{?}{=} \sum_h P(z|h) \frac{P(g|h, \boldsymbol{b}) P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?
3. How do we deal with missing data?
4. How do we incorporate selection effects?
5. What is our prior?

Selection Effects

- How to deal with selection effects? $P(g|h)$

Selection Effects

- How to deal with selection effects? $P\left(\hat{\boldsymbol{F}}_{\boldsymbol{g}} \middle| g\right)$

Selection Effects

Selection effect(s)

- How to deal with selection effects? $P\left(\hat{\boldsymbol{F}}_{\boldsymbol{g}} \middle| g, s_g = 1, \boldsymbol{S}_g\right)$

Binary selection flag
(1=in/0=out)

Selection Effects

Selection effect(s)

- How to deal with selection effects? $P\left(\hat{\boldsymbol{F}}_{\boldsymbol{g}} \middle| g, s_g = 1, \boldsymbol{S}_g\right)$

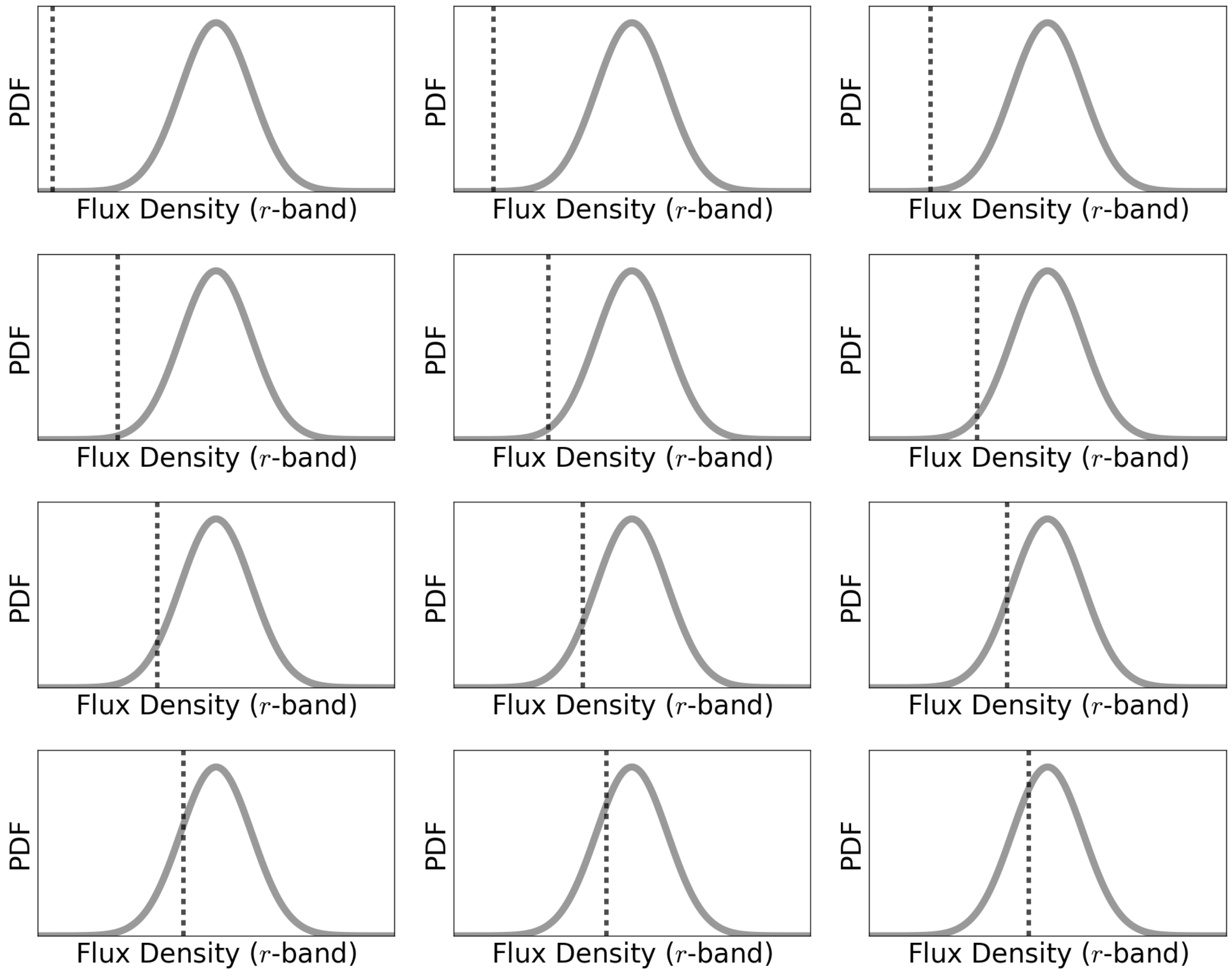
Binary selection flag
(1=in/0=out)

Selection Probability Original PDF

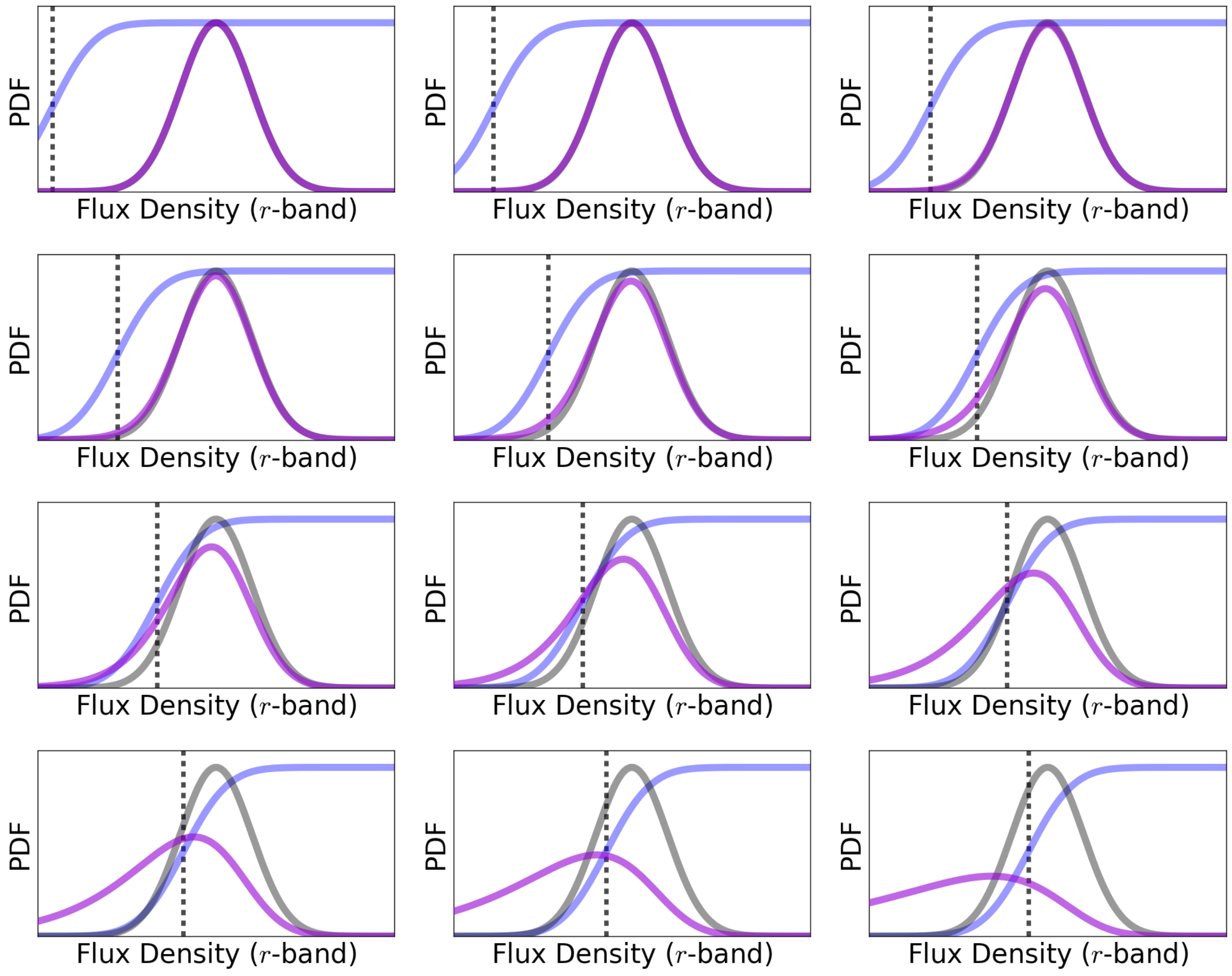
$$\frac{P\left(s_g = 1 \middle| \hat{\boldsymbol{F}}_g, \boldsymbol{S}_g\right) P\left(\hat{\boldsymbol{F}}_g \middle| g\right)}{P\left(s_g = 1 \middle| g, \boldsymbol{S}_g\right)}$$

Marginalized
Selection Probability

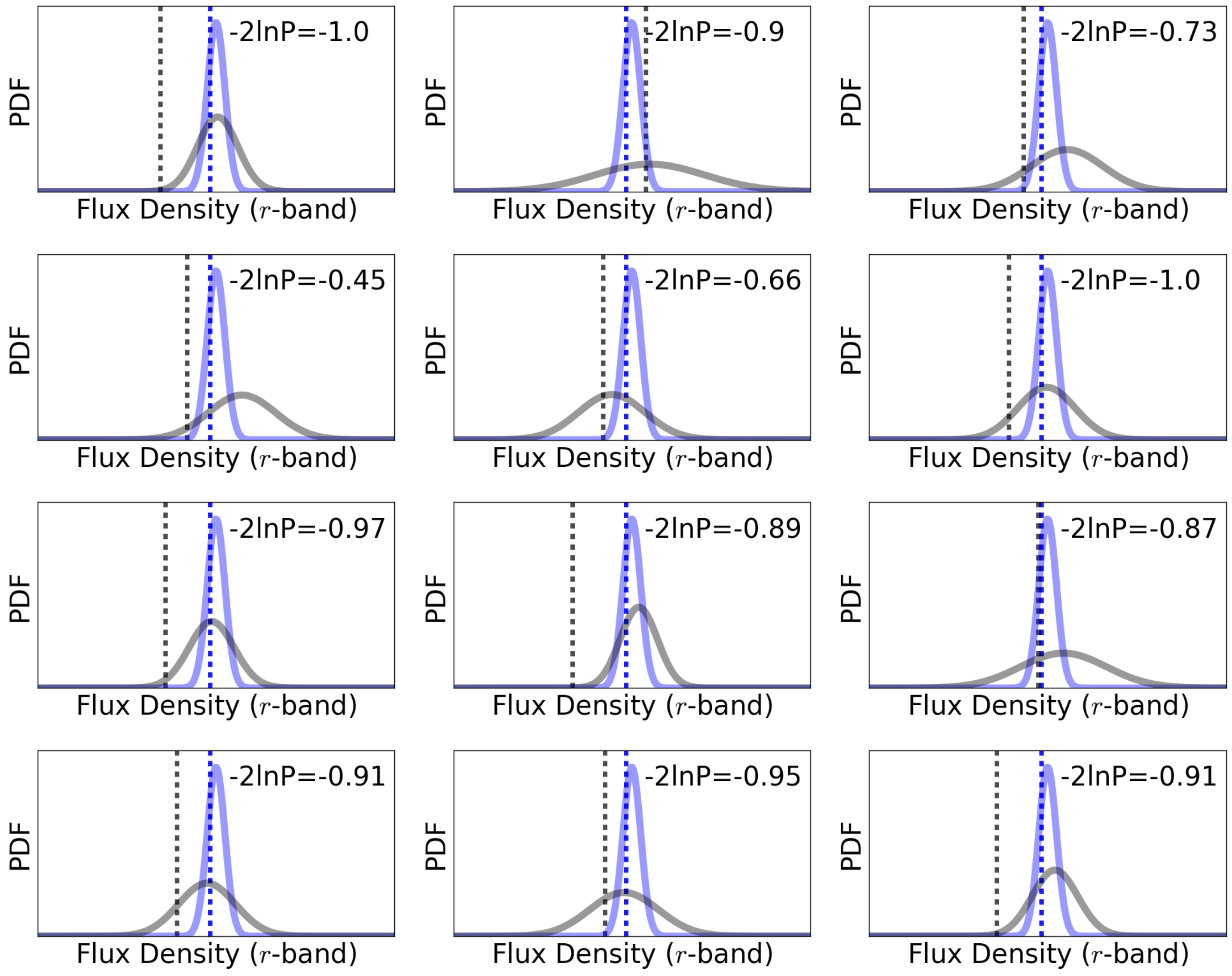
Application: Signal-to-Noise Cut



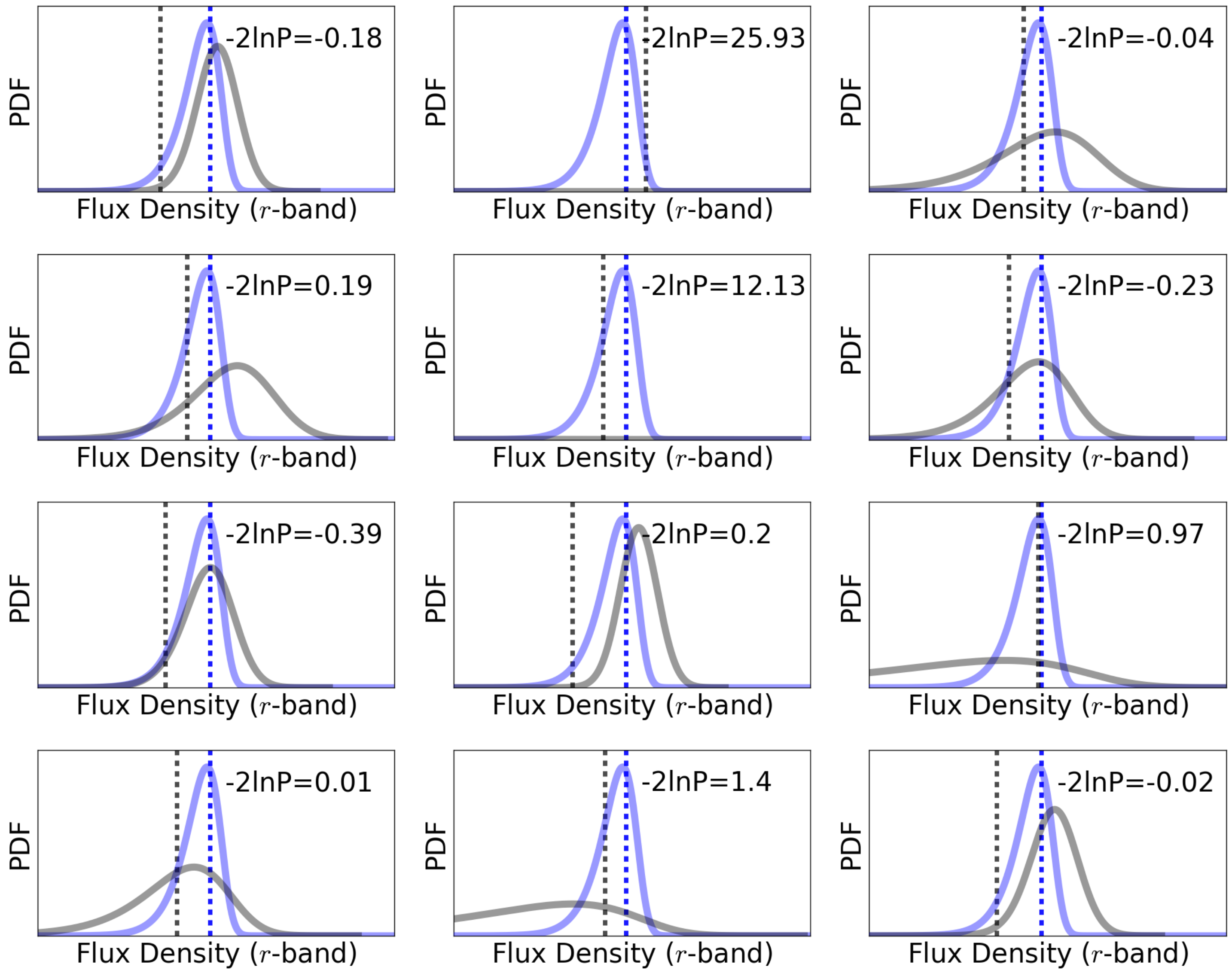
Application: Signal-to-Noise Cut



Application: Signal-to-Noise Cut



Application: Signal-to-Noise Cut



The “Big Data” Assumption

$$P(z|g) = \sum_h P(z|h) P\left(h \middle| g, s_g = 1, \boldsymbol{S_g}\right)$$

$$\stackrel{?}{=} \sum_h P(z|h) \frac{P(g|h, \boldsymbol{b}) P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?
3. How do we deal with missing data?
4. How do we incorporate selection effects?
5. What is our prior?

The “Big Data” Assumption

$$P(z|g) = \sum_h P(z|h) P\left(h \middle| g, s_g = 1, \boldsymbol{S_g}\right)$$

$$\overset{?}{=} \sum_h P(z|h) \frac{P(g|h, \boldsymbol{b}) P(h)}{P(g)}$$

1. What is this likelihood?
2. How do we compute it?
3. How do we deal with missing data?
4. How do we incorporate selection effects?
5. What is our prior?

Mixing and Matching Galaxies

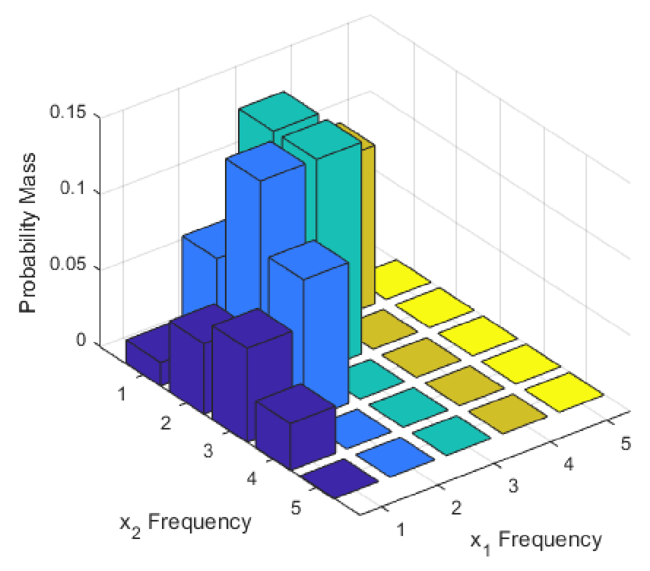
- How to deal with population mismatch?

Based on Leistedt, Peiris, & Mortlock (2016)

Mixing and Matching Galaxies

- How to deal with population mismatch?

$$\boldsymbol{p}'_g \sim \mathrm{Mult}\left(n = 1, \boldsymbol{p} = \boldsymbol{p}_g\right)$$



Trinomial distribution from Mathworks.

N_h categories with probability vector

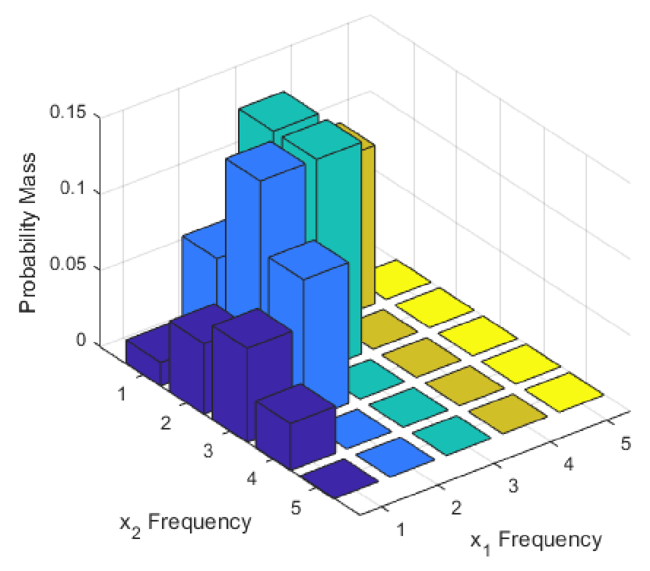
$$\boldsymbol{p}_g = \{P(h|g)\}$$

Based on Leistedt, Peiris, & Mortlock (2016)

Mixing and Matching Galaxies

- How to deal with population mismatch?

$$\boldsymbol{p}'_g \sim \mathrm{Mult}\left(n = 1, \boldsymbol{p} = \boldsymbol{p}_g\right)$$



Trinomial distribution from Mathworks.

N_h categories with probability vector

$$\boldsymbol{p}_g = \{P(h|g)\}$$

$$\boldsymbol{h} = \{0,1,2\}$$

$$\boldsymbol{p}'_g = \{0,0,1\} \rightarrow h'_g = 2$$

Based on Leistedt, Peiris, & Mortlock (2016)

Mixing and Matching Galaxies

- How to deal with population mismatch?

$$\boldsymbol{p}'_g \sim \mathrm{Mult}\left(n = 1, \boldsymbol{p} = \boldsymbol{p}_g\right)$$

$$P(\boldsymbol{h}) \overset{?}{=} \boldsymbol{n} \sim \sum\nolimits_{g \in \boldsymbol{g}} \boldsymbol{p}'_g$$

Based on Leistedt, Peiris, & Mortlock (2016)

Mixing and Matching Galaxies

- How to deal with population mismatch?

$$\boldsymbol{p}'_g \sim \mathrm{Mult}(n = 1, \boldsymbol{p} = \boldsymbol{p}_g)$$

$$\boldsymbol{n} \sim \sum_{g \in \boldsymbol{g}} \boldsymbol{p}'_g$$

Concentration

$$\boldsymbol{P}(\boldsymbol{h}) \sim \mathrm{Dir}(\boldsymbol{w} | \boldsymbol{\alpha} = \boldsymbol{n} + 1)$$

Population weights

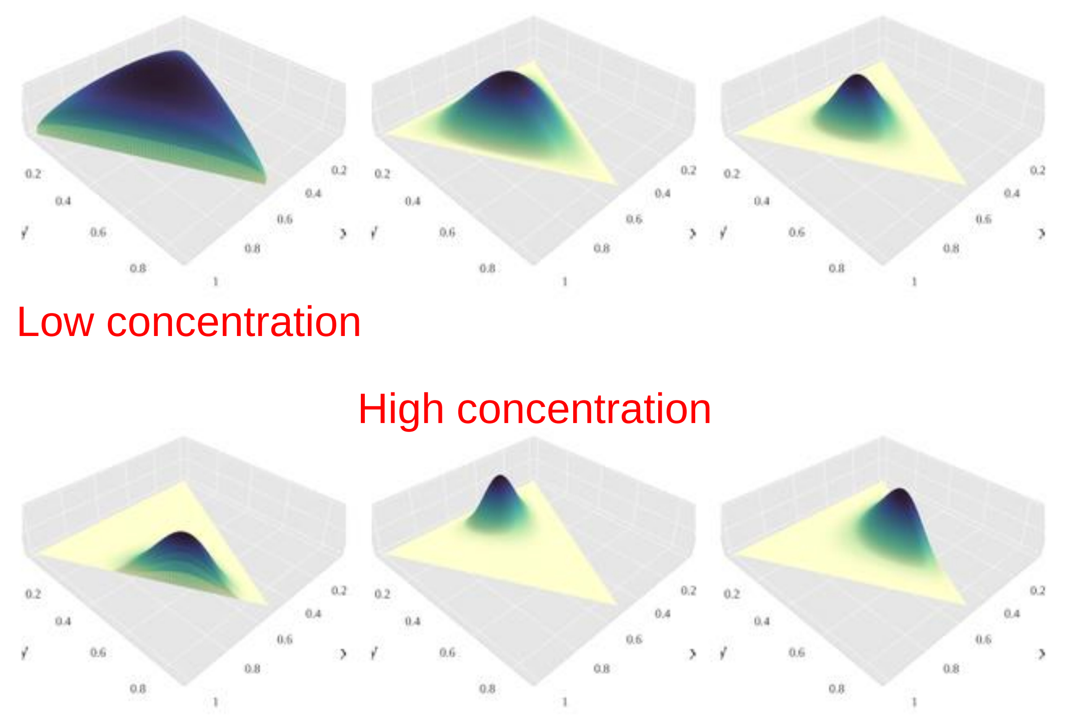
Counts

Based on Leistedt, Peiris, & Mortlock (2016)

Mixing and Matching Galaxies

Courtesy of Wikipedia.

- How to de



Concentration

$$\boldsymbol{P}(\boldsymbol{h}) \sim \mathrm{Dir}(\boldsymbol{w}|\boldsymbol{\alpha} = \boldsymbol{n} + 1)$$

Population weights

Counts

Based on Leistedt, Peiris, & Mortlock (2016)

Hierarchical Modeling

Hierarchical
posteriors

- How to deal with population mismatch? $P\left(\boldsymbol{w}, \left\{\wp_g\right\} \middle| \boldsymbol{D}, \boldsymbol{S}\right)$

Population weights
(training set)

Based on Leistedt, Peiris, & Mortlock (2016)

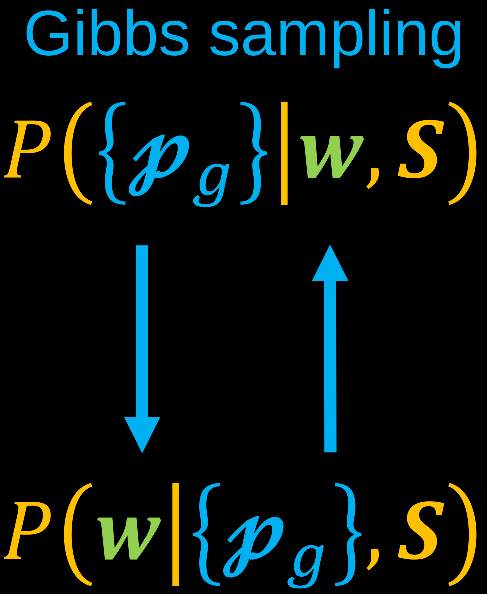
Hierarchical Modeling

- How to deal with population mismatch?

$$P\left(\boldsymbol{w}, \{\boldsymbol{\wp}_g\} \middle| \boldsymbol{D}, \boldsymbol{S}\right)$$

Hierarchical posteriors

Population weights (training set)



Gibbs sampling

$$P\left(\{\boldsymbol{\wp}_g\} \middle| \boldsymbol{w}, \boldsymbol{S}\right)$$

$$P\left(\boldsymbol{w} \middle| \{\boldsymbol{\wp}_g\}, \boldsymbol{S}\right)$$

Based on Leistedt, Peiris, & Mortlock (2016)

Hierarchical Modeling

Hierarchical posteriors

- How to deal with population mismatch? $P\left(\boldsymbol{w}, \{\boldsymbol{\wp}_g\} \middle| \boldsymbol{D}, \boldsymbol{S}\right)$

Population weights (training set)

Gibbs sampling

$$P\left(\{\boldsymbol{\wp}_g\} \middle| \boldsymbol{w}, \boldsymbol{S}\right)$$

$$\downarrow \quad \uparrow$$

$$P\left(\boldsymbol{w} \middle| \{\boldsymbol{\wp}_g\}, \boldsymbol{S}\right)$$

1. Sample hierarchical posteriors: $\boldsymbol{\wp}_g^{(i)} \sim \mathrm{Mult}\left(\mathrm{n} = 1, \mathbf{p} = \boldsymbol{p}_g \boldsymbol{w}^{(i-1)}\right)$

Based on Leistedt, Peiris, & Mortlock (2016)

Hierarchical Modeling

Hierarchical posteriors

- How to deal with population mismatch? $P\left(\boldsymbol{w}, \{\wp_g\} \middle| \boldsymbol{D}, \boldsymbol{S}\right)$

Population weights
(training set)

Gibbs sampling

$$\begin{tikzcd}
P\left(\{\wp_g\} \middle| \boldsymbol{w}, \boldsymbol{S}\right) \arrow[d, shift right=3] \\
P\left(\boldsymbol{w} \middle| \{\wp_g\}, \boldsymbol{S}\right) \arrow[u, shift right=3]
\end{tikzcd}$$

1. Sample hierarchical posteriors: $\wp_g^{(i)} \sim \mathrm{Mult}\left(\mathrm{n} = 1, \mathbf{p} = \boldsymbol{p}_g \boldsymbol{w}^{(i-1)}\right)$
2. Compute counts: $\boldsymbol{n}^{(i)} \sim \sum_{g \in \boldsymbol{g}} \wp_g^{(i)}$

Based on Leistedt, Peiris, & Mortlock (2016)

Hierarchical Modeling

Hierarchical posteriors

- How to deal with population mismatch? $P(\boldsymbol{w}, \{\wp_g\} | \boldsymbol{D}, \boldsymbol{S})$

Population weights (training set)

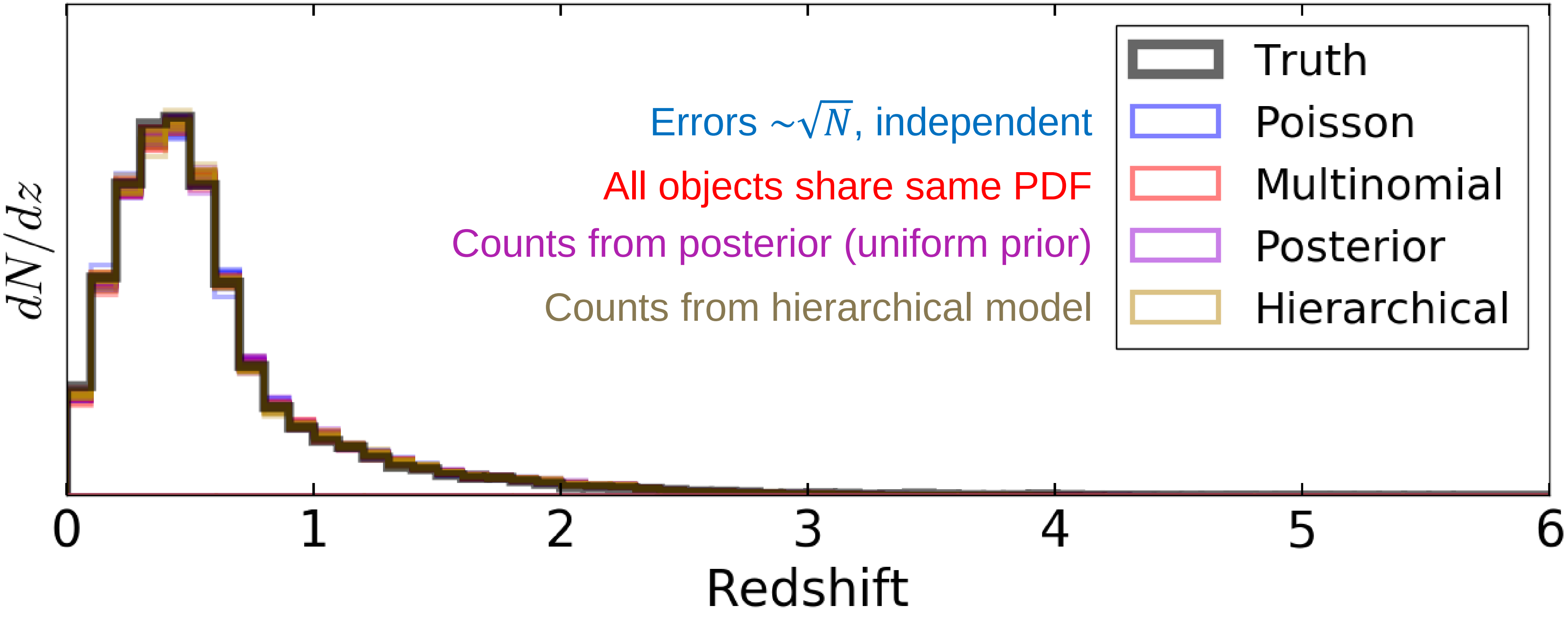
Gibbs sampling

$$\begin{tikzcd} P(\{\wp_g\} | \boldsymbol{w}, \boldsymbol{S}) \arrow[d, shift right=2] \\ P(\boldsymbol{w} | \{\wp_g\}, \boldsymbol{S}) \arrow[u, shift right=2] \end{tikzcd}$$

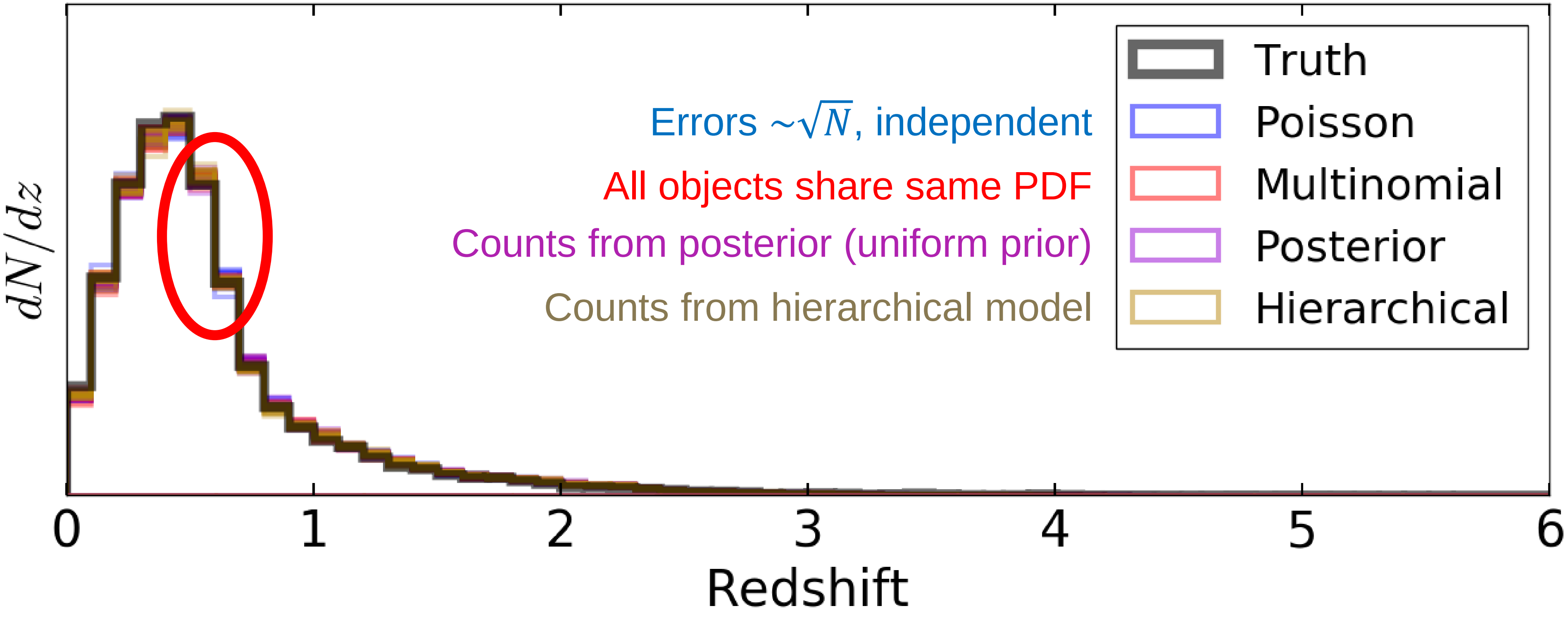
1. Sample hierarchical posteriors: $\wp_g^{(i)} \sim \mathrm{Mult}(\mathrm{n} = 1, \mathbf{p} = \boldsymbol{p}_g \boldsymbol{w}^{(i-1)})$
2. Compute counts: $\boldsymbol{n}^{(i)} \sim \sum_{g \in \boldsymbol{g}} \wp_g^{(i)}$
3. Sample weights: $\boldsymbol{w}^{(i)} \sim \mathrm{Dir}(\boldsymbol{w} | \boldsymbol{n}^{(i)} + 1)$

Based on Leistedt, Peiris, & Mortlock (2016)

Application to Mock Data



Application to Mock Data



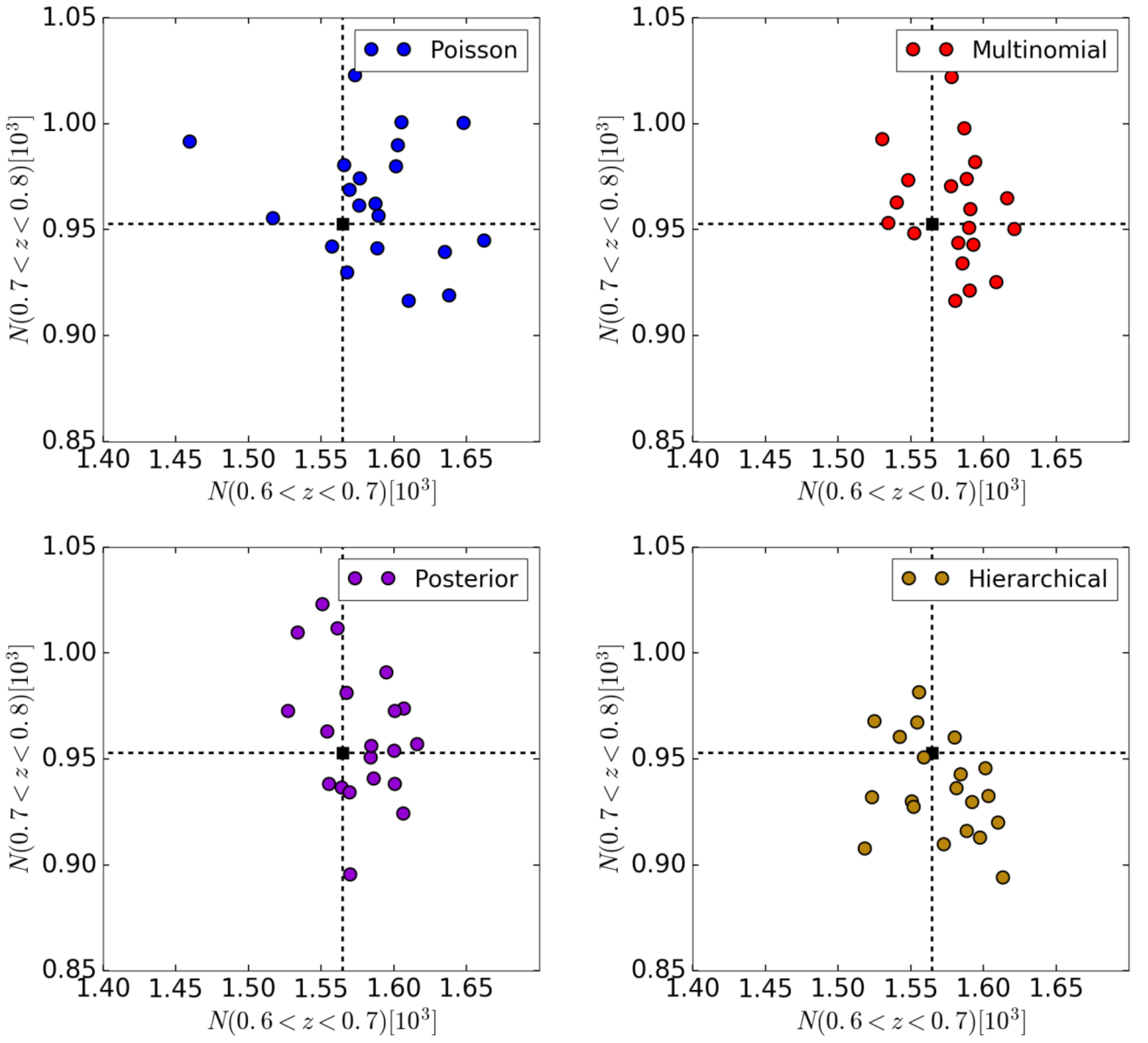
Errors $\sim\sqrt{N}$, independent

All objects share same PDF

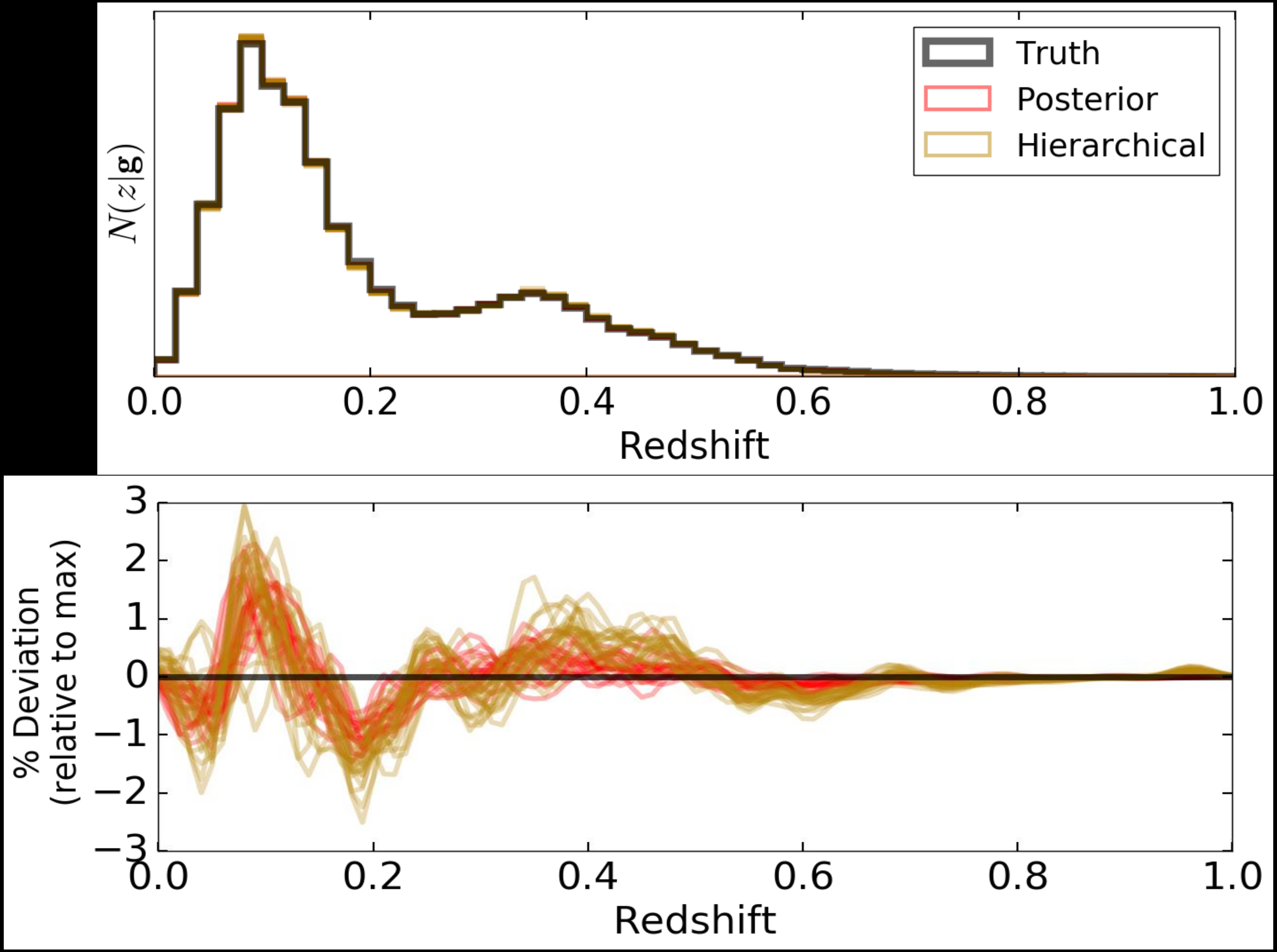
Counts from posterior (uniform prior)

Counts from hierarchical model

Application to Mock Data



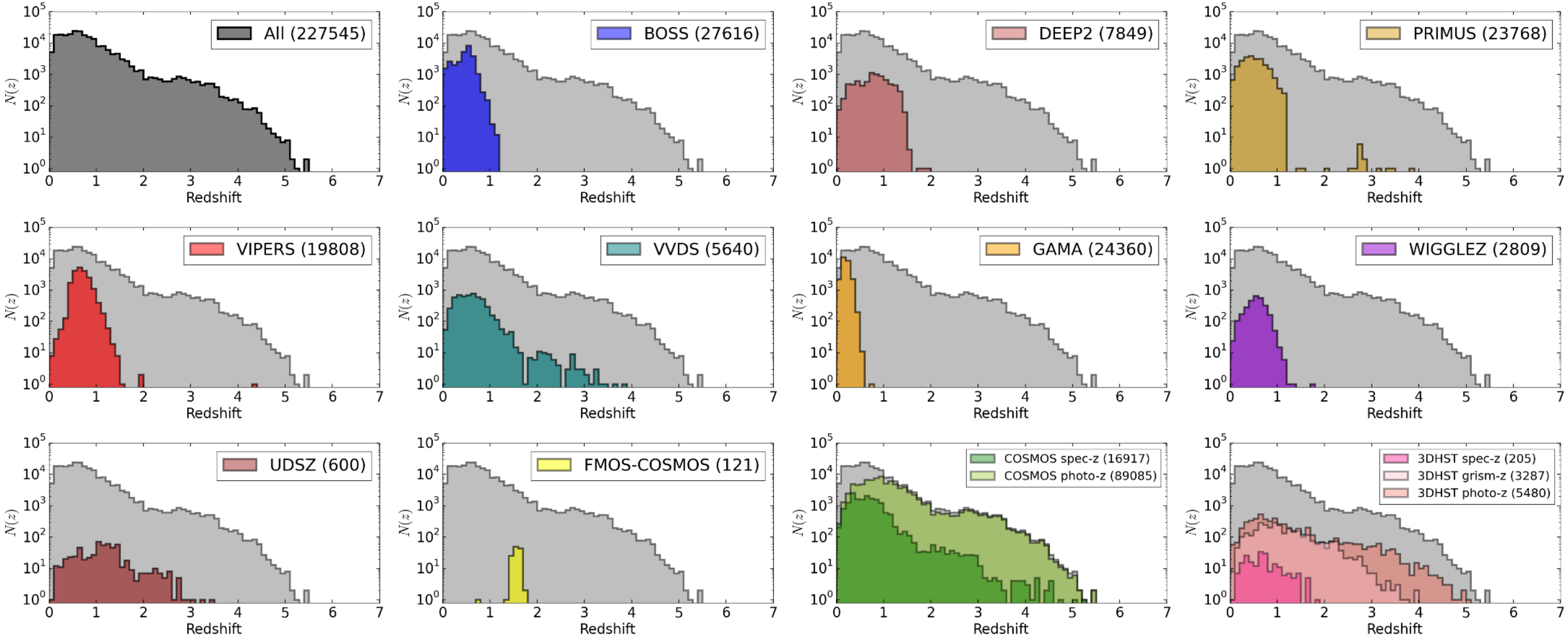
Tests on ~100k SDSS Galaxies



Next Steps

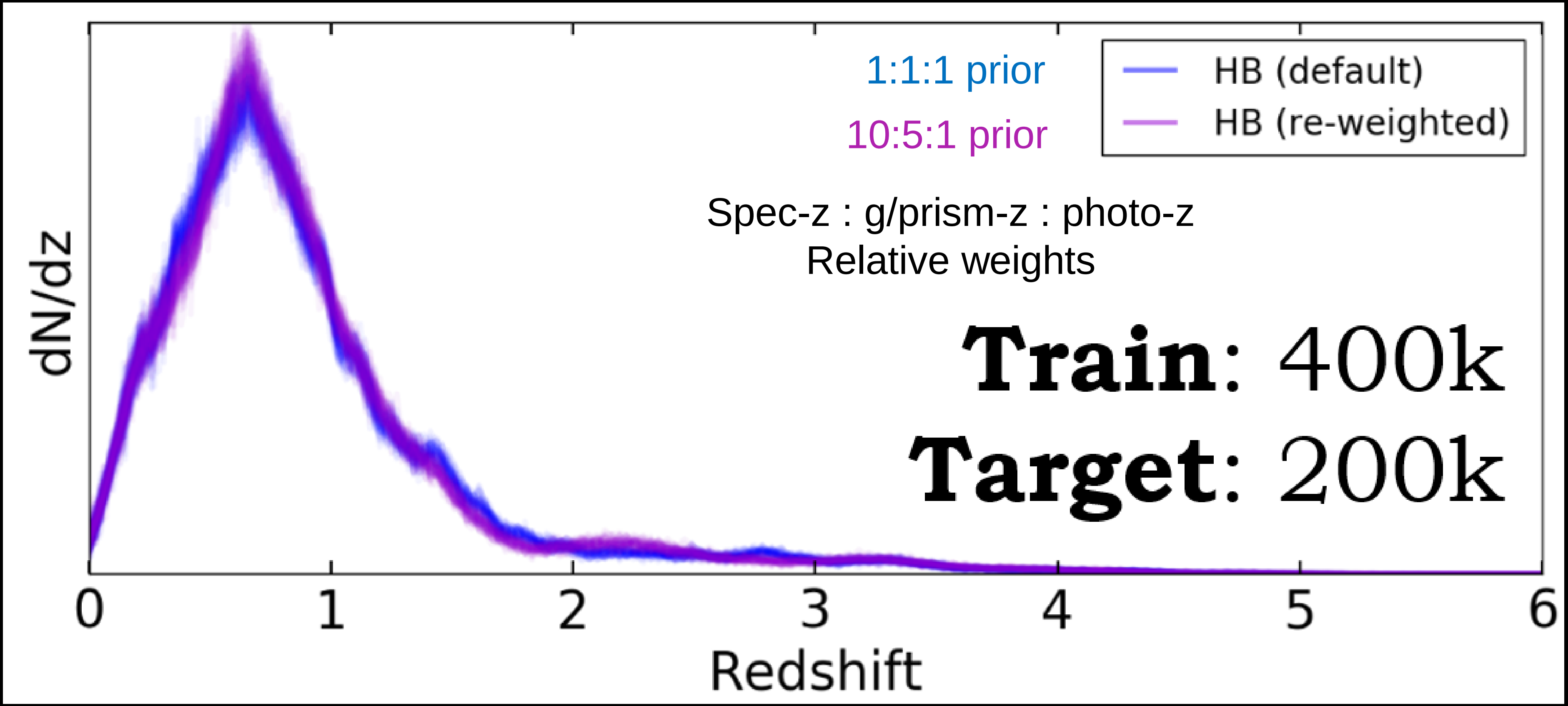
Robustness of Training Data/Priors

- Hyper Suprime-Cam (HSC) SSP has ~380k objects taken from 11 surveys.
- Wide variety of selection criteria, data quality/reliability.



Note: old plot from early internal data release.

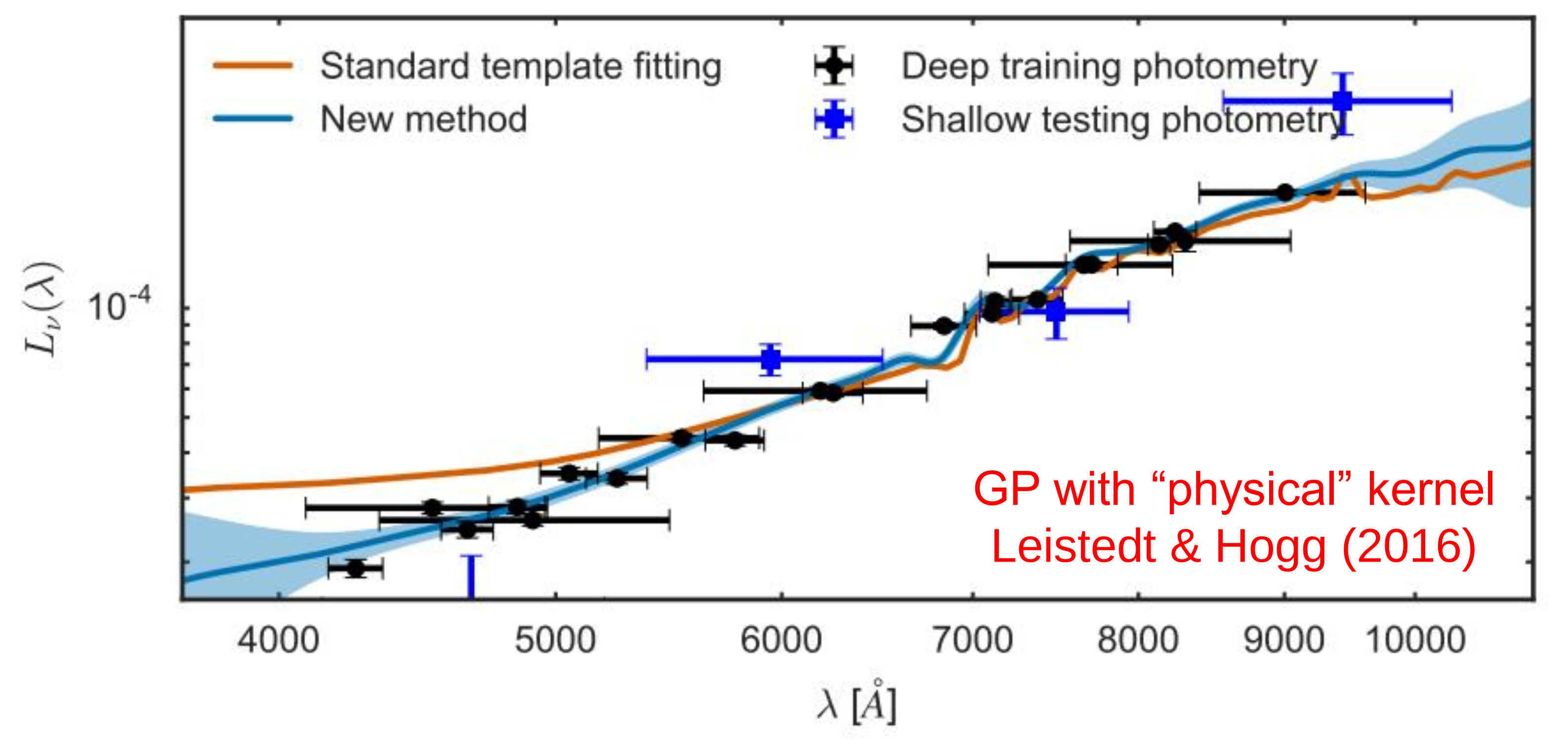
Robustness of Training Data/Priors



Preliminary

Incorporating “Physical” Priors

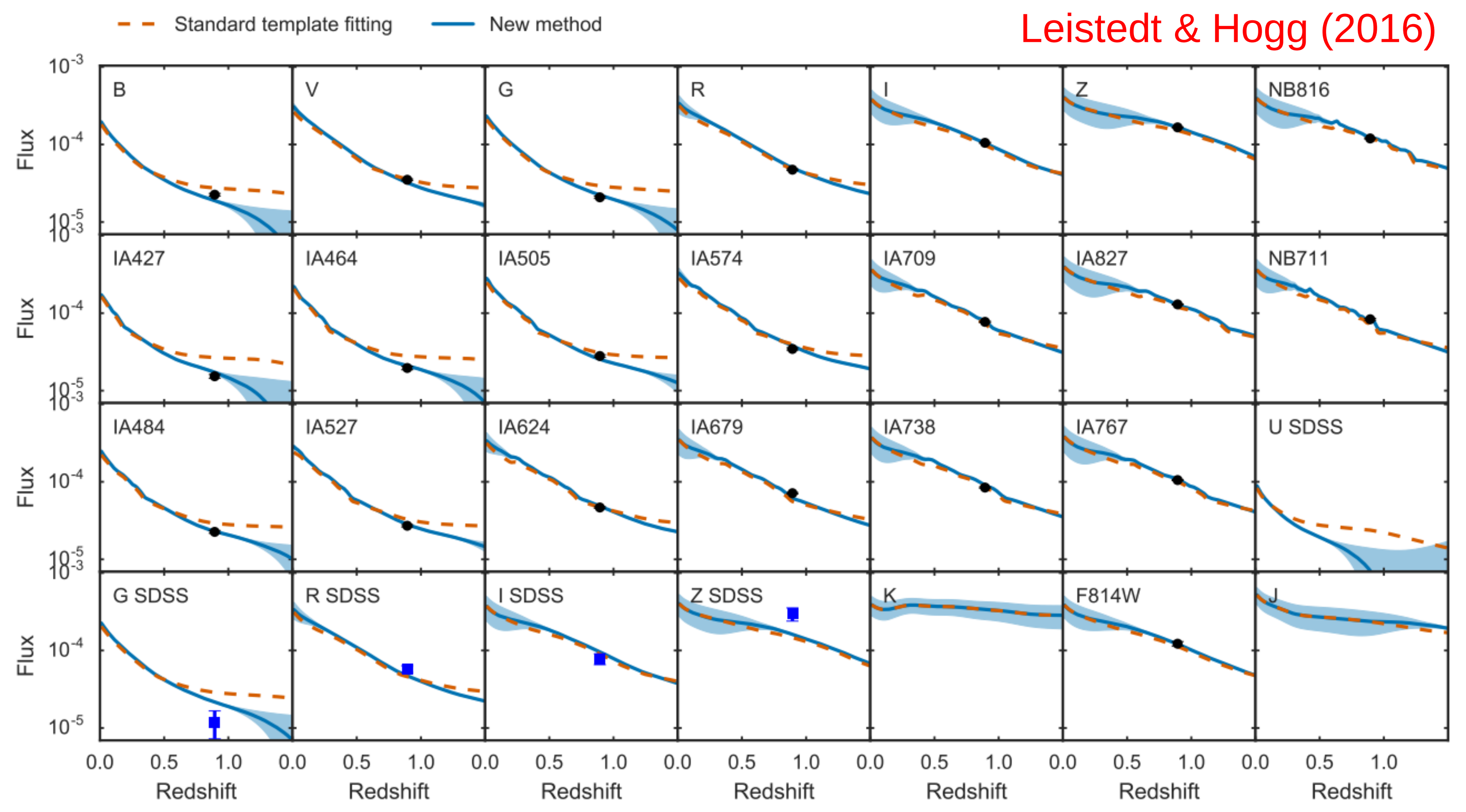
- Moving a galaxy from one redshift to another is a smooth, physical process that is well-understood.
- Want to incorporate this into our priors/predictions.



GP with “physical” kernel
Leistedt & Hogg (2016)

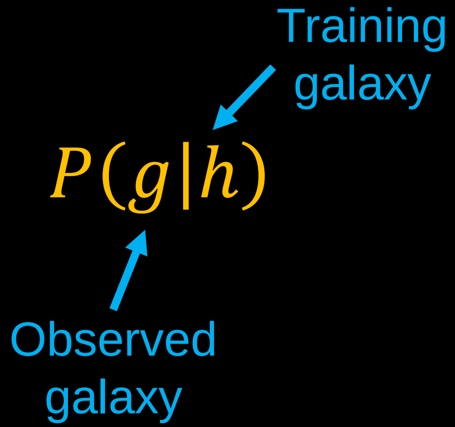
Incorporating “Physical” Priors

- Can use to augment training data.
- Straightforward to “impute” missing values.



Likelihood / Distance Metric

- What metric(s) to use?
  - More sophisticated machine learning methods could be used to compute posterior samples (or possibly likelihoods) over complex domains.



$P(g|h)$

Training galaxy

Observed galaxy

NGC 4414

NGC 5457

Summary

- Photometric redshifts (photo-z’s) are an integral part of modern “big data” extragalactic science.
- Large training datasets gives new opportunities to develop Bayesian, data-driven photo-z’s.
- Taking advantage of these datasets requires dealing with real-world problems (e.g., biased training data) using a variety of statistical methods (e.g., hierarchical Bayes).
- Early results look promising!

Code is available!

Although still under active development, code, tutorials, and rough draft of a paper are online at:

github.com/joshspeagle/frankenz