Bayesian Analysis of Reverberation Mapping Data

Brendon J. Brewer

Abstract Reverberation mapping is a powerful technique for studying the broad line regions (BLR) and the masses of the central black holes in distant active galactic nuclei (AGN). By monitoring the temporal variations of the continuum emission and the broad emission lines, it is possible to measure the size of the broad line region, and combining this with velocity information from the line widths yields an estimator for the black hole mass. However, this estimator depends on an unknown dimensionless proportionality constant called the virial coefficient. Recently, we have developed an alternative, direct approach to analyzing reverberation mapping data that infers details of the astrophysical situation from the data, bypassing the need for a virial coefficient and providing information about the physical configuration of the BLR. In this contribution I will outline the method and discuss how it differs from traditional reverberation mapping analysis.

1 Introduction

Reverberation Mapping is an important technique for measuring the masses of the black holes that power Active Galactic Nuclei (AGN). The distribution of matter surrounding the black hole can also be studied, yielding constraints on AGN physics [3]. The measurement of black hole masses enables the study of the relations between supermassive black holes and their host galaxies [9].

The method makes use of the variability of the central continuum source [11], and the subsequent response of the broad lines, emitted by orbiting gas (Figure 1). In the traditional method, the typical time delay, or lag, τ between the continuum variations and the broad line response is measured by cross correlating the continuum light curve with the line flux light curve. The time delay τ measures the size of

Brendon J. Brewer
Department of Physics, University of California, Santa Barbara
e-mail: brewer@physics.ucsb.edu

the broad line region (BLR), and the width of the broad lines, σ_l gives their typical orbital velocities. These measurements combined can give an estimate of the black hole mass, according to the formula [18]:

$$M_{BH} = f\frac{\sigma_l^2 c\tau}{G} \quad (1)$$

Here, the black hole mass is given in terms of physical constants and measurable quantities, but also depends on the dimensionless *virial coefficient* f. The virial coefficient is meant to encode the effect of the geometrical configuration of the BLR: for example, whether it is spherically symmetric, disky and face-on (this would imply a high value for f), disky and edge-on (implying low f), or whatever. However, if the value of f for for any individual system is unknown, the black hole mass inherits this uncertainty. Typically, the distribution of f values for a population of AGN is used to indicate the uncertainty, implying that the uncertainty in an individual black hole mass is influenced by the diversity of f values across all systems, rather than on the data for that particular system.

The standard reverberation mapping procedure has been used with great success [2, 7, 19, 1], and has provided the basis for the calibration of less costly methods [8]. However, reverberation mapping data do not really arrive to us in the form of a value for τ and σ_l. These numbers are the results of procedures performed on the full data set, and there is no reason to think that they are sufficient statistics. The full data set consists of time series of the observed continuum flux, and spectral time series of the broad line response (i.e. the shape and flux of the chosen broad emission line, over time). To make the most of the data, we should perform an inference calculation based on the full data set. For more details about our approach, please see [17] and [5].

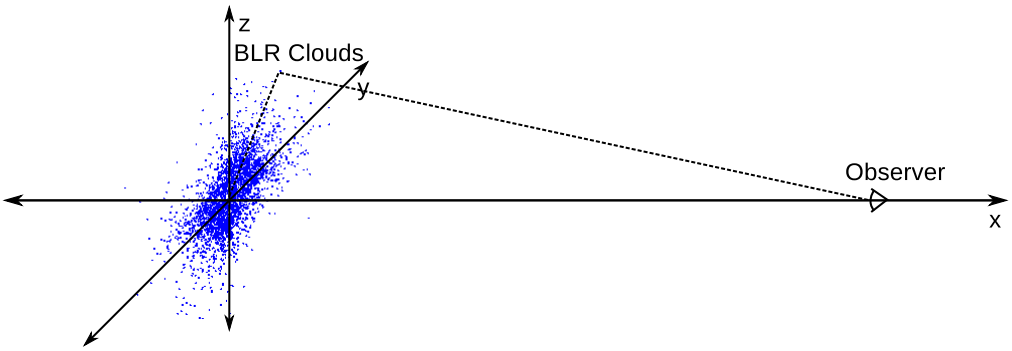

Fig. 1 The distribution of extra path lengths the light must travel from the central engine to a BLR cloud and then to the observer is the cause of the delayed response of the emission-line flux, and the variations in line shape. The distribution of BLR gas in this diagram corresponds to a probable configuration inferred from the Arp 151 data.

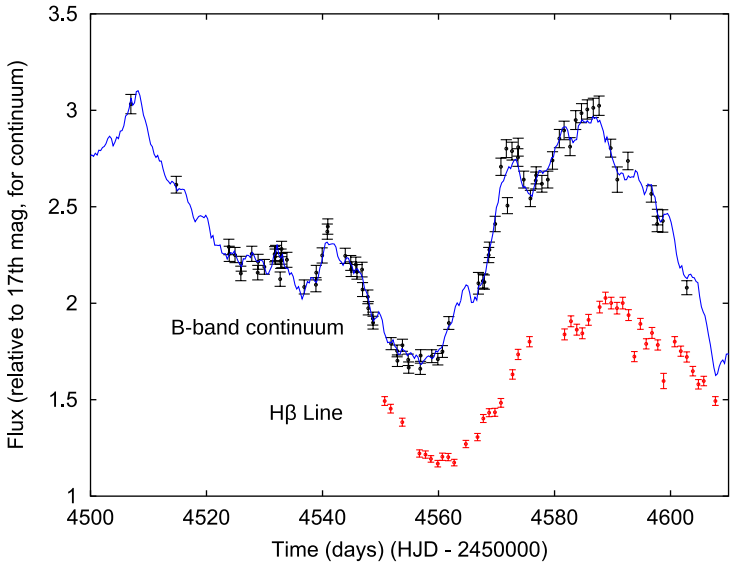

Fig. 2 Continuum flux time series for Arp 151, observed as part of the Lick AGN Monitoring Project (LAMP), and the corresponding Hβ flux time series. Note that the full Hβ data set actually consists of a spectrum at each time. i.e. there is also shape information, in addition to the flux plotted here.

2 An Inference Approach

To best exploit the data, we should directly answer the question "what possible physical situations are plausible, in light of the data?". We begin by defining a hypothesis space and prior probabilities over that space, describing possible physical scenarios that might describe the system. In particular, we consider the following properties to be unknown: the black hole mass, the physical configuration of the relevant broad line region gas, the velocity distribution of the gas, and the full continuum light curve $y_{\rm cont}(t)$ at all times, not just the measured times. All of these properties can be described by parameters denoted collectively by θ. We then define sampling distributions, $p(D|\theta)$, describing what data we would expect to observe if a particular value for θ, i.e. a particular physical situation were known to be true. The sampling distributions describe prior knowledge about how the data are related to the parameters. By Bayes' rule, the posterior distribution for the parameters given the data is then given by:

$$p(\theta|D=D^*) \propto p(\theta)p(D|\theta)|_{D=D^*} \tag{2}$$

where D^* is the actual data set that was observed. The posterior distribution describes what is known about the parameters after taking into account the data, and suitable summaries can be derived from the full distribution (best estimates and error bars, credible intervals, etc).

Note that, in traditional reverberation mapping analysis, one obtains either the transfer function (the distribution of lags required to map the continuum variations into the line variations), or the mean lag τ from the cross-correlation of the continuum time series and the line flux time series. The value for the virial coefficient f is then estimated based on separate, population-based criteria. However, in reality, the transfer function and the appropriate value of f both depend on the geometry of the BLR, and thus there should be a prior correlation between them. In other words, if we are capable of reconstructing the transfer function from data, then we should also be more informed about the appropriate f-value. In our method, all of this happens automatically as a result of inferring the physical situation from the data.

2.1 The Hypothesis Space

In our method, we numerically represent the geometry and kinematics of the BLR by a large number of point-like clouds (although we also have an implementation that describes the density function on a spatial grid). We do not aim to infer the position and velocity of every cloud from the data. Instead we parameterize the distribution of the clouds by a small number of hyperparameters, and infer those hyperparameters. To generate a description of a 3D distribution of BLR clouds, we start by generating an axisymmetric distribution in the *x*-*y* plane, and then apply rotations to "puff up" the model into a 3D configuration. Finally, we weight the clouds by a non-axisymmetric illumination function to model non-axisymmetric distributions of gas.

In 2D, the position of a cloud is assigned in plane polar coordinates. The distance of a cloud from the black hole is prescribed according to

$$r = F\mu + (1-F)\mathscr{G} \tag{3}$$

where $F \in [0,1]$ and $\mathscr{G}$ is drawn from a gamma distribution with mean μ and standard deviation $\beta\mu$. With this prescription μ is the overall mean radius, F is the fraction of the mean radius that is due to the hard lower limit, and $\beta \in [0,1]$ describes the shape of the distribution; $\beta \approx 1$ is an exponential distribution, and $\beta \approx 0$ is a narrow normal distribution[1]. The polar coordinate ϕ of a cloud is chosen uniformly from $[0, 2\pi]$.

We then assign cloud velocities in a probabilistic manner (note that we can assign multiple velocities to each cloud in order to improve sampling of the phase space in a computationally efficient way; throughout this study we adopted 100 velocities for each of the 1,000 clouds) that depends on one additional hyperparameter. We assume that the only force acting on the BLR clouds is gravity from the central black hole. This is clearly a simplification, but enables us to create a parameterization for the phase space distribution of the clouds that depends on the black hole mass.

[1] We chose μ and β as the coordinates on the gamma distribution parameter space for clarity of meaning.

The total energy of a cloud at a distance r from the black hole, moving with angular momentum L, is given by

$$E = \frac{1}{2}\left(\dot{r}^2 + \frac{L^2}{r^2}\right) - \frac{GM_{BH}}{r}, \tag{4}$$

which has the minimum possible value

$$E_{\min} = -\frac{GM_{BH}}{r}. \tag{5}$$

If we knew the position, energy, and angular momentum of a cloud, we could solve for the radial velocity,

$$\dot{r} = \pm\sqrt{2\left(E + \frac{GM_{BH}}{r}\right) - \frac{L^2}{r^2}}. \tag{6}$$

We choose the negative (inbound) solution with probability q and the outbound with probability $1-q$, another free parameter. For solutions to exist, the angular momentum must satisfy

$$L^2 \leq L^2_{\max} = 2r^2\left(E + \frac{GM_{BH}}{r}\right). \tag{7}$$

Circular orbits are obtained if we set the energy and angular momentum to

$$E_{\rm circ} = -\frac{1}{2}\frac{GM_{BH}}{r} \text{ and} \tag{8}$$

$$L_{\rm circ} = \pm L_{\max}. \tag{9}$$

To get elliptical orbits, instead of assigning E and L the exact circular values above, we assign them at random from the following probability distributions:

$$E = \left(\frac{1}{1+\exp(-\chi)}\right)E_{\min}, \tag{10}$$

$$\text{where } \chi \sim \mathcal{N}(0,\lambda^2), \tag{11}$$

$$\text{and} \tag{12}$$

$$p(L) \propto \exp\left(\frac{|L|}{\lambda}\right), \quad |L| < L_{\max}. \tag{13}$$

These probability distributions are centered around the values for circular velocities, but the parameter λ describes the dispersion, or how noncircular typical orbits will be. The circular orbit formulae are reproduced when $\lambda \to 0$. The probability distributions for E and L given three different values for λ are shown in Figure 3, along with the corresponding line shapes.

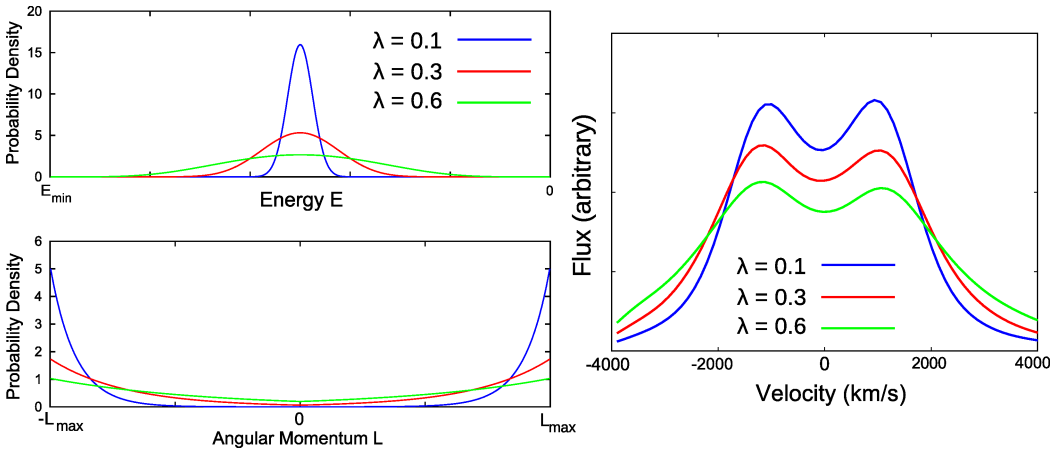

Fig. 3 Probability distributions for energy and angular momentum that generate elliptical orbits for the clouds. The effects on the predicted line shape are illustrated for an edge-on disk for three different values of the "noncircularity parameter" λ.

We then rotate the models (positions and velocities of the clouds) by an appropriate distribution of angles to generate an axisymmetric distribution of angular momentum vectors. The first rotation is about the y axis by a small random angle; the typical size of these angles determines the opening angle $\phi_{\rm open}$ of the disk or torus. We then rotate around the z axis by random angles to restore the axisymmetry of the model. Finally, we rotate again about the y axis, by the inclination angle $\phi_{\rm inc}$ to model the inclination of the system with respect to the line of sight.

To obtain non-axisymmetric models, in order to reproduce possible asymmetries in the emission line shape, we weight each cloud by a simple spatially varying illumination function. In spherical polar coordinates, this function is

$$W = \frac{1}{2} + \kappa \cos\phi, \tag{14}$$

where $\kappa \in [-\frac{1}{2}, \frac{1}{2}]$ is a free parameter and ϕ is the azimuthal angle. Positive κ illuminates the front portion of the BLR, negative κ illuminates the back. We also implemented, as a secondary check, a model with a different (linear in x) functional form for the illumination. For the application to Arp 151 discussed in this contribution, the linear model provided a poorer fit to the data. When simulating mock data, we also add the narrow-line component as a narrow line which does not respond to the continuum variations.

2.2 The Prior Distribution

Before we take into account a data set, we must describe our prior knowledge of the parameters. The parameter list is

$$\left\{M_{BH}, \mu, F, \beta, \phi_{\rm open}, \phi_{\rm inc}, \kappa, q, \lambda, y_{\rm cont}(t)\right\} \tag{15}$$

These are the black hole mass, the mean radius of the BLR, the fraction of the mean radius that is caused by a lower cutoff, the shape parameter of the radial BLR profile, the opening angle of the BLR, the inclination angle, the strength of the front/back asymmetry, the fraction of inflowing clouds, the departure from circularity of the orbits, and the continuum light curve at all times.

The prior distribution for $y_{\rm cont}(t)$, before conditioning on the continuum data, was assigned to be a Gaussian process with mean y_0 and covariance function

$$C(\Delta t) = \sigma^2 \exp\left(-\left|\frac{\Delta t}{L}\right|^{\alpha}\right) \tag{16}$$

which introduces additional hyperparameters σ (the typical variation amplitude), L (the typical variation timescale) and $\alpha \in [1,2]$ (describing the smoothness of the variations). This Gaussian process model, in the case $\alpha = 1$ has been studied extensively in the context of AGN variability [11, 12, 13] and reverberation mapping [22].

This expands the full parameter list to

$$\left\{M_{BH}, \mu, F, \beta, \phi_{\rm open}, \phi_{\rm inc}, \kappa, q, \lambda, y_{\rm cont}(t), y_0, \sigma, L, \alpha, A\right\} \tag{17}$$

where we have also added a response coefficient A, whose purpose will be described in Section 2.3. The priors for all of these parameters are listed in Table 1.

2.3 The Sampling Distribution

The sampling distribution $p(D|\theta)$ can be assigned by considering mock data. Specifically, if we knew all of the details of the physical situation, we would be able to predict mock noise-free spectra, up to the effects of observational noise. The sampling distribution can then be assigned as a multivariate normal distribution, with mean values equal to the simulated noise-free data, and variances given by the supplied "error-bars" on the data. Of course, in reality, our model does not account for all effects apart from noise, but this common assumption provides a useful starting point.

To predict mock data, we apply the simple physical model of the broad line region discussed above. The BLR clouds positions $\{x_i, y_i, z_i\}$ and velocities $\{u_i, v_i, w_i\}$ are assigned. The time delay associated with each cloud is calculated using

Parameter	Prior
M_{BH}	Uniform given μ
μ	log-Uniform
F	Uniform(0,1)
β	Uniform(0,1)
$\phi_{\rm open}$	Uniform(0, π)
$\phi_{\rm inc}$	Uniform(0, $\pi/2$)
κ	Uniform(-1/2, 1/2)
q	Uniform(0,1)
λ	Uniform(0,1)
$y_{\rm cont}(t)$	Gaussian Process(y_0, Eqn 16)
y_0	Uniform
σ	log-Uniform
L	log-Uniform($10^{-3}\times$ data length, data length)
α	Uniform(1, 2)
A	log-Uniform

Table 1 List of the prior probability distributions for the parameters. If any prior refers to another parameter, it is a conditional prior given the other parameter. If no bounds are given for a uniform or log-Uniform ("Jeffreys") prior, broad bounds were used.

$$\tau_i = \sqrt{x_i^2 + y_i^2 + z_i^2} - x_i \tag{18}$$

as the observer is assumed to lie on the x-axis at a large distance.

The predicted broad line spectrum as a function of velocity at time t is given by

$$S(v,t) = A\sum_{i=1}^{N} w_i \delta(v - u_i) y_{\rm cont}(t - \tau_i) \tag{19}$$

where A is a proportionality constant, and w_i is the weight of cloud i. This equation states that each cloud produces a spike in the spectrum at the cloud's recession velocity, and the flux of the spike is proportional to the continuum flux at a time τ_i in the past. To produce a smooth predicted spectrum, the spikes are convolved with an instrumental blurring kernel, and then binned at the same resolution as the actual data.

3 Sampling the Posterior Distribution

With any complex Bayesian Inference problem in more than a few dimensions, the best way to summarise the posterior distribution is to generate random samples from it. From the samples, the marginal distributions of parameters and the posterior probability of any hypothesis of interest can be computed trivially. One of the most popular Markov Chain Monte Carlo (MCMC) methods for drawing samples from a probability distribution, given the ability to compute the probability density at any point, is the Metropolis-Hastings algorithm [14]. However, if the posterior

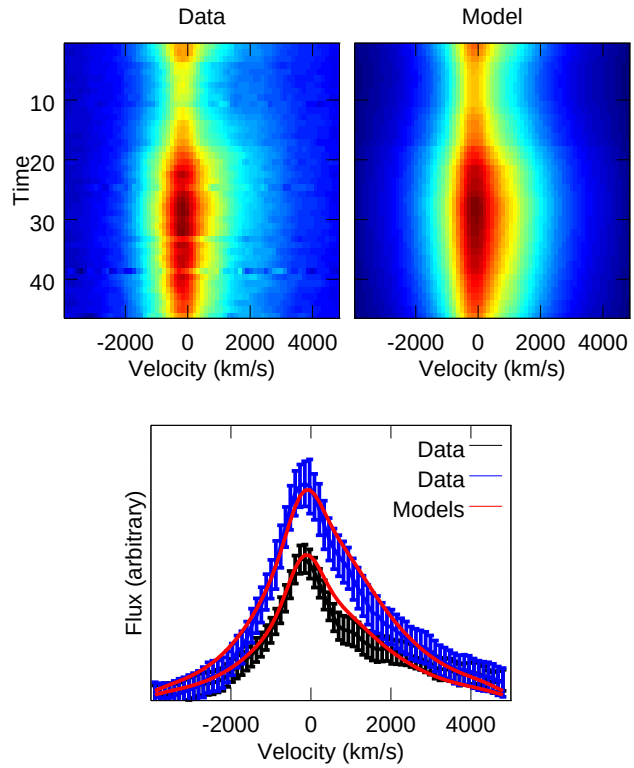

Fig. 4 *Top left*: Measured spectrum of the broad Hβ emission line as a function of epoch; these are the data used for our inference. *Top right*: Model-predicted spectrum as a function of time, using parameter values chosen at random from the posterior distribution. The major features of the data (time variation, line widths) are reproduced by our model. *Bottom*: Example of the spectral line shape at two times, along with model fits.

distribution contains multiple modes or strong correlations, the Metropolis-Hastings algorithm can run into difficulty, becoming trapped for long times in a single mode or a small part of a degeneracy curve. In this situation, the theoretical guarantee that the Metropolis-Hastings algorithm will eventually converge is not useful.

Many techniques exist for overcoming this problem. A large class of techniques is based on *annealing* the likelihood function, i.e. raising it to some power less than one, to soften the effect of the data. Another recent method is Nested Sampling, which has the advantage that no annealing schedule needs to be specified [20]. However, the standard implementation of Nested Sampling destroys the diversity of the initial population of models prematurely, because of its copying operation.

To implement our parameter space exploration, we used Diffusive Nested Sampling [6] (DNS). DNS is an efficient[2] MCMC-based version of Nested Sampling that works by exploring a mixture of the prior and a sequence of more constrained distributions that are created as the algorithm proceeds. For the purposes of exploring the parameter space efficiently, the main advantage of DNS is that it continually revisits the prior, "forgetting" its location along degeneracy curves. See Figure 5 for an illustration.

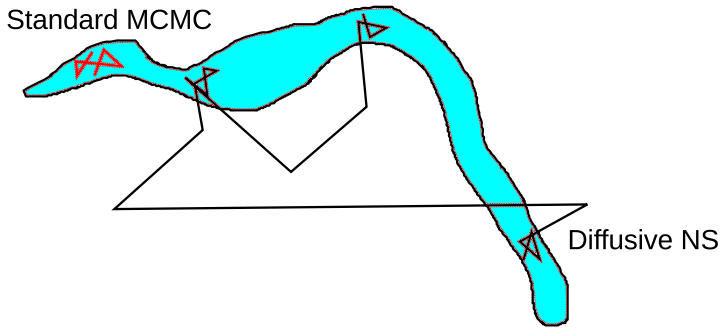

Fig. 5 An illustration of MCMC sampling of a strongly dependent probability distribution. Standard Metropolis-Hastings can easily become stuck in one part of the parameter space if the proposal distribution is not aware of the shape of the target distribution. Diffusive NS, on the other hand, samples a mixture of constrained and unconstrained distributions and is able to "forget" its location and try again on a different part of the target distribution.

4 Arp 151

We have applied our modelling strategy to Lick AGN Monitoring Project (LAMP) data for the active galaxy Arp 151, and are in the process of applying it to more AGNs from this sample. The results are discussed in more detail in [5]. Here, we discuss an issue relating to using simplistic models to fit data.

4.1 Systematic Errors and Nested Sampling

In the application to the Arp 151 data, it was discovered that our BLR model could not fit the data to within the very small supplied error bars. Of course, the best solution to this is to develop more complex models with more freedom and more

[2] The error on the log evidence is a factor of four less in DNS compared to the original MCMC-based Nested Sampling method, for the same computational effort.

realistic physics, so that the data could be fit more exactly. However, the fit of this simple model to the data is still of some utility. In an attempt to obtain reliable inference of the black hole mass from an oversimplified model, we experimented with decreasing the constraining effect of the data, making the posterior distribution more conservative, i.e. more like the prior. Figure 6 shows the posterior distribution for the black hole mass as a function of compression. At $x = 0$ in this plot, the prior for the black hole mass is shown. As the parameter space is compressed by Nested Sampling (finding the best $1/e$ of the remaining prior mass), the x-value advances by 1. After about 20 compressions, the posterior inference on the black hole mass is remarkably insensitive to further compression. We discovered that all models above a compression of 20 reproduce the major qualitative aspects of the data (see Figure 4), despite not fitting to within the error bars. However, most models below a compression of 20 do not resemble the data at all. Therefore, we selected models between a compression of 20 and 35 to form the posterior distribution. Note that the effect of this selection is very similar to the effect of raising the temperature of the likelihood function, or increasing the size of the error bars on the data. These equivalent inflated error-bars are shown in the lower panel of Figure 4. An important advantage of Nested Sampling over other sampling techniques is that all values for the temperature or compression are obtained in a single run, making it computationally trivial to consider the consequences of weakening the effect of the data.

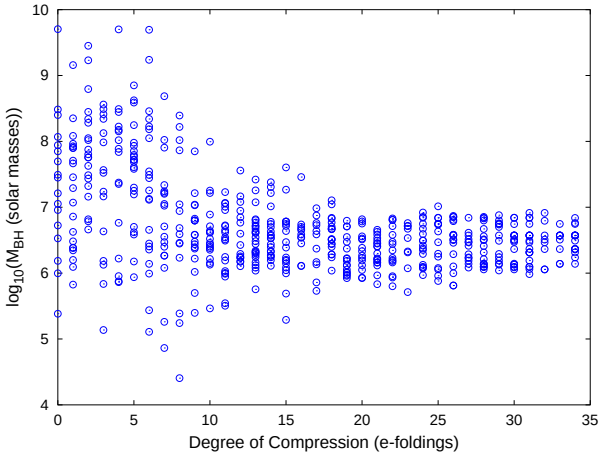

Fig. 6 The probability distribution for the black hole mass as a function of the compression of parameter space. Each step along the x-axis corresponds to selecting the best $1/e \approx 37\%$ of the remaining parameter space from the previous step, in terms of likelihood.

With the posterior samples in hand, the marginal distributions can be calculated. These marginal posterior distributions, and some joint distributions, are shown in Figure 7. In particular, there is a strong correlation between the inclination angle

$\phi_{\rm inc}$ of the BLR and the black hole mass, and this correlation runs in the expected direction: a more face-on configuration (higher inclination angle) lowers the cloud velocities along the line of sight, and thus requires a larger black hole mass to explain the data. A plausible BLR configuration for Arp 151 is shown in Figure 1.

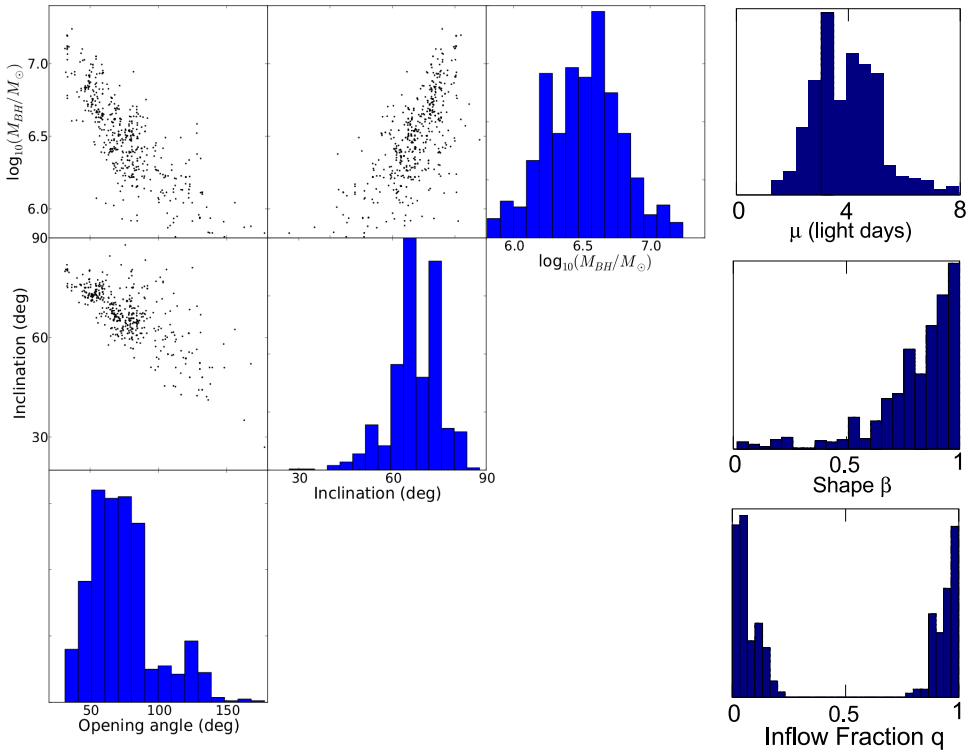

Fig. 7 Joint and marginal posterior distributions for the parameters of the BLR geometry and kinematics. The strongest correlation is between the inclination angle and the opening angle of the disk. Both of these parameters are also strongly correlated with the black hole mass.

5 Conclusions

The median and 68% credible interval for the black hole mass in Arp 151 was found to be $10^{6.51\pm0.28}$ $M_\odot$. This is lower than, but overlaps with, the value of $10^{6.85\pm0.07}$ $M_\odot$ obtained by [4] assuming $\log_{10} f = 0.74$ based on requiring active and inactive galaxies to obey the same correlation between M_{BH} and host-galaxy stellar velocity dispersion σ_* [15], and neglecting uncertainty in f. Recent measurements suggest that the intrinsic uncertainty in f from the standard method is at

least 0.4 dex [21, 10], 33% higher than our uncertainty. Reversing the traditional argument, our measurement implies that $\log_{10} f = 0.40 \pm 0.28$, a low value, for this particular system. This low value agrees with the updated estimate of $\bar{f}$ from [9], although the low value may also just apply to this single system. Our method also allows for the inference of more structural parameters of the BLR, not just the mean radius and the black hole mass. Although the basic philosophy is sound, future work is needed to improve the realism of the physics and the flexibility of the BLR distribution in our model. This will result in more robust, and hopefully smaller, black hole mass uncertainties, and a more detailed picture of the physics of AGN.

6 Acknowledgements

I would like to thank Tommaso Treu and Anna Pancoast for their substantial contributions to this work. I would also like to thank Aaron Barth for valuable discussions, and the entire LAMP team for the Arp 151 data and the encouragement. I acknowledge support by the NSF through CAREER award NSF-0642621.

References

1. Barth A. J., et al., 2011, ApJ, 732, 121
2. Bentz M. C., et al., 2008, ApJ, 689, L21
3. Bentz M. C., et al., 2010, ApJ, 720, L46
4. Bentz, M. C., et al. 2009, ApJ, 705, 199
5. Brewer B. J., et al., 2011, ApJ, 733, L33
6. Brewer, B., Partay, L., & Csanyi, G. 2010, Statistics and Computing, doi:10.1007/s11222-010-9198-8
7. Denney K. D., et al., 2010, ApJ, 721, 715
8. Denney K. D., Peterson B. M., Dietrich M., Vestergaard M., Bentz M. C., 2009, ApJ, 692, 246
9. Graham A. W., Onken C. A., Athanassoula E., Combes F., 2011, MNRAS, 412, 2211
10. Greene, J. E., et al. 2010, ApJ, 721, 26
11. Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009, ApJ, 698, 895
12. Kozłowski, S., et al. 2010, ApJ, 708, 927
13. MacLeod, C. L., et al. 2010, ApJ, 721, 1014
14. Neal, R. M., 1993, Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRGTR931, Dept. of Computer Science, Univ. of Toronto.
15. Onken, C. A., Ferrarese, L., Merritt, D., Peterson, B. M., Pogge, R. W., Vestergaard, M., & Wandel, A. 2004, ApJ, 615, 645
16. Onken, C. A., et al. 2007, ApJ, 670, 105
17. Pancoast, A., Brewer, B. J., & Treu, T. 2011, ApJ, 730, 139
18. Peterson, B. M., et al. 2004, ApJ, 613, 682
19. Peterson B. M., Bentz M. C., 2006, NewAR, 50, 796
20. Skilling, J., Nested Sampling for General Bayesian Computation, Bayesian Analysis 4, pp. 833-860 (2006)
21. Woo, J., et al. 2010, ApJ, 716, 269
22. Zu, Y., Kochanek, C. S., & Peterson, B. M. 2010, ArXiv: 1008.0641