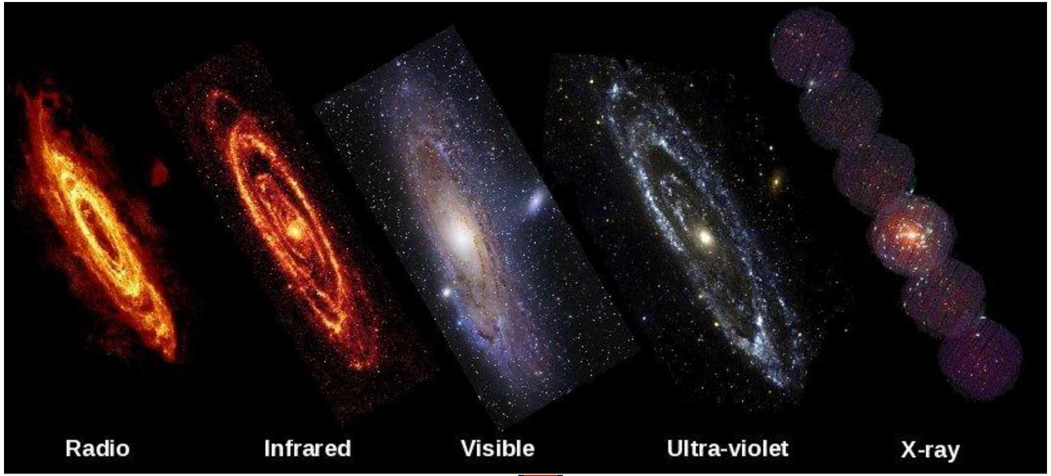
A Practical Guide to Bayesian SED Fitting

Joel Leja, Adam Carnall

Galaxy properties are inferred by fitting spectral energy distributions (SEDs) with models. Take beautiful galaxy data:



The Andromeda Galaxy
Planck / NASA / ESA

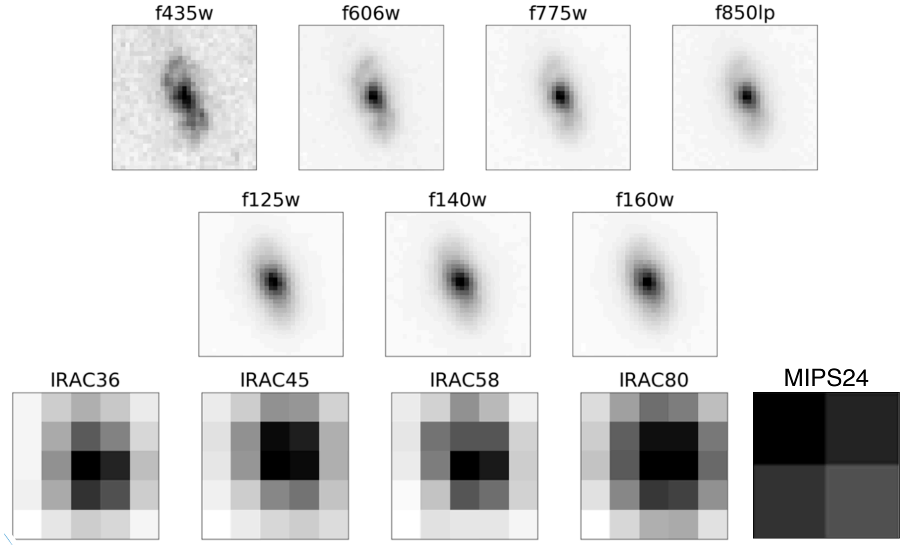
… and use models to turn them into (*even more beautiful*) inferred parameters.

stellar mass	dust content
star formation history	chemical abundances
nebular properties	active black holes

Key Idea: Stellar Populations in Distant Galaxies are (almost) always Unresolved

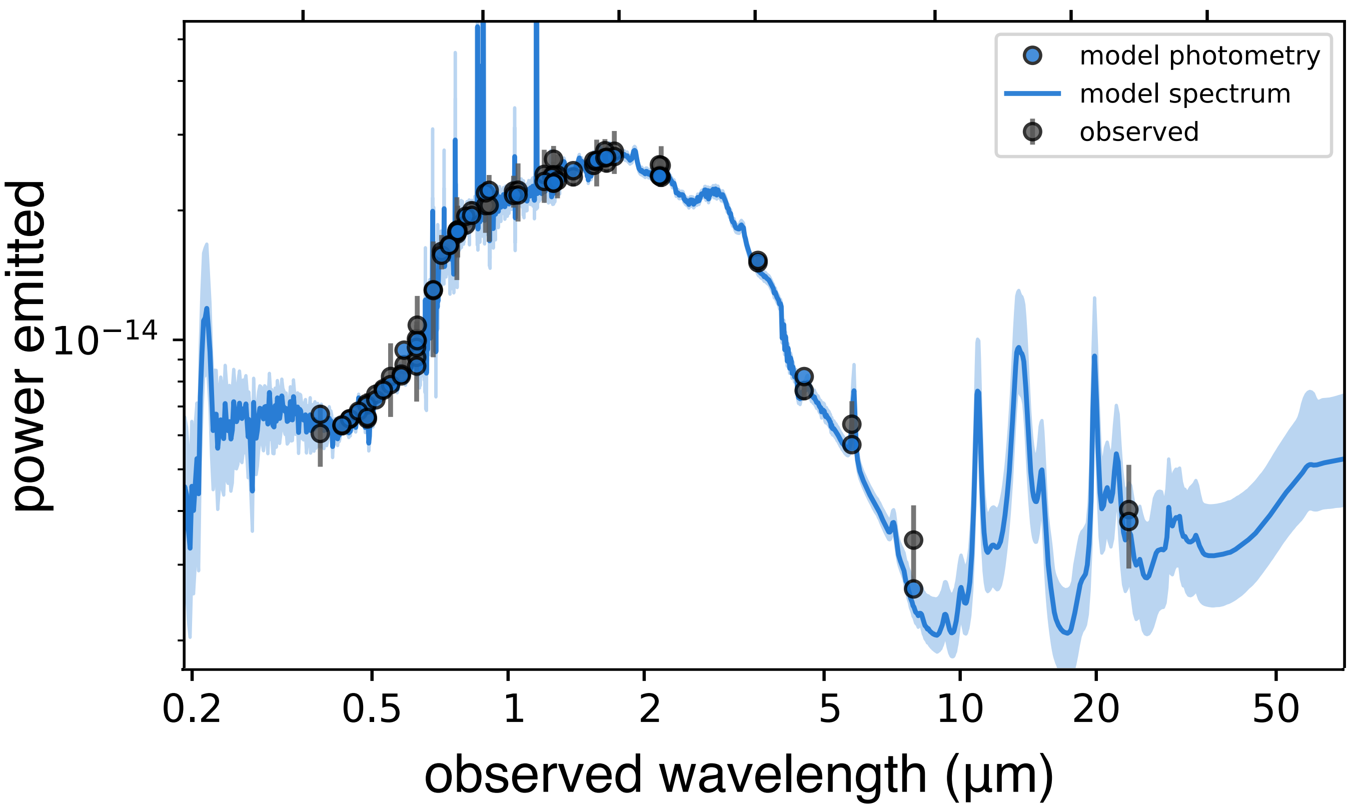
Andromeda

A typical distant galaxy

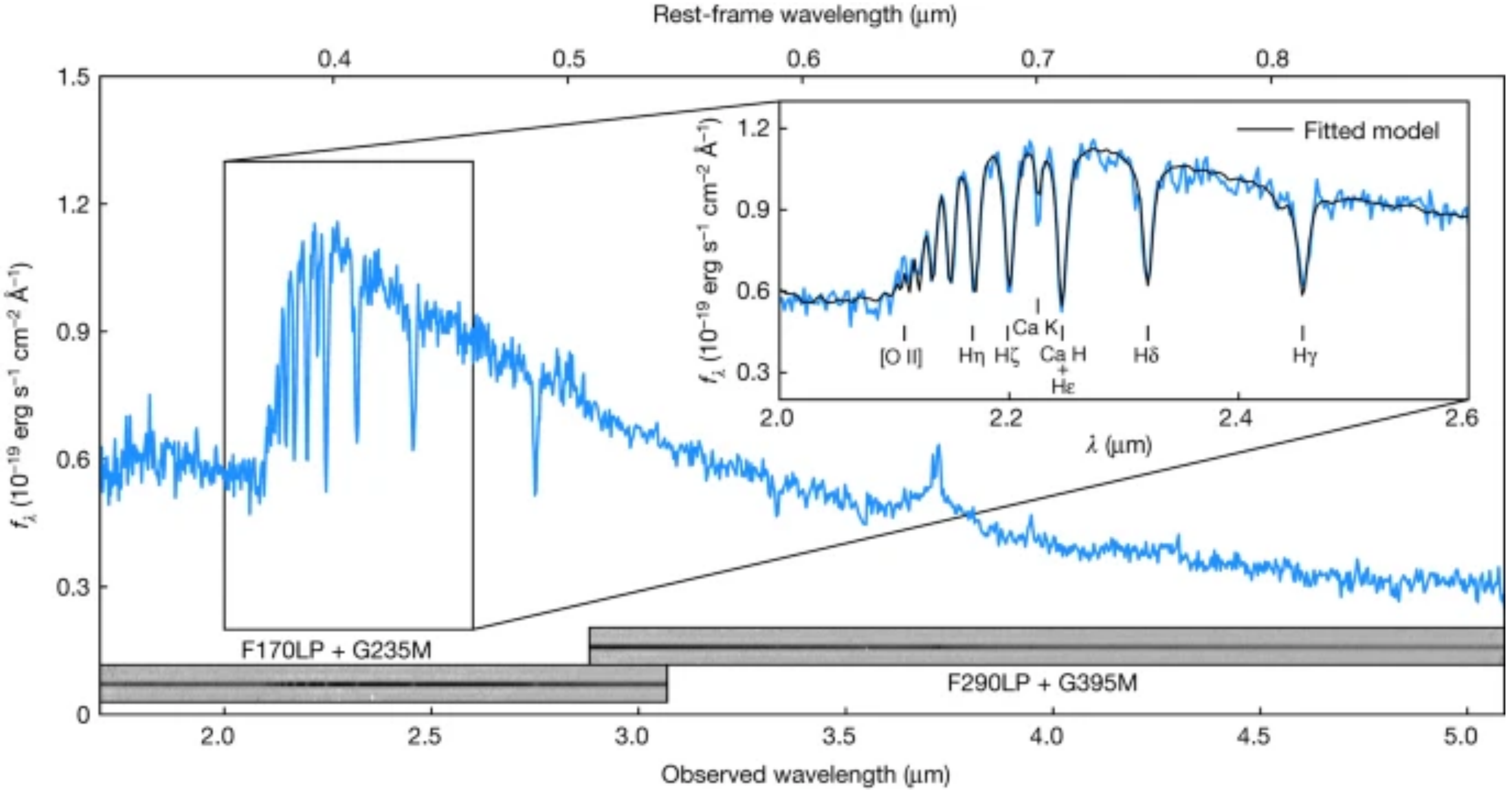


In most cases we model the **sum of the light** — hard to reconstruct constituents.

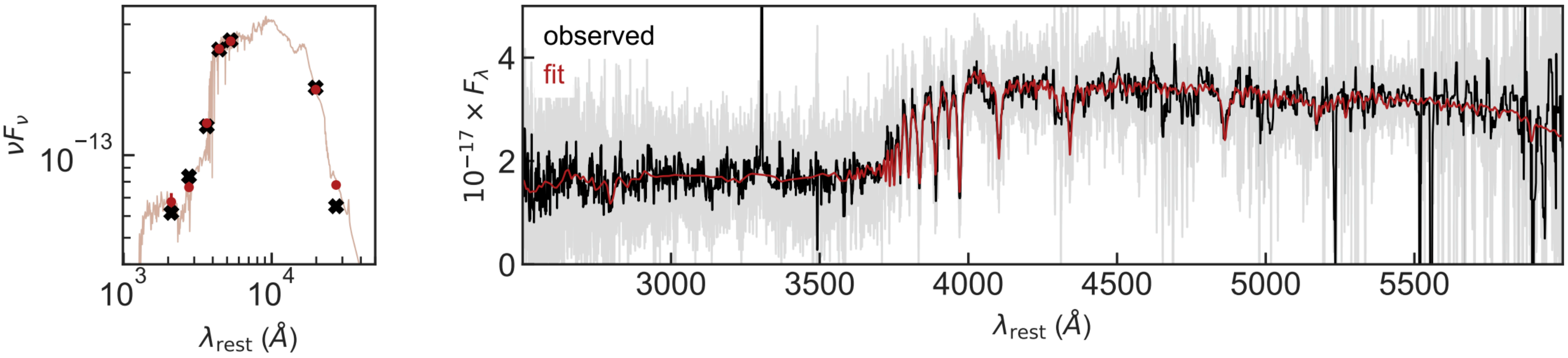
Galaxy surveys are often multiwavelength. Can **fit photometry** across many wavelengths to measure properties.



More rarely, spectroscopy is available. **Fitting spectroscopy** can give detailed constraints on galaxy formation histories and physical conditions.



The best scenario is a **joint spectrophotometric fit**. The photometry gives good wavelength coverage and precise normalization, while the spectroscopy gives detailed constraints.



Prominent Spectral Features

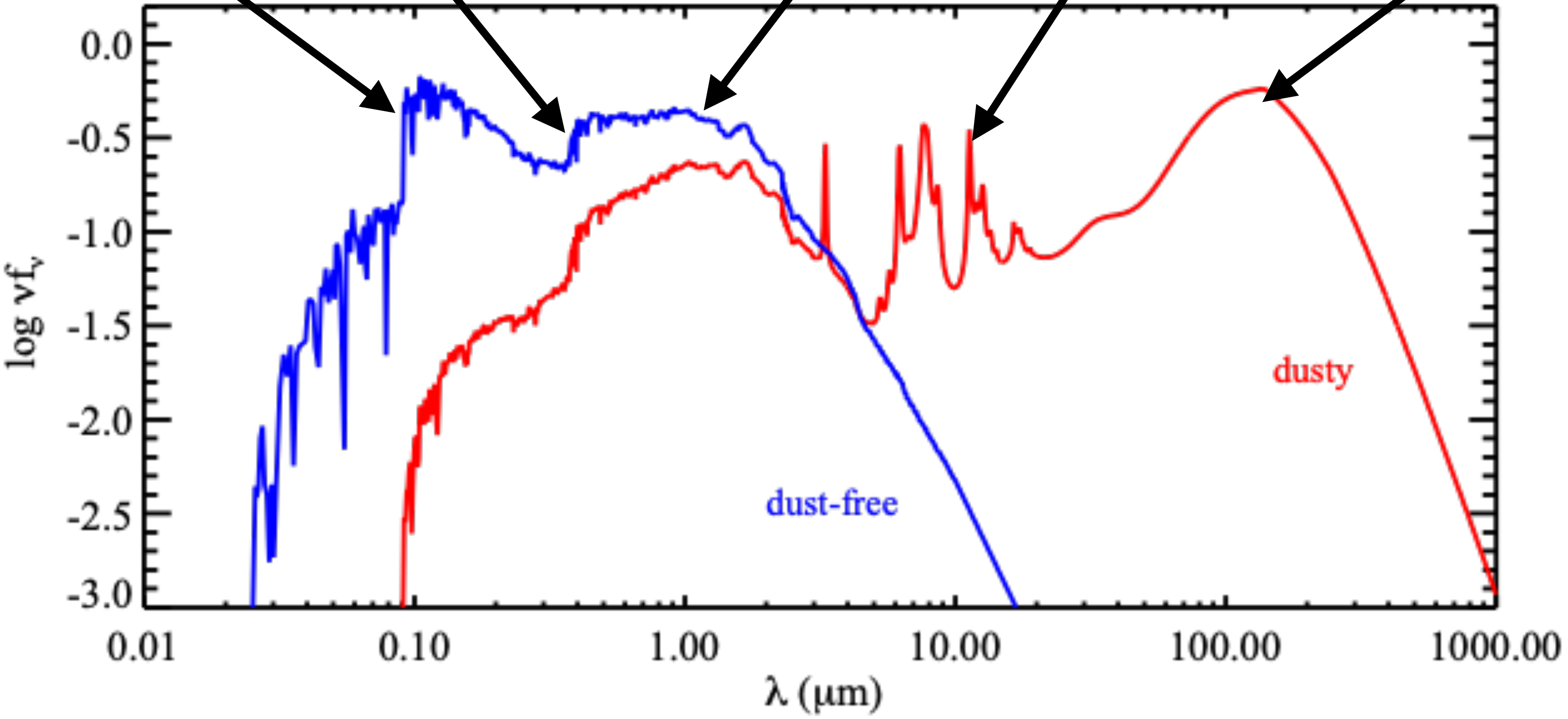
Lyman break: caused by neutral hydrogen in ISM or galaxy

Balmer break: strength sensitive to **age**, **metallicity**

near-IR bump: sensitive to mass in old stars (unlike UV/optical)

PAH line emission: from large hydrocarbons, strong/variable

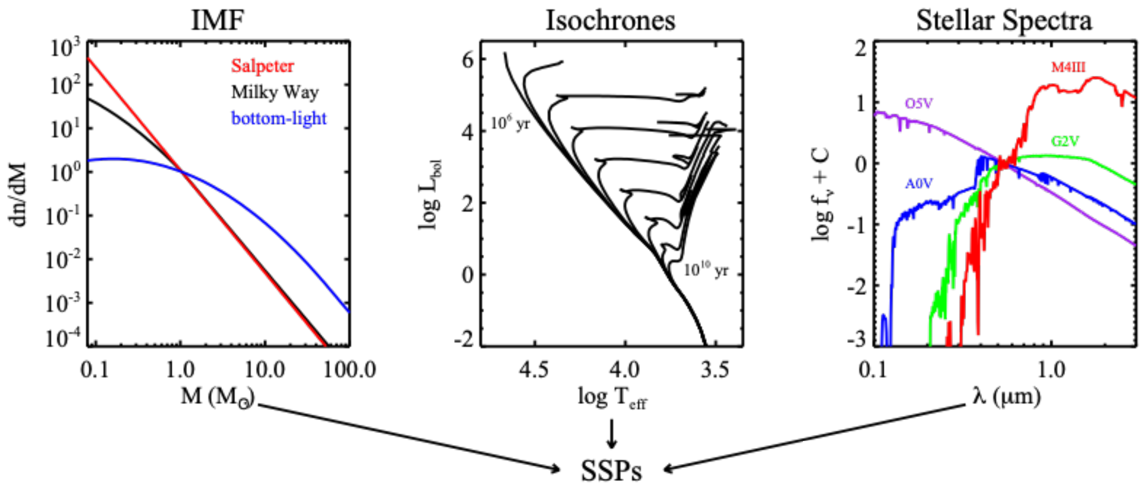
thermal dust bump: cold dust, typically heated by star formation



Conroy 2013 ARAA

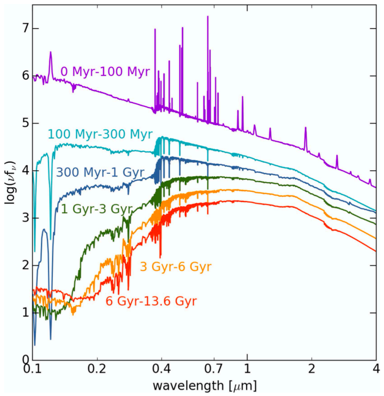
Simple Stellar Populations

A **simple stellar population (SSP)** is a set of stellar spectra at a fixed age and metallicity summed following a **stellar initial mass function**. Globular clusters constitute reasonable real-life approximations of an SSP.



$$f_{SSP}(t, Z) = \int_{m_{lo}}^{m_{up}(t)} f_{star}[T_{eff}(M), \log g(M)|t, Z]\, \Phi(M)\, dM$$

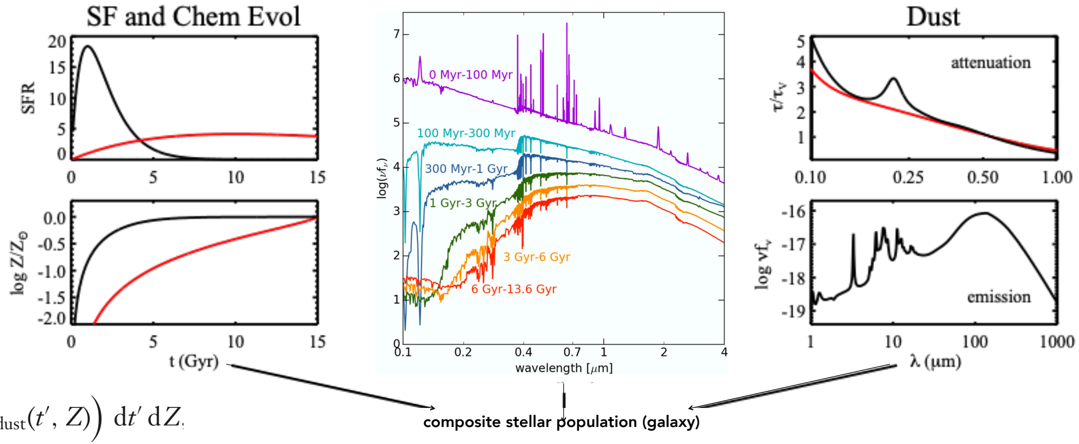
IMF x spectra(stellar mass, t, Z)



Conroy 2013 ARAA
Leja et al. 2017

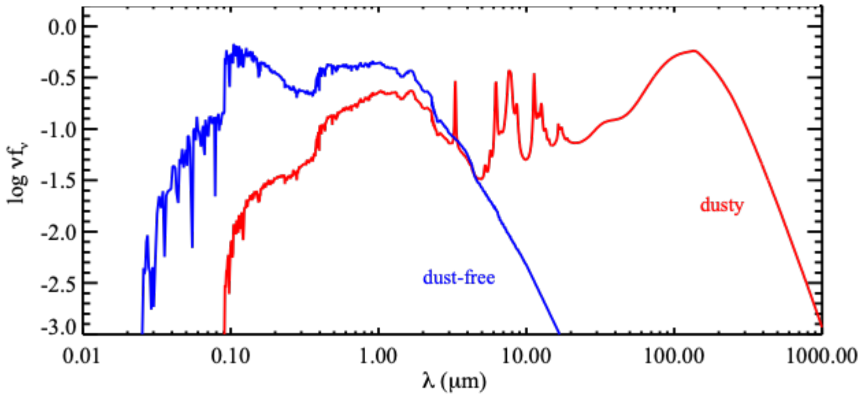
Complex Stellar Populations

A complex stellar population (CSP) is a sum of SSPs at different ages/metallicities, often combined with dust attenuation and re-emission. Galaxies constitute real-life approximations of a CSP.



$$f_{\rm CSP}(t) = \int_{t'=0}^{t'=t} \int_{Z=0}^{Z_{\rm max}} \left({\rm SFR}(t-t') P(Z, t-t') f_{\rm SSP}(t', Z) e^{-\tau_d(t')} + A\, f_{\rm dust}(t', Z) \right) {\rm d}t'\, {\rm d}Z,$$

SFR x SSP x dust + dust emission
(+ nebular emission + AGN + …)

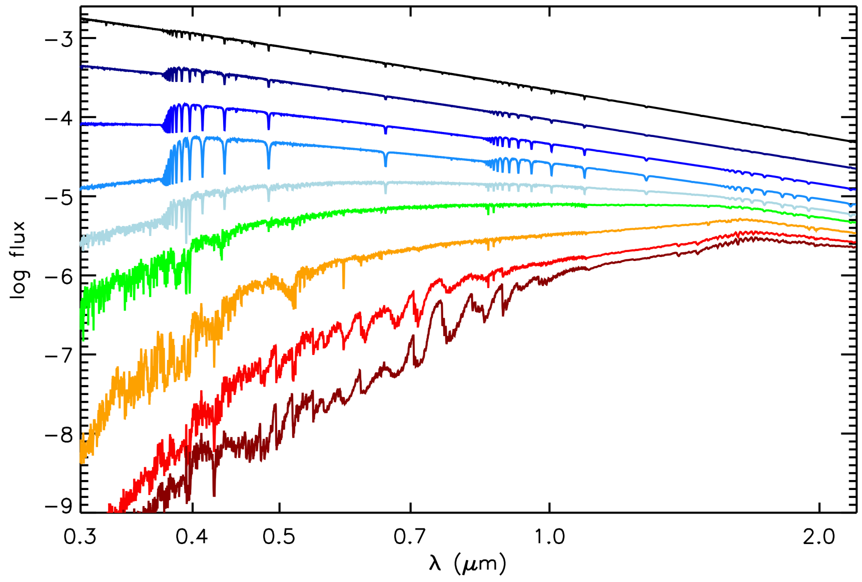


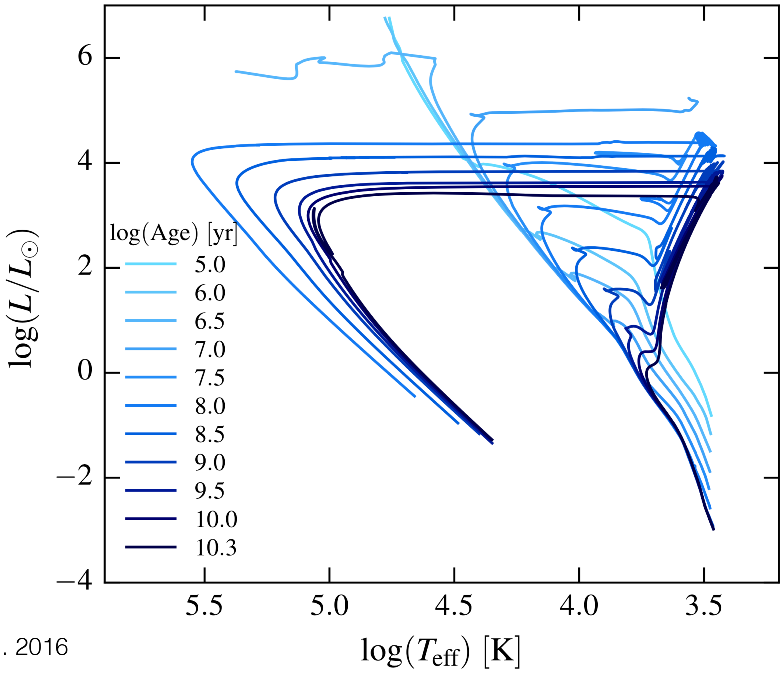
Conroy 2013 ARAA
Leja et al. 2017

Stellar Isochrones and Libraries

Whether you are modeling stars in the Galaxy or a galaxy of stars, you use the same ingredients:

- Models of the **stellar surface**: photometry, spectra
- Models of the **stellar interior**: stellar tracks and isochrones





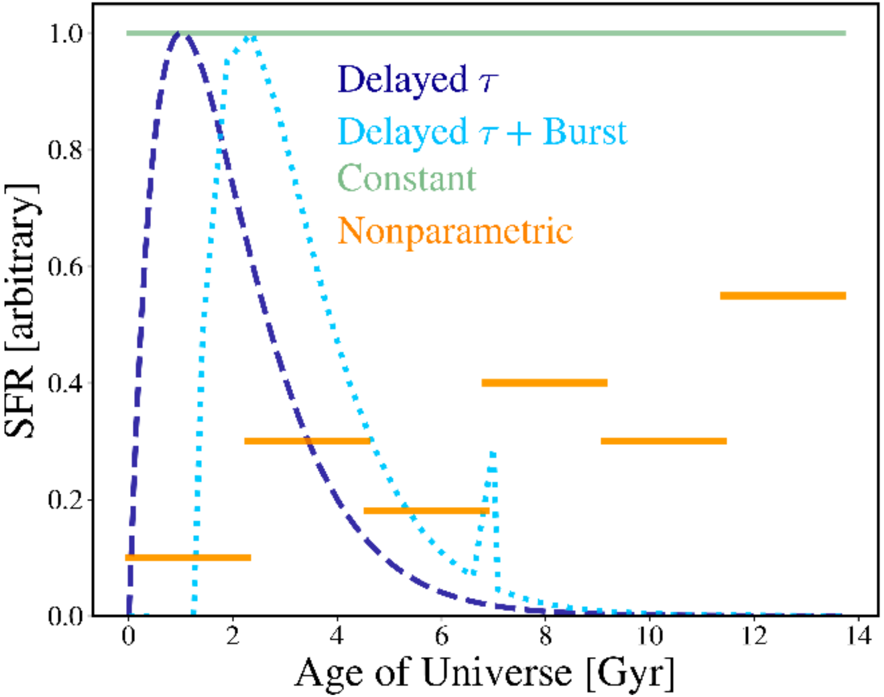
Choi et al. 2016

This week we will explore the *variation between different models* (different treatment of exotic stellar evolution phases, rotation, binaries…), and *their origins* (empirical or theoretical, and associated drawbacks)

Star Formation History Recovery: Parametric SFH Models

How do we model galaxy star formation histories? The most straightforward way is to **fit an equation** describing SFR(t). There are a few families of common options.

constant SFR — one parameter! very simple, not flexible	$\mathrm{SFR} \propto constant$
exponentially declining — also known as a "tau" model, very popular in literature	$\mathrm{SFR} \propto e^{-t/\tau}$
delayed tau model — has a rising component at early times, thought to be more realistic	$\mathrm{SFR} \propto te^{-t/\tau}$
lognormal model — rises at early times, falls at late times, can quench rapidly — flexible but not very common	$\log \mathrm{SFR} \propto te^{\frac{(x-\mu)^2}{\sigma^2}}$



Lower et al. 2020

Outline: Modeling Tips & Tricks

1. Building a Model
    a. Framing: what does your model do?
    b. Choosing priors
2. Fitting a Model
    a. Multimodality
    b. Fitting mock data (bug testing)
3. Evaluating your Fit
    a. Look at residuals, posteriors, chains/traces of the sampler
    b. Predict data which are not included in the fit
    c. Fit using a different prior and see how the results change

Building a Model

What model elements do you need to generate/predict your data?

Do you have rest-frame infrared data? If so, do you want to fit the dust emission SED? Do you want to predict the dust SED? Is AGN emission important?

Do you have a spectrum? Do you know or do you have to fit for the redshift, velocity dispersion, etc… Are the stellar libraries appropriate for your data? Is nebular emission important?

Do you trust the calibration of the data? Do you trust the uncertainties? Are there outliers? Or do you need to include these as parameters of the model that you will fit?

What parameters can you fix? (Keeping in mind this is equivalent to a delta-function prior, and if the parameter is truly unknown and important then your posterior PDF will be too “narrow”)

Building a Model

What is your scientific question?

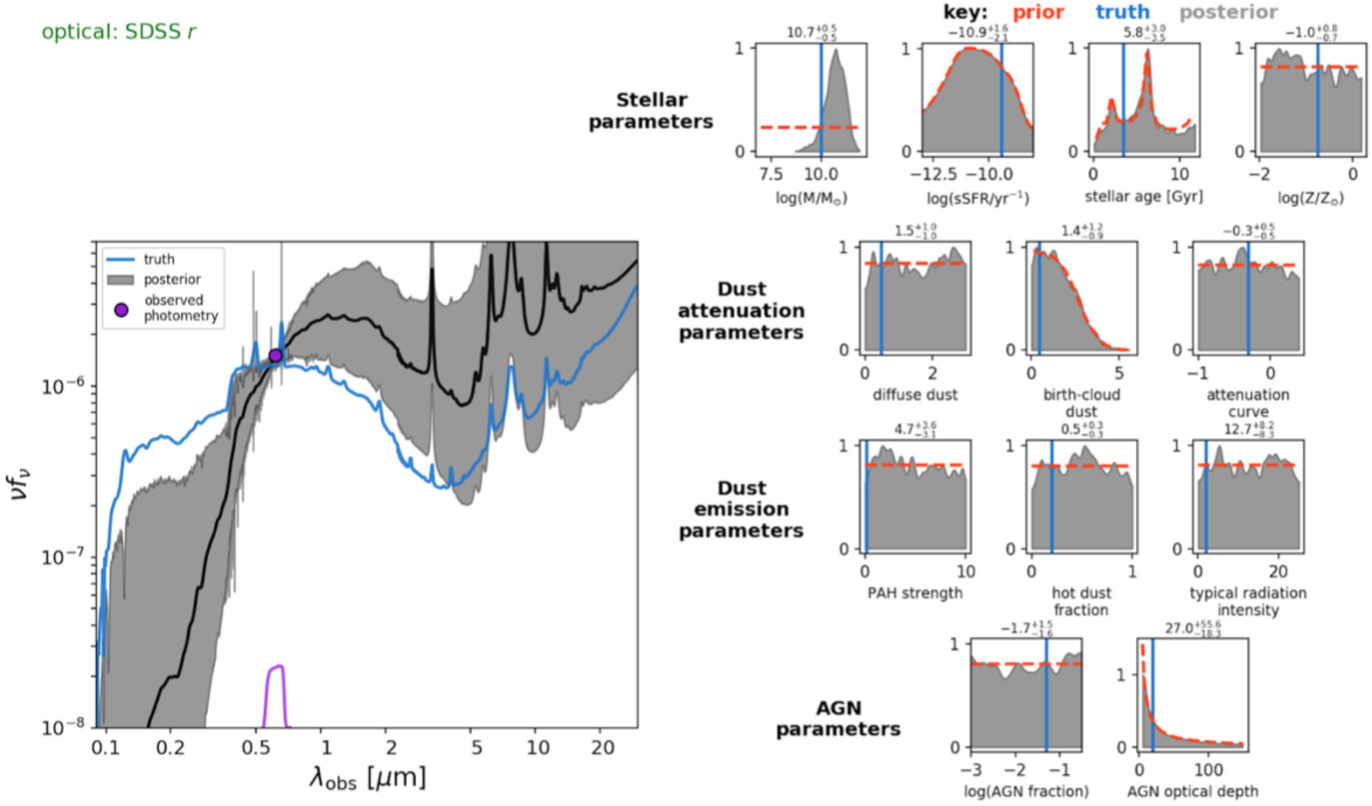
In galaxy SEDs nearly everything affects everything else. But, sometimes your scientific question will dictate what model components to include, or which data to fit, or which priors to adopt.

How do you want to parameterize the SFH?

Should you include a mass-metallicity prior? (E.g., if you are interested in inferring the mass-metallicity relation, then probably not!)

Is it ok to fit a psuedo-narrowband (i.e. an index) instead of the full spectrum? Maybe, if you don’t care about line velocities

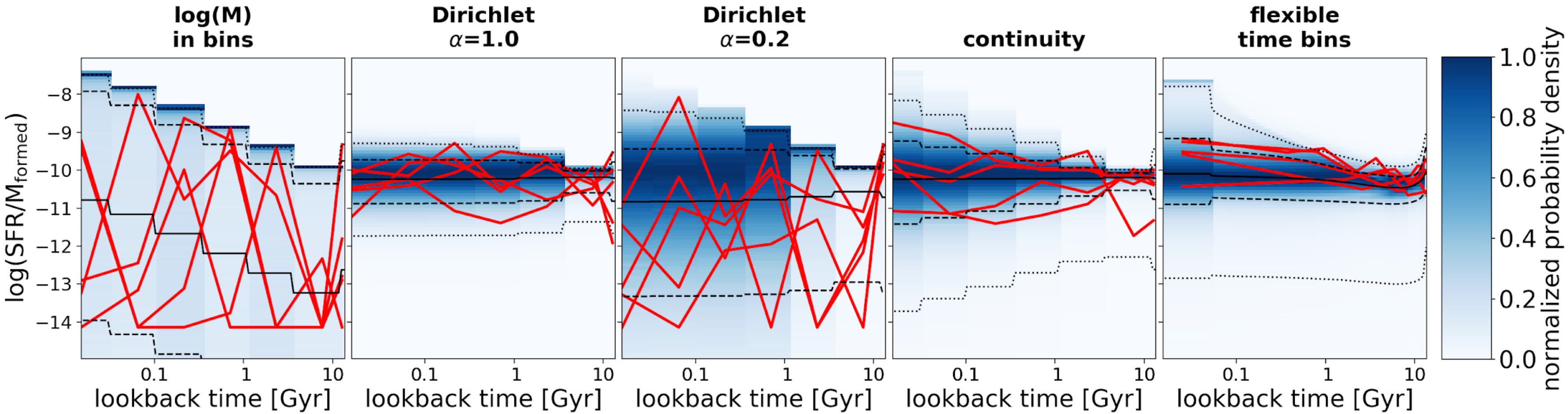
Choosing Priors



Choosing Priors

Prior density should track occurrence in population; check the literature

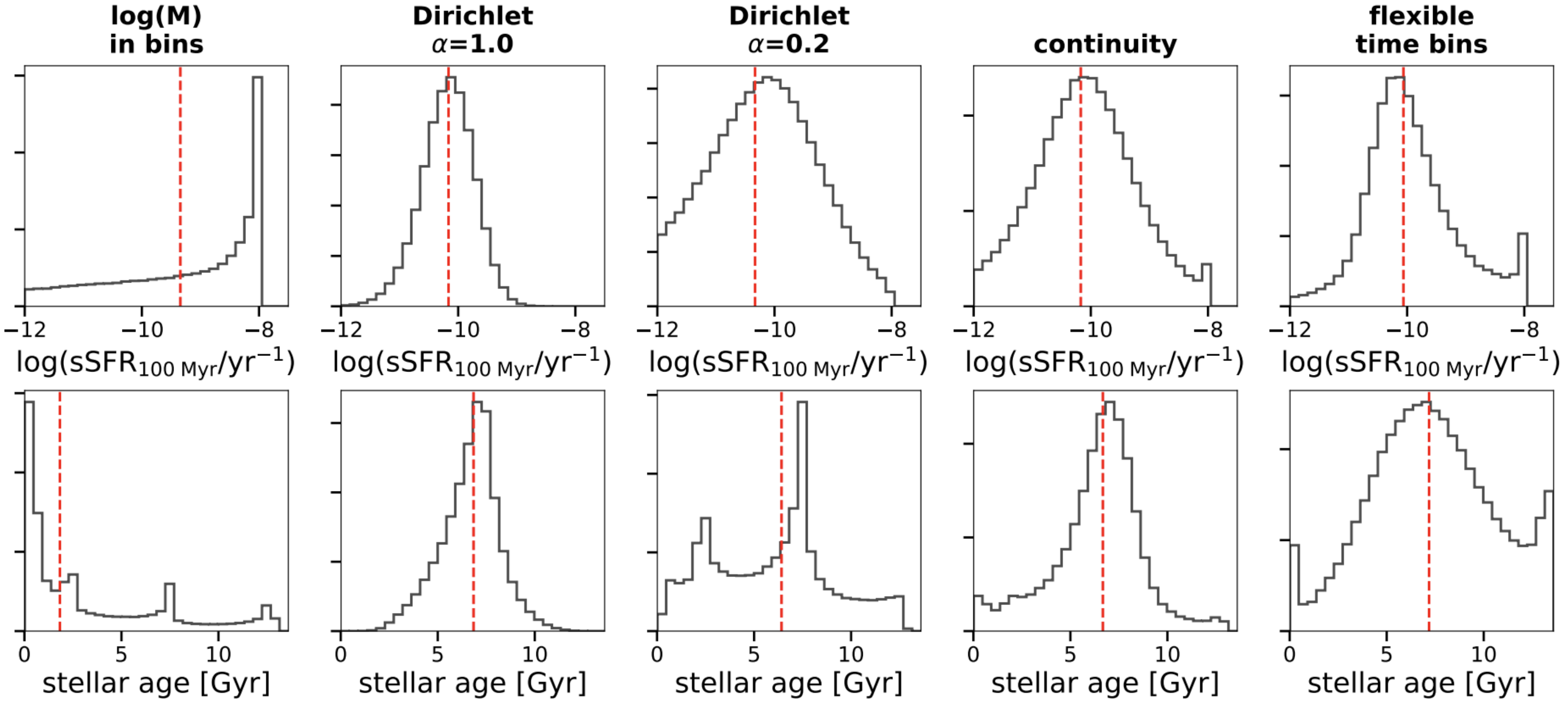
Plot your priors -- know what you are putting into the model



Leja+19

Choosing Priors

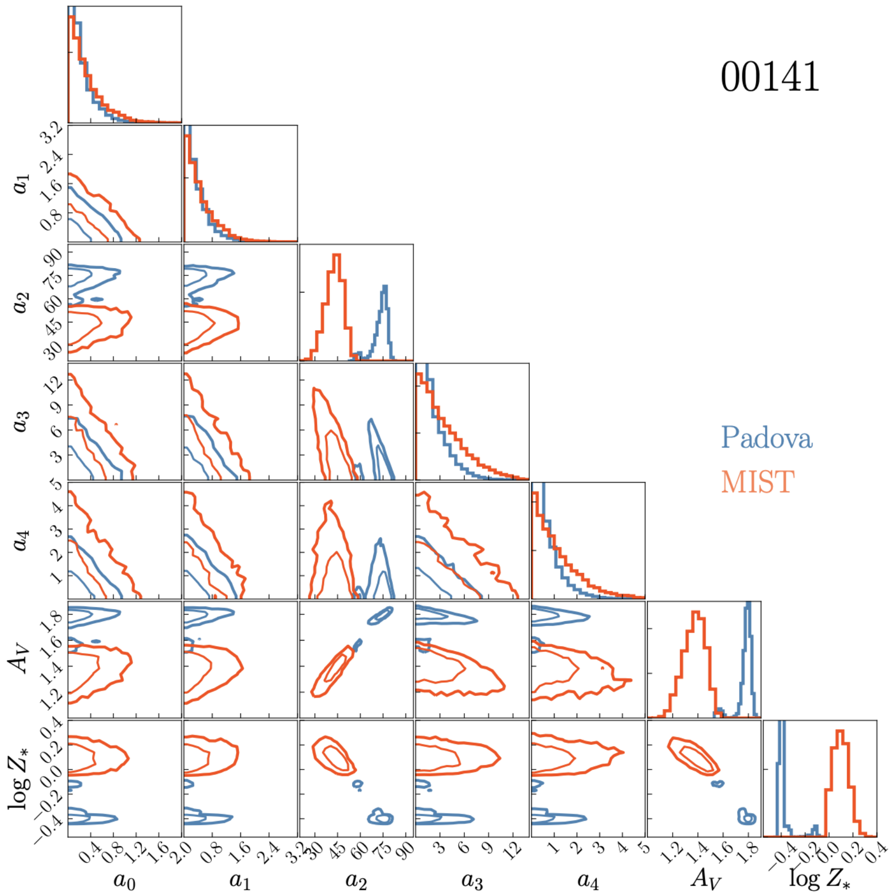
Transform your priors into target parameters; are they reasonable?



Leja+19

Fitting a Model

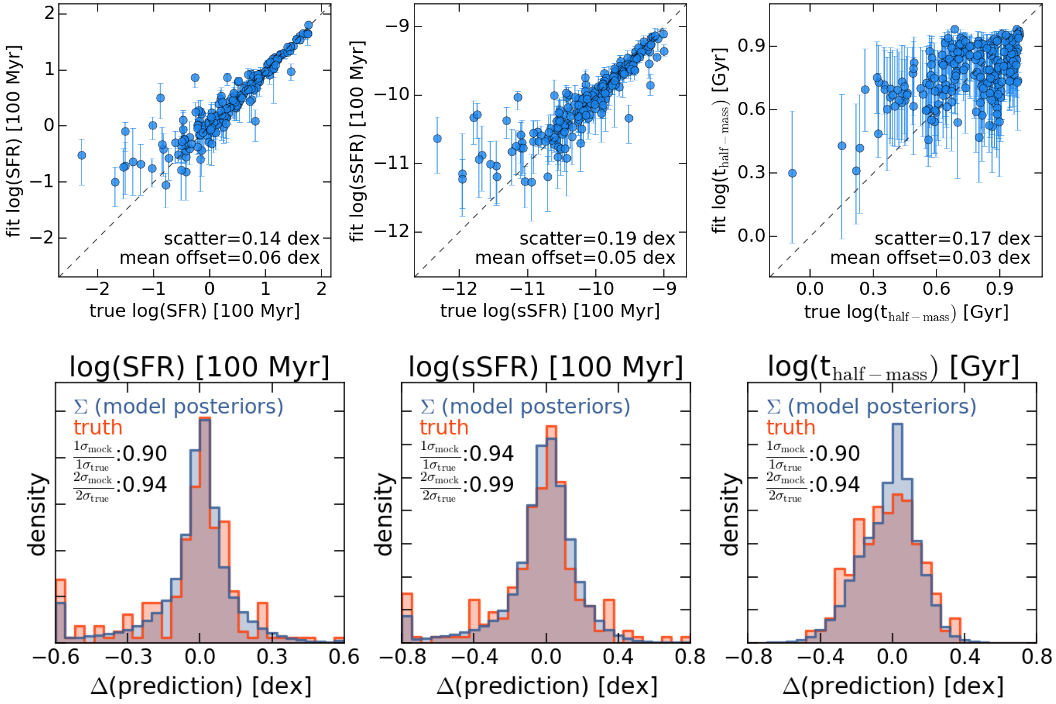
Be aware that multi-modal posteriors can magnify even small model differences into significant parameter offsets.



Morishita et al. 2018

Fitting a Model

Successful mock tests demonstrate (a) that your code works, and (b) how much information it recovers. Keep in mind that this is the ideal situation for your code!



Leja+17

Two types:

(a) **Noiseless**: the best-fit solutions should be *exactly* the input value.

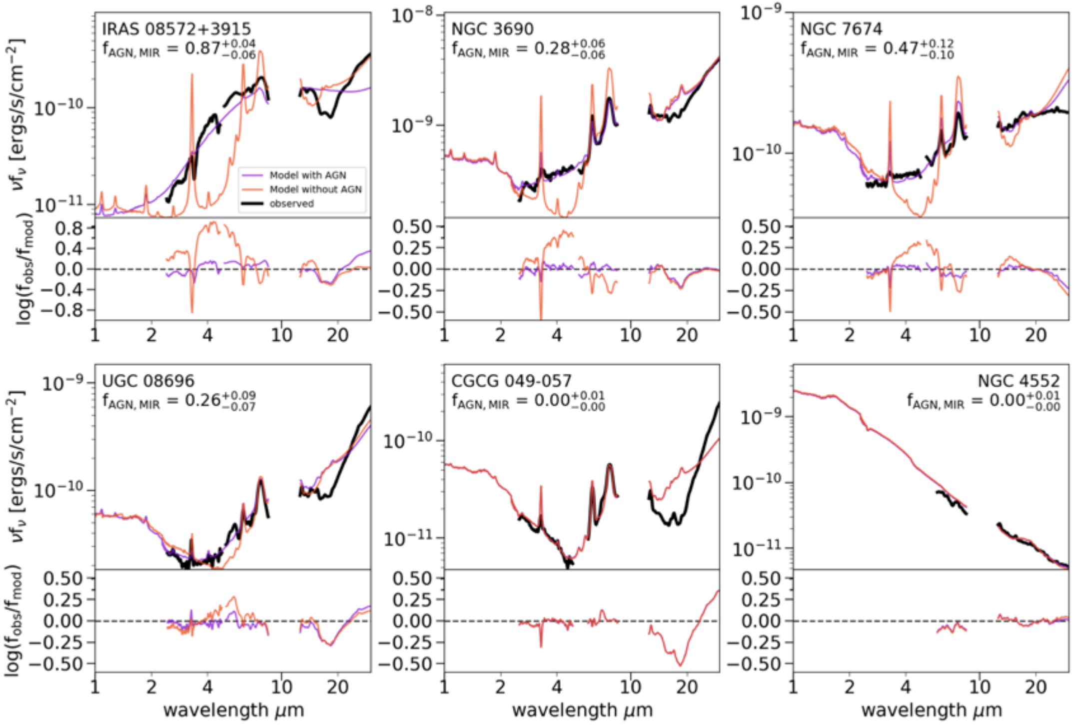
(b) **Noised**: in a suite of tests, the posteriors should be statistically consistent with the inputs

Evaluating Your Fit

(1) Find the maximum posterior probability sample from your chain, and
(2) generate the prediction for your data using your model and the parameters of this sample

(3) Compare the prediction to the actual data,
(4) look at residuals and at ***uncertainty normalized*** residuals.

Is there some data feature you can't model well (missing ingredient)? Maybe the uncertainties are underestimated?

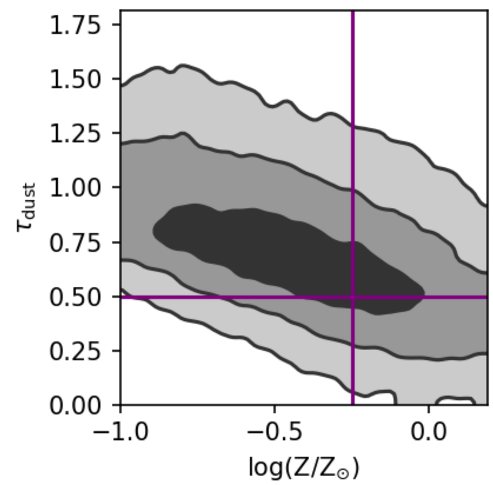


Leja+18

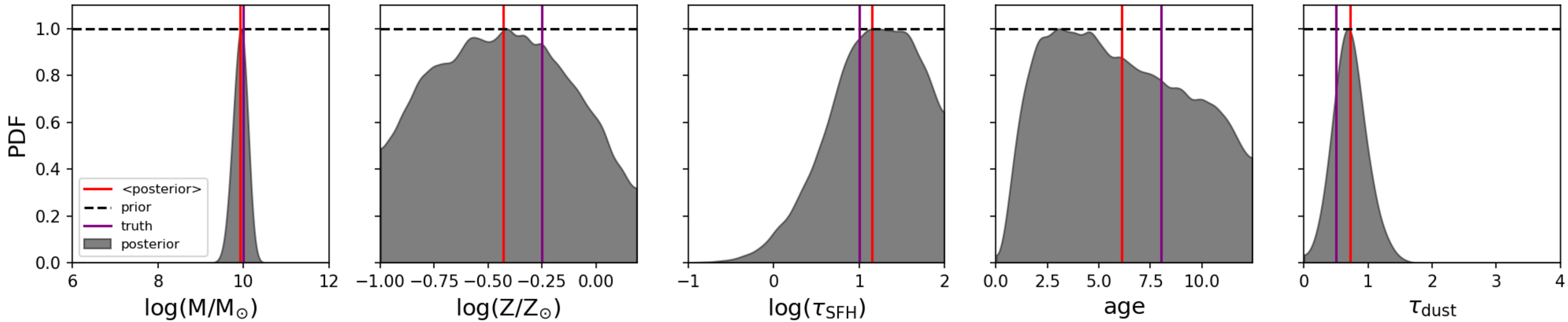
Evaluating Your Fit

Check the posterior PDFs:

Are you bumping up against a prior? Do the covariances you see make sense? How well are your parameters constrained (i.e. is the posterior different than the prior?)



Good fit

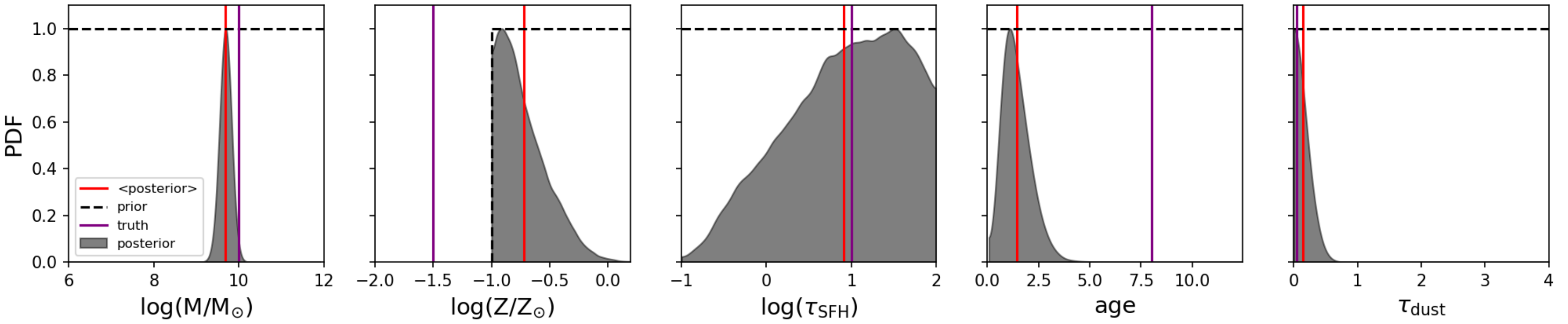


Evaluating Your Fit

Check the posterior PDFs:

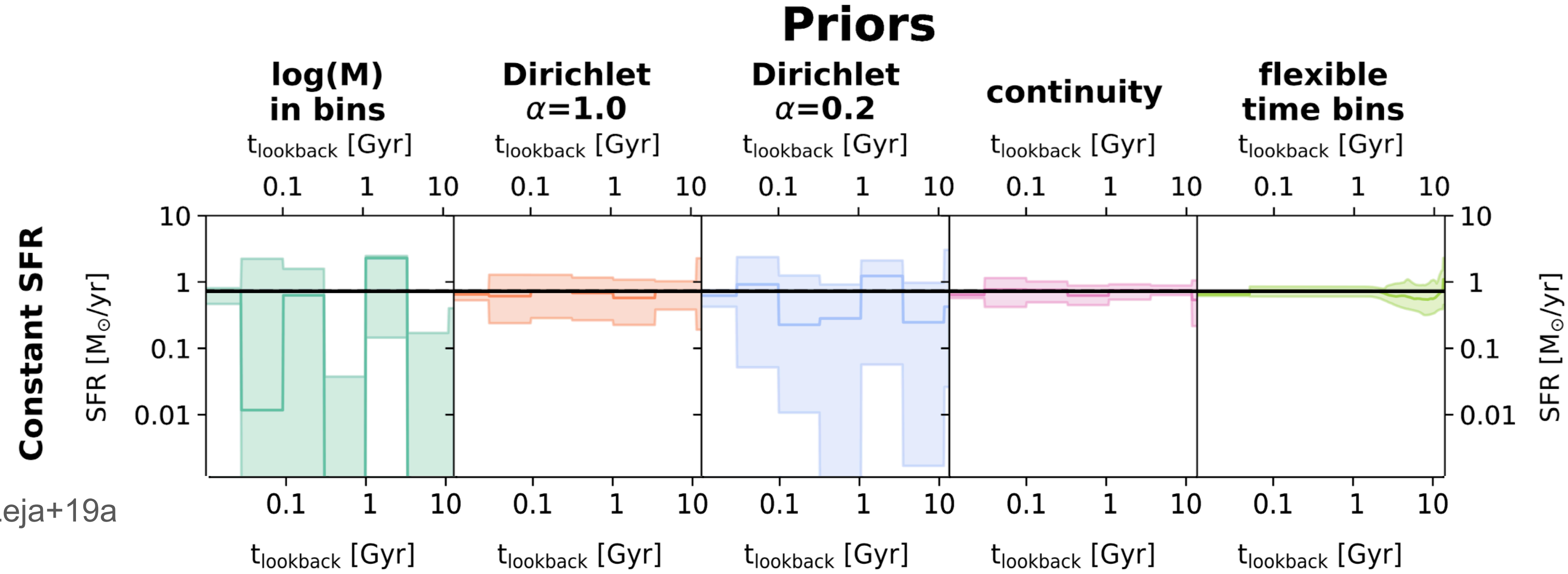
Are you bumping up against a prior? Do the covariances you see make sense? How well are your parameters constrained (i.e. is the posterior different than the prior?)

Bad fit (metallicity prior rules out truth)



Evaluating Your Fit

Change priors and re-fit. If your results are sensitive to the prior, make sure your prior is well-motivated (if not, go back to the drawing board!).

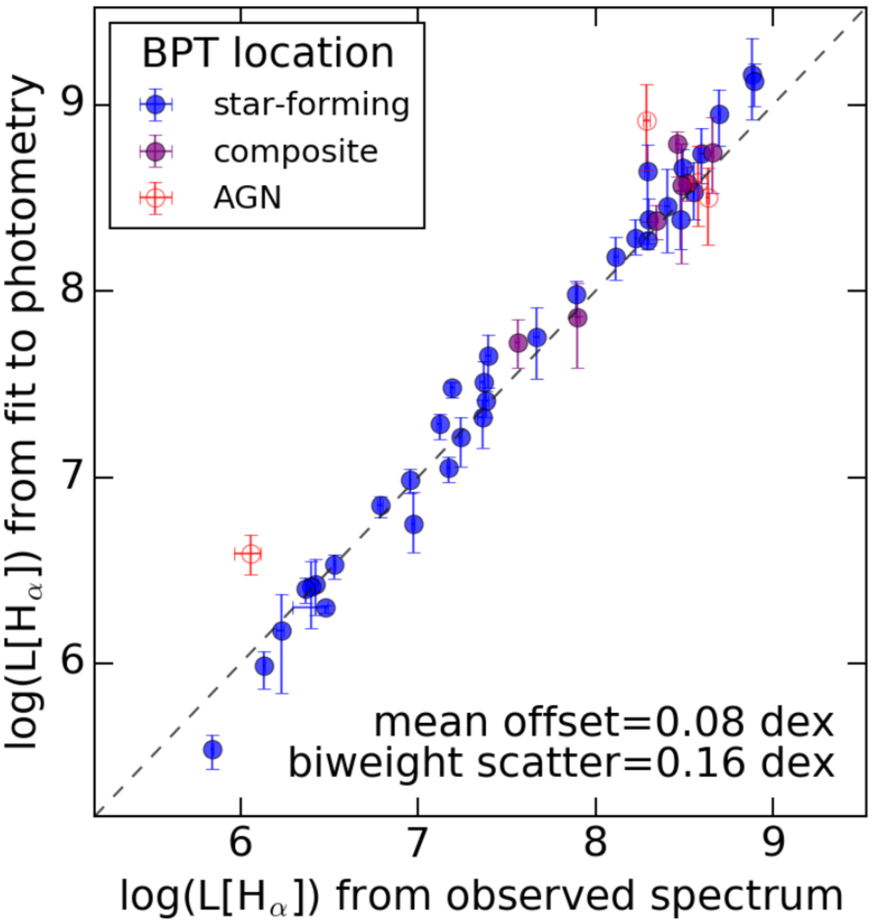


Leja+19a

Evaluating Your Fit

Real data is never paired with a ground truth. To evaluate your model:

(1) Predict data that weren't fit (e.g., predict spectroscopy after fitting photometry, or predict dynamical mass)
(2) Predict galaxy scaling relationships -- mass-metallicity, cosmic star formation rate density, fundamental plane...

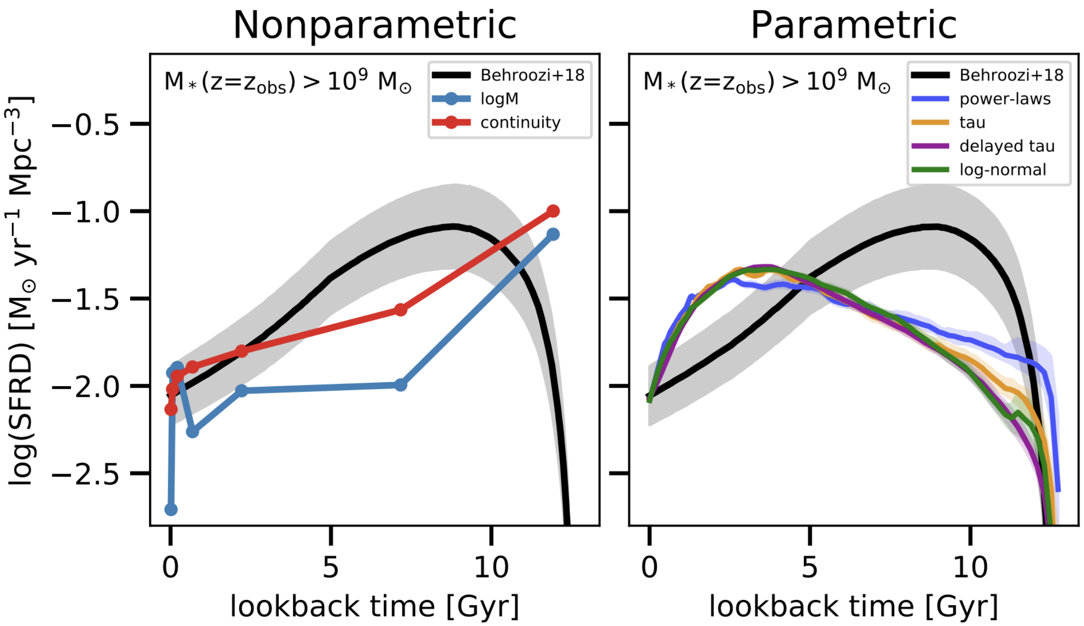


Leja+17

Evaluating Your Fit

Real data is never paired with a ground truth. To evaluate your model:

(1) Predict data that weren't fit (e.g., predict spectroscopy after fitting photometry, or predict dynamical mass)
(2) Predict galaxy scaling relationships -- mass-metallicity, cosmic star formation rate density, fundamental plane...



Leja+19a, Carnall+19a

Prospector Links

Installation:
https://prospect.readthedocs.io/en/stable/installation.html

Quick start:
https://prospect.readthedocs.io/en/stable/quickstart.html

Demonstrations, examples of phot / spec fits:
https://prospect.readthedocs.io/en/stable/demo.html

Github:
https://github.com/bd-j/prospector

BAGPIPES Links

Installation:
https://bagpipes.readthedocs.io/en/latest/

Notebooks & basics:
https://drive.google.com/drive/folders/1eYntNnCeYDEbwG9fNK-L6LPfXsWExo9I?usp=sharing

Github:
https://github.com/ACCarnall/bagpipes