Galaxy Formation: Bayesian SAMs Methods

Neal Katz
UMass Astronomy

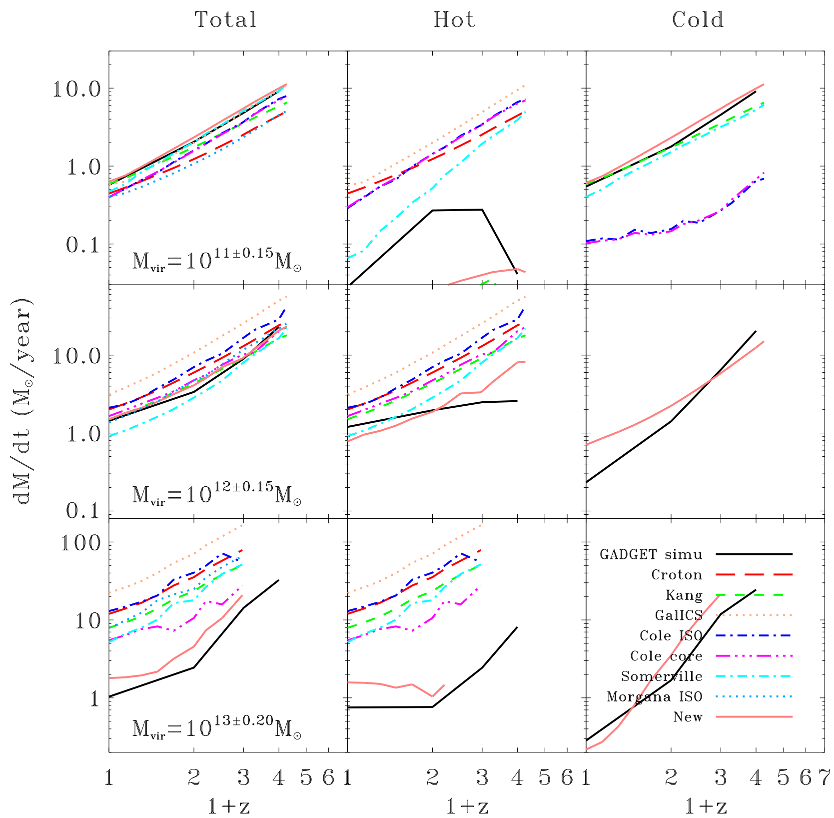
Collaborators: Yu Lu, Houjun Mo, Martin Weinberg

September, 2013

A Simple SAM

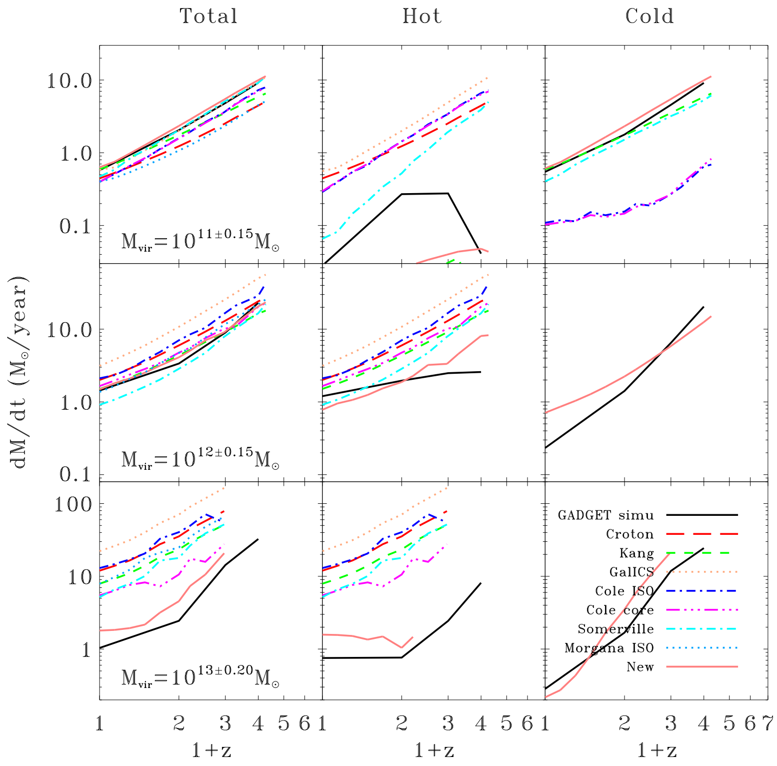
- Create a simple SAM that encompasses most published SAMs.
- Assume a cosmology and dark matter type.
  - Six hidden parameters: Ω_m, Λ,H_0,σ_8,n, & Ω_b.
  - Determines power spectrum.
- Begin with dark matter merger trees.
  - Can be extracted from N-body simulations.
  - Can be generated by a Monte-Carlo method using extended Press-Schechter.
- Next model the radiative cooling of the gas.
  - Stop radiative cooling above a halo mass of M_{cc} to mock up "Velvet Rope" feedback, e.g. AGN

There are Many Ways to Be Cool



- Different methods give quite different results.
- Accretion is either hot or cold; no mixed accretion.
- Developed new model that has mixed accretion.

There are Many Ways to Be Cool



- Compared to simulations:
  - Total cooling rates too small in low mass, cold mode halos.
  - Total cooling rates are too high in higher mass, hot mode halos.
- In the work presented here assume Croton cooling law.

SAM Star Formation

- Assume the cold gas forms an exponential disk with scalelength $r_{disk} = 0.035 r_{vir}$.
- Stars only form in gas, $m_{\rm SF}$, above density threshold of $f_{\rm SF}\ {\rm M}_\odot\ {\rm pc}^{-2}$.
- SFR inversely proportional to disk dynamical time:

$$\tau_{disk} = \frac{r_{disk}}{v_{vir}},$$

$$\dot{m}_* = \epsilon_* \frac{m_{\rm SF}}{\tau_{disk}},$$

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

SAM Star Formation Feedback

- We assume that a fraction α_{SN} of the supernova energy can either reheat the gas or make a wind.
- The mass of reheated gas is $f_{\mathrm{rh}} \Delta t \dot{m}_*$.

$$f_{\mathrm{rh}} = \alpha_{\mathrm{RH}} \left(\frac{220\,\mathrm{km/s}}{v_{vir}} \right)^{\beta_{\mathrm{RH}}},$$

$$f_{\mathrm{rh,max}} = \alpha_{\mathrm{SN}} \left(\frac{V_{\mathrm{SN}}}{v_{vir}} \right)^2,$$

- The mass of gas in the wind is

$$m_{\mathrm{wind}} = \varepsilon_{\mathrm{W}} \Delta t \dot{m}_* \left\{ \alpha_{\mathrm{SN}} \left(\frac{V_{\mathrm{SN}}}{v_{\mathrm{esc}}} \right)^2 - f_{\mathrm{rh}} \left[\left(\frac{v_{vir}}{v_{\mathrm{esc}}} \right)^2 \right] \right\}.$$

- A wind fraction of f_{RI} returns as hot gas on a dynamical time.

SAM Mergers

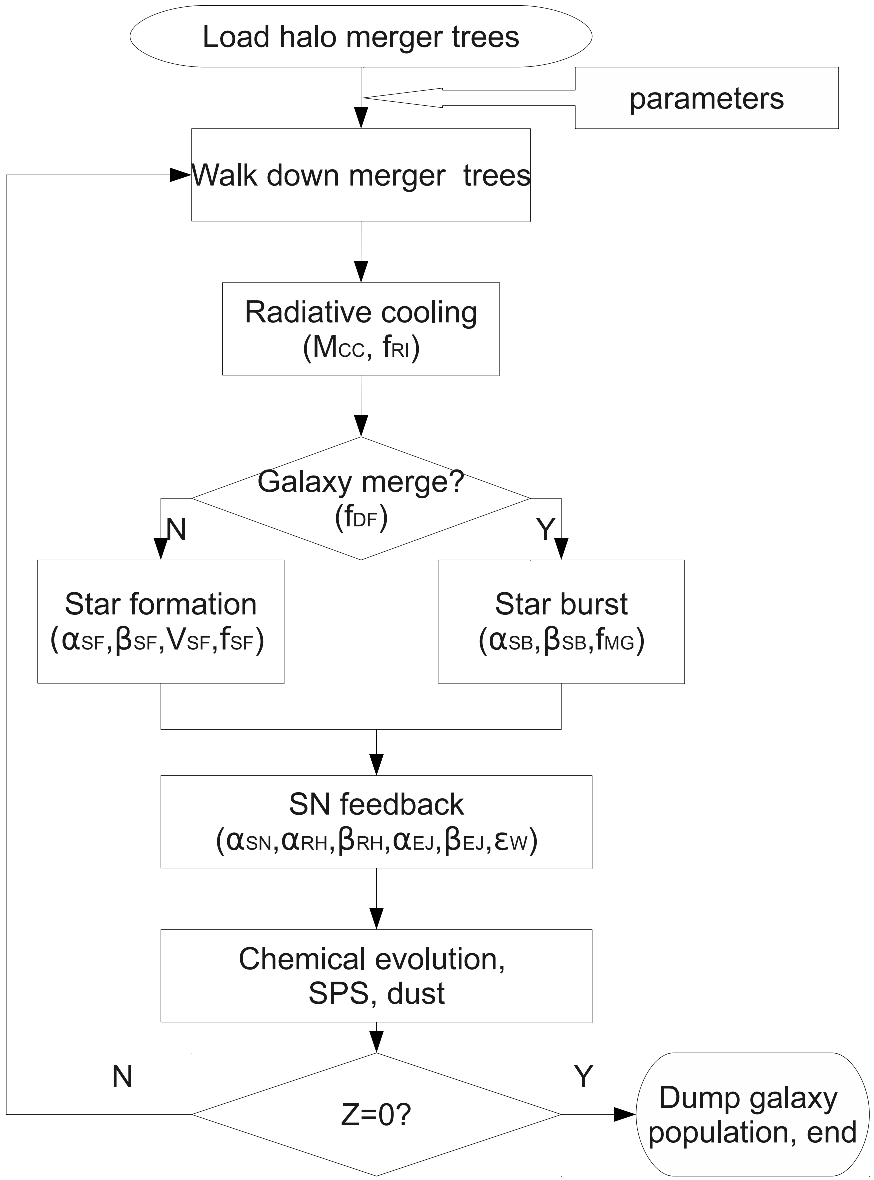
- When dark halos merge the smaller central galaxy becomes a satellite of the larger central along with all previous satellites of both galaxies.
- The satellites start at r_{vir} and sink by dynamical friction with a time scale

$$t_{\rm fric} = f_{\rm DF} \frac{1.17 r_{vir}^2 v_{vir}}{\ln(1 + M_{vir}/M_{sat}) G M_{sat}}.$$

- When galaxies merge a starburst occurs consuming a fraction $e_{\rm burst}$ of the combined cold gas,

$$e_{\rm burst} = \alpha_{\rm burst} \left(\frac{M_{sat}}{M_{central}} \right)^{\beta_{\rm burst}}.$$

SAM Flowchart



Play it SAM: the Old Song

- Usually the SAM is adjusted "to match" a fundamental data set by adjusting the parameters "by hand' and the fit is judged "by eye."
- This is not probabilistically rigorous; there are no confidence intervals for the parameters given the data.
- Predictions for other observables are made using the "fit" parameters instead of the full range of allowed values and again the "fits" are assessed "by eye."
- Some parameter values are fixed arbitrarily.
- To assess the effect of some physical parameter one holds the others fixed and varies the one parameter.
- When adding new physical effects, i.e. new parameters, the values of the old parameters are held at their old values.
- Use a Bayesian Inference approach to get around these problems.

Bayes Theorem

- Bayes theorem states the probability of a model characterized by its parameter vector θ, given some data set D.

$$P(\theta|D) = \frac{L(D|\theta)\pi(\theta)}{\int L(D|\theta)\pi(\theta)d\theta}$$

- $P(\theta|D)$: posterior distribution.
- $L(D|\theta)$: likelihood function; probability of the data given θ
- $\pi(\theta)$: prior distribution of the parameter vector θ, our prior knowledge about the parameter.

Advantages of Bayesian Approach

- Maximum Likelihood (ML) assigns the best-fit parameter value to the model that has the highest probability of generating the observed data.

- Really want to know: what is the probability of the model for the observed data?
- The best fit model suffers from intrinsic covariance and the possibility of complex topologies leading to multiple, non-Gaussian modes.
- Need the full posterior as provided by Bayesian MCMC.
- Can use statistics like Bayesian Evidence to discriminate between models. e.g. Does one need AGN feedback?

Bayesian Evidence

- Apply Bayes theorem to give the probability of the theory M based on the data D given a prior probability of the theory.

$$P(M|D) = \frac{P(D|M)P(M)}{\int P(D|M)P(M)\,dM}$$

where

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

- $P(M|D)$ is the Bayesian Evidence.

Bayes Factor

- Can estimate the posterior odds of two different theories M_1 and M_2 parametrized by different parameter vectors θ_1 and θ_2:

$$\frac{P(M_1|D)}{P(M_2|D)} = \frac{P(M_1)}{P(M_2)} K_{12} \quad \text{where} \quad K_{12} \equiv \frac{P(D|M_1)}{P(D|M_2)}. \tag{1}$$

- $P(D|M_i)$: the marginal likelihood for model i.
- If one does not favor either theory a priori, $\frac{P(M_1)}{P(M_2)} = 1$ since $P(M_1) = P(M_2)$.
- K_{12}: Bayes factor–odds in favor of one theory over another for the data.
- Example: Is data better fit with a model that adds AGN feedback?

The Bayesian Approach in Practice

- Goal is to characterize the posterior distribution by sampling $P(\theta|D)$.
- Not practical to solve analytically or by evaluating the posterior probability over a grid in parameter space for complex problems.
- Sample the posterior using Markov chain Monte Carlo (MCMC).
- MCMC algorithms sample from probability distributions using a Markov chain that has the desired distribution as its equilibrium distribution.
- A Markov chain is a random process where the next state θ^{i+1} depends only on the current state θ^{i} and not on the past.
- Metropolis–Hastings is a common MCMC algorithm.

Metropolis–Hastings Algorithm

- Use proposal function $Q(\theta^p;\theta^i)$ to generate proposed sample θ^p.
- $Q(\theta^p;\theta^i)$ must be symmetric, $Q(\theta^p;\theta^i) = Q(\theta^i;\theta^p)$, e.g. $\theta^p = \mathcal{N}(\theta^i,\sigma^2)$.
- Proposal is accepted, i.e. $\theta^{i+1} = \theta^p$ if

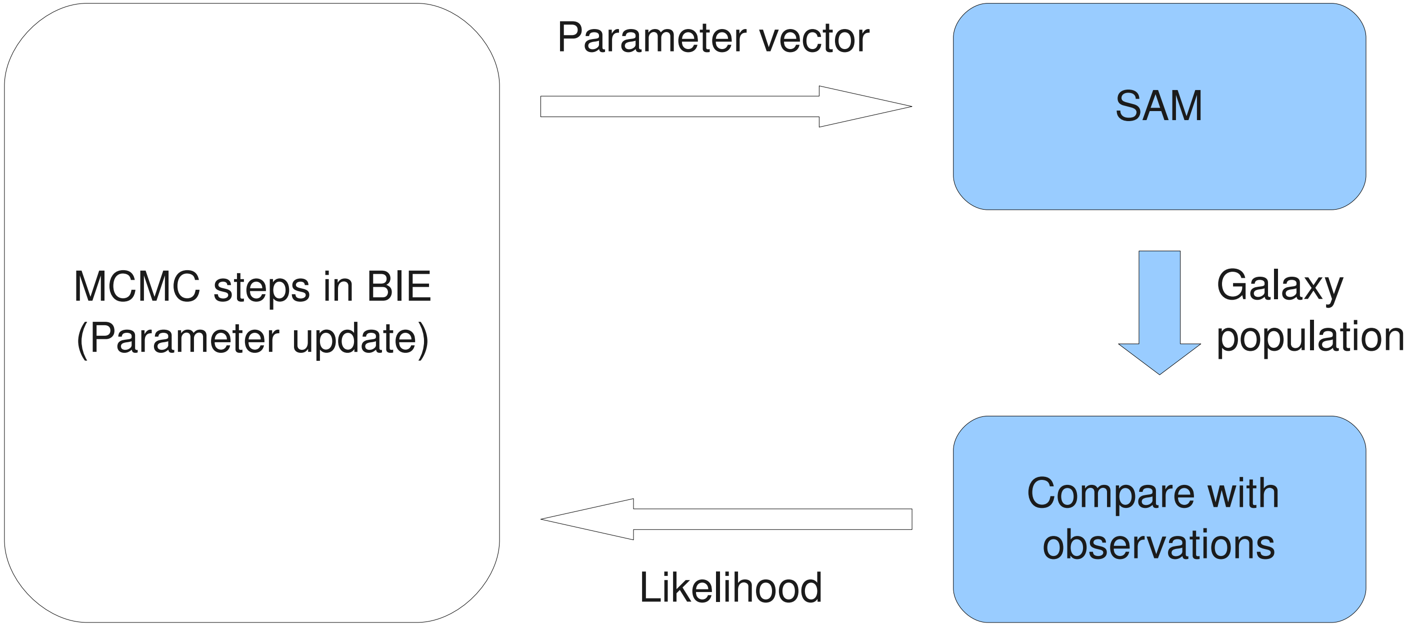
$$\alpha < \min\left(\frac{P(\theta^p)Q(\theta^i;\theta^p)}{P(\theta^i)Q(\theta^p;\theta^i)}, 1\right) = \min\left(\frac{L(D|\theta^p)\pi(\theta^p)Q(\theta^i;\theta^p)}{L(D|\theta^i)\pi(\theta^i)Q(\theta^p;\theta^i)}, 1\right)$$

  where α is a random number $\alpha \sim \mathcal{U}(0,1)$.
- If the proposal is rejected then $\theta^{i+1} = \theta^i$.
- Start from a random initial value θ^0; run for many iterations until initial state forgotten, called *burn-in*.
- Adjust σ to get good acceptance rate, $\sim 25\%$ to get good mixing.

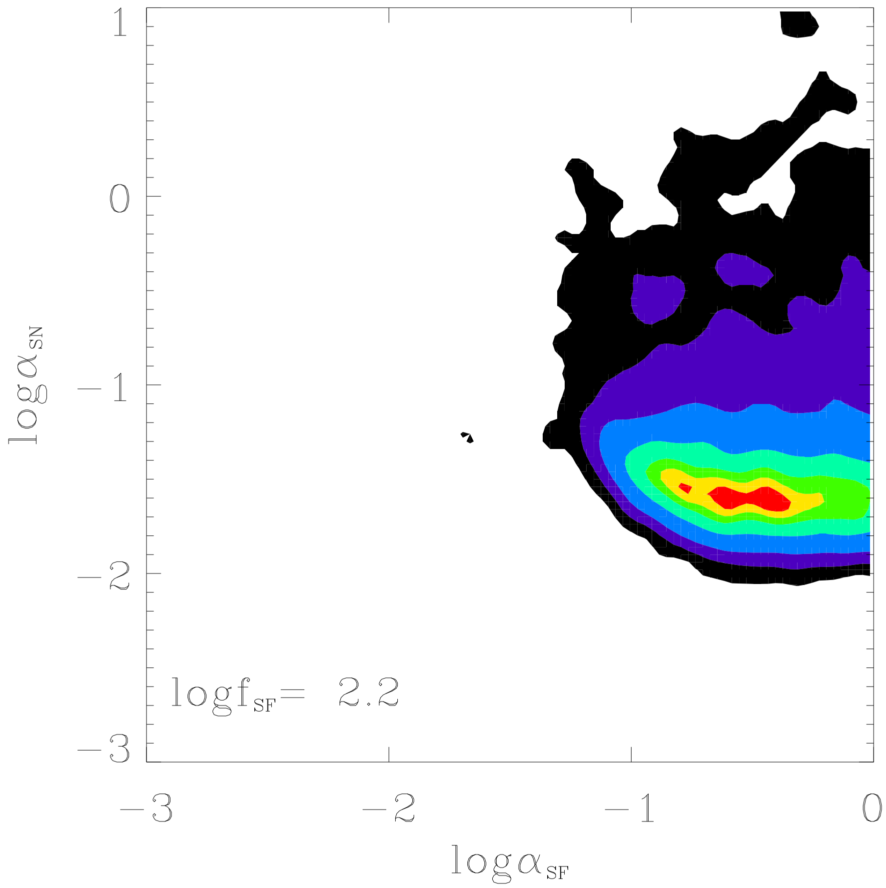
Bayesian Inference Engine (BIE)

- An MCMC parallel software platform for performing Bayesian inference over very large data sets.
- Developed by multi-disciplinary team from Astro and Comp Sci at UMass led by Martin Weinberg.
- Uses scalable multiprocessor software architecture and operates on modest cost hardware.
- Uses standard MPI and POSIX threads so runs on a broad spectrum of parallel or scalar machines.
- Includes: standard Metropolis-Hastings, simulated tempering, parallel tempering, parallel hierarchical sampling, differential evolution, and independent multiple chains.
- Saves *checkpoint* images so can restart from last MCMC step.
- Available at: `www.astro.umass.edu/~weinberg/bie`

Bayesian SAM Flowchart

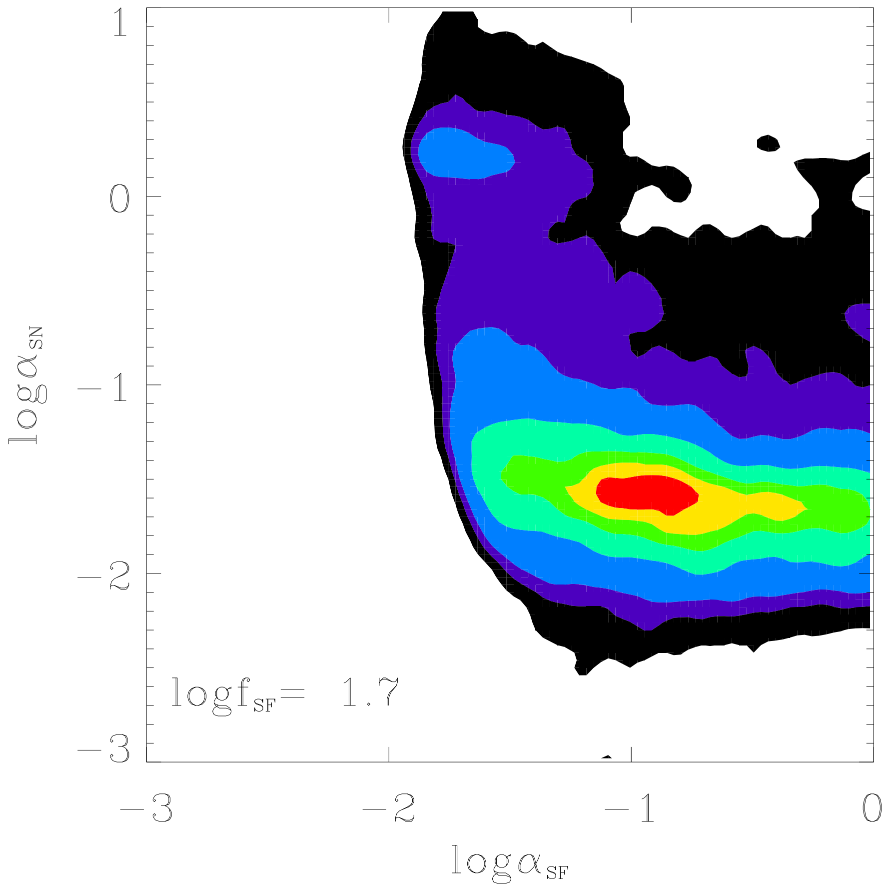


Posterior Madness: Why we need Fancy MCMC



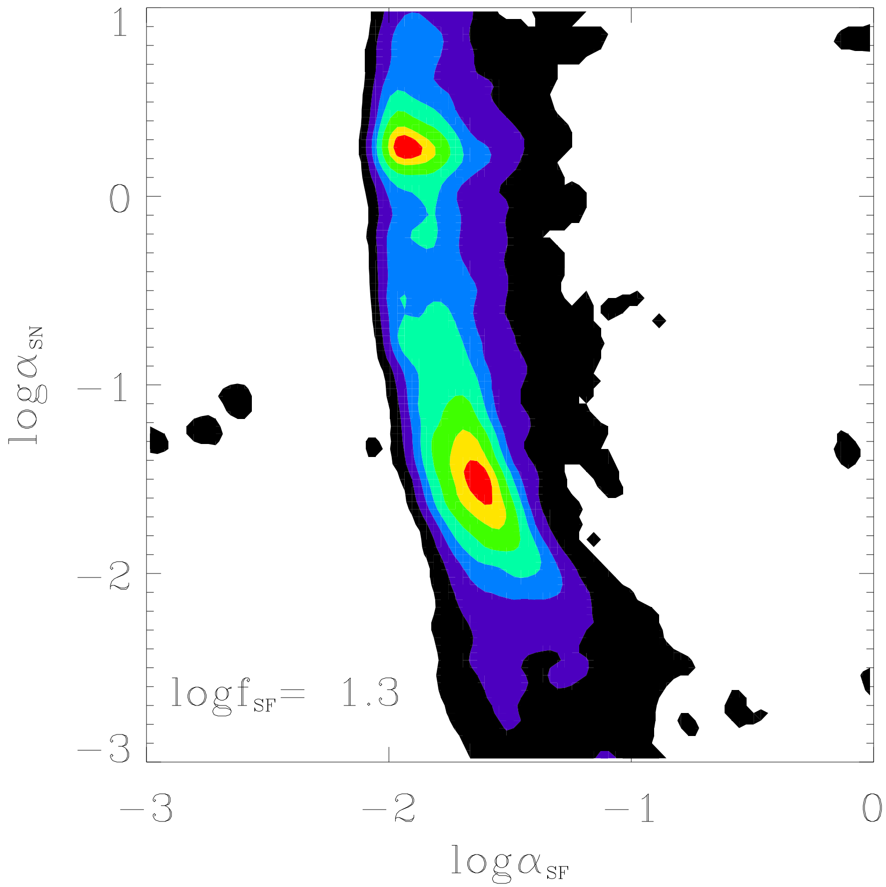
- Marginalized over 10 of 13 parameters with slices in $\log f_{SF}$.
- Posterior is very thin and twisted.
- Impossible to find global maximum by hand.

Posterior Madness: Why we need Fancy MCMC



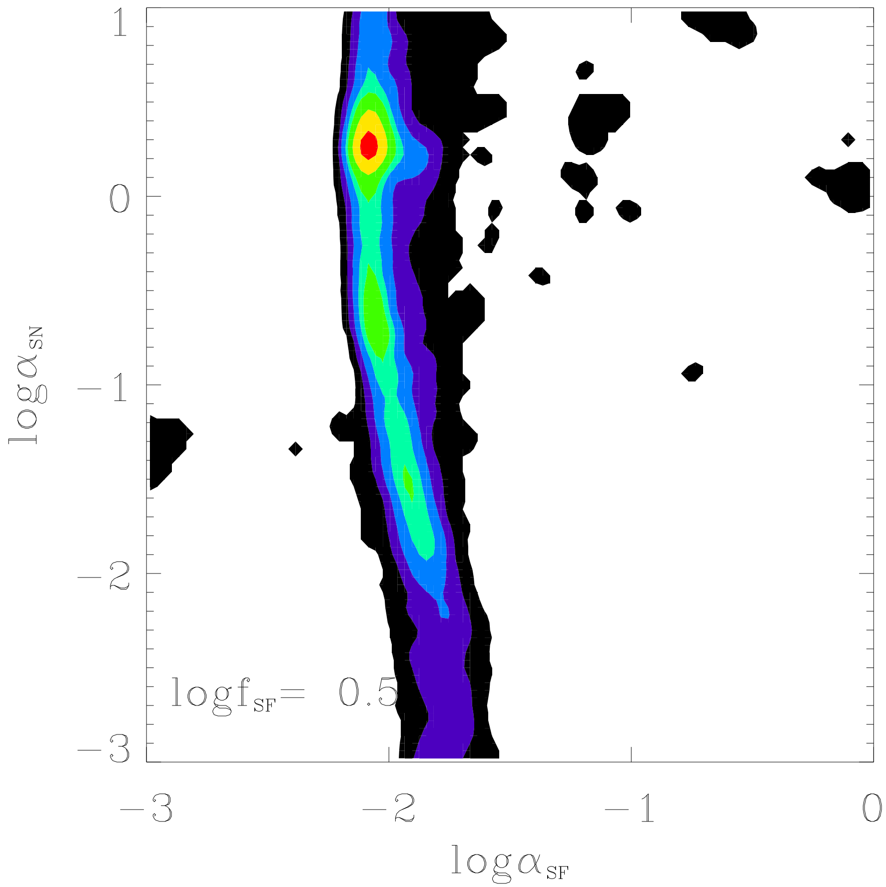
- Marginalized over 10 of 13 parameters with slices in $\log f_{SF}$.
- Posterior is very thin and twisted.
- Impossible to find global maximum by hand.

Posterior Madness: Why we need Fancy MCMC



- Marginalized over 10 of 13 parameters with slices in log f_{SF}.
- Posterior is very thin and twisted.
- Impossible to find global maximum by hand.

Posterior Madness: Why we need Fancy MCMC

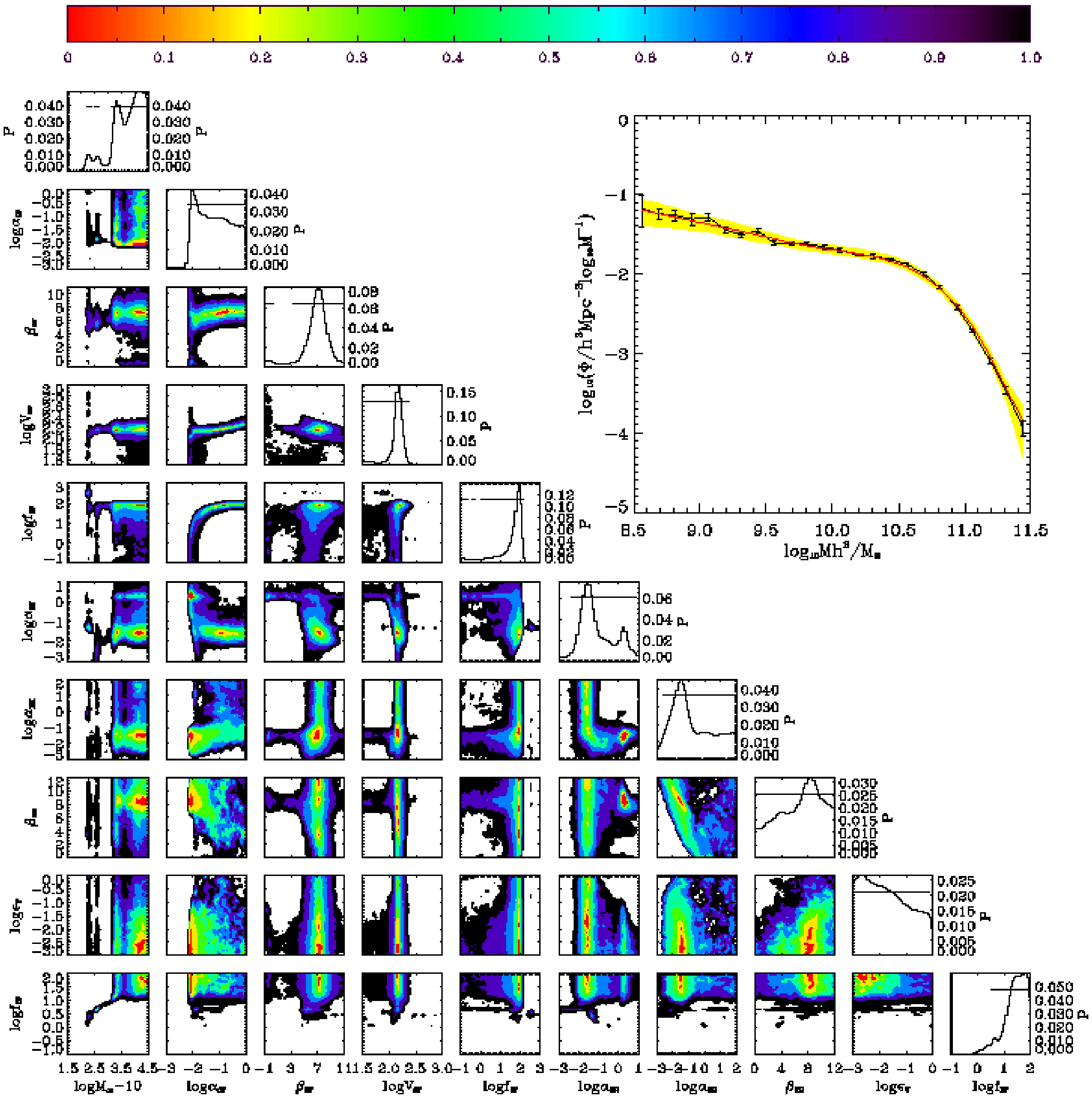


- Marginalized over 10 of 13 parameters with slices in $\log f_{SF}$.
- Posterior is very thin and twisted.
- Impossible to find global maximum by hand.

Play it SAM: The New Song

- Use galaxy stellar mass function as data constraint.
- Can fit the data well.
- Posterior is complex and multi-modal.
  - Some modes are equivalent to previously published SAMs.
- Some parameters are covariant, e.g. f_{SF}–α_{SF}; α_{RH}–β_{RH}; and M_{CC}–f_{DF}.
  - Either M_{CC} is about 10^{12} and f_{DF} is about one or M_{CC} is large and f_{DF} is very large making the time for mergers to occur very long.

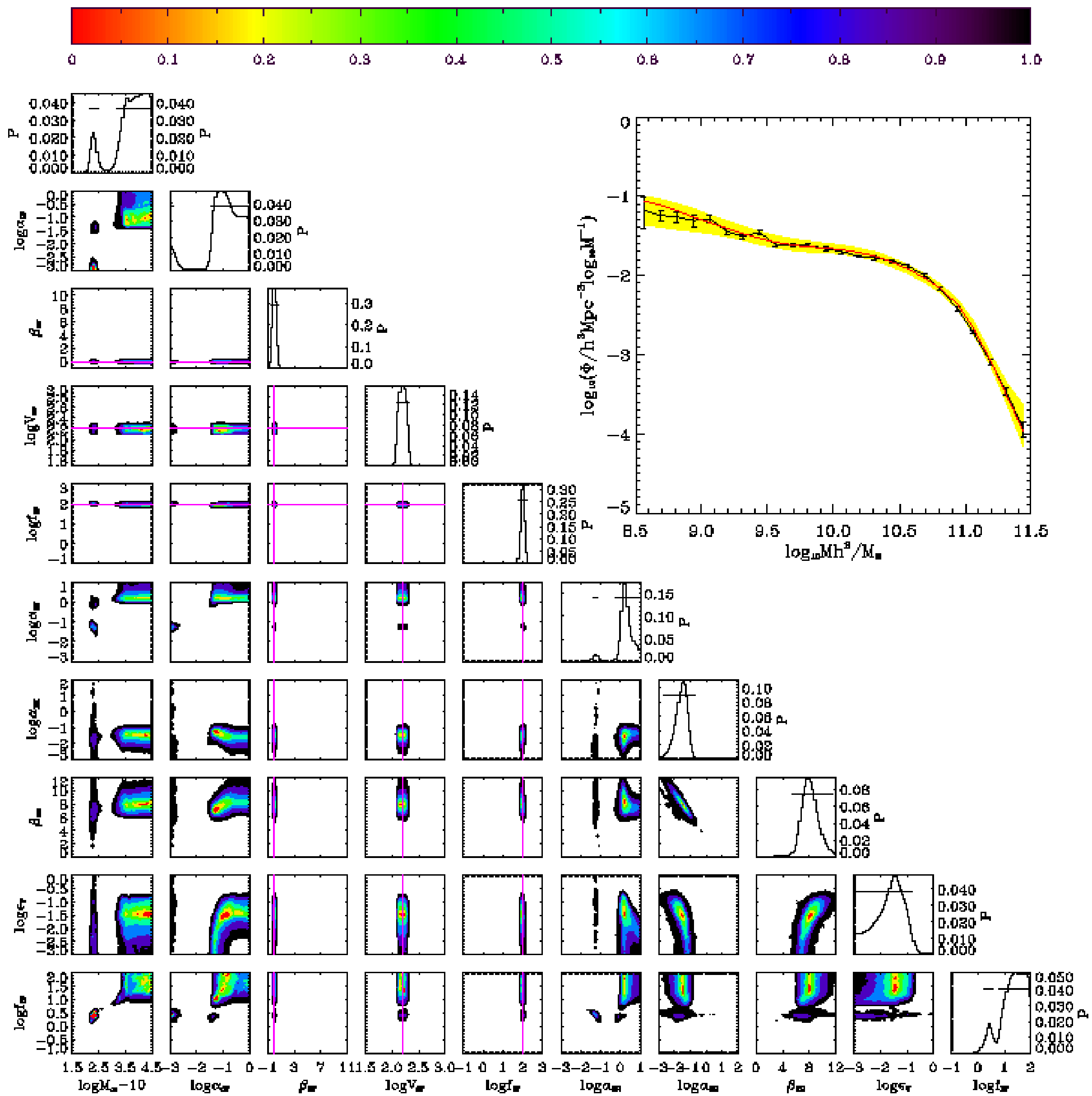
SAM marginalized Posterior



Play it SAM: The New Song (With Restrictions)

- Restrict the prior to make our SAM more like the Croton SAM.
  - $\beta_{\rm SF} = 0$.
  - $V_{\rm SF} = 160\,{\rm km\,s^{-1}}$.
  - $f_{\rm SF} = 100$, equivalent to $\Sigma_{\rm sf} \approx 10\,{\rm M_\odot/pc^2}$.
- Still fits the data well.
- Greatly reduces covariances and the allowed parameter range.
- The main mode in the posterior barely overlaps the main mode for the unrestricted prior.
- Now requires $\beta_{\rm RH}$ to be large.
- Previous published work claimed that $\beta_{\rm RH}$ must be large to match low mass end.

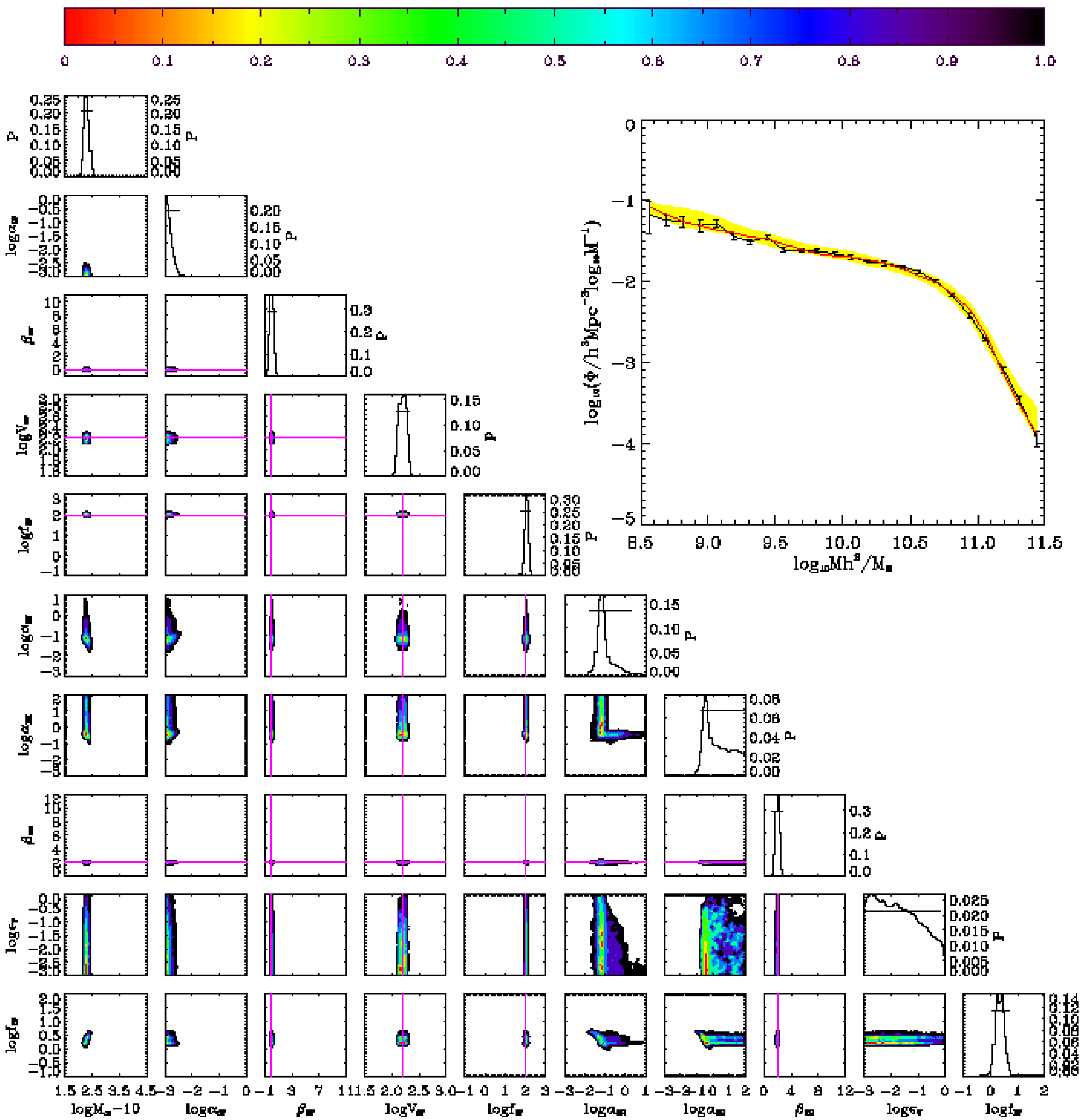
SAM marginalized Posterior: Restricted Prior



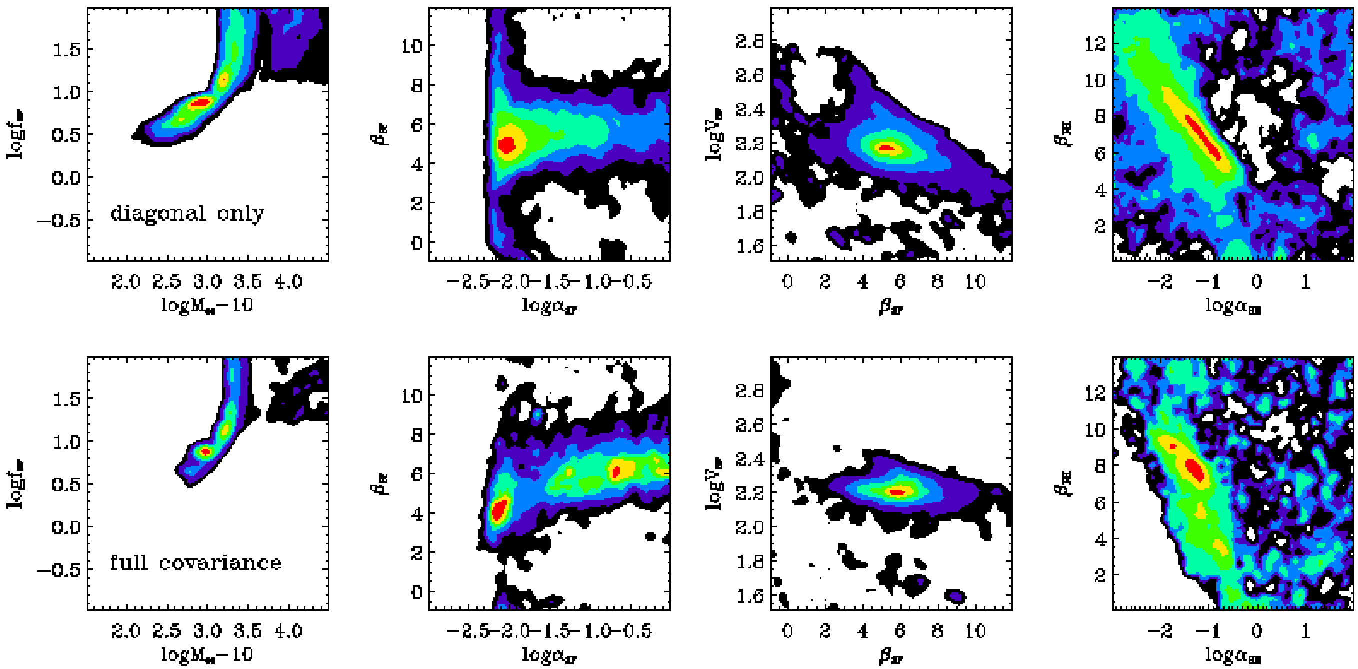
Play it SAM: The New Song (With Even More Restrictions)

- Set $\beta_{RH} = 2$.
- Still able to match the data even though β_{RH} is not large.
  - Moral: Cannot hold one parameter fixed to see the allowed range of another parameter.
- Again the main mode has moved.
- Further reduces covariances and the allowed parameter range.

SAM: Further Restricted Prior



Covariant Errors in the Data

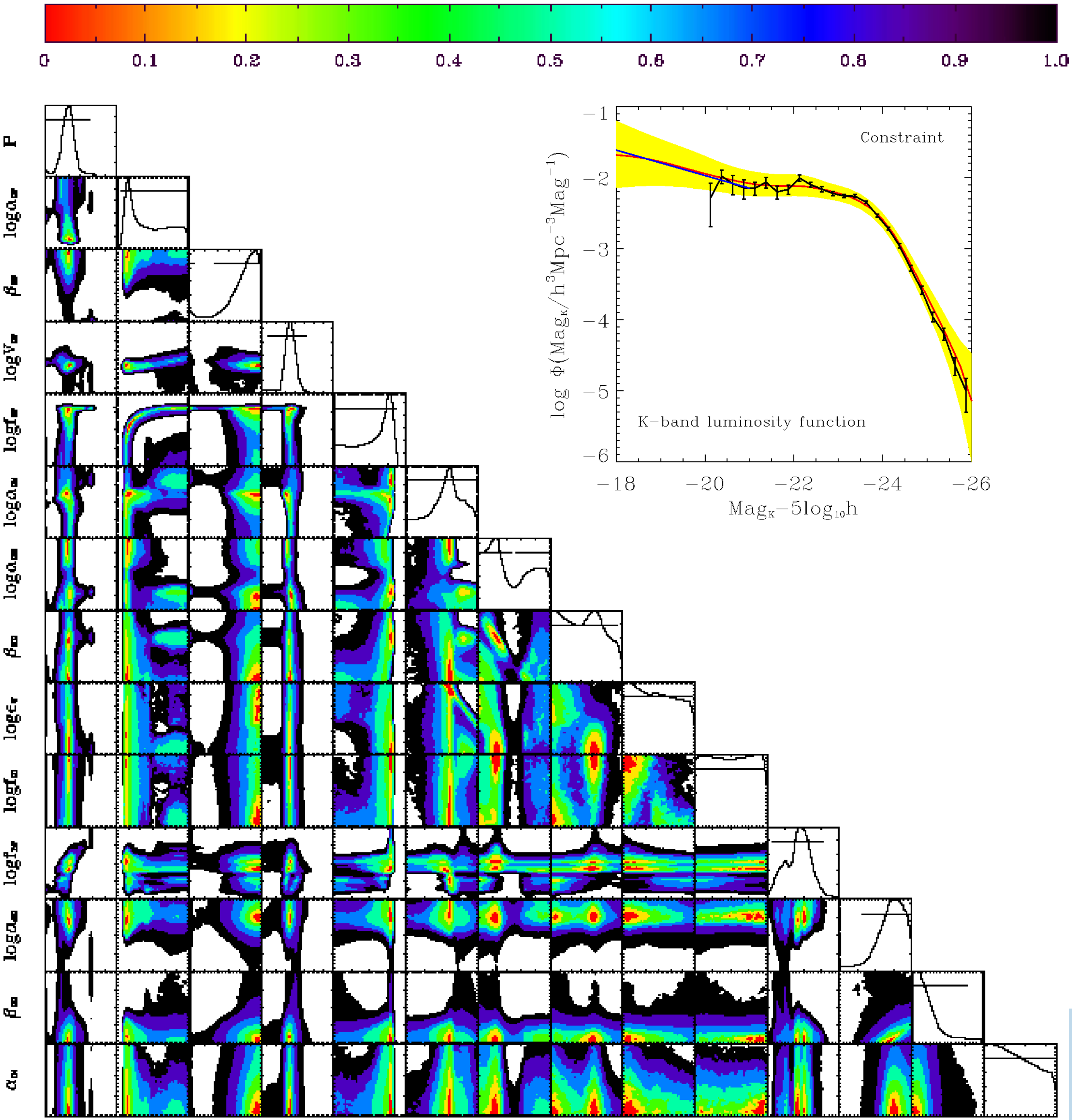


- The error covariance matrix of the stellar mass function is not really diagonal, i.e. the errors in each bin are not independent.
- Including the full error covariance matrix reduces the allowed region.
- Using the full covariance matrix has also allowed a new mode.

Play it SAM: K-band Constrained

- K-band errors should be diagonal.
- The faint end completeness is not well understood so we parametrize it with an additional parameter and then marginalize over it.
- Again we can fit the data and find other SAMs in some of the posterior modes.
- Some differences with the stellar mass function constrained inference.
  - Now only the mode with M_{CC} about 10^{12} and f_{DF} about one is allowed.
  - Allowed ranges of β_{SF} and β_{RH} have also changed.

SAM: K-band Constrained



Conclusions

- It is possible and highly desirable to use Bayesian Inference with MCMC when using SAMs.
- The solutions to SAMS are multi-modal.
- Many parameters are highly covariant.
- One should use the entire posterior when making predictions.
- Observers should always publish the full error covariance matrix of their data to make it useful for Bayesian Inference.
- Need to use multiple data constraints simultaneously in the future.
- May need to add more physical processes.
- “Full” SAMs will have over 50 parameters!
- Much work remains to be done.