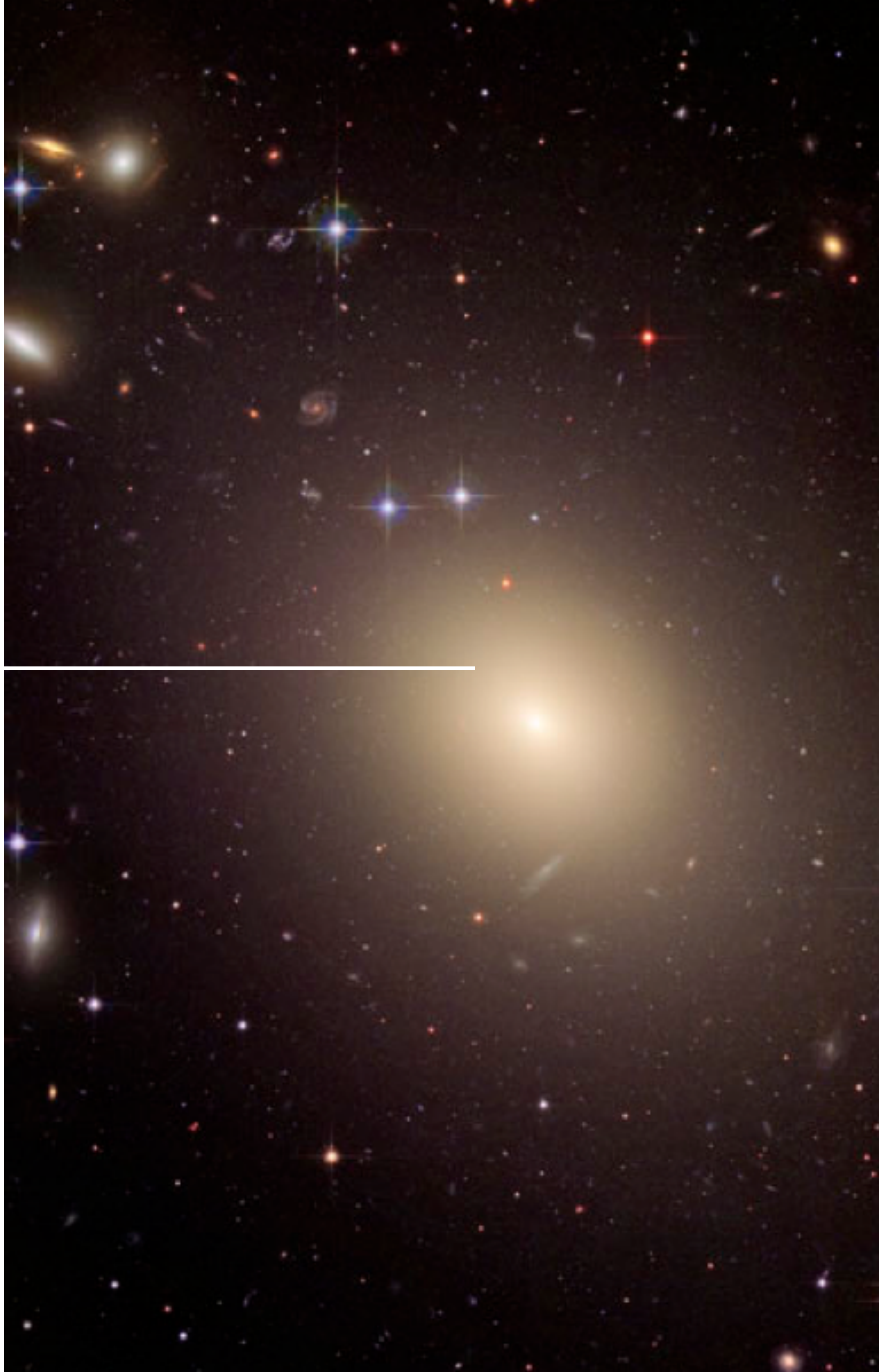
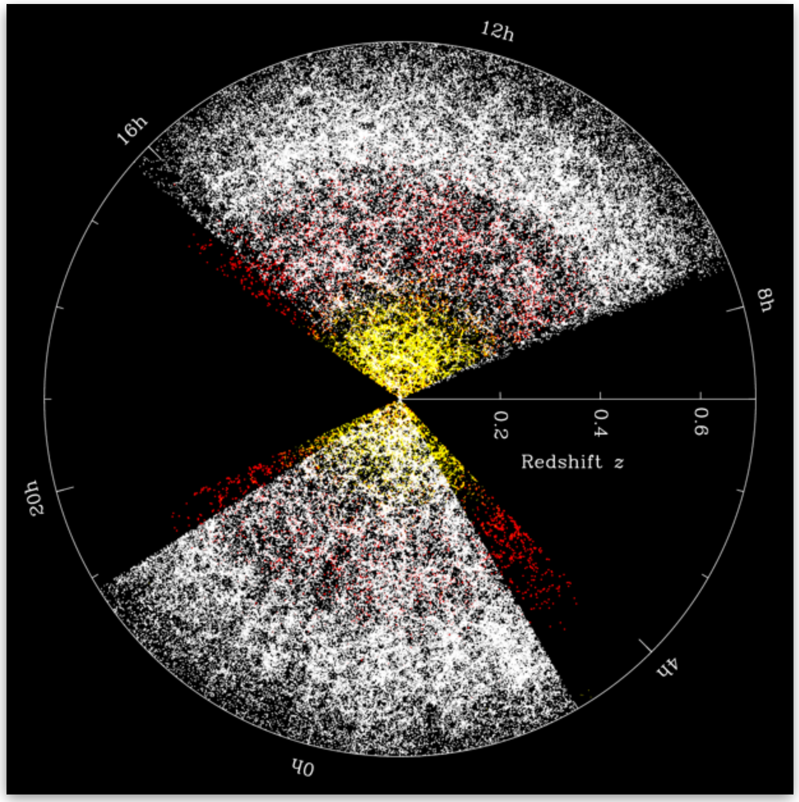
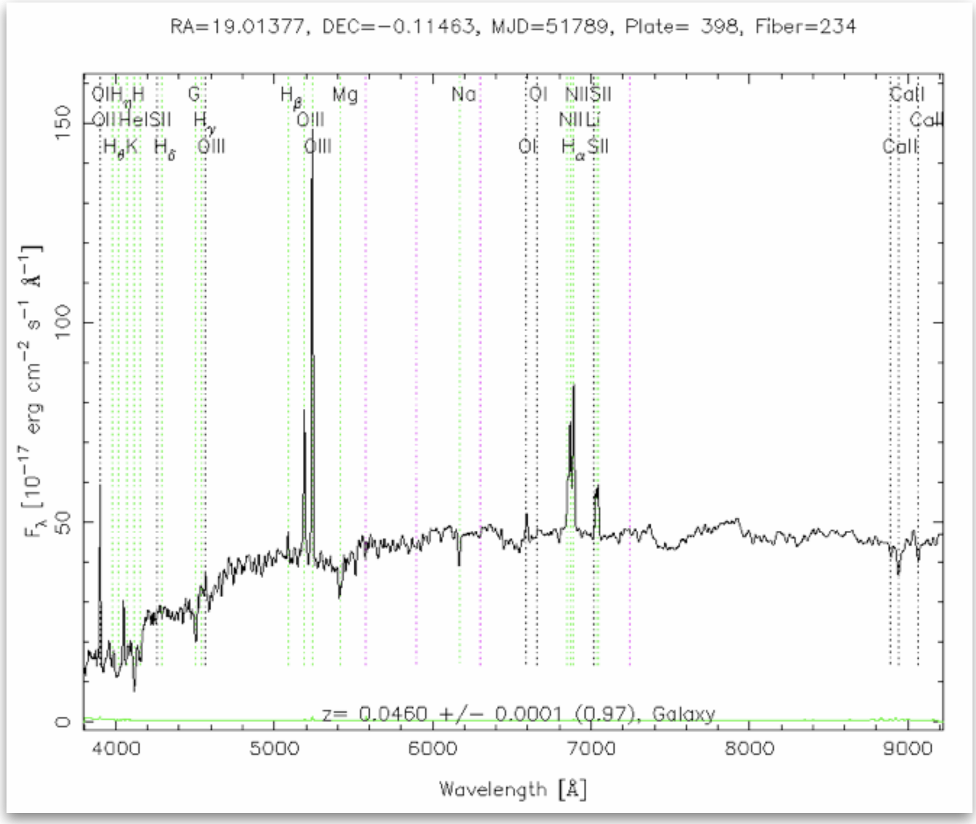
Unlocking the full potential of galaxy spectroscopic surveys

Rita Tojeiro
University of St Andrews
STFC Ernest Rutherford Fellow
L'Óreal-UNESCO For Women in Science UK&I Fellow



galaxy redshift surveys = precise 3-dimensional positions
+ spectra

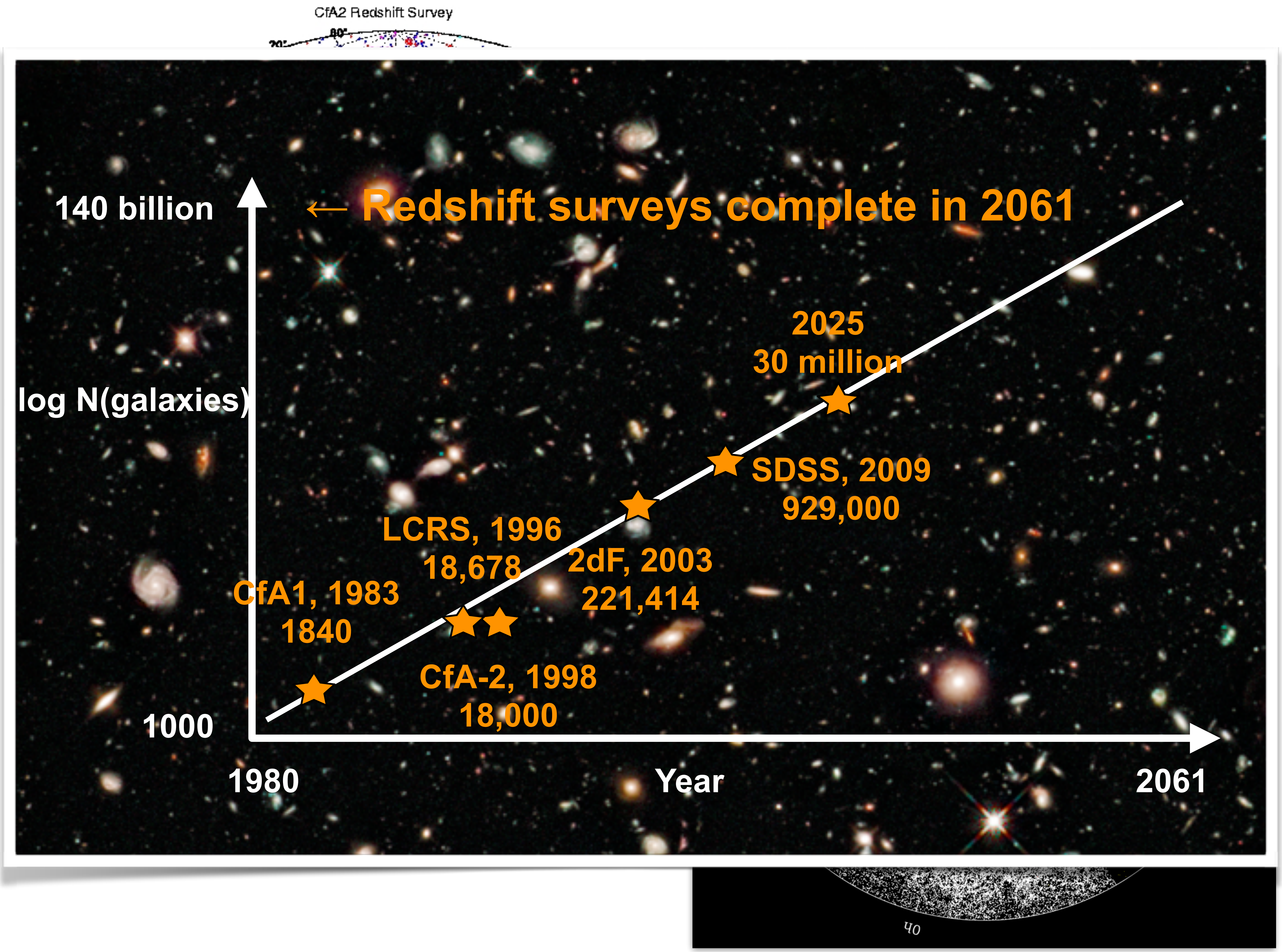




two-point statistics
higher-point statistics
small-scale clustering
topology / voids
cross-correlations with [...]

galaxy physics
quasar physics
LyA cosmology

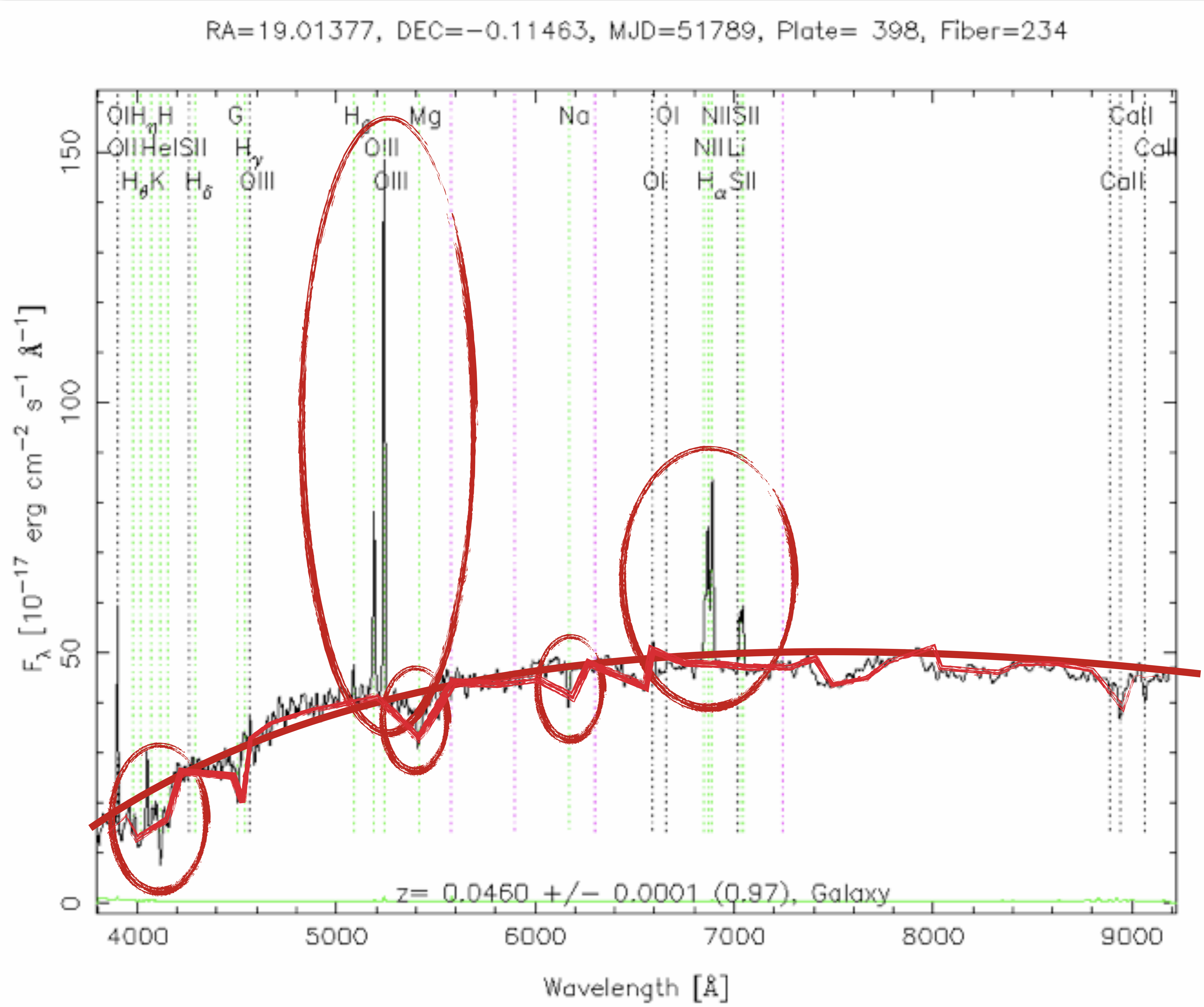
-> expansion, neutrinos, inflation, gal evolution, gravity, composition, etc


CfA2 Redshift Survey
140 billion
← Redshift surveys complete in 2061
log N(galaxies)
2025
30 million
SDSS, 2009
929,000
LCRS, 1996
18,678
2dF, 2003
221,414
CfA1, 1983
1840
CfA-2, 1998
18,000
1000
1980
Year
2061

I - Galaxy spectra and non-parametric SFHs

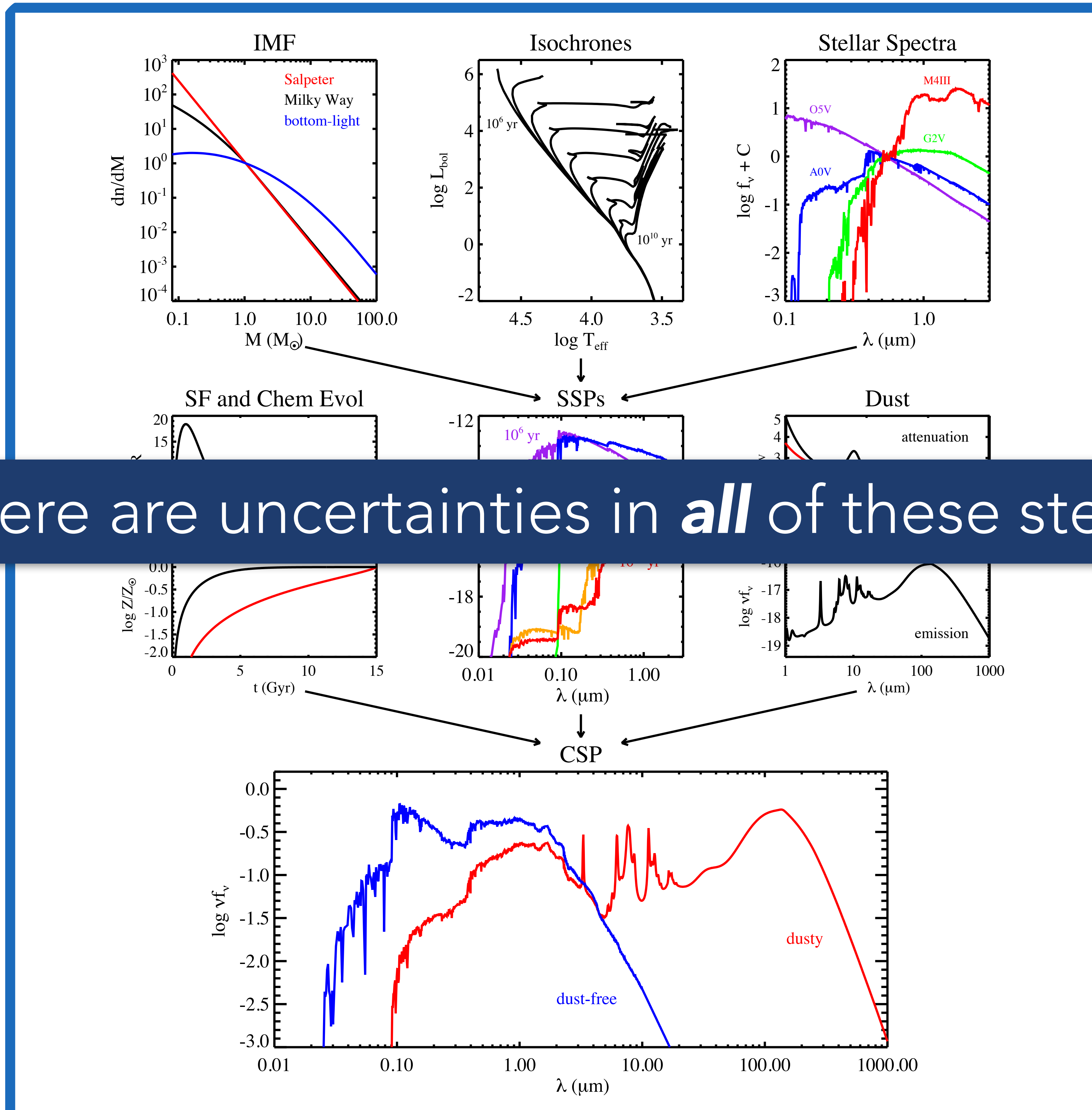
II - Measuring redshift space distortions on large scales

III - The formation time of halos



RA=19.01377, DEC=-0.11463, MJD=51789, Plate= 398, Fiber=234
OI Hγ H
OII HeI SII
Hθ K
Hδ
G
Hγ
OIII
Hβ
OIII
OIII
Mg
Na
OI
OI
NII SII
NII Li
Hα SII
CaII
CaII
CaII
z= 0.0460 +/- 0.0001 (0.97), Galaxy
Fλ [10^-17 erg cm^-2 s^-1 Å^-1]
Wavelength [Å]
0
50
100
150
4000
5000
6000
7000
8000
9000

Simple stellar populations as galaxy building blocks



there are uncertainties in ***all*** of these steps

The problem to solve

$$F_\lambda = \int_0^t f_{dust}(\{p\}, t)\psi(t, Z)S_\lambda(t, Z)dt$$

The problem to solve

$$F_\lambda = \int_0^t f_{dust}(\{p\}, t)\psi(t, Z)S_\lambda(t, Z)dt$$

Simple in principle (especially in the optical)!

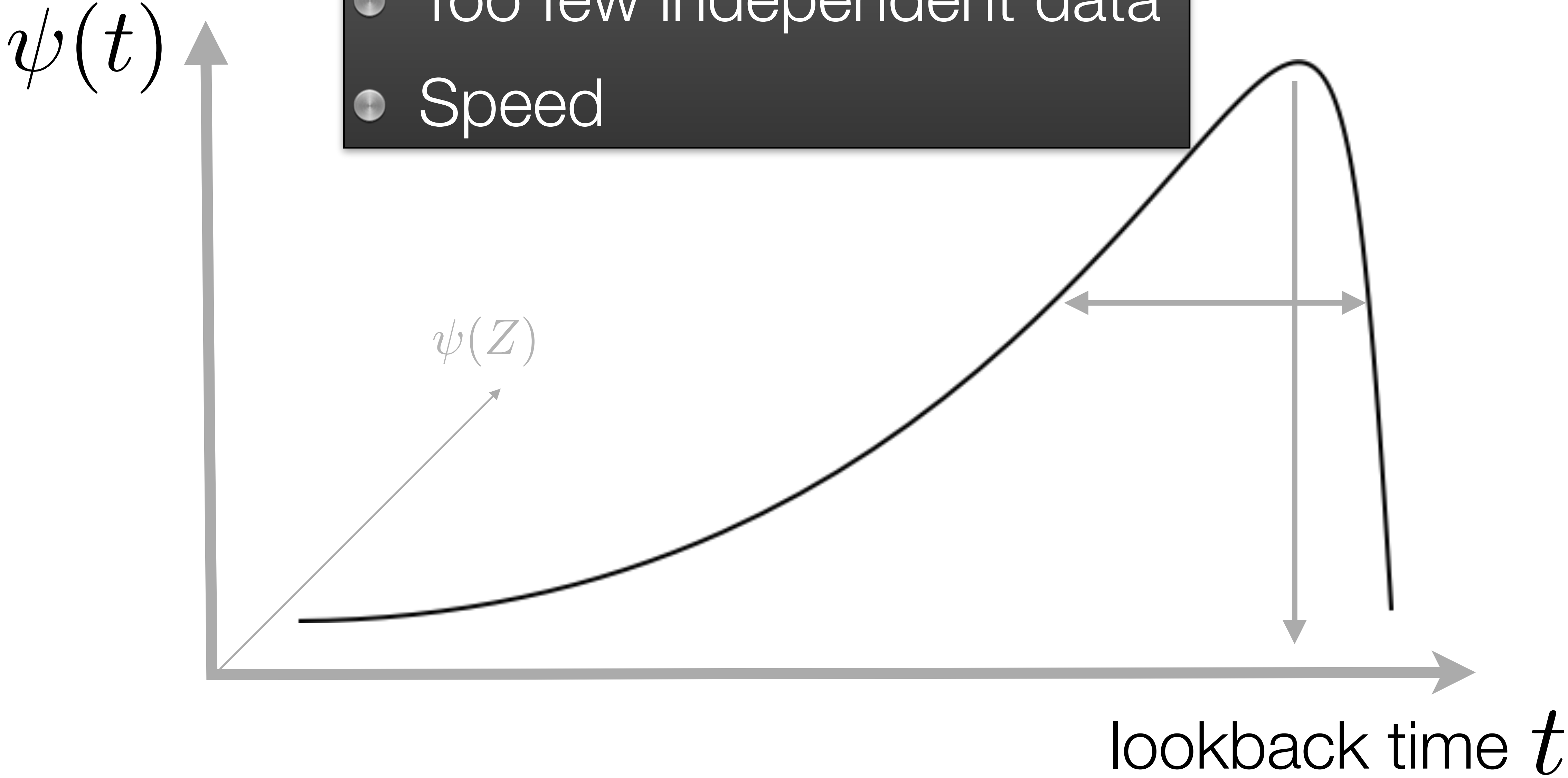
Complications due to data quality, poor modelling, large degeneracies and very large datasets.

Describing galaxies

Parametric star formations histories

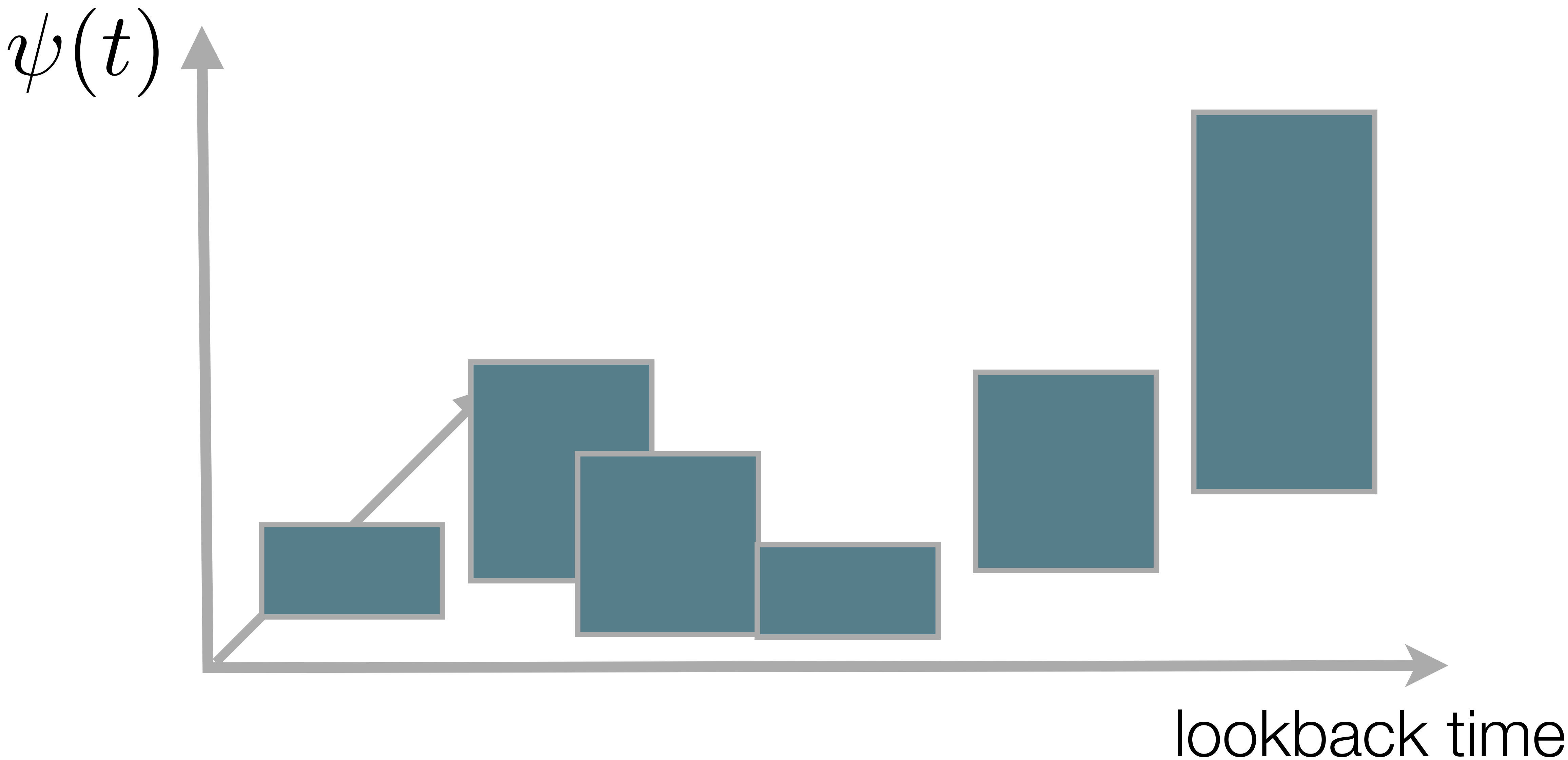
Why reduce dimensionality?

- Too few independent data
- Speed



Parametric SFHs reduce dimensionality,
but are effectively a prior and demonstrably lead to biases

Non-parametric star formations histories



Non-parametric SFHs fit many *more* parameters, but are *less* dependent on choice of parametrisation.
Huge demand on data and models.

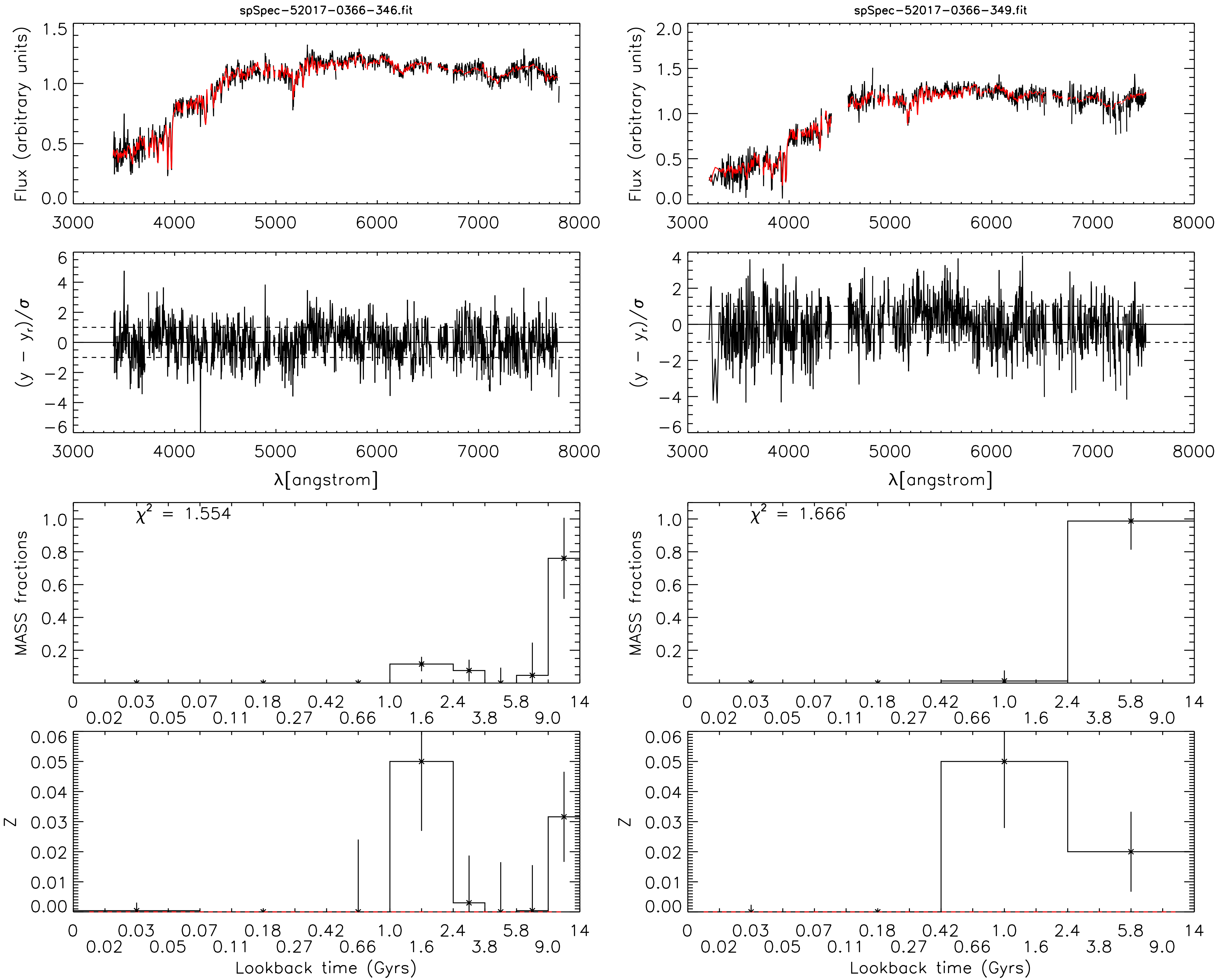
Full-spectral fitting

(we've been doing this a while and getting good at it: see also e.g. MOPED Heavens et al., *STARLIGHT* Cid Fernandes et al., STECMAP Ocvirk et al., Koleva et al., MacArthur et al., *FIREFLY* Wilkinson et al.)

VESPA

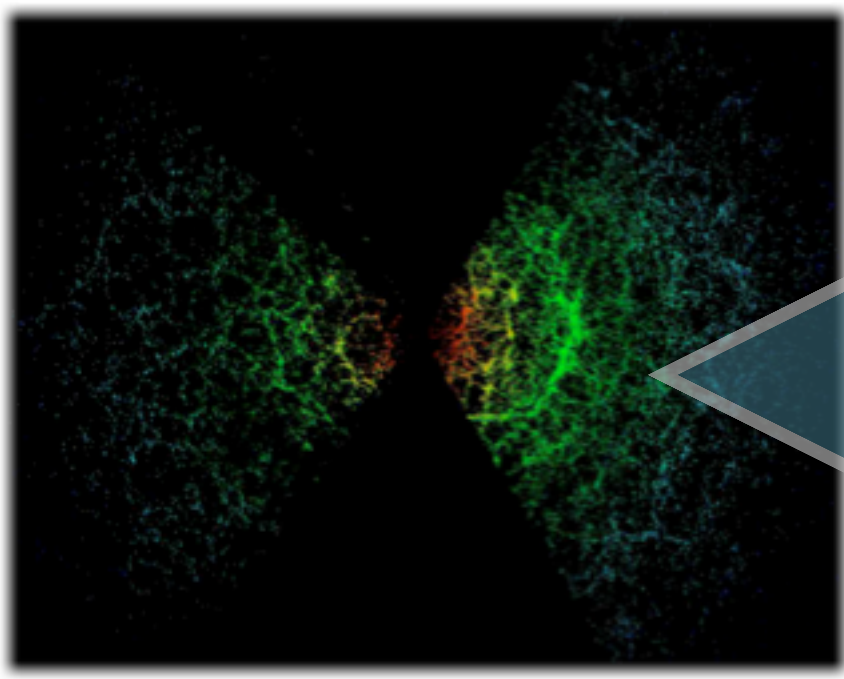
[Tojeiro et al. 2007,2009]

- Adaptable age/Z grid depending on quality and range of data.
- Z free for each age.
- Chooses the 'right' matrix to invert and does it - fast.
- Fits spectroscopy (absorption line only) and photometry
- Dust attenuation modelled with mixed-slab 2-parameter dust model of Charlot & Fall 2000
- Works with any SSP model



spSpec-52017-0366-346.fit
spSpec-52017-0366-349.fit
Flux (arbitrary units)
(y − y_r)/σ
λ[angstrom]
χ² = 1.554
χ² = 1.666
MASS fractions
Z
Lookback time (Gyrs)

What would you do with 800,000+ SFHs?



How do galaxies assemble their stellar mass? [Tojeiro et al 2011a,b, Tojeiro et al. 2012a]

Why are red and blue spirals or ellipticals different? [Tojeiro et al. 2013]

How does environment affect the formation of dark matter halos and the assembly of stellar mass? [Eardley, Tojeiro, Peacock in prep]

How does the population of type Ia supernovae progenitors evolve with redshift? [Aubourg, Tojeiro et al. 2008; Brandt, Tojeiro et al. 2010]

Does the stellar initial mass function (IMF) evolve with redshift? [Wilkins, et al. 2008]

Is the Universe homogeneous? [Hoyle, Tojeiro et al. 2012]

Matching progenitors across cosmic time and measuring RSD on large scales [Tojeiro et al 2012b]

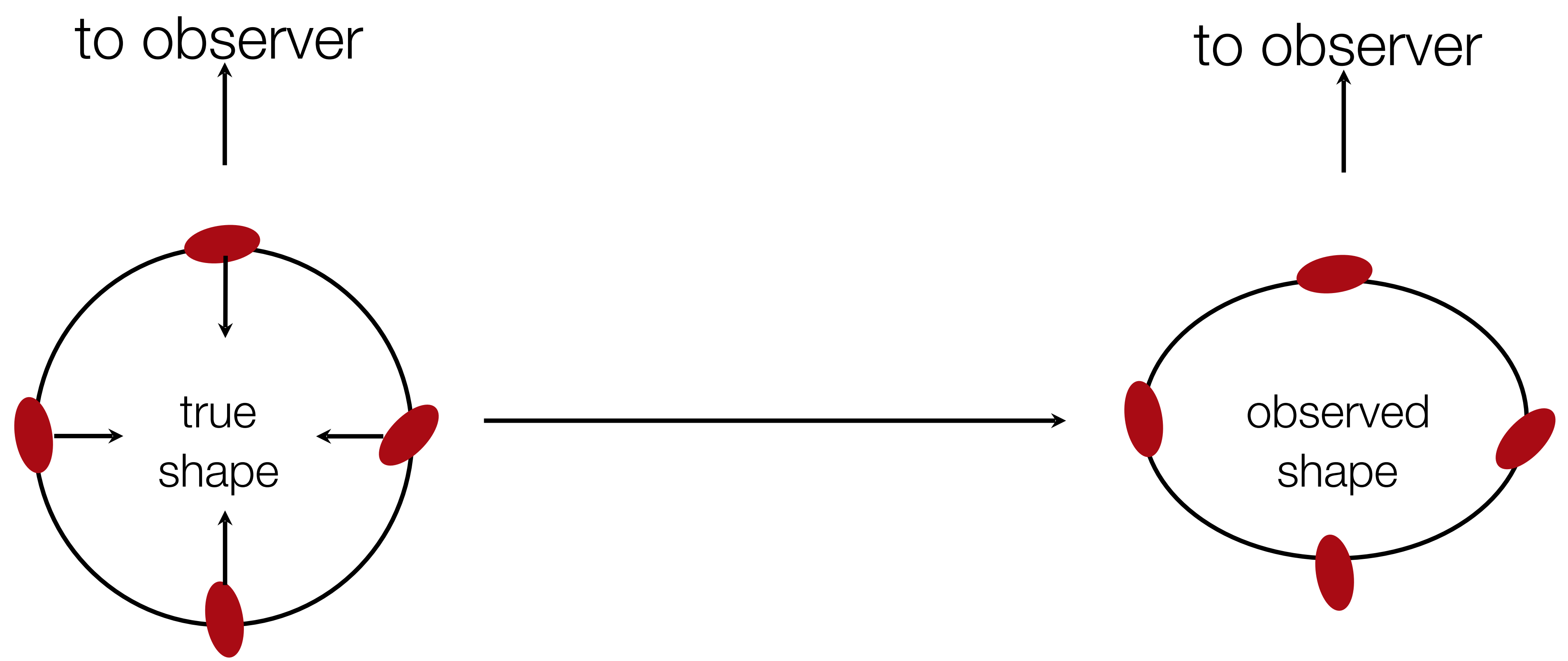
Can we use this extra dimension to help us interpret simulations? [Tojeiro, Thomas, Henriques in prep]

I - Galaxy spectra and non-parametric SFHs

II - Measuring redshift space distortions on large scales

III - The formation time of halos

Redshift-Space Distortions

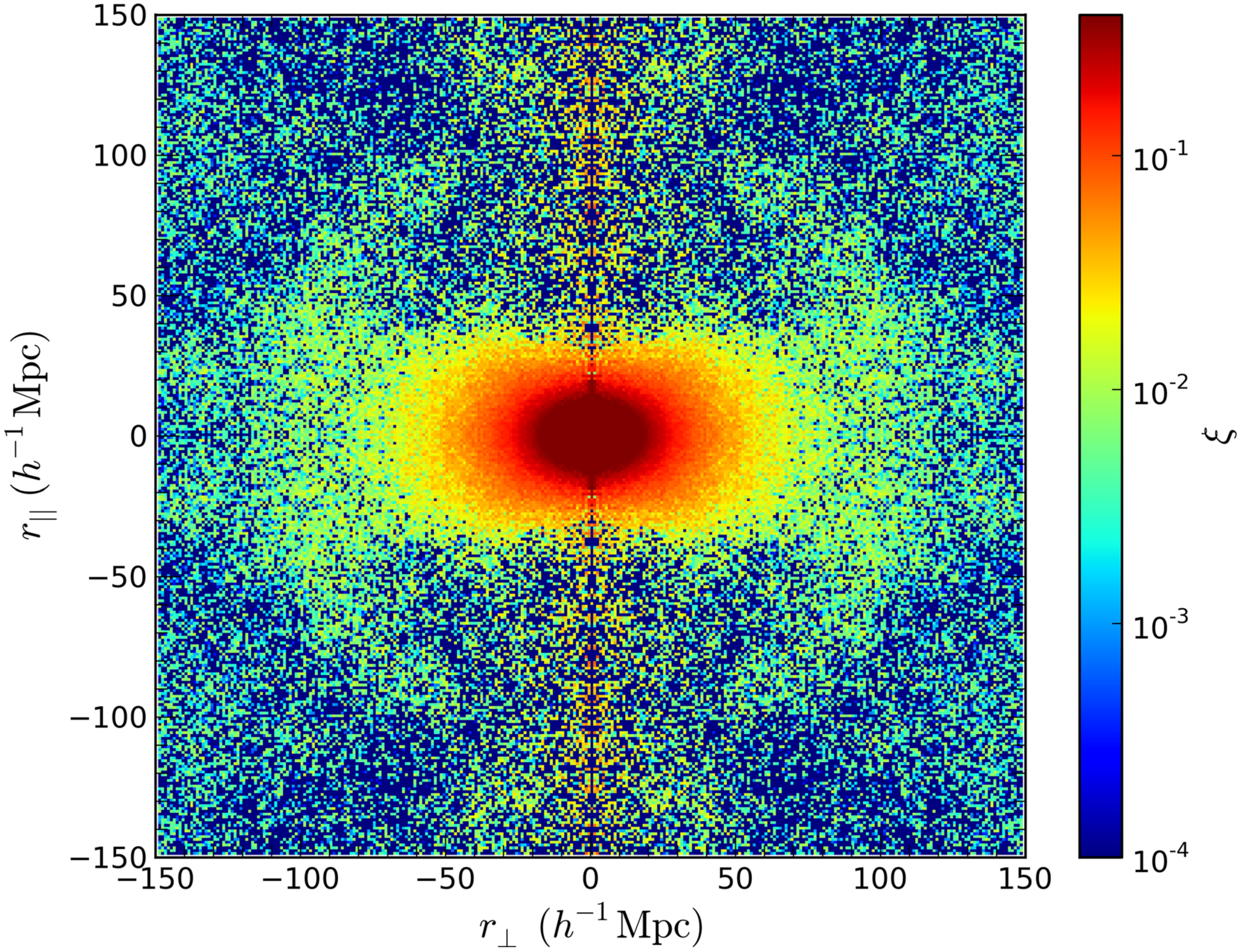


Coherent motion of galaxies as they trace the large scale gravitational potential leaves an **anisotropic** imprint in their correlation function.

This anisotropic imprint gives us the **growth rate of structure**, and is **sensitive** to the theory of **gravity**.

Probing gravity via the growth rate of structure

$$f = \frac{d \log D}{da}$$



Shape affected by observational systematics.
Small (non-linear) scales difficult to model.

Why measure RSD on large (linear) scales?

The non-linear regime is *hard.*

$$P_g(k,\mu) = (b + f\mu^2)^2 P_m(k) \qquad f = \frac{dlogD}{da}$$

[assuming no velocity bias, $\nabla.\mathbf{v} = -f\delta_m$].

The large-scale amplitude changes due to:

- the growth of the matter perturbations in the density field
- the evolution of the bias, set by the velocity field

$$P_g(k,\mu,z) = (b(z) + f(z)\mu^2)^2 \sigma_8^2(z_0) \frac{D^2(z)}{D^2(z_0)} P_m(k,z_0)$$

Simplify with multipoles:

$$\xi(\mu, r) = \xi_0(r)P_0(\mu) + \xi_2(r)P_2(\mu) + \xi_4(r)P_4(\mu)$$

$$\xi_0(r, z) = \left(b(z)^2 + \frac{2}{3}f(z)b(z) + \frac{1}{5}f(z)^2\right)\sigma_8^2(z)\xi(r)$$

$$\xi_2(r, z) = -\left(\frac{4}{3}b(z)f(z) + \frac{4}{7}f(z)^2\right)\sigma_8(z)^2\left[\xi(r) - \bar{\xi}(r)\right]$$

Monopole + quadrupole trivially measure the combinations:

$$f(z)\sigma_8(z)$$

$$b(z)\sigma_8(z)$$

…but linear scales are noisy.

The evolution of galaxy bias

For a conserved sample of tracers (no mergers), the evolution of linear galaxy bias is known exactly: [Fry 1996, Tegmark & Peebles 1998, Chan 2012, etc]

$$b(z) = [b(z_0) - 1]\,\frac{D(z_0)}{D(z)} + 1$$

- Knowing how much of the evolution of the large-scale power is due to galaxy bias, we know how much it's due to growth.

$f(z)$ $\sigma_8(z)$ $b(z)$

A practical application to data

- Measure the large-scale amplitude of a *suitable sample* of galaxies as a function of redshift, which we describe as:

$$A_0(z) = \left(b^2(z) + \frac{2}{3}f(z)b(z) + \frac{1}{5}f^2(z)\right)\sigma_8^2(z)$$

$$A_2(z) = -\left(\frac{4}{3}b(z)f(z) + \frac{4}{7}f^2(z)\right)\sigma_8^2(z)$$

- Model s8 as a smooth function (spline, polynomial, etc), parametrised by n nodes, from which growth factor and growth rate can be computed.

$$f(z) = \frac{d\log D(z)}{d\log a(z)} \qquad D(z) = \frac{\sigma_8(z)}{\sigma_8(z=0)}$$

- Full parameters are: $b_0, \sigma_8(z_{i=0,\ldots,n})$

So. We need a passively evolving sample of galaxies.

How do I find the progenitors of today's galaxies at higher redshift?

How do I measure their merger rate?

Yes, we left the complicated, non-linear structure formation behind only to enter the murky, treacherous waters of galaxy evolution, stellar population synthesis and complicated selection functions.

Welcome, to an entirely different can of worms.

stellar evolution != dynamical evolution

stellar evolution + dynamical evolution ~ galaxy evolution

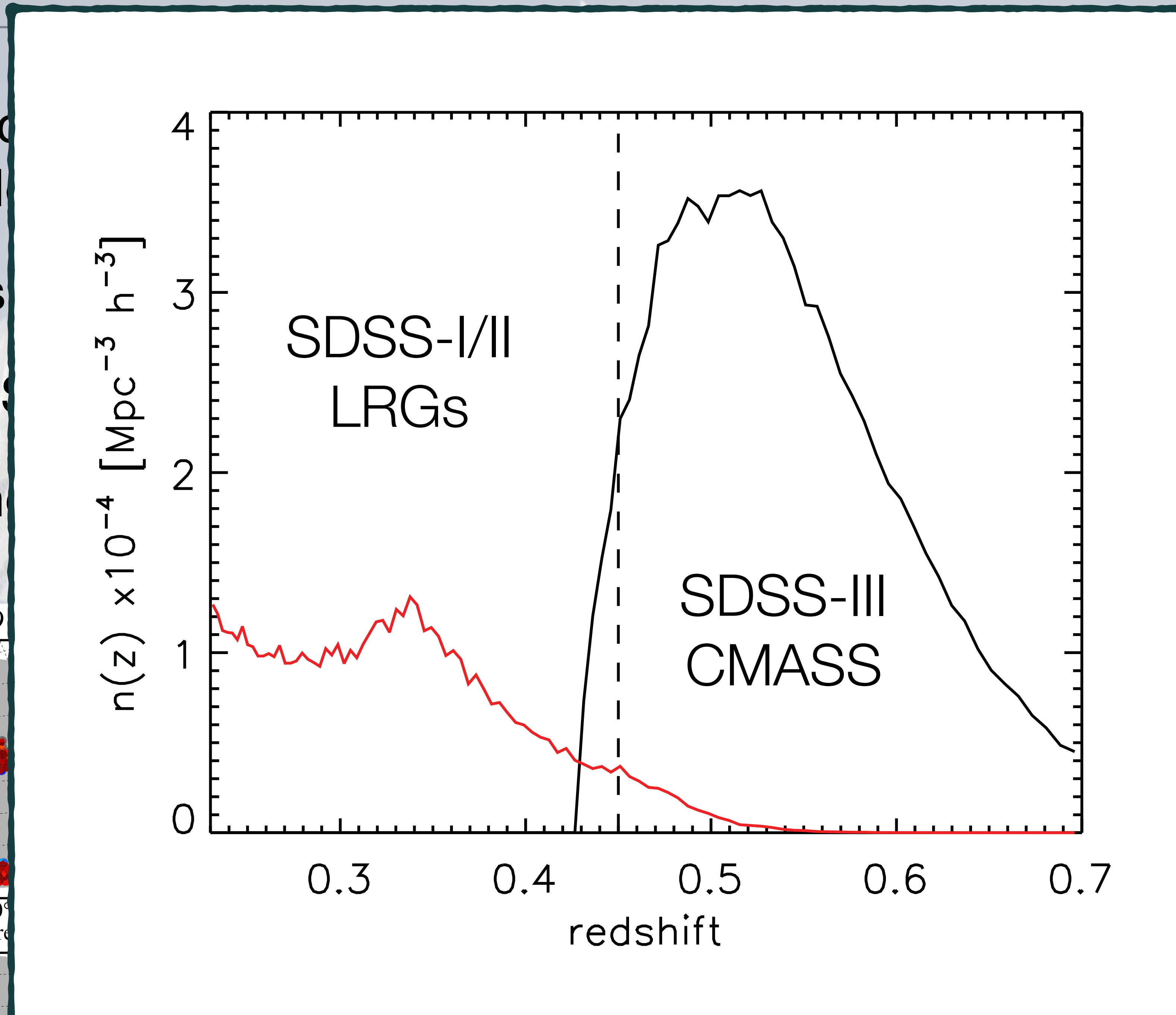
Using the past star-formation and chemical history of a local sample, and having exact knowledge of the survey selection function, one can predict what the luminosity and number density of objects should be at larger redshifts, in the absence of mergers.

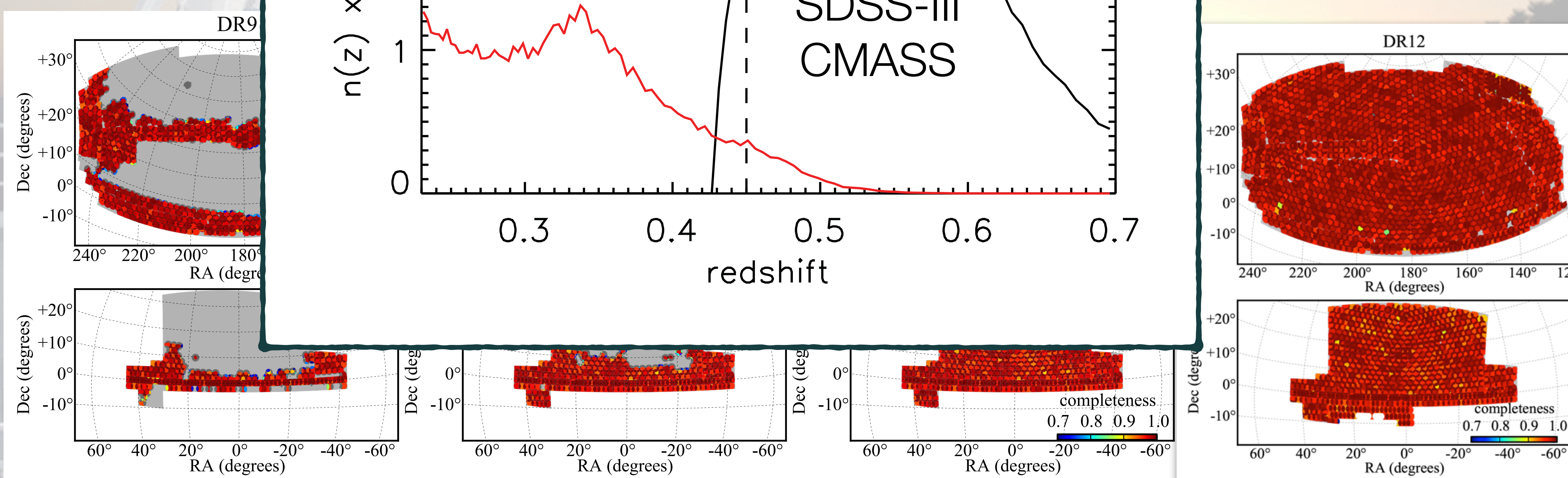
The ***difference*** can be interpreted as a merger history (this requires some assumptions).

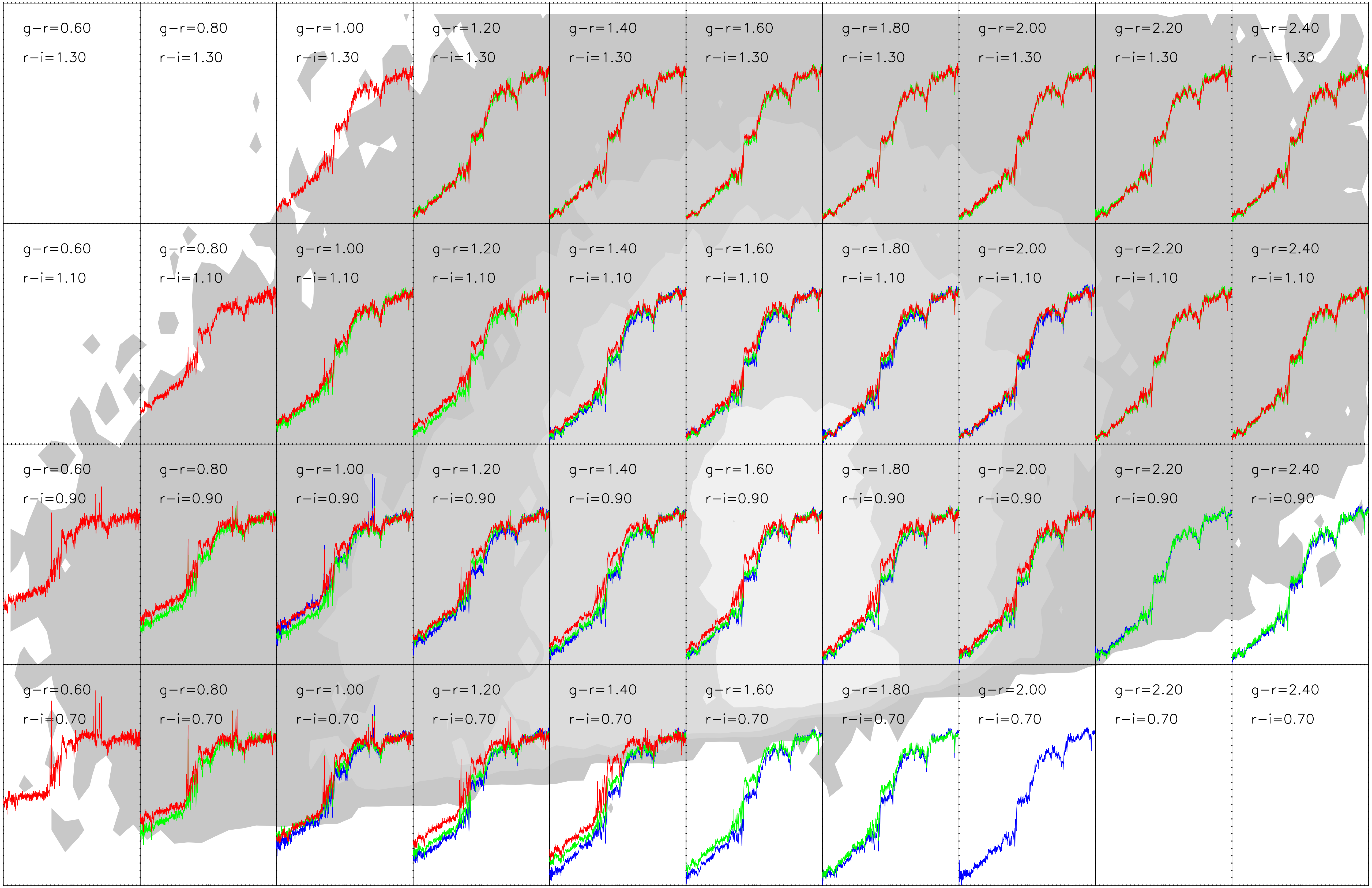
We need: non-parametric star-formation histories and lots and lots of galaxies.

The state of the art in LSS:
The Baryon Oscillations Spectroscopic Survey (BOSS)

- Spectrosc ost of the sky accessibl
- Redshifts
- + 220k Q
- All data n et al. 2015]

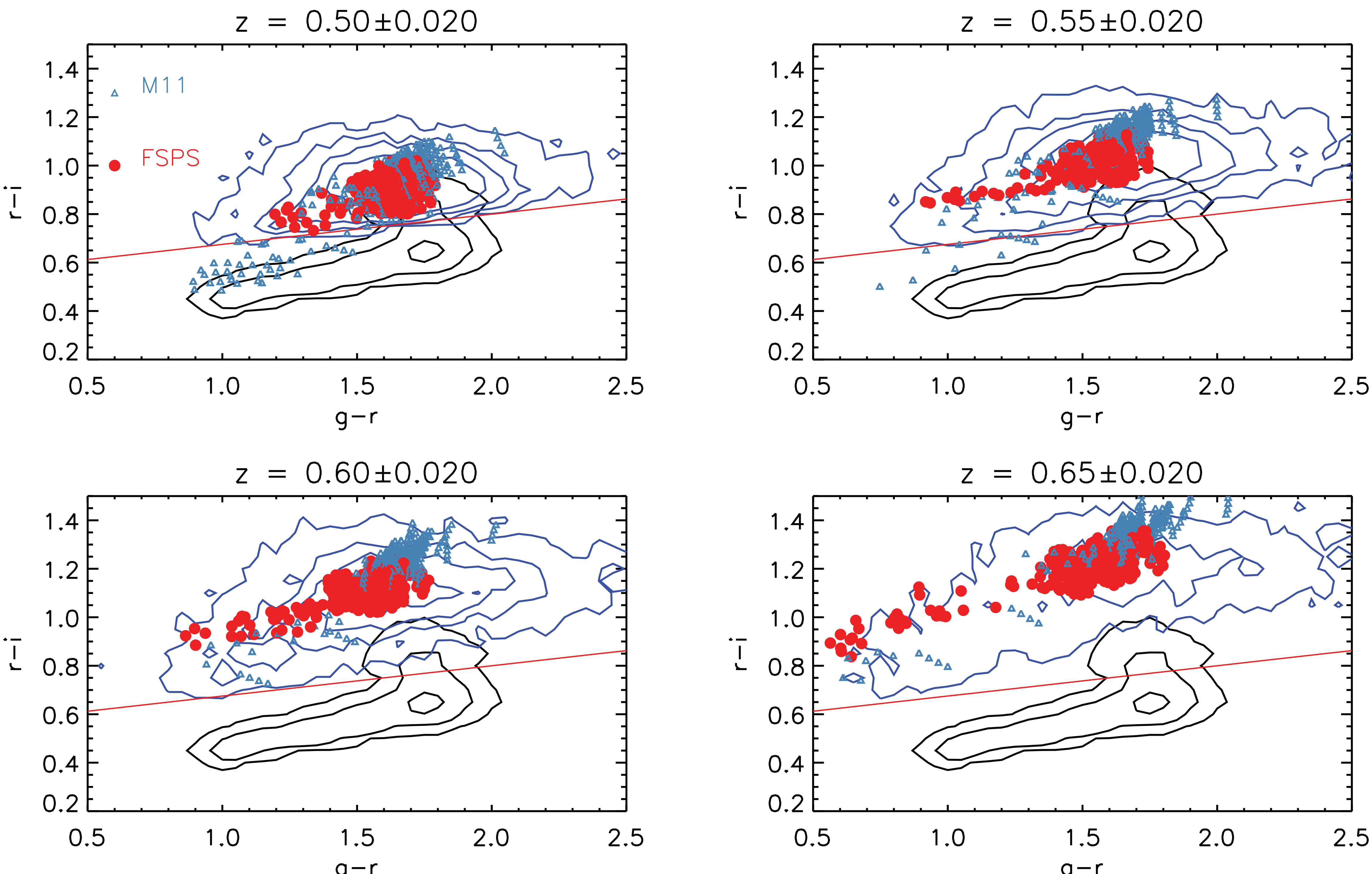






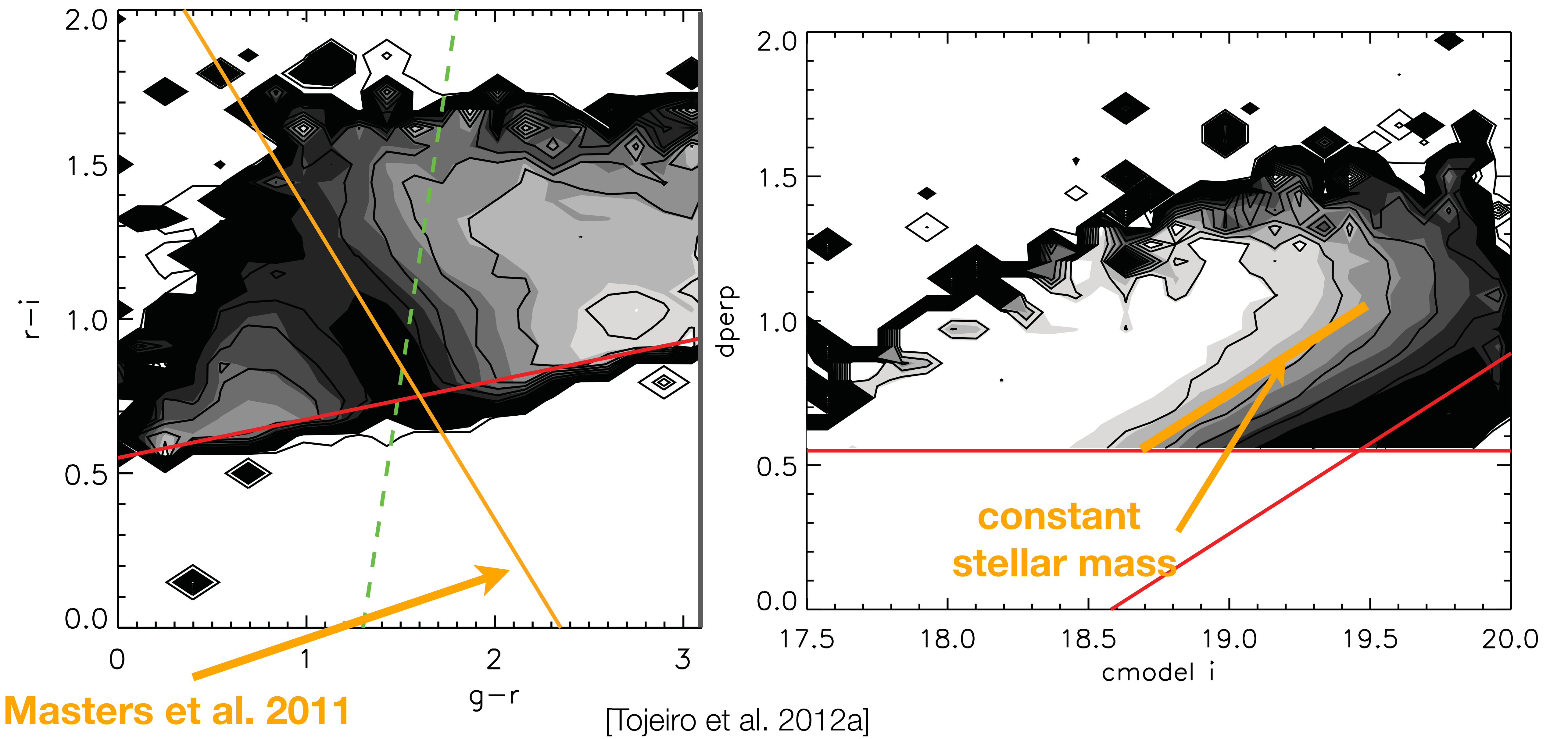
g−r=0.60 r−i=1.30
g−r=0.80 r−i=1.30
g−r=1.00 r−i=1.30
g−r=1.20 r−i=1.30
g−r=1.40 r−i=1.30
g−r=1.60 r−i=1.30
g−r=1.80 r−i=1.30
g−r=2.00 r−i=1.30
g−r=2.20 r−i=1.30
g−r=2.40 r−i=1.30
g−r=0.60 r−i=1.10
g−r=0.80 r−i=1.10
g−r=1.00 r−i=1.10
g−r=1.20 r−i=1.10
g−r=1.40 r−i=1.10
g−r=1.60 r−i=1.10
g−r=1.80 r−i=1.10
g−r=2.00 r−i=1.10
g−r=2.20 r−i=1.10
g−r=2.40 r−i=1.10
g−r=0.60 r−i=0.90
g−r=0.80 r−i=0.90
g−r=1.00 r−i=0.90
g−r=1.20 r−i=0.90
g−r=1.40 r−i=0.90
g−r=1.60 r−i=0.90
g−r=1.80 r−i=0.90
g−r=2.00 r−i=0.90
g−r=2.20 r−i=0.90
g−r=2.40 r−i=0.90
g−r=0.60 r−i=0.70
g−r=0.80 r−i=0.70
g−r=1.00 r−i=0.70
g−r=1.20 r−i=0.70
g−r=1.40 r−i=0.70
g−r=1.60 r−i=0.70
g−r=1.80 r−i=0.70
g−r=2.00 r−i=0.70
g−r=2.20 r−i=0.70
g−r=2.40 r−i=0.70

Evolving LRGs back in time

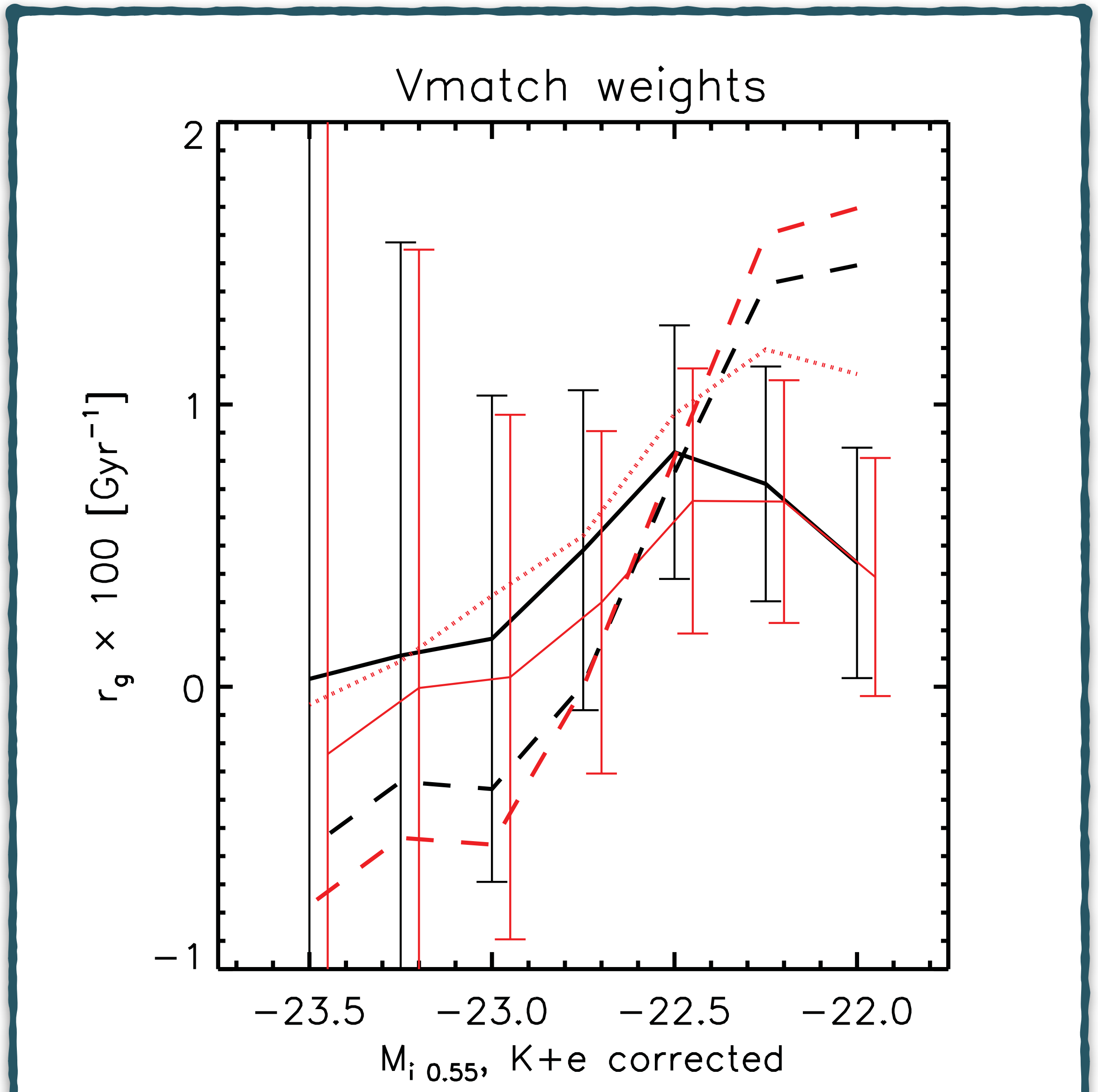


[Tojeiro et al. 2012a]

Identifying LRG progenitors at higher redshift



[Tojeiro et al. 2012a]



[Tojeiro et al. 2012a]

$$r_g = \left(1 - \frac{n'_{\rm LRG}/\ell'_{\rm LRG}}{n'_{\rm CMASS}/\ell'_{\rm CMASS}}\right)\frac{1}{\Delta t}$$

change in luminosity per galaxy from z=0.57 to z=0.3:

less than 2% Gyr^{-1} for all magnitudes: small growth due to mergers at the high mass end.

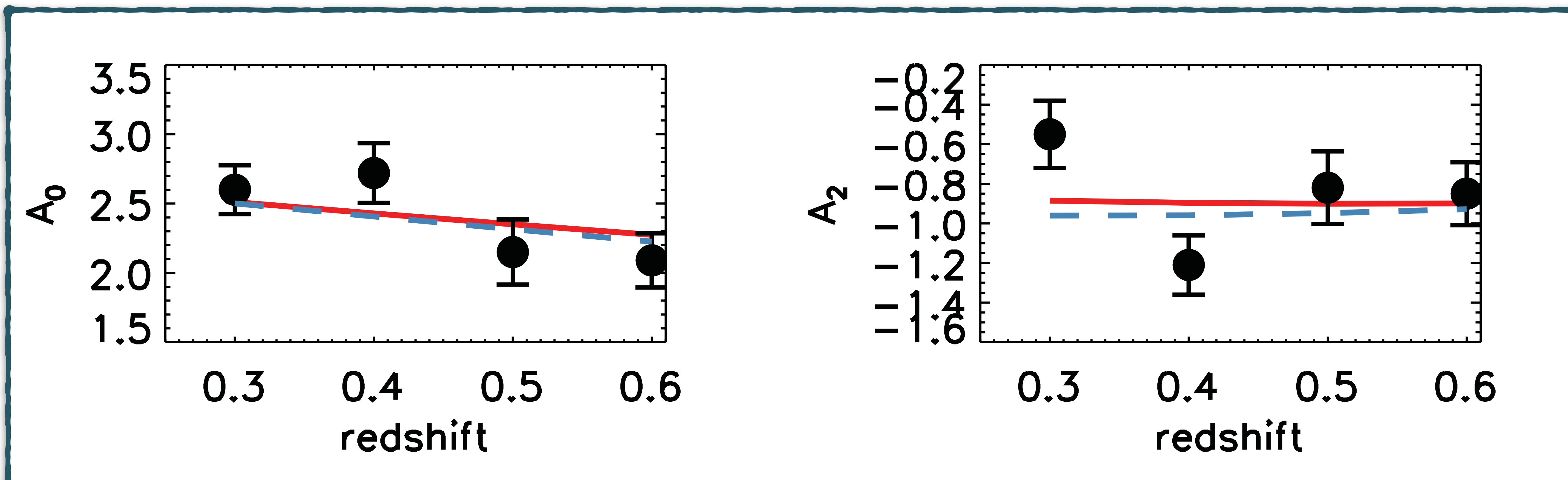
Larger rates in literature can be attributed to assuming a passive stellar evolution.

Done!

Measured amplitude of monopole + quadrupole on 4 redshift slices [fitted scales 30-80 Mpc/h].

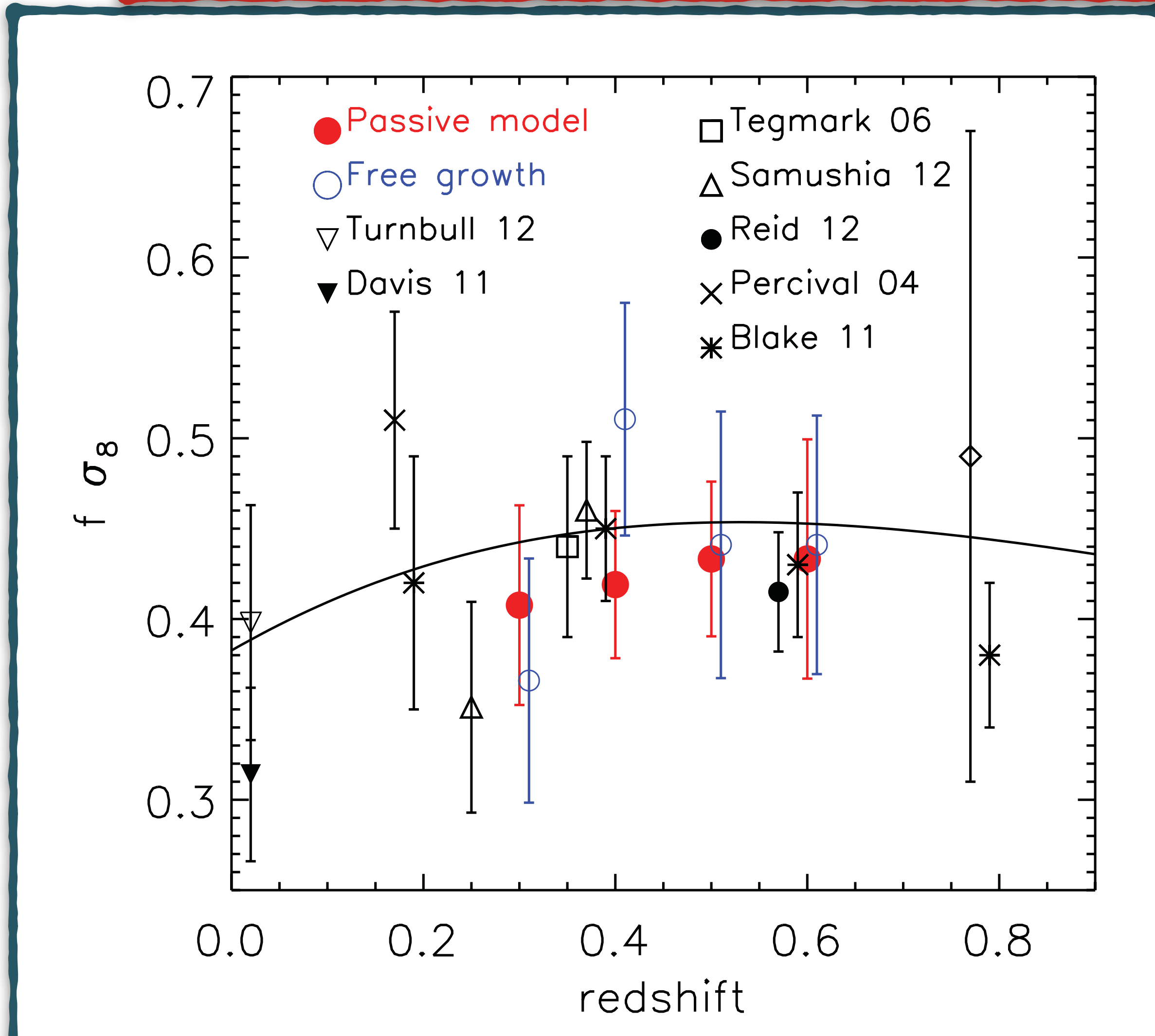
$$A_0(z) = \left(b^2(z) + \frac{2}{3}f(z)b(z) + \frac{1}{5}f^2(z)\right)\sigma_8^2(z)$$

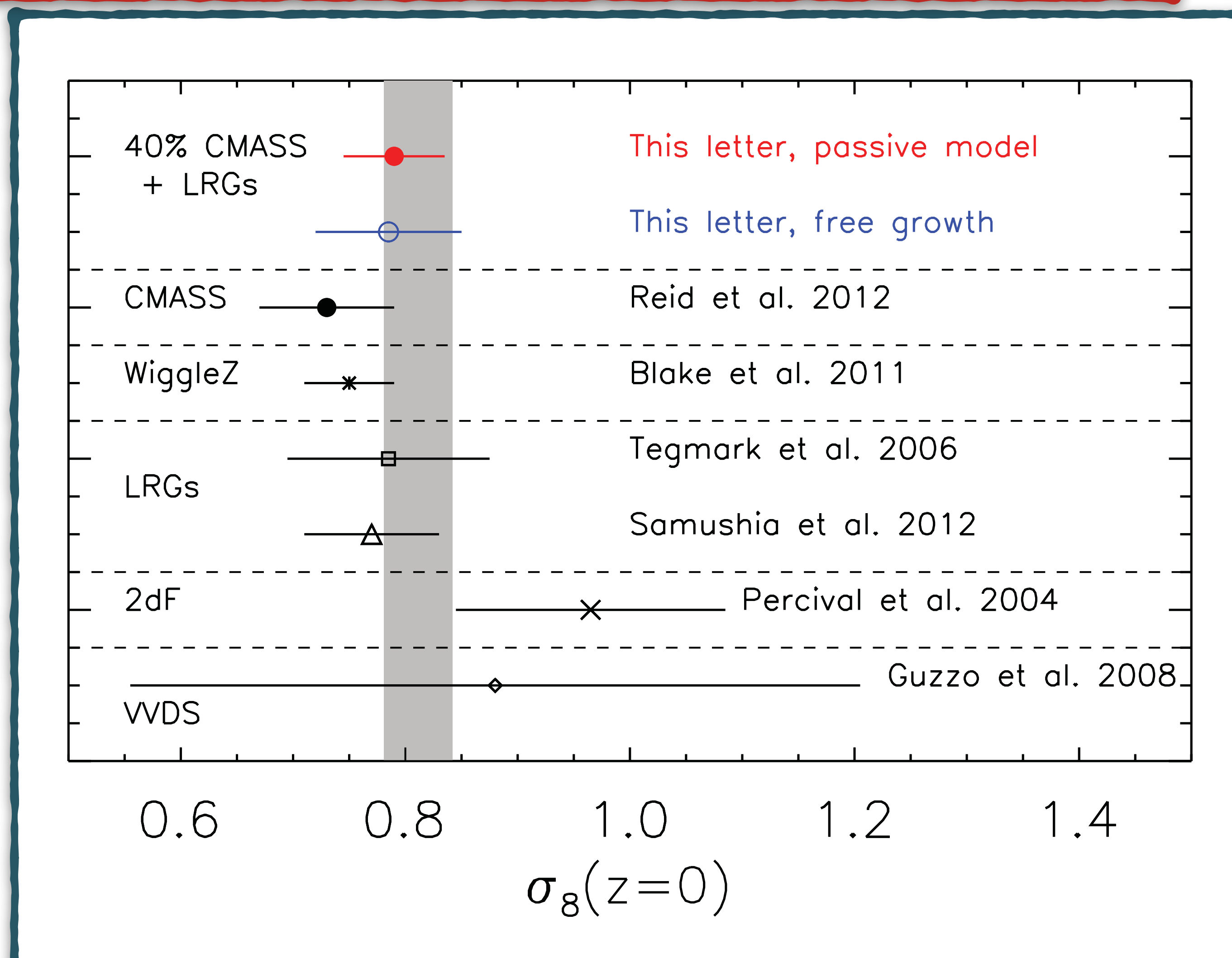
$$A_2(z) = -\left(\frac{4}{3}b(z)f(z) + \frac{4}{7}f^2(z)\right)\sigma_8^2(z)$$



fraction of samples

- Constraining power:
  - 1.5 better than a free-growth on same data and scales.
  - Comparable to state-of-the-art measurements on smaller scales.
- Get f(z) directly.
- Potential systematics: very different.





What next? Extended redshift range in upcoming surveys is rather tempting.

But - at lower luminosities, bluer colours and higher-z, passive evolution is an increasingly poor assumption. Can we learn from a more general bias model?

$$b_1(t) = 1 + \frac{\bar{n}_{gi}^{(c)}}{\bar{n}_g^{(c)}}(b_{1i} - 1)\frac{D_i}{D} + \frac{1}{\bar{n}_g^{(c)} D}\int_{\bar{n}_{gi}^{(c)}}^{\bar{n}_g^{(c)}} dn_*(b_1^* - 1)D_*$$

[Chan et al. 2012; Duckworth et al., in prep]

non-constant number density

galaxy growth/loss

The evolving number density:

$$\bar{n}_g^{(c)} = \int_{a_{int}}^{a} \frac{da}{a} \left[e^{-(lna - lna_0)^2/2\sigma_0^2} (\alpha_1 a^{-3} + \alpha_2 a^{-6}) \right]$$

$$A(t) = \frac{1}{a^3} e^{-(\ln a - \ln a_0)^2/2\sigma_0^2}$$

parametrising galaxy formation across cosmic time

$$j(\rho) = \alpha_1 \frac{\rho}{\rho_0} + \alpha_2 \left(\frac{\rho}{\rho_0} \right)^2$$

parametrising the dark matter density field

For eBOSS, we need a further w(z) to account for selection window -> completeness via clustering redshifts will help (Dominic Bates et al., in prep)

Part II: summary & future work

- RSD using large (> 30 Mpc/h) scales is a possibility and promising complementary route, given a carefully chosen and weighed galaxy sample.
- Passive galaxies have yielded the first measurement of this sort, using the simple bias model of Fry 1996.
- Extension to higher-z, fainter or bluer galaxies will require a more accommodating bias model, now under exploration.
- Future work: fitting expansion history simultaneously; quasi-linear effects; velocity bias; HOD model; MG tests with f(z).

Tojeiro et al. 2012 MNRAS 424 136
Tojeiro et al. 2012 MNRAS 424 2339
Bates, Tojeiro et al. in prep
Duckworth, Tojeiro et al. in prep

I - Galaxy spectra and non-parametric SFHs

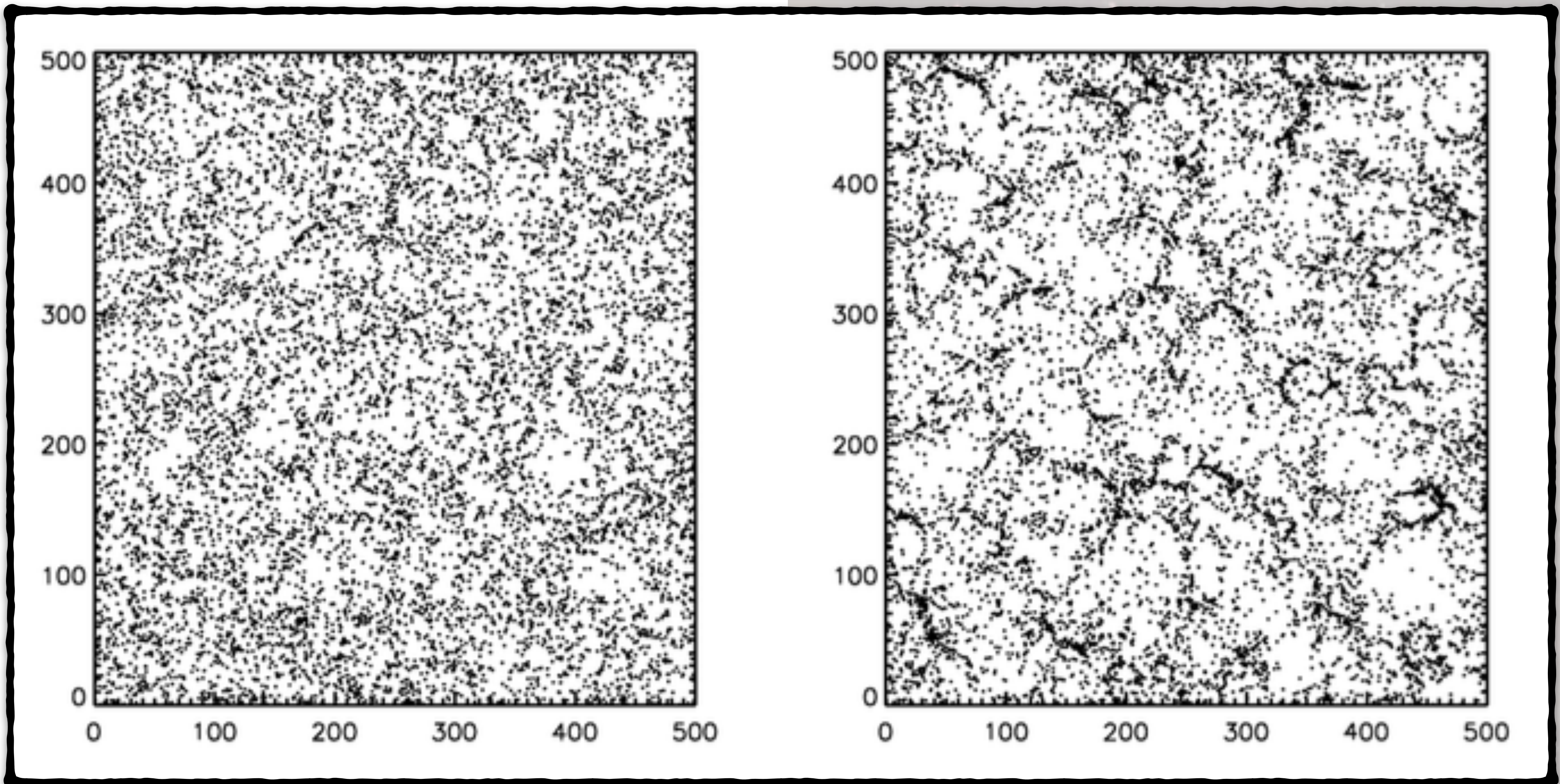
II - Measuring redshift space distortions on large scales

III - The formation time of halos

Halo/galaxy assembly bias

- Theoretical models of the halo-galaxy relationship assume that *galaxy populations in DM halos depend* ***only*** *on halo mass.*
  - very successful at describing the clustering of galaxies of different luminosity, colour or environment.
- However, simulations shows that the clustering of DM halos depends not only on their mass but also - often in a complex way - on their *assembly history.* I.e. halos of the same mass cluster differently according to how long ago they assembled their mass: **assembly bias**.

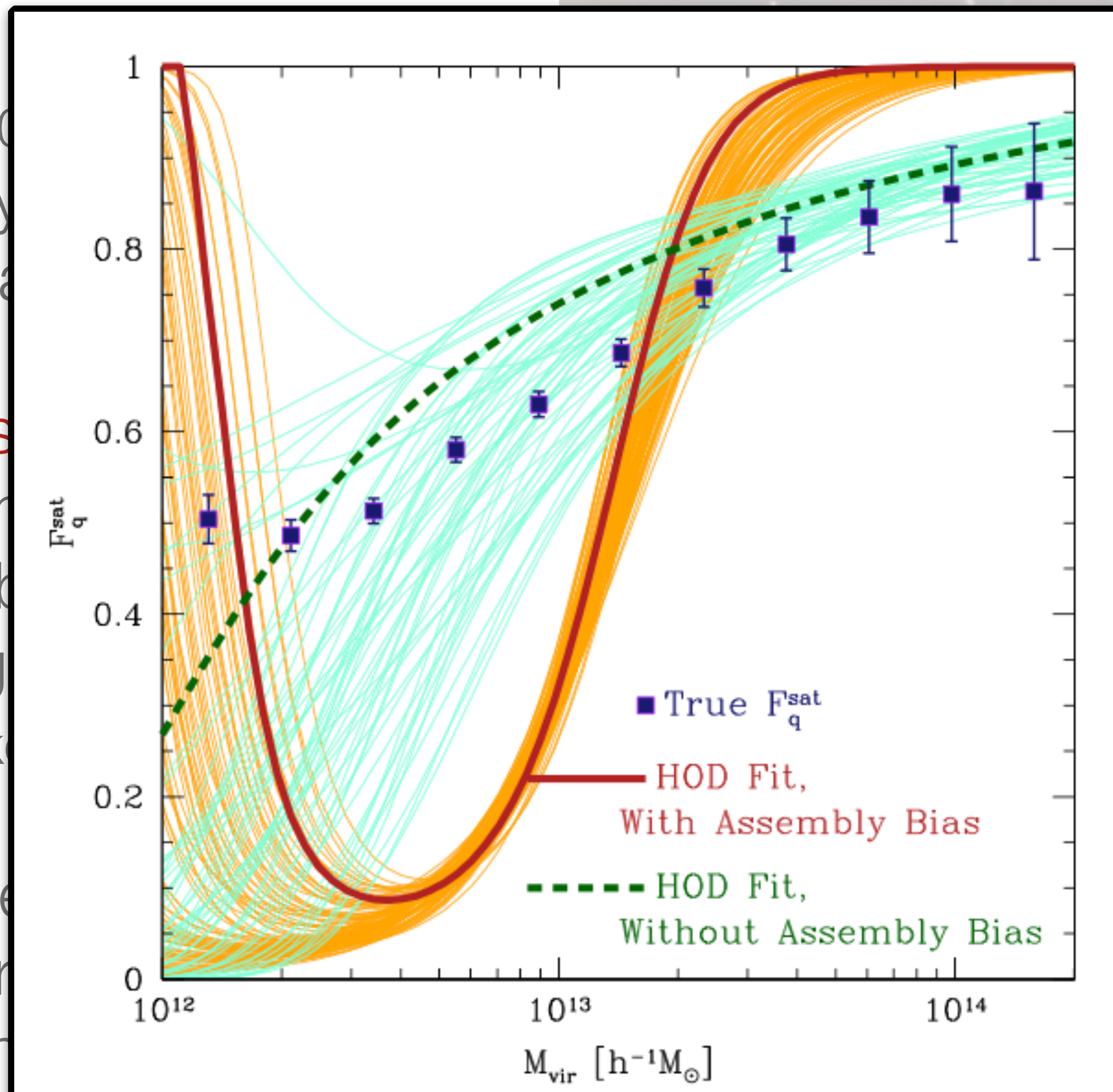
Simulations and observations



[Gao et al. 2005]

Simulations and observations

- Halo assembly ... & White 2005; Wechsler et al. 20... 7; Li, Mo & Gao 2008], usually by ... on of halo assembly time ...



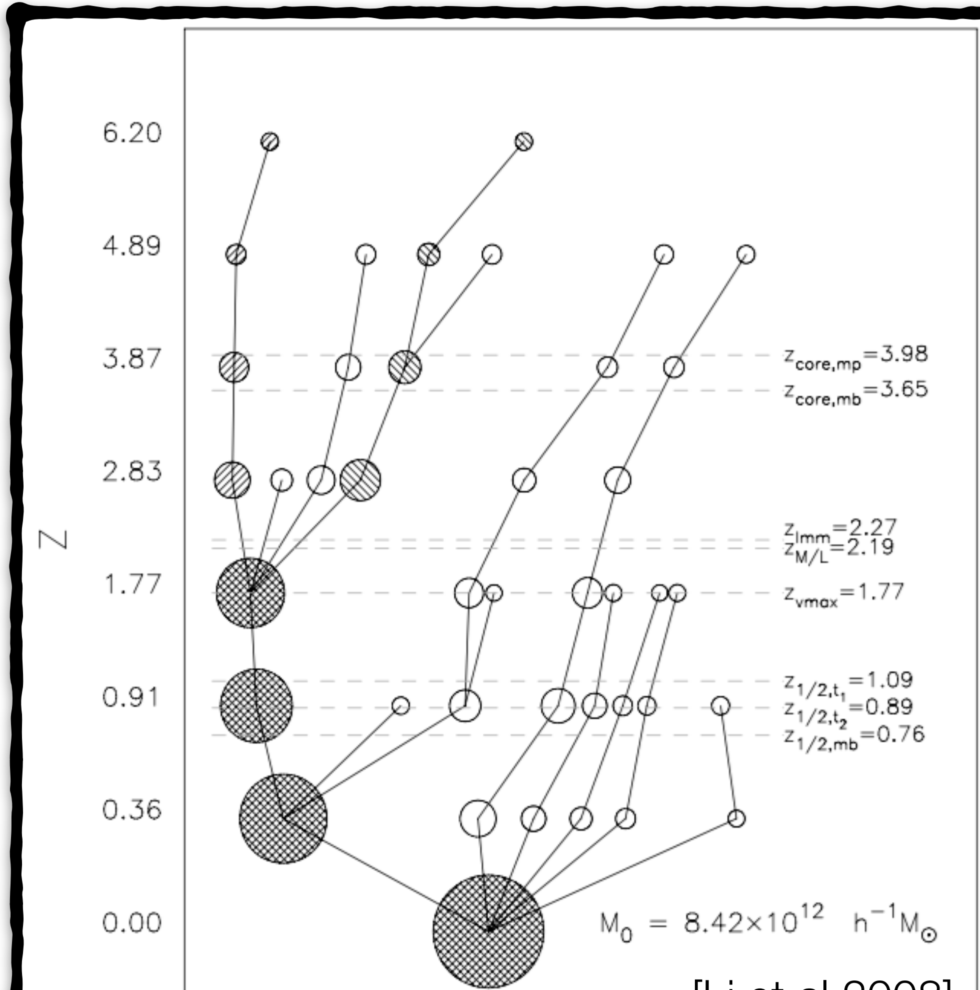
- Results are less ... mplitude some have foun... consistent with assembly b... g et al. 2008, 2013], but using ... . Blanton & Berlind 2007; Tink...

- Recently Zentn... ignoring halo assembly bias ... d galaxy-halo relationship fro...

Halo formation times

- Part of the difficulty is finding a good observational proxy for halo formation time: might resolved SFHs help?

- We investigate this using the SAM of Henriques et al. 2015, run on the Millennium simulation.



[Li et al 2008]

When has a halo formed?

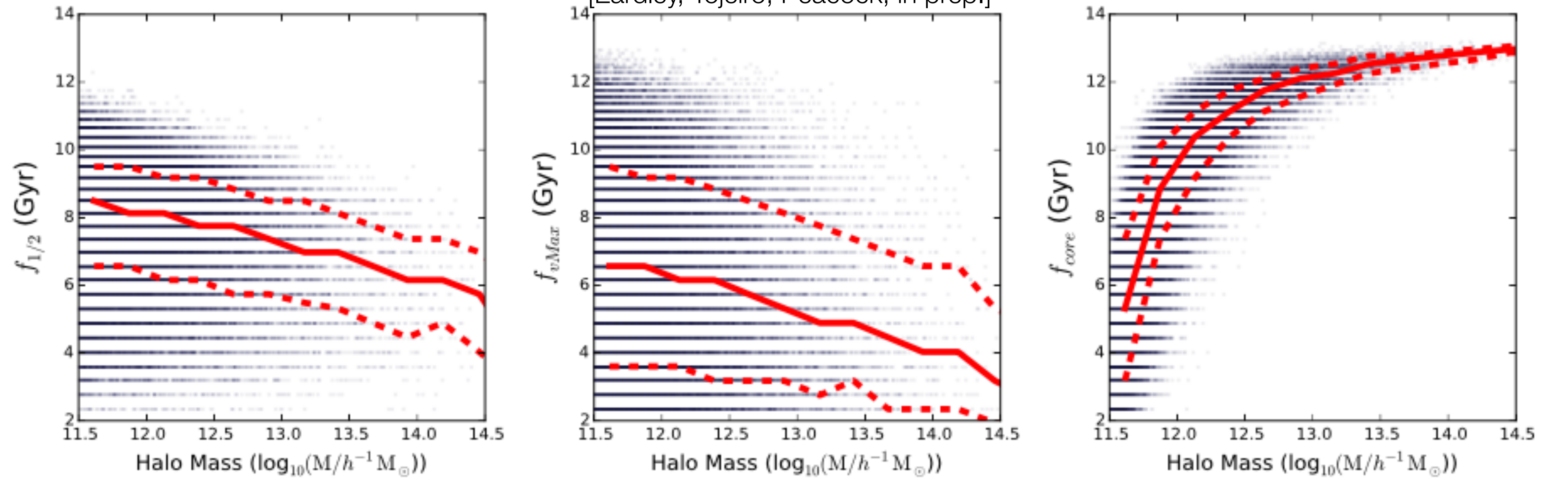
Halo formation time is not a well defined quantity – how can we best characterize a halos assembly history with one number? Below are some of the approaches that have previously been used (*Li 2008 et al.*).

$f_{1/2}$ Earliest time at which a progenitor had at least half the final halo mass,

f_{vMax} The time at which the halos virial velocity reaches its maximum value over the entire accretion history

f_{core} The time at which a progenitor reaches a fixed mass, $M_c = 10^{11.5} h^{-1} M_\odot$

[Eardley, Tojeiro, Peacock, in prep.]



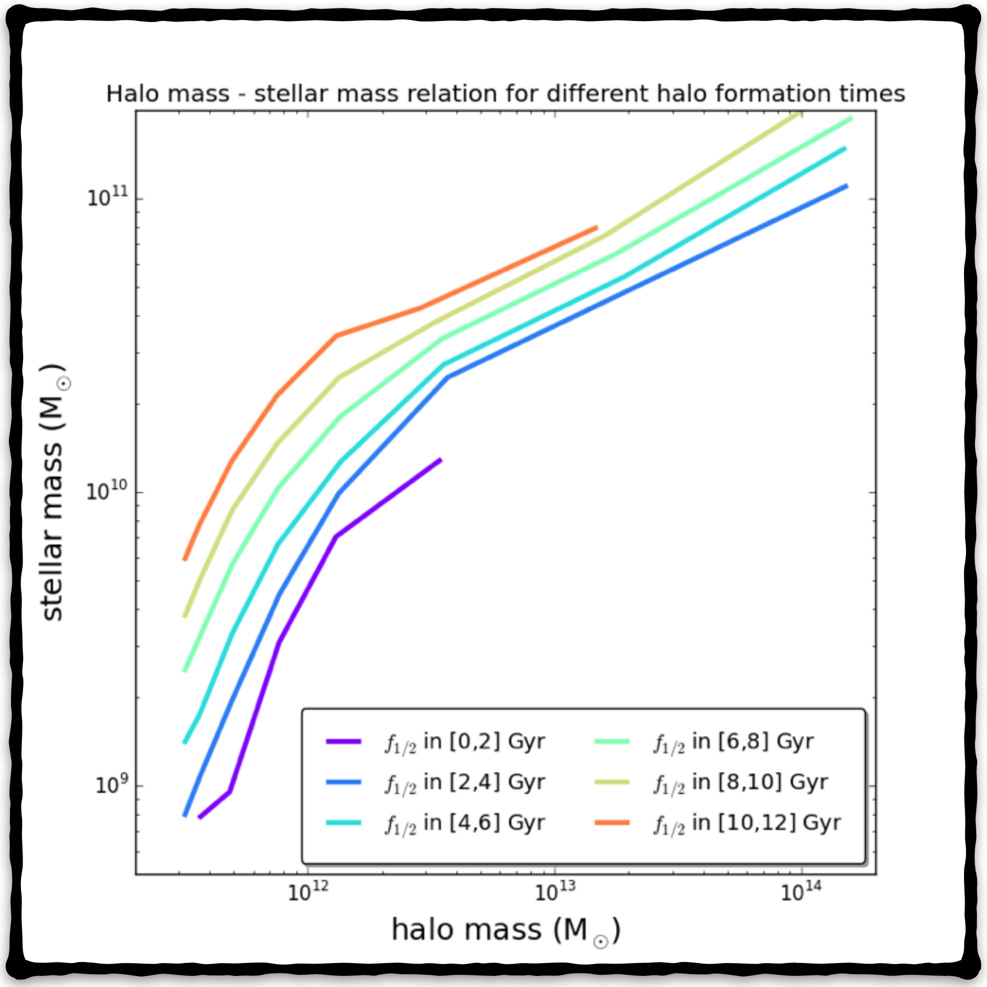
Potential observational proxies:

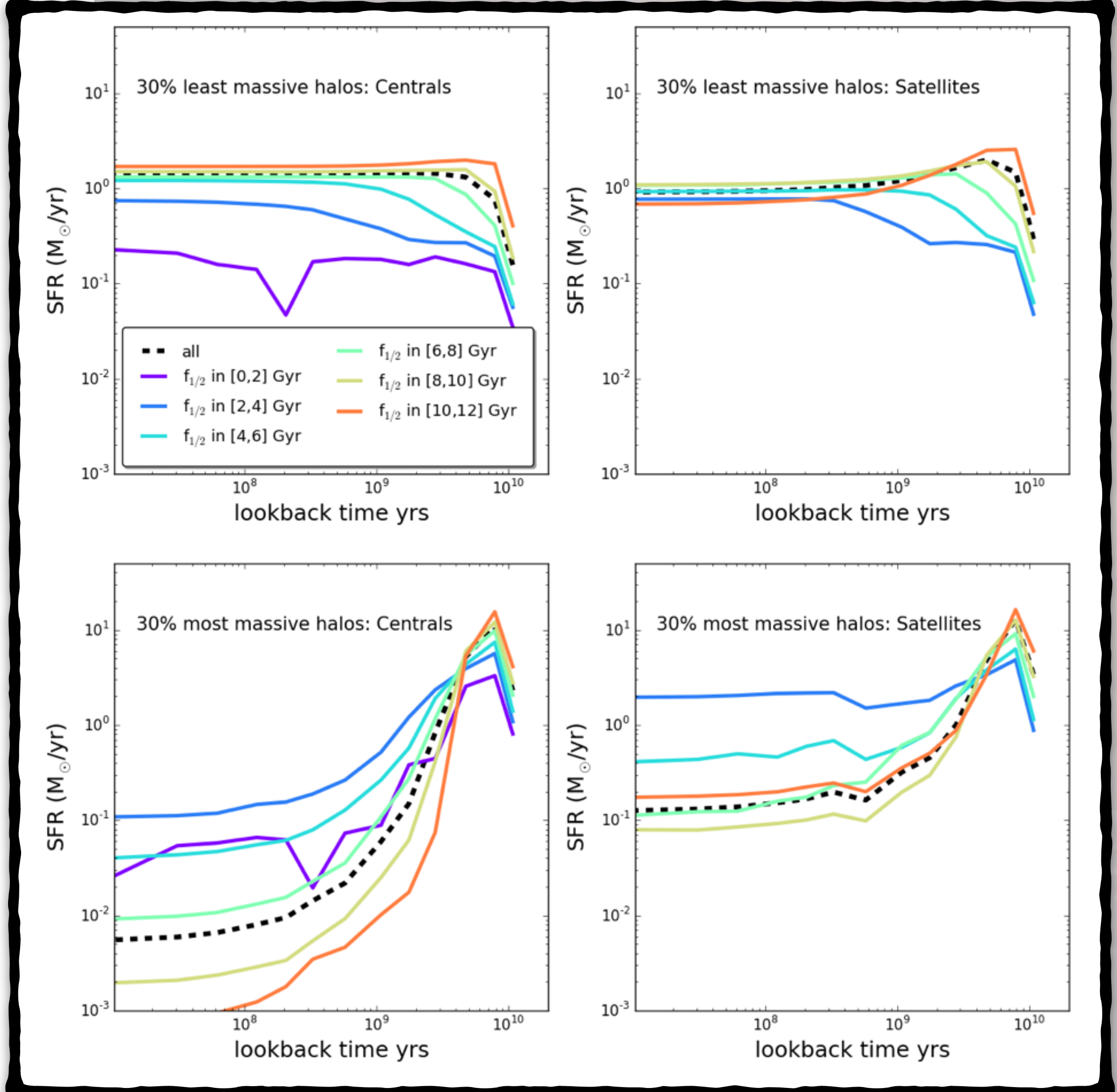
Instantaneous SFR

Mass-weighted age

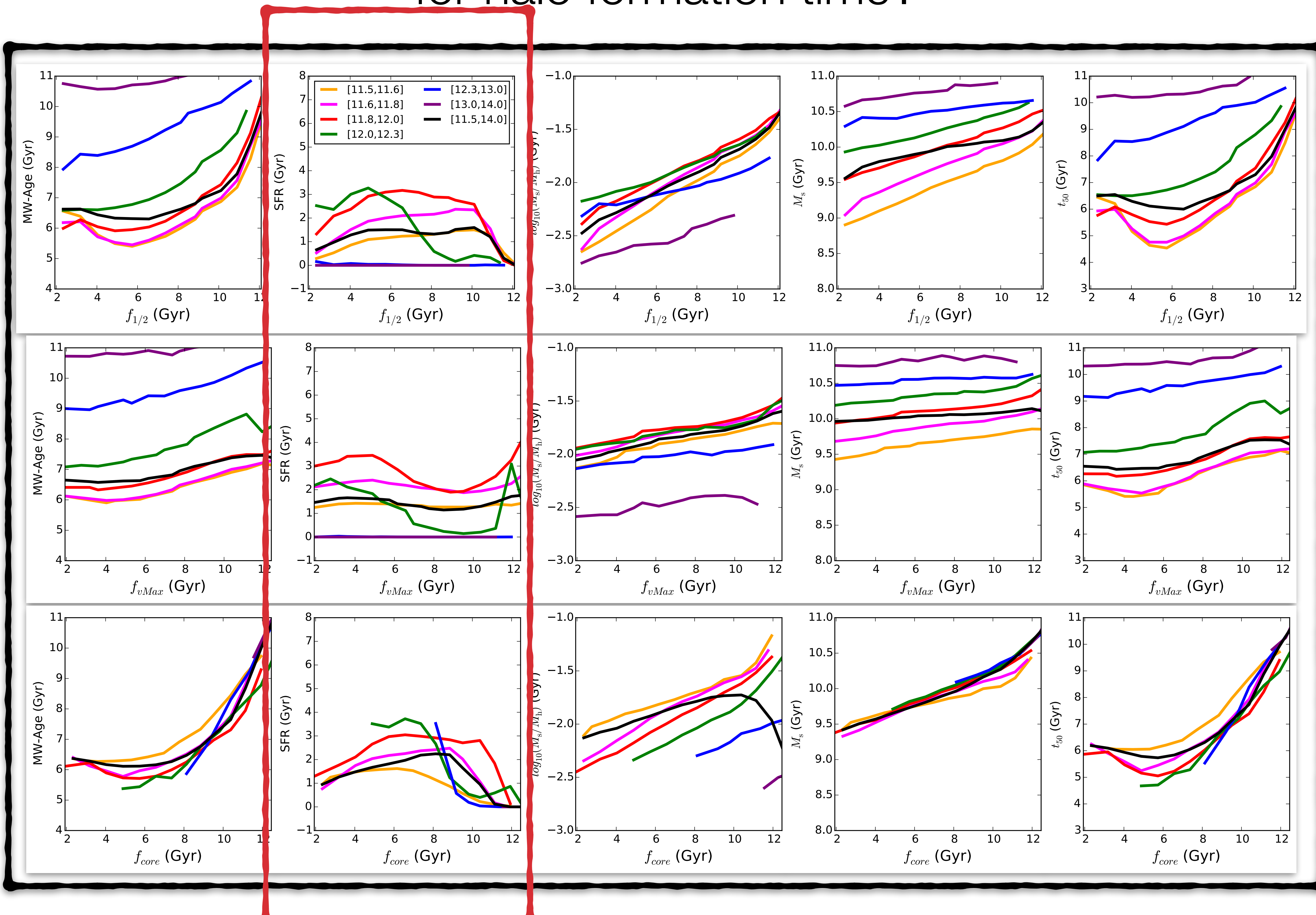
$M_{stellar} / M_{halo}$

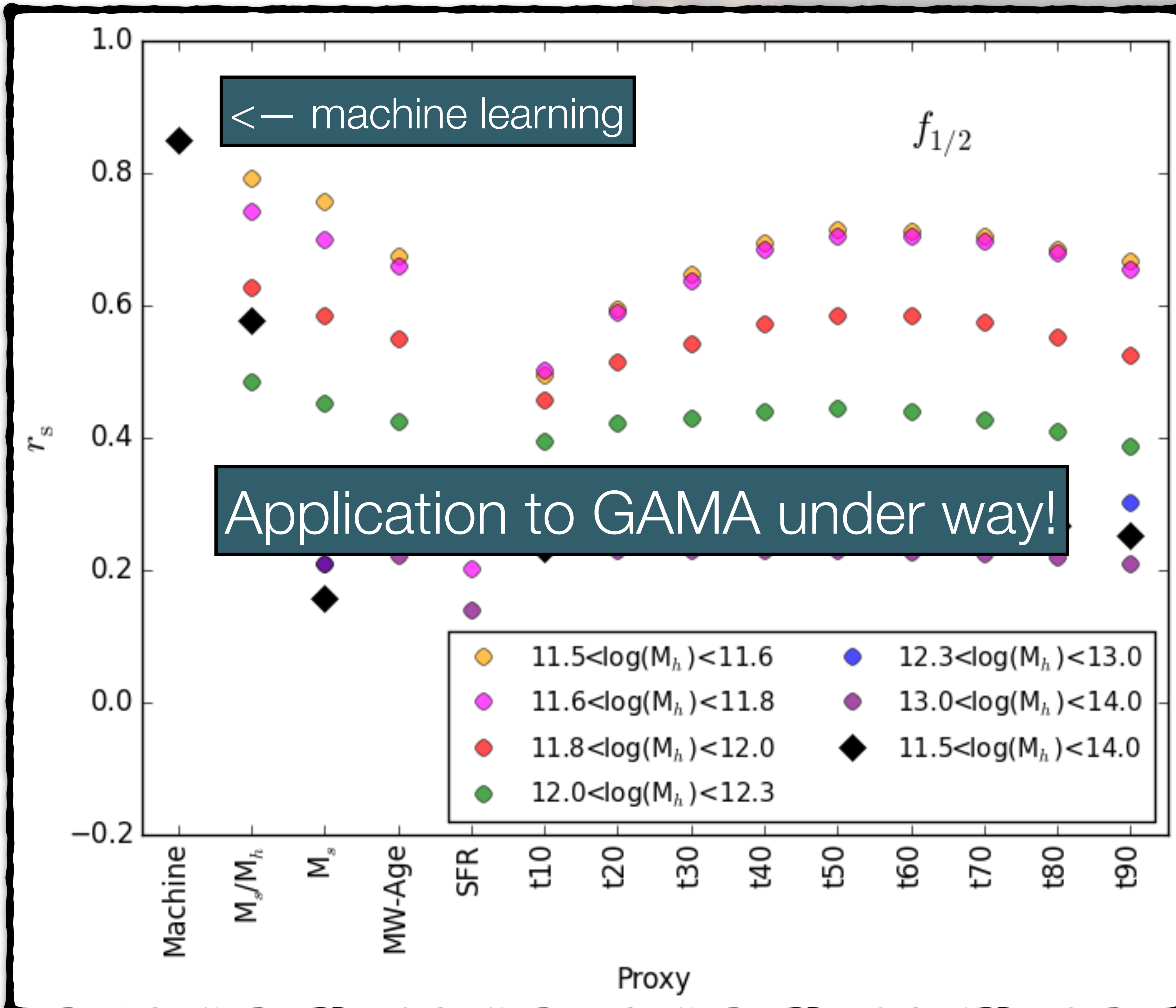
t_{xx} - time at which xx% of the stellar mass formed





Can we find a good observational proxy for halo formation time?





<— machine learning
Application to GAMA under way!
$f_{1/2}$
r_s
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
11.5<log(M$_h$)<11.6
11.6<log(M$_h$)<11.8
11.8<log(M$_h$)<12.0
12.0<log(M$_h$)<12.3
12.3<log(M$_h$)<13.0
13.0<log(M$_h$)<14.0
11.5<log(M$_h$)<14.0
Machine
M$_s$/M$_h$
M$_s$
MW-Age
SFR
t10
t20
t30
t40
t50
t60
t70
t80
t90
Proxy

Part III - summary & future work

- Assembly bias has the potential to affect many studies in galaxy evolution and cosmology. Lots of beautiful work on the theory side, but ***observationally very hard to study***.

- We investigate proxies for halo formation time using a SAM.

    - $\mathbf{M_{stellar}/M_{halo}}$ seems like the **best** predictor of formation time, but a measurement of the **shape of the SFH helps**.

    - Large dependency on halo mass regardless.

    - From simulations, we expect other halo properties (e.g. concentration) to affect the statistical properties of the halos and galaxies within them - can we ever hope to observationally disentangle them? ***What observables should we be focusing on?***

- Application to GAMA underway (where photometry and well matched multi-wavelength aperture photometry is now available).

Overall summary and take-home messages

- Abundant information in optical spectra that is comfortably sufficient to answer a number of interesting and unexplored questions.
- Full-spectral fitting is challenging (ask me why) but worth it.

Rita Tojeiro
STFC Ernest Rutherford Fellow
L’Oréal for Women in Science UK&I Fellow
University of St Andrews

University of St Andrews

Science & Technology Facilities Council

For Women in Science

FONDATION L'ORÉAL

- Non-parametric SFHs open up a vast parameter space space that is still largely unexplored - and are the *only way* to tackle some important problems.
- Simulations are wonderful validation tools.
- Non-parametric SFHs from panchromatic + spectral data now a real possibility.

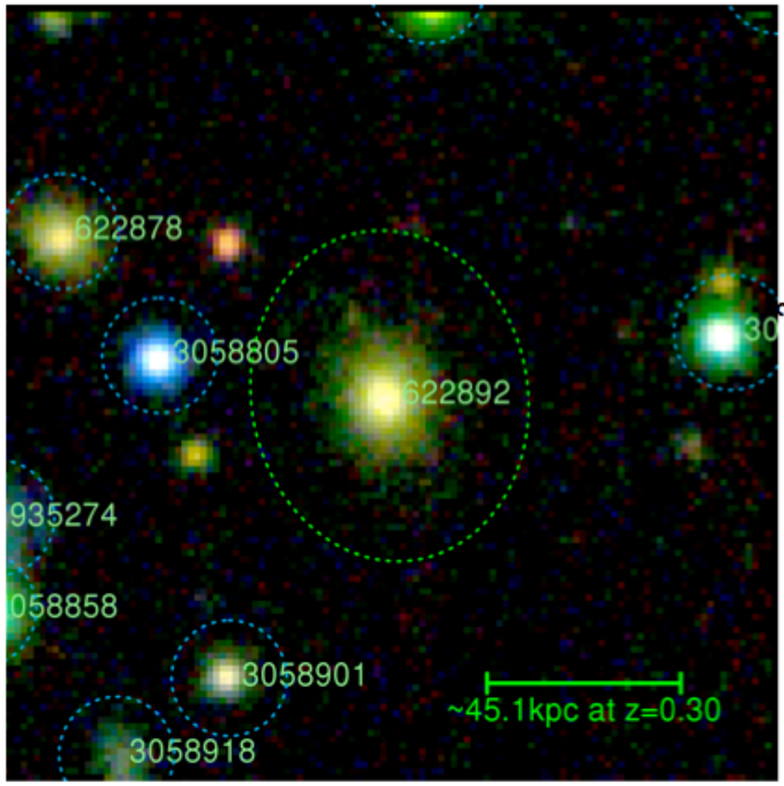
Thanks!

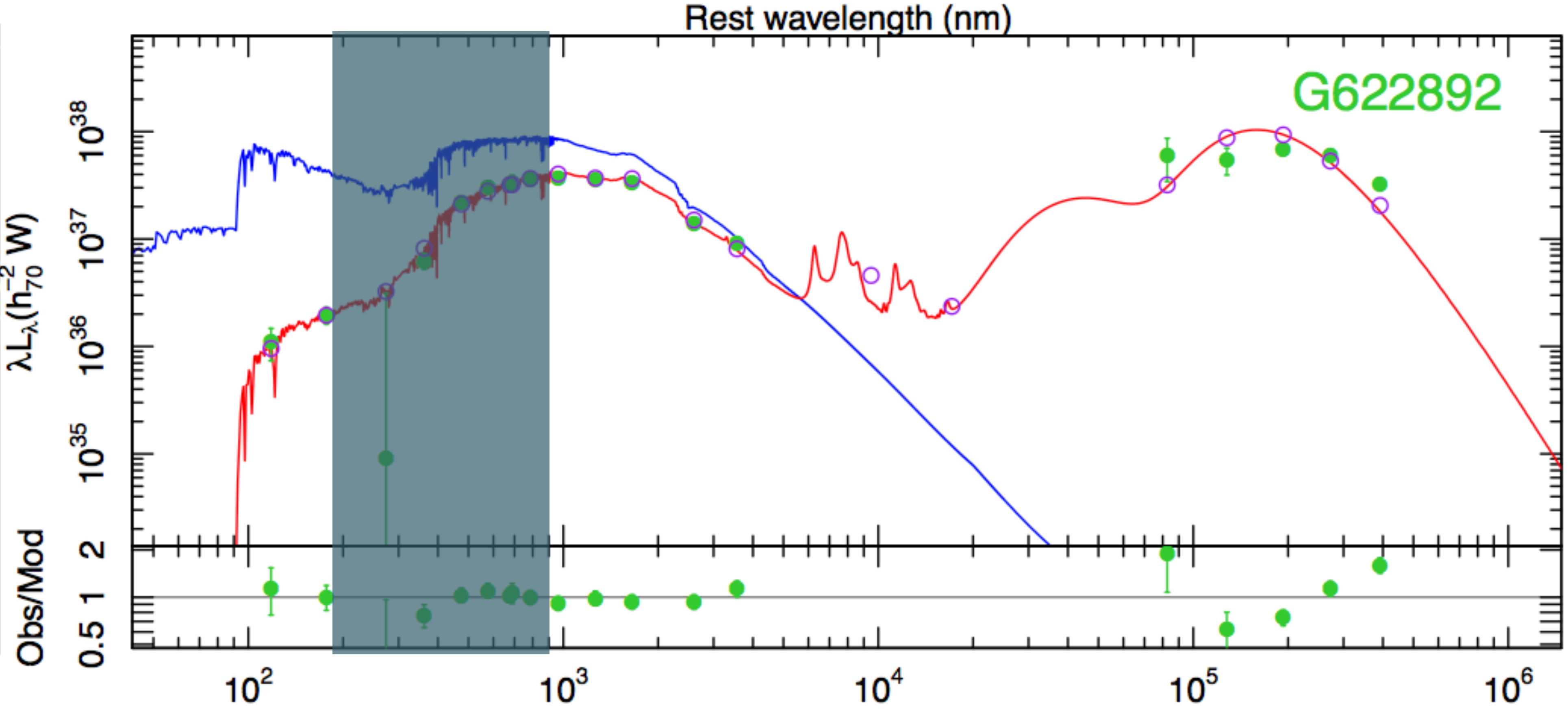
more stuff

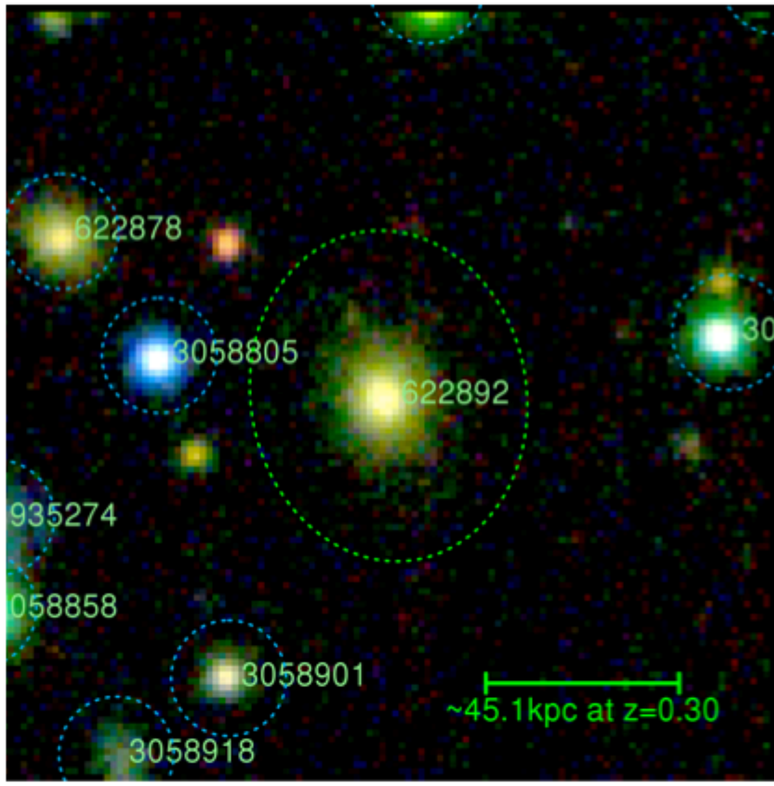
What next?

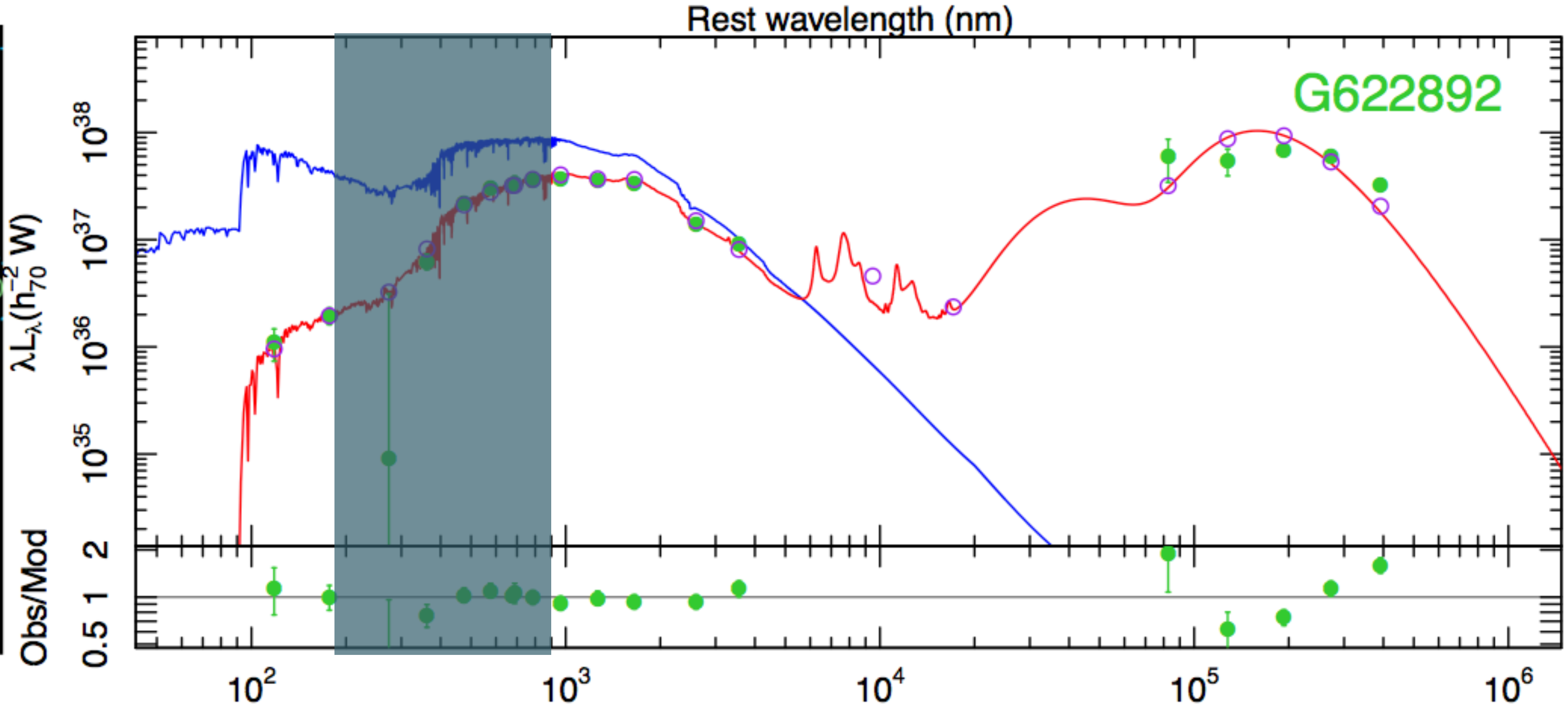
- Suite of "spectral observations" from Hen15 light-cone (and EAGLE to follow) to mimic a range of surveys. *Fundamental tool to understand and quantify limitations of any methodology.*

We need to do this beyond the optical!



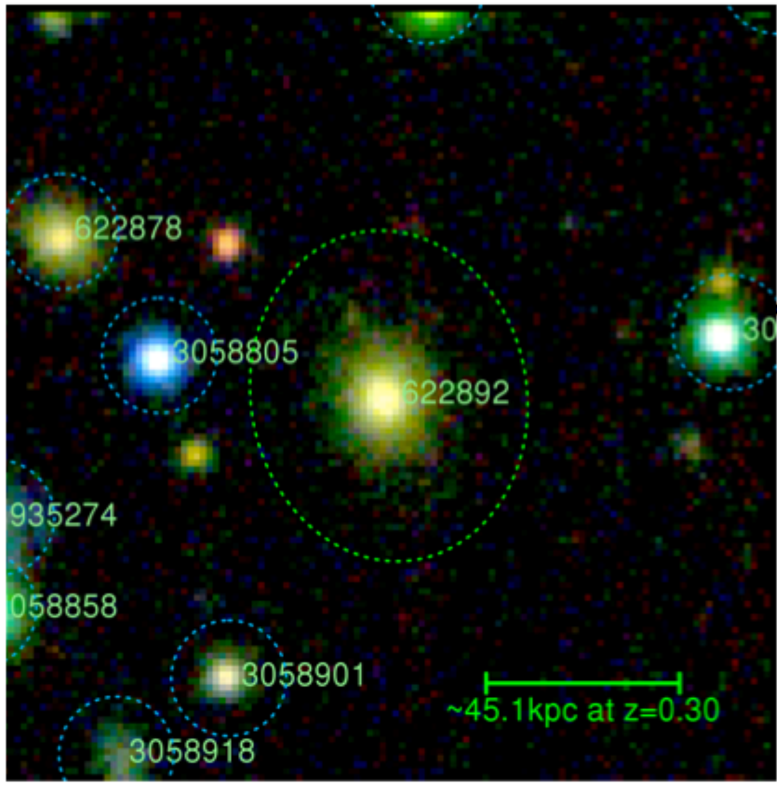


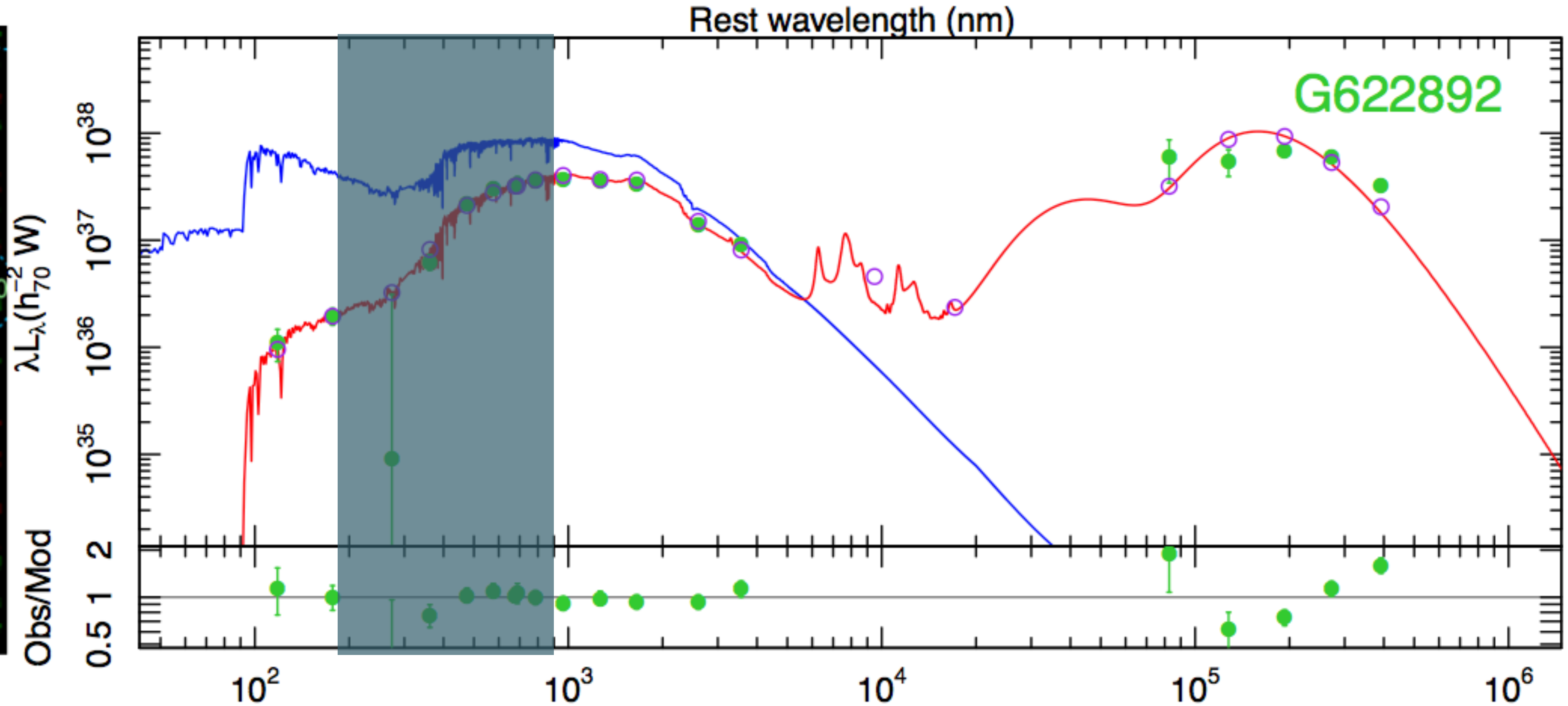




Why?

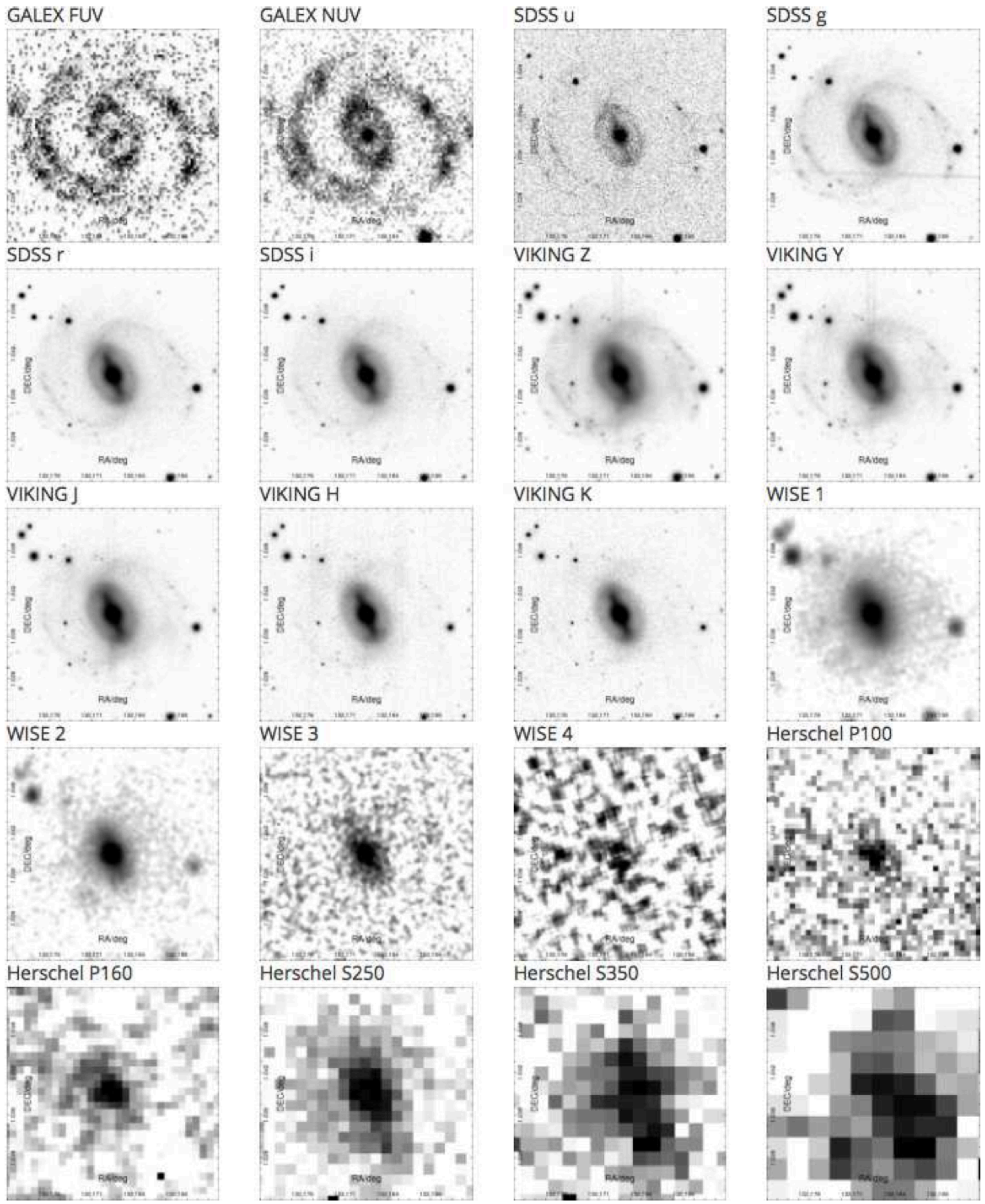
- Ability to pin down young and old populations.
- Good treatment of dust (stellar masses independent of inclinations).
- Higher resolution in lookback time: science!

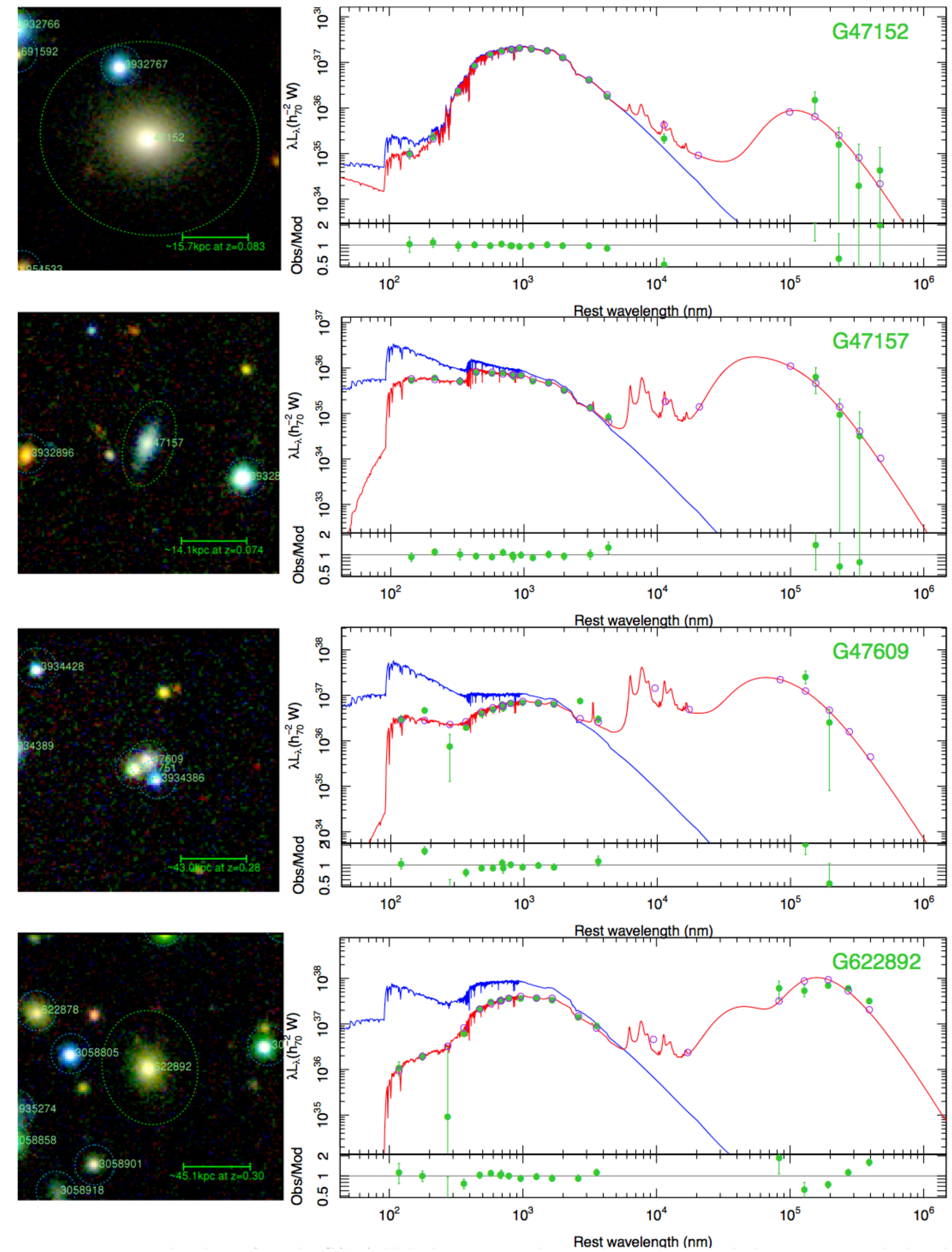




Why is it difficult?

- Requires energy-balance or radiative transfer treatment.
- Non-linear - current approaches use parametric SFHs only.
- Takes time.
- More models to depend on!
- Ideal: Bayesian approach with an MCMC, an adaptive-grid in lookback time and a way to marginalise over issues of parametrisation. If you have ideas - see me!





GAMA panchromatic data release is here: 221K galaxies, 21-band aperture matched, deblended photometry + spectra.

If not now, when?