

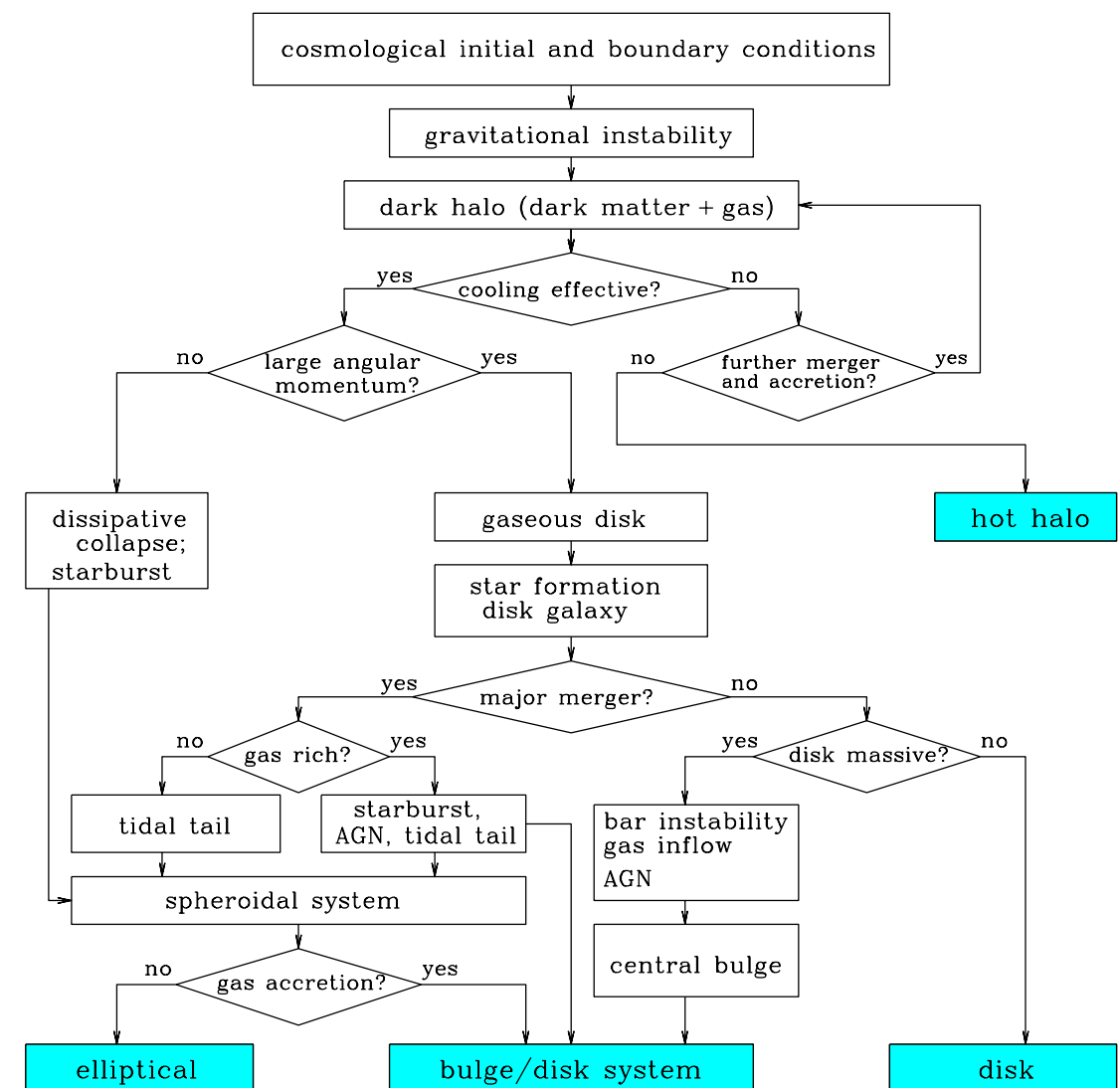
A Bayesian Approach to the Semi-analytical Model of Galaxy Formation

Houjun Mo
(University of Massachusetts)

With
Yu Lu, N. Katz and Martin Weinberg

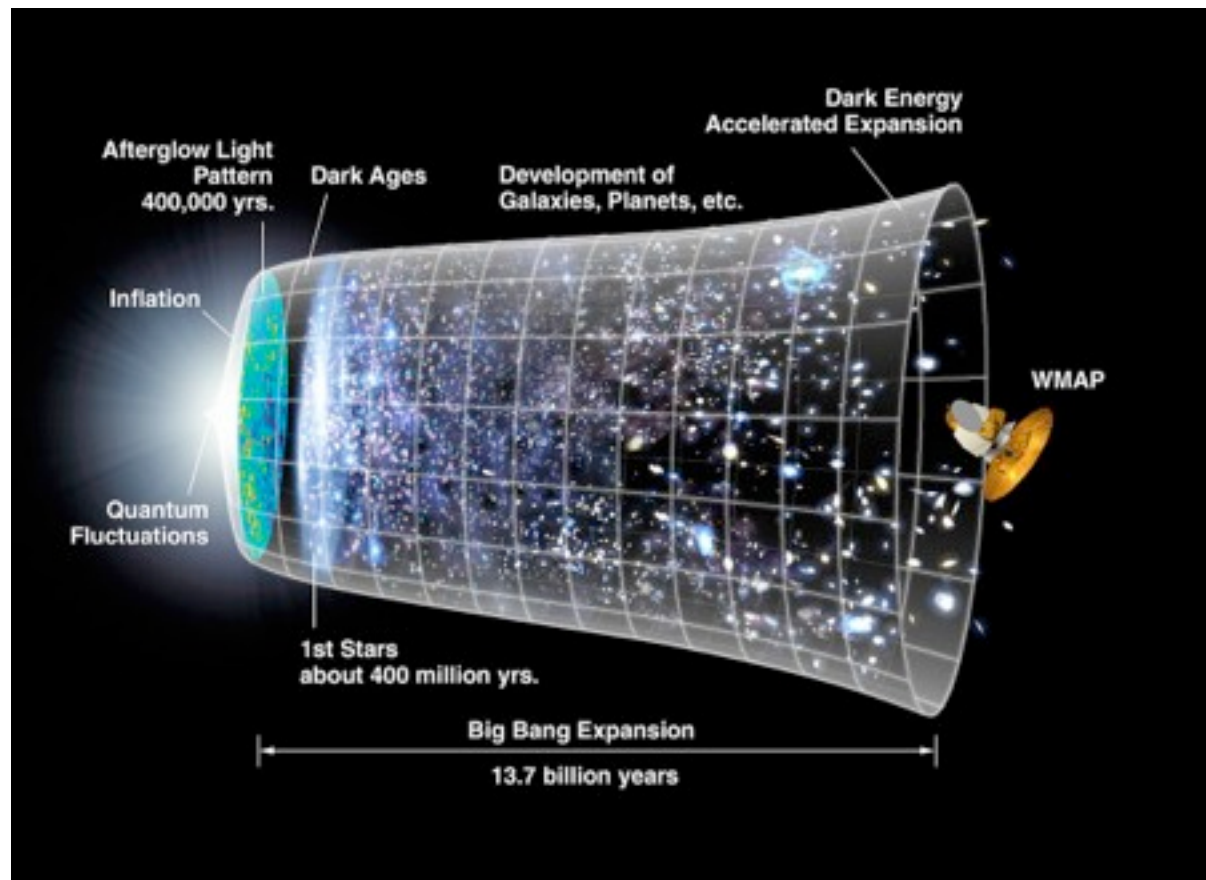
General framework of galaxy formation

- Cosmology: provides the initial and boundary conditions for galaxy formation
- Gravitational instability: drives large scale structure formation
- Formation of dark matter halos in the cosmic density field
- Formation of galaxies in dark matter halos: gas cooling, heating and accretion; star formation; feedback; galaxy merger
- Galaxy evolution: spectral synthesis; chemical evolution; dynamical evolution
- Galaxy populations: both low and high z



Can we understand the galaxy population in current cosmology?

The standard paradigm

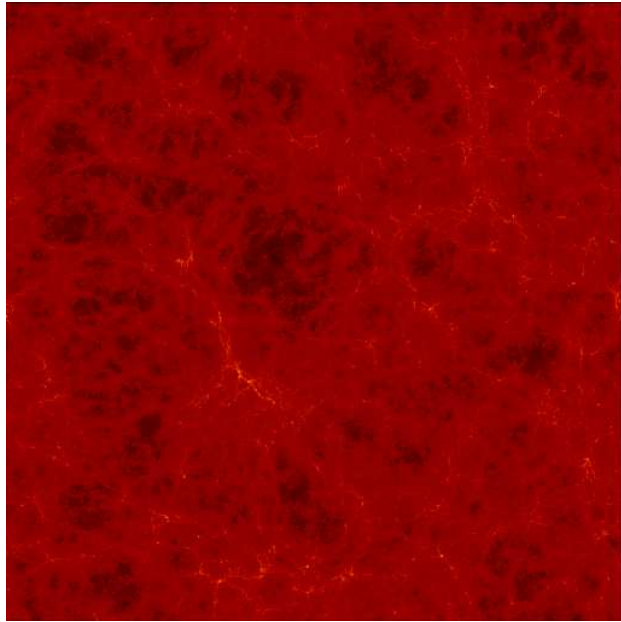


- Big Bang cosmology
- The Universe is dominated by cold dark matter and dark energy
- Primordial perturbations, probably generated by inflation, well constrained by observation (CMB; LSS etc)
- Cosmological models well constrained

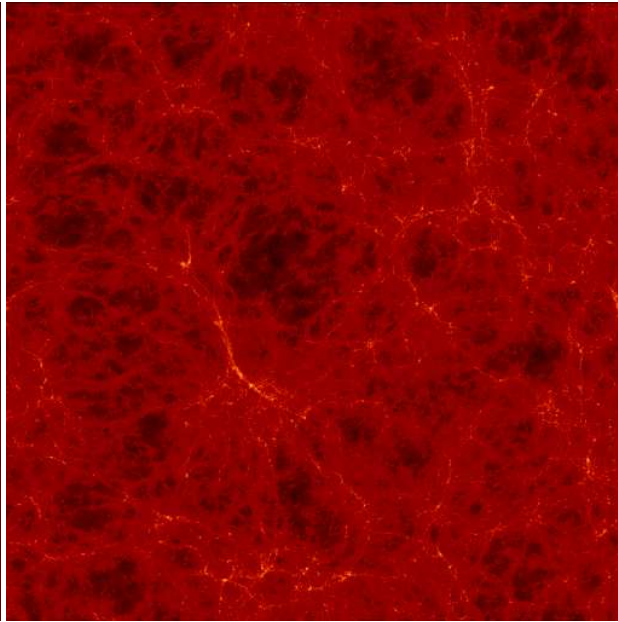
$$k = 0; \Omega_m = 0.25; \Lambda = 0.75; h = 0.7; \Omega_b = 0.04; n = 1$$

Gravitational instability well understood

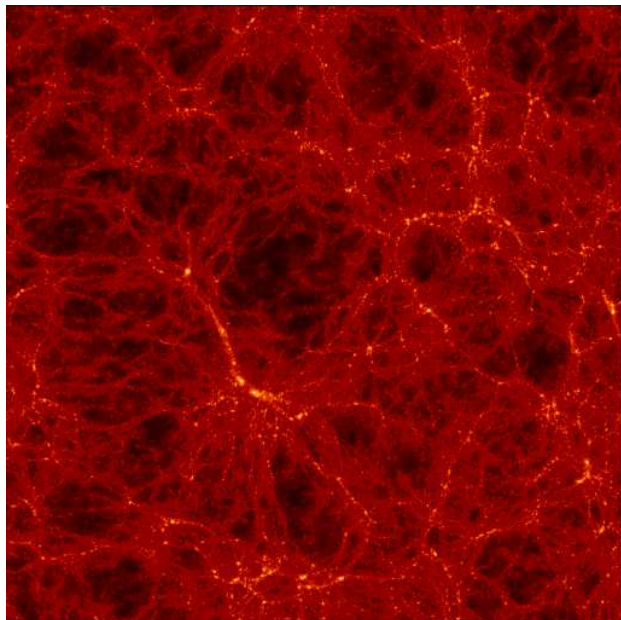
$z=3$



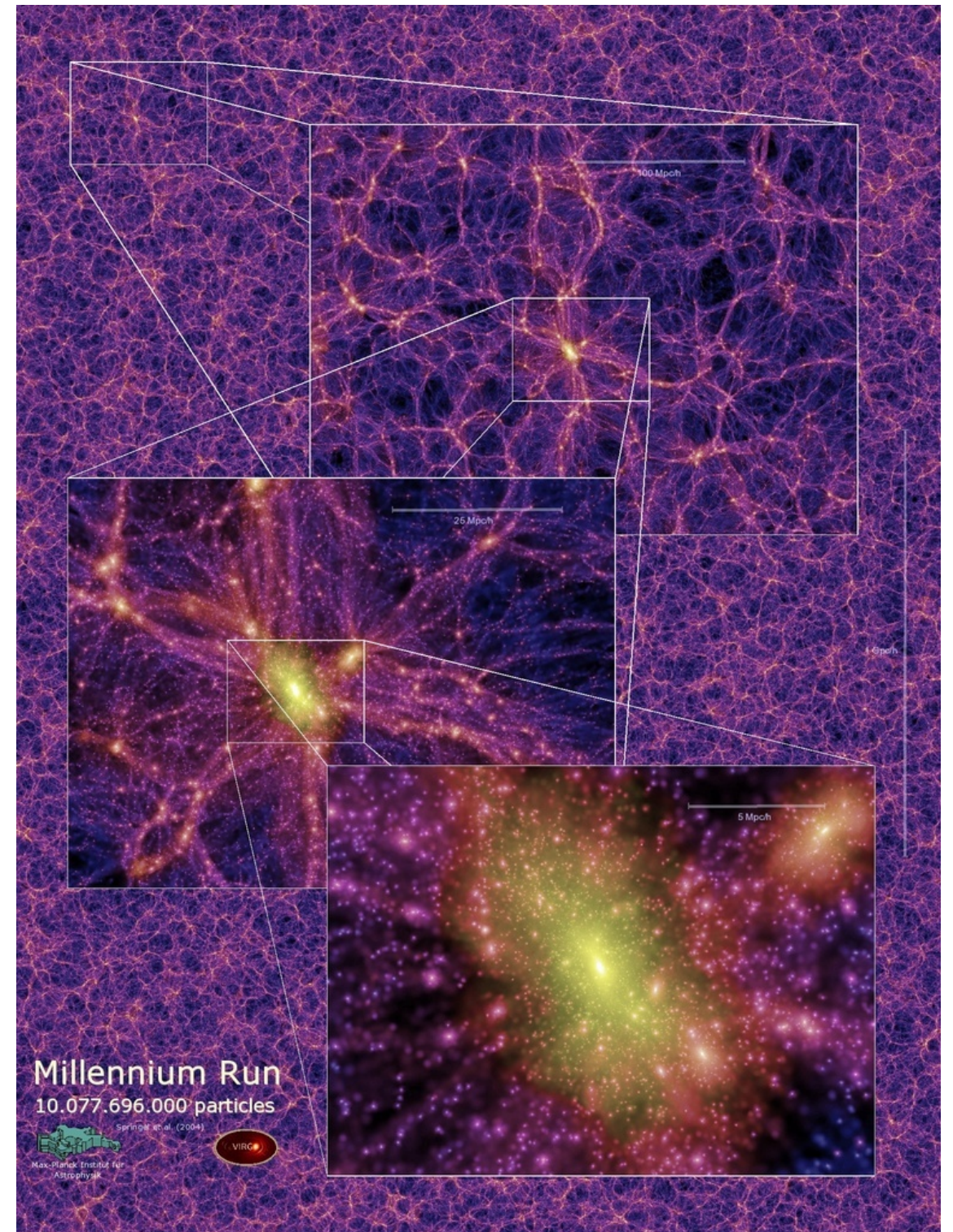
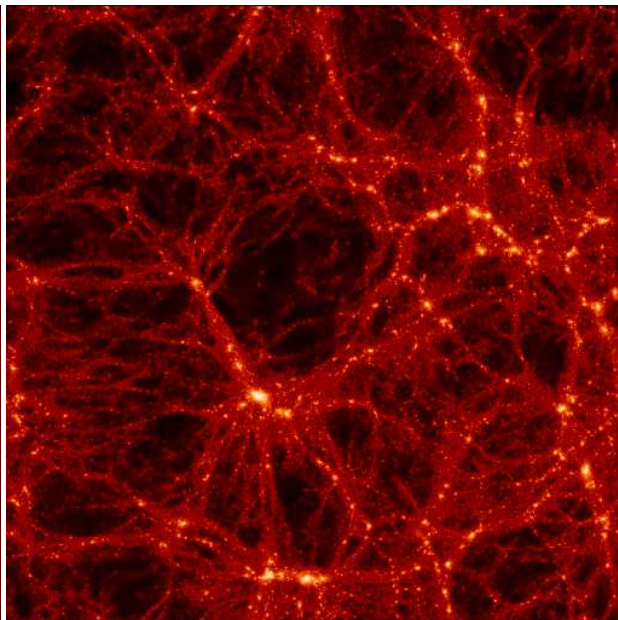
$z=2$



$z=1$



$z=0$



Basic processes

- Hierarchical formation of DM halos (halo merger histories)
- Baryons get **shock heated**
- Hot gas cools and settles in a **disk** in the center of the halo potential well
- Cold gas in disk is transformed into stars (**star formation**)
- Energy output from stars (**feedback**) reheats some of cold gas
- After haloes merge, galaxies sink to center by **dynamical friction**
- After halos merge, galaxies sink to center by **dynamical friction**
- Chemical evolution and stellar population

How to make model predictions

- Full numerical simulations: limited by numerical resolution; many processes, such as star formation, feedback have to be treated by uncertain sub-grid prescriptions. Since many of the details are not well understood, the parameter space is too large to be explored numerically.
- Semi-analytical models: processes are modeled as a set of 'prescriptions' that carry a number of free parameters. The free parameters in the models are then tuned to reproduce certain observational data of the galaxy population.

Some examples of SAM prescriptions

Radiative cooling:

Modeled by cooling function $\Lambda(T, Z)$, $du/dt = \mathcal{C}n_{\text{H}}^2\Lambda$, where u is the internal energy density of the gas.

The cooling timescale:

$$t_{\text{cool}} \equiv \frac{\rho\mathcal{E}}{\mathcal{C}} = \frac{3nk_{\text{B}}T}{2n_{\text{H}}^2\Lambda(T)} \approx 3.3 \times 10^9 \frac{T_6}{n_{-3}\Lambda_{-23}(T)} \text{yr},$$

$\mathcal{E}$: internal energy per unit mass. This should be compared with free-fall time scale of the cloud:

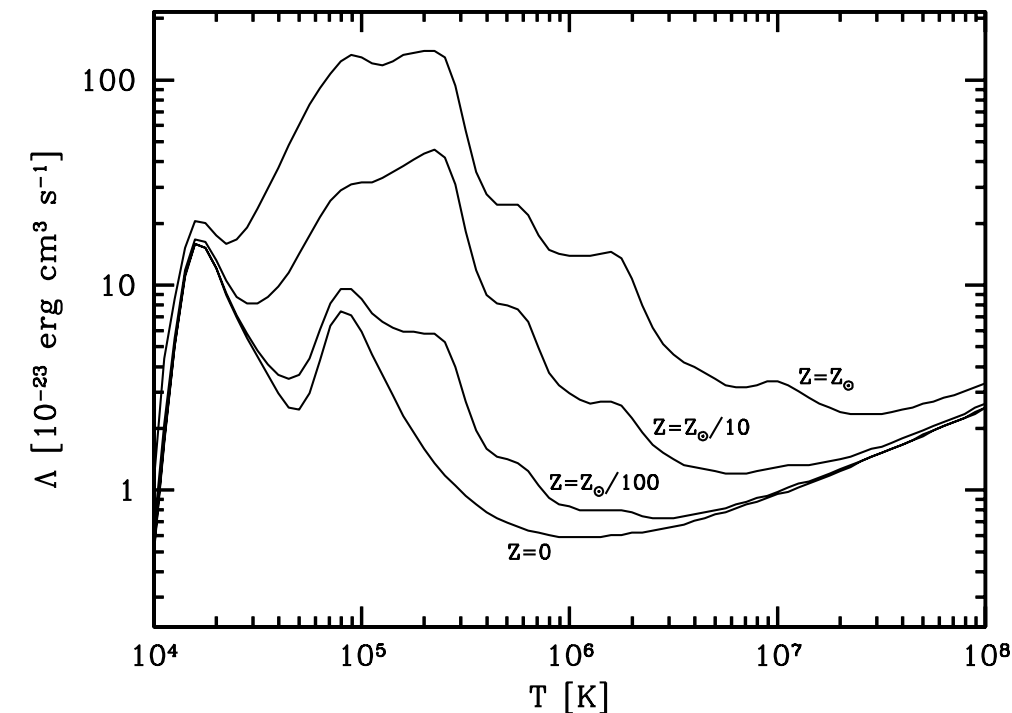
$$t_{\text{ff}} = \sqrt{\frac{3\pi}{32G\rho}} = \sqrt{\frac{3\pi f_{\text{gas}}}{32Gn\mu m_{\text{p}}}} \approx 2.1 \times 10^9 f_{\text{gas}}^{1/2} n_{-3}^{-1/2} \text{yr}.$$

Cooling is effective if $t_{\text{cool}} \ll t_{\text{ff}}$.

Define a cooling radius through $t_{\text{cool}}(r_{\text{cool}}) = t_{\text{age}}$, and a free-fall radius through $t_{\text{ff}}(r_{\text{ff}}) = t_{\text{age}}$, the cold gas accretion rate is

$$\dot{M}_{\text{acc}} = 4\pi\rho_{\text{gas}}(r_{\text{acc}})r_{\text{acc}}^2 \frac{dr_{\text{acc}}}{dt}, \quad r_{\text{acc}} = \min[r_{\text{cool}}, r_{\text{ff}}].$$

Uncertainties in choosing t_{age} and in the gas density profile, ρ_{gas} .



Star formation prescription

Star formation rate in gaseous disks

$$\dot{M}_* = \epsilon_* \frac{M_{\text{cold}}}{\tau_d},$$

where τ_d is the dynamical time scale of the disk, and star formation efficiency is parameterized as

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

where α_{SF} and β_{SF} are parameters and their values are not known a priori. For example, the value of β_{SF} ranges from 0 to 2.5 in early models.

Supernova feedback

The SN energy released by a mass of ΔM_* of star formation

$$E_{\text{fb}} = \alpha_{\text{SN}} \frac{1}{2} \Delta M_* V_{\text{SN}}^2 ,$$

where $V_{\text{SN}} = 630 \text{ km/s}$ and α_{SN} describes the feedback efficiency.

The effects of feedback:

- (i) heating the ISM to hot halo gas;
- (ii) ejecting the ISM from the halo;
- (iii) driving out hot wind from halo.

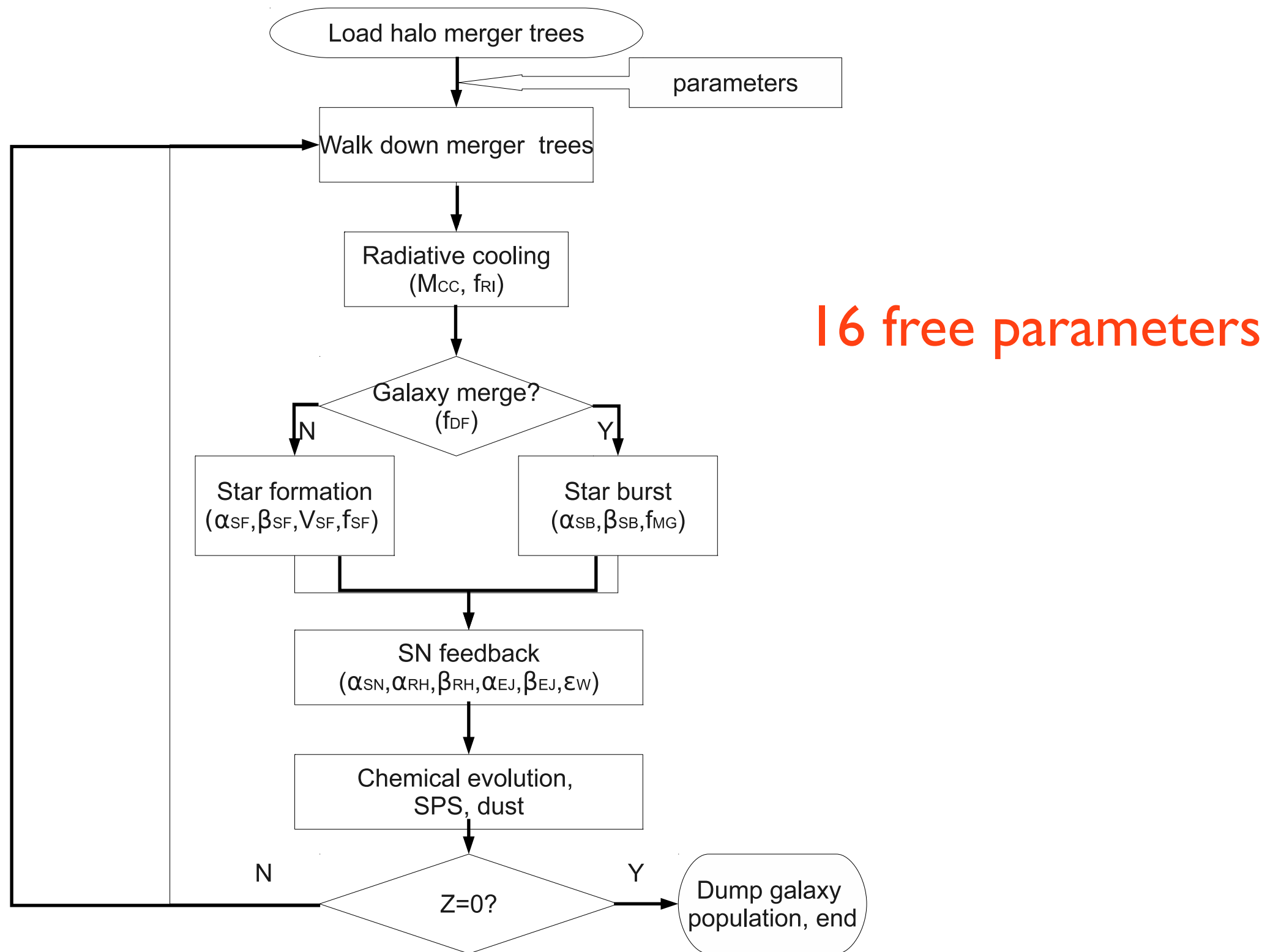
$$E_{\text{fb}} = \frac{1}{2} (1 - f_{\text{ej}}) f_{\text{rh}} \Delta M_* V_{\text{vir}}^2 + \frac{1}{2} f_{\text{ej}} f_{\text{rh}} \Delta M_* v_{\text{esc}}^2 + \frac{1}{2} \Delta M_{\text{wind}} v_{\text{esc}}^2 .$$

$$f_{\text{rh}} = \alpha_{\text{RH}} \left(\frac{V_0}{V_{\text{vir}}} \right)^{\beta_{\text{RH}}} ; \quad f_{\text{ej}} = \alpha_{\text{EJ}} \left(\frac{V_0}{v_{\text{esc}}} \right)^{\beta_{\text{EJ}}} .$$

$$\Delta M_{\text{wind}} = \epsilon_{\text{W}} \Delta M_* \left\{ \alpha_{\text{SN}} \left(\frac{V_{\text{SN}}}{v_{\text{esc}}} \right)^2 - f_{\text{rh}} \left[\left(\frac{V_{\text{vir}}}{v_{\text{esc}}} \right)^2 + f_{\text{ej}} \right] \right\} .$$

All the model parameters are quite uncertain.

A fiducial model family



How to use SAM to study galaxy formation?

- Early SAM applications: a subset of model parameters is held fixed while other parameters are adjusted to match some observational properties; the goodness of fit is often assessed ``by eye".
- Problems with this approach:
 - (i) many model parameters are uncertain, fixing parameters is equivalent to imposing unsubstantiated priors;
 - (ii) no statistically rigorous inference can be made about models.
- Given the uncertainties in models, and the large amounts of observational data, galaxy formation is best studied with the Bayesian inference approach.

The Bayesian Approach

Bayesian Theorem: The posterior probability of a set of parameters Θ in a model (or hypothesis) M for given data D is

$$P(\Theta|D, M) \propto P(\Theta|M)L(D|\Theta, M),$$

where $P(\Theta|M)$ is the prior probability distribution describing the knowledge about the parameters acquired before seeing the data, and $L(D|\Theta, M)$ is the likelihood of data D for the given model parameter set Θ .

This not only allows to derive posterior distributions of model parameters for a given model family, it also allows to compare different model families using Bayesian evidence.

Bayesian Evidence

The normalization of the posterior,

$$P(D|M) = \int P(\Theta|M)L(D|\Theta, M)d\Theta$$

is the total likelihood of the data for the given model. The probability of the model for the given data is

$$P(M|D) = \frac{P(M)P(D|M)}{P(D)}.$$

Thus, for two model families M_1 and M_2 , their relative odd for the given data is

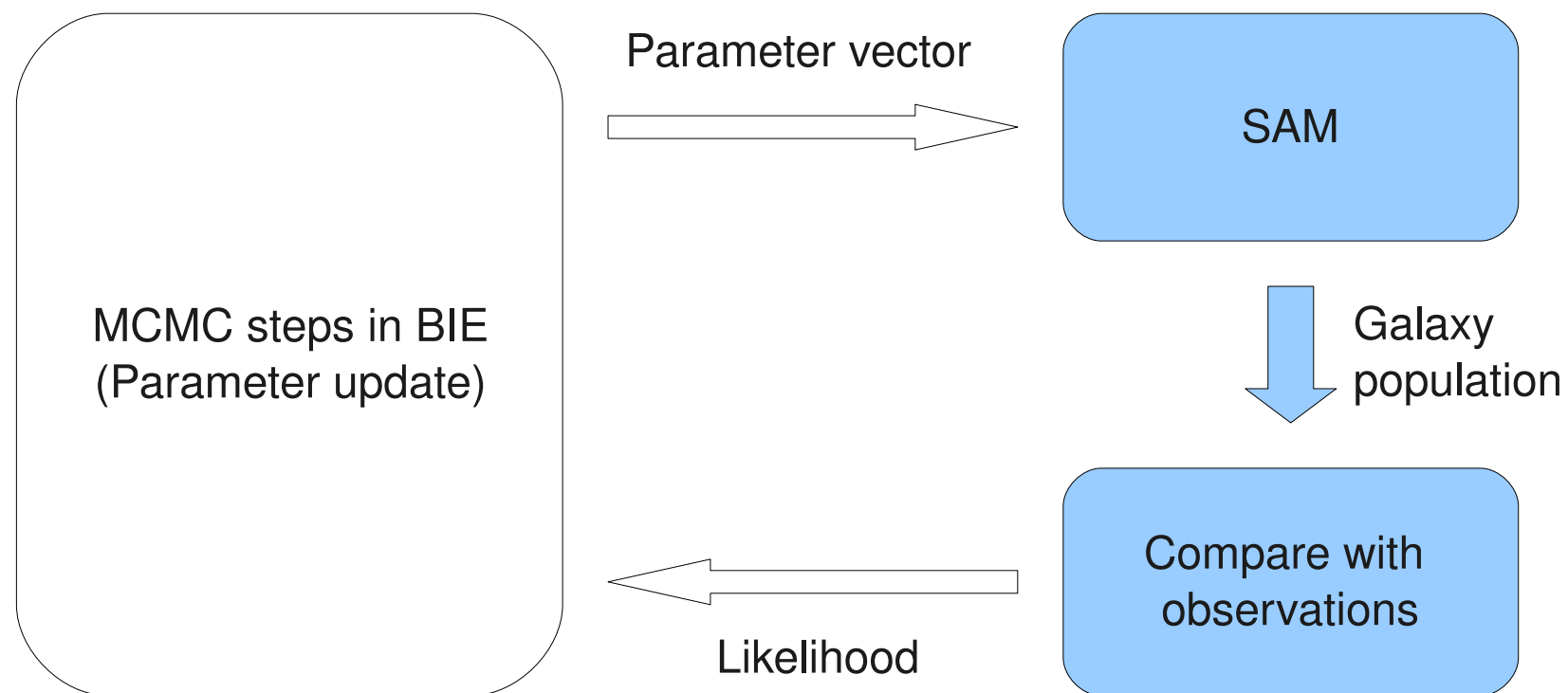
$$\frac{P(M_1|D)}{P(M_2|D)} = \frac{P(M_1)P(D|M_1)}{P(M_2)P(D|M_2)}$$

The ratio $P(M_1)/P(M_2)$ is the relative prior probability, and $P(D|M_1)/P(D|M_2)$ is the Bayesian evidence and can be obtained from the posterior distribution.

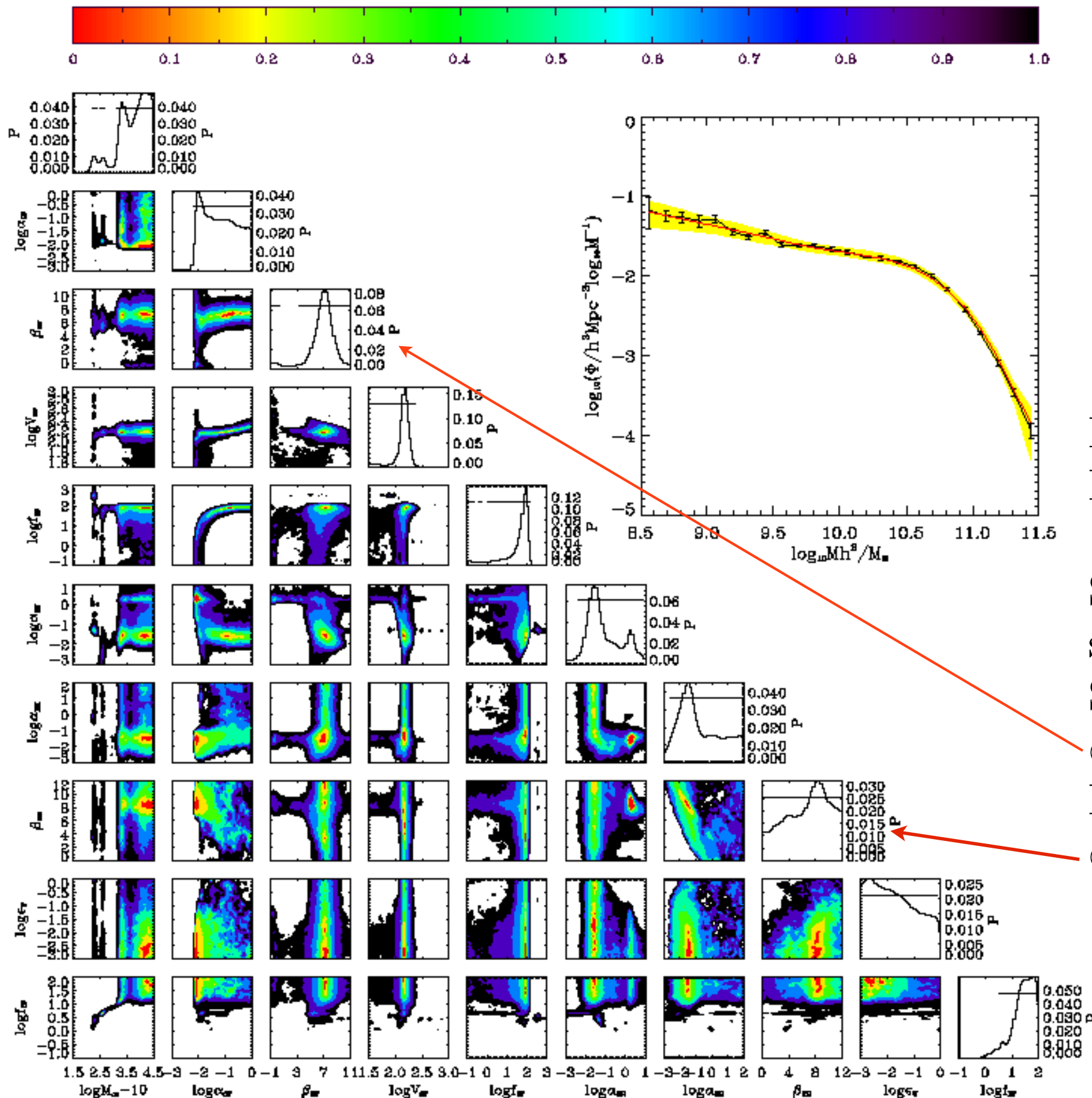
The Markov-Chain Monte-Carlo (MCMC)

- For our problem, it is impossible to derive the posterior probability distribution analytically, A MCMC approach is adopted to sample the posterior.
- Bayesian Inference Engine, a software package, developed at UMass, which includes advanced MCMC algorithms (differential evolution, tempering scheme, etc) and supports parallel computation.

The Bayesian SAM



Inference from galaxy stellar mass function



Posterior distribution complex
Degeneracy among parameters
(e.g. $\alpha_{\text{SF}} - f_{\text{SF}}$)

Some parameters extreme,
suggesting new physics:

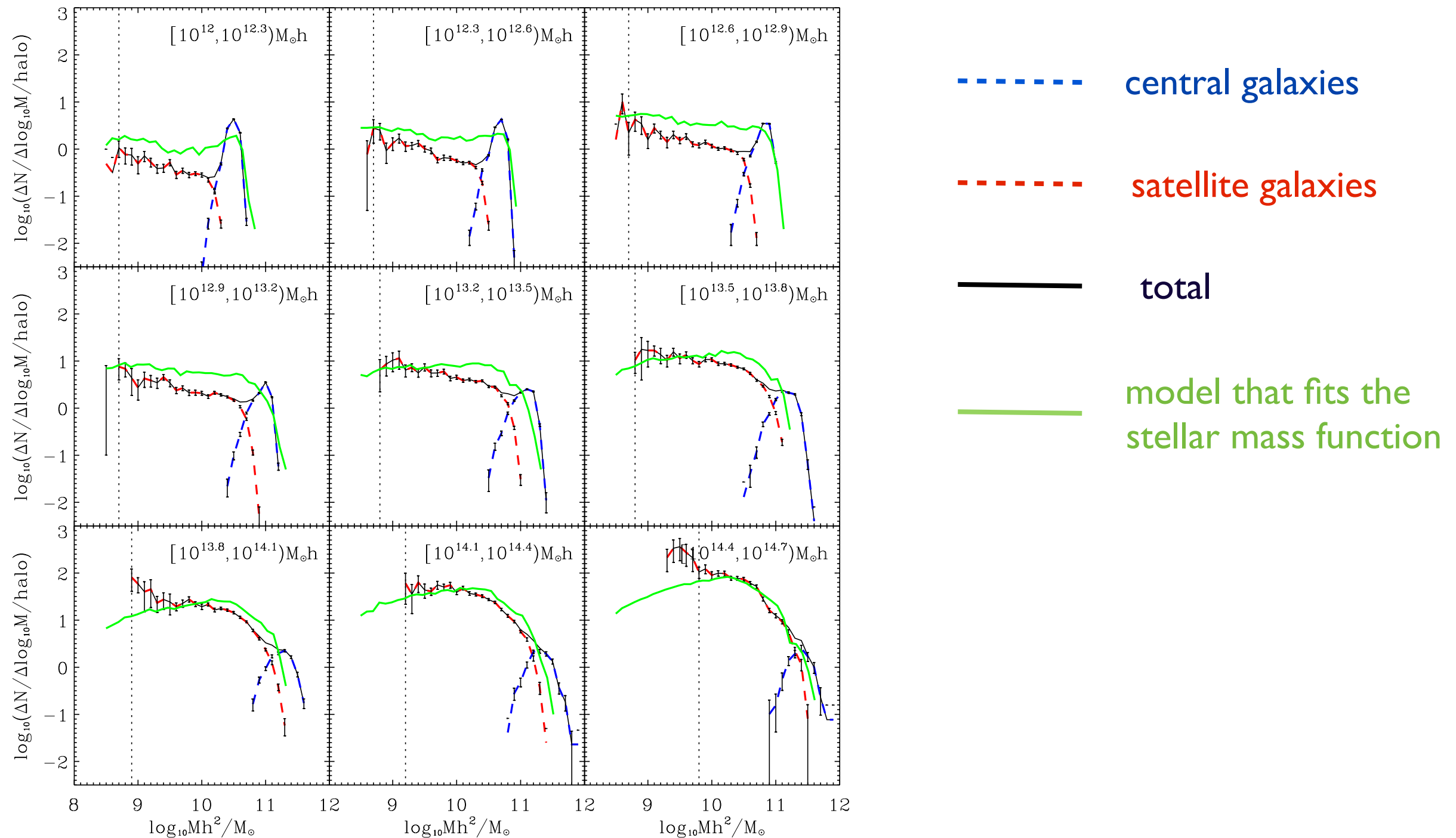
Star formation efficiency,

$$\propto V_{\text{vir}}^{\beta_{\text{SF}}}, \beta_{\text{SF}} \sim 7;$$

Feedback efficiency,

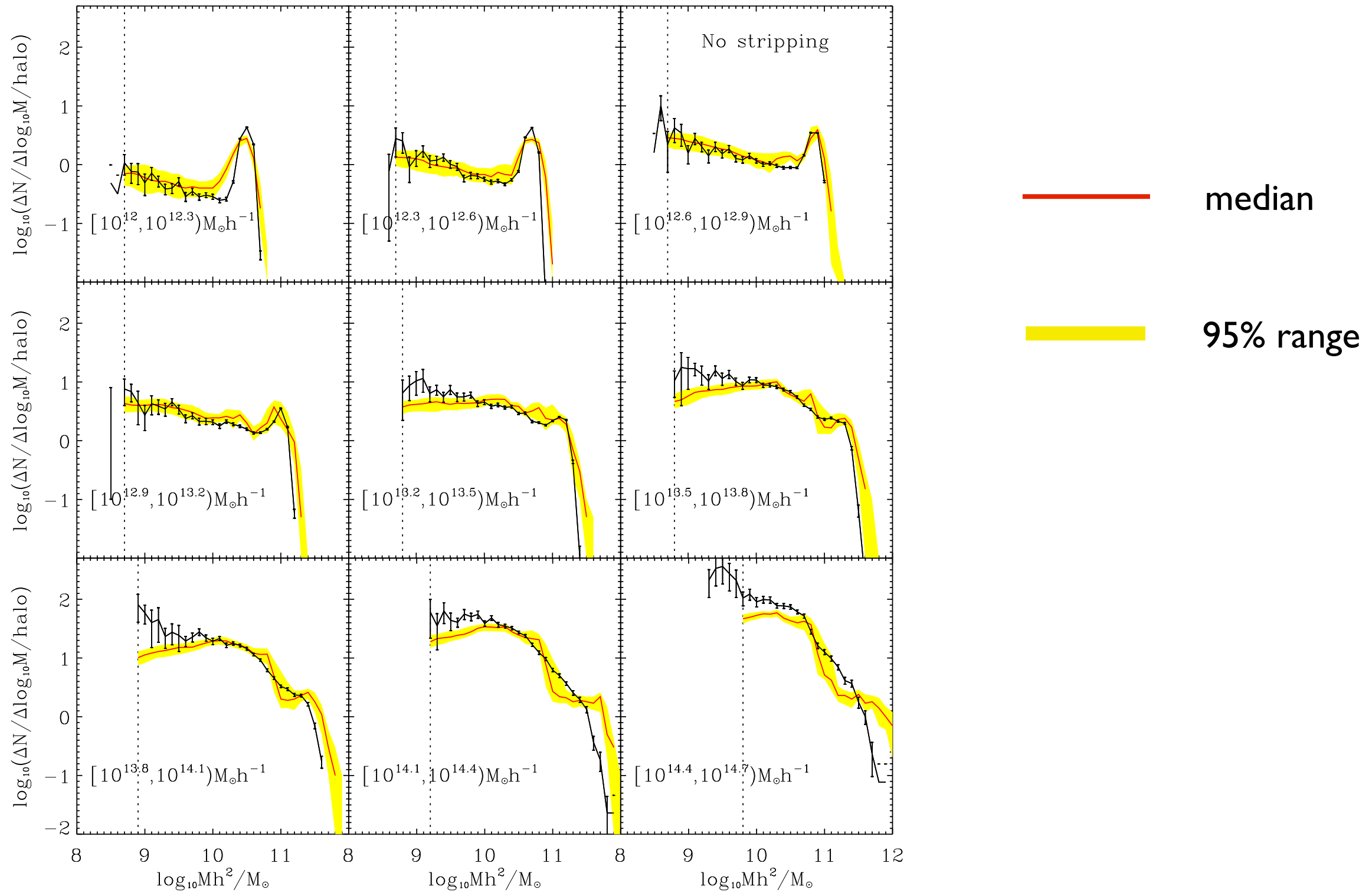
$$\propto V_{\text{vir}}^{-\beta_{\text{RH}}}, \beta_{\text{RH}} \sim 8.$$

Inference from conditional stellar mass function

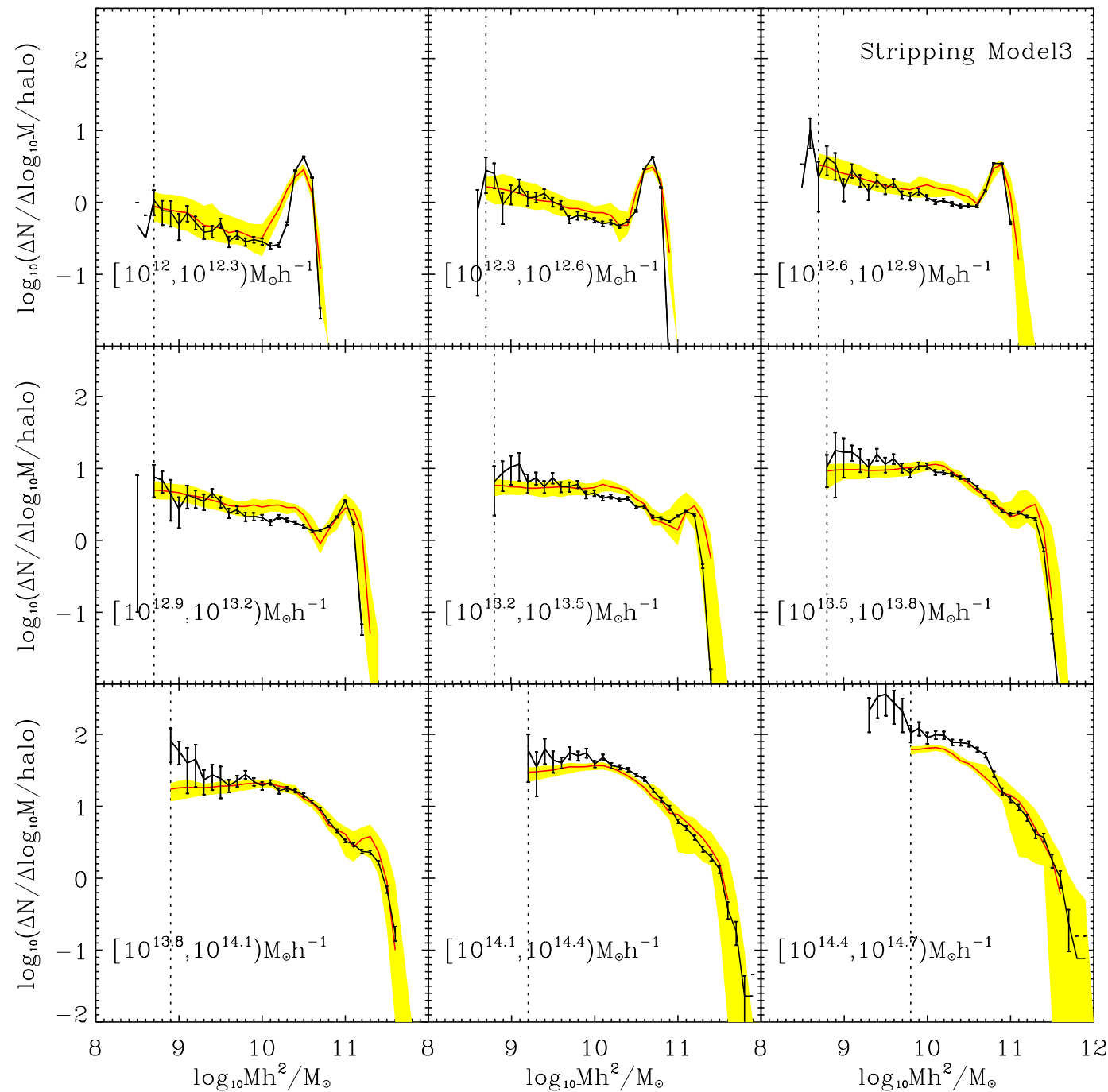


Data: Yang et al (2009)

The fiducial model family



Model including tidal stripping



Bayesian evidence

This model family:

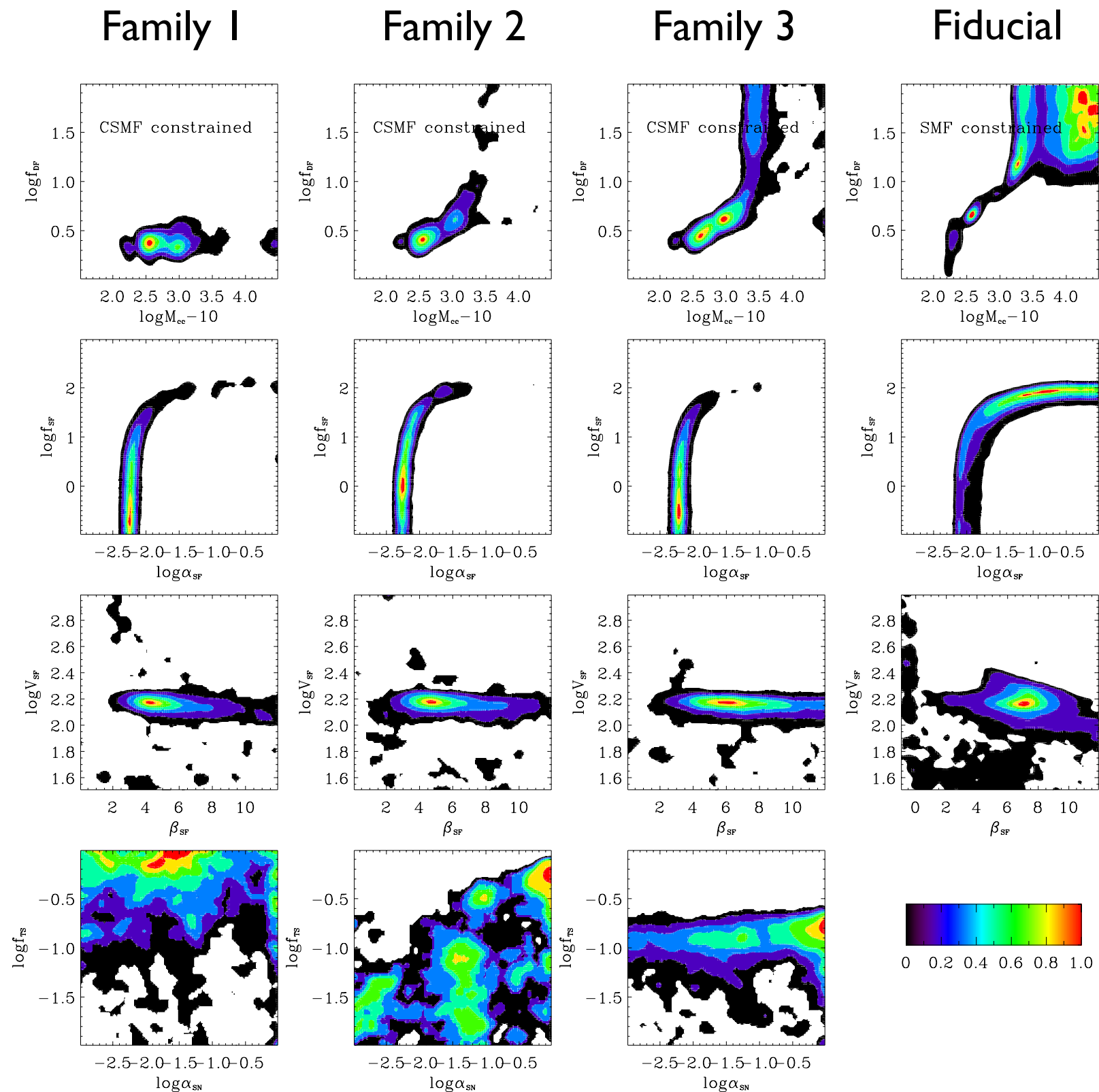
$$\log P(D|M) = -45.6^{+3.40}_{-7.48};$$

Fiducial family (no stripping):

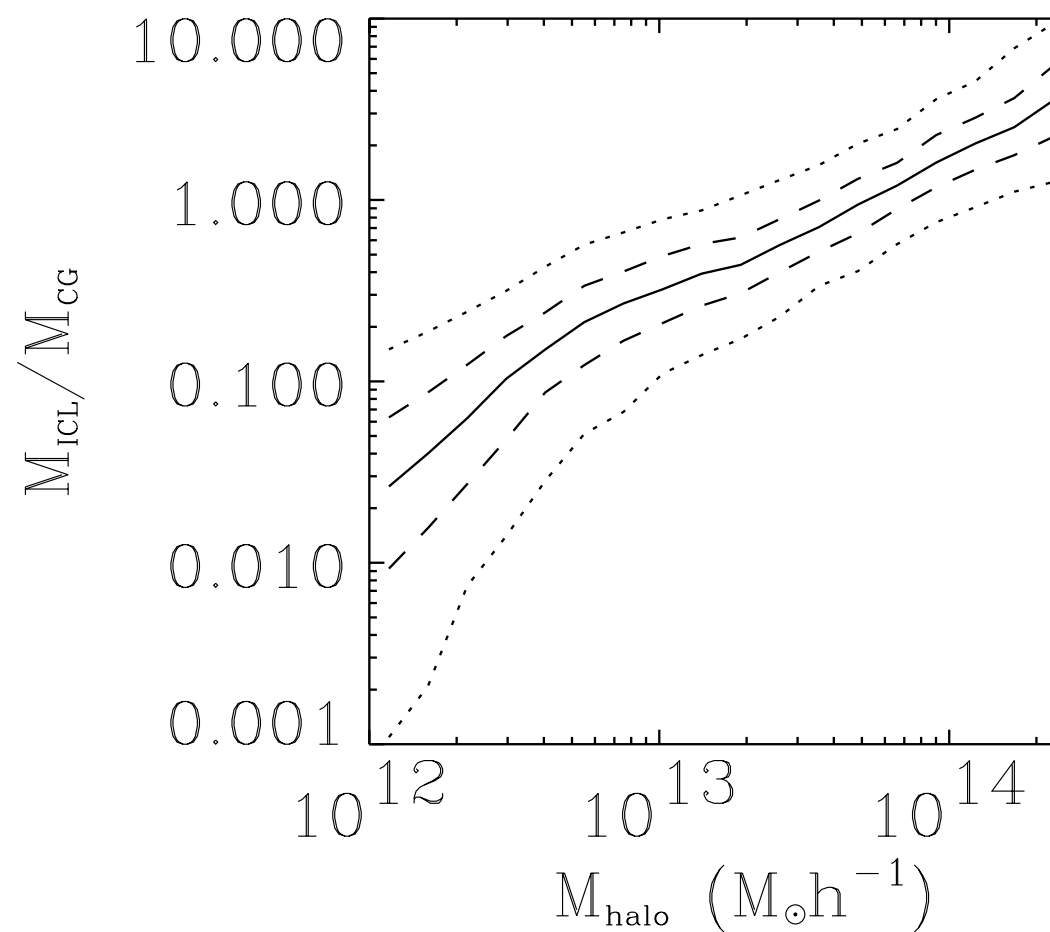
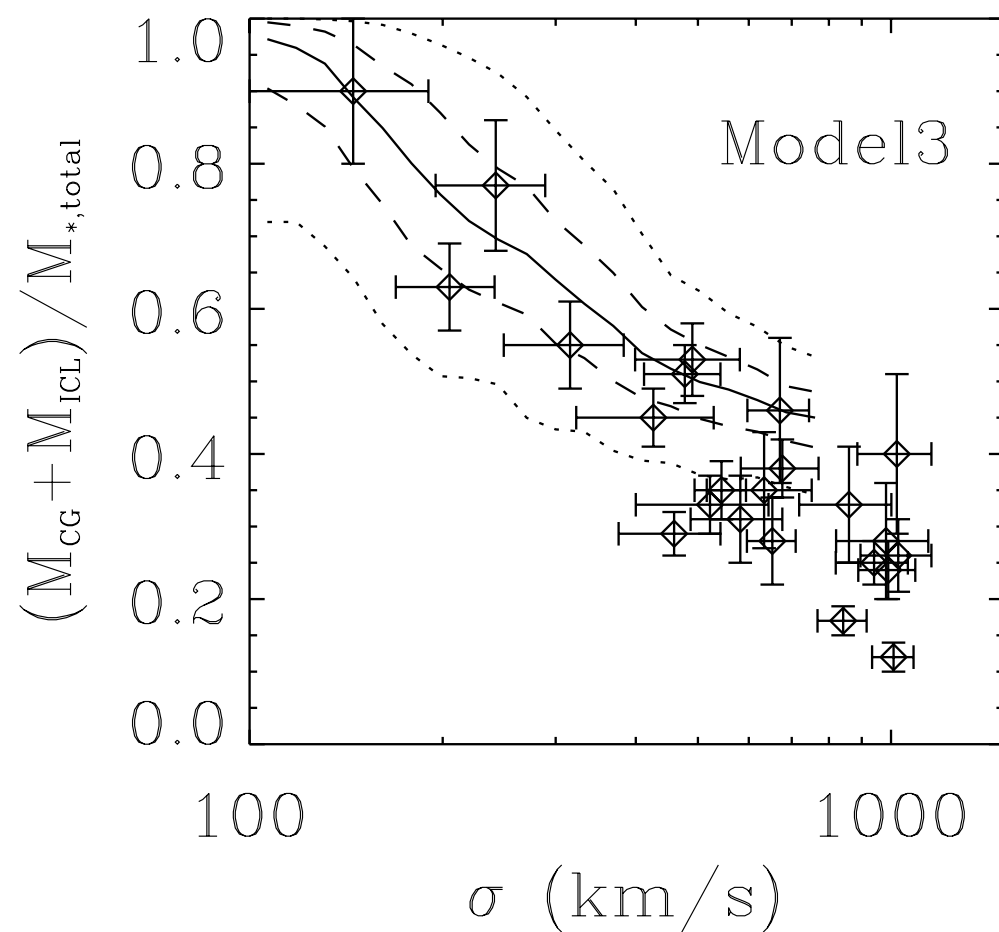
$$\log P(D|M) = -69.15^{+2.19}_{-0.89}.$$

Stripping fraction: $df_s = \alpha_{TS} \frac{dt}{t_{\text{dyn}}}$

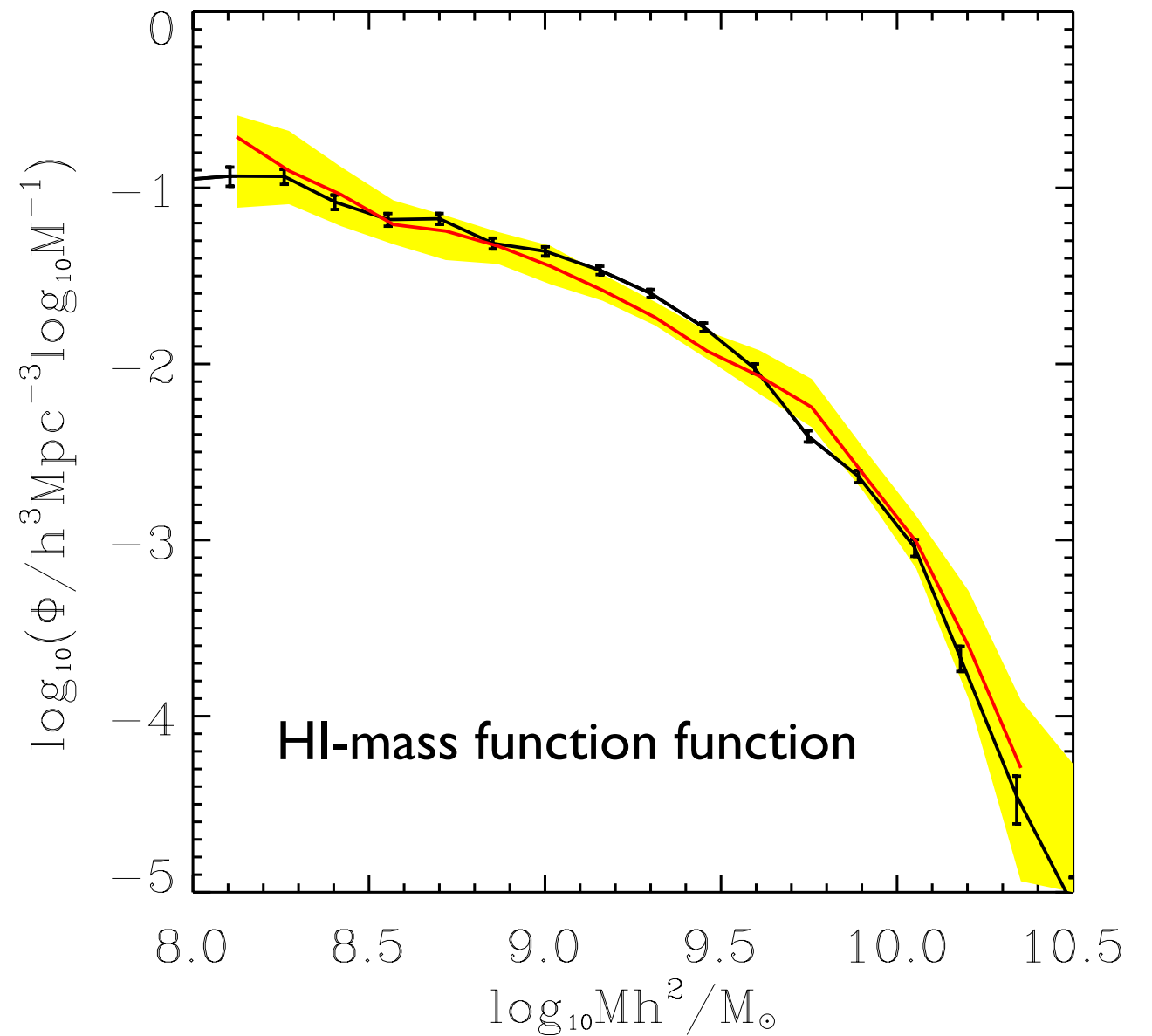
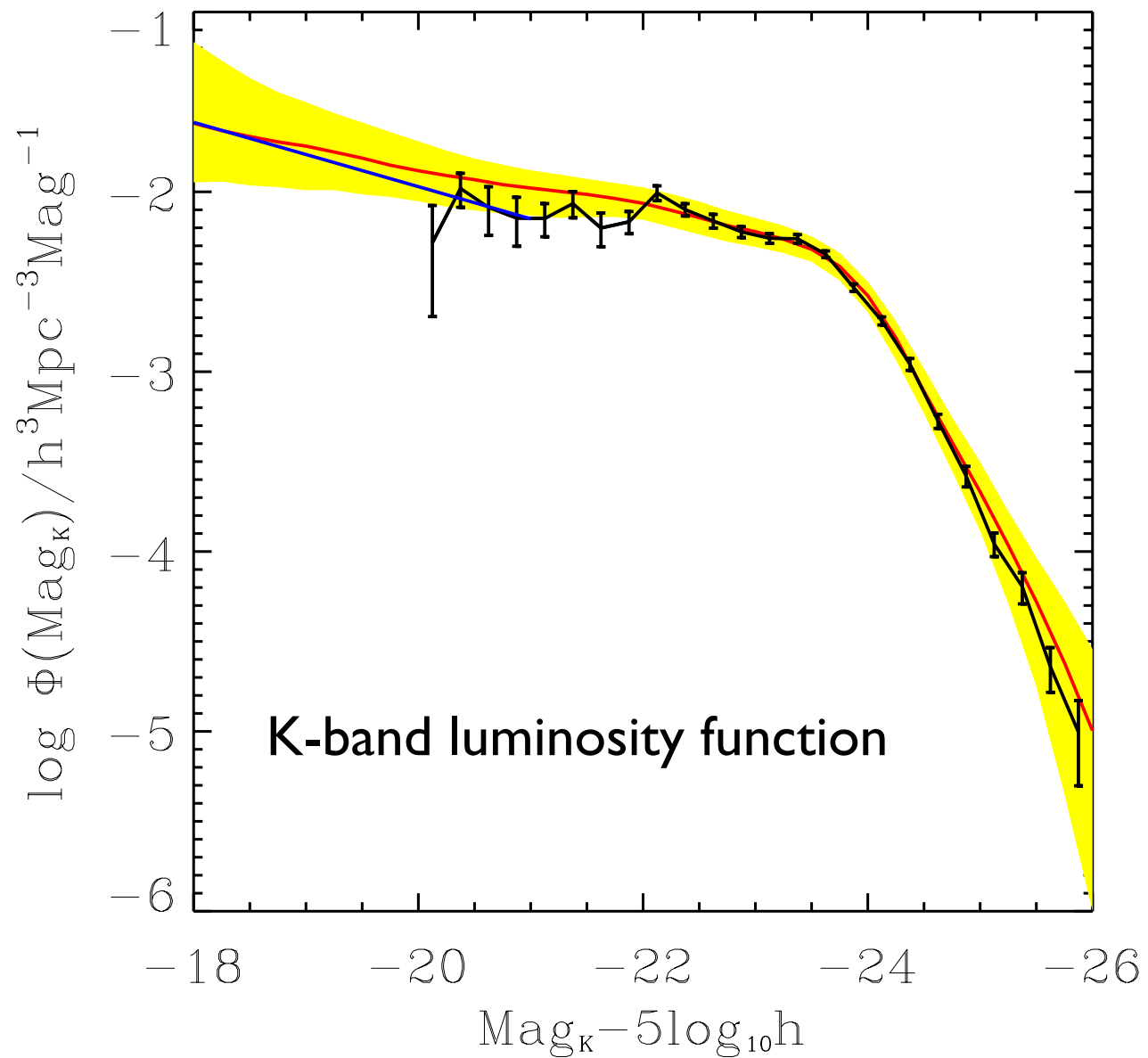
Posterior distribution



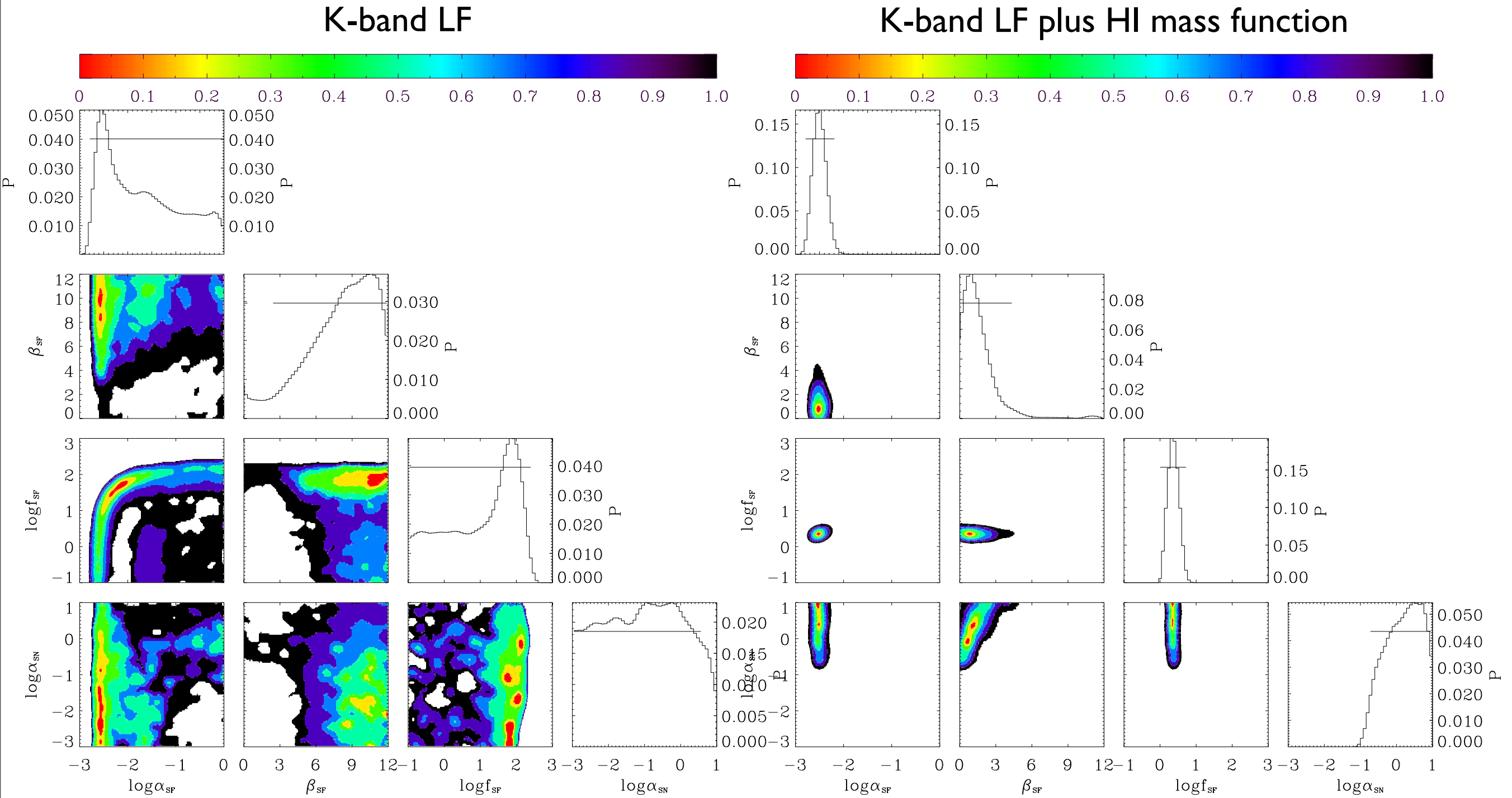
Intra-cluster stars



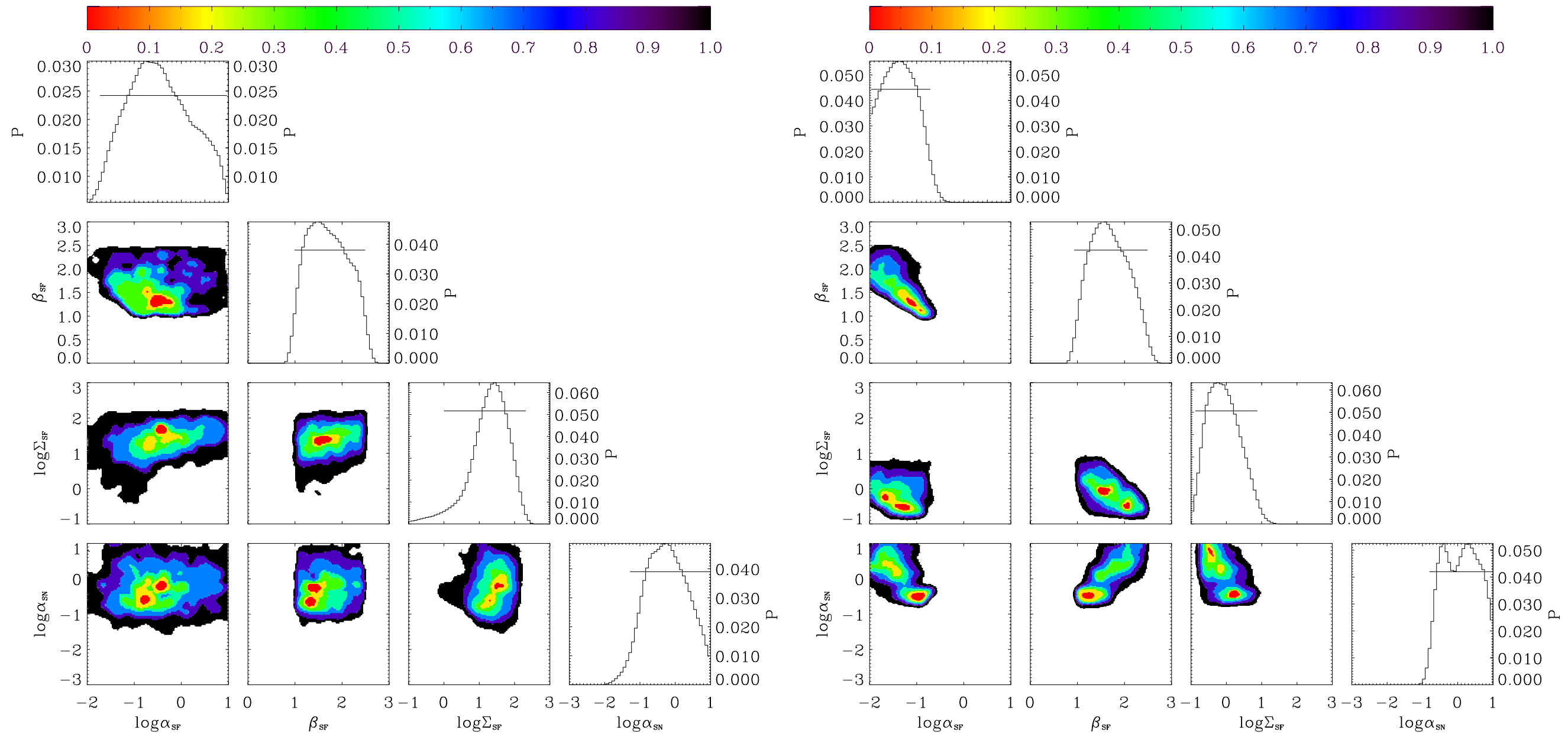
Inference from HI mass function



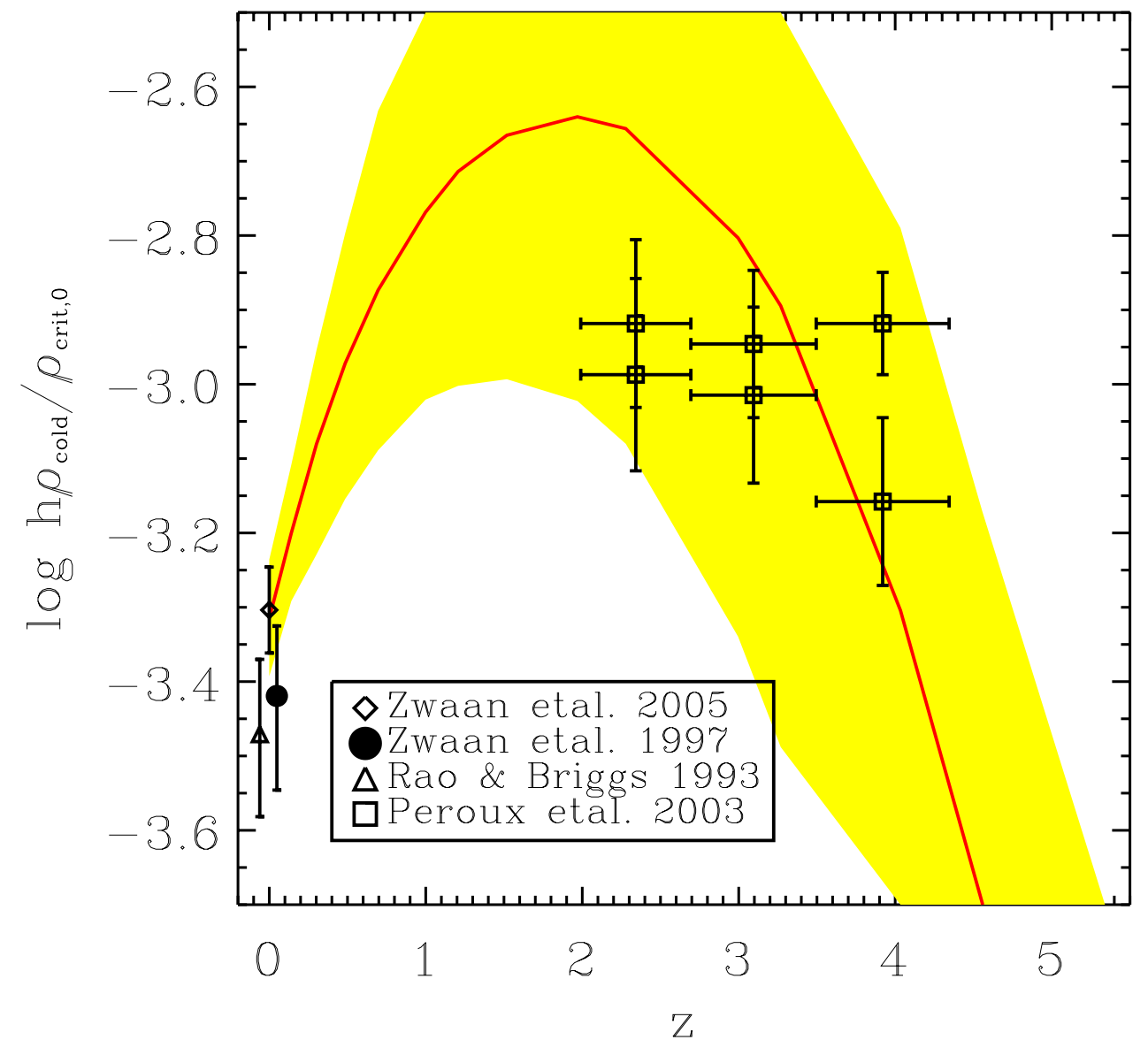
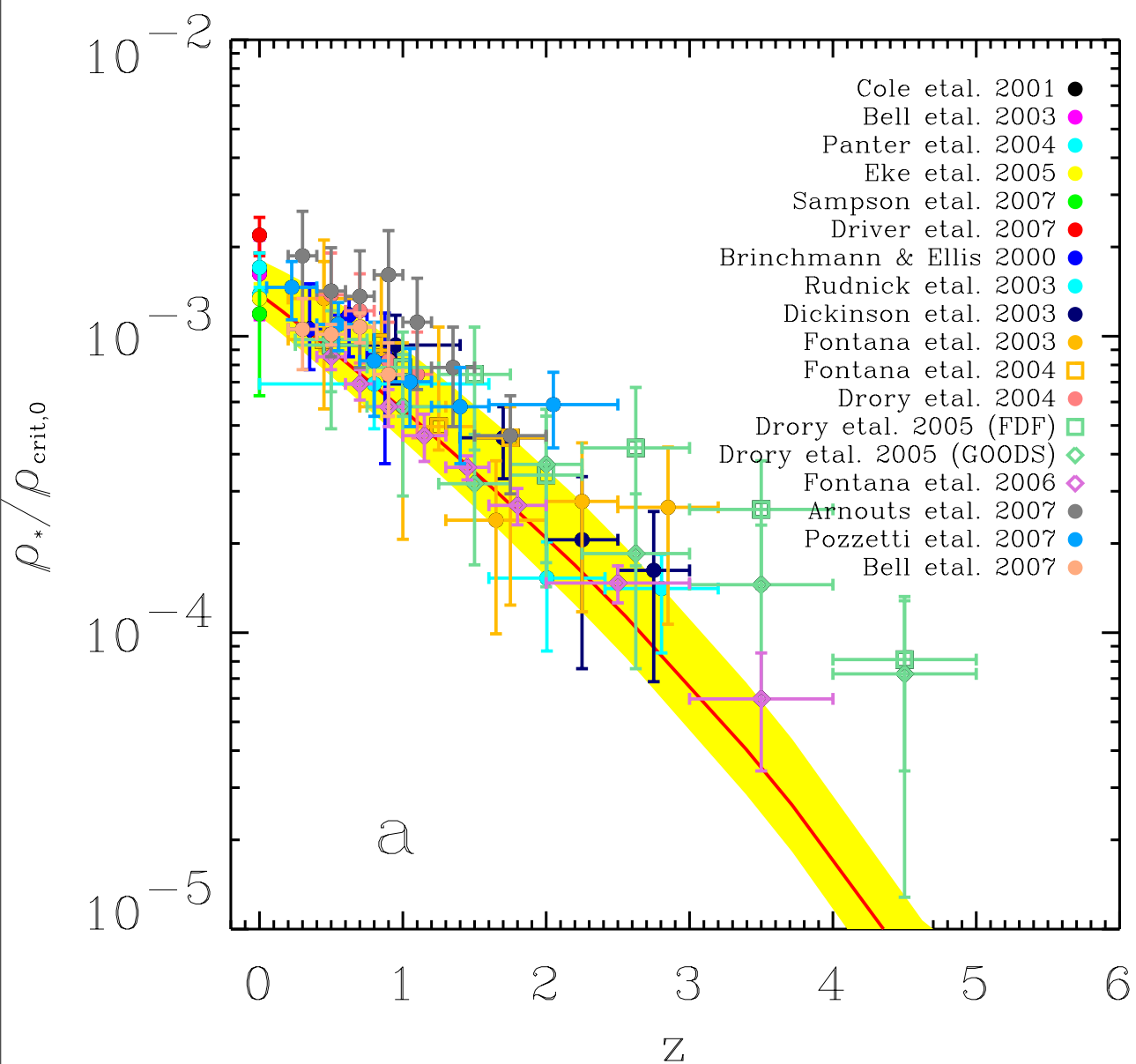
Posterior Distribution



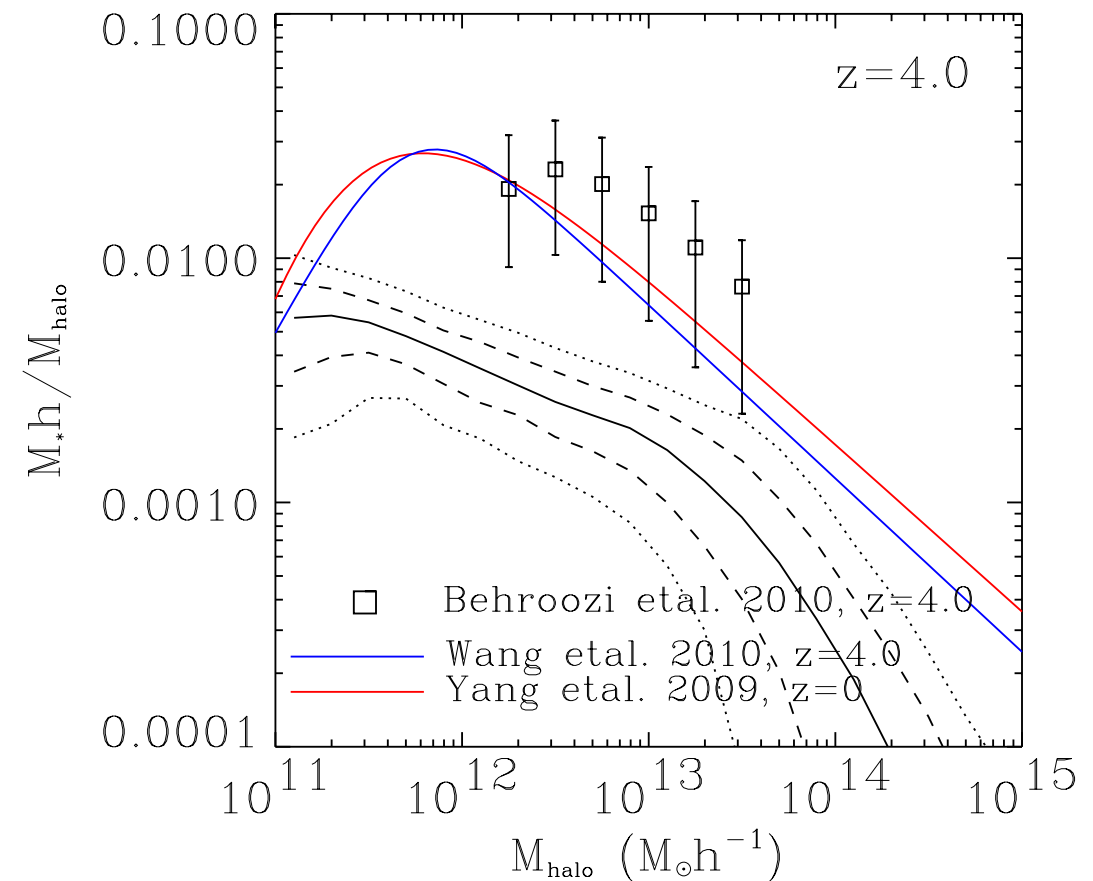
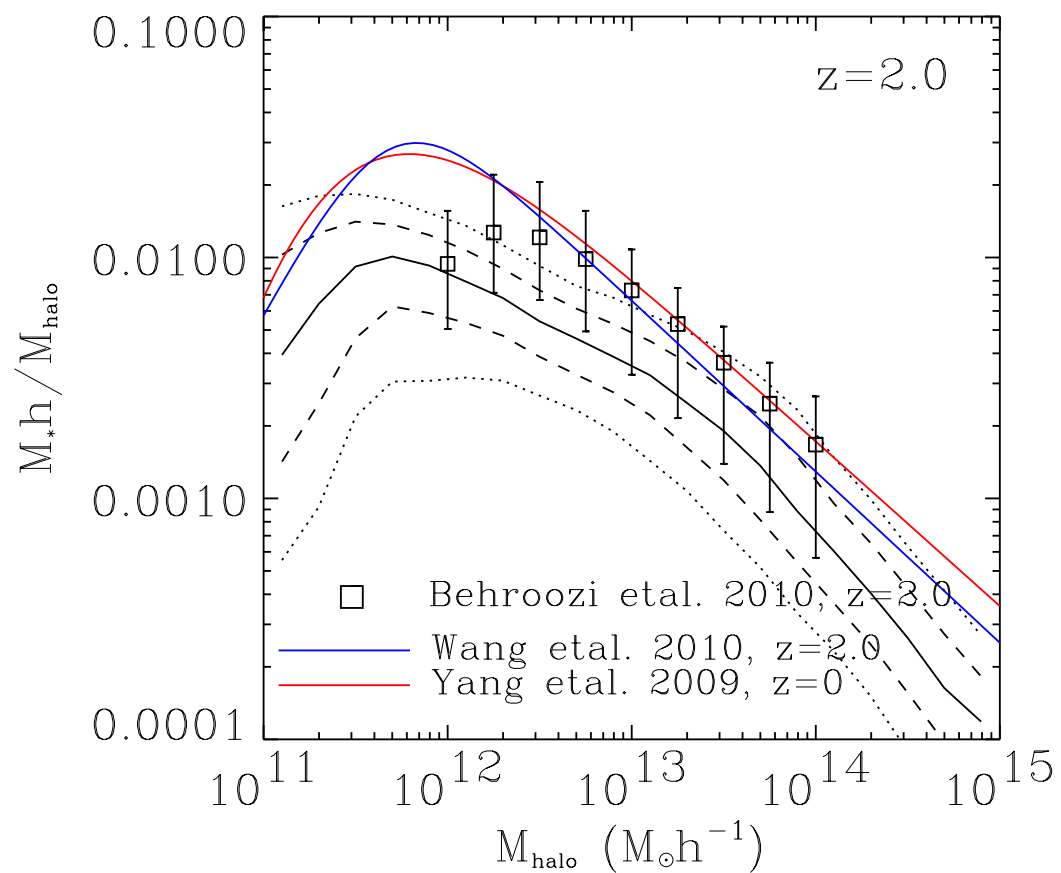
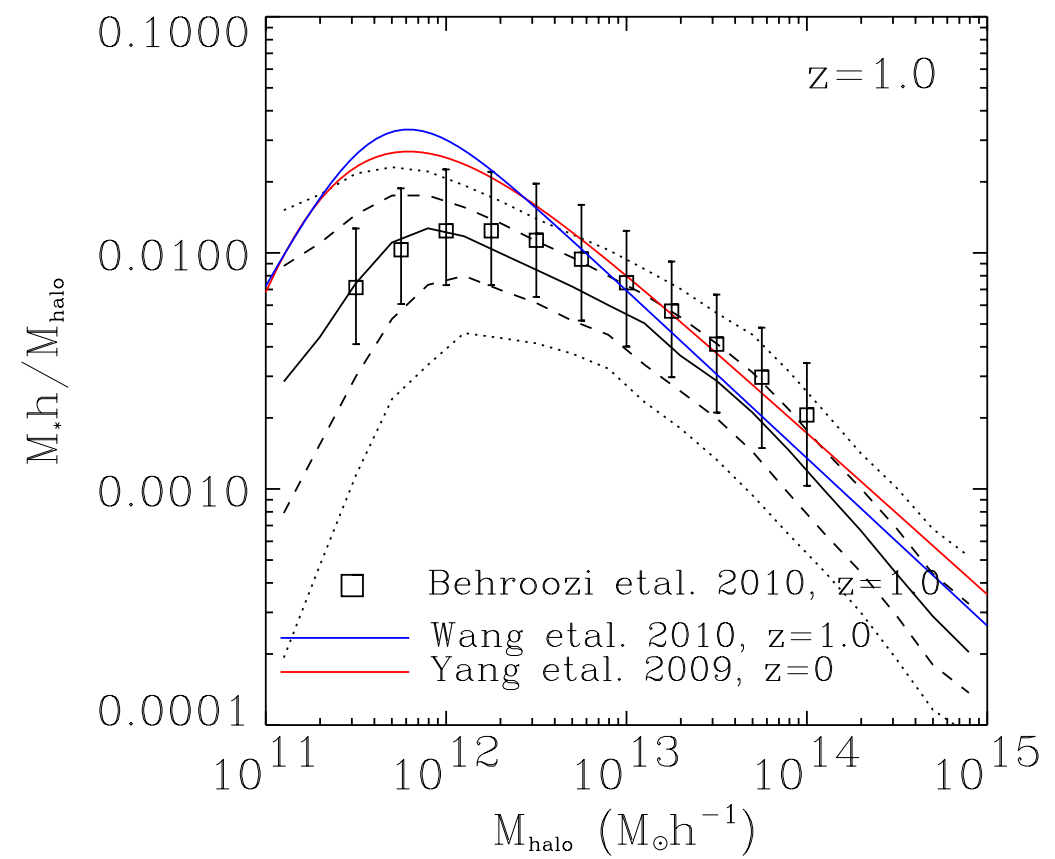
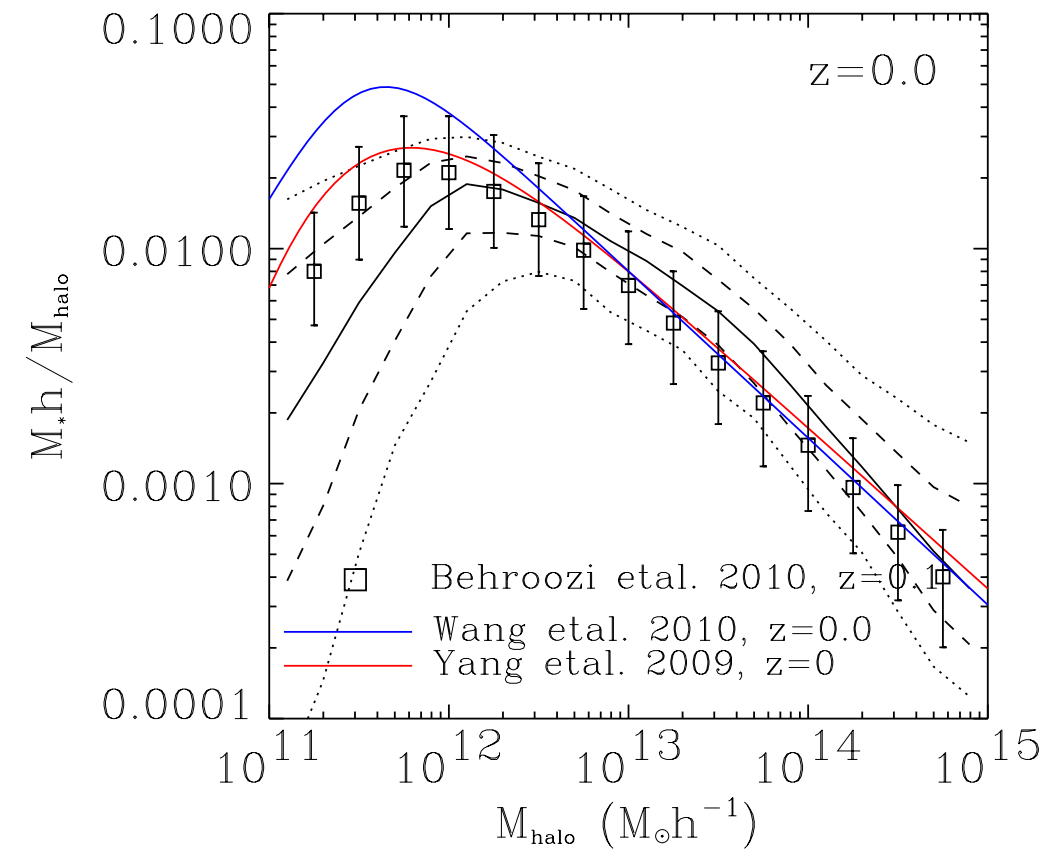
Another model family



Model predictions for high-z galaxies



Stellar mass - halo mass relation



Conclusions

- Many physical processes affecting galaxy formation are not yet well understood while copious observational data are available to constrain models. The problem is best tackled with the Bayesian inference approach.
- It is possible to develop a SAM in a rigorous Bayesian inference frame (needs advanced MCMC algorithm and powerful computation).
- Bayesian inference-based SAM is very powerful in constraining galaxy formation models with observational data in a statistically rigorous way.
- It is ideal in dealing with a variety of data constraints, because it explicitly builds on previous results by incorporating the constraints from previous inferences into new data sets.
- Bayesian evidence allows one to discriminate models.
- The Bayesian approach has the ability to probe the tension between data and model, helping identify missing physics.
- For a given model family, the posterior distribution of model parameters obtained for a given set of data can be used to predict observables that include the inferential uncertainties. Such predictions can be used to assess the power of new observations.