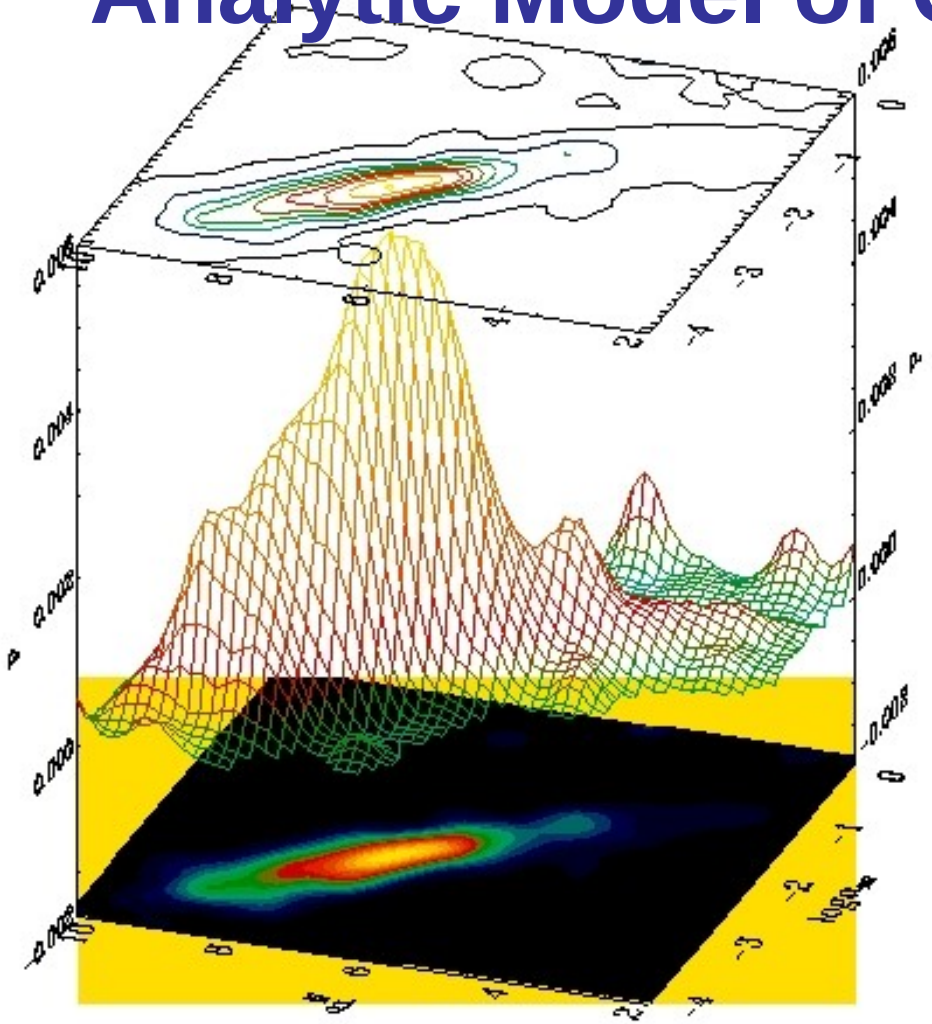
Bayesian Approach Based Semi-Analytic Model of Galaxy Formation



Yu Lu

collaborate with:
Houjun Mo
Martin Weinberg
Neal Katz

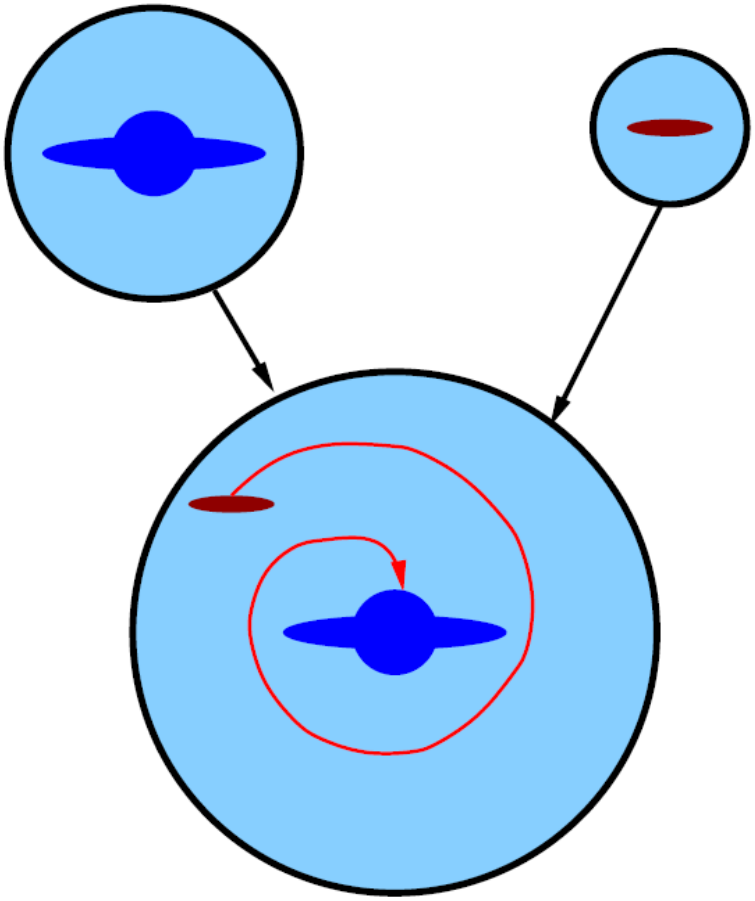
UMass, Amherst

Outline

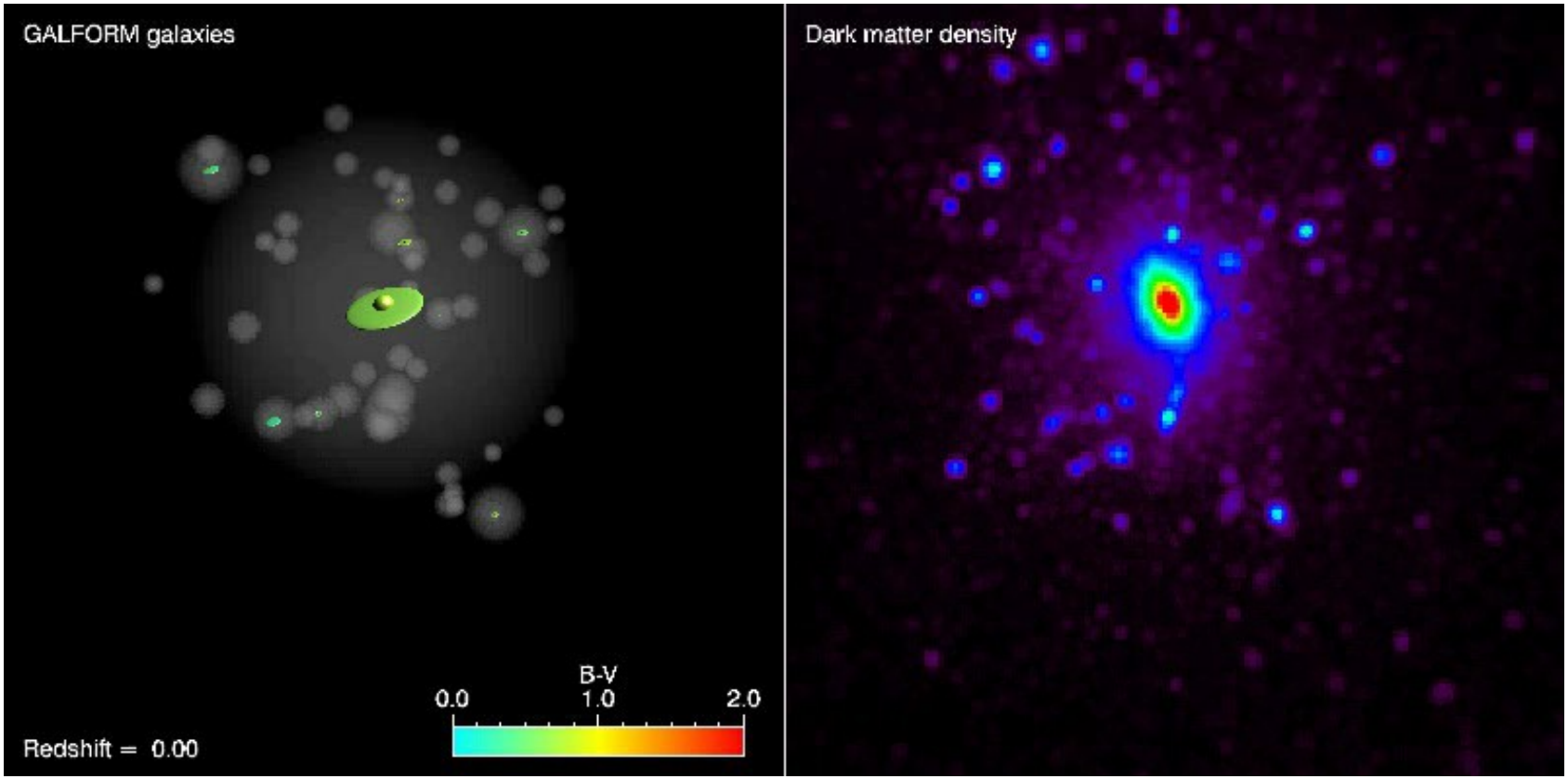
- Why do we want to build Semi-Analytic Model of galaxy formation on the Bayesian basis?
- How do we do it?
  - Bayesian model inference
  - Bayesian approach based SAM
- What can we learn from the Bayesian approach based SAM?
  - Stellar mass function
  - Condition stellar mass function
  - HI mass function

Semi-Analytic Model (SAM)

- Galaxy formation – a challenging problem
- Processes in SAM:
  1. Dark matter halos: distribution, growth
  2. Hot gas distribution
  3. Radiative cooling
  4. Star formation and supernova feedback
  5. Galaxy merger
  6. AGN, reionization, environmental effects…
- Phenomenological model, parameterizations, random realizations



SAM



Credit to Andrew Benson and Galform team

SAM

- Useful method
  - Predictive:
    - Luminosity function (stellar mass function), Tully-Fisher relation, Color-magnitude diagram
  - Low computation cost:
    - Model inference, hypothesis test
- Productive method (in terms of the number of citations)
- Problems
  - Tweaking parameters by hand
  - Fitting by eye
  - No way to rule out/in a model
  - No way to compare two models

SAM - as a problem of model inference

Given:

- model (hypothesis),
- a reasonable range of parameters (prior)
- data

We ask for:

- the probability distribution of the model parameters for explaining the data (confidence range of model parameters - posterior)
- the degree of belief that the model is supported by the data

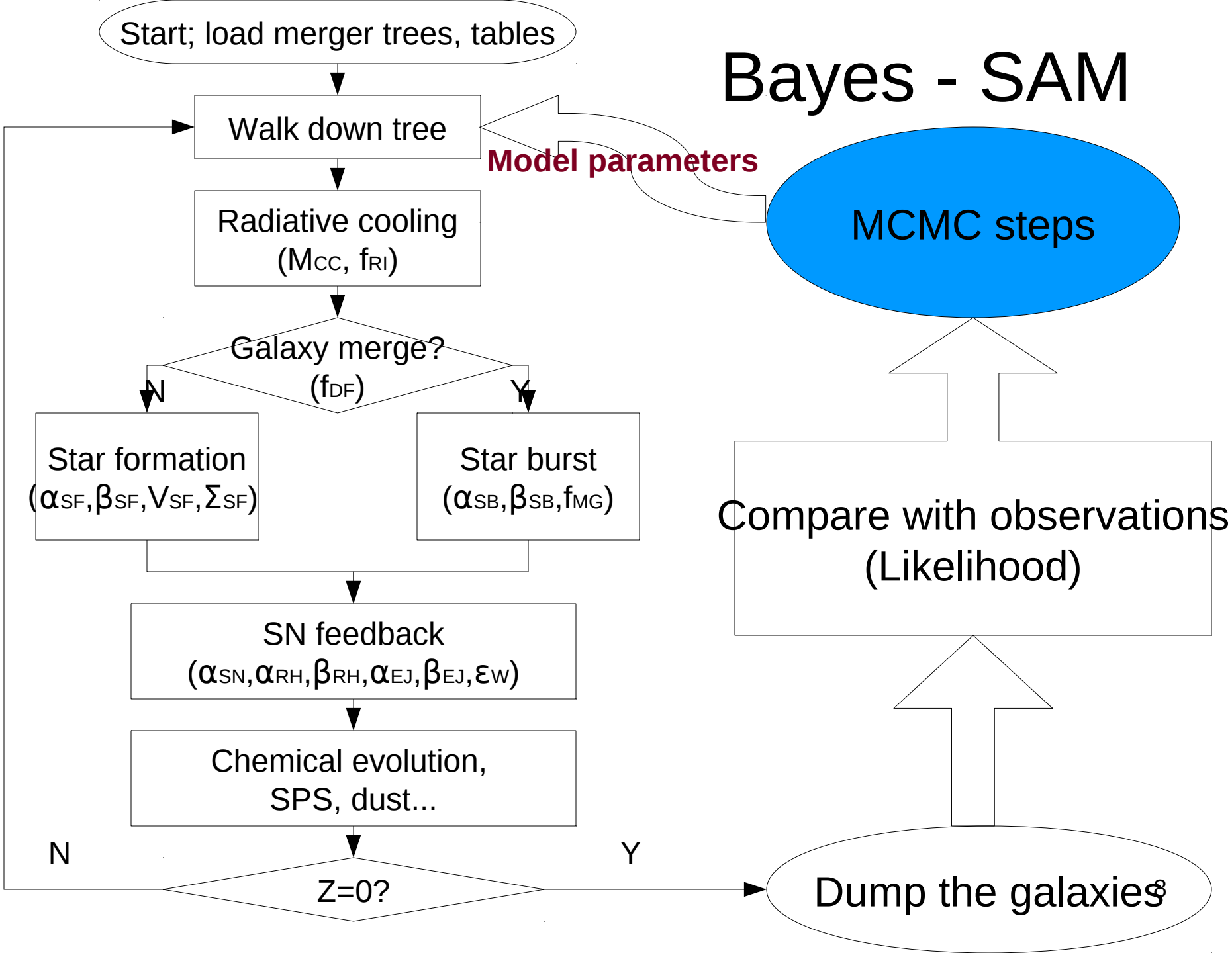
Bayesian approach provides a general framework for SAMing!

Bayesian model inference

- Bayes theorem: $$p(\boldsymbol{\theta} \mid D) = \frac{p(\boldsymbol{\theta}, D)}{p(D)} = \frac{p(\boldsymbol{\theta}) L(D \mid \boldsymbol{\theta})}{\int p(\boldsymbol{\theta}) L(D \mid \boldsymbol{\theta}) d\boldsymbol{\theta}} \propto p(\boldsymbol{\theta}) L(D \mid \boldsymbol{\theta})$$
- Marginalized posterior: $$p(\boldsymbol{\theta}_s \mid D) = \int p(\boldsymbol{\theta} \mid D) d\overline{\boldsymbol{\theta}_s}$$
- Bayesian evidence (Occam's razor): $$E(D \mid M) = \int p(\boldsymbol{\theta}) L(D \mid \boldsymbol{\theta}) d\boldsymbol{\theta}$$
- Prediction: $$p(K \mid M) = \int K(\boldsymbol{\theta} \mid M) p(\boldsymbol{\theta} \mid D) d\boldsymbol{\theta}$$

The goal is to put SAM on a probabilistic footing.

Bayes - SAM


Start; load merger trees, tables
Walk down tree
Model parameters
MCMC steps
Radiative cooling
(M_{CC}, f_{RI})
Galaxy merge?
(f_{DF})
N
Y
Star formation
(α_{SF},β_{SF},V_{SF},Σ_{SF})
Star burst
(α_{SB},β_{SB},f_{MG})
Compare with observations
(Likelihood)
SN feedback
(α_{SN},α_{RH},β_{RH},α_{EJ},β_{EJ},ε_W)
Chemical evolution,
SPS, dust...
N
Y
Z=0?
Dump the galaxies

Generalized parameterization

- More than 15 parameters
- AGN: $M_{\rm h} > M_{\rm CC}$
- Merger timescale:

$$\tau_{\rm merg} = f_{\rm DF}\tau_{\rm df}$$

- Star formation efficiency:

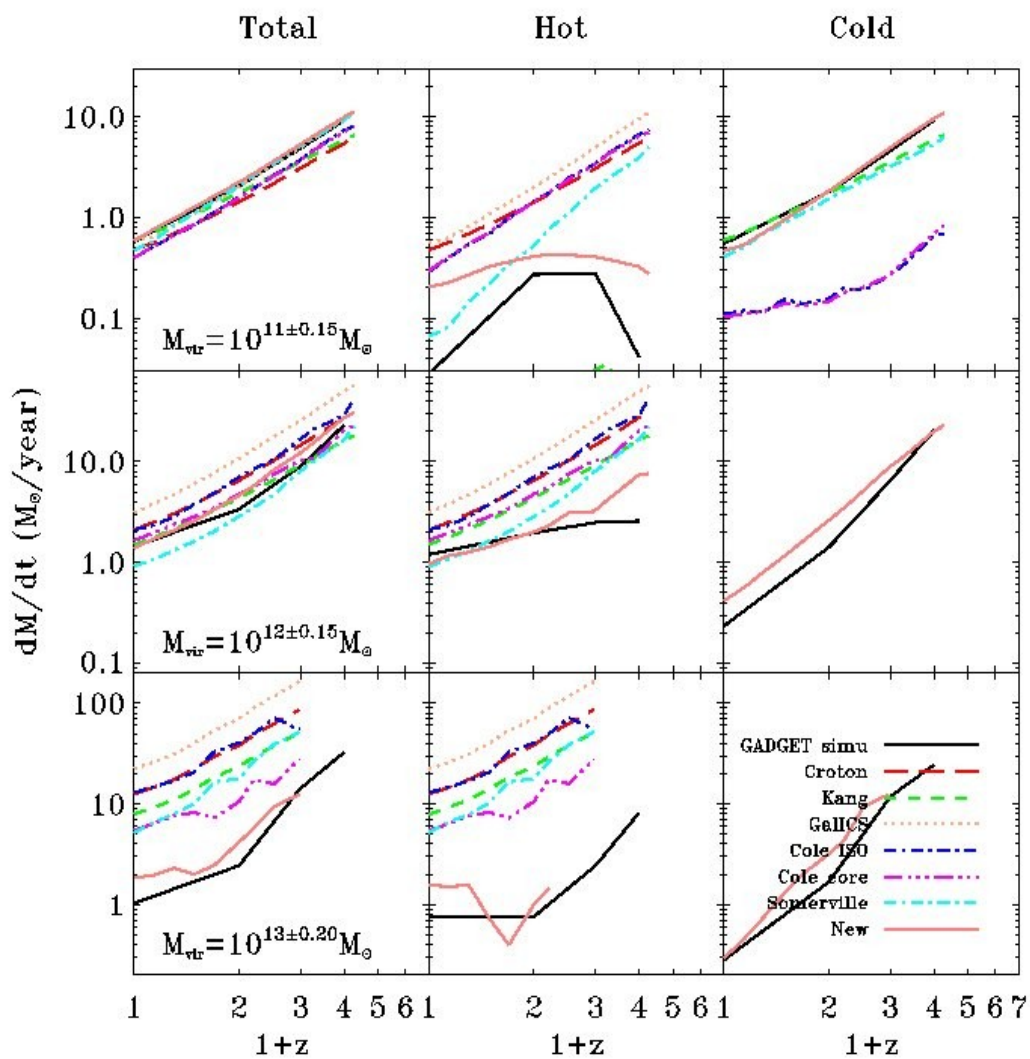
$$\dot{m}_* = \epsilon_* \frac{m_{\rm sf}}{\tau_{\rm disc}}$$

$$\epsilon_* = \begin{cases} \alpha_{\rm SF} & V_{\rm vir} \geq V_{\rm SF} \\ \alpha_{\rm SF}\left(\frac{V_{\rm vir}}{V_{\rm SF}}\right)^{\beta_{\rm SF}} & V_{\rm vir} < V_{\rm SF} \end{cases}$$

- SN reheating:

$$\dot{m}_{\rm rh} = f_{\rm rh}\dot{m}_*$$

$$f_{\rm rh} = \alpha_{RH}\left(\frac{V_0}{V_{\rm vir}}\right)^{\beta_{\rm RH}}$$



Lu etal in preparation

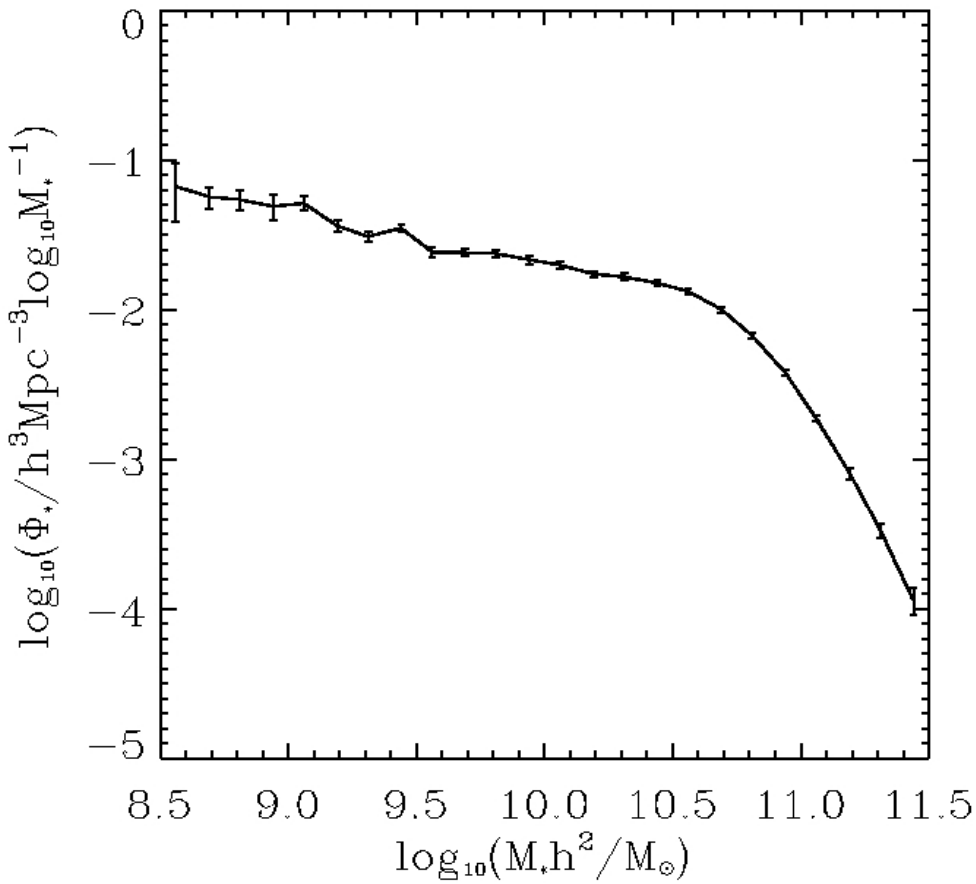
What can we do about SAM with Bayesian approach?

- Automatically search for good fit in a large high-D parameter space – spend computation in the places of good fit
- Directly sample the posterior probability distribution; study the internal working mechanism of SAM, degeneracy, multi-modal…; study how the given data constrain the model
- Test models
- Make predictions

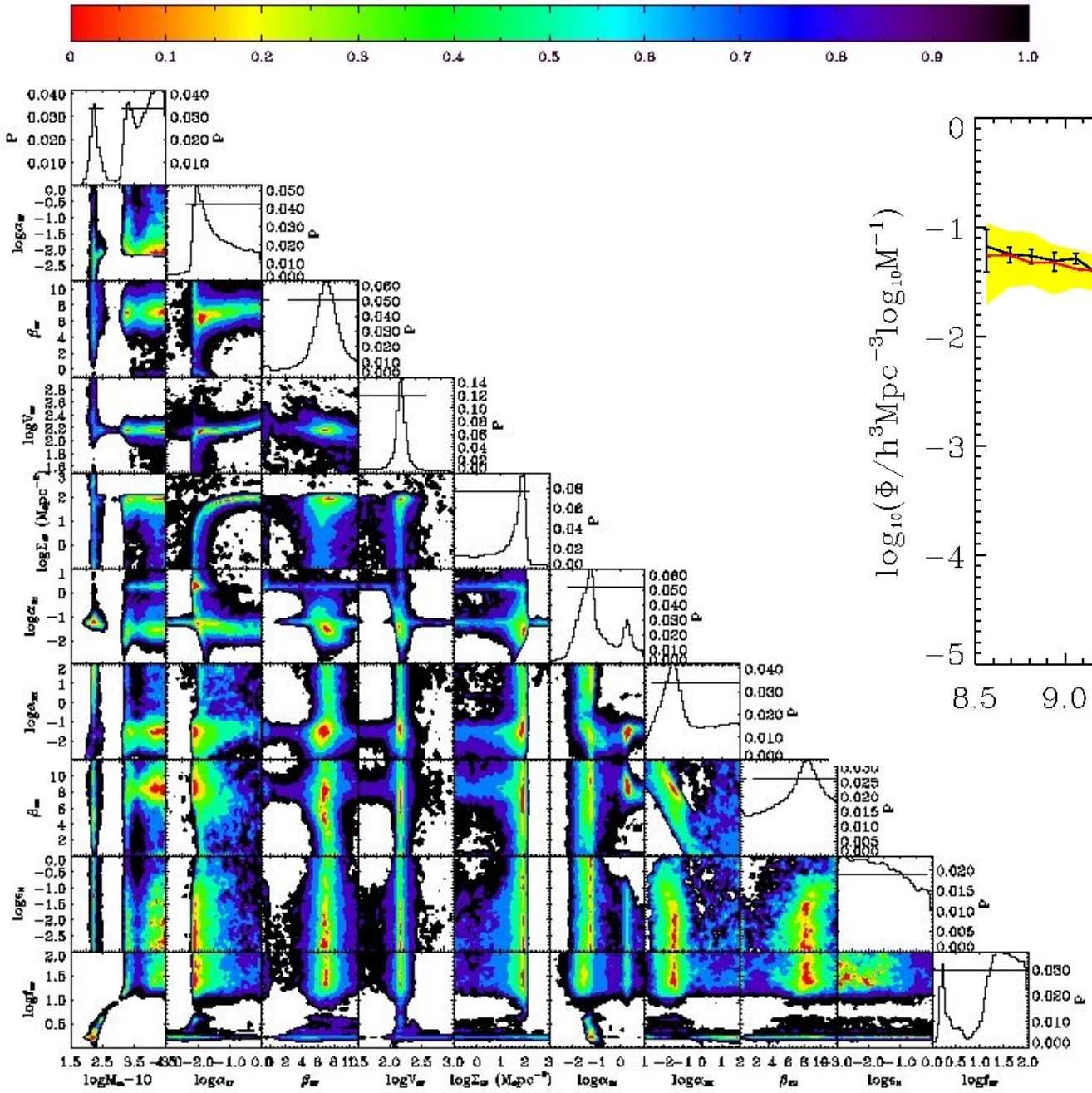
What can we learn from stellar mass function?

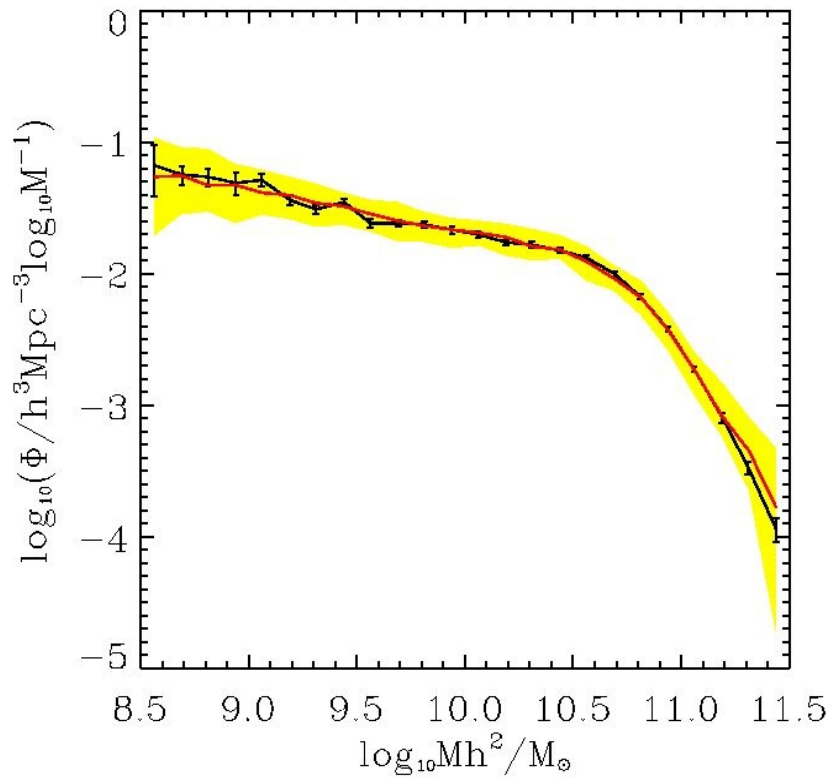
- Data: SDSS+2MASS (Bell etal 2003)
- Likelihood

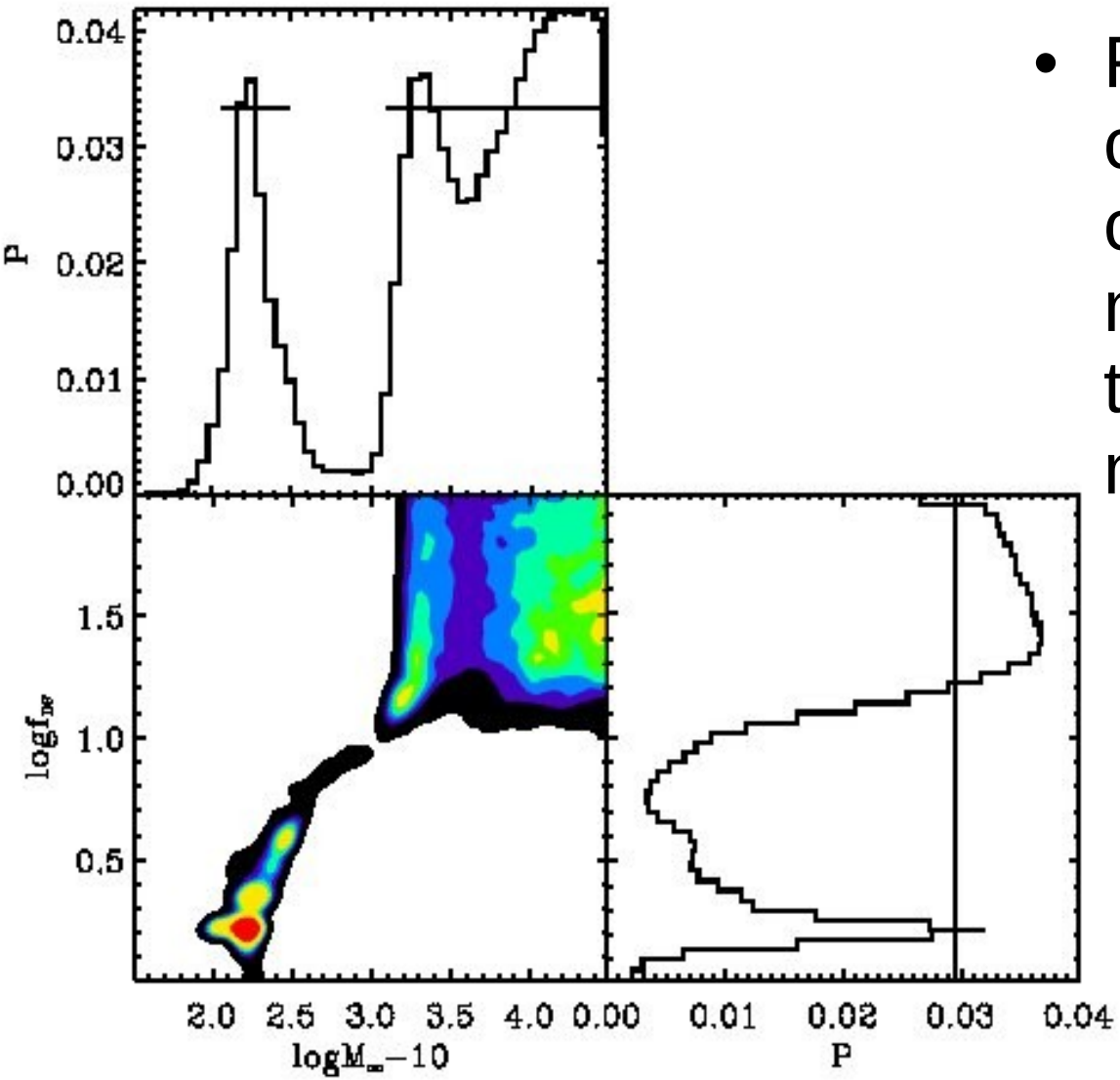
$$L = \exp[-\frac{1}{2}\sum_{i}^{n}\frac{(\phi_i - \phi_i^{'})^2}{\sigma_i^2 + \omega_i^2}]$$



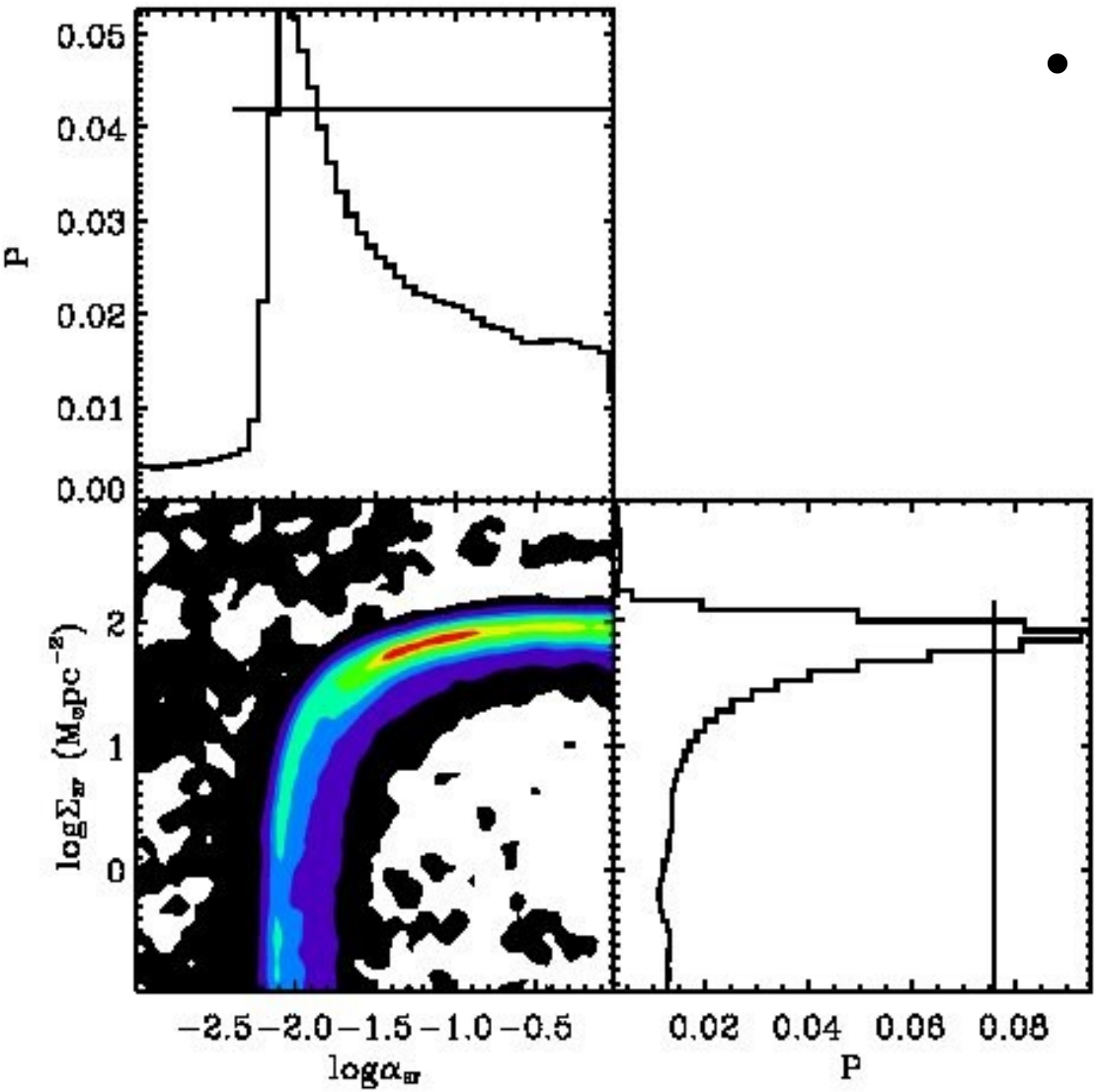
Bell etal 2003


0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
P
logM* −10
logα*
β*
logV*
logΣ* (M⊙pc⁻²)
logα*
logσ*
β*
logσ*
logf*

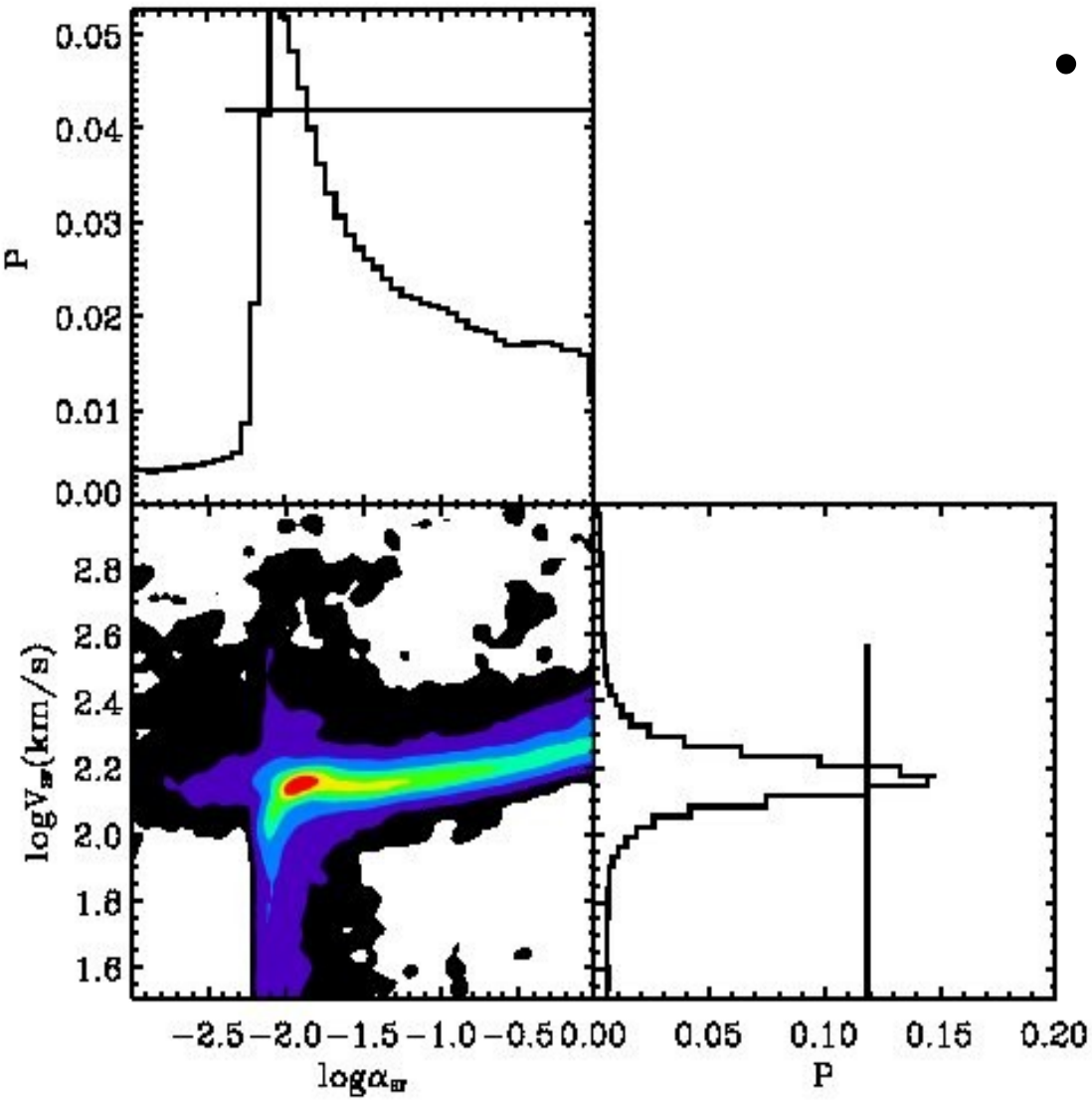

log10(Φ/h³Mpc⁻³log10M⁻¹)
log10Mh²/M⊙
0 −1 −2 −3 −4 −5
8.5 9.0 9.5 10.0 10.5 11.0 11.5



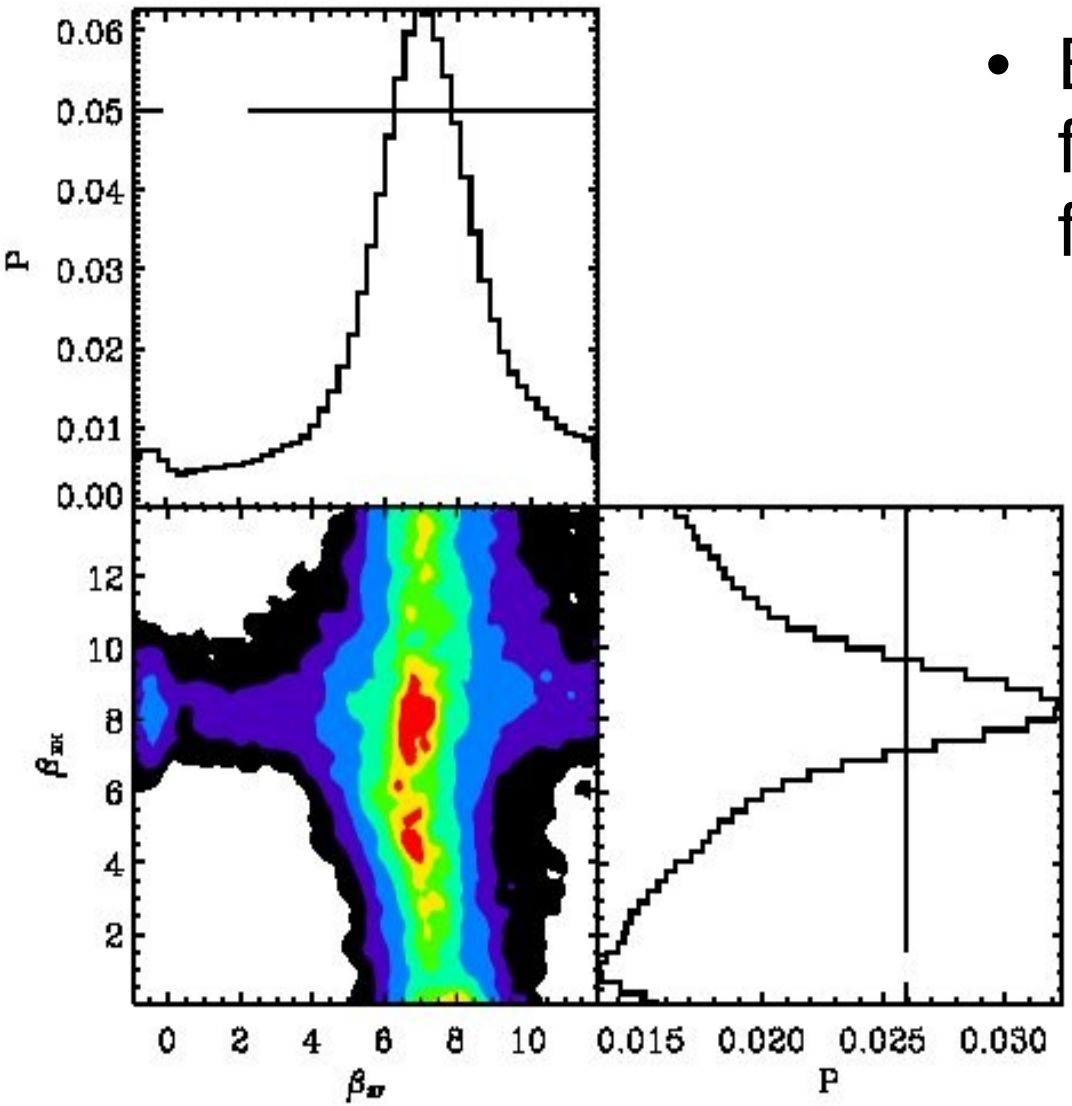
- Fairly strong constraints on cooling cut-off halo mass and merger time scale, but bi-modal appears.



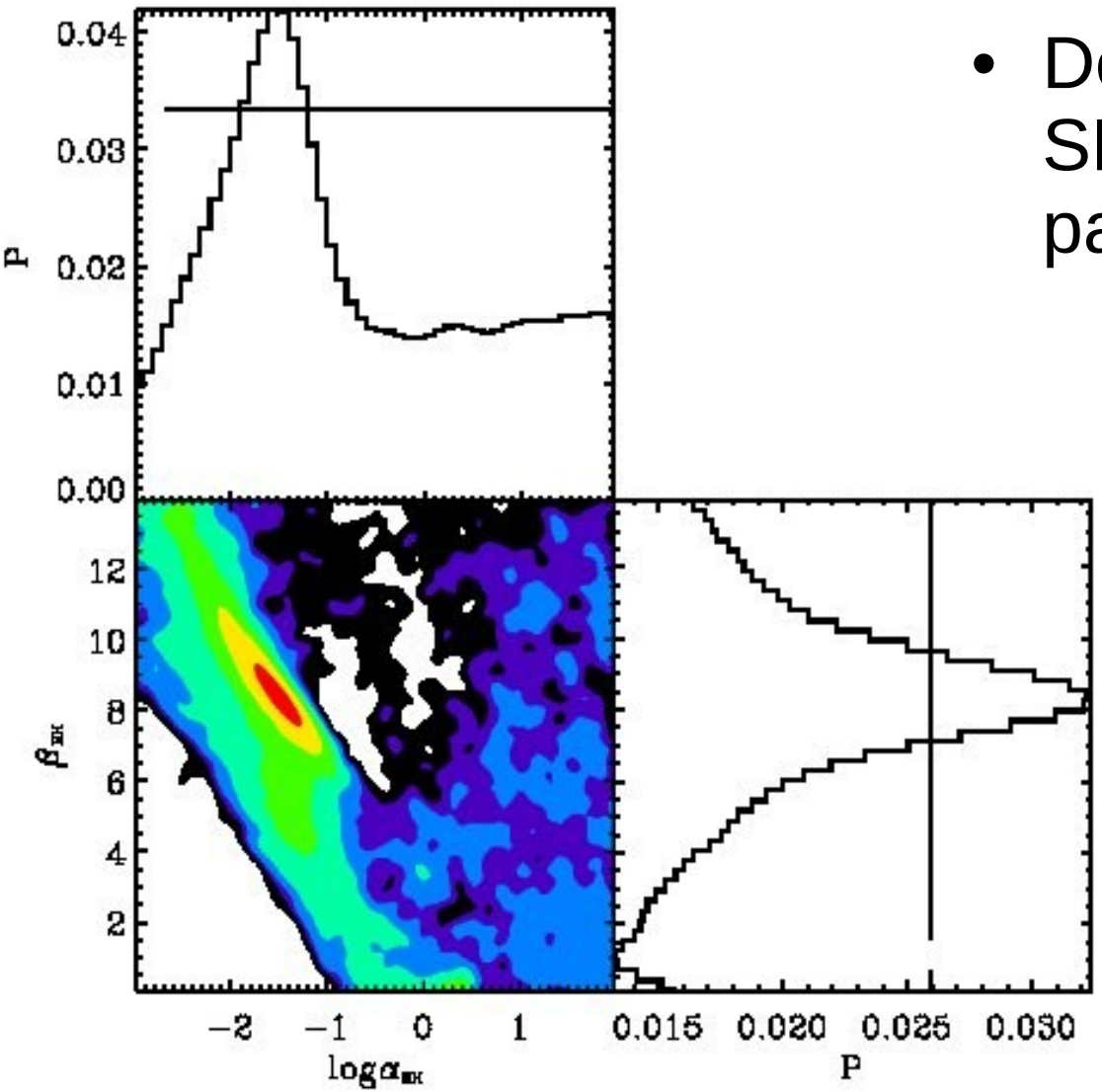
- Strong degeneracy between the star formation efficiency and star formation threshold (and disk size - $\Sigma_{SF} R_d^2$)



- Strong constraint on star formation transition halo circular velocity

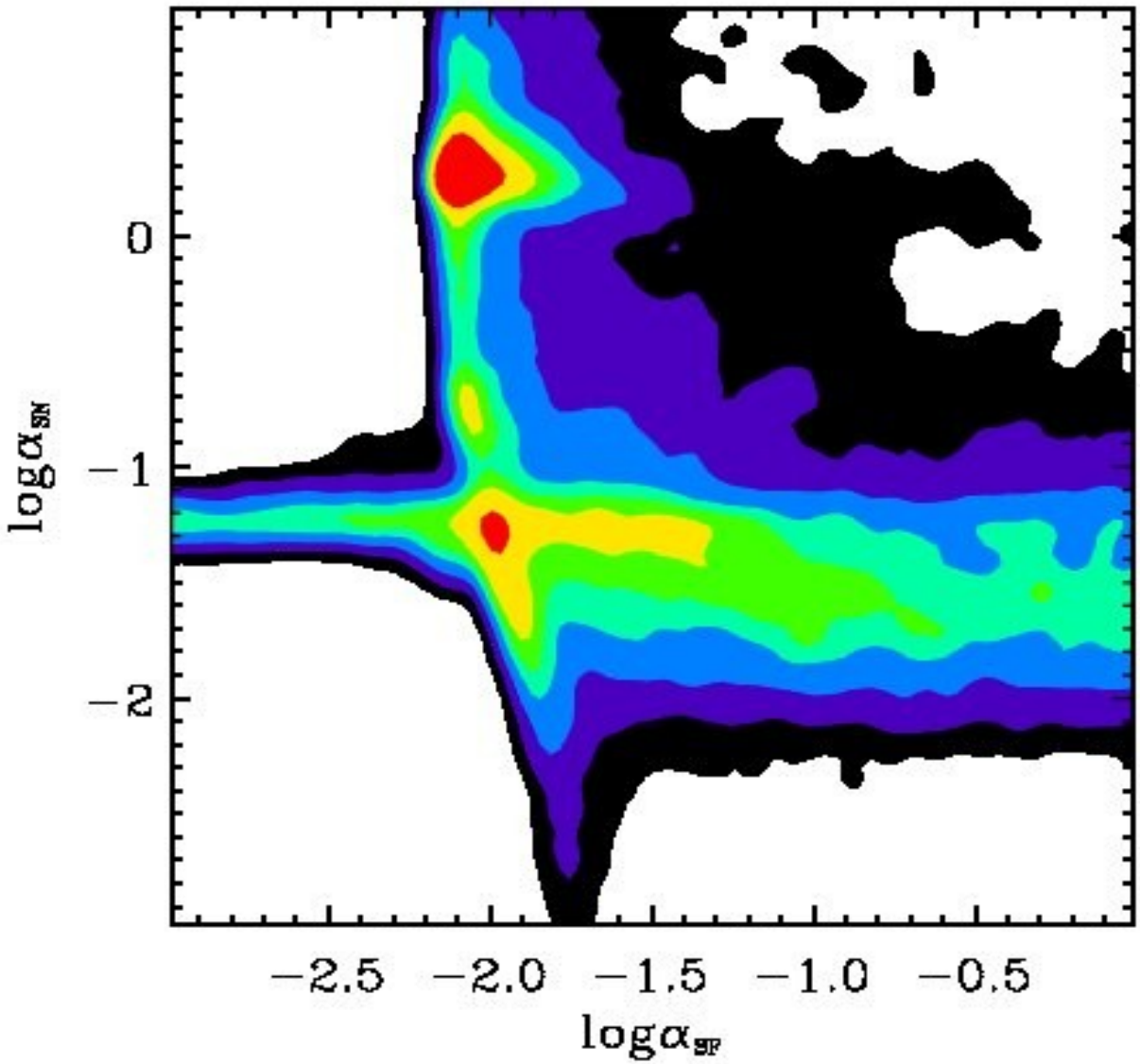


- Bimodal for star formation and SN feedback parameters



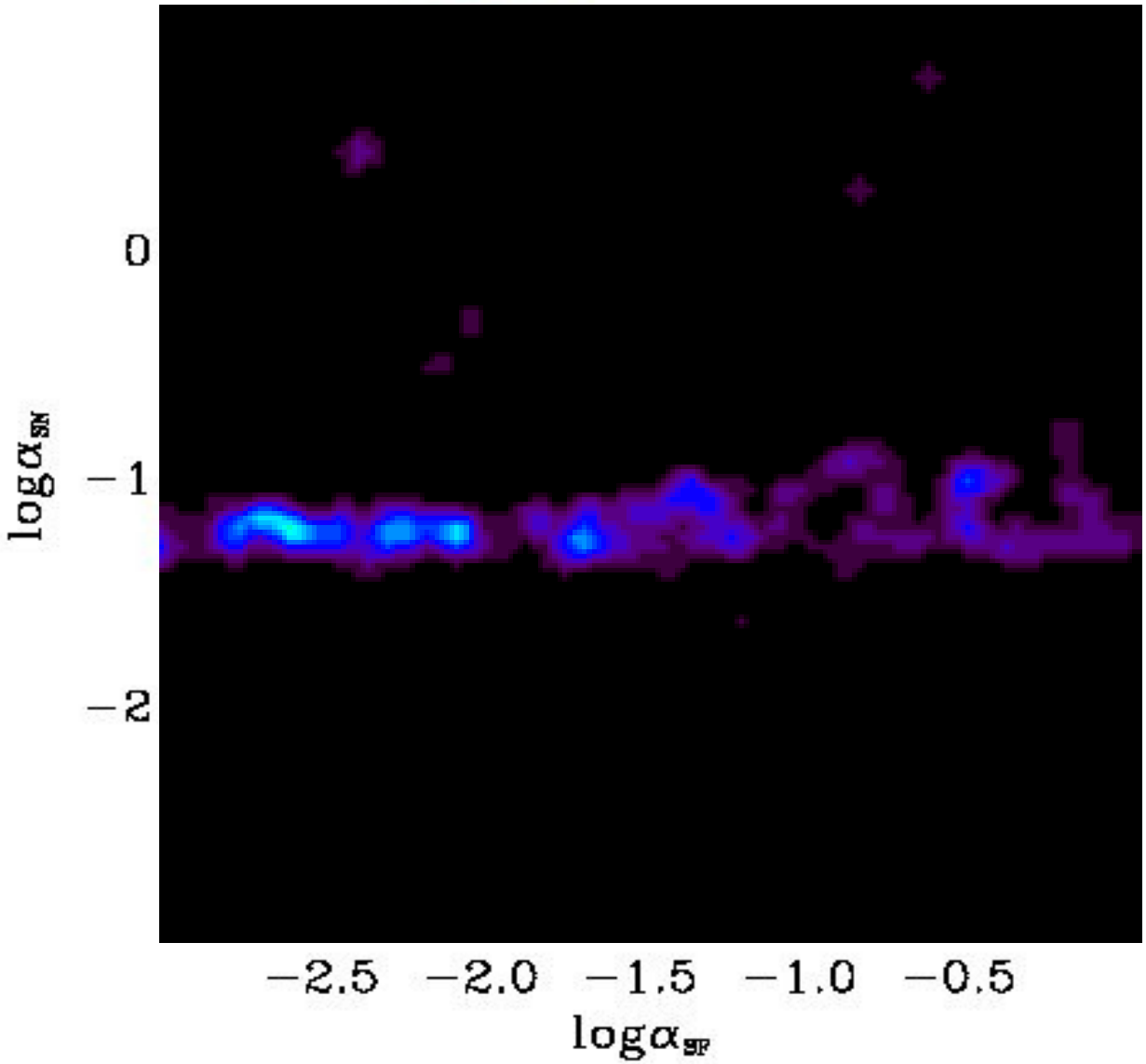
- Degeneracy between SN feedback parameters

Structure of the likelihood function



- 10-D parameter space
- Marginalize 8 Ds

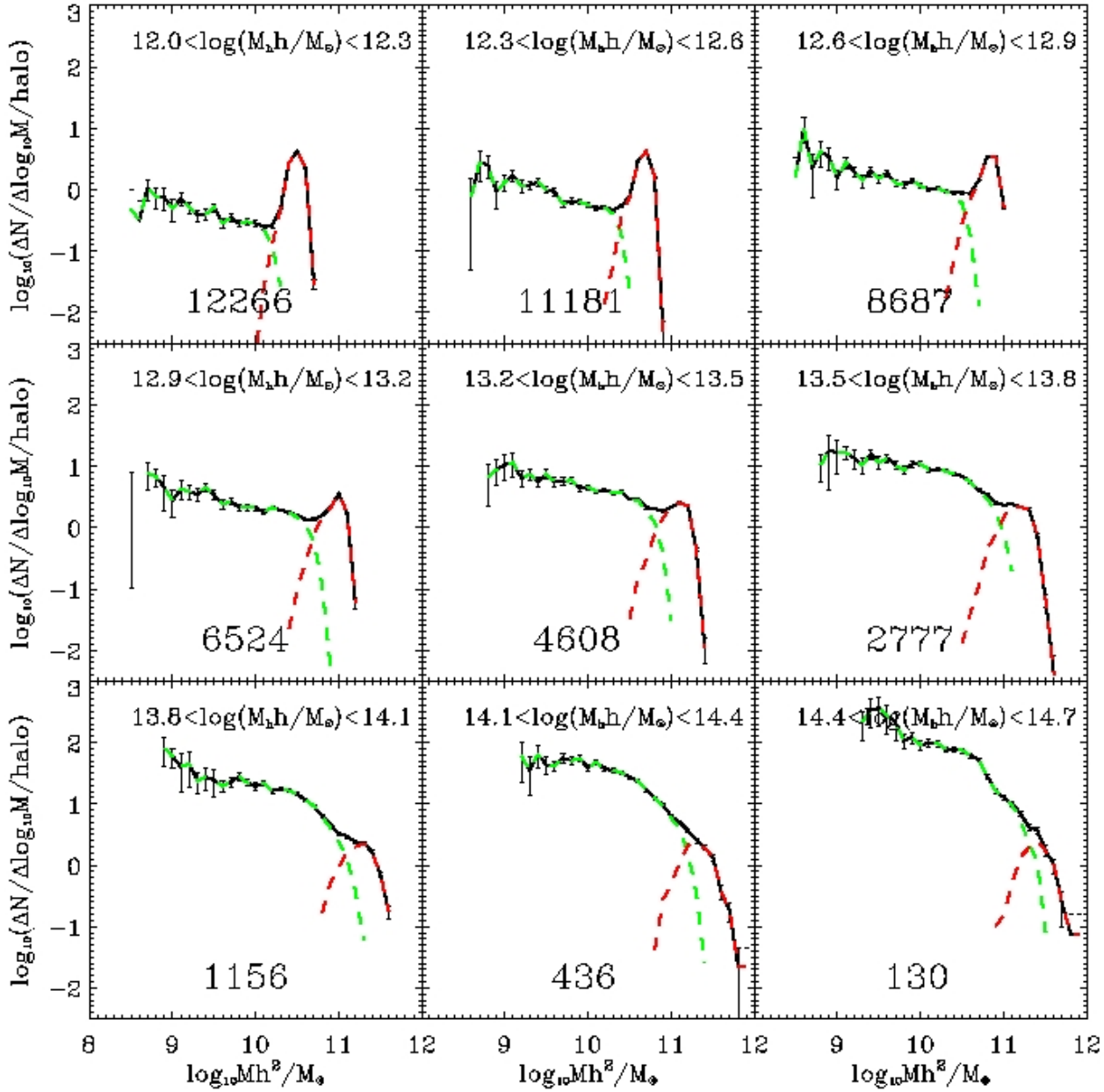
Structure of the likelihood function



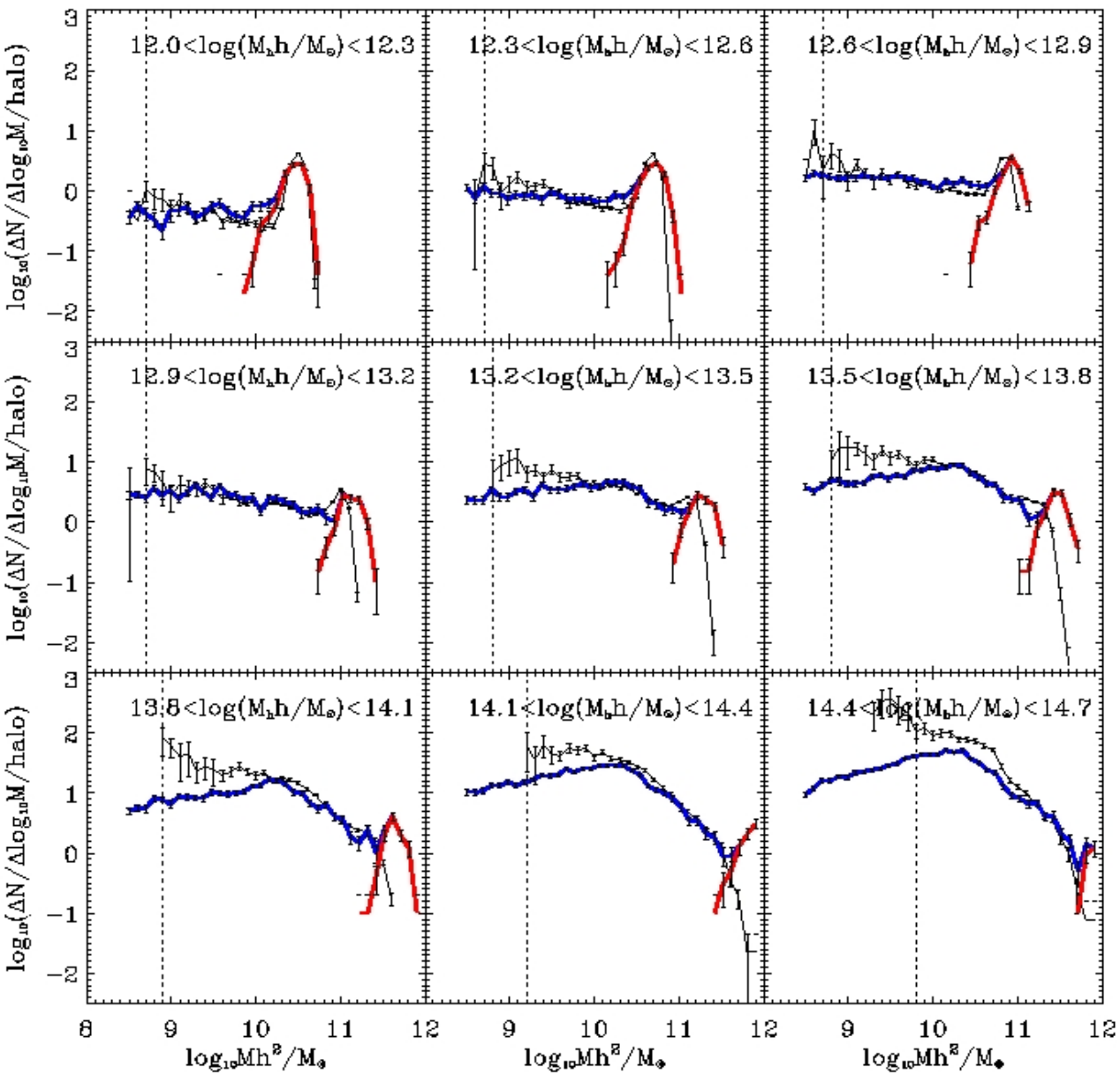
- 10-D parameter space
- Marginalize 7 Ds
- 2-D slice

Conditional stellar mass function

- Data (Yang etal 2009) from halo catalog for SDSS galaxies.
- More information in data
  - Break degeneracy
  - Test model further

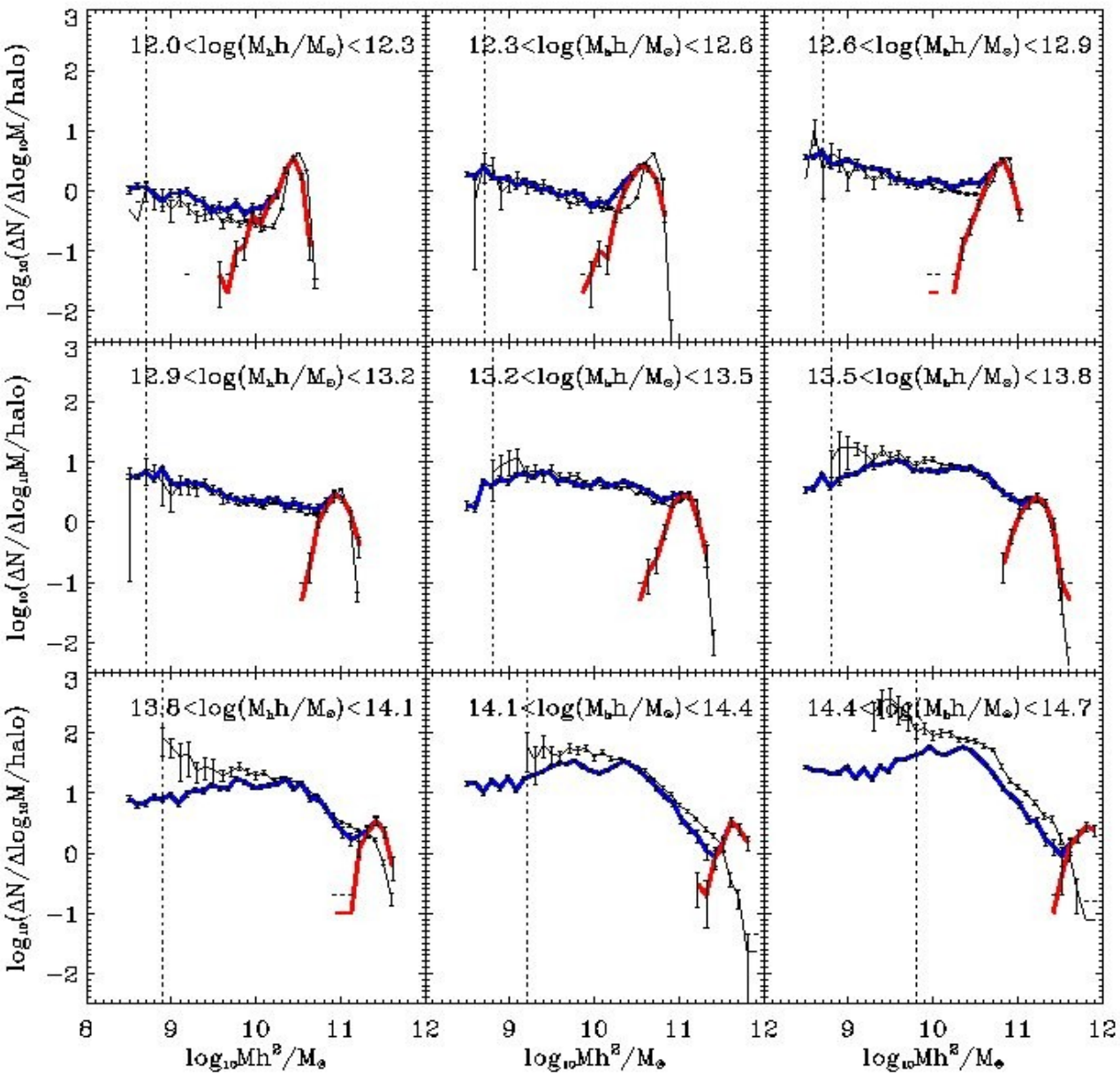


Fiducial model



- 13 free parameters
- Cooling: Munich model (Croton 2006)

New cooling model

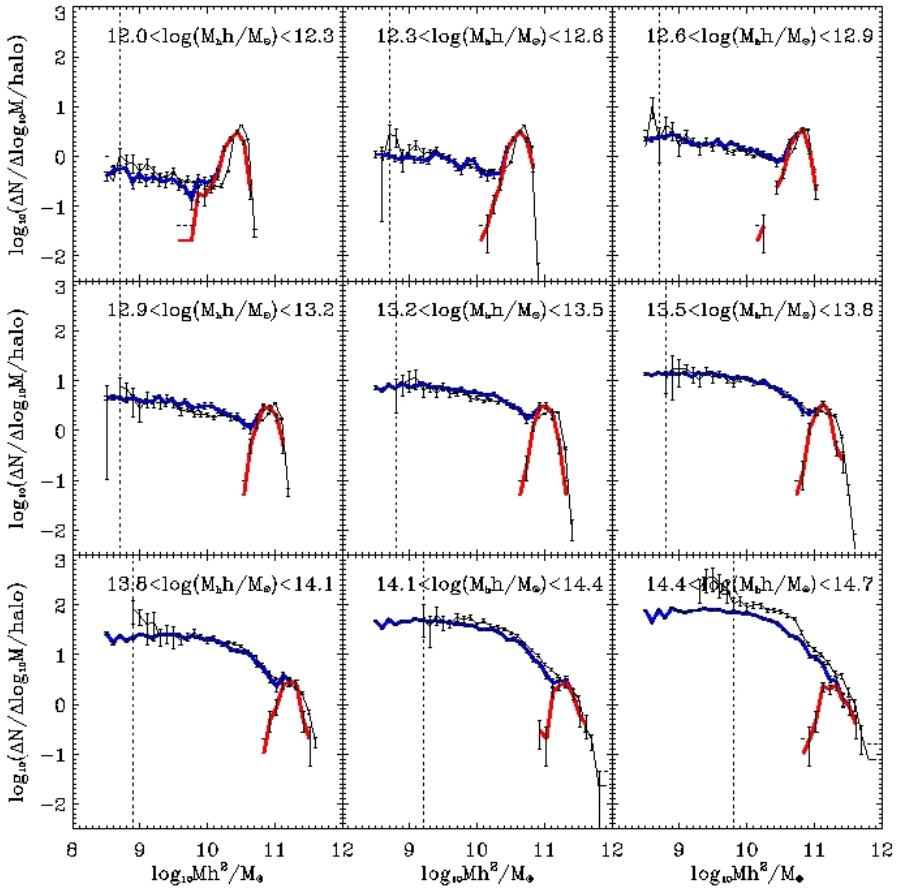


- New cooling model (cools faster in small halos, slower in massive halos

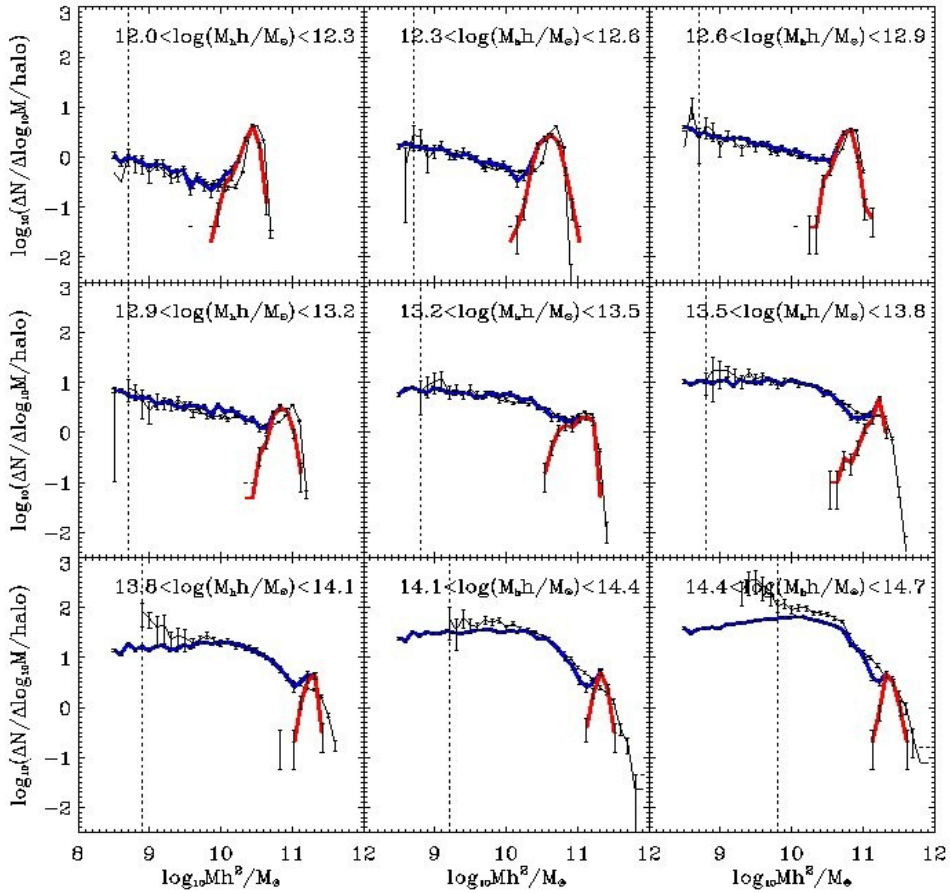
Satellite stripping models

- Hints:
  - The models over-predict central galaxy mass by a factor of 2 for cluster size halos.
  - In observation, clusters have significant fraction of stellar mass not in galaxies.
- Tidal stripping model: a fraction of stellar mass, f_{TS} , of a satellite galaxy is stripped in one orbital time.

Munich cooling model+stripping



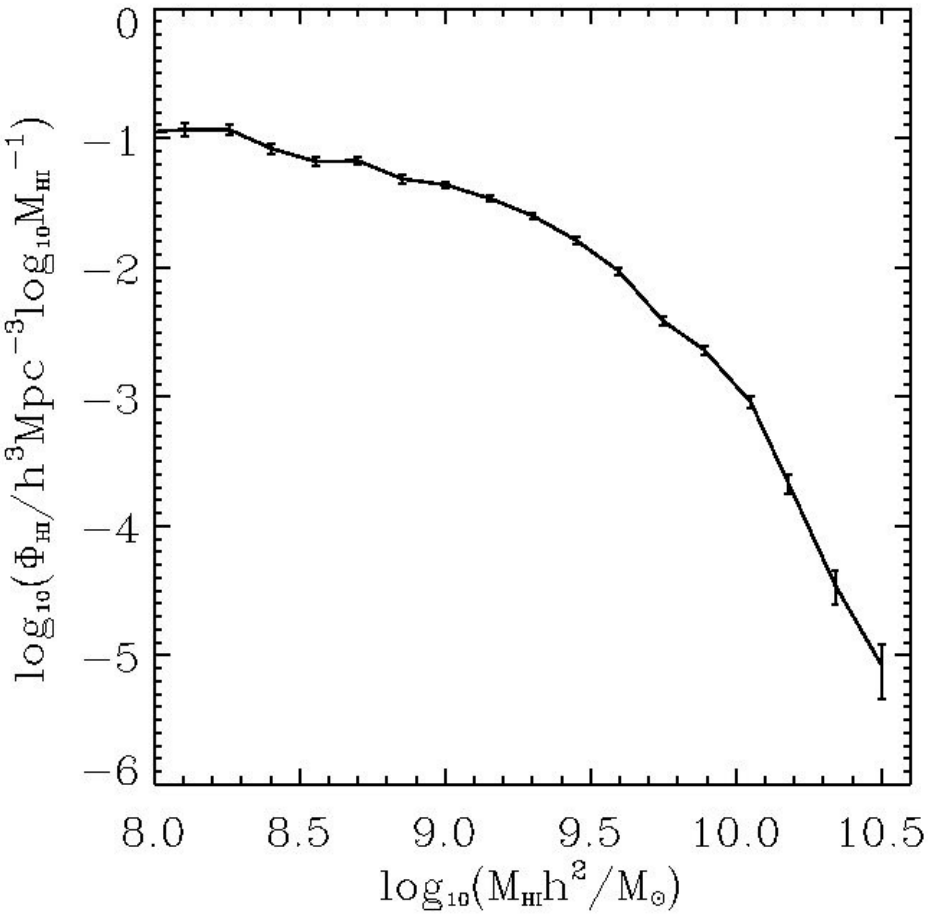
New cooling model+stripping



$$\frac{E(M_1)}{E(M_0)} = 3.5\times10^{9}$$

$$\frac{E(M_2)}{E(M_0)} = 2.1\times10^{10}$$

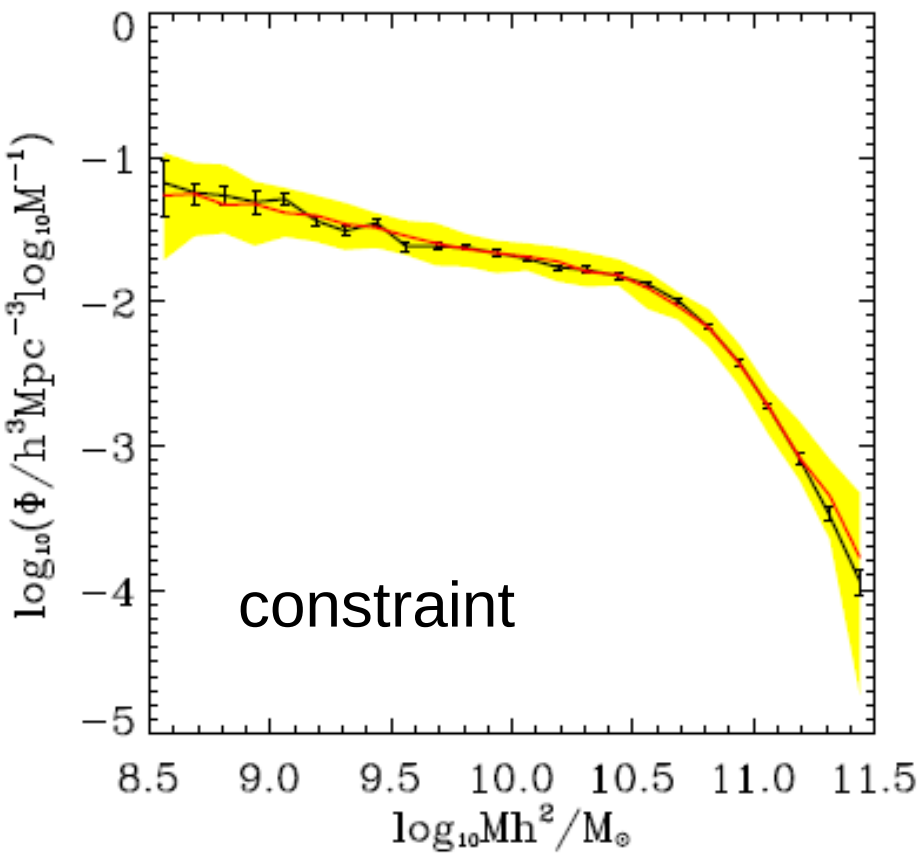
HI mass function

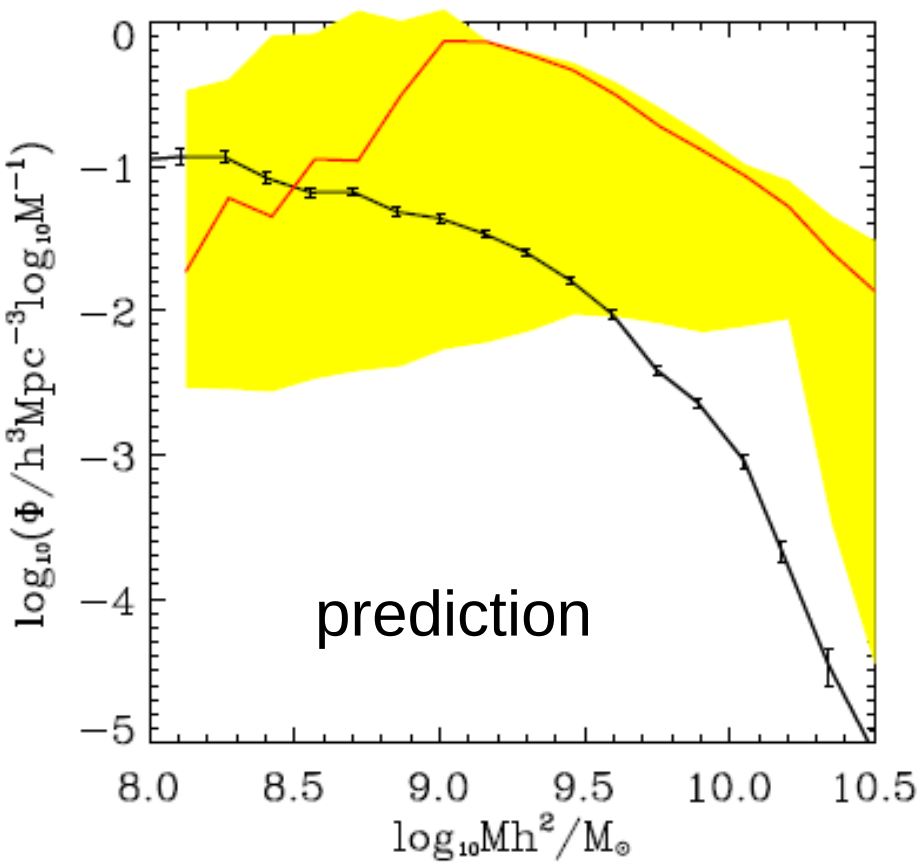


- The mass function of HI gas at z=0 (HIPASS, Zwaan etal 2005)
- Constrain the cold gas mass

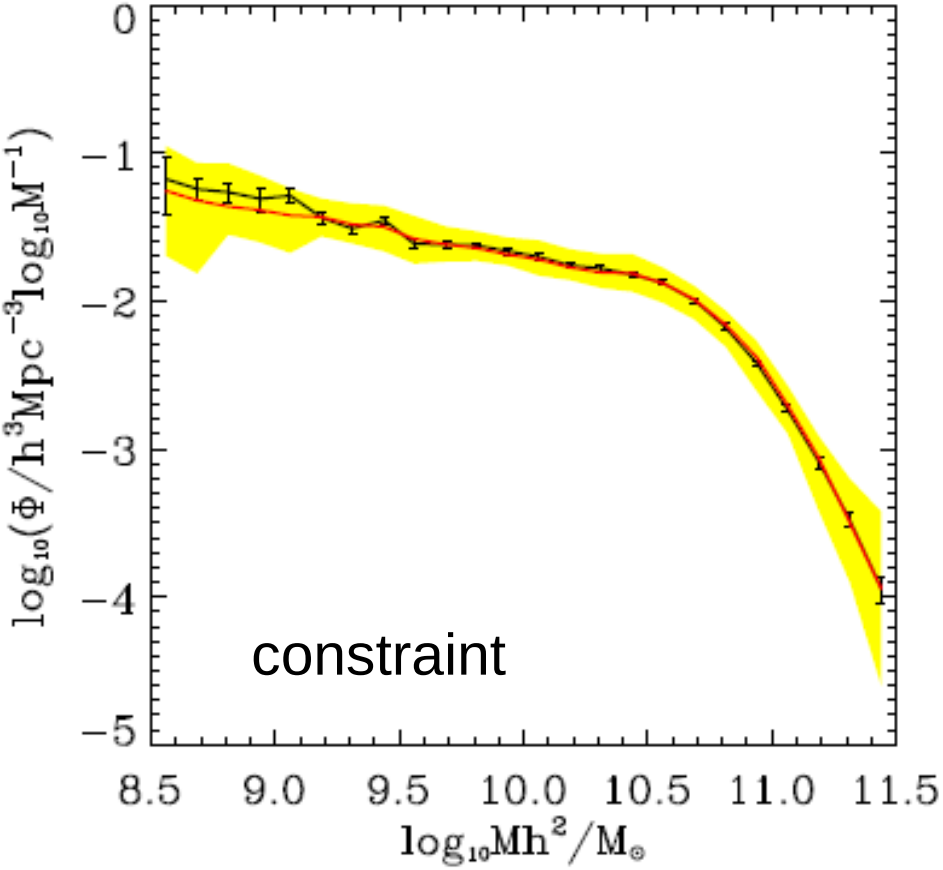
Is HI mass function compatible with stellar mass function?

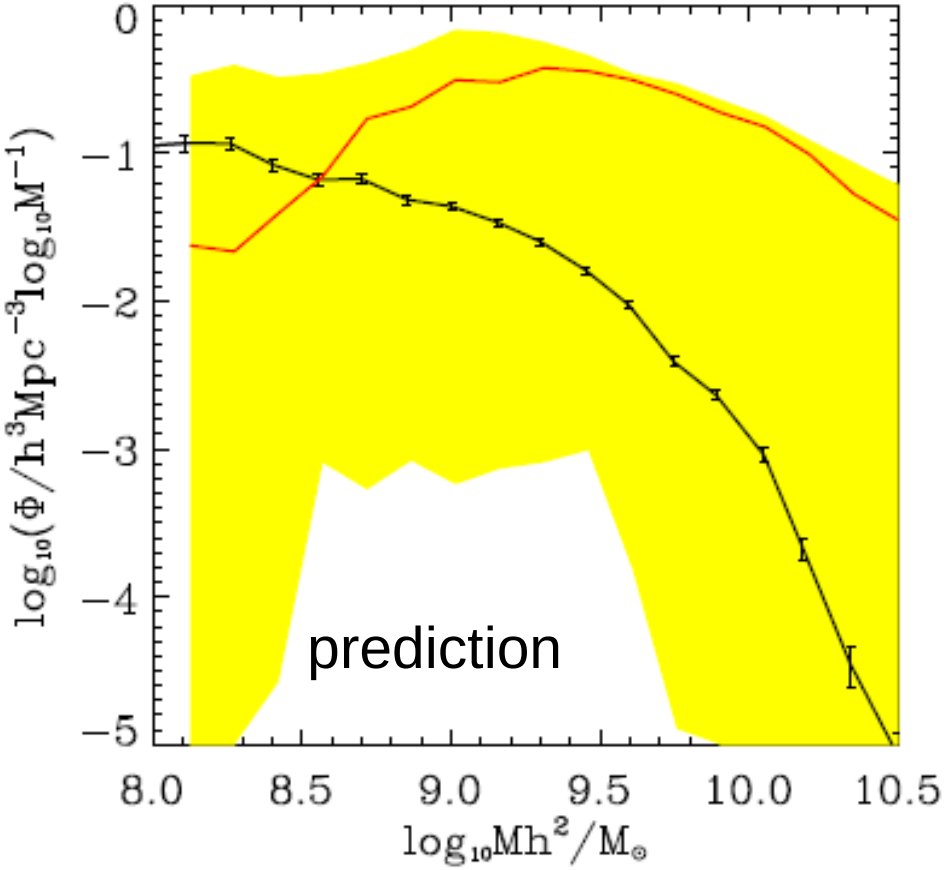
Munich cooling model

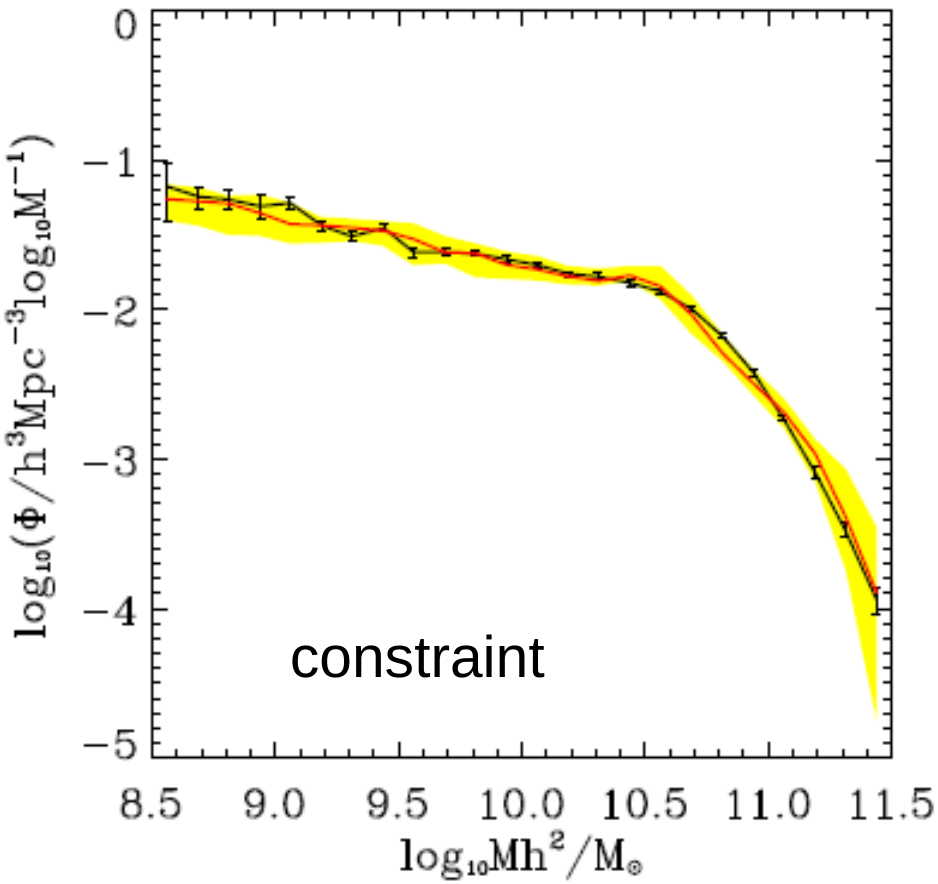


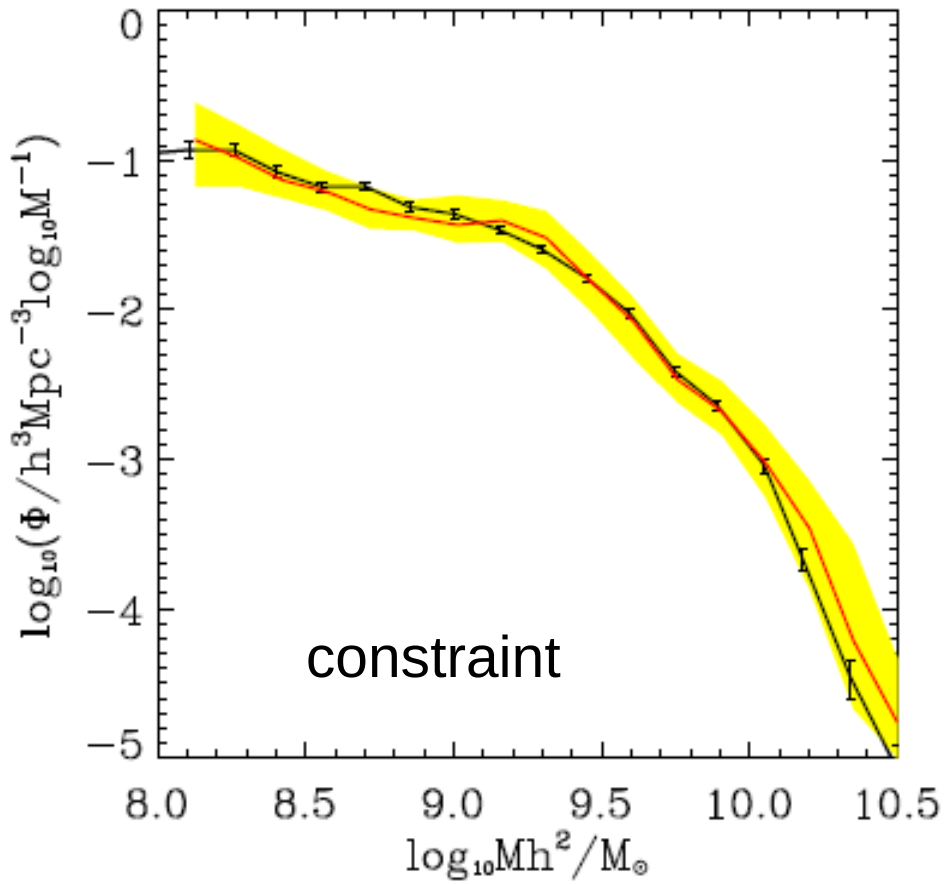


New cooling model

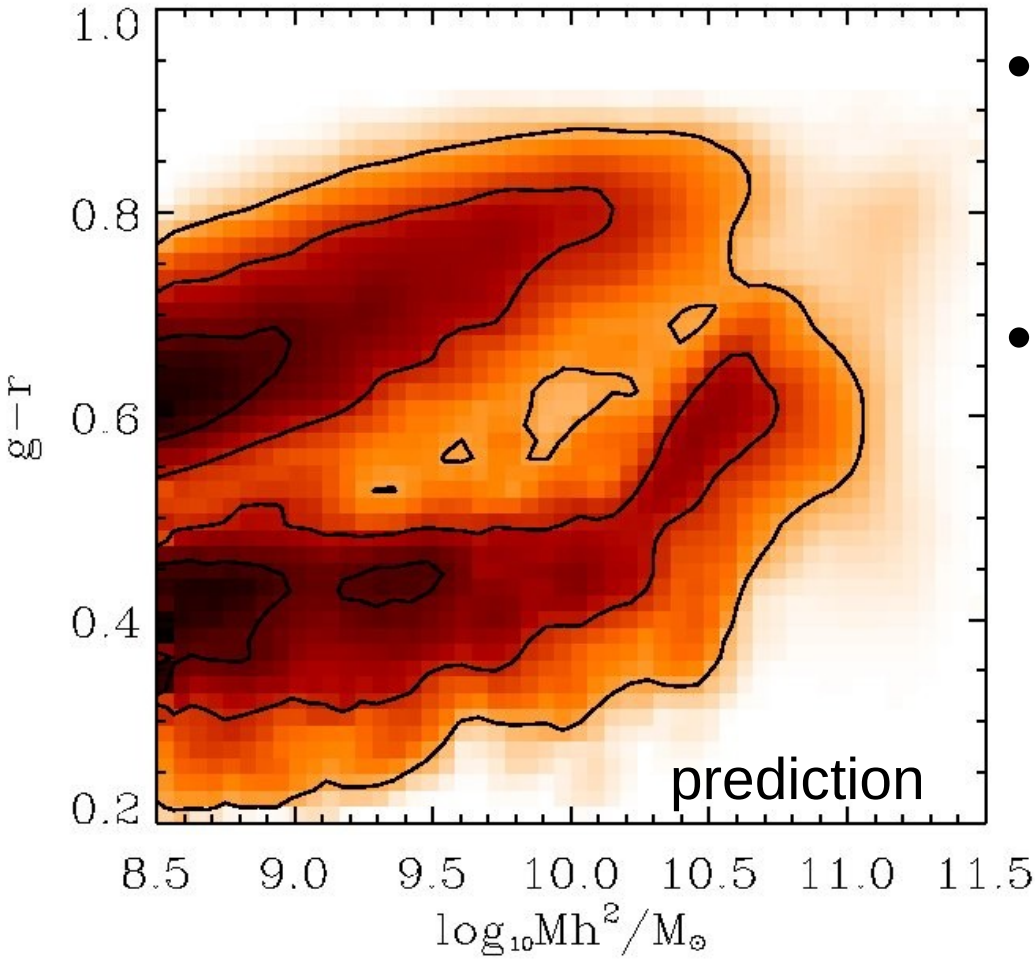





constraint
$\log_{10}(\Phi/h^3\mathrm{Mpc}^{-3}\log_{10}M^{-1})$
$\log_{10}Mh^2/M_\odot$
0
−1
−2
−3
−4
−5
8.5
9.0
9.5
10.0
10.5
11.0
11.5

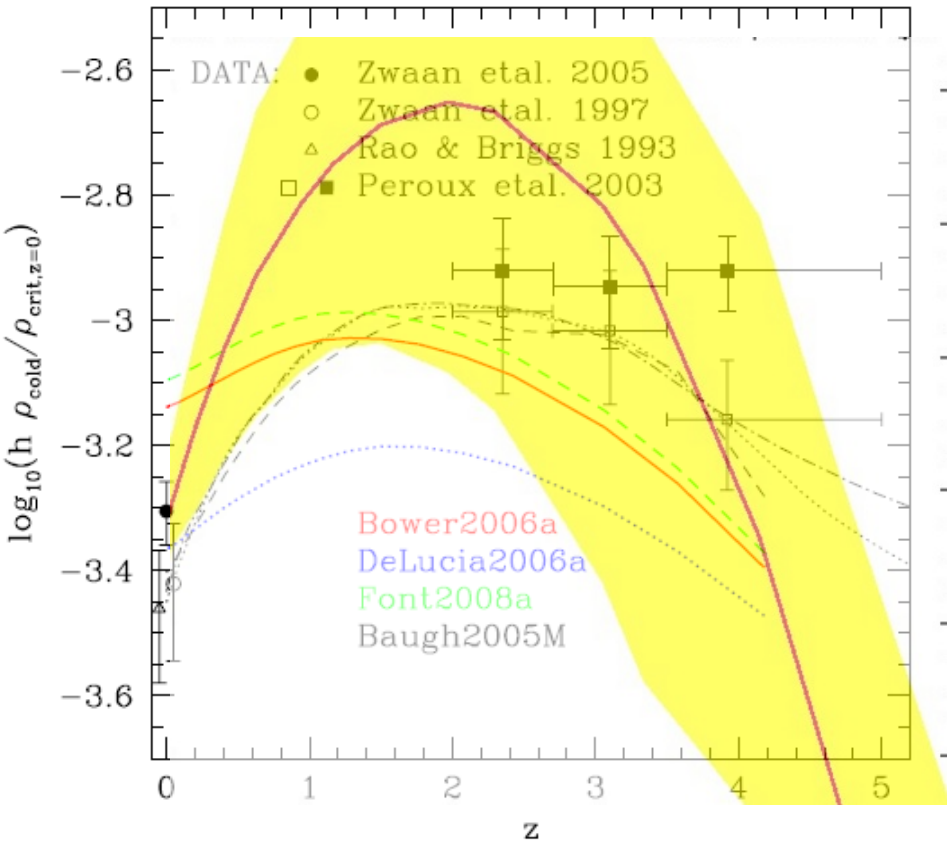

constraint
$\log_{10}(\Phi/h^3\mathrm{Mpc}^{-3}\log_{10}M^{-1})$
$\log_{10}Mh^2/M_\odot$
0
−1
−2
−3
−4
−5
8.0
8.5
9.0
9.5
10.0
10.5

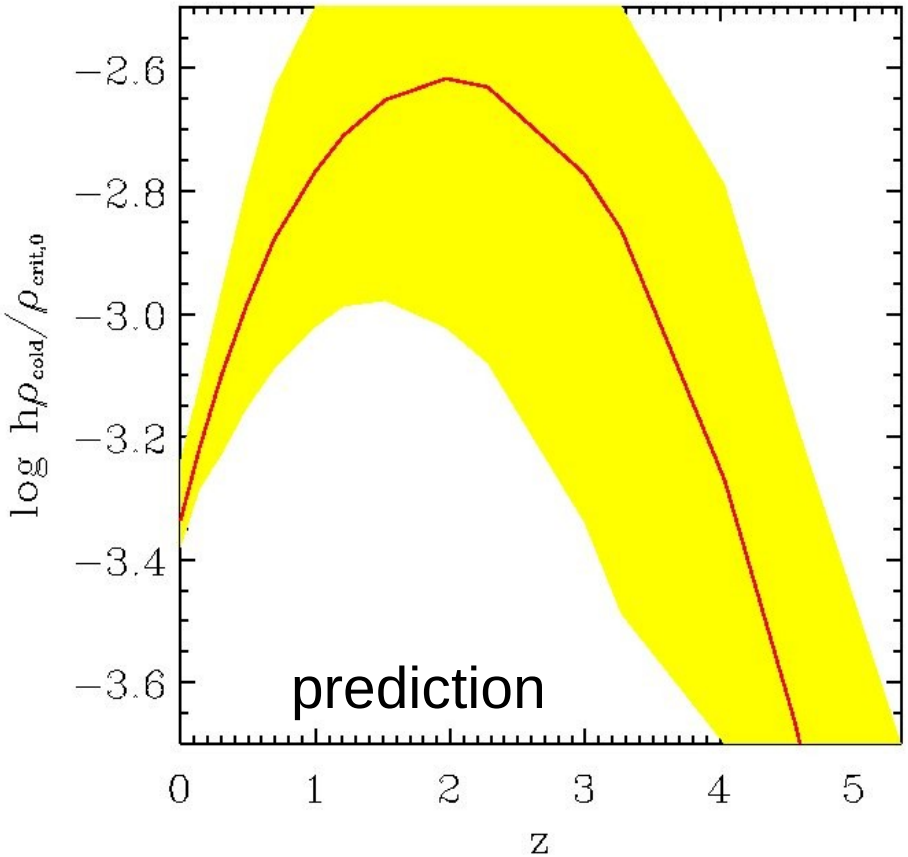
Color-stellar mass diagram



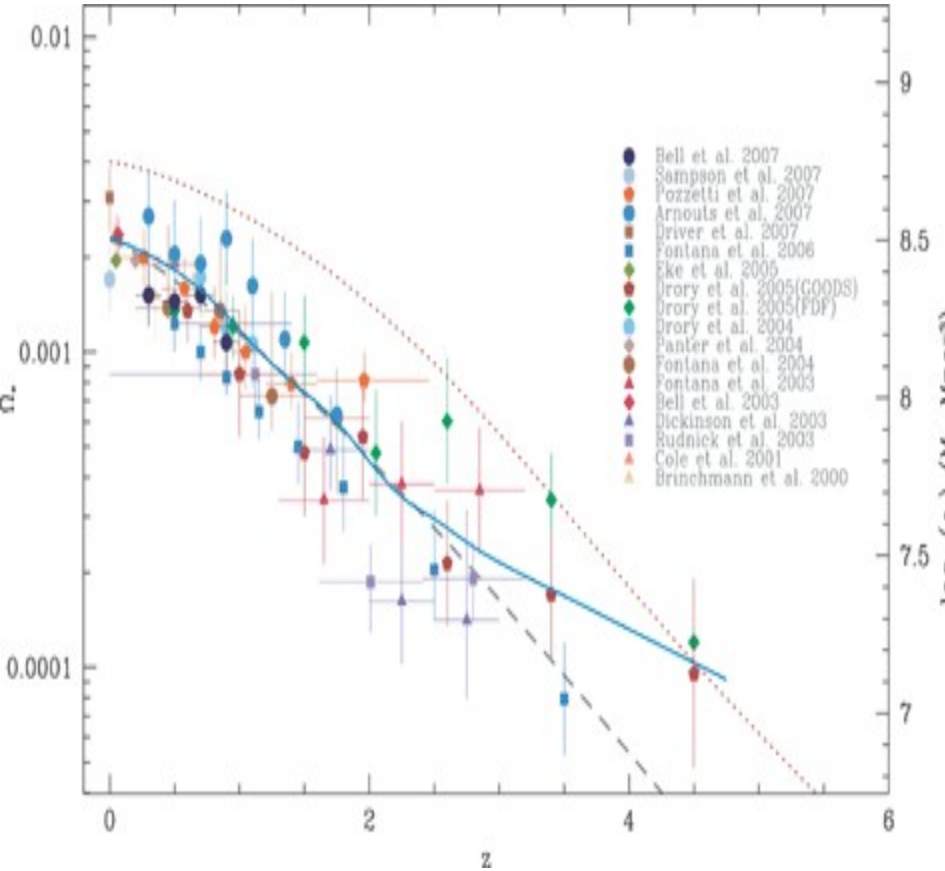
- Bimodality arises without AGN feedback.
- AGN feedback might be helpful to quench star formation in massive galaxies.

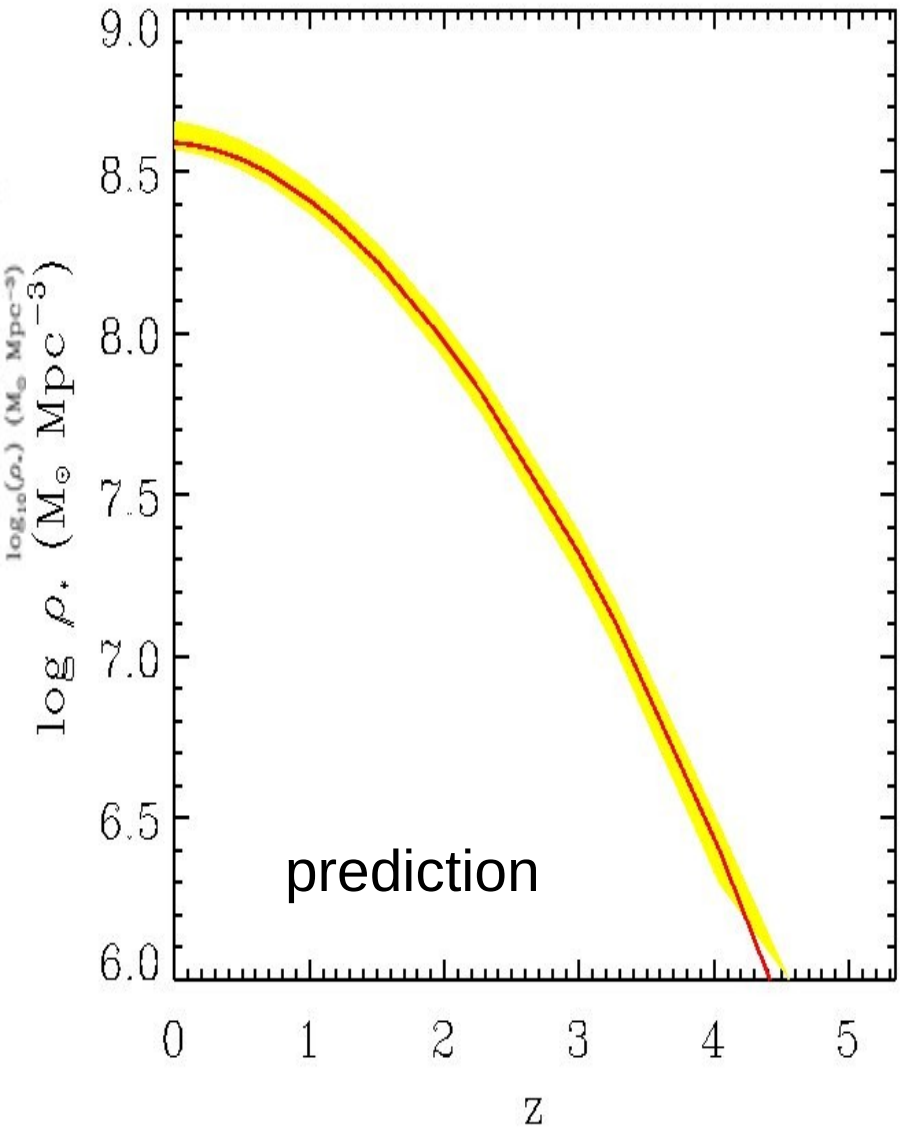
Cold gas mass density





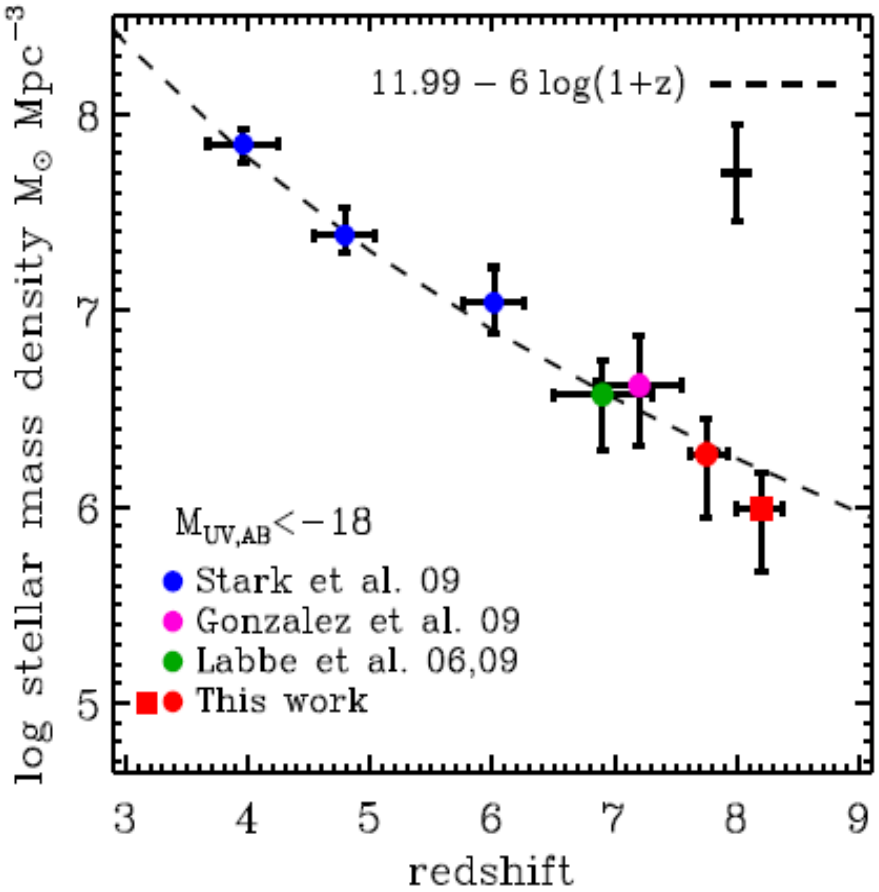
Stellar mass density



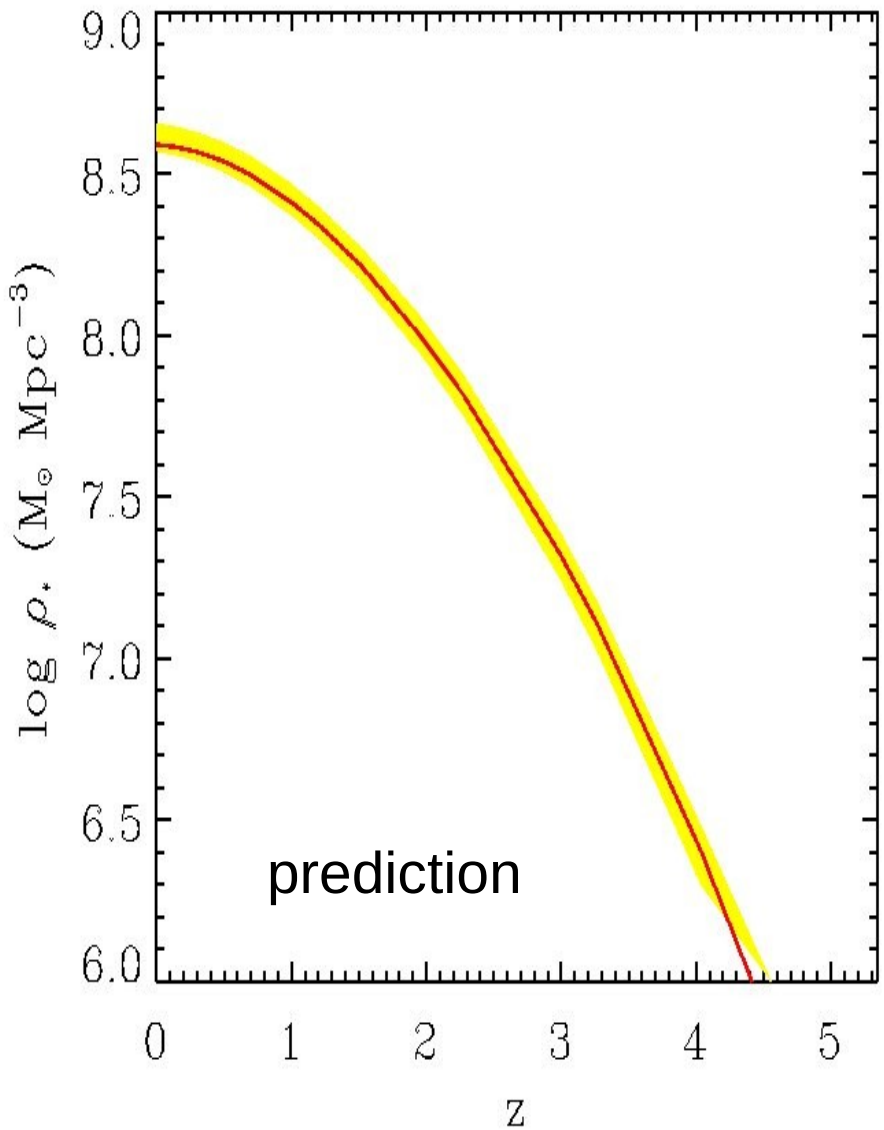


Wilkins etal 2008

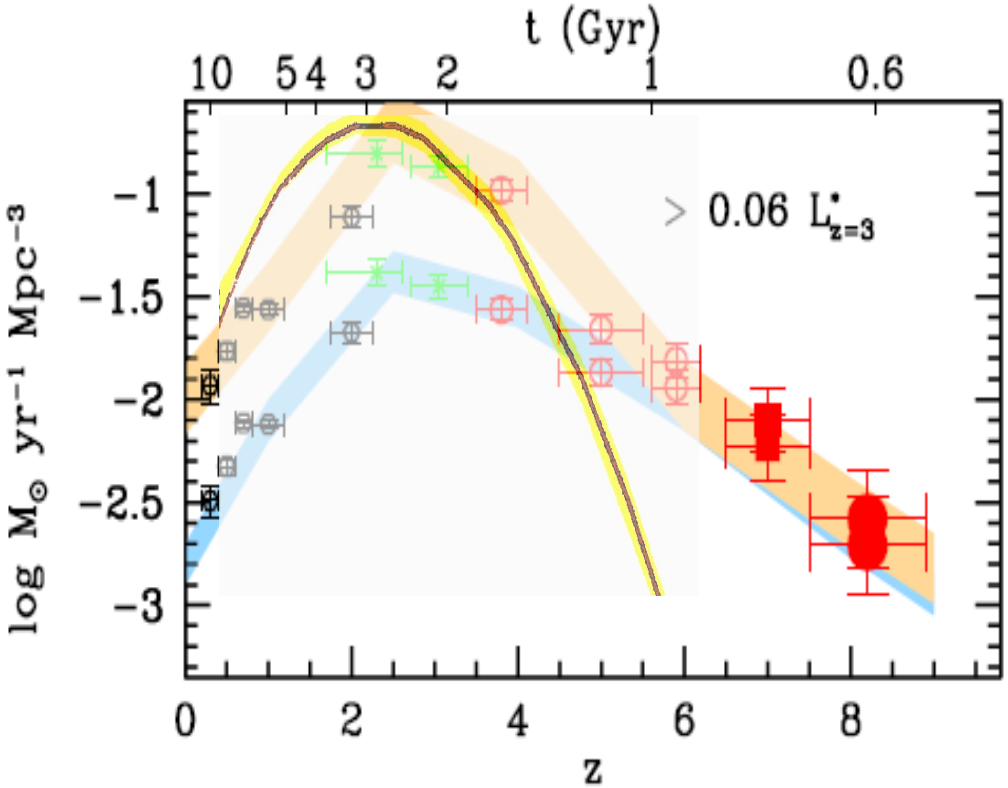
Stellar mass density

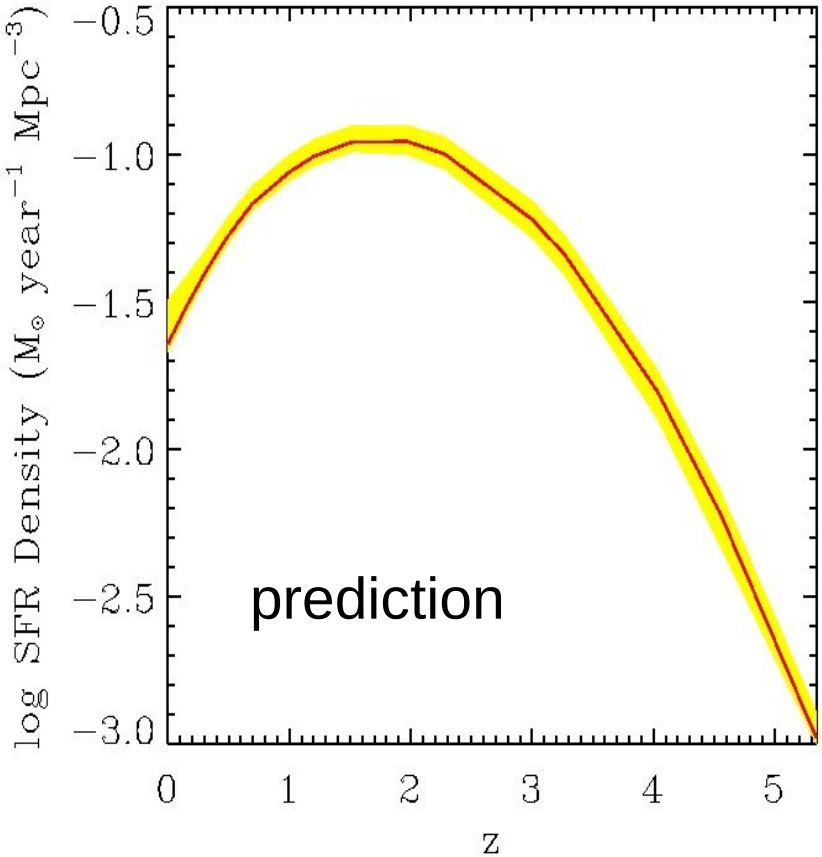


Labbe etal 2009

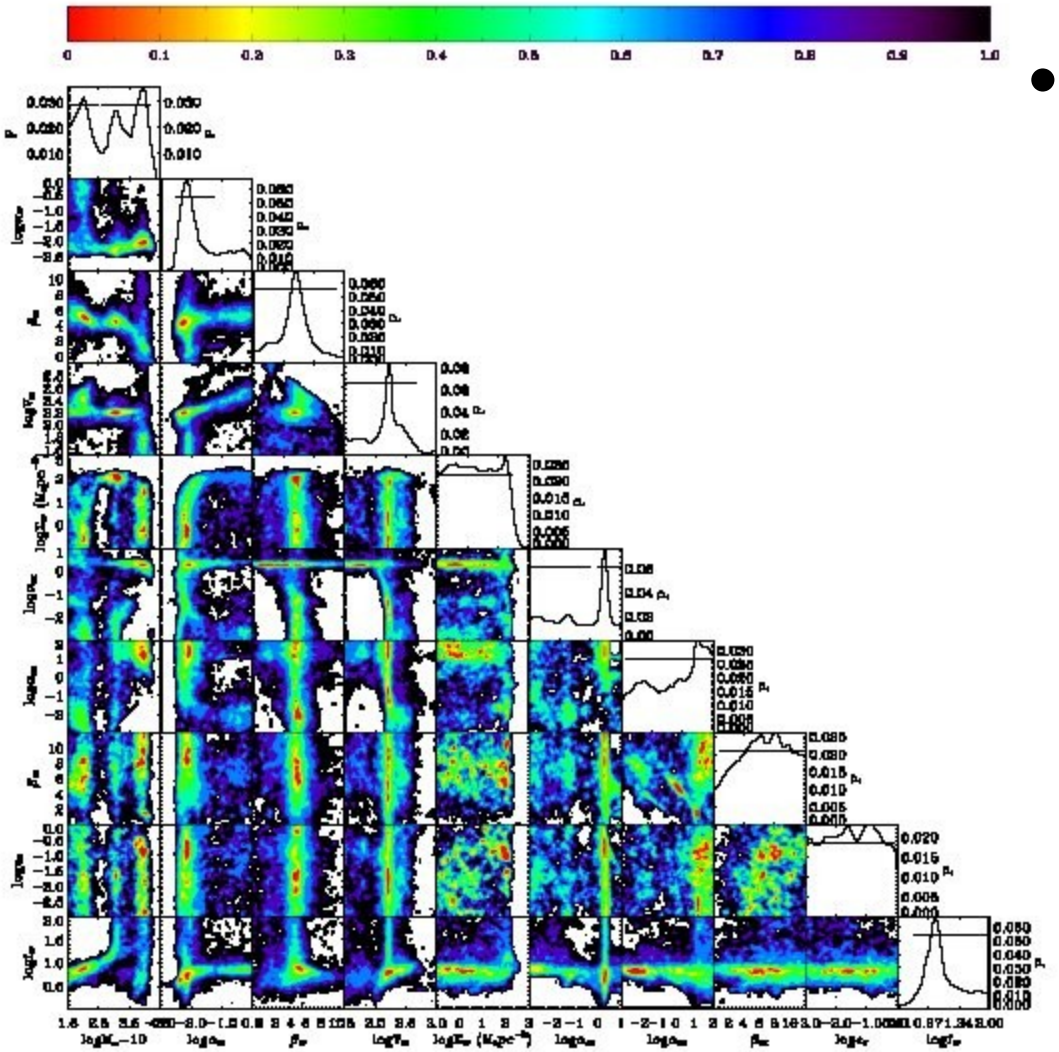


Star formation rate density



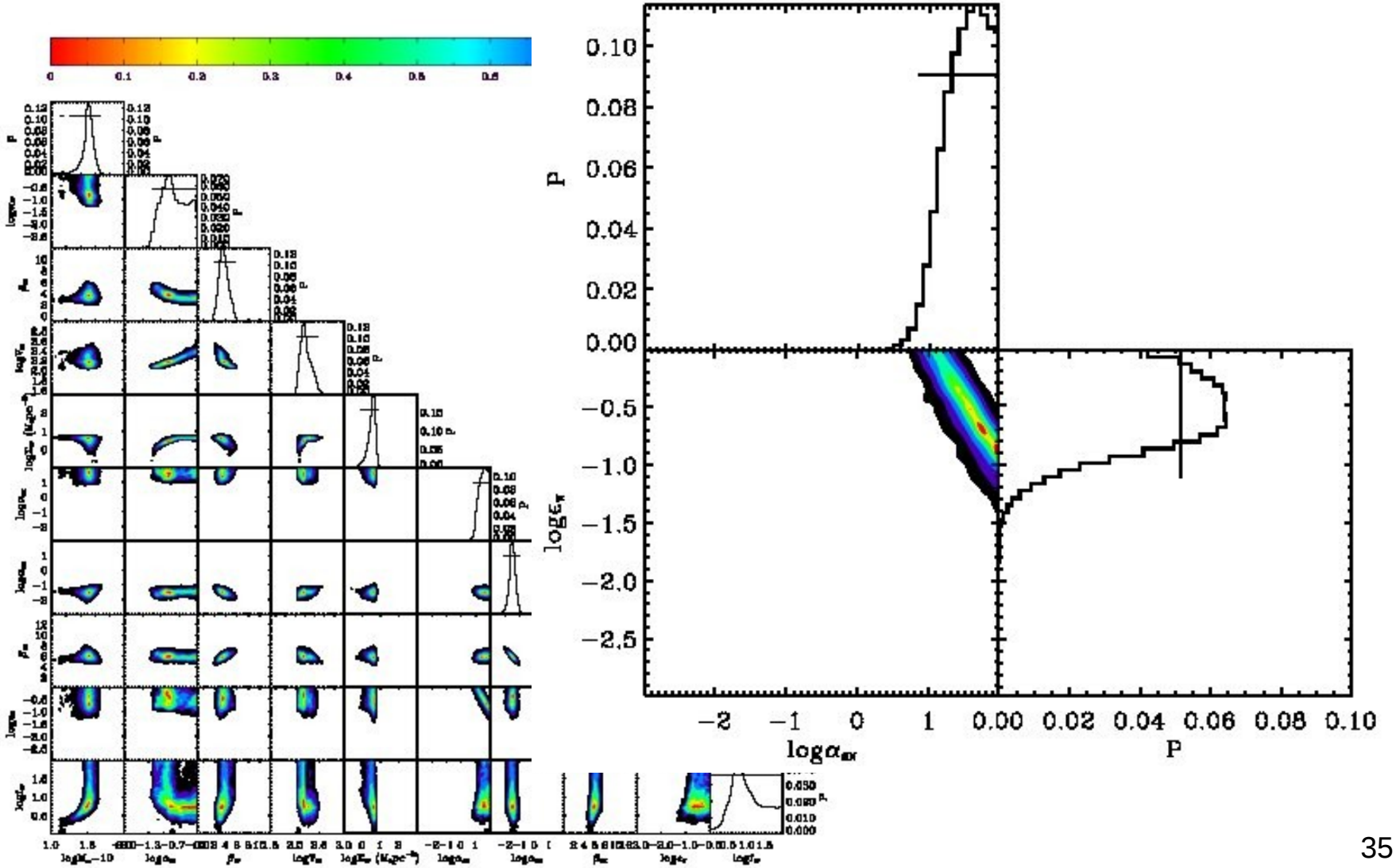


Bouwens etal 2009



- Posterior of the new cooling model, constrained by stellar mass function

Too much SN feedback energy!



Summary

- SAM contains huge degeneracy.
- The degeneracy can be broken by using more data of different observations.
- Some important physics are missing from the model.
- For modelers: models need to either follow the real physics or be broad enough.
- For observers: errors need to be estimated with more care.

Summary

- Bayesian approach allows us to use the real power of SAM.
  - Constrain models
  - Select models
  - Predict observables for future observations
- Constraints from other observations: clustering, colors, high-z
- Other applications: deep galaxy surveys+HOD model

Thank you!