Parameter Estimation for JSPAM Galactic Merger Simulations Using a Modified Metropolis-Hastings Method

Graham West, John Wallin, Zach Sinkala

Abstract. Gravitational N-body models can used to model the dynamical properties of interacting galaxies. These models represent states in a six-dimensional phase space with three-dimensional projections that can observed with astronomical telescopes. Although these models have thousands of particles, the dynamical state that describes the system can be represented with twelve dynamical parameters. In this paper, we present the implementation and results of an MCMC-type optimization method to solve the problem of estimating these dynamical parameters for galaxy mergers. Using an adaptive Metropolis-Hastings algorithm combined with the appropriate error function, we are able to find a merger simulation of minimum error with respect to a given merger target. We also explore the connections between the convergence of the method and the sensitivity of the parameters in this system.

Key words. non-linear dynamics, cosmology, MCMC, galaxy mergers, simulation, optimization

AMS subject classifications. 68Q25, 68R10, 68U05

1. Introduction.

Understanding the morphology and evolution of galactic mergers is a central issue in contemporary astrophysicis–in particular, the inverse problem of tracing a merger back in time to determine its dynamical history. Due to the highly non-linear nature of galactic systems as well as limited observational knowledge of a given system's current state, this problem cannot be solved by simply reversing time in an N-body simulation, necessitating the use search algorithms which require a great number of merger simulations to fully explore the parameter space.

In order to resolve this problem, we need three things: 1) a fast simulation code, 2) an error function which accurately measures the differences between two mergers, and 3) an optimization algorithm which can minimize that error function. Step 1 was completed by Wallin et al. [3] with the development of the simulation code entitled JSPAM. Steps 2 and 3 both will be described in this paper since an optimization problem is intimiately linked to its error function.

In what follows, we will discuss the tools and data we had at our disposal, including JSPAM, *Galaxy Zoo: Mergers* and *Merger Wars*, and the SDSS and MaNGA surveys. Then– in the majority of the paper–we will examine our method of calculating the error function and the MCMC algorithm which minimizes it. We will show how this algorithm converges on some sample systems, and explore how the sensitivity of the dynamical parameters that define our system are related to the convergence rate of the MCMC method. Finally, we will discuss several future steps which we will take to integrate the MaNGA data into our process so that we can begin fitting real mergers.

1.1. JSPAM.

Due to the nature of gravitation as a pair-wise interaction and the structure of galaxies being primarily a cloud of massive, interacting particles, the most direct way to simulate a merger would be to set up a full $O(n^2)$ N-body code. Approximations can be

made which reduce the complexity to $O(n \log(n))$–hierarchical tree codes, for example–but, even this is too slow for our purposes. Due to the great quantity of simulations needed to be performed, we required a code which was $O(n)$. Thus, we used the restricted three-body JSPAM code developed by Wallin et al. [3].

This code makes several significant simplifications for the sake of runtime, but since the mergers are not simulated over a long time interval, accuracy is retained. A detailed description of the code is contained in [3], so we will provide only a brief overview of its functionality. Within the code, galaxies themselves are comprised of two parts: a center of mass and a swarm of massless particles. The massless particles only interact with the masses of the two galaxies while the centers of mass interact with both each other and the particles. Initially these particles are placed in a single disk of circular orbits in specified orbital plane around the center of mass of their respective galaxy. The gravitational fields around the center of masses of the galaxies are spherically symmetric. These fields are modeled after the disk, bulge and halo potential described by Hernquist (1988?) used to model the self-consistent gravitational fields of disk galaxies.

The reference frame of the simulation is setup so that one galaxy (called the primary) is initially at the origin with zero velocity and the other galaxy (called the secondary) is given a relative position and velocity. The equation that describes the acceleration of the ith particle is at position r_i is:

$$\vec{a}_i = -\frac{GM_1(|\vec{r}_i - \vec{R}_1|)}{(\vec{r}_i - \vec{R}_1)^2}\hat{r}_{i1} - \frac{GM_2(|\vec{r}_i - \vec{R}_2|)}{(\vec{r}_i - \vec{R}_2)^2}\hat{r}_{i2} \tag{1}$$

where $M_1(r)$ and $M_2(r)$ are the cumulative mass distributions of the two galaxies, R_1 and R_2 and the center of mass positions, and G is the universal gravitational constant.

The list of dynamical parameters which describe the merger include these relative positions and velocities, the masses of the galaxies, their radii, and their orientations in terms of altitude and azimuth (w.r.t. the z-axis), for a total of 14 parameters. Since the projected position of the two galaxies are constrained by astronomical observations (x, y are always known), there are 12 free parameters in the simulations.

Given a list of *final* parameters post-merger, the centers of mass are placed at their respective positions and integrated backward in time, past the point of closest approach to a fixed starting time. Then, particle positions and velocities as well as the galactic potential are then initialized. Finally, the system is integrated forward in time, with the centers of mass back to their original positions. Since the galactic potential is a fairly complicated function which must be called many times throughout a simulation, it is assumed to be spherically symmetric, pre-calculated at a set of fixed radial distances, and stored. During simulations, the value of the potential in between these radii is calculated via linear interpolation. Since the potentials are stored, they do not change over time; however, since they originate from the centers of mass, they do translate through space. In addition to the potential, JSPAM also includes a force due to dynamical friction. This phenomenon slows objects traveling through a medium of gravitating particles by causing an increase in particle density behind the object, thus increasing the gravitational force in the retrograde direction.
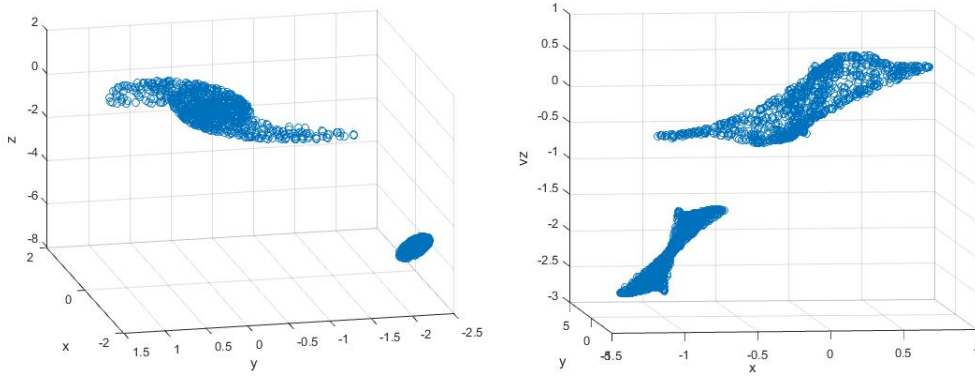
Figure 1. *Final time-step of a JSPAM merger simulation (1000 particles per galaxy). Left: X-Y-Z plot. Right X-Y-Vz plot. NOTE: the graphs are viewed at two different orientations.*

1.2. Galaxy Zoo: Mergers and Merger Wars. From a computational perspective, the problem of finding an appropriate function for comparing morphologies is a real challenge. It is difficult to find a method that is robust enough to perform on a level equivalent to that of a human. By nature, humans are excellent at pattern recognition; and where it can be difficult to "train" an algorithm to spot similarities between objects, humans can do it almost instinctively. For this reason, Holincheck et al. [1] employed the help of thousands of citizen scientists to assist the *Galaxy Zoo: Mergers* project in which they volunteered their pattern recognition abilities to determine best-fit models for 62 actual mergers.

The volunteers analyzed the morphologies of many thousands of simulations via a vote-based tournament scheme called the *Merger Wars* algorithm. In a typical session, a volunteer is shown an image of the target (the actual merger) along with two simulations which model the target. The volunteer then chooses which simulation is most similar to the target, which counts as a "win" for one and a "loss" for the other. (They can also select neither image, thereby affording both images a loss.) Since it is entirely possible for images to be judged poorly, every simulation participates in multiple rounds of the tournament so that a bad round does not have a large impact on the results of the competition. The fitness of a particular simulation is then calculated as the percentage of "wins" it achieved, with the maximum being 1. Since an integral part of our research involved find an explicit mathematical fitness function for the merger morphologies, it was very useful to have the *Merger Wars* data with which we could assess the validity of our own data.

1.3. SDSS and MaNGA. The end goal of our research is to construct an automated pipeline which can compare the morphologies of simulated mergers with images of actual mergers. Once we reach this stage, we will use the Sloan Digital Sky Survey's (SDSS) extensive catalog of observational data. Over the years, SDSS has provided many large, detailed data releases on collections of galactic mergers, the latest of which includes data from the Mapping Nearby Galaxies at APO (MaNGA) project. Thanks to the use of new observational

107 tools, MaNGA has been able to capture spectra over the surface of nearly 2,000 galaxies, in-
108 cluding many mergers. With this information, it is possible to derive the z-velocity fields and
109 dispersions of the galaxies (z being the axis from the telescope to the galaxy with positive z
110 in the direction of the telescope) and using these velocity fields, we will be able to apply our
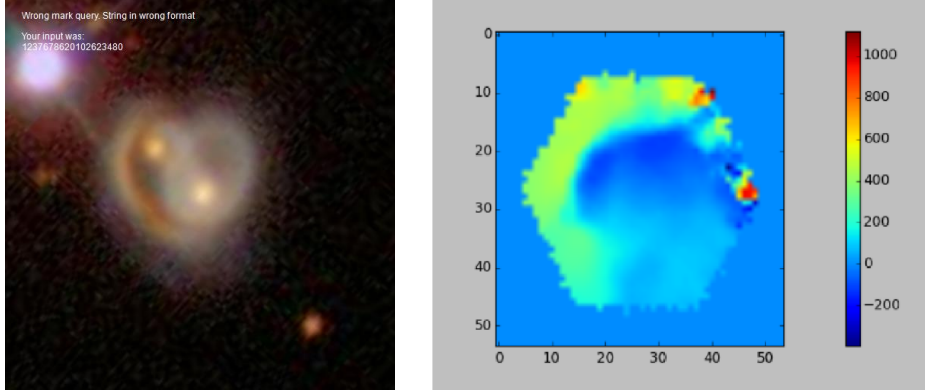111 MCMC method for estimating the mergers' dynamical parameters.

Figure 2. *Visible image (512 × 512 pixels) and z-velocity (54 × 54 pixels) of a merger observed by MaNGA.*

112 **2. Methods.** At its heart, a parameter estimation problem is essentially an optimization
113 problem in which one minimizes a given error funciton w.r.t. a set of parameters. The
114 two fundamental parts to the solution of such an optimization problem are 1) a meaningful
115 error function and 2) a robust optimization algorithm. In this section, we will examine the
116 reasoning behind the function we developed as well as the modifications we made to the
117 canonical Metropolis-Hastings method.

118 **2.1. Error Function.** There are essentially two types of information which our error func-
119 tion must capture: particle positions and velocities. Obviously, however, we can't evalute
120 our function on a per-particle basis because–among many reasons–once we begin using real
121 merger targets, we will not have per-star information. It was very important that we kept
122 the context of real merger targets in our minds as we experimented with different functions
123 because whatever we decided on in the testing phase must still be applicable to real targets.
124 Fortunately, this restriction led us to an elegant solution to the problem. The MaNGA data
125 is in the format of a 54×54 image for the z-velocity and a 512×512 image for visible light
126 intensity. Therefore, simply binning the particle number and z-velocity of a simulation will
127 give us an equivalent format.
128 Binning the particle number is trivial; however, when binning the z-velocity, we must
129 account for the opacity of gas and dust within the galaxies. To do so, we bin the velocities in
130 3D, averageing the values of all the particles within each bin to get its value. The equation
131 which governs the decrease in intensity of light as it passes through a medium of density ρ

132 from z_- to z_+ is:

(2)
$$I(x,y,z_-,z_+) = I_0(x,y,z_-)\exp(-\kappa \int_{z_-}^{z_+} \rho(x,y,z')dz')$$

134 where I_0 is the the actual intensity of light generated at a point and I is the intensity observed
135 after the light has passed through Now, we use this as a kernel weighting for the velocity to
136 map it from 2D to 3D:

(3)
$$v^{2D}(x,y) = \frac{\int_{z_-}^{z_+} v^{3D}(x,y,z)I(x,y,z,z_+)dz}{\int_{z_-}^{z_+} I(x,y,z,z_+)dz}$$

138 These formulae are discretized by treating the function I_0 as proportional to the the particle
139 count in a single bin:

(4)
$$v_{i,j}^{2D} = \frac{\sum_{k=0}^{n_{bin}-1} v_{i,j,k}^{3D} n_{i,j,k} w_{i,j,k}}{\sum_{k=0}^{n_{bin}-1} n_{i,j,k} w_{i,j,k}}$$

(5)
$$w_{i,j,k} = \exp(-\kappa \sum_{l=k}^{n_{bin}-1} \frac{n_{i,j,l}/n_{pts}}{dxdydz}dz)$$

143 where κ is the coeffcient of opacity, n_{bin} is the number of bins on one axis, n_{pts} is the total
144 number of particles (included for normalization purposes), and dx, dy, dz are the dimensions
145 of a single bin.
146   Having binned the simulation, we have two grids representing the x-y distribution of
147 particle density and z-velocity. We now divide the error function into two parts, one repre-
148 senting the similarity of the mergers' particle distributions and the other the RMS error of
149 the z-velocities between them:

(6)
$$f(X,X_0) = (1 - OvrFrac) \cdot RMSE$$

151 where,

(7)
$$OvrFrac = \frac{A_{X \cap X_0}}{A_X + A_{X_0} - A_{X \cap X_0}}$$

153 and

(8)
$$RMSE^2 = \frac{1}{A_{X \cap X_0}} \sum_{i,j=0}^{n_{bin}-1} w_{i,j}|v_{i,j}^X - v_{i,j}^{X_0}|^2$$

155   In these equations, X, X_0, and $X \cap X_0$ are indices referring to the *simulation*, *target*,
156 and their intersection, respectively. So, A_X is the number of *simulation* bins with particles
157 present, A_{X_0} the number of *target* bins, and $A_{X \cap X_0}$ the number of *overlapping* bins they
158 share. The overlap fraction, $OvrFrac$, is the fraction of bins/pixels/area in which both the
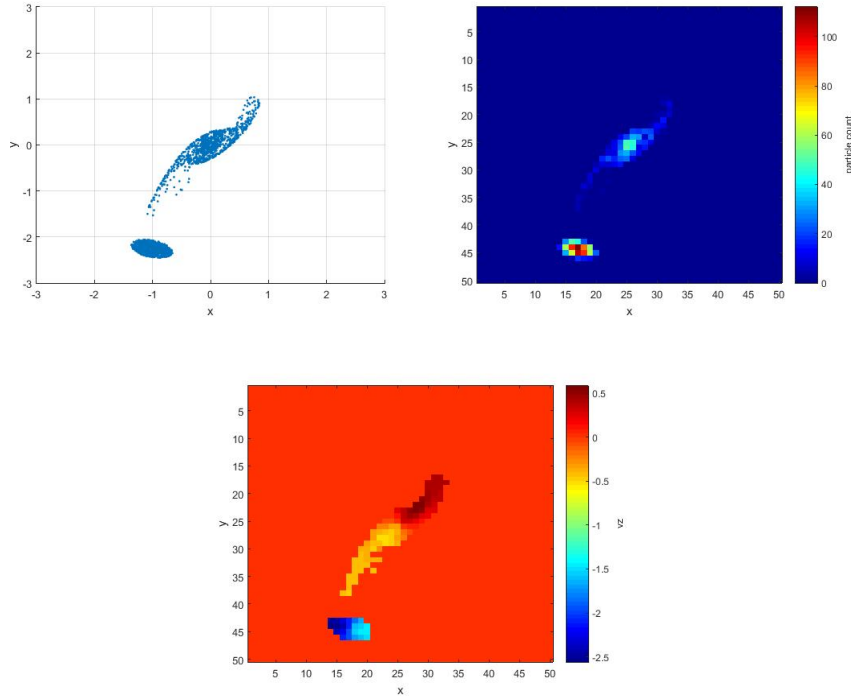159 simulation and the target have particles present. $RMSE$ is the weighted root-mean-squared

Figure 3. *Left: a per-particle image of a merger simulation. Right: a 50x50 square-binning of the left image (high-density regions are red and low-density regions are blue). Bottom: a 50x50 square-binning of the z-velocities.*

error between the velocity fields only calculated over the overlapping bins and normalized by the number of overlapping bins (thus, $w_k = 1$ if the pixels overlap, $w_k = 0$ otherwise).

We experimented with different weighting schemes...

Although this weighting system is not perfect, it works adequately for systems where was have a measured velocity across the interacting galaxy system. The MaNGA project in the Sloan Digital Sky Survey **REFERENCE NEEDED** was designed specifcally to measure these quantities for real galaxies. This project will sample the velocity fields for about 10,000 galaxies over its mission using integrated field spectroscopy. Although most of the galaxies in the sample will not be gravitationally interacting with a neighbor, we expect to have a sample of a few hundred interacting systems within the next several years. In addition to providing the velocity field, the mission also provides the ionization state for several different atomic species to allow measurements of quantites such as star formation rate, internal absorption, and temperature of the interstellar medium.

It is important to have both the overlap and RMSE terms in the function. Consider the fact that morphologically similar mergers will have have high overlap and low velocity error, but the converse of this statement is not necessarily true. If the overlap is high, there is a proability that the velocity error could still be large, or vice versa. This probability is decreased if we include both velocity and overlap in the calculation.

178   **2.2. Adaptive Metropolis-Hastings.** Due to the chaotic nature of the merger system,
179 the error lanscape of the parameters space is incredibly complicated, having countless local
180 minima and maxima. Consequently, any naive gradient-based optimization scheme would
181 be completely unreliable. Therefore, we chose a more robust MCMC method called the
182 Metropolis-Hastings algorithm. Given an initial set of parameter values, this method applies
183 random–in our case normally-distributed–perturbations to the parameters. Each parameter i
184 has its own specified σ_i for its associated distribution. At each subsequent step, the pertur-
185 bations are accepted with probability α, given by:

186 (9)
$$\alpha(X', X) = \min(1, \frac{P(X', X_0)}{P(X, X_0)})$$

187 where X' is the current, perturbed simulation and X is last previously accepted simulation
188 and P is given by its logairthm:

189 (10)
$$\log(P(X, X_0)) = \frac{1}{2}\log(2\pi\Delta^2) + \frac{f(X, X_0)^2}{2\Delta^2}$$

190 where Δ is a value representing the scale of the error. Also, the larger Δ becomes, the more
191 relaxed the method will be to accept poorer simulations. In the canonical algorithm, the
192 σ_i values are fixed; however, we included additional functionality to allow them to adapt if
193 multiple simulations are consequtively rejected. Let $\sigma_{i,j}$ be the σ value for the i-th parameter
194 after j consequtive rejections. Then,

195 (11)
$$\sigma_{i,j} = \begin{cases} \sigma_i & j \bmod 3 = 0 \\ \sigma_i \cdot M_j & j \bmod 3 = 1 \\ \sigma_i \cdot m_j & j \bmod 3 = 2 \end{cases}$$

196 where,

197 (12)
$$M_j = (1 + a)^{(j+2)/3}$$

198
199 (13)
$$m_j = (1 - b)^{(j+1)/3}$$

200 and $0 < a, b < 1$. These two factors serve to alternate between increasing, decreasing, and
201 resetting the σ values so that the MCMC can escape local minima or fall into a thin global
202 minimum quicker. We will discuss the values used for a and b in the next section, along with
203 the impact this modification has on convergence.

204   **2.3. Probablity of Convergence Within n-Steps.** One of the most important problems
205 we faced was that of determining how well our method converges. We wanted a technique
206 which would account for both how accurately and how quickly convergence occurs as we
207 varied the magnitude of the initial perturbation. Our final product was a technique which
208 could calculate the probability of convergence within a fixed number of steps given an initial
209 perturbation.
210   Since MCMC methods are stochastic, individual runs tell us nothing about their conver-
211 gence. We must run many tests and calculate statistics on the collective results. Therefore,

	min param value	actual param value	max param value	max error value
z	2.85849	3.35734	3.85636	0.36256
v_x	-0.35995	-0.12979	0.10037	1.19831
v_y	0.41617	0.55971	0.70234	0.82497
v_z	3.17224	4.02072	4.86919	0.60629
mass ratio	0.46377	0.79796	1.13216	0.46780
total mass	1.28367	1.7633	2.24293	0.30094
rad prim	0.26547	0.30541	0.34535	0.08838
rad sec	0.37291	0.41196	0.45101	0.03025
phi prim	147.39649	152.77952	158.16255	0.01729
phi sec	190.09539	195.62801	201.16063	0.01444
theta prim	140.86124	147.22542	153.5896	0.039687
theta sec	133.44108	138.42519	143.4093	0.02756

Table 1
test

our technique for testing consisted of running many instances of the MCMC, all with the same target and initial perturbation and calucating the percentage which converged (i.e., reached a minimum error threshold) so as to calculate an approximation to the convergence probability. This was done on a univariate basis by choosing an interval of evenly-spaced points centered on the target parameter value. The interval's width was chosen to be equal to twice the standard deviation of the top 25% runs' value of that parameter. These points were then repeatedly used as the initial perturbation for the MCMC algorithm. By calcuating this probability and plotting it over the interval, we can visually inspect how much the shape of the error curve impacts the rate of convergence, as well as compare the convergence of different parameters.

3. Results. Intro...

3.1. Error vs. Parameters. We began with a univariate analysis of the error function, sampling it over the interval dscribed in section 2.3. The following plots illustrate the error function's dependence on the 12 parameters. Notice how the minimum is located in the center of the interval, coincident with the target's parameter value.

Notice how the prim/sec's (one of them) variance in error tends to be much larger than the other. This is because that galaxies is the one that is nearer to the observer and thus blocks the view of the other one. Were the opacity coefficient lowered, this effect would be reduced and the galaxies would have more similar error profiles.

3.2. Example 1 - Heart shaped galaxy. Describe the galaxy we are modeling Interacting ring galaxy - di - number of models that were run by volunteers - number of "good models" - fitness function results - issues with the best fit - real velocity field has been observed by MaNGA

For our purposes, we will use the best fit model as our target system rather than the observations. This allows us to do additional sampling and to know the exact dynamical parameters.
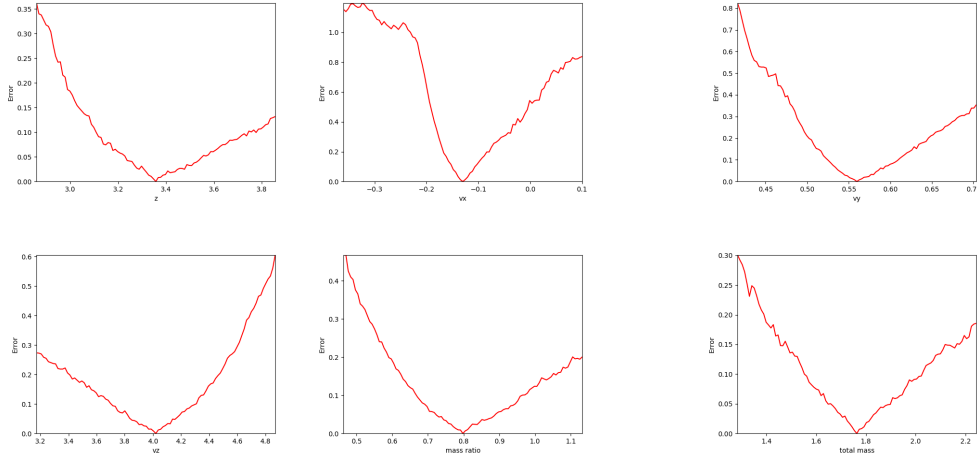
Figure 4. *Plots.*

To ex

3.3. Convergence Probabilities. MCMC plots

3.4. Physical Convergence - what does it look like?. Density/velocity heat maps, scatter plots

3.5. Example 2.

3.5.1. parameter values. To measure the sensitivity, we sweep through a range of physically reasonable values for each the the parameters and calculate the fitness as a function fo the parameter's value. The peak of the fitness function should occur when we have a perfect match.

For this galaxies, we list the values we are using along with the range allowed in the models. Make a table parameter 1, actual value, minimum value, maximum value. parameter 2, actual value, minimum value, maximum value. parameter 3, actual value, minimum value, maximum value. parameter 4, actual value, minimum value, maximum value. parameter 5, actual value, minimum value, maximum value.

4. Analysis.

5. Conclusions.

REFERENCES

[1] A. Holincheck, J. Wallin, K. Borne, L. Fortson, C. Lintott, A M. Smith, S. Bamford, W. Keel, and M. Parrish, "Galaxy Zoo: Mergers - Dynamical Models of Interacting Galaxies." *MNRAS*, (2016). Vol. 459, Is. 1, 720-745.

[2] H. Mo, F. Bosch, and S. White, *Galaxy Formation and Evolution*, 2010, Cambridge University Press.

[3] J. Wallin, A. Holincheck, and A. Harvey, "JSPAM: A restricted three-body code for simulating interacting galaxies." *Astronomy and Computing* 16, (2016). 26-33.

[Bizyaev et al.(2017)] Bizyaev, D., Walterbos, R. A. M., Yoachim, P., et al. 2017, apj, 839, 87

[Jin et al.(2016)] Jin, Y., Chen, Y., Shi, Y., et al. 2016, mnras, 463, 913

[Law et al.(2016)] Law, D. R., Cherinka, B., Yan, R., et al. 2016, aj, 152, 83

[Sánchez et al.(2016)] Sánchez, S. F., Pérez, E., Sánchez-Blázquez, P., et al. 2016, rmxaa, 52, 171

[Sánchez et al.(2016)] Sánchez, S. F., Pérez, E., Sánchez-Blázquez, P., et al. 2016, rmxaa, 52, 21

[Drory et al.(2015)] Drory, N., MacDonald, N., Bershady, M. A., et al. 2015, aj, 149, 77

[Bundy et al.(2015)] Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, apj, 798, 7

[York et al.(2000)] York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, aj, 120, 1579