Sorting galaxy histories

JD Cohn, UC Berkeley

Have a wealth of galaxy data

- stellar mass functions, many z
- luminosity functions, many z
- colors/sizes/shapes
- spectra
- clustering (mostly low z)
- gas measurements (CGM, etc)
- etc…

And much more is coming!!

From these-what sorts of things can we learn about how galaxies form?

Understanding galaxy formation

- galaxy formation= how galaxy properties evolve
- large number of physical processes/scales
- very complex

Understanding galaxy formation

- galaxy formation= how galaxy properties evolve
- large number of physical processes/scales
- very complex

Approaches:

Could try to simulate every measurement in detail

try to match all the observations in detail

Understanding galaxy formation

- galaxy formation= how galaxy properties evolve
- large number of physical processes/scales
- very complex

Approaches:

Could try to simulate every measurement in detail
try to match all the observations in detail

Many efforts in this direction right now
(see Nick B's slide!)

Understanding galaxy formation

- galaxy formation= how galaxy properties evolve
- large number of physical processes/scales
- very complex

Here--go minimal

observationally and in detailed simulations, emergent statistical trends are seen

- perhaps from "self-regulation"
- properties/physics conspire to give simple relations

e.g. $\Delta M^* \sim f(M_h,z)\, \Delta M_h$

look for simple descriptions of galaxy evolution

predict a few basic statistical properties over time, for a sample (e.g. M*,SFR)

??*Why* would one characterize galaxy histories in simple ways??

- *identifying* emergent causes/effects
  - and comparing to theories for emergent causes/effects
- intercompare properties
  - e.g. halo mass to SFR history
- to relate galaxy properties to lss
  - for understanding
  - to make mock catalogues

look for simple descriptions of galaxy evolution

predict a few basic statistical properties over time, for a sample (e.g. M*,SFR)

- any successes suggest essential parts captured/ summarized
- any failures point to additional required physics

Context—simulations

- Give whole (theoretical) history in detail
- case here: use dark matter halo histories
  - using as little as possible from these as well
  - Here Millennium simulation + L-galaxies model
  - (Springel++05, Lemson++06, Henriques++15)
  - tuned to observations mentioned earlier

look for simple descriptions of galaxy evolution

predict a few basic statistical properties over time, for a sample (e.g. M*,SFR)

look for simple descriptions of galaxy evolution

predict a few basic statistical properties over time, for a sample (e.g. M*,SFR)

Which properties??

Can be too simple:

"Galaxies start out small, form stars, then merge with other galaxies and/or stop forming stars. The End." ☺

look for simple descriptions of galaxy evolution

predict a few basic *statistical properties over time*, for a sample (e.g. M*,SFR)

(*statistical properties over time:* look at ensembles of galaxy histories)

look for simple descriptions of galaxy evolution

predict a few basic *statistical properties over time*, for a sample (e.g. M*,SFR)

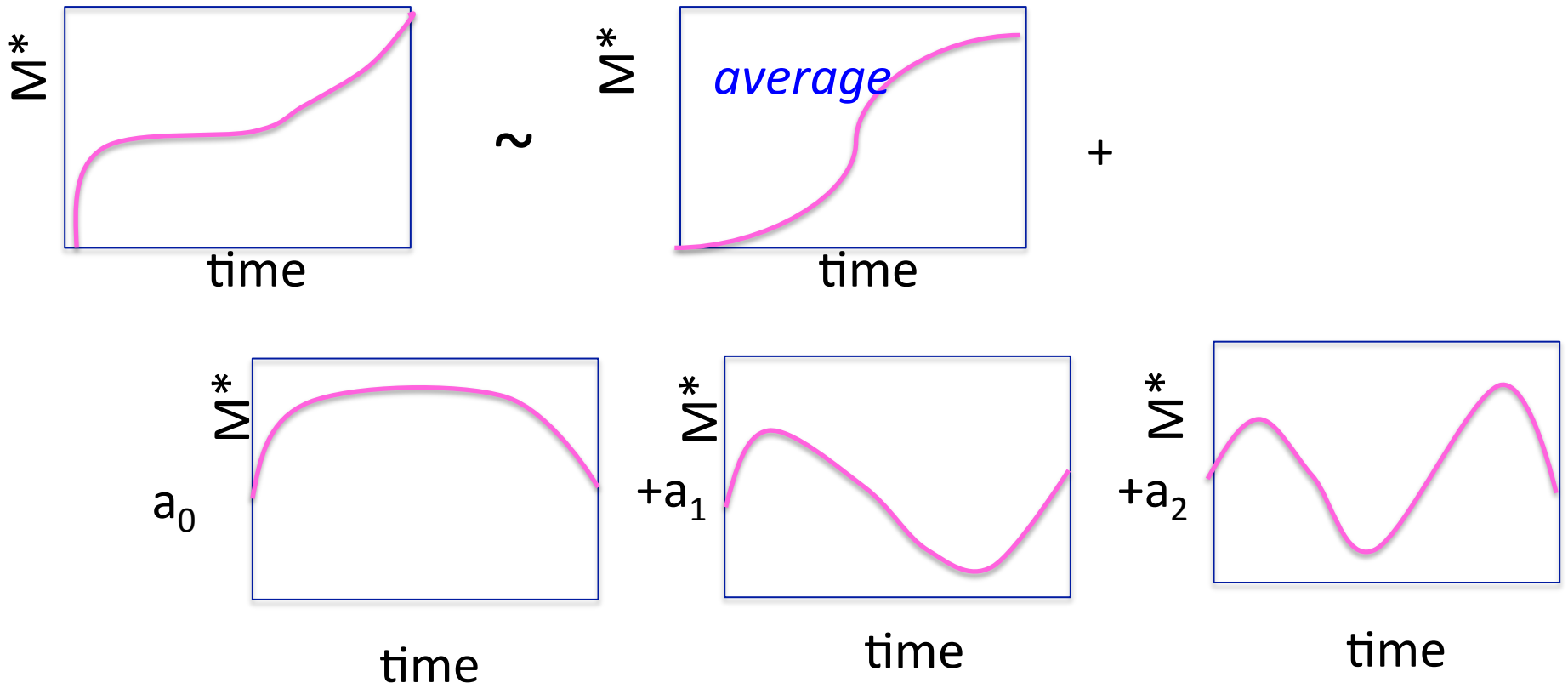
Which properties??

Previously: *Simple histories for M**

-- for each galaxy, write M*(t) as average plus ~3 fluctuations (PCA)

--coefficients a0,a1,a2 capture *most* of scatter

(fixed M* final, hydro and SAM examples, van de Voort &JDC)



Try PCA for SFR histories?

- PCA did not work well at all!
  - Needed *many* fluctuations to describe 90% scatter for histories (~might as well give terms for whole history)
  - see also Shamshiri++15

Try PCA for SFR histories?

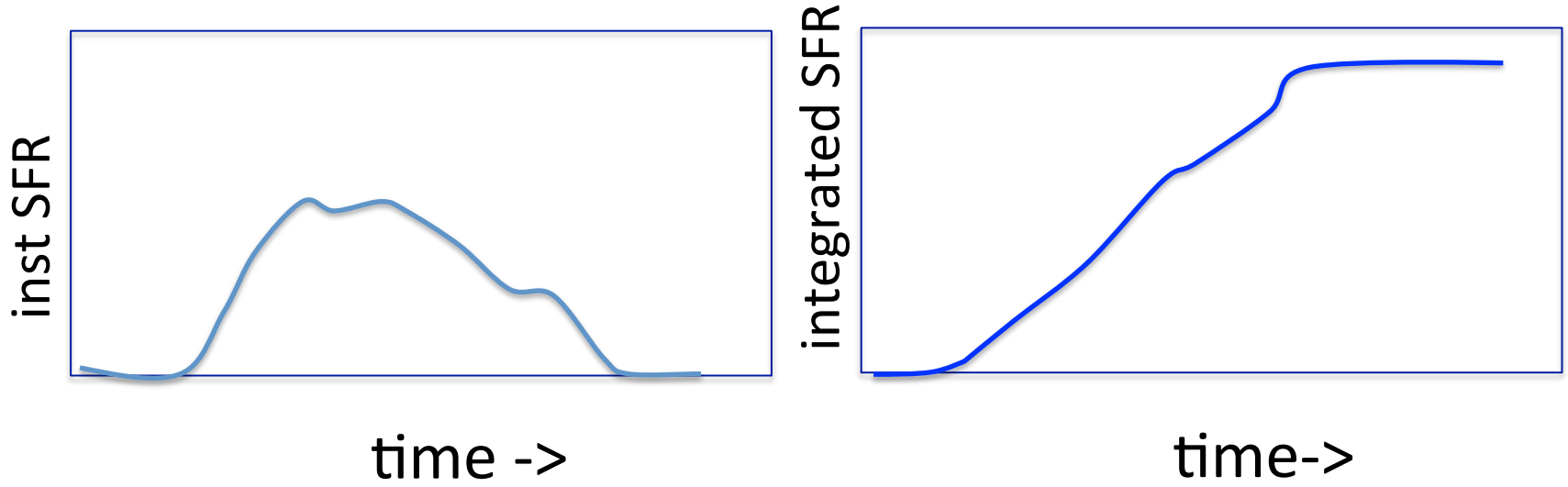
- PCA did not work well at all!
  - Needed *many* fluctuations to describe 90% scatter for histories (~might as well give terms for whole history)
  - see also Shamshiri++15
- New piece: Diemer++ ('17):
  - fit to **integral** of SFR
    - this is **choice** of what to focus on
    - small changes in sfr for high M* halo not as important

(Context: use lognormal fits to SFR, following Gladders++

$SFR = A/[\sqrt{2\pi}\, t\, \tau] \exp(- \ln(t)-T_o)^2 /2\tau^2$

3 parameters A, T_o, τ)

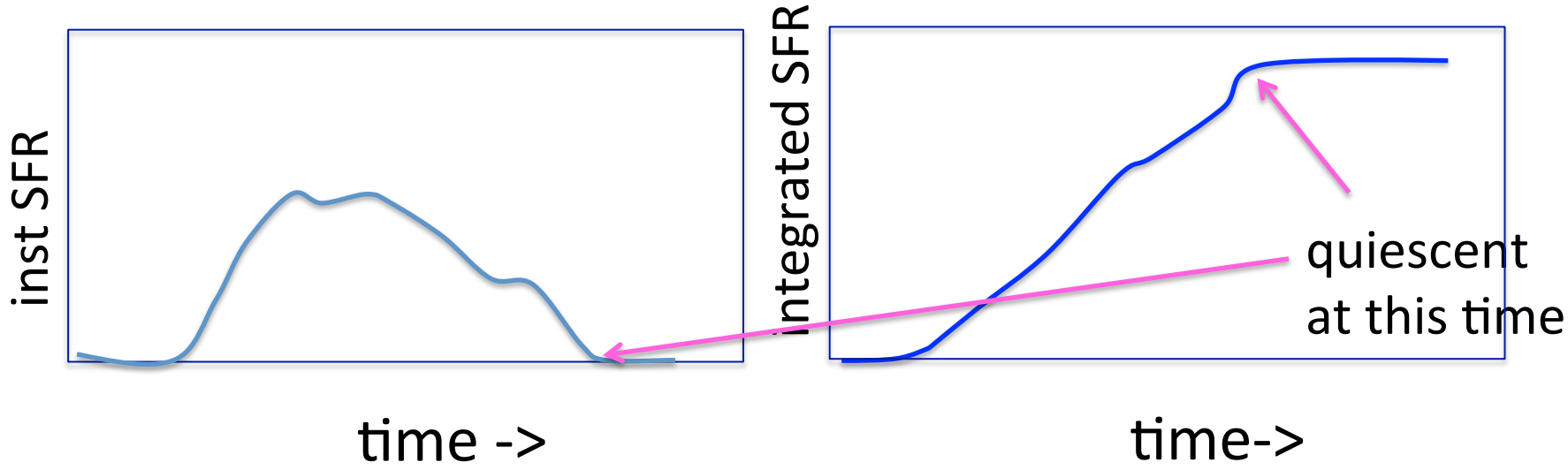
SFR Histories



Characterize these!

Beyond "integrated SFR goes up and then flat at some point…."☺

SFR Histories

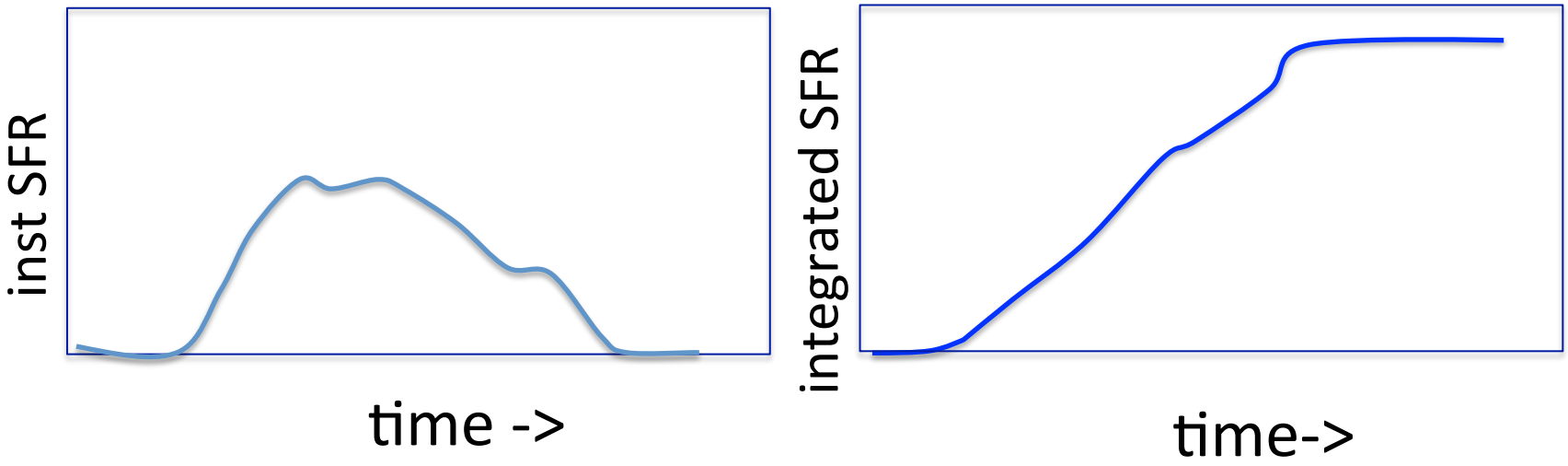


Characterize these!

Beyond "integrated SFR goes up and then flat at some point…."☺

SFR Histories

Do PCA on *integral* of all SFR histories in sample

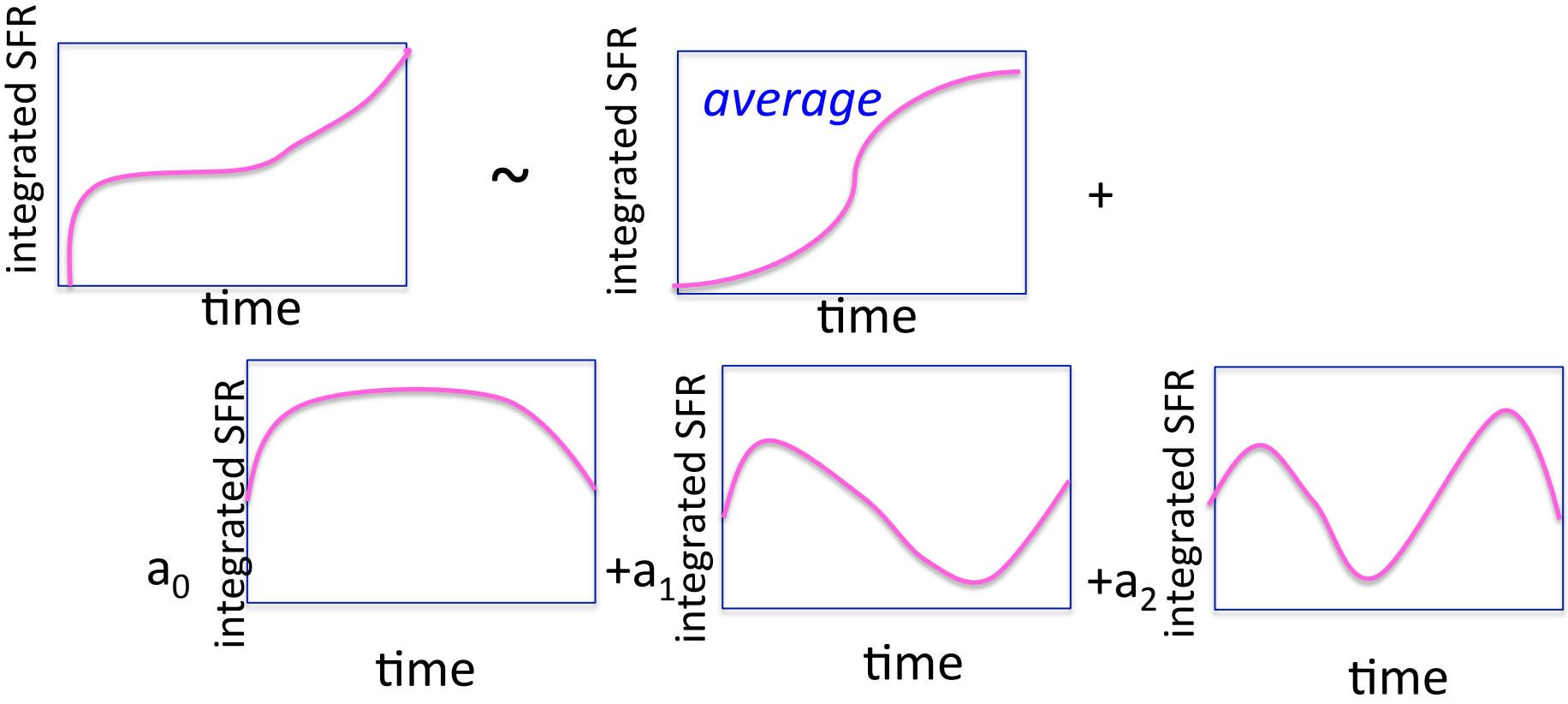


(**Rescale all histories to have same integrated SFR at final time)

Find: most of scatter in first 3 fluctuations (good!)

But: scatter is *large*!

i.e., for integrated SFR



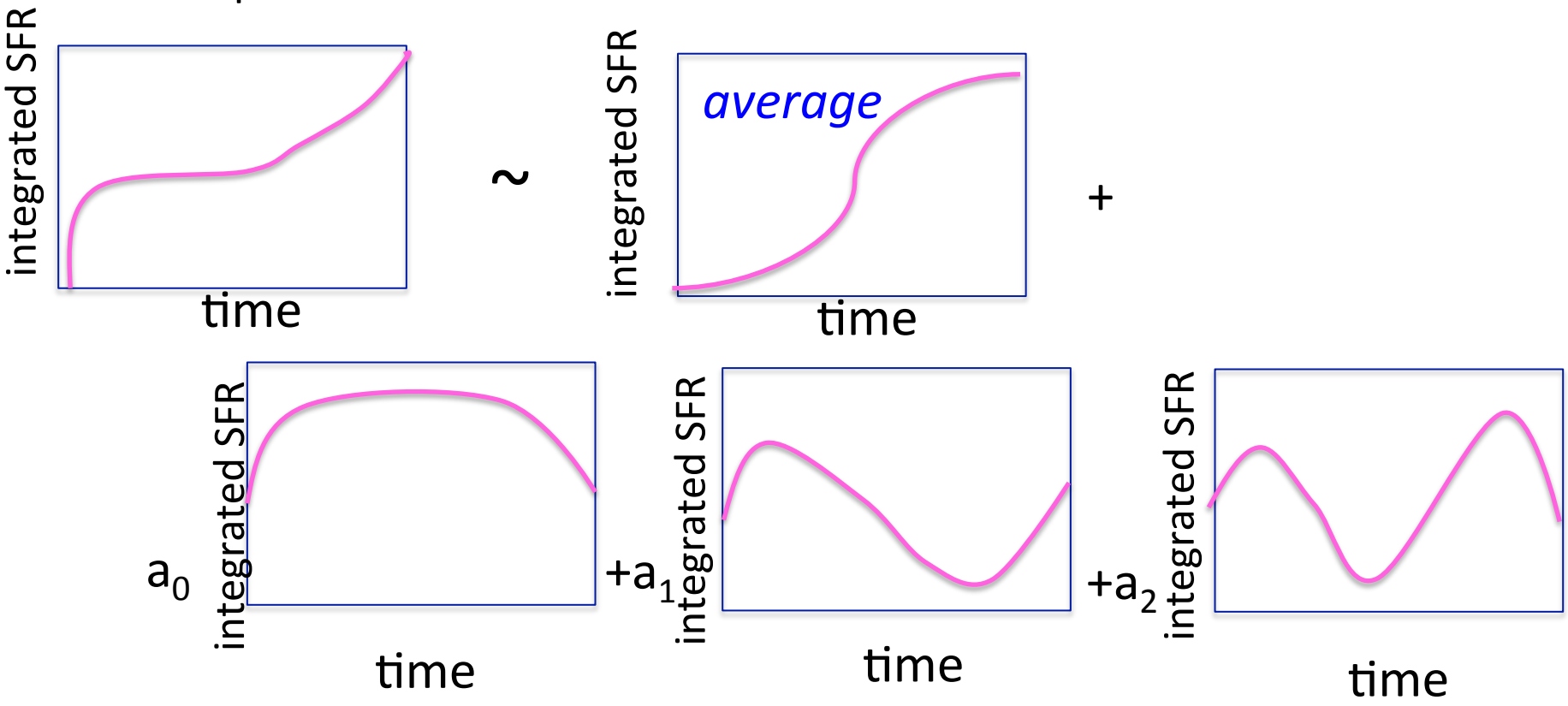
a_0, a_1, a_2 give > 95% of scatter
but scatters are big!

Split apart sample

- For stellar mass histories, used galaxies of similar final M*, not whole sample
- hint from Pacifici++16
  - stacked SFR histories of quenched galaxies of same M^*_{final}
  - dominated by galaxies which quenched most recently
- Try:
  - *stack by sfr peak (from Diemer++17 fit)*
  - highly correlated with PC_0 coefficient
- This works much better!
  - *again* scatter dominated by leading few fluctuations
  - *and* total fluctuations around average history *much smaller*

i.e., for integrated SFR,

fixed t_{peak} in range, n +/- 0.5 , n = 1,2,.....



a_0, a_1, a_2 give >~ 90% of scatter in subsample

scatters relatively “small”

SFR histories

Many caveats/fine print

- random samples
  - sum of 20 equal log mass bins (M_h or M*)
- no starbursts in Millennium SAM outputs, so some histories missing SFR contributions ☹, also some SFR goes to ICL
- overall rescalings are used when doing stacking

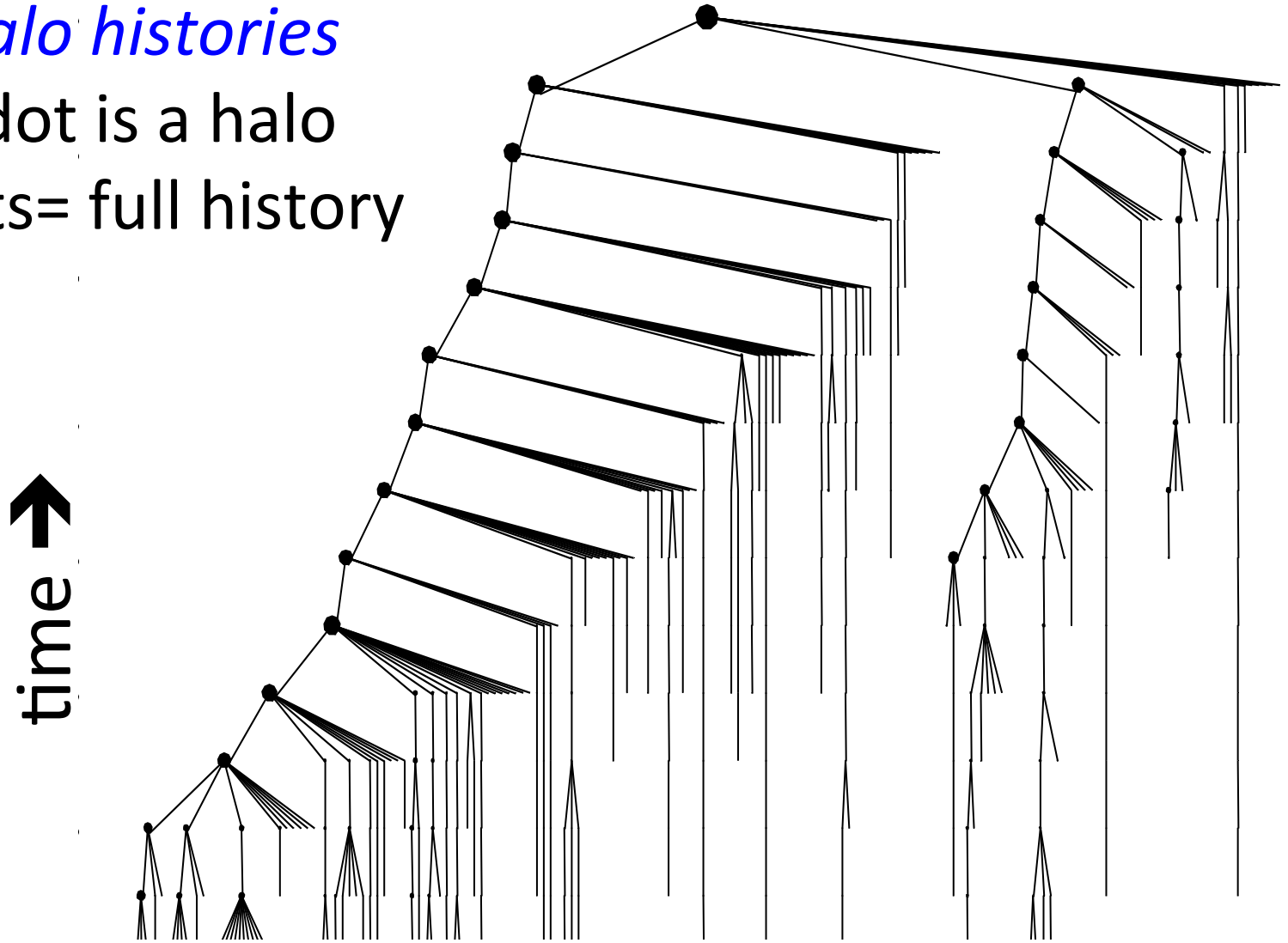
- SFR histories can be different things to different people
  - *Which* SFR histories?

Which SFR histories?

DM halo histories

Each dot is a halo

all dots= full history



M. White

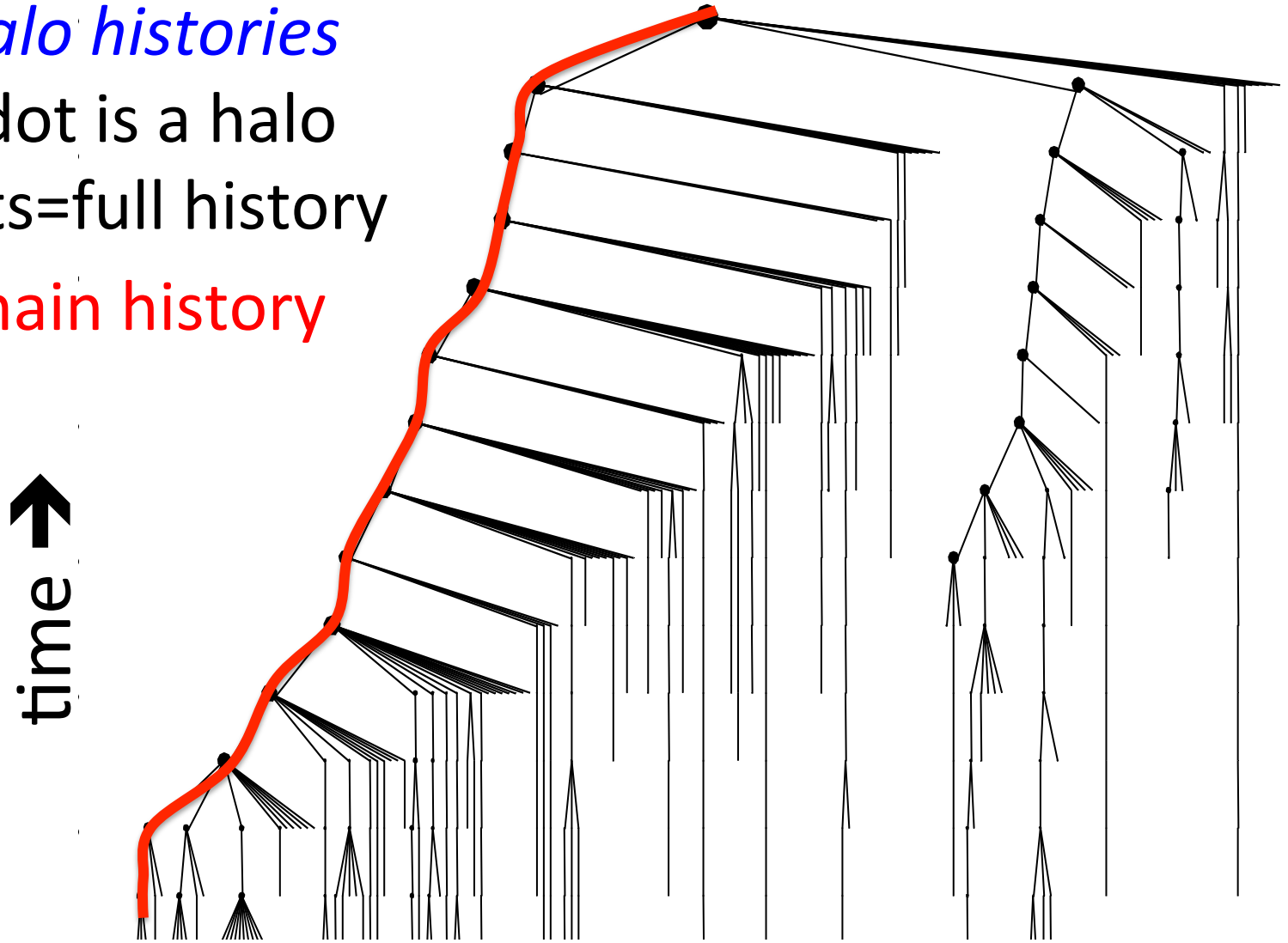
Which SFR histories?

DM halo histories

Each dot is a halo

all dots=full history

— main history



M. White

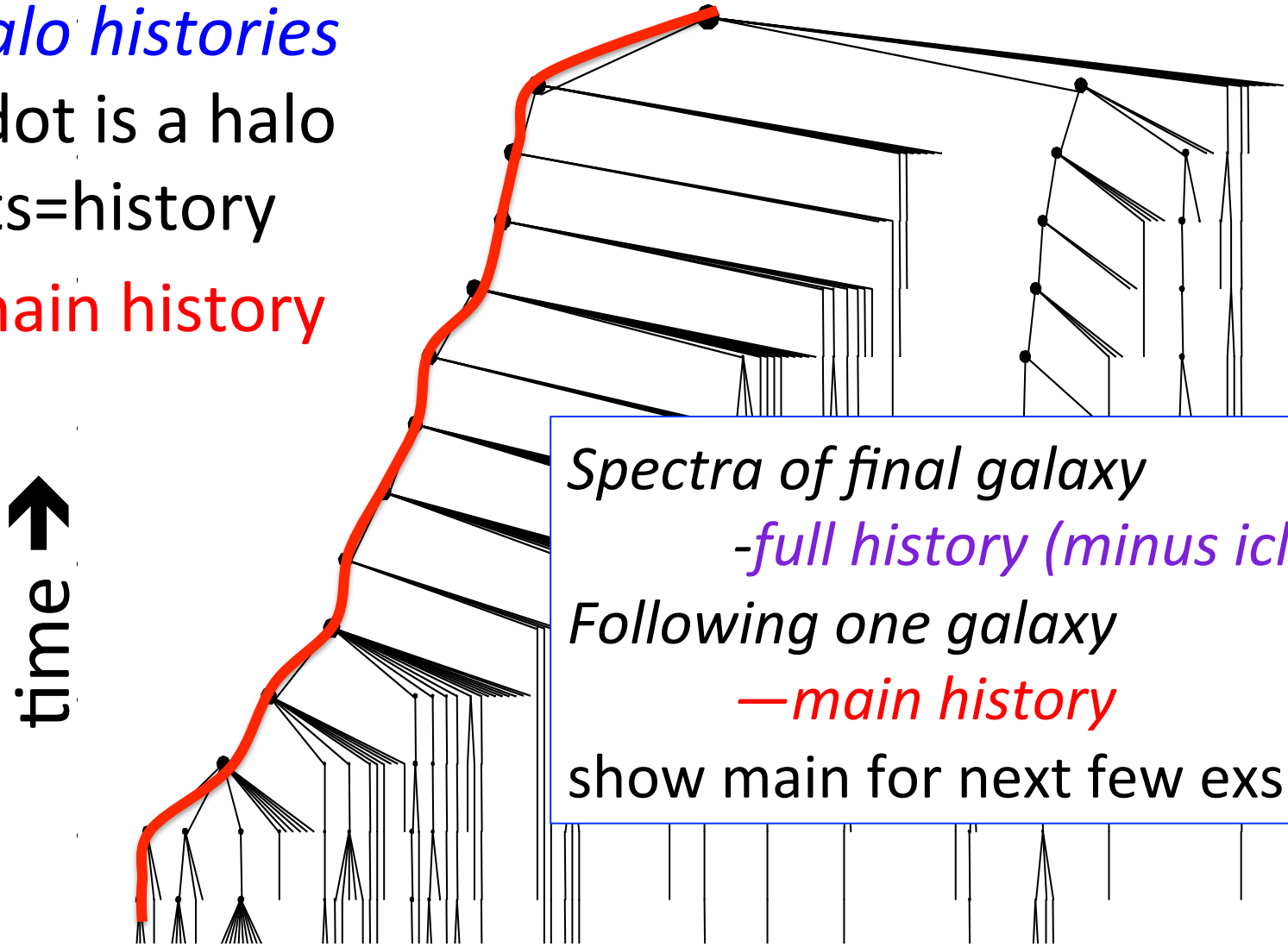
Which SFR histories?

DM halo histories

Each dot is a halo

all dots=history

—— main history



Spectra of final galaxy

-full history (minus icl)

Following one galaxy

—main history

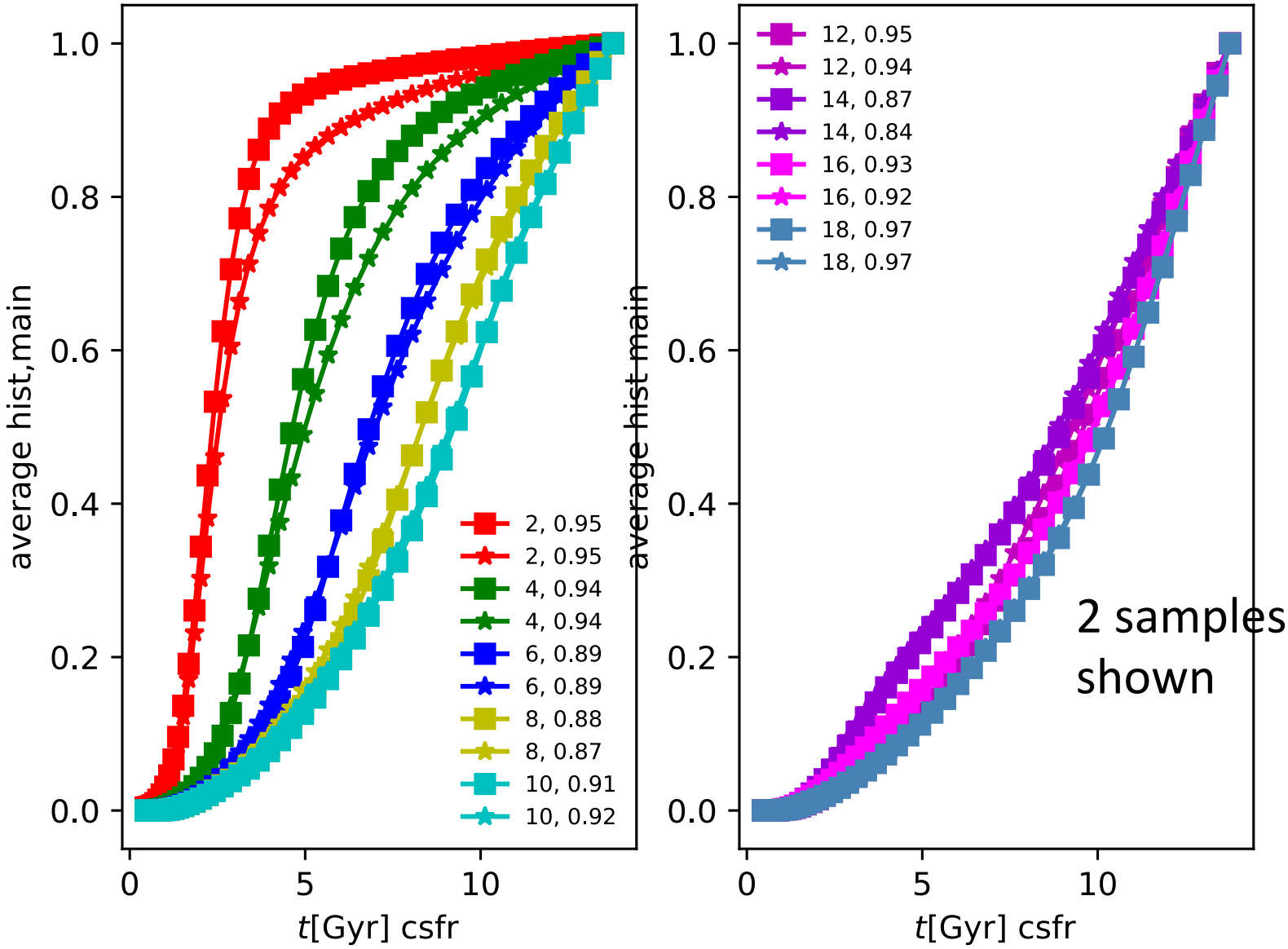
show main for next few exs

*main SFR often peaks slightly later than full, +/-1

M. White

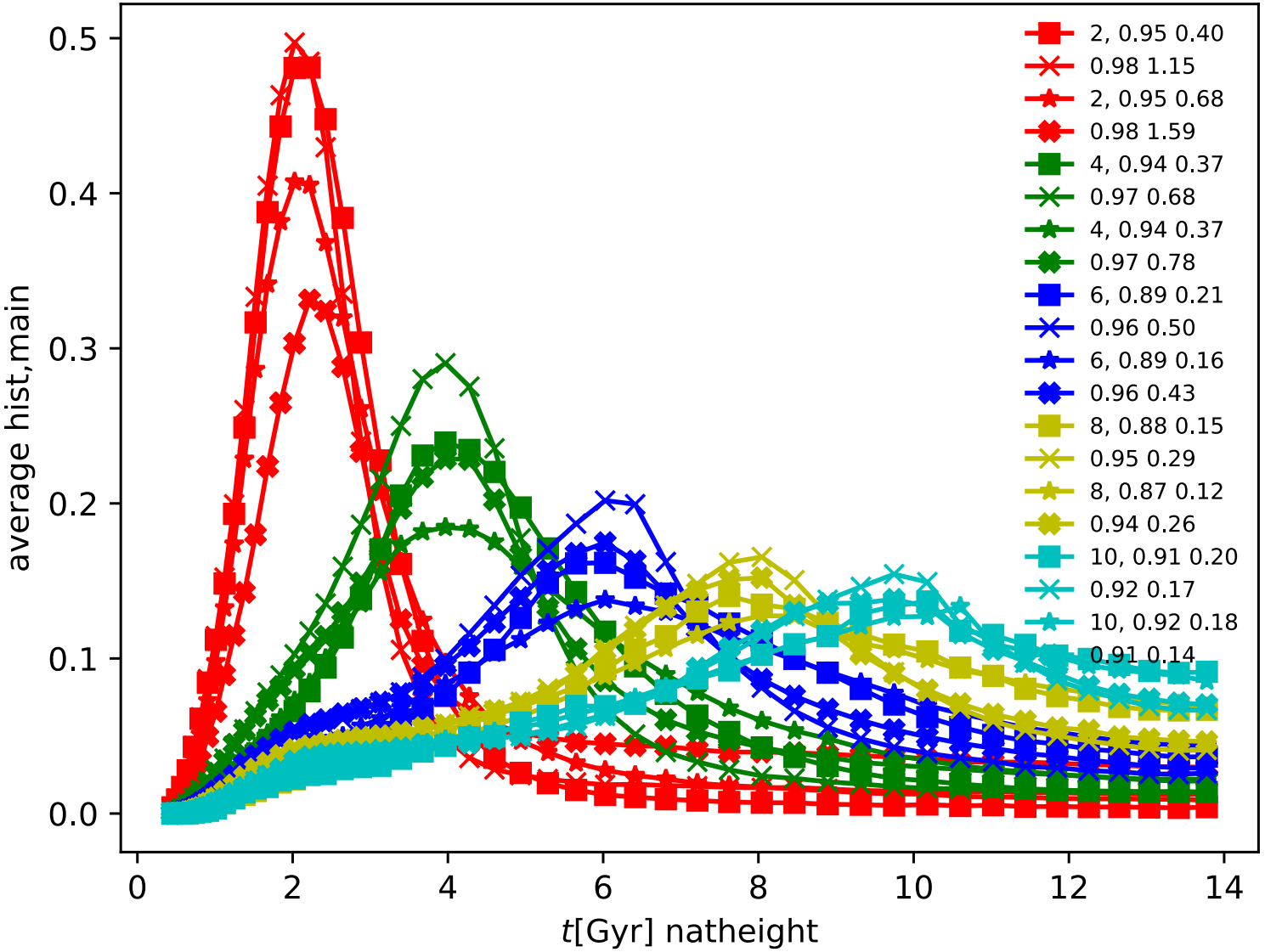
Average main histories for different t_{peak}, 2 samples, plus fraction of scatter from first 3 perturbations*

*norm by each gal's int of sfr



average instantaneous main histories

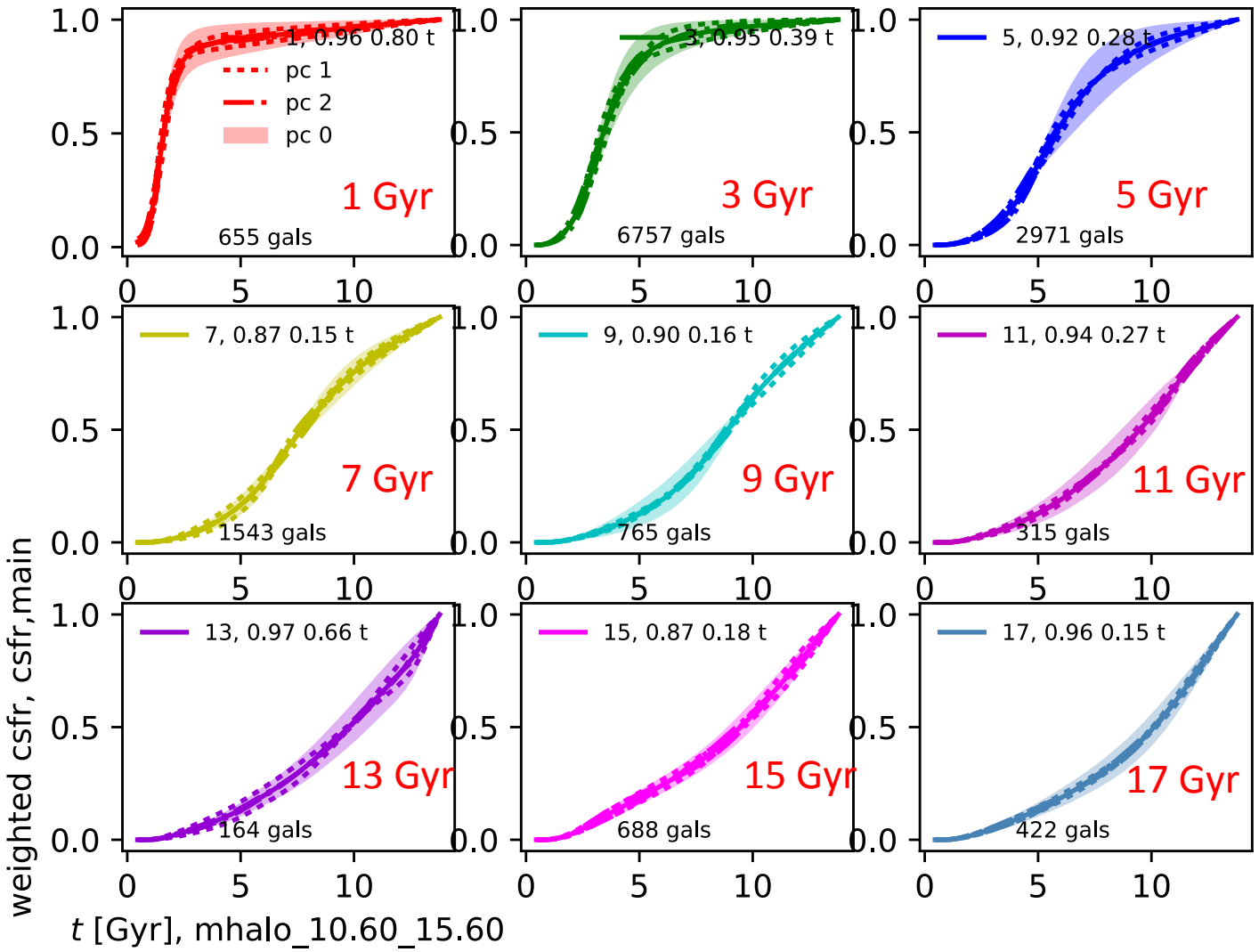
different samples and fits ← → different widths



PCA results, stacked on t_{peak} as shown

average + first 3 perturbations (*x* median coefficient)

**Scatter is not Gaussian!



Scatter around averages

- small
- controlled by a few major perturbations
  - 90% or more in those
  - maybe can treat rest as random scatter??
    - this would be nice for creating mock catalogues ☺

Can look at each stack in detail

which galaxies are sharing same t_{peak} ?

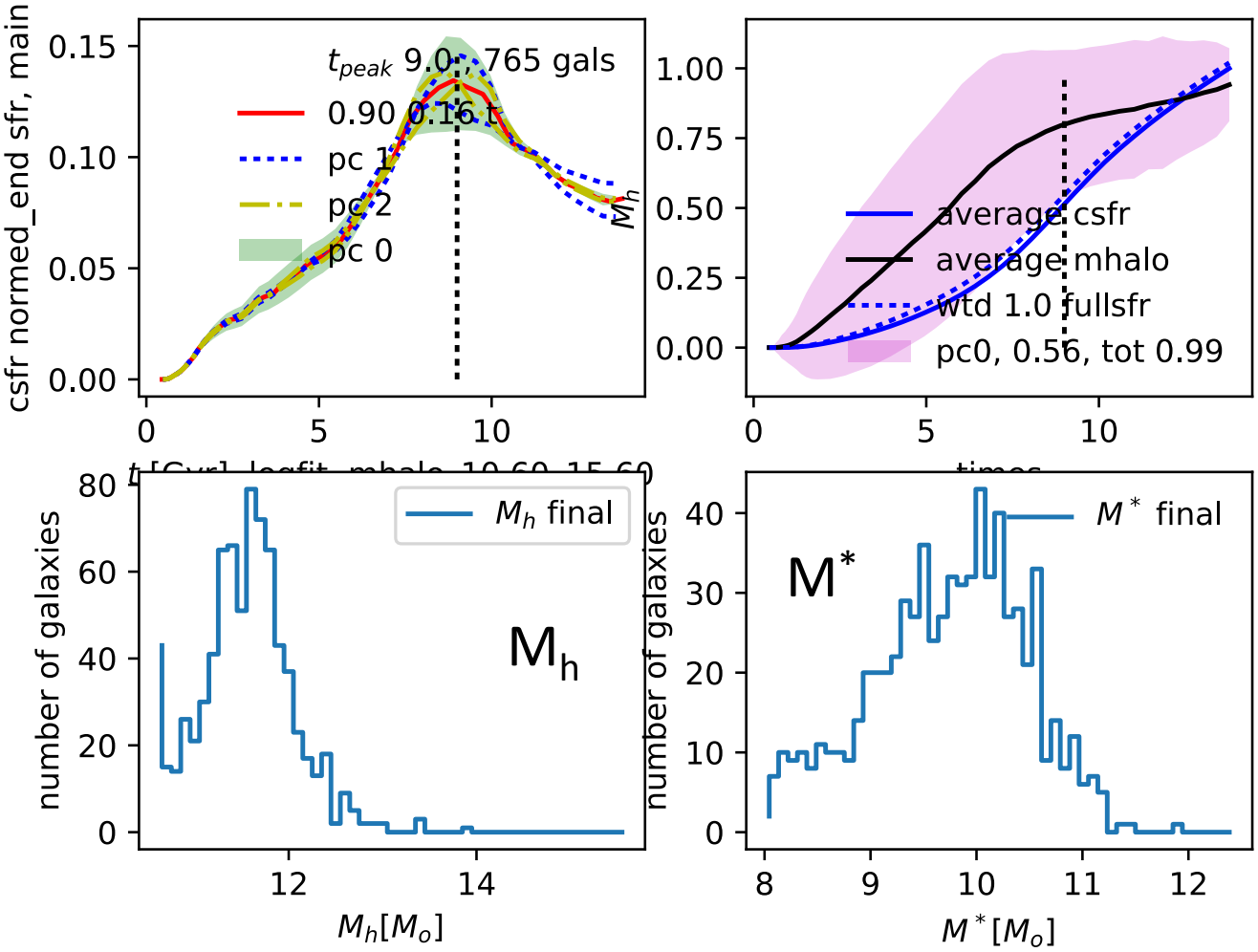
- following Diemer++ analysis for Illustris

  (Note: Bluck++16--quenching relative to SDSS

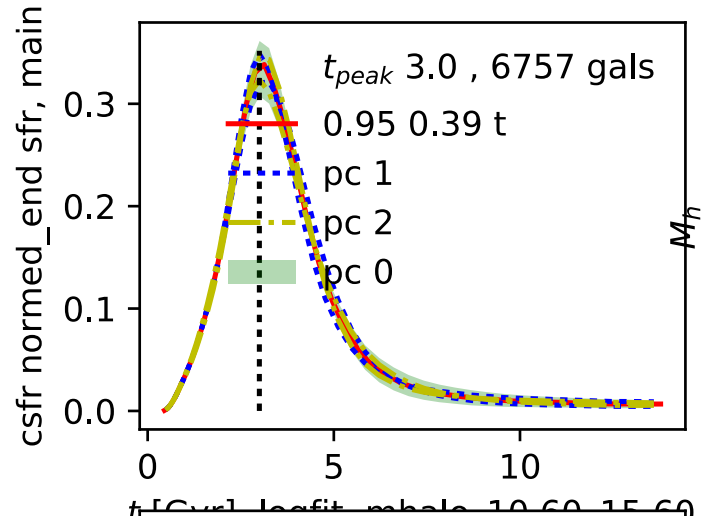
  Illustris not enough, MS too much)

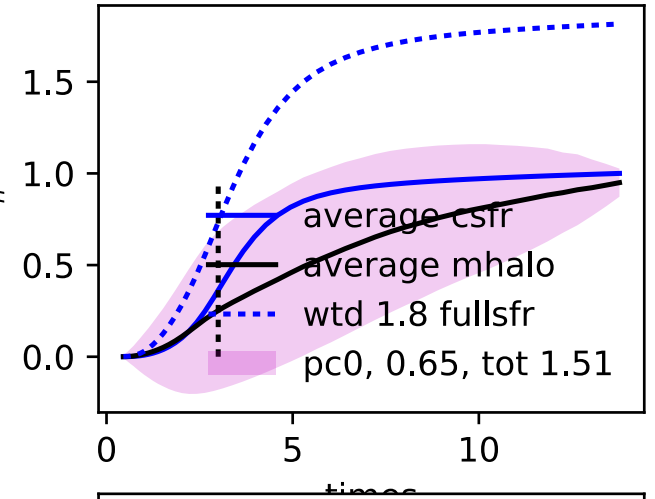
- consider stacks and
  - average SFR history (main & full),
  - integrated SFR history (main & full),
  - M* and M_h distributions

Example, peak at t=9 Gyrs

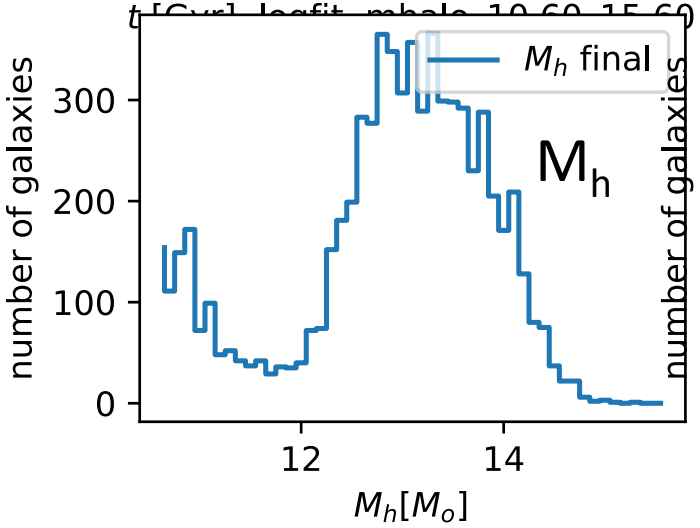


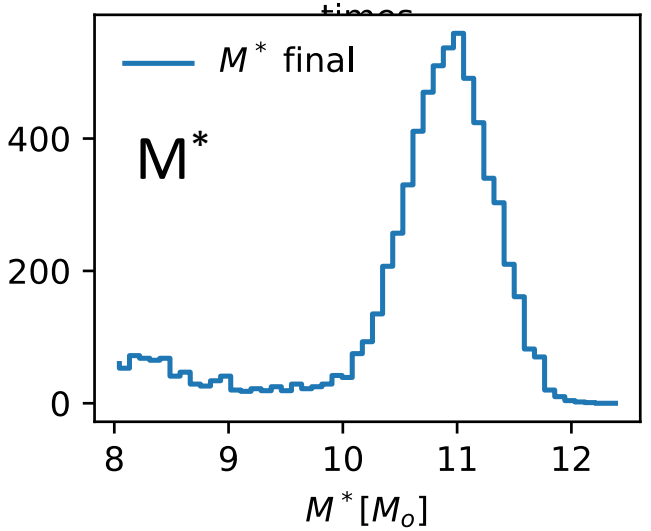
Example, peak at t=3 Gyrs





main and full very different



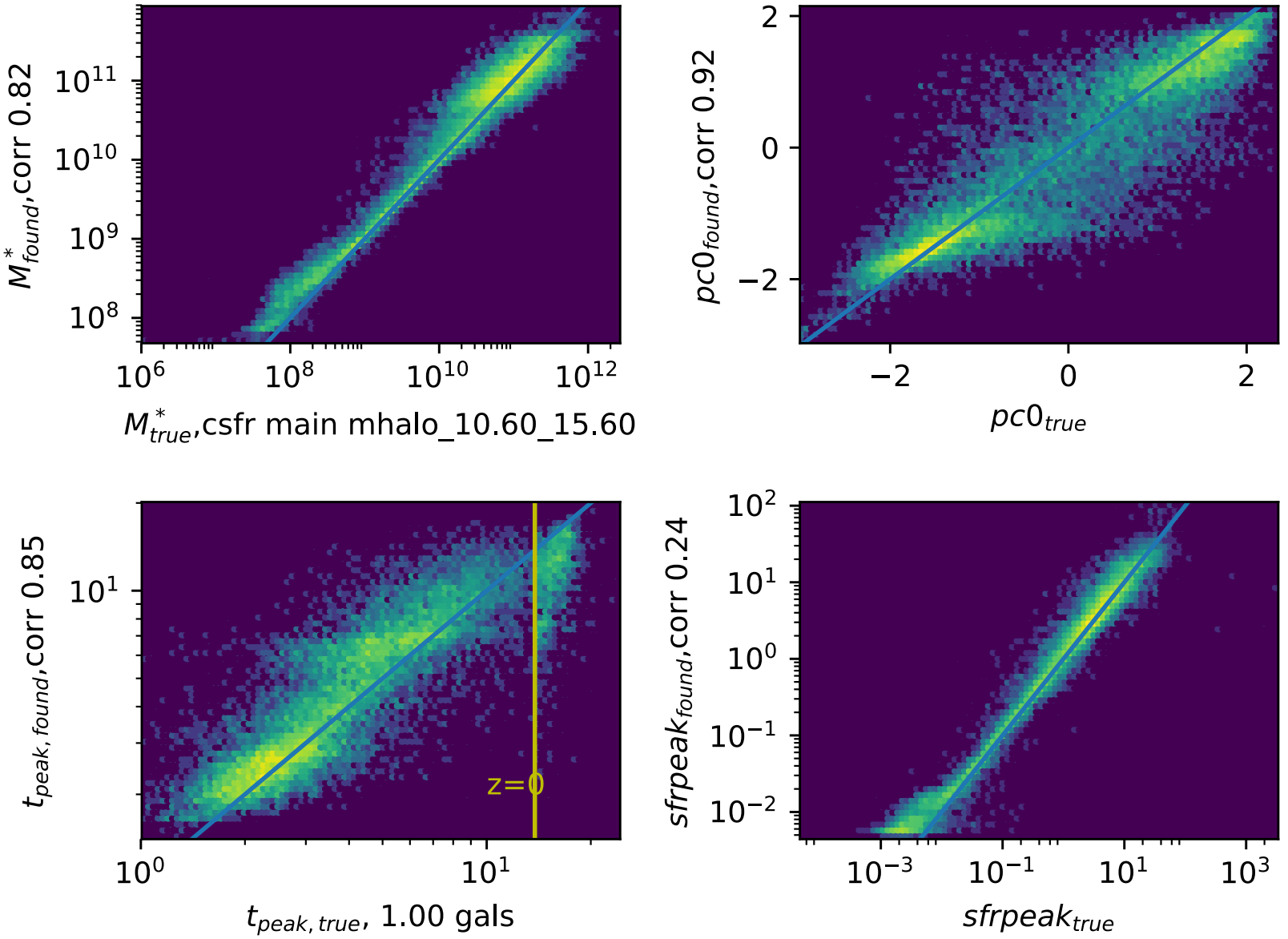


M* and M halo larger

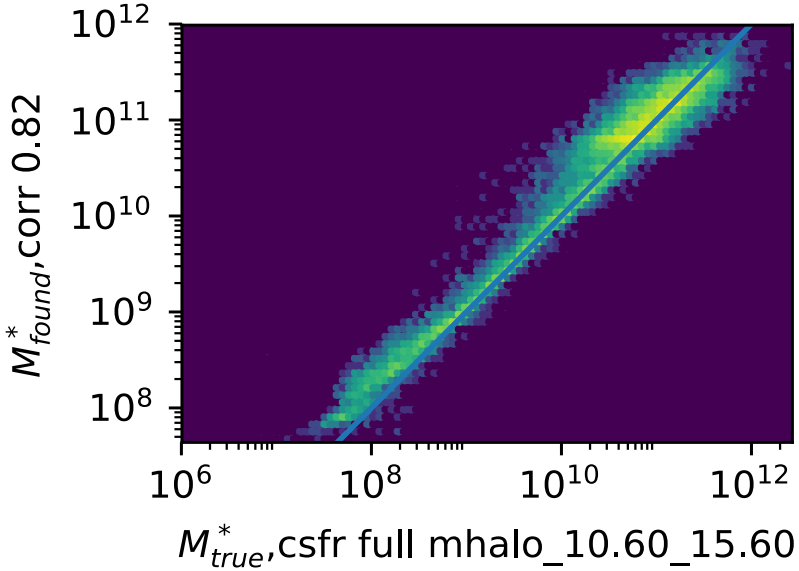
Tie these SFR histories to halo histories?

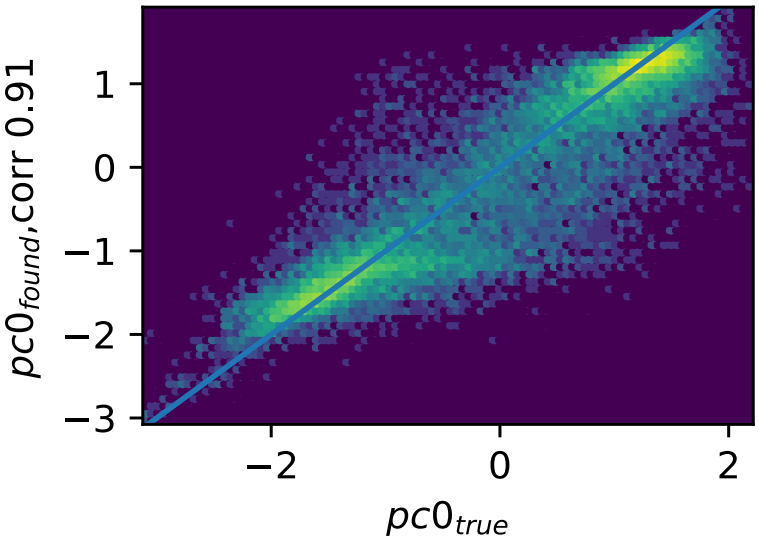
- First pass--throw into machine learning
- Follow Kamdar, Turk, Brunner
  - they got many galaxy properties just using dm histories or fixed time detailed dm properties
  - used only central galaxies, but all central galaxies (so low mass dominated)
  - got really cool results (and codes are on github)!

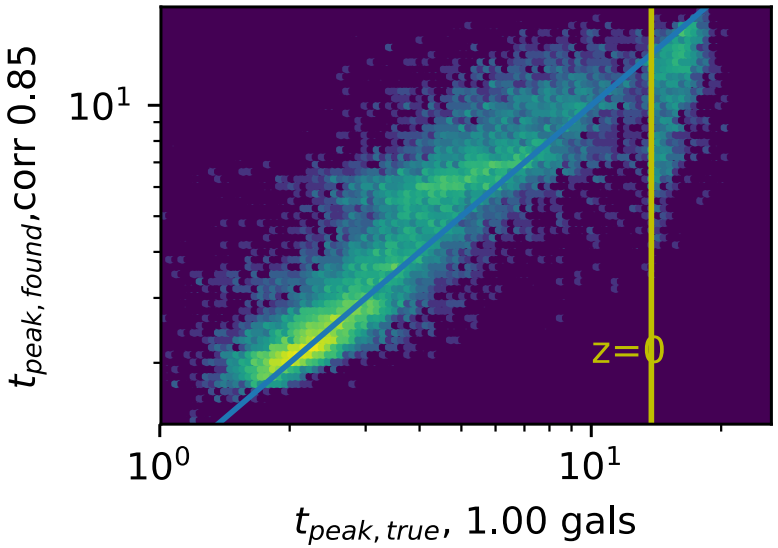
Correlations between true/found for main SFR's

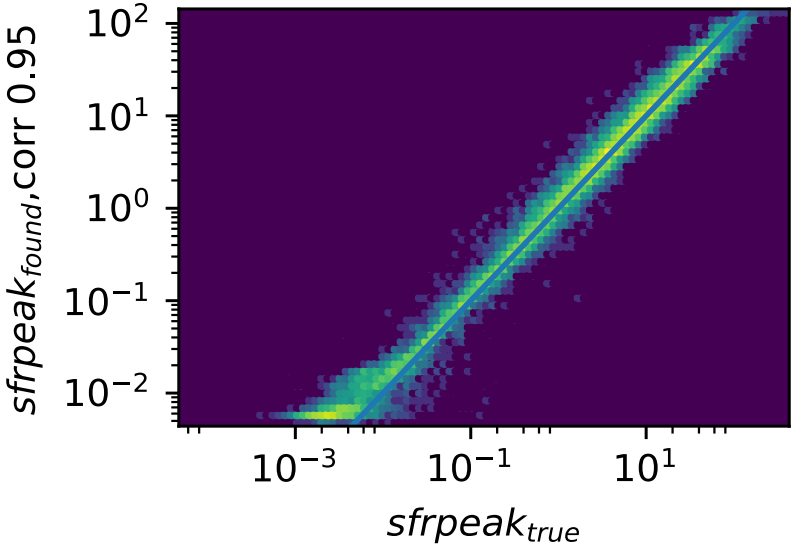


Correlations between true/found for full SFR's



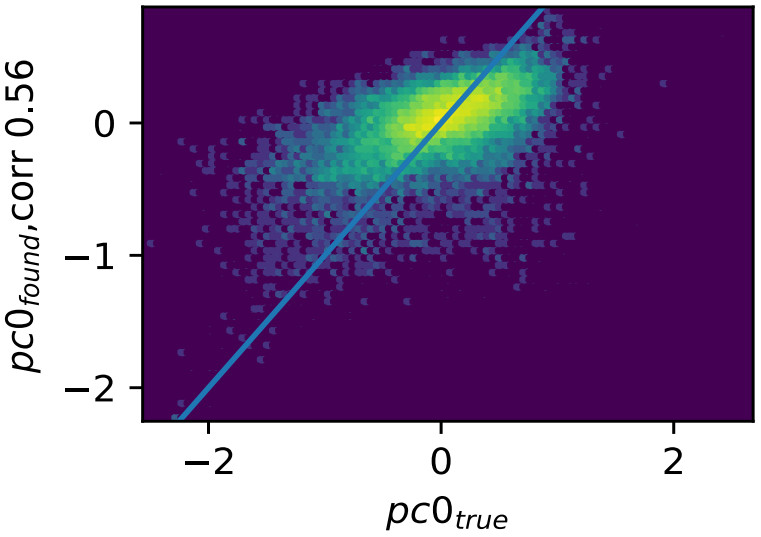




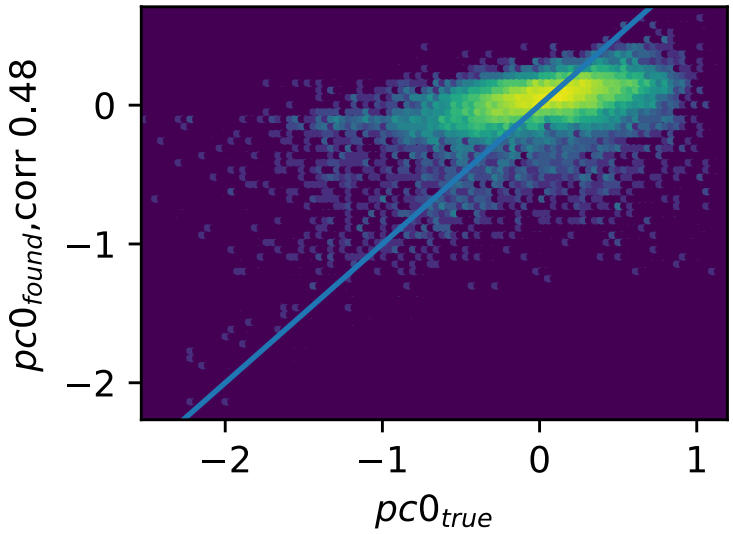


PC_0 was for stacking all galaxies--
much worse for recovering PC_0 for separate t_{peak} stacks!

main SFRs



full SFRs



This is largest fluctuation for fixed t_{peak} stacks
--seems to be related to width of lognormal fit?
--width not as closely tied to halo histories in first pass

summary

Look for simple parameterizations of galaxy formation histories

- use to discuss/analyze trends (causes/effects)

Follow Diemer++17 and use *integrated* SFR histories

Find:

Scatter around average int sfr history dominated by ~3 fluctuations

Grouping galaxies by t_{peak} from Diemer++17 lognormal fit

- lowers total scatter, ~3 fluctuations still give 90% of scatter*
- seems to give nice simplification of histories

*(did rescale all to same z=0 integrated sfr)

- machine learning can get some parameters well from halo histories

Need to investigate more:

- picked equal logM*, $logM_h$ samples, experiment with uses
- lognormal t_{peak} not 100% correl with flucts around full average →is another parameterized fit better?
- relations of Diemer++17 or relations to M* PCA from this angle

thank you