Enabling Bayesian Inference for the Astronomy Masses

Performance Report submitted by M. D. Weinberg, PI
Period: 3/15/06–3/14/07
Grant Number: NNG-06-GF25G

1 Executive summary

Active development proceeded in four of the five defined research topics:

1. Development of statistical methodology
    - Implemented the *Reversible Jump* Markov chain as an alternative model selection tool for multicomponent mixture models
    - Implemented the *Parallel chains* algorithm for accelerating convergence
    - Implemented the *Particle filter* algorithm as an alternative to pure MCMC simulation
    - Updated the posterior visualization tool using the GL-based Visualization Tool Kit with a GTK+ GUI.
    - Augmented the BIE architecture to for user defined likelihood routines.
    - Updated the read/write methods to allow arbitrary SQL table output with named fields
2. Development of persistence technology
    - Reviewed and redesigned current experimental persistence design based on preprocessor macros
    - Developed conceptual design for persistent store based on SVN repositories
    - Began re-implementation of serialization and persistence methods

3. Astronomical applications

   - Developed and implemented a GALFIT-based galaxy image analyzer, in discussion with the GALFIT author, Chien Peng. Tests show that the new application, BIE–GALFIT, is significantly less biased. Moreover, we have shown that standard approaches to model selection using GALFIT and χ^2 statistics fails where Bayes ratio methods succeed. We anticipate releasing a stand alone BIE–GALFIT package in the upcoming year.
   - Developed and implemented semi-analytic method routine.

I will detail some of advances below and end with a list of Milestones for Year 2.

Administrative issues

Although the start date is 3/1/06, we were notified of the award in June, 2006. Work began in early Summer by the PI followed by the entire team in September. The work reported here covers approximately the first 7 months of activity.

Because of the award timing, we advertised the post-doctoral position in Fall 2006 and have recently hired Jörg Colberg, who will begin in September 2007.

2 Research milestones and summary

2-1 Persistence subsystem development

In the area of persistence and work flow management, we have three objectives at present: 1) restore inter-command state save/restore to functionality, while simplifying the programmer's interface and making it more robust from a software engineering perspective; 2) add support for checkpointing while running Monte-Carlo Markov chains; and 3) begin design of the high level persistence tool and interface that will present and manage the various computations and lines of work a researcher is investigating, providing support analogous to that offered by integrated development environments to programmers.

We have nearly met the first objective. The new research assistant has now mostly climbed the learning curve on approaches to adding persistence (save/restore) to Java. He picked apart our previous code and determined that it broke because it was relying on properties of particular compilers that tend to change from release to release. We redesigned that part and are also moving towards using the better persistence support available from the Boost C++ libraries, which are widely used and well-maintained. We reworked the way in which the programmer indicates which fields of a class must be saved and in which the system works in the necessary automatically generated code for the actual saving and restoring. The result is easier for programmers to use and will also better help them avoid certain possible mistakes. We are now well positioned to tackle checkpointing, which should be relatively straightforward given working save/restore, and then to dig in to the intellectually more challenging issues of the high level tool.

2-2 BIE–GALFIT

2-2.1 Motivation

The galaxy structure is evolving due to gravitational and gas dynamical physics in the expanding Universe. To understand the evolution of galaxy structure based on their morphology has been done by human eye, which led to the systems in use today such as Hubble type. As galaxy surveys have become deeper and more voluminous, researchers have explored a variety of automatic classification schemes.

There are two main approaches towards describing galaxy structure from the two dimensional images. Non-parametric approaches estimate several quantities such as total brightness, galaxy half-light radius, concentration and asymmetry. However the results are sensitive to the depth of image. Thus one can underestimate the flux and size of faint galaxies on noisy background (Blanton et al., 2003). On the other hand, parametric approaches use particular functional light profiles(observationally motivated or sometimes physically motivated) for modeling galaxy light distribution in the image. Although parametric approach has less flexible than non-parametric one, it can capture the light which is at larger radius and still significant contribution but not seen clearly in the image. Also since we know that there are several common types of luminous components(disk, bulge, bar and spiral arms) which consist of the galaxy light distribution, the parametric approach using different light profiles modeling each component can provide the information of galaxy structures which vary over cosmic time scale and depend on density environment.

The various researches about galaxy structural parameters have been done using two popular parametric galaxy fitting code, GALFIT and GIM2D . Recently Häußler et al. have done exhaustive tests and comparisons between GALFIT and GIM2D . They show that GALFIT offers a number of important advantages over GIM2D for galaxy fitting on large moderate depth *HST/ACS* data, foremost its much higher speed and its robustness to nearby galaxies (Häußler, 2007).

GALFIT is a modular package written to perform two dimensional image decompositions for galaxies which are from nearby to distant (Peng et al., 2002). GALFIT takes an input image and outputs a model-subtracted images as well as a catalog of structural parameters for an arbitrary number of components. The predefined components include most of the commonly used profiles. Each predefined component has up to ten parameters but allows for an arbitrary number of user-defined profiles and components. Some parameters may be fixed depending on one's application but a typical fit will require greater than 12 parameters. GALFIT optimizes the parameters of the likelihood function using Levenberg-Marquardt downhill algorithm. However it is possible that it converges on a local minimum of likelihood. This becomes severe when the number of parameter is large since the topology of likelihood can be multi-modal. Also if the image quality is poor, there may be no strong mode in likelihood function and it is hard to know which combination of parameter should describe the galaxy structure.

This motivates our Bayesian Inference Engine back end, which will allow GALFIT-based investigations of the full posterior not just the extremum mode, and will establish proper prior distributions, which allow inferences using Bayes Factors over a wide variety of competing models and hypotheses.

2-2.2 Galaxy image modeling

All tests and investigations described here use *synthetic* data. We generate two simple, simulated galaxies using MIDAS package. They are idealized galaxies with very high *S/N* . Secondly, we select several high and low *S/N* isolated galaxy images from large, simulated ensemble of galaxies in Häußler (2007). They simulated the galaxy light profiles and putting them in an empty space. This image was convolved with a real F850LP-band PSF derived from GEMS dataset and appropriate amount of noise was added to it. For more details, see Häußler (2007). We estimate mean *S/N* based on Häußler (2007)) using mean surface brightness(mag/arcsec^2) within half-light radius, r_e

$$\mu = \mathrm{mag} + 2.5\log(2\frac{b}{a}\pi r_e^2) \tag{1}$$

where b/a is axis ratio. For example, $\mu = 20.5, 24.0$ correspond to $S/N = 10, 0.9$ respectively. See their figures for the scaling between *S/N* and μ. The typical surface brightness of sky background is $\mu = 22.5$. All images are kindly provided by Dr. Daniel McIntosh and Mr. Yicheng Guo. Last, we use one image with three galaxies with same Sérsic index, $n = 4$. The light profiles of those galaxies are blended.

The most critical issue about galaxy fitting is to estimate sky background. Although we can fit sky background as an extra parameter, this can lead to the biased result if the model light profile does not exactly describe the real galaxy light distribution. The estimation of sky background is also affected by the relative size of galaxy to the image, especially for the Sérsic profile with long tail(i.e. $n = 4$). Thus it is usually better to fix the sky background based on independent measurement(Häußler (2007)). We thus select relatively small galaxies($r_e < 10$) with the image size of from 240 by 240 or 400 by 400.

For modeling these data, we use single Sérsic model from GALFIT and the different priors for each parameters. *BIE* currently provides 7 types of different prior. See *BIE* website for more information.

2-2.3 Synthetic galaxy images

All images are generated following Sérsic light profile with two different indices, $n = 1$ and $n = 4$, which correspond exponential disk and de Vaucouleur galaxy respectively. The radial surface brightness profile of Sérsic function is given by

$$\Sigma(r) = \Sigma_e exp[-\kappa(\frac{r}{r_e})^{1/n} - 1)] \tag{2}$$

where Σ_e is the surface brightness at effective radius r_e which is such that half of the total flux is within r_e. The parameter *n* is Sérsic index or often called concentration parameter. When *n* is large, it has steep inner profile, and a highly extended outer wing. Inversely, when it is small, it has a shallow inner profile and a steep truncation at large radius. The κ is not a independent variable and related with *n*. Usually a good approximation for κ for $n > 0.5$ is $\kappa = 2n - 1/3 + 0.009876/n$ (MacArthur et al. (2003)). Figure 1 shows Sérsic profile for different *n* using this approximation for κ.

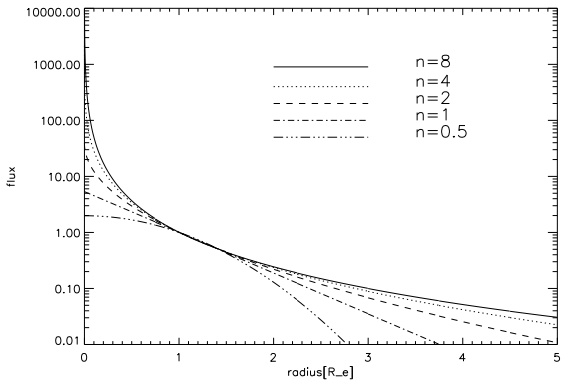

Figure 1: Sérsic surface brightness profiles for n=0.5, 1, 2, 4 and 8 (equation 2). The profiles have been normalized at $r_e = 1$.

For an exponential profile (n=1), 99.1% of the flux resides within the inner $4r_e$ and 99.8% of the flux resides within the inner $5r_e$. For an n=4 profile, 84.7% of the flux resides within the inner $4r_e$ and 88.4% of the flux resides within the inner $5r_e$ (Graham and Driver, 2005).

2-2.4 Bayesian approach for modelling data

For the likelihood function, we construct the likelihood function using models in GALFIT .

$$P(D \mid \theta) = \frac{exp(-\frac{1}{2}[\mathbf{D}-\mathbf{M}(\theta)]^t \mathbf{W}[\mathbf{D}-\mathbf{M}(\theta)])}{(2\pi)^{Npix/2} \mid \mathbf{W} \mid^{-1/2}} \tag{3}$$

where $\mathbf{D}$ is data vector($N_x \times N_y$), $\mathbf{M}(\theta)$ is a model vector and $\mathbf{W}$ is a weight matrix for pixel value.

For the prior for parameters, we mostly adopt the uniform prior with a range(top-hat) which leads the likelihood dominated posterior probability distribution and basically the same case with the maximum likelihood method, a least informative case of Bayesian statistics. As we shall see in later, the effect of prior becomes more significant when we have data where the information is weak and degenerated. For example, in case of low *S/N* data, the informative priors for some parameters help to obtain the robust estimate for those parameters.

We use GALFIT model with single Sérsic profile and different prior for each parameter. In GALFIT , each Sérsic function has 8 free parameters in the fit: centroid of the profile(x_c, y_c), integrated magnitude(M_{tot}) which is related with Σ_e, effective radius(r_e), Sérsic index(n), axis ratio(b/a), position angle(PA) and diskiness/boxiness(c). Also GALFIT can add another profile for setting up the sky background with 3 free parameter:sky level, sky gradient in X,Y direction. Some of parameters can be hold to fix while fitting. See Peng et al. (2002) for more general information. This is the model $\mathbf{M}(\theta)$ used in the likelihood function.

2-2.5 Selected results

In this section, we illustrate and interpret the Bayesian MCMC results for all simulated galaxies with different structural parameters. First we show very ideal cases, which are two isolated galaxies

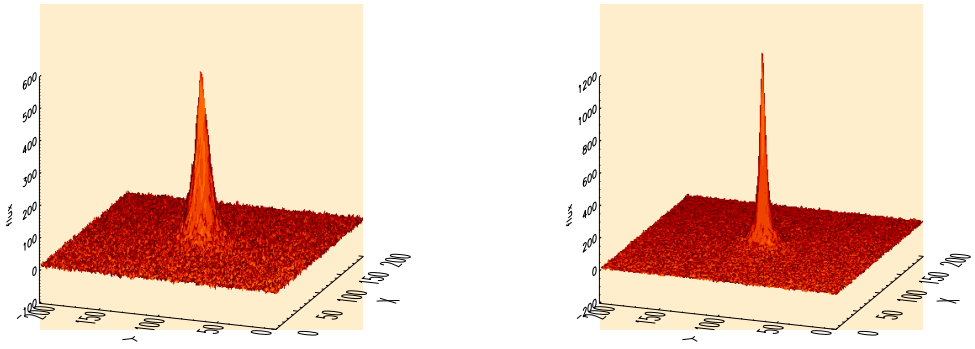

Figure 2: The surface plot of light profiles from disk0 and sph0. Left is exponential disk(n=1) and right is spheroid(n=4).

Table 1: PARAMETER VALUES OF DISK0 AND SPH0

Parameter	disk0	sph0
image size	200×200	200×200
position X		
position Y		
Total magnitude[mag], M_{tot}	20.0	20.0
Effective radius[pix], R_e	10.0	10.0
Sersic index, n	1.0	4.0
Axis ratio, b/a	1.0	1.0
Position Angle, PA[deg]	0.0	0.0
GALFIT best fit		
Total magnitude[mag], M_{tot}	20.0 ± 0.00	20.0 ± 0.01
Effective radius[pix], R_e	10.12 ± 0.04	10.13 ± 0.13
Sersic index, n	1.01 ± 0.01	3.99 ± 0.05
Axis ratio, b/a	0.99 ± 0.00	1.00 ± 0.01
Position Angle, PA[deg]	–	–

with very strong signal. we check if *BIE* is working as we expect and try different techniques for improving chain mixing and convergence. Then we compare the result with GALFIT and study the parameter correlation and uncertainties. We have also explored more realistic galaxy images with high and low *S/N* and characterize how *BIE* works for the strongly or weakly informative data. We also show the effect of strong prior over weakly informative likelihood. Last we model the multiple galaxies in one image, where their light profiles are blended.

Two ideal galaxies: Disk and Spheroid

We show two galaxy images, disk0 and sph0 in Figure 2. Their input parameter values are listed in Table 1. The galaxies are modelled by Sérsic profile with 7 free parameters. The diskiness vs. boxiness parameter is fixed to zero and the sky background level is also hold to the known value. However the sky background fitting is very robust for these galaxies. We generate MCMC using

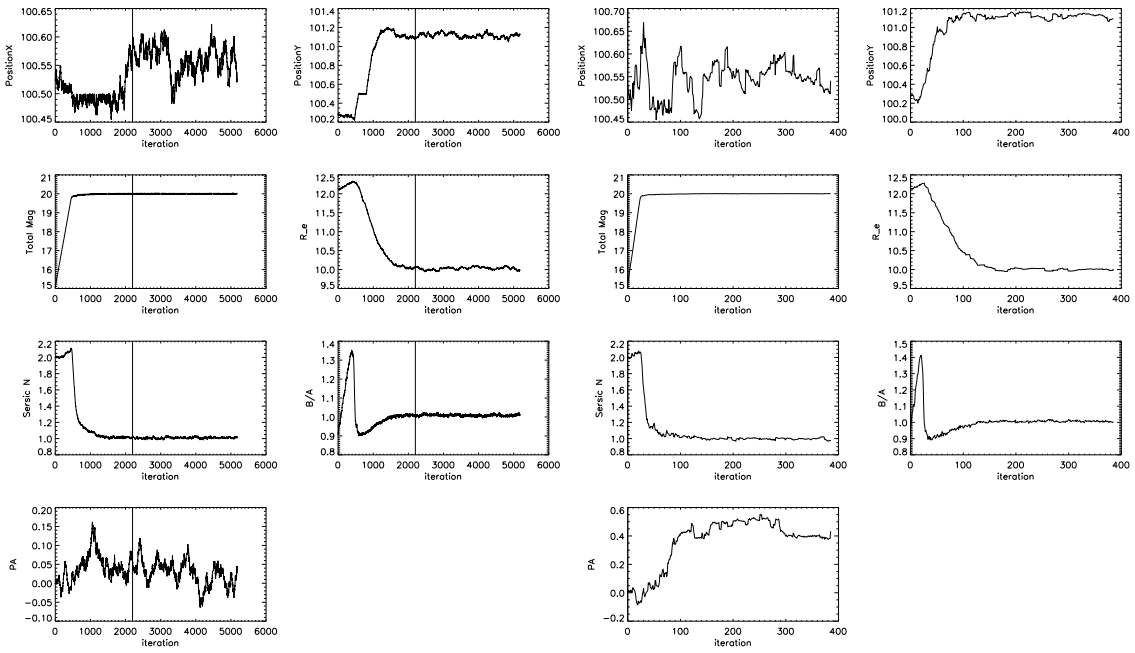

Figure 3: The trace of states for disk0. Left is for pure Metropolis-Hastings(MH) and right is for simulated tempering. Although the simulated tempering algorithm takes longer time than pure MH for advancing one step, it converges with smaller number of iteration than pure MH. PA is not converging since it could have any value.

different sampling algorithms in *BIE* and confirm its feasibility. Three different algorithms are Metropolis-Hastings, simulated tempering and parallel chain.

Comparison between different algorithms

For the exponential disk, we run *BIE* with Metropolis-Hastings and tempered simulation algorithm. The chains for all model parameters are shown in Figure 3. After burn-in period (marked as black vertical line in Figure 3) , all values are closely converging to true input parameters. The left and right panels in Figure 3 are respectively MCMC with Metropolis-Hastings and simulated tempering algorithms for 7 free parameters. Since their axis ratios are both 1.0, there is no preferred values for PA. Although the simulated tempering requires Metropolis-Hastings 'internal' steps for advancing one step, it converges more quickly than Metropolis-Hastings algorithm.

For spheroid, we show MCMC with simulated tempering and parallel chain algorithms in Figure 4. In parallel chain simulation, *BIE* runs several chains with different starting points and temperatures. Each chains probe different regions of parameter posterior probability distribution and the swaps between different chains start to occur based on relative probability of each chain state. In this experiments, parallel chain algorithm is generally more robust and faster than Metropolis-Hastings and simulated tempering.

We show 1D marginalized distributions of parameters for disk0 and sph0 in Figure 5 and Figure 6 respectively. The parameter distribution in Figure 5 is sampled from the chain with Metropolis-Hastings and the parameter distribution in Figure 6 is from the simulations with simulated tempering (black) and parallel chain (blue) algorithms. There are two different error bars in the figures.

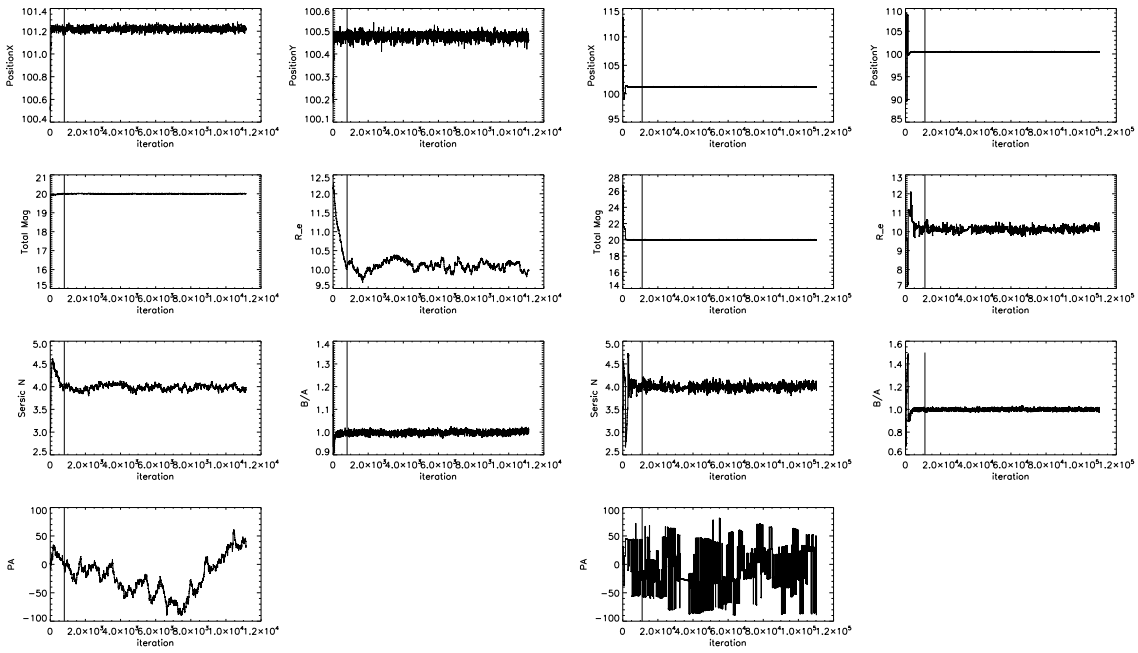

Figure 4: The trace of states for sph0. Left is for simulated tempering and right is for parallel chain algorithm. PA is not converging since it could have any value.

The dotted lines are 1, 2 & 3 standard deviations from the expected values. The solid lines are 68.3, 95.4 and 99.73% percentiles from the median, which correspond to 1,2 & 3 standard deviations in one dimensional parameter space. The solid vertical lines indicate the true input parameters.

Random sampling from the posterior is distributed around the true input parameters but the peak is not located at the exact input value. In Figure 6 the posterior probability distributions for each parameters are relatively smooth in case of simulated tempering algorithm. This is because one single chain probes the parameter space and chain transition is smooth. In parallel chain algorithm, several chains start from different initial conditions probing different parameter spaces and meanwhile, there is chain swapping which is based on the relative posterior probability of each chain. This makes discrete jumps in the trace of parameters and peaks in the marginalized parameter distributions. There are slight offsets of median/expected values for M_{tot},r_e and n from the true input parameters in simulated tempering algorithm. However, the expected and median values from parallel chain simulation are very close to the true input parameters although the offset of r_e is slightly larger than that from simulated tempering simulation. This indicates that the parallel chain effectively probes the parameter space. In Figure 5 we also see slight offsets of expected/median values from the true parameters. Pure Metropolis-Hastings probes parameter space less efficiently than simulated tempering and parallel chain algorithm.

In general the topology of parameter space for exponential disk is smoother than spheroid given the same *S/N* . Since spheroid has a light profile with long tail, the correct modeling including the outer part of the profile is severely hampered by sky background and noise. For example, it is hard to reveal clear uni-modality of r_e and n around true values for spheroid. This can be more clearly seen in the 2D marginalized parameter distributions in next section. For both disk0 and sph0, magnitude is the most robust parameter even though there is slight offset from the true magnitude.

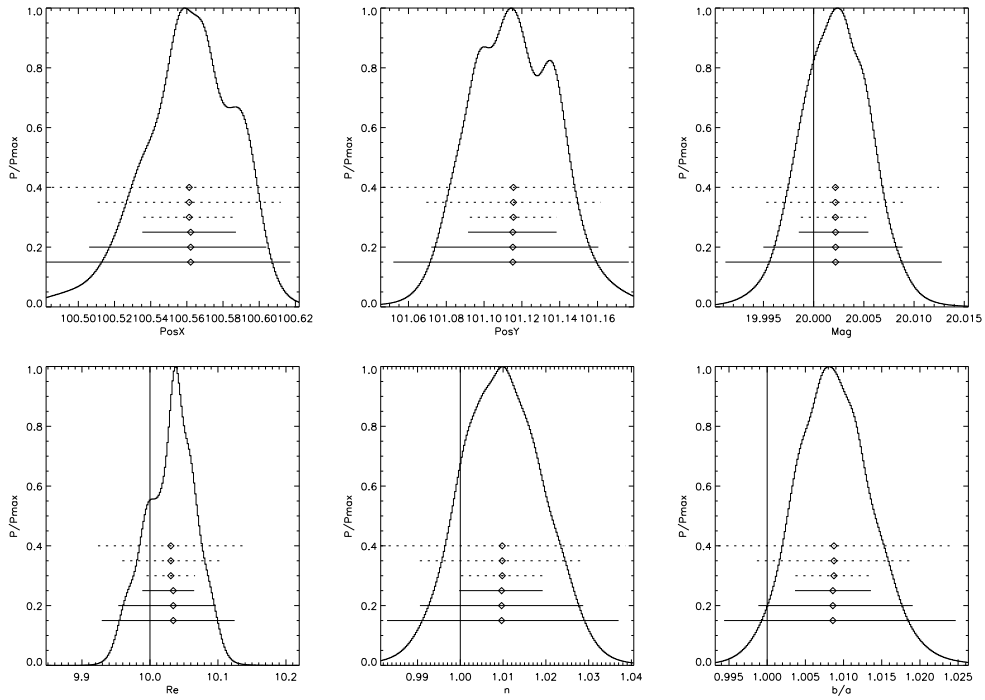

Figure 5: 1D marginalized parameter posterior probability distribution for disk0 with arbitrary normalized factor. It is sampled by Metropolis-Hastings algorithm.

Uncertainties and correlation of parameters

The errors from *BIE* are consistent; all true input parameters are enclosed within at least two standard deviations (see Figs. 5, 6). As a comparison, GALFIT best fits and errors are listed in Table 1. The marginalized contours of parameter distribution with offsets from the true input values for M_{tot},r_e, n and b/a are shown in Figure 7 and Figure 8. The distribution in Figure 7 is sampled from the simulation for exponential disk with Metropolis-Hastings algorithm and the distribution in Figure 8 is sampled from the simulation for spheroid with parallel chain algorithm. In these figures, black diamonds and error bars are the GALFIT best fit and uncertainty. Each contour level corresponds to 5, 10, 20, 30, 40, 50, 68.5, 90, 95, 99% cumulative marginal percentages respectively. The 5% contour means that 95% of samples are within the contour. Thus these levels correspond to confidence levels. GALFIT errors are estimated from parameter covariance matrix, which is standard way of estimating parameter errors.

For disk0, the probability density peaks of *BIE* and GALFIT best fits have offsets from the true input parameters(see Figure 7). They are both mutual close in value, however GALFIT estimates for r_e has larger offset from the true value $r_e = 10$ than *BIE* . GALFIT estimated zero errors for M_{tot} and b/a (see Table 1). The one standard deviation error bar for n barely encloses the true value, but the error bar for r_e does not. On the other hand, in case of *BIE* , the true input values are within 50% marginalized confidence level except for the axis ratio b/a, which should be inverse if it is larger than 1. The size of 50% marginalized confidence level is comparable to the size of GALFIT error bar. However GALFIT error quotation for r_e is too small and the true $r_e = 10.0$ is barely enclosed by three standard deviations.

Table 2: PARAMETER VALUES OF HIGH *S/N* GALAXIES

	disk310	disk309	sph545	sph438
image size	401×401	401×400	400×400	400×400
S/N	~ 8.0	~ 7.0	~ 8.0	~ 8.0
Input parameter				
position X				
position Y				
Total magnitude[mag], M_{tot}	23.51	22.41	24.42	25.41
Effective radius[pix], R_e	8.91	7.55	3.07	2.14
Sersic index, n	1.0	1.0	4.0	4.0
Axis ratio, b/a	0.19	0.94	0.59	0.45
Position Angle, PA[deg]	75.8	13.1	131.9	140.1
GALFIT best fit				
Total magnitude[mag], M_{tot}	23.55 ± 0.02	22.43 ± 0.01	24.42 ± 0.06	24.89 ± 0.4
Effective radius[pix], R_e	9.11 ± 0.28	7.29 ± 0.11	3.25 ± 0.3	7.05 ± 7.66
Sersic index, n	1.25 ± 0.09	0.97 ± 0.03	3.47 ± 0.85	13.99 ± 11.65
Axis ratio, b/a	0.14 ± 0.01	0.98 ± 0.01	0.46 ± 0.06	0.12 ± 0.06
Position Angle, PA[deg]	75.8 ± 0.51	60.39 ± 32.52	141.6 ± 4.37	167.9 ± 3.37

For sph0, like the case of disk0, the density distribution peak of *BIE* and GALFIT best fits are close to each other but offset from the true input parameters. All parameter but r_e are enclosed within 50% marginalized confidence level. Thus *BIE* provides statistically *realistic* errors of model parameters for disk0 and sph0.

One of the distinct features of Figure 7 and Figure 8 is the correlation of parameters. In contrast to the parameter fitting as GALFIT does, Bayesian inference naturally reveals the correlation of parameters by measuring correlation of two MCMCs for corresponding parameters.

For disk0, M_{tot}, r_e, n and b/a are weakly correlated. Their absolute correlation coefficients range from 0.17 to 0.55. M_{tot} has negative correlation with r_e and n. r_e has positive correlation with n and negative correlation with b/a. For sph0, M_{tot}, r_e and n are strongly correlated. Their absolute correlation coefficients range from 0.65 to 0.82. M_{tot} has negative correlation with r_e and n. r_e has positive correlation with n. The parameter of Sérsic profile with large n which has highly concentrated central cusp and very shallow long tail in outer region significantly varies with changing other parameters. This strong correlation of parameters is the main reason to hamper to fit the data correctly.

Isolated galaxies with high S/N from large ensemble image

We select isolated galaxy images with high *S/N* , which means that it's mean surface brightness is lower than sky surface brightness. As in last section, we model those galaxy with single Sérsic profile with the sky background level as a free parameter. True input sky background level is 18.14 with r.m.s. deviation of pixel values, 3.78. de Jong et al. emphasizes the importance of determining sky background in 2D galaxy image fitting(de Jong (1996)). Häußler et al. compared several

methods to determine sky background and found that, for simulated galaxies, allowing GALFIT to determine the sky background appears as reliable as their isophotal sky method, for real galaxies, complex structures that deviate from profile assumptions may affect sky estimation(Häußler (2007)). Thus we let *BIE* determine the sky background in this study. The images of two exponential disks and spheroids are shown in Figure 9. And the all true input parameters for those galaxies are listed in Table 2. We use parallel chain algorithm for these galaxy images.

Isolated galaxies with low S/N from large ensemble image

We selected isolated galaxy images with low *S/N* , typically 2.0, which means that it's mean surface brightness is higher than sky surface brightness. We model them with the same way in section 4.2. The galaxies are barely seen by eye in the image center. We intentionally choose these extreme cases for characterizing the power and limit of *BIE* . Since the data quality of low *S/N* image is poor, signal and noise contribution to data are almost equal. The Bayesian inference has significant advantage compared with fitting technique in this case. In Bayesian approach, the equal or even more important question is not what the best parameters are, but, for a given data, what the relative significance of model parameter sets, one of which may be the true one, is. Since there is more severe degeneracy of parameters in low *S/N* data than high *S/N* data, it is hard to model the signal correctly. Then downhill method to find χ^2 minimum is highly probable to fail to reach the global minimum. However the Bayesian approach with MCMC can probe and reveal the global structure of probability distribution of parameters.

In addition to this, Bayesian approach has another advantage, which is the usage of prior information for model parameters. The parameter posterior is multiplication of the likelihood and the prior. Therefore the non-uniform prior can change the posterior as different from the posterior with uniform prior. In many cases, the likelihood dominates the prior. That is why the maximum likelihood method works well. However if the likelihood is not a strong function of parameters, the prior can severely affect the result. The carefully selected prior can be a very useful information or cause a bias in the result. The low *S/N* image is weakly informative and the likelihood may not be a very strong function of parameters.

We will present a summary of these results in a later report.

Multiple galaxies

Here, we study more complicatedly structured data, which may be highly degenerated in parameter space. If the galaxies are close to each other in the image, their light profile may be blended. Then the model with increased dimension in parameter space for those multiple galaxy may have a degeneracy and complicated topology in parameter posterior probability distribution. In this case, the standard χ^2 minimization technique using downhill method may fail to find global minimum.

In real galaxy survey data, we often encounter the case of blended light profile from multiple galaxies. The solution is to simultaneously fit those galaxies with multicomponent models or mask other galaxies with suitable masking map. GALFIT can do these. However neither of them can give us a way to fully understand the parameter probability space. Simultaneous fit by χ^2 minimization with downhill method may be stuck to local minimum and image masking is only helpful in a limited case. Bayesian inference on this case can be a very powerful way to probe parameter probability space and to determine the significance of other neighbor peaks and parameter uncertainties.

We show the example with three spheroids close to each other in Figure 10. The image size is 321×269 with plate scale, 0.03 [arcsec/pixel]. Thus the physical scale of this image is 9.63×8.07 arcsec^2. They are all spheroids with n=4, but other parameters are not known. These galaxies are modelled using single Sérsic profile with 7 free parameters each and fixed sky background level. Thus total free parameters are 21. MCMC is constructed by parallel chain algorithm.

The 1D marginalized distributions of n for these three galaxies are shown in Figure 11. x_c, y_c of each galaxies are shown on the top of each panel. The galaxy at upper right corner is masked.

GALFIT result for n of these galaxies(galaxy1, galaxy2, galaxy3) are $n_1 = 3.96 \pm 0.04, n_2 = 4.52 \pm 0.14, n_3 = 3.27 \pm 0.09$ respectively. the estimation for galaxy1 is reasonable, however even three standard deviation error for galaxy2 is not large enough to enclose 4.0, moreover, best fit Sérsic index for galaxy3 is very different from 4.0 and error is too small. *BIE* estimations for these galaxies are $n_1 = 3.87, n_2 = 4.06$ and $n_3 = 4.15$. They also have offsets from 4.0. However the errors indicate that the estimations are realistic. The true input $n = 4.0$ is enclosed within one standard deviation for galaxy2 and two standard deviations for galaxy1, galaxy3.

As we go higher dimensional parameter space, the parameter probability space becomes more complex and it is hard to reach the global minimum using downhill method since more parameters are probably degenerated. Bayesian MCMC shows the real power on this problem. Although *BIE* takes much longer time than the low dimensional case, it successfully samples the parameter posterior.

2-2.6 Conclusions

- BIE–GALFIT is both feasible and advantageous for studying galaxy parameters.
- Under a wide variety of conditions, the parallel chain algorithm is more effective that the other algorithms.
- M_{tot}, r_e and n are strongly correlated in spheroid and weakly correlated in disk.
- For high *S/N* image, the marginalized posterior has multi-mode. *BIE* error quotation is more realistic than GALFIT and all true parameters are enclosed by two standard except for b/a of disk310, which is enclosed by three standard deviations.
- For low *S/N* image, the marginalized posterior has multiple mode. The robust parameters, x_c, y_c, M_{tot}, of which posterior are smooth and uni-modal in high *S/N* image, have multi-mode. *BIE* error quotation is much better than GALFIT and all true parameters are enclosed by two standard deviations except for disk115, of which $r_e, b/a$ are not included in three standard deviations.
- We test the effect of prior to posterior. For low *S/N* image, the posterior is sensitive to prior. Since the results can be biased using prior, we should carefully choose the prior or uniform prior is better.
- For multiple galaxy with same n, GALFIT fails to fit simultaneously however *BIE* recover n which is close to the true n within two standard deviations.

2-3 SAMS–GALFIT

Semi-Analytic Models (SAMs) have been extensively used to study the formation and evolution of galaxies (e.g. Kauffmann et al., 1999; Somerville and Primack, 1999; Cole et al., 2000). In SAMs, one starts with a catalog of merger trees which describe the assembly of individual dark matter halos, and all other additional physical processes, e.g. gas cooling, star formation and feedback, AGN, galaxy mergers, etc., are also added into SAMs through empirical functions. With SAMs, one models the formation and evolution of a large number of galaxies. As the output of SAMs, a catalog of the properties of the modeled galaxies is produced. Base on the catalog many statistical properties, for instance luminosity function, Tully-Fisher relation, etc., can be obtained and such properties can be directly compared with observed sample galaxies. We build a SAM and incorporate Bayesian Inference and Markov Chain to explore the model parameters space.

2-3.1 Dark matter halo merger trees

We have developed sophisticated programs to generate the merger trees using Monte-Carlo methods. For a given halo mass M_2 at a given redshift z_2, we calculate the conditional probability for such a halo having a progenitor with with mass $M_1 < M_2$ at an earlier redshift z_1. We generate random numbers according to the conditional probability to allocate progenitor halo masses. We implement a number of merger tree generating schemes and study the properties of these different schemes. We briefly describe the methods and the comparison in this section.

2-3.2 Binary tree without accretion

In the binary tree without accretion, at each time step a halo either splits into two progenitors or does not fragment but retain its mass. In practice, to make the Monte-Carlo method more efficient, we change variables. In stead of redshift and mass, we choose $\omega \equiv \delta_c(z) = \delta_{c,0}/D(z)$ as our time variable, and $S(M) \equiv \sigma^2(M)$ as our mass variable. The probability for taking a new step ΔS in a time-step $\Delta\omega$ is

$$P(\Delta S, \Delta\omega)\mathrm{d}\Delta S = \frac{1}{\sqrt{2\pi}} \frac{\Delta\omega}{(\Delta S)^{3/2}} \exp\left[-\frac{(\Delta\omega)^2}{2\Delta S}\right] \mathrm{d}\Delta S. \quad (4)$$

If we make a change in variables further, $x \equiv \Delta\omega/(2\sqrt{\Delta S})$, the variable x becomes a Gaussian distribution with zero mean and unit variance. By generating a Gaussian random number, we produce a new mass for one of the two progenitor halos and the rest mass if any is assigned for the other progenitor.

In Figure 12, we show the conditional mass functions at four redshifts of halos generated by this merger tree scheme and compare them with theoretical model. Clearly, this scheme over produces halos in high redshifts. The reason for that is because this method does not consider another chancel for halos to gain their their mass, smooth accretion.

2-3.3 binary tree with accretion

In the binary tree with accretion, at each time step there is also a mount of mass which is smaller than the mass resolution is allocated as smooth accretion. We follow the method of Somerville et al. (1999) which is described below to generate such merger trees.

(i) Pick a mass M from the mass-weighted probability distribution equation 4. This mass can be anywhere in the range $0 \leq M \leq M_2$. If $M < M_{\rm res}$, we count it as accreted mass. If $M \geq M_{\rm res}$, we count it as a progenitor.

(ii) Compute the unallocated mass $\Delta M = M_2 - M$.

(iii) If the unallocated mass ΔM is larger than $M_{\rm res}$, then it may or may not contain a progenitor. To determine this, pick another mass M from the distribution, but with the restriction $M < \Delta M$. Depending on its mass, count it as accreted mass or a progenitor as before. In either case, subtract M from the mass reservoir.

(iv) Repeat this process until either

(a) the mass reservoir ΔM falls below the minimum halo mass $M_{\rm res}$, in which case it must abandon any aspirations of harboring a real progenitor and must be accreted mass, or

(b) we have found a total of two progenitors ($M > M_{\rm res}$), in which case the remaining mass is considered to be accreted mass in accord with our ansatz.

(v) Each progenitor now becomes a parent, we calculated a new time-step, and repeat the whole process.

We compare the realization of such a merger tree with theory in Figure 13. This method as pointed out by Somerville et al. (1999) under produces halos.

2-3.4 n-branch trees

In the n-branch tree, a halo can fragment into an arbitrary number of progenitors and in the same time some mass may be also allocated as accretion. To make n-branch trees, we follow the scheme of the binary tree with accretion but continue picking progenitor masses until the unallocated mass ΔM is less than $M_{\rm res}$, loosening requirement of having less than two progenitors used previously. In Figure 14, we show the conditional mass function of such merger trees. It follows the theory nicely especially at high redshifts, but has some over production in low mass bins at low redshifts.

2-3.5 Two-branch tree

We modify the program based on the one for the n-branch trees. To make such a two-branch tree, we only allow two or less progenitors to be generated. Once a progenitor halo M_1 and an amount of accretion mass $M_{\rm acc}$ have already been randomly picked up, the rest mass $\Delta M = M_2 - M_1 - M_{\rm acc}$ is directly assigned as the other progenitor halo. If both of the previously randomly picked two masses are larger than $M_{\rm res}$, these two masses are allocated as the progenitors and the rest, no matter it is smaller or larger than $M_{\rm res}$, is assigned to be accretion. We show the conditional mass function in Figure 15. The realization recover the theory nicely at all redshifts.

2-3.6 Galaxy formation model

We developed a semi-analytic galaxy formation model base on halo merger trees, including recipes for hot gas and radiative cooling, star formation and feedback, chemical evolution, stellar population synthesis, and galaxy mergers.

After two halos merge, the galaxies contained by the halos are expected to merge in a timescale about dynamical friction timescale. In observation, galaxy mergers are found to be triggers of star bursts and structure transformations. The physical processes involved in mergers depend on the

mass ratio of the two progenitors. When a small halo mergers into a big halo, dynamical friction and tidal stripping work; when two halos with comparable masses merger, the old structure in each of the objects will be destroyed. Because of this, people categorize mergers into two groups, major mergers and minor mergers. Usually, people use a characteristic mass ratio to classify major/minor mergers – if the smaller halo has mass lower than 1/3 (or 0.3) of the bigger halo, the merger is classified as a minor merger; otherwise it is treated as a major merger. For different mergers, SAMs have completely different treatments.

In our model, we assume the most massive galaxy in a newly merged halo becomes the central galaxy. In stead of stick on 0.3 as the merger threshold, we set it to be a free parameter. When two halos merge, we start the merger clocks for all galaxies other than the central galaxy in the halos. In merger timescale, the galaxy will merge into the central galaxy. For minor mergers, we add all the gas mass in the small galaxy into the primary galaxy and add its stellar mass into the bulge of the primary galaxy. For major mergers, we add dark matter, hot gas, cold gas and stellar mass from the two galaxies together. The shape of the remnant is spheroid. A star burst is assumed to be associated with a merger. It converts a fraction of cold gas into stars instantaneously.

In summary, we have 13 free parameters in total in our current model. They are:

- $V_{\rm cut}$: cooling cut-off
- α_0: amplitude of star formation efficiency
- $V_{\rm sf}$: turn-over circular velocity of star formation efficiency
- n: power index of circular velocity dependence of star formation efficiency
- ε_0: amplitude of SN feedback
- $V_{\rm fb}$: turn-over circular velocity of SN feedback
- $\beta_{\rm fb}$: power index of circular velocity dependence of SN feedback
- $f_{\rm merg}$: merger threshold for mass ratio
- $k_{\rm merg}$: merger timescale in units of dynamical friction time
- $\beta_{\rm burst,min}$: star burst amplitude for minor mergers
- $\alpha_{\rm burst,min}$: power index of mass ratio dependence of star burst in minor mergers
- $\beta_{\rm burst,maj}$: star burst amplitude for major mergers
- $\alpha_{\rm burst,maj}$: power index of mass ratio dependence of star burst in major mergers

With properly chosen parameters, our model reproduces the galaxy luminosity functions at $z = 0$ in K-band and SDSS-bands presented in published literatures (Somerville and Primack, 1999; Kang et al., 2005; Croton et al., 2006).

2-3.7 Preliminary results

We have jointed our SAM with BIE. The likelihood of a parameter set is evaluated by comparing the predicted K-band luminosity function with observation. Firstly, we use mock luminosity function which is generated by the model with a given set of parameters. We release only a few parameters to study the behavior of the BIE-SAM. We found that 1) the BIE-SAM works nicely for searching the true parameter if only one parameter was set to be free, 2) some parameters in the model are highly degenerate, so that the BIE-SAM hardly converges to true parameters if those parameters were released. 3) With a big number of free parameters and big range of prior for the parameters, the model is still not able to reproduce the observed K-band luminosity function.

Single parameter models

Using our model, with fixed parameters we generate a mock K-band luminosity function (solid line in Fig. 16). To test our BIE-SAM, we set only one of those 13 parameters free to fit the mock luminosity function. The tests show that our code converges to the true value very quickly.

multiple parameter fit

We now increase the dimensionality of the models and use the BIE-SAM to search the true parameter values by comparing the K-band luminosity function with the mock. In the tests which involve α_0 and V_{sf} (the amplitude and the turn-over circular velocity of star formation efficiency), a strong degeneracy shows up. The code spends long time to converge into the true value. Figure 17 shows the posterior distribution of these two parameters from a test which only has these two free parameters; Figure 18 shows the same distribution from a three free parameter test (these two and n). In both cases, a number of modes and a long degeneracy valley present. Figure6 shows a predicted K-band luminosity function with α_0 and V_{sf} very different from the true value. These result show clearly that the SAM has strong degeneracy among the parameters.
/emphfitting observed luminosity function

We adopt our BIE-SAM to fit the real observed K-band luminosity function with all the 13 free parameters. With very broad prior distribution of the parameters, the model is still not able to fit the data well, especially in the faint end (see Fig. 20). Although we may have poor mixing problem, the failure to reproduce the faint end slope is a robust result.

3 Milestones for Year 2

1. Statistical & MCMC development
   Continued testing and benchmarking of Bayesian model selection methods, including Bayes factors and Reversible Jump algorithms. further MP optimization.

2. Persistence subsystem
   We anticipate a working implementation of our persistence subsystem. this will support recording computations and the relationships between inputs and outputs, in a research log, so that one can always go back and determine the origin of data and how it was processed, replaying from a previous state, but with different commands or parameters—what we call

what-if exploration. One can always go back to some previous time or step and compute forward in new directions, and checkpointing and recovery.

3. Star count modeling with SQL databases
   We anticipate beginning on the final project, star count analyses of the Galaxy and its satellites, interfacing to 2MASS and/or SDSS catalogs with large SQL databases. Because our own galaxy and nearby local group companions can be studied in careful detail, we can probe the features of Milky Way structure to refine theories of galactic interaction. We anticipate working with a new graduate student colleague in the upcoming year.

4. BIE–GALFIT
   We are currently testing idealized data sets and benchmarking the efficiency of BIE in hypothesis testing. During Year 2, we anticipate moving on to inference on real astronomical data and publications demonstrating the methods and application. In addition, we are currently implementing computational optimizations that will allow production analysis. We anticipate a full-up stand-alone version of BIE to be released to the public in the upcoming year.

5. Semi-analytic models
   We will continue to improve the performance of our SAM implementation and test its performance. In Year 2, we plan to apply the Bayes Factor methodology to test specific hypotheses about the importance of various parameters in the underlying physical mechanisms or used to test the effect of different physical hypotheses, i.e., different parameterizations and combinations of physical processes, without the constraint that their prescriptions be nested. We have begun discussions with other SAM practitioners hope to test BIE with their codes as well.

4 Bibliography

Blanton, M. R., Hogg, D. W., Bahcall, N. A., Baldry, I. K., Brinkmann, J., Csabai, I., Eisenstein, D., Fukugita, M., Gunn, J. E., Ivezić, Ž., Lamb, D. Q., Lupton, R. H., Loveday, J., Munn, J. A., Nichol, R. C., Okamura, S., Schlegel, D. J., Shimasaku, K., Strauss, M. A., Vogeley, M. S., and Weinberg, D. H. 2003, ApJ, 594, 186.

Cole, S., Lacey, C. G., Baugh, C. M., and Frenk, C. S. 2000, MNRAS, 319, 168.

Croton, D. J., Norberg, P., Gaztanaga, E., and Baugh, C. M. 2006, ArXiv Astrophysics e-prints.

de Jong, R. S. 1996, A&A Suppl., 118, 557.

Graham, A. W. and Driver, S. P. 2005, Publications of the Astronomical Society of Australia, 22, 118.

Häußler, B. 2007, *Ph.D. thesis*, Max-Planck-Institut für Astronomie, Heidelberg, Germany.

Kang, X., Jing, Y. P., Mo, H. J., and Börner, G. 2005, ApJ, 631, 21.

Kauffmann, G., Colberg, J. M., Diaferio, A., and White, S. D. M. 1999, MNRAS, 303, 188.

MacArthur, L. A., Courteau, S., and Holtzman, J. A. 2003, ApJ, 582, 689.

Peng, C. Y., Ho, L. C., Impey, C. D., and Rix, H.-W. 2002, AJ, 124, 266.

Somerville, R. S. and Primack, J. R. 1999, MNRAS, 310, 1087.

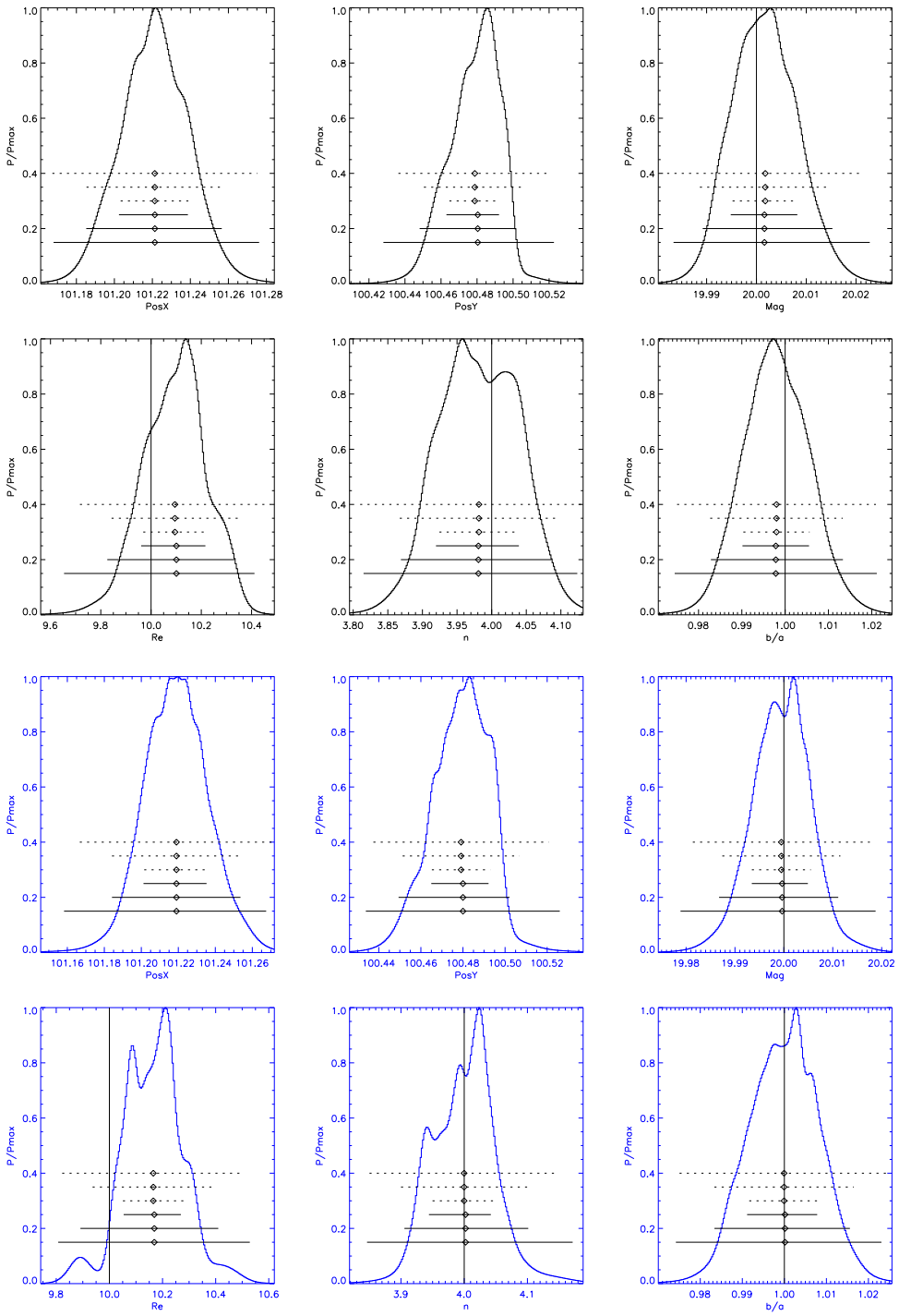

Figure 6: 1D marginalized parameter posterior probability distribution for sph0 with arbitrary normalized factor. Black line corresponds to the sampling by simulated tempering algorithm and blue line corresponds to the sampling by parallel chain algorithm.

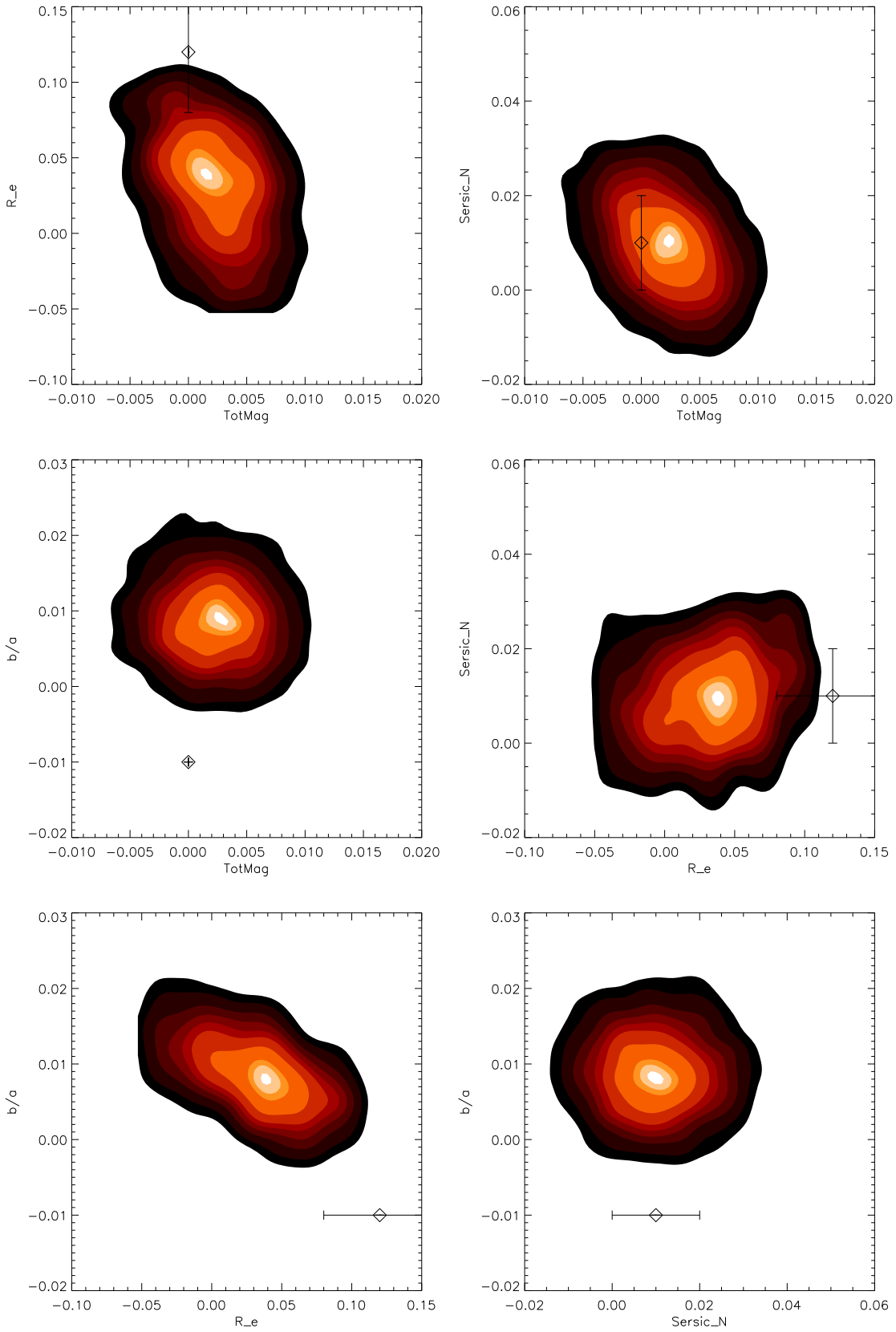

Figure 7: Two dimensional marginalized distribution of parameters for disk0. Sampling algorithm is Metropolis-Hastings. There are slight correlations between M_{tot}, r_e and n.

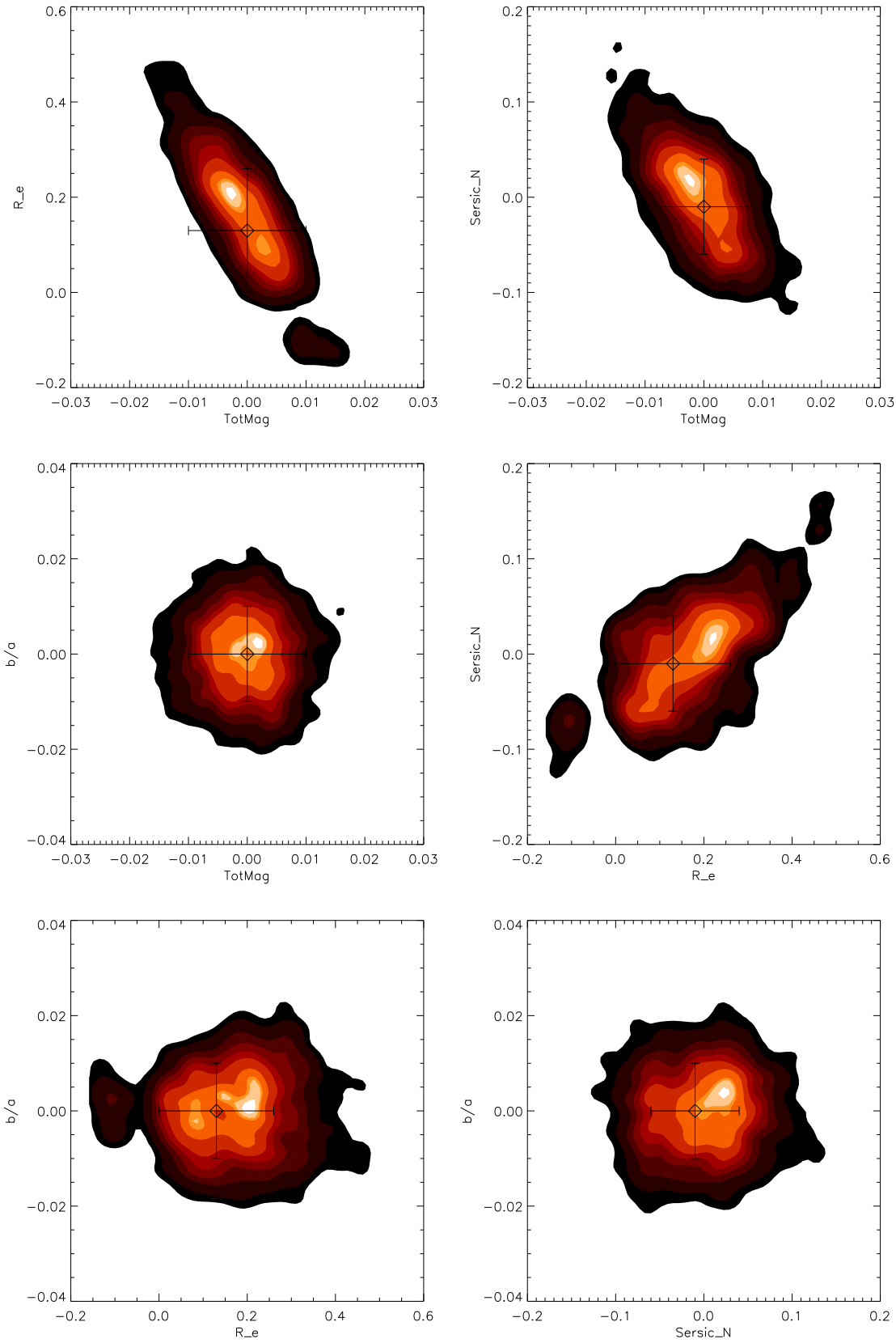

Figure 8: Two dimensional marginalized distribution of parameters for sph0. Sampling algorithm is Metropolis-Hastings. There are clear correlations between M_{tot}, r_e and n.

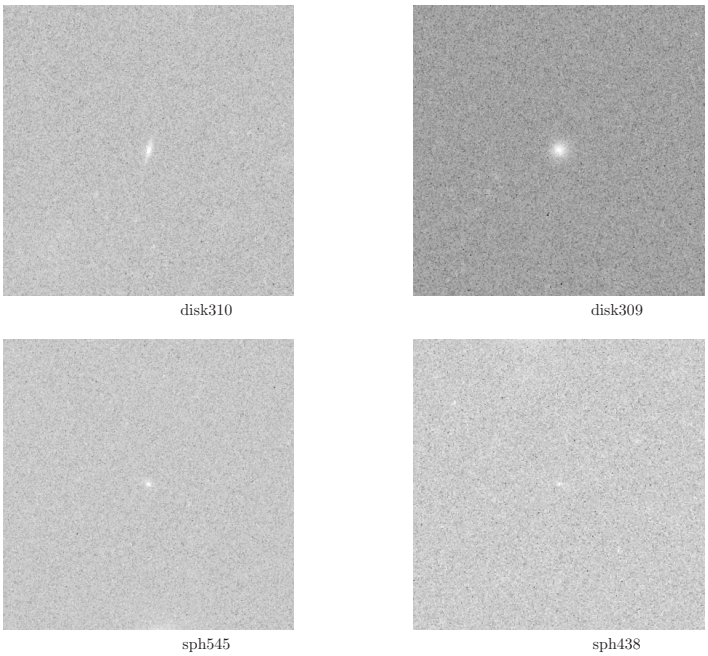

Figure 9: High *S/N* galaxy images. The top two is exponential disk and the bottom two is spheroid.

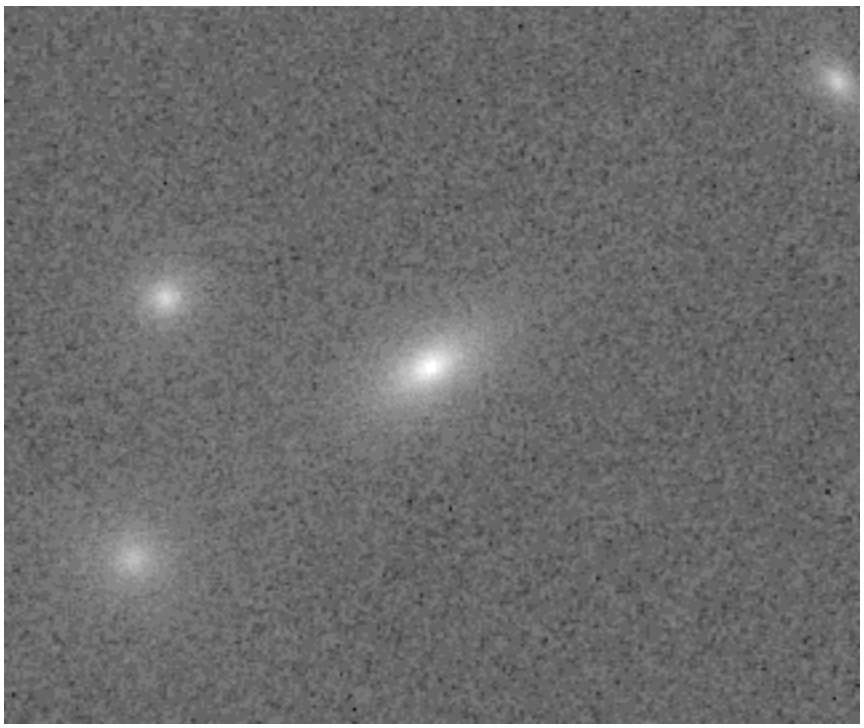

Figure 10: The simulated image of multiple galaxies with $n = 4$.

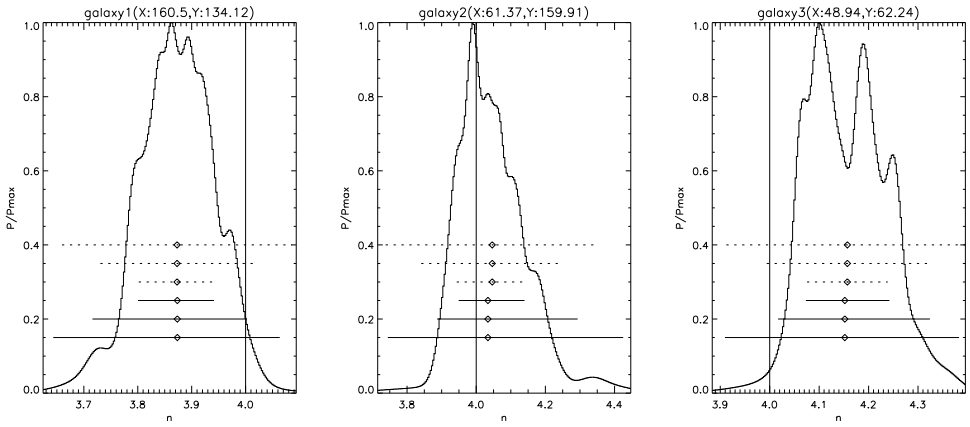

Figure 11: The marginalized distribution of *n* for the galaxies

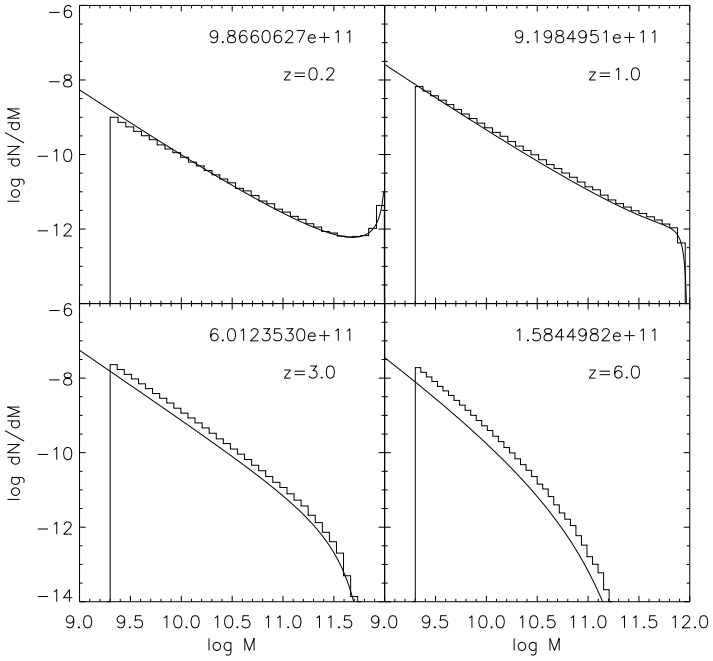

Figure 12: The conditional mass functions of the binary merger trees without accretion at the redshifts indicated at the upper-right corner of each panel. The solid lines are the eps predictions.

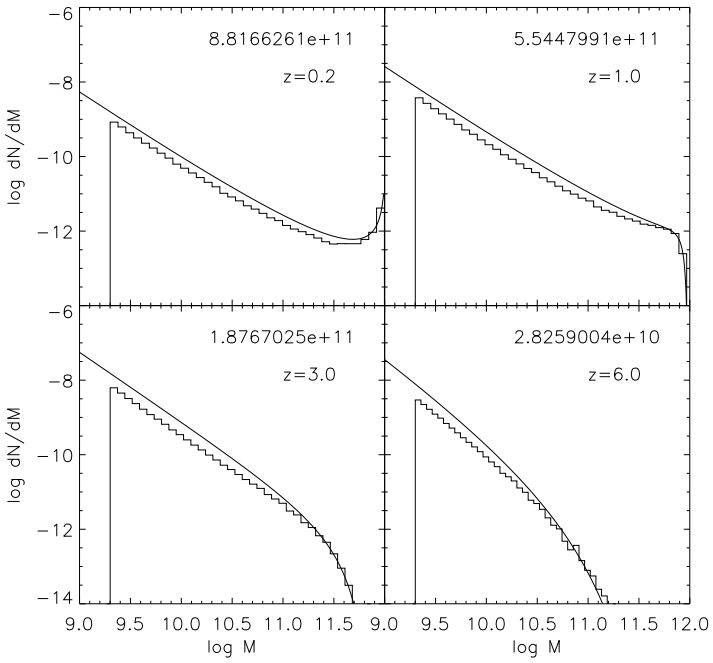

Figure 13: The conditional mass functions of the binary merger trees with accretion.

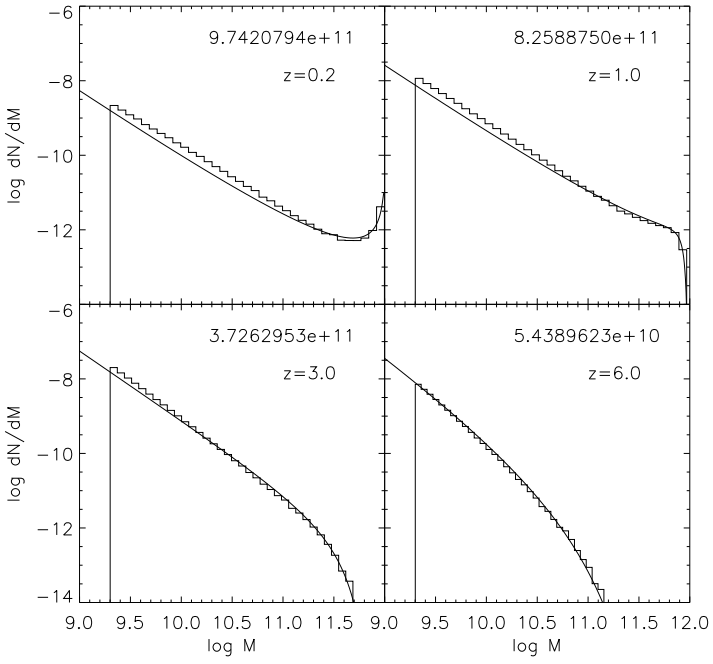

Figure 14: The conditional mass functions of the n-branch merger trees.

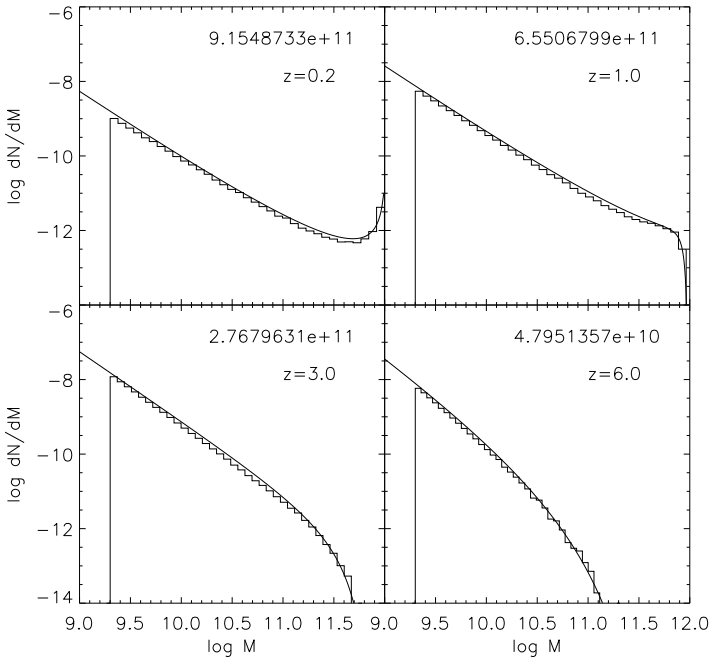

Figure 15: The conditional mass functions of the two-branch merger trees.

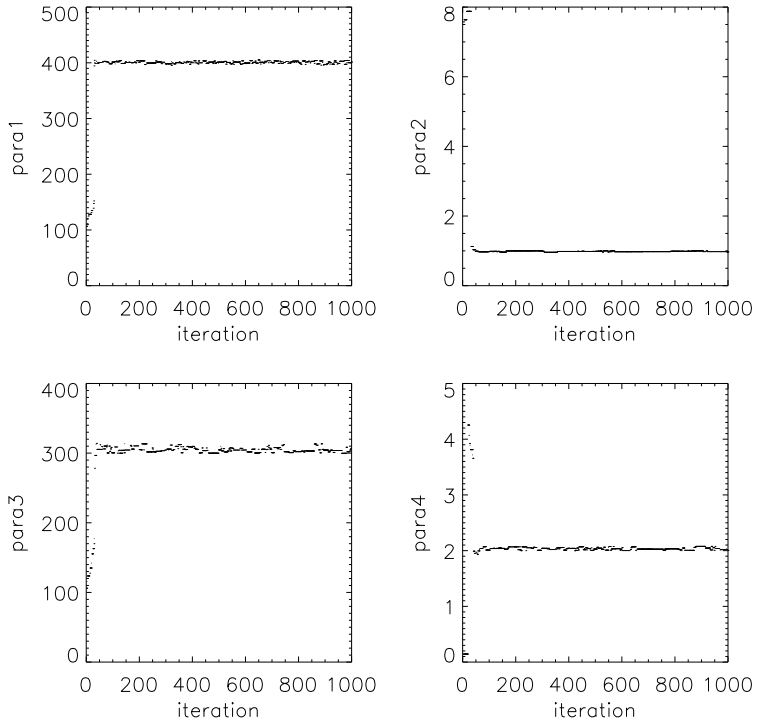

Figure 16: Each panel shows a MCMC trace of the free parameter of a single free parameter test. The true value of the parameters are, respectively, 400, 1, 300, and 2 for $V_{\rm cut}$, α_0, $V_{\rm sf}$, and n.

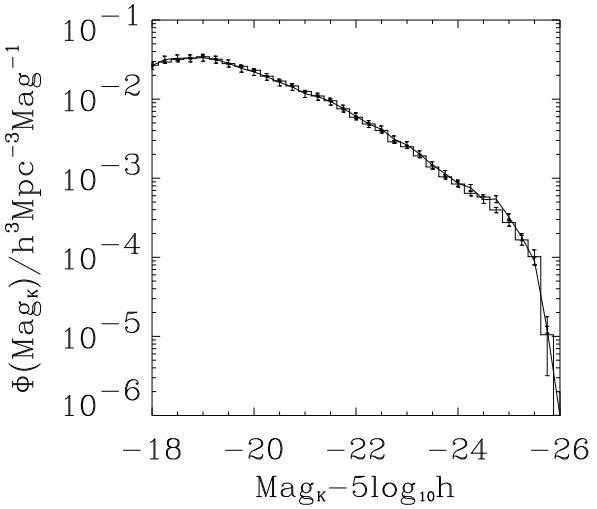

Figure 17: The predicted K-band luminosity function compared with the mock. To produce the shown curve, we adopt $\alpha_0 = 9.0$ and $V_{\rm sf} = 900$, which are very different from the true values ($\alpha_0 = 1.0$ and $V_{\rm sf} = 300$).

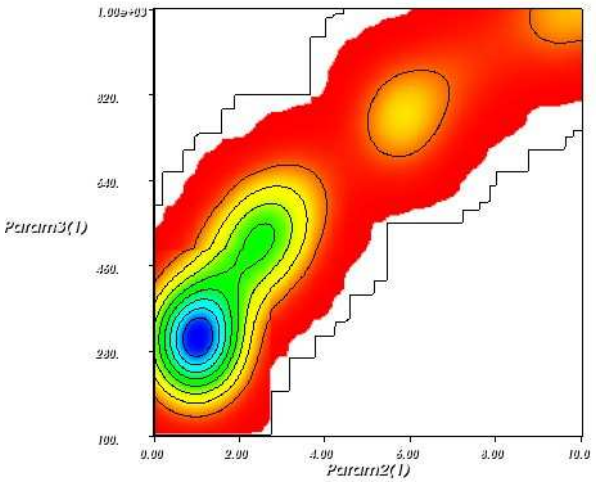

Figure 18: Posterior distribution of α_0 (x-axis) and $V_{\rm sf}$ (y-axis) from a two free parameter test. The contours enclose from 10% to 90% of the distribution with 10% increment.

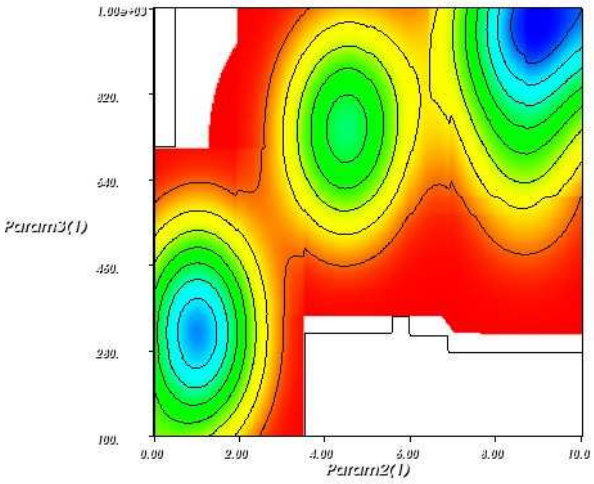

Figure 19: Posterior distribution of α_0 (x-axis) and V_{sf} (y-axis) from a three free parameter test.

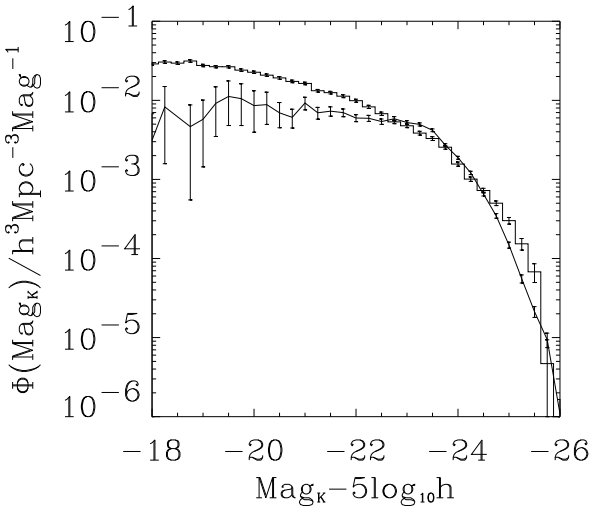

Figure 20: The K-band luminosity function produced by the best fitting parameters (in histogram) is compared with real observation (in solid curve, Cole et al. 2001). In big range of freedom of the parameters, the model can not reproduce the luminosity function in the faint end.