

A PIPELINE FOR CONSTRUCTING A CATALOG OF  
MULTI-METHOD MODELS OF INTERACTING GALAXIES

by

Anthony Holincheck  
A Dissertation  
Submitted to the  
Graduate Faculty  
of  
George Mason University  
In Partial fulfillment of  
The Requirements for the Degree  
of  
Doctor of Philosophy  
Computational Sciences and Informatics

Committee:

\_\_\_\_\_ Dr. John Wallin, Dissertation Director  
\_\_\_\_\_ Dr. Kirk Borne, Committee Chair  
\_\_\_\_\_ Dr. Jessica Rosenberg, Committee Member  
\_\_\_\_\_ Dr. Daniel Carr, Committee Member  
\_\_\_\_\_ Dr. Michael Summers, Director, School of  
Physics, Astronomy, and Computational  
Sciences  
\_\_\_\_\_ Dr. Timothy L. Born, Associate Dean for  
Student and Academic Affairs, College of  
Science  
\_\_\_\_\_ Dr. Vikas Chandhoke, Dean, College of  
Science

Date: \_\_\_\_\_ Summer Semester 2013  
George Mason University  
Fairfax, VA

A Pipeline for Constructing a Catalog of Multi-method Models of Interacting Galaxies

A dissertation submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy at George Mason University

By

Anthony Holincheck  
Master of Science  
George Mason University, 2002  
Bachelor of Science  
The College of William & Mary, 1998

Director: Dr. John Wallin, Professor  
Middle Tennessee State University  
Department of Physics and Astronomy

Summer Semester 2013  
George Mason University  
Fairfax, VA

Copyright © 2013 by Anthony Holincheck  
All Rights Reserved

## Dedication

I dedicate this dissertation to my family. For my parents, my wife Nancy, and for my children, Will, Emma, and Sophie, I want you to know the inspiration and support you have provided is immeasurable.

## Acknowledgments

One of the most enjoyable parts of working with the amazing Zooniverse organization is that I get to thank thousands of people for their contributions. The *really* fun part is that I can do that without hyperbole. The Citizen Scientists who volunteered to participate in this research made it possible to accumulate an impressive amount of information about the interacting galaxies we studied. The tremendous number of simulations performed and evaluated, over 3 million, has allowed us to address the longstanding issue of convergence with some confidence. The logistics of recognizing the contributions of so many require that we direct you to the website<sup>1</sup> for details. I thank you all for taking the time to visit our little corner of the Zooniverse, the Merger Zoo, and crash some galaxies together.

Beyond the Citizen Scientists, many other people deserve recognition for their support to this research. I would like to thank John Wallin for all of his guidance throughout this process. Thank you for believing in my ability to perform the work that I did. Thank you Jessica Rosenberg for going beyond the call of duty by providing me with the equivalent of a course in Galactic and Extragalactic Astronomy. I want to thank Daniel Carr for his many discussions on how to present the multidimensional results. Kirk Borne, thank you for all of your advice and for making me aware of the need for projects like this to play a part in dealing with the data tsunami that is impacting science.

The Merger Zoo website evolved, like many endeavors, through a combination of necessity born of laziness and the help of many others. John Wallin's SPAM simulation software was the focus of his own dissertation. At the beginning of this research that software was compiling and running in excellent fashion from its initial language of FORTRAN. However, my inability to get any of the third party visualization libraries to compile on my laptop lead me to port the software to Java. This was a selfish choice for me which made it possible to view realtime results for simulations. This in turn lead to the rapid development of the Explore and Enhance views for the software. The version of the interface displayed on the Merger Zoo site was the result of many rounds of feedback. For their help in refining the tool I want to thank the following people: Philip Murray, Geza Gyuk, Mark SubbaRao, Nancy Ross Dribin, and Jordan Raddick. I would also like to thank Rebecca Ericson and the Astronomy laboratory instructors at George Mason University for allowing me to debut the tool for use with their lab students.

We partnered with the Zooniverse for the National Science Foundation's Cyber-Enabled Discovery and Innovation (CDI) program. I want to thank all of our partners from that program in particular the leaders that supported this project, Lucy Fortson and Chris Lintott. In addition to making resources available, they both provided a great deal of direct feedback. I want to thank Arfon Smith for the technical assistance needed to implement a website and support it so that thousands of volunteers could contribute their work. I want to thank Daniel Darg for his work with the Galaxy Zoo data for mergers and his

---

<sup>1</sup><http://mergers.galaxyzoo.org/authors>

assistance with handling “deblending”. Thank you to Allen Harvey for his comments along the way. I would like to thank Lior Shamir both for his excellent WNDCHRM software and his insightful collaboration.

Working on a project such as this one can lead to other areas of life being neglected. I want to thank my family for their patience and understanding. I would also like to thank my employer Janet Cathell for her support and flexibility.

I accessed data from a number of locations. The relevant scientific and technical references have been cited in the body of the dissertation. However, several of the data providers require an additional acknowledgment, which I do here.

The Digitized Sky Surveys were produced at the Space Telescope Science Institute under U.S. Government grant NAG W-2166. The images of these surveys are based on photographic data obtained using the Oschin Schmidt Telescope on Palomar Mountain and the UK Schmidt Telescope. The plates were processed into the present compressed digital form with the permission of these institutions.

The National Geographic Society - Palomar Observatory Sky Atlas (POSS-I) was made by the California Institute of Technology with grants from the National Geographic Society.

The Second Palomar Observatory Sky Survey (POSS-II) was made by the California Institute of Technology with funds from the National Science Foundation, the National Geographic Society, the Sloan Foundation, the Samuel Oschin Foundation, and the Eastman Kodak Corporation. Additional acknowledgements are available at the DSS website <http://stdata.stsci.edu/dss/acknowledging.html>.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

This research has made use of the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

I would like to acknowledge support from the National Science Foundation grant DRL #0941610.

# Table of Contents

	Page
List of Tables . . . . .	xi
List of Figures . . . . .	xiv
Abstract . . . . .	xxiv
1 Introduction . . . . .	1
1.1 A Short History of the Study of Interacting Galaxies . . . . .	2
1.2 Some Important Physical Processes in Interacting Systems . . . . .	4
1.2.1 Gravity . . . . .	5
1.2.2 Gas Dynamics . . . . .	6
1.3 Comparing Models to Observations . . . . .	9
1.4 Automating the Search for Models of Interacting Galaxies . . . . .	10
1.4.1 Current Methods . . . . .	13
1.4.2 The Role of Citizen Scientists . . . . .	14
1.4.3 A Pipeline for Constructing Models of Systems of Interacting Galaxies . . . . .	16
2 Modeling Galaxies . . . . .	18
2.1 Restricted Three-Body . . . . .	19
2.2 N-Body . . . . .	20
2.2.1 Tree Codes . . . . .	21
2.3 Additional Physical Processes . . . . .	22
2.3.1 Dynamical Friction . . . . .	22
2.3.2 Modeling Gas . . . . .	24
2.4 Specific Codes . . . . .	24
2.4.1 SPAM . . . . .	25
2.4.2 MASS99 . . . . .	29
2.5 Validity of Restricted Three-Body Approximation . . . . .	29
3 The Population of Interacting Galaxies . . . . .	34
3.1 Sky Surveys . . . . .	35
3.1.1 Digitized Sky Survey . . . . .	35
3.1.2 Sloan Digital Sky Survey . . . . .	36

3.1.3	Wide-field Infrared Survey Explorer . . . . .	36
3.2	Catalogs of Interacting Galaxies . . . . .	37
3.2.1	Arp . . . . .	37
3.2.2	Galaxy Zoo - Mergers Catalog . . . . .	38
3.2.3	The overlap between Arp and SDSS . . . . .	39
3.2.4	Hubble Space Telescope - ‘Galaxies gone Wild!’ . . . . .	39
3.2.5	Other catalogs . . . . .	40
3.3	The Sample for this Study . . . . .	40
3.4	Photometric Observations . . . . .	41
3.4.1	Deblending . . . . .	42
3.4.2	WISE Photometry . . . . .	43
3.5	Matching SDSS Value-added Catalog with WISE . . . . .	43
3.5.1	SDSS / WISE Regression Analysis . . . . .	47
4	Identifying Morphological Matches . . . . .	49
4.1	Target Preparation . . . . .	49
4.1.1	Simulation Filter . . . . .	53
4.2	Automated Methods . . . . .	54
4.3	Citizen Scientists . . . . .	55
4.3.1	Merger Wars . . . . .	58
4.3.2	Finale Activities . . . . .	60
4.3.3	Merger Zoo Contributions . . . . .	61
4.4	Current Process . . . . .	63
4.4.1	Potential for a New Automated Method . . . . .	63
5	Simulation Results . . . . .	67
5.1	Results for each Pair of Galaxies . . . . .	67
5.1.1	SDSS 587722984435351614 . . . . .	68
5.1.2	SDSS 587724234257137777 . . . . .	78
5.1.3	SDSS 587726033843585146 . . . . .	82
5.1.4	SDSS 587727177926508595 . . . . .	86
5.1.5	SDSS 587727178988388373 . . . . .	90
5.1.6	SDSS 587727222471131318 . . . . .	94
5.1.7	SDSS 587728676861051075 . . . . .	99
5.1.8	SDSS 587729227151704160 . . . . .	104
5.1.9	SDSS 587731913110650988 . . . . .	109
5.1.10	SDSS 587732136993882121 . . . . .	113
5.1.11	SDSS 587732772130652231 . . . . .	117
5.1.12	SDSS 587733080814583863 . . . . .	122

5.1.13	SDSS 587734862680752822	126
5.1.14	SDSS 587735043609329845	130
5.1.15	SDSS 587735665840881790	135
5.1.16	SDSS 587736941981466667	139
5.1.17	SDSS 587738569246376675	144
5.1.18	SDSS 587738569249390718	149
5.1.19	SDSS 587739153356095531	153
5.1.20	SDSS 587739407868690486	157
5.1.21	SDSS 587739505541578866	161
5.1.22	SDSS 587739646743412797	165
5.1.23	SDSS 587739647284805725	169
5.1.24	SDSS 587739707420967061	173
5.1.25	SDSS 587739721376202860	177
5.1.26	SDSS 587739721900163101	182
5.1.27	SDSS 587739810496708646	187
5.1.28	SDSS 587739845393580192	191
5.1.29	SDSS 587741391565422775	195
5.1.30	SDSS 587741532784361481	200
5.1.31	SDSS 587741534400217110	205
5.1.32	SDSS 587741602030026825	209
5.1.33	SDSS 587741722819493915	214
5.1.34	SDSS 587741817851674654	218
5.1.35	SDSS 587741829658181698	222
5.1.36	SDSS 587742010583941189	226
5.1.37	SDSS 587742014353702970	230
5.1.38	SDSS 587742571610243080	234
5.1.39	SDSS 587745402001817662	238
5.1.40	SDSS 587746029596311590	242
5.1.41	SDSS 588011124116422756	246
5.1.42	SDSS 588013383816904792	250
5.1.43	SDSS 588017604696408086	254
5.1.44	SDSS 588017604696408195	258
5.1.45	SDSS 588017702948962343	262
5.1.46	SDSS 588017978901528612	266
5.1.47	SDSS 588018055130710322	271
5.1.48	SDSS 758874299603222717	275
5.1.49	Arp 148	279
5.1.50	CGCG 436-030	283
5.1.51	Arp 272	287

5.1.52	ESO 77-14 . . . . .	291
5.1.53	NGC 5331 . . . . .	295
5.1.54	NGC 6786 . . . . .	299
5.1.55	SDSS 587736523764334706 . . . . .	303
5.1.56	SDSS 587747120521216156 . . . . .	307
5.1.57	SDSS 588007005230530750 . . . . .	311
5.1.58	SDSS 758877153600208945 . . . . .	315
5.1.59	SDSS 587739720308818095 . . . . .	319
5.1.60	Arp 273 . . . . .	323
5.1.61	SDSS 1237678620102623480 . . . . .	327
5.1.62	Arp 244 . . . . .	332
5.2	Convergence of Model Parameters . . . . .	336
5.2.1	Morphological Group 1 . . . . .	340
5.2.2	Morphological Group 2 . . . . .	341
5.2.3	Morphological Group 3 . . . . .	343
5.2.4	Morphological Group 4 . . . . .	344
5.2.5	Morphological Group 5 . . . . .	346
5.2.6	Morphological Group 6 . . . . .	347
5.2.7	Morphological Group 7 . . . . .	349
5.2.8	Morphological Group 8 . . . . .	350
5.2.9	Morphological Group 9 . . . . .	352
5.2.10	Morphological Group 10 . . . . .	353
5.2.11	Morphological Group 11 . . . . .	355
5.2.12	Morphological Group 12 . . . . .	356
5.3	Coverage of Parameter Space . . . . .	358
5.3.1	Distribution of Simulation Parameters . . . . .	359
5.3.2	$\beta$ Parameter . . . . .	360
6	Comparing Simulation Values to Observations . . . . .	366
6.1	Stellar Mass . . . . .	366
6.2	Star Formation Rate . . . . .	368
6.2.1	Projected Separation Distance . . . . .	371
7	Conclusion . . . . .	378
7.1	The Pipeline for Constructing Models . . . . .	378
7.2	The Merger Zoo Catalog of Interacting Galaxy Models . . . . .	380
7.2.1	Machine Learning Training Set . . . . .	380
A	A Method for Photometry of Irregular Galaxies . . . . .	382

A.1	Introduction . . . . .	382
A.2	Obtaining Imagery . . . . .	382
A.3	Background Subtraction . . . . .	383
A.3.1	Source Identification . . . . .	383
A.3.2	Background Estimation . . . . .	384
A.4	Measuring Flux in an Aperture . . . . .	386
A.5	Radial Profile . . . . .	388
A.5.1	Multiple Bands . . . . .	390
B	The Effectiveness of the Merger Wars Algorithm . . . . .	391
B.1	Abstract from Submitted Draft . . . . .	391
B.2	Introduction . . . . .	391
B.2.1	Motivation . . . . .	393
B.3	Methods . . . . .	393
B.3.1	Comparison Function . . . . .	395
B.3.2	Simulation Method . . . . .	396
B.3.3	Accuracy Metric . . . . .	397
B.4	Results . . . . .	398
B.5	Conclusion . . . . .	404
C	Target Preparation and Simulation Search Software . . . . .	405
C.1	Introduction . . . . .	405
C.1.1	Running the Software . . . . .	405
C.2	Target Preparation . . . . .	406
C.2.1	Target Tab . . . . .	406
C.2.2	Disk Info . . . . .	408
C.2.3	Disk Orientation . . . . .	409
C.2.4	Parameter Ranges . . . . .	411
C.3	Simulation Search . . . . .	411
C.3.1	Simulate . . . . .	412
C.3.2	Preferences . . . . .	413
C.3.3	Review . . . . .	414
C.3.4	Enhance . . . . .	414
C.3.5	Evaluate . . . . .	415
C.3.6	Statistics . . . . .	416
C.3.7	Saving . . . . .	416
C.3.8	About the Software . . . . .	417
	Bibliography . . . . .	418

## List of Tables

Table	Page
3.1 Linear regressions of mass, SFR, specific SFR, with WISE . . . . .	48
5.1 Summary for SDSS 587722984435351614 . . . . .	68
5.2 Summary for SDSS 587724234257137777 . . . . .	78
5.3 Summary for SDSS 587726033843585146 . . . . .	82
5.4 Summary for SDSS 587727177926508595 . . . . .	86
5.5 Summary for SDSS 587727178988388373 . . . . .	90
5.6 Summary for SDSS 587727222471131318 . . . . .	94
5.7 Summary for SDSS 587728676861051075 . . . . .	99
5.8 Summary for SDSS 587729227151704160 . . . . .	104
5.9 Summary for SDSS 587731913110650988 . . . . .	109
5.10 Summary for SDSS 587732136993882121 . . . . .	113
5.11 Summary for SDSS 587732772130652231 . . . . .	117
5.12 Summary for SDSS 587733080814583863 . . . . .	122
5.13 Summary for SDSS 587734862680752822 . . . . .	126
5.14 Summary for SDSS 587735043609329845 . . . . .	130
5.15 Summary for SDSS 587735665840881790 . . . . .	135
5.16 Summary for SDSS 587736941981466667 . . . . .	139
5.17 Summary for SDSS 587738569246376675 . . . . .	144
5.18 Summary for SDSS 587738569249390718 . . . . .	149
5.19 Summary for SDSS 587739153356095531 . . . . .	153
5.20 Summary for SDSS 587739407868690486 . . . . .	157
5.21 Summary for SDSS 587739505541578866 . . . . .	161
5.22 Summary for SDSS 587739646743412797 . . . . .	165
5.23 Summary for SDSS 587739647284805725 . . . . .	169
5.24 Summary for SDSS 587739707420967061 . . . . .	173
5.25 Summary for SDSS 587739721376202860 . . . . .	177
5.26 Summary for SDSS 587739721900163101 . . . . .	182

5.27	Summary for SDSS 587739810496708646 . . . . .	187
5.28	Summary for SDSS 587739845393580192 . . . . .	191
5.29	Summary for SDSS 587741391565422775 . . . . .	195
5.30	Summary for SDSS 587741532784361481 . . . . .	200
5.31	Summary for SDSS 587741534400217110 . . . . .	205
5.32	Summary for SDSS 587741602030026825 . . . . .	209
5.33	Summary for SDSS 587741722819493915 . . . . .	214
5.34	Summary for SDSS 587741817851674654 . . . . .	218
5.35	Summary for SDSS 587741829658181698 . . . . .	222
5.36	Summary for SDSS 587742010583941189 . . . . .	226
5.37	Summary for SDSS 587742014353702970 . . . . .	230
5.38	Summary for SDSS 587742571610243080 . . . . .	234
5.39	Summary for SDSS 587745402001817662 . . . . .	238
5.40	Summary for SDSS 587746029596311590 . . . . .	242
5.41	Summary for SDSS 588011124116422756 . . . . .	246
5.42	Summary for SDSS 588013383816904792 . . . . .	250
5.43	Summary for SDSS 588017604696408086 . . . . .	254
5.44	Summary for SDSS 588017604696408195 . . . . .	258
5.45	Summary for SDSS 588017702948962343 . . . . .	262
5.46	Summary for SDSS 588017978901528612 . . . . .	266
5.47	Summary for SDSS 588018055130710322 . . . . .	271
5.48	Summary for SDSS 758874299603222717 . . . . .	275
5.49	Summary for Arp 148 . . . . .	279
5.50	Summary for CGCG 436-030 . . . . .	283
5.51	Summary for Arp 272 . . . . .	287
5.52	Summary for ESO 77-14 . . . . .	291
5.53	Summary for NGC 5331 . . . . .	295
5.54	Summary for NGC 6786 . . . . .	299
5.55	Summary for SDSS 587736523764334706 . . . . .	303
5.56	Summary for SDSS 587747120521216156 . . . . .	307
5.57	Summary for SDSS 588007005230530750 . . . . .	311
5.58	Summary for SDSS 758877153600208945 . . . . .	315
5.59	Summary for SDSS 587739720308818095 . . . . .	319
5.60	Summary for Arp 273 . . . . .	323

5.61	Summary for SDSS 1237678620102623480 . . . . .	327
5.62	Summary for Arp 244 . . . . .	332
5.63	Simulation results and observed parameters . . . . .	361
5.64	Simulation results and orbit parameters . . . . .	362
5.65	Remaining fraction for the twelve simulation parameters . . . . .	363
A.1	Background $\mu$ and $\sigma$ before and after background subtraction. . . . .	386
A.2	Sersic profile parameters for Arp 290 primary and secondary galaxies. . . . .	390
B.1	Performance for a List of 1000 Items with Weber Comparison Function . . . . .	400
B.2	“Best of N” Number of Comparisons for a List of 1000 . . . . .	402
B.3	“Best of N” Performance for a List of 1000 Items . . . . .	402

## List of Figures

Figure		Page
1.1	Gridded fitness function . . . . .	15
1.2	Pipeline overview. . . . .	17
2.1	Rotational velocity curves for restricted three-body potentials. . . . .	30
2.2	MASS99 and SPAM trajectories with varying dynamical friction. . . . .	32
2.3	SPAM vs. MASS99 . . . . .	33
3.1	Normalized distribution of redshifts for three merger catalogs. . . . .	41
3.2	MPA/JHU Mass as a function of W1 absolute magnitude. . . . .	45
3.3	MPA/JHU SFR as a function of W3 absolute magnitude. . . . .	45
3.4	SFR as a function of Mass for different morphological types. . . . .	46
3.5	Relative frequency of W1 - W3 for various samples. . . . .	47
4.1	Explore. . . . .	56
4.2	Enhance. . . . .	57
4.3	Merger Wars. . . . .	58
4.4	Merger Zoo Process. . . . .	60
4.5	Simulation Showdown. . . . .	61
4.6	Best of the Best. . . . .	61
4.7	Cumulative count of simulations viewed. . . . .	62
4.8	The distribution of simulations viewed among volunteers. . . . .	62
4.9	Merger Wars fitness as a function of Zernike distance. . . . .	65
4.10	Clusters selected by Zernike distances . . . . .	66
5.1	Fitness distribution for SDSS 587722984435351614 . . . . .	69
5.2	Images for SDSS 587722984435351614 . . . . .	70
5.3	Trajectories for SDSS 587722984435351614 . . . . .	72
5.4	Parallel Coordinate plot for SDSS 587722984435351614 . . . . .	76
5.5	Glyph plot for SDSS 587722984435351614 . . . . .	77
5.6	Fitness distribution for SDSS 587724234257137777 . . . . .	78
5.7	Images for SDSS 587724234257137777 . . . . .	79
5.8	Trajectories for SDSS 587724234257137777 . . . . .	80

5.9	Parallel Coordinate plot for SDSS 587724234257137777 . . . . .	81
5.10	Fitness distribution for SDSS 587726033843585146 . . . . .	82
5.11	Images for SDSS 587726033843585146 . . . . .	83
5.12	Trajectories for SDSS 587726033843585146 . . . . .	84
5.13	Parallel Coordinate plot for SDSS 587726033843585146 . . . . .	85
5.14	Fitness distribution for SDSS 587727177926508595 . . . . .	86
5.15	Images for SDSS 587727177926508595 . . . . .	87
5.16	Trajectories for SDSS 587727177926508595 . . . . .	88
5.17	Parallel Coordinate plot for SDSS 587727177926508595 . . . . .	89
5.18	Fitness distribution for SDSS 587727178988388373 . . . . .	90
5.19	Images for SDSS 587727178988388373 . . . . .	91
5.20	Trajectories for SDSS 587727178988388373 . . . . .	92
5.21	Parallel Coordinate plot for SDSS 587727178988388373 . . . . .	93
5.22	Fitness distribution for SDSS 587727222471131318 . . . . .	95
5.23	Images for SDSS 587727222471131318 . . . . .	96
5.24	Trajectories for SDSS 587727222471131318 . . . . .	97
5.25	Parallel Coordinate plot for SDSS 587727222471131318 . . . . .	98
5.26	Fitness distribution for SDSS 587728676861051075 . . . . .	100
5.27	Images for SDSS 587728676861051075 . . . . .	101
5.28	Trajectories for SDSS 587728676861051075 . . . . .	102
5.29	Parallel Coordinate plot for SDSS 587728676861051075 . . . . .	103
5.30	Fitness distribution for SDSS 587729227151704160 . . . . .	105
5.31	Images for SDSS 587729227151704160 . . . . .	106
5.32	Trajectories for SDSS 587729227151704160 . . . . .	107
5.33	Parallel Coordinate plot for SDSS 587729227151704160 . . . . .	108
5.34	Fitness distribution for SDSS 587731913110650988 . . . . .	109
5.35	Images for SDSS 587731913110650988 . . . . .	110
5.36	Trajectories for SDSS 587731913110650988 . . . . .	111
5.37	Parallel Coordinate plot for SDSS 587731913110650988 . . . . .	112
5.38	Fitness distribution for SDSS 587732136993882121 . . . . .	113
5.39	Images for SDSS 587732136993882121 . . . . .	114
5.40	Trajectories for SDSS 587732136993882121 . . . . .	115
5.41	Parallel Coordinate plot for SDSS 587732136993882121 . . . . .	116
5.42	Fitness distribution for SDSS 587732772130652231 . . . . .	118
5.43	Images for SDSS 587732772130652231 . . . . .	119

5.44	Trajectories for SDSS 587732772130652231 . . . . .	120
5.45	Parallel Coordinate plot for SDSS 587732772130652231 . . . . .	121
5.46	Fitness distribution for SDSS 587733080814583863 . . . . .	122
5.47	Images for SDSS 587733080814583863 . . . . .	123
5.48	Trajectories for SDSS 587733080814583863 . . . . .	124
5.49	Parallel Coordinate plot for SDSS 587733080814583863 . . . . .	125
5.50	Fitness distribution for SDSS 587734862680752822 . . . . .	126
5.51	Images for SDSS 587734862680752822 . . . . .	127
5.52	Trajectories for SDSS 587734862680752822 . . . . .	128
5.53	Parallel Coordinate plot for SDSS 587734862680752822 . . . . .	129
5.54	Fitness distribution for SDSS 587735043609329845 . . . . .	131
5.55	Images for SDSS 587735043609329845 . . . . .	132
5.56	Trajectories for SDSS 587735043609329845 . . . . .	133
5.57	Parallel Coordinate plot for SDSS 587735043609329845 . . . . .	134
5.58	Fitness distribution for SDSS 587735665840881790 . . . . .	135
5.59	Images for SDSS 587735665840881790 . . . . .	136
5.60	Trajectories for SDSS 587735665840881790 . . . . .	137
5.61	Parallel Coordinate plot for SDSS 587735665840881790 . . . . .	138
5.62	Fitness distribution for SDSS 587736941981466667 . . . . .	140
5.63	Images for SDSS 587736941981466667 . . . . .	141
5.64	Trajectories for SDSS 587736941981466667 . . . . .	142
5.65	Parallel Coordinate plot for SDSS 587736941981466667 . . . . .	143
5.66	Fitness distribution for SDSS 587738569246376675 . . . . .	145
5.67	Images for SDSS 587738569246376675 . . . . .	146
5.68	Trajectories for SDSS 587738569246376675 . . . . .	147
5.69	Parallel Coordinate plot for SDSS 587738569246376675 . . . . .	148
5.70	Fitness distribution for SDSS 587738569249390718 . . . . .	149
5.71	Images for SDSS 587738569249390718 . . . . .	150
5.72	Trajectories for SDSS 587738569249390718 . . . . .	151
5.73	Parallel Coordinate plot for SDSS 587738569249390718 . . . . .	152
5.74	Fitness distribution for SDSS 587739153356095531 . . . . .	153
5.75	Images for SDSS 587739153356095531 . . . . .	154
5.76	Trajectories for SDSS 587739153356095531 . . . . .	155
5.77	Parallel Coordinate plot for SDSS 587739153356095531 . . . . .	156
5.78	Fitness distribution for SDSS 587739407868690486 . . . . .	157

5.79	Images for SDSS 587739407868690486 . . . . .	158
5.80	Trajectories for SDSS 587739407868690486 . . . . .	159
5.81	Parallel Coordinate plot for SDSS 587739407868690486 . . . . .	160
5.82	Fitness distribution for SDSS 587739505541578866 . . . . .	161
5.83	Images for SDSS 587739505541578866 . . . . .	162
5.84	Trajectories for SDSS 587739505541578866 . . . . .	163
5.85	Parallel Coordinate plot for SDSS 587739505541578866 . . . . .	164
5.86	Fitness distribution for SDSS 587739646743412797 . . . . .	165
5.87	Images for SDSS 587739646743412797 . . . . .	166
5.88	Trajectories for SDSS 587739646743412797 . . . . .	167
5.89	Parallel Coordinate plot for SDSS 587739646743412797 . . . . .	168
5.90	Fitness distribution for SDSS 587739647284805725 . . . . .	169
5.91	Images for SDSS 587739647284805725 . . . . .	170
5.92	Trajectories for SDSS 587739647284805725 . . . . .	171
5.93	Parallel Coordinate plot for SDSS 587739647284805725 . . . . .	172
5.94	Fitness distribution for SDSS 587739707420967061 . . . . .	173
5.95	Images for SDSS 587739707420967061 . . . . .	174
5.96	Trajectories for SDSS 587739707420967061 . . . . .	175
5.97	Parallel Coordinate plot for SDSS 587739707420967061 . . . . .	176
5.98	Fitness distribution for SDSS 587739721376202860 . . . . .	178
5.99	Images for SDSS 587739721376202860 . . . . .	179
5.100	Trajectories for SDSS 587739721376202860 . . . . .	180
5.101	Parallel Coordinate plot for SDSS 587739721376202860 . . . . .	181
5.102	Fitness distribution for SDSS 587739721900163101 . . . . .	183
5.103	Images for SDSS 587739721900163101 . . . . .	184
5.104	Trajectories for SDSS 587739721900163101 . . . . .	185
5.105	Parallel Coordinate plot for SDSS 587739721900163101 . . . . .	186
5.106	Fitness distribution for SDSS 587739810496708646 . . . . .	187
5.107	Images for SDSS 587739810496708646 . . . . .	188
5.108	Trajectories for SDSS 587739810496708646 . . . . .	189
5.109	Parallel Coordinate plot for SDSS 587739810496708646 . . . . .	190
5.110	Fitness distribution for SDSS 587739845393580192 . . . . .	191
5.111	Images for SDSS 587739845393580192 . . . . .	192
5.112	Trajectories for SDSS 587739845393580192 . . . . .	193
5.113	Parallel Coordinate plot for SDSS 587739845393580192 . . . . .	194

5.114	Fitness distribution for SDSS 587741391565422775 . . . . .	196
5.115	Images for SDSS 587741391565422775 . . . . .	197
5.116	Trajectories for SDSS 587741391565422775 . . . . .	198
5.117	Parallel Coordinate plot for SDSS 587741391565422775 . . . . .	199
5.118	Fitness distribution for SDSS 587741532784361481 . . . . .	201
5.119	Images for SDSS 587741532784361481 . . . . .	202
5.120	Trajectories for SDSS 587741532784361481 . . . . .	203
5.121	Parallel Coordinate plot for SDSS 587741532784361481 . . . . .	204
5.122	Fitness distribution for SDSS 587741534400217110 . . . . .	205
5.123	Images for SDSS 587741534400217110 . . . . .	206
5.124	Trajectories for SDSS 587741534400217110 . . . . .	207
5.125	Parallel Coordinate plot for SDSS 587741534400217110 . . . . .	208
5.126	Fitness distribution for SDSS 587741602030026825 . . . . .	210
5.127	Images for SDSS 587741602030026825 . . . . .	211
5.128	Trajectories for SDSS 587741602030026825 . . . . .	212
5.129	Parallel Coordinate plot for SDSS 587741602030026825 . . . . .	213
5.130	Fitness distribution for SDSS 587741722819493915 . . . . .	214
5.131	Images for SDSS 587741722819493915 . . . . .	215
5.132	Trajectories for SDSS 587741722819493915 . . . . .	216
5.133	Parallel Coordinate plot for SDSS 587741722819493915 . . . . .	217
5.134	Fitness distribution for SDSS 587741817851674654 . . . . .	218
5.135	Images for SDSS 587741817851674654 . . . . .	219
5.136	Trajectories for SDSS 587741817851674654 . . . . .	220
5.137	Parallel Coordinate plot for SDSS 587741817851674654 . . . . .	221
5.138	Fitness distribution for SDSS 587741829658181698 . . . . .	222
5.139	Images for SDSS 587741829658181698 . . . . .	223
5.140	Trajectories for SDSS 587741829658181698 . . . . .	224
5.141	Parallel Coordinate plot for SDSS 587741829658181698 . . . . .	225
5.142	Fitness distribution for SDSS 587742010583941189 . . . . .	226
5.143	Images for SDSS 587742010583941189 . . . . .	227
5.144	Trajectories for SDSS 587742010583941189 . . . . .	228
5.145	Parallel Coordinate plot for SDSS 587742010583941189 . . . . .	229
5.146	Fitness distribution for SDSS 587742014353702970 . . . . .	230
5.147	Images for SDSS 587742014353702970 . . . . .	231
5.148	Trajectories for SDSS 587742014353702970 . . . . .	232

5.149	Parallel Coordinate plot for SDSS 587742014353702970 . . . . .	233
5.150	Fitness distribution for SDSS 587742571610243080 . . . . .	234
5.151	Images for SDSS 587742571610243080 . . . . .	235
5.152	Trajectories for SDSS 587742571610243080 . . . . .	236
5.153	Parallel Coordinate plot for SDSS 587742571610243080 . . . . .	237
5.154	Fitness distribution for SDSS 587745402001817662 . . . . .	238
5.155	Images for SDSS 587745402001817662 . . . . .	239
5.156	Trajectories for SDSS 587745402001817662 . . . . .	240
5.157	Parallel Coordinate plot for SDSS 587745402001817662 . . . . .	241
5.158	Fitness distribution for SDSS 587746029596311590 . . . . .	242
5.159	Images for SDSS 587746029596311590 . . . . .	243
5.160	Trajectories for SDSS 587746029596311590 . . . . .	244
5.161	Parallel Coordinate plot for SDSS 587746029596311590 . . . . .	245
5.162	Fitness distribution for SDSS 588011124116422756 . . . . .	246
5.163	Images for SDSS 588011124116422756 . . . . .	247
5.164	Trajectories for SDSS 588011124116422756 . . . . .	248
5.165	Parallel Coordinate plot for SDSS 588011124116422756 . . . . .	249
5.166	Fitness distribution for SDSS 588013383816904792 . . . . .	250
5.167	Images for SDSS 588013383816904792 . . . . .	251
5.168	Trajectories for SDSS 588013383816904792 . . . . .	252
5.169	Parallel Coordinate plot for SDSS 588013383816904792 . . . . .	253
5.170	Fitness distribution for SDSS 588017604696408086 . . . . .	254
5.171	Images for SDSS 588017604696408086 . . . . .	255
5.172	Trajectories for SDSS 588017604696408086 . . . . .	256
5.173	Parallel Coordinate plot for SDSS 588017604696408086 . . . . .	257
5.174	Fitness distribution for SDSS 588017604696408195 . . . . .	258
5.175	Images for SDSS 588017604696408195 . . . . .	259
5.176	Trajectories for SDSS 588017604696408195 . . . . .	260
5.177	Parallel Coordinate plot for SDSS 588017604696408195 . . . . .	261
5.178	Fitness distribution for SDSS 588017702948962343 . . . . .	262
5.179	Images for SDSS 588017702948962343 . . . . .	263
5.180	Trajectories for SDSS 588017702948962343 . . . . .	264
5.181	Parallel Coordinate plot for SDSS 588017702948962343 . . . . .	265
5.182	Fitness distribution for SDSS 588017978901528612 . . . . .	267
5.183	Images for SDSS 588017978901528612 . . . . .	268

5.184	Trajectories for SDSS 588017978901528612 . . . . .	269
5.185	Parallel Coordinate plot for SDSS 588017978901528612 . . . . .	270
5.186	Fitness distribution for SDSS 588018055130710322 . . . . .	271
5.187	Images for SDSS 588018055130710322 . . . . .	272
5.188	Trajectories for SDSS 588018055130710322 . . . . .	273
5.189	Parallel Coordinate plot for SDSS 588018055130710322 . . . . .	274
5.190	Fitness distribution for SDSS 758874299603222717 . . . . .	275
5.191	Images for SDSS 758874299603222717 . . . . .	276
5.192	Trajectories for SDSS 758874299603222717 . . . . .	277
5.193	Parallel Coordinate plot for SDSS 758874299603222717 . . . . .	278
5.194	Fitness distribution for Arp 148 . . . . .	279
5.195	Images for Arp 148 . . . . .	280
5.196	Trajectories for Arp 148 . . . . .	281
5.197	Parallel Coordinate plot for Arp 148 . . . . .	282
5.198	Fitness distribution for CGCG 436-030 . . . . .	283
5.199	Images for CGCG 436-030 . . . . .	284
5.200	Trajectories for CGCG 436-030 . . . . .	285
5.201	Parallel Coordinate plot for CGCG 436-030 . . . . .	286
5.202	Fitness distribution for Arp 272 . . . . .	287
5.203	Images for Arp 272 . . . . .	288
5.204	Trajectories for Arp 272 . . . . .	289
5.205	Parallel Coordinate plot for Arp 272 . . . . .	290
5.206	Fitness distribution for ESO 77-14 . . . . .	291
5.207	Images for ESO 77-14 . . . . .	292
5.208	Trajectories for ESO 77-14 . . . . .	293
5.209	Parallel Coordinate plot for ESO 77-14 . . . . .	294
5.210	Fitness distribution for NGC 5331 . . . . .	295
5.211	Images for NGC 5331 . . . . .	296
5.212	Trajectories for NGC 5331 . . . . .	297
5.213	Parallel Coordinate plot for NGC 5331 . . . . .	298
5.214	Fitness distribution for NGC 6786 . . . . .	299
5.215	Images for NGC 6786 . . . . .	300
5.216	Trajectories for NGC 6786 . . . . .	301
5.217	Parallel Coordinate plot for NGC 6786 . . . . .	302
5.218	Fitness distribution for SDSS 587736523764334706 . . . . .	303

5.219	Images for SDSS 587736523764334706 . . . . .	304
5.220	Trajectories for SDSS 587736523764334706 . . . . .	305
5.221	Parallel Coordinate plot for SDSS 587736523764334706 . . . . .	306
5.222	Fitness distribution for SDSS 587747120521216156 . . . . .	307
5.223	Images for SDSS 587747120521216156 . . . . .	308
5.224	Trajectories for SDSS 587747120521216156 . . . . .	309
5.225	Parallel Coordinate plot for SDSS 587747120521216156 . . . . .	310
5.226	Fitness distribution for SDSS 588007005230530750 . . . . .	311
5.227	Images for SDSS 588007005230530750 . . . . .	312
5.228	Trajectories for SDSS 588007005230530750 . . . . .	313
5.229	Parallel Coordinate plot for SDSS 588007005230530750 . . . . .	314
5.230	Fitness distribution for SDSS 758877153600208945 . . . . .	315
5.231	Images for SDSS 758877153600208945 . . . . .	316
5.232	Trajectories for SDSS 758877153600208945 . . . . .	317
5.233	Parallel Coordinate plot for SDSS 758877153600208945 . . . . .	318
5.234	Fitness distribution for SDSS 587739720308818095 . . . . .	319
5.235	Images for SDSS 587739720308818095 . . . . .	320
5.236	Trajectories for SDSS 587739720308818095 . . . . .	321
5.237	Parallel Coordinate plot for SDSS 587739720308818095 . . . . .	322
5.238	Fitness distribution for Arp 273 . . . . .	323
5.239	Images for Arp 273 . . . . .	324
5.240	Trajectories for Arp 273 . . . . .	325
5.241	Parallel Coordinate plot for Arp 273 . . . . .	326
5.242	Fitness distribution for SDSS 1237678620102623480 . . . . .	328
5.243	Images for SDSS 1237678620102623480 . . . . .	329
5.244	Trajectories for SDSS 1237678620102623480 . . . . .	330
5.245	Parallel Coordinate plot for SDSS 1237678620102623480 . . . . .	331
5.246	Fitness distribution for Arp 244 . . . . .	332
5.247	Images for Arp 244 . . . . .	333
5.248	Trajectories for Arp 244 . . . . .	334
5.249	Parallel Coordinate plot for Arp 244 . . . . .	335
5.250	Kurtosis vs. Skewness of the fitness distribution. . . . .	336
5.251	The galaxy pairs with the three lowest fitness distribution skewnesses. . . . .	337
5.252	The galaxy pairs with the three highest fitness distribution skewnesses. . . . .	338
5.253	Skewness vs. Number of Simulations Viewed. . . . .	339

5.254	Morphological Group 1 . . . . .	340
5.255	Eccentricity and inclination for Group 1 . . . . .	341
5.256	Morphological Group 2 . . . . .	342
5.257	Eccentricity and inclination for Group 2 . . . . .	342
5.258	Morphological Group 3 . . . . .	343
5.259	Eccentricity and inclination for Group 3 . . . . .	344
5.260	Morphological Group 4 . . . . .	345
5.261	Eccentricity and inclination for Group 4 . . . . .	345
5.262	Morphological Group 5 . . . . .	346
5.263	Eccentricity and inclination for Group 5 . . . . .	347
5.264	Morphological Group 6 . . . . .	348
5.265	Eccentricity and inclination for Group 6 . . . . .	348
5.266	Morphological Group 7 . . . . .	349
5.267	Eccentricity and inclination for Group 7 . . . . .	350
5.268	Morphological Group 8 . . . . .	351
5.269	Eccentricity and inclination for Group 8 . . . . .	351
5.270	Morphological Group 9 . . . . .	352
5.271	Eccentricity and inclination for Group 9 . . . . .	353
5.272	Morphological Group 10 . . . . .	354
5.273	Eccentricity and inclination for Group 10 . . . . .	354
5.274	Morphological Group 11 . . . . .	355
5.275	Eccentricity and inclination for Group 11 . . . . .	356
5.276	Morphological Group 12 . . . . .	357
5.277	Eccentricity and inclination for Group 12 . . . . .	357
5.278	The distribution of simulation parameters for all galaxies . . . . .	364
5.279	The galaxy pairs with the lowest $\beta$ (top) and highest $\beta$ bottom. . . . .	365
6.1	Photometric Mass vs. Simulation Mass . . . . .	367
6.2	Photometric Mass Ratio vs. Simulation Mass Ratio . . . . .	368
6.3	Star formation rate vs. $t_{min}$ for all galaxies. . . . .	369
6.4	Normalized star formation rate vs. $t_{min}$ for all galaxies. . . . .	370
6.5	SFR for primary galaxies divided by mass ratio. . . . .	372
6.6	SFR for secondary galaxies populations divided by mass ratio. . . . .	373
6.7	SFR for primary galaxies divided by mass ratio. . . . .	374
6.8	Projected separation distance as a function of $T_{min}$ . . . . .	375

6.9	SFR vs. Projected Separation Distance . . . . .	376
6.10	SFR vs. Binned Separation Distance . . . . .	377
A.1	SDSS color thumbnail for Arp 290. . . . .	385
A.2	Arp 290 source pixels with background set to white. . . . .	385
A.3	Arp 290 ackground mask with 4 pixel padding. . . . .	385
A.4	Custom WISE photometry tool measuring Arp 290. . . . .	387
A.5	Custom WISE photometry tool measuring UGC 11751. . . . .	388
A.6	Radial profile and model fits for W1 of the Arp 290 primary galaxy. . . . .	389
A.7	Radial profile and model fits for W1 of the Arp 290 secondary galaxy. . . . .	389
B.1	Spearman Rho for $n = 1000$ . . . . .	398
B.2	Spearman Rho for $n = 100\ 000$ . . . . .	399
B.3	Spearman Rho for Repeated Sorts of $N = 1000$ , Accuracy = 0.75 . . . . .	401
B.4	Number of Comparisons vs. Comparison Accuracy . . . . .	403
C.1	The target identification screen. . . . .	407
C.2	Target tab with downloaded image. . . . .	407
C.3	Disk Info with primary center selected. . . . .	408
C.4	Disk Orientation with background level raised. . . . .	409
C.5	Image after the star to the right of the primary has been erased. . . . .	410
C.6	Automated disk parameters. . . . .	410
C.7	Disks after user adjustments. . . . .	410
C.8	Parameter ranges. . . . .	411
C.9	Simulate tab with target loaded. . . . .	412
C.10	Simulate screen with selected simulations. . . . .	413
C.11	Preferences. . . . .	413
C.12	Review tab. . . . .	414
C.13	Enhance tab. . . . .	415
C.14	Evaluate screen with Merger Wars. . . . .	416
C.15	Statistics tab with table view of states . . . . .	417
C.16	Statistics tab with histogram view. . . . .	417

## Abstract

### A PIPELINE FOR CONSTRUCTING A CATALOG OF MULTI-METHOD MODELS OF INTERACTING GALAXIES

Anthony Holincheck, PhD

George Mason University, 2013

Dissertation Director: Dr. John Wallin

Galaxies represent a fundamental unit of matter for describing the large-scale structure of the universe. One of the major processes affecting the formation and evolution of galaxies are mutual interactions. These interactions can include gravitational tidal distortion, mass transfer, and even mergers. In any hierarchical model, mergers are the key mechanism in galaxy formation and evolution. Computer simulations of interacting galaxies have evolved in the last four decades from simple restricted three-body algorithms to full n-body gravity models. These codes often included sophisticated physical mechanisms such as gas dynamics, supernova feedback, and central blackholes. As the level of complexity, and perhaps realism, increases so does the amount of computational resources needed. These advanced simulations are often used in parameter studies of interactions. They are usually only employed in an ad hoc fashion to recreate the dynamical history of specific sets of interacting galaxies. These specific models are often created with only a few dozen or at most a few hundred sets of simulation parameters being attempted.

This dissertation presents a prototype pipeline for modeling specific pairs of interacting galaxies in bulk. The process begins with a simple image of the current disturbed morphology and an estimate of distance to the system and mass of the galaxies. With the use of an updated restricted three-body simulation code and the help of Citizen Scientists, the pipeline is able to sample hundreds of thousands of points in parameter space for each system. Through the use of a convenient interface and innovative scoring algorithm, the pipeline aids researchers in identifying the best set of simulation parameters. This dissertation demonstrates a successful recreation of the disturbed morphologies of 62 pairs of interacting galaxies. The pipeline also provides for examining the level of convergence and uniqueness of the dynamical properties of each system. By creating a population of models for actual systems, the current research is able to compare simulation-based and observational values on a larger scale than previous efforts. Several potential relationships between star formation rate and dynamical time since closest approach are presented.

## Chapter 1: Introduction

In terms of the large-scale structure of the universe, galaxies represent a fundamental unit of matter. Galaxies tend to be found in groups called clusters, which in turn form a hierarchy of super-clusters tracing the distribution of matter in the universe. Understanding the formation of galaxies and the manner in which they evolve over their lifetimes is important for addressing a number of important questions (Springel et al., 2005). These include what role inflation may have played in the early universe, what is dark matter, and how will the universe as a whole evolve. One of the major processes affecting the formation and evolution of galaxies are mutual interactions. These interactions can include gravitational tidal distortion, mass transfer, and even mergers. In any hierarchical model mergers are the key mechanism in galaxy formation and evolution.

The various interaction effects lead to what one important review (Struck, 2006b) summarizes as five major themes to be studied: the generation of tidal morphologies, induced nuclear activity, induced star formation, the role of collisions in galaxy evolution, and the dependence of these effects on the clustering environment. This dissertation will focus mainly on the generation of tidal morphologies and address in part some aspects of induced star formation. When constructing models of specific systems, achieving a match to the tidal morphology is a necessary prerequisite for obtaining an accurate model. While it might be possible to have matched the shape correctly but have unrealistic predictions related to other characteristics such as star formation rate, it is unrealistic to claim an accurate model for a specific system without a match to the morphology.

Galaxy interactions take place on timescales of a billion years or more. Even though we are able to look back through time to earlier epochs and see galaxies at many stages of interaction, we cannot hope to observe any particular system for more than just a single

instant in time. Because of this static view provided by observations, researchers have turned to simulations of interacting systems.

Previous progress in developing detailed models of specific, observed systems has been ad hoc. The time to model individual systems has been quoted as taking as much as six months (Hammer et al., 2009). Since 2000, a number of researchers have developed semi-automated methods for trying to speed the process, (Theis and Kohle, 2001), (Wahde and Donner, 2001), (Barnes and Hibbard, 2009), (Smith et al., 2010). These methods have seen success in matching artificial systems, but their application to real sets of interacting galaxies usually requires detailed observation data beyond a simple image as well as customized fitness functions.

This dissertation contributes to the study of interacting galaxies by presenting a set of software tools and processes that were used to develop models for 62 pairs of interacting galaxies. These tools incorporate the input of a large number of Citizen Scientist volunteers who helped search for and review candidate models for particular systems. As part of the research done for this dissertation, a website called Merger Zoo (Holincheck et al., 2010b) was launched to enable the contributions of volunteers from the Zooniverse community. The specific models developed with the help of the volunteers, the tools for processing and analyzing the results, and the set of evaluated results will all contribute to future work in this field.

## **1.1 A Short History of the Study of Interacting Galaxies**

There are several excellent reviews in the literature discussing the history, observational evidence, and an overview of simulation techniques for studying interacting galaxies. Three of them are Barnes and Hernquist (1992), Struck (1999), and Struck (2006b). Noting the growth and diversification in the field since the early 1970s, the authors of each paper make note of how it is not practical to give a comprehensive overview in a single volume.

In the early part of the 20th century, astronomers came to realize that many of the nebulous objects they had been observing were actually galaxies of similar significance to

the Milky Way. With a universe full of “island universes” to catalog, Hubble proposed a classification scheme where galaxies were categorized based on their morphologies. His scheme, with modifications, is still the most widely used system. To illustrate his categories, Hubble famously proposed a tuning fork diagram with elliptical galaxies on the handle, two types of spirals on the forks, and lenticular galaxies at or near the intersection. If the fork is oriented horizontally, the handle is on the left with barred spirals on the bottom fork and non-barred spirals on the top fork. Hubble suggested a possible evolutionary sequence where galaxies would become more organized as they transitioned from ellipticals to lenticulars and then into spirals of increasingly loose spiral arms. This speculated order of galaxy types resulted in astronomers referring to ellipticals as early-type galaxies and spirals as late-type ones. This particular evolutionary path has been discredited, but yet the terminology remains.

Galaxies that were unsymmetrical were called irregular galaxies by Hubble. The first type of irregulars have a rough or clumpy appearance. The second type have a smooth appearance. De Vaucouleurs added refinements such as intermediate categories on the prongs of the tuning fork and a third prong for weakly barred systems.

Equipped with this classification system, astronomers of the early- to mid-20th century began amassing catalogs of galaxies and their Hubble-types. Among the important realizations to come from this effort were that galaxies tended to exist in large clusters, and that not all galaxies could be assigned an exact Hubble type. The second realization, that some galaxies had peculiar, though not necessarily irregular, shapes led some to speculate as to the origins of these morphologies. The fact that galaxies were often found in dense clusters provided a possible explanation: that these isolated, island universes may on occasion pass close to one another and interact. These interactions may be simple close passages, collisions, or even captures resulting in mergers.

One of the first to speculate that tidal forces caused peculiar morphologies was Zwicky in the 1950s. At the time his theory was considered unlikely because collisions between galaxies were thought to be improbable (Struck, 2006b) and that gravity alone would not

be able to cause such narrow features (Barnes and Hernquist, 1992). In the 1960s, working with the data collected from the Palomar sky survey, Arp collected images of over 300 peculiar galaxies. He published this atlas of objects with his attempt at a categorization scheme.

Several attempts in the 1960s and 1970s were successful in demonstrating that tidal distortions cause by gravity could produce the thin tails and bridges seen in interacting galaxies. The researchers used a simple model for pairs of interacting galaxies. Each galaxy was treated as a point mass and is surrounded by a disk of mass-less test particles. The particles in each disk are influenced by the gravity of both galaxies. These models are referred to by several names, but will be called restricted three-body methods in this work. The results from these models showed that the more dramatic features of interacting galaxies are produced by slow, close passages along parabolic or elliptical orbits. Faster passages usually result in smaller features being generated.

The most often cited of the restricted three-body papers from this time is Toomre and Toomre (1972). The authors of this work were able to produce simple, yet credible models of four well-known pairs of interacting galaxies. This seemed to answer conclusively that it was possible to recreate the general morphology of interacting systems, especially the tails and bridges, with gravitational tidal disturbances. Additionally, Toomre and Toomre (1972) predicted tidal that effects on the gas in the galaxies could lead to enhanced star formation as well as gas transfer to nuclear regions. From the 1970s onward, the understanding of interacting systems was influenced by both observations and simulations.

## **1.2 Some Important Physical Processes in Interacting Systems**

Galaxies are composed of luminous stars, an interstellar medium of gas and dust, and some form of non-luminous but gravitationally-charged, dark matter. In general, stars and dark matter are treated as collisionless. The interstellar medium (ISM) is generally treated as

a compressible fluid (Barnes and Hernquist, 1992), the primary component being the gas. In each of the interacting galaxies gas can collide, be compressed, dissipate, heat, or cool. All three components feel the effects of gravity. However, the gas experiences a number of interesting phenomena in response to being disturbed when two galaxies interact. For modeling purposes, the physics is generally divided into gravity and gas dynamics.

### 1.2.1 Gravity

Most restricted three-body methods ignore gas. Most use an analytic formula to describe the gravity potential for the disk of each galaxy. Depending on the nature of the potential, they may either ignore or include contributions from dark matter. Usually, the entire mass for the galaxy is assigned to a particle located at the center of the galaxy. Massless test particles are then distributed randomly around the disk to sample this potential. At each time step, the acceleration of a test particle is calculated by adding the acceleration from each of the two disks. The three bodies in this calculation are the two galaxy centers of mass and the test particle. This is repeated for all test particles. With the acceleration information, the particle positions and velocities are updated for the next time step. Toomre and Toomre (1972) demonstrated that this simple model will generate simulations with tidal features matching real systems.

Real galaxies do not have all of their mass concentrated exactly at their centers. Also, in real interacting systems, as the orbits of stars and dark matter are disturbed, their responses cause the distribution of mass to change which represents a change to the gravitational potential. Restricted three-body methods fail to account for these changes and are said to have rigid potentials.

A more realistic treatment of gravity is to use an N-body method. In these simulations multiple massive particles are used to represent each galaxy. To achieve a self-consistent representation of the potential, typically the number of particles will exceed 10 000. With the mass in the system distributed throughout, the potential will respond along with the mass as the orbits of particles are tidally disturbed. The mutual interactions between the

particles also allows for the modeling of smaller pieces of the galaxies experiencing self-gravity. For example, a tidal tail formed during the close passage of two galaxies may initially appear rather broad. The gravity experienced between simulation particles in the tail may cause it to contract into a thinner shape. In real systems, dense regions of star formation are sometimes observed near the tips of tidal tails. These regions may be forming tidal dwarf galaxies (Struck, 2006b). This behavior is largely missing from restricted three-body models. Wallin et al. (1990) demonstrated that even in the absence of self gravity in tidal features, restricted three-body models could produce realistic density enhancements. Combining these density increases with a simple Schmidt law allows even these simple models to estimate induced star formation.

The most straight forward way to calculate the mutual gravitational acceleration in N-body simulations is with direct summation. Iterate over all particles in the simulation. For each particle, iterate over all other particles and calculate the potential between the two particles, adding to the total felt by the particle. This requires  $\sim O(N^2)$  operations to compute the potential for all particles at each time step. In the mid 1980s the hierarchical tree method was developed. In this method, particles are organized in a hierarchical tree data structure. When calculating the potential experienced by a particle, proceed as before with direct summation. However, for distant particles, use low-order multipole expansions to approximate the potential. This expansion is aided by use of the tree structure. These tree codes can compute the potential at each time step with  $\sim O(N \log N)$  operations (Barnes and Hernquist, 1992). Other methods for calculating the potential are grid-based and can be a hybrid of grids and particles.

### 1.2.2 Gas Dynamics

In general, much of the gas in galaxies will be in the form of neutral hydrogen atoms, HI. This gas emits photons at a wavelength of 21cm due to electrons jumping between split, hyperfine levels in the ground state. The radiation observed due to this quantum effect is an effective means of tracing the presence of HI gas in a galaxy. When treated as an emission

line in the spectrum of a galaxy or region of a galaxy, it is useful to look for any broadening or shift in this line. Such effects allow measurements of the line-of-sight velocity of this gas with respect to the observer to be measured. This velocity information describes how the gas in the system is moving.

The gas in interacting galaxies will follow the same trajectories as stars only as long as pressure forces remain small. However, shocks in the gas will redistribute the angular momentum between various components causing the path of the gas to be very different than the stars. Shocks can form through several means. One could be direct collisions between gas in one galaxy and the other for encounters when the two galaxies actually pass through one another. Another shock forming effect are the various gravitational torques and tides in the system. Any time two different parcels of gas experience different forces, they can develop a relative velocity leading to one parcel overtaking the other and causing a shock. Originally it was suspected that large-scale shocks would end up stripping gas from interacting galaxies. However, numerical results in 1980s and 1990s showed that the shocked gas would remain in the galaxy and may even be driven towards the center (Barnes and Hernquist, 1992).

Gas that has been shocked, or merely had its density increased, may experience localized, gravitational collapse. This collapse may lead to the formation of new stars. Stars with a range of masses will form in this manner, with new stars being surrounded by gas. The more massive stars will be O and B type. They are much hotter and radiate intensely in the UV range. This radiation will interact with the surrounding gas. The UV radiation from hot, young stars will ionize the hydrogen atoms turning clouds of HI into regions of HII. The protons and electrons will recombine to form neutral hydrogen again with the electron in a random energy level. The subsequent cascade will often include the transition from  $n=3$  to  $n=2$ , known as the Balmer  $\alpha$  or  $H\alpha$ . This causes the electron to emit a photon with a wavelength of about 656nm, which produces an emission line. By measuring the strength of the  $H\alpha$  emission line one can estimate the star formation rate in that location.

Star formation converts gas into stellar mass. Some of this mass will be in the form

of low mass stars which will have very long lifetimes  $\sim 10$  Gyrs. Other portions of this mass will be in the form of high mass stars which will have short lifetimes. Towards the later stages of their lifetime, these stars will have strong stellar winds which will disperse the gas in the vicinity. This could lead to localized density increases, triggering more star formations. But if the winds are strong enough they could blow the gas into a lower density configuration which would actually decrease or cut off star formation. These interactions with the surrounding gas can affect the temperature along with the density.

In addition to triggering star formation, the effect on the gas in interacting systems can cause large portions of it to lose angular momentum. This gas will then be driven towards the nuclear regions of the galaxy. Interactions may be an important mechanism for fueling nuclear activity.

Clearly, the behavior of gas in interacting galaxies must be included in any model that wishes to accurately describe the system (Barnes and Hernquist, 1992). Two broad categories of methods are particle-based or grid-based. For particle-based methods, the gas can be treated as a set of discrete clouds that undergo inelastic collisions. These methods are referred to as "sticky particle". Another particle method is called smoothed particle hydrodynamics (SPH). Even though it is particle-based, it treats the gas as a continuous medium. The physical properties tracked by gas particles in this method are calculated using smoothing kernels that weight the contributions of nearby particles. The second broad category of methods is grid-based. In the last several years, some researches have been using adaptive mesh refinement (AMR) to make grid hydrodynamic calculations more spatially adaptable. AMR allows localized regions of complex interactions to be modeled with a relatively fine grid while less active regions can be modeled with a coarse grid.

### 1.3 Comparing Models to Observations

In computational science it is important to relate simulations to empirical data<sup>1</sup>. This is similar to theoretical sciences that develop models which make predictions or attempt to explain observed data. Both the restricted three-body and the full N-body methods produce simulation output which can be evaluated based on morphology. If the shape, position, and orientation of tidal features matches the observed values, then the simulation is a plausible approximation of the real system. This is the first and most important constraint on a model of a specific system of interacting galaxies.

Researchers are able to provide a qualitative assessment of how well a simulation matches an image of a pair of interacting galaxies. Toomre and Toomre (1972) included a relatively small number of test particles in the graphical presentation of their simulation results. As computers increased in CPU speed and graphics capabilities, the number of particles was increased. An important early effort to incorporate interactive graphics was Borne (1988). The authors in that work demonstrated the utility of creating a simulated surface density plot of particle mass. They in turn were able to construct simulated isophotes for comparison with observations.

Other constraints on the models can involve the gas dynamics. If a model includes a treatment of the gas in the system, it may be possible to calculate the line-of-sight velocity of the gas for various regions in the simulated galaxies. These velocities can be compared to observed HI velocity data. One of the first to incorporate this approach was Borne (1988). Additionally, some models can predict a star formation rate (SFR) based upon the density of gas in the system. This predicted SFR can be compared to values based on measurements of the H $\alpha$  emission line.

In addition to using observational data to constrain a model solution, the dynamical parameters from a simulation can be compared to observed values in an attempt to determine any correlations. For example, the closest distance between two galaxies in a recent passage

---

<sup>1</sup>Barnes (2011) gives a useful definition of the process of modeling interacting galaxies as a search “to define a mapping from the current morphology and kinematics of a tidal encounter back to the initial conditions.”

might correlate with the SFR. To determine if this is true, the SFR and closest distance for multiple systems can be used to construct a scatterplot. This type of correlation, while potentially more meaningful with a full N-body model solution, can be performed with the dynamical parameters from restricted three-body models.

## 1.4 Automating the Search for Models of Interacting Galaxies

Determining the dynamical parameters for a model of a real system of interacting galaxies can be a time-consuming process. Toomre and Toomre (1972) offered a series of coarse, yet revealing, parameter studies. For example they showed the different morphologies resulting by varying the inclination angle while holding other values fixed. Using the physical intuition gained from such studies, researchers attempting to model a specific system will narrow the range of simulation parameters to be used. However, there is a tremendous amount of trial and error involved in finding a best-fit orbit. This is especially true if one is trying to match the kinematics data from the simulation to observations. Hammer et al. (2009) claims that "[t]he accurate modelling of both morphology and kinematics takes several months, from two to six months for a well-experimented user."

Several attempts at speeding and even automating this process have been published. Wahde (1998) was one of the first to demonstrate the use of a genetic algorithm for optimizing models of interacting galaxies. A genetic algorithm (GA) uses the evolutionary processes of crossover and mutation to randomly assemble new offspring from an existing population of solutions. The parent solutions are chosen to generate offspring in proportion to their fitness. The more fit, or better matched to the target system, an individual model is, the more often it will contribute its genetic information to subsequent generations. The genes in this GA approach are simply the dynamical model parameters like inclination, mass ratio, disk orientations, etc. The fitness function to be evaluated and optimized needs to provide a meaningful qualitative value for how well a given simulation result matches the

target system. A value of 1 means a perfect match, lower values indicate worse matches.

With a fitness function defined, a GA is seeded with an initial population and then set to evolve for some number of generations. Researchers trying to optimize galaxy models will use a population size of around 50 and will evolve for 50 generations. There is an extensive body of research on the convergence behavior of GAs in terms of the nature of the fitness landscape being studied and the particular evolutionary mechanisms invoked (De Jong, 2006).

At least three groups have published results of GA optimization of models of interacting galaxies: Wahde and Donner (2001), Theis and Harfst (2000), and Smith et al. (2010). They all demonstrate convergence to one or few best-fit models for real systems. However, the convergence radius for these systems is not well documented however. A large radius of convergence (perhaps even global in scale) is demonstrated by Wahde (1998) and others when they are modeling *artificial systems*. These systems use the simulation code itself to generate a high-resolution simulated observation of a hypothetical system of interacting galaxies. The researchers are then able to use their GA to optimize and find a close fit to the known dynamical parameters. Additionally, to demonstrate convergence as well as some amount of uniqueness, it is customary to take the resulting best-fit models, apply a set of random alterations to the dynamical parameters, and then use these altered models as the initial population in a new GA run. If this population converges to the same best-fit model, then some confidence in the local uniqueness of the model is gained. However, Smith et al. (2010) found four distinct best-fit models for the pair of galaxies NGC 7714/5. Even though, they used some kinematics data in their fitness calculation, this demonstrates the potential degeneracy within the models when the fitness is based primarily on the morphology.

The performance of GAs as a method for optimizing a fitness function over a particular domain is much better than an exhaustive search. Consider a model that requires 12 parameters. In order to sample each parameter at 10 values along its range,  $10^{12}$  simulation runs will have to be performed. If a single simulation runs in about 1 sec, one would need over 30 000 years (assuming no increase in computing power) to perform an exhaustive

search for a single system. Reducing the number of free parameters obviously brings down the number of simulations to be performed, but doing so might over constrain the system and not find a good match. GAs allow for a reasonable approximation to a full parameter search. By randomly distributing the initial population and evolving for several generations, there is also no restriction on the resolution used to sample the parameter space. Further confidence in the convergence of the model can be gained by performing multiple runs with different initial random populations. If one considers the time to perform 50 runs of a population of 50 models for 50 generations, it would take less than 36 hours to perform the needed simulations at a rate of one per second.

For both cases, exhaustive search and GA, the total time needed to compute the simulations can be sped up by distributing the work across multiple machines. Both methods can be treated as embarrassingly parallel. However, it is much more reasonable to expect access to 36 machines for one hour than to get access to 30 000 machines for one year. These estimates assume a time per simulation of about 1 sec. This is only reasonable for a restricted three-body simulation. A single full N-body simulation can take hours to days to run. After finding the best-fit model using the GA or exhaustive search, the researcher is likely to want to use these parameters as an initial guess in a full N-body simulation. The intention is that by restricting the ranges of each of the parameters with a lower fidelity model, the optimized full N-body model can be found with fewer trials. This approach is known as multi-method-modeling (Wahde and Donner, 2001). The combination of GAs, parallel computing, and multi-method-modelling provides extreme time savings. It has the potential to reduce the time it takes to find an accurate full N-body model of a system from several months with manual methods to several days. The GA also provides some assurance that a reasonable volume of parameter space was sampled and that the optimized solution is unique.

One would expect that the prospect of applying these automated matching systems to a large number of real galaxies would have these research groups modeling large populations of galaxies. Unfortunately, this has not yet happened. For a single study (Hammer et al.,

2009), their population of 33 is the largest number of models attempted for specific systems, of which we are aware. They used the Identikit software from Barnes and Hibbard (2009) and related simulation code to run full n-body simulations of just 48 different sets of initial conditions. They then altered the viewing angles of the results to find qualitative matches to the morphology and observed kinematics of their sample of 33 galaxy mergers. They fit 33 mergers with a total of 48 simulations run, not 48 per merger. The authors of that study (Hammer et al., 2009) correctly indicate that uniqueness of their solutions is not guaranteed. There are several drawbacks to these current automated methods that hinder widespread application.

#### 1.4.1 Current Methods

The main reason these gridded methods are not generally applicable is due to the characteristics of the fitness functions used. The methods are relying on the use of kinematics data as part of the fitness calculation. This data is not available for the vast majority of systems where disturbed morphologies are observed. So a reliable fitness function that depended mainly on morphological information would be more widely applicable. Also, the scoring of close matches is not always performed correctly with existing fitness calculations.

At the present time, it appears that all researchers using GAs have implemented some form of the same basic fitness function. These groups compute some type of distance map between an image of the target system and the simulation output. If you treat the plane of the sky projection of particle positions (and gas) as an image, one can scale and align it to the target image. From there one computes the difference between the two images in each region. The number of regions ranges from a coarse five by five grid on each image to a pixel by pixel comparison. The difference can be a simple subtraction or a chi-squared value. One modification to this basic fitness function involves identify regions where tidal features occur and granting them additional weights in the difference calculation. Another change is to use a mask to block out any pixels of the central disks regions from the differencing. Both measures allow the tidal features to have more impact on the final score for a model.

The latest enhanced fitness functions also make use of the kinematics data for galaxies. A similar difference calculation can be performed for the simulated and observed plots of this data. Using the kinematics data is thought to better guarantee an accurate best-fit model.

Another problem with these grid-based methods is that they can produce non-sensical relative results. These occur when two non-exact matches are scored. One is judged subjectively to be much worse than the other by human experts yet still receives a higher fitness score by the algorithm. Consider a system with symmetrical tidal tails running north and south from the disk of the galaxy. Two different simulation results are generated. One also has symmetrical tidal tails. However they are rotated away from the north-south line. Another simulation produces two tidal tails that are not symmetric. One points east, the other points south. This situation is depicted in Figure 1.1. While neither model is an exact match, which one is closer to the real system? An expert might judge the symmetrical system to be closer. It can be made into an exact match simply by rotating the system a few degrees. The south/east tail combination would have to undergo much more substantial changes to become a match, so it should score worse. Unfortunately, when using the difference calculation described above, the system with the symmetrical tails will receive a lower score because neither tail lines up. The other system will be scored higher because at least the southern tail matches the target. One could imagine additional pathologies of this type for other morphologies as well.

#### **1.4.2 The Role of Citizen Scientists**

A human evaluator reviewing simulation output can be seen as applying a more robust fitness function. Our visual processing capability will allow us to see similar morphologies where the simple difference calculation will not. Also, by focusing on people's ability to match similar shapes, additional observational data is not needed to achieve initial convergence on morphologies. Using a human fitness function allows this method to be applied to a large number of systems that may lack additional observational data, such as kinematics. A single reviewer may lose interest after viewing 125 000 simulations of the same system

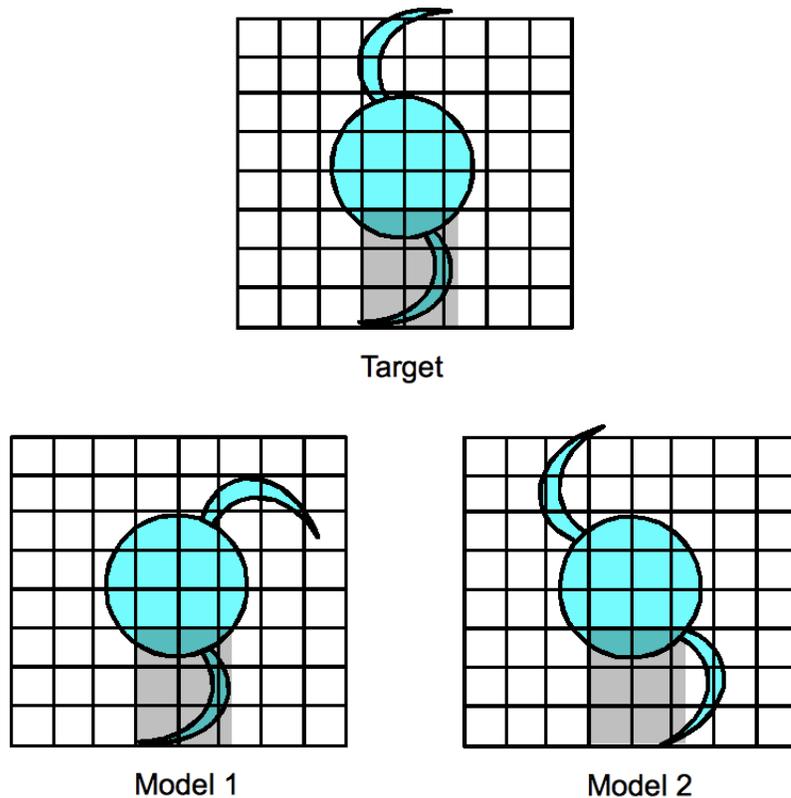


Figure 1.1: An example where the gridded fitness function fails. The highlighted gray area shows the grid method would score Model 1 higher than Model 2.

looking for morphological matches. However, if the work is distributed across 1 000 volunteers, each one would only need to review 125 simulations. This would achieve the same number of samples of parameter space as the 50/50/50 GA, though purely random samples would not necessarily guarantee convergence.

Volunteers can be trained to recognize morphological similarities between between simulations and images of real systems. By applying this training, when presented with a set of simulations based on randomly chosen parameters, they can identify the ones that are plausible matches. Additionally they can be presented with subsequent rounds of pre-vetted simulations to identify the best-of-the-best of each round. In this way the volunteers, acting now as Citizen Scientists, can help explore parameter space and characterize the fitness of simulation outputs at each location in that space that is sampled. By combining the efforts

of thousands of Citizen Scientists, detailed knowledge of the fitness landscape is gained. This knowledge is useful for the direct identification of the best-fit model. However, a collection of human-scored simulations also provides an opportunity to train new automated fitness functions. These fitness functions could use algorithms that better recreate the human judged scores for simulation outputs. Once automated, these fitness calculators could be used with existing GAs to achieve more reliable best-fit matches.

### **1.4.3 A Pipeline for Constructing Models of Systems of Interacting Galaxies**

This dissertation will focus on building a pipeline for constructing multi-method models of interacting galaxies. The pipeline will start with target identification and processing. It will proceed to presenting the results of restricted three-body simulations to Citizen Scientist volunteers and collecting their scores for each simulation. The best-fit models will be determined from these scores. The dynamical parameters from these models can be compared to observations and derived values, such as SFR. They can also be used as initial guesses in full N-body simulations to produce more detailed simulation outputs. Finally, the collection of annotated simulation outputs will be accumulated into a training set of data to improve automated fitness functions.

Figure 1.2 presents an overview of the pipeline. The input is an image of a pair of interacting galaxies. The physical scale is set with some additional information including a distance estimate (usually from redshift) and a mass estimate (usually from luminosity or color information). The stages of the pipeline will result in a simulation model and orbit parameters.

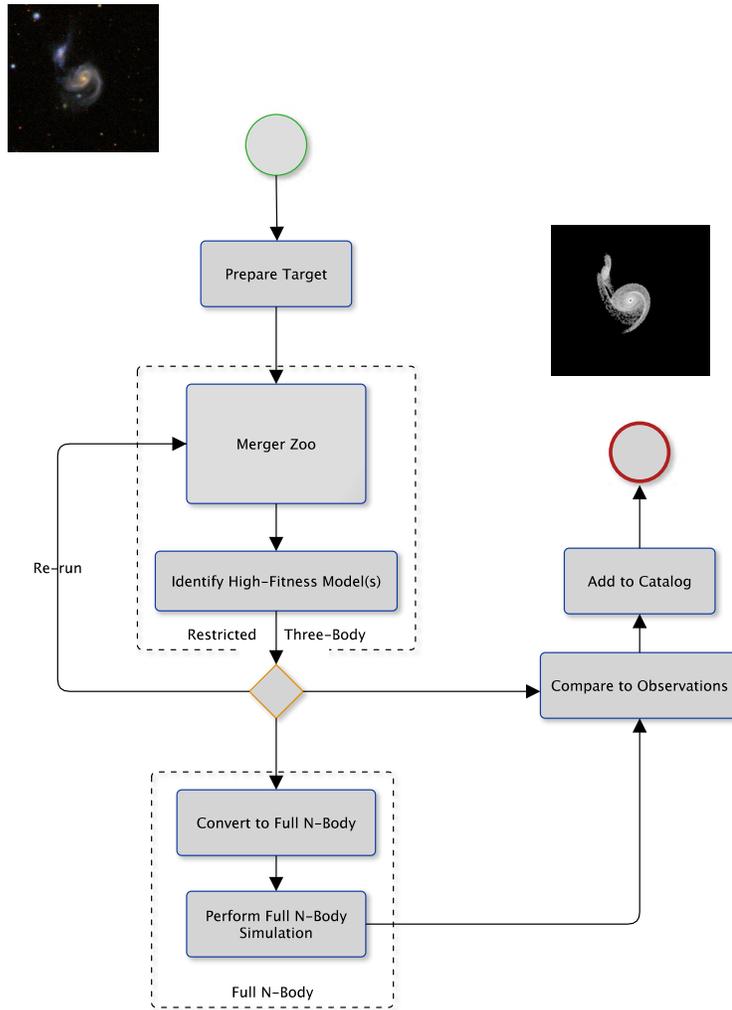


Figure 1.2: Pipeline overview.

## Chapter 2: Modeling Galaxies

Galaxies are host to most of the observable matter in the universe. They have sizes and masses that range several orders of magnitude with dwarf galaxies having a few hundred million to a billion stars, large spiral galaxies like our own Milky Way with over 100 billion stars and a diameter of 100,000 light years, and giant ellipticals that can have ten times as much mass. Binney and Tremaine (2008) name four principal constituents for galaxies like the Milky Way: stars, interstellar medium, central black hole, and dark halo. Using our own galaxy as the scale, there are on the order of  $10^{11}$  stars with a total mass close to  $5 \times 10^{10}$  times that of our sun. The interstellar medium, or ISM, consists of gas (mostly atomic and molecular hydrogen) and dust (the common name for small solid particles). The ISM has a mass of about 10% of that of the stars. The central black hole, a feature recognized as common to most galaxies like the Milky Way, has a mass on the order of  $10^6$  solar masses. The fourth component, the dark halo, is composed of weakly interacting dark matter particles. It is the largest component of a galaxy both in mass and size with on the order of  $10^{12}$  solar masses of matter and a radius on the order of 200 kpc. The thin disk where the stars orbit the center of the galaxy is only on the order of 10 kpc in radius and a thickness of about 0.5 kpc. Stars orbit in this disk in roughly circular orbits with velocities of  $\sim 100 \text{ kms}^{-1}$ . That equates to an orbital period around the galaxy on the order of 100 Myr. This period, along with the short lifetime of type O stars, establishes the key time scale for simulating galaxies.

Modeling all four constituents simultaneously is still a difficult proposition. Hopkins et al. (2013) continue to add additional feedback mechanisms between formation of new stars and the gas dynamics in simulations. Tools such as these are often used in parameter

studies where hypothetical systems are studied by holding most parameters fixed and varying a few key simulation inputs to study their effects. These studies continue to enhance our understanding of detailed physical processes in interacting galaxies. The approach in this research, is somewhat different. Our objective is to determine realistic orbit parameters to describe the relative orbits of a large number of actual systems. Toomre and Toomre (1972) showed that the first order effects of gravity are the most important factor in the formation of large scale tidal distortions in interacting galaxies. The multi-method approach advocated here focuses on the use of simpler models first to narrow down the orbit parameters. With a much reduced volume of parameter space to explore, more sophisticated, and computationally intensive, models can be constructed.

We use two types of particle codes to simulate gravity in interacting galaxies. The first is the restricted three-body method in which only one particle with mass is assigned to each galaxy and the remaining massless test particles trace the evolution of the potential. The other is a full n-body simulation in which each particle is assigned mass and the mutual interactions are summed to produce a self consistent and more realistic model of interacting galaxies.

## 2.1 Restricted Three-Body

The gravitational potential produces an attractive force between every pair of massive objects. In a restricted three-body simulation<sup>1</sup> a single massive particle is used to represent the entire mass of each galaxy. The massive particles are placed at the center of each galaxy. The disks of the galaxies are then populated with a set of massless test particles. The acceleration acting on each massless particle is the sum of the accelerations produced by each of the galaxy centers. Test particles are randomly distributed around the disk of the galaxies with respect to some distribution function, usually chosen to mimic the actual mass distribution, and then assigned an initial velocity. The second order differential equation for the acceleration due to gravity as a function of position is usually decomposed into two

---

<sup>1</sup>Sometimes also referred to as multiple three-body or restricted multi-body simulations.

coupled, first order differential equations. The first equation sets the time derivative of the position equal to the particle velocity, and the second equation defines the time derivative of the velocity as the acceleration due to galaxy. The simulation can be advanced at each time step by using the previous velocity and computed acceleration to advance the position and velocity respectively of each particle. Many numerical techniques exist that can be applied to solve these equations such as the Euler method of leapfrog integration.

One important reason for the impact of Toomre and Toomre (1972) was the success they had in recreating the disturbed morphologies for four actual pairs of disk galaxies. Other researchers applied a similar approach to modeling elliptical galaxies (Borne, 1984). Simulations of interacting galaxies using the restricted three-body method produce realistic and visually appealing results with only a few thousand particles and can run in under one second on modern personal computers.

## 2.2 N-Body

The mechanics of n-body simulations are similar to the restricted three-body situation. Particles are distributed throughout the disks of each galaxy and assigned initial velocities. However, instead of concentrating the mass of each galaxy into a single particle, every particle in the simulation is assigned a portion of the total galaxy mass. Most simulation techniques assign the same mass to each particle and simply include more or fewer particles for higher or lower mass galaxies in the same simulation. The mutual interaction between every pair of particles has to be considered to compute the acceleration needed to advance the simulation. The simplest approach is to perform this calculation by iterating over each pair of particles and calculating the gravitational acceleration. This method has a computation complexity of  $O(n^2)$ . Increasing the number of particles by a factor of ten increases the simulation time by a factor of 100. This is discouraging because it requires on the order of 10,000 particles to represent the disk of a galaxy. Also, in order to simulate stable disks at equilibrium for a galaxy in isolation, the dark matter halo must be simulated as well. Without the halo, the disk will become unstable on its own and collapse. Simulating an

individual galaxy requires many tens of thousands of particles. Simulating an interacting pair of galaxies obviously requires double the number of particles.

The advantage of N-body simulations is that they include self-gravity. Self-gravity refers to the effects of mutual gravitational attraction between localized groups of particles. This is important in regions that have been disturbed tidally, as during close passages of interacting galaxies. Self-gravity is thought to play an important role in the formation and evolution of tidal features including through induced star formation.

In order to benefit from the enhanced realism of n-body simulations without having to devote unaffordably large amounts of computational resources at a single model, some optimization needs to be performed. One method of speeding up the direct summation of forces in the n-body problem is to use specialized hardware such as the GRAPE boards (Baumgardt and Makino, 2003) or modern Graphical Processing Units or GPUs (Portegies Zwart et al., 2007). Another approach is to find approximations for the force of gravity.

### 2.2.1 Tree Codes

One popular approximation technique distinguishes between close and distant interactions between pairs of particles. Consider the force felt by a particle due to two other particles that are very distant to the first particle. These two particles are located close to one another. The force felt by the first particle could be approximated by calculating the center of mass of the two particles and treating them like one new particle. The advantage of this method comes when considering larger groups of distant particles. Their combined force on a distant particle can be approximated by computing their center of mass. To further improve this approximation, an expansion of the gravitational moments up to and including the quadrupole moment is calculated for each group. The error in this approach is usually controlled through a single parameter,  $\theta$ . This parameter is also called the opening angle and is used to determine distance and size of a group of particles relative to a particle of interest before the approximation can be applied. To aid in the identification of groups and the tracking of their parameters, a tree data structure is used. This hierarchical data

structure divides the simulation volume up using a recursive binary splitting algorithm, or other balancing technique, to identify groups of particles. The tree code offers performance of  $O(n \log n)$ . This is a significant improvement over the direct summation technique. This technique was first proposed by Barnes and Hut (1986). Tree Codes have since been parallelized and also ported to run on GPU hardware. Simulations with billions of particles are now possible. We are approaching a time, within a few iterations of Moore’s Law, when simulations with one particle per star will be possible.

## 2.3 Additional Physical Processes

Section 1.2 describes several physical processes that are important for simulating interacting galaxies. Here we describe two. The first is dynamical friction, which is inherent in  $n$ -body simulations but is missing in restricted three-body ones. The second is the potential inclusion of gas dynamics.

### 2.3.1 Dynamical Friction

Self-consistent  $N$ -body codes have demonstrated that the orbits of secondary galaxies will decay over time. One important process that leads to the loss of orbital energy is scattering in the form of dynamical friction. These codes can also produce other multi-body effects like violent relaxation. These effects are absent in restricted three-body codes. The orbital decay, even during a first passage encounter, can be significant. Dynamical friction plays a key role in galaxy evolution through other interactions such as between a bar and the dark matter halo. A parameterized version of this effect leading to orbital decay is derived in Binney and Tremaine (2008)

A massive body  $M$  moving through a field of other massive particles will interact with them through the gravitational force. The field particles have individual masses much less than  $M$ . However, these field particles are part of an overall system that is very massive and large. It is customary to approximate this system as infinite and homogeneous, with the distribution of velocities taken to be Maxwellian. As the body  $M$  moves through this

field of stars, the field stars will be deflected resulting in an enhanced density behind the massive body sometimes referred to as a wake. The attraction of this wake on the moving body is opposite in direction compared to its velocity resulting in dynamical friction.

For a set of background masses of density  $\rho$  and a Maxwellian distribution of velocities with dispersion  $\sigma$ , Chandrasekhar's dynamical friction formula (Chandrasekhar and von Neumann, 1943) for the acceleration becomes

$$\frac{d\mathbf{v}}{dt} = \frac{4\pi G^2 M \ln \Lambda}{v^3} \left[ \operatorname{erf}(X) - \frac{2X}{\sqrt{\pi}} e^{-X^2} \right] \mathbf{v}$$

$$X = \frac{v}{\sigma} \tag{2.1}$$

We define the following values useful for calculating the velocity dispersion  $\sigma$

$$p(r) = G \int_0^r \frac{\rho(r)m(r)}{r^2} dr$$

$$v_r^2 = \frac{p(r)}{\rho(r)}$$

If our multi-model method approach is to succeed, we need to be able to convert models derived using restricted three-body methods to full n-body simulations. Ignoring the effects of dynamical friction will make this process much more difficult. By including it, we reduce

the likelihood of needing to perform redundant optimization of the orbit parameters again when the full n-body simulations are computed.

### 2.3.2 Modeling Gas

Barnes and Hernquist (1992) emphasizes the importance of including gas dynamics in simulations. In addition to self-gravity, gas dynamics produce important effects that result from close passages of interacting galaxies. Modeling these effects, including induced star formation, is important to ensure accurate models. The technique considered for simulating gas dynamics here is a particle-based technique known as smoothed particle hydrodynamics (SPH). The particles in SPH represent regions of gas in which the values representing the state of the gas, such as density and temperature, are smoothed with a kernel. The SPH particles track the gas properties at each location and interact with one another during the simulation allowing the dynamics of the gas to be modeled. SPH can be used with both restricted three-body and n-body simulations.

For this dissertation no simulations with gas particles were run. The inclusion of the discussion of gas dynamics is intended to highlight its importance to producing realistic simulations of galaxies and to afford an opportunity to say that nothing about the multi-method approach inhibits the inclusion of gas effects.

## 2.4 Specific Codes

Multiple galaxy simulation codes exist. For performing N-body simulations GADGET, along with its derivatives GADGET-2 and GADGET-3, is the most commonly used code by researchers today. One drawback of the current GADGET code is that it does not provide a built-in mechanism for setting the initial conditions of simulations. Researchers must establish particle masses, positions, and velocities elsewhere with another tool and then use GADGET to perform the simulation. Multiple restricted three-body codes exist as well. However, having fallen out of favor with the advent of tree codes, many researchers have left these simpler codes behind to focus on full n-body simulations. One code currently

being used by at least one research group is called MINGA (Petsch and Theis, 2008). It includes a treatment of dynamical friction but the source code does not appear to have been published for use by others. Similarly Mihos (2002) has produced an online Java Applet for the GalCrash code which allows users to run restricted three-body simulations in real time. Again, the source code does not appear to have been made available.

This research makes use of two codes produced by Wallin and other researchers. The first is a restricted three-body code called SPAM<sup>2</sup> (Wallin, 1990). The other is a full n-body tree code called MASS99<sup>3</sup> (Antunes and Wallin, 2001). The benefits of using these two codes in a multi-model method is that both codes were developed by the same researcher using similar coordinate systems. This simplifies the process of converting between the restricted three-body and the full n-body models.

#### 2.4.1 SPAM

The single greatest advantage of the SPAM code is that it is formulated to make the center of mass of the primary galaxy the origin of the coordinate system. This makes it a very simple task to match up each simulation time step with a target image. It also requires that we define the relative orbit of the two galaxies in terms of the position and velocity of the secondary galaxy with respect to the primary. This is somewhat different from the usual step of setting the origin at the center of mass computed from both galaxies together.

#### Included Potentials

The SPAM code includes two potentials<sup>4</sup>. The first is a purely analytic potential for gravity using a softened point mass. The expression for acceleration of a massless test particle due to gravity is modified to include a softening length,  $\epsilon$ , as shown in Equation 2.2.

---

<sup>2</sup>SPAM is the Stellar Particle Animation Module.

<sup>3</sup>MASS is the Multi-physics Astrophysical Simulation Software.

<sup>4</sup>A third undocumented potential exists as well. It is a Modified Newtonian Dynamics (MOND) potential. It is based on the analytic expressions in Scarpa (2006)

$$\vec{a} = -\frac{GM}{(r^2 + \epsilon^2)} \hat{r} \quad (2.2)$$

The gravitational potential is given below.

$$\phi(r) = \frac{GM}{\epsilon} \left[ \frac{\pi}{2} - \tan^{-1} \left( \frac{r}{\epsilon} \right) \right] \quad (2.3)$$

One important enhancement made to the SPAM code was the addition of a more realistic potential. This potential includes contributions from a halo, disk and bulge. It is referred to as the Three Component Potential. These components are the same three used in Hernquist (1993) and the expressions below follow from that work as well as the MASS99 implementation described in Antunes and Wallin (2007).

For the Halo

$$q_{halo} = \frac{\gamma_{halo}}{r_c} \quad (2.4)$$

$$\alpha_{halo} = \frac{1}{1 - q_{halo} \sqrt{\pi} \exp[q_{halo}^2] (1 - \text{erf}(q_{halo}))} \quad (2.5)$$

$$M(r)_{halo} = \frac{4\pi\alpha_{halo}}{r_{halo}^2} \frac{1}{2\sqrt{\pi^3}} \int_0^r \frac{\exp(-x^2)}{x^2 + q_{halo}^2} x^2 dx \quad (2.6)$$

For the bulge

$$M(r)_{bulge} = 4\pi\rho_{local} \int_0^r \exp\left[\frac{-x^2}{h_{bulge}^2}\right] x^2 dx \quad (2.7)$$

And for the disk, make the approximation that the mass is distributed with

$$M(r)_{disk} = \frac{1}{2h_{disk}^2} \int_0^r \exp\left[\frac{-x}{h_{disk}}\right] x^2 dx \quad (2.8)$$

Summing the mass of all the components, we can then calculate the acceleration

$$M_{total}(r) = M_{disk}(r) + M_{halo}(r) + M_{bulge}(r) \quad (2.9)$$

$$\vec{a} = \frac{GM_{total}(r)}{r^2} \hat{r} \quad (2.10)$$

Within the plane of the disk, this approximation performs well. However, for particles above or below the plane, the modeled force is not as accurate.

The three component potential is used to initialize an n-body simulation by distributing a large number of particles, at least 10 000, according to the analytic potential described above. The resulting potential is then approximated by sampling the velocity dispersion and radial force as function of radial distance. These sampled values are stored in large arrays. Later, during subsequent force calculations, the radial distance between each particle and the respective centers-of-mass is used to lookup the force and velocity dispersion in this array. The values are then scaled by multiplying by the appropriate mass and used in the force calculations. It is this lookup process that gives the three component potential its

alternate name, of n-body interpolated potential (NBI).

### **Setting Initial Conditions**

The SPAM code was modified for use with Merger Zoo to accept the following inputs: x-, y-, z- components of position and velocity, mass of each galaxy, radius of each galaxy, and two orientation angles for each galaxy. The orientation angles are similar to an inclination with respect to the reference plane and a position angle. These 14 parameters are needed to model the relative orbit trajectory as well as the orientations of the galaxy disks with respect to the reference plane. These values are specified for the current epoch. To perform a simulation, the orbit must first be integrated in reverse to an earlier time, a process we call backward integration. This is performed by the PERTURBER\_POSITION routine.

Once the trajectory has been backward integrated, the disk of mass-less test particles for each galaxy is created. Particles are assigned to a radial distribution based on the user's choice of radial profile. Particles can be distributed with a probability proportional to  $1/r$ ,  $\exp(-r/r_0)$ , or  $\exp(-r^2/a + r \times b + c)$ . After determining a particle's radial position the circular velocity is calculated according the user-specified central potential. The particle's azimuthal position is then randomized. Finally, an optional random "heat" value can be applied to the particle's velocity. The user specifies the maximum magnitude of the velocity offset which is chosen from a uniform random distribution independently for each of the three components. This is handled by the PROFILE. As each particle's position and velocity are assigned, they are rotated with respect to the corresponding galaxies orientation angles.

### **Integration**

Updated velocities and positions are calculated using a fixed time step, fourth-order Runge-Kutta integrator. The force calculation for the selected potential is performed by the DIFFEQ routine. This routine operates on an array of all particles, the two centers of mass and the test particles, with each call.

### 2.4.2 MASS99

MASS99 is an “N-body plus SPH” code. It simulates older stars, gas, and the formation of new stars with SFR feedback mechanisms (Antunes and Wallin, 2007). It is implemented as an optimized tree code with multiple trees for gravity and gas. The software can be run in a cluster environment through the use of an MPI runtime. It has also been updated to support Open MP for running on single machines with a large number of cores. The software includes separate utilities to assist in initializing simulations, running them, and then projecting the output for convenient visualization.

The MASS99 code is a good candidate for being able to match restricted three-body simulations with our modified potential because it uses a similar potential to determine its own initial conditions. This should help to simplify the process of converting a best-fit restricted three-body model to a full n-body model, and then later, a full n-body model with gas dynamics.

## 2.5 Validity of Restricted Three-Body Approximation

The restricted three-body approach was used by Toomre and Toomre (1972) to recreate plausible orbit trajectories necessary to create the disturbed morphologies of several well know interacting galaxies. However, researchers soon moved to more sophisticated n-body codes to simulate galaxies. Does our advocacy of a return to three-body codes represent a step backwards? No. By updating the potential to something more sophisticated than a softened point mass and adding dynamical friction, we are providing a very useful approximation to n-body simulations. Again, in our multi-method approach, we are able to use this approximation to significantly reduce the total parameter space to be searched for best-fit simulation initial conditions and at greatly reduced computational cost. After using a restricted three-body approximation for the first level search, researchers should then proceed to optimize full n-body simulations to determine the best-fit models for galaxies.

In Figure 2.1 we present the rotational velocity curve for a model galaxy using our two

potentials. Within the disk radius, where  $r$  is less than 1, we see a flatter rotation curve for the three component potential. For comparison, we plot a theoretical velocity curve based on a NFW profile (Navarro et al., 1996). The two potentials that account for dark matter provide similar rotation curves. This is expected because our three component potential is derived from Hernquist (1993). All three potentials converge asymptotically.

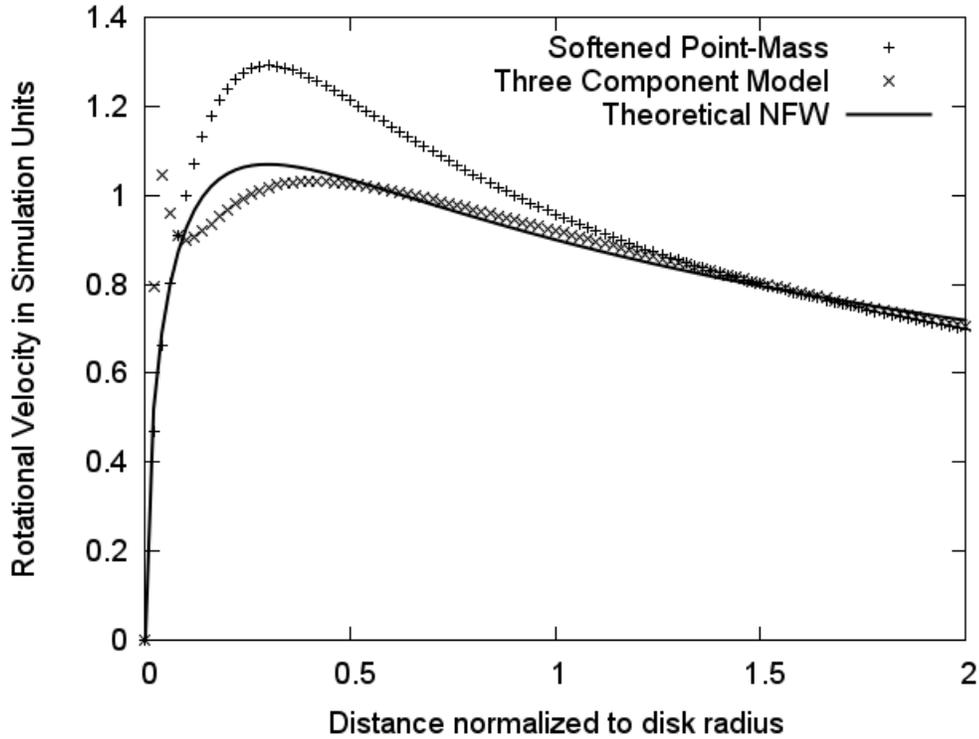


Figure 2.1: Rotational velocity curves for restricted three-body potentials.

The other approximation that needs to be validated is the dynamical friction treatment. At the present state, the strength of the dynamical friction in our restricted three-body code is a configurable parameter. During the Merger Zoo project, we left that parameter set to a relatively low value, meaning that most of our simulations had little to no dynamical friction. When attempting to build a full  $n$ -body simulation that matches the restricted three-body simulation, it is important to have an appropriate amount of dynamical friction.

Otherwise, the full n-body simulation will lose orbital angular momentum and the secondary galaxy will follow a different trajectory from the approximation. This in turn is likely to yield substantially different morphology than the restricted three-body approximation.

In Figure 2.2 we show the trajectory of the secondary galaxy from a MASS99 n-body run with a decaying orbit. Each of the subpanels from top to bottom include a SPAM trajectory for the same initial conditions, but with increasing dynamical friction. The top panel shows almost no orbit decay for SPAM, clearly a mismatch with the MASS99 trajectory. The middle parameter shows some significant decay, but not at the same rate. The bottom panel shows a qualitatively similar decay pattern. The analytic treatment of dynamical friction in SPAM is effective at shedding orbital angular momentum. Unfortunately, it remains a parameter that requires manual tuning. This dissertation does not present a full solution for converting restricted three-body models to full n-body ones due to the difficulty with automatically aligning the trajectories. That work has now become the focus of another research project.

Figure 2.2 demonstrates the ability to match trajectories between MASS99 and SPAM by tuning the dynamical friction term. The next step is to demonstrate similar morphology between the restricted three-body and the full n-body simulations. For a particular set of SPAM parameters a simulation was run using 2000 particles and a dynamical friction coefficient of 0.01. The backwards propagated SPAM position was used as the starting position for a MASS99 simulation with over 4800 particles. Figure 2.3 shows the results of the two simulations. The first row shows the particle positions and trajectory of the secondary galaxy for each simulation. The second row compares the two trajectories and the two sets of particle positions. We see that the trajectories are a good match for most of the simulation with some divergence apparent near the time of closest approach, when dynamical friction is strongest. The particle positions are a qualitative match. The shape and size of the tail on the primary galaxies are a good match. For the secondary galaxy, MASS99 produces a tail that is slightly more tightly curved than SPAM does. This demonstrates the feasibility of converting SPAM parameters to full n-body, MASS99 simulations.

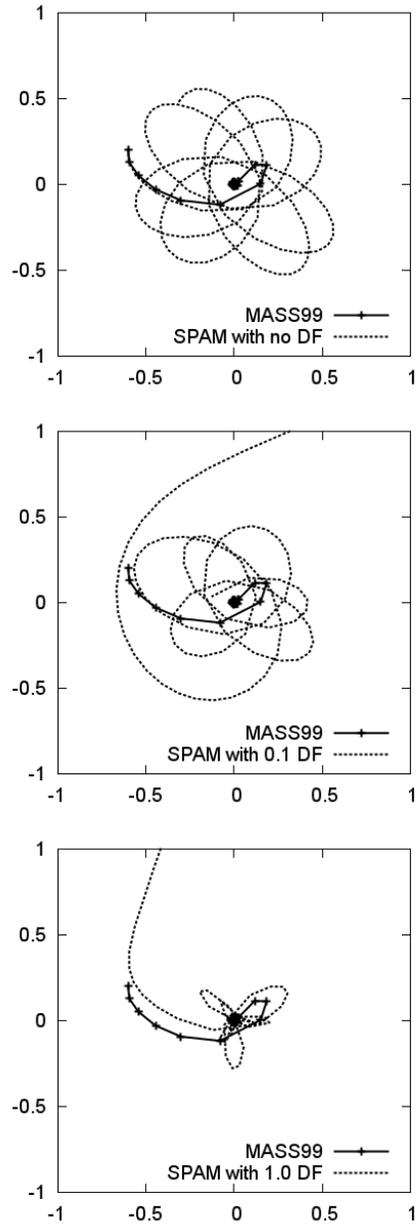


Figure 2.2: MASS99 and SPAM trajectories with varying dynamical friction.

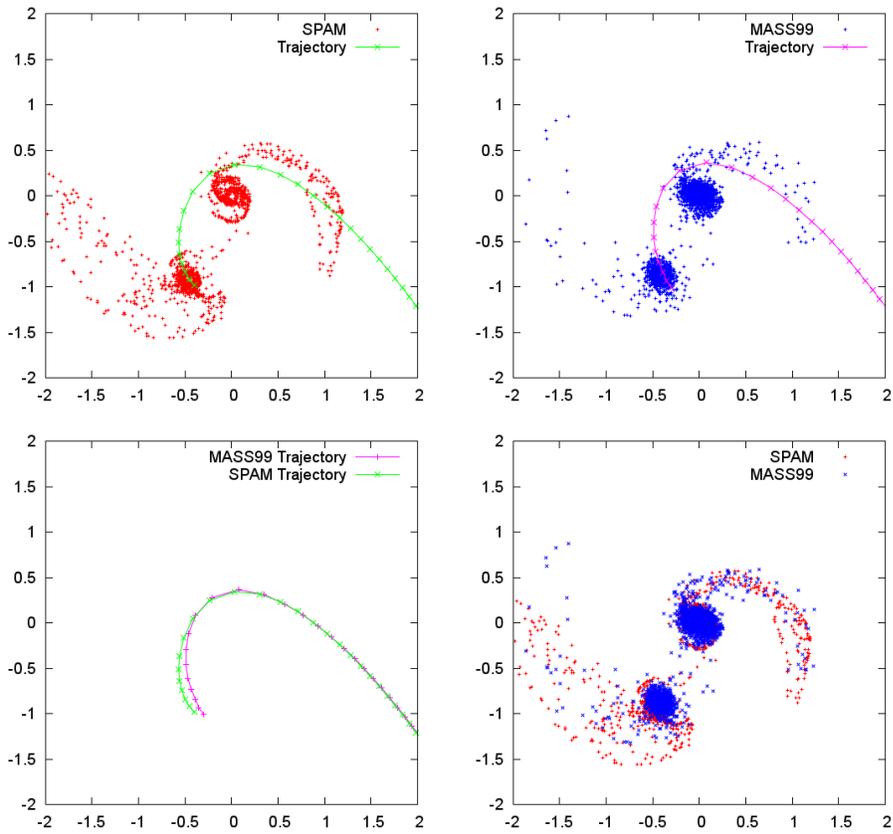


Figure 2.3: The particle positions and trajectory are plotted for SPAM on the top left and MASS99 on the top right. The two trajectories are compared on the bottom left. The particle positions are compared on the bottom right.

## Chapter 3: The Population of Interacting Galaxies

During the 20th century astronomers cataloged tens of thousands of galaxies and classified them according to Hubble's morphologically-based system. They were able to identify many objects from previously compiled lists of nebulae, such as the New General Catalog (NGC), as being galaxies outside of the Milky Way. As astronomers revisited the thousands of NGC objects, and discovered many thousands of new galaxies, they continued to assign Hubble classifications to them. Over time, the asymmetric, irregular galaxies stood out from the more common elliptical and disk galaxies. As early as the first few decades of the 20th century it was noticed that many of these irregular galaxies had nearby companions. They were often described as double and triple galaxies. Due to the relative proximity of irregular galaxies to one another, it was thought that perhaps they were able to influence, or interact, with each other. Struck (2006a) names Holmberg and Zwicky as early investigators into the possibility that galaxy collisions could play a role in the creation of irregular galaxies. Holmberg published a catalog of multiple galaxies in 1937, Zwicky published another one in 1959. Also in 1959, the National Geographic Society Palomar Observatory Sky Survey (POSS) was published. This large set of photometric plates provide observations of the entire northern sky with significant coverage to declinations to  $-30^\circ$ . By reviewing the POSS, Vorontsov-Velyaminov produced a catalog of 355 objects that he called interacting galaxies. In 1966, Arp produces his Atlas of Peculiar Galaxies with 338 objects. Both Vorontsov-Velyaminov and Arp included their own classification schemes to attempt to categorize the interacting galaxies. Both authors also made use of a sky survey to provide the initial data for selecting their objects of interest.

In the later years of the 20th century and the beginning of the 21st century, digital versions of photographic plates and later CCD images were made available on the Internet. One of the first and most important combined the northern sky data from POSS with

southern sky data from the UK Schmidt telescope. It is known as the Digitized Sky Survey and became available in 1994. Other projects to observe the entire sky or as much of it as possible from a given observatory have been conducted as well. These include ground-based efforts such as the Sloan Digital Sky Survey and orbital observatories like the Wide-field Infrared Survey Explorer. Most surveys after POSS also were conducted with automated image processing pipelines that identified photometric sources in the images and applied algorithms to measure their properties. These properties are usually made available in databases. This gives researchers the opportunity to study catalogs of millions of objects by querying for automatically measured values and to also access science images for direct photometric measurements as well.

## **3.1 Sky Surveys**

Several sky surveys were used in this research. Some of them included mechanisms for querying for images, both thumbnail and FITS files. These capabilities were often consumed by writing software to submit HTTP requests to specific URLs to automatically download images. The sky surveys also usually include a database of catalog properties of objects. Similar software was developed to query these databases.

### **3.1.1 Digitized Sky Survey**

The physical plates produced as part of the POSS were very useful to astronomers producing the first large-scale catalogs of interacting galaxies. It took until 1994 for these images to become widely available in digital form on the internet. In that year, the Catalogs and Survey Branch (CASB) of the Space Telescope Science Institute (STScI) completed the first generation of the Digitized Sky Survey (DSS). This survey included the data from the Palomar observatory and was combined with observations of the southern sky taken with the UK Schmidt telescope at the Anglo-Australian Observatory. A second generation of the DSS was made available in 2006. It consists of images captured in followup surveys at both Palomar Observatory and the Anglo-Australian Observatory. The first DSS had

plates for both blue and red filters. The second DSS has both of those filters as well as an IR filter. The angular resolution of the images is between 1" and 1.7". The images may be accessed via a simple HTML form hosted by STScI<sup>1</sup>. The POSS is extremely useful because it provides whole-sky coverage at sufficient resolution to study hundreds of nearby interacting galaxies.

### 3.1.2 Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS) has been in operation since 2000 and has observed over one quarter of the entire sky with deep exposures with five different filters. Additionally, the 2.5 meter telescope at Apache Point Observatory, New Mexico, includes an ability to use optical fibers fed through custom metal plates to feed spectrographs. By conducting multiple passes over the observation area, the telescope is able to acquire alternate photometric and spectroscopic observations. Not all objects are observed spectroscopically. SDSS also includes automated pipelines for processing the raw observations to identify likely sources and catalog their photometric and spectroscopic properties. The photometric properties for galaxies include their magnitudes in each of the five filters: u, g, r, i, z as well as radial brightness profiles. The pipelines attempt to fit two types of profiles, exponential and de Vaucouleurs, the first being a reasonable fit for disk galaxies and the second being used to model elliptical galaxies. The spectroscopic data includes a number of useful spectral lines that can aid researchers study metallicities of galaxies, but is most useful for providing highly accurate redshift measurements for galaxies. The SDSS is extremely useful because it provides photometric properties and redshift measurements for a large number of galaxies.

### 3.1.3 Wide-field Infrared Survey Explorer

The Wide-field Infrared Survey Explorer (WISE) mission was launched into low earth orbit in December 2009 to survey the entire sky at four different infrared wavelengths (Wright et al., 2010). The telescope has an angular resolution of  $\sim 6''$  for the first three bands know

---

<sup>1</sup>[http://archive.stsci.edu/cgi-bin/dss\\_form/](http://archive.stsci.edu/cgi-bin/dss_form/)

as W1, W2, and W3 with wavelengths of 3.4, 4.6 and 12  $\mu\text{m}$ . The fourth band, W4, with a wavelength of 22  $\mu\text{m}$  and has an angular resolution of  $\sim 12''$ . WISE was able to complete one complete survey of the sky plus an additional 20 percent coverage before the loss of cryogenic coolant began to affect operations. A further 30 percent coverage of the sky in bands W1, W2, and W3 was performed before coolant loss prevented further operations. The loss of cryogenic coolant is a normal part of infrared observing missions in orbit. The infrared observations in WISE are particularly useful as proxies for estimating stellar mass and star formation rate in galaxies.

## 3.2 Catalogs of Interacting Galaxies

Irregular, or peculiar, galaxies have very interesting shapes. It is easy to see that their disturbed morphologies attract attention from many different researchers. The introduction to this chapter gave a brief overview of some of the catalogs of interacting galaxies that were created in the last century. Below are details for the catalogs used in this research.

### 3.2.1 Arp

Halton Arp published his Atlas of Peculiar Galaxies in 1966 (Arp, 1966). He selected the set of 338 galaxies by reviewing the lists constructed by Zwicky, Vorontsov-Velyaminov, and others. The Atlas contained a combination of reproductions of POSS plates and blue-band imagery that he collected. The classification scheme proposed by Arp had several high-level classes of peculiar galaxies. They were spiral galaxies, spirals with companions on arms, elliptical galaxies, galaxies, and double galaxies. The scheme itself may not be an ideal one. For example, one sub-category for spirals with companions is “elliptical companions” and a sub-category for elliptical galaxies is “close to and perturbing spirals”. This makes it difficult for independent researchers to uniquely classify additional peculiar galaxies according to this scheme. The catalog does present a set of nearby galaxies that are well resolved by SDSS, HST, and WISE. The telescopes from those surveys can produce high resolution imagery useful for identifying important morphological details like tidal tails

and bridges.

### 3.2.2 Galaxy Zoo - Mergers Catalog

SDSS Data Release 6 (DR6) provided a catalog of photometric properties for approximately 100 million galaxies. Around 1 million of them also had spectral data. This massive data set far exceeds the capacity for a single researcher, or even a small team, to visually inspect each galaxy. The Galaxy Zoo project was launched in July 2007. It provided a simple to use HTML interface that would present volunteers with an image of a galaxy and then a set of buttons to click for the classification. These volunteers, acting as Citizen Scientists, were tasked with identifying whether they were looking at an elliptical galaxy, a spiral galaxy rotating clockwise, a spiral galaxy rotating anti-clockwise, an edge on disk, a star, and image artifact, or a potential galaxy merger. The site was stood up with the expectation that it would take years to classify the 1 million galaxies in the sample. By the end of the projects first year, over 150,000 people signed up to contribute a total of more than 50 million classifications. This means that each galaxy was voted on more than an average of 50 times. Lintott et al. (2008) presents the results of these classifications. That paper also describes previous efforts at producing catalogs of morphological classifications. Previous efforts ranged from a few hundred to over 10,000 with usually one or only a few people classifying each image. The Galaxy Zoo data set was several orders of magnitude larger in size, but also contained a distribution of opinions for each galaxy. This allows researchers to study how elliptical- or spiral-like a given galaxy is. The advent of convenient access to galaxy images by SDSS made this large scale effort possible, and the use of modern web browsers made contribution by Citizen Scientists convenient. Automated processes going beyond those manually constructed catalogs often use observed values such as colors as a proxy for morphology and are likely to have unknown biases. Galaxy Zoo helps to address these issues by providing the distribution of user selections for each galaxy. By producing classifications of such a large set of galaxies by direct visual inspection, the Galaxy Zoo catalog has become not only an extremely useful data set on its own but also a remarkable

training set for researchers wishing to develop their own automatic classification techniques. The details of the data release for the catalog are presented in Lintott et al. (2011).

One of the possible choices when classifying galaxies was a button labelled “Merger” showing an icon of a galaxy with a tidal tail. Out of the million galaxies classified by the Galaxy Zoo volunteers, thousands were identified by at least a portion of the users as being mergers. The catalog presented in Darg et al. (2009) contains over 3003 visually selected pairs of interacting galaxies at various stages of merging.

### **3.2.3 The overlap between Arp and SDSS**

There are 205 of the 338 Arp galaxies covered by SDSS DR7. The Darg catalog contains 32 Arp galaxies. The discrepancy is due in part to the fact that the Galaxy Zoo sample of approximately 1 million galaxies included only those galaxies with spectroscopic data as well. So a large number of Arp galaxies, though observed by SDSS, do not have spectroscopic data within the survey.

### **3.2.4 Hubble Space Telescope - ‘Galaxies gone Wild!’**

In 2008, near the time of the 18th anniversary of the launch of the Hubble Space Telescope (HST) aboard the Space Shuttle Discovery, several versions of a press release were released by various NASA and ESA organizations to mark the occasion. One version<sup>2</sup> was titled ‘Galaxies gone Wild!’ and featured 59 images of interacting galaxies that were observed by HST. Two years later, during the celebration of the 20th anniversary of the launch of HST, the Galaxy Zoo project was promoting a version of their website that was used to classify HST imagery of galaxies. In conjunction with this celebration, several upgrades to the Merger Zoo website and target preparation software were made so that galaxy targets without any SDSS data could be studied. Of these 59 pairs of galaxies, nine of them were presented as targets on the Merger Zoo website. Eight of these pairs were further studied with the Merger Wars activity. The only target not processed by Merger Wars was Arp 65.

---

<sup>2</sup><http://www.spacetelescope.org/news/heic0810/>

This set of galaxies is referred to later as HST GGW or just the HST galaxies.

### 3.2.5 Other catalogs

As mentioned above, other catalogs of interacting galaxies such as Vorontsov-Velyaminov do exist. The Arp Atlas was sufficient for identifying the targets studied as part of the Merger Zoo. However, another online database containing information about galaxies, much more than just interacting ones, proved critical to this research. That database is the NASA Extragalactic Database (NED) hosted by the Infrared Processing and Analysis Center (IPAC) at CalTech. The information in this database is made available through a set of HTML forms. It was a simple matter to write software that could submit HTTP requests matching the form parameters sent through the web page and programmatically retrieve information about galaxies such as the name of an object at a given set of RA and Dec coordinates, as well as basic object information such as representative magnitude and redshift.

## 3.3 The Sample for this Study

The sample for this study consists of 54 pairs of interacting galaxies drawn from SDSS and eight pairs of interacting galaxies from HST GGW for a total of 62 interacting pairs. Figure 3.1 shows the distribution of redshift values for the Merger Zoo sample studied here compared to the full Arp catalog and the Galaxy Zoo sample of mergers studied in Darg et al. (2009). The pairs in our sample tend to have a significantly smaller redshift than those in the Darg catalog. There is obvious overlap with the the Arp catalog. However, as pictured in Figure 3.1, the Arp catalog has a median redshift lower still than our catalog. The redshifts for the Arp galaxies were acquired by querying the NED database with the names of each galaxy (Arp 1, Arp 2, . . . , Arp 338). The database returned no redshift measurement at all for 45 of the Arp galaxies. For 51 other galaxies in the Arp catalog, a query of the NED database returned a measured redshift of exactly 0. So even these somewhat “famous” galaxies do not have complete coverage in current astronomical databases.

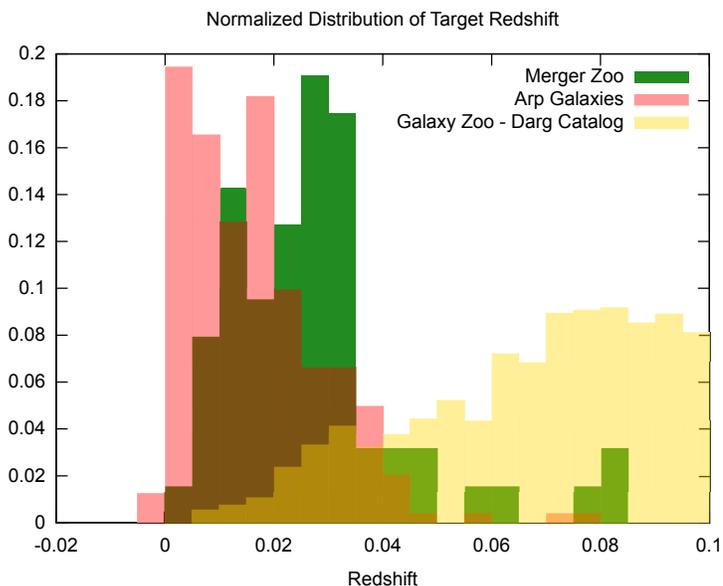


Figure 3.1: Normalized distribution of redshifts for three merger catalogs.

### 3.4 Photometric Observations

Galaxies are studied by inferring physical properties from the observed amount and distribution of light that they produce. SDSS is a very useful survey for measuring the light from galaxies because it has recording imagery of them through five different filters, u,g,r,i, and z spanning the visible spectrum and into the near infrared. The photometric pipeline used by SDSS automatically detects sources of light in each image, measures the amount of flux for each source in each filter, determines whether the source is likely a star or galaxy, and saves the information to a database. When the measurements are accurate, this provides a wealth of information for researchers. The five photometric magnitudes can be used to form an approximate spectral energy distribution (SED). This in turn can be used to fit a theoretical population of stars to estimate the stellar mass of the galaxy. SEDs can also be used, when combined with a large number of redshift measurements, to calculate a photometrically determined redshift for galaxies. For estimating the mass of galaxies to be used in our sample, we obtain the database values for the ugriz magnitudes and apply the relationships in Bell et al. (2003) to calculate multiple mass values. We used the minimum,

mean, and maximum values to provide the initial target and range of masses for our simulations. For HST targets our mass estimates were based on a simple 1 to 1 mass-to-light ratio. We determined the luminosity of the galaxy based on the magnitude value returned by NED and the redshift. We then divided by the solar luminosity to arrive at an estimate for number of solar masses.

### 3.4.1 Deblending

The algorithms used in SDSS through DR7 for automated processing had a preference for chopping up large, nearby galaxies (like our entire sample) into smaller pieces through a process known as deblending. This process has been improved for DR8 and later. For Arp 82, the clearly larger primary galaxy is estimated to have a mass of  $1.7 \times 10^9 M_{\odot}$  and the secondary a mass of  $1.0 \times 10^{10} M_{\odot}$ . The mass ratio is 1 to 5.9 for the primary to secondary. Using values from DR8, the mass of the primary becomes  $4.6 \times 10^{10} M_{\odot}$  and the secondary changes slightly to  $9.9 \times 10^9 M_{\odot}$ . A mass ratio of 4.6 to 1 is a realistic value for this system. The NED-based mass estimates are  $2.3 \times 10^{10} M_{\odot}$  and  $7.2 \times 10^9 M_{\odot}$  respectively, with a mass ratio of 3.2 to 1. Values in DR8 are much improved, but they were not available when we began this research and still suffer from the issue of deblending for some targets.

However, researchers studying SDSS galaxies have taken to redoing the photometric measurements for large galaxies. For our sample, we initially considered performing our own re-measurement of the galaxies in our sample. The first issue we encountered was that most existing tools for measuring galaxies did not have a built-in capability for setting apertures for irregular and overlapping galaxies. We solved with problem by building our own tool for measuring irregular galaxies. The details are in Appendix A. Rather than using traditional circular or elliptical apertures, we use a combination of image thresholding and contour finding algorithms to find the pixels containing flux from our galaxy. We perform this operation on images that have had the background subtracted by using a combination of source masking and two dimensional linear regression similar to West et al. (2010). One additional problem remained. For large sky surveys, the CCD imaging arrays tend to be

very big compared to other telescopes. This size allows the survey to image large fields of view. However, even with large frame sizes, some galaxies will fall on the border between frames. For a survey such as SDSS, which makes a point of repeated observations of the exact same frame, this means some of the flux is in one frame and some flux is in another. The two frames must be combined in order to measure the total integrated flux of a galaxy. The process of combining multiple frames is sometimes referred to as mosaicking. It can be accomplished in multiple ways, but each of them adds complexity to the process of making photometric observations.

### **3.4.2 WISE Photometry**

WISE atlas imagery was available for each of our target galaxies from both SDSS and HST without the need to perform mosaicking. The angular resolution for WISE is noticeably degraded compared to SDSS, 6" vs. 0.5", but for large galaxies such as ours the key morphological features are resolved. The additional advantage of using WISE imagery is that simple relationships exist to convert magnitude and color information into mass, star formation rate, and specific star formation rate. The photometric tool, originally intended for use with SDSS imagery, was able to work with work WISE image data. It provided the ability to measure magnitudes in all four IR bands measured by the survey, W1, W2, W3, and W4.

## **3.5 Matching SDSS Value-added Catalog with WISE**

The SDSS database is a valuable resource for astronomers. A number of researchers have sought to extend its utility by producing catalogs of derived values matched to SDSS objects. These lists are referred to in general as value-added catalogs. Here we refer to one in particular, Kauffmann et al. (2003), as the MPA/JHU value-added catalog, or simply the value-added catalog in the following sections. The research team who produced this dataset used the ugriz magnitudes, redshift, and two important stellar absorption line indices to estimate the stellar mass and star formation histories for over 100,000 in the SDSS DR1.

They have since updated their value-added catalog for subsequent data releases. Here we consider their data for SDSS DR7. Their process is a very useful one, but cannot be applied directly to our targets for a number of reasons. First, many of our targets were deblended in the SDSS catalog, so any derived values calculated by Kauffmann et al. (2003) will be inaccurate. Second, we cannot replicate their analysis for many of our targets because we may not have individual redshift measurements for both galaxies in a pair and we definitely do not have spectral line data for many of our galaxies. Additionally, the calculation of star formation rate depends on assuming a specific star formation history (or set of histories). Eventually, the star formation history is a property we expect to model with full n-body simulations with gas dynamics. Assuming it as part of the photometric process could be problematic. Finally, not all of our targets were observed by SDSS.

With these issues in mind, we decided to investigate the relationship between infrared magnitudes measured by WISE and stellar mass and SFR for galaxies. To accomplish this, we queried the WISE database for information on the approximately 700,000 galaxies that were in both the Galaxy Zoo dataset and the SDSS value-added catalog. This gave us the opportunity to compute absolute magnitudes for WISE pass bands and compare them with model values for mass and SFR from the value-added catalog. We discovered several interesting correlations. During the same time we were performing this analysis in the summer of 2012, another group was pursuing a similar, more in-depth study as well (Yan et al., 2013). That work also attempted to incorporate information about metallicities derived from spectral data. Many of our targets lacked spectral data, so we were unable to gain any direct value from replicating their method. We also learned of an as-yet, unpublished analysis from 2012 as well that took a similar approach (Shi et al., 2012).

Here we present the results of our three way catalog matching for SDSS, Galaxy Zoo, and WISE. Figure 3.2 shows the linear regression fit to mass as a function of W1. Figure 3.3 shows the linear regression fit to SFR as a function of W3.

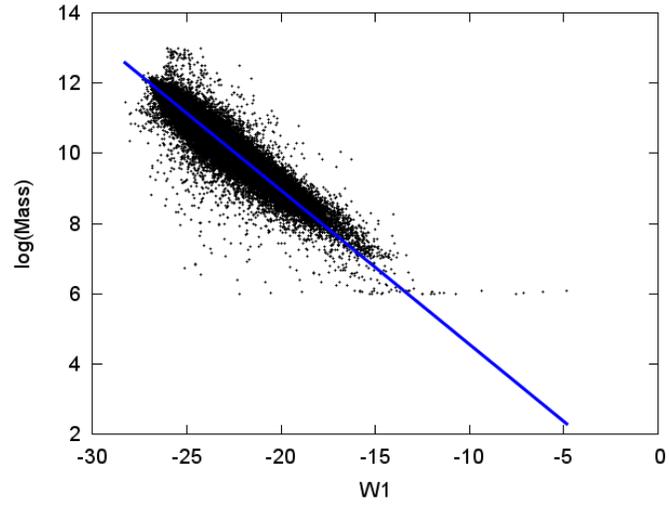


Figure 3.2: MPA/JHU Mass as a function of W1 absolute magnitude.

$$\log(Mass) = -0.439 \times W1 + 0.164 \quad (3.1)$$

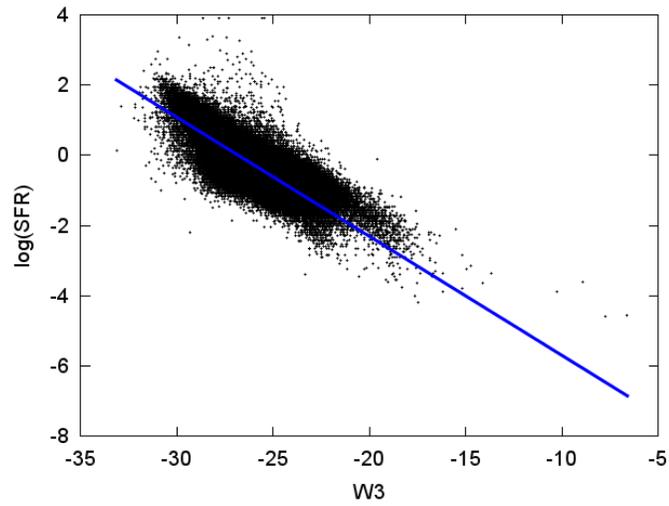


Figure 3.3: MPA/JHU SFR as a function of W3 absolute magnitude.

$$\log(SFR) = -0.339 \times W3 - 9.091 \quad (3.2)$$

We compare our sample to two morphological types from Galaxy Zoo, ellipticals and spirals. We have prepared a “clean” version of the Galaxy Zoo sample by only including those galaxies where the vote fraction was greater than or equal to 0.8. For example a spiral where only half the Citizen Scientists voted it as a spiral would be excluded. In Figure 3.4 we see the results of plotting the SFR as a function of mass shows that spirals and ellipticals from the Galaxy Zoo catalog still have some overlap but are reasonably separated. Plotting the Merger Zoo targets on top shows that they tend to look like the spirals, but again, with overlap in the ellipticals.

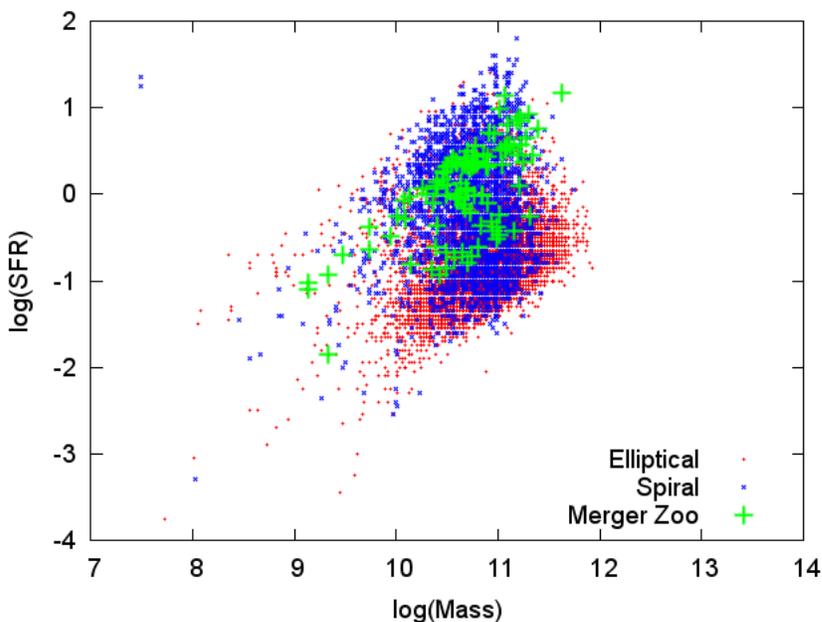


Figure 3.4: SFR as a function of Mass for different morphological types.

The  $W1 - W3$  color is a equivalent to SFR normalized by stellar mass. Plotting the three populations again shows the separation go the spirals in blue and ellipticals in red. Figure

3.5 shows that the Merger Zoo galaxies have a higher normalized SFR than ellipticals. Our sample tends to have a higher normalized SFR than most spirals as well.

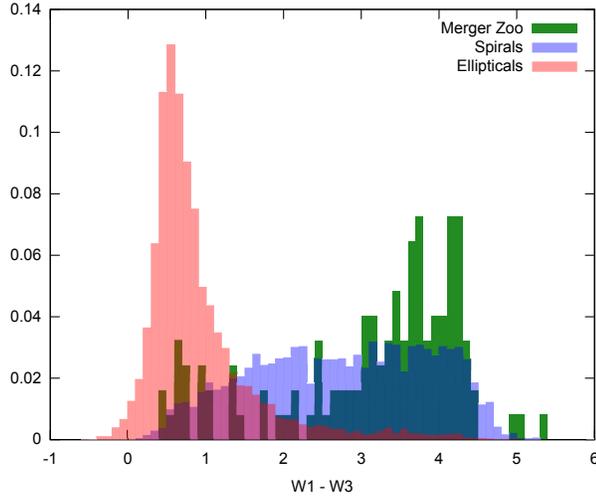


Figure 3.5: Relative frequency of W1 - W3 for various samples.

### 3.5.1 SDSS / WISE Regression Analysis

The relationships chosen above are the ones with the best Pearson’s correlation coefficient,  $\rho$ . Strong relationships have  $\rho$  values close to +1 or -1, indicating positive or negative correlations. Completely uncorrelated data has a  $\rho$  value of 0. For Mass, W1 was the best predictor, with W2 a reasonable second choice. For SFR, W3 was significantly better than others. For specific SFR, SFR per unit mass, W1-W3 would be a logical choice since it represents a ratio of mass and SFR. It was actually the second best choice, with W2-W3 being just slightly better choice with a lower RMS. Table 3.1 contains the RMS residuals and  $\rho$  values obtained by performing linear regressions on all WISE magnitudes and colors. The cells representing the best fits for each quantity have a darker background. The best RMS residual is 0.19 dex which is substantially larger than the uncertainty of the measured flux within a given aperture.

Table 3.1: RMS residuals and Pearson's  $\rho$  for linear regressions of mass, SFR, specific SFR, with all WISE magnitudes and colors

	Mass		SFR		Specific SFR	
	RMS (dex)	$\rho$	RMS (dex)	$\rho$	RMS (dex)	$\rho$
w1	0.19	-0.97	0.60	-0.33	0.71	0.50
w2	0.22	-0.94	0.59	-0.38	0.73	0.46
w3	0.54	-0.56	0.38	-0.80	0.81	-0.17
w4	0.47	-0.70	0.51	-0.59	0.82	0.10
w1-w2	0.64	0.19	0.51	0.59	0.78	0.31
w1-w3	0.58	-0.46	0.48	0.65	0.41	0.87
w1-w4	0.57	-0.50	0.58	0.39	0.59	0.70
w2-w3	0.56	-0.51	0.51	0.60	0.40	0.87
w2-w4	0.54	-0.57	0.60	0.31	0.59	0.69
w3-w4	0.65	0.14	0.53	-0.55	0.69	-0.53

## Chapter 4: Identifying Morphological Matches

Since the middle of the twentieth century catalogs with hundreds of interacting galaxies have existed. Starting with the publication of Toomre and Toomre (1972), researchers have had a tool to use to attempt to model the specific relative orbit for two interacting galaxies. That seminal paper itself contained plausible models based on the authors' ability to recreate the morphologies for four systems: Arp 295, M51, NGC 4676 and NGC 4038/9. However, in the ensuing decades, no single paper has attempted to provide models for a large number of systems. Researchers have tended to focus on small numbers of galaxies at a time, attempting to produce more detailed models for well studied systems, such as M51. No attempt to model many dozens of actual systems by matching their morphologies appears to have been made.

This chapter presents the pipeline built to generate plausible models for a large number of systems. For a given pair of interacting galaxies, the input to the pipeline is primarily an image of the two galaxies showing any tidal distortions. Next, an estimate of the distance to the pair and an estimate of the mass of each galaxy are required. With this minimal information some constraints on the initial conditions can be applied. This allows us to select sets of initial conditions and run simulations. The output of the simulations are then compared to the image of the system. The quality, or fitness, of the match is evaluated and then the best simulation is selected.

### 4.1 Target Preparation

The target preparation process for Merger Zoo began by selecting an interesting pair of interacting galaxies. The criteria used were that the galaxies had to have obvious tidal distortions and that some idea of what the progenitor disks were had to be discernible in the

image. Once a pair was selected and the approximate sky coordinates determined from the Arp catalog or a NED query, a color thumbnail was downloaded from the SDSS server. The thumbnail was then converted to a gray scale. A simple threshold was applied by manually raising the brightness “floor” of the image until most of the image was replaced with black pixels, leaving just the galaxy pair. This gray scale thumbnail was made available, along with the color one, for users in the Merger Zoo project to compare with their simulations. The center of the image was selected to be the center of the galaxy identified as primary. The selection of primary was made based on which galaxy appeared to be larger.

Combining the sky coordinates of the center and the angular size of the image, it is simple to use bilinear interpolation to identify approximate RA and dec values for other points in the image, such as the center of the secondary galaxy. Initially, automated routines were developed to identify contiguous groups of pixels in the thresholded image. The largest group would correspond to the primary and the next largest group would then correspond to the secondary galaxy. Next, the pixel groups would be fitted with a minimum bounding box. From this box a rotated and inclined ellipse could be fit to each pixel group. The center of each galaxy could be estimated by a selection of either the brightest pixel in a group, or the center of the fitted ellipse. Because the initial process lacked the ability to explicitly mask stars in the image, occasionally a group of pixels associated with a star was the first or second largest group. A change to a manual process was made where the researcher preparing the target would simply click on the image at the location of the center of each galaxy. This allowed for consistent identification of the primary and secondary galaxies. The researcher also had the ability to adjust the size, shape, and orientation of each of the fitted ellipses.

The simulation parameters used in this study included the three dimensional position and velocity vectors needed to describe the relative orbit of the two galaxies. Two of those parameters, the x and y separation distance of the two galaxies are determined by locating the galaxy centers in the image. The next parameters to describe the simulation include two orientation angles and size of each disk. One angle is the inclination with respect to

the sky. The other angle is the position angle of the galaxy disk. These two angles, along with the size of the disk are estimated from the image by the automated process described above. The remaining two parameters to be estimated are the masses of the two galaxies. For SDSS targets, the CAS server was queried for the galaxy target nearest to the center of each image. This query usually returned two types of information. The first was the redshift to the galaxy, which allows us to set a specific scale for separation distance and disk size. The other type of information is photometric values, u, g, r, i, z. Following citemasspaper, it is possible to estimate the mass of the galaxies using the relative color values contained in the five bands. For HST targets, NED was queried for redshift and a photometric magnitude, usually a B magnitude. The mass was estimated by converting the magnitude to a luminosity, using the redshift information to first estimate distance. The luminosity was then converted to a mass with a simply mass-to-light ratio of one. Of the six orbit parameters, six disk parameters, and two masses, the image and database information provided reasonable estimates for ten out of fourteen.

Initial attempts were made to use the difference between the redshifts of the two galaxies to constrain the line-of-sight velocity component, the z direction. However, not all pairs had observed redshifts for both galaxies

The next stage in target preparation is to determine the appropriate range that the simulation parameters are allowed to vary. The x and y components of the relative position vector are held fixed. However, the disk orientations, masses, and all three velocity components are allowed to vary over a range of values.

- The masses are each allowed to vary over two orders of magnitude from  $0.1 \times \text{mass}$  to  $10 \times \text{mass}$  as determined above.
- The x and y velocities are allowed to vary between  $\pm$  the escape velocity as computed using the sum of the two maximum mass values determined in the previous step for a particle located at the current x and y separation of the two galaxies.
- The z velocity was originally allowed to range between 0 and the line of sight velocity

determined from the redshifts. However, not all galaxies had redshift values so the z velocity varied the same as the x and y components.

- The z position is allowed to vary between  $\pm 5 \times$  the diameter of the disk of the primary galaxy.
- The position angles are allowed to vary  $\pm 20$  degrees.
- The inclination angle is used to describe the rotation direction, so with a four-fold degeneracy, the inclination angle is allowed to vary  $\pm 20$  degrees.
- The disk radius for each galaxy is allowed to vary from 0.5 to 1.5 the value estimated from the image.

After the initial range of parameters are selected, we review the simulation results of several hundred randomly selected input parameters. Each parameter within a randomly generated set is selected at random from a uniform distribution scaled to match the minimum and maximum values of that parameter. If during this review phase fewer than ten results with at least some matching morphology are identified, the parameter ranges are adjusted manually. Usually these adjustments are to restrict the ranges. Once we are able to find at least 10 useful candidate simulation matches within a set of 100 to 200 sets of randomly generated simulation parameters, the target is considered ready to be presented to the Citizen Scientists.

In this manner, the fourteen simulation parameters were assigned allowed maximum and minimum values. These ranges were stored in a simulation parameter file. This file was bundled into a ZIP archive with the color and gray scale thumbnail images and uploaded to an Amazon S3 storage bucket where they could be accessed by the Merger Zoo website. With the specified ranges, simulation parameters can be selected by drawing a random number from a uniform distribution between 0 and 1 and scaling by the min and max values.

### 4.1.1 Simulation Filter

An additional filter is imposed on the randomly selected sets of parameters. During initial testing that included half a dozen different systems, the above ranges of parameter values allowed for a large number of randomly generated simulation inputs that resulted in simulations that showed no tidal features. In order to estimate whether a given simulation will result in tidal features we calculated a form of the tidal approximation parameter from Binney and Tremaine (2008). They provide an estimate for the change in velocity for a particle at a location in the primary galaxy due to tidal forces resulting from the passage of the secondary galaxy.

$$\Delta v \approx \frac{2GM_2}{b^2V_2}(x, y, 0) \quad (4.1)$$

We adopted a similar form for the coefficient which we called  $\beta$ , and in our canonical units it looks like

$$\beta = \frac{M_1 + M_2}{r_{min}^2 V_{r_{min}}} \quad (4.2)$$

Before running a full simulation with all of the test particles, we perform a backwards integration of just the two galaxy centers of mass. During the backwards integration we determine the closest approach distance,  $r_{min}$ , and the relative velocity at the time of closest approach,  $V_{r_{min}}$ . The  $\beta$  parameter captures two important quantities, the first is the mutual gravitational attraction. This is important because we wish to observe tidal distortions to both the primary and secondary in some systems. The second key component is the inverse velocity at the time of closest approach. This incorporates the sense of interaction time, during which one galaxy can impose a force or impart an impulse. Even though the units are not quite the same as the tidal approximation parameter we believe it contains sufficient information to predict whether or not there will be noticeable tidal distortions. The  $\beta$  parameter increases with increasing mass, it decreases with increasing distance, and

decreases with increasing velocity. More massive systems have a chance to cause greater tidal distortions. Systems that pass farther apart from one another will have less distortion. Systems that are only close for a short time due to a high relative velocity will not show as much distortion.

For our threshold, we set an arbitrary minimum value of  $\beta = 0.5$ . Any set of simulation parameters with a  $\beta$  greater than 0.5 was considered to have a significant chance of displaying tidal distortion. For systems with a  $\beta$  value less than the specified minimum, we accepted them with an exponentially decreasing probability.

$$p = \exp(-0.05 \frac{\beta}{\beta_{min}}) \quad (4.3)$$

We would then draw a number from a uniform random distribution between 0 and 1. If the number exceeded the probability then the parameter set was kept. If not, a new randomly generated set of parameters was tested. A parameter set with  $\beta$  close to but less than 0.5 will be accepted with probability 0.05. At  $\beta = 0.01$  the probability falls to 0.01 and at  $\beta = 0.001$  the probability is only 0.0001. This decreasing probability allows us to sample parameter space with sets of parameters that do not exceed our minimum  $\beta$  while at the same time we avoid having to review a large number of simulations that will likely not show any tidal distortions. With these thresholds there are an average of 12 parameter sets rejected for every one that passes. Of the parameter sets where  $\beta$  was below the threshold, on average 7% are accepted after passing the probability filter.

## 4.2 Automated Methods

Before the Merger Zoo was developed, we attempted to build an automated system to run simulations and compare them to the target images. The optimization problem we were attempting to solve was to identify the best set of values for the remaining twelve simulation parameters. An exhaustive search of this parameter space would be difficult. Consider testing just ten different values of each of the twelve parameters. You would still

have to run ten to the twelfth, or one trillion, simulations. At one second per simulation, this would take over 30,000 years.

The approach we took initially was to use Genetic Algorithms similar to the work of Theis and also Wahde. We used a GA code known as PIKAIA. We encoded each of the twelve parameters as genes on the chromosome for each simulation. Chromosomes were generated by randomly selecting gene values from the allowed range. Populations as small as 50 and as large as 5000 were generated. The populations were allowed to evolve through the use of cross-over and mutation in each generation. In order for the GA to proceed, we needed an automated method to evaluate the fitness of each simulation. Following a similar approach to other researchers, we applied a low resolution grid to the simulation output and the image. The algorithm computes the percentage of light, or simulation particles, in each grid cell. A distance between the simulation and the image can be computed by calculating the difference between each grid cell and summing over all cells. Unfortunately, this approach led to many GA-derived solutions that were only locally, and not globally, optimal. This is because the algorithm does not properly differentiate between mediocre matches correctly (see Figure 1.1). After running several thousand different initial populations for several thousand generations each, we were unable to find any plausible matches for the handful of systems we attempted to model this way. The automated fitness function being used by other researchers to select between simulations that were already close to a match, was not applicable for situations where large portions of parameter space needed to be searched.

### **4.3 Citizen Scientists**

Humans are able to rapidly review simulation output and compare to target imagery. The slight rotation example above is a good example of how we perceive a close match. We built the Explore interface to allow for the review of thousands of simulations by an individual researcher. The screen presents the results of restricted three-body simulations run in real time to the user. There are eight simulation outputs surrounding the target image in the

center. The user can click on a simulation image to indicate that they believe it is a possible match to the target in the center. If not a match, then clicking on the simulation image can be used to indicate that it at least shares one or more important tidal features with the target. After reviewing the current set of eight images, the user clicks “More” to see eight more images. In this fashion, a single user can review a thousand simulations an hour.

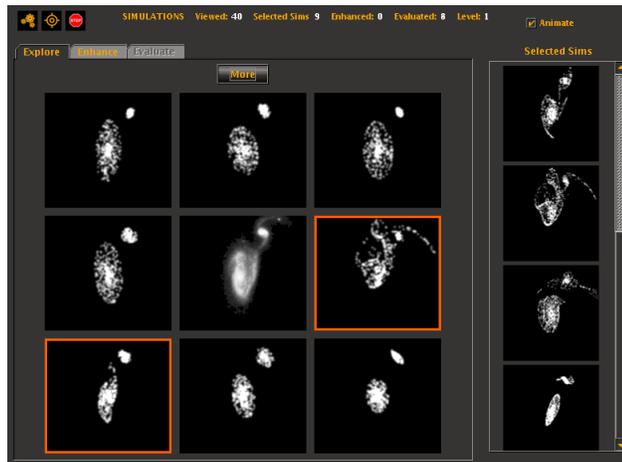


Figure 4.1: Explore.

Once the interface was constructed we realized it was an ideal way to solicit input from Citizen Scientist volunteers. Using our partnership with the Galaxy Zoo project, we launched the Merger Zoo. In addition to the Explore interface, we also developed the Enhance interface. This screen allowed the user to adjust each of the twelve simulation parameters to attempt to improve how well the simulation matched the target image. Whenever the user had determined that they had found a somewhat better match, they could save the simulation to their selected set.

After each set of eight simulations were added to the users selected set, we changed their screen to the Evaluate activity. Here they were asked to select, in order, the three

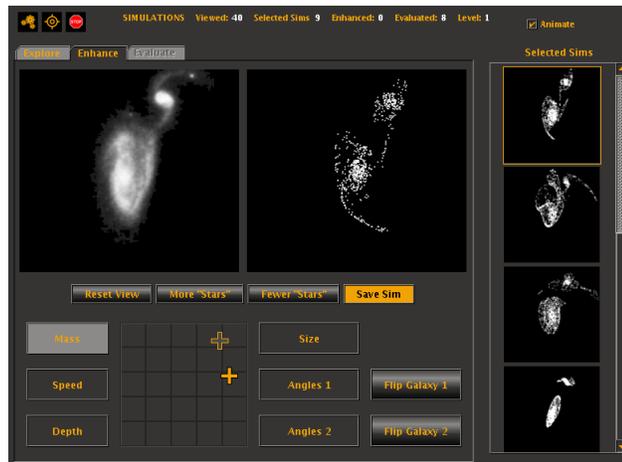


Figure 4.2: Enhance.

best images from the set of eight simulations that they had already selected. After eight first round images were selected, the user was asked to select the top three in a second level “tournament” of sorts. Dedicated users worked their way up through four or more levels of the tournament.

User selections were saved in session files by the applet which uploaded them to Amazon S3 bucket. This was the Merger Zoo interface that was implemented for the beta phase of the website and presented in Holincheck et al. (2010a). The number of sessions were less than a hundred or so for the beta. The research team downloaded the session files and manually reviewed the selected simulations to identify good matches. An attempt was made to use clustering and kernel density estimations to identify useful sets of simulation parameters from all sets that advanced to at least level three in the tournament. Wallin et al. (2010) presented a model for Arp 86 that was found by Merger Zoo beta users.

Soon after the full Merger Zoo website was launched, Citizen Scientists were reviewing simulations at the rate of several hundred sessions a day, where each session could include 100 to 1000 simulations. After a few thousand user sessions were completed, the files were downloaded and examined. The expectation was that the Evaluate activity would place the best simulation from each session at the top of the list. When the research team reviewed the sessions however, they noticed two interesting facts. First was that the users were on

track to review hundreds of thousands of simulations for each galaxy pair and were selecting tens of thousands of simulations. The second was that the Evaluate activity did not leave the simulations sorted in any kind of useful order. The initial attempts to cluster simulation sets on this larger volume of data did not yield useful results. Another method for evaluating the selected simulations was needed. With so many thousands of simulations to review for each galaxy pair, would it be possible to have the Citizen Scientists help us review them? The answer was yes, and the method for accomplishing this is called Merger Wars.

### 4.3.1 Merger Wars

The Merger Wars algorithm is used to sort the large lists of simulation results. For each target image, users would be presented simulation images two at a time. The user simply clicks on the simulation image that is a better match to the target image. When a simulation image is presented, that counts as participating in a competition. When a simulation image is selected, that counts as a win in the competition. The overall score, or fitness, for a simulation image is the simple ratio of number of wins to number of competitions. The images with a higher winning percentage are considered better matches than those with a lower winning percentage.



Figure 4.3: Merger Wars.

The idea of ranking by pairwise comparisons is not a new one. The study of similar methods dates back to at least the 18th century and they are used today to rank chess players, online XBox gamers, and cat pictures. However, the Merger Wars algorithm includes two novel enhancements. The first is the inclusion of a third choice labelled “Neither is a good match”. In situations where users feel that the simulation images are both rather poor, then can click the “Neither” button to record a loss for both images. The other enhancement has to do with image selection and shallow tournaments. Rather than simply compete all the images in a single large tournament in a winner take all style, or simply apply user selections as the comparison function in a traditional sorting algorithm, the method makes use of shallow tournaments. In the current implementation, images are competed against each other in randomly selected sets of eight. In a given tournament, the image can lose in the first round, second round or third round and accumulate zero to three wins. These shallow tournaments soften the impact of an incorrectly judged competition. For example, if a user clicks on the wrong image, a good image is scored down and a bad image is scored up. However, both images will be competed again in other tournaments. There is no single elimination in the larger process. Simulation images are selected for inclusion in a tournament in such a way as to keep the total number of competitions for all images close to equal.

In Appendix B we demonstrate that the Merger Wars algorithm has comparable performance to traditional sorting algorithms,  $O(n \log n)$ , and that in the presence of inaccurate comparisons, where users click the wrong image, the algorithm is more accurate than traditional sorting algorithms. The Merger Wars method was implemented as a JavaScript interface in the Merger Zoo website. This allowed Citizen Scientists that did have the Java plugin installed and enabled, and who then were not able to run simulations, to contribute to the project. Additionally, it gave volunteers a chance to see the types of simulations being selected by others.

The Explore and Enhance features in the Merger Zoo applet allowed Citizen Scientists to review over 3 million simulations and select around 66,000 for followup. The Merger Wars

algorithm provided an accurate method to sort these results with over 1 million competitions and identify the top simulations for each pair. Figure 4.4 gives a schematic view of the Merger Zoo application supported by the research team and the Citizen Scientists. This is a sub view of the larger pipeline in Figure 1.2.

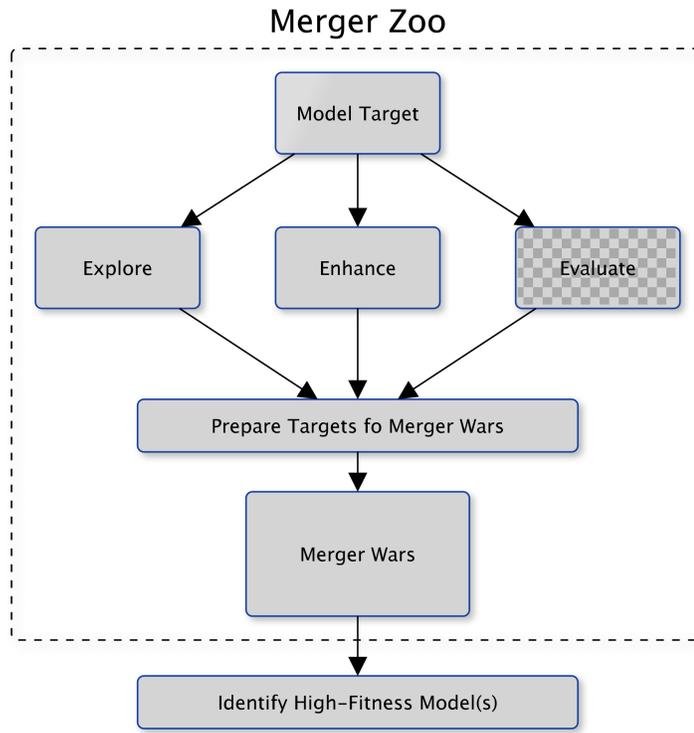


Figure 4.4: Merger Zoo Process.

### 4.3.2 Finale Activities

In the last few months of the Merger Zoo project, a set of finale activities were launched. For each of the 54 SDSS targets, the top 20 or so Merger Wars results were reviewed by the research team. They selected 4 to 8 simulations per target to represent the best simulations. Two activities were then launched to further rate these top simulations with the goal of selecting a single overall best simulation for each pair. The first activity was

called “Simulation Showdown”. Here, the user was presented with two sets of images. The first set included the target image and simulation image from one galaxy pair, the other set included the two images for another pair. The user needed to identify which image was a better match to its respective target. This compared simulations from different pairs against each other. The second finale activity was called “Best of the Best”. It presented the target image in the center. The best images were distributed at random around the target image. The volunteer was asked to select the best image for each target. Each activity generated a new fitness score. The top simulations for each pair were ranked by these scores. For about half of the systems the top-ranked simulation for “Simulation Showdown” and “Best of the Best” were the same. For the other half, the research team picked a consensus best simulation from the two candidates for each target.

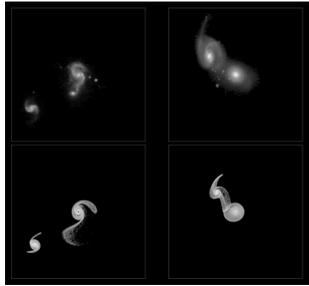


Figure 4.5: Simulation Showdown.

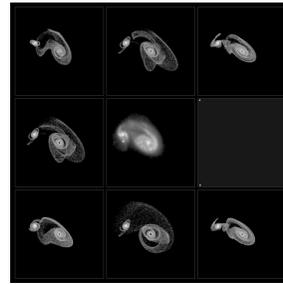


Figure 4.6: Best of the Best.

### 4.3.3 Merger Zoo Contributions

The Galaxy Zoo Mergers project, or Merger Zoo, was launched on November 23, 2009. The last simulation submissions and Merger War clicks were collected on June 7, 2012. The site remains available on the Internet with a notice that data collection is no longer occurring. In the two and a half year period that the site was active, 6081 Citizen Scientists with the Zooniverse logged in<sup>1</sup> and ran a combined  $3.31 \times 10^6$  restricted three-body simulations in 4765 hours of session time. The volunteers also judged  $10^6$  Merger Wars competitions.

<sup>1</sup>A total of 30305 registered Zooniverse volunteers visited the site, but only 6081 completed the tutorial and saved results.

In addition to the more than 3 million simulations viewed by the volunteers, the  $\beta$  filter described in Section 4.1 was used to exclude an estimated 300 million ( $3 \times 10^8$ ) sets of initial conditions that did not produce significant tidal distortions. Of the simulations that the volunteers viewed, they selected over 66000 simulations as being of potential interest and spent time trying to refine the parameters for 13000 simulations in the Enhance activity. This means on average that each pair of galaxies had 4.8 million sets of initial conditions rejected by the  $\beta$  filter with over 50000 simulations reviewed by volunteers who selected, again on average, over 1000 simulations per system to be evaluated with over 16000 Merger Wars competitions.

Figure 4.7 shows the cumulative number of simulations viewed by Citizen Scientists with respect to the time since the site launched. For the first six months the rate at which volunteers reviewed simulations was notably higher than the last two years the site was active. Figure 4.8 groups volunteers into bins by the log of the number of simulations they viewed. Most volunteers viewed at least 64 simulations. There were  $\sim 500$  users that viewed at least 1000 simulations, and 34 users viewed 10000 simulations or more. The two most active volunteers viewed  $\sim 325000$  simulations over 250 hours and  $\sim 553000$  over 100 hours respectively. Based on the time zone information submitted with their results, over 90% of volunteers were from Europe or the United States.

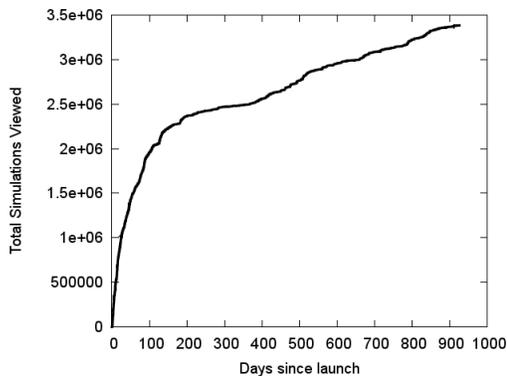


Figure 4.7: Cumulative count of simulations viewed.

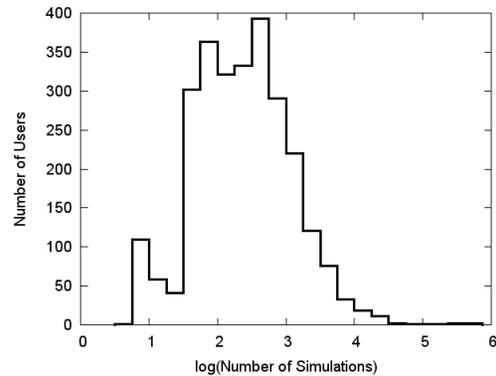


Figure 4.8: The distribution of simulations viewed among volunteers.

## 4.4 Current Process

The capabilities developed for Merger Zoo required the use of various web services ranging from hosting of a web site backed by a database, access to cloud storage services, and manual processing of batches of data to initiate the Merger Wars process. Using lessons learned along the way, a set of two applications has been developed that replace the functionality of the Merger Zoo. They are described in detail in Appendix C. The target preparation package allows a researcher to start with either the RA/Dec coordinates or just the name of a pair of interacting galaxies. It then steps the researcher through each of the target preparation pieces: identifying disk locations, setting disk size and shape, thresholding the image, masking stars, estimating mass and distance, and setting the range of simulation parameters. The tool saves a set of files, including the parameter range file and thumbnail images, for use by a second piece of software. The second tool combines many features of the Merger Zoo applet with the Merger Wars algorithm and some statistical summary information. This tool allows the researcher to perform the rapid search of the Explore activity and the fine tuning of the Enhance activity. However, the selected simulations are made available in realtime for sorting and fitness evaluation by the Merger Wars activity. The two tools together make it possible for a single researcher to prepare a target, view several thousand simulations, and select the best set of simulation parameters in under an hour. These two tools are completely independent of the Merger Zoo and allow a single researcher to perform the same parts of the pipeline previously fulfilled by Citizen Scientists.

### 4.4.1 Potential for a New Automated Method

Attempts were made to develop a new automated fitness evaluator while the Merger Zoo project was active. Efforts focused on the application of computer vision techniques in an ad hoc fashion. Techniques such as contour matching and simple image moments were applied, but they yielded no meaningful correlation with user-assigned fitness values. After the Merger Zoo finale period, we began collaborating with the creator of the WNDCHRM software (Shamir et al., 2008). This software applies many dozens of image measurement

algorithms with multiple scales and orders of moments. In all, it can extract almost 3000 features from a single image. The software also contains a powerful classification tool.

In Shamir et al. (2013) we applied the tool to random selections of user selected states for each galaxy to classify simulation images by which target image they belong too. The simulations images were also rotated about their centers by a randomly selected angle. This was done to ensure that any features identified as useful for classifying galaxies would be rotation independent. The WNDCHRM classifier computes a Fisher score for each image feature indicating how well each feature can separate the classes. The output includes a sorted list of features with the ones having the largest Fisher score on the top of the list.

During this classification activity, the software reported that the Zernike image moments had the highest Fisher scores. Zernike moments are computed from a set of complex polynomials forming an orthogonal basis. Following similar data structures to Boland (1999) we implemented an independent Zernike moment calculation tool and extracted all moments up to order 20 for each target image and each simulation image. The moments were then assigned to a feature vector for each image. The distance between the feature vector of the target image and that of the simulation image is an indication of how similar the images are. Low fitness images have a wide range of Zernike distances. However, high fitness images tend to only have small Zernike distances. A large Zernike distance is a strong indicator of poor fitness. This means that at the very least the Zernike moments can be used to filter out images that are likely to have low fitness, even if some low fitness images will still pass the filter. Some moments correlated more strongly with Merger Wars fitness than others. If we select a subset of moments to compute the distance, we see an even greater correlation. Figure 4.9 below shows Merger Wars fitness as a function of weighted Zernike distance. The overall correlation is not quite strong enough to make a specific prediction about fitness. However, the horizontal and vertical lines demonstrate the lower, right quadrant of low-fitness, high-distance simulations can be easily excluded by an automated Zernike distance calculation.

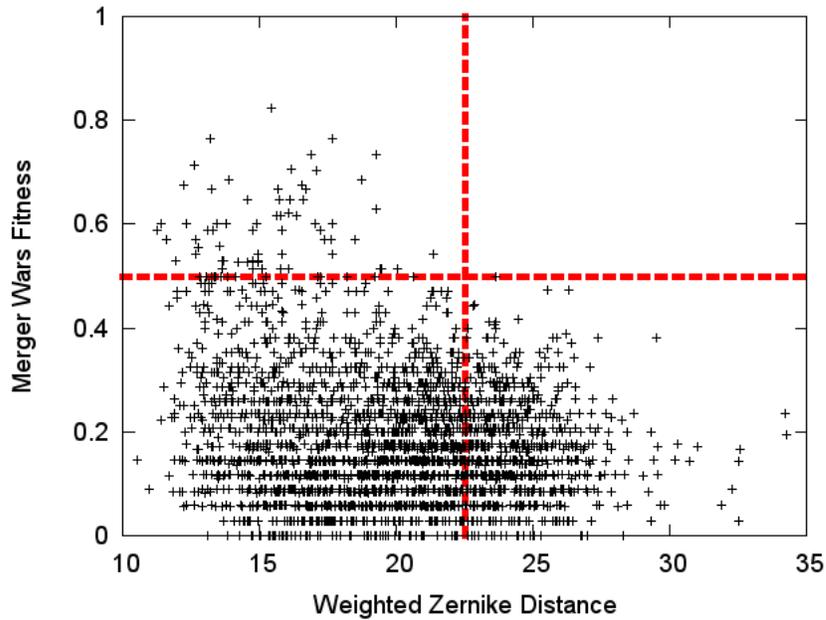


Figure 4.9: Merger Wars fitness as a function of Zernike distance.

Moving beyond the fitness evaluation uses of Zernike moments, we attempted clustering of the target images in that feature space. The objective was to separate the targets into groups with similar morphologies. The results of K-means clustering were somewhat disappointing because the majority of images were assigned to one or two clusters. We then implemented an agglomerative hierarchical clustering routine. The algorithm is seeded by assigning each target image to its own cluster. Each subsequent pass of the algorithm computes which pair of clusters is closest to each other and combines them. The dendrogram plot of the results confirmed that most of the targets were located close to one another in the feature space with a number of outliers. Several of the morphological clusters are presented in Figure 4.10. These images were selected by the clustering algorithm based on their relative distances in Zernike feature space.

The potential correlation between Zernike distance and Merger Wars fitness and the success of clustering similar morphologies are strong indicators that the computer vision

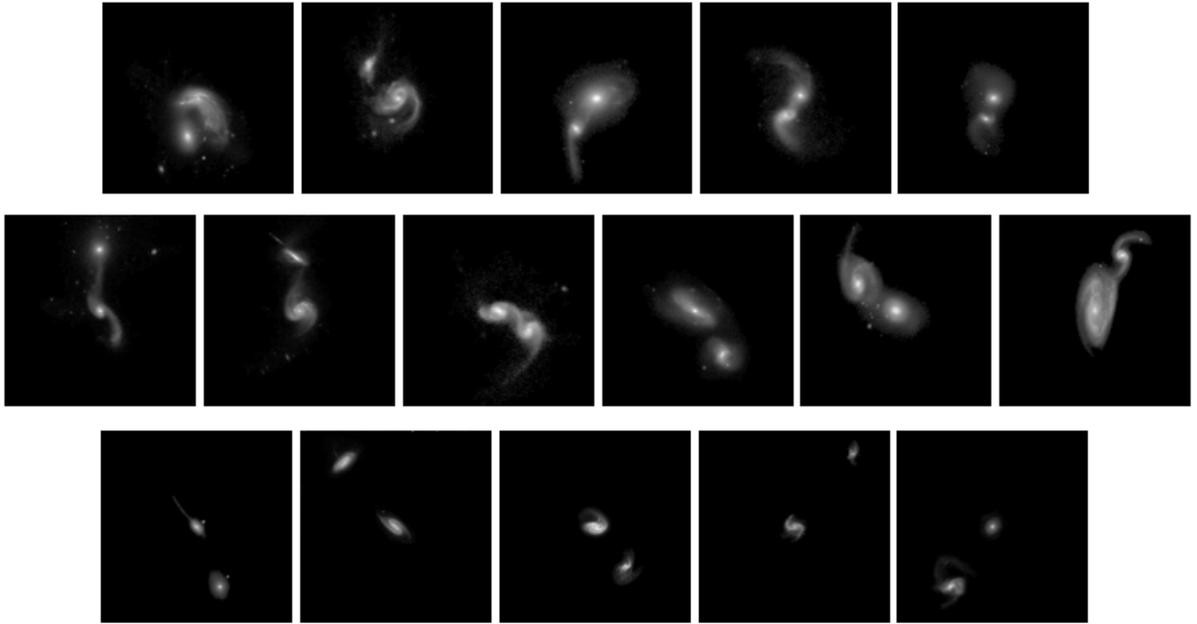


Figure 4.10: Several clusters of similar interacting galaxies selected automatically based on relative Zernike distances.

approach of developing an automated fitness evaluator holds promise. Future work will attempt to modify our genetic algorithm optimizer to make use of a Zernike moment-based fitness function.

## Chapter 5: Simulation Results

In this Chapter we present the simulation results for each system. We then quantify how well the simulation parameters for each system have converged to a unique best-fit.

### 5.1 Results for each Pair of Galaxies

In the next 62 subsections of this Chapter, each of the 62 Merger Zoo targets is presented. The subsection for each target includes summary information. We include a simple histogram of the Merger Wars fitness to indicate the distribution of volunteer-judged “quality” for the simulations. Next we attempt to demonstrate the convergence of the best-fit orbit through several means. The first is to present the target image along with the simulation results of the best 3 targets. Next, we present plots of the trajectories from the simulations for several different fitness populations. Similar trajectories indicate convergence. As the fitness level is increased, the diversity of trajectories should decrease if the model has converged. The next set of plots include information about how much of the total parameter space remains for each fitness population.

We acknowledge the large number of tables and figures that follow is typically considered material for an appendix. However, in this instance, we seek to highlight the quality of results of each system in order to inform later discussion in the Chapter. We believe that we have developed a succinct method for presenting information about several populations of dozen of parameters for tens of thousands of simulations for each system.

### 5.1.1 SDSS 587722984435351614

Each subsection like this one is labelled with the primary name of the target. This target's name indicates that it was selected from the list of targets in SDSS. The list of aliases in Table 5.1.1 includes Arp 240 and a pair of NGC catalog numbers. For other targets, if there is no corresponding Arp or NGC catalog numbers, we provide a cross match to the IC, UGC, and 2MASS catalogs. The right ascension, declination, and redshift are provided as well. The second part of the table summarizes the Merger Zoo activity for the target. It lists the total number of simulations viewed by all volunteers, how many they rejected, how many they selected, and the number that were enhanced. The next three columns describe the Merger Wars outcome for the simulation images for this target. There were over 22000 Merger Wars competitions, but only  $\sim 7000$  winners. That means that for more than 15000 Merger Wars competitions, the volunteers clicked the neither button.

Table 5.1: Identification Information and Merger Zoo summary for SDSS 587722984435351614.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587722984435351614	Arp 240, NGC 5257/5258	13:39:52.8	+0:50:23.4	0.022676

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
74697	71868	2829	603	22745	7463	15282

The low number of Merger Wars winners resulted in a large number of simulations receiving a low Merger Wars fitness score. We see in Figure 5.1 that almost 70% of simulations were assigned a fitness of 0. Looking towards the higher fitness value, we see a relatively low fraction of states with fitness scores above 0.4, and only a few above 0.8. The distribution of fitness values is different for each target.

Next, we present the target image used on Merger Zoo for the pair of interacting galaxies. It is in the upper left corner of Figure 5.2. The single best simulation for each target is

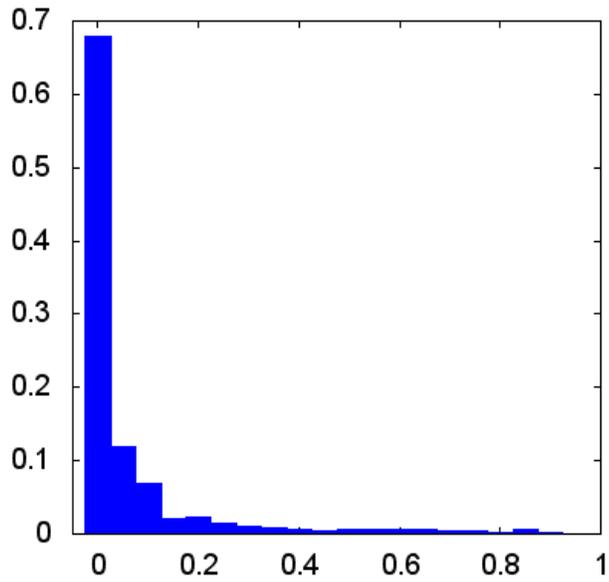


Figure 5.1: Relative frequency of fitness for all selected states of SDSS 587722984435351614

located at the upper right panel of these quad plots. The next two highest fitness simulations occupy the bottom row. In this manner we can view how well the best simulations match the tidal features and overall morphology of the target image. For this particular pair of galaxies, the volunteers have done an excellent job. Each galaxy has symmetric tails that are recreated in the simulation with the proper size and orientation. The quad plot for the target image and top simulations is presented on its own page for each system.



Figure 5.2: Target image and top 3 simulations for SDSS 587722984435351614

Next we present a quad plot in Figure 5.3 for the trajectory of the secondary galaxy relative to the primary galaxy in all of the selected simulations. Each plot shows a set of trajectories. The trajectories are calculated as part of the simulation. They are rotated from the plane of the sky to be in the plane of the primary disk. This is a different rotation for each simulation because the orientation angles for the primary disk,  $\theta_1$  and  $\phi_1$ , are allowed to vary. The black circle represents the size of the primary disk computed from the average of all  $r_1$  values for the set of trajectories plotted in that panel. The circle is the same size in each plot, so the overall scale for each panel is adjusted accordingly. An individual blue line traces the path of the secondary galaxy for a single simulation. The

top left panel shows the paths of all simulations that the volunteers selected. The top right panel shows the trajectories for the top 50% of the population, by fitness. The lower left panel shows the paths for the top 10%. Finally, on the lower right, we see the trajectories for the top 3 simulations. These are the same simulations plotted in Figure 5.2. The top three trajectories pass the primary disk in roughly the same location. They have similar shapes, but one trajectory appears somewhat shorter in this projection than the other two. This means it has a different inclination relative to the plane of the primary disk than the others. Trajectories that are very different from one another in the bottom right panel indicate a non-unique orbit.

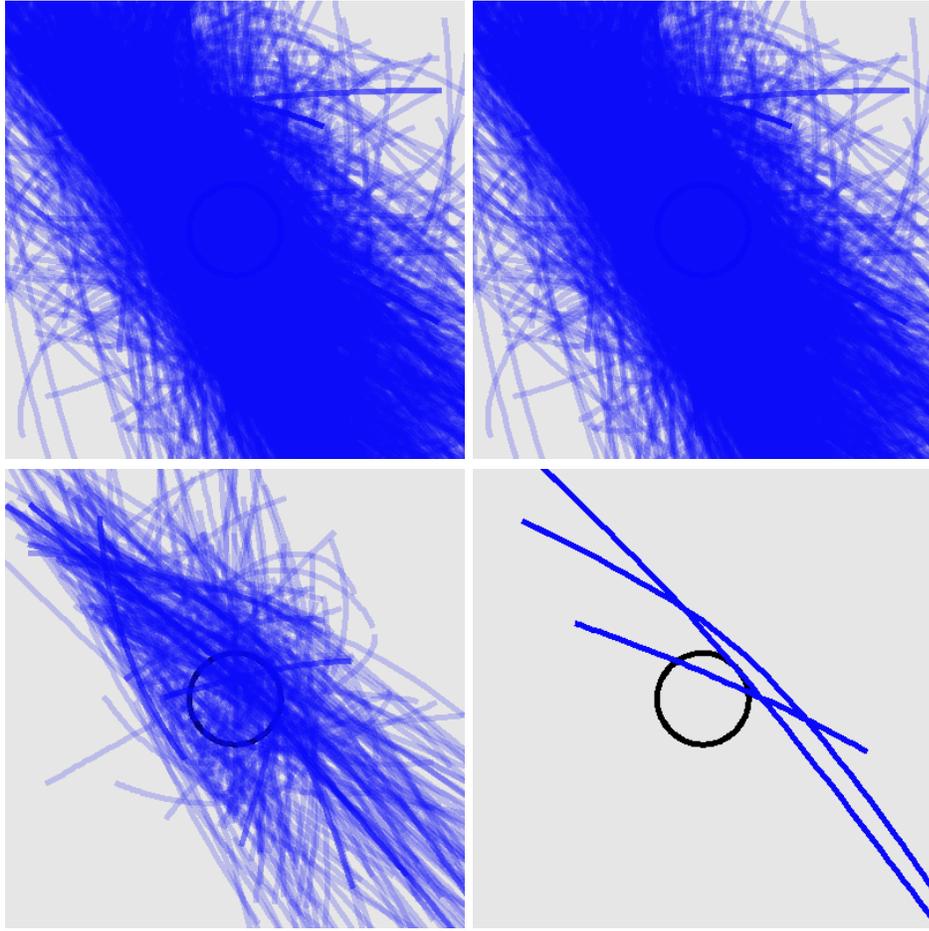


Figure 5.3: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587722984435351614

The final pair of plots for this system describes the level of convergence for each parameter. The parameters are broken up into two different sets. The first set of parameters are the 12 simulation parameters that were varied as part of the Merger Zoo process. This occurred either by random selection by the software during the Explore activity or by the users' selections during the Enhance activity. The 12 parameters are:

- $r_z$  - the z-component of the orbit position vector in plane-of-the-sky frame.
- $v_x$ ,  $v_y$ , and  $v_z$  - the components of the orbit velocity vector in plane-of-the sky frame.
- $m_1$ ,  $m_2$  - the mass of the primary and secondary galaxies

- $r_1, r_2$  - the disk radius of the primary and secondary galaxies
- $\phi_1, \phi_2$  - the position angle of the primary and secondary galaxies
- $\theta_1, \theta_2$  - the disk inclination angle of the primary and secondary galaxies

For each parameter, four populations are considered: all selected states, top 50% by fitness, top 10% by fitness, and the top states selected by the experts. The set of expert states were the ones evaluated in the Merger Zoo finale activity and included between 4 and 8 high fitness simulations for each target. The y-axis represents the fraction of parameter space remaining. For each population, this fraction is computed by dividing the full range of parameters in that population by the full range of parameters for all simulations viewed for that target. For example, consider a target where  $r_z$  was allowed to vary between -10 and 10 simulation units. For the selected states, the value ranged between -5 and 5. The fraction of that parameter remaining would be 10 divided by 20 or 0.5. If the set of expert states had  $r_z$  values between 0 and 2, then the fraction of parameter space remaining for that parameter would be 0.1. For each population, the fraction of parameter space remaining is plotted above each parameter. A line is then drawn connecting these points and the area under this line is filled in. As the fitness of the population is increased, the color darkens. Target systems that are converging to a unique orbit will have a small area under the curve. For this target we see that  $v_x$  and  $v_y$  as well as  $\phi_1$  and  $\phi_2$  have a small fraction of parameter space remaining. That means these values have converged more than other simulation values in the top panel of Figure 5.4.

The bottom panel in Figure 5.4 contains a similar plot for the orbit parameters. These values included classical orbit elements such as eccentricity as well as orientation angles relative to the plane of the sky and the plane of each disk. The orbit parameters shown in the plot are:

- $t_{min}$  - the time since closest approach of the two galaxies.

- $r_{min}$  - the closest approach distance.
- $p$  - the semi-paramater for conic section orbits.
- $ecc$ - the eccentricity of the orbit.
- $inc, lan, \omega$  - the inclination angle for the orbit, longitude of ascending node, and argument of pericenter in plane-of-the-sky frame.
- $dinc, dlan, d\omega$  - the inclination angle for the orbit, longitude of ascending node, and argument of pericenter in frame of the primary disk.
- $dinc2, dlan2, d\omega2$  - the inclination angle for the orbit, longitude of ascending node, and argument of pericenter in frame of the secondary disk.
- $mr$  - mass ratio.
- $vtca$  - the velocity at time of closest approach.
- $cv_1, cv_2$  - the orbital velocity of a particle at the edge of the disk of the primary and secondary galaxy.
- $\beta$  - the interaction parameter.

For this target, most of the orbit parameters are well constrained with little of the parameter space remaining. The orbit angles referenced to the plane of the disk of the secondary galaxy have the highest remaining fraction. This is consistent with the fact that the simulation parameter  $\theta_2$  was not well constrained because that angle is needed to perform the necessary rotation for  $dinc2, dlan2, d\omega2$ .

The information in the parallel coordinate plots can also be represented in a pair of glyph plots. In Figure 5.5 we plot the convergence information for each parameter along its own radial line. Here we have chosen to represent the radial distance in a given direction not by how much of the parameter space remains, but by how much was eliminated. Well converged values have a large radius. This means a well converged set will have a large area.

We plot two populations for each glyph. The bottom layer is the population of the expert states. The top layer is the population of all selected simulations. The area of the glyph for all selected simulations is smaller than the area for the top simulations. This is consistent with the top fitness population having better convergence than the large population.

The glyph plots are not repeated in the individual results for each system. They are however used in a later section as a more compact representation of similar information as what is provided in the parallel coordinate plots.

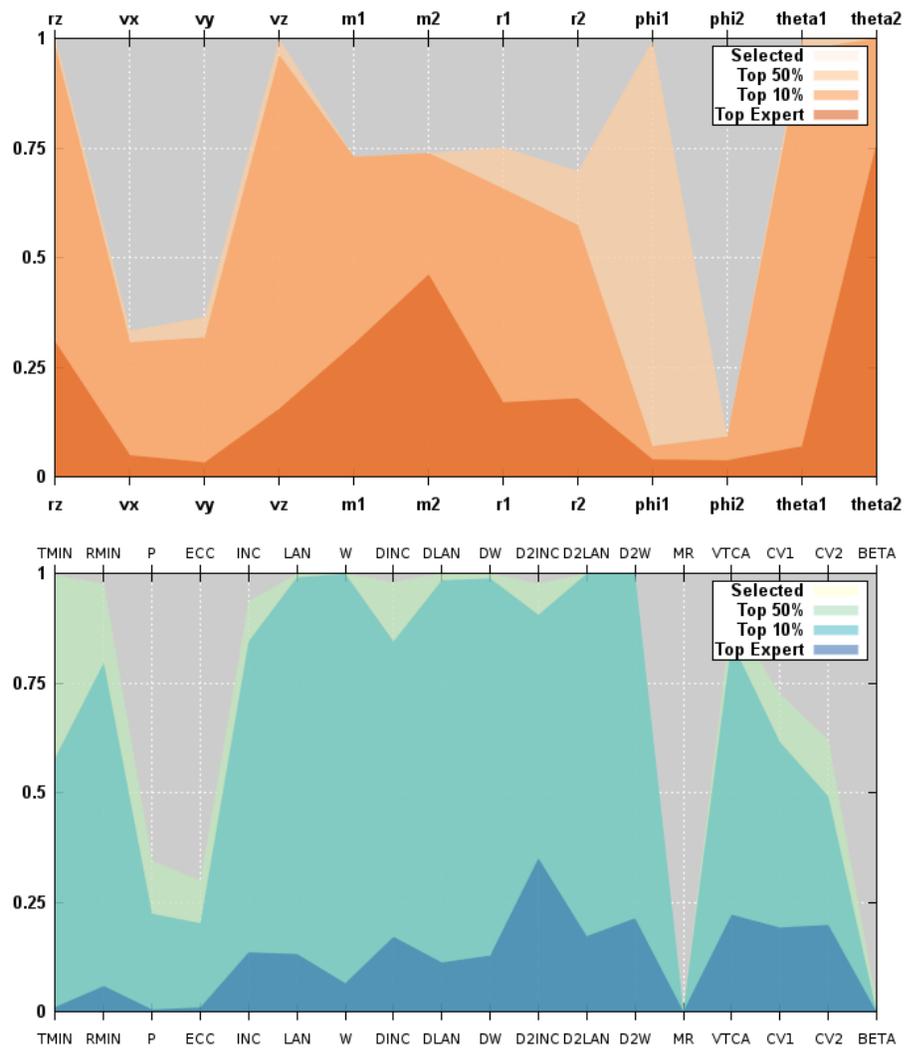


Figure 5.4: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587722984435351614

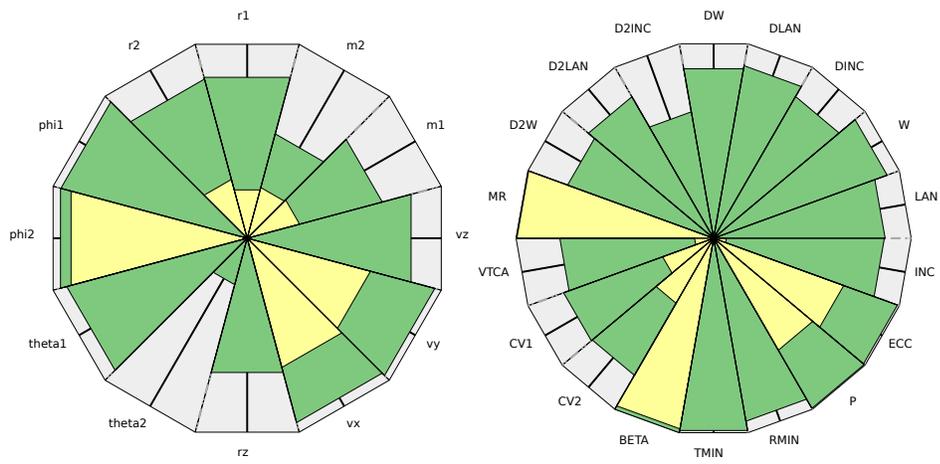


Figure 5.5: Glyph plots for convergence of simulation and orbit parameters for SDSS 587722984435351614

### 5.1.2 SDSS 587724234257137777

Arp 290 has a steadily declining fitness distribution with no simulations with a fitness above 0.75. The top simulations do capture the general shape of the primary galaxy including the enhanced density in the southern region. The trajectories show moderate convergence. The simulation parameters are largely converged but with uncertainty in the disk orientation parameters that translates to uncertain orbit angle parameters.

Table 5.2: Identification Information and Merger Zoo summary for SDSS 587724234257137777.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587724234257137777	Arp 290, IC 195/196	2:03:49.7	+14:44:19.1	0.012162

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
80157	78432	1725	347	26568	12848	13720

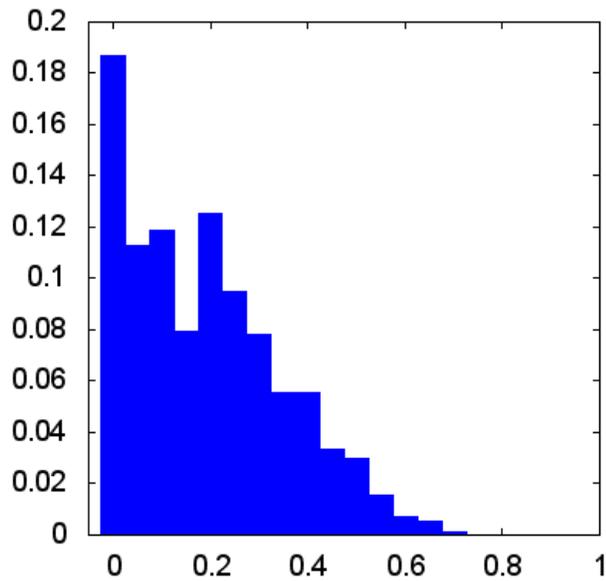


Figure 5.6: Relative frequency of fitness for all selected states of SDSS 587724234257137777

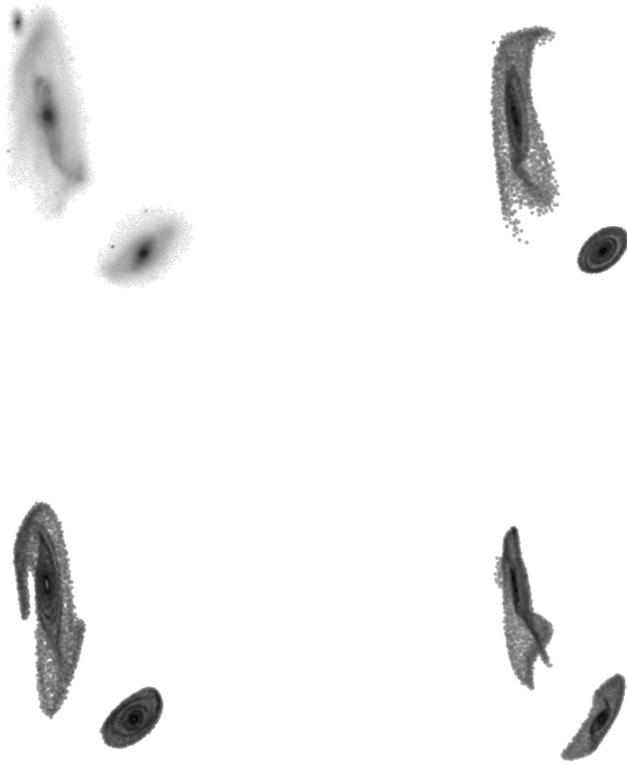


Figure 5.7: Target image and top 3 simulations for SDSS 587724234257137777

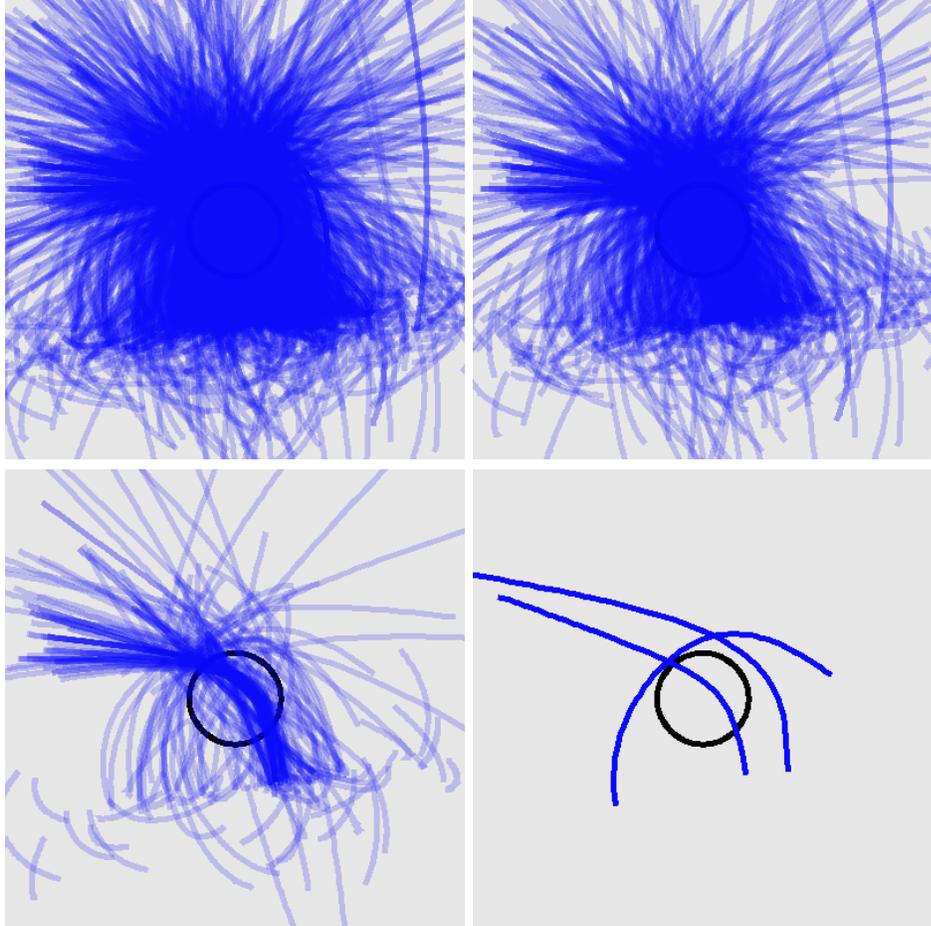


Figure 5.8: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587724234257137777

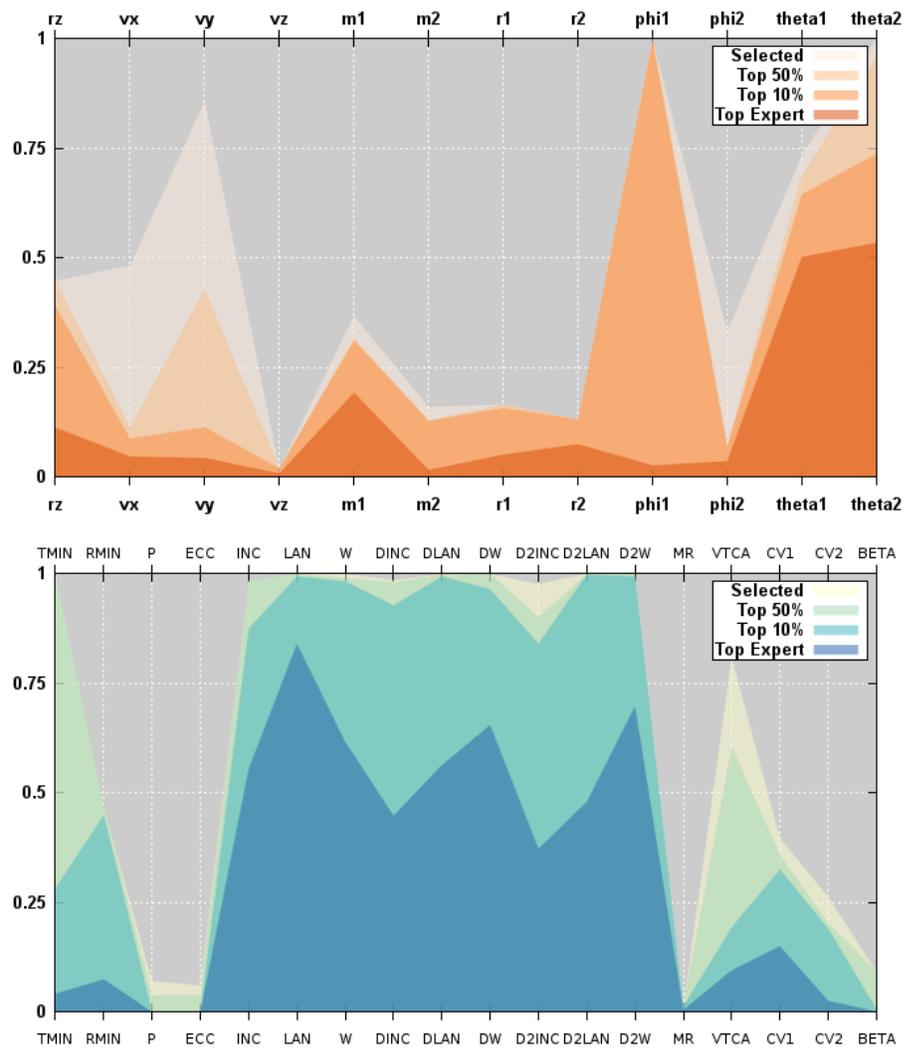


Figure 5.9: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587724234257137777

### 5.1.3 SDSS 587726033843585146

Arp 142, sometimes called the Penguin Galaxy, has a sharp drop off in fitness. However the high fitness simulations have impressively captured the asymmetric tidal features. The top 3 trajectories are almost identical. The convergence of the simulation parameters is good except for the mass of the secondary and the disk orientation parameters. The uncertain secondary mass also leaves the  $cv_2$  parameter somewhat unconstrained.

Table 5.3: Identification Information and Merger Zoo summary for SDSS 587726033843585146.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587726033843585146	Arp 142, NGC 2936/2937	9:37:44.0	+2:45:36.5	0.023313

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
28253	27909	344	80	6399	2391	4008

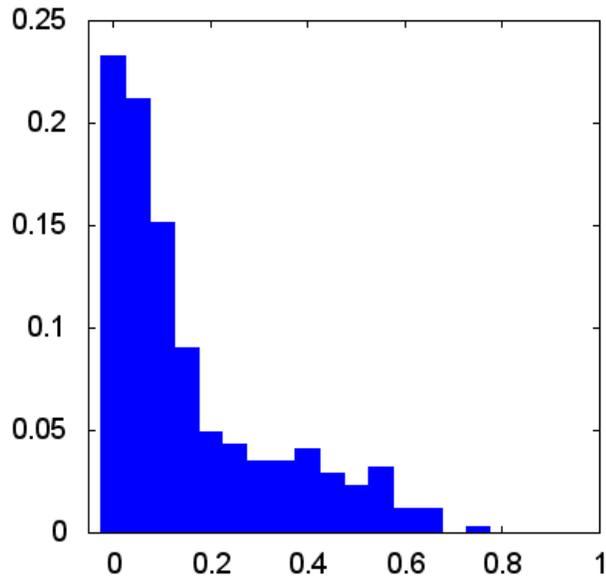


Figure 5.10: Relative frequency of fitness for all selected states of SDSS 587726033843585146

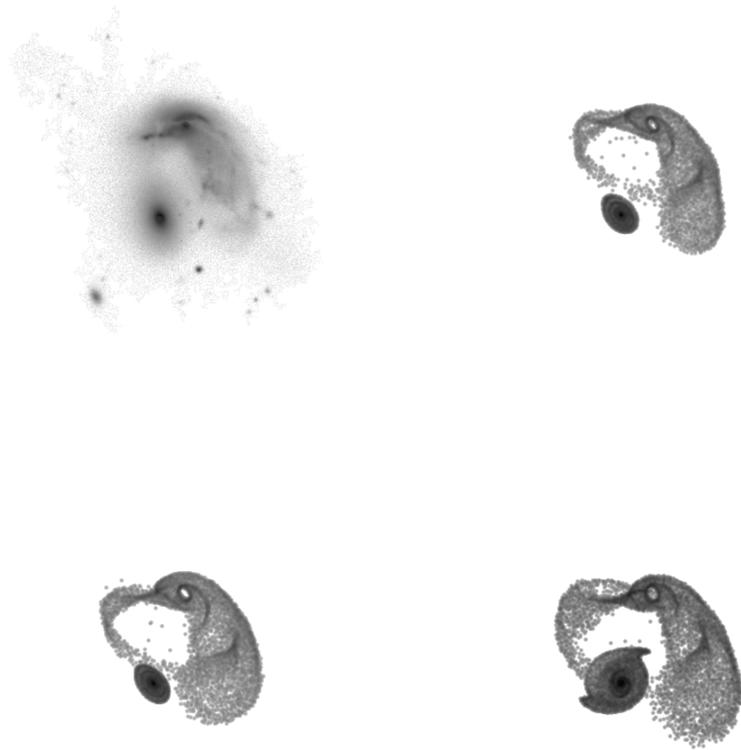


Figure 5.11: Target image and top 3 simulations for SDSS 587726033843585146

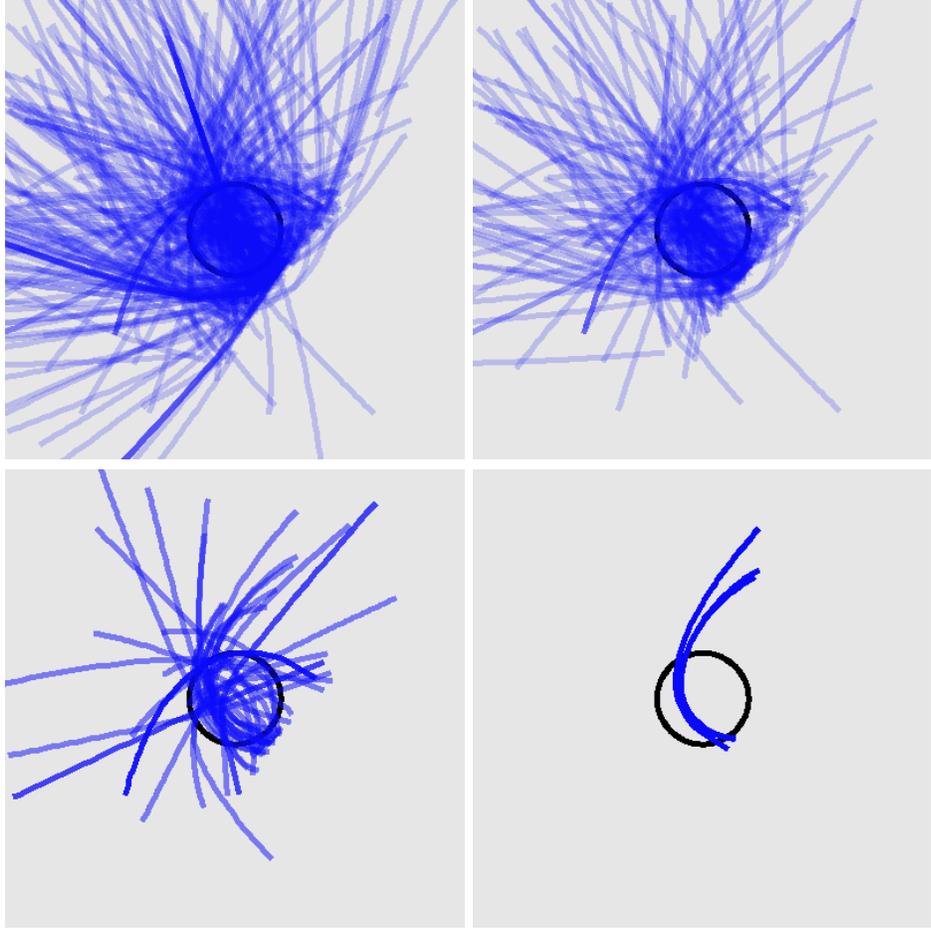


Figure 5.12: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587726033843585146

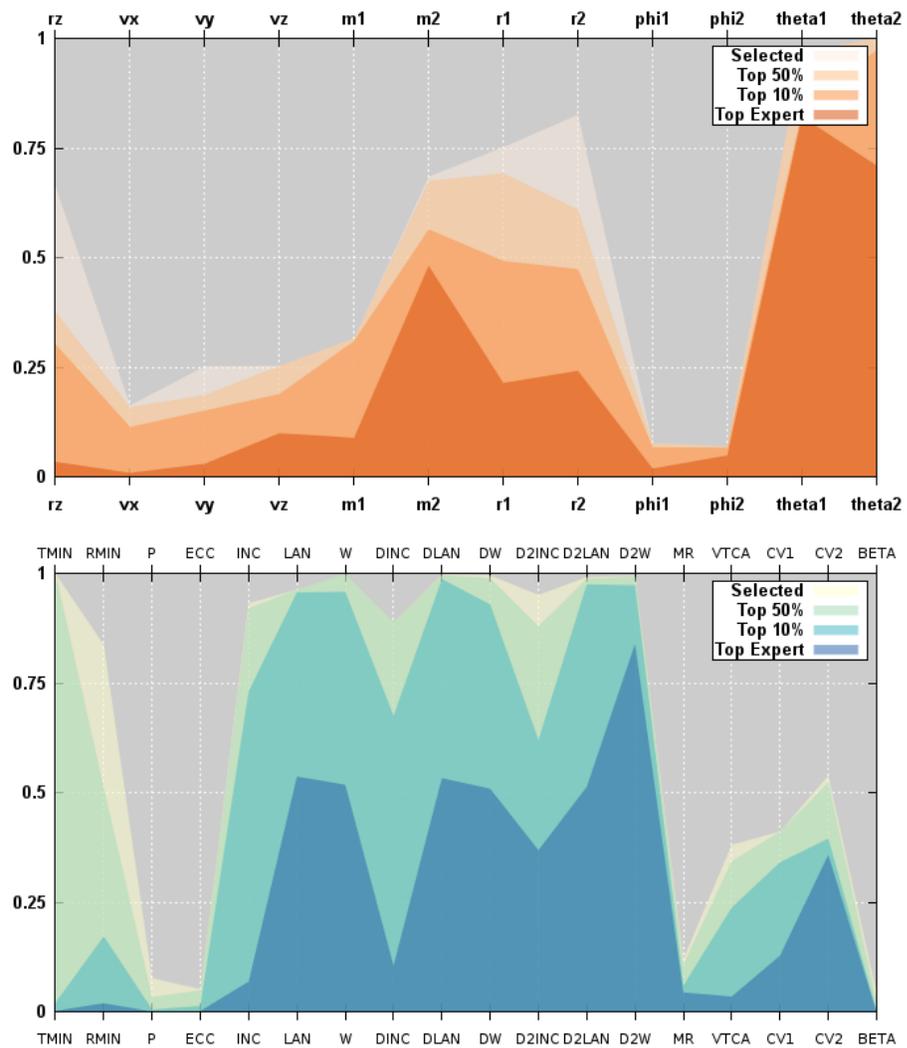


Figure 5.13: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587726033843585146

### 5.1.4 SDSS 587727177926508595

Arp 318 has few high fitness states. The top simulations show a reasonable recreation of the western tail of the primary. The top 3 trajectories are very similar. The convergence on simulation parameters is excellent except for  $\theta_2$ . The orbit parameters are also well constrained except for some of the orientation angles.

Table 5.4: Identification Information and Merger Zoo summary for SDSS 587727177926508595.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587727177926508595	Arp 318, NGC 833/835	2:09:24.5	-10:08:09.6	0.013586

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
34267	33761	506	100	9101	3884	5217

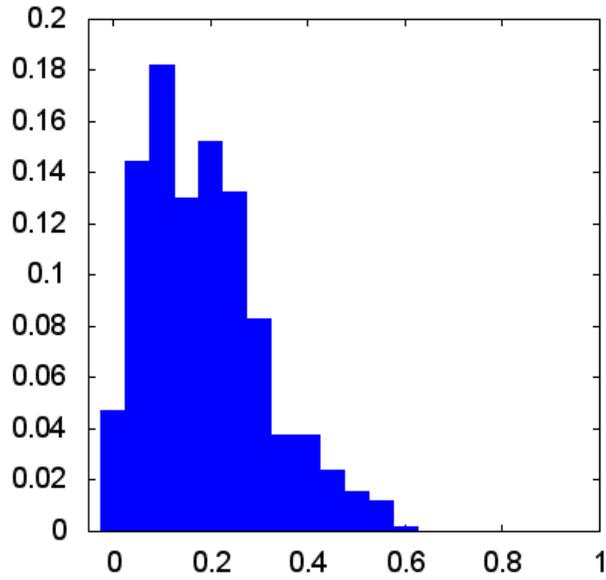


Figure 5.14: Relative frequency of fitness for all selected states of SDSS 587727177926508595

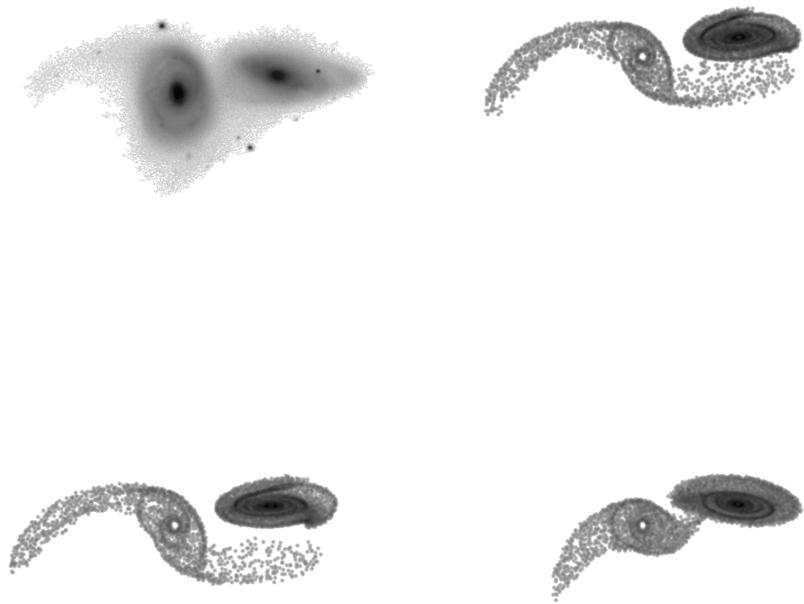


Figure 5.15: Target image and top 3 simulations for SDSS 587727177926508595

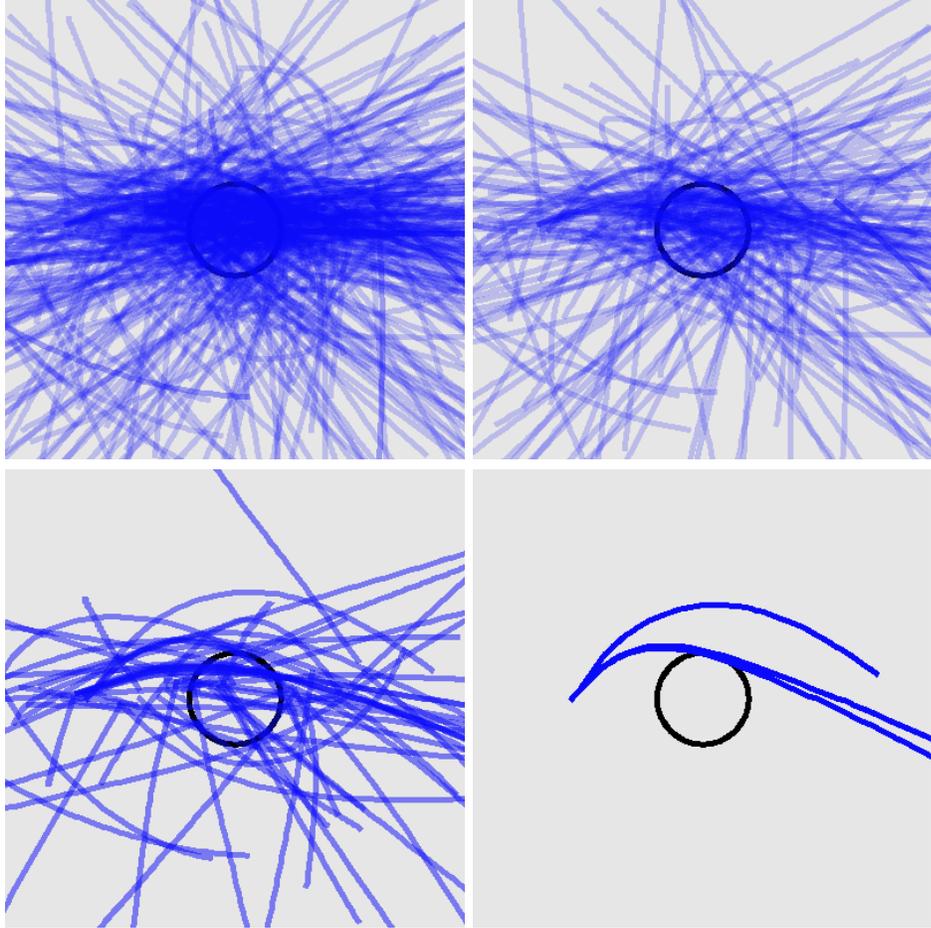


Figure 5.16: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587727177926508595

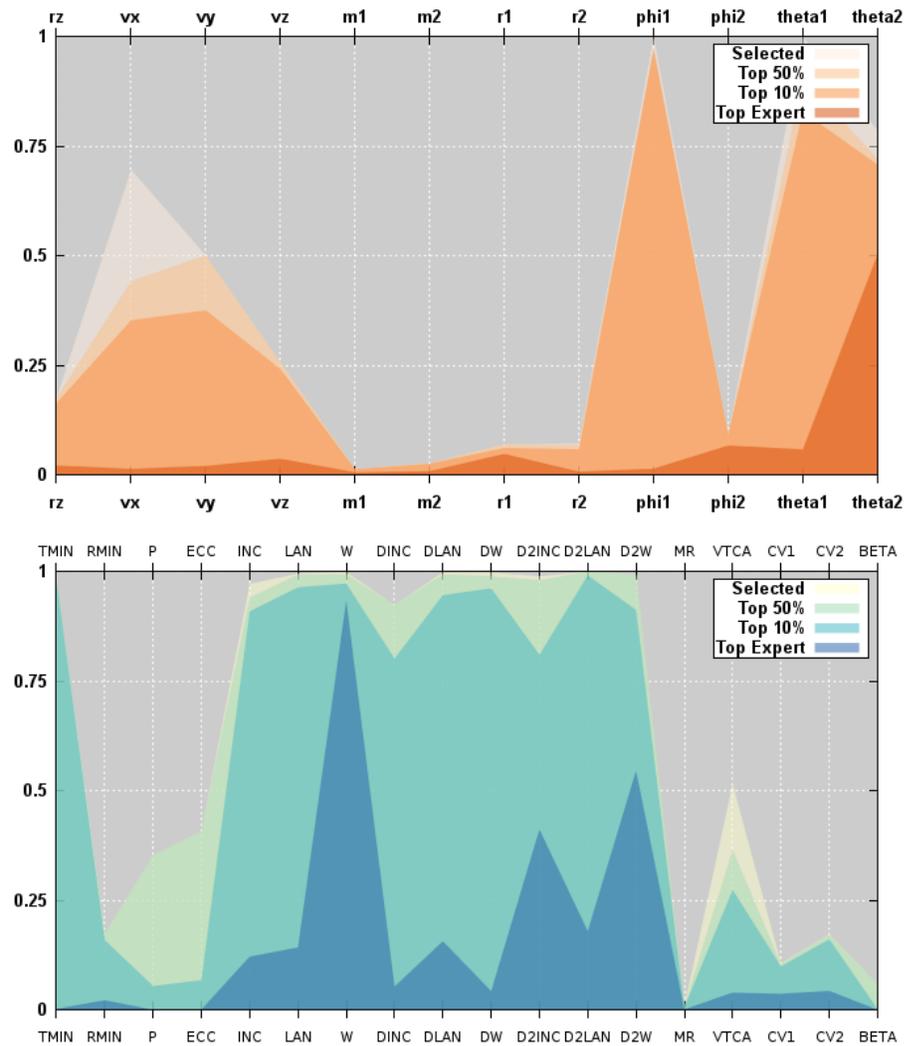


Figure 5.17: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587727177926508595

### 5.1.5 SDSS 587727178988388373

Arp 256 has a maximum fitness near 0.75. The simulations show that the tidal features of the primary and the secondary are both recreated. The top 3 trajectories pass by the primary disk in almost the same location. For the simulation parameters the mass values remain unconstrained which adds uncertainty to a number of the orbit parameters.

Table 5.5: Identification Information and Merger Zoo summary for SDSS 587727178988388373.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587727178988388373	Arp 256	0:18:50.0	-10:21:41.8	0.027329

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
28823	28199	624	215	11105	5055	6050

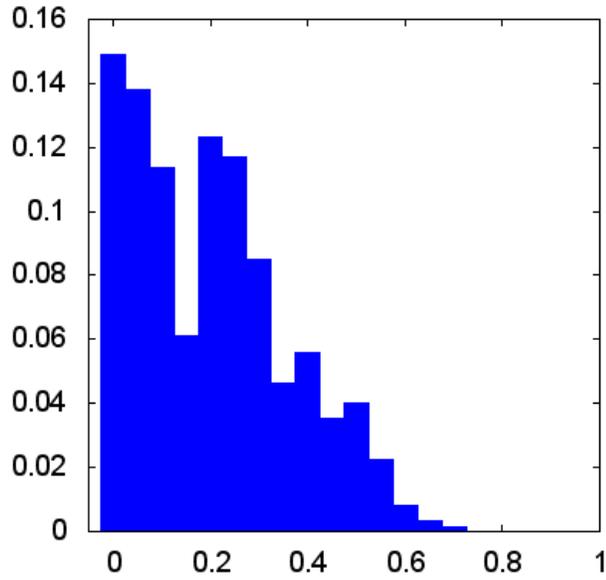


Figure 5.18: Relative frequency of fitness for all selected states of SDSS 587727178988388373

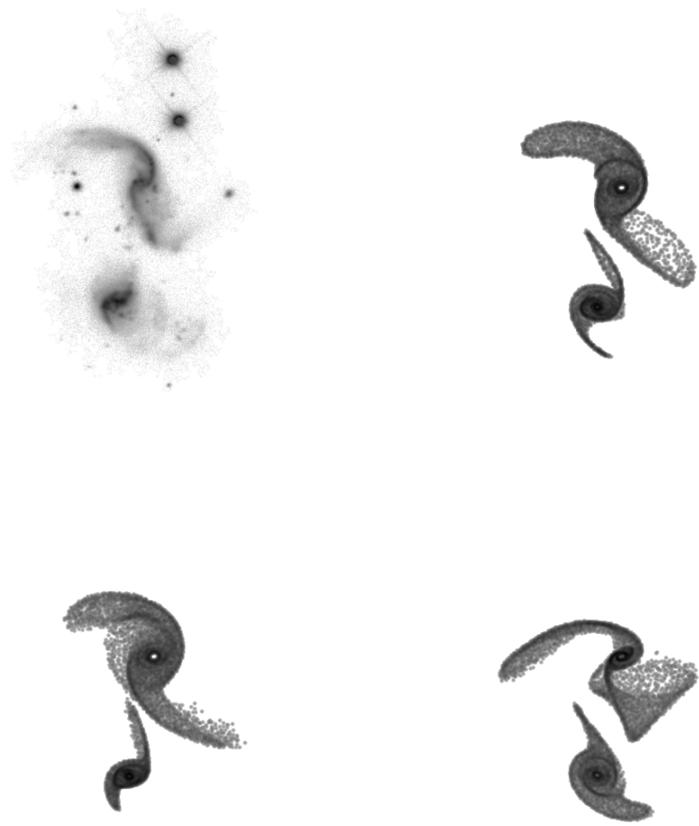


Figure 5.19: Target image and top 3 simulations for SDSS 587727178988388373

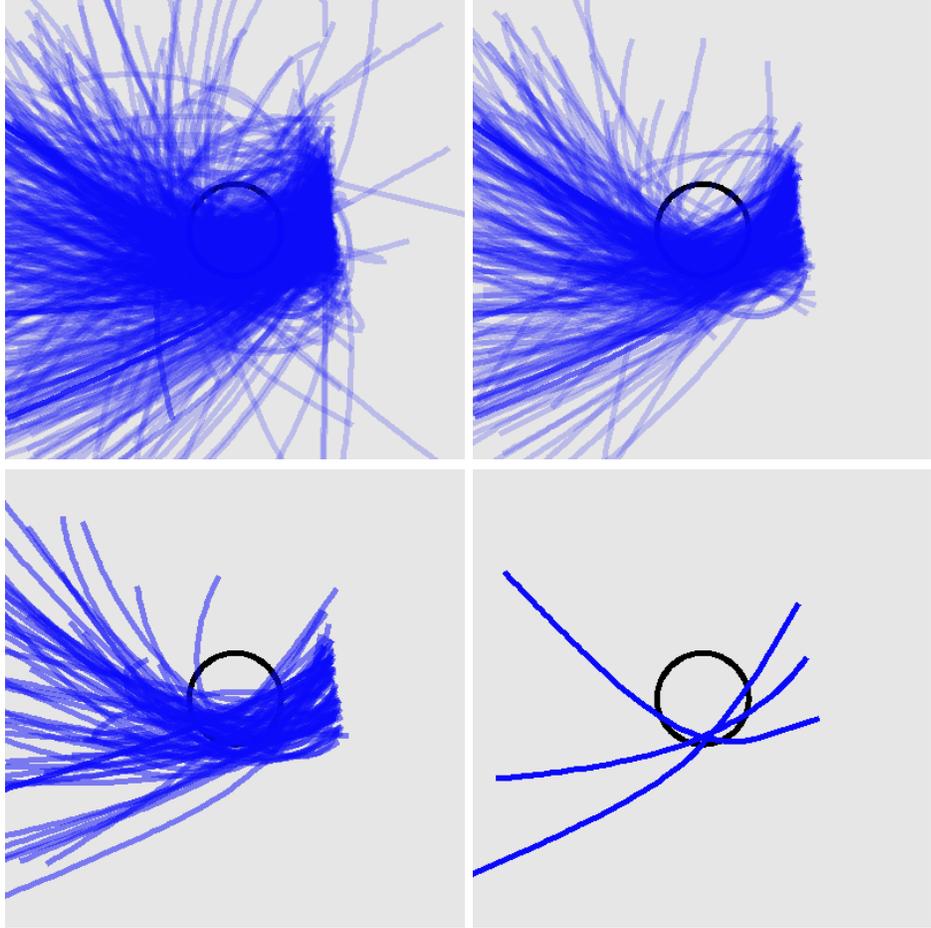


Figure 5.20: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587727178988388373

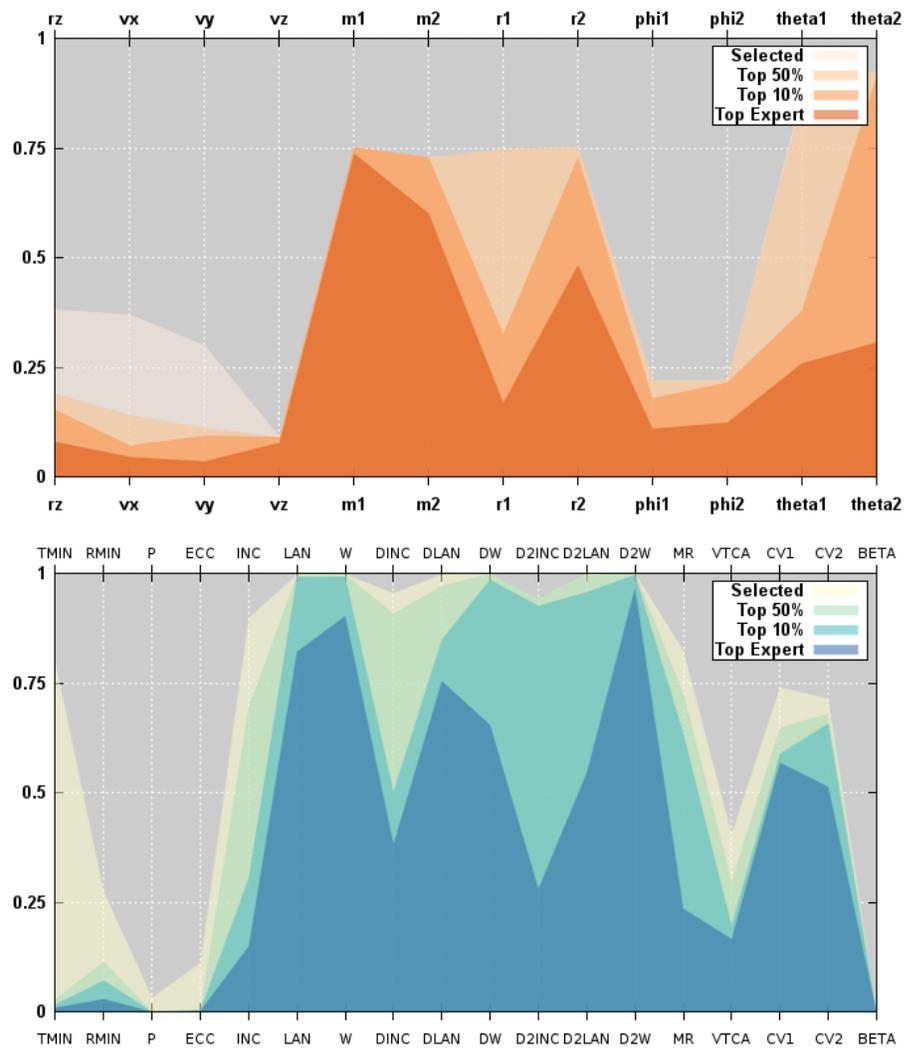


Figure 5.21: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587727178988388373

### 5.1.6 SDSS 587727222471131318

This target has a steep fitness decline. Only a small number of states, 259, were selected by the volunteers. The top three simulations have an accurate recreation of the tidal features of the primary. The bar feature in the secondary is not well modelled by a restricted three-body simulation. The trajectories for the top to fitness populations appear converged. The simulation parameters are well constrained with the exception of  $\phi_1$ . This is understandable because the inclination angle for this galaxy is low, its plane is largely the same as the plane of the sky, so the position angle is not well defined.

Table 5.6: Identification Information and Merger Zoo summary for SDSS 587727222471131318.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587727222471131318	UGC 11751	21:28:59.3	+11:22:55.0	0.029087

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
30080	29821	259	71	4899	1914	2985

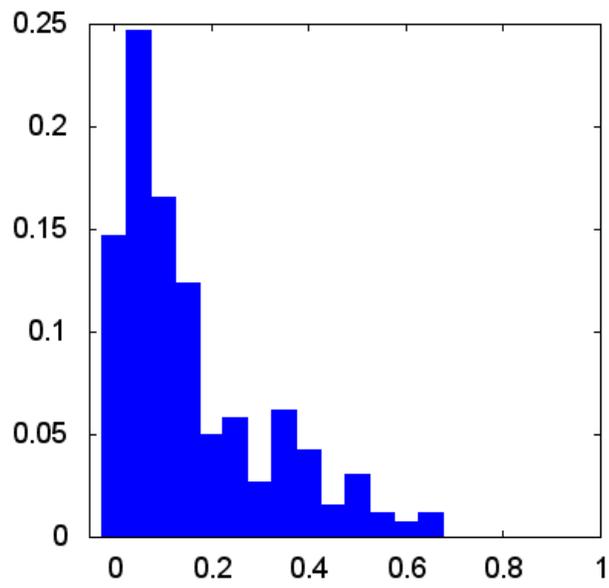


Figure 5.22: Relative frequency of fitness for all selected states of SDSS 587727222471131318

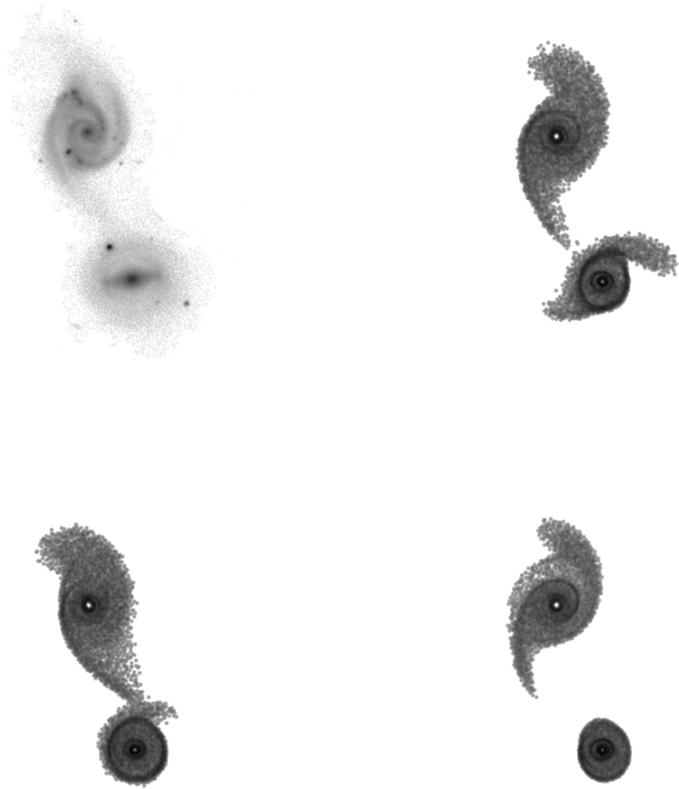


Figure 5.23: Target image and top 3 simulations for SDSS 587727222471131318

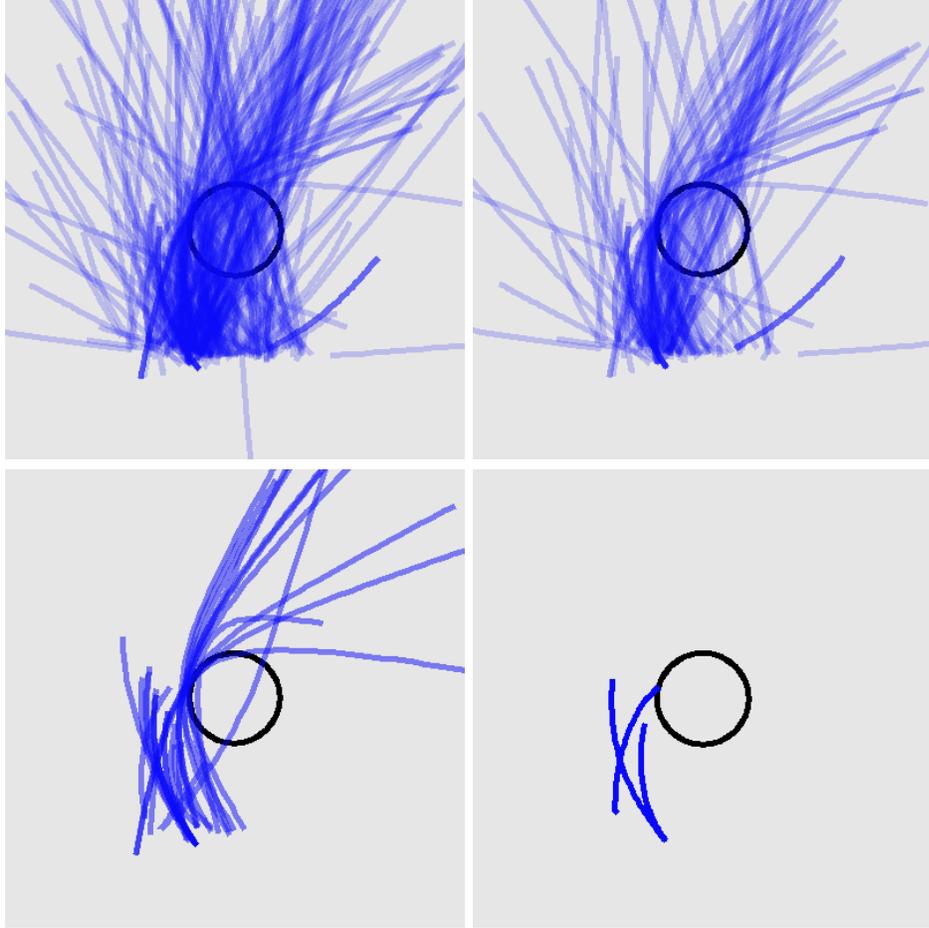


Figure 5.24: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587727222471131318

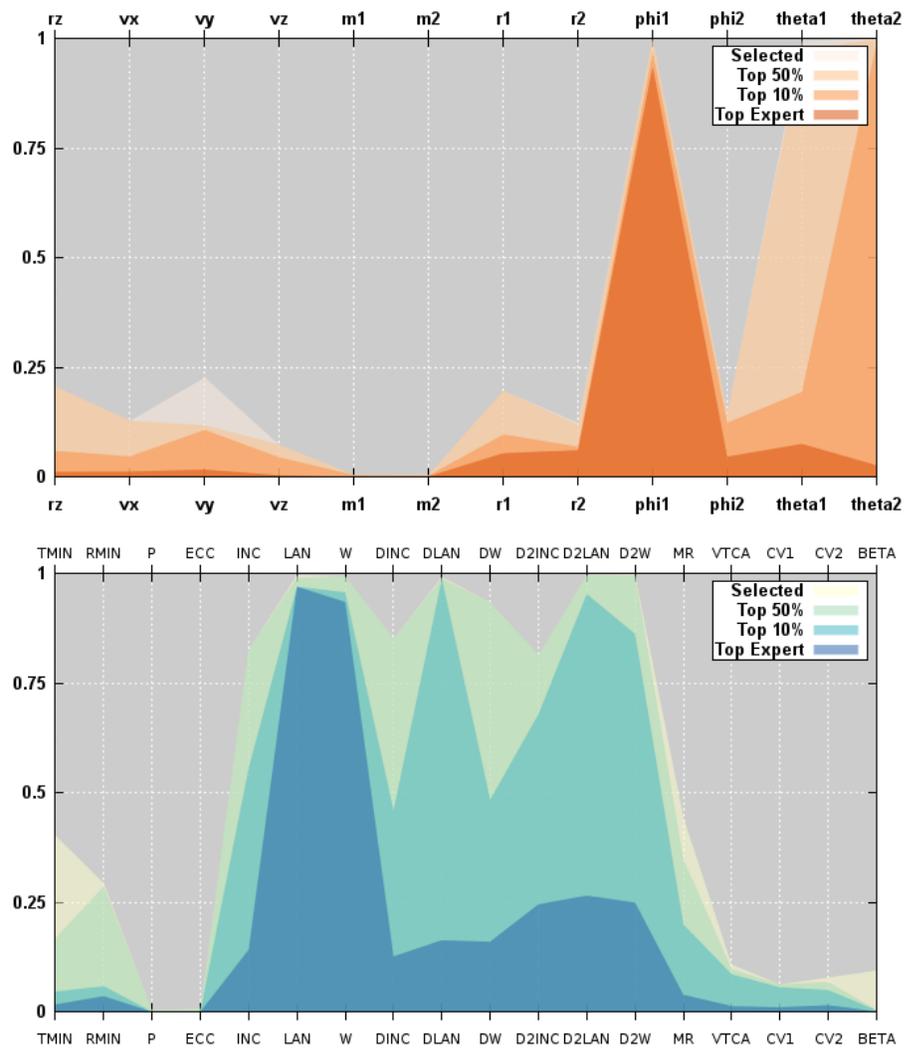


Figure 5.25: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587727222471131318

### 5.1.7 SDSS 587728676861051075

For Arp 104, over 0.75 of simulations have a Merger Wars fitness of 0. Regardless of the small number of higher fitness states, two of the top 3 simulations have recreated a thin tidal bridge between the two galaxies. However, they also show a southern tail on the secondary galaxy that is not in the target image. This is probably not a correct match to the morphology. The trajectories show a lack of specific convergence. However the top 3 trajectories all show a multiple passsage orbit. The masses and orientation angles are not well constrained, nor are the orbit parameters.

Table 5.7: Identification Information and Merger Zoo summary for SDSS 587728676861051075.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587728676861051075	Arp 104, NGC 5216/5218	13:32:10.2	+62:46:02.4	0.009783

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
65717	64074	1643	348	8161	3164	4997

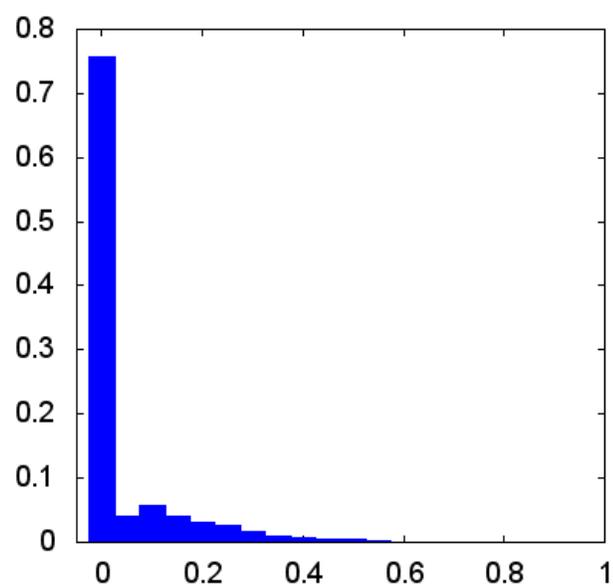


Figure 5.26: Relative frequency of fitness for all selected states of SDSS 587728676861051075

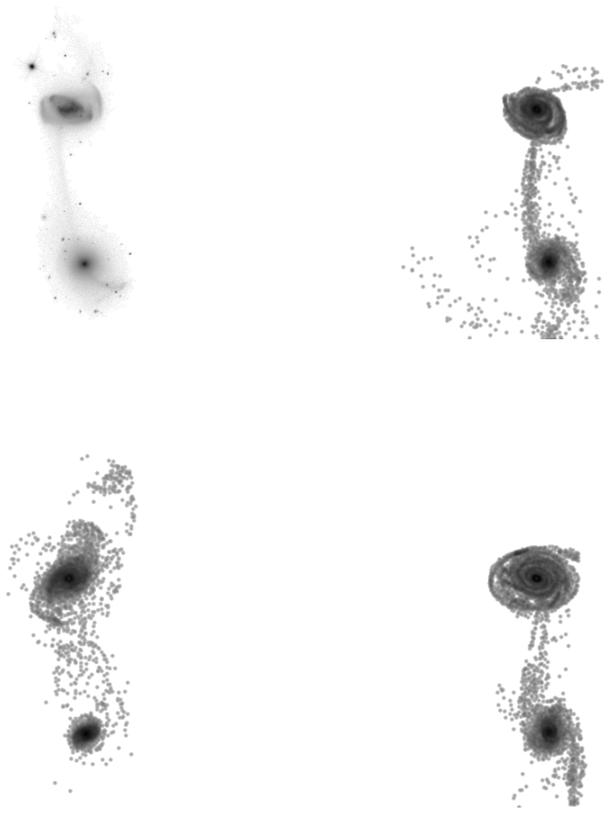


Figure 5.27: Target image and top 3 simulations for SDSS 587728676861051075

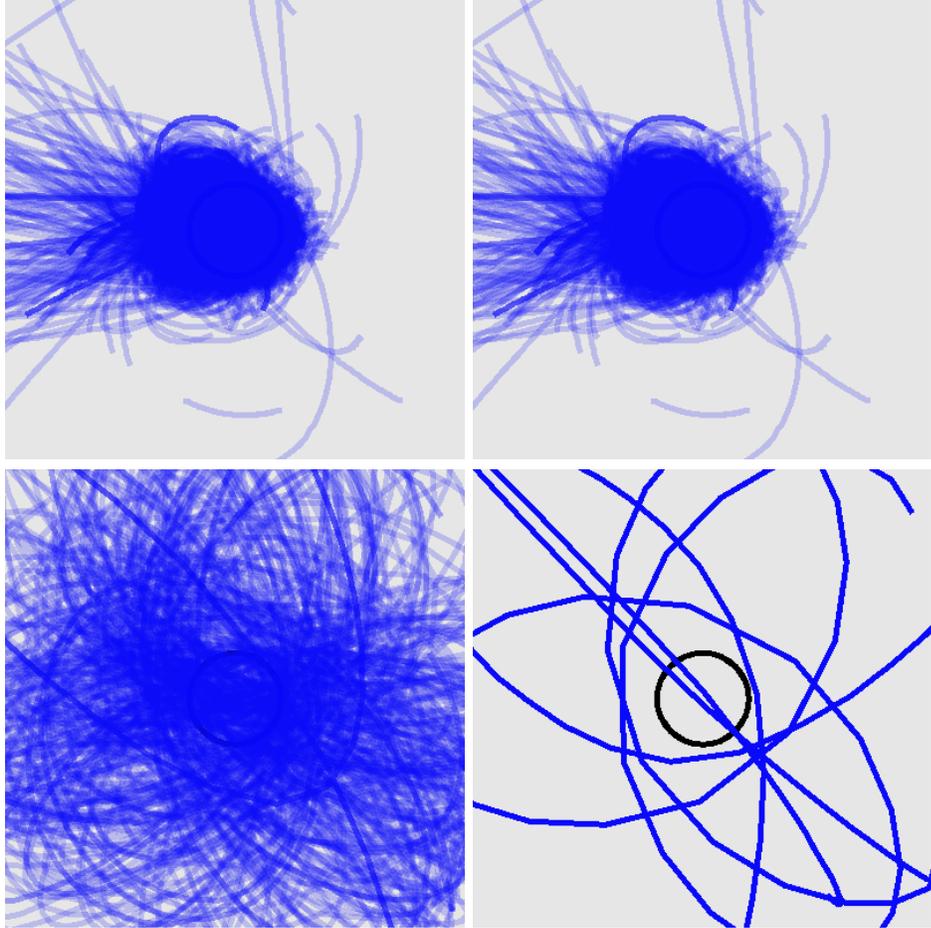


Figure 5.28: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587728676861051075

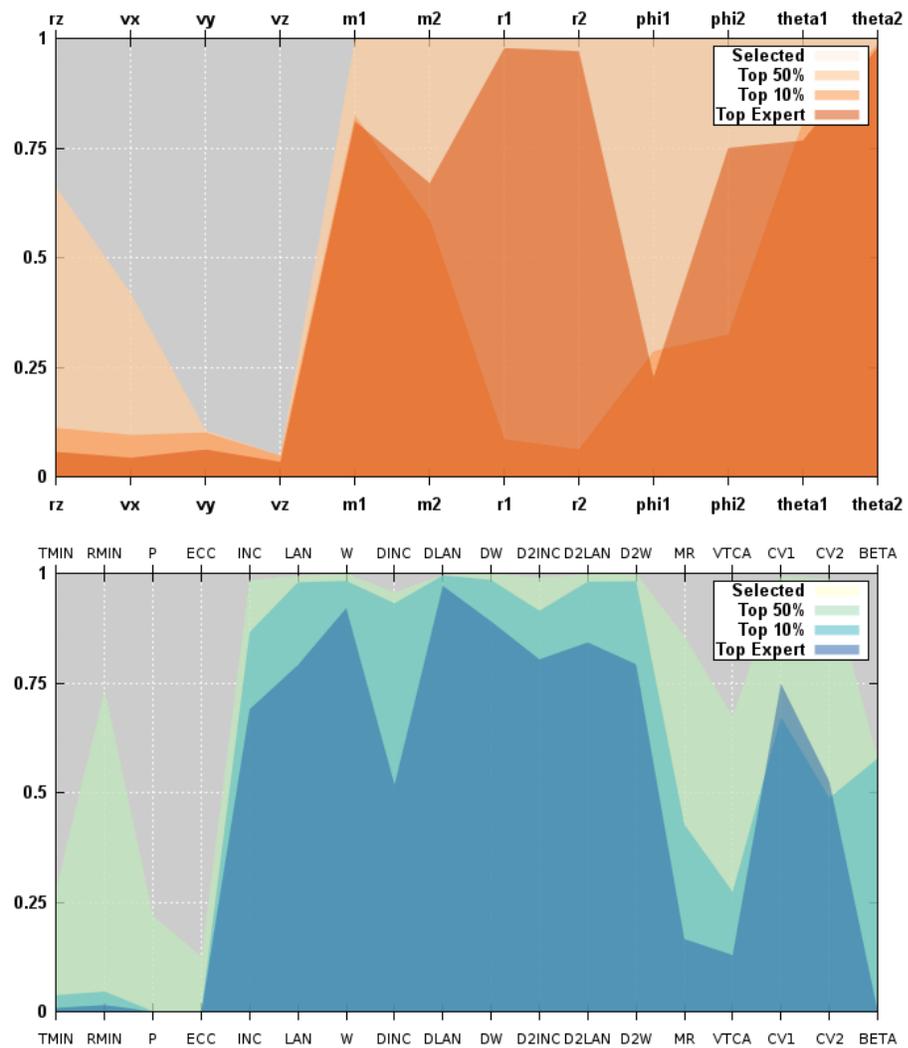


Figure 5.29: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587728676861051075

### 5.1.8 SDSS 587729227151704160

This double ring galaxy, or Heart Galaxy, was one of the more popular Merger Zoo targets. Over 177000 simulations were viewed by volunteers. All of the top 3 simulations have recreated at least one ring, with the other galaxy showing a thinning out of the disk, if not a fully formed ring. The top trajectories all pass through the disk of the primary, which is expected for a ring galaxy. The primary mass was not well constrained, but the remainder of simulation and orbit parameters were.

Table 5.8: Identification Information and Merger Zoo summary for SDSS 587729227151704160.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587729227151704160	2MASS 15530935+5408557	15:53:08.6	+54:08:50.4	0.046756

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
177399	174398	3001	554	50678	19256	31422

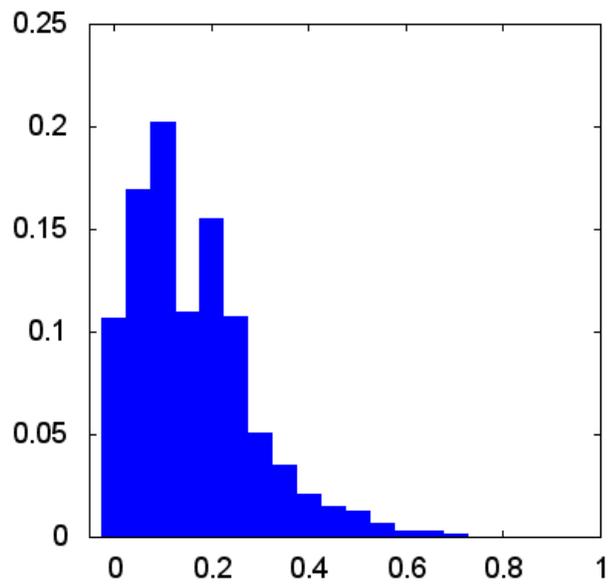


Figure 5.30: Relative frequency of fitness for all selected states of SDSS 587729227151704160

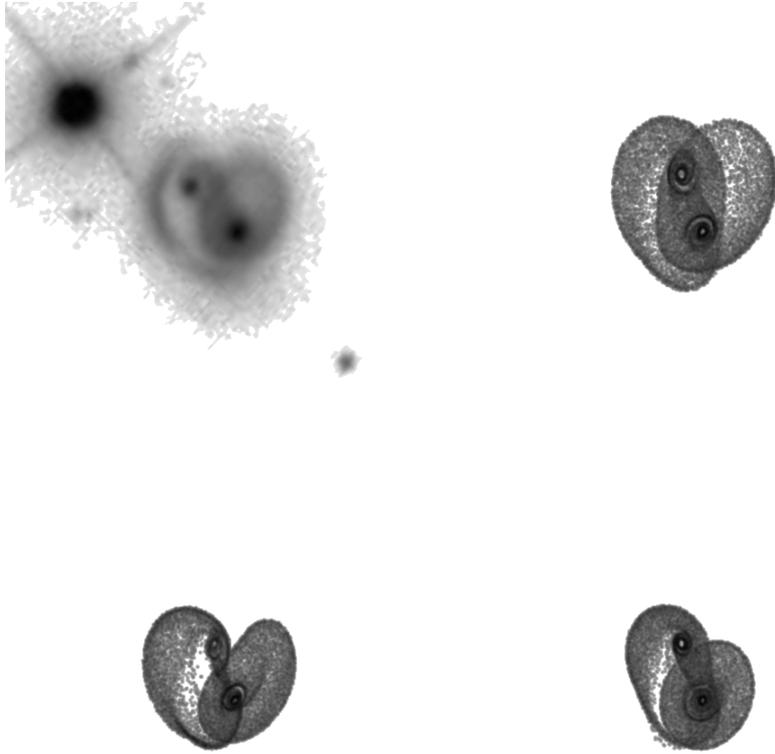


Figure 5.31: Target image and top 3 simulations for SDSS 587729227151704160

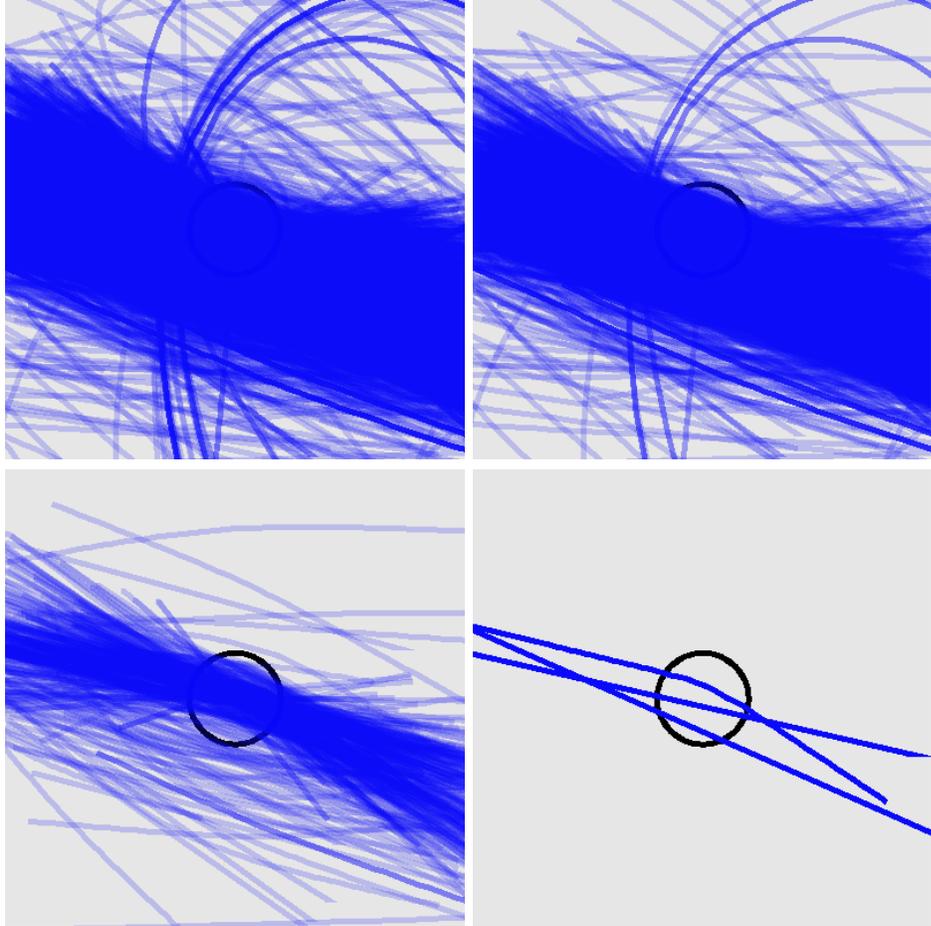


Figure 5.32: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587729227151704160

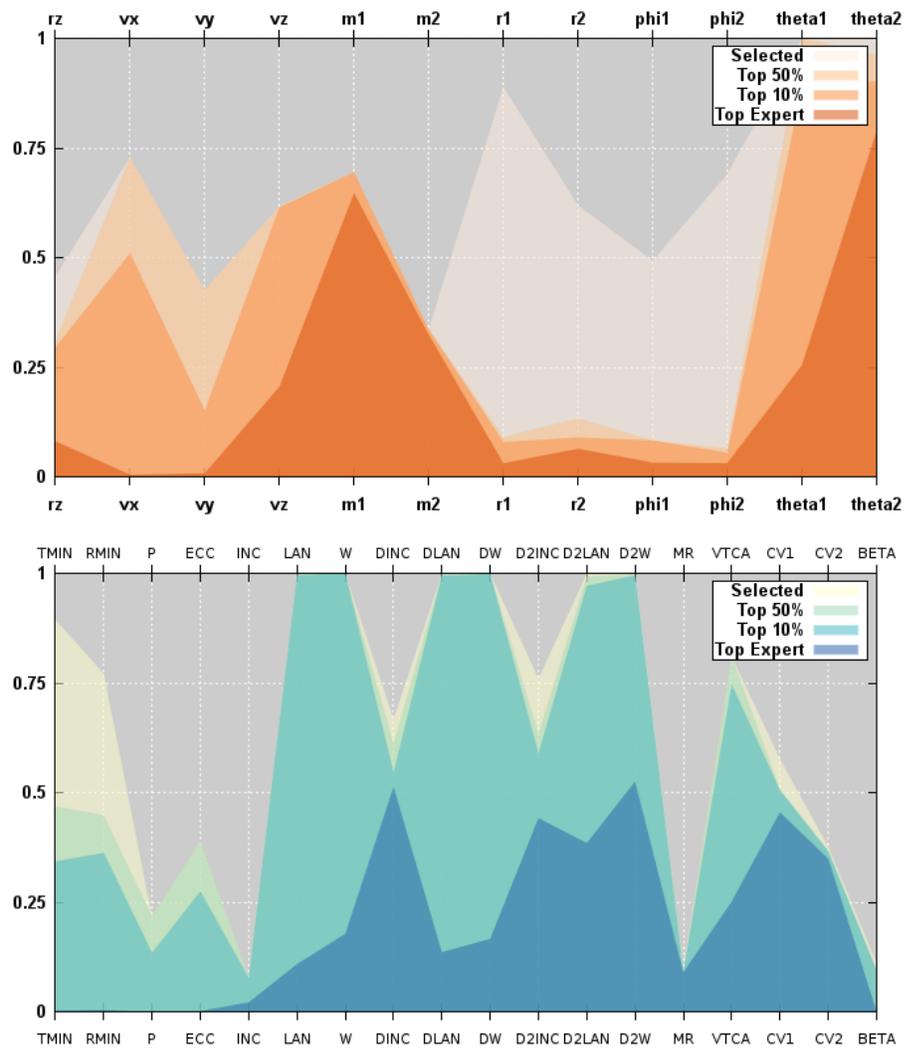


Figure 5.33: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587729227151704160

### 5.1.9 SDSS 587731913110650988

Arp 285 has a high number of mediocre fitness states near 0.3. The top 3 simulations recreate the symmetric tails of the primary. The jet-like feature in the secondary is not recreated. The trajectories are close to identical. The masses and disk sizes are not well constrained, but yet the orbit parameters seem relatively converged.

Table 5.9: Identification Information and Merger Zoo summary for SDSS 587731913110650988.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587731913110650988	Arp 285, NGC 2854/2856	9:24:02.9	+49:12:14.1	0.009243

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
17971	17412	559	82	10201	5590	4611

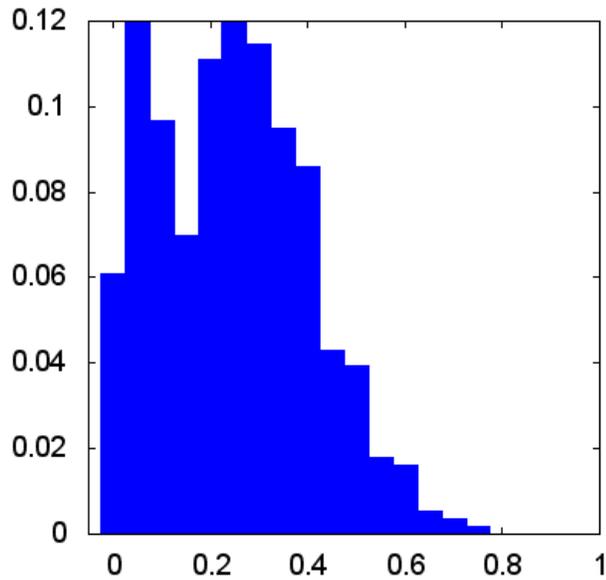


Figure 5.34: Relative frequency of fitness for all selected states of SDSS 587731913110650988



Figure 5.35: Target image and top 3 simulations for SDSS 587731913110650988

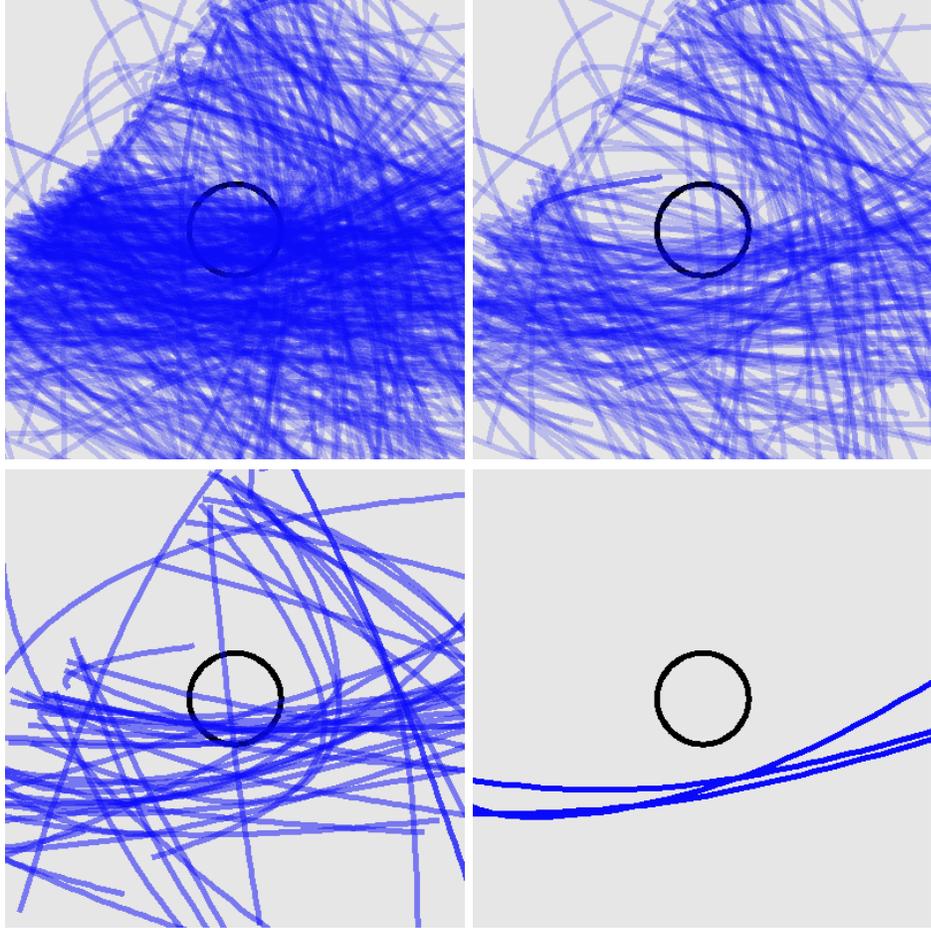


Figure 5.36: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587731913110650988

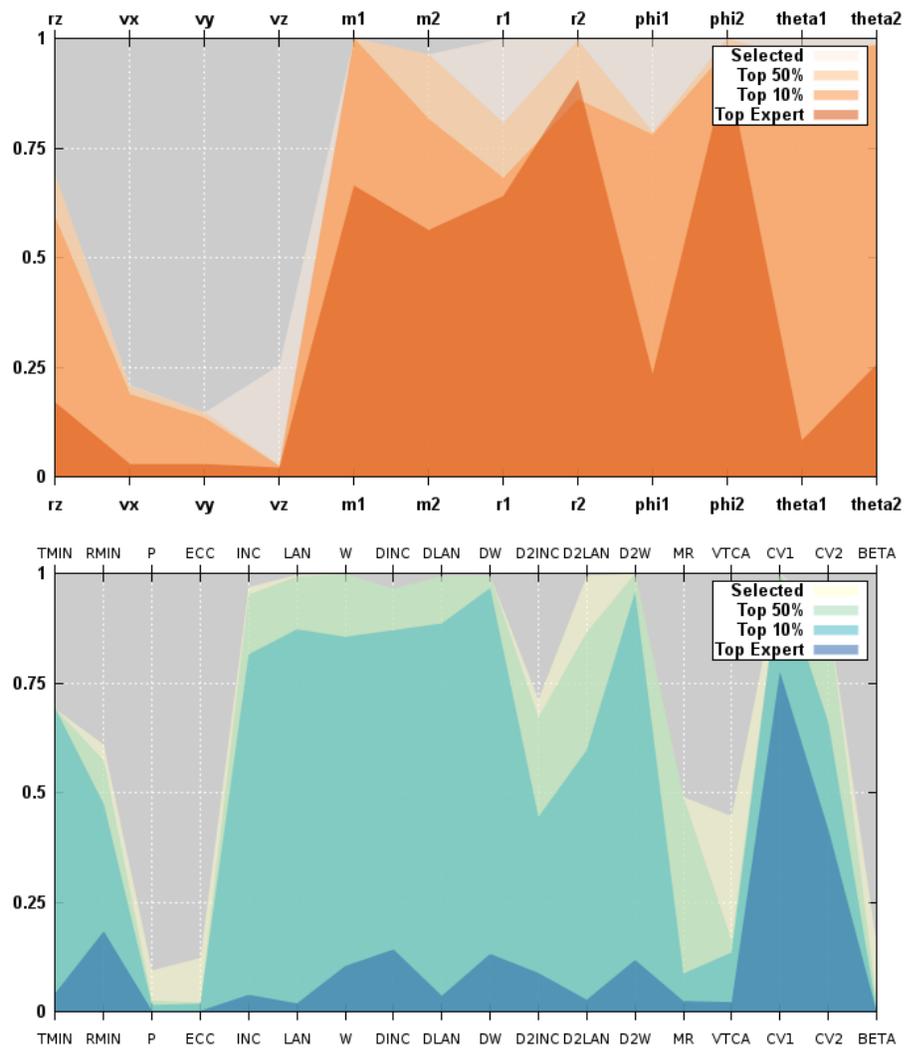


Figure 5.37: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587731913110650988

### 5.1.10 SDSS 587732136993882121

Arp 214 has a large portion of low fitness simulations, but not all 0 fitness. The simulations recreate the tails of the primary but do not generate them with quite the same length. The top trajectories are similar. The simulation parameters are well constrained with the exception of disk sizes and orientation angles. This is reflected in the lack of convergence in the orbit angles as well.

Table 5.10: Identification Information and Merger Zoo summary for SDSS 587732136993882121.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587732136993882121	Arp 214, NGC 3718/3729	11:32:35.4	+53:04:00.0	0.003312

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
26424	25706	718	91	13063	4319	8744

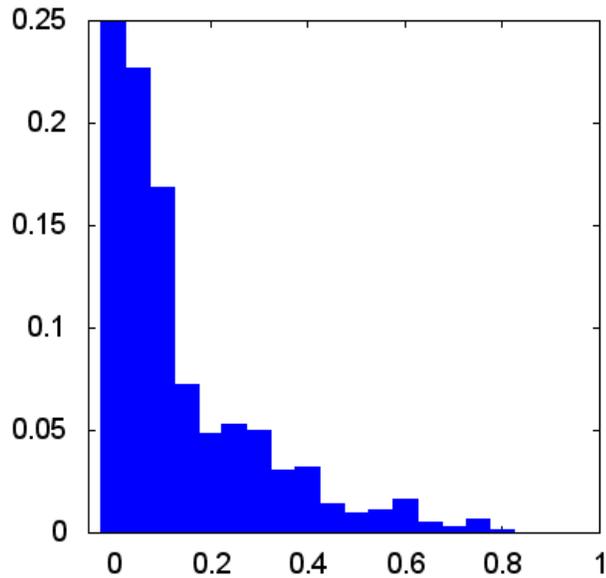


Figure 5.38: Relative frequency of fitness for all selected states of SDSS 587732136993882121



Figure 5.39: Target image and top 3 simulations for SDSS 587732136993882121

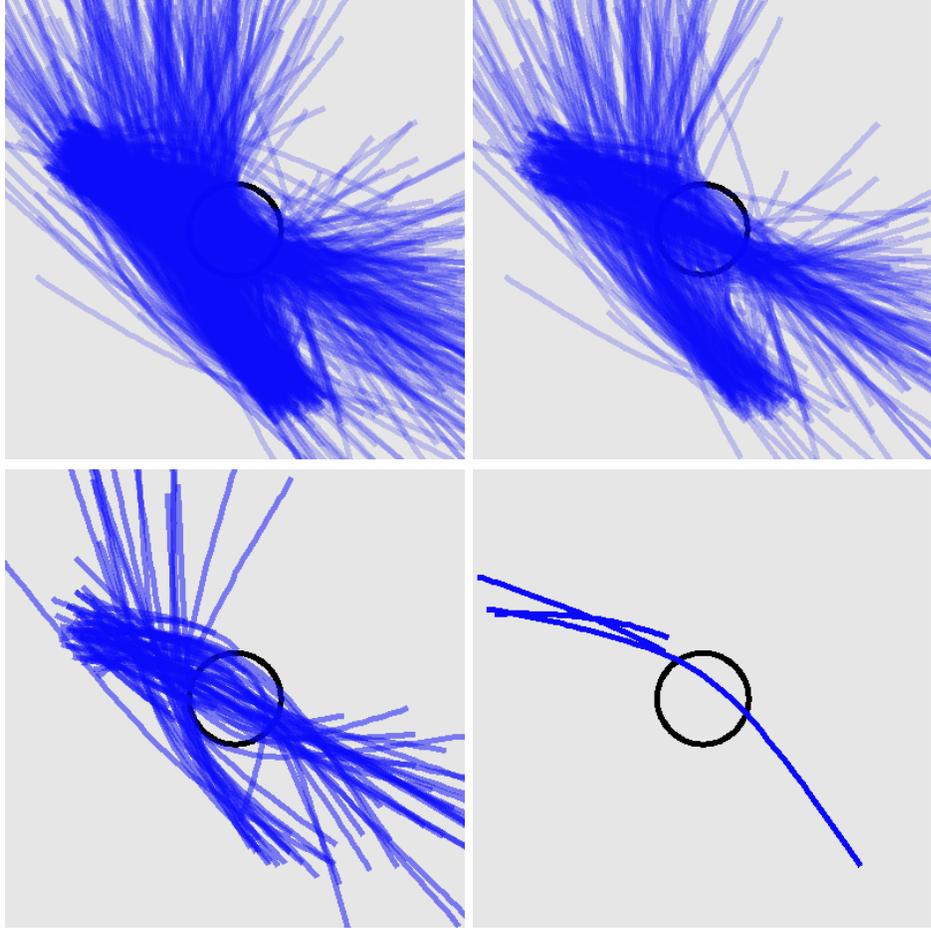


Figure 5.40: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587732136993882121

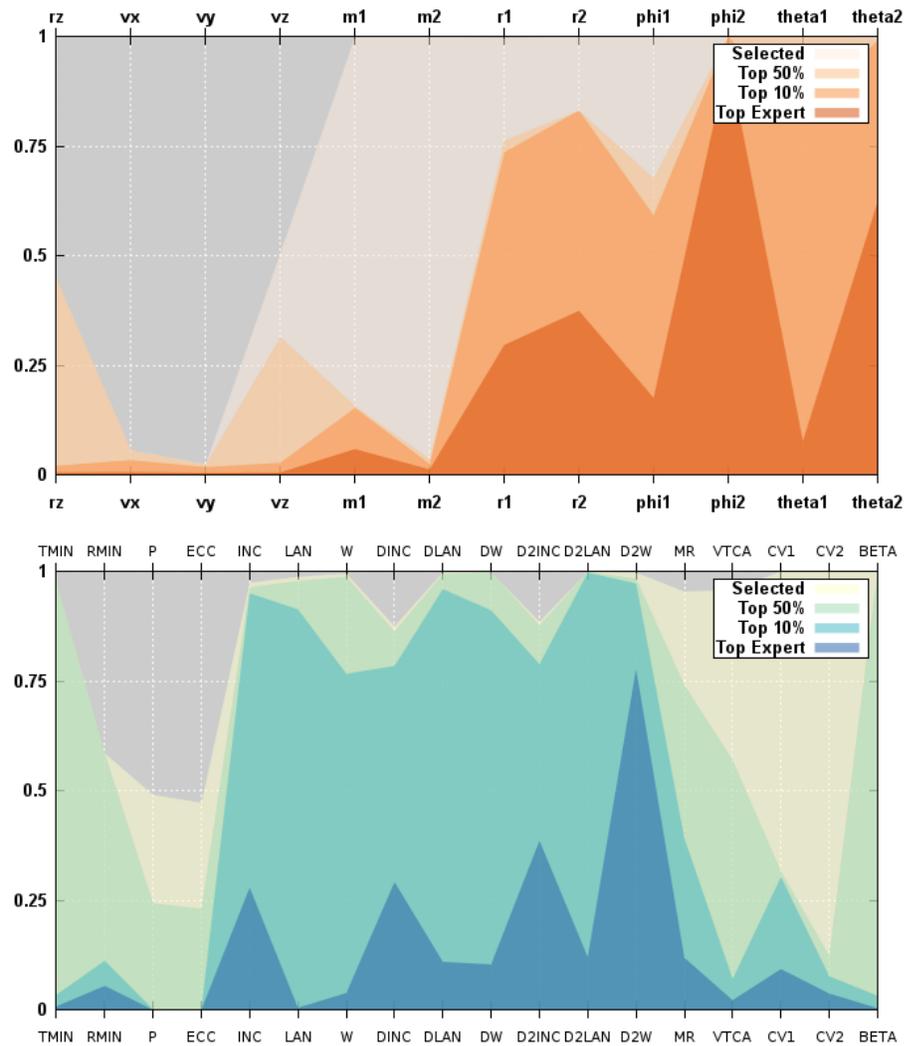


Figure 5.41: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587732136993882121

### 5.1.11 SDSS 587732772130652231

NGC 4320 has a declining fitness profile. There were relatively few simulations selected, 594, out of  $\sim 80000$  that were viewed by volunteers. The best simulation is a reasonably good recreation of the merger, especially considering it is difficult to identify two separate galaxies. The southern tail is not quite correct, but the general asymmetry displayed is a match. The top trajectories still has some diversity. The least constrained simulation parameter besides the usual disk angle is the mass of the secondary galaxy. The orbit parameters are unconstrained as a result.

Table 5.11: Identification Information and Merger Zoo summary for SDSS 587732772130652231.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587732772130652231	NGC 4320, UGC 7452	12:22:57.6	+10:32:52.7	0.026675

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
79931	79337	594	163	10736	3598	7138

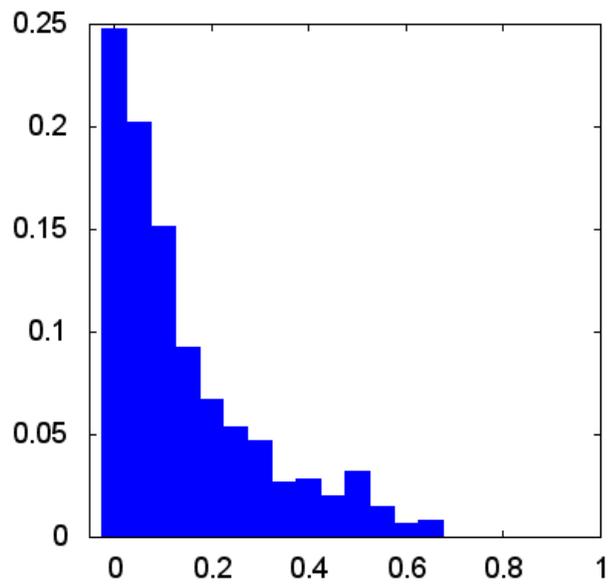


Figure 5.42: Relative frequency of fitness for all selected states of SDSS 587732772130652231

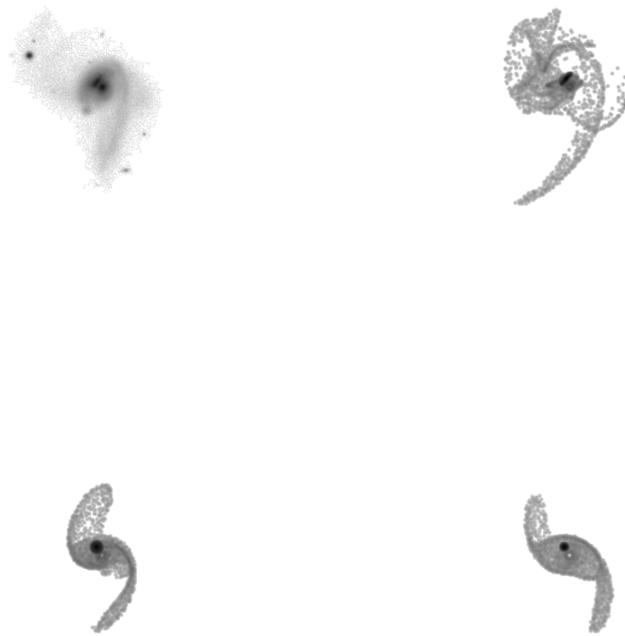


Figure 5.43: Target image and top 3 simulations for SDSS 587732772130652231

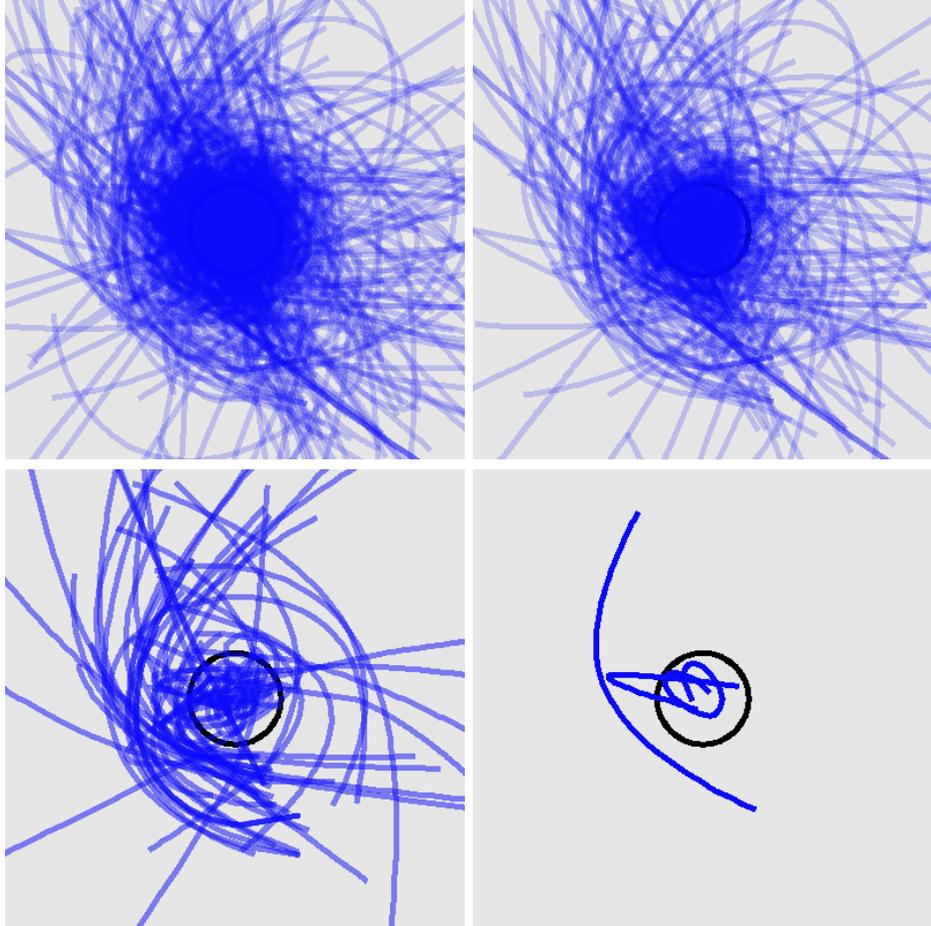


Figure 5.44: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587732772130652231

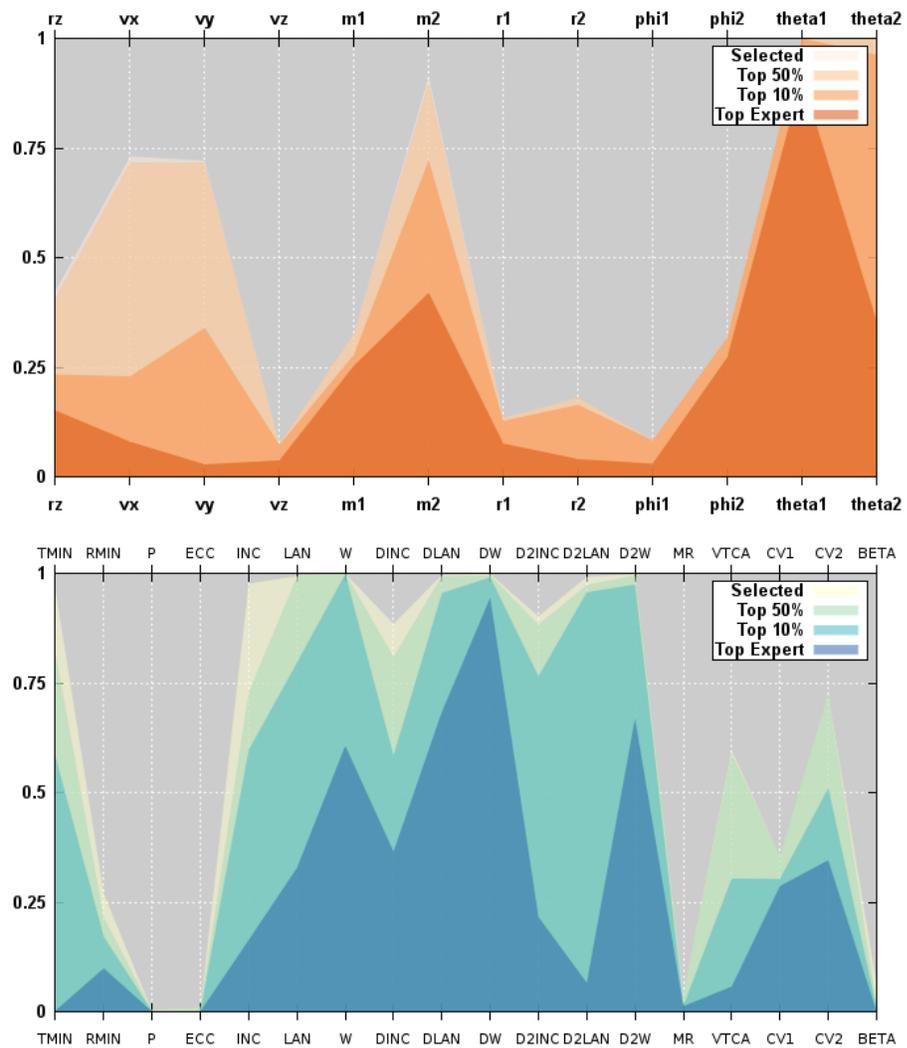


Figure 5.45: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587732772130652231

### 5.1.12 SDSS 587733080814583863

NGC 4669 has a number of simulations with fitnesses above 0.6. The broad, asymmetric tidal tail of the primary is recreated in the simulations. The top trajectories are similar. The masses, disk sizes, and orientation angles are not well constrained. The mass ratio is constrained even though the individual masses retain some variance.

Table 5.12: Identification Information and Merger Zoo summary for SDSS 587733080814583863.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587733080814583863	NGC 4669, UGC 7905	12:43:49.4	+54:54:16.4	0.016261

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
25603	24999	604	124	11039	4682	6357

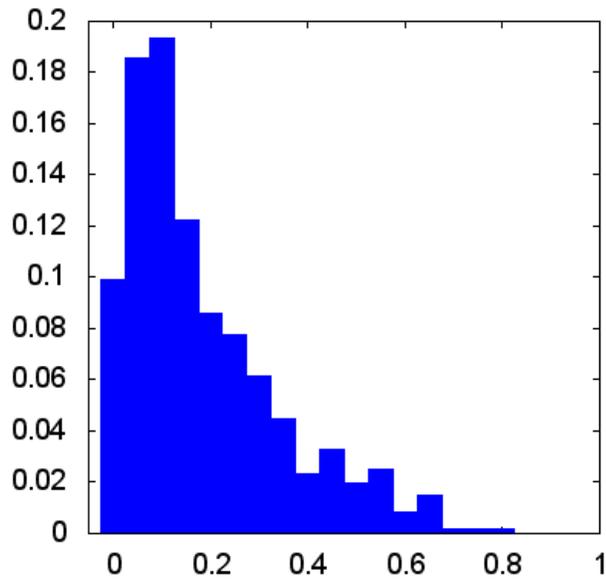


Figure 5.46: Relative frequency of fitness for all selected states of SDSS 587733080814583863

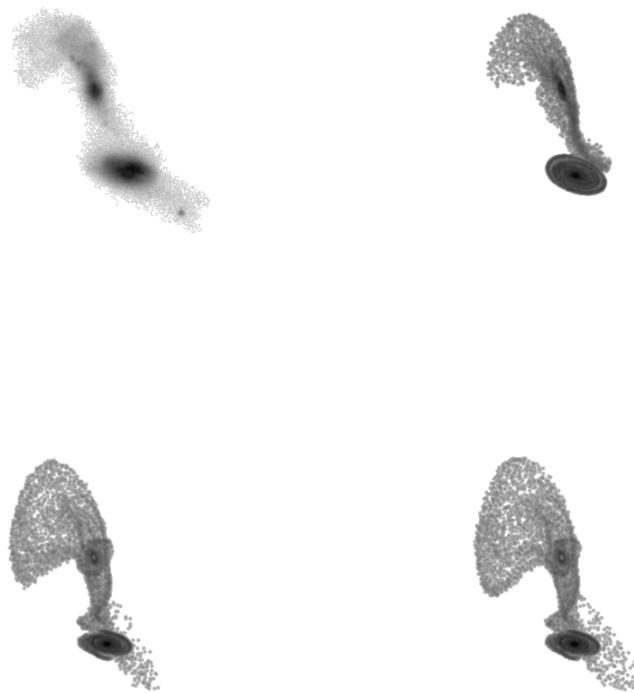


Figure 5.47: Target image and top 3 simulations for SDSS 587733080814583863

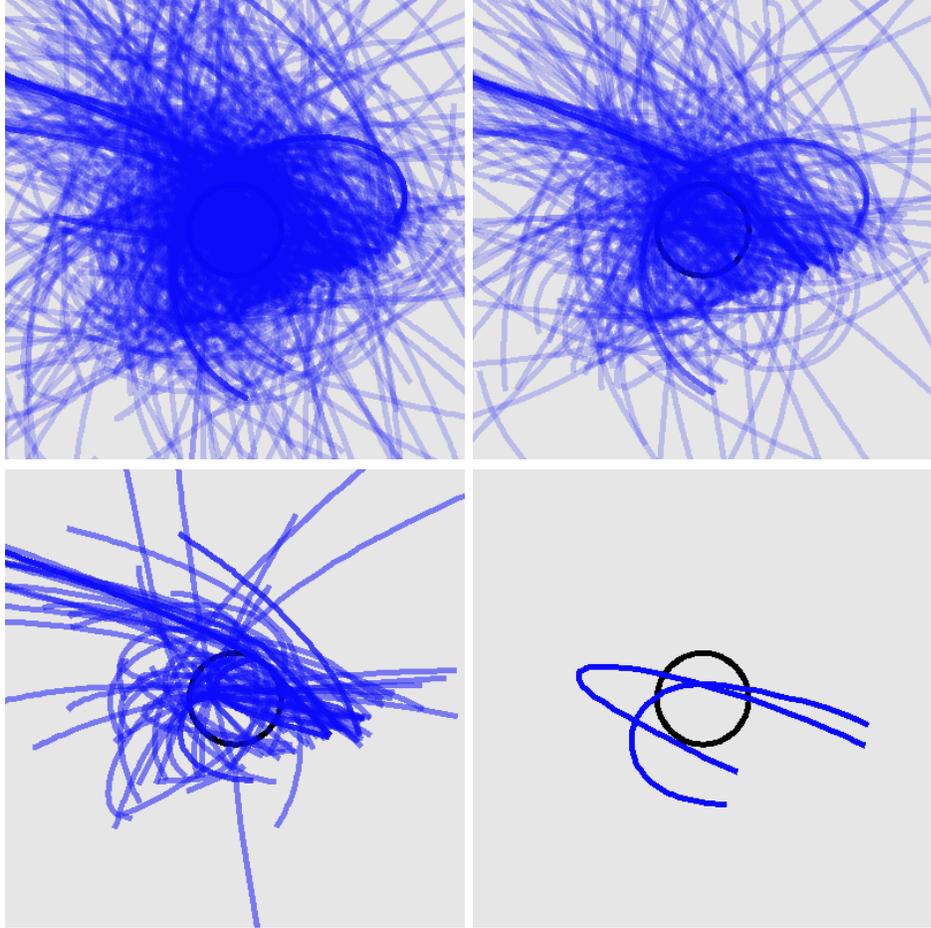


Figure 5.48: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587733080814583863

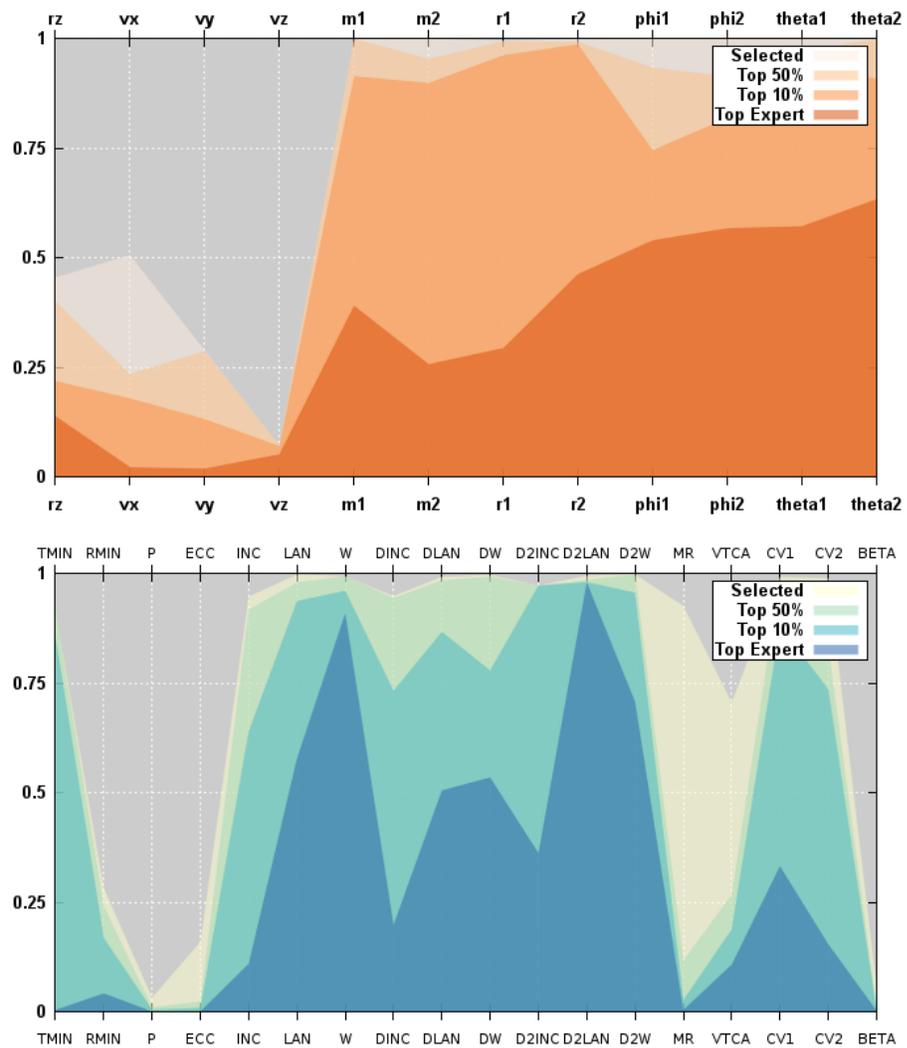


Figure 5.49: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587733080814583863

### 5.1.13 SDSS 587734862680752822

Arp 255 has a number of medium and high fitness simulations. The simulations show accurate reconstructions of tidal features for both the primary and the secondary galaxies. The top trajectories are very similar. The masses and disk parameters appear to not be well constrained. However, the non angular orbit parameters are converged.

Table 5.13: Identification Information and Merger Zoo summary for SDSS 587734862680752822.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587734862680752822	Arp 255, UGC 5304	9:53:08.8	+7:51:58.2	0.040858

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
37782	37270	512	143	9247	4606	4641

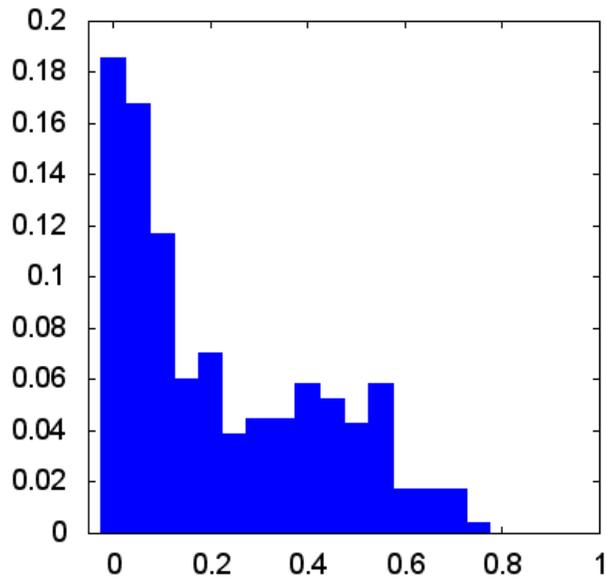


Figure 5.50: Relative frequency of fitness for all selected states of SDSS 587734862680752822

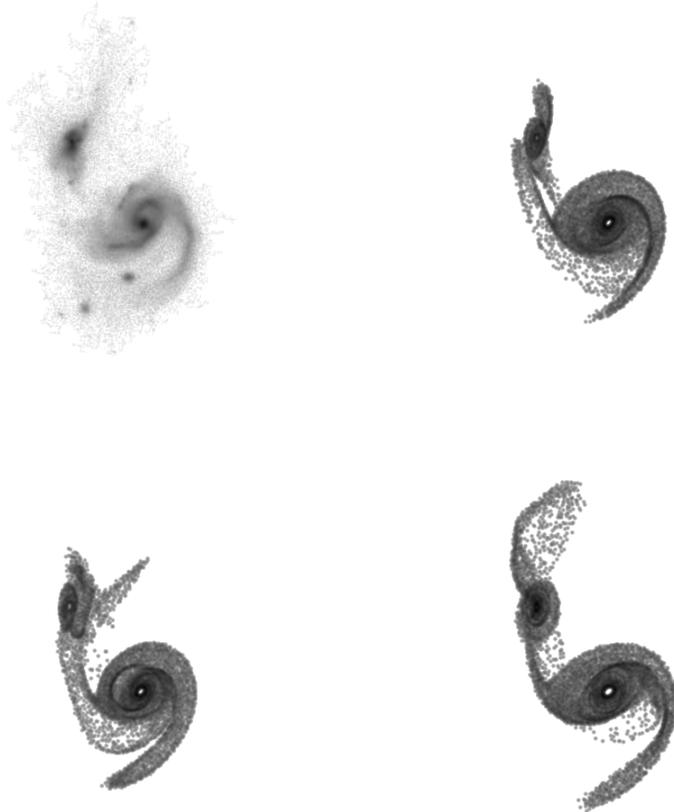


Figure 5.51: Target image and top 3 simulations for SDSS 587734862680752822

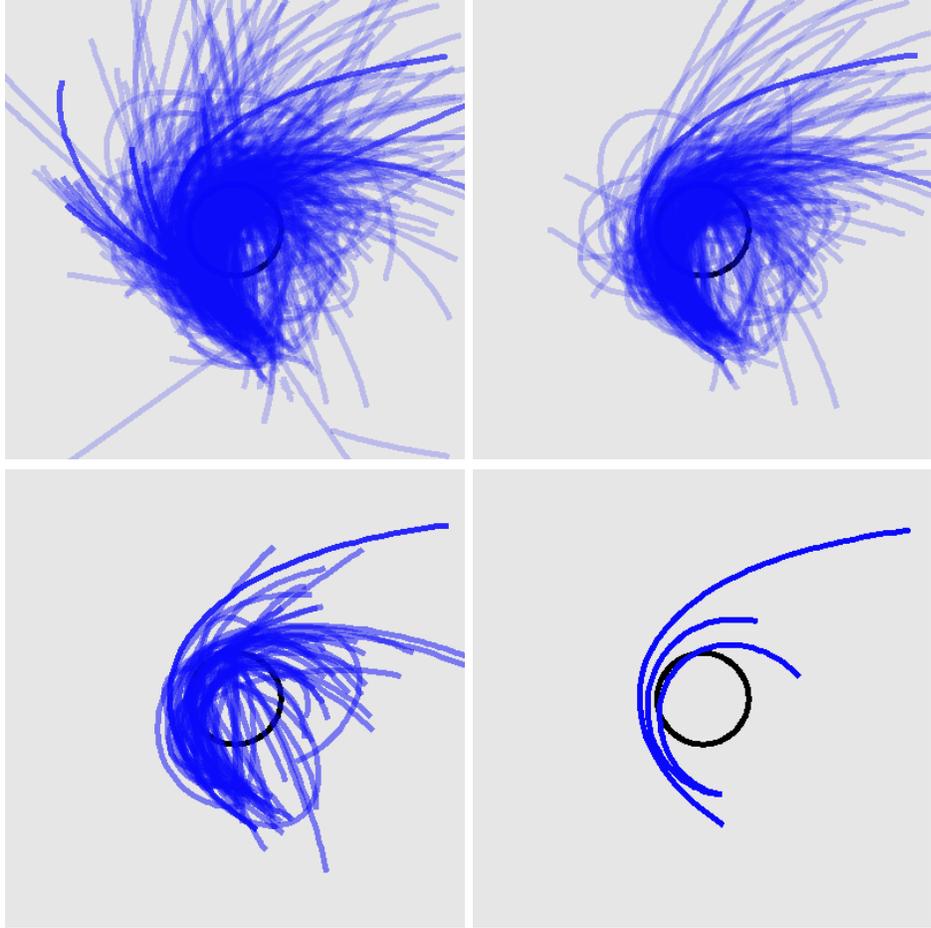


Figure 5.52: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587734862680752822

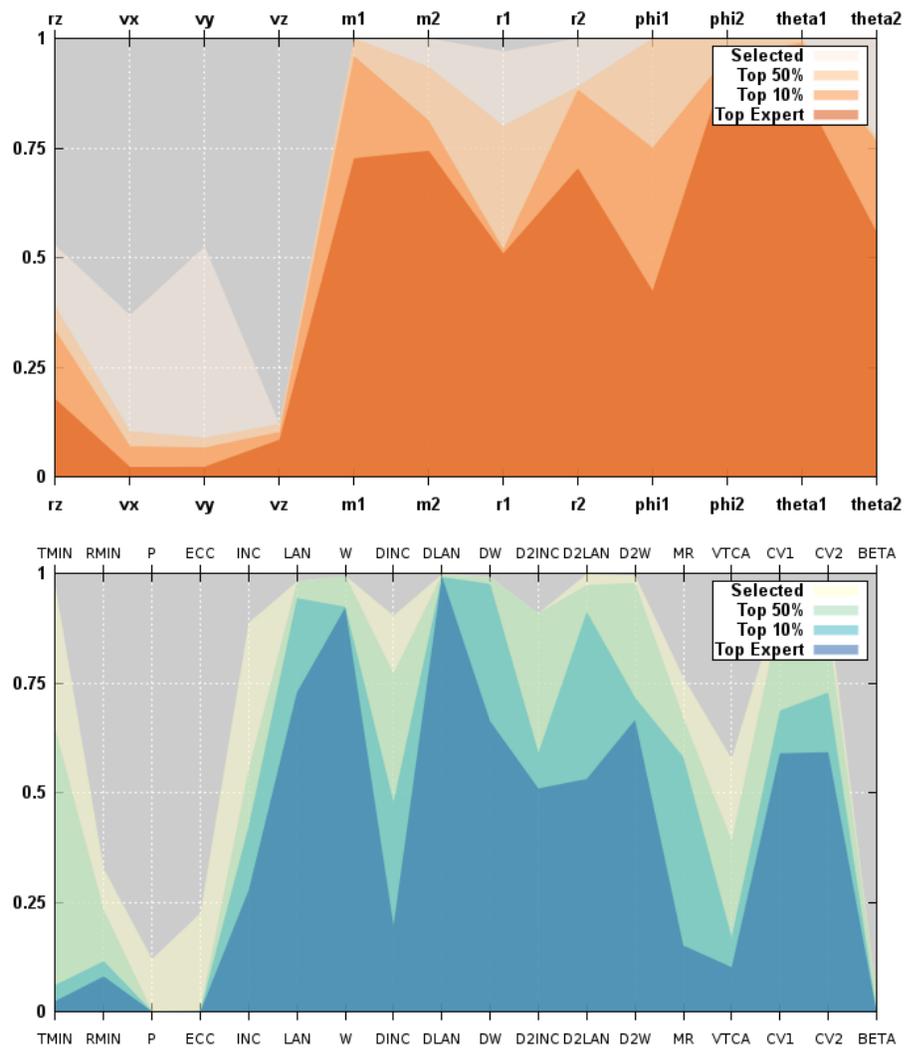


Figure 5.53: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587734862680752822

### 5.1.14 SDSS 587735043609329845

Arp 82 was another popular target for Merger Zoo with over 150000 simulations viewed by the volunteers. They identified simulations with fitnesses above 0.8. The long tidal tails of the primary are well matched even if the length and orientation of the northern tail is not a perfect recreation. The trajectories are similar showing a bound orbit. The simulation parameters are well constrained except for mass of the secondary and the disk inclination angles. The orbit parameters match the expected behavior.

Table 5.14: Identification Information and Merger Zoo summary for SDSS 587735043609329845.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587735043609329845	Arp 82, NGC 2535/2536	8:11:13.5	+25:12:23.8	0.013666

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
152318	149410	2908	457	44774	19761	25013

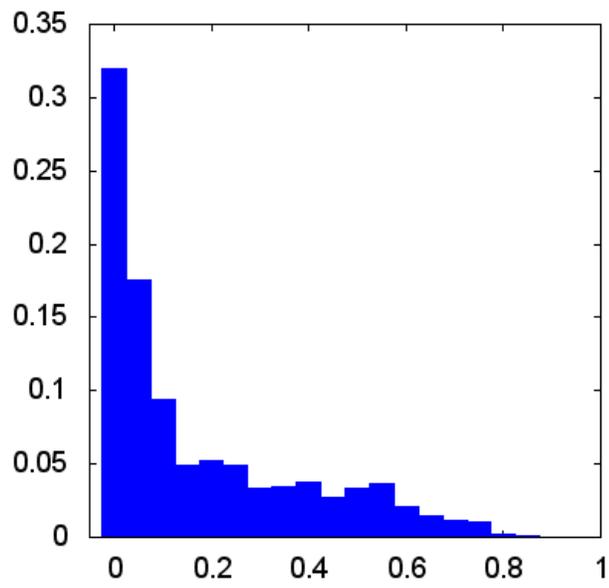


Figure 5.54: Relative frequency of fitness for all selected states of SDSS 587735043609329845



Figure 5.55: Target image and top 3 simulations for SDSS 587735043609329845

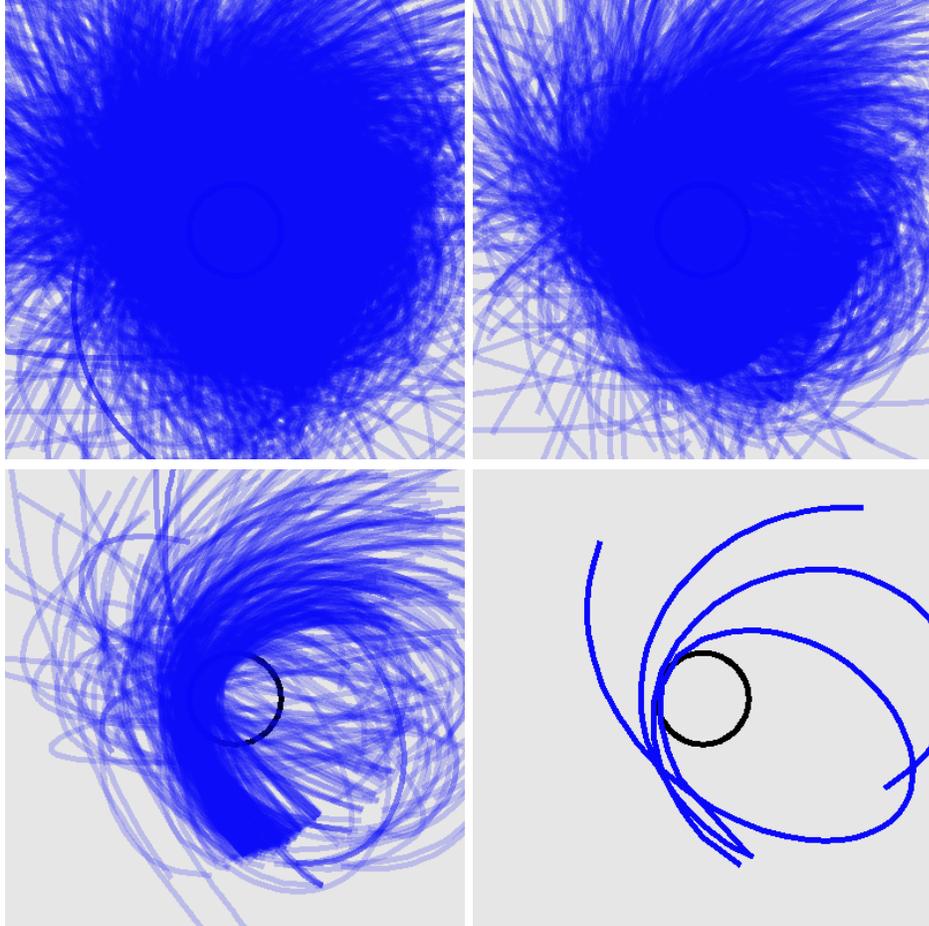


Figure 5.56: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587735043609329845

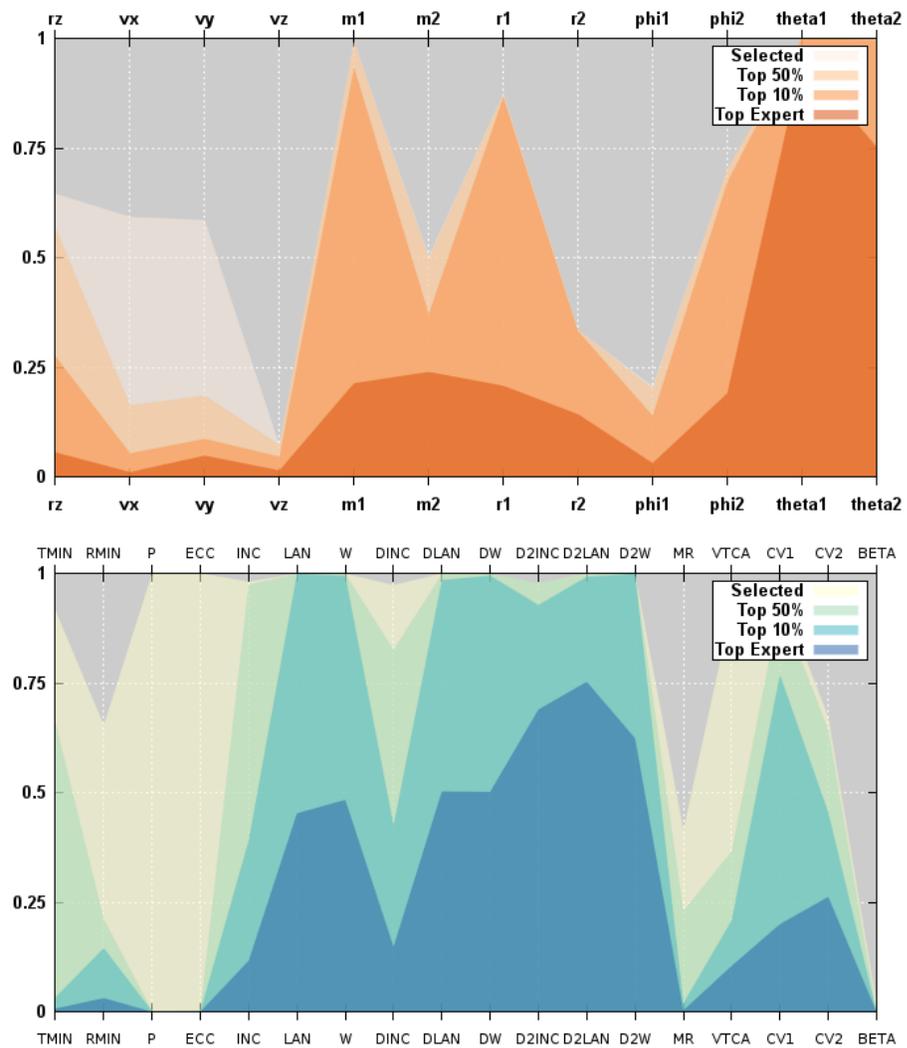


Figure 5.57: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587735043609329845

### 5.1.15 SDSS 587735665840881790

Volunteers identified a few high-fitness simulations for Arp 239. The best simulation does not quite connect the tidal features of the primary and secondary, but the other top simulations do a better job of that. The top trajectories are similar. The mass of the secondary is not well constrained causing uncertainty in derived orbit parameters as well.

Table 5.15: Identification Information and Merger Zoo summary for SDSS 587735665840881790.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587735665840881790	Arp 239, NGC 5278	13:41:39.3	+55:40:14.6	0.02521

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
18576	18262	314	72	5940	2289	3651

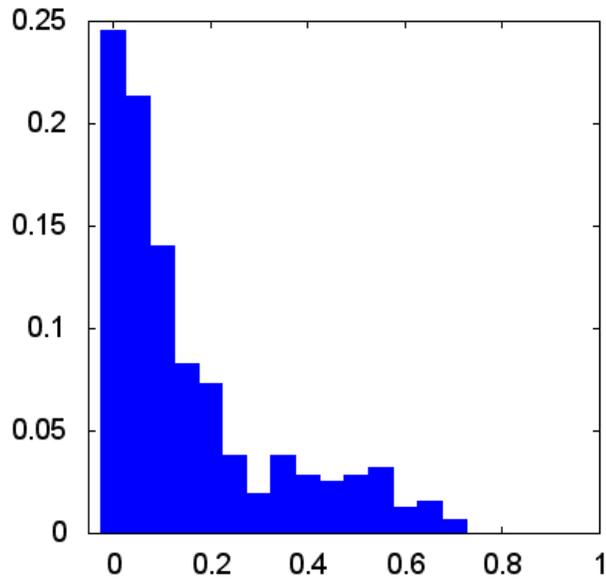


Figure 5.58: Relative frequency of fitness for all selected states of SDSS 587735665840881790

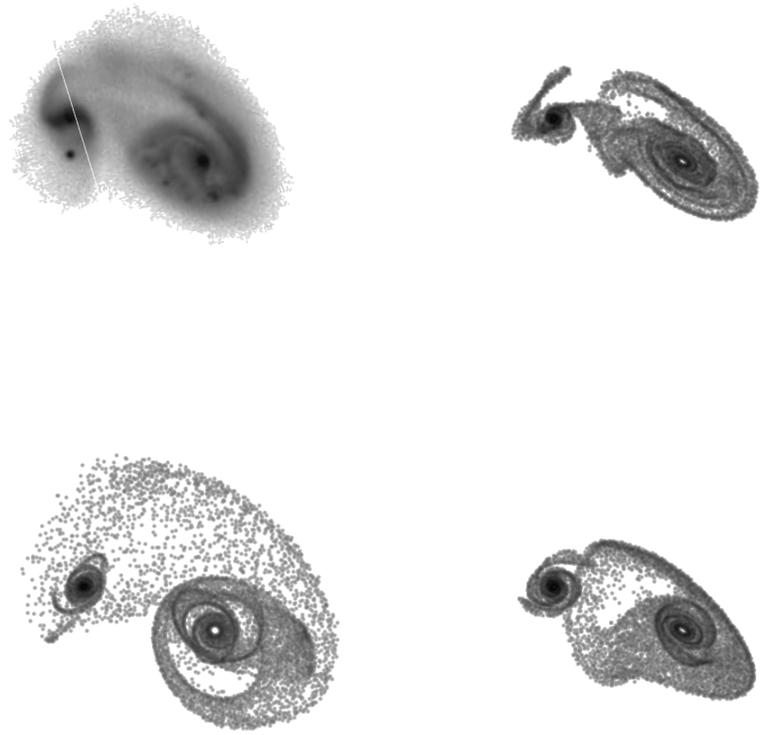


Figure 5.59: Target image and top 3 simulations for SDSS 587735665840881790

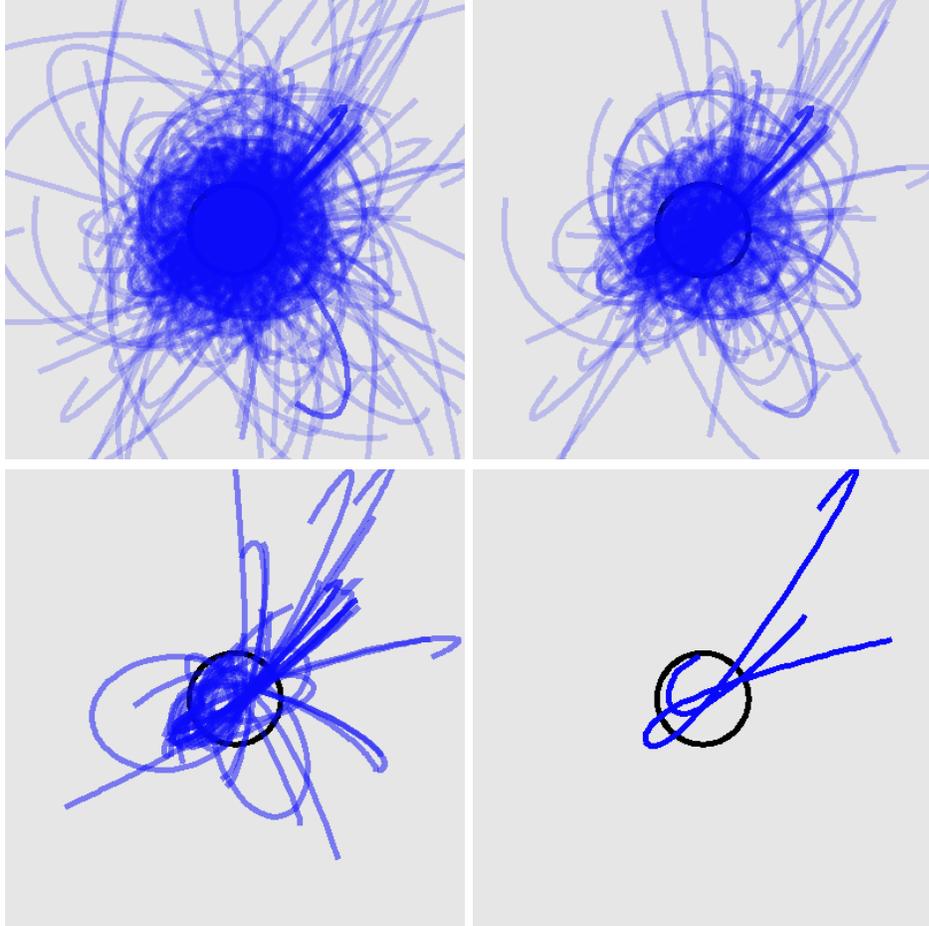


Figure 5.60: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587735665840881790

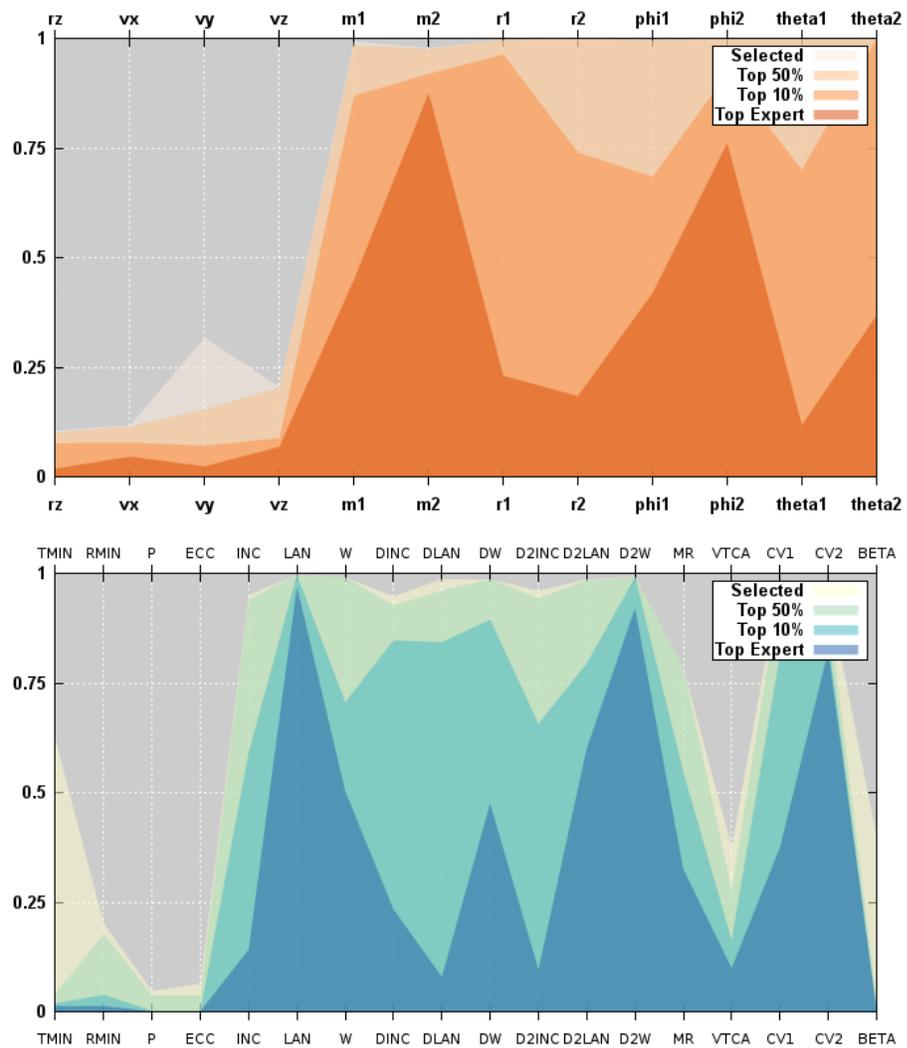


Figure 5.61: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587735665840881790

### 5.1.16 SDSS 587736941981466667

The fitness distribution for Arp 199 is more uniform than for other targets. The simulations show that there are no strong tidal features for users to follow for guidance. There is some density enhancement in the inclined galaxy disk that matches the target image. The top trajectories are not identical but they are relatively similar compared to the entire set of trajectories considered. Surprisingly, most of the simulation and orbit parameters show convergence. It may be that the lack of features caused volunteers to reject any disturbed morphologies which may have been useful for constraining the simulation.

Table 5.16: Identification Information and Merger Zoo summary for SDSS 587736941981466667.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587736941981466667	Arp 199, NGC 5544/5545	14:17:02.5	+36:34:16.6	0.01014

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
13668	12765	903	102	16210	11512	4698

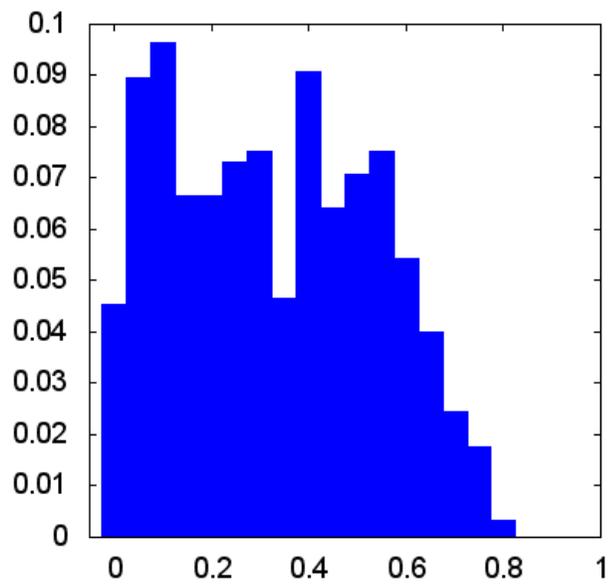


Figure 5.62: Relative frequency of fitness for all selected states of SDSS 587736941981466667

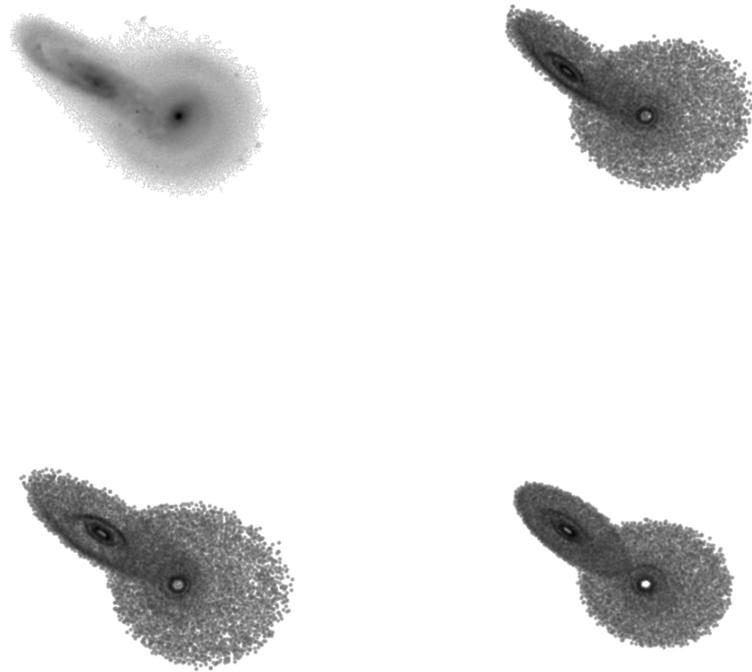


Figure 5.63: Target image and top 3 simulations for SDSS 587736941981466667

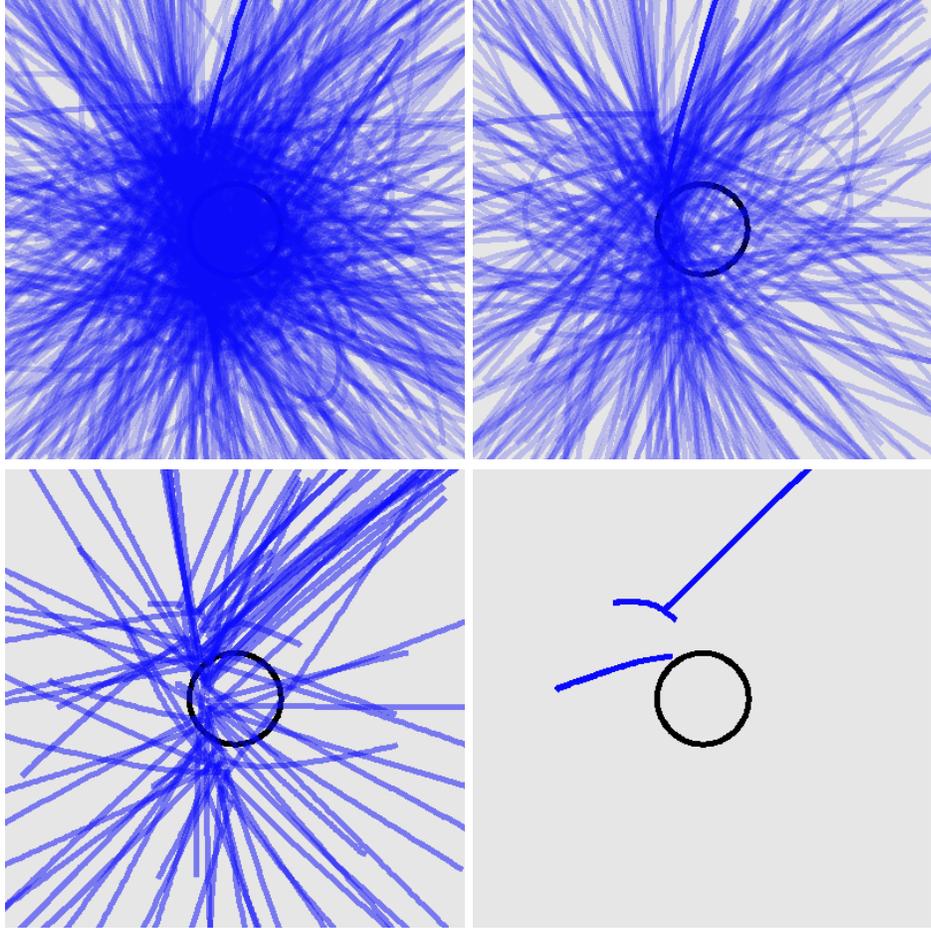


Figure 5.64: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587736941981466667

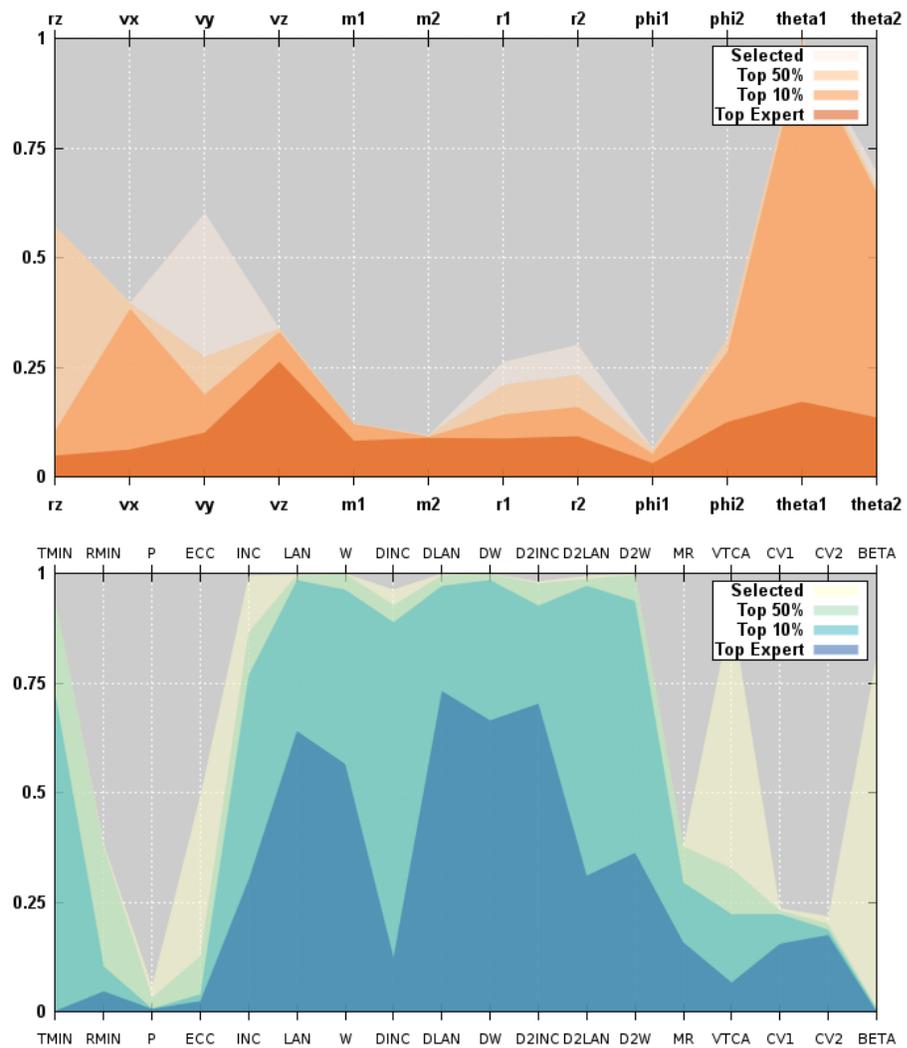


Figure 5.65: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587736941981466667

### 5.1.17 SDSS 587738569246376675

There is a steep decline in the fitness distribution of Arp 57. The simulations recreate the tidal features of both the primary and the secondary. One detail not well matched is the sharp “kink” in the northern tail of the primary. The top trajectories are similar. The small number of selected states, 231, has resulted in a situation where the expert population has worse convergence than the top 10% for the primary disk radius. There is also poor convergence for the masses and orientation angles. However, many of the orbit parameters show decent convergence.

Table 5.17: Identification Information and Merger Zoo summary for SDSS 587738569246376675.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587738569246376675	Arp 57	13:16:47.4	+14:25:39.6	0.060208

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
20418	20187	231	129	4453	1822	2631

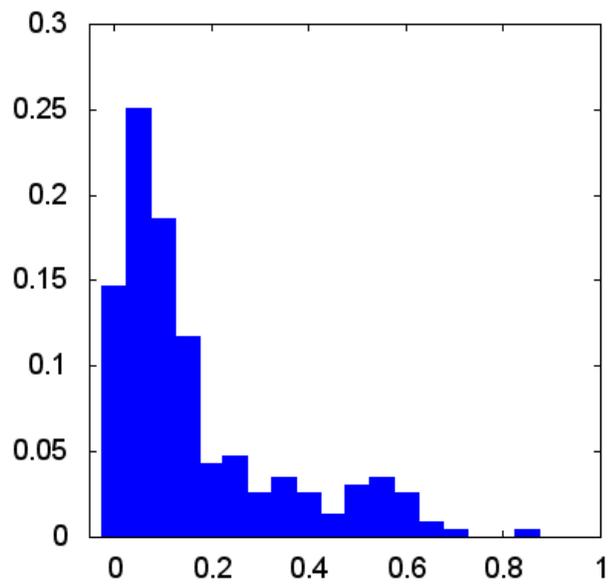


Figure 5.66: Relative frequency of fitness for all selected states of SDSS 587738569246376675

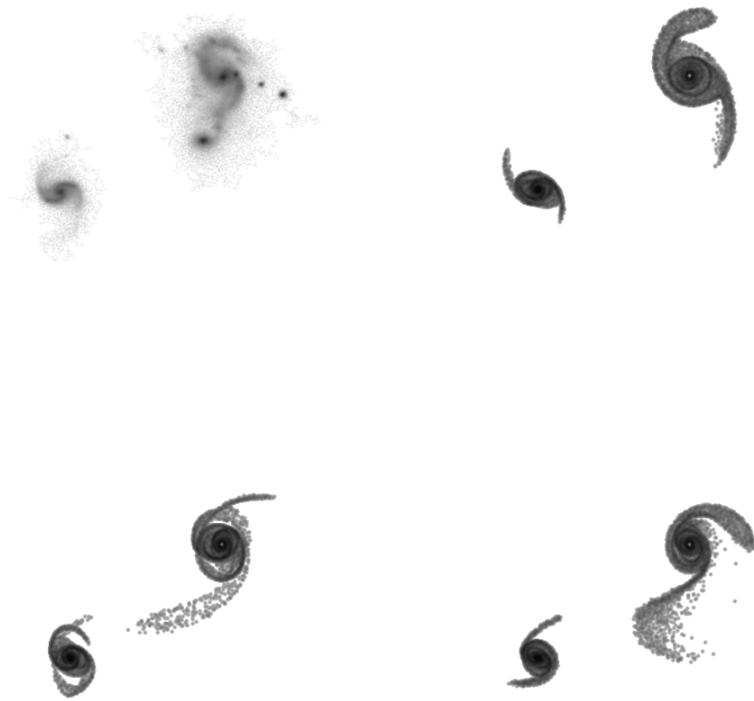


Figure 5.67: Target image and top 3 simulations for SDSS 587738569246376675

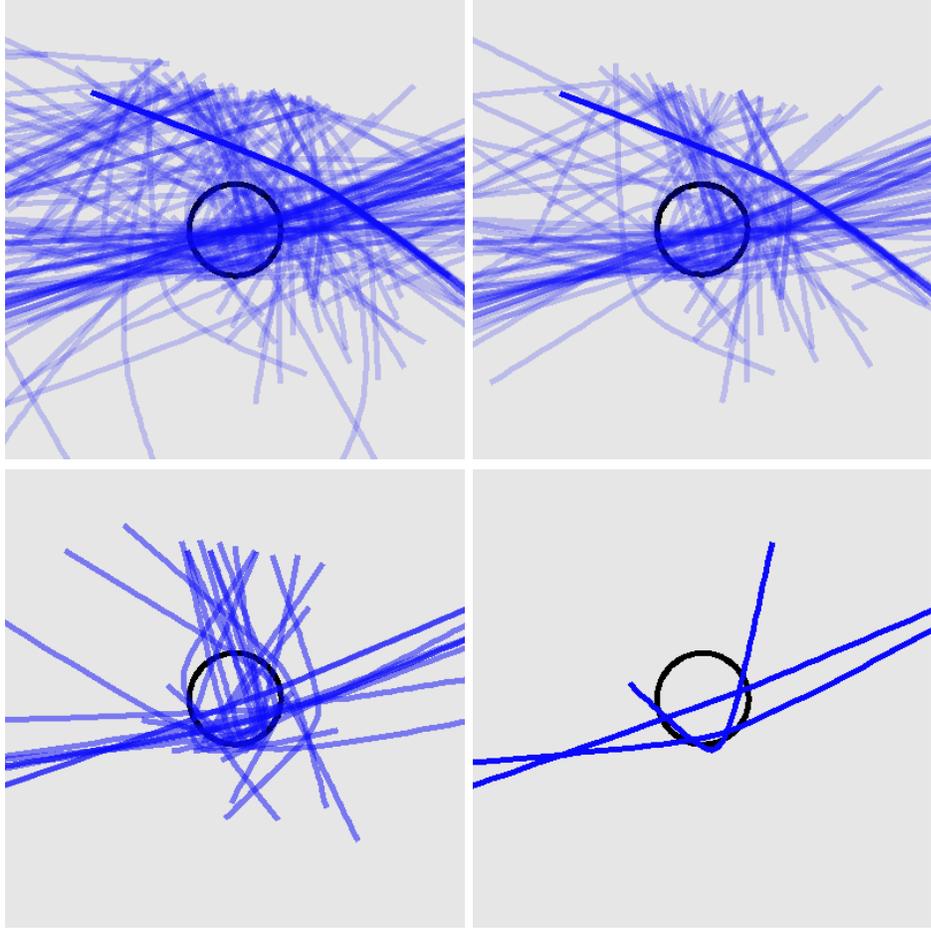


Figure 5.68: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587738569246376675

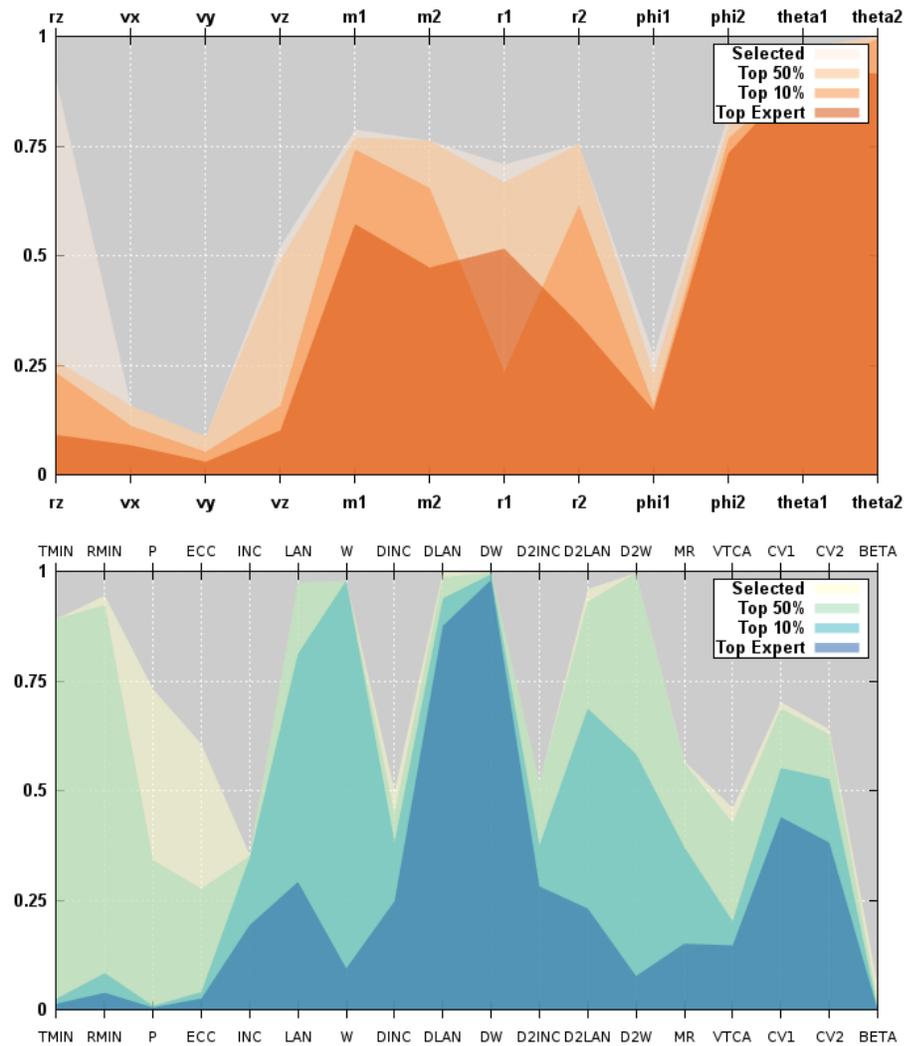


Figure 5.69: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587738569246376675

### 5.1.18 SDSS 587738569249390718

This target has a lopsided fitness distribution. The top simulations recreate the tidal tails of the primary. The trajectories are similar to one another. The simulation parameters are well converged with the exception of the inclination angle for the secondary disk. It is possible that the secondary galaxy in this pair is actually elliptical.

Table 5.18: Identification Information and Merger Zoo summary for SDSS 587738569249390718.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587738569249390718	2MASS 13445034+1355178	13:44:50.3	+13:55:16.9	0.082453

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
21792	21390	402	161	7453	3161	4292

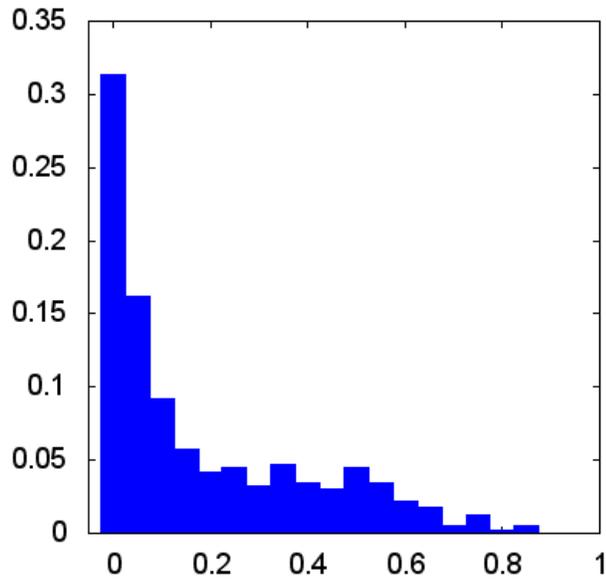


Figure 5.70: Relative frequency of fitness for all selected states of SDSS 587738569249390718

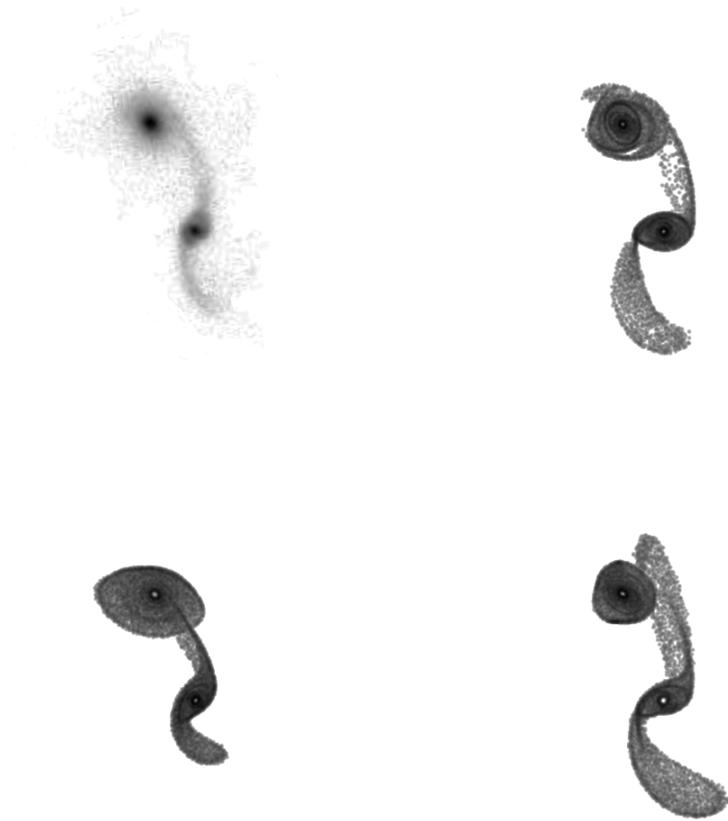


Figure 5.71: Target image and top 3 simulations for SDSS 587738569249390718

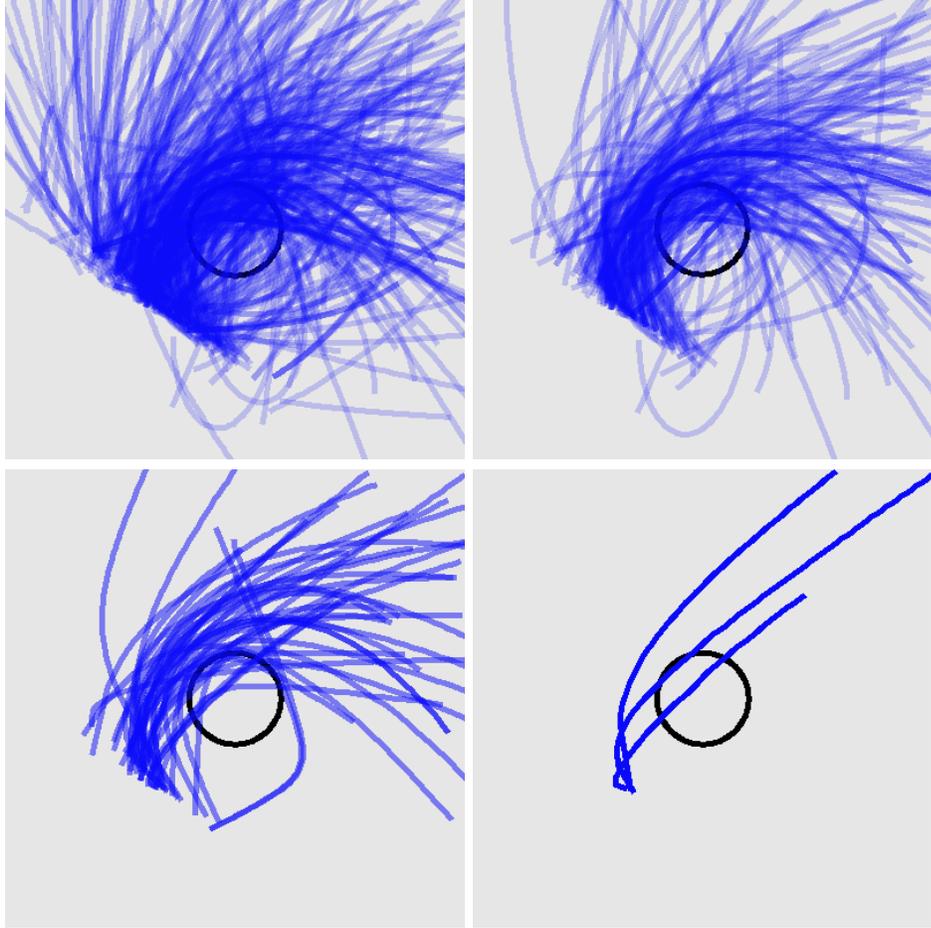


Figure 5.72: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587738569249390718

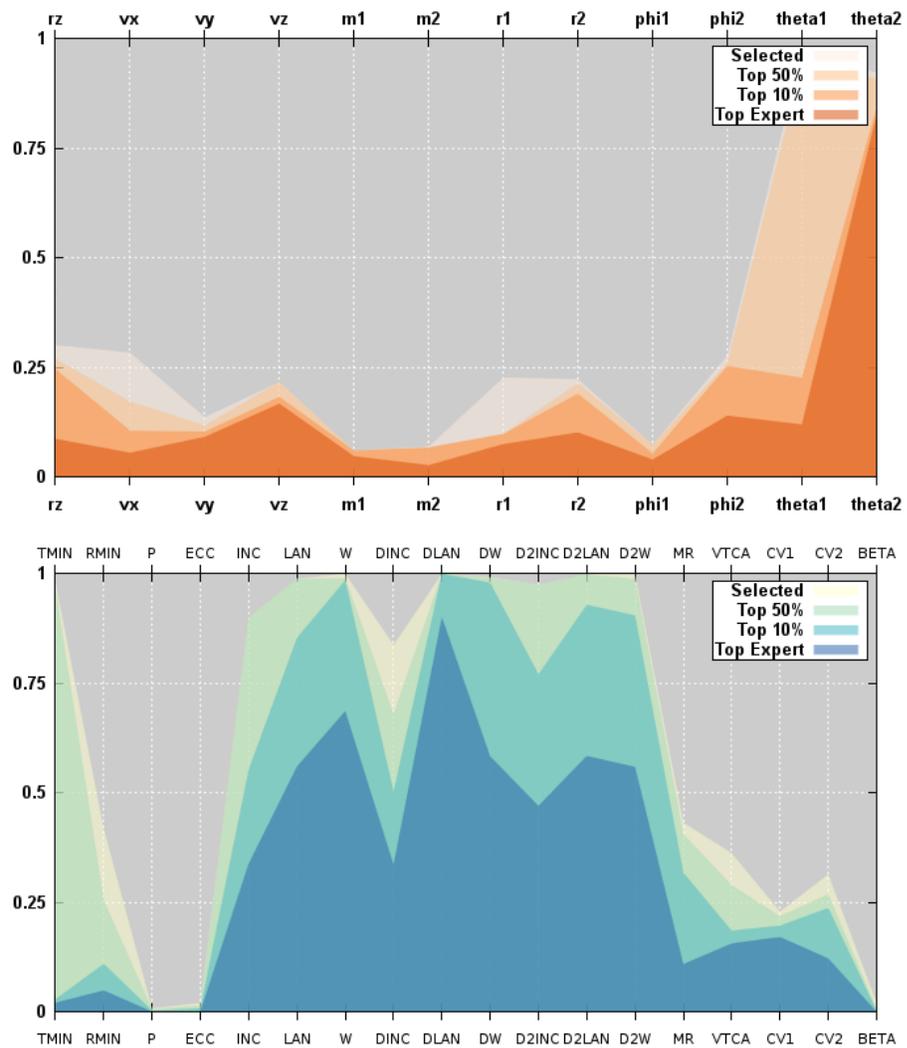


Figure 5.73: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587738569249390718

### 5.1.19 SDSS 587739153356095531

Arp 247 has few high fitness simulations. The best ones are not all able to recreate the presence of connecting material for the two galaxies. The trajectories are of a similar shape by they are not oriented the same way. The simulation parameters show good convergence except for the mass of the secondary. However the orbit parameters are not well converged.

Table 5.19: Identification Information and Merger Zoo summary for SDSS 587739153356095531.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587739153356095531	Arp 247, IC 2338/2339	8:23:34.0	+21:20:50.3	0.018079

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
36286	35186	1100	193	19460	8528	10932

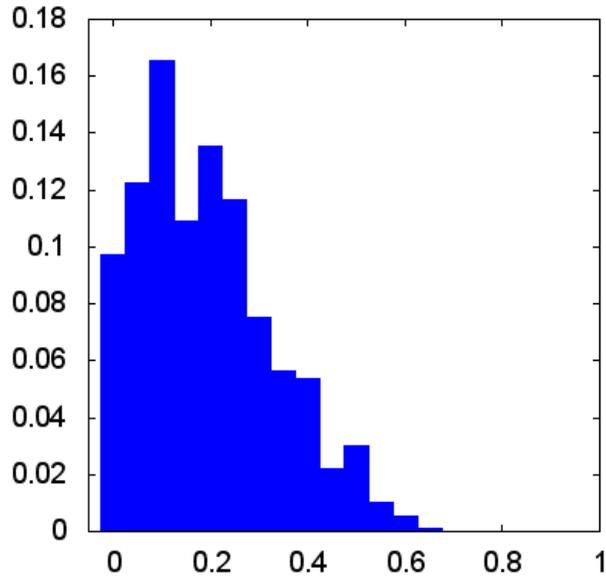


Figure 5.74: Relative frequency of fitness for all selected states of SDSS 587739153356095531

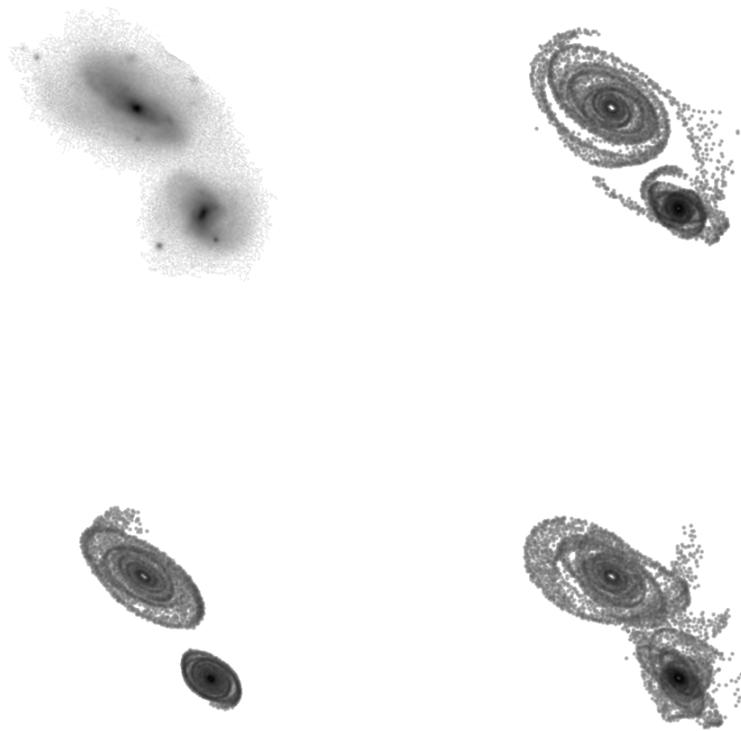


Figure 5.75: Target image and top 3 simulations for SDSS 587739153356095531

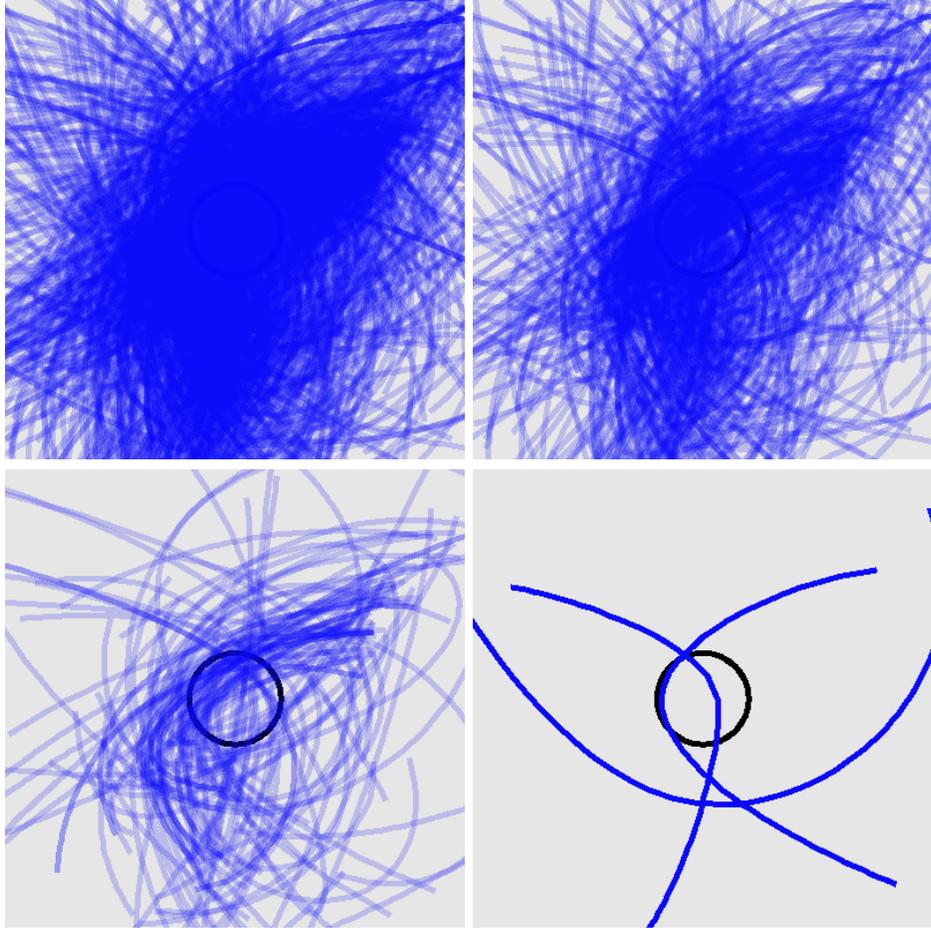


Figure 5.76: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587739153356095531

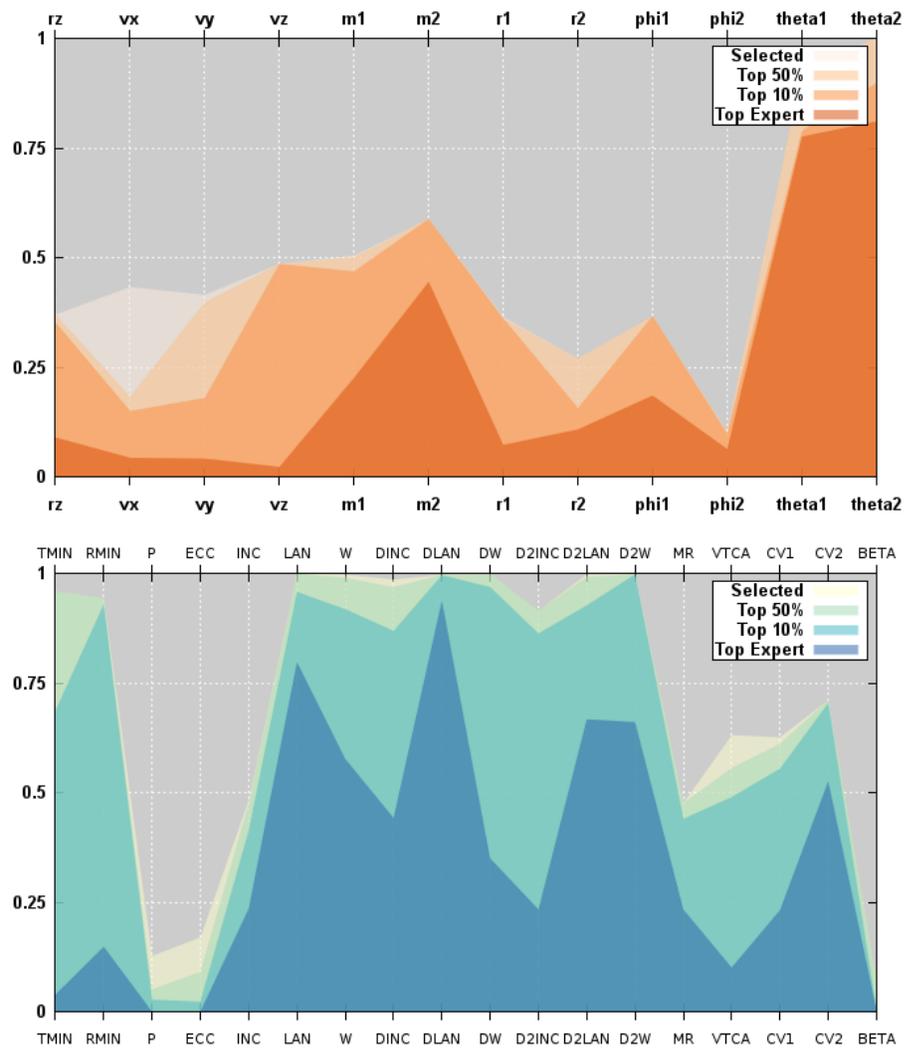


Figure 5.77: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587739153356095531

### 5.1.20 SDSS 587739407868690486

The fitness for Arp 241 shows a sharp decline. However, the volunteers were very successful in recreating the tails, one each, for the primary and secondary galaxies. The top 3 trajectories are similar. There is an appearance of a family of trajectories in the bottom left trajectory panel for the top 10% of simulations. The disk masses, radii, and angles were not well constrained.

Table 5.20: Identification Information and Merger Zoo summary for SDSS 587739407868690486.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587739407868690486	Arp 241, NGC 5699	14:37:50.4	+30:28:59.5	0.034521

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
42380	41944	436	187	8001	3184	4817

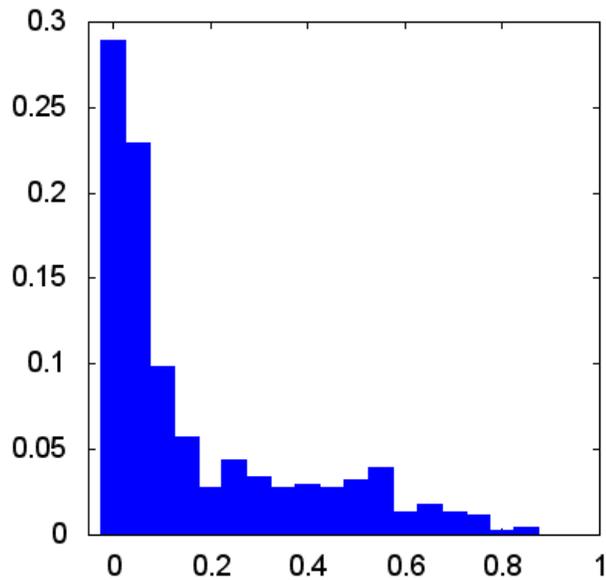


Figure 5.78: Relative frequency of fitness for all selected states of SDSS 587739407868690486

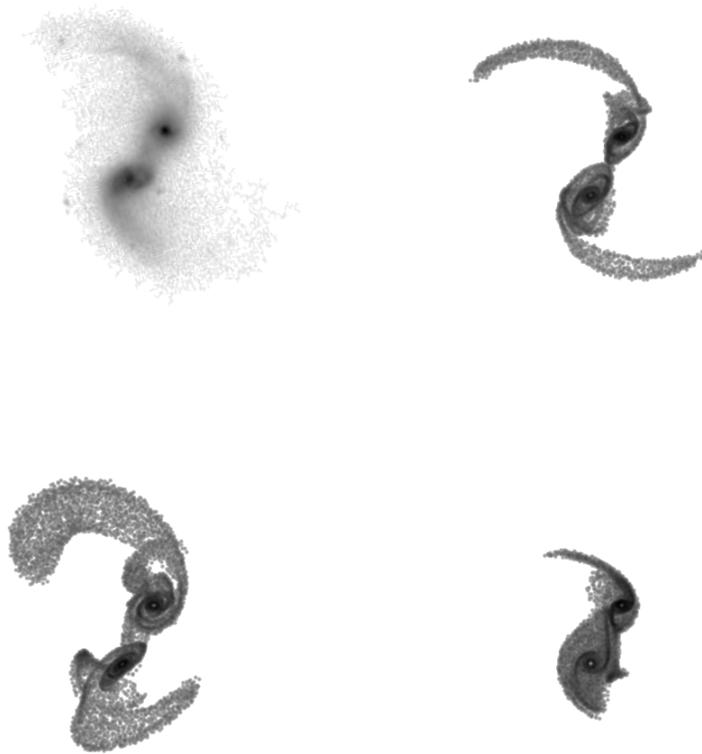


Figure 5.79: Target image and top 3 simulations for SDSS 587739407868690486

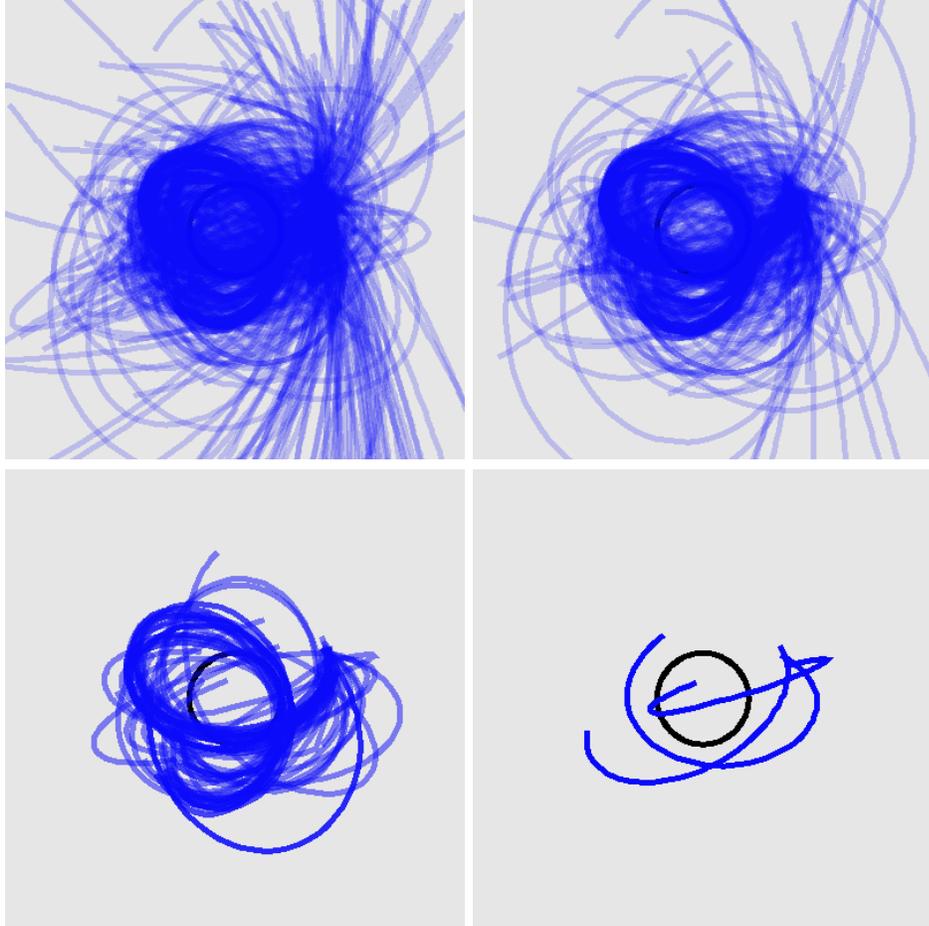


Figure 5.80: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587739407868690486

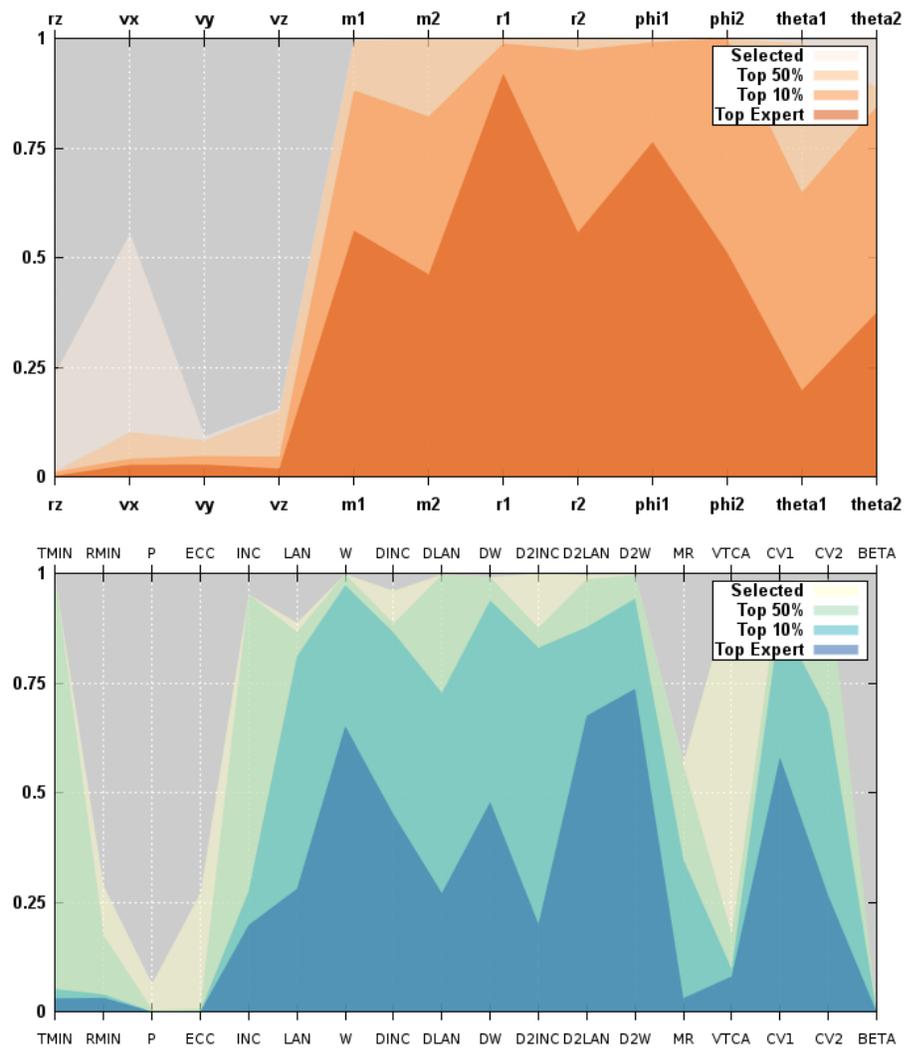


Figure 5.81: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587739407868690486

### 5.1.21 SDSS 587739505541578866

There are several high fitness simulations for Arp 313. The tidal features are reasonably matched in the simulation images. The diffuse nature of the northern tail is recreated. The trajectories are similarly parabolic. The simulation parameters are well converged except for the disk sizes and angles.

Table 5.21: Identification Information and Merger Zoo summary for SDSS 587739505541578866.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587739505541578866	Arp 313, NGC 3994/3995	11:57:36.4	+32:16:39.8	0.010854

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
26294	25947	347	56	6474	2905	3569

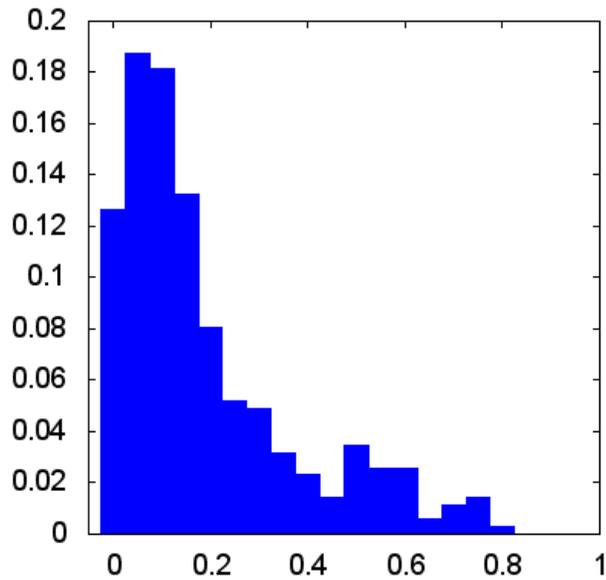


Figure 5.82: Relative frequency of fitness for all selected states of SDSS 587739505541578866

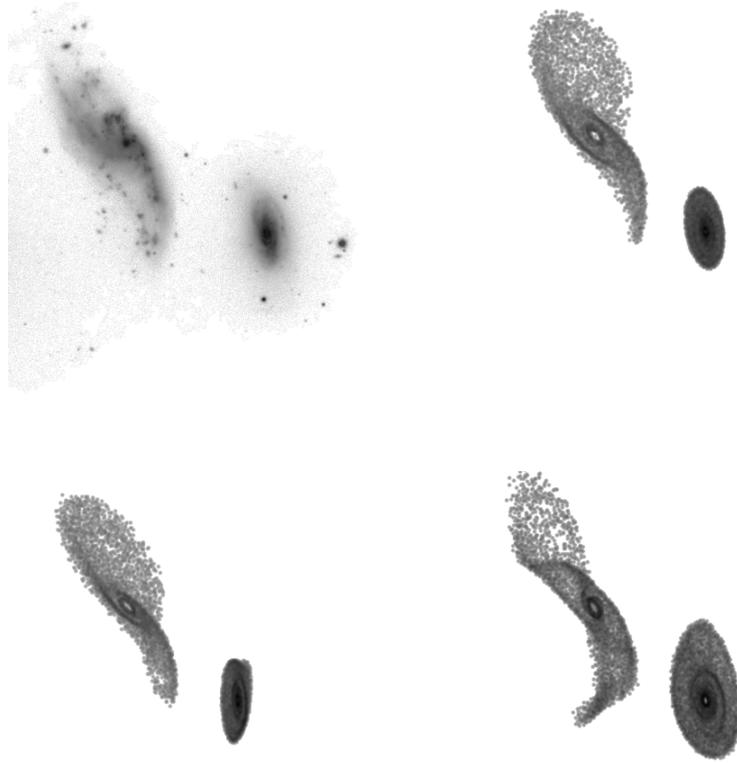


Figure 5.83: Target image and top 3 simulations for SDSS 587739505541578866

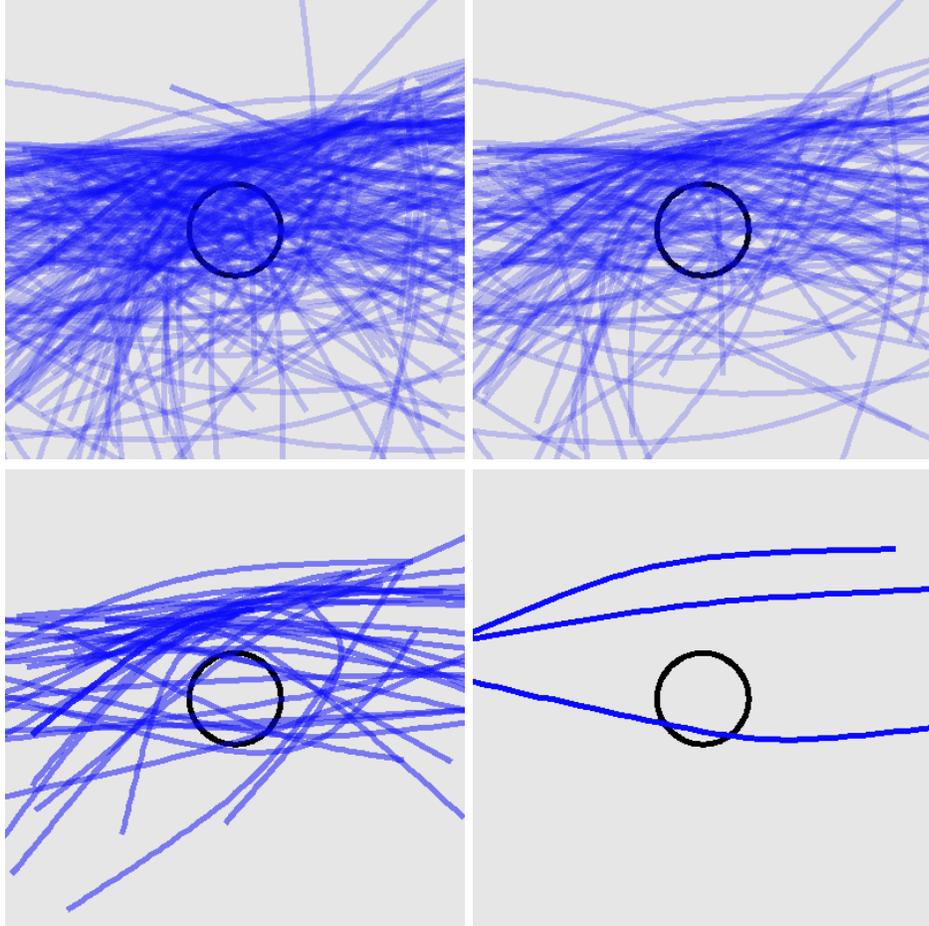


Figure 5.84: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587739505541578866

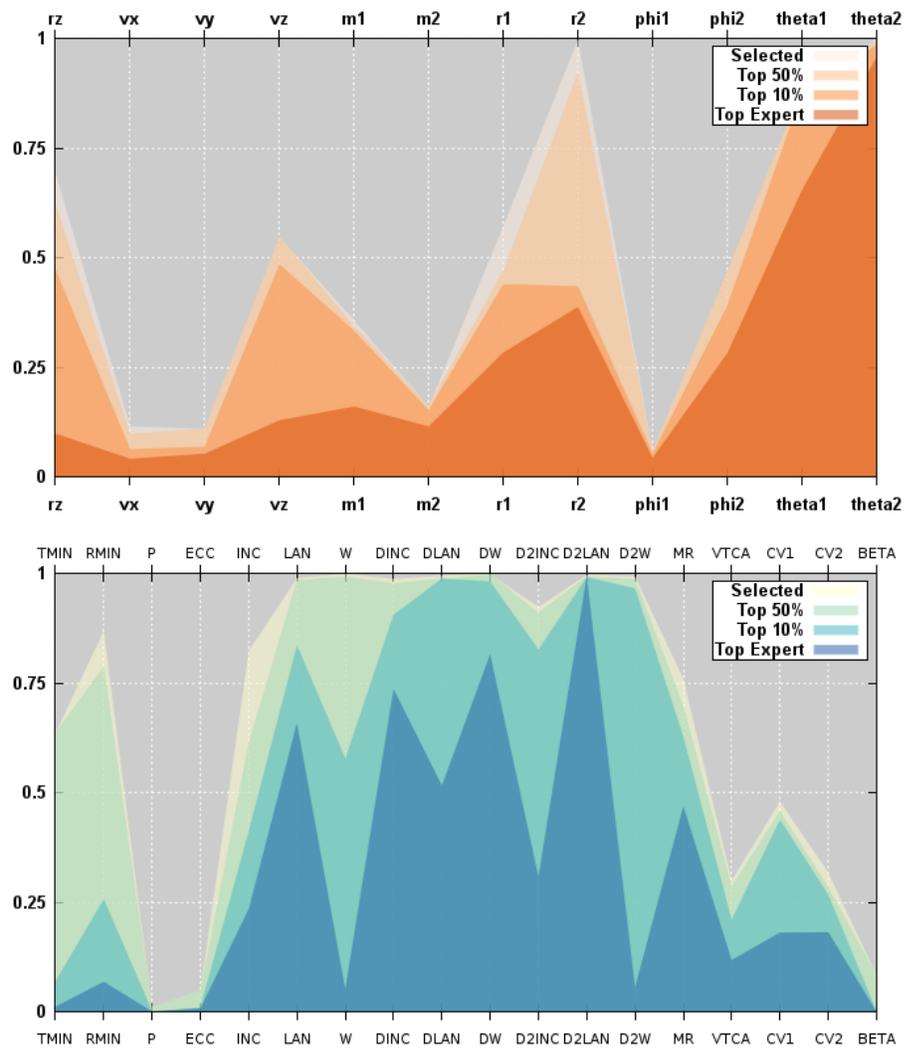


Figure 5.85: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587739505541578866

### 5.1.22 SDSS 587739646743412797

Arp 107 has a large concentration of low percentage states. The large prominent tidal tail of the primary was a popular feature for our volunteers to attempt to match. They viewed over 100000 simulations. The curvature of that tail is a good match, as well as the existence of the bridge between the two galaxies. The top trajectories are very similar. The simulation parameters are well constrained except for the orientation angles of the secondary galaxy, yet the orbit parameters still show a wide range of parameter space remaining.

Table 5.22: Identification Information and Merger Zoo summary for SDSS 587739646743412797.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587739646743412797	Arp 107, UGC 5984	10:52:14.8	+30:03:28.3	0.034597

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
106516	103925	2591	517	45337	20228	25109

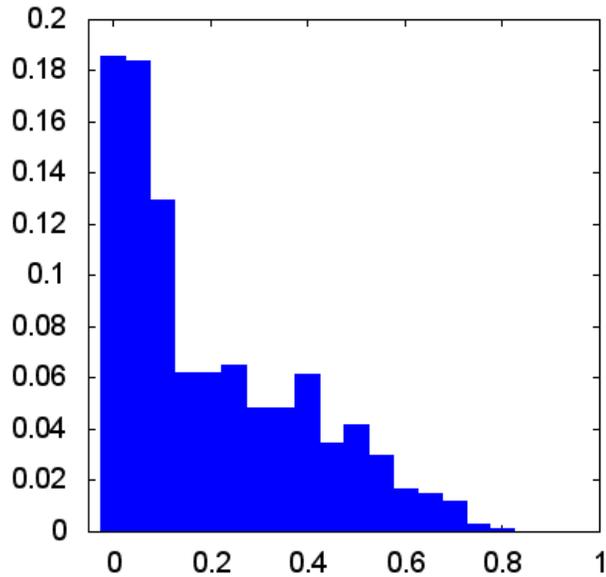


Figure 5.86: Relative frequency of fitness for all selected states of SDSS 587739646743412797

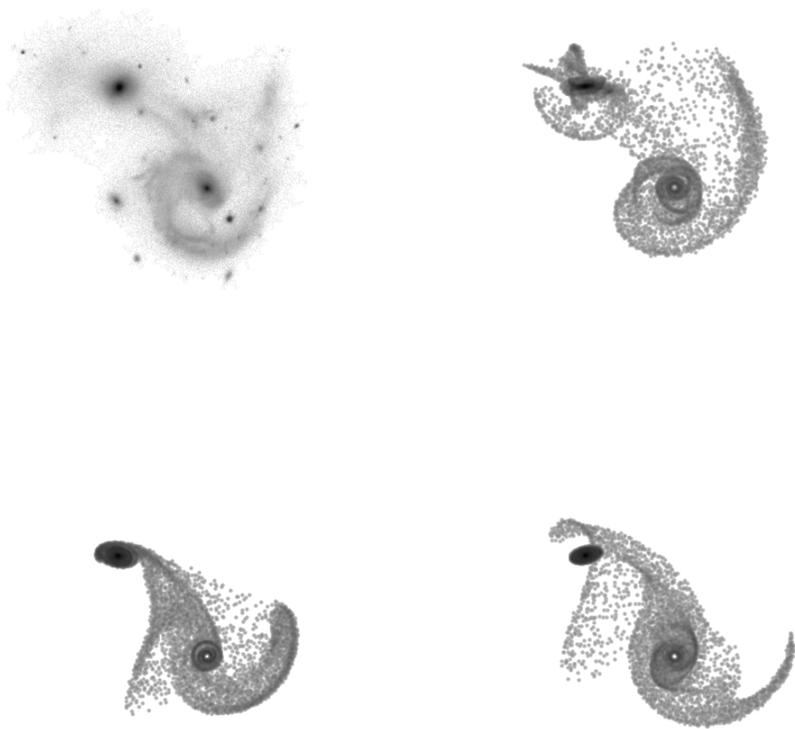


Figure 5.87: Target image and top 3 simulations for SDSS 587739646743412797

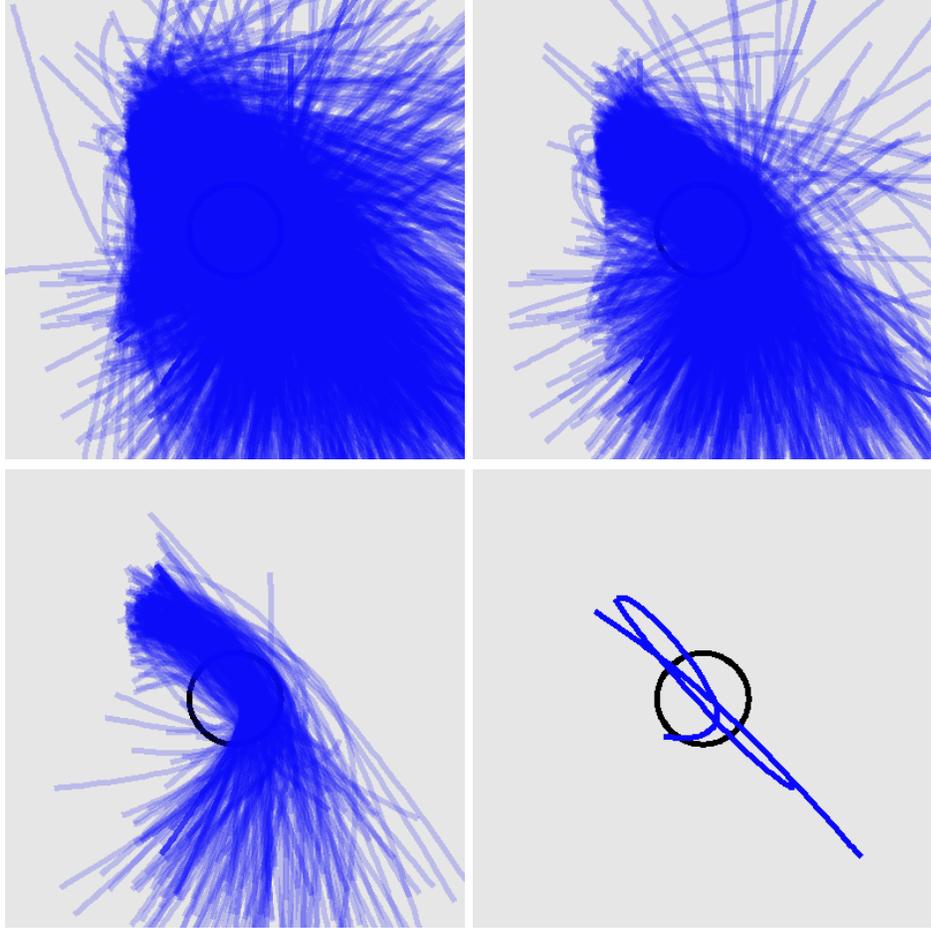


Figure 5.88: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587739646743412797

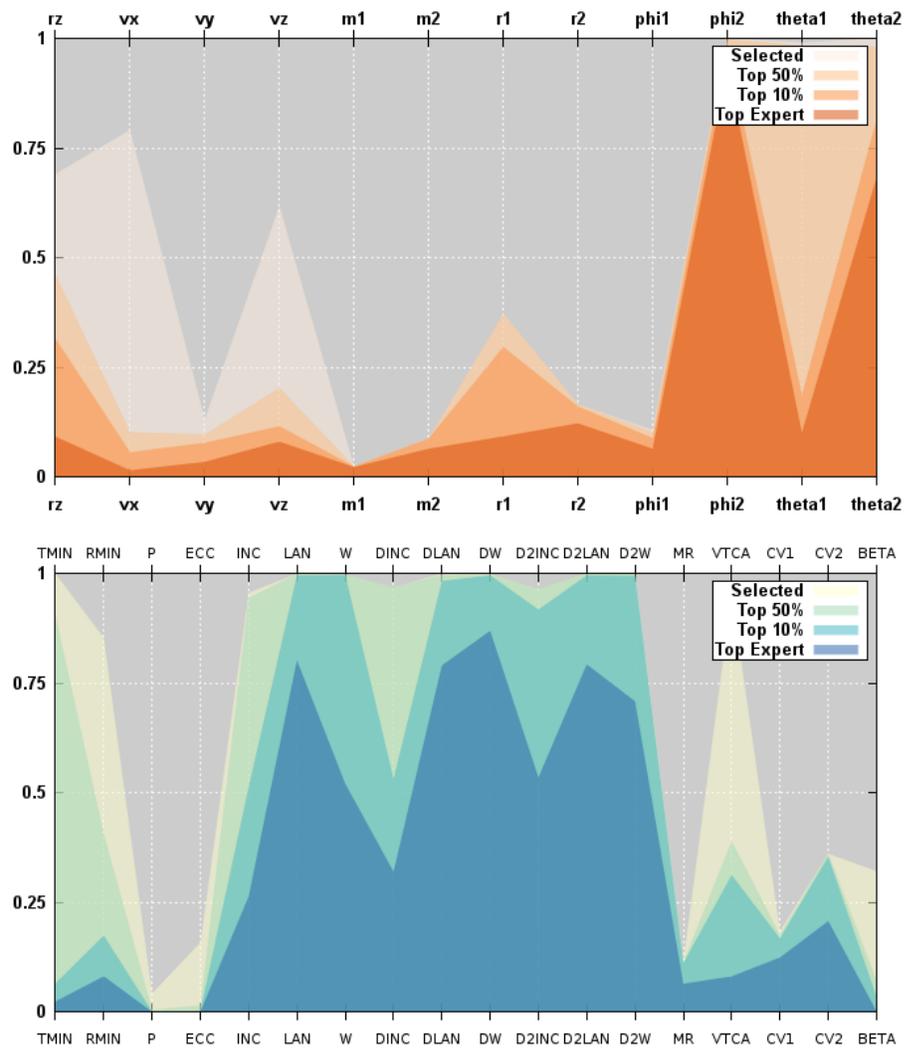


Figure 5.89: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587739646743412797

### 5.1.23 SDSS 587739647284805725

Arp 294 has a steadily declining fitness distribution. The target image has no strong tidal features for users to match. The volunteers recreated some density enhancements in the two disks. The top trajectories were similar. The simulation parameters were modestly converged with the exception of the disk inclination angles.

Table 5.23: Identification Information and Merger Zoo summary for SDSS 587739647284805725.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587739647284805725	Arp 294, NGC 3786/3788	11:39:42.4	+31:54:33.4	0.008933

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
20239	19441	798	101	14045	8276	5769

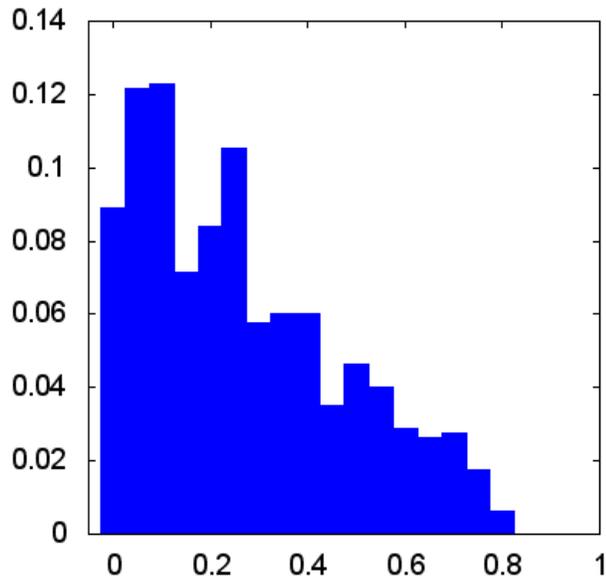


Figure 5.90: Relative frequency of fitness for all selected states of SDSS 587739647284805725

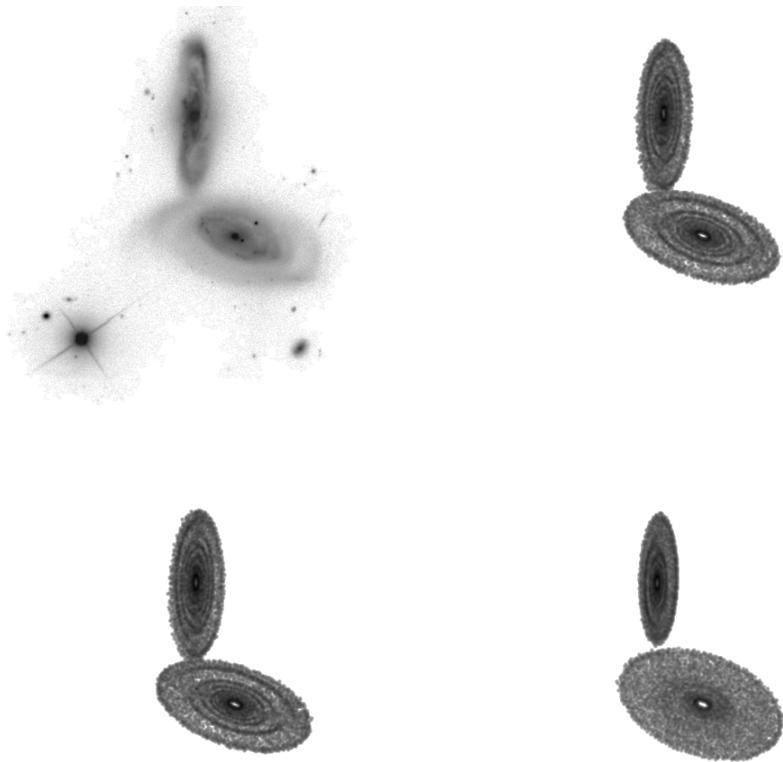


Figure 5.91: Target image and top 3 simulations for SDSS 587739647284805725

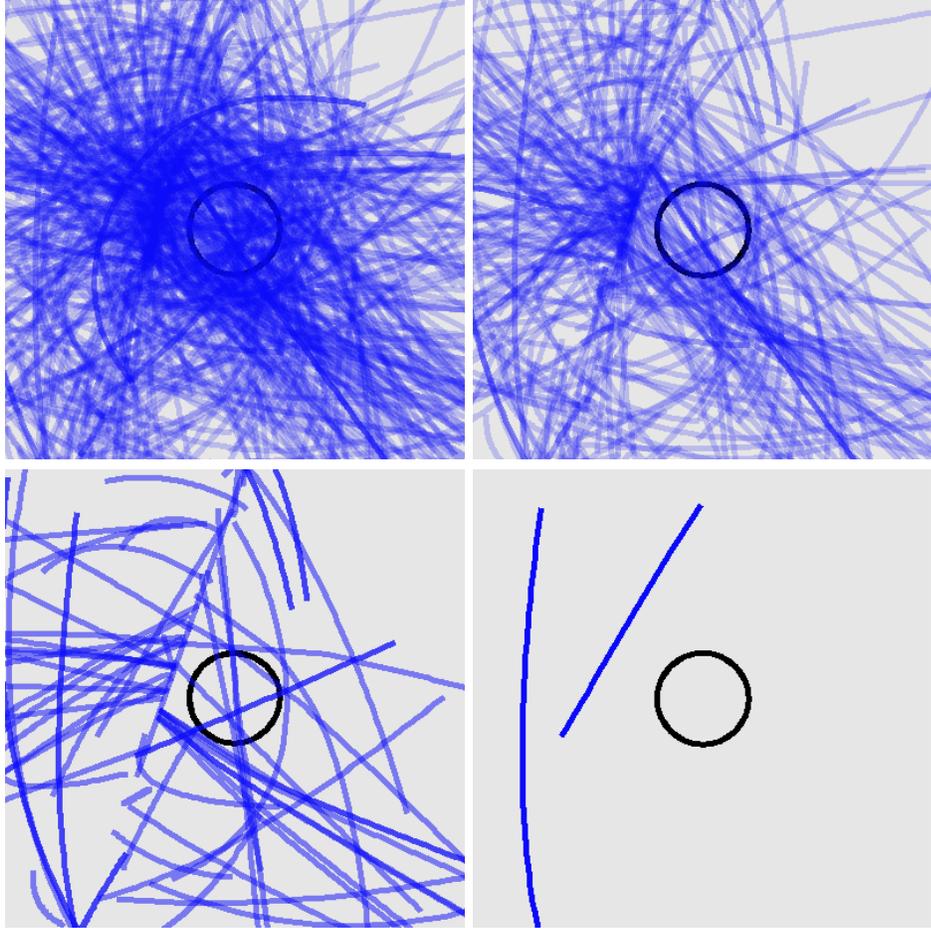


Figure 5.92: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587739647284805725

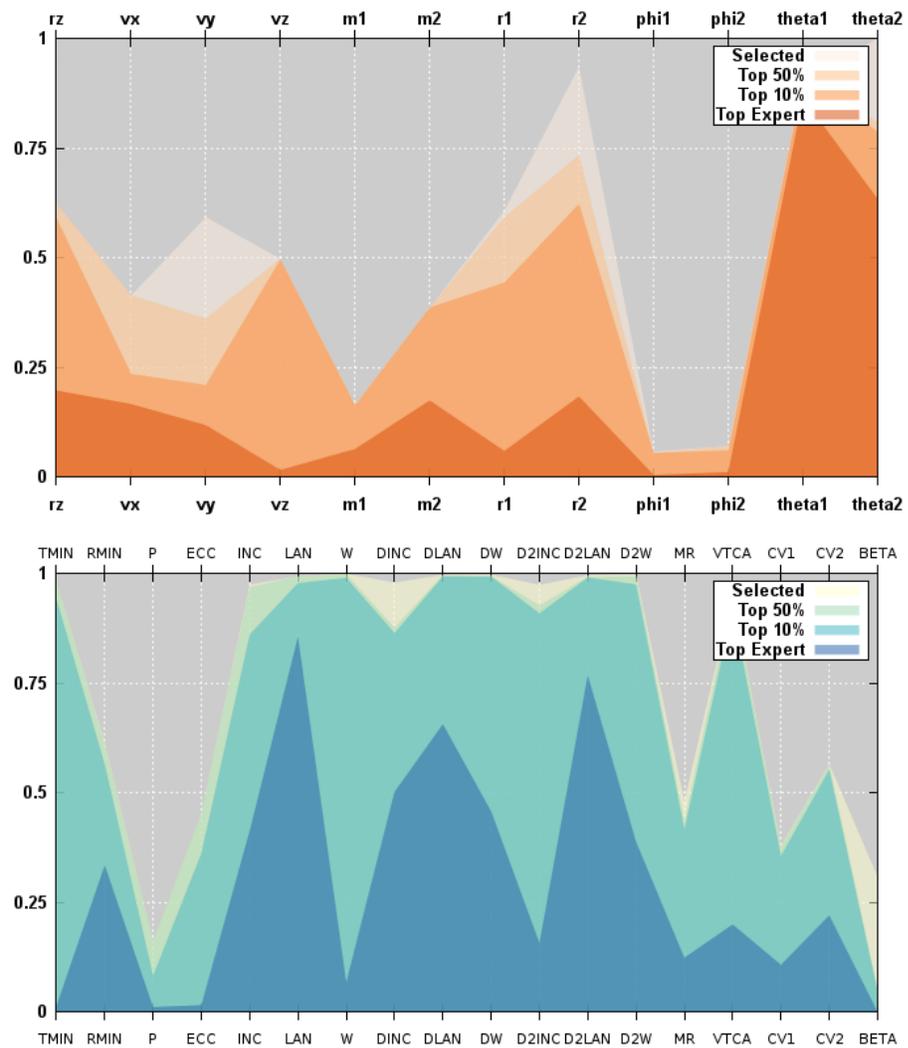


Figure 5.93: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587739647284805725

### 5.1.24 SDSS 587739707420967061

The steadily declining fitness distribution for Arp 172 is similar to that of some other systems. However, the target image does contain tidal features, even if they are somewhat diffuse. The top simulations recreate the two very loose “tails” in the image. The top 3 trajectories are very similar. The simulation parameters are modestly converged.

Table 5.24: Identification Information and Merger Zoo summary for SDSS 587739707420967061.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587739707420967061	Arp 172, IC 1178/1181	16:05:33.1	+17:36:04.6	0.033687

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
53962	53487	475	134	8647	4019	4628

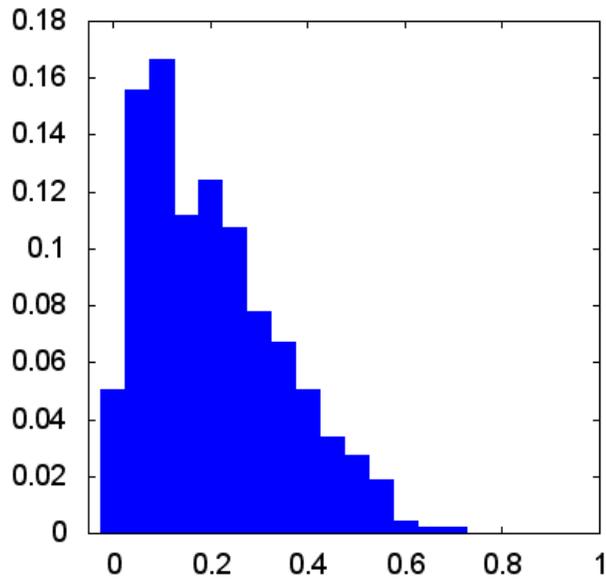


Figure 5.94: Relative frequency of fitness for all selected states of SDSS 587739707420967061

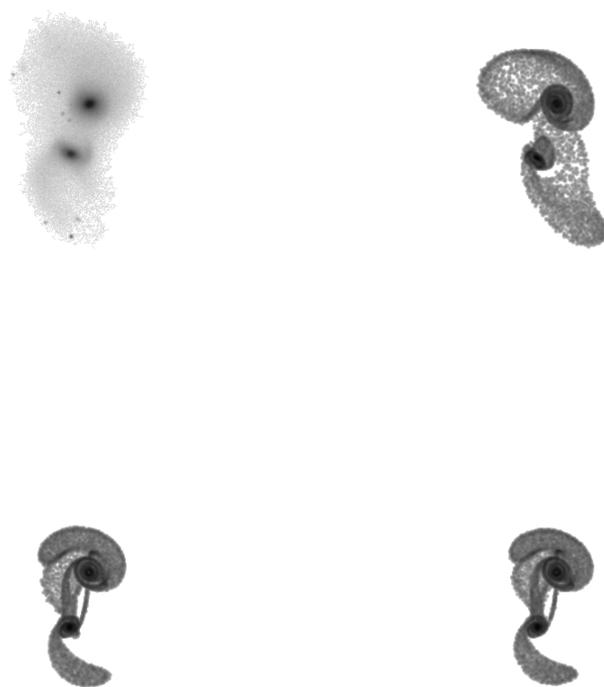


Figure 5.95: Target image and top 3 simulations for SDSS 587739707420967061

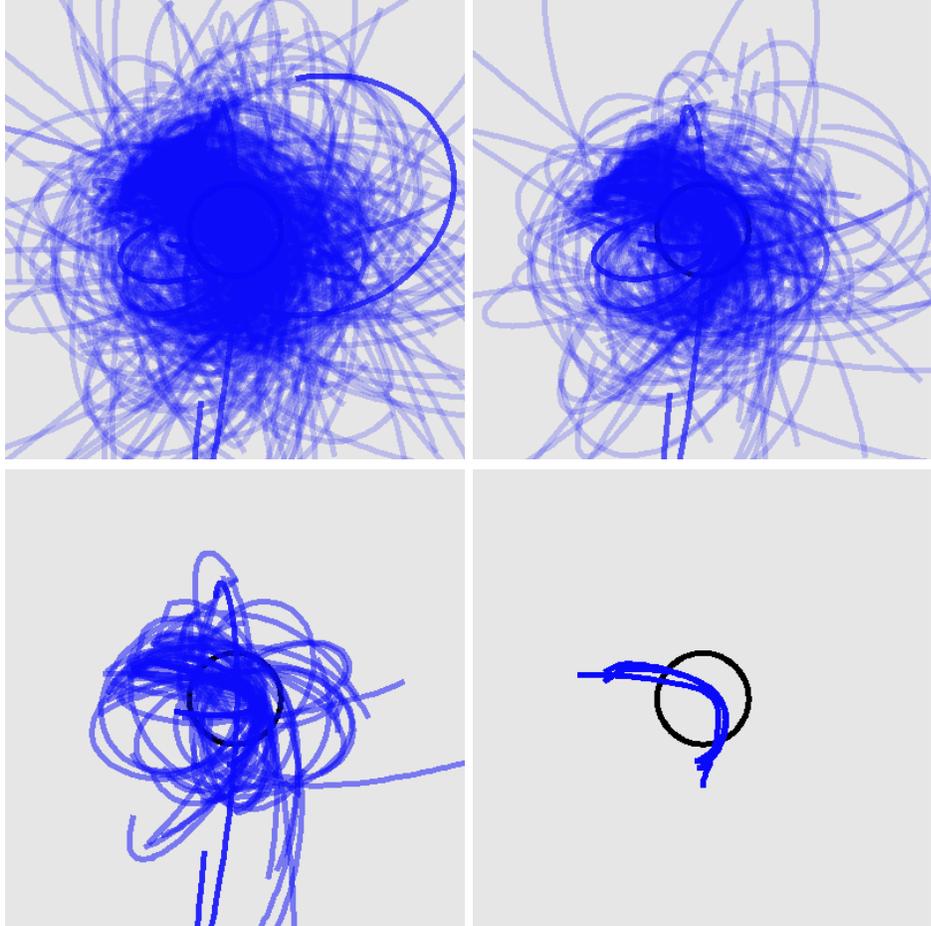


Figure 5.96: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587739707420967061

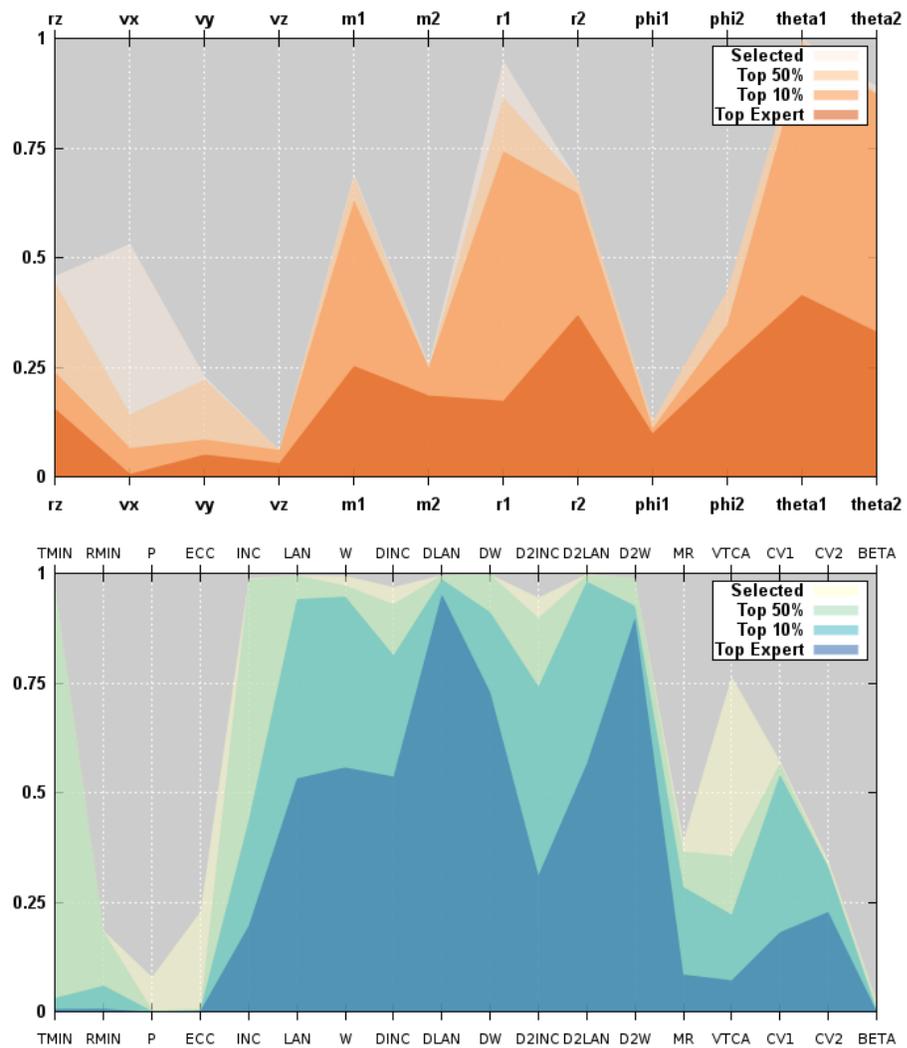


Figure 5.97: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587739707420967061

### 5.1.25 SDSS 587739721376202860

For Arp 302, a few high-fitness simulations were identified, though the steadily declining distribution indicates convergence may not have been achieved. The target image is another relatively bland system. The volunteers were able to create some density enhancements in the two disks. The top 3 trajectories are not that similar. The mass of the primary is very poorly constrained. This cascaded uncertainty to orbit parameters such as the time and distance of closest approach and the orbital velocity of stars in the outer disk of the primary.

Table 5.25: Identification Information and Merger Zoo summary for SDSS 587739721376202860.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587739721376202860	Arp 302, UGC 9618	14:57:00.6	+24:37:03.3	0.033669

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
14177	13772	405	46	7366	4150	3216

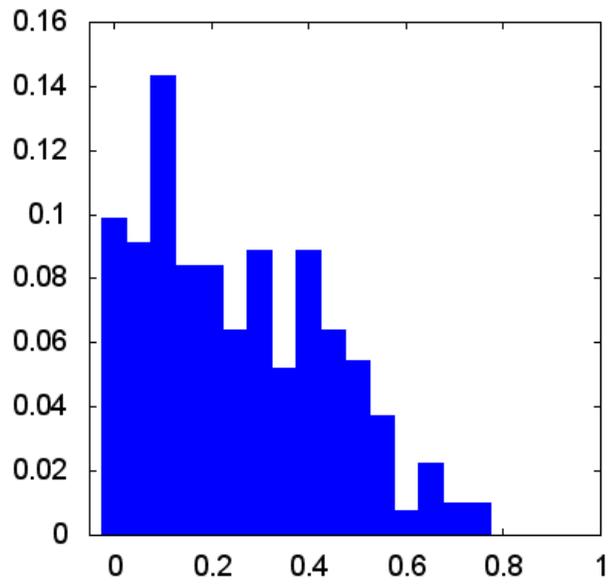


Figure 5.98: Relative frequency of fitness for all selected states of SDSS 587739721376202860

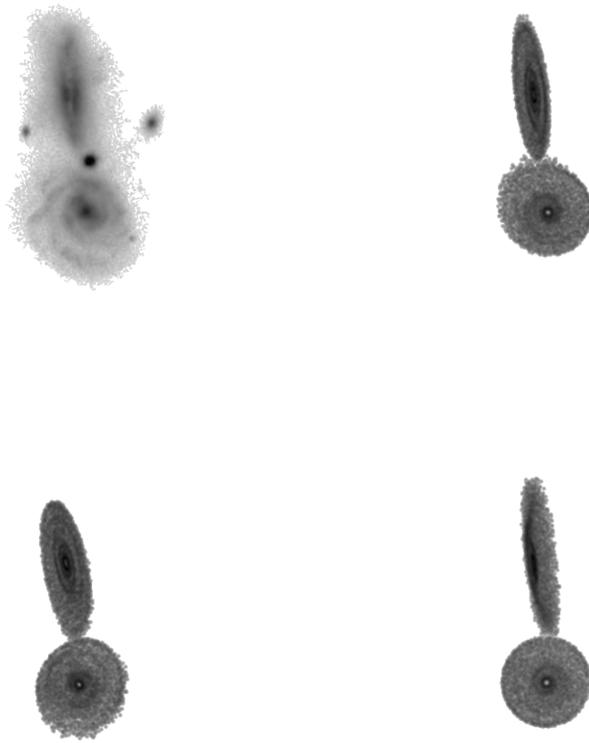


Figure 5.99: Target image and top 3 simulations for SDSS 587739721376202860

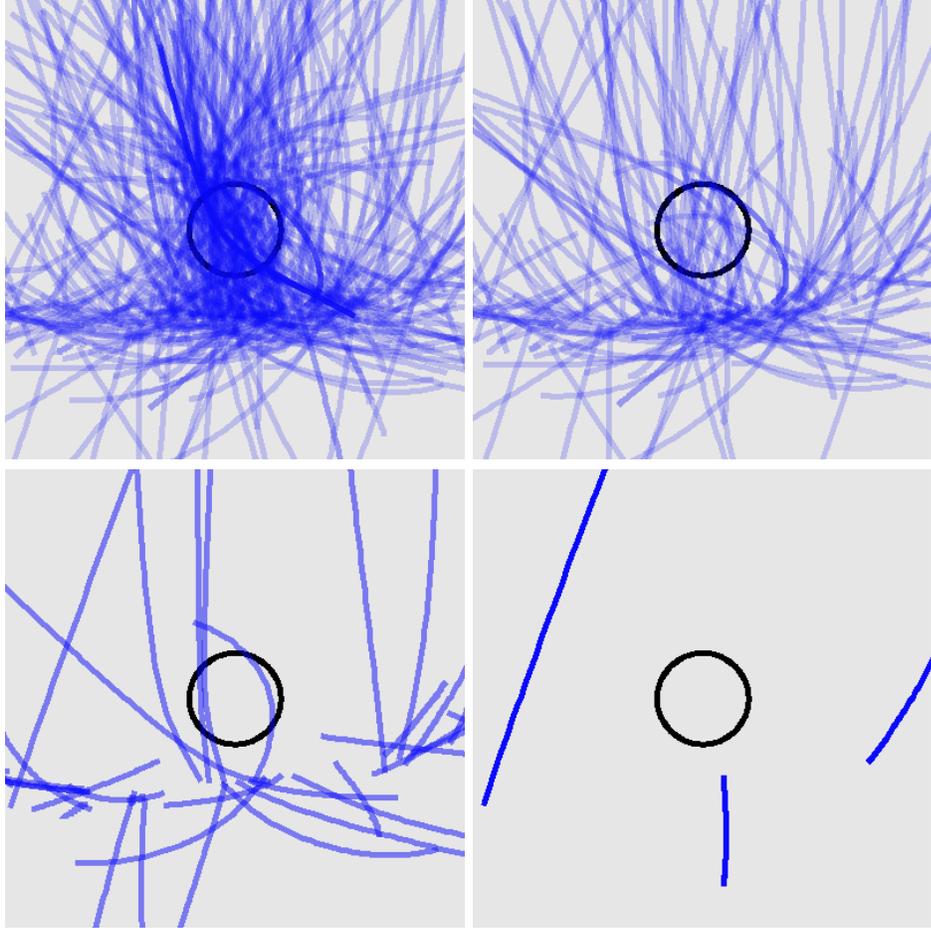


Figure 5.100: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587739721376202860

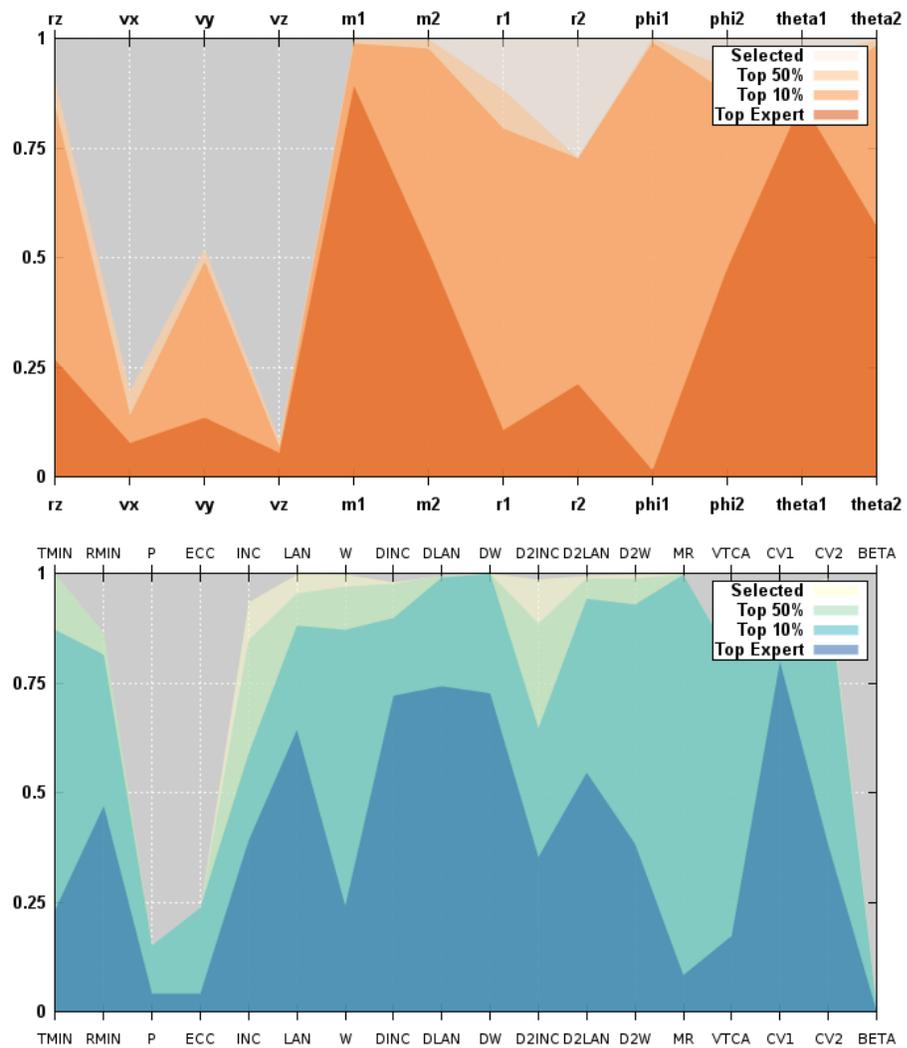


Figure 5.101: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587739721376202860

### 5.1.26 SDSS 587739721900163101

The steep decline of fitness values for Arp 242 is easily noticed. The long, elegant tidal features of the “Mice” can be difficult to recreate. Our volunteers submitted simulations that capture the orientation, length, and thickness of the tails. The top 3 trajectories are very similar. There is also a family of trajectories visible in the lower left trajectory panel. For the simulation parameters, the convergence of the mass, size, and orientations of the disks is low. However the orbit eccentricity, semi-parameter, and inclination are well constrained.

Table 5.26: Identification Information and Merger Zoo summary for SDSS 587739721900163101.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587739721900163101	Arp 242, NGC 4676	12:46:10.1	+30:43:52.6	0.022059

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
55366	54706	660	161	11887	4815	7072

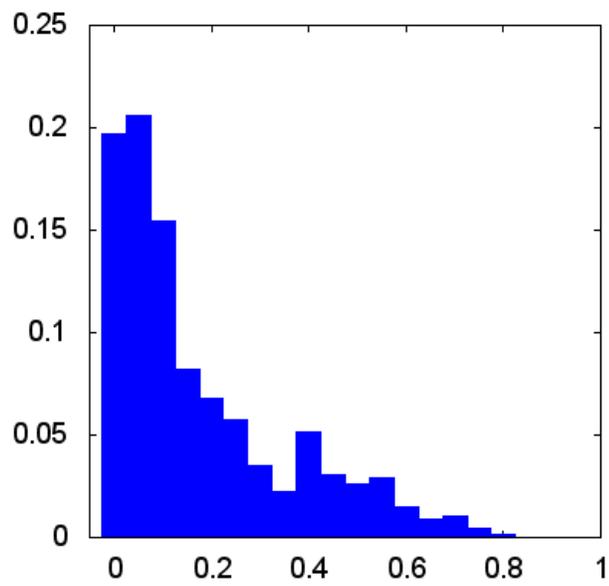


Figure 5.102: Relative frequency of fitness for all selected states of SDSS 587739721900163101



Figure 5.103: Target image and top 3 simulations for SDSS 587739721900163101

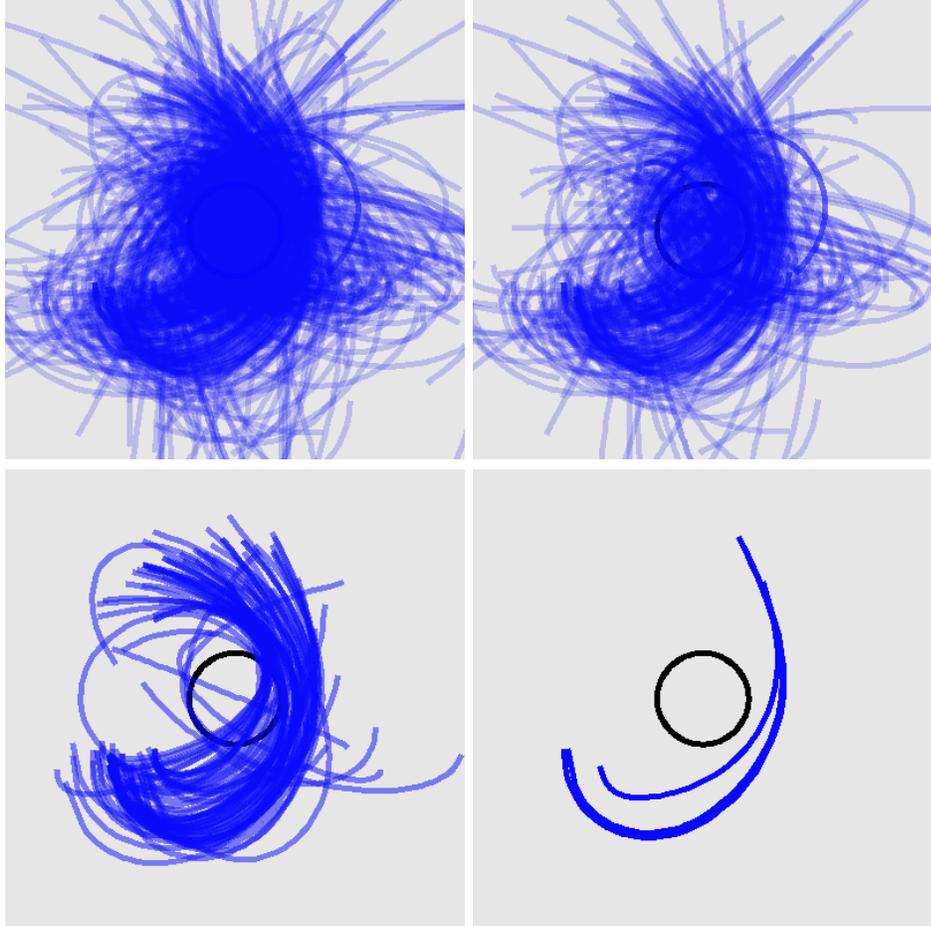


Figure 5.104: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587739721900163101

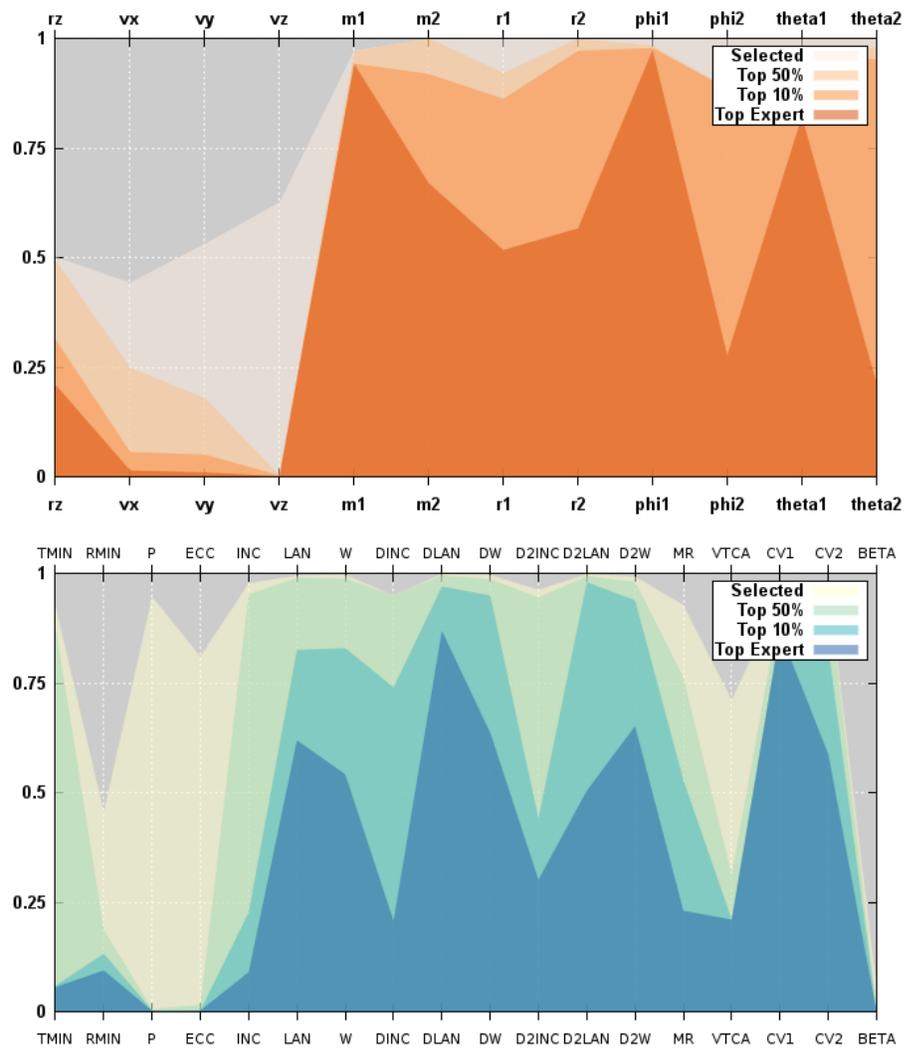


Figure 5.105: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587739721900163101

### 5.1.27 SDSS 587739810496708646

The small percentage of high-fitness simulations for Arp 72 is apparent. The top simulations present moderate matches to the southern tail and diffuse northern feature of the primary galaxy. The trajectories are similar to each other. The simulation parameters appear well converged, yet the orbit inclination is not.

Table 5.27: Identification Information and Merger Zoo summary for SDSS 587739810496708646.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587739810496708646	Arp 72, NGC 5994/5996	15:46:58.2	+17:53:04.4	0.010974

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
44249	43849	400	76	7350	2416	4934

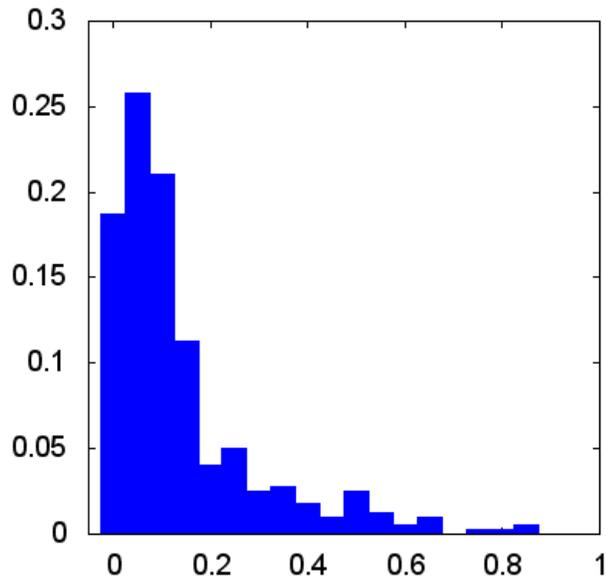


Figure 5.106: Relative frequency of fitness for all selected states of SDSS 587739810496708646

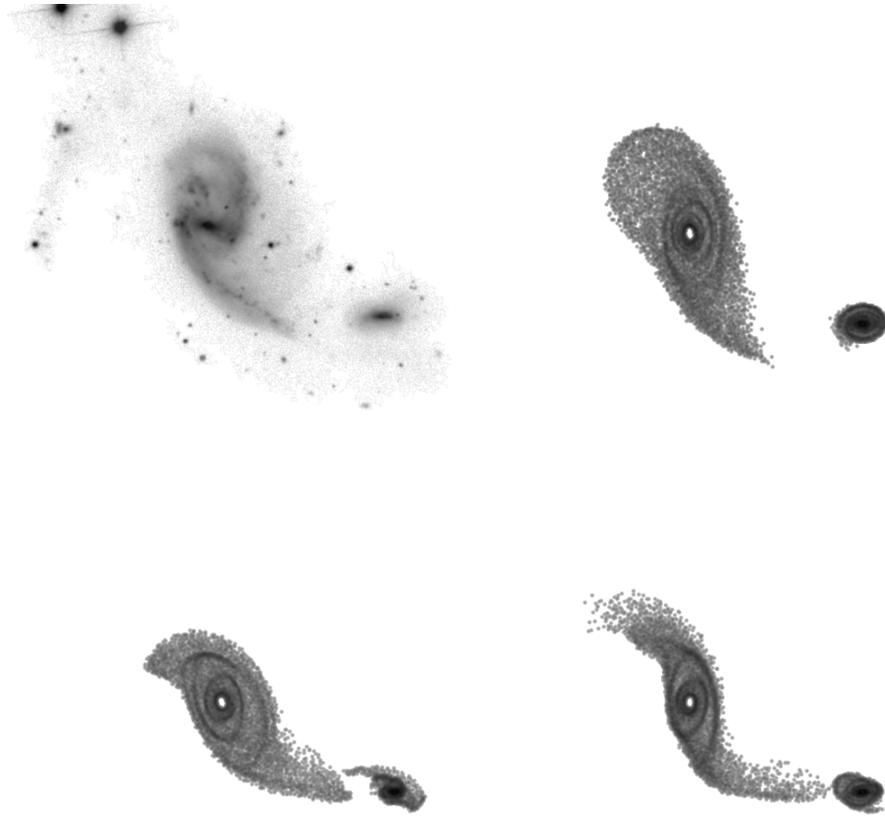


Figure 5.107: Target image and top 3 simulations for SDSS 587739810496708646

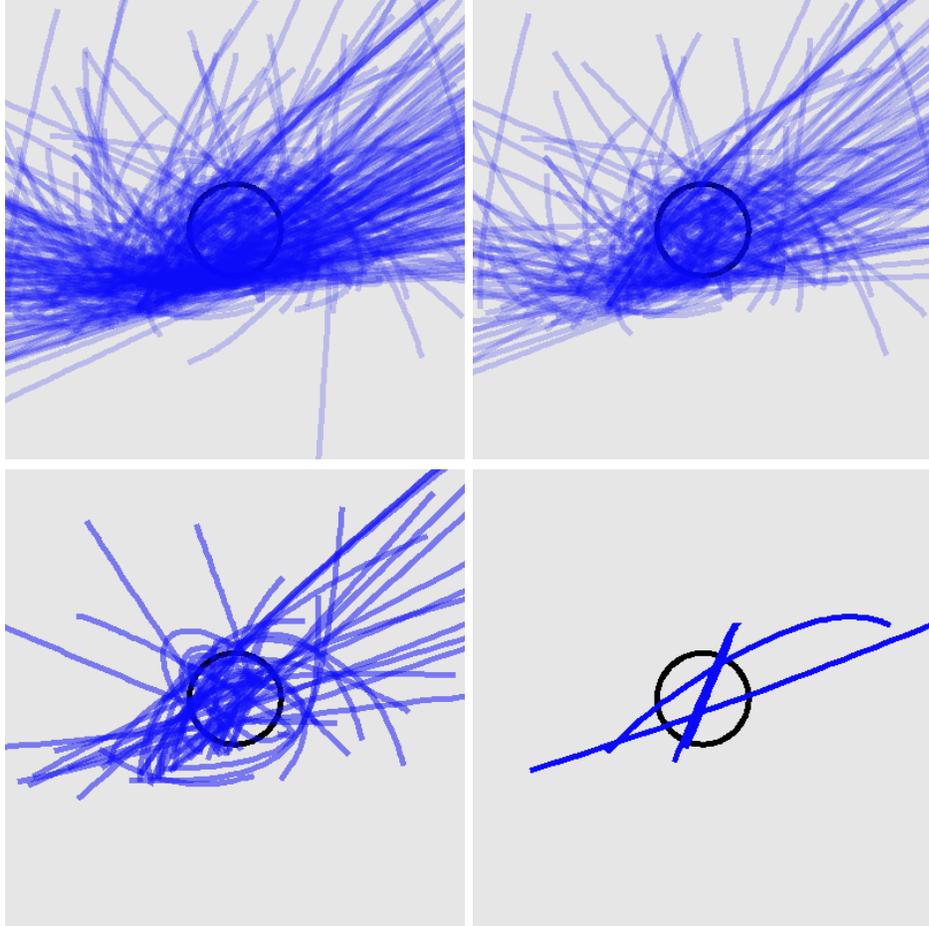


Figure 5.108: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587739810496708646

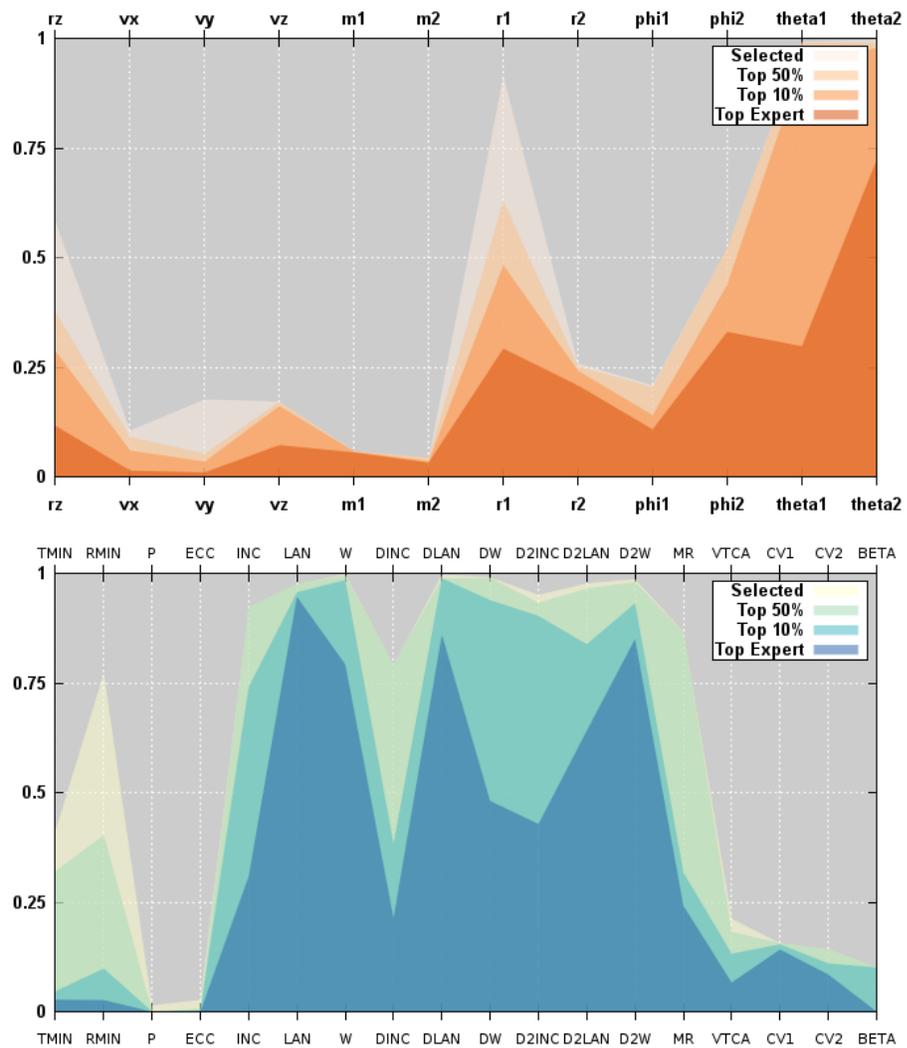


Figure 5.109: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587739810496708646

### 5.1.28 SDSS 587739845393580192

There are relatively few high-fitness simulations for Arp 101. The asymmetric tidal feature of the norther galaxy is not well recreated by the top simulations. The top trajectories are similar to each other though. The disk sizes and orientations are not well converged.

Table 5.28: Identification Information and Merger Zoo summary for SDSS 587739845393580192.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587739845393580192	Arp 101, 10164/10169	UGC 16:04:31.7	+14:49:08.8	0.015274

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
30553	30095	458	130	8378	2951	5427

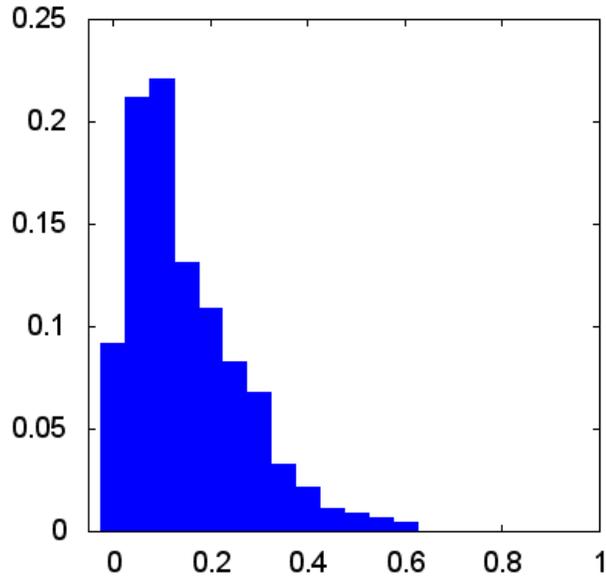


Figure 5.110: Relative frequency of fitness for all selected states of SDSS 587739845393580192



Figure 5.111: Target image and top 3 simulations for SDSS 587739845393580192

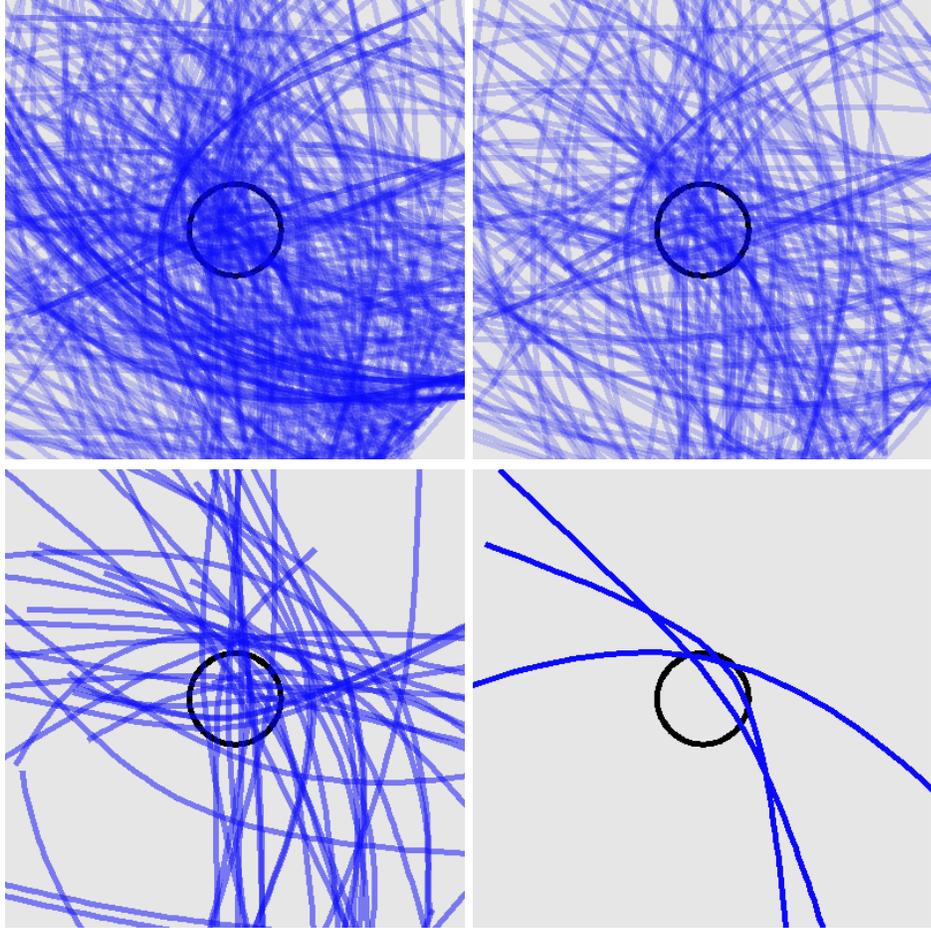


Figure 5.112: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587739845393580192

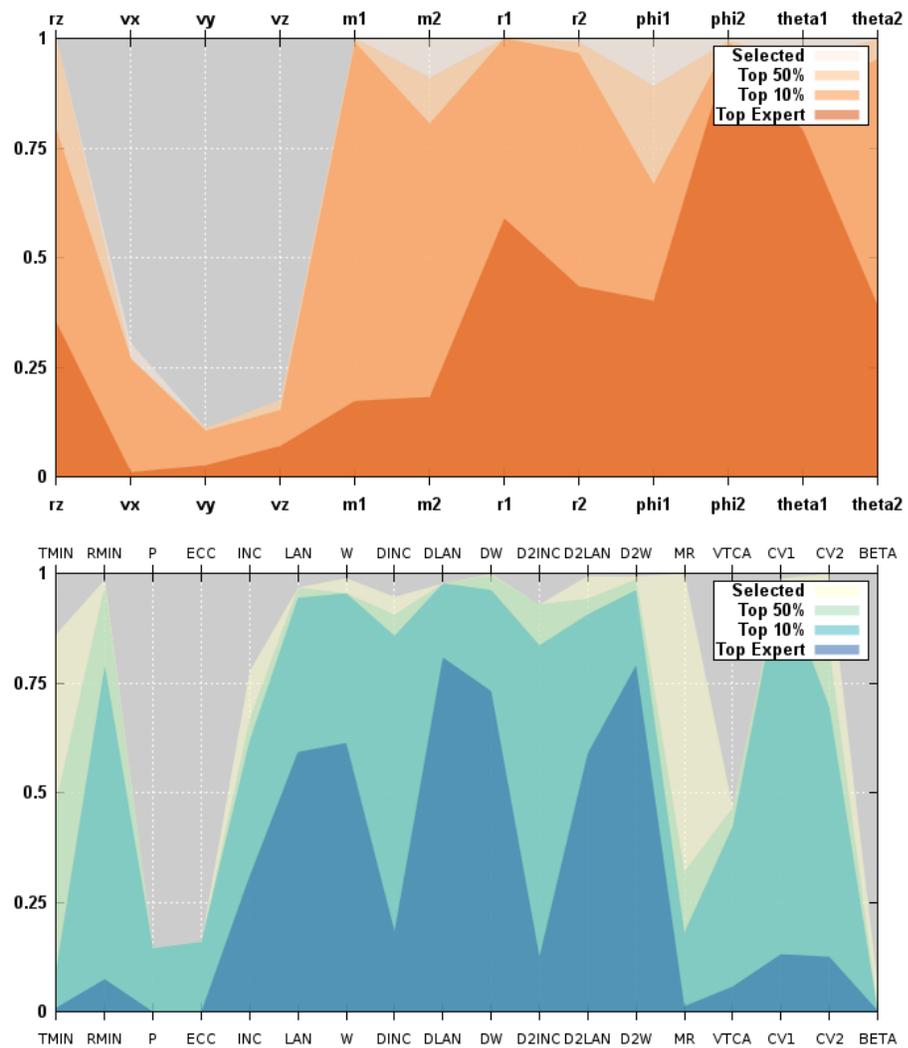


Figure 5.113: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587739845393580192

### 5.1.29 SDSS 587741391565422775

Arp 58 has a fitness distribution with a large fraction of low fitness states combined with a flat distribution extending to high-fitness values greater than 0.8. The simulation images show a very accurate recreation of the large tidal tails of the primary. There is an additional density enhancement near the secondary that is at least present in some of the simulations. The top trajectories are almost identical. The simulation parameters are impressively well converged with the exception of the mass of the primary. That uncertainty transfers only to the mass ratio as other orbit parameters seem constrained.

Table 5.29: Identification Information and Merger Zoo summary for SDSS 587741391565422775.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587741391565422775	Arp 58, UGC 4457	8:31:57.6	+19:12:40.4	0.037222

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
29718	29336	382	69	7085	3279	3806

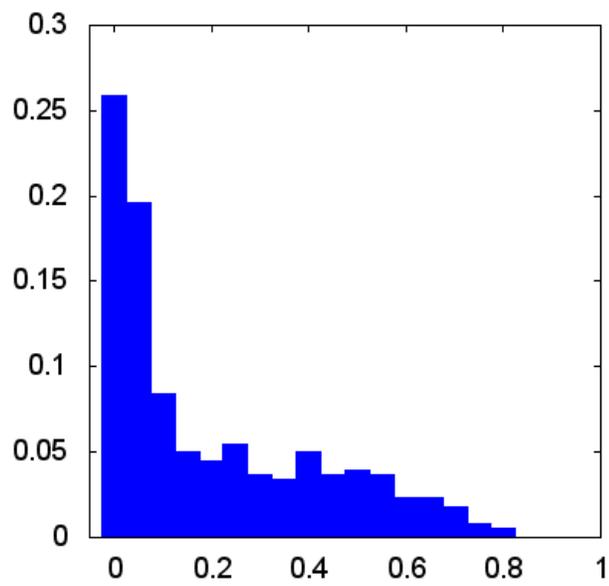


Figure 5.114: Relative frequency of fitness for all selected states of SDSS 587741391565422775

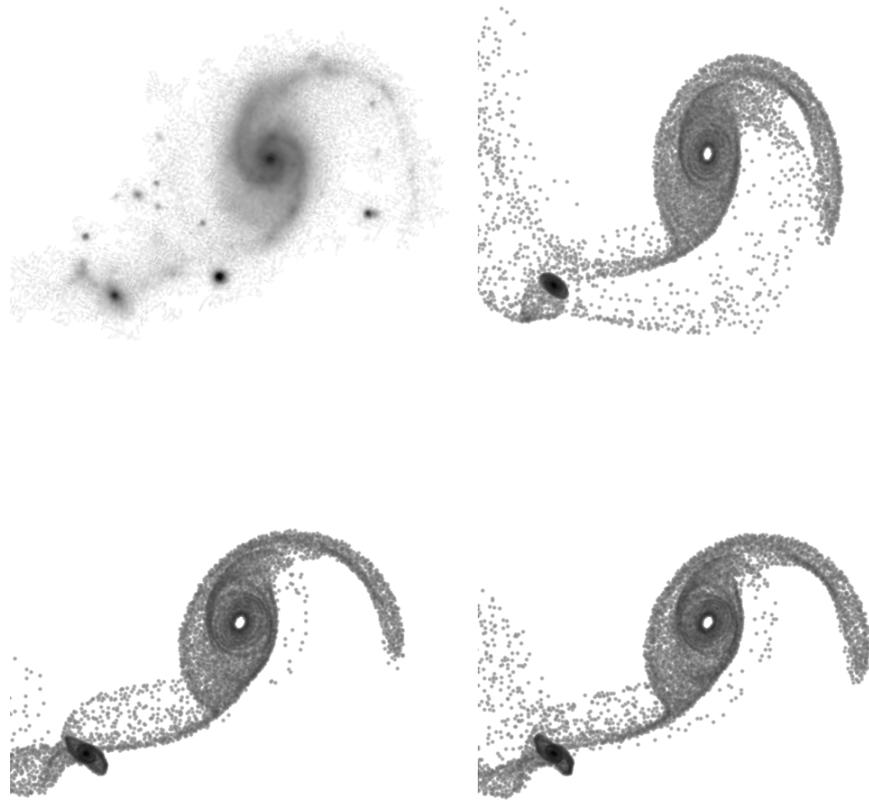


Figure 5.115: Target image and top 3 simulations for SDSS 587741391565422775

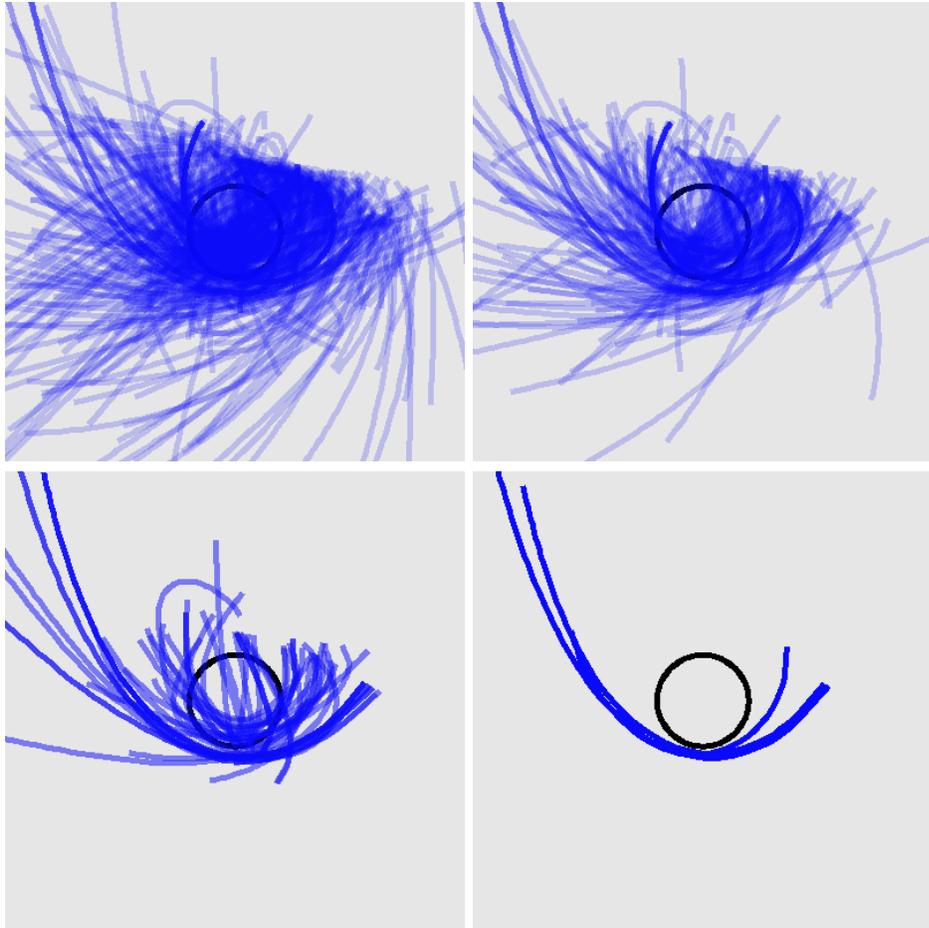


Figure 5.116: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587741391565422775

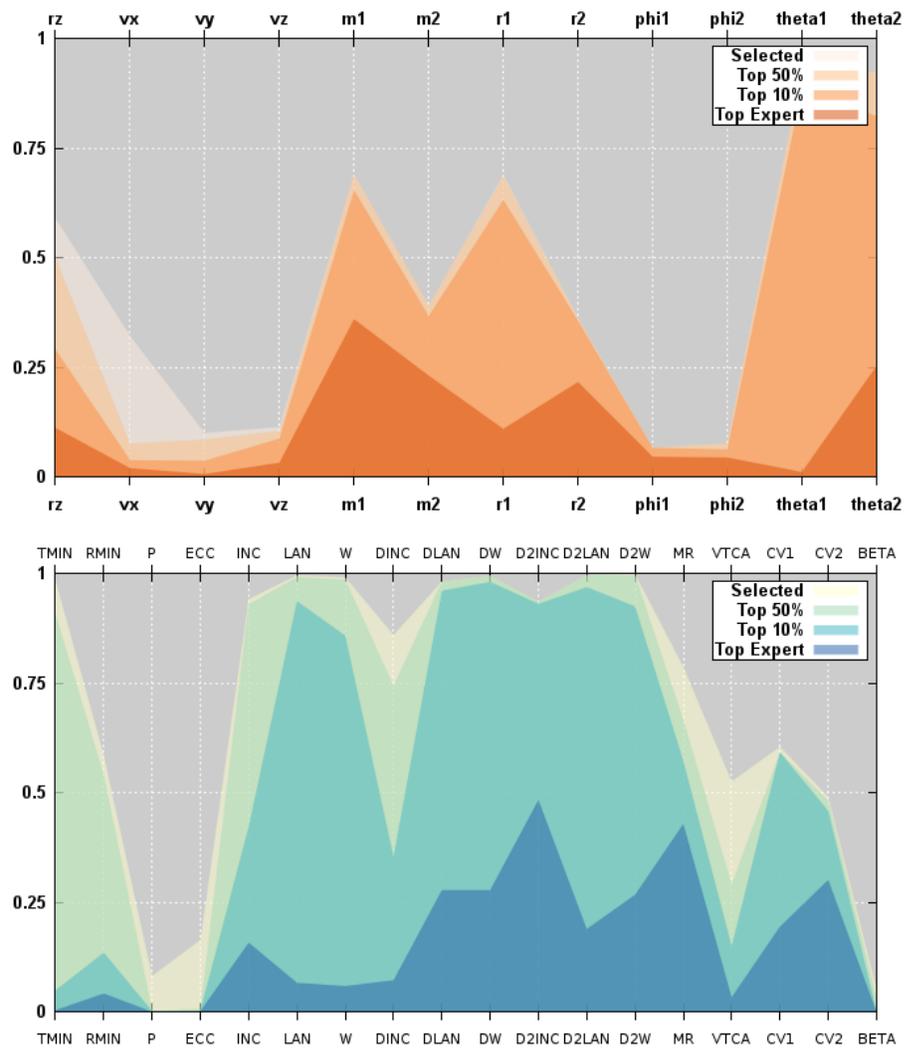


Figure 5.117: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587741391565422775

### 5.1.30 SDSS 587741532784361481

Arp 105 was moderately popular with  $\sim 60000$  simulations viewed by volunteers. The top simulations all recreate the long, thin tidal tail of the primary, though the orientation is not quite correct. The simulations also predict a southernly tail that is behind the secondary galaxy. The trajectories show a family structure with the top 3 being almost identical. The simulation parameters are well constrained with the exception of the disk orientation parameters. The orbit parameters are nicely converged.

Table 5.30: Identification Information and Merger Zoo summary for SDSS 587741532784361481.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587741532784361481	Arp 105, NGC 3561	11:11:12.9	+28:42:42.4	0.029356

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
59402	58505	897	224	15814	6402	9412

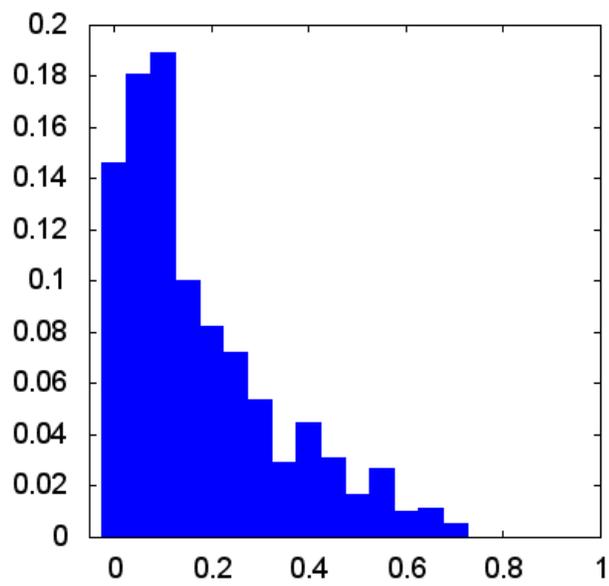


Figure 5.118: Relative frequency of fitness for all selected states of SDSS 587741532784361481

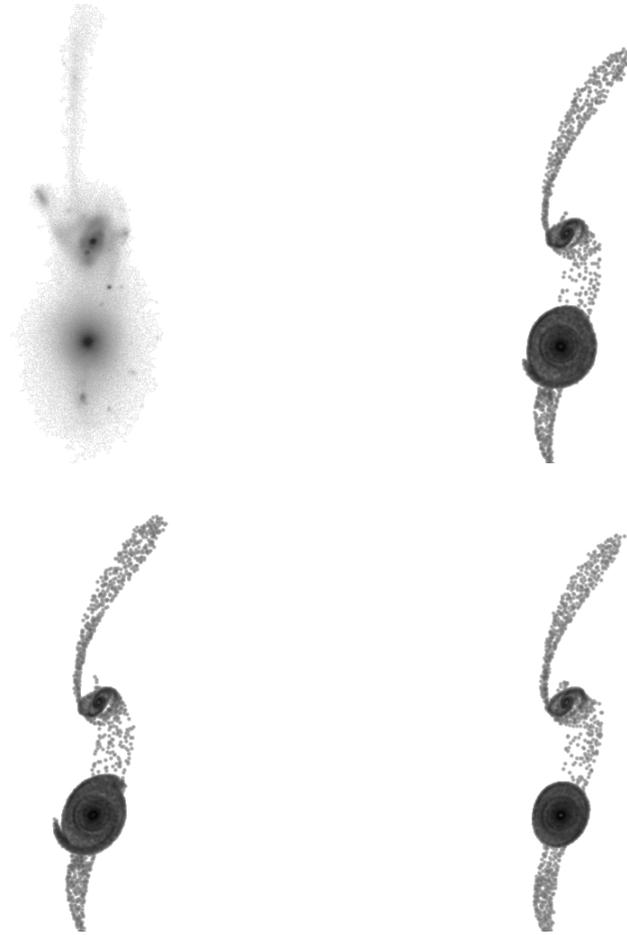


Figure 5.119: Target image and top 3 simulations for SDSS 587741532784361481

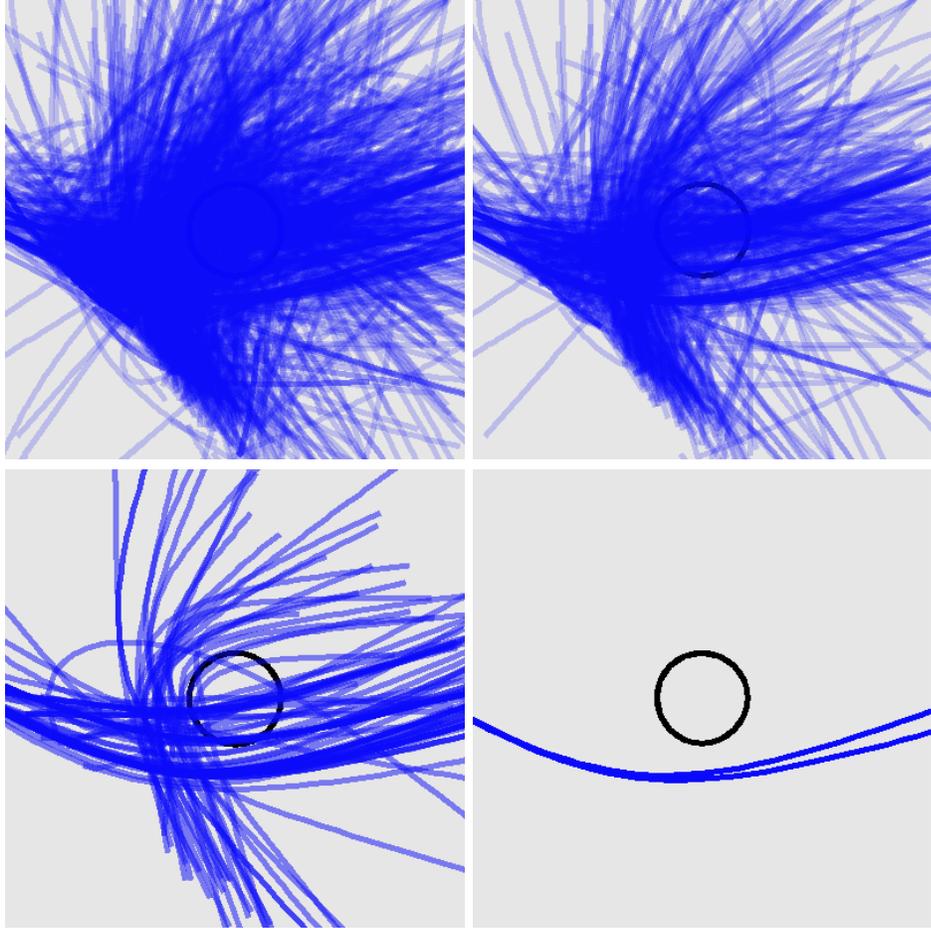


Figure 5.120: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587741532784361481

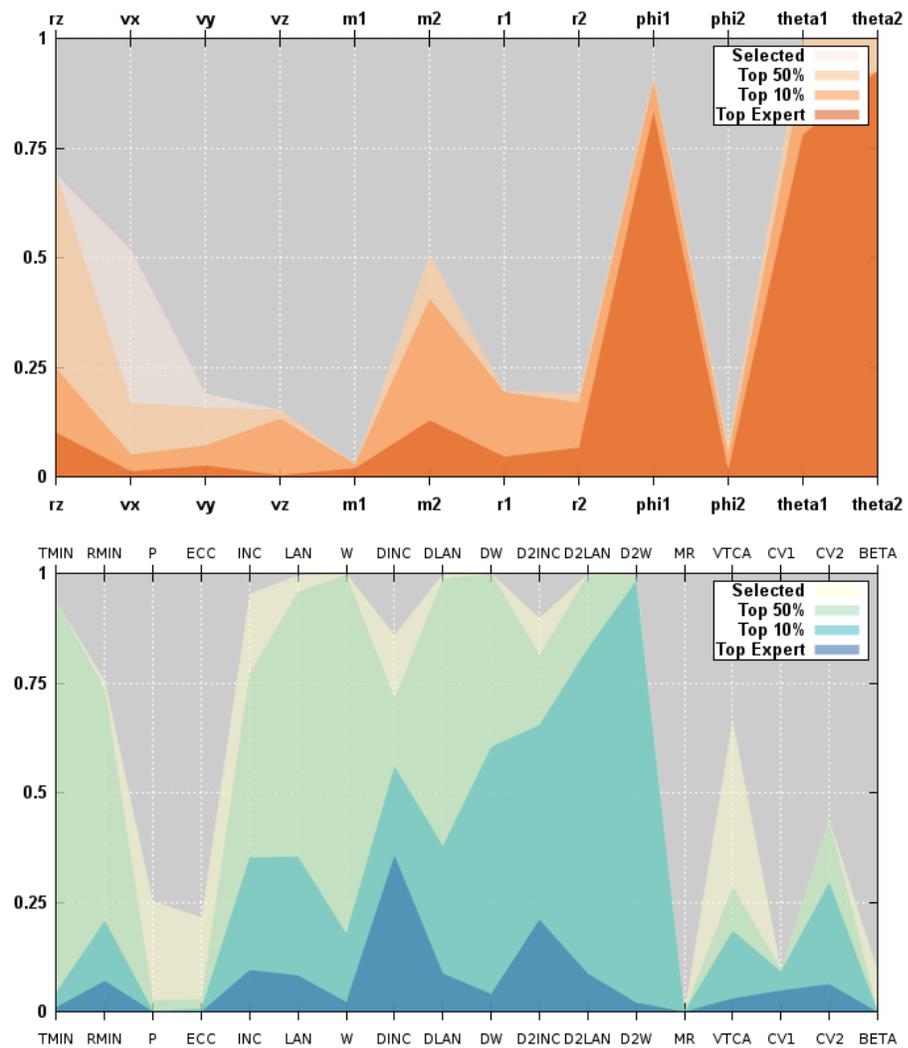


Figure 5.121: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587741532784361481

### 5.1.31 SDSS 587741534400217110

There are asymmetric tidal features in Apr 97. The L-shaped fitness distribution is easily observed. Both the curvature of the southern tail and straightness of the northern one are well matched to the target image. The top trajectories are similar. The simulation parameters seem well converged with the exceptions of the radius and inclination of the primary disk.

Table 5.31: Identification Information and Merger Zoo summary for SDSS 587741534400217110.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587741534400217110	Arp 97, UGC 7085A	12:05:45.4	+31:03:31.0	0.023366

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
50875	50251	624	177	11177	4866	6311

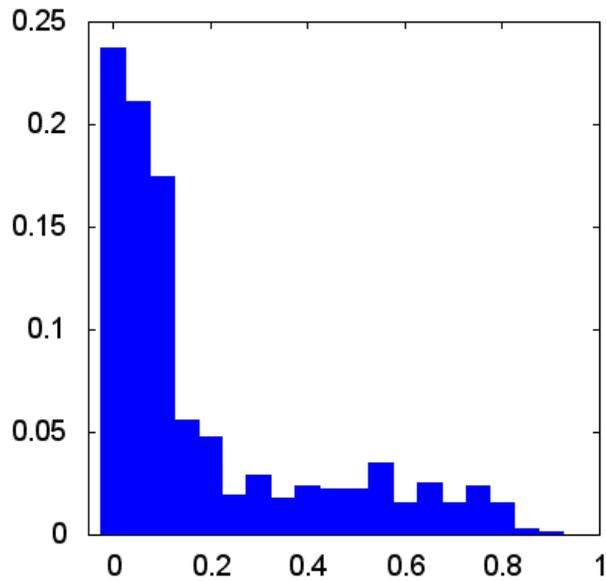


Figure 5.122: Relative frequency of fitness for all selected states of SDSS 587741534400217110

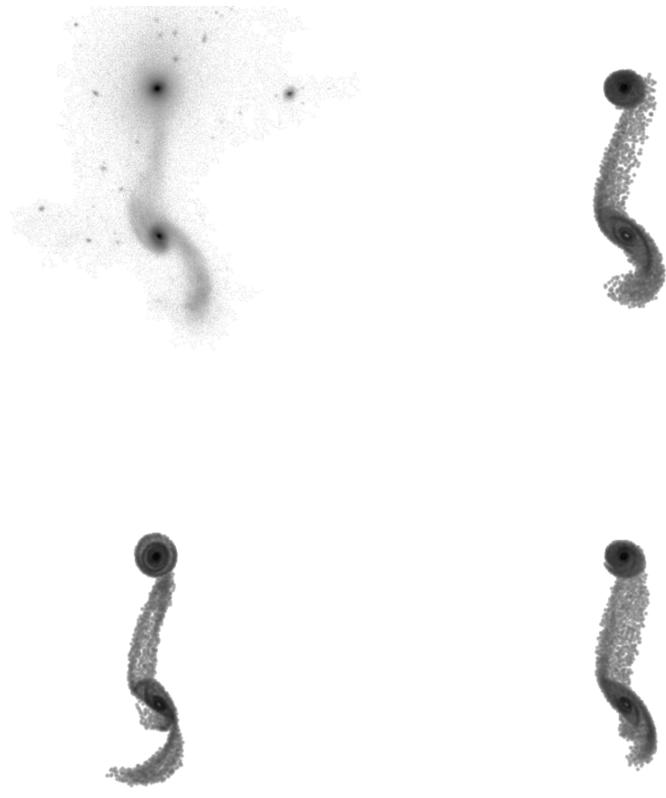


Figure 5.123: Target image and top 3 simulations for SDSS 587741534400217110

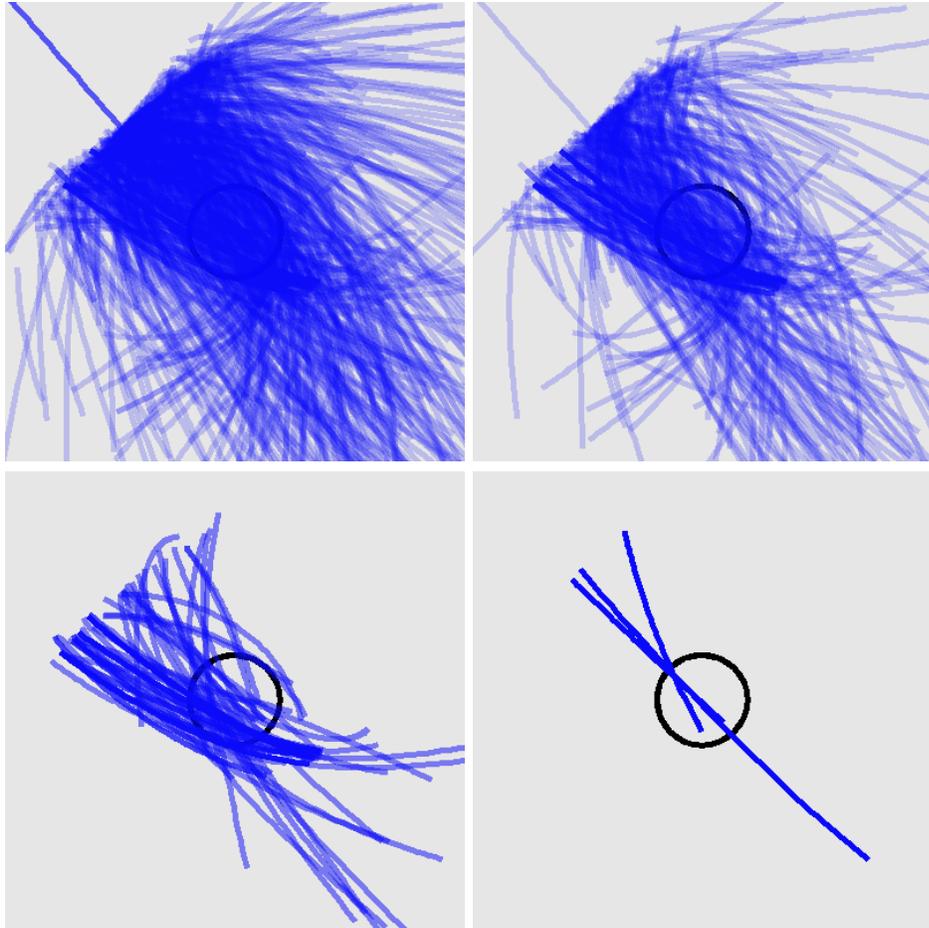


Figure 5.124: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587741534400217110

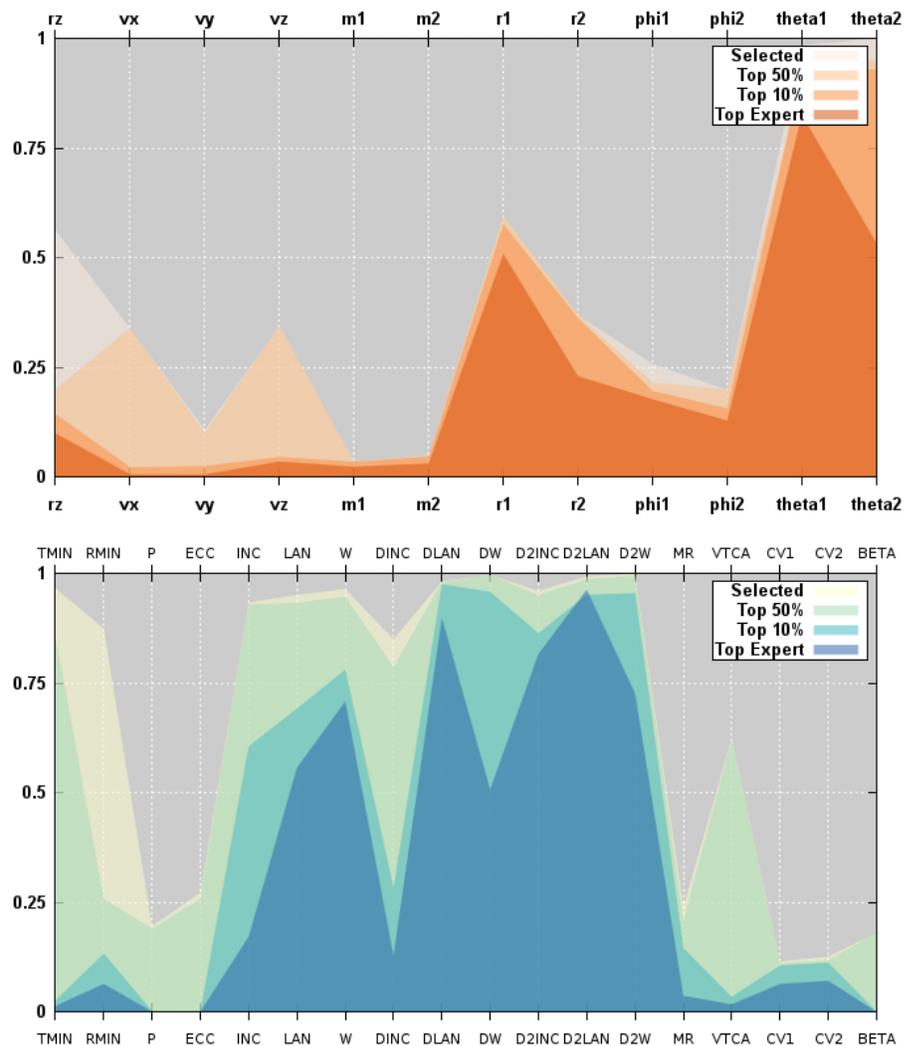


Figure 5.125: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587741534400217110

### 5.1.32 SDSS 587741602030026825

The volunteers only viewed 7300 simulations for Arp 305. This system contains two small, widely-spaced galaxies. However, both galaxies have tidal features. The top simulations seem to have achieved a moderate match to the tidal tails of the primary. The orientation of the tails is not quite correct. For the secondary, the top simulations produce similar distortion to what is visible in the target image. The trajectories are very similar. The simulation parameters are well constrained except for the radius of the secondary disk.

Table 5.32: Identification Information and Merger Zoo summary for SDSS 587741602030026825.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587741602030026825	Arp 305, NGC 4016/4017	11:58:45.6	+27:27:07.3	0.011521

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
7367	7101	266	144	4998	2023	2975

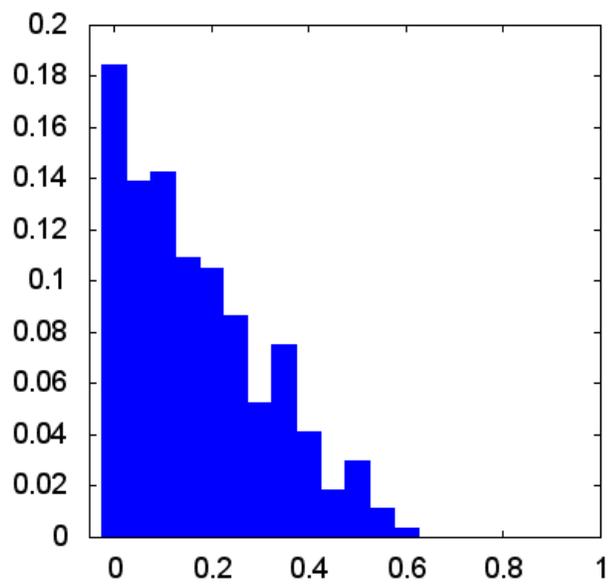


Figure 5.126: Relative frequency of fitness for all selected states of SDSS 587741602030026825



Figure 5.127: Target image and top 3 simulations for SDSS 587741602030026825

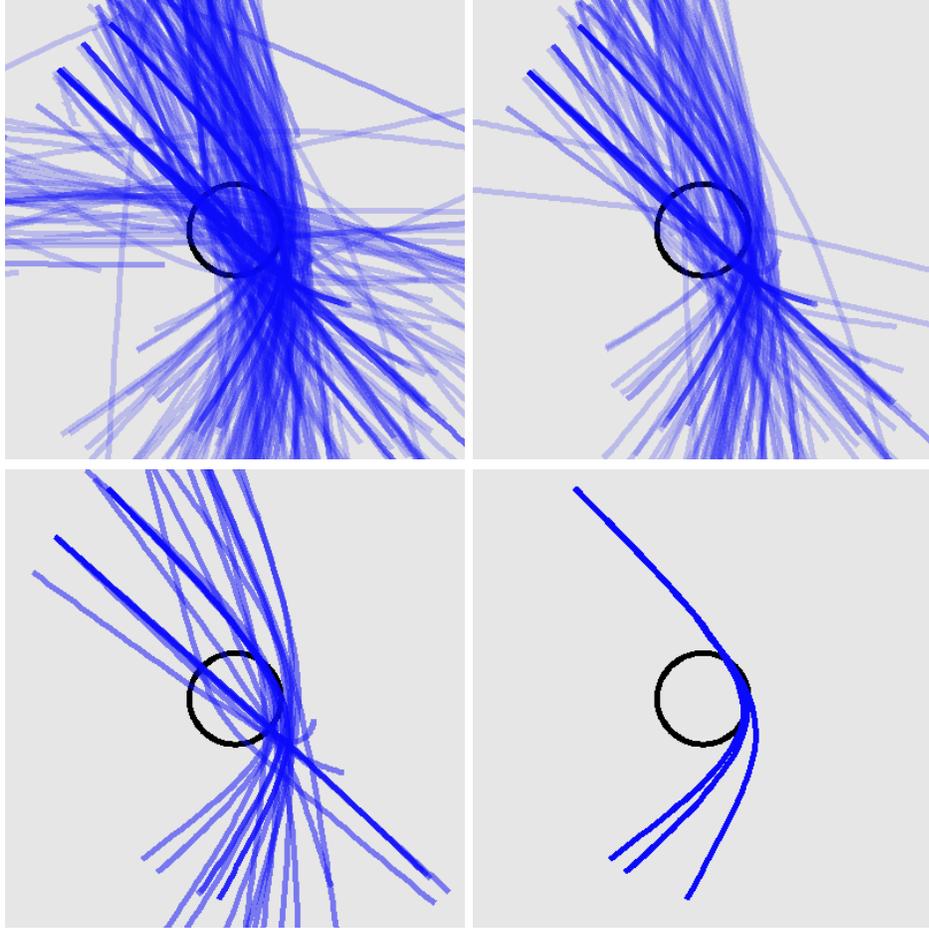


Figure 5.128: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587741602030026825

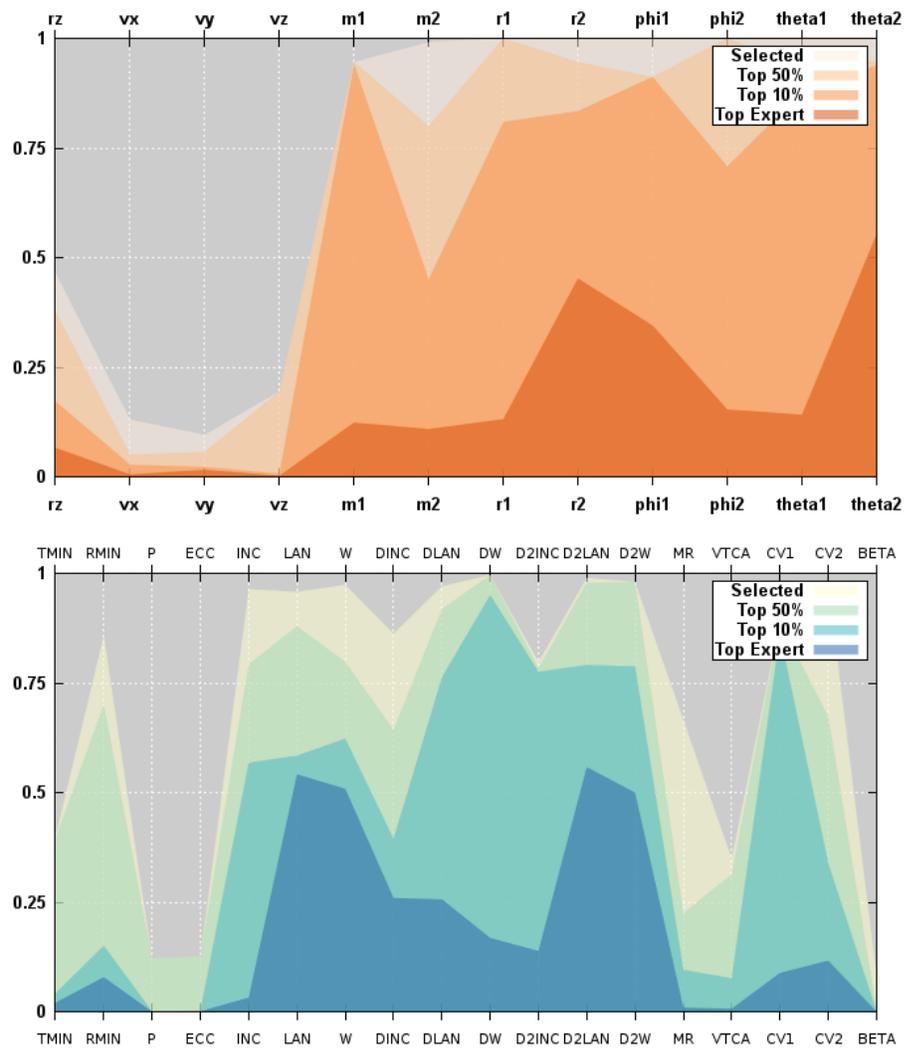


Figure 5.129: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587741602030026825

### 5.1.33 SDSS 587741722819493915

Arp 106 has an L-shaped fitness distribution. The warped disk and tidal tails are well matched in the simulation images. The top trajectories are not identical but do appear to be bound orbits. The simulation parameters are moderately converged. The orbit parameters appear better converged than the simulation values.

Table 5.33: Identification Information and Merger Zoo summary for SDSS 587741722819493915.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587741722819493915	Arp 106, NGC 4211	12:15:35.8	+28:10:39.8	0.022012

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
40835	40216	619	196	11050	4825	6225

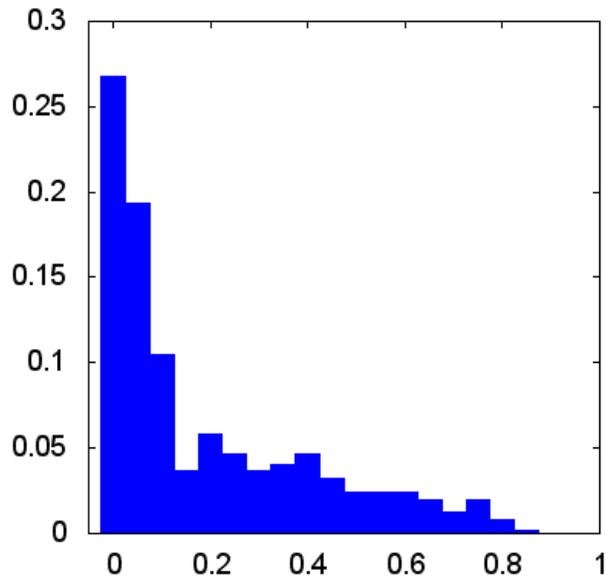


Figure 5.130: Relative frequency of fitness for all selected states of SDSS 587741722819493915

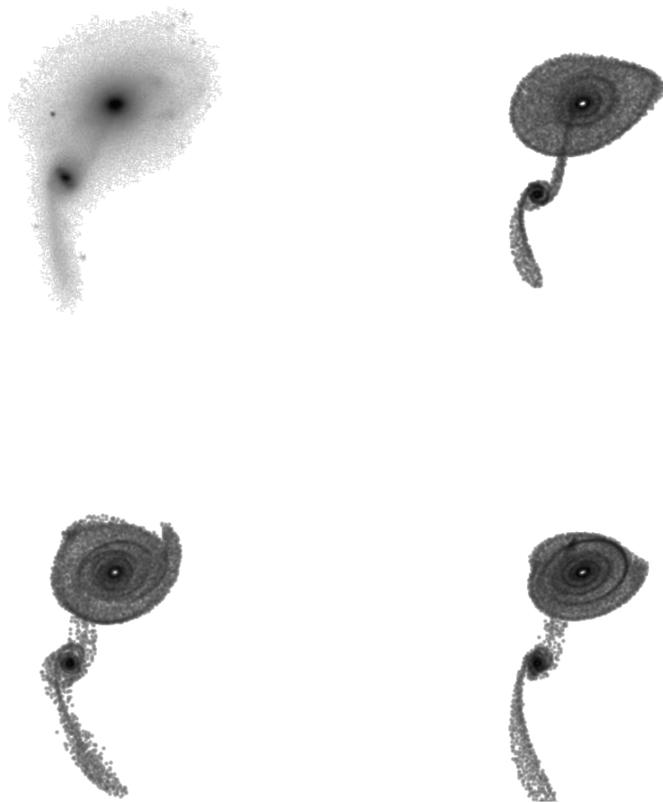


Figure 5.131: Target image and top 3 simulations for SDSS 587741722819493915

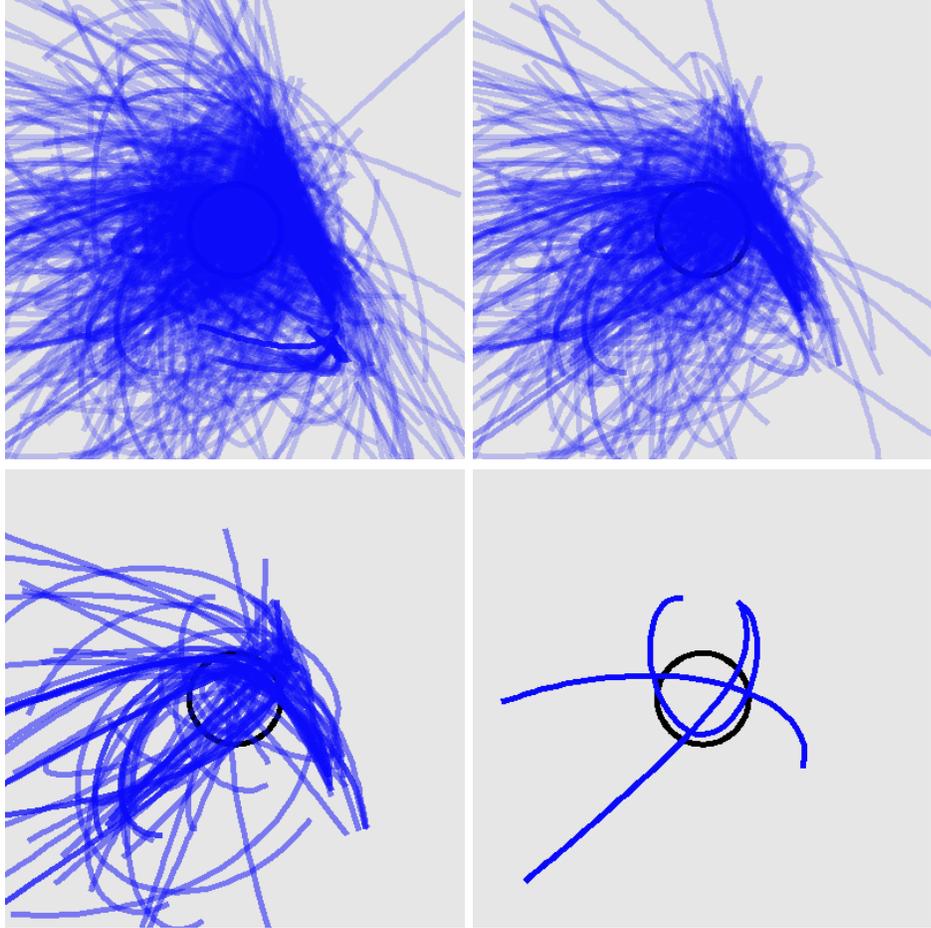


Figure 5.132: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587741722819493915

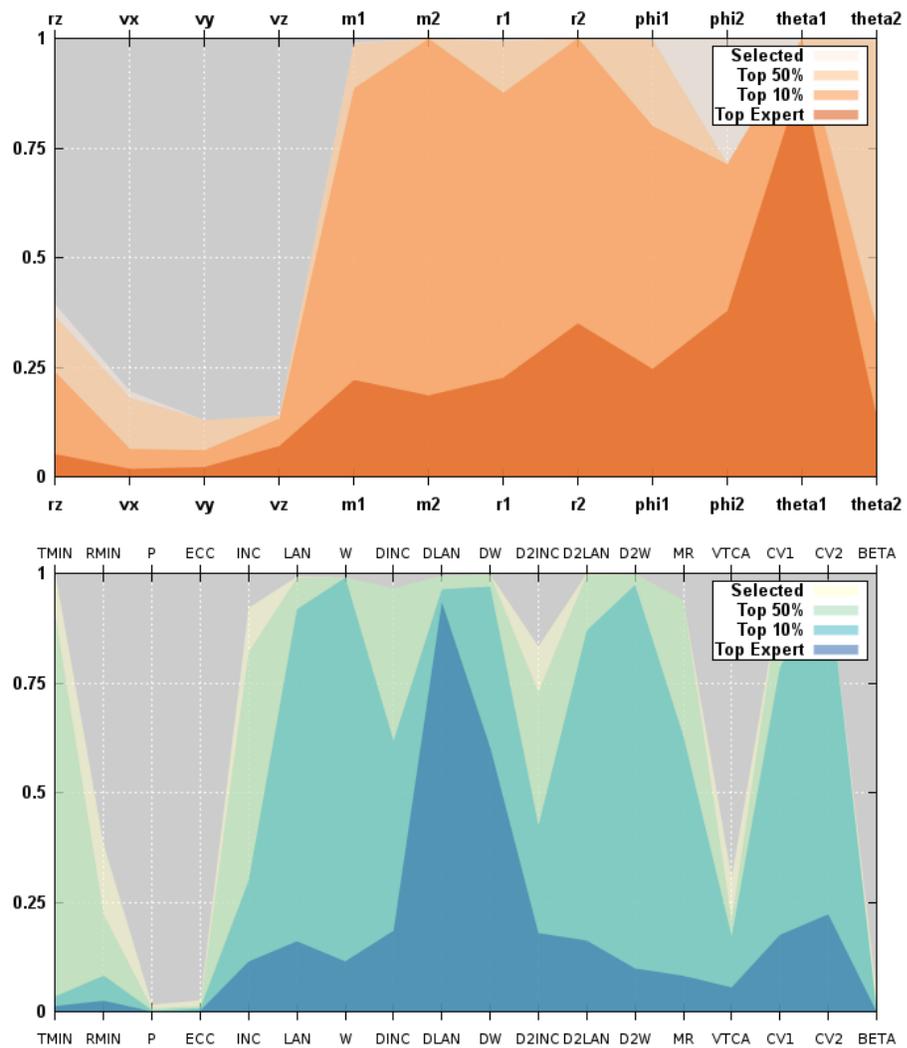


Figure 5.133: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587741722819493915

### 5.1.34 SDSS 587741817851674654

The NGC 2802/2803 pair does not exhibit clear tidal features. The smoothly declining fitness distribution is apparent. The top simulations do produce a morphological fuzziness, and in one case a tail, in the general area of distortion in the target image. The top trajectories are similar. The convergence of the simulation parameters is not good for masses, radii, or orientation angles.

Table 5.34: Identification Information and Merger Zoo summary for SDSS 587741817851674654.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587741817851674654	NGC 2802/2803, UGC 4897	9:16:41.4	+18:57:49.3	0.02914

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
26368	25584	784	145	13959	4632	9327

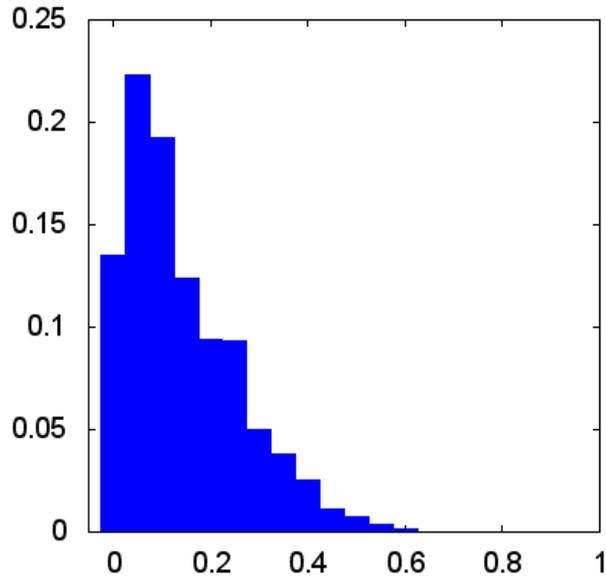


Figure 5.134: Relative frequency of fitness for all selected states of SDSS 587741817851674654

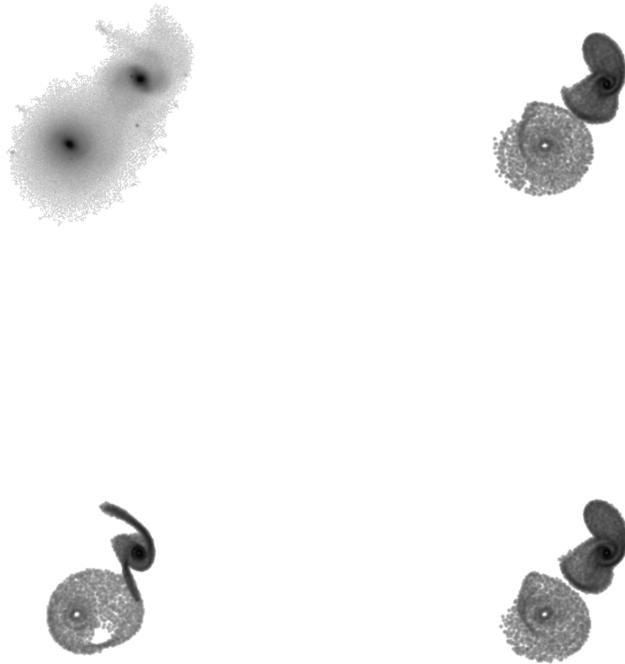


Figure 5.135: Target image and top 3 simulations for SDSS 587741817851674654

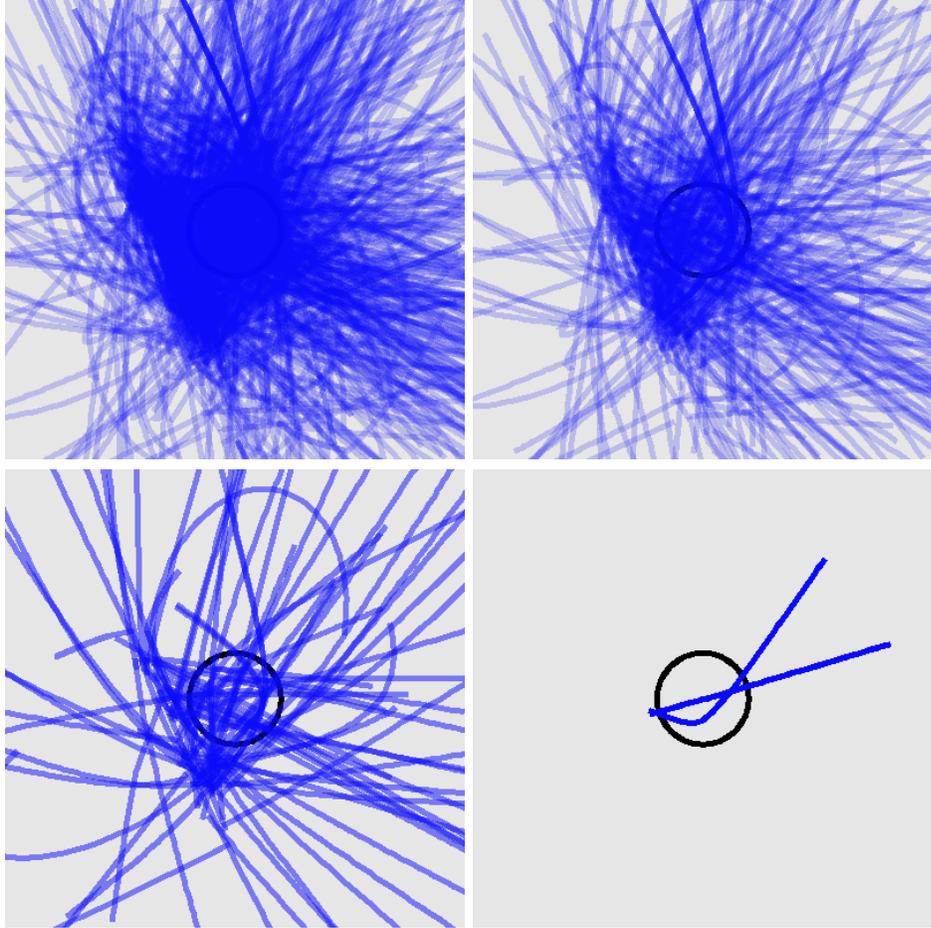


Figure 5.136: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587741817851674654

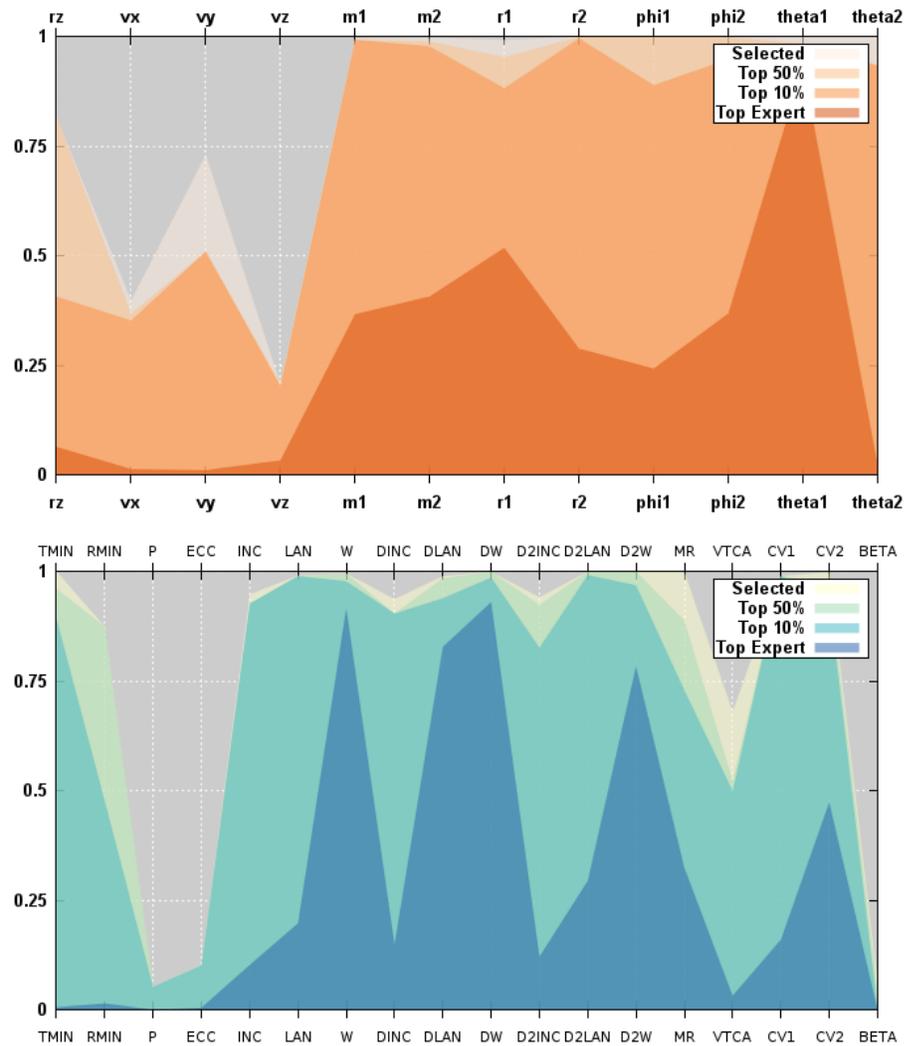


Figure 5.137: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587741817851674654

### 5.1.35 SDSS 587741829658181698

Another popular merger, Arp 301 was viewed in over 100000 simulations by the volunteers. The symmetric, stubby tails of the primary galaxy are well matched by the top simulations. The trajectories are similar. The convergence for simulation parameters is moderate to good except for  $\phi_1$ .

Table 5.35: Identification Information and Merger Zoo summary for SDSS 587741829658181698.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587741829658181698	Arp 301, UGC 6207	11:09:51.4	+24:15:41.8	0.020591

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
104801	103002	1799	317	31705	12970	18735

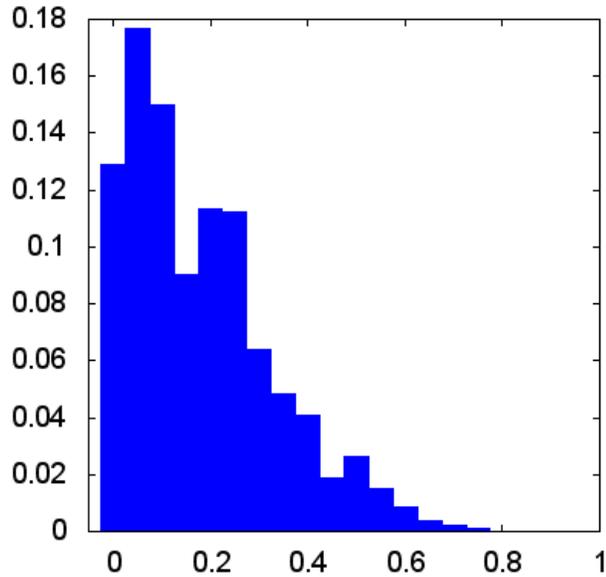


Figure 5.138: Relative frequency of fitness for all selected states of SDSS 587741829658181698

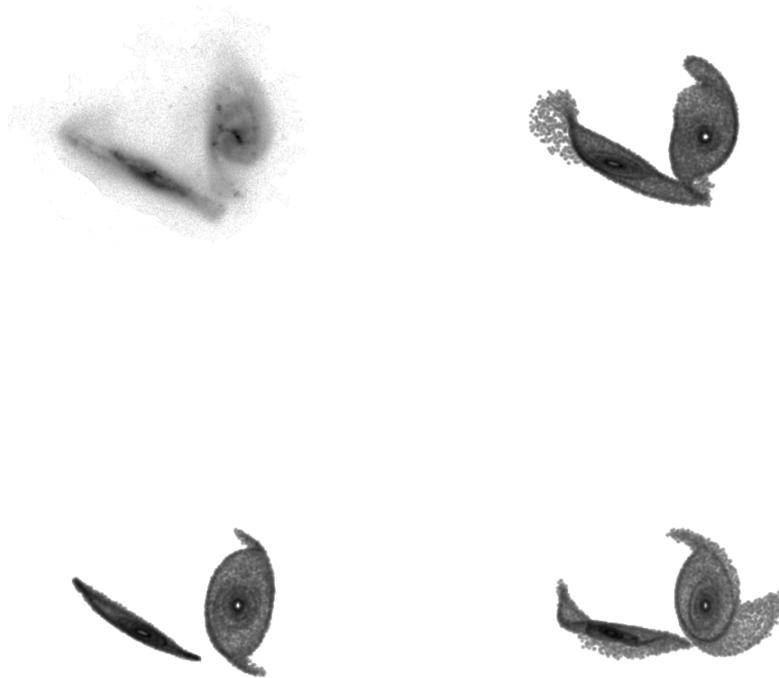


Figure 5.139: Target image and top 3 simulations for SDSS 587741829658181698

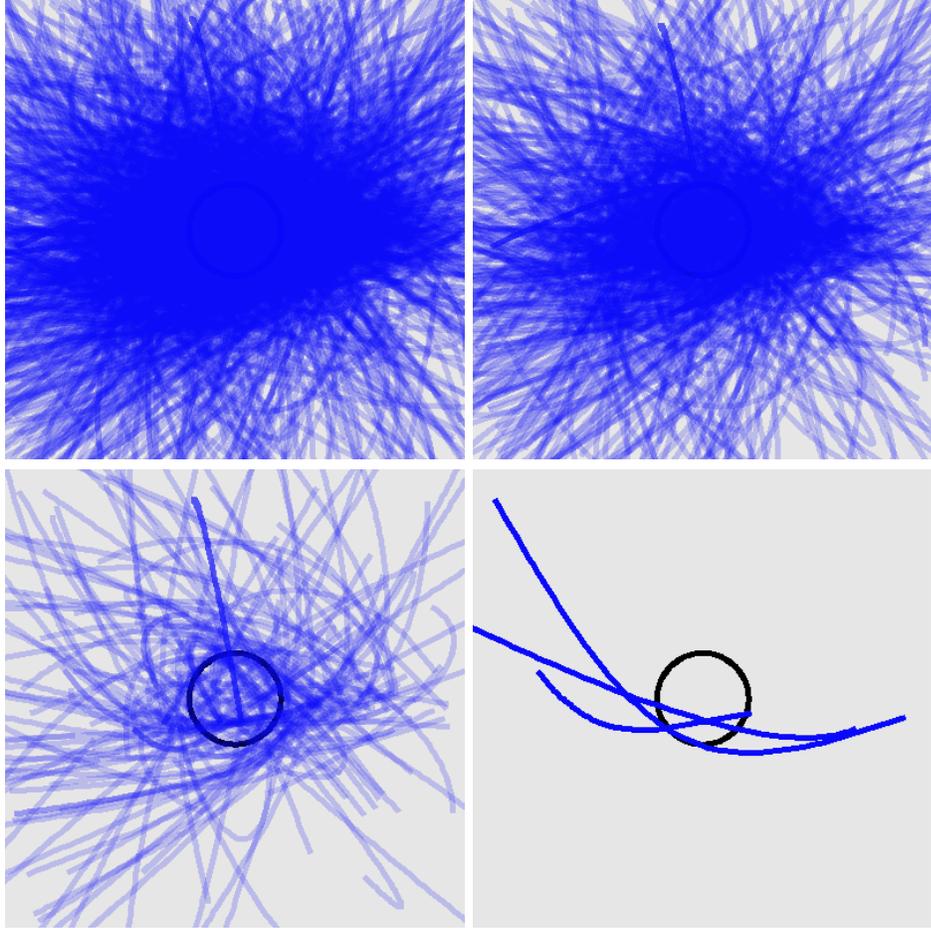


Figure 5.140: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587741829658181698

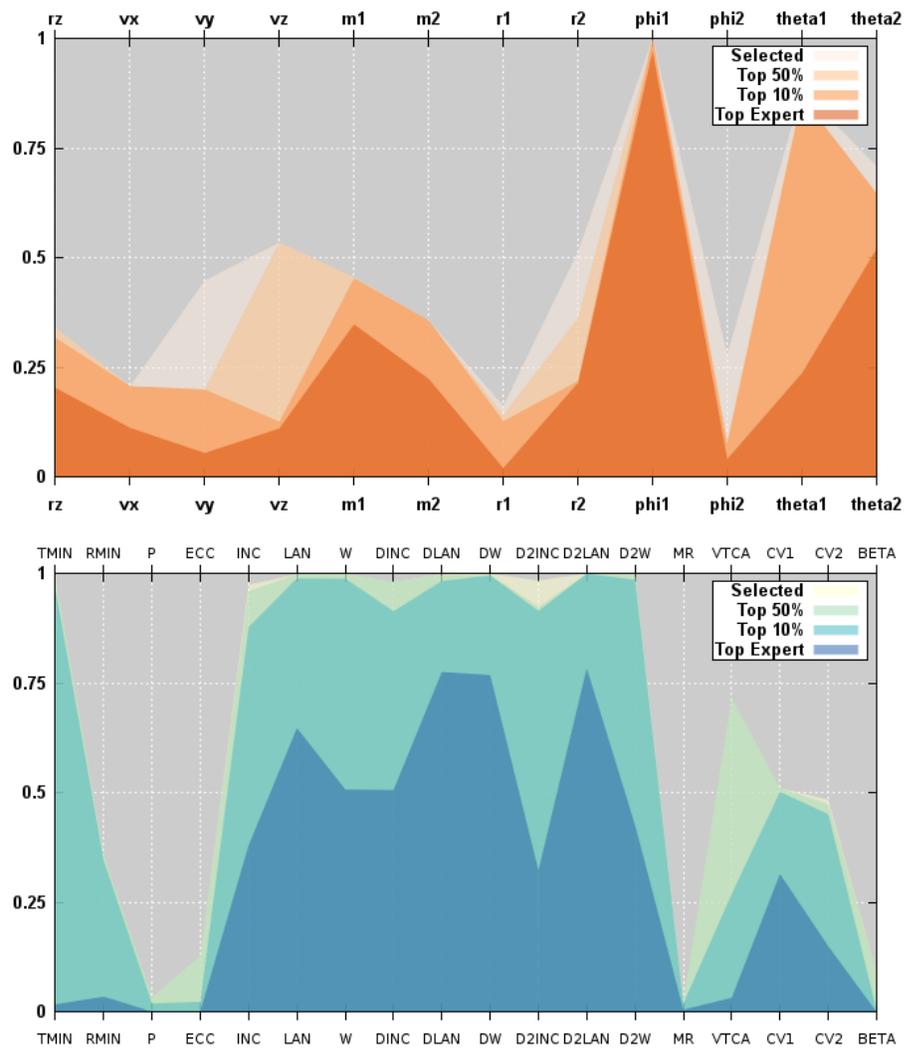


Figure 5.141: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587741829658181698

### 5.1.36 SDSS 587742010583941189

The tidal features are sometimes difficult to view for Arp 89. The top simulations recreated the subtle hook in the northern feature. The existence of a potential feature connected to the secondary was not well matched. Only two of the top three trajectories are similar. The convergence of the masses is poor.

Table 5.36: Identification Information and Merger Zoo summary for SDSS 587742010583941189.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587742010583941189	Arp 89, NGC 2648	8:42:39.8	+14:17:08.2	0.006871

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
22360	21193	1167	219	20663	11208	9455

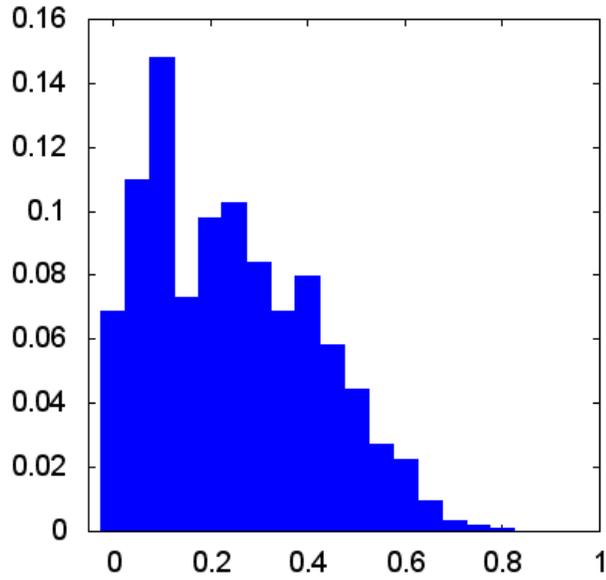


Figure 5.142: Relative frequency of fitness for all selected states of SDSS 587742010583941189



Figure 5.143: Target image and top 3 simulations for SDSS 587742010583941189

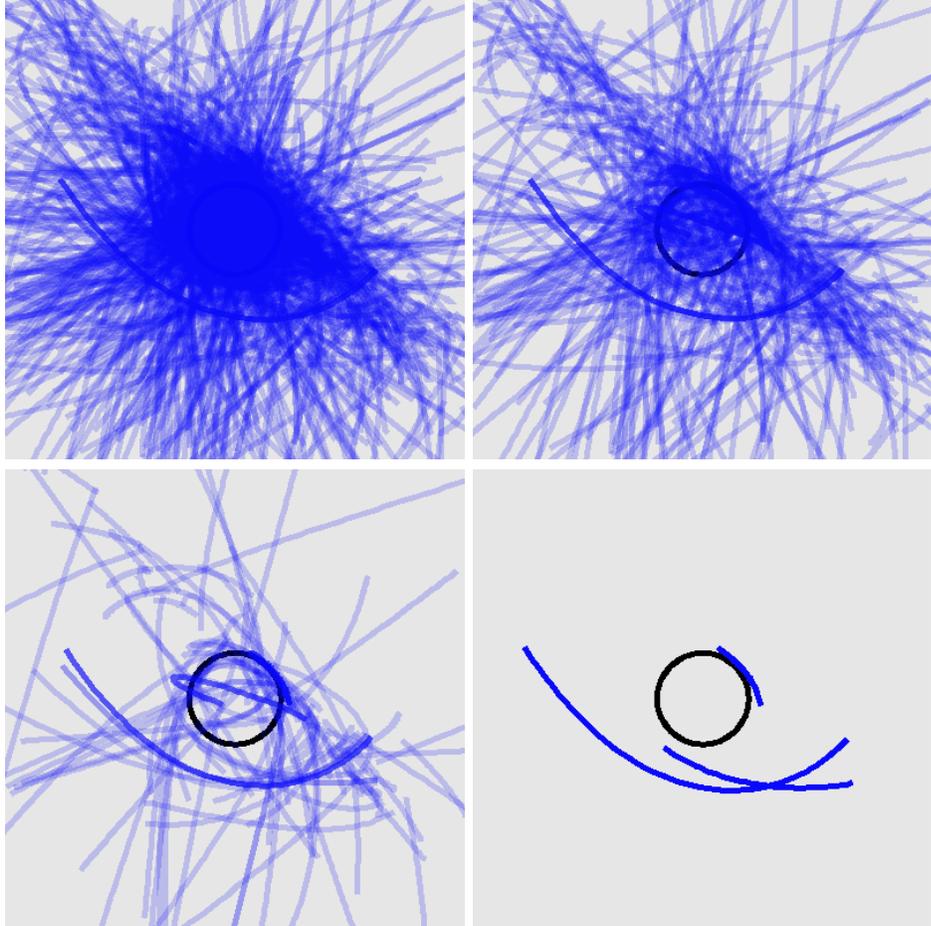


Figure 5.144: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587742010583941189

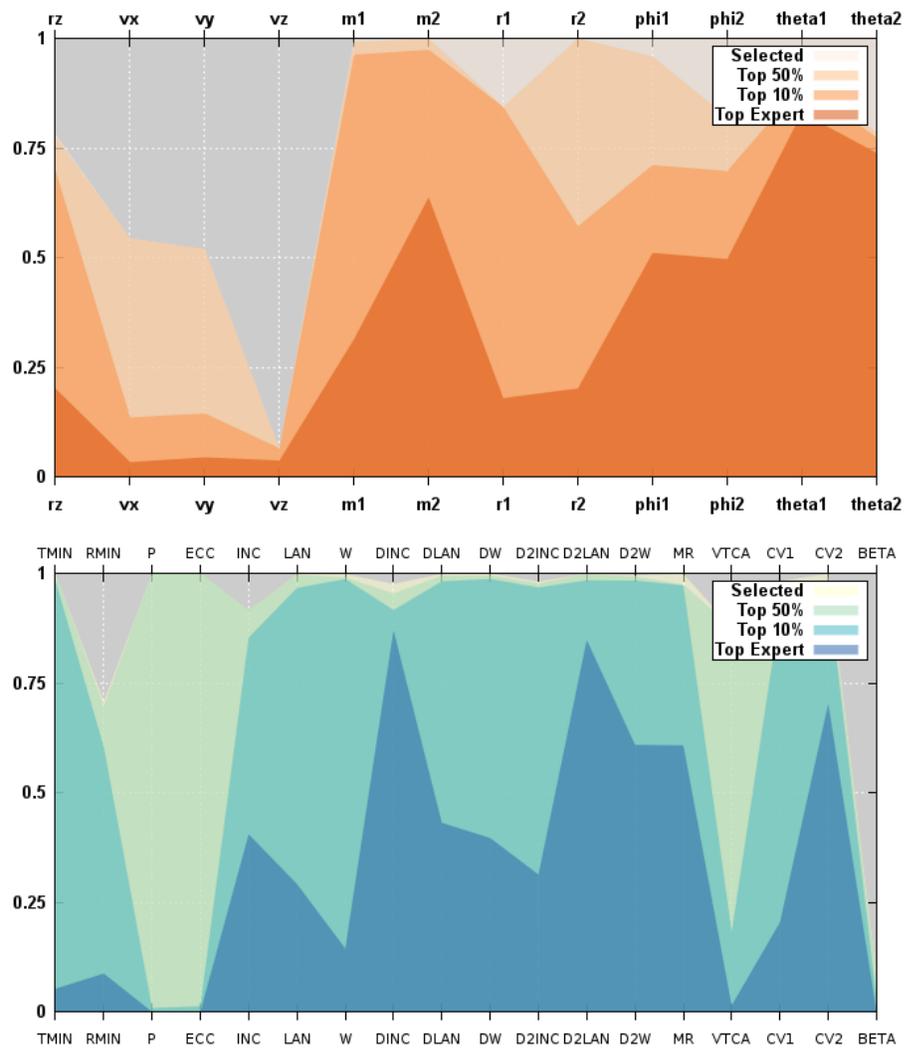


Figure 5.145: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587742010583941189

### 5.1.37 SDSS 587742014353702970

Arp 87 presents another L-shaped fitness distribution. The tidal tails of the primary galaxy are well matched by the volunteers. The trajectories are similar. The simulation parameters converge except for the  $\theta$  angles. The orbit parameters are reasonably well converged.

Table 5.37: Identification Information and Merger Zoo summary for SDSS 587742014353702970.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587742014353702970	Arp 87, NGC 3808	11:40:44.0	+22:25:45.8	0.023603

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
42154	41733	421	213	7672	2873	4799

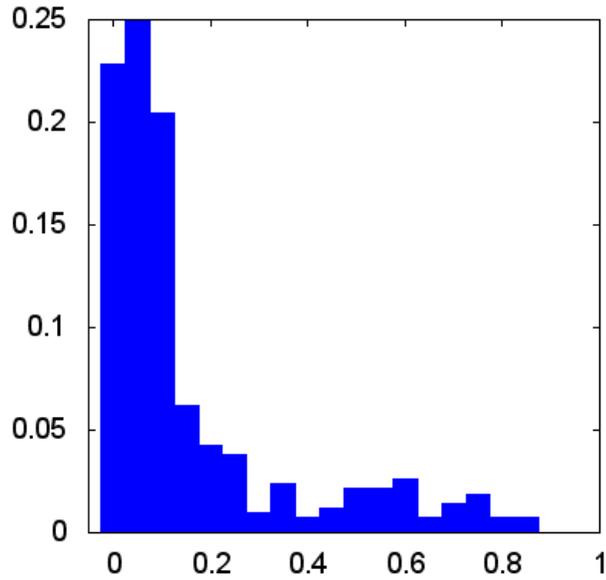


Figure 5.146: Relative frequency of fitness for all selected states of SDSS 587742014353702970

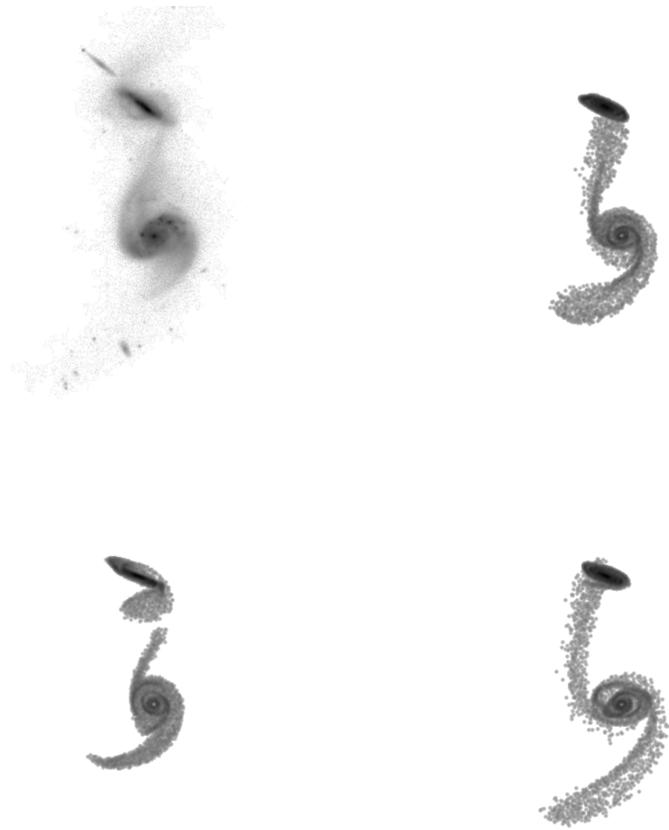


Figure 5.147: Target image and top 3 simulations for SDSS 587742014353702970

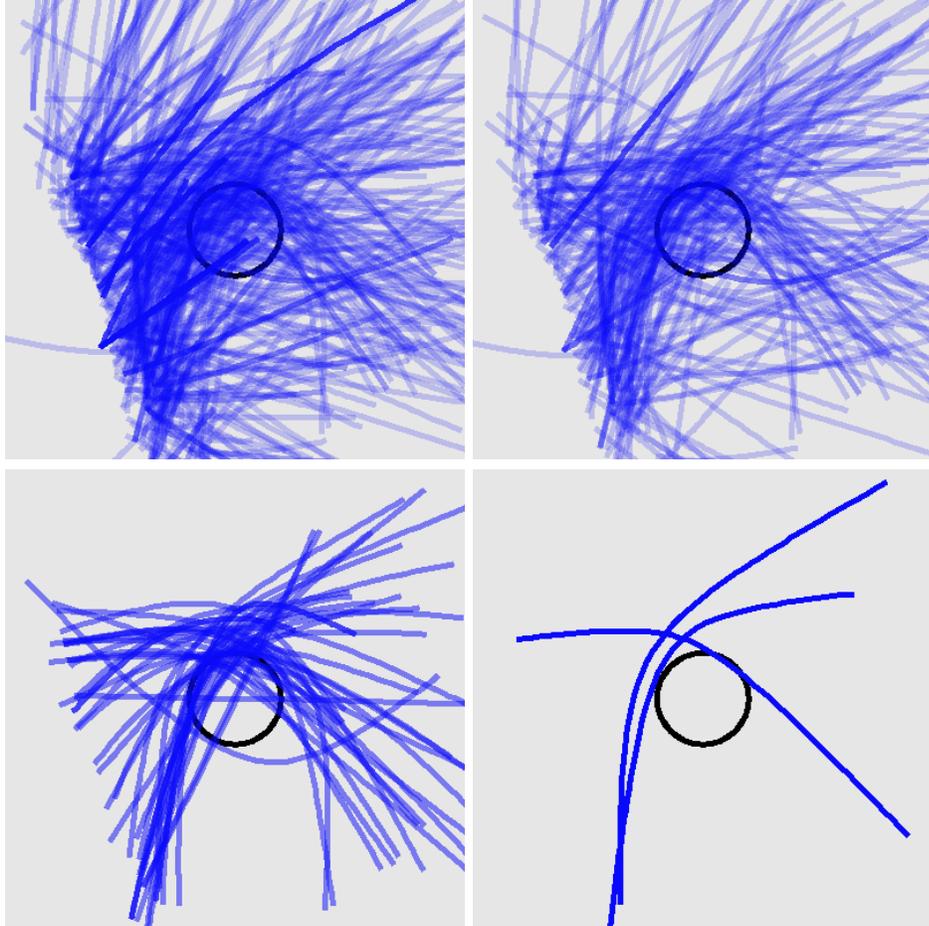


Figure 5.148: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587742014353702970

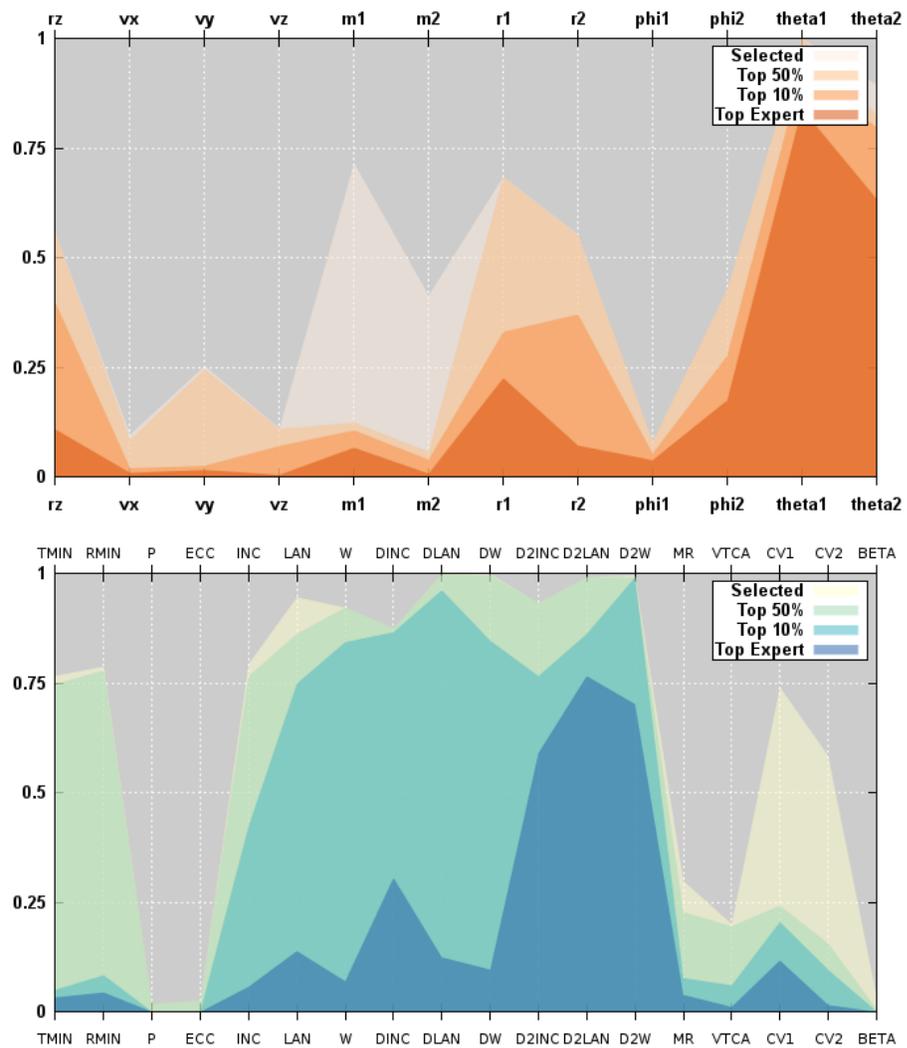


Figure 5.149: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587742014353702970

### 5.1.38 SDSS 587742571610243080

The L-shaped fitness distribution for Arp 191 is not as pronounced as it is for some targets. The western tidal tail is well matched in the simulations. One also predicts an eastern tail. The trajectories are similar. The simulation parameters are well constrained as well as the orbit values.

Table 5.38: Identification Information and Merger Zoo summary for SDSS 587742571610243080.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587742571610243080	Arp 191, UGC 6175	11:07:20.8	+18:25:58.6	0.026548

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
62068	61604	464	191	8589	3633	4956

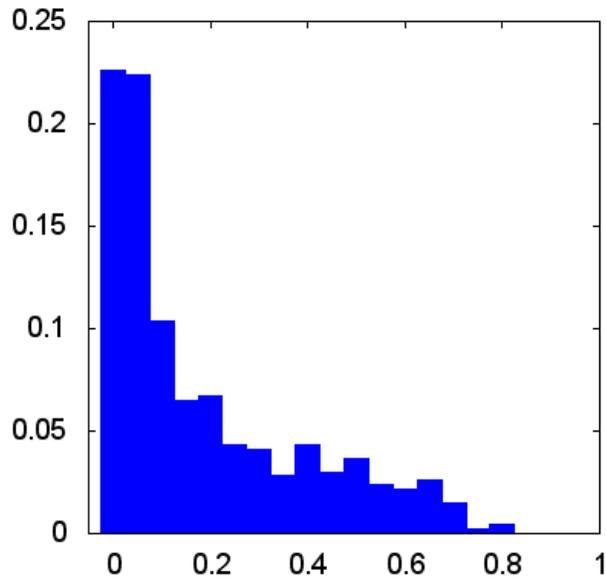


Figure 5.150: Relative frequency of fitness for all selected states of SDSS 587742571610243080

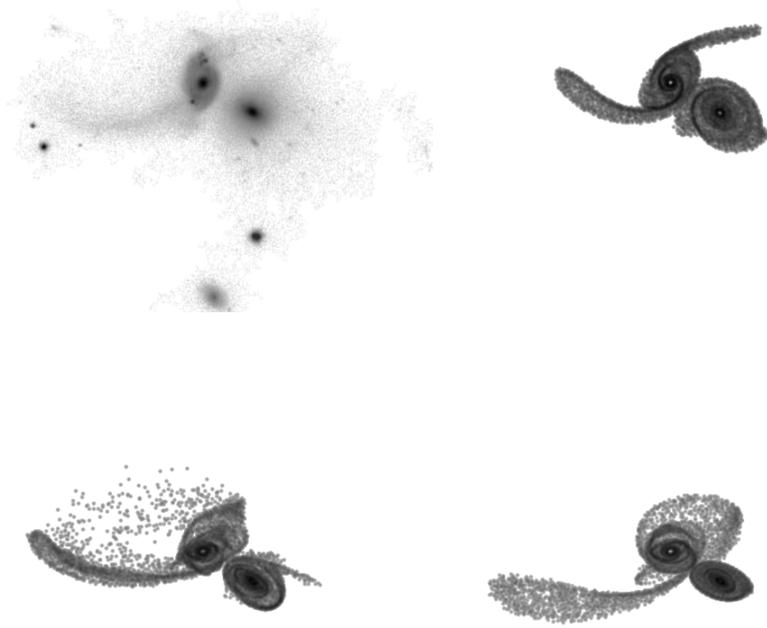


Figure 5.151: Target image and top 3 simulations for SDSS 587742571610243080

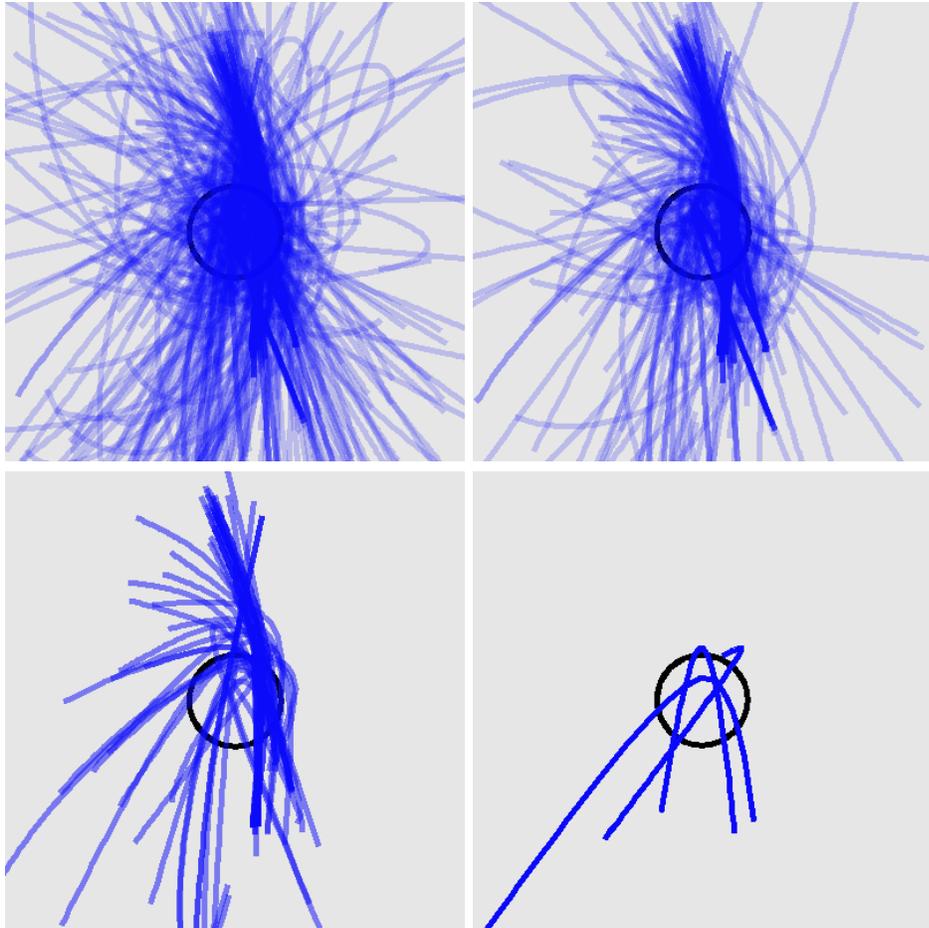


Figure 5.152: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587742571610243080

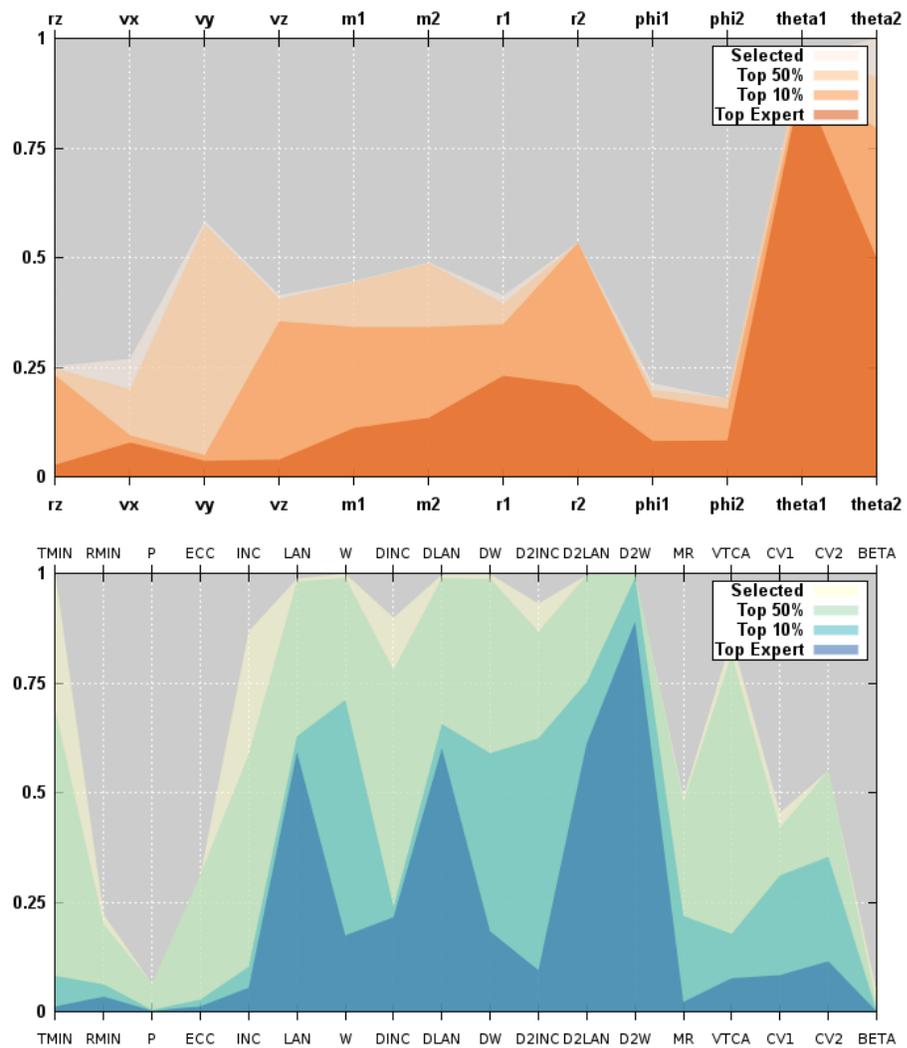


Figure 5.153: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587742571610243080

### 5.1.39 SDSS 587745402001817662

Arp 237 consists of two galaxies very close to one another. The fitness distribution is roughly L-shaped. The southern tail is well matched. The fuzziness of the top features is only roughly matched. The top trajectories are all closed orbit, multiple passages. The mass of the secondary galaxy is not well constrained nor is the size of the primary galaxy.

Table 5.39: Identification Information and Merger Zoo summary for SDSS 587745402001817662.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587745402001817662	Arp 237, UGC 5044	9:27:43.4	+12:17:14.1	0.028833

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
25173	24927	246	95	4662	1588	3074

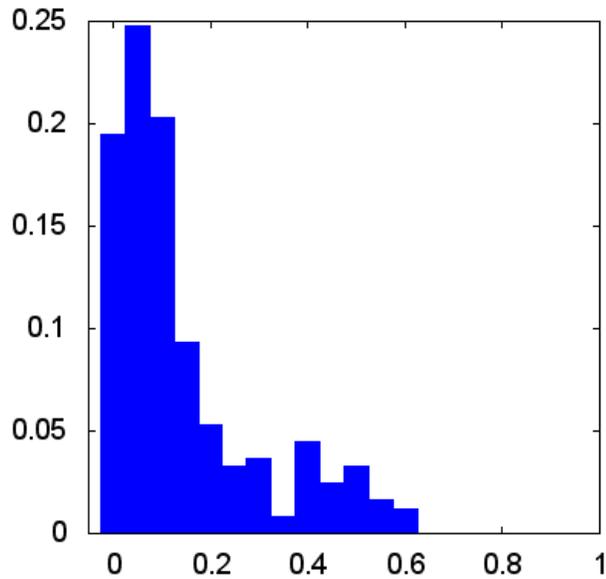


Figure 5.154: Relative frequency of fitness for all selected states of SDSS 587745402001817662

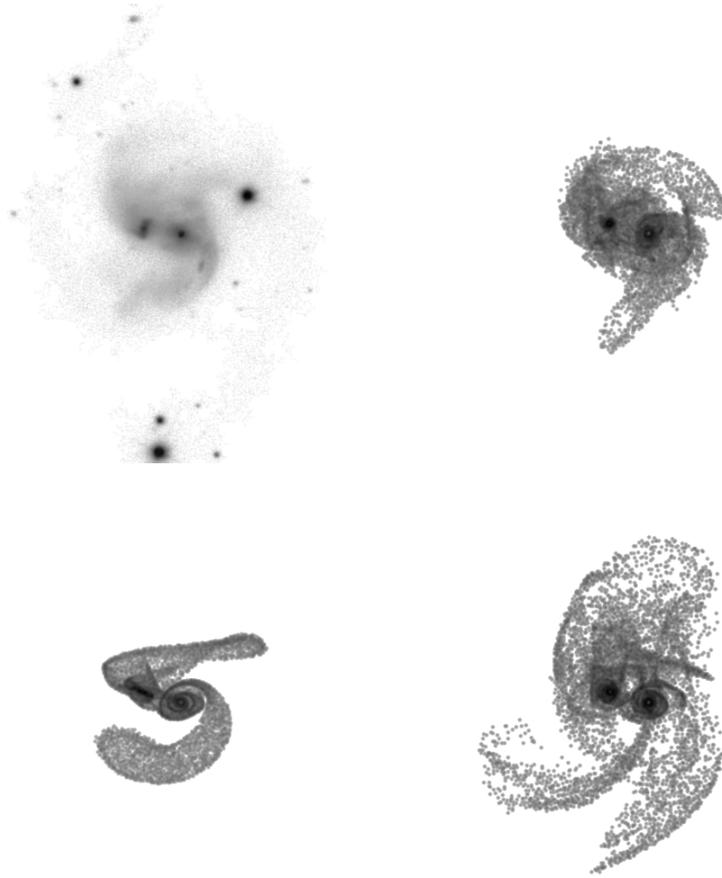


Figure 5.155: Target image and top 3 simulations for SDSS 587745402001817662

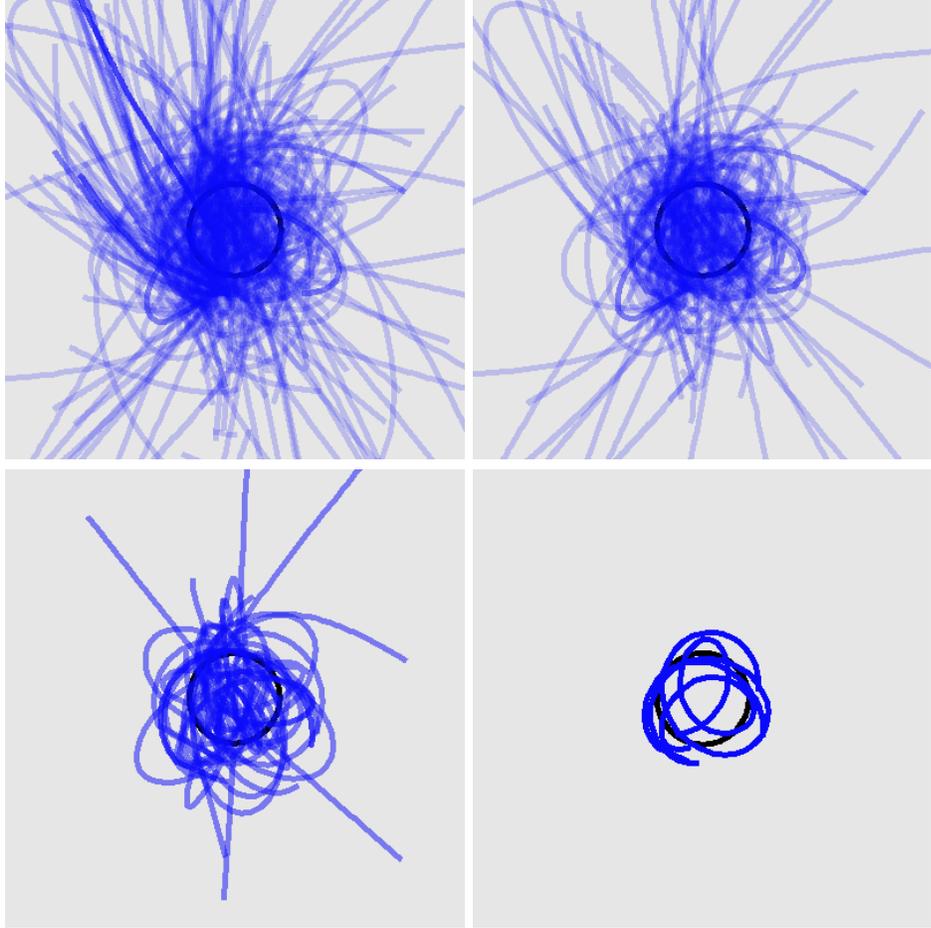


Figure 5.156: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587745402001817662

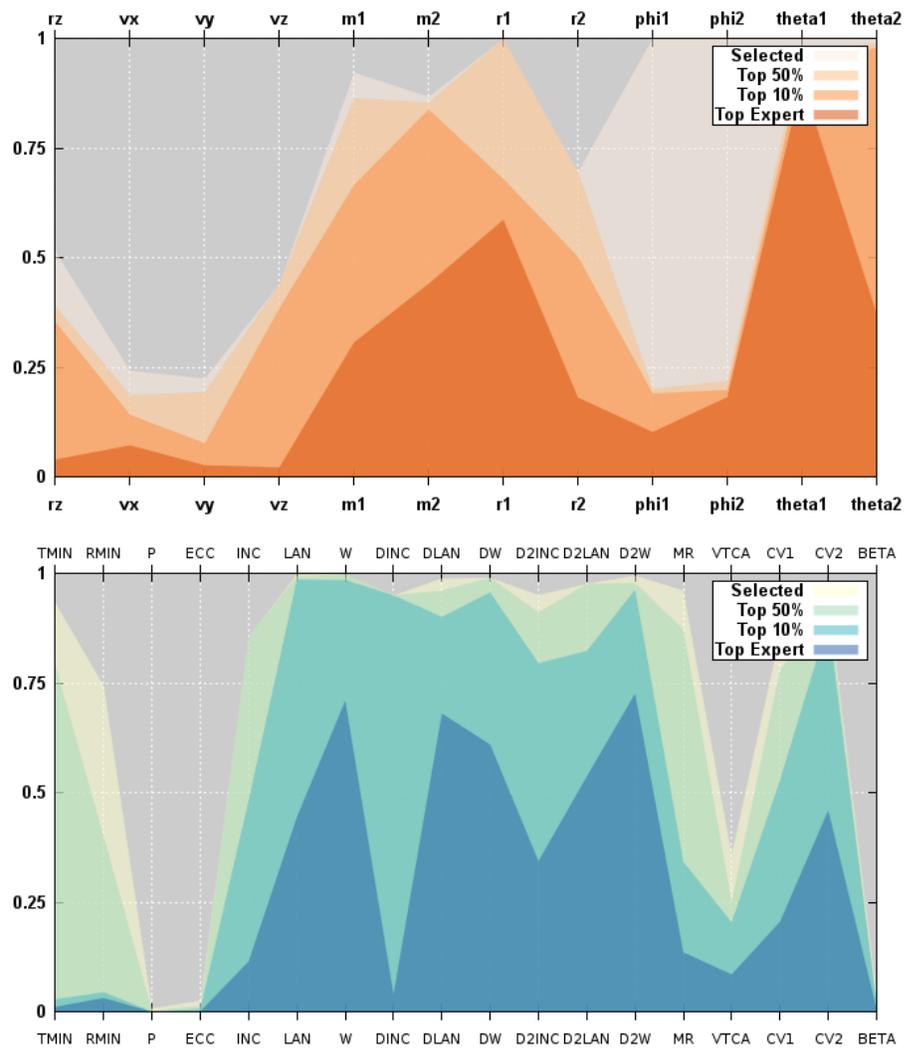


Figure 5.157: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587745402001817662

### 5.1.40 SDSS 587746029596311590

Arp 181 was a popular target with over 126000 simulation views. The sharp, thin tidal features of the primary are moderately matched. The internal spiral structure of the secondary is only loosely matched by the density enhancements in the simulations. The trajectories are similar. Other than the mass of the primary, the simulation parameters and orbit parameters are well converged.

Table 5.40: Identification Information and Merger Zoo summary for SDSS 587746029596311590.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587746029596311590	Arp 181, NGC 3212	10:28:16.6	+79:49:24.5	0.032402

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
126303	124050	2253	370	39427	15970	23457

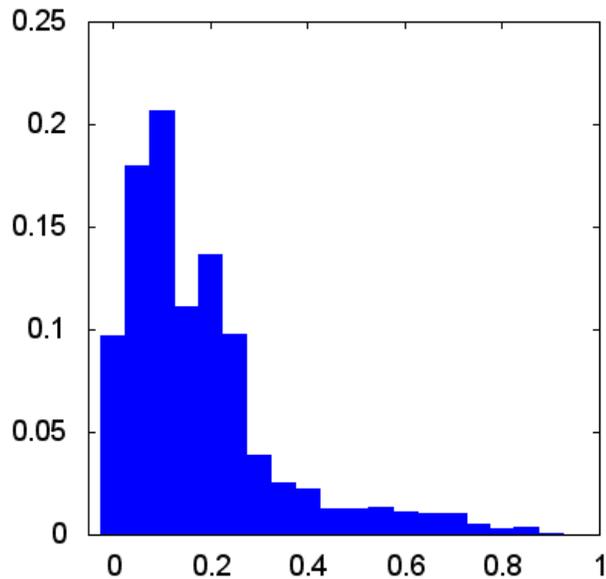


Figure 5.158: Relative frequency of fitness for all selected states of SDSS 587746029596311590

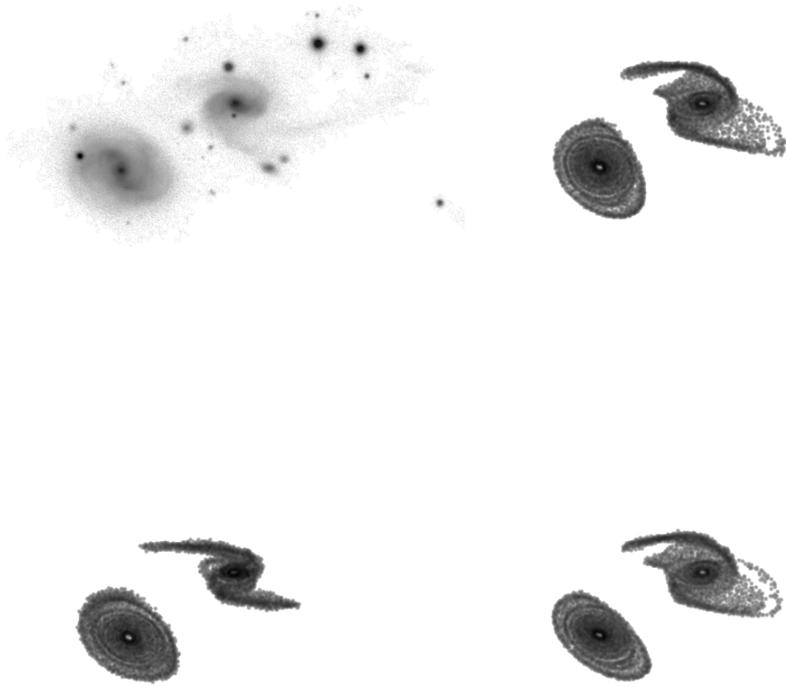


Figure 5.159: Target image and top 3 simulations for SDSS 587746029596311590

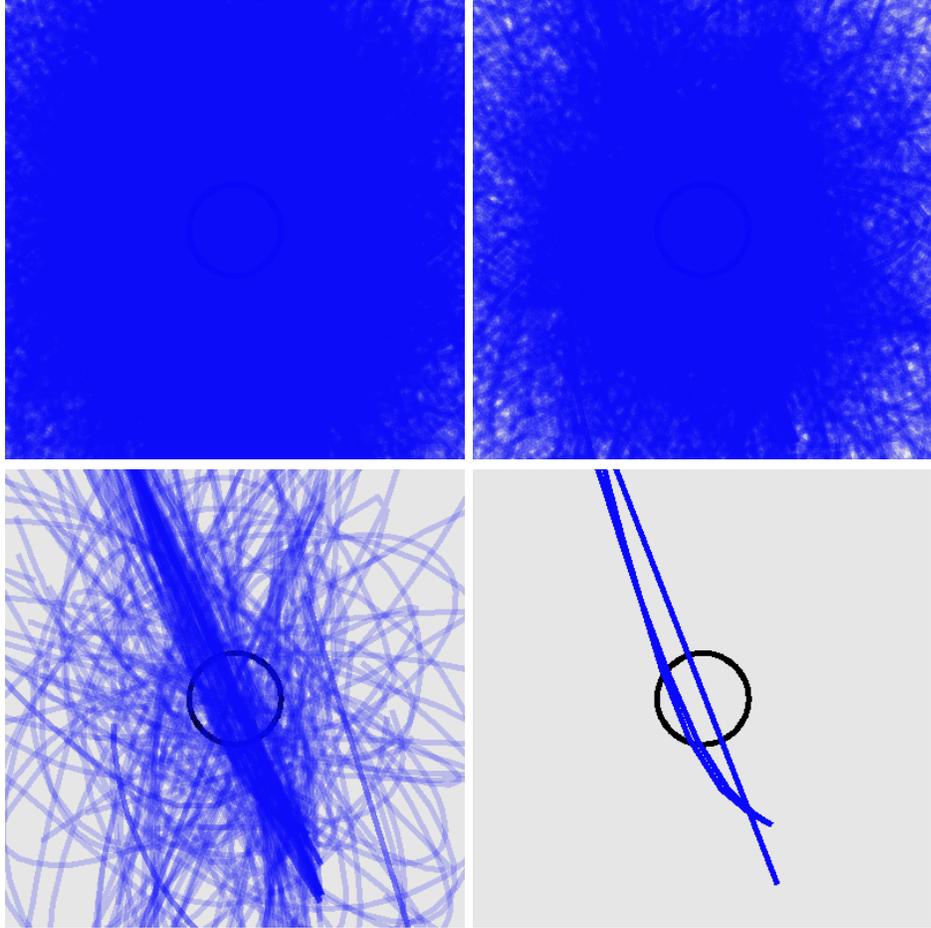


Figure 5.160: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587746029596311590

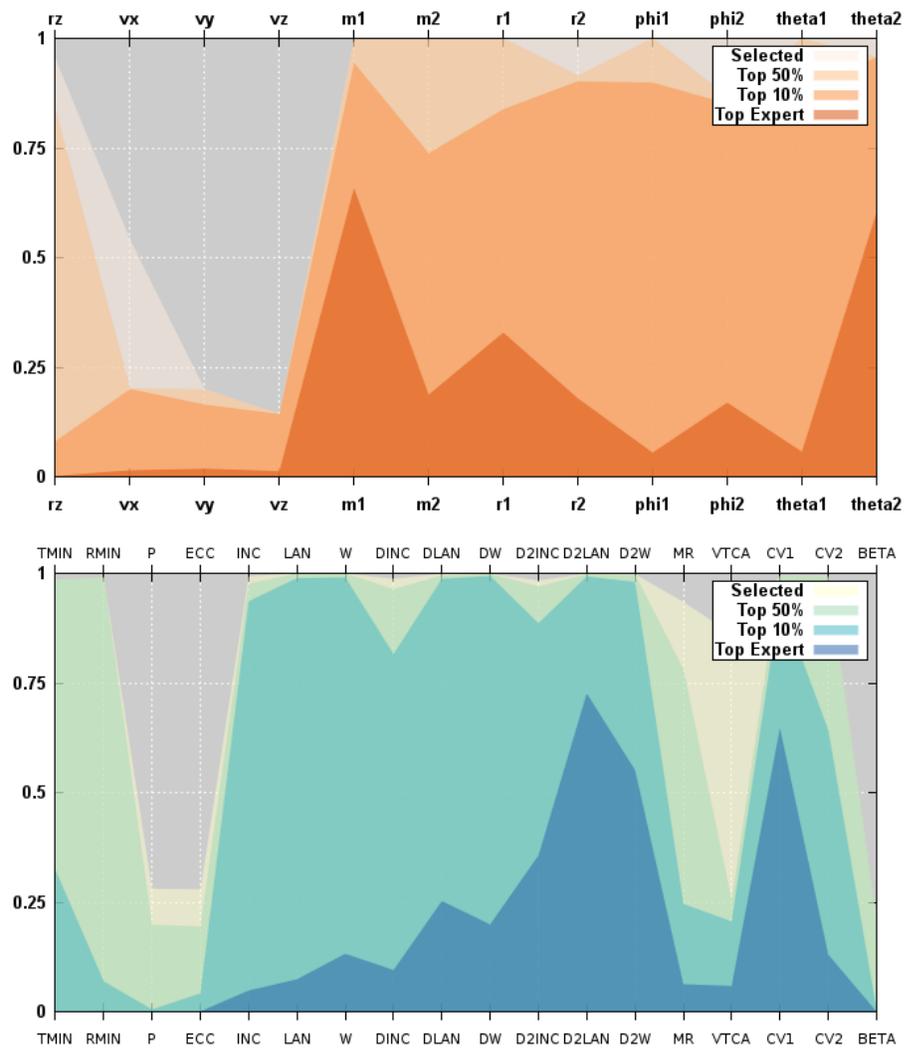


Figure 5.161: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587746029596311590

### 5.1.41 SDSS 58801124116422756

Another L-shaped fitness distribution is present for Arp 238. This system presents two galaxies each with obvious tidal features. The top simulations recreate the top and bottom tails. They also match some of the distortion near the disk of the secondary galaxy. The trajectories are very similar. The simulation parameters are very well converged.

Table 5.41: Identification Information and Merger Zoo summary for SDSS 58801124116422756.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 58801124116422756	Arp 238, UGC 8335	13:15:31.1	+62:07:45.1	0.030788

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
49305	48767	538	118	9796	3790	6006

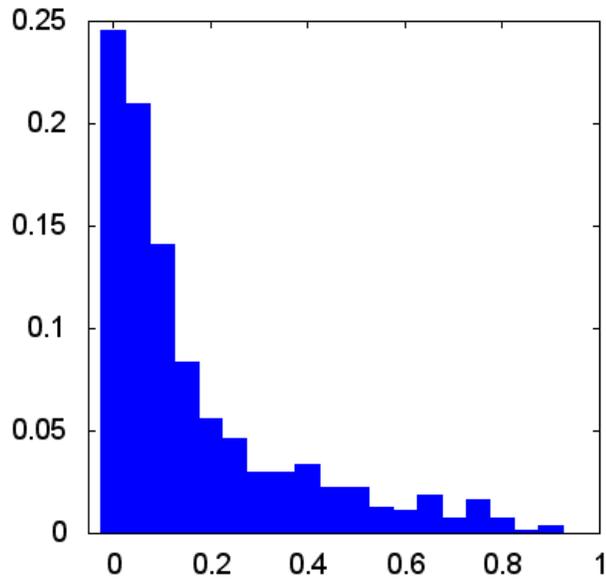


Figure 5.162: Relative frequency of fitness for all selected states of SDSS 58801124116422756

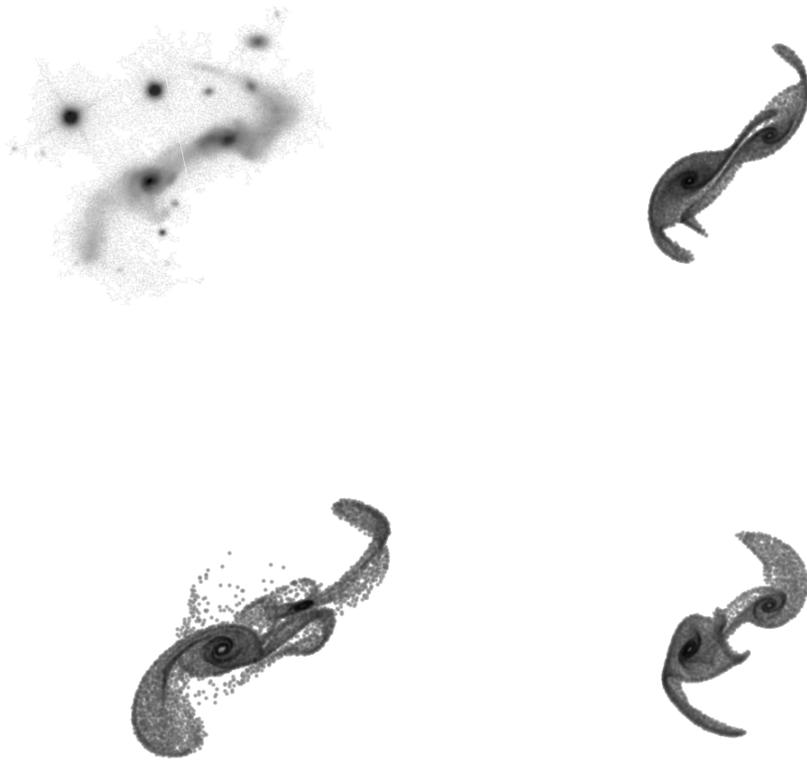


Figure 5.163: Target image and top 3 simulations for SDSS 588011124116422756

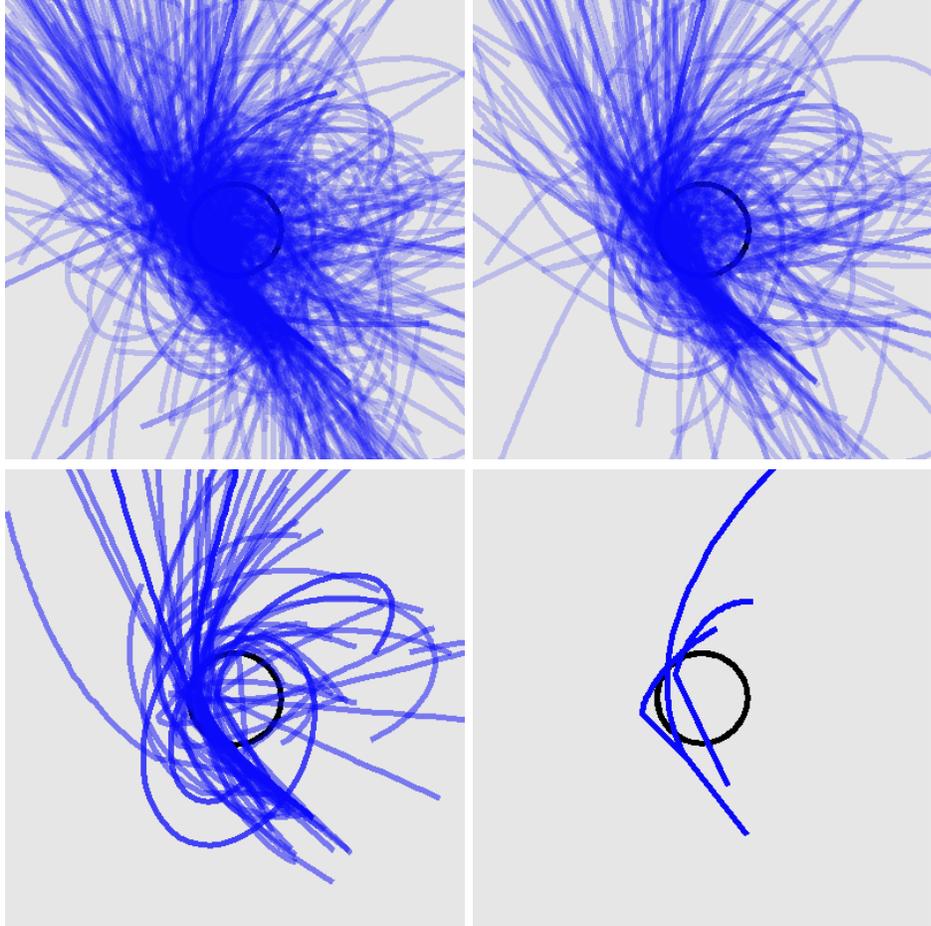


Figure 5.164: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 588011124116422756

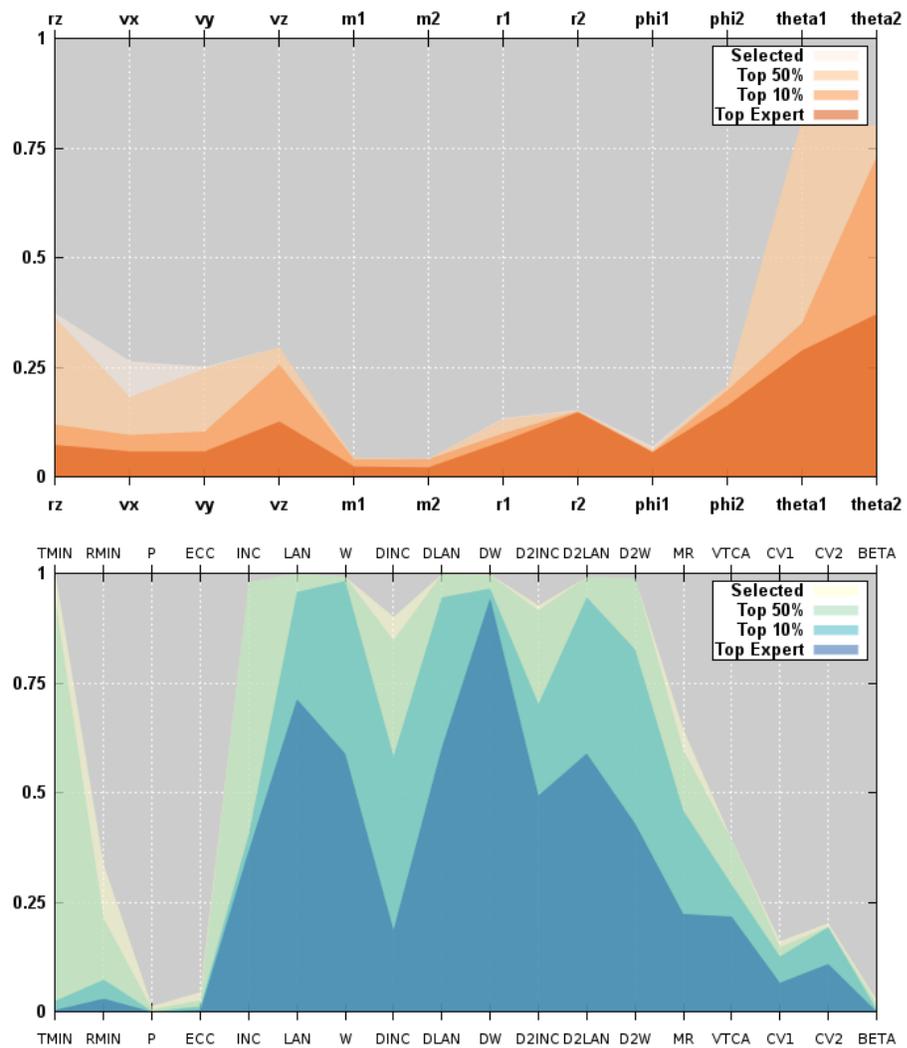


Figure 5.165: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 588011124116422756

### 5.1.42 SDSS 588013383816904792

This target has a roughly L-shaped fitness distribution. Both galaxies have distinct, though not necessarily symmetric, tidal features. The top simulations are all good matches. The trajectories are similar, but one of them shows a sharp kink that is probably due to an error in the smoothing done to generate the plot. The simulation parameters are very well converged. The convergence of the orbit parameters is better than in most other systems.

Table 5.42: Identification Information and Merger Zoo summary for SDSS 588013383816904792.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 588013383816904792	2MASS 12043959+5257265	12:04:39.5	+52:57:25.8	0.080803

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
59488	59132	356	122	6641	2709	3932

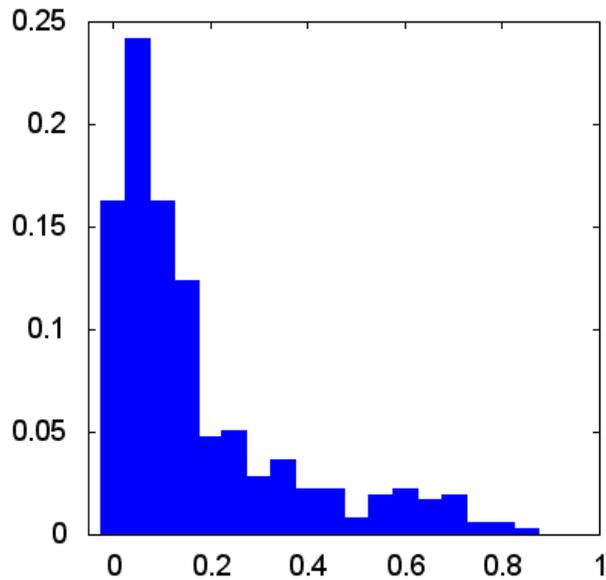


Figure 5.166: Relative frequency of fitness for all selected states of SDSS 588013383816904792



Figure 5.167: Target image and top 3 simulations for SDSS 588013383816904792

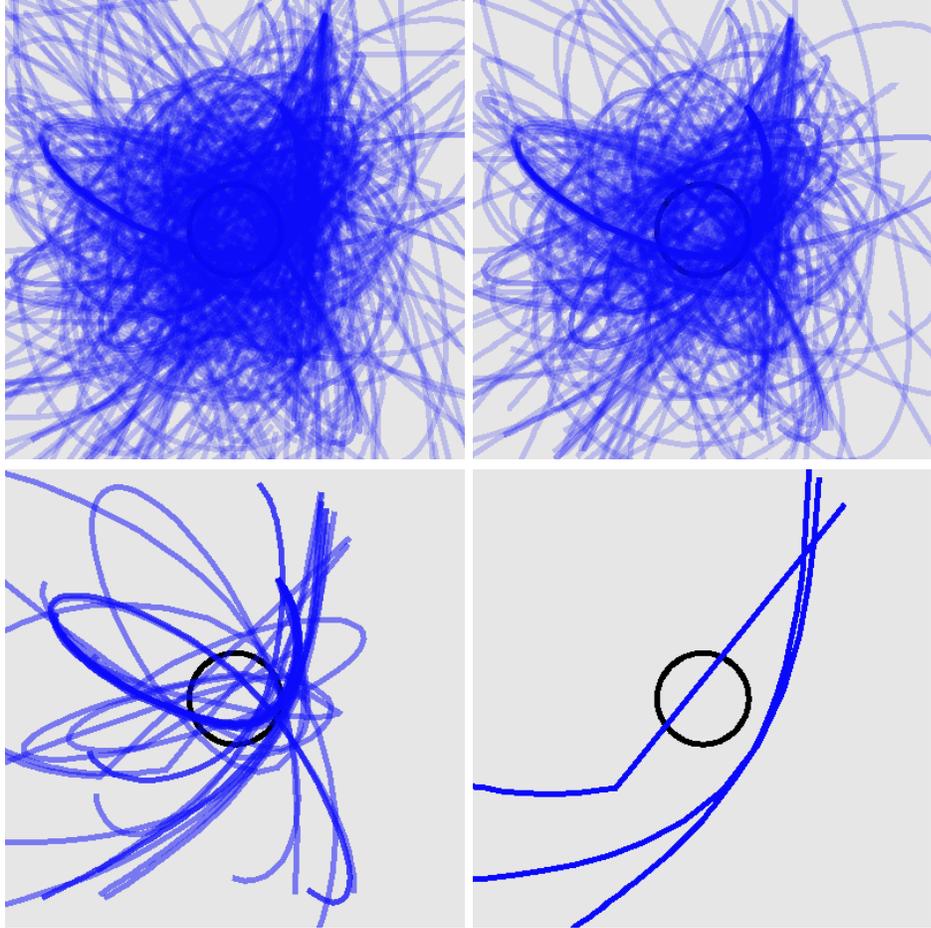


Figure 5.168: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 588013383816904792

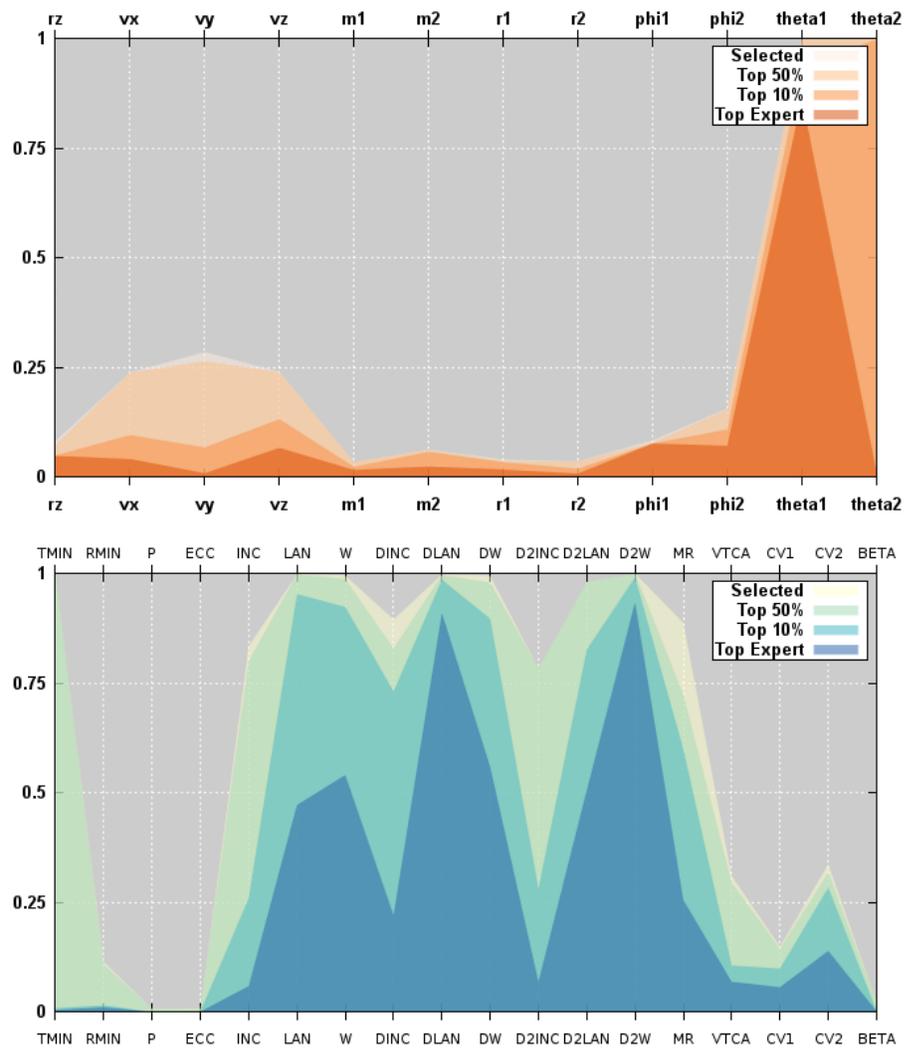


Figure 5.169: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 588013383816904792

### 5.1.43 SDSS 588017604696408086

The fitness distribution is not as L-shaped as it is for others. Arp 297 has a long, thin tail that wraps around the primary galaxy. The volunteers did a very good job matching that feature. The top trajectories are similar. There is good convergence of simulation and orbital parameters. This target differs from most in that the disk orientation angles are well constrained.

Table 5.43: Identification Information and Merger Zoo summary for SDSS 588017604696408086.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 588017604696408086	Arp 297, 5752/5754 NGC	14:45:19.6	+38:43:52.5	0.015214

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
37553	36934	619	142	11201	5446	5755

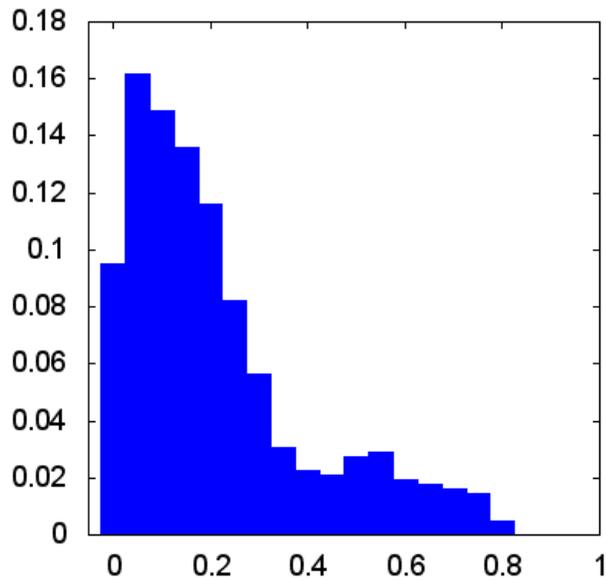


Figure 5.170: Relative frequency of fitness for all selected states of SDSS 588017604696408086

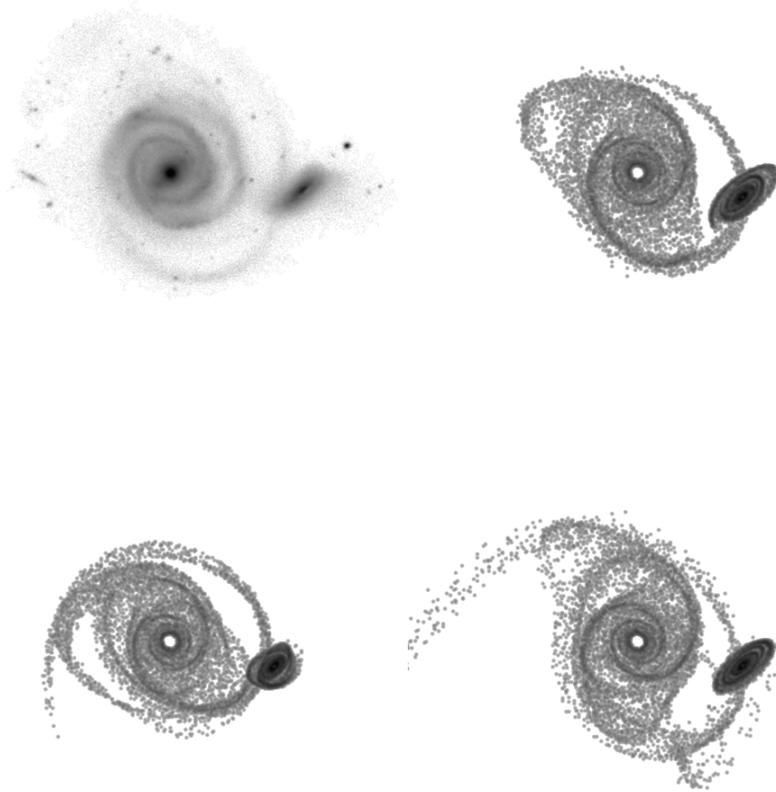


Figure 5.171: Target image and top 3 simulations for SDSS 588017604696408086

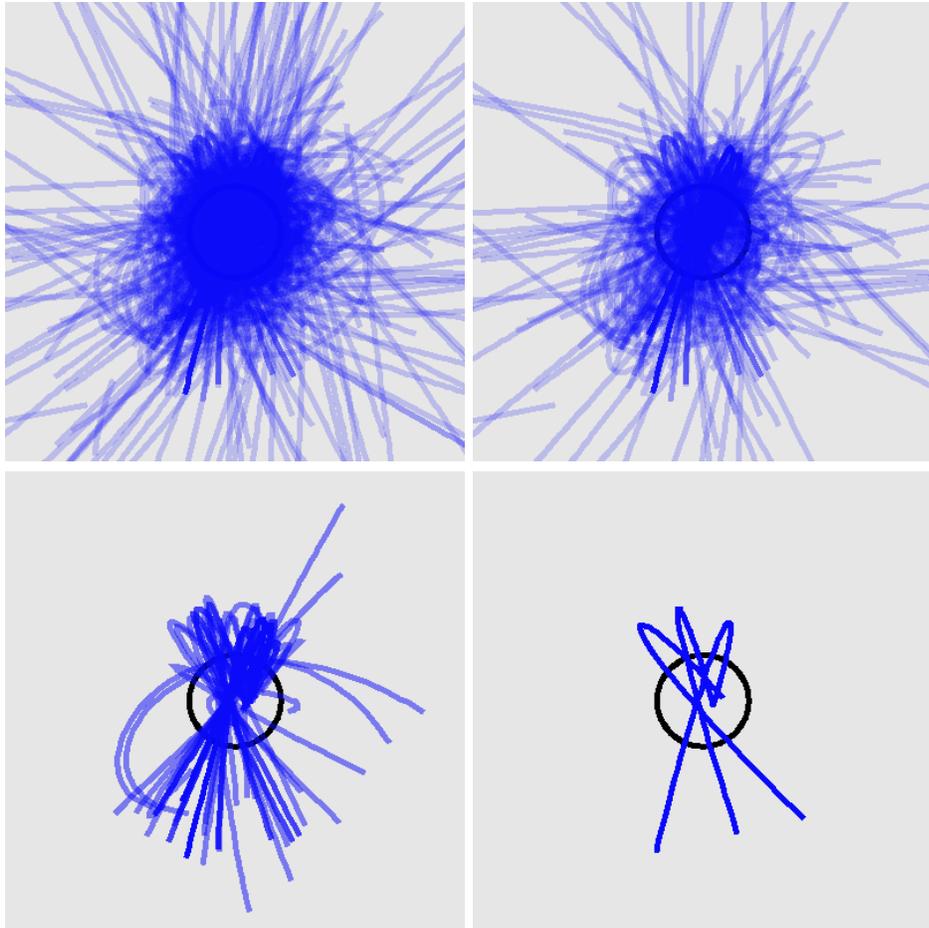


Figure 5.172: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 588017604696408086

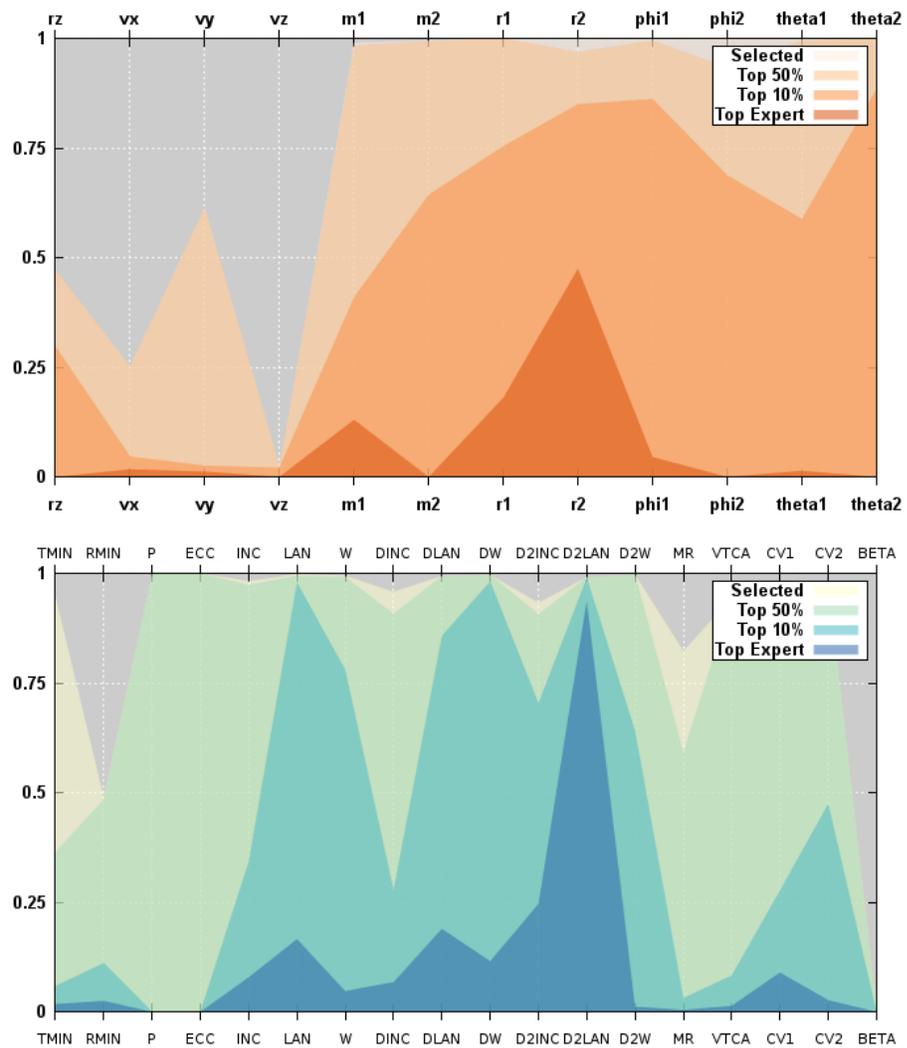


Figure 5.173: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 588017604696408086

### 5.1.44 SDSS 588017604696408195

The NGC 5753/5755 pair of galaxies only displays strong tidal features on the larger primary. The simulation images demonstrate the volunteers ability to achieve a modest match to the tidal tails. The northern tail has a “kink” that is not well matched. Two of the top trajectories are similar, with the lower left panel showing a family of paths. The simulation parameters for the disks have moderate to poor convergence.

Table 5.44: Identification Information and Merger Zoo summary for SDSS 588017604696408195.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 588017604696408195	NGC 5753/5755, UGC 9507	14:45:18.8	+38:48:20.6	0.032099

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
51971	51386	585	195	10717	4578	6139

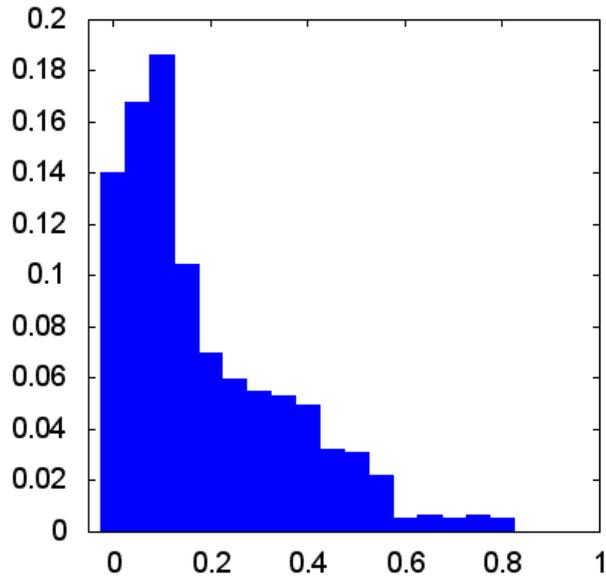


Figure 5.174: Relative frequency of fitness for all selected states of SDSS 588017604696408195



Figure 5.175: Target image and top 3 simulations for SDSS 588017604696408195

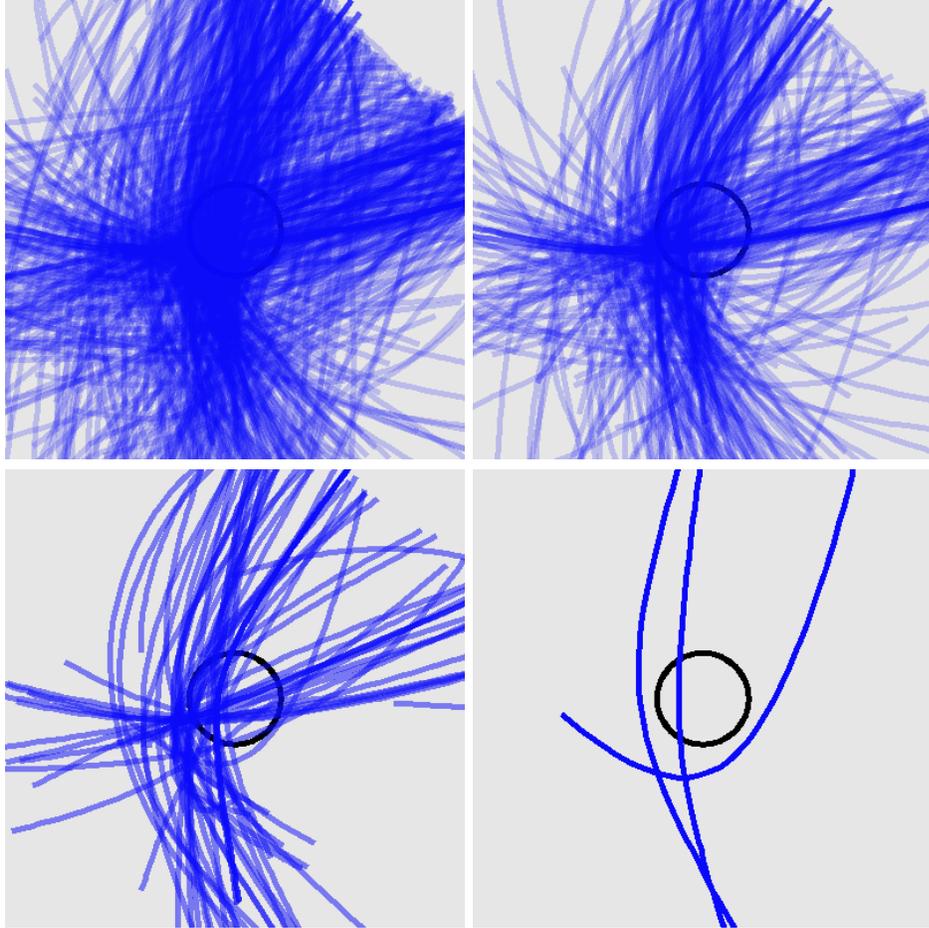


Figure 5.176: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 588017604696408195

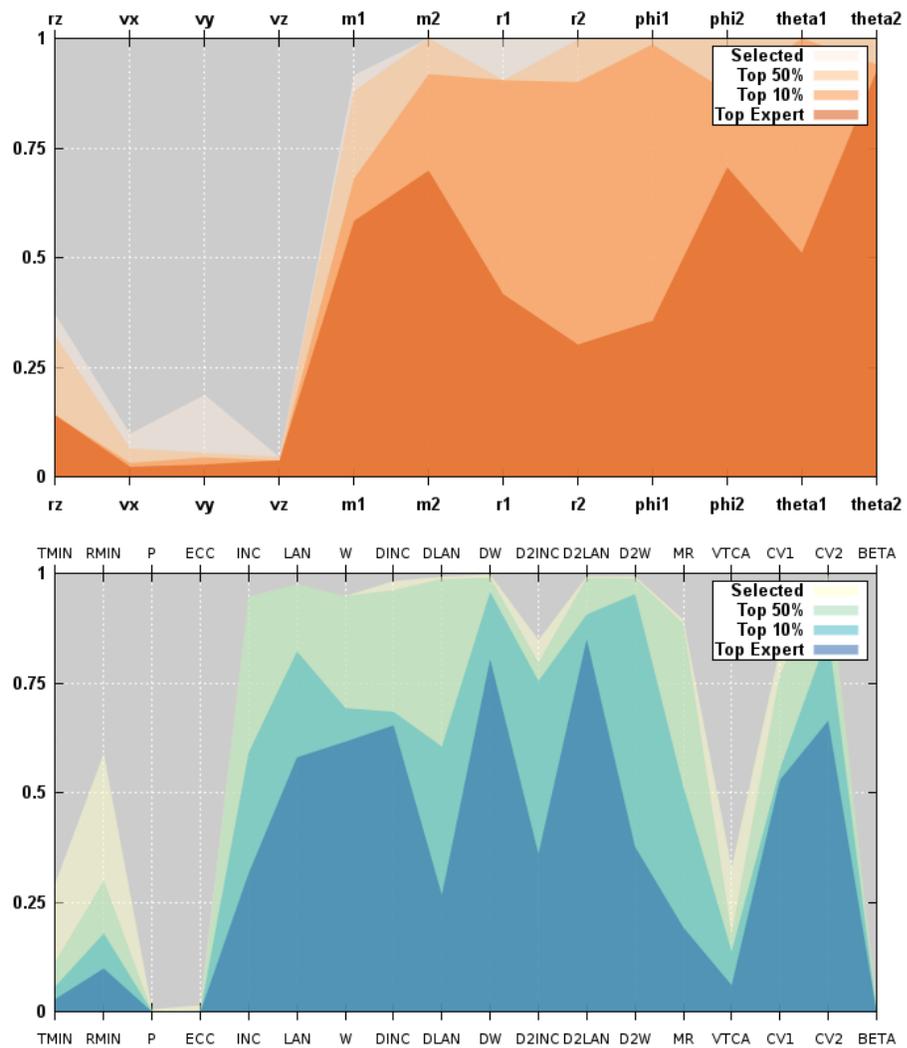


Figure 5.177: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 588017604696408195

### 5.1.45 SDSS 588017702948962343

The fitness distribution for Arp 173 presents a rough L shape. The very thin tidal features of the primary are well matched, including the curvature and the tip of the southern tail. The top trajectories are not very similar. The convergence for orbit and simulation parameters is poor. Here we have a plausible morphological match, but poor convergence.

Table 5.45: Identification Information and Merger Zoo summary for SDSS 588017702948962343.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 588017702948962343	Arp 173, UGC 9561	14:51:29.2	+9:20:05.3	0.029441

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
28770	28359	411	94	7546	3151	4395

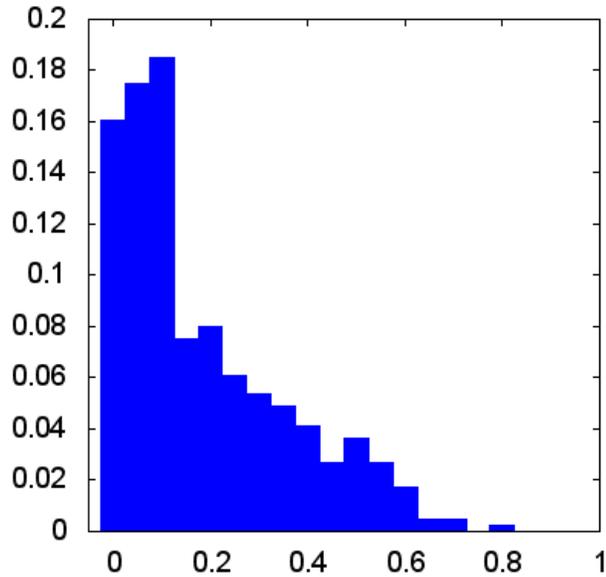


Figure 5.178: Relative frequency of fitness for all selected states of SDSS 588017702948962343

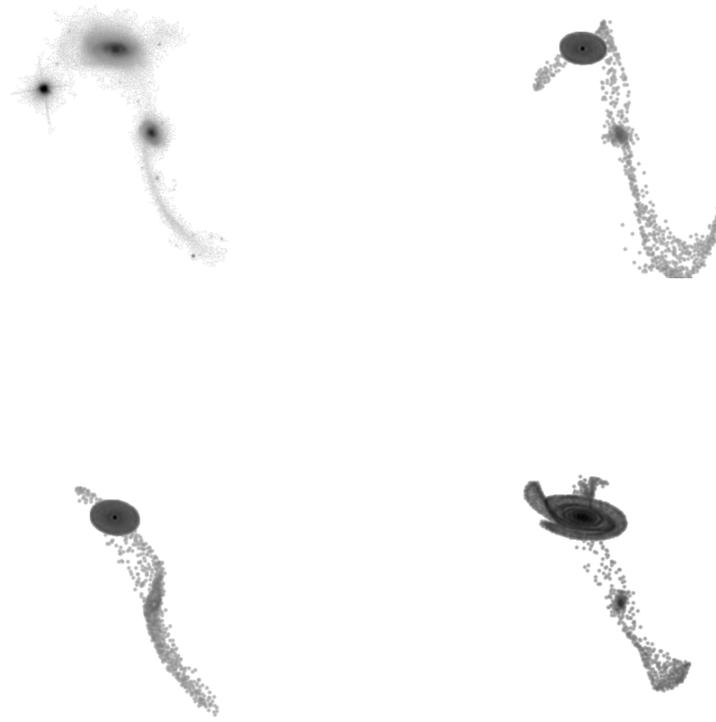


Figure 5.179: Target image and top 3 simulations for SDSS 588017702948962343

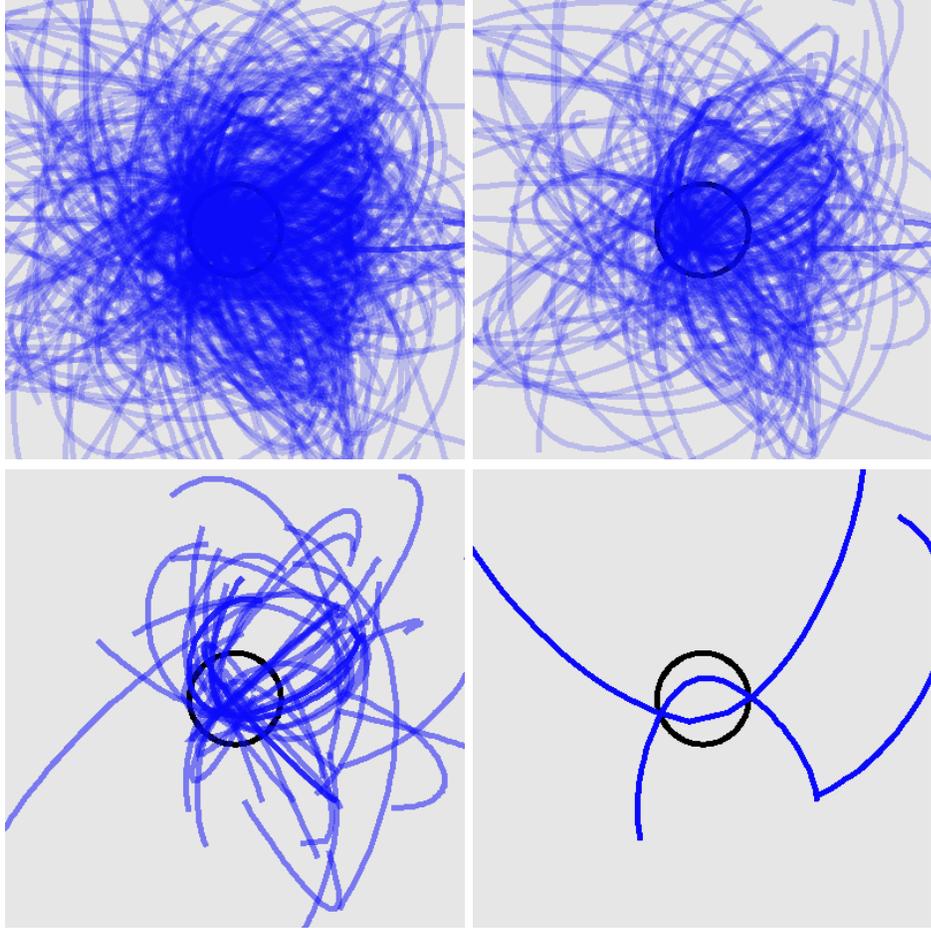


Figure 5.180: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 588017702948962343

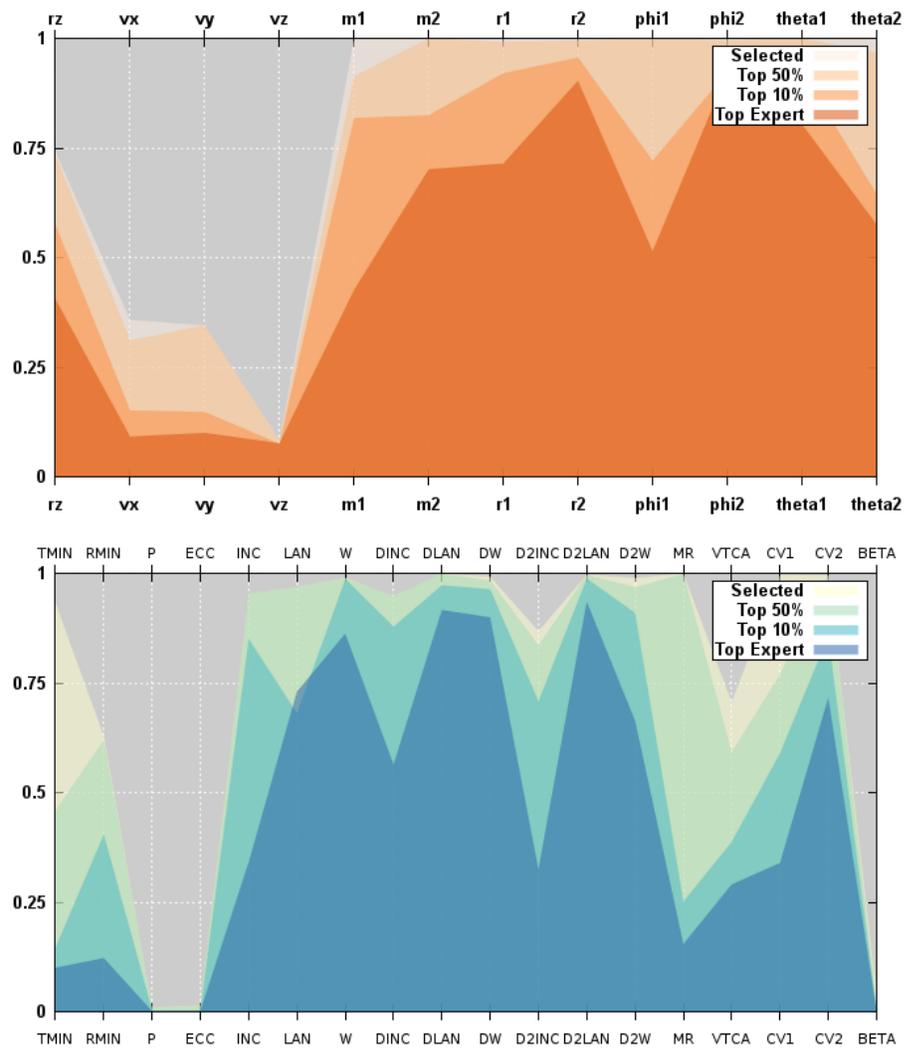


Figure 5.181: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 588017702948962343

### 5.1.46 SDSS 588017978901528612

For Arp 84, the tidal features of the secondary galaxy are more prominent. The fitness distribution is in between a smoothly declining one and an L-shaped one. The tails of the secondary galaxy are well matched in the simulation images. The volunteers have also produced a reasonable match to the disturbed disk of the primary galaxy. The trajectories are moderately similar. The simulation parameter convergence is good except for disk sizes and orientations.

Table 5.46: Identification Information and Merger Zoo summary for SDSS 588017978901528612.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 588017978901528612	Arp 84, NGC 5394/5395	13:58:37.9	+37:25:28.8	0.011711

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
18924	18511	413	119	7643	3596	4047

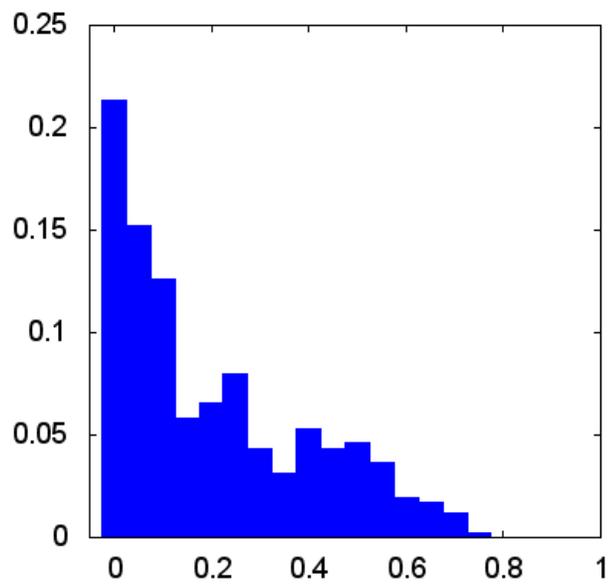


Figure 5.182: Relative frequency of fitness for all selected states of SDSS 588017978901528612

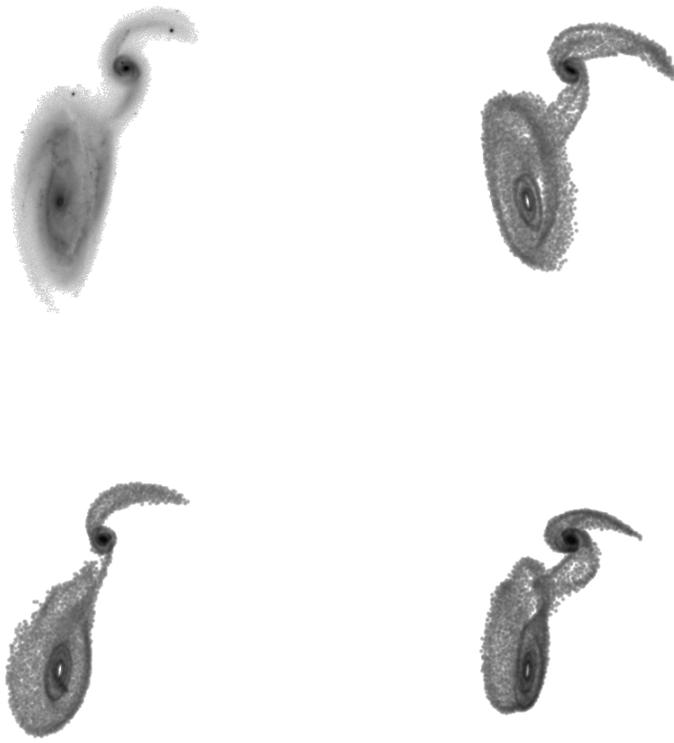


Figure 5.183: Target image and top 3 simulations for SDSS 588017978901528612

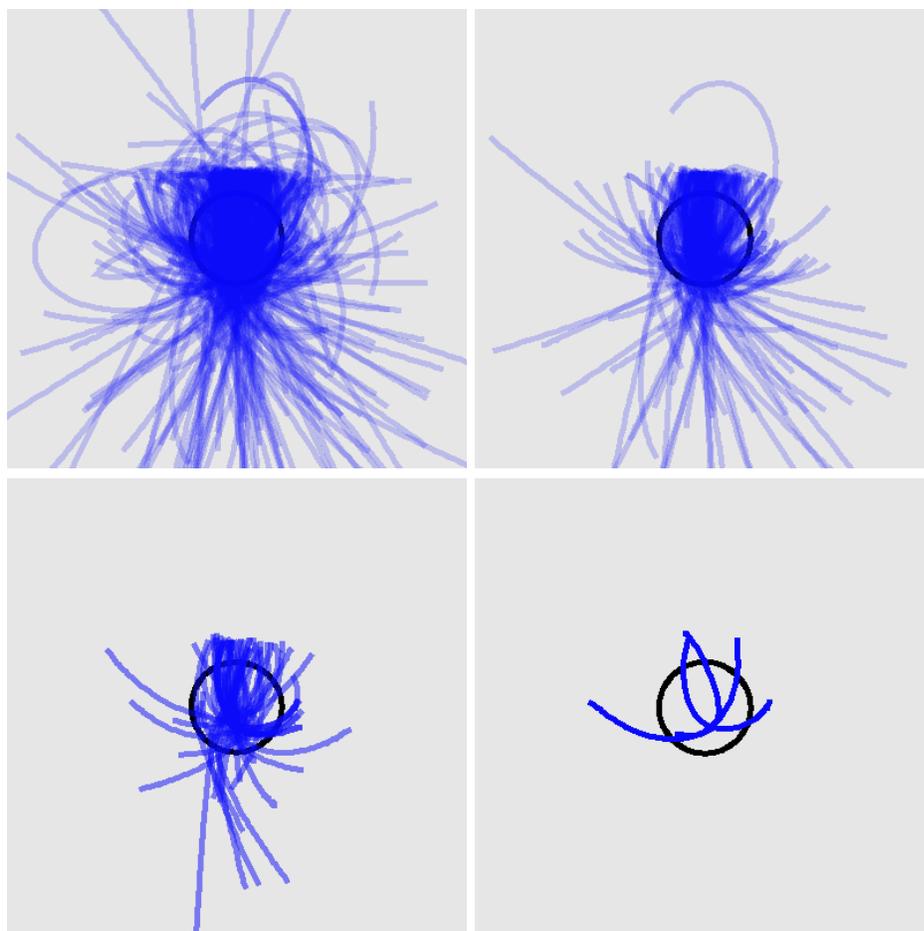


Figure 5.184: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 588017978901528612

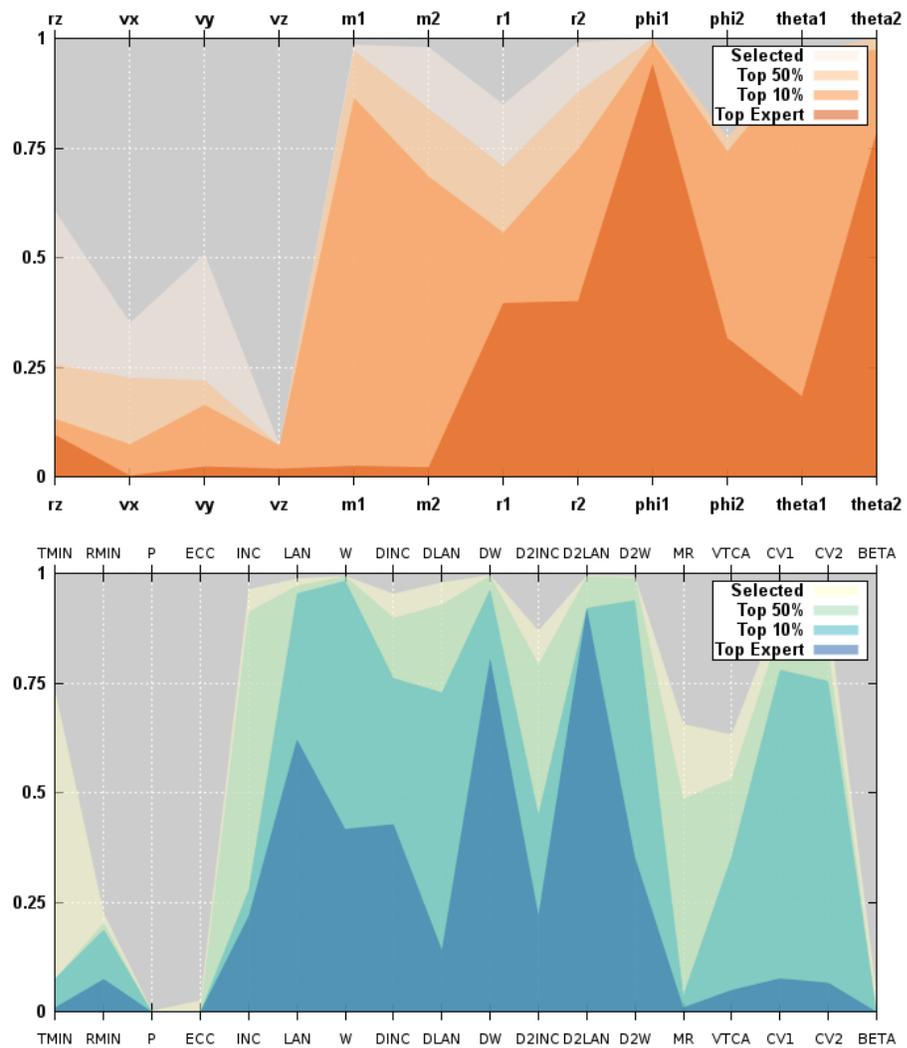


Figure 5.185: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 588017978901528612

### 5.1.47 SDSS 588018055130710322

UGC 10650 has a long northern tail. The top simulations provide three distinct versions of that tail, each of which is a moderate to good match. The trajectories are roughly similar. The convergence of simulation masses is moderate while disk sizes and orientations is closer to poor convergence.

Table 5.47: Identification Information and Merger Zoo summary for SDSS 588018055130710322.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 588018055130710322	UGC 10650	17:00:06.8	+23:07:53.5	0.056835

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
47369	46938	431	108	7917	3769	4148

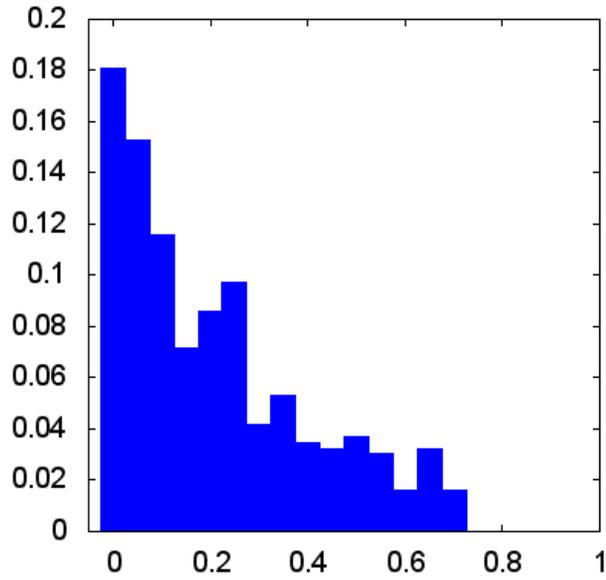


Figure 5.186: Relative frequency of fitness for all selected states of SDSS 588018055130710322



Figure 5.187: Target image and top 3 simulations for SDSS 588018055130710322

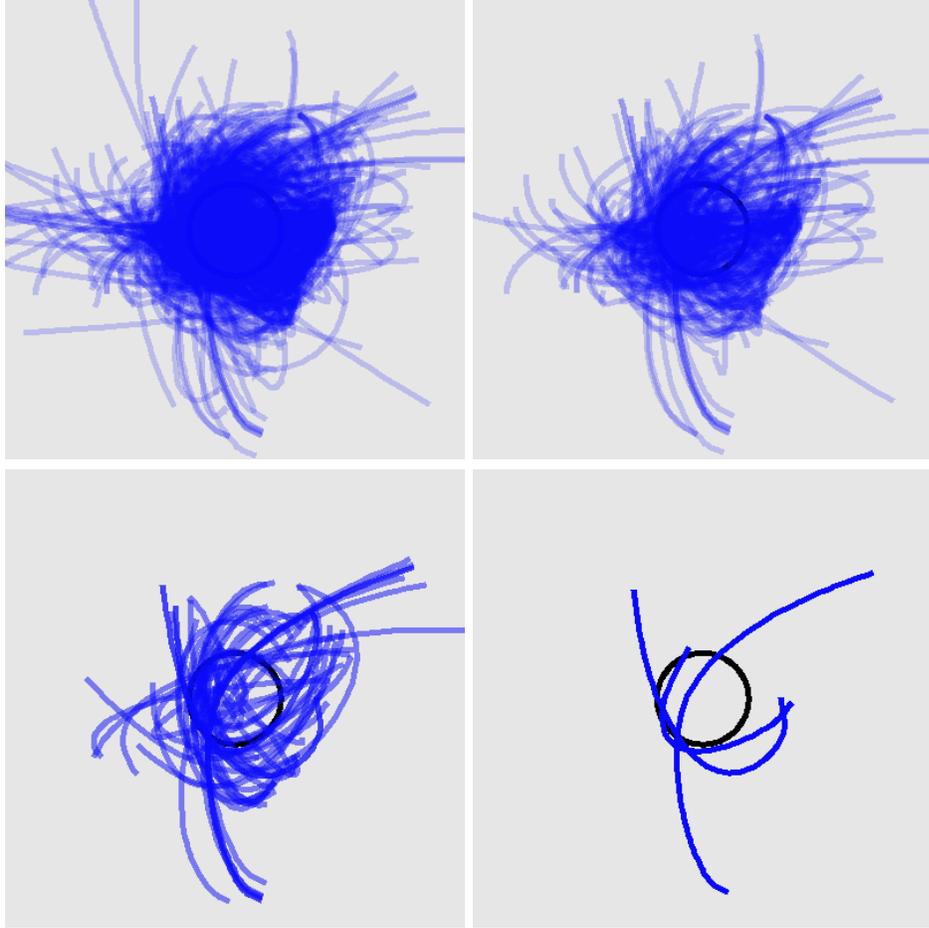


Figure 5.188: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 588018055130710322

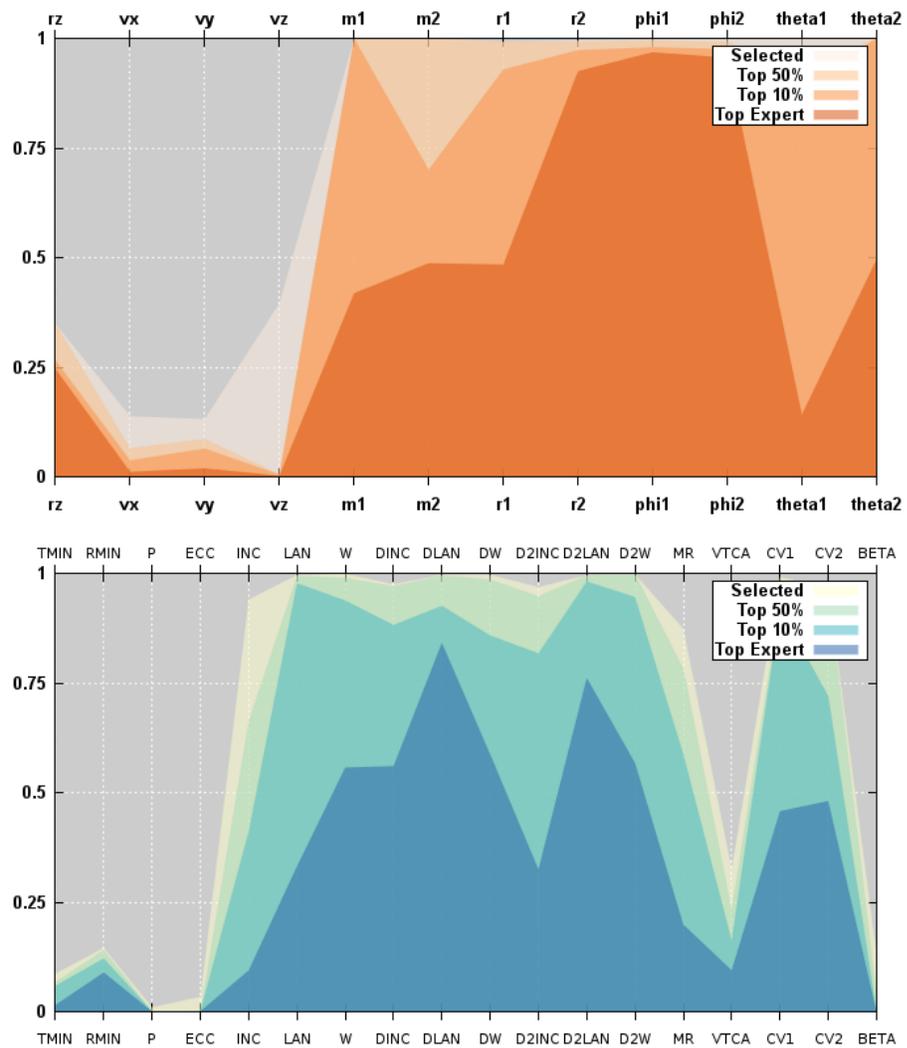


Figure 5.189: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 588018055130710322

### 5.1.48 SDSS 758874299603222717

The fitness distribution for Arp 112 hints at an L shape. The simulations present a moderately good match to the asymmetric norther tail. The enhanced density in the other galaxy is somewhat matched by the volunteers. The trajectories are similar to each other. The simulation parameters have moderate convergence except for the size of the primary disk.

Table 5.48: Identification Information and Merger Zoo summary for SDSS 758874299603222717.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 758874299603222717	Arp 112	0:01:26.7	+31:26:00.2	0.016048

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
113395	112327	1068	272	18994	7353	11641

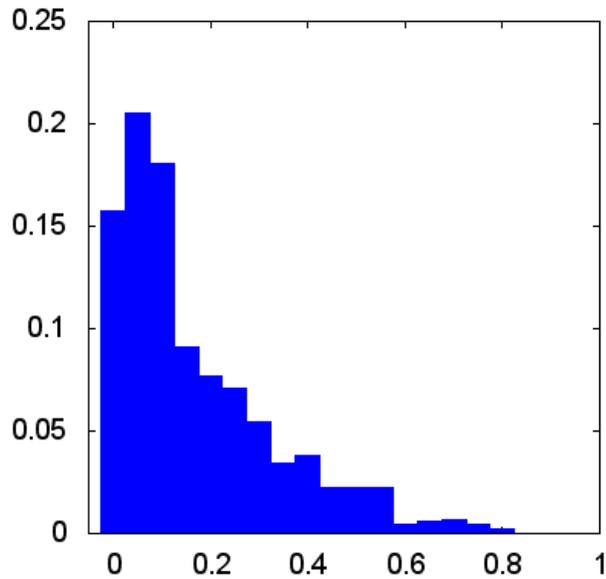


Figure 5.190: Relative frequency of fitness for all selected states of SDSS 758874299603222717

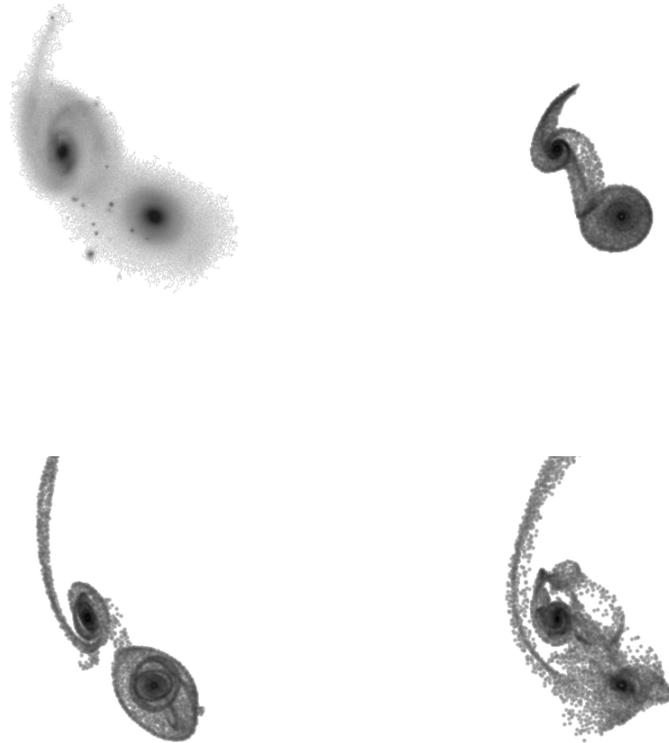


Figure 5.191: Target image and top 3 simulations for SDSS 758874299603222717

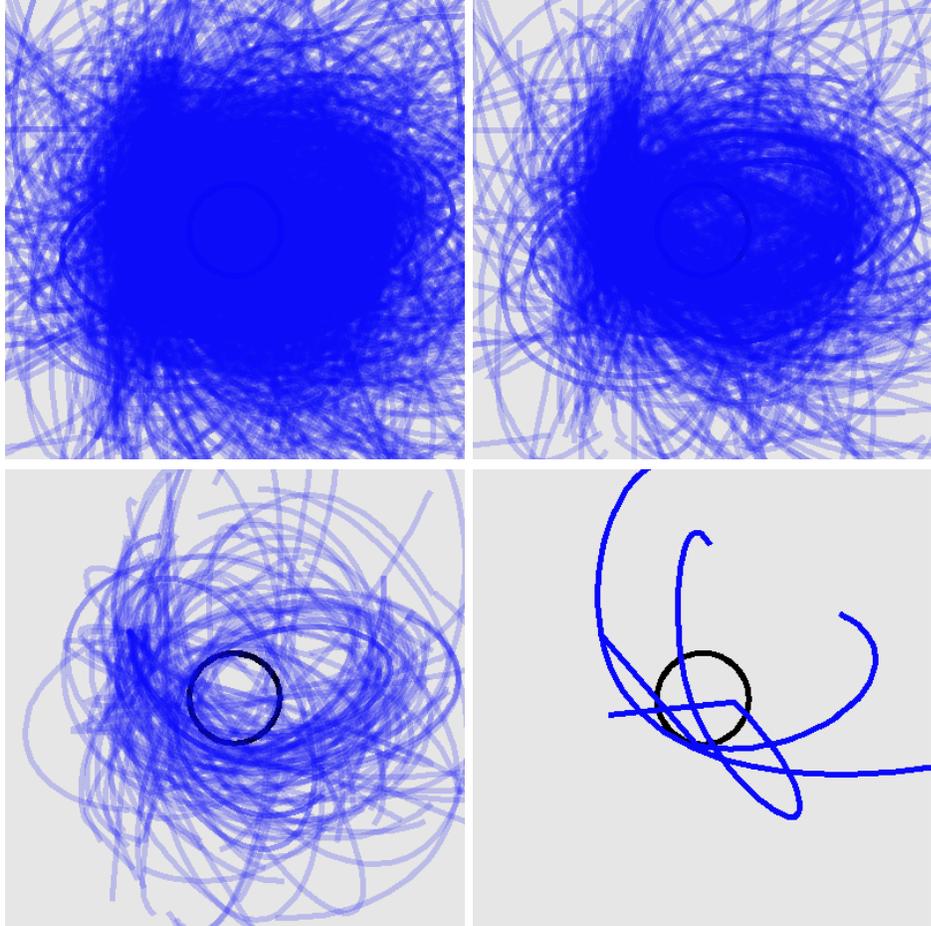


Figure 5.192: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 758874299603222717

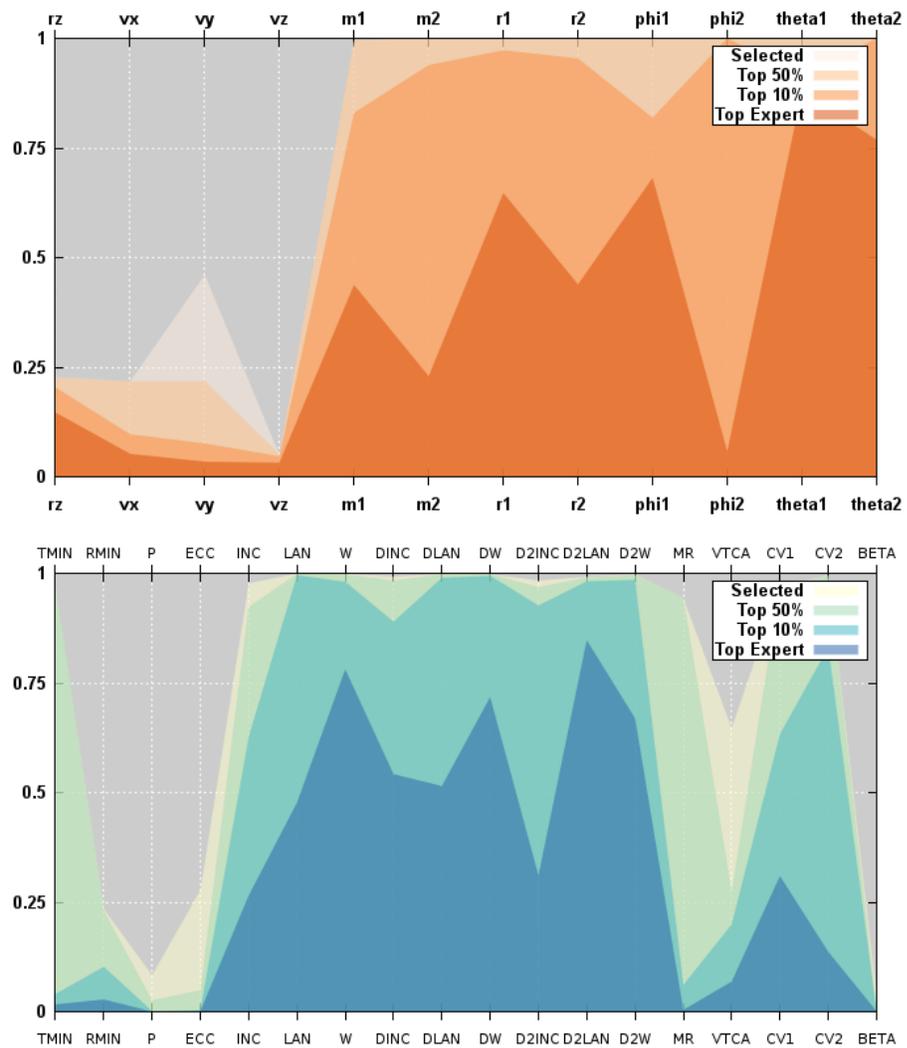


Figure 5.193: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 758874299603222717

### 5.1.49 Arp 148

Arp 148 has an unexcited fitness distribution. This is matched by the unexciting morphology with no easy tidal features for volunteers to match. The top three simulations all look like unperturbed disks. Two of the top three trajectories are similar. The simulation and orbit parameters have poor convergence.

Table 5.49: Identification Information and Merger Zoo summary for Arp 148.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
Arp 148	Arp 148	11:03:54.1	+40:50:57.7	0.034524

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
33460	31676	1784	182	24073	15206	8867

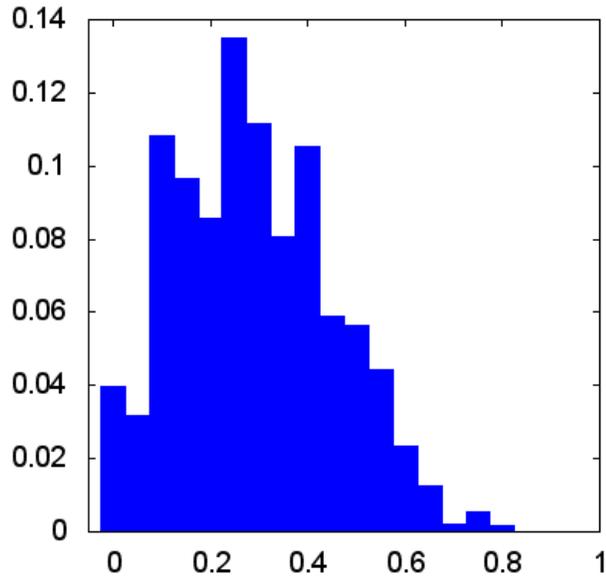


Figure 5.194: Relative frequency of fitness for all selected states of Arp 148

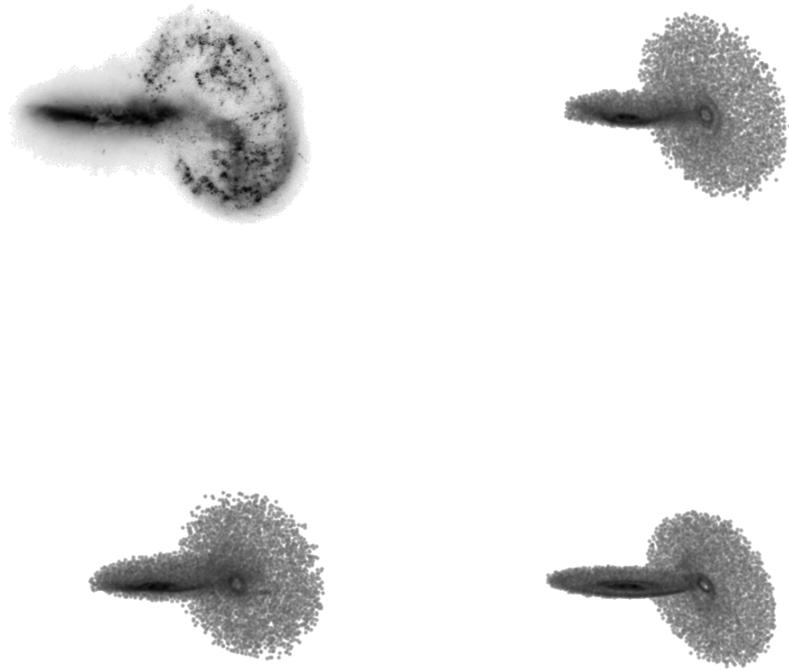


Figure 5.195: Target image and top 3 simulations for Arp 148

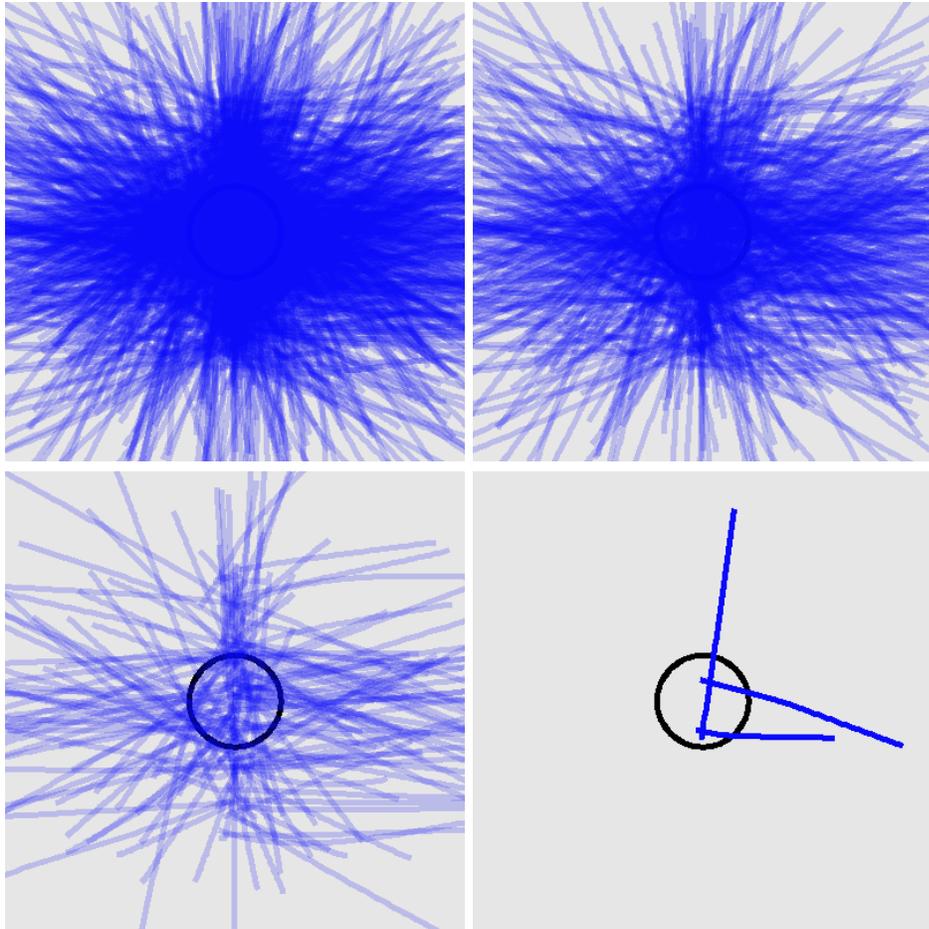


Figure 5.196: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for Arp 148

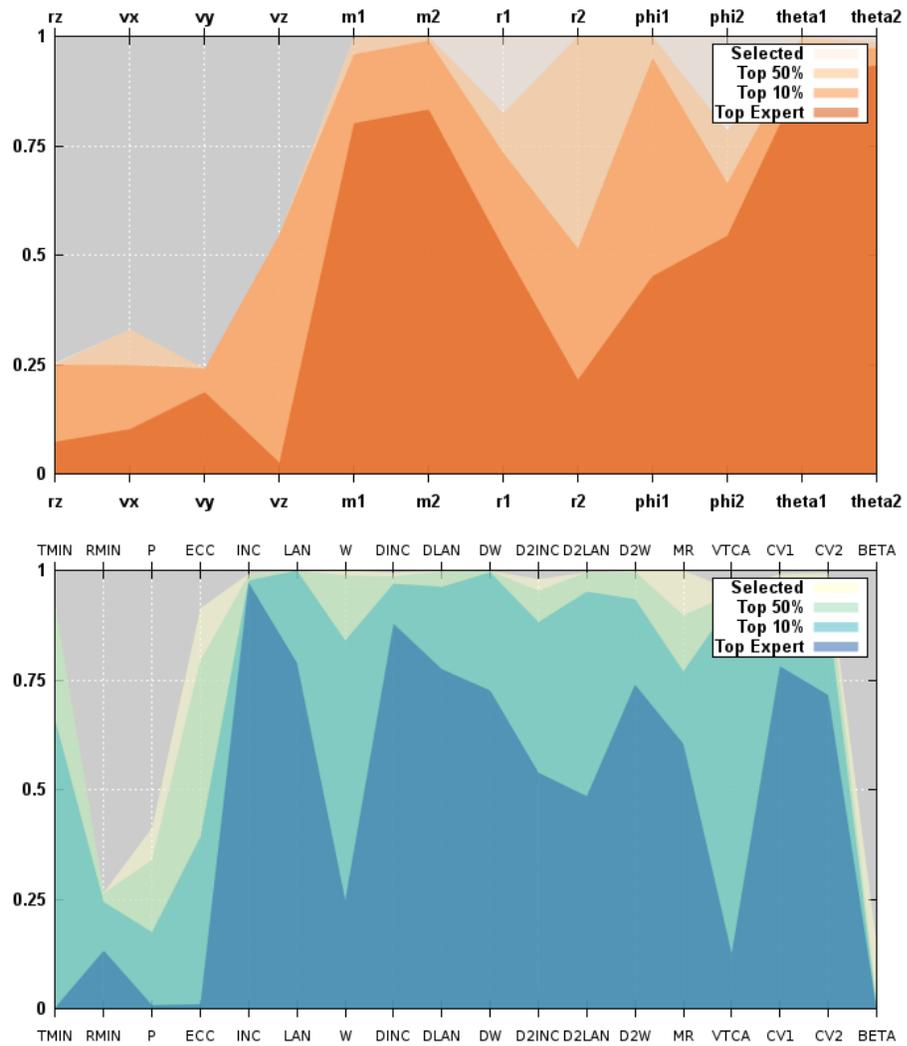


Figure 5.197: Parallel coordinates for convergence of simulation and orbit parameters for Arp 148

### 5.1.50 CGCG 436-030

This HST GGW target has asymmetric features on the primary and the secondary. The volunteers were able to match the tidal features. However, the top simulations usually have corresponding symmetric tails that are not present in the image. The trajectories are very similar. The simulation and orbit parameter convergence is poor.

Table 5.50: Identification Information and Merger Zoo summary for CGCG 436-030.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
CGCG 436-030	2MASS 01200265+1421417	1:20:02.7	+14:21:43.4	0.031569

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
27351	26885	466	84	6384	3201	3183

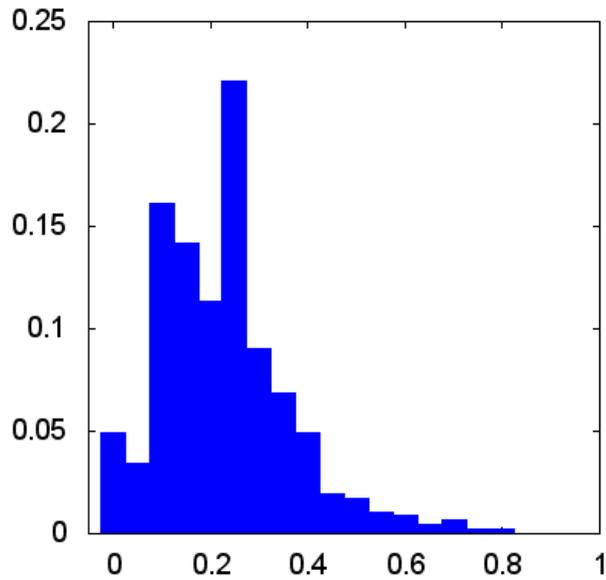


Figure 5.198: Relative frequency of fitness for all selected states of CGCG 436-030



Figure 5.199: Target image and top 3 simulations for CGCG 436-030

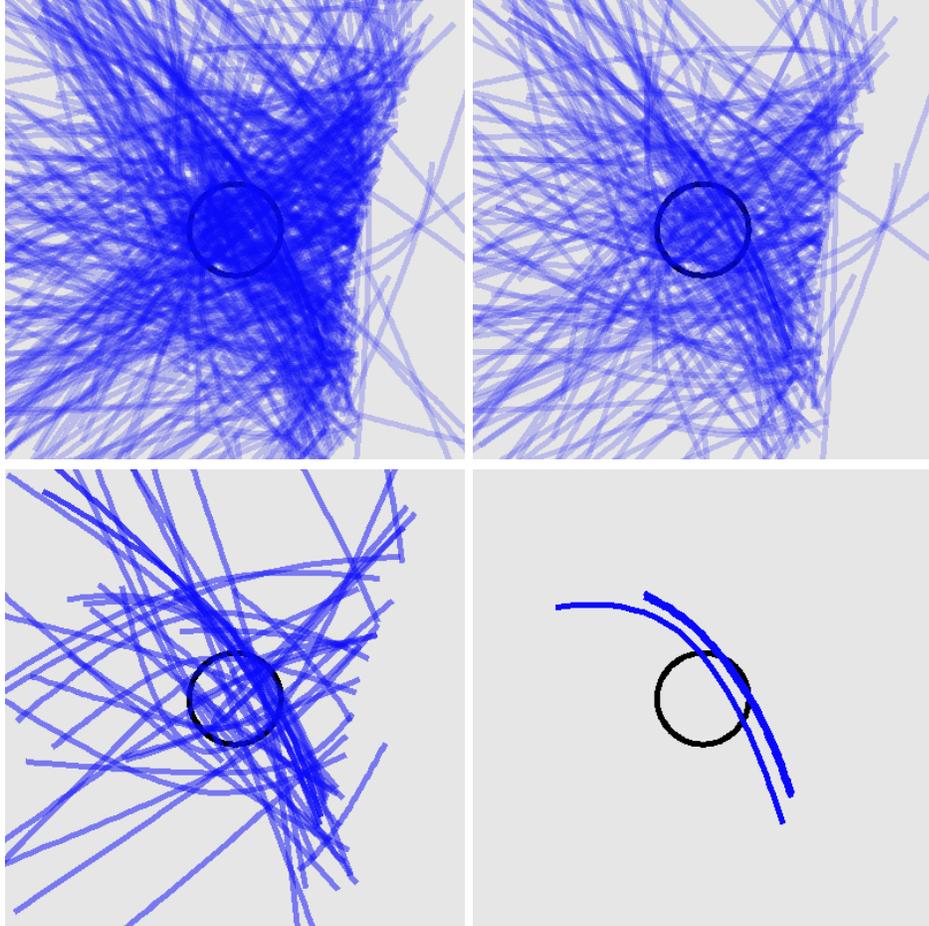


Figure 5.200: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for CGCG 436-030

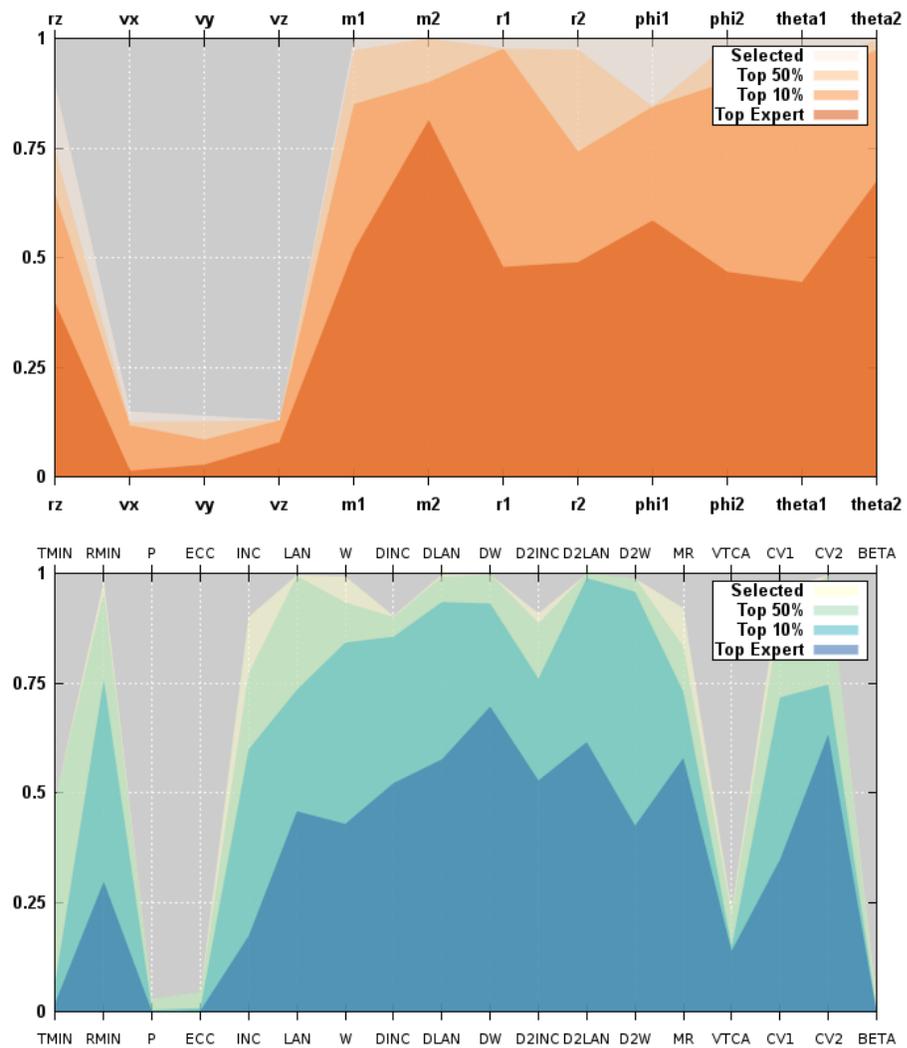


Figure 5.201: Parallel coordinates for convergence of simulation and orbit parameters for CGCG 436-030

### 5.1.51 Arp 272

Arp 272 has distinct tidal features. However, they are within the plane of the disk and are difficult to model with our restricted three-body code. Additionally, there is what appears to be a strong collection of material in a knot between the two galaxies. This may be a dwarf galaxy, the creation of which cannot be modeled without self-gravity. The top simulations each produced density enhancements in the two disks. The match is very rough. The top trajectories show that the two galaxies remain close to each other during the simulation period. The simulation parameter convergence is poor, particularly for the galaxy masses.

Table 5.51: Identification Information and Merger Zoo summary for Arp 272.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
Arp 272	Arp 272, NGC 6050	16:05:23.4	+17:45:25.8	0.031928

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
11901	11031	870	130	11697	5770	5927

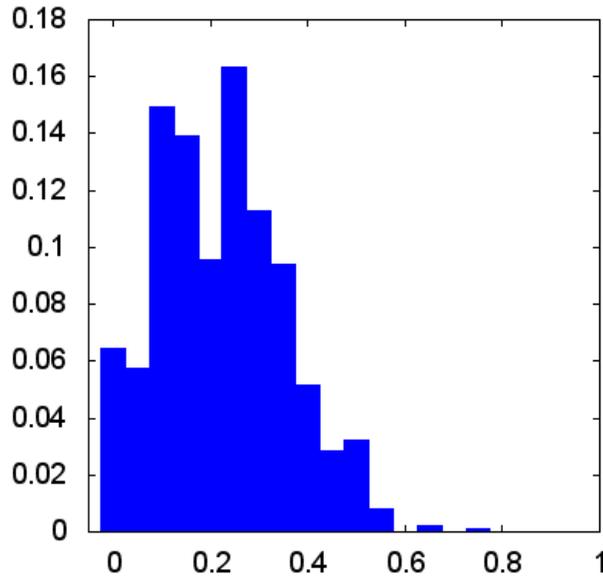


Figure 5.202: Relative frequency of fitness for all selected states of Arp 272

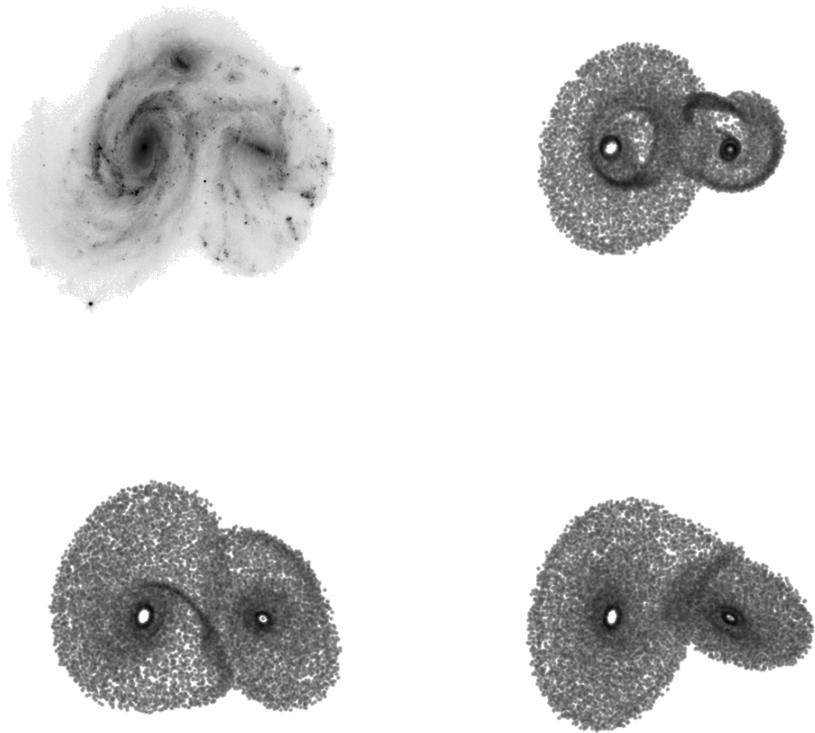


Figure 5.203: Target image and top 3 simulations for Arp 272

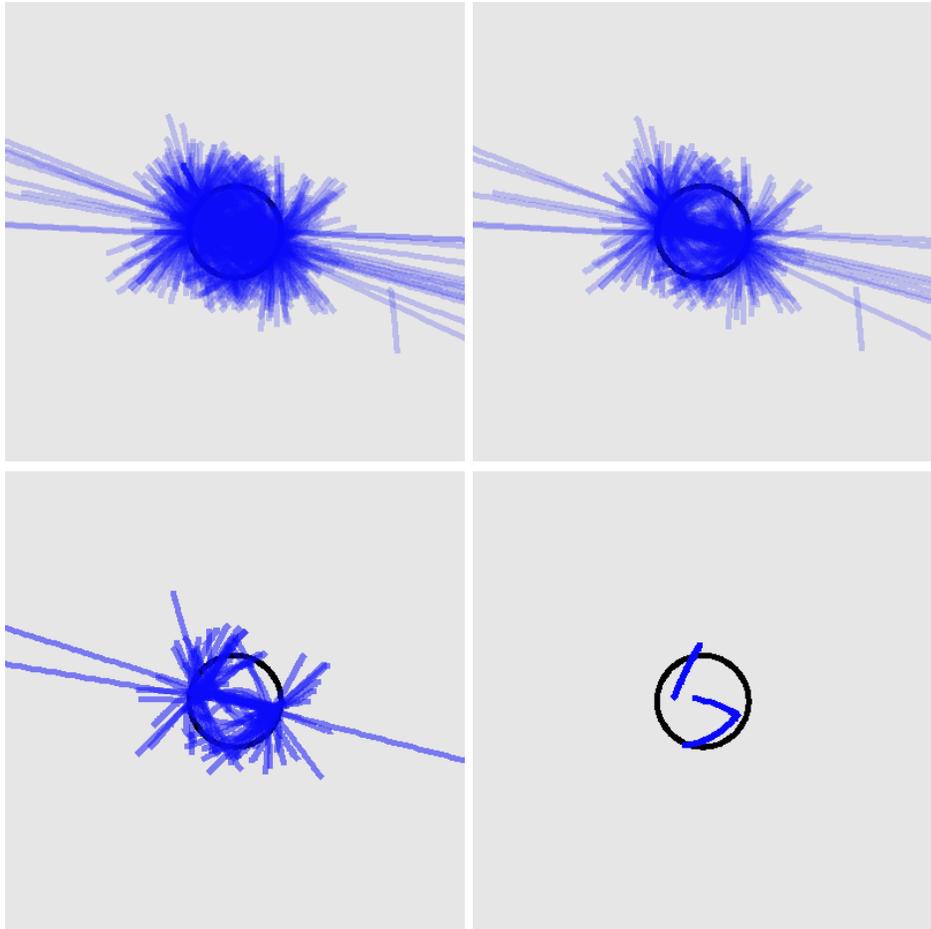


Figure 5.204: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for Arp 272

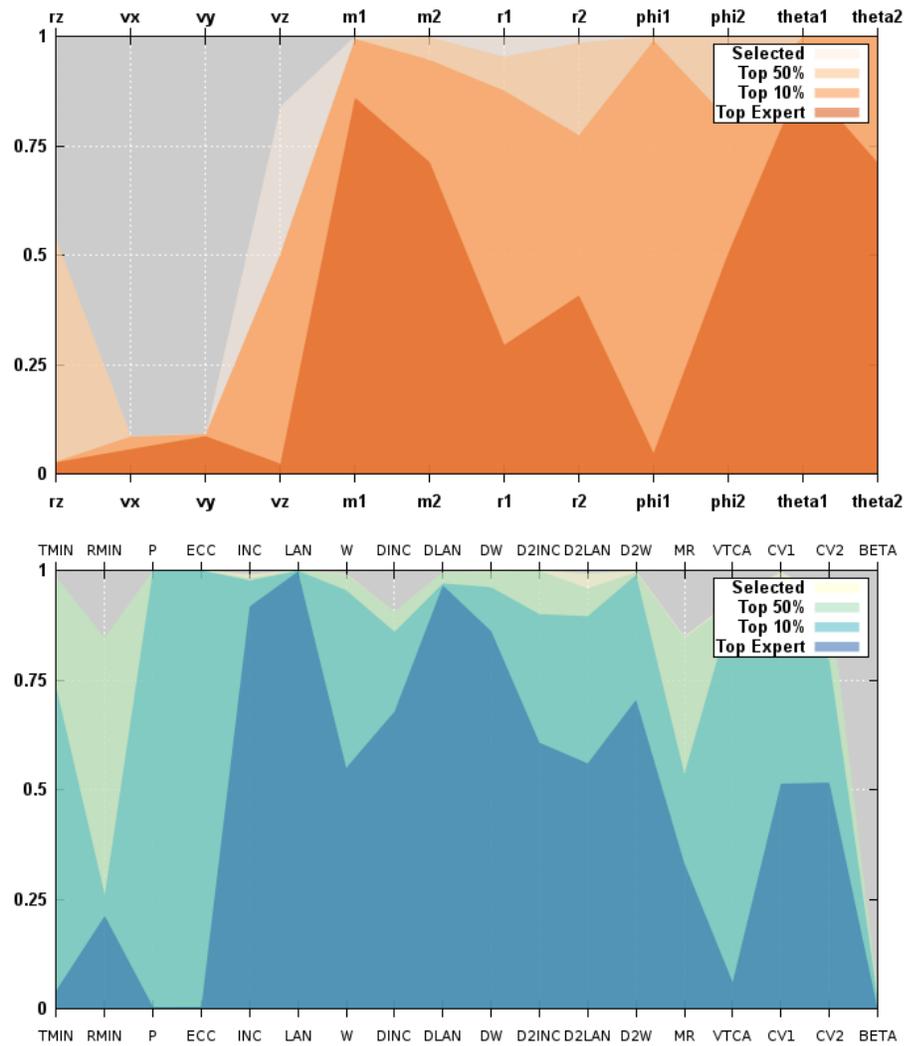


Figure 5.205: Parallel coordinates for convergence of simulation and orbit parameters for Arp 272

### 5.1.52 ESO 77-14

This HST GGW target was more popular than other HST targets with volunteers viewing over 51000 simulations. Each galaxy has a tidal tail. Those tails are both present in each of the top simulations. The best-fit simulation seems to have an additional ring of debris not seen in the target image. The trajectories are all closed, multi-passage orbits. The simulation and orbit parameters have moderate to poor convergence.

Table 5.52: Identification Information and Merger Zoo summary for ESO 77-14.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
ESO 77-14	2MASS 23210539-6912472	23:21:04.6	-69:12:47.3	0.042326

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
51759	50977	782	128	10759	4335	6424

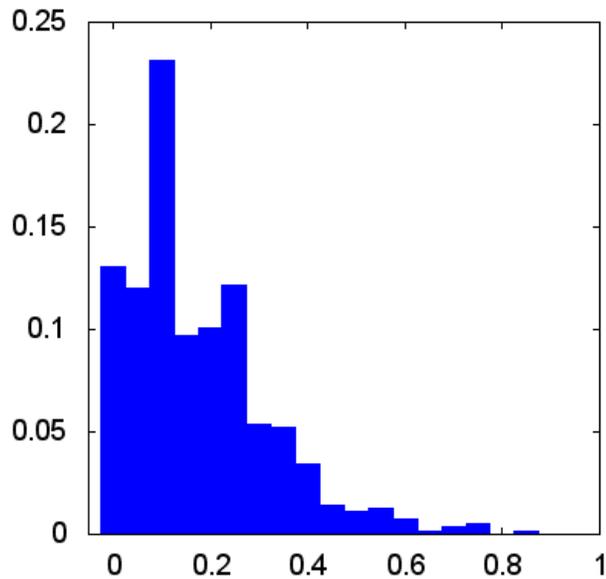


Figure 5.206: Relative frequency of fitness for all selected states of ESO 77-14

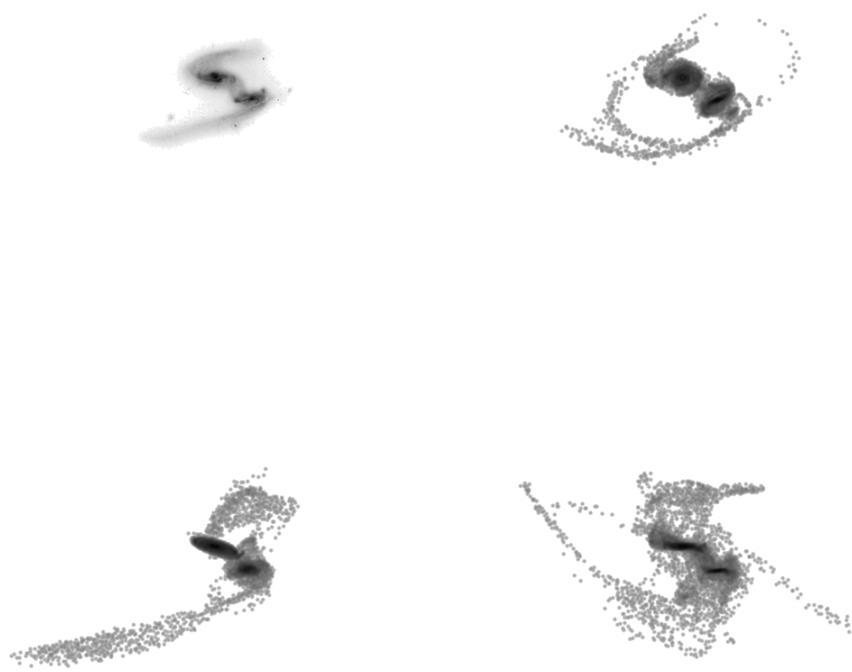


Figure 5.207: Target image and top 3 simulations for ESO 77-14

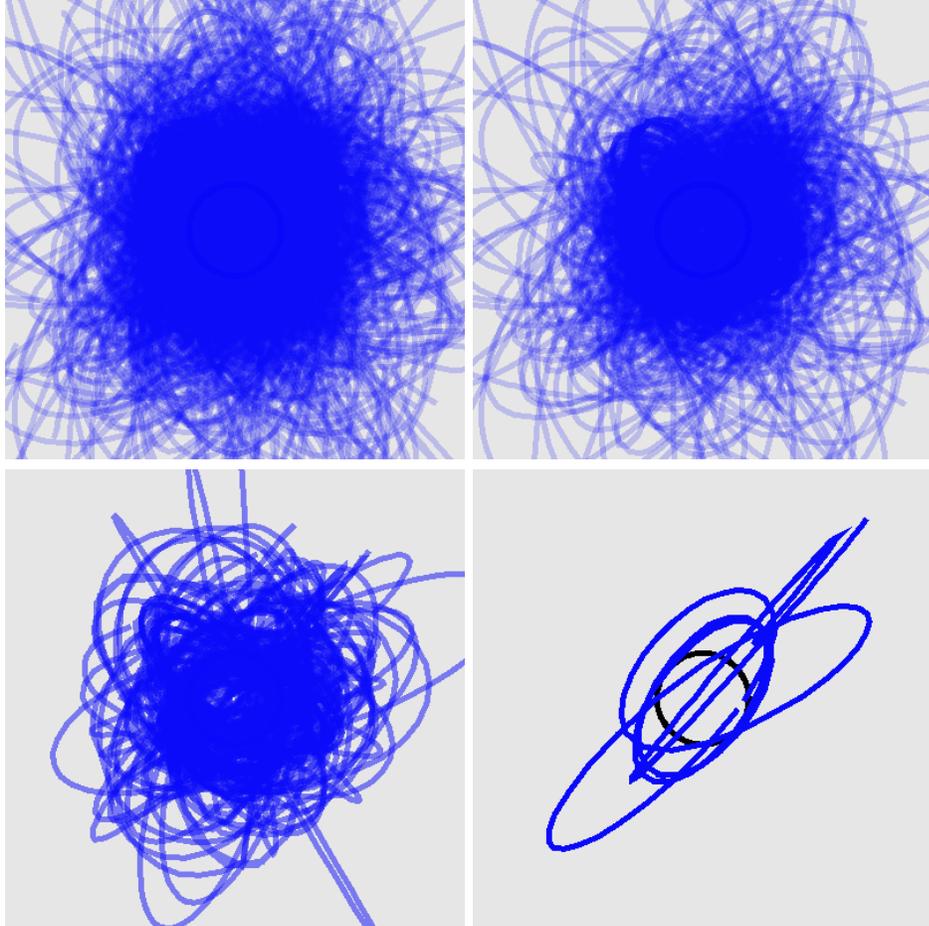


Figure 5.208: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for ESO 77-14

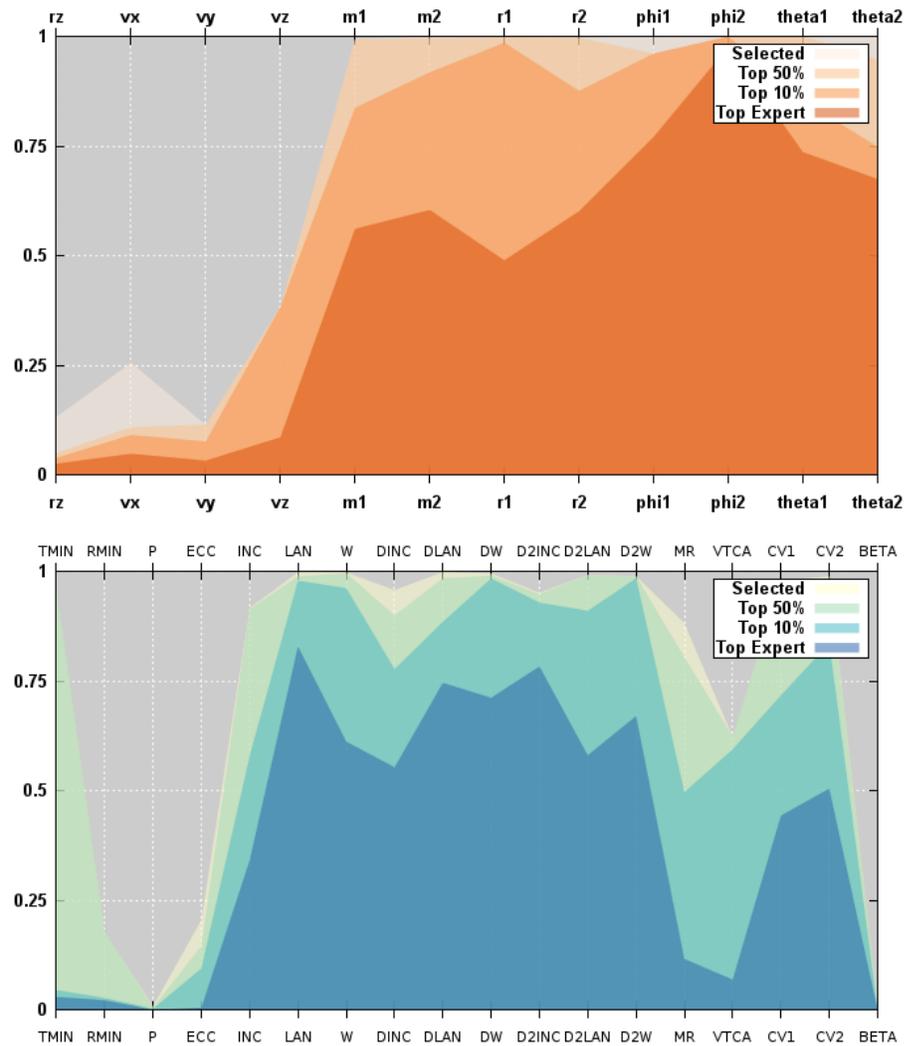


Figure 5.209: Parallel coordinates for convergence of simulation and orbit parameters for ESO 77-14

### 5.1.53 NGC 5331

NGC 5331 does have obvious features, but not the more popular long tails. The small northern tail in the image is moderately matched by the best simulation. The trajectories are not very similar. The convergence of orbit and simulation parameters is poor.

Table 5.53: Identification Information and Merger Zoo summary for NGC 5331.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
NGC 5331	NGC 5331, UGC 8774	13:52:16.2	+2:06:01.1	0.032799

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
6595	6302	293	68	4137	2479	1658

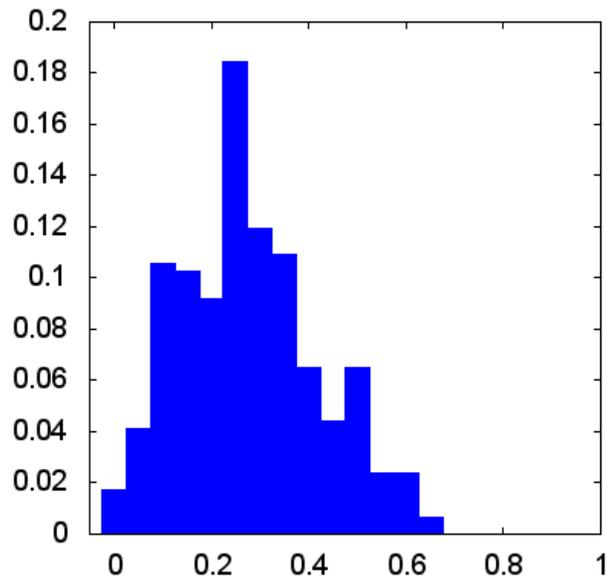


Figure 5.210: Relative frequency of fitness for all selected states of NGC 5331

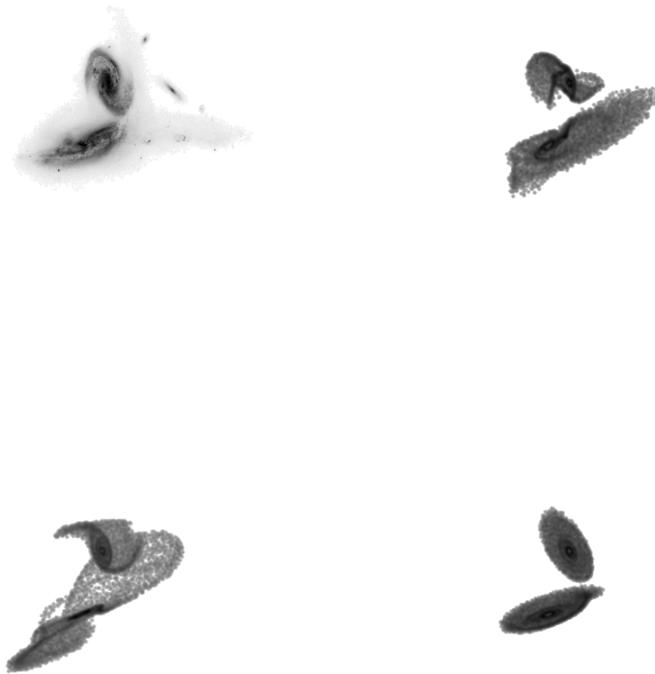


Figure 5.211: Target image and top 3 simulations for NGC 5331

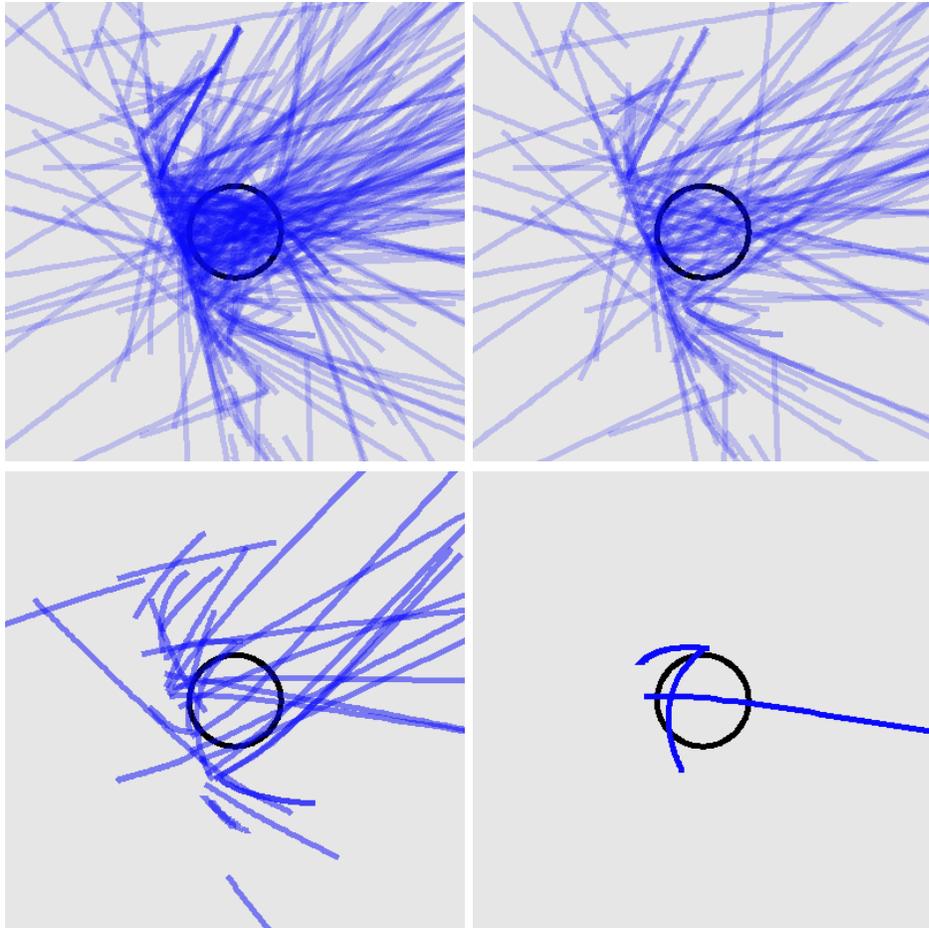


Figure 5.212: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for NGC 5331

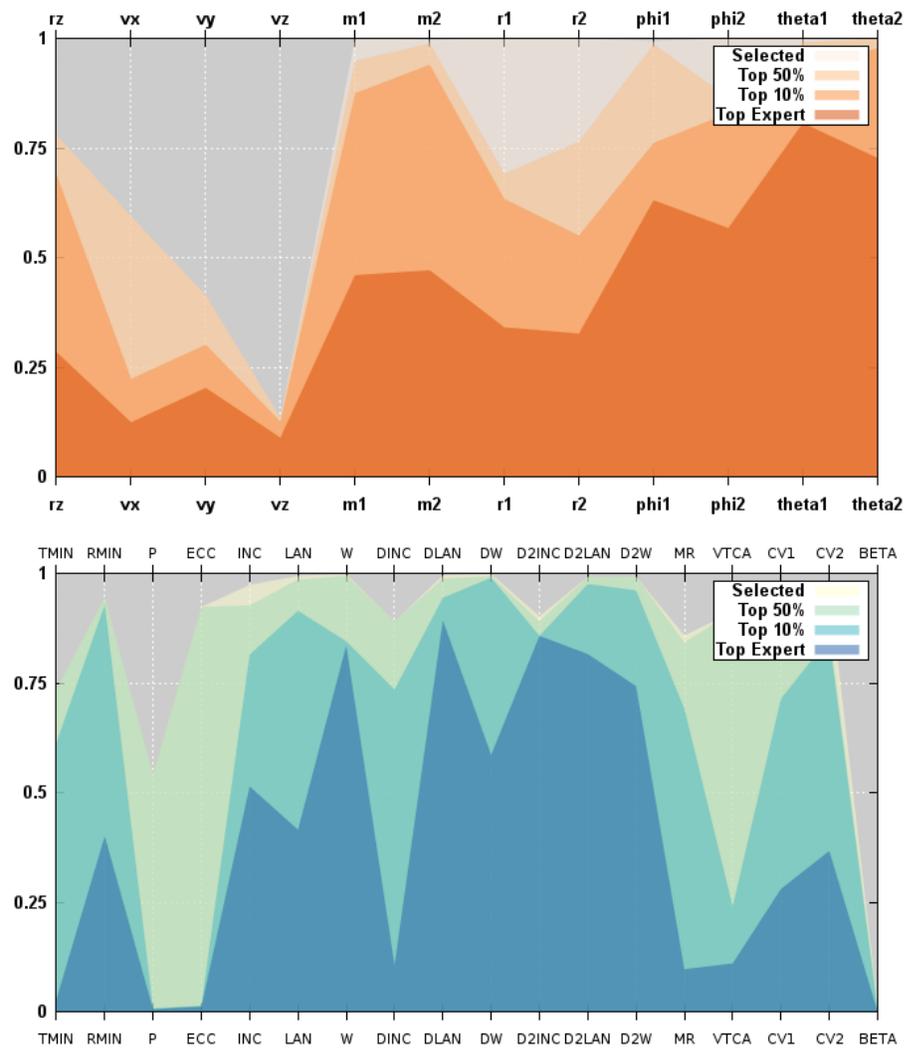


Figure 5.213: Parallel coordinates for convergence of simulation and orbit parameters for NGC 5331

### 5.1.54 NGC 6786

NGC 6786 definitely has an L-shaped fitness distribution. It was one of the most popular targets with over 250000 simulations viewed. The primary and secondary each have two tidal tails. The top simulations are all moderate to good matches for all four key tidal features. The convergence of the trajectories is impressive. Even though the top paths are not identical, they represent a substantially reduced set of possibilities from the largely saturated image of trajectories for even the top 50% of simulations. The simulation parameters are well converged with the exception of the disk inclination angle.

Table 5.54: Identification Information and Merger Zoo summary for NGC 6786.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
NGC 6786	NGC 6786, UGC 11415	19:10:53.8	+73:24:37.0	0.025017

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
253392	246370	7022	1389	40144	17348	22796

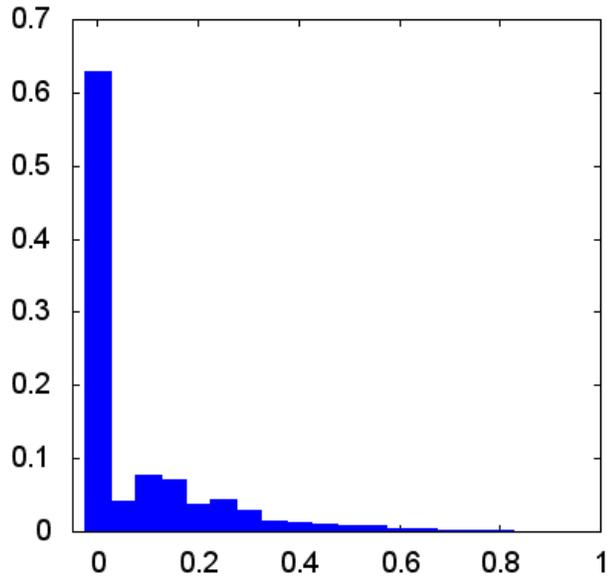


Figure 5.214: Relative frequency of fitness for all selected states of NGC 6786

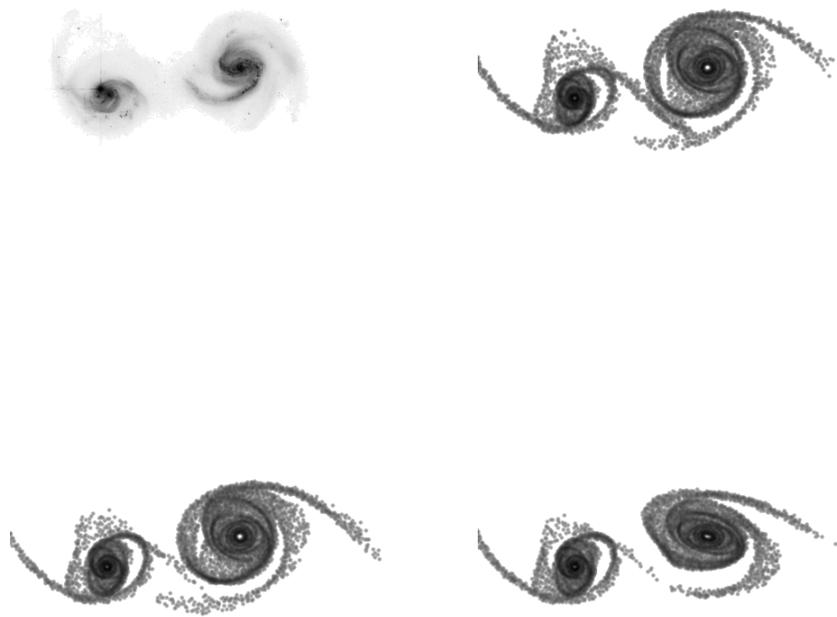


Figure 5.215: Target image and top 3 simulations for NGC 6786

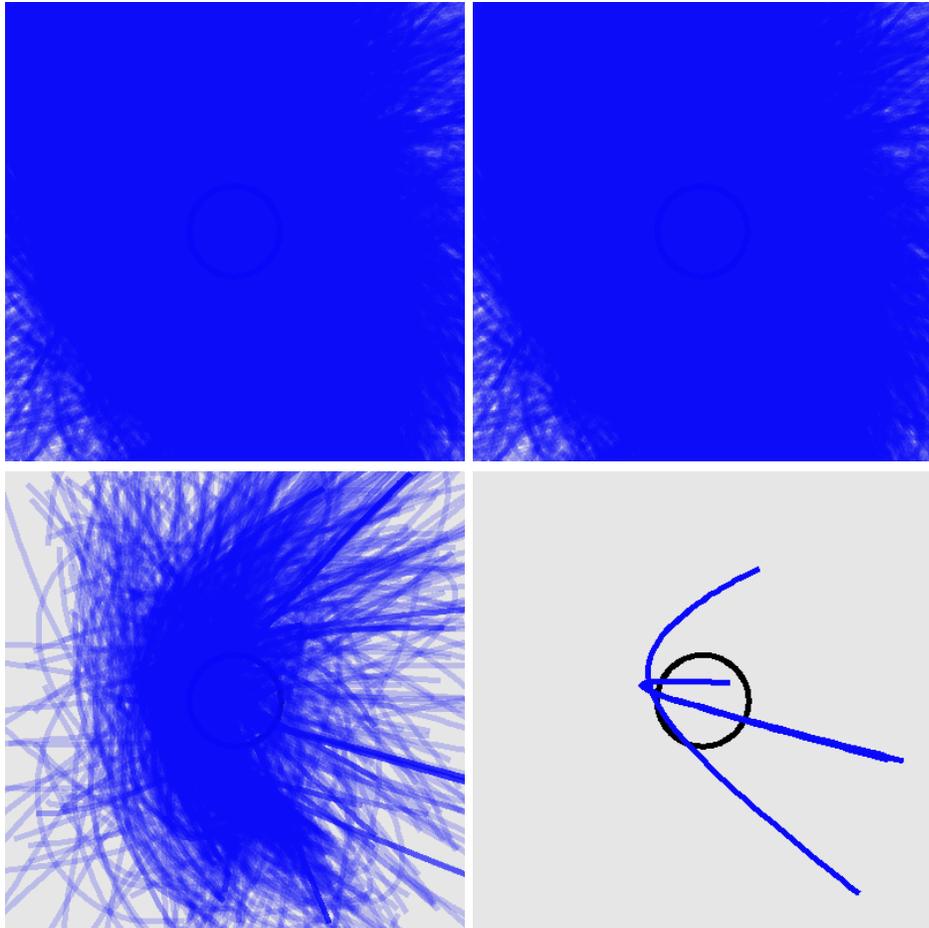


Figure 5.216: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for NGC 6786

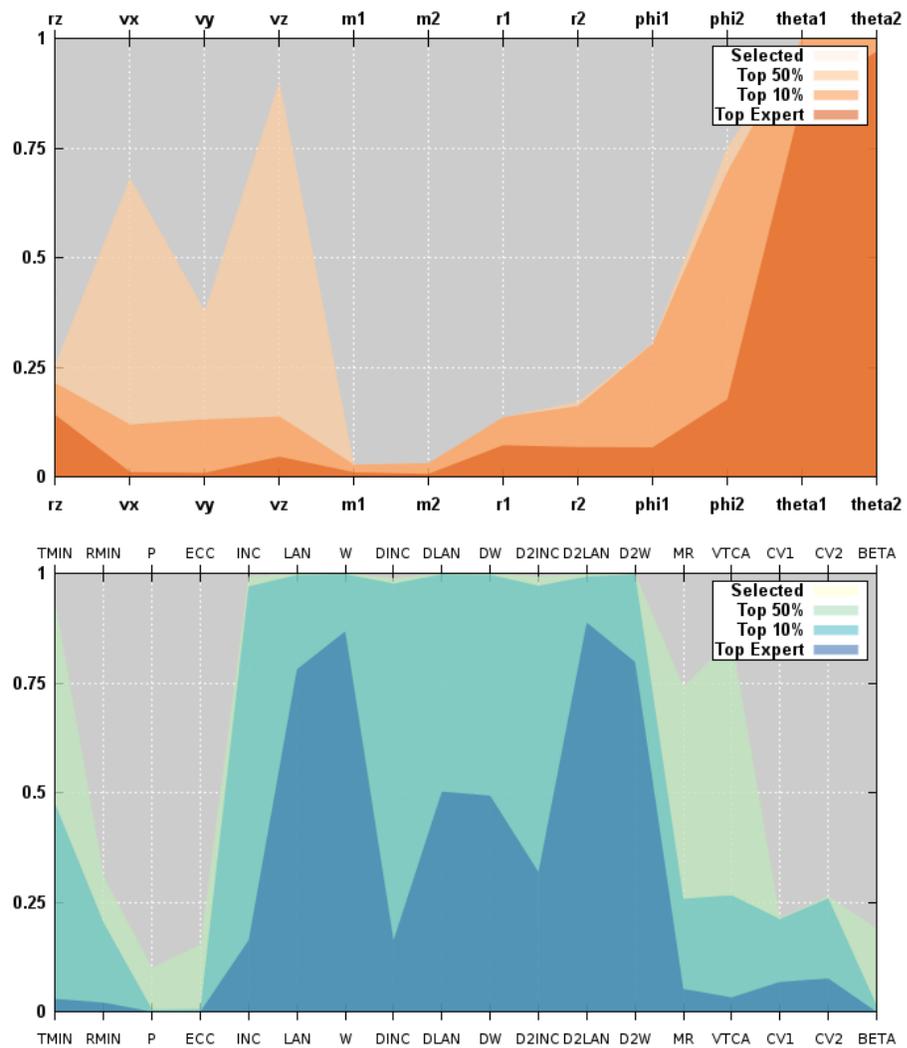


Figure 5.217: Parallel coordinates for convergence of simulation and orbit parameters for NGC 6786

### 5.1.55 SDSS 587736523764334706

Arp 274 is another example of a less than exciting pair of largely unperturbed disks. The internal spiral structure was not well matched. The top simulations simply produce density enhancements. The trajectories are not very similar to each other. The simulation parameter convergence is moderate to poor.

Table 5.55: Identification Information and Merger Zoo summary for SDSS 587736523764334706.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587736523764334706	Arp 274, NGC 5679	14:35:08.7	+5:21:31.7	0.028867

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
30401	28236	2165	206	37950	24428	13522

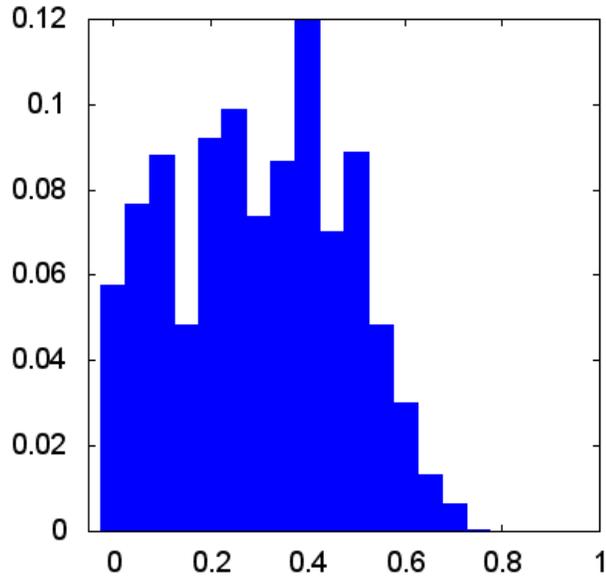


Figure 5.218: Relative frequency of fitness for all selected states of SDSS 587736523764334706

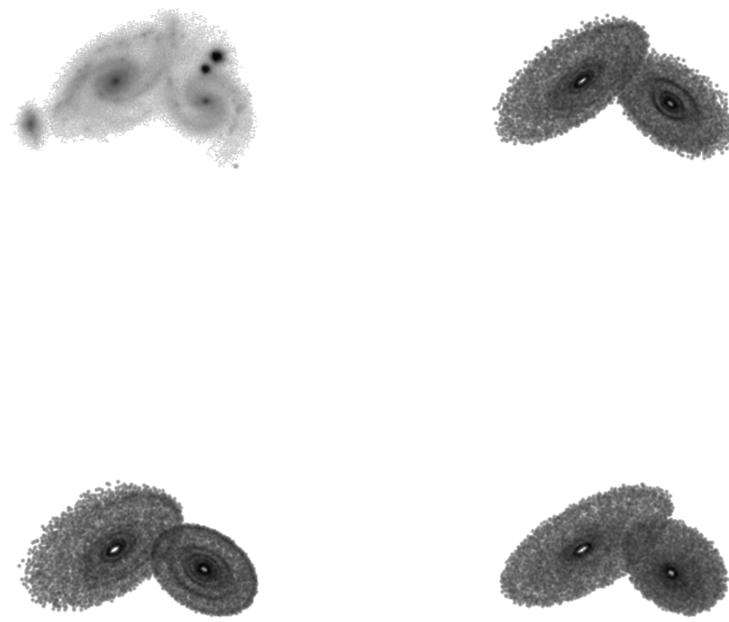


Figure 5.219: Target image and top 3 simulations for SDSS 587736523764334706

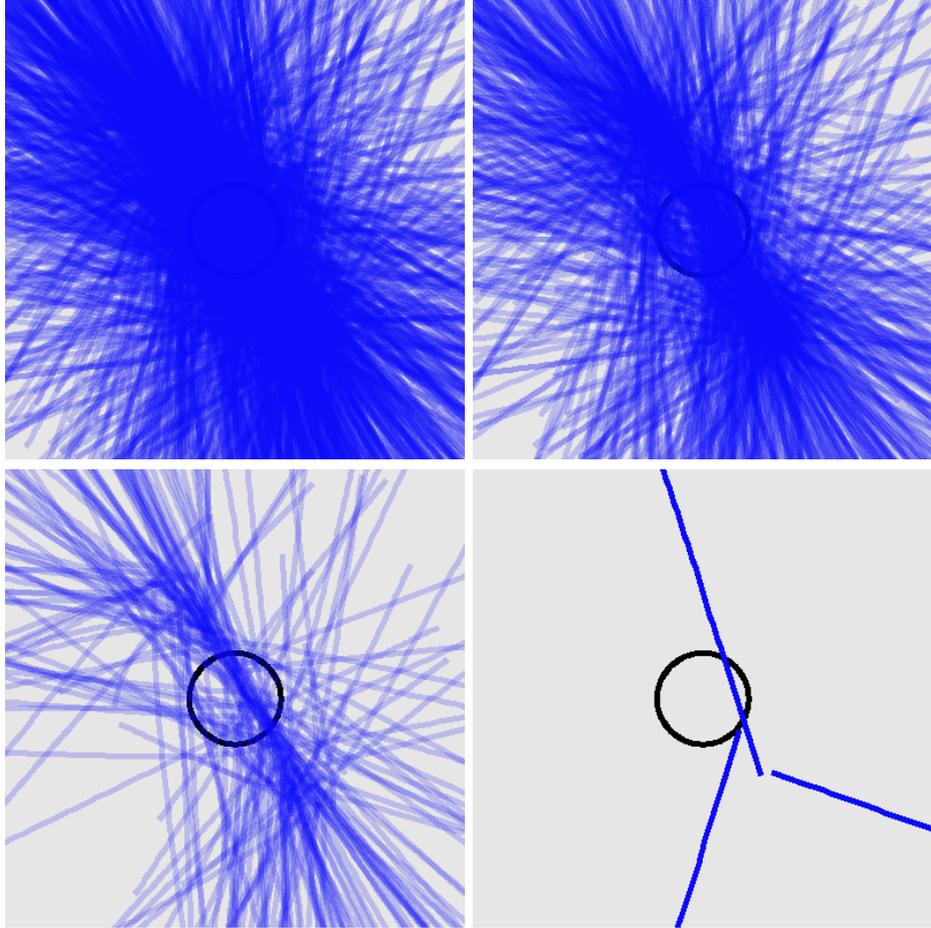


Figure 5.220: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587736523764334706

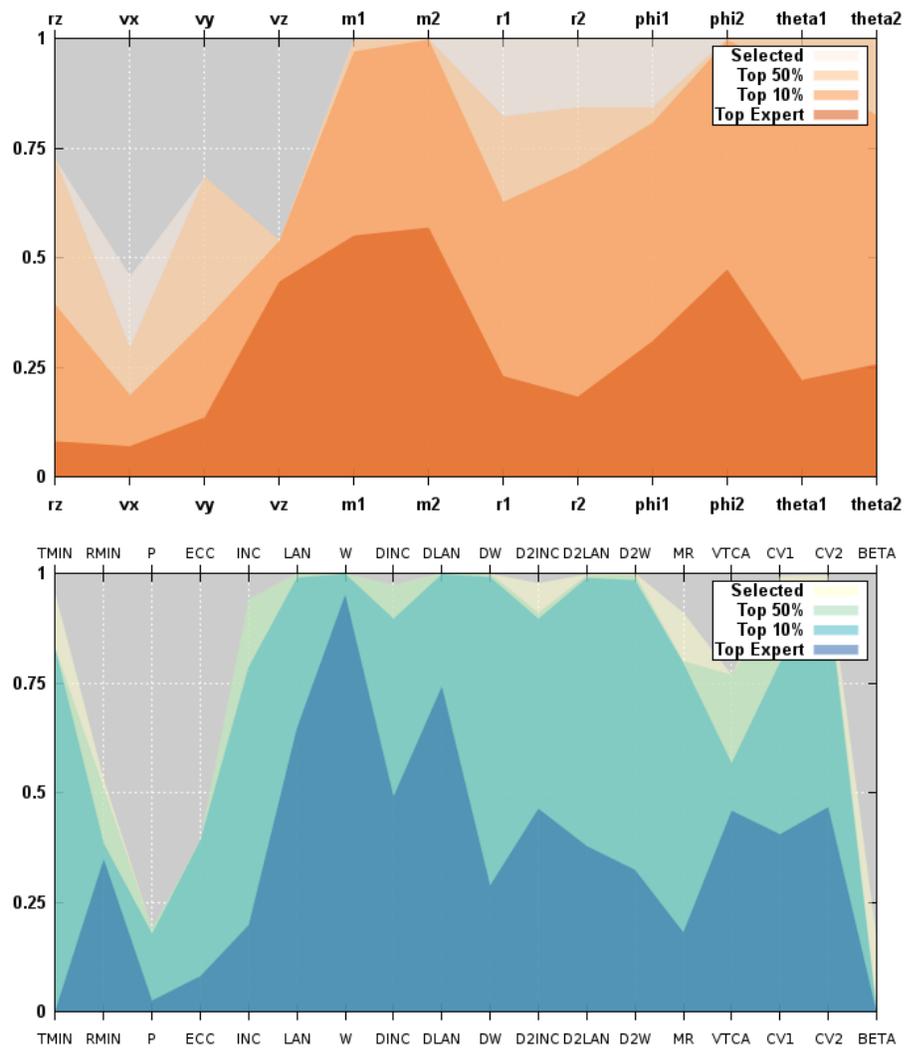


Figure 5.221: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587736523764334706

### 5.1.56 SDSS 587747120521216156

The fitness distribution for Arp 146 is between L-shaped and smoothly declining. The creation of the ring galaxy is matched in all simulations. The shape is not a good match in all simulations. The features of the non-ring galaxy are not well matched. One of the top simulations even has the rotation going in the opposite direction. The trajectories are not very similar to each other. The simulation parameter convergence is moderate.

Table 5.56: Identification Information and Merger Zoo summary for SDSS 587747120521216156.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587747120521216156	Arp 146	0:06:44.7	-6:38:12.9	0.07544

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
51786	50560	1226	191	21576	7028	14548

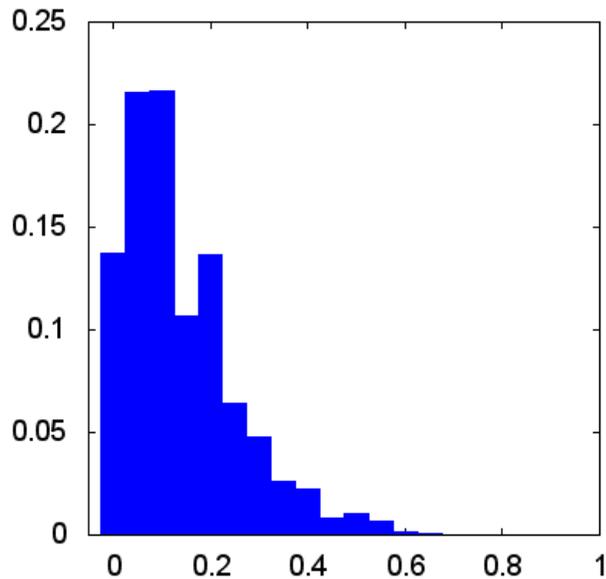


Figure 5.222: Relative frequency of fitness for all selected states of SDSS 587747120521216156

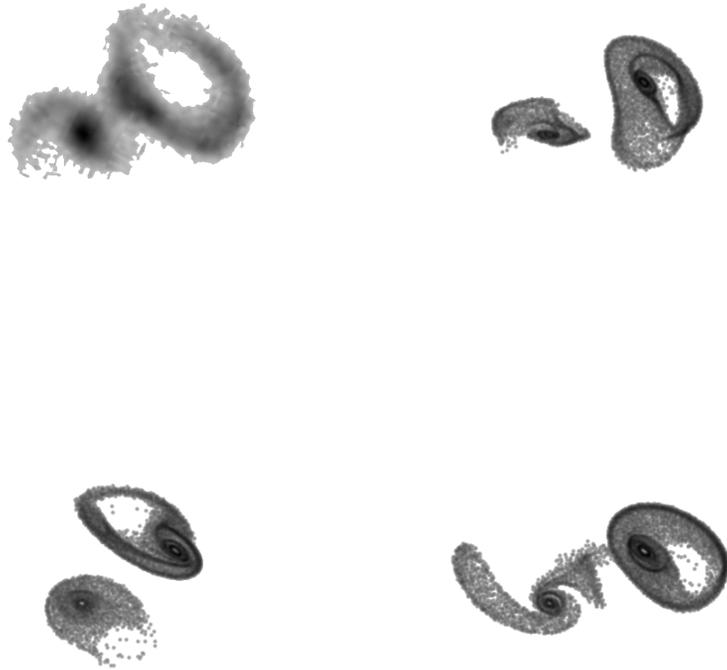


Figure 5.223: Target image and top 3 simulations for SDSS 587747120521216156

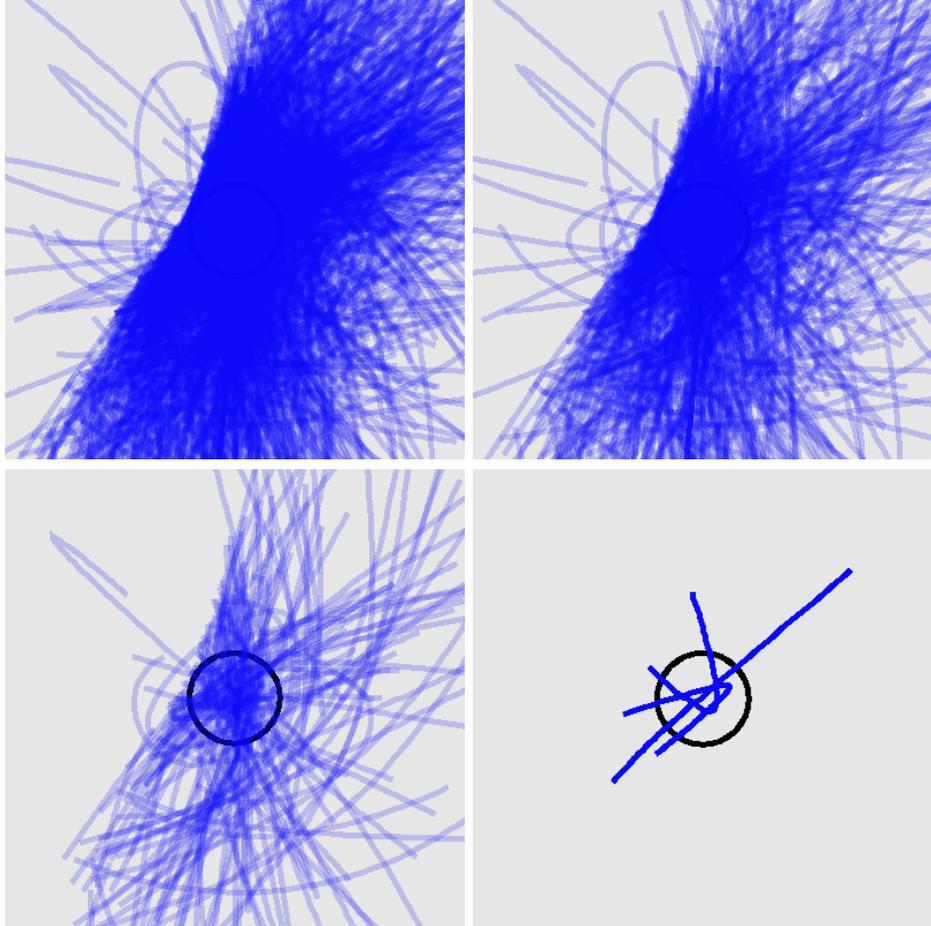


Figure 5.224: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587747120521216156

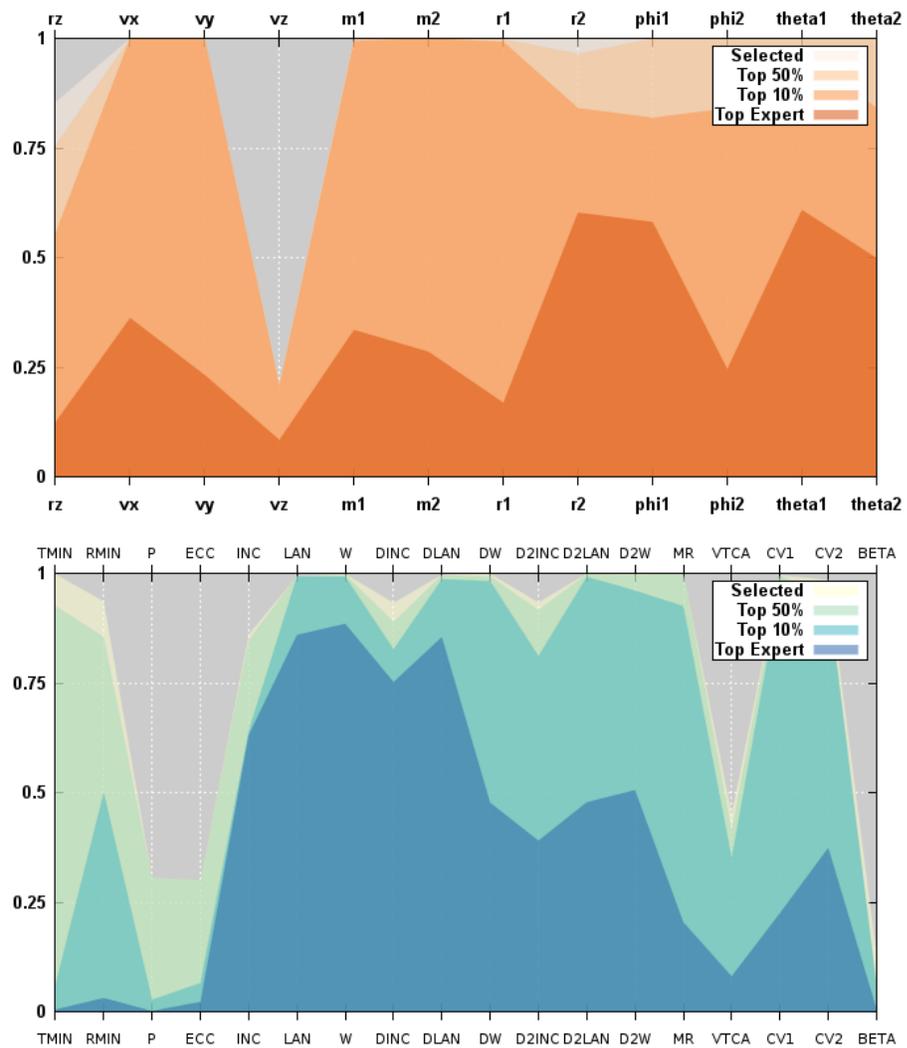


Figure 5.225: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587747120521216156

### 5.1.57 SDSS 588007005230530750

Arp 143 has an almost L-shaped fitness distribution. There is not distinct tidal feature, however, the elongated disk of the primary galaxy is moderately well matched by the top simulations. The trajectories are similar. The simulation parameter convergence is good except for the sizes of the two disks.

Table 5.57: Identification Information and Merger Zoo summary for SDSS 588007005230530750.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 588007005230530750	Arp 143, NGC 2444/2445	7:46:52.9	+39:01:55.6	0.013503

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
75426	73228	2198	375	38320	14153	24167

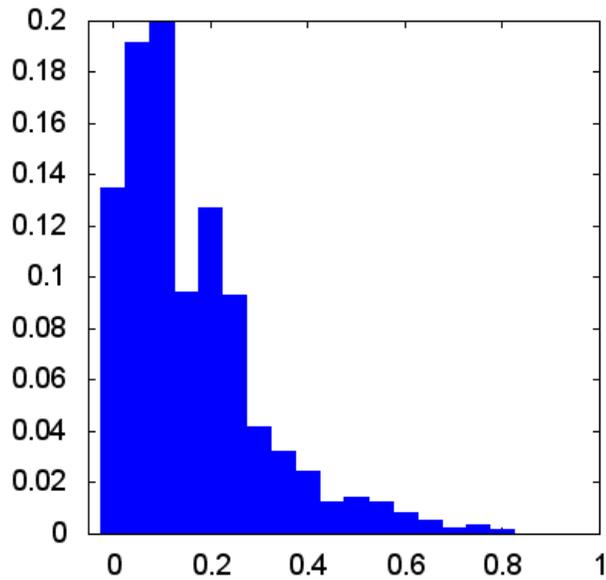


Figure 5.226: Relative frequency of fitness for all selected states of SDSS 588007005230530750

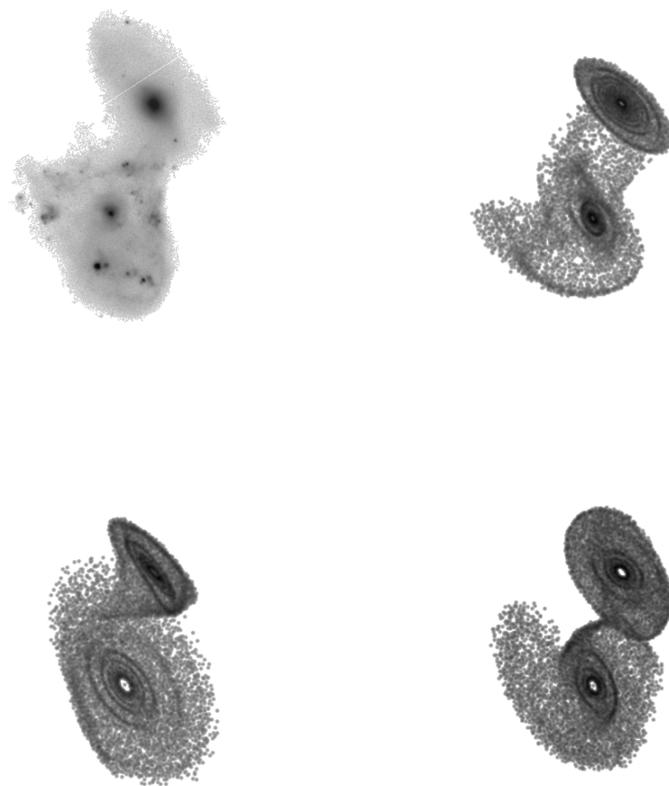


Figure 5.227: Target image and top 3 simulations for SDSS 588007005230530750

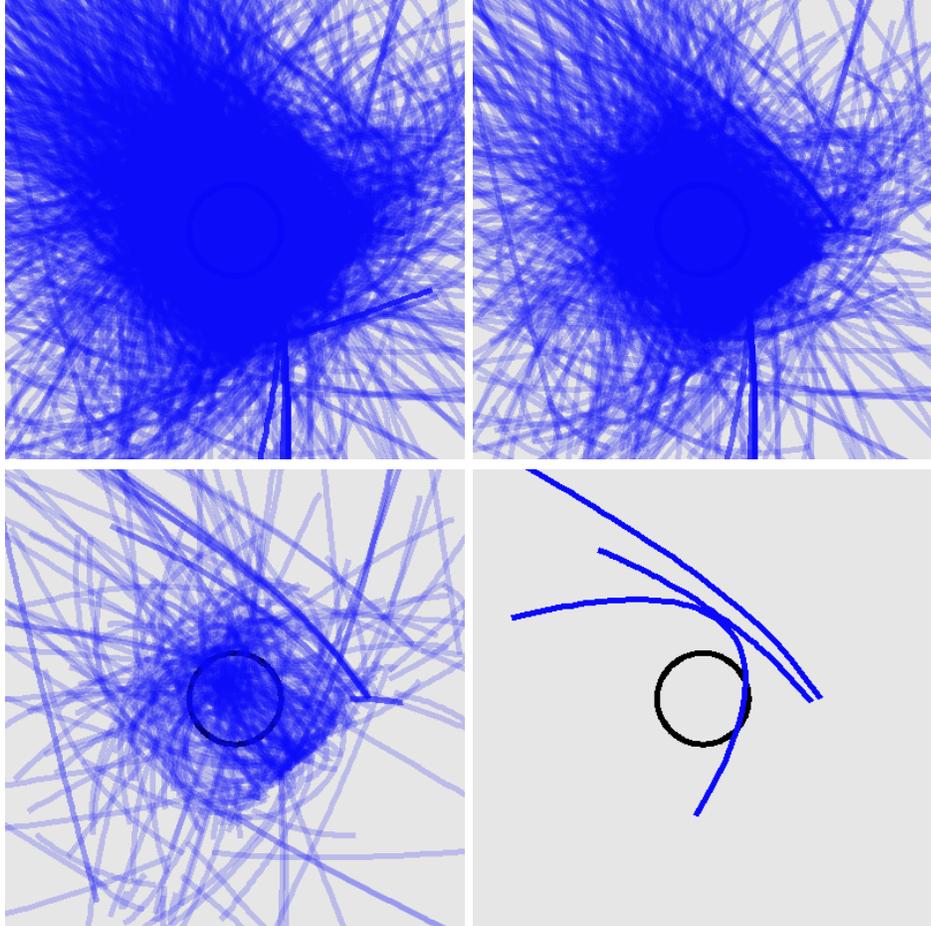


Figure 5.228: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 588007005230530750

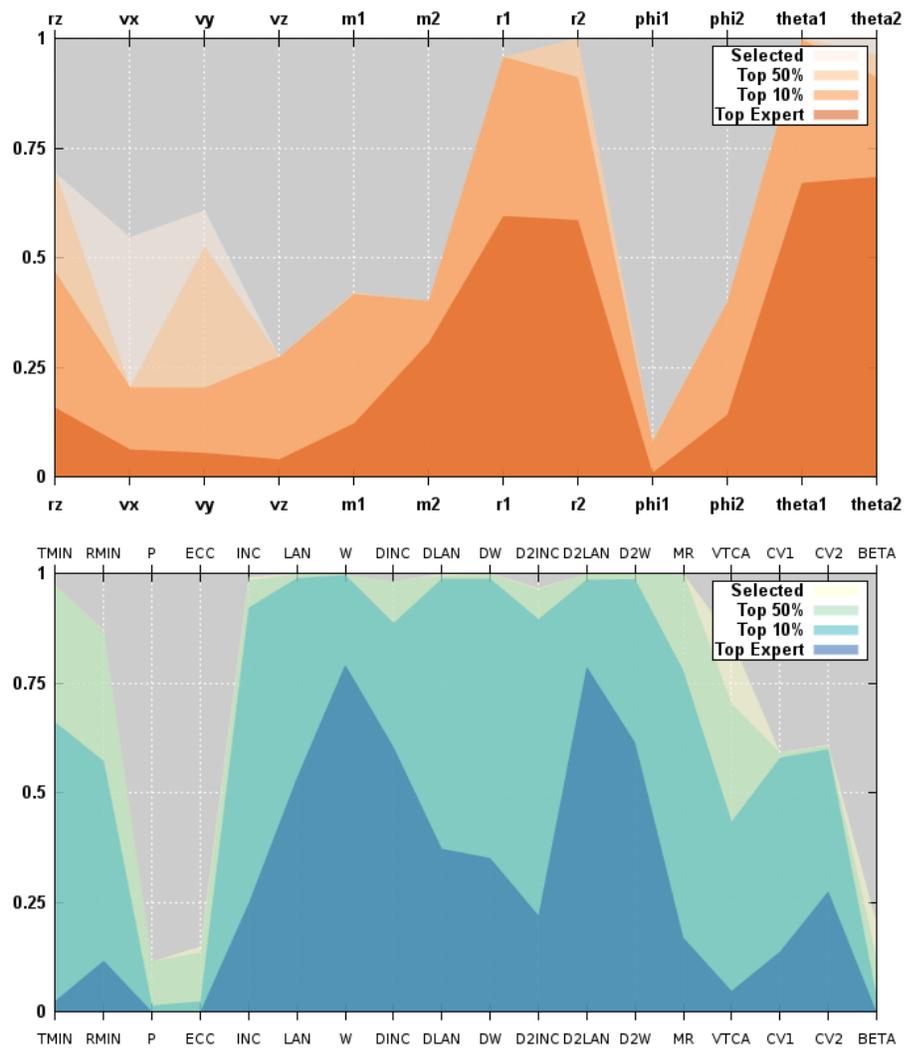


Figure 5.229: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 588007005230530750

### 5.1.58 SDSS 758877153600208945

Arp 70 was viewed in over  $\sim 177000$  simulations. The primary galaxy has a nicely hooked tidal tail that was well matched by the volunteers. The trajectories are very similar. The simulation parameters are not well converged due to the masses.

Table 5.58: Identification Information and Merger Zoo summary for SDSS 758877153600208945.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 758877153600208945	Arp 70, UGC 934	1:23:28.2	+30:47:04.0	0.035004

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
176999	173502	3497	429	60763	28395	32368

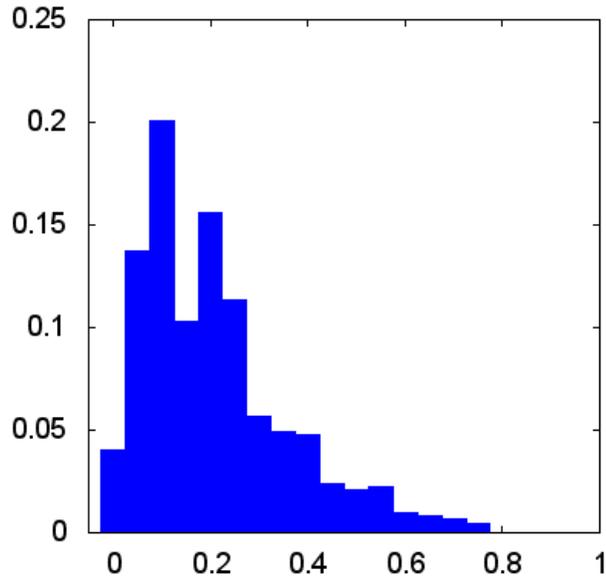


Figure 5.230: Relative frequency of fitness for all selected states of SDSS 758877153600208945

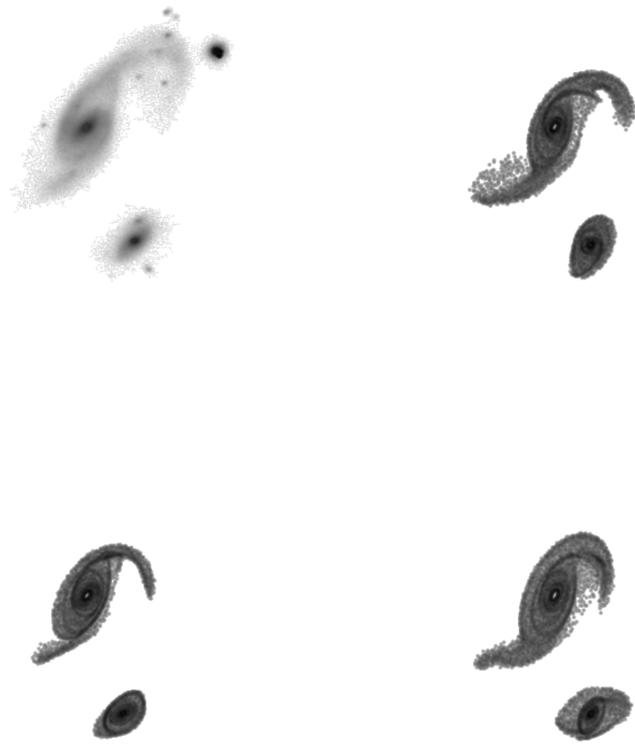


Figure 5.231: Target image and top 3 simulations for SDSS 758877153600208945

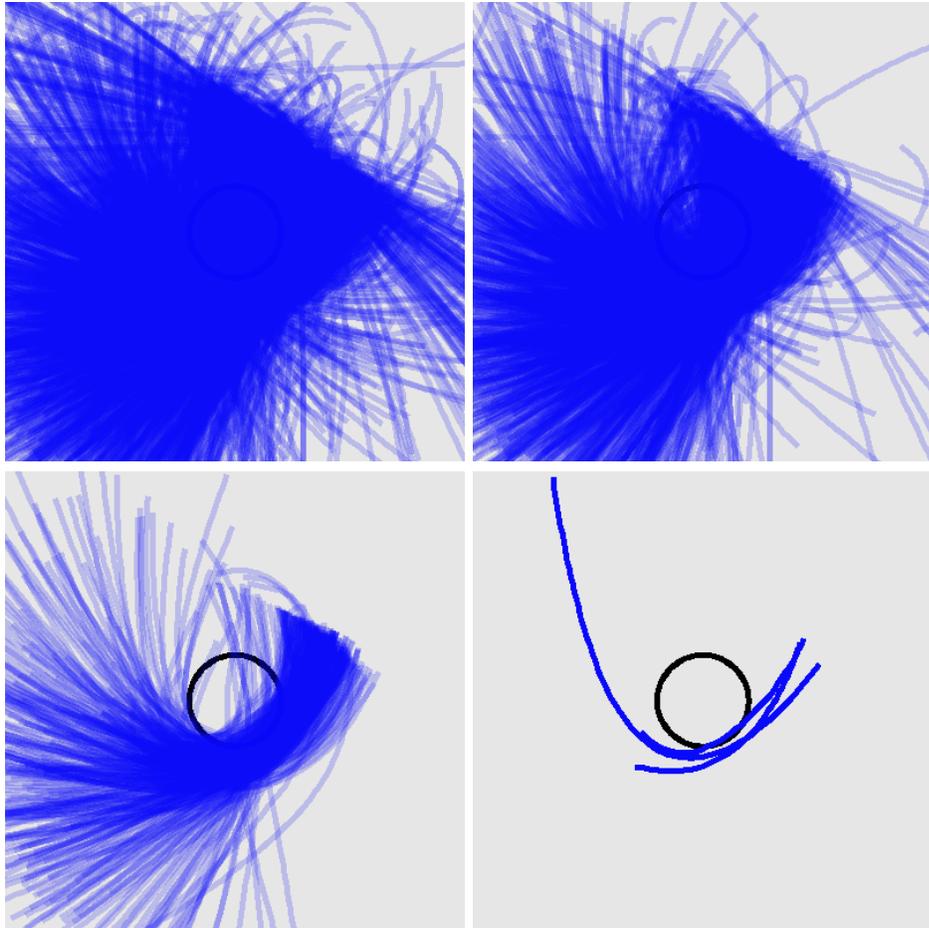


Figure 5.232: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 758877153600208945

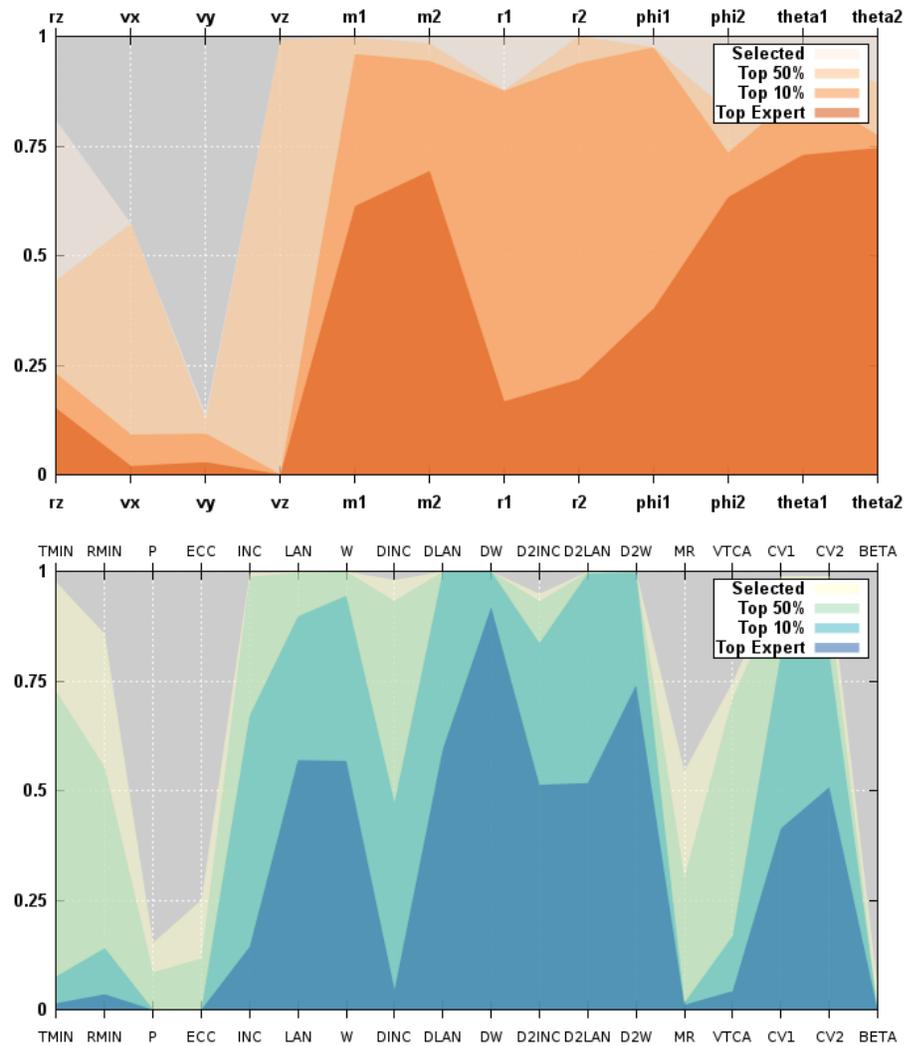


Figure 5.233: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 758877153600208945

### 5.1.59 SDSS 587739720308818095

Arp 218 has a hooked tidal tail. That feature was only reproduced well by the best simulation. The trajectories are not very similar. However, the simulation parameters are very well converged.

Table 5.59: Identification Information and Merger Zoo summary for SDSS 587739720308818095.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 587739720308818095	Arp 218, UGC 10084	15:53:36.8	+18:36:34.5	0.049944

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
47867	46288	1579	302	25282	9410	15872

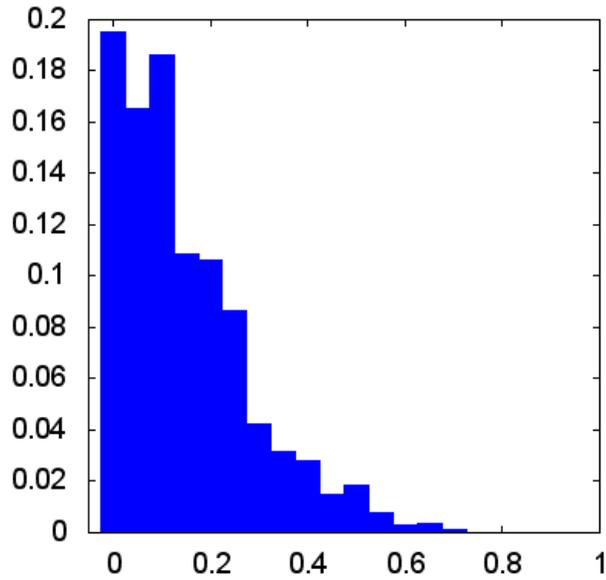


Figure 5.234: Relative frequency of fitness for all selected states of SDSS 587739720308818095

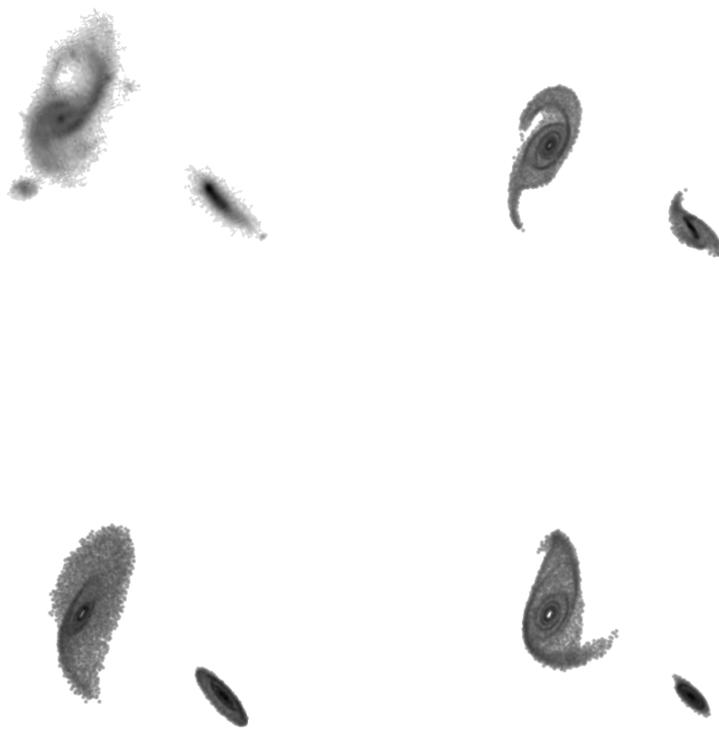


Figure 5.235: Target image and top 3 simulations for SDSS 587739720308818095

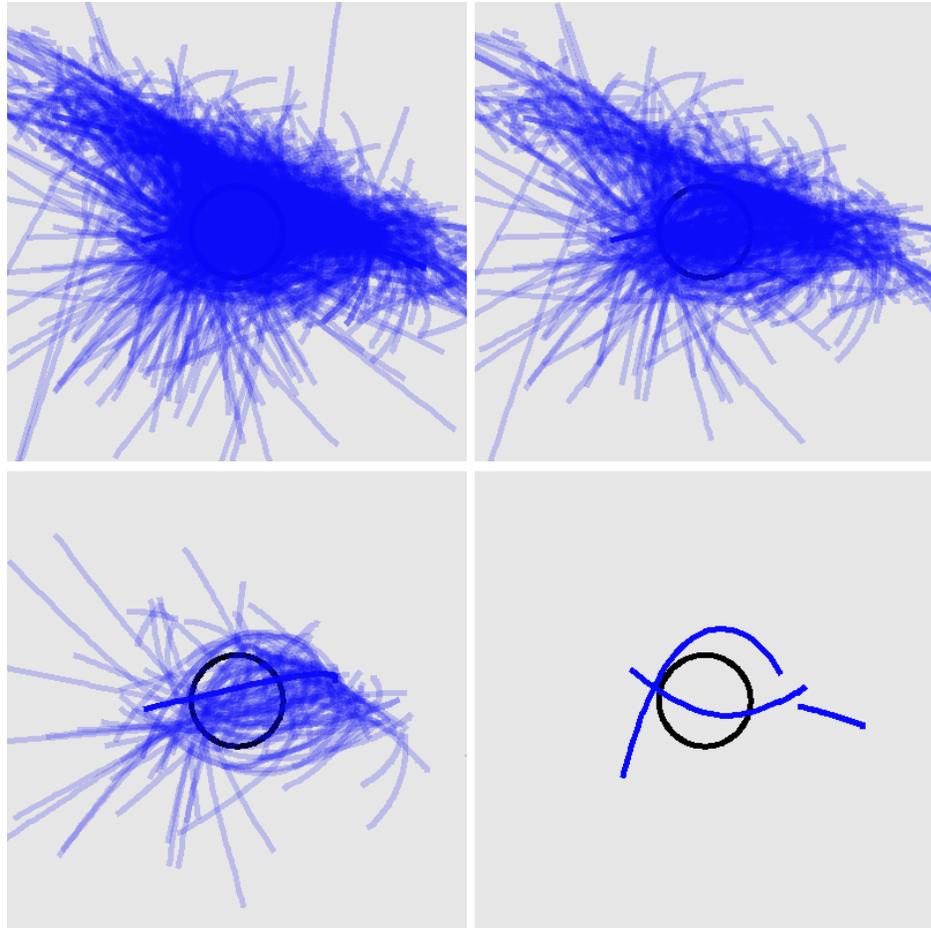


Figure 5.236: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 587739720308818095

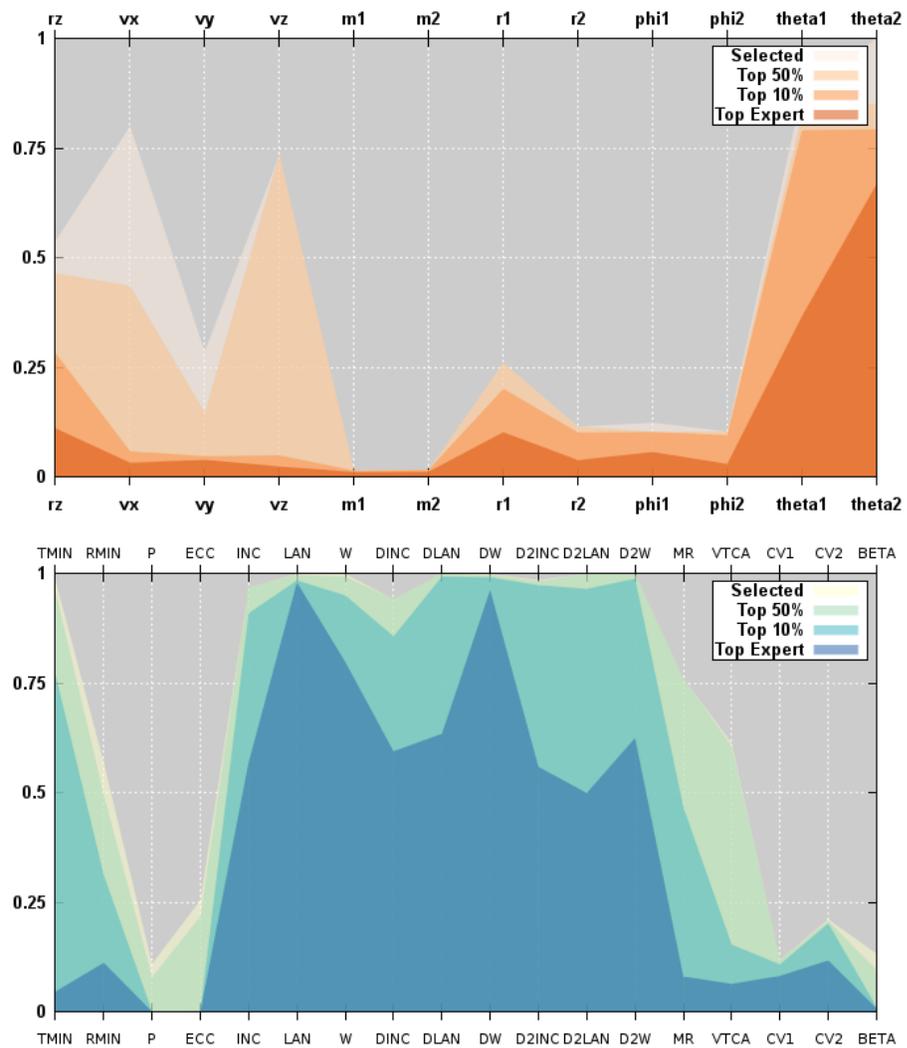


Figure 5.237: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 587739720308818095

### 5.1.60 Arp 273

Arp 273 has an L-shaped fitness distribution. The distortion of the primary galaxy is often described as a rose. The simulations failed to reproduce the void in the disk of the primary. However, the trajectories were similar in that they all passed close to the actual disk of the primary. The orbit parameters will well constrained with the usual exceptions of the the disk orientation angles.

Table 5.60: Identification Information and Merger Zoo summary for Arp 273.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
Arp 273	Arp 273, 1810/1813	UGC 2:21:28.6	+39:22:31.0	0.025227

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
71836	70249	1587	198	18067	7012	11055

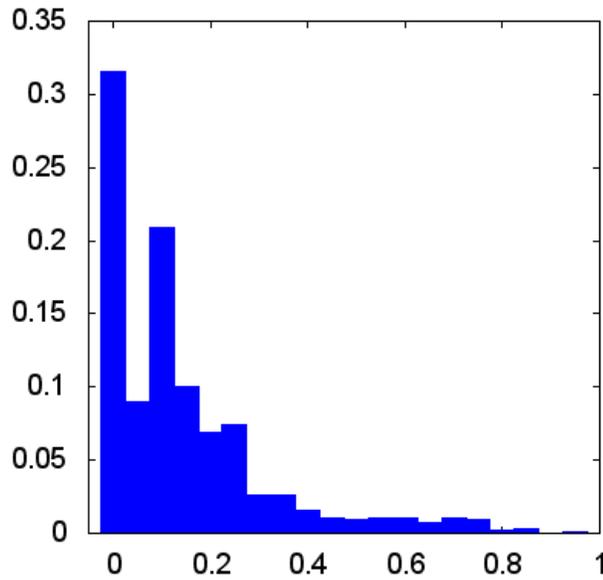


Figure 5.238: Relative frequency of fitness for all selected states of Arp 273

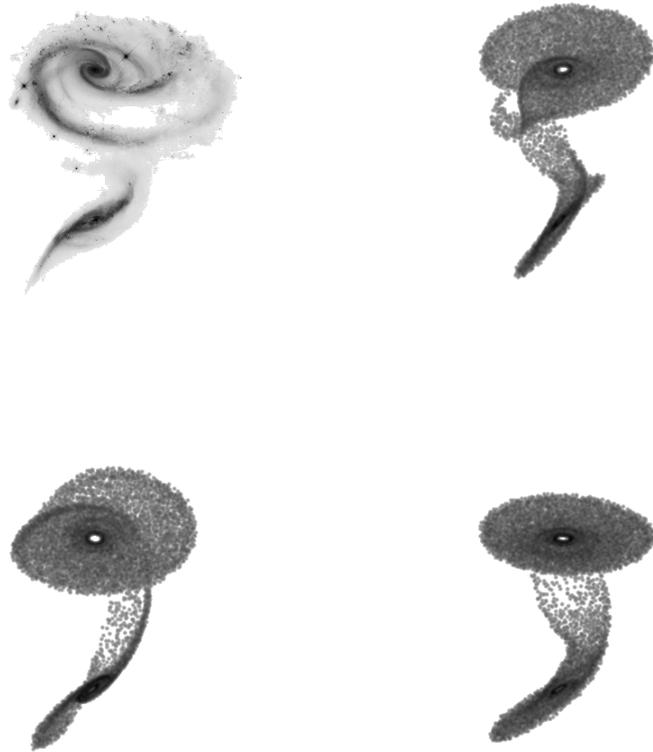


Figure 5.239: Target image and top 3 simulations for Arp 273

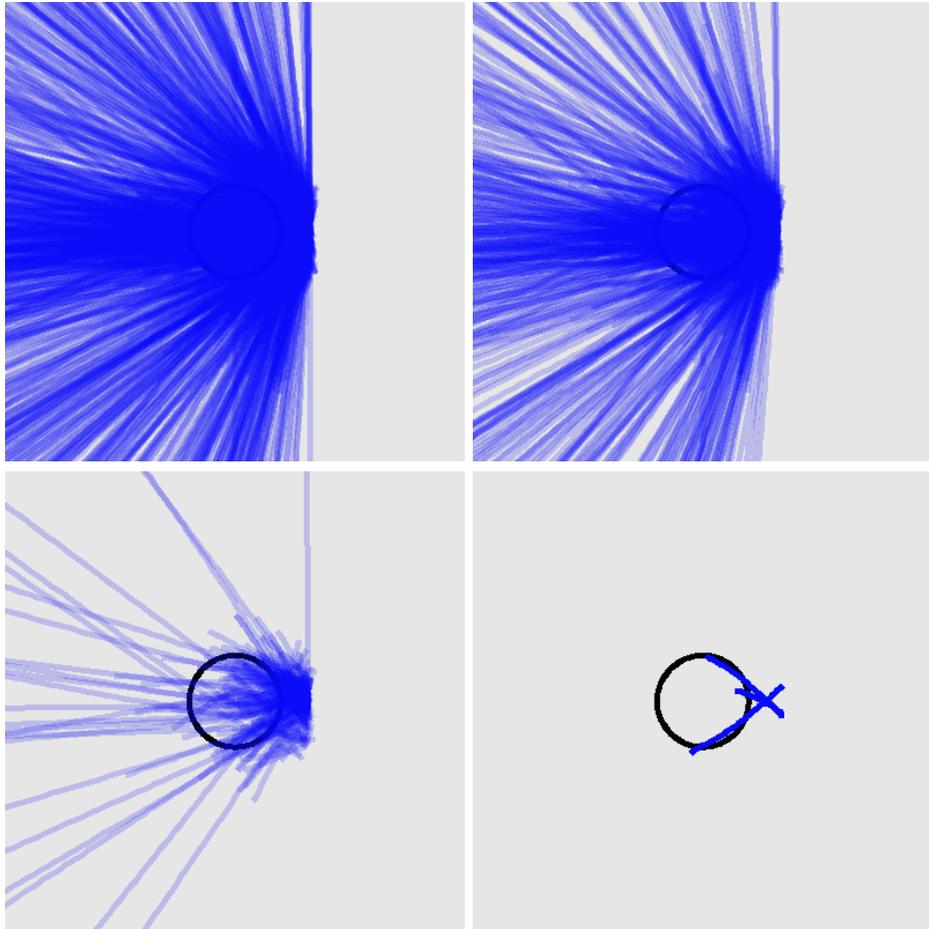


Figure 5.240: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for Arp 273

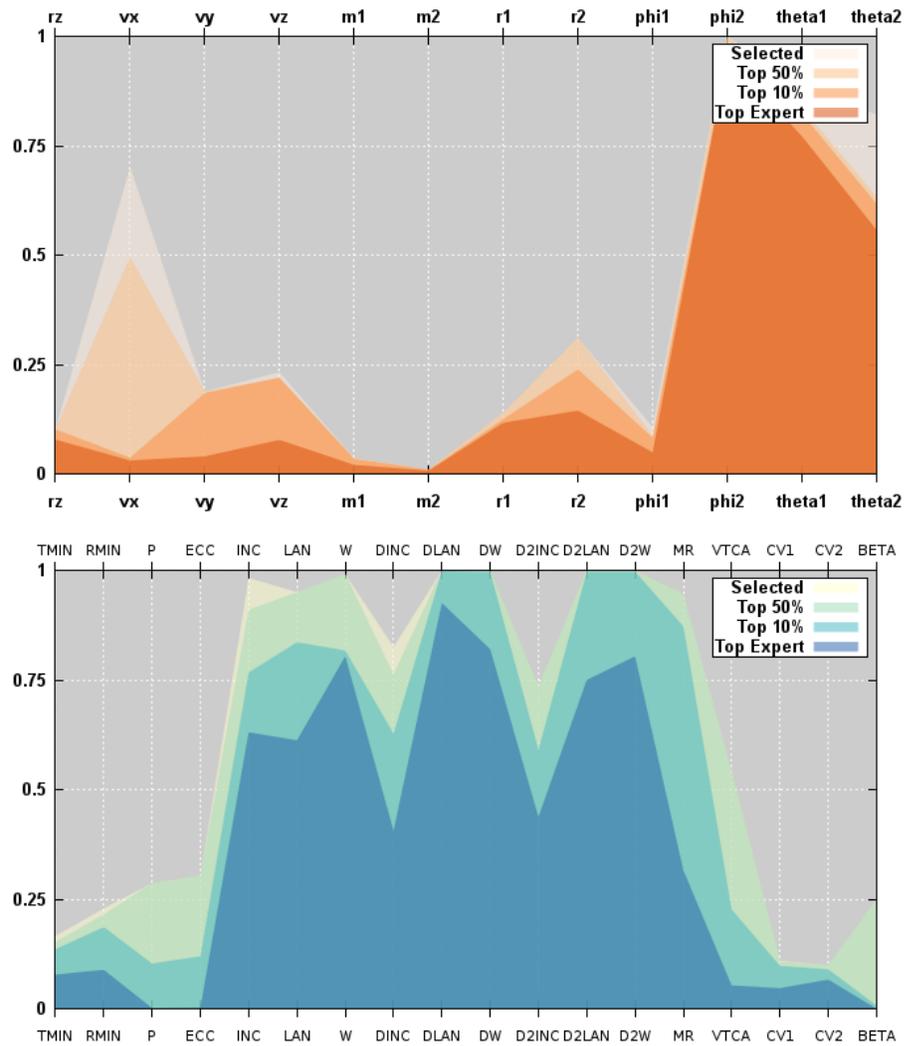


Figure 5.241: Parallel coordinates for convergence of simulation and orbit parameters for Arp 273

### 5.1.61 SDSS 1237678620102623480

This target was the only target from SDSS DR8. It is sometimes referred to as the Violin Clef galaxy. It is thought to be part of a multiple merger system with a total of four galaxies at the same approximate redshift (Willett, 2011). There was no redshift value available from SDSS when this galaxy was first discovered by a Citizen Scientist from the Galaxy Zoo project. We modeled the system as a simple pair. The fitness distribution resulted in a partial match to the tidal tails. The length of the northern tail was not well matched by the simulations. The trajectories are very similar. The simulation parameter convergence is moderate to poor, though the velocity does seem well constrained.

Table 5.61: Identification Information and Merger Zoo summary for SDSS 1237678620102623480.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
SDSS 1237678620102623480	2MASS 00040703+0320165	0:04:15.4	+3:23:01.8	0.02288

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
72808	71767	1041	257	18558	5706	12852

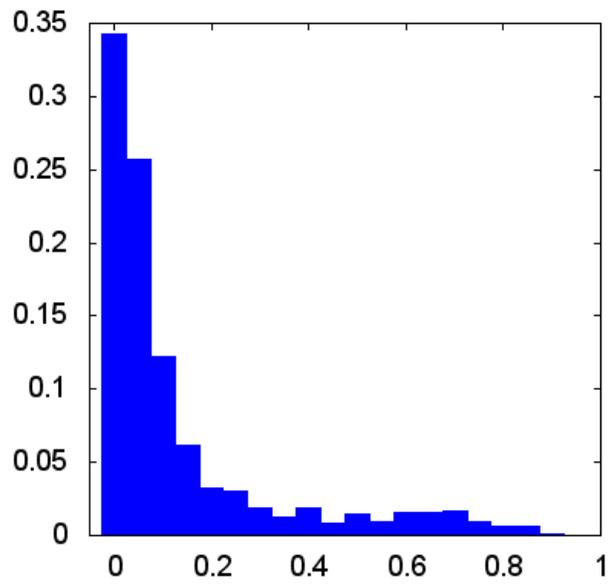


Figure 5.242: Relative frequency of fitness for all selected states of SDSS 1237678620102623480



Figure 5.243: Target image and top 3 simulations for SDSS 1237678620102623480

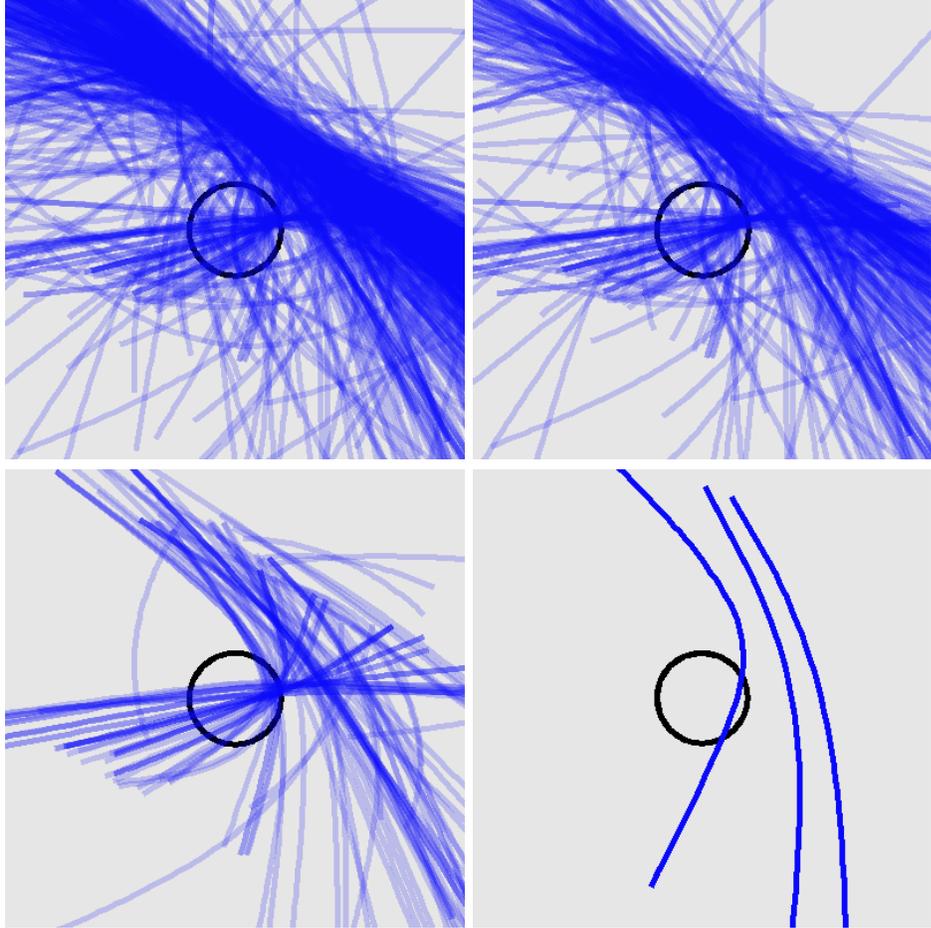


Figure 5.244: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for SDSS 1237678620102623480

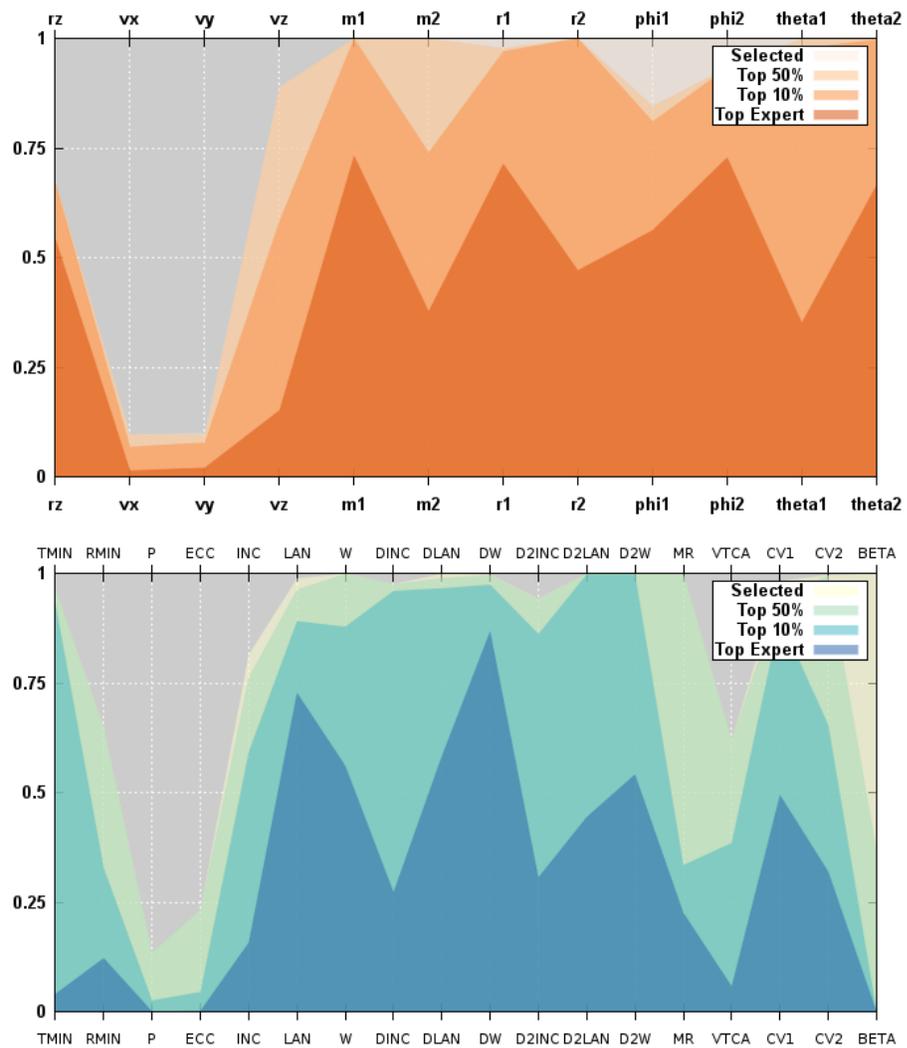


Figure 5.245: Parallel coordinates for convergence of simulation and orbit parameters for SDSS 1237678620102623480

### 5.1.62 Arp 244

Arp 244, the Antennae Galaxies, has an L-shaped fitness distribution. The general shape and orientation of the tidal tails is matched by two of the top three simulations. However the length of the features is usually shorter in the simulations. The trajectories are very similar and appear to be closed orbits. The simulation parameters have moderate convergence. Interestingly, the  $v_y$  parameter is not well converged in addition to the disk masses.

Table 5.62: Identification Information and Merger Zoo summary for Arp 244.

Name	Aliases	RA (hms)	Dec (dms)	Redshift
Arp 244	Arp 244, NGC 4038/4039	12:01:53.0	-18:52:00.8	0.005601

Viewed	Rejected	Selected	Enhanced	MW Comps	MW Wins	Neither
35758	34562	1196	218	16093	5555	10538

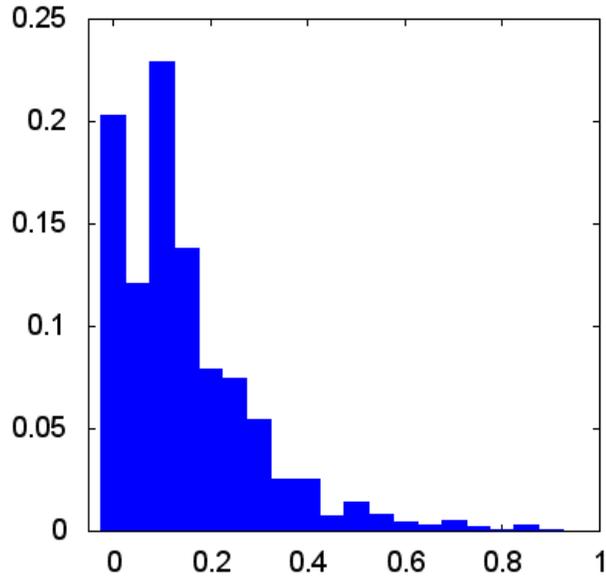


Figure 5.246: Relative frequency of fitness for all selected states of Arp 244



Figure 5.247: Target image and top 3 simulations for Arp 244

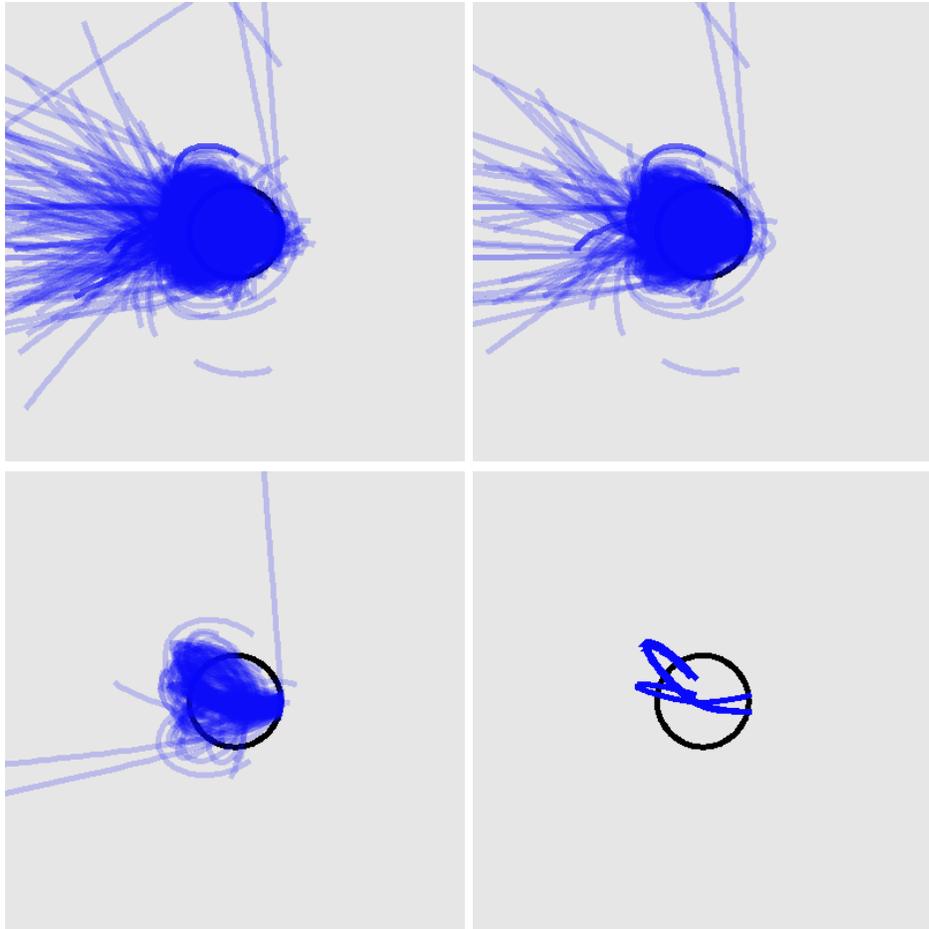


Figure 5.248: Trajectories for all selected states, the top 50%, the top 10%, and the top 3 states for Arp 244

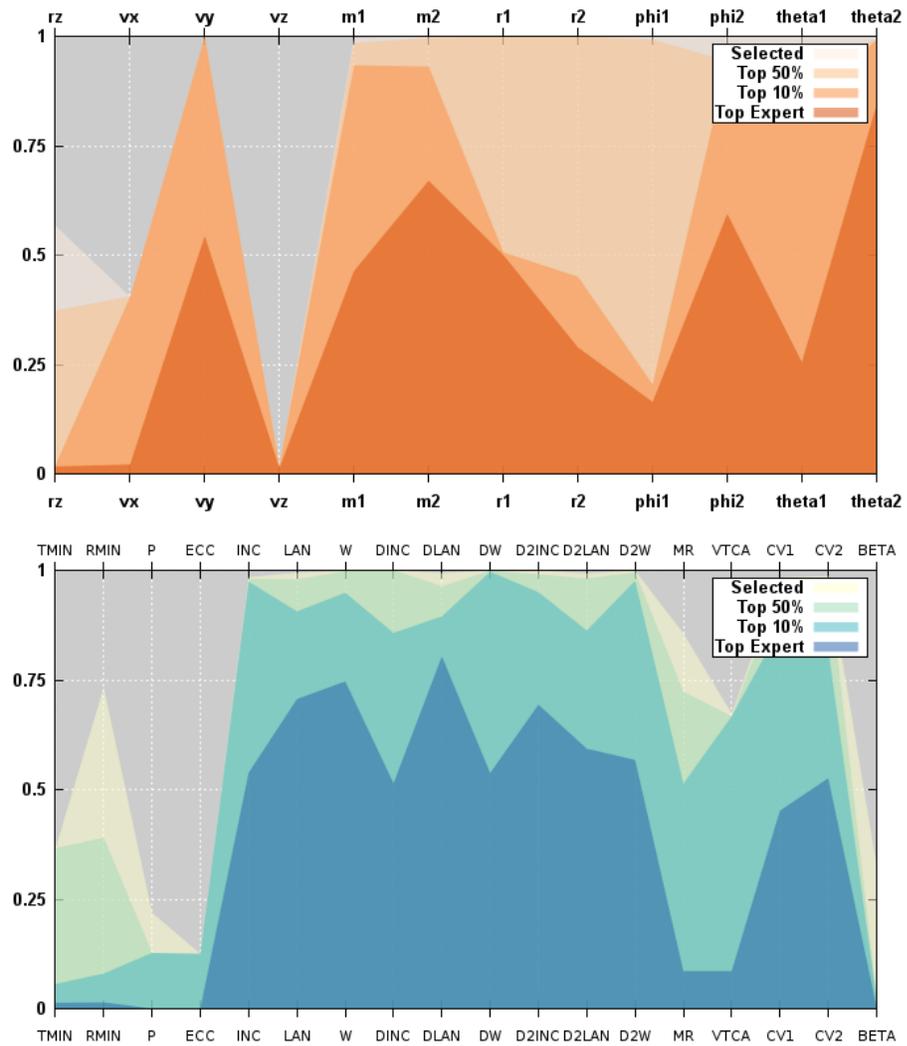


Figure 5.249: Parallel coordinates for convergence of simulation and orbit parameters for Arp 244

## 5.2 Convergence of Model Parameters

The qualitative review of the results for each pair of galaxies revealed an interesting pattern. The pairs where the fitness distributions were L-shaped tended to have a good simulation match to the target image. They also tended to have a high level of convergence. This L-shape is characteristic of power law distributions with long tails. In order to quantify the shape of the fitness distributions we computed the first four statistical moments for the fitness population of each galaxy pair. From those moments we calculated the skewness and kurtosis. Figure 5.250 shows the kurtosis and skewness for each fitness distribution.

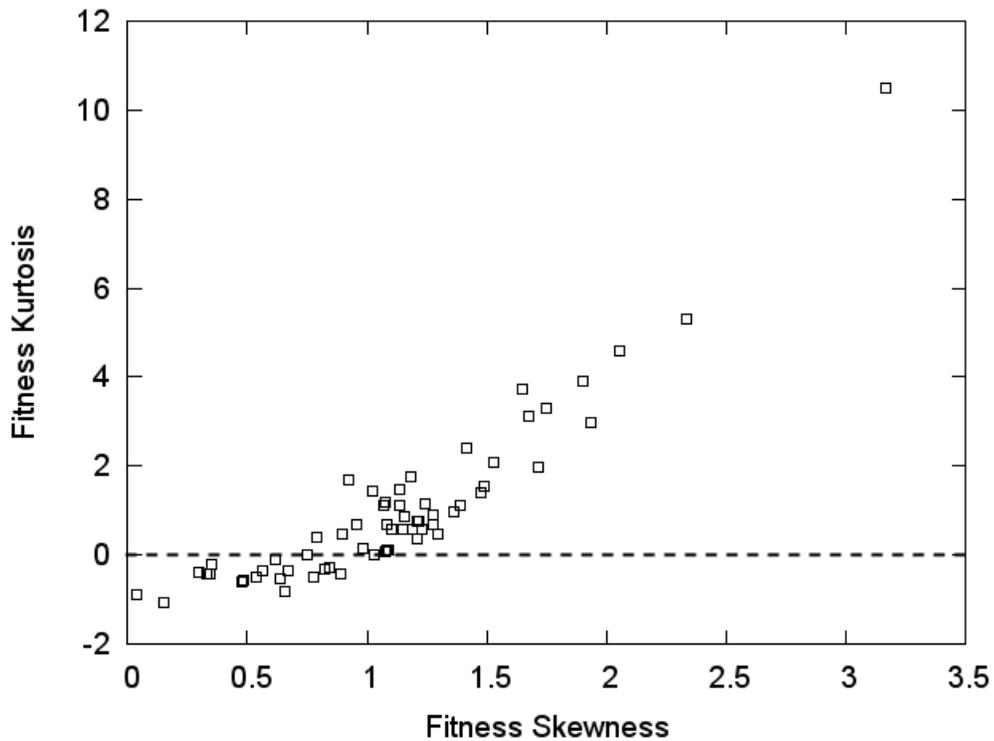


Figure 5.250: Kurtosis vs. Skewness of the fitness distribution.

The kurtosis value is negative for the fitness distribution of 18 pairs. This indicates that their fitness distributions are flatter than a normal distribution. These distributions also have low skewness resulting in less differentiation between low- and high- fitness simulations.

For comparison we show the target image, best fit simulation, fitness distribution, top 3 trajectories and glyph plots for several galaxy pairs. Figure 5.251 shows the three pairs with the lowest skewness. All three have almost no distinguishable tidal features. The fitness distributions are relatively flat. The trajectories are divergent and the glyph plots do not indicate a high level of convergence for simulation and orbit parameters.

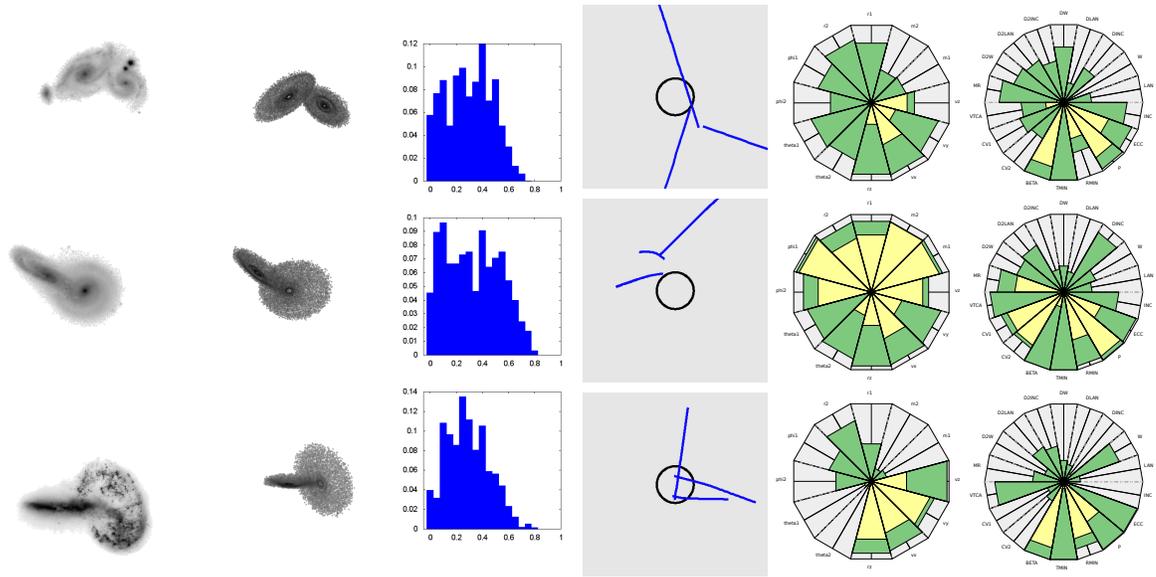


Figure 5.251: The galaxy pairs with the three lowest fitness distribution skewnesses.

Figure 5.252 shows the three pairs with the highest skewness. All three have very obvious tidal features. The fitness distributions are very skewed with most simulations given a fitness score of zero. The trajectories are consistent. However, the glyph plots are not remarkably different from those of the low skewness galaxy pairs. This seems to indicate that the parallel coordinate and glyph plots are not a good indicator of which sets of simulations have converged to a high fitness match of the simulation. One possible explanation is that not every target was prepared identically. In some instances simulation parameter ranges were edited by hand. For example, the double-ring galaxy in our sample had the ranges for  $v_x$  and  $v_y$  restricted to keep their relative magnitudes low compared to

$r_z$ . This was done to ensure more simulation parameter sets would be selected with an overall velocity perpendicular to the plane of the sky in order to encourage formation of rings<sup>1</sup>. Other simulation ranges were edited on an ad hoc basis. Also, because some of the parameters are based on intrinsic properties of each galaxy pair (velocity ranges were constrained by estimated mass), not every simulation parameter was sampled across the same range of values for each galaxy.

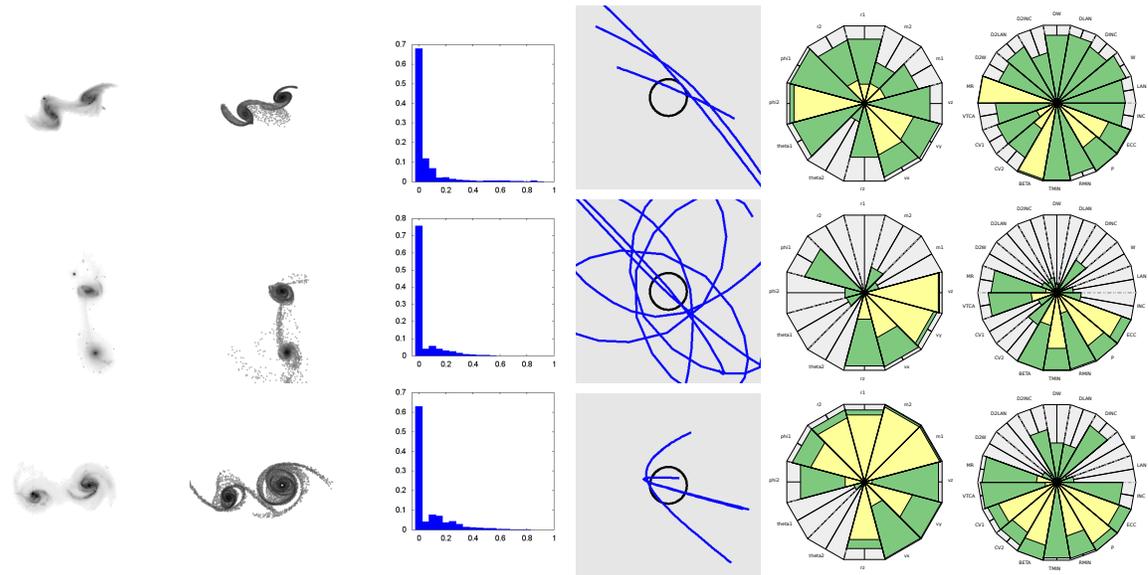


Figure 5.252: The galaxy pairs with the three highest fitness distribution skewnesses.

Another trend hinted at during the qualitative review of all of the galaxy pairs was that target images with larger, more distinct tidal features tended to be more popular. We wanted to know if the activity level for a target influenced the skewness. In Figure 5.253 we show how the skewness varies with the number of simulations reviewed. Systems with the most number of simulations do not have the highest skewness. There is no definite trend overall, though a claim could be made that the targets with the fewest number of simulations viewed tended to have smaller skewness values.

<sup>1</sup>Rings are likely to form when one galaxy passes perpendicularly through the plane, and within the disk, of the other galaxy. Here both galaxies disks had relatively low inclination angles with respect to the plane of the sky, so a velocity vector perpendicular to the sky is likely to be perpendicular to each disk as well.

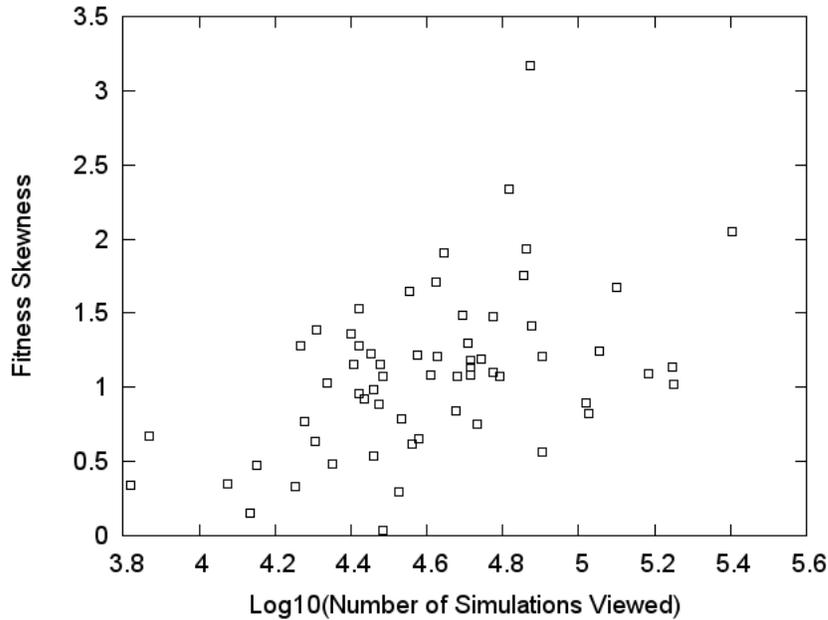


Figure 5.253: Skewness vs. Number of Simulations Viewed.

If the activity level does not heavily influence the fitness skewness, then it remains possible that the distribution of fitness values is related to inherit properties of the interacting galaxies. We group the target images into clusters as described in Section 4.4.1. For each group we discuss any potential similarities between parameter values and convergences for members. It is important to remember that these morphological groupings are not uniquely defined. It is possible that some galaxy pairs belong more to other groups than as defined here. One important benefit of arbitrarily specifying a maximum cluster size is that we can study properties for a finite set of related galaxies at the same time. The last two figures for each group are each a set of box plots for the eccentricity and inclination of the relative orbit with respect to the disk of the primary. For the box plots, the upper and lower adjacent values are included. Several orbits are extremely hyperbolic with very large eccentricities. They are simply off the scale, so assume any missing eccentricity box refers to a hyperbolic orbit.

### 5.2.1 Morphological Group 1

The galaxies in this group have a modest skewness to their fitness distributions, eccentricities around 0.5 and inclinations between  $40^\circ$  and  $60^\circ$ . The exception is the second target with a lower skewness, and larger interquartile range (IQR), and smaller glyph plots.

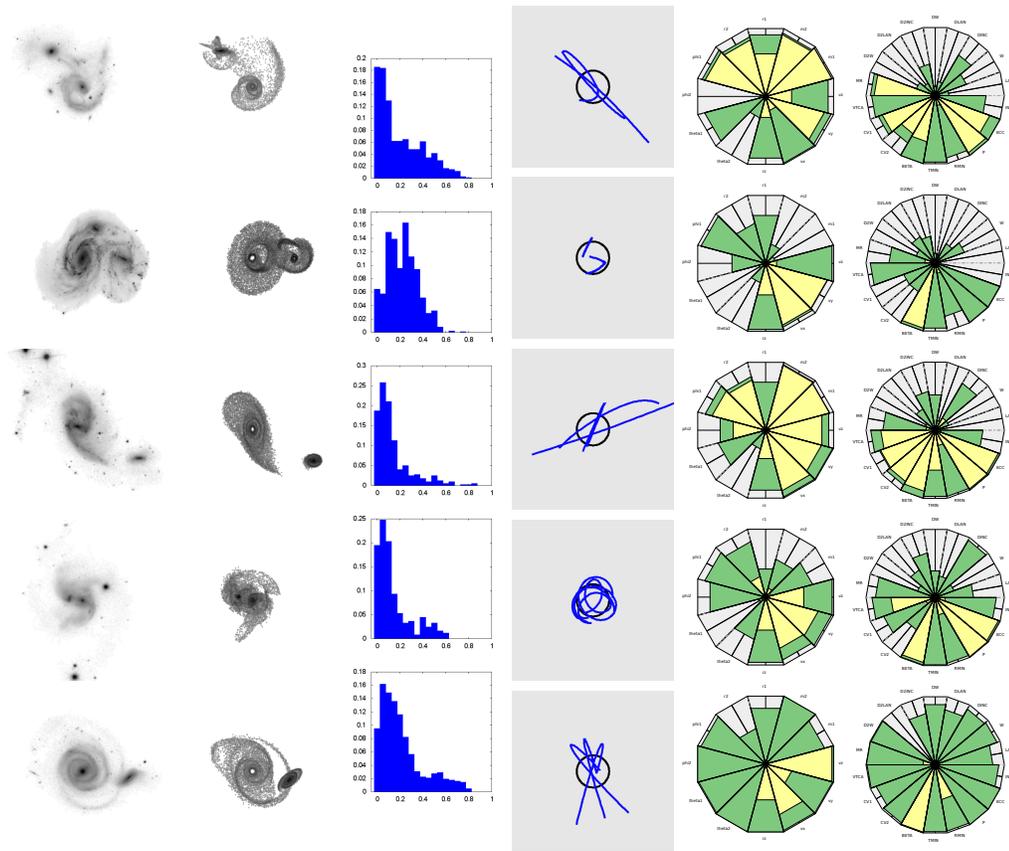


Figure 5.254: Morphological Group 1

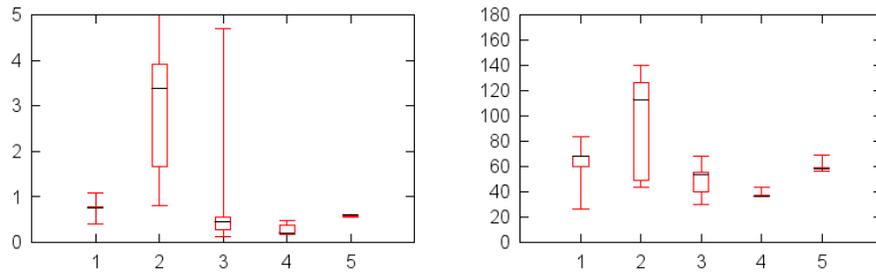


Figure 5.255: The distribution of eccentricity and inclination for the top simulations in group 1

### 5.2.2 Morphological Group 2

Three out of five galaxies in this group have parabolic orbits. The inclinations are in a range of  $30^\circ$  and  $80^\circ$ . The first two galaxies have a low skewness to their fitness distributions while the last three have a higher value.

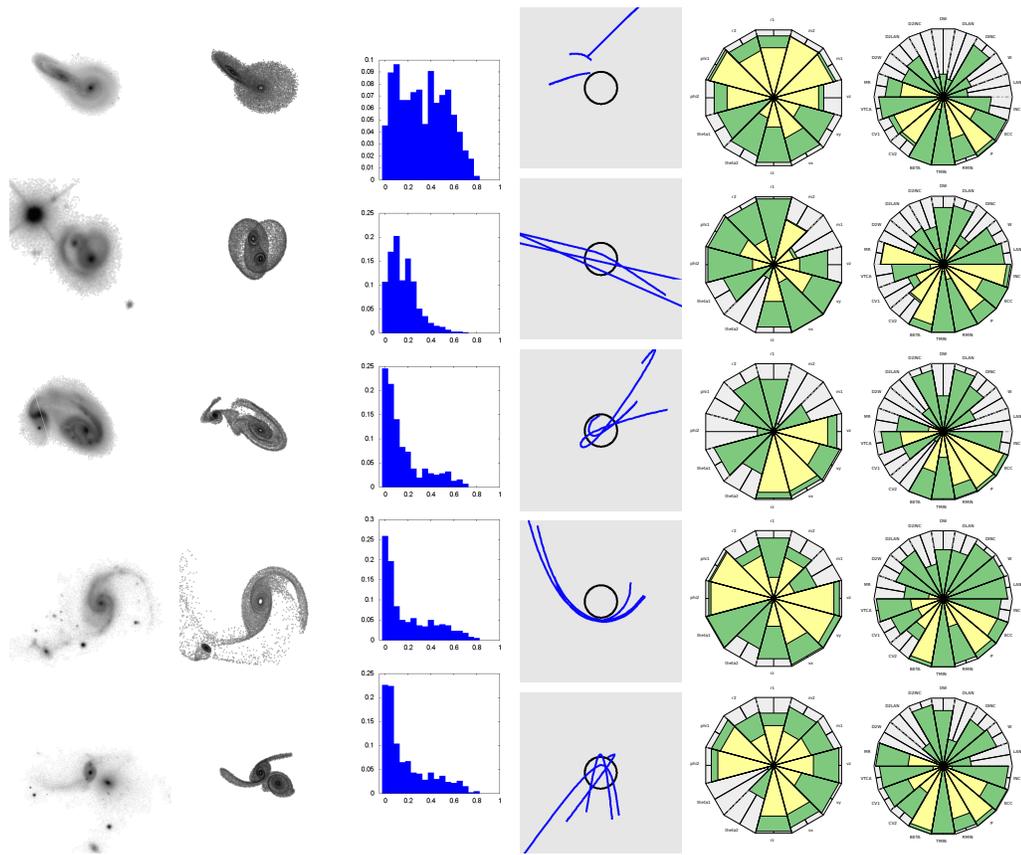


Figure 5.256: Morphological Group 2

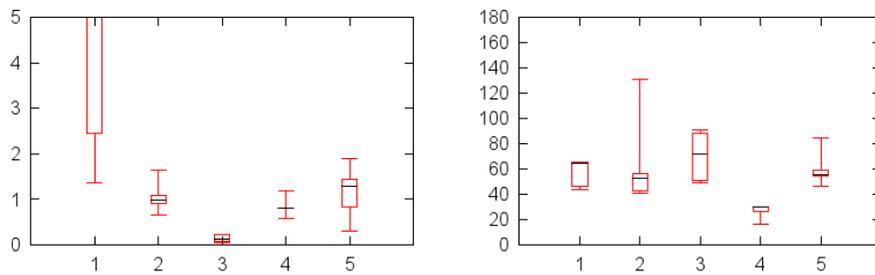


Figure 5.257: The distribution of eccentricity and inclination for the top simulations in group 2

### 5.2.3 Morphological Group 3

All five galaxies in this group have a tightly converged eccentricity less than 1.0 meaning they are closed orbits. The inclinations range from  $40^\circ$  and  $90^\circ$ .

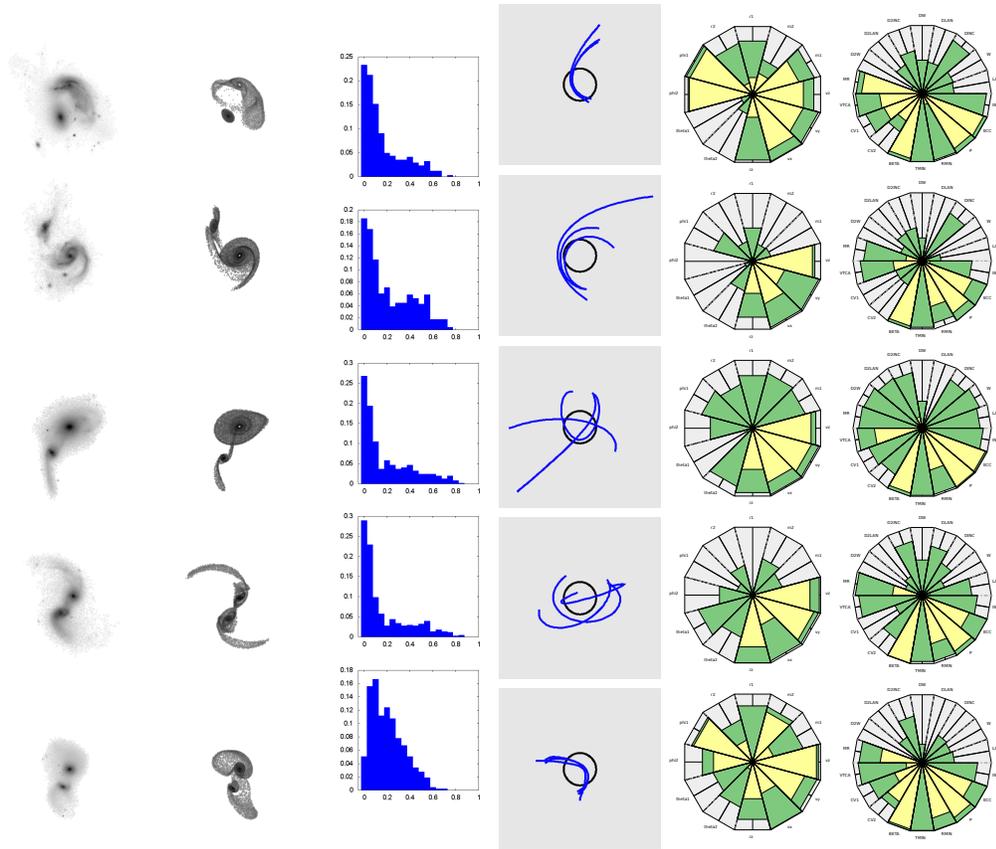


Figure 5.258: Morphological Group 3

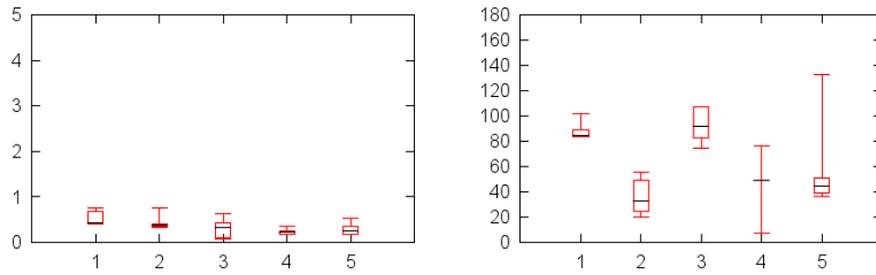


Figure 5.259: The distribution of eccentricity and inclination for the top simulations in group 3

#### 5.2.4 Morphological Group 4

The five galaxies in this group all have low skewness for their fitness distributions. Neither the eccentricities nor inclinations are well constrained with all systems having a median value near the mean of all possible inclinations,  $90^\circ$ .

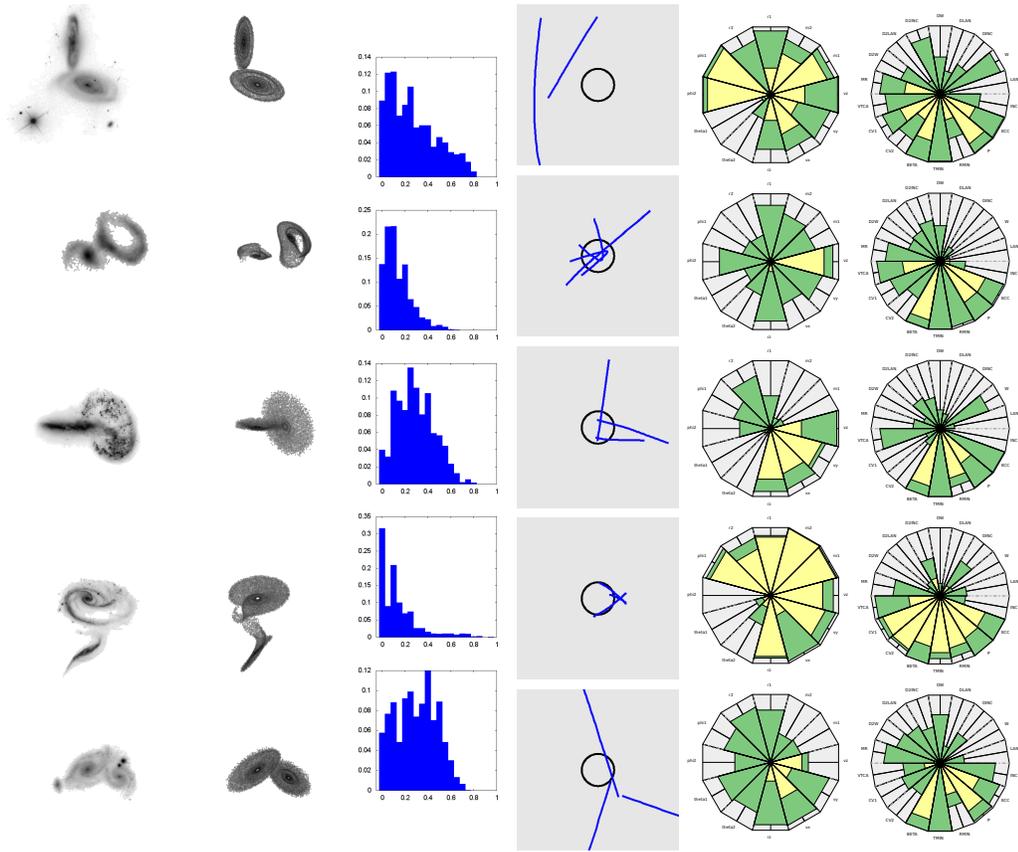


Figure 5.260: Morphological Group 4

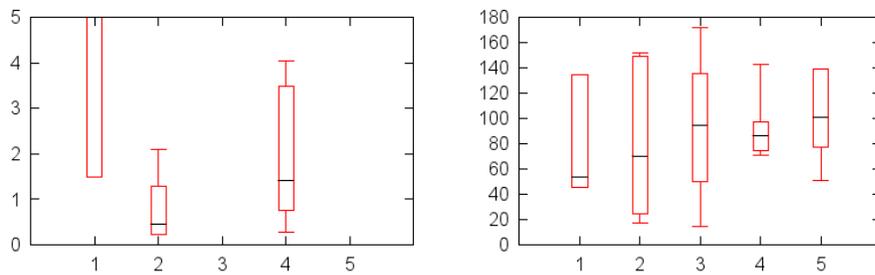


Figure 5.261: The distribution of eccentricity and inclination for the top simulations in group 4

### 5.2.5 Morphological Group 5

Five of six galaxies have low to moderate skewness. The sixth value has a high skewness and a closed orbit while the other five have parabolic orbits. The inclinations range from  $20^\circ$  and  $60^\circ$ , with the least constrained orbit having a high IQR for both inclination and eccentricity.

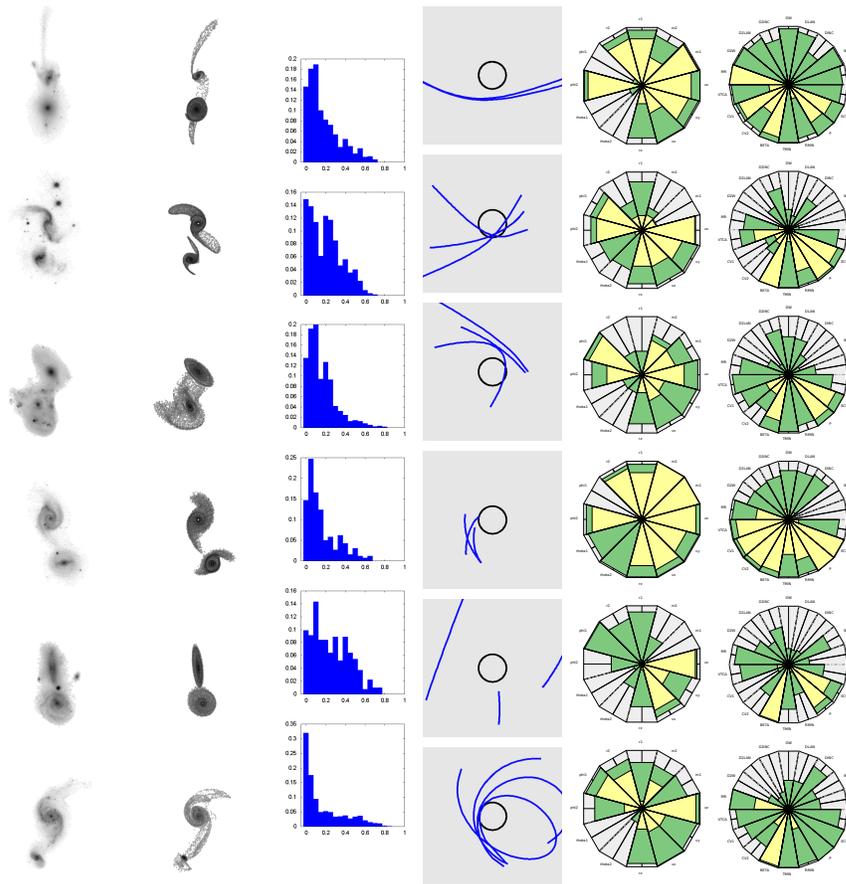


Figure 5.262: Morphological Group 5

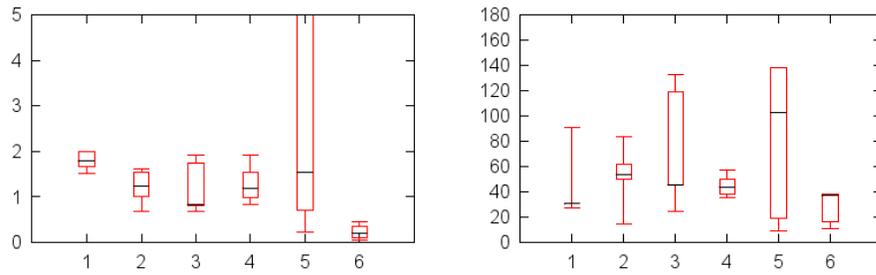


Figure 5.263: The distribution of eccentricity and inclination for the top simulations in group 5

### 5.2.6 Morphological Group 6

These galaxies have moderate to high skewness with five out of six having closed orbits, with tightly constrained eccentricities. The inclinations are less well converged and range between  $40^\circ$  and  $100^\circ$ .

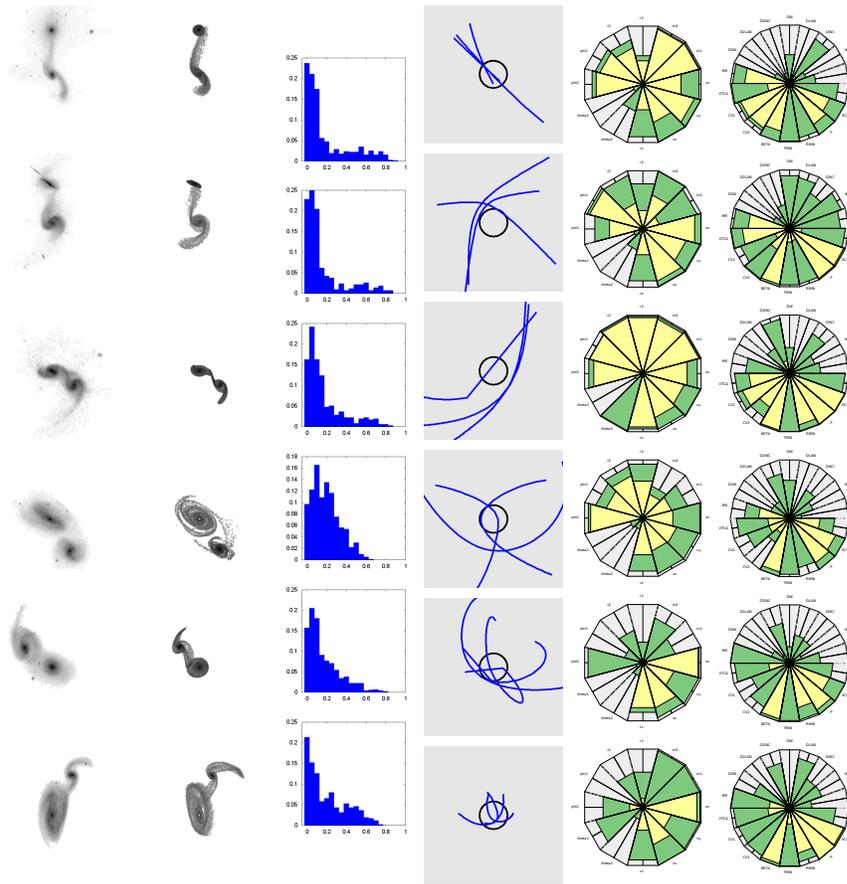


Figure 5.264: Morphological Group 6

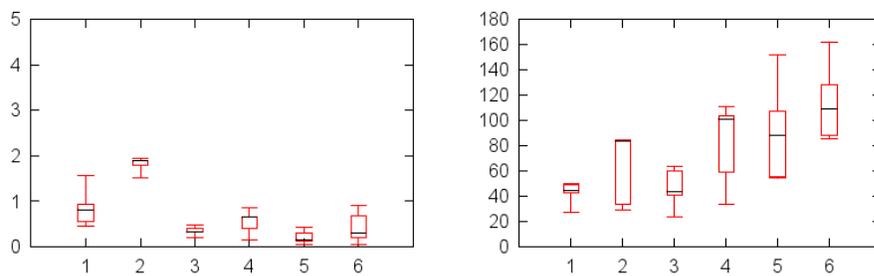


Figure 5.265: The distribution of eccentricity and inclination for the top simulations in group 6

### 5.2.7 Morphological Group 7

The skewness is low for all the fitness distributions in this group. The eccentricities are not well constrained and tend towards parabolic orbits. The median inclinations range from  $40^\circ$  and  $120^\circ$ .

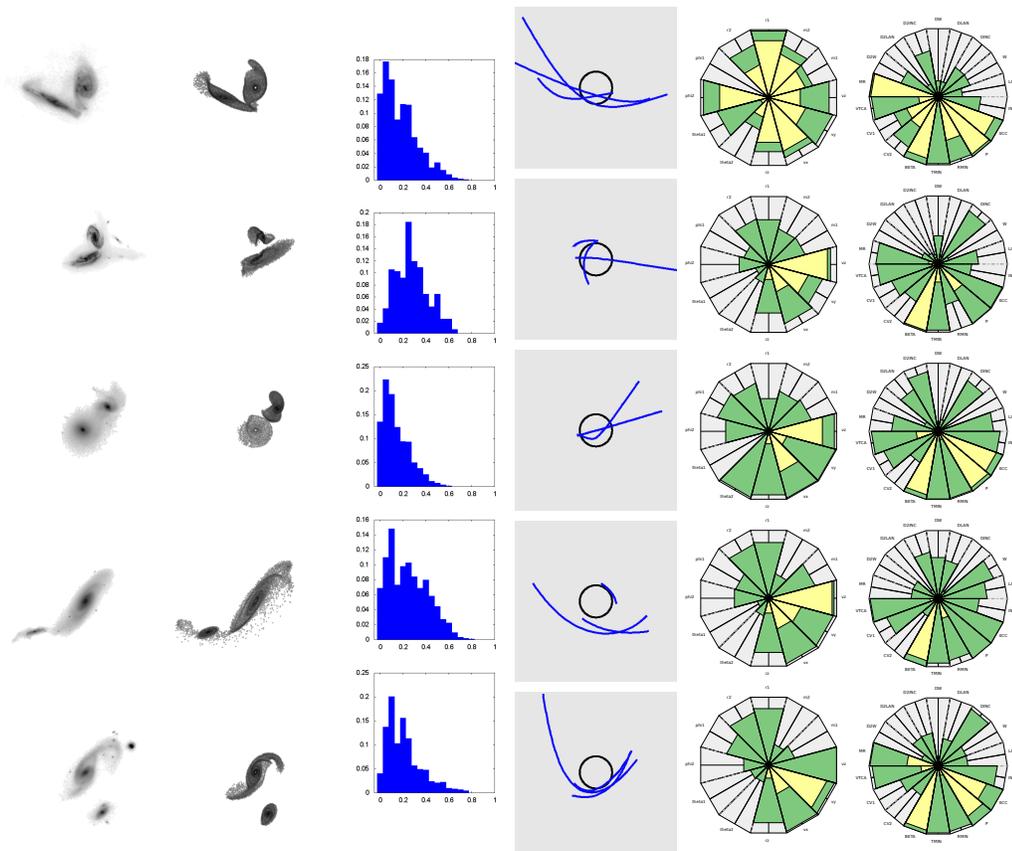


Figure 5.266: Morphological Group 7

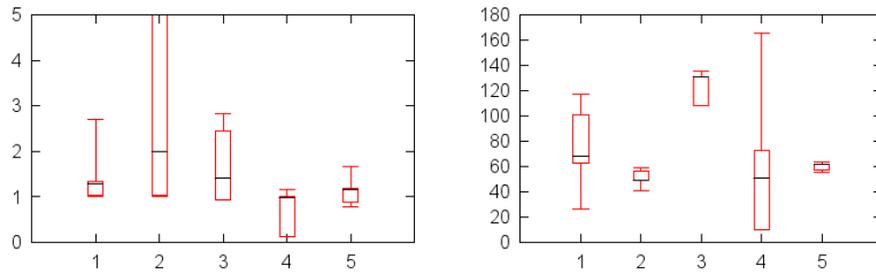


Figure 5.267: The distribution of eccentricity and inclination for the top simulations in group 7

### 5.2.8 Morphological Group 8

The galaxies in this group all have interesting tails. Three of five eccentricities are constrained and represent closed orbits. The inclination are less well constrained and range between  $40^\circ$  and  $80^\circ$ .

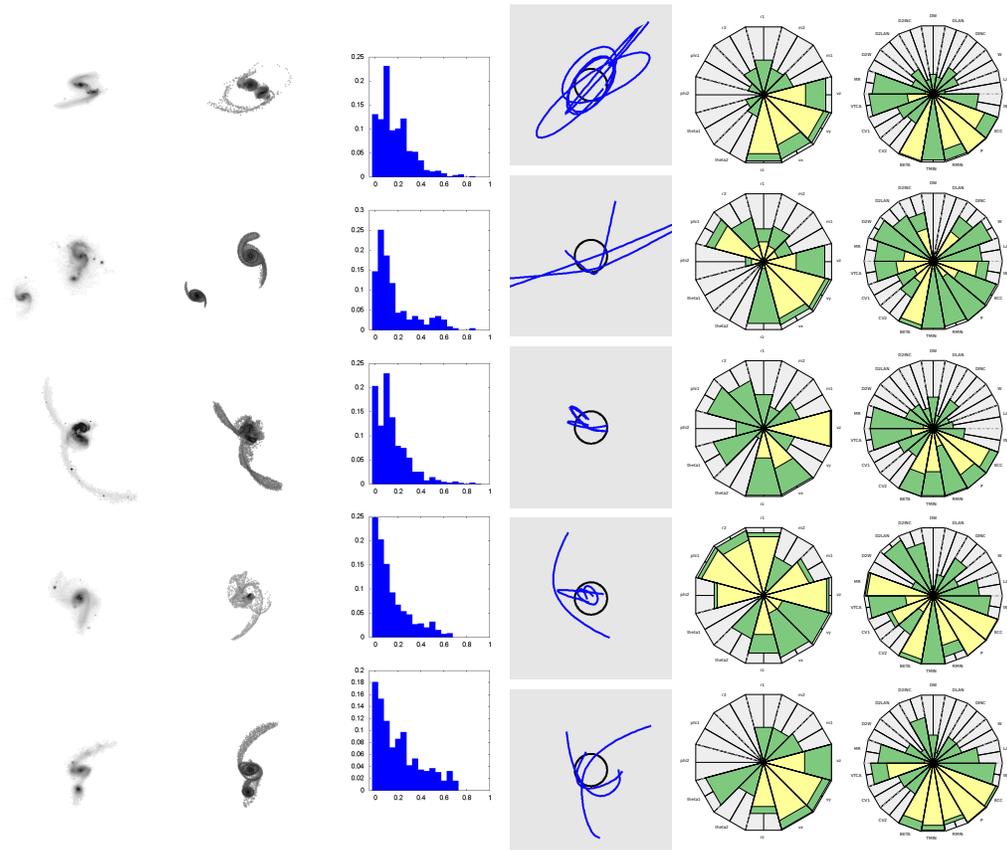


Figure 5.268: Morphological Group 8

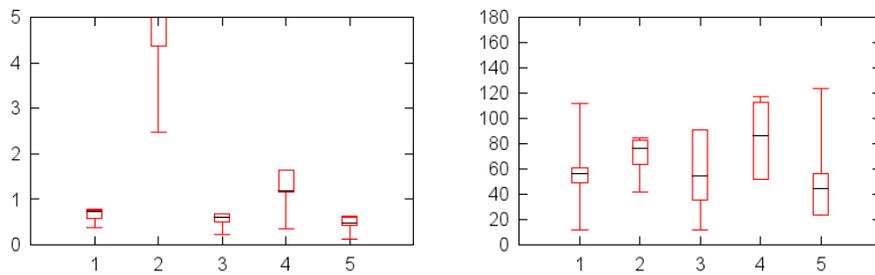


Figure 5.269: The distribution of eccentricity and inclination for the top simulations in group 8

### 5.2.9 Morphological Group 9

This group is a mixed bag with three pairs having parabolic orbits. The other two orbits are closed and hyperbolic. The inclinations are at different ranges with some well constrained near  $10^\circ$  and others nearly perpendicular or retrograde in the range of  $90^\circ$  and  $120^\circ$ .

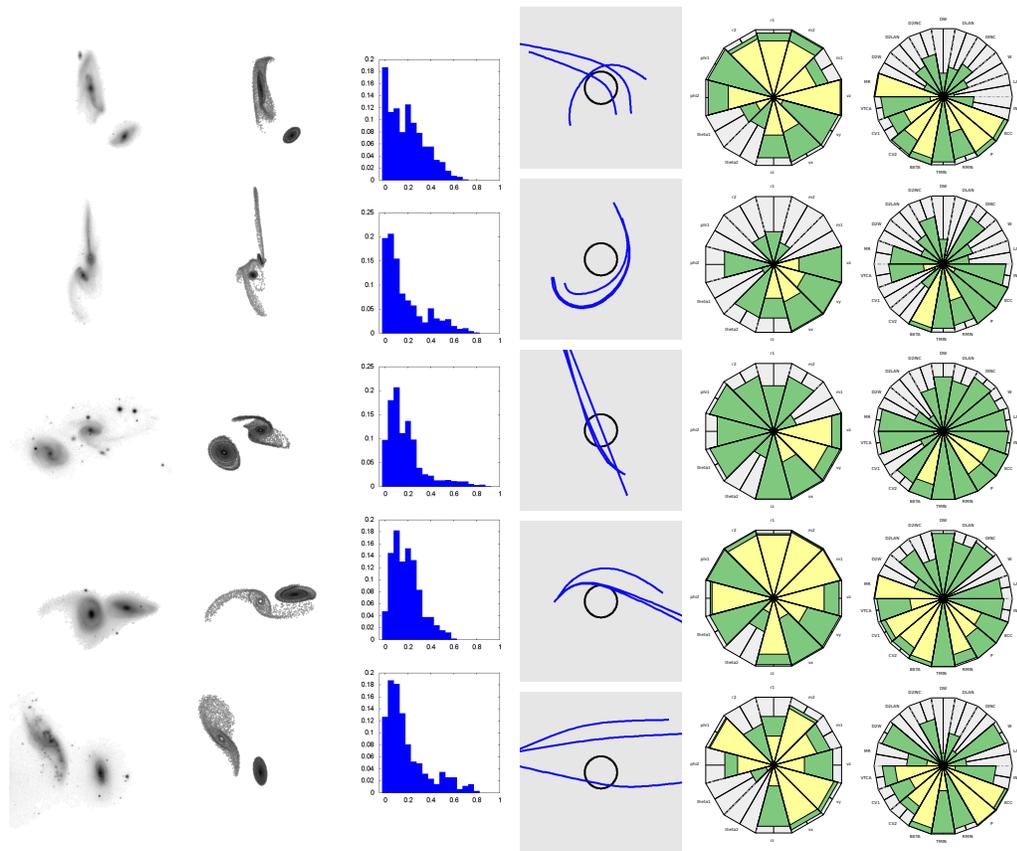


Figure 5.270: Morphological Group 9

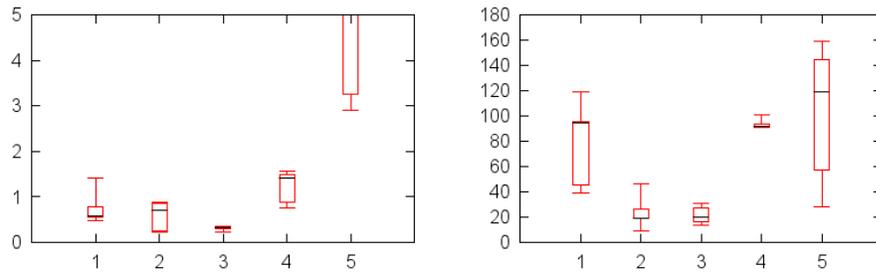


Figure 5.271: The distribution of eccentricity and inclination for the top simulations in group 9

### 5.2.10 Morphological Group 10

All galaxies have a high skewness value for their fitness distributions. This group has mostly closed, elliptical orbits with one hyperbolic orbit. The inclinations are in the range of  $40^\circ$  and  $70^\circ$ .

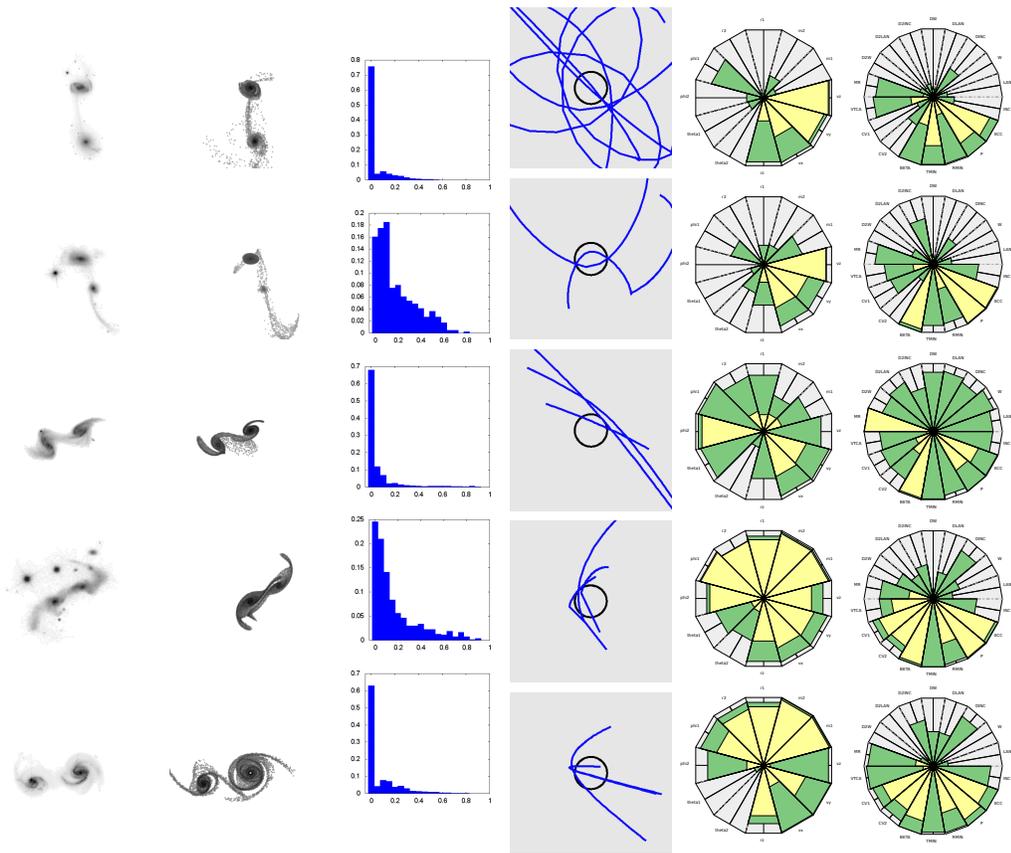


Figure 5.272: Morphological Group 10

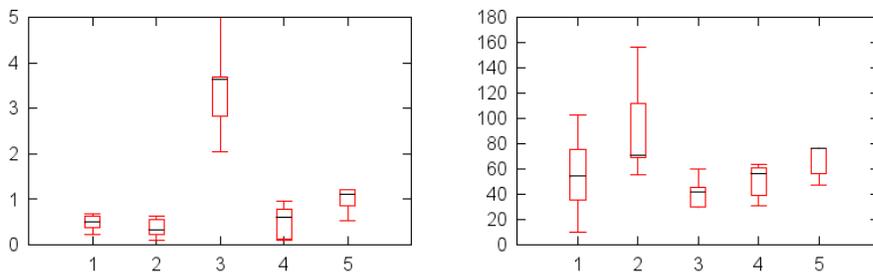


Figure 5.273: The distribution of eccentricity and inclination for the top simulations in group 10

### 5.2.11 Morphological Group 11

This is another group of high skewness galaxies with primarily parabolic orbits and most inclinations between  $50^\circ$  and  $90^\circ$ .

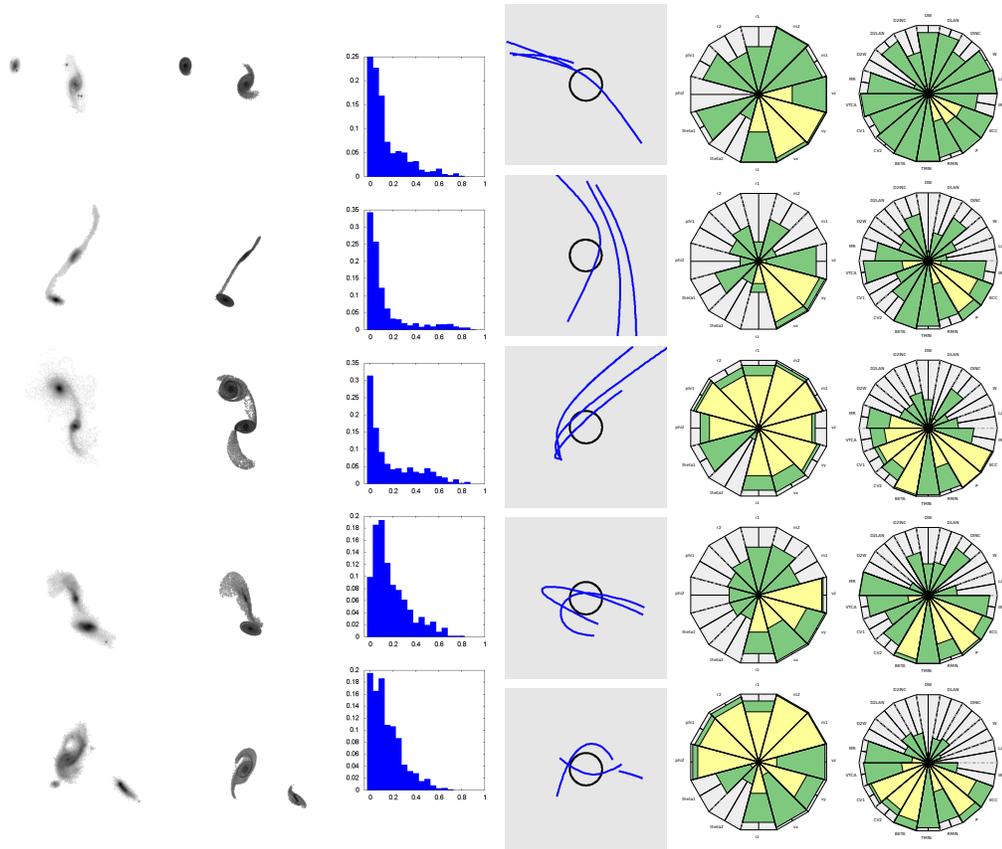


Figure 5.274: Morphological Group 11

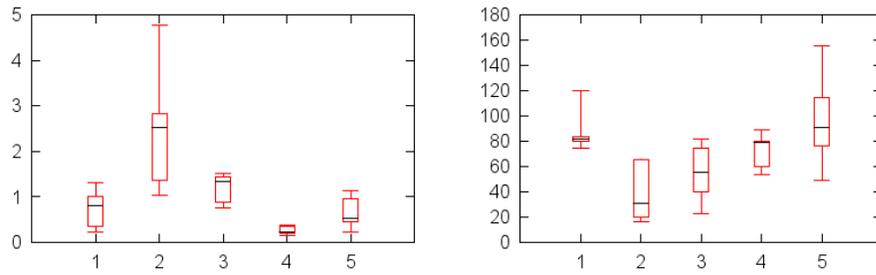


Figure 5.275: The distribution of eccentricity and inclination for the top simulations in group 11

### 5.2.12 Morphological Group 12

This last group consists of pairs with short tidal features and galaxies that are relatively far apart compared to their sizes. The skewness values for the fitness distributions are low and the IQR for inclinations are high for most of the pairs. The orbits tend to be parabolic to hyperbolic which is consistent with widely spaced galaxies. The inclinations are not well converged an range from  $0^\circ$  and  $180^\circ$ .

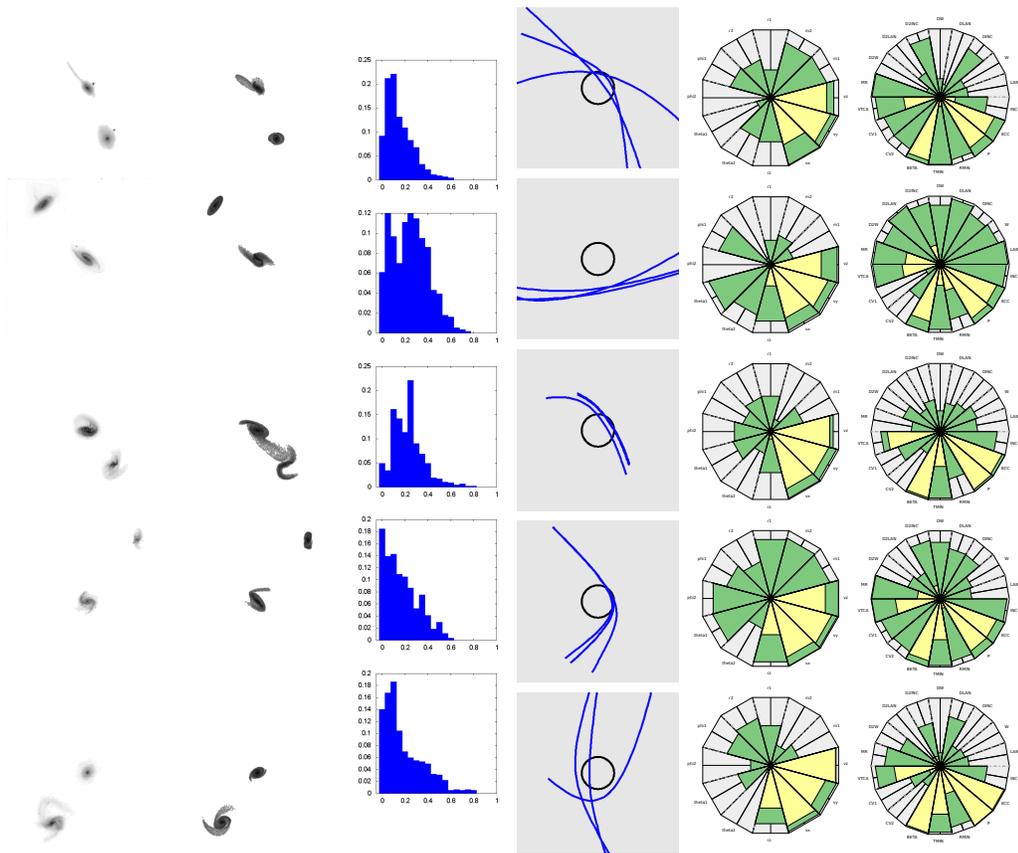


Figure 5.276: Morphological Group 12

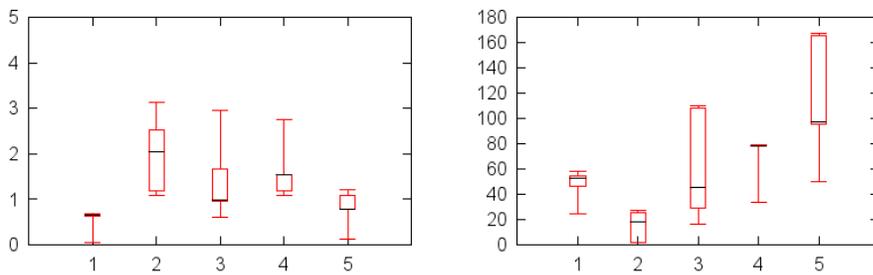


Figure 5.277: The distribution of eccentricity and inclination for the top simulations in group 12

These morphological groups were selected by an agglomerative hierarchical clustering algorithm using Zernike moments to calculate separation in feature space. The link between morphology and orbit parameters is demonstrated by most groups having similar values for eccentricities and inclinations.

### 5.3 Coverage of Parameter Space

The discussion above focuses on how well the pipeline narrowed the total range of parameter values. This section presents a brief summary of results in order to demonstrate that a wide range of parameter space was covered for each target. Table 5.63 presents the observed values and simulation parameters for each primary and secondary galaxy for all 62 target pairs. The columns include the star formation rate derived from the measured W3 magnitude and the stellar mass derived from the measured W1 magnitude. The rest of the columns include simulation parameters such as mass, disk radius, and orbit inclination with respect to the plane of the disk. For each of the simulation parameters, we have included the value from the best-fit simulation as well as the minimum and maximum values from all simulations presented to the Citizen Scientists. The simulation mass values were usually allowed to vary over three orders of magnitude. The disk radius usually ranged from approximately 1 kpc to tens of kiloparsecs. For most systems, the inclination with respect to the plane of the disk ranged over the entire  $0^\circ$  to  $180^\circ$  range.

A summary of simulation parameters referring to the pair, rather than the individual galaxies, is included in Table 5.64. The mass ratios, minimum distance, time since closest approach, orbital eccentricity, and  $\beta$  are included. Like in Table 5.63, this table includes the value from the best-fit simulation as well as the minimum and maximum values from all simulations presented to the Citizen Scientists. The mass ratios cover a large range of values due to the large range in individual mass values that were sampled. The minimum separation distance varies approximately two orders of magnitude. The time since closest approach ranged from 0, the current epoch, to over 8 Gyr. However, the oldest best-fit  $T_{min}$  is less than 0.57 Gyr. Orbit eccentricities range from circular, 0, through parabolic, 1, and

very hyperbolic, greater than 1000. The  $\beta$  parameter had a wide range from a minimum less than 0.001 to a maximum value of over 50000 for the most extreme case.

For each of the twelve simulation parameters we can compute the remaining fraction of parameter space. This is done by computing the range of parameter values for the top expert states and then dividing by the range of parameter values for all states shown to the Citizen Scientists. This ratio indicates how much of parameter space remains, and by inversion, how much of parameter space was ruled out for producing a good match to the target image. The distribution of remaining fraction values are shown in Table 5.65. Overall, the components of the relative velocity vector were best constrained, along with the z-component of the relative position vector. This means the process places tight constraints on the path through space that the secondary galaxy travels with respect to the primary. The disk radii are the next best constrained along with the disk inclination angle. The two masses typically have remaining fractions near 0.3. Eliminating 70% of a parameter that spans two orders of magnitude is a significant reduction. The least well constrained values are the two position angles,  $\theta$ .

### 5.3.1 Distribution of Simulation Parameters

Figure 5.278 contains seven histograms and one scatter plot. The histograms show the distribution of some important simulation parameters across all pairs. The values shown are for the best-fit value for each pair. The masses range over more than three orders of magnitude. Most of the projected separations are less than 40 kpc. The  $R_{min}$  values tend to be smaller by a factor of two. The  $T_{min}$  values range from 0 to 570 Myr. The  $\beta$  values range from 0.01 to 100. Most of the eccentricities are between 0 and 2 including a mix of elliptical and nearly parabolic orbits. However a number of orbits are very hyperbolic with eccentricities much greater than 1. The inclination with respect to the sky covers most of the full range. There is a peak in the distribution near  $90^\circ$ . The scatter plot in the bottom right of Figure 5.278 offers a potential explanation. The orbits with very high eccentricities tend to have inclinations nearly perpendicular to the plane of the sky. This is a result of our

selection bias for picking galaxies that have relatively small projected separation distances. A hyperbolic passage would need to be nearly perpendicular to the plane of the sky for us to have a high probability of perceiving the galaxies as still being close to each other.

### 5.3.2 $\beta$ Parameter

The  $\beta$  filter described in Section 4.1.1 removed from view of the Citizen Scientists approximately 93% of the simulations with a  $\beta$  value less than 0.5. Of the 62 best-fit simulations, 17 had  $\beta$  values less than the threshold of 0.5. It could be argued that the filter was too aggressive in throwing out “boring” simulations if more than one quarter of our best-fit parameter sets would have been eliminated. However, this was the justification for putting in the probabilistic exception to the filter. That exception means that rather than not cover the volume of parameter space with low  $\beta$ , we simply cover it less densely than we otherwise would without a filter. In Figure 5.279 the three galaxy pairs with the lowest best-fit  $\beta$  are plotted along with the the three highest. The low  $\beta$  pairs are largely unperturbed disks. These simulations are not well constrained by the morphology alone and it is likely that low  $\beta$  simulations, indicating a weak interaction so far, would reproduce the disks. The high  $\beta$  simulations show strong tidal features, either by the inclusion of long thin tidal tails, or a wide debris field as in the middle image in the bottom row.

**Table 5.63:** Simulation results for primary and secondary galaxies comparing observed SFR and mass with simulation mass, disk radius, and orbit inclination relative to the disk. Simulation parameters include best value and the min and max values.

Target	Galaxy	W3 - SFR (log M <sub>⊙</sub> yr <sup>-1</sup> )	W1 - Mass (log M <sub>⊙</sub> )	Sim. Mass (log M <sub>⊙</sub> )	Min Sim. Mass (log M <sub>⊙</sub> )	Max Sim. Mass (log M <sub>⊙</sub> )	R <sub>disk</sub> (kpc)	Min R <sub>disk</sub> (kpc)	Max R <sub>disk</sub> (kpc)	i <sub>orb</sub> (deg)	Min i <sub>orb</sub> (deg)	Max i <sub>orb</sub> (deg)
58772298443551614	pri	0.899 ± 0.386	11.214 ± 0.195	12.537	12.479	12.593	44.687	1.818	17.077	60.268	49.375	79.407
58772298443551614	sec	0.752 ± 0.386	11.305 ± 0.195	12.470	12.499	12.499	63.751	1.560	95.152	86.190	75.288	178.788
587724242427137777	pri	-0.205 ± 0.387	10.723 ± 0.195	10.713	9.122	11.284	14.106	1.130	48.972	119.153	1.011	179.610
5877242427137777	sec	-0.937 ± 0.390	10.431 ± 0.196	10.559	8.780	11.485	6.582	1.200	38.404	114.84	0.300	179.404
587726033843585146	pri	0.657 ± 0.386	11.272 ± 0.195	11.040	10.211	11.503	19.164	5.228	23.587	88.892	0.825	178.982
587726033843585146	sec	-0.427 ± 0.390	10.948 ± 0.196	11.188	10.302	11.204	5.014	3.513	13.849	54.434	0.903	179.360
58772717926508595	pri	0.567 ± 0.386	11.119 ± 0.194	10.565	9.121	12.560	9.870	1.130	70.951	91.308	0.316	178.646
58772717926508595	sec	-0.358 ± 0.387	10.830 ± 0.195	11.011	8.792	12.498	9.209	1.072	75.115	124.526	0.434	179.492
58772717926508595	sec	-0.427 ± 0.390	10.948 ± 0.196	11.115	10.301	11.427	14.111	3.349	28.701	83.272	0.116	178.940
58772717926508595	sec	0.990 ± 0.386	11.011 ± 0.197	11.100	10.391	11.280	8.156	3.023	13.439	148.756	0.328	179.236
58772722471131318	pri	0.335 ± 0.387	10.876 ± 0.198	10.283	9.602	12.563	12.222	4.757	71.018	35.522	1.927	177.934
58772722471131318	sec	-0.609 ± 0.393	10.804 ± 0.198	10.192	9.300	12.459	11.336	1.810	75.174	16.750	0.648	178.680
587728676861051075	pri	0.300 ± 0.386	10.731 ± 0.194	10.942	9.627	11.174	1.440	1.132	13.848	75.842	0.526	179.458
587728676861051075	sec	-0.671 ± 0.387	10.490 ± 0.195	10.676	9.574	10.750	0.884	0.641	15.300	135.593	0.739	179.376
587729227151704160	pri	-0.011 ± 0.390	10.847 ± 0.201	10.894	9.267	11.255	4.581	1.130	48.583	56.139	1.274	176.349
587729227151704160	sec	-0.011 ± 0.390	10.847 ± 0.201	10.992	9.178	11.568	6.179	1.060	37.755	130.053	2.018	179.624
58773191311050988	pri	0.007 ± 0.386	10.295 ± 0.195	9.204	9.204	10.584	5.009	4.104	7.136	18.273	0.174	179.389
58773191311050988	sec	0.265 ± 0.386	10.463 ± 0.195	10.505	9.079	10.526	5.706	4.130	6.486	96.195	3.396	177.324
58773213099882121	pri	-0.616 ± 0.386	10.385 ± 0.194	10.070	9.353	10.887	8.015	5.299	17.664	81.345	10.264	163.829
58773213099882121	sec	-0.481 ± 0.386	9.948 ± 0.194	10.009	9.578	11.368	6.312	2.355	7.532	80.180	20.127	179.684
587732727130652231	pri	-0.046 ± 0.388	10.610 ± 0.199	10.898	9.358	11.280	7.073	1.130	48.896	86.691	0.084	179.045
587732727130652231	sec	0.031 ± 0.388	10.704 ± 0.198	11.048	9.057	11.569	2.573	1.057	38.236	124.826	0.540	179.628
58773308014583863	pri	-0.693 ± 0.389	9.477 ± 0.206	10.314	8.735	10.910	4.187	3.703	10.292	79.241	19.436	179.436
58773308014583863	sec	0.064 ± 0.387	10.256 ± 0.198	11.065	9.087	11.546	6.031	2.881	7.525	144.864	0.338	179.291
587734862806752822	pri	0.344 ± 0.388	11.038 ± 0.199	11.363	10.311	11.503	15.292	9.294	23.050	49.500	0.688	178.858
587734862806752822	sec	0.421 ± 0.387	10.562 ± 0.203	10.642	10.100	10.819	11.735	6.257	15.574	27.872	0.296	179.728
58773504309829845	pri	0.335 ± 0.386	10.698 ± 0.195	11.172	9.130	11.526	5.901	1.821	14.419	11.270	2.252	179.964
58773504309829845	sec	-0.255 ± 0.387	10.007 ± 0.198	10.874	8.885	11.371	2.984	1.577	13.916	80.180	0.735	179.491
587735665840881790	pri	0.570 ± 0.386	11.220 ± 0.195	10.968	10.470	11.223	7.449	5.420	7.644	49.338	0.581	179.738
587735665840881790	sec	0.000 ± 0.388	10.594 ± 0.198	10.272	10.149	10.985	2.214	1.146	2.700	122.185	0.517	179.330
587736941981466667	pri	-0.815 ± 0.388	10.135 ± 0.196	10.246	10.029	11.255	4.546	1.815	14.298	43.800	1.609	179.116
587736941981466667	sec	-0.279 ± 0.386	10.347 ± 0.196	10.653	9.665	11.000	6.629	3.669	13.323	113.226	0.909	179.349
587738569246376675	pri	0.761 ± 0.387	11.387 ± 0.199	11.149	10.605	11.423	17.700	30.346	84.466	3.674	176.850	179.330
587738569246376675	sec	0.322 ± 0.389	10.749 ± 0.206	10.990	10.387	11.281	9.771	2.937	13.448	83.263	4.576	175.886
587738569246376675	sec	-0.165 ± 0.400	10.674 ± 0.213	10.961	9.621	12.562	11.017	1.828	70.763	82.243	1.952	178.886
587738569246376675	sec	-0.381 ± 0.400	10.497 ± 0.197	10.857	9.819	12.408	16.629	2.355	73.154	120.800	1.688	178.688
587739153536995531	pri	0.102 ± 0.387	10.405 ± 0.198	10.870	10.241	11.223	7.173	3.327	19.074	59.088	2.275	177.846
587739153536995531	sec	-0.023 ± 0.387	10.128 ± 0.199	10.555	10.308	11.234	3.330	1.654	12.301	109.910	8.427	163.535
587739153536995531	sec	0.327 ± 0.387	10.507 ± 0.202	10.199	9.671	10.847	3.721	2.918	6.236	49.638	5.482	160.447
587739407868690486	pri	0.714 ± 0.387	10.912 ± 0.198	10.369	9.391	10.567	5.169	1.390	6.692	128.969	4.272	175.802
587739407868690486	sec	0.330 ± 0.386	10.650 ± 0.195	10.712	10.192	11.123	4.642	4.501	12.055	159.444	0.498	178.381
58773956541578866	pri	0.140 ± 0.386	10.398 ± 0.195	10.436	9.999	11.175	11.104	5.702	17.768	128.896	1.524	178.656
58773956541578866	sec	-0.448 ± 0.386	10.197 ± 0.197	10.655	9.847	10.685	4.691	2.865	45.602	83.363	0.340	179.372
58773964743412797	sec	-0.475 ± 0.393	11.012 ± 0.198	10.856	9.837	12.144	6.551	1.595	47.282	114.965	0.328	179.290
58773964743412797	sec	-0.005 ± 0.386	10.488 ± 0.195	10.633	10.280	11.452	8.691	5.225	18.275	45.263	1.066	179.498
58773964743412797	sec	-0.038 ± 0.386	10.588 ± 0.195	10.821	10.390	11.171	8.198	3.512	13.933	110.188	0.363	179.627
587739707420967061	pri	-0.255 ± 0.394	10.143 ± 0.196	11.202	10.470	11.410	8.806	1.148	10.696	51.055	0.300	179.491
587739707420967061	sec	-0.240 ± 0.390	11.012 ± 0.198	11.115	10.153	11.568	8.086	1.083	9.452	64.253	0.651	179.389
587739721376202860	pri	1.167 ± 0.386	11.621 ± 0.195	10.750	9.811	10.986	14.163	10.088	18.626	138.654	0.303	179.845
587739721376202860	sec	0.419 ± 0.387	10.675 ± 0.200	10.320	9.487	10.663	9.522	3.978	9.944	72.933	0.744	178.881
587739721376202860	sec	-0.538 ± 0.386	11.077 ± 0.195	10.519	10.452	11.267	10.655	5.499	14.508	19.476	0.735	179.491
587739721376202860	sec	-0.108 ± 0.388	10.894 ± 0.196	10.743	10.431	11.217	5.858	4.908	9.906	85.375	0.383	179.678
587739810496708646	pri	0.252 ± 0.386	10.462 ± 0.195	11.119	9.819	11.421	9.184	5.756	12.758	67.939	1.411	179.123
587739810496708646	sec	-1.097 ± 0.390	9.137 ± 0.205	9.979	9.306	11.281	2.587	1.761	12.826	141.753	0.744	179.699
587739810496708646	sec	-0.056 ± 0.387	10.636 ± 0.196	10.714	9.835	10.812	6.597	3.466	8.663	54.694	1.176	179.391
587739810496708646	sec	-0.875 ± 0.391	10.344 ± 0.197	10.777	9.852	10.850	4.452	3.405	8.819	96.761	1.436	178.578
587741391565422775	pri	0.540 ± 0.387	11.061 ± 0.198	11.325	9.940	11.339	19.353	3.403	23.954	16.813	0.436	179.403
587741391565422775	sec	-0.381 ± 0.393	9.731 ± 0.217	10.551	10.330	11.201	2.848	2.249	9.979	71.923	1.439	179.200
587741391565422775	sec	0.796 ± 0.386	11.084 ± 0.197	10.775	9.240	10.270	9.231	2.966	16.958	30.661	0.481	179.514
587741391565422775	sec	-0.093 ± 0.388	11.212 ± 0.196	11.450	9.221	11.569	11.876	1.058	38.307	97.662	1.184	179.505
587741534400217110	pri	0.319 ± 0.387	10.494 ± 0.199	9.983	9.778	11.423	3.136	4.694	16.516	42.937	0.810	179.627
587741534400217110	sec	-0.859 ± 0.394	10.494 ± 0.199	10.090	9.310	11.227	3.621	1.807	12.903	111.965	0.611	179.292
587741620300282825	pri	0.179 ± 0.386	10.847 ± 0.195	10.529	9.633	10.645	10.505	6.467	8.466	78.609	0.764	177.616
587741620300282825	sec	-0.642 ± 0.388	9.729 ± 0.199	10.271	9.201	10.377	7.027	5.076	12.072	77.726	1.913	178.647
587741620300282825	sec	-0.310 ± 0.388	10.929 ± 0.196	11.125	10.699	11.220	8.976	5.085	10.140	107.297	0.738	179.889
58774172819493915	pri	0.067 ± 0.387	10.382 ± 0.199	10.300	10.066	10.625	2.944	1.880	4.701	128.344	0.996	179.213
58774172819493915	sec	-0.422 ± 0.391	11.154 ± 0.196	10.751	9.015	10.178	13.780	2.514	15.780	108.392	0.758	179.504
58774172819493915	sec	-0.521 ± 0.392	10.986 ± 0.197	10.364	9.900	11.380	7.522	4.195	12.485	147.982	0.946	179.239
587741829658181998	pri	0.370 ± 0.386	10.581 ± 0.197	11.055	9.115	11.409	10.237	1.130	48.995	68.150	0.583	179.406
587741829658181998	sec	0.377 ± 0.386	10.796 ± 0.196	11.175	8.825	11.567	16.684	0.771	38.397	36.383	0.249	179.575
58774201058941189	pri	-0.738 ± 0.387	10.575 ± 0.194	10.709	9.016	10.803	12.049	8.157	15.319	50.688	0.980	179.688
58774201058941189	sec	-1.017 ± 0.388	9.135 ± 0.200	10.491	9.479	10.655	4.657	2.069	4.896	97.860	1.148	179.311
58774201058941189	sec	0.337 ± 0.387	10.776 ± 0.197	10.164	9.944</							

Table 5.64: Simulation results for each target include the mass ratio, minimum distance, time since closest approach, orbital eccentricity, and  $\beta$  parameters. The best value, min and max values are included.

Target	Best MR	Min Mr	Max MR	$R_{min}$ (kpc)	Min $R_{min}$ (kpc)	Max $R_{min}$ (kpc)	$T_{min}$ (Myr)	Min $T_{min}$ (Myr)	Max $T_{min}$ (Myr)	$ecc$	Min $ecc$	Max $ecc$	$\beta$	Min $\beta$	Max $\beta$
587722984435351614	1.167	0.307	889.801	81.443	0.402	409.274	249.449	0.000	8602.542	3.700	0.001	677.694	0.291	0.001	3920.725
587724234257137777	1.427	0.020	272.820	18.813	0.128	199.207	479.844	0.000	8612.936	0.790	0.001	3080.303	0.499	0.001	4361.565
587726033843585146	0.712	0.147	15.843	8.947	0.105	142.032	77.953	0.000	8657.975	0.442	0.001	168.073	2.937	0.009	7040.448
587727177926508595	0.358	0.018	217.774	15.036	0.190	383.156	169.764	0.000	8656.243	1.579	0.001	838.669	0.732	0.002	2763.580
587727178988388373	1.036	0.372	5.458	17.250	0.102	231.411	122.992	0.000	8657.975	1.628	0.002	400.671	0.801	0.002	12355.246
587727222471131318	1.234	0.185	8.147	16.299	0.207	379.329	502.363	0.000	8628.526	0.991	0.012	2272.374	0.369	0.000	606.179
587728676861051075	1.845	0.076	16.571	1.069	0.057	148.908	67.559	0.000	8635.455	0.640	0.001	464.457	101.600	0.002	31287.255
587729227151704160	0.798	0.074	29.031	1.527	0.075	141.077	65.827	0.000	8628.526	0.657	0.007	483.573	32.806	0.003	14902.455
587731913110650988	0.050	0.050	29.772	18.315	0.285	83.746	242.520	0.000	8614.668	2.533	0.002	841.036	0.228	0.003	496.124
587732136993882121	1.151	0.183	3.646	9.324	0.170	152.956	495.434	0.000	8635.455	1.007	0.004	3181.084	0.675	0.001	833.729
587732772130652231	0.709	0.060	50.696	4.117	0.077	155.099	133.386	0.000	8651.046	0.365	0.002	1180.444	12.037	0.002	10825.800
587733080814583863	0.177	0.004	79.719	8.363	0.112	140.481	102.205	0.000	8652.778	0.218	0.001	176.206	2.548	0.006	6797.398
587734862680732822	5.266	1.578	15.792	20.493	0.188	150.749	247.717	0.000	8657.975	0.375	0.002	309.546	0.885	0.005	3510.090
587735043609329845	1.986	0.021	416.918	6.374	0.047	151.156	131.654	0.000	8657.975	0.155	0.001	931.630	5.021	0.004	47950.387
587735665840881790	4.957	0.314	11.597	4.592	0.138	129.854	131.654	0.000	8652.778	0.119	0.000	118.606	5.899	0.009	3637.425
587736941981466667	1.569	0.177	6.620	12.080	0.140	141.443	3.465	0.000	8595.613	7.365	0.005	244.303	0.257	0.002	2238.622
587738569246376675	1.441	0.382	5.886	14.738	0.648	231.592	289.292	0.000	8654.510	4.378	0.008	182.603	0.683	0.002	337.144
587738569249300718	0.395	0.204	6.945	24.142	0.223	325.138	263.308	0.000	8657.975	0.748	0.001	394.143	0.665	0.002	2593.574
587739153356095531	2.069	0.154	7.558	6.605	0.147	95.170	320.473	0.000	8657.975	0.413	0.001	780.897	3.025	0.002	3912.572
587739407868690486	0.677	0.130	28.202	7.389	0.064	150.308	329.135	0.000	8637.188	0.219	0.001	328.179	2.049	0.002	17007.528
587739505541578866	1.887	0.150	6.748	17.512	0.226	127.950	109.134	0.000	8602.542	6.253	0.002	1065.720	0.261	0.001	2176.819
587739646743412797	1.353	0.167	21.220	6.741	0.203	201.900	420.946	0.000	8657.975	0.400	0.001	732.706	3.651	0.001	2552.681
587739647284805725	0.648	0.146	4.625	62.351	0.231	142.814	115.928	0.000	8656.243	1.491	0.001	325.053	0.078	0.003	1459.349
587739707420967061	1.400	0.103	10.923	2.146	0.111	148.919	159.506	0.000	8656.243	0.523	0.001	151.476	37.757	0.013	8839.703
587739721376202860	2.692	0.143	30.496	68.384	0.098	141.524	100.473	0.000	8649.314	28.310	0.002	668.570	0.017	0.002	8194.417
587739721900163101	0.597	0.174	6.803	15.990	0.081	150.089	429.607	0.000	8657.975	0.702	0.001	198.824	0.655	0.005	12756.872
587739810496708646	1.379	0.599	10.777	14.658	0.213	157.070	162.835	0.000	8652.778	4.699	0.002	5138.944	0.207	0.000	736.442
587739845393580192	0.864	0.097	9.110	4.873	0.137	71.771	214.804	0.000	8642.384	0.636	0.014	680.411	4.460	0.003	3324.229
587741391565422775	5.944	0.088	9.070	21.807	0.245	148.497	256.379	0.000	8657.975	0.580	0.001	304.237	0.713	0.004	1843.127
587741532784361481	0.211	0.019	393.639	21.427	0.069	120.486	278.898	0.000	8652.778	1.784	0.002	804.968	0.649	0.002	8977.519
587741534400217110	0.781	0.415	9.752	17.261	0.085	146.805	325.670	0.000	8656.243	0.824	0.001	1434.159	0.281	0.002	3633.394
587741602030026825	1.809	0.194	27.380	12.595	0.164	91.844	472.915	0.000	6423.324	1.549	0.006	1323.764	0.590	0.001	1759.599
5877417228194393915	6.686	1.194	14.136	7.088	0.083	147.802	124.725	0.000	8657.975	0.626	0.000	179.424	2.988	0.004	11209.197
587741817851674654	2.436	0.035	6.708	8.734	0.150	140.689	91.811	0.000	8657.975	2.821	0.014	383.936	1.026	0.002	3378.716
587741829658181698	0.759	0.016	284.456	16.455	0.092	197.674	65.827	0.000	8657.975	1.301	0.000	882.459	0.930	0.002	11681.497
587742010583941189	1.655	0.185	21.114	25.408	0.153	147.430	452.127	0.000	8657.975	0.987	0.001	1136.358	0.292	0.001	2150.949
587742014353702970	0.868	0.136	19.277	11.147	0.122	138.027	329.135	0.000	8538.447	1.806	0.003	2021.211	0.522	0.001	2485.946
587742571610243080	0.602	0.053	8.506	9.236	0.160	230.897	147.244	0.000	8657.975	1.295	0.000	126.375	1.916	0.003	4413.198
587745402001817662	2.413	0.214	6.025	3.621	0.172	137.550	103.937	0.000	8656.243	0.386	0.001	241.766	13.322	0.004	3083.534
587746029596311590	0.563	0.109	6.013	2.169	0.065	148.946	20.787	0.000	8657.975	0.336	0.001	109.481	24.801	0.002	21124.506
588011241146422756	1.098	0.197	5.019	13.540	0.174	275.404	58.898	0.000	8657.975	0.106	0.001	192.612	1.900	0.003	5201.787
588013383816904792	2.099	0.197	4.140	4.953	0.110	263.861	50.236	0.000	8657.975	0.471	0.001	596.985	5.908	0.002	11168.961
588017604696408086	2.322	0.068	77.628	18.567	0.069	146.923	562.994	0.000	8656.243	0.587	0.001	937.365	0.576	0.002	14854.786
588017604696408195	1.761	0.082	18.461	8.824	0.103	132.723	337.796	0.000	8657.975	0.794	0.000	2069.538	1.374	0.001	5754.884
588017702948962343	30.202	0.138	1075.938	4.775	0.223	98.984	174.961	0.000	8637.188	0.341	0.000	171.560	16.792	0.007	4792.330
588017978901528612	4.089	0.037	271.690	4.606	0.078	133.241	230.394	0.000	8657.975	0.054	0.004	2237.097	4.418	0.001	10030.263
588018055130710322	0.945	0.910	6.015	13.918	0.067	143.050	452.127	0.000	8654.510	0.604	0.000	345.020	0.788	0.003	19540.015
758874299603222717	0.950	0.007	149.897	5.867	0.063	149.720	58.898	0.000	8657.975	0.050	0.000	171.140	5.684	0.002	21353.400
Arp 148	0.388	0.274	19.651	8.497	0.123	146.496	38.110	0.000	8495.140	12.908	0.001	597.817	0.239	0.001	3202.354
CGCG 436-030	3.722	0.323	11.068	19.466	0.134	95.978	263.308	0.000	8647.581	0.977	0.011	835.493	0.426	0.002	2848.546
Arp 272	1.838	0.140	18.966	8.654	0.180	151.840	363.780	0.000	8543.644	0.810	0.003	913.581	1.002	0.001	1273.594
ESO 77-14	1.311	0.112	21.132	3.018	0.026	149.208	93.544	0.000	8656.243	0.782	0.002	80.297	15.930	0.002	51774.985
NGC 5331	1.608	0.259	7.622	7.681	0.220	173.170	119.528	0.000	8590.416	1.998	0.010	1432.837	1.202	0.001	971.091
NGC 6786	2.438	0.039	25.521	13.392	0.064	346.368	511.025	0.000	8657.975	1.108	0.001	980.623	0.763	0.003	18724.424
587736523764334706	3.239	0.067	14.916	59.210	0.312	142.069	0.000	0.000	8638.920	16.797	0.007	658.300	0.025	0.003	773.381
587747120521216156	0.833	0.049	1.376	2.776	0.180	146.276	102.205	0.000	8657.975	0.233	0.003	82.315	14.192	0.004	2833.376
588007005230530750	2.979	0.063	13.720	12.117	0.153	138.341	230.394	0.000	8657.975	0.694	0.000	1996.395	0.792	0.001	2912.730
758877153600208945	1.289	0.016	319.669	19.671	0.045	212.174	271.969	0.000	8652.778	1.181	0.001	1842.282	0.468	0.001	47999.308
587739720308818095	4.121	0.149	43.927	6.002	0.187	328.818	285.828	0.000	8657.975	0.457	0.001	1183.856	2.704	0.001	1910.213
Arp 273	2.004	0.040	23.218	37.833	0.221	408.622	284.095	0.000	8652.778	4.036	0.007	1215.151	0.114	0.001	1590.700
1237678620102623480	0.643	0.128	1.807	25.472	0.262	148.960	159.370	0.000	8652.778	2.523	0.005	1369.829	0.240	0.001	1213.433
Arp 244	0.758	0.076	16.571	1.572	0.063	148.908	55.433	0.000	2929.299	0.493	0.001	464.457	36.021	0.002	18465.811

Table 5.65: The mean, median, min, and max remaining fraction for the twelve simulation parameters.

Name	Mean RF	Median RF	Best RF	Worst RF
$r_z$	0.12	0.09	0.0	0.54
$v_x$	0.04	0.02	0.003	0.36
$v_y$	0.05	0.02	0.005	0.54
$v_z$	0.06	0.03	0.0	0.44
$m_1$	0.33	0.30	0.0007	0.94
$m_2$	0.32	0.28	0.0	0.87
$r_1$	0.29	0.22	0.01	0.97
$r_2$	0.31	0.22	0.007	0.97
$\phi_1$	0.29	0.17	0.004	0.97
$\phi_2$	0.37	0.27	0.0	0.97
$\theta_1$	0.53	0.65	0.01	0.95
$\theta_2$	0.56	0.62	0.0	0.97

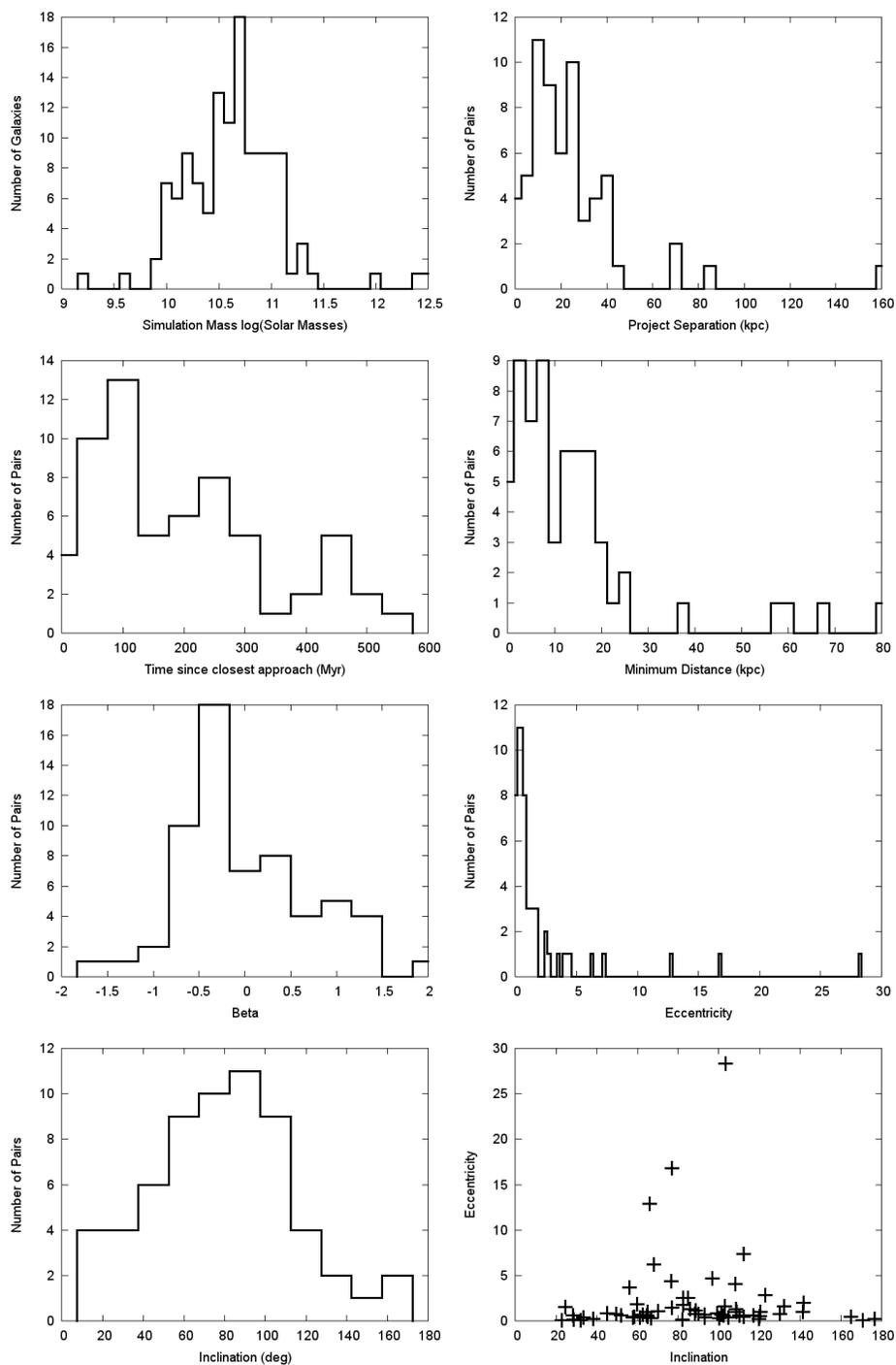


Figure 5.278: The distribution of simulation parameters including mass,  $T_{min}$ , projected separation,  $R_{min}$ ,  $\beta$ ,  $ecc$ ,  $inc$ , and  $ecc$  vs.  $inc$ . The bottom right plot shows that the highest eccentricities are nearly perpendicular to the plane of the sky.

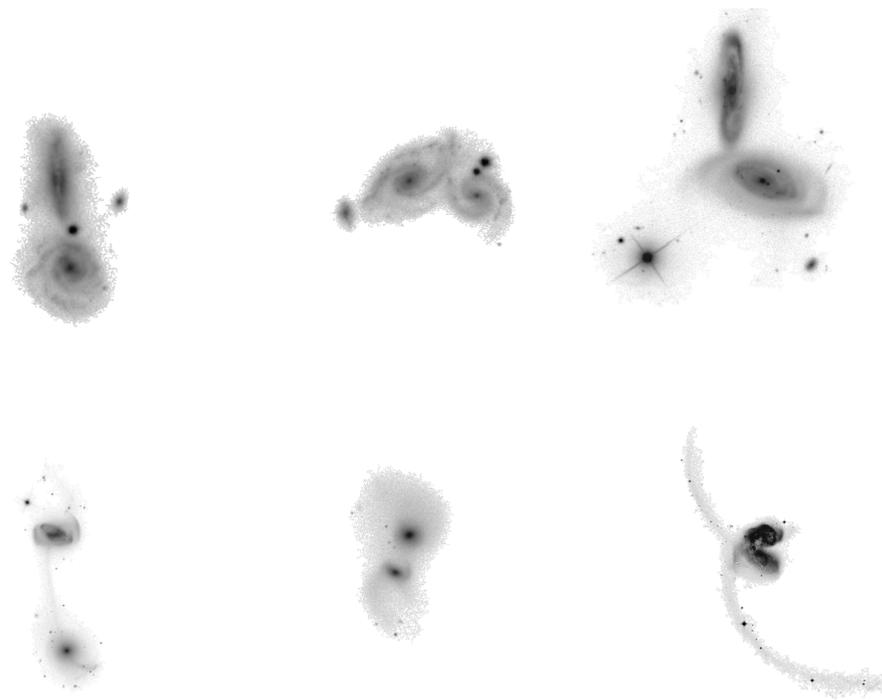


Figure 5.279: The galaxy pairs with the lowest  $\beta$  (top) and highest  $\beta$  bottom.

## Chapter 6: Comparing Simulation Values to Observations

In this Chapter we present comparisons between simulation values and physical quantities derived from photometric observations of the galaxies.

### 6.1 Stellar Mass

The WISE photometry provides us with estimates for stellar mass and star formation rate for each of our galaxies. We begin comparing our simulation values to observations with the galaxy masses. For the best-fit simulation, we have the final mass for the primary and secondary galaxies. We convert those values to physical units and plot them along with the photometric masses for each galaxy. Photometric uncertainties are due largely to the uncertainty in the linear regression between W1 magnitude and mass. The uncertainty for simulation values was estimated by sampling the values from the set of “top” fitness states 10000 times in accordance to their fitness values from the two finale activities. The variance in the sampled population is calculated in order to estimate the one  $\sigma$  errors for each simulation, and derived orbit, parameter.

Figure 6.1 shows that the two different approaches to estimating mass give similar results to within  $\sim 0.5$  dex. We have separated the SDSS galaxies, that had their initial simulation masses estimated with ugriz colors, from the HST galaxies, that had their masses estimated from a simple mass-to-light ratio from NED magnitude values.

The agreement for SDSS and HST galaxies seems to be of a similar quality. The simulation masses for HST galaxies may be slightly underestimated compared to the SDSS galaxies. Figure 6.1 confirms that our best-fit simulations converged to realistic values for the galaxy masses even though they were randomly selected from a set of values that varied over two

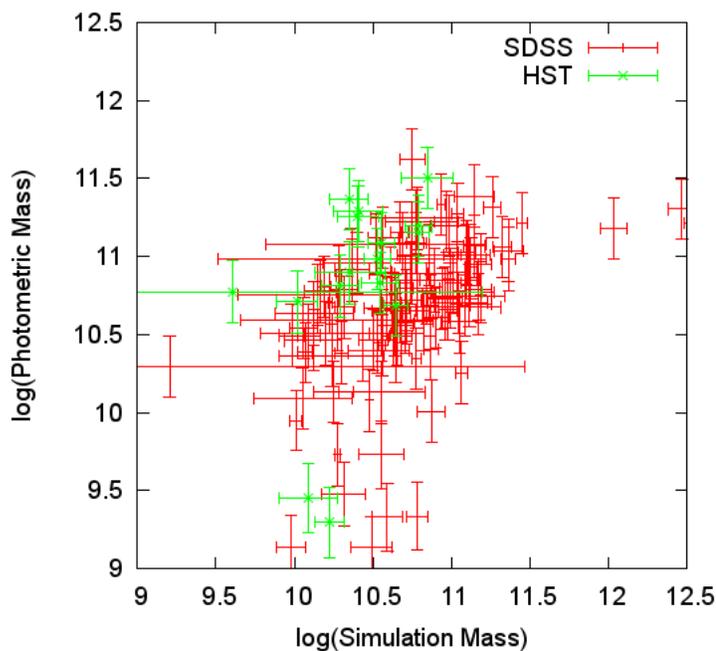


Figure 6.1: Photometric Mass vs. Simulation Mass

orders of magnitude. The photometric mass is due to the stellar population. The simulation mass is due to the total dynamical mass of each galaxy. This mass includes the dark matter halo. The agreement between the stellar population mass and our dynamical mass indicates that the simulation mass is probably too low. We would expect the total mass to be higher than just the stellar mass. This is another reason to consider incorporating observed velocity information into our pipeline to help constrain the dark matter mass in addition to the stellar mass.

The relationship between photometrically derived mass ratio and simulation mass ratio is shown in Figure 6.2. The correlation between photometric and simulation mass ratio seems consistent from mass ratios close to 1 up to 10. Most of the outliers have correspondingly large uncertainties in their simulation values. This overall agreement between the two values for mass ratio are consistent with the previous agreement between simulation mass and photometric mass.

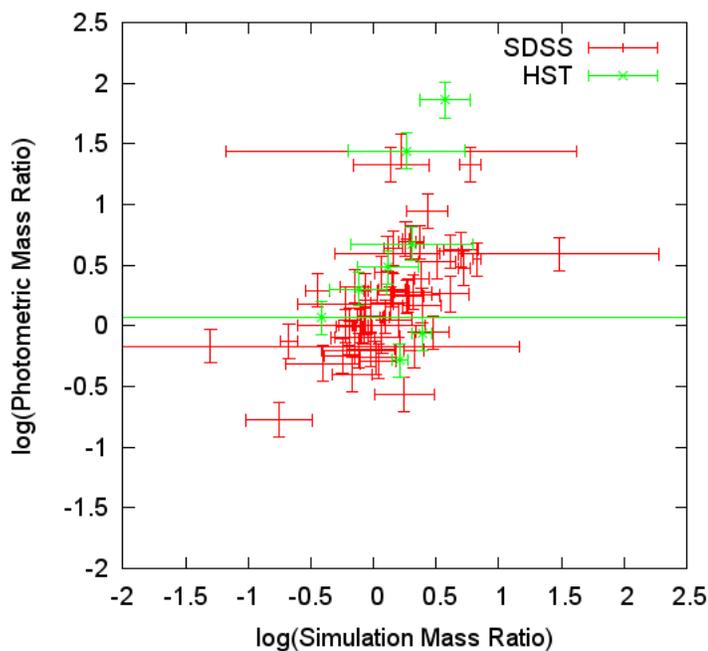


Figure 6.2: Photometric Mass Ratio vs. Simulation Mass Ratio

## 6.2 Star Formation Rate

We did not perform the density compression measurements from Wallin et al. (1990) for our restricted three-body models. We were also not able to advance the pipeline far enough to conduct a large number of full n-body simulations. So we do not have any simulation-based estimates for the star formation rate. However, we can attempt to verify an expected relationship between the photometrically observed star formation rate and time since the star formation was triggered. As an estimate of the time since starburst we select our simulation time since closest approach. Bruzual A. and Charlot (1993) propose an exponentially decreasing star formation rate with time. We adopt an expression of the same form as Kennicutt (1998) for our star formation rate history in Equation 6.1.

$$R(t) = R_0 \exp -\frac{t}{\tau} \tag{6.1}$$

We convert the time since closest approach,  $t_{min}$ , from simulation units to years and then compute the log value. We then plot the log of SFR as computed by our relationship with the W3 magnitude. The SFR as a function of time for all galaxies<sup>1</sup> is shown in Figure 6.3 along with error estimates. The exponential decay we expect from Bruzual A. and Charlot (1993) should appear as a decreasing linear trend on the log-log scale.

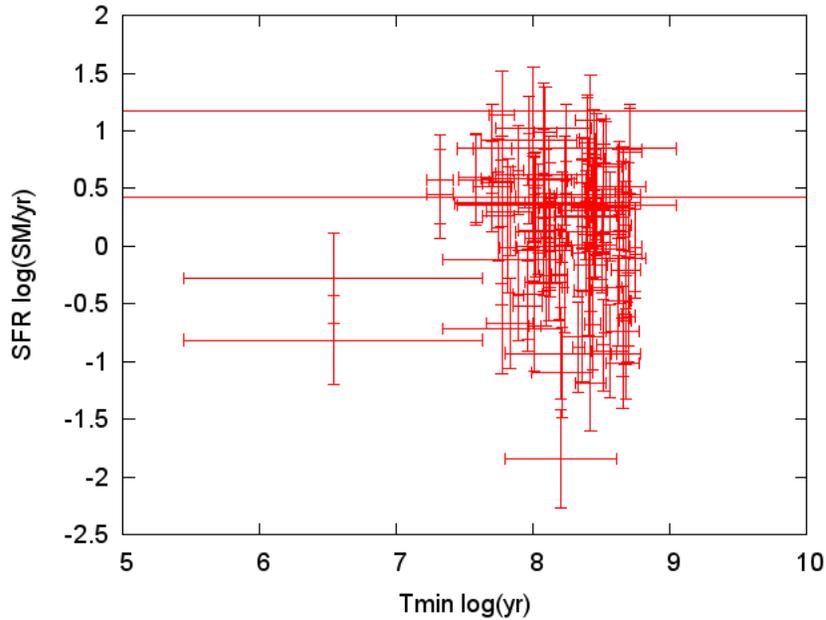


Figure 6.3: Star formation rate vs.  $t_{min}$  for all galaxies.

Any claim that Figure 6.3 presents an obvious trend between SFR and  $t_{min}$  is likely to be met with some skepticism. One possible refinement is to normalize the SFR so that it is per unit mass of the host galaxy. Figure 6.4 shows this normalized SFR. Again, no definitive relationship is present.

We applied to additional refinements. First we separated the galaxies into separate populations for the primary and secondary galaxies. Within each sub-population, we divided

<sup>1</sup>The values for SDSS 587736523764334706 were excluded because its  $t_{min}$  was so close to zero it stretches the log scale excessively to -18.

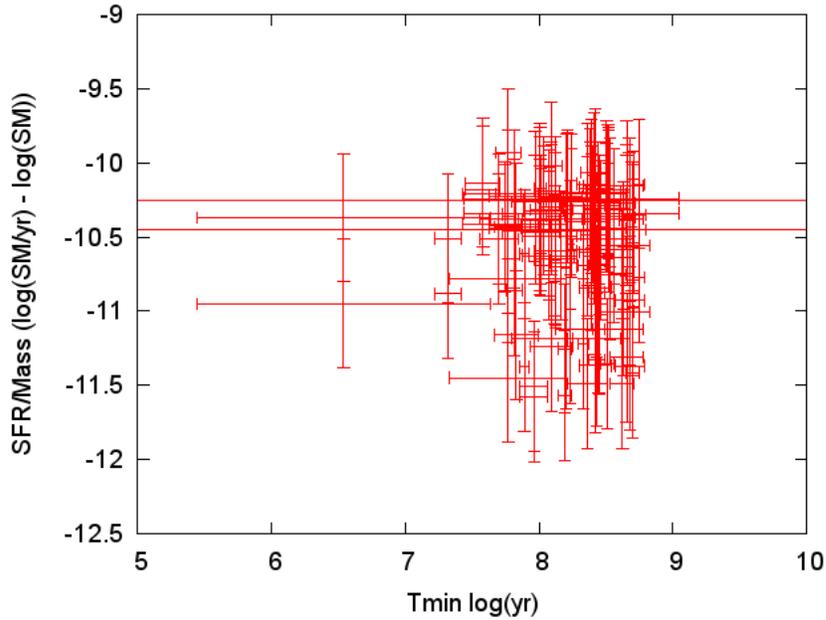


Figure 6.4: Normalized star formation rate vs.  $t_{min}$  for all galaxies.

the galaxies into three groups based upon mass ratio. The first group was for nearly equal masses with a mass ratio less than or equal to 1.5. The next group covered the mass ratios less than or equal to 3, the traditional limit for major mergers. The final group was for the minor mergers with mass ratio greater than 3. Figure 6.5 shows the SFR and normalized SFR for the primary galaxies. Figure 6.6 shows the same values for the secondary galaxies.

In general the SFRs for the secondary galaxies appear lower than for the primaries. This seems to hold even for the normalized SFR. As one last refinement, we plot the SFR for the primary galaxies but exclude the values for mass ratios less than or equal to 1.5. In Figure 6.7 we see a recognizable decline in SFR as  $t_{min}$  increases. We have plotted an exponential decay for comparison. The slope of the line was determined by simply linear regression (no weights). The characteristic timescale  $\tau$  for the decay expression that corresponds to the line is 0.81 Gyr.

### 6.2.1 Projected Separation Distance

When astronomers observe pairs of galaxies, especially mergers, one useful quantity that is possible to measure is the projected separation distance. This is simply the distance in the plane of the sky separating the two galaxies. Because redshift information is often lacking, or lacking sufficient precision and accuracy, it is usually not possible to determine the full three-dimensional distance between the two galaxies. Additionally, most populations of pairs are not modeled with simulations, so dynamical values are not available. This leaves the two-dimensional projected separation distance as an estimate of how close the galaxies are at present epoch as well as a proxy for time since closest approach. Many papers in the literature attempt to study the star formation rates of a population of galaxy pairs as a function of this distance. One recent example is Scudder et al. (2012) where they present evidence for decreasing star formation rate with increasing separation distance.

We present the projected separation distance and time since closest approach in the Figure 6.8. There does appear to be a correlation. This suggests that  $r_{sep}$  may be a good proxy for  $t_{min}$ .

Figure 6.9 shows the normalized star formation rate as a function of projected separation distance,  $r_{sep}$ . There does not appear to be any strong relationship. We would expect to see a similar trend of decreasing SFR for increasing  $r_{sep}$  similar to Scudder et al. (2012), but that is not the case. If we adopt a binning approach similar to those authors, we also do not see a similar relationship. In contrast to the expected decreasing relationship, Figure 6.10 seems to show an increasing relationship. The population size for each bin is small. The bin near 70 kpc only has two galaxies. Such a small population is not likely to yield as useful of a result as the that displayed in Figure 3 of Scudder et al. (2012).

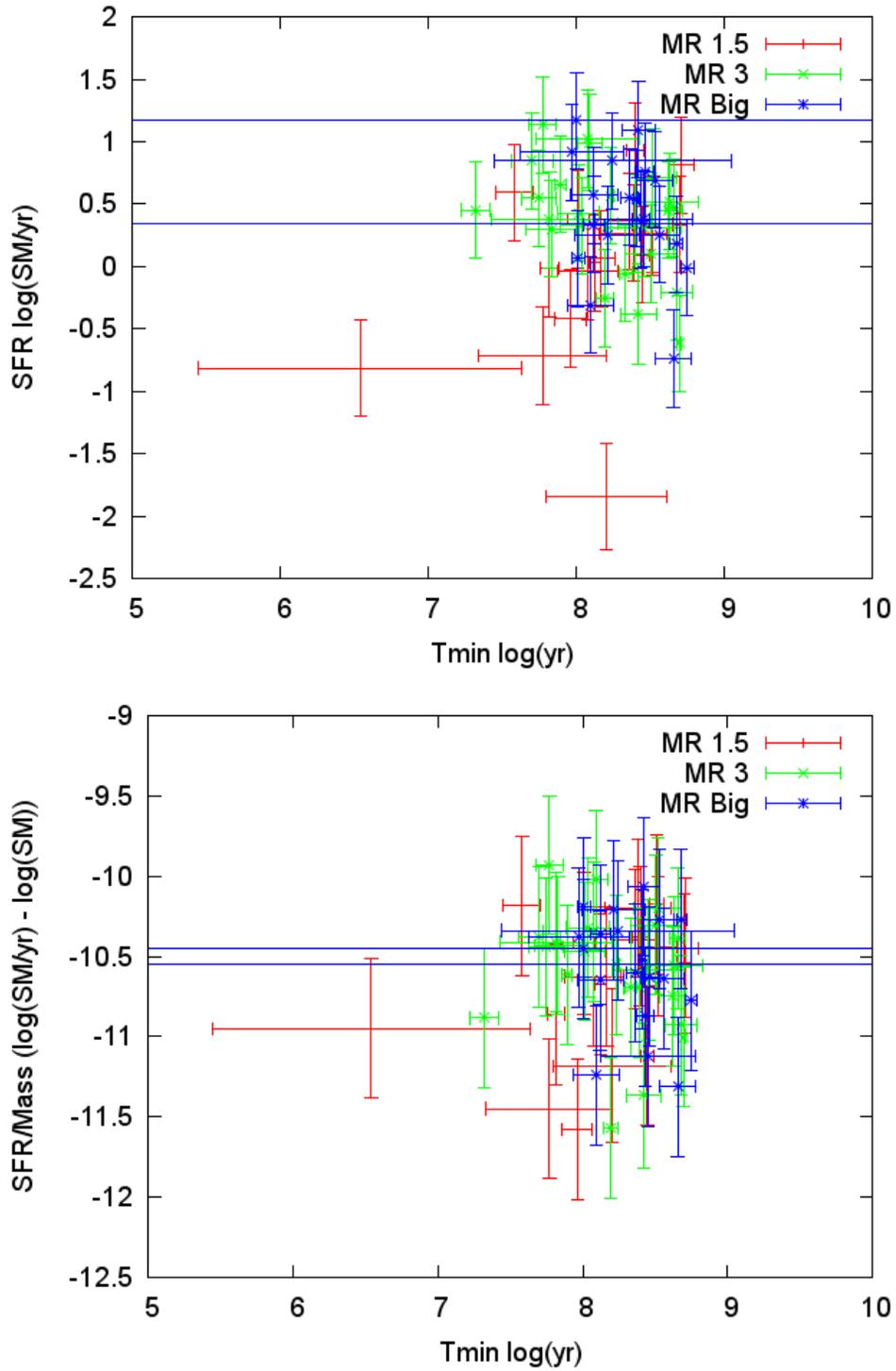


Figure 6.5: SFR for primary galaxies divided by mass ratio.

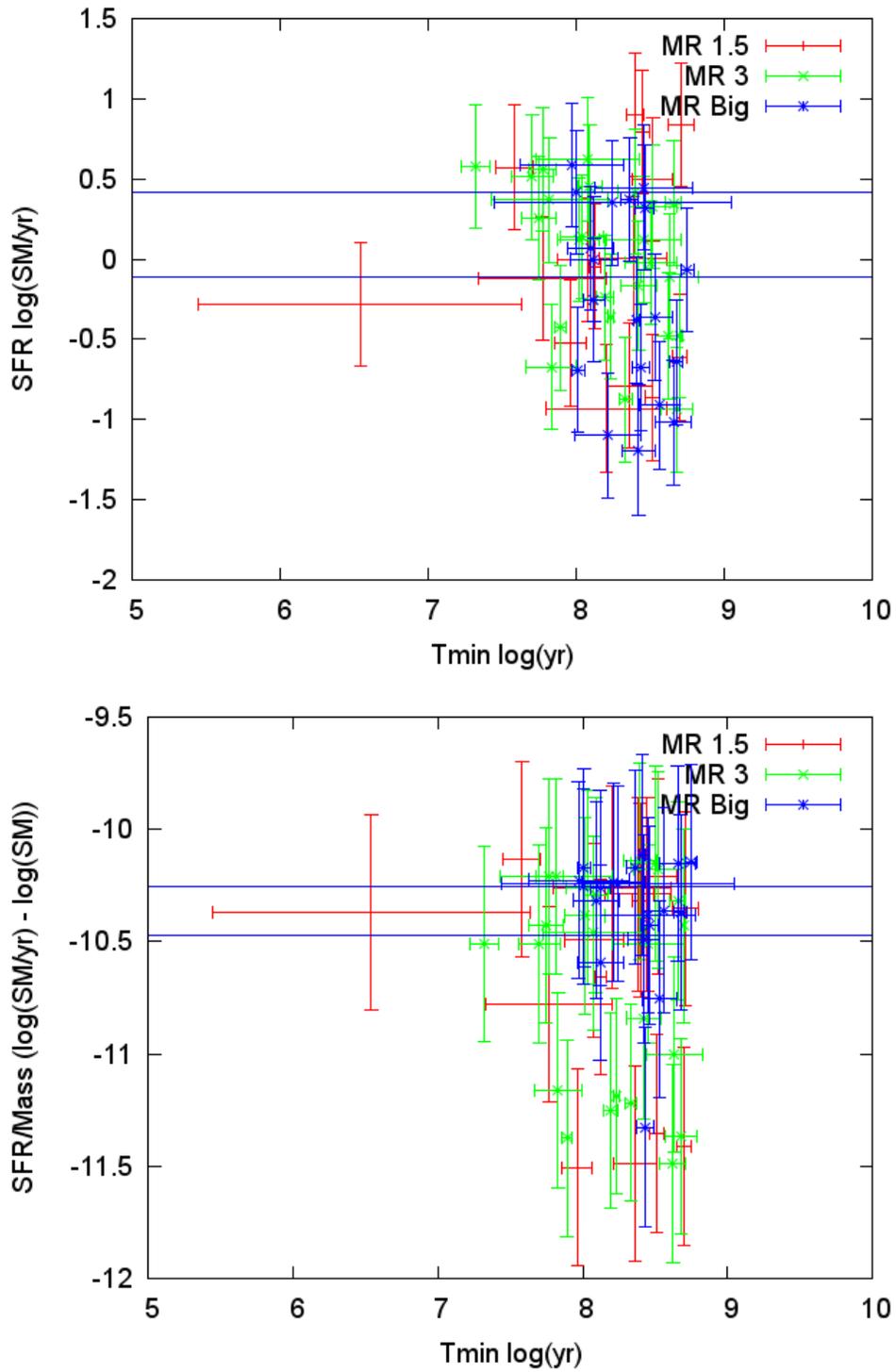


Figure 6.6: SFR for secondary galaxies populations divided by mass ratio.

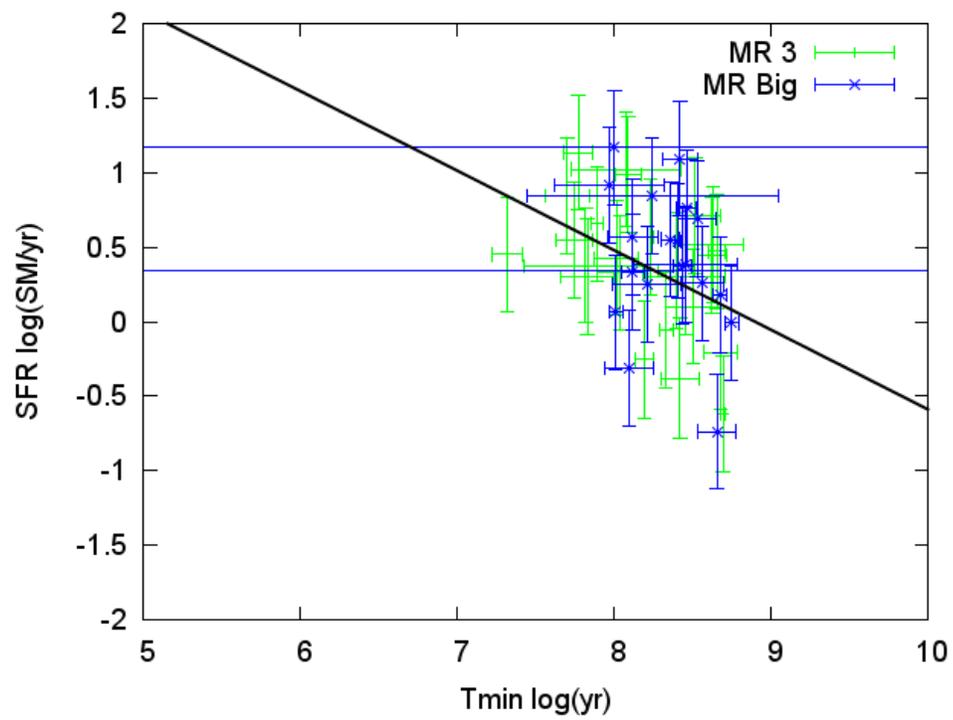


Figure 6.7: SFR for primary galaxies divided by mass ratio.

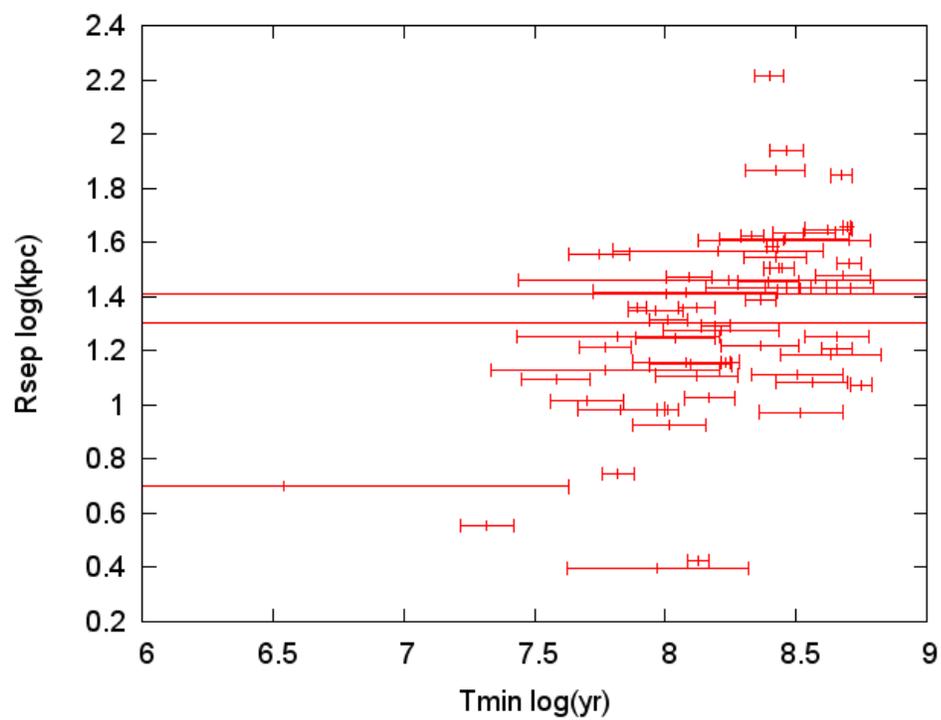


Figure 6.8: Projected separation distance as a function of  $T_{min}$ .

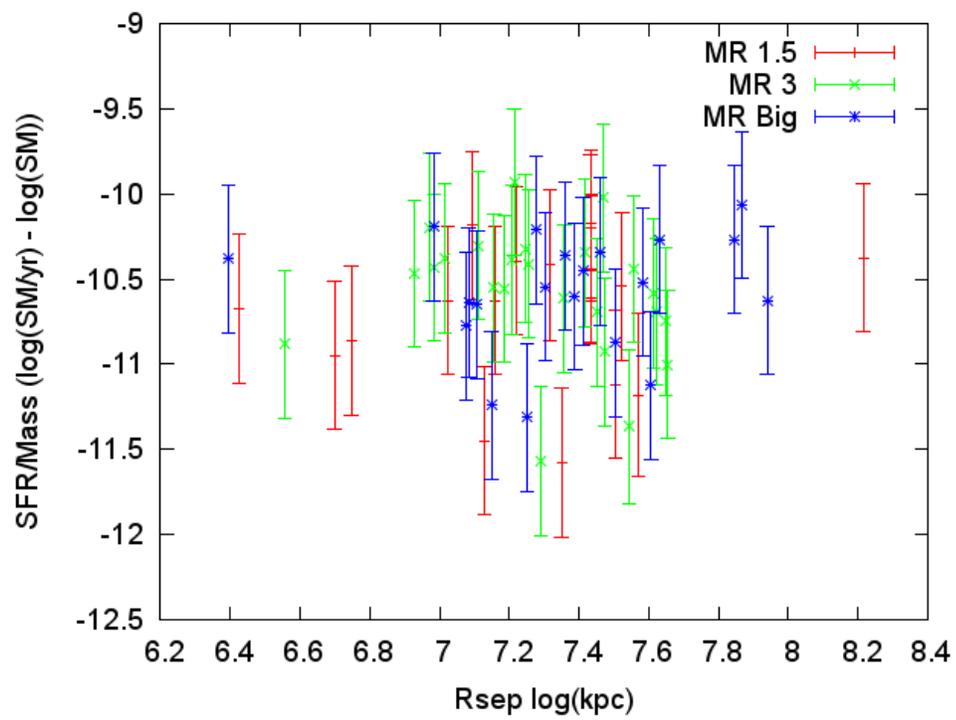


Figure 6.9: Normalized star formation rate as a function of projected separation distance.

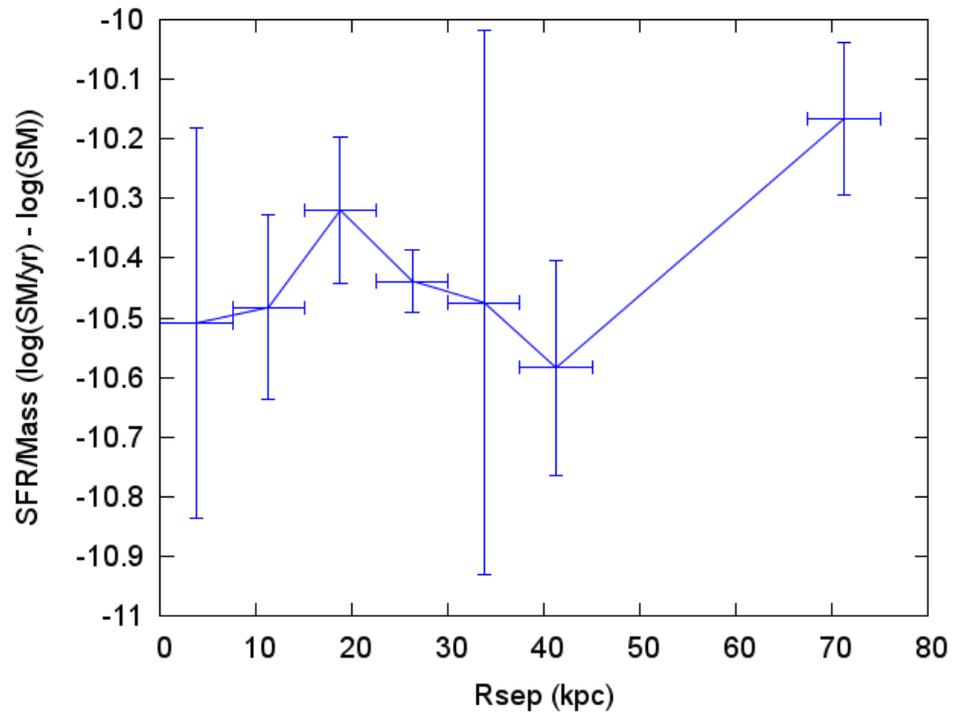


Figure 6.10: Normalized star formation rate as a function of binned projected separation distance.

## Chapter 7: Conclusion

The prototype pipeline was able to perform simulations for over 300 million samples of parameter space for 62 pairs of interacting galaxies. The pipeline was able to present over 3 million of those parameter sets to Citizen Scientist volunteers who selected 66000 of them for followup evaluation. Those same Citizen Scientists were able to perform 1 million Merger Wars competitions to assign fitness scores to the selected set of simulations and rank them accordingly. The final result is a best-fit, restricted three-body dynamical model of each of our 62 target interacting pairs. In the absence of a uniform observationally-derived data for all of our systems, we developed useful expressions for mass and star formation rate based on data from the WISE survey. Finally we have produced a training set of 66000 human-evaluated simulation results. These images have already proven useful as a training set for machine learning and computer vision algorithms.

Our dynamical mass values and mass ratios were a reasonable match to their photometrically-derived counterparts. We were able to produce an exponential decline in star formation rate with time when we exclude interacting galaxies with mass ratios very close to one.

### 7.1 The Pipeline for Constructing Models

The pipeline as presented in Figure 1.2 was implemented with the exception of the Full N-Body sub processes. Those processes were prototyped during the conduct of this research, however, they are now the focus of a separate dissertation which is an indicator of the level of work expected to be able to perform those tasks. The comparison to observations is somewhat more limited than originally planned. Even relatively famous and close galaxies such as those in the Arp catalog do not have a homogeneous set of observed quantities available. The most glaring omission is the lack of consistent measurement of the redshifts

to these galaxies. As a step towards providing uniform measurements, we developed a tool for performing photometry of infrared imagery from WISE. We also developed relationships for stellar mass and star formation rate. This provided us the consistent set of observational data we needed in the comparison step of the pipeline.

Applying experience gained by running a website for thousands of Citizen Scientists, we were able to refactor our software tools into a pair of applications that now run independently of Merger Zoo. These applications replace the previous version of Prepare Target and Merger Zoo processes in the overall pipeline. For interacting pairs similar to those from the Arp catalog processed by Merger Zoo, the new tool allows a researcher to perform and review several thousand simulations in under an hour. The efficiency of the Merger Wars algorithm and the meaningful constraints we place on the 14 dimensional simulation parameter space allow us to rapidly model interacting galaxies. The new process could potentially duplicate the converge of simulation parameters for a single system in only a few hours of work by a single researcher. The new pipeline is more efficient than Merger Zoo. The Evaluate task has been eliminated, as well as the Prepare Targets for Merger Wars task. This tool will be used in the future to model even more of the Darg catalog of mergers identified by Galaxy Zoo.

There are exceptions to the reasons for such optimism. Consider the possibility of the existence of systems that simply cannot be modeled realistically with restricted three-body approximations or where the morphology may not uniquely constrain the orbit trajectory. Barnes and Hibbard (2009) stresses the importance of incorporating velocity information, such as derived from HI observations, to constrain model results. They discuss the history of full n-body models for NGC 7252 where initially the best-fit orbit was thought to be retrograde (Borne and Richstone, 1991), but as improved velocity measurements were made, the best-fit orbit was changed to be a prograde orbit.

We provided several methods for visualizing convergence of simulation and orbit parameters. We found a strong correlation between the skewness of the fitness distribution of simulations and how well the best fit simulation matched the target image. We believe

the large sample and convergence information generated by our population will provide guidance to researchers studying such difficult systems to model.

## 7.2 The Merger Zoo Catalog of Interacting Galaxy Models

We have created a set of 62, best-fit, dynamical models using a restricted three-body code. The level of convergence for each system varies somewhat and is evaluated visually. The sheer number of simulations reviewed by Citizen Scientists, over 50000, is orders of magnitude more than viewed by current researches simulating individual systems. It is important to note, that current researchers are using the more sophisticated and computationally intensive full n-body codes. The hundred or so simulation runs they review represented an increased level of realism over the restricted three-body simulations run here. However, by reviewing so few simulations, they are unable to estimate the uniqueness of their final models. The restricted three-body portion of our multi-model process is currently the best mechanism available for exploring a wide volume of parameter space in order to achieve an estimate of uncertainty in final simulation parameters.

Our catalog also includes measurements of stellar mass and star formation rate derived from WISE photometry. Our dynamical masses for the galaxies correspond well with those derived from WISE data. The mass ratios were comparable as well. When we examine the star formation rates for the primary galaxies and exclude those pairs with nearly equal masses, we see an obvious trend of exponential rate of decay of star formation with time. The characteristic decay timescale for our population was 0.81 Gyr.

### 7.2.1 Machine Learning Training Set

The 66000 images with fitness score can be used by other researchers to develop a better automated fitness function. Our initial attempts at performing computer vision analysis on this data set has identified Zernike moments as potentially useful image characteristics. We will make the data set available to the public in accordance with applicable Zooniverse data release policies. We expect to be able to make use of these images to produce an

automated fitness function that will allow genetic algorithms to compete with our current human-driven pipeline.

# Appendix A: A Method for Photometry of Irregular Galaxies

## A.1 Introduction

Large Sky Surveys such as the Sloan Digital Sky Survey (SDSS) and Wide-field Infrared Survey Explore (WISE) (Wright et al., 2010) provide imagery and automated photometric measurements of millions of galaxies. However, the automated routines are often optimized for detection and processing of point sources like stars or galaxies with sizes on the order of a few arc seconds (Lupton et al., 2001) (Cutri et al., 2012). Additionally, where these automated routines can accommodate resolved galaxies, they usually do not have algorithms to address highly irregular shapes. Often galaxies are modelled by an exponential or de Vacoulers profile. The deviance from these profiles are an indication that a galaxy is irregular, but does not measure its actual photometric properties. While it is correct to say that these automated pipelines are not optimized for close, irregular galaxies, they are still quite successful in extracting photometric properties of many millions of objects.

This appendix presents the details of the photometric procedure developed to measure images of interacting galaxies. It presents the background subtraction process, flux measurement, and flux apportionment methods used to assign flux to both galaxies in each pair. This method was initially developed to work with SDSS imagery, but was extended to work with WISE imagery.

## A.2 Obtaining Imagery

Image data suitable for photometric measurements is provided by SDSS and WISE. The SDSS imagery is available from their Data Archive Server (DAS). This web interface provides access to FITS files for specific frames. A frame refers to a particular portion of the sky observed by the telescope. The DAS provides fpC files that are bias subtracted, flat fielded, and with bad pixels replaced by interpolated values. These are referred to as corrected frames.

The WISE images are available from the NASA/IPAC Infrared Science archive. The WISE atlas images are intensity images formed by coadding multiple images of the same portion of the sky. In addition to the intensity image, the service also provides uncertainty maps for each image which represent the  $1 \sigma$  noise for each pixel in the coadded image. This noise includes known detector noise and the Poisson noise.

Both surveys provide online forms for performing bulk query and download of images either by frame identifiers or source positions. The images are downloaded as FITS files and are registered to the appropriate location in the sky allowing for astrometry in addition to photometry.

### **A.3 Background Subtraction**

Flux in an image comes from both the sources of interest and the more diffuse sky background. The flux from the sky can be greater than some of the sources of interest, or at least brighter than the low surface brightness regions of a source. In order to properly measure the flux from sources, the background flux level must be estimated and subtracted from the images. To develop the approach presented here, three separate methods for improving upon the background estimation and deblending resulting from the automated pipeline were reviewed. They were West et al. (2010), Simard et al. (2011), and Blanton et al. (2011). In West et al. (2010), the authors were attempting to produce more reliable photometry for a sample of about 200 HI-selected galaxies. The authors in Simard et al. (2011) were attempting to fit combined bulge-disk models to 1.2 million galaxies. The last paper (Blanton et al., 2011) was attempting to provide improved background estimates for all photometric frames in SDSS. They tested their method on SDSS DR7 and have had their results incorporated into the imaging servers supporting SDSS DR8.

#### **A.3.1 Source Identification**

The background subtraction process begins by first identifying the sources in an image. The pixels containing flux from the sources are masked in the image. The unmasked pixels will

then contain only background flux. Sources can be identified through pre-existing catalogs or by measuring the image directly. For source identification, we use the SExtractor tool (Bertin and Arnouts, 1996). The automated tool works by applying a series of brightness thresholds to the image and then scanning for regions of 8 connected pixels. These regions can be combined at different thresholds to identify all of the source pixels. The tool also includes the ability to deblend the flux from multiple sources that have been grouped together. For the purposes of generating a source mask, the deblending feature is not needed.

### A.3.2 Background Estimation

In West et al. (2010) the authors fit a tilted plane to the background sky to estimate its value within the masked regions. Tilted plane is another name for a two-dimensional, linear regression. In Simard et al. (2011) they use the combined galaxy and background fitting routines within the GIM2D software package. Blanton et al. (2011) fit a spline model to the background. The tilted plane method was first introduced by West et al. (2010) to improve on the order of only 0.01 magnitudes compared to higher order models such as a fourth order polynomial fit. It is adopted here for use with WISE frames. The tilted plane values are calculated over the entire frame and subtracted from the unmasked science image. Fitting the plane involves performing the two-dimensional regression required to estimate  $\alpha$ ,  $\beta_1$ , and  $\beta_2$  as described in Equation A.1. The coordinate  $X_1$  corresponds to the pixel x coordinate and  $X_2$  corresponds to the y coordinate.

$$Y = \alpha + \beta_1 \times X_1 + \beta_2 \times X_2 \tag{A.1}$$

We apply the source identification and background estimation to the WISE imager for Arp 290 shown in the thumbnail in Figure A.1. This system shows two irregular disk galaxies that are well separated.

For the W1 imagery we used the SExtractor tool to identify the sources in the image. The tool also provides a FITS file that indicates which pixels in the image belong to sources. This source mask is padded by 4 pixels around the outside of all sources by convolving a 4 x

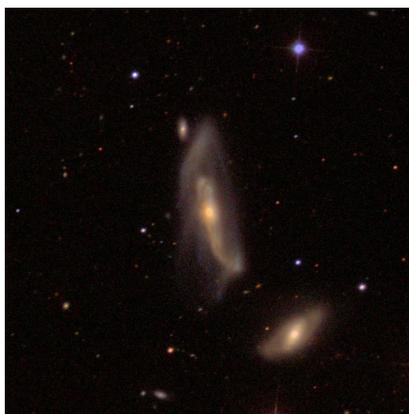


Figure A.1: SDSS color thumbnail for Arp 290.

4 pixel kernel with the mask. This padding acts to associate low surface brightness features with the sources and keep those pixels out of the background estimation. The mask is then inverted to represent the known background pixels. Figure A.2 shows the sources identified by SExtractor with the source centers highlighted in pick and the background pixels set to white. Figure A.3 shows this same mask after padding and inversion. The background is estimated from the pixels in this image.

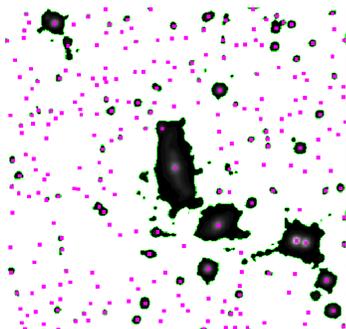


Figure A.2: Arp 290 source pixels with background set to white.



Figure A.3: Arp 290 background mask with 4 pixel padding.

The same mask is used to obtain background pixels in all four bands, W1, W2, W3, and W4. The results of the background estimation process are shown in Table A.1. The background is modeled as a fitted plane. For convenience, we have provided the estimated mean and standard deviation of the background flux before and after the tilted plane was

subtracted. The final column shows the relative difference between the  $\sigma$ s before and after subtraction. In all cases, the tilted plane mode reduced the size of  $\sigma$ , and in the best case, W2, reduced it by 28%. This is better than would result from simply subtracting the mean or median background flux. The resulting mean background flux in all four bands is very close to zero.

Table A.1: Background  $\mu$  and  $\sigma$  before and after background subtraction.

Band	$\mu_{pre}$	$\sigma_{pre}$	$\mu_{post}$	$\sigma_{post}$	Rel. $\Delta_\sigma$
W1	4.44	0.0739	-4.12E-16	0.0721	0.0249
W2	11.4	0.0694	-8.52E-17	0.0691	0.00463
W3	812	0.417	8.58E-14	0.369	0.113
W4	268	0.0476	-2.85E-14	0.0342	0.282

## A.4 Measuring Flux in an Aperture

Typically apertures are circles or ellipses. To measure flux, simply add up all of the pixels inside of the aperture. Some amount of attention can be paid to pixels that are on the boundary. For our apertures, pixels on the aperture boundary are included in the flux measurement.

For our irregular galaxies we use a custom tool written in Java to do the processing. All four bands, W1, W2, W3, and W4 are downloaded to the same directory. The background estimation and subtraction process is applied. With the background subtracted, a contour finding algorithm is applied. The user clicks on the center of the primary galaxy and then the secondary galaxy to indicate where they are. The user is allowed to have an additional constant amount subtracted from the image. This is equivalent to raising the floor of the image in a thresholding operation. Whenever the background is adjusted, the contour finding algorithm is reapplied. After first identifying the galaxy centers, the software will help to estimate which threshold level is needed in order to draw separate contours around each galaxy. If the two galaxies are separated into distinct contours at the lowest level, the sum of all pixels inside each contour is computed and stored for later use. If the two

galaxies are in the same contour, the total flux is computed. Then, using the relative pixel area for the separated contour level is used to assign a proportional amount of the total contour flux to each galaxy.

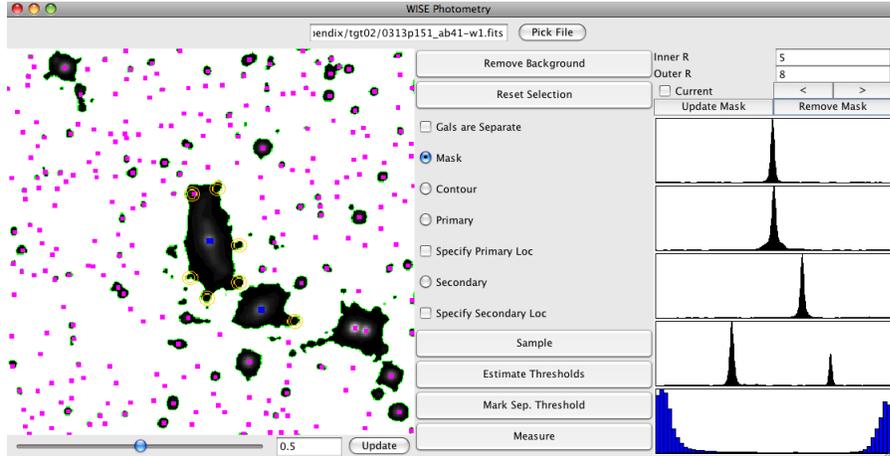


Figure A.4: Custom WISE photometry tool measuring Arp 290.

Figure A.4 shows the custom photometry tool measuring Arp 290. The yellow circles represent stars that we want to mask. Their flux is replaced with an estimate of the local background. That value is estimated by calculating the median flux value between the two concentric circles. The tool also shows the sample estimates. The four histograms on the right side are, from top to bottom the flux values in the same row as the primary source, the flux from the column of the primary, the flux values for the row of the secondary, the flux values for the column of the secondary, and the flux values on the straight line joining the two galaxies. For galaxies that are well separated, the flux drops down to the background level on this final plot. An example of what this looks like when the two galaxies have overlapping flux see Figure A.5 below. That figure shows the WISE photometry tool measuring UGC 11751. The line plot does not return all the way to the background level. The proportional assignment of flux will be necessary in this case.

After masks have been applied, primary and secondary identified, and minimal threshold necessary to separate the galaxies is identified, the software will integrate the total flux for each galaxy and convert to magnitudes. The results are written to a directory where the

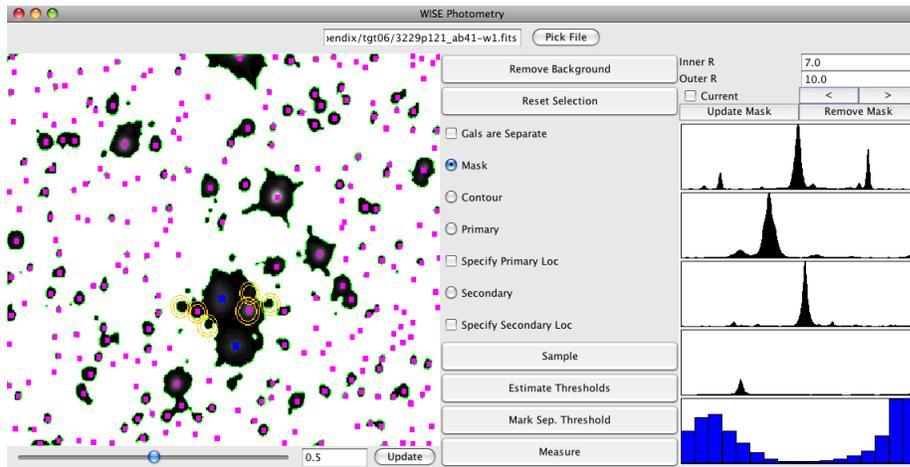


Figure A.5: Custom WISE photometry tool measuring UGC 11751.

WISE images are stored. Each galaxy pair has its imagery in a separate directory with the imagery for all four bands together. The same background masks, source masks, and galaxy contour apertures are used for each WISE band.

## A.5 Radial Profile

With the total integrated flux computed for each galaxy and the center of each galaxy located in pixel coordinates, the software identifies the brightest level in each galaxy. It then attempts to set a number of distinct threshold levels by raising the floor of the image repeatedly until the contour finder detects a smaller contour. The flux inside the next contour is measured. The process is repeated until the floor reaches the brightest level.

The resulting set of contours each have their own fluxes and areas. These apertures form the first step in a differential light measurement. In an attempt to compute a radial profile similar to a set of nested elliptical apertures, the software computes the square-root of the area of each contour to estimate an effective radius. The differential light curve can then be produced. Analytic radial intensity profiles have been known for sometime such as the exponential of de Vaucouleurs  $R^{1/4}$  law. Then general form of these laws is known as a Sersic profile.

$$I(R) = I_0 10^{-b_n [(R/R_e)^{1/n} - 1]} \quad (\text{A.2})$$

The parameter  $b_n$  is not a free parameter it depends on the others and is selected in such a way that half of the total light comes from  $R < R_e$ . The analytic expression for  $b_n$  is from Ciotti and Bertin (1999) is used. That leaves three parameters to fit,  $I_0$ ,  $R_e$ , and  $n$ . The exponent  $n$  is closely linked to morphology of galaxies. Values close to 1 are associated with spiral galaxies. Values close to 4 are associated with elliptical galaxies. We used the Levenberg-Marquardt nonlinear least-squares method to fit the Sersic profiles (Lampton, 1997).

It is possible to fit multiple Sersic profiles simultaneously for a galaxy, such as in the case of an elliptical bulge embedded in a spiral disk. In Figures A.6 and A.7 we show the intensity profile as a function of effective radius for W1 for the primary and secondary galaxies of Arp 290. The effective radius is computed by simply taking the square root of the area of a given flux contour.

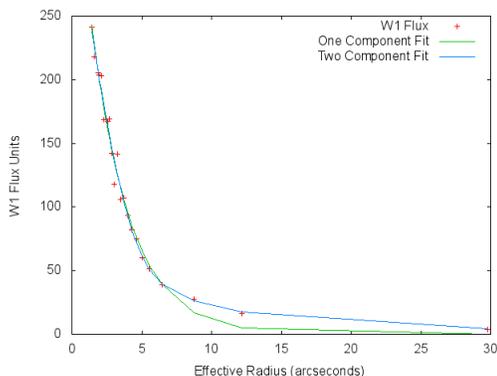


Figure A.6: Radial profile and model fits for W1 of the Arp 290 primary galaxy.

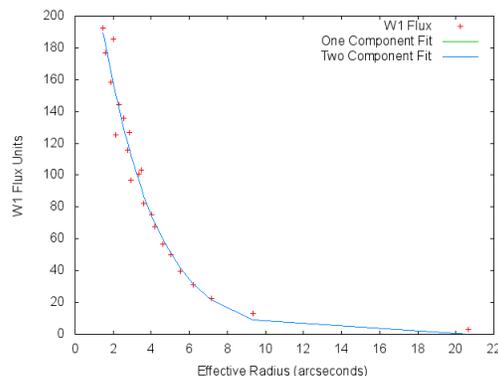


Figure A.7: Radial profile and model fits for W1 of the Arp 290 secondary galaxy.

The two component fit is a slightly better match for the primary, but adds nothing to the fit for the secondary. The single Sersic profile for the primary had an exponent of 1.00 and an  $R_e$  of  $10.7''$ . The exponent for the secondary was 0.92 and  $R_e$  was  $9.17''$ . These galaxies have brightness profiles similar to spiral galaxies.

### A.5.1 Multiple Bands

The Sersic profile can be fit for each band, W1, W2, W3, and W4. Table A.2 presents the exponent and  $R_e$  for the primary and secondary galaxies of Arp 290.

Table A.2: Sersic profile parameters for Arp 290 primary and secondary galaxies.

Band	Primary		Secondary	
	$R_e$	$n$	$R_e$	$n$
W1	10.7	1.00	9.17	0.91
W2	11.3	0.95	8.99	0.82
W3	166	1.70	13.0	0.80
W4	46.5	0.85	63.6	1.14

Similar results were obtained for each of the 62 target galaxies. These values will be used in followup photometric analysis of the targets. Not only for the traditional spiral or elliptical dichotomy but also to examine cause of the different values for  $n$  and  $R_e$  for some spectral bands.

## Appendix B: The Effectiveness of the Merger Wars Algorithm

### B.1 Abstract from Submitted Draft

The method of pairwise comparisons for ranking is often applied to relatively small populations,  $n = 100$  or even  $n = 10$ . Many methods of finding optimal rankings involve performing comparisons between all pairs resulting in  $n(n - 1)/2$  comparison operations. For accurate comparison functions, traditional fast sorting methods can produce a total rank order of the population in  $O(n \log n)$  comparison operations. There has been significant study of methods for comparing numbers or items with values that contain uncertainty. These methods are often called fuzzy sorting. However, the performance of sorting methods when used with errors or uncertainties in the comparison function, rather than the number itself, does not appear to have been studied extensively. This work presents the results of experimental simulations to determine the effect on overall rank accuracy of the error rate in the comparison function used for sorting. For comparison, the performance of an updated method for ranking by scoring is also presented. For large error rates in the comparison functions, the ranking by scoring method produces more accurate results than sorting and preserves  $O(n \log n)$  performance. This is an improvement over traditional pairwise ranking algorithms.

### B.2 Introduction

The study of pairwise comparisons to produce ranked sets began at least as early as the 18th century with the work of the Marquis de Condorcet. Thurstone (1927) demonstrated that it was possible to recover a psychological scale to quantify items that have undergone a series of pairwise comparisons. Pairwise comparison methods are used to rank items such as chess players or competing hypotheses. In the field of machine learning, methods exist

to evaluate the results of pairwise comparisons in order to learn preferences. The results of these algorithms can then be used to rank objects directly or to rank candidate labels in classification problems.

When ranking objects according to some criteria, be it something physical like length or abstract like quality, two obvious approaches are evaluation and comparison. Under an evaluation approach, each item is evaluated based on the criteria and assigned a score. The scores are then used as keys to sort the items. The second approach is to compare pairs of items as if they were in a competition. Typical comparison methods involve performing all possible combinations of pairs,  $n(n - 1)/2$ .

In preference learning methods, usually the preferences are gathered and then an error minimization technique is applied to this set of comparison results (sometimes referred to as a preference matrix) to account for potential inconsistencies including the violation of transitivity. These inconsistencies, which are not necessarily errors, can arise for a multitude of reasons: differing opinions between evaluators, changing one's mind, fatigue, or an actual circular paradox where  $A > B > C > A$ . The preference matrix can then be used to induce a utility function,  $f$ , that can be used to determine if  $f(a) > f(b)$  (Hüllermeier et al., 2008).

In this paper, we propose two schedules of pairwise comparisons that can avoid having to populate the entire preference matrix. The first schedule follows from Silverstein and Farrell (2001) where it was observed that a list sorting algorithm performs comparisons between pairs in an efficient way. The other schedule of comparisons proposed here is a ranking by scoring method (Pyle, 1999). Rather than using round-robin or purely random selection of pairs, the proposed method groups items into small sets to be evaluated tournament style.

We evaluate the performance of each method type in the presence of several different types and levels of errors. Each of these schedules can produce accurate ranks of the entire set of items with only  $O(n \log n)$  comparisons.

### B.2.1 Motivation

In our work with Galaxy Zoo - Mergers (Wallin et al., 2010), a Zooniverse Project, we are tasked with determining which of the thousands of simulations submitted by volunteers best recreates the features seen in an image of interacting galaxies. At the current time algorithms for comparing simulations results to actual images only provide accurate results over a limited range. Such techniques can easily determine a poor match and an exact match, however, they often fail to determine the correct relative accuracy of two good, but not perfect, simulation results. For now, humans are better able to assess quality in those situations.

To harness the innate image processing abilities of our volunteers, we created an activity known as Merger Wars. Volunteers are presented with a series of pairwise comparisons where they are asked to determine which of the two simulation results is a better match to the target image. In the short term, the volunteers are able to filter the submitted simulations so that we can focus follow-up, high-resolution simulations on the small fraction of high-fitness results. Our long term objective is to produce a utility or fitness function that covers the entire range of preferences learned from our volunteers. The ranking by scoring method presented here follows from what was implemented to satisfy a number of other constraints related to operating an online web site. Such constraints prevented us from easily implementing a traditional sorting approach such as Silverstein and Farrell (2001). The tournament method was our solution for operating under these constraints. We present our evaluation of the performance of both types of methods.

### B.3 Methods

For sorting, three algorithms with  $O(n \log n)$  performance were considered: merge sort, quick sort, and tree sort. Two algorithms with  $O(n^2)$  were considered: bubble sort and insertion sort. For an overview of sorting in general, and these algorithms in particular, consult Knuth (1973). The ranking by scoring method was used with a simple win fraction

score and also with an Elo score (Elo, 1978). The win fraction score is obtained by taking the number of times an item “wins” a comparison and dividing it by the number of comparisons in which that item participated. The Elo score for two opponents in a competition makes use of their current rating or score to calculate the expected outcome of a match between the two. The difference between the actual outcome and the expected outcome is used to update the players’ ratings or scores. The expected Elo score for Player 1 and Player 2 is given by

$$E_1 = \frac{1}{1 + 10^{(R_2 - R_1)/400}}, \quad (\text{B.1})$$

and

$$E_2 = \frac{1}{1 + 10^{(R_1 - R_2)/400}}. \quad (\text{B.2})$$

After a match, or comparison, the score for each player is calculated as 0 for a loss, 1 for a win, and 0.5 for a tie (which is not considered in this paper). The update to each player’s rating is then calculated by  $R'_1 = R_1 + K(S_1 - E_1)$  and  $R'_2 = R_2 + K(S_2 - E_2)$ . Here we used  $K = 30$  and seeded all items with an initial score of 200.

The ranking by scoring method (RSM) and RSM with ELO (RSM-ELO) requires that the number of comparisons, number of wins, and score for each item be stored. For RSM-ELO, the ELO formula is used to update the score for both items following each comparison. Some implementations apply the ELO formula to batch results following a tournament. We did not consider that variation in the present work. For the basic RSM implementation the score is updated after each comparison simply by incrementing the number of comparisons, number of wins, and calculating the win fraction. The scheduling for RSM and RSM-ELO is done as follows:

1. Select 8 items to compete in a tournament from the set of items with the fewest total comparisons

2. Perform 4 pairwise comparisons for Round 1
3. Using the Round 1 winners, perform 2 pairwise comparisons for Round 2
4. Using the Round 2 winners, perform a final pairwise comparison for Round 3

This very simple tournament separates the 8 items into 4 groups: Round 1 losers, Round 1 winners, Round 2 winners, and a Round 3 winner. Generally, low ranking items are eliminated in the first round and high ranking items advance to the final round. This tournament is single elimination, or knockout, but because it is small, the impact of erroneous comparisons is minimized. For further protection against errors, the selection in the first step is used to ensure a uniform number of comparisons for all items in the set.

For a first pass through the population of items, they are randomly selected 8 at a time from the set of items with no comparisons. After all items have competed in at least one tournament, some will have 1 comparison, some 2, and some 3. The next pass through the population will select random sets of 8 items with only 1 comparison. These are the Round 1 losers from the first pass. The next pass will take random sets of 8 from the population with only 2 comparisons, etc. As the method progresses, the items are competed in multiple short tournaments. In each tournament, the winners gain more wins and comparisons within a tournament, and the losers will be re-competed sooner on the next large pass.

### B.3.1 Comparison Function

The comparison symbols for greater than and less than can be combined into a comparison function:

$$c(a, b) = \begin{cases} -1, & a < b \\ 1, & a > b \end{cases} \quad (\text{B.3})$$

A complete comparison function would include a case where  $a=b$  returns 0. That case is not considered here. We constructed an erroneous comparison function by selecting a

random number,  $x$ , from a uniform distribution between 0 and 1. For random errors our comparison function looked like

$$er(a, b, x, Accuracy) = \begin{cases} c(a, b), & x \leq Accuracy \\ -c(a, b), & x > Accuracy \end{cases} \quad (\text{B.4})$$

We used several different accuracy levels: 1, 0.99, 0.95, 0.9, 0.8, 0.75, 0.7, 0.6, and 0.5. In addition to random errors, we considered the impact of errors arising from the fact that people are less able to discriminate between two nearly equal choices than they are for two choices that are very different. We called this type of error a Weber error after Weber's Law. The form of Weber's Law we used states that unless the relative difference between two objects is greater than a certain threshold, the person making the comparison will be unable to distinguish between the two choices. We require the relative distance between the two items to be greater than 0.25. Less than that threshold and either value of the comparison function is returned with equal likelihood.

$$ew(a, b) = \begin{cases} c(a, b), & \frac{a-b}{a} > 0.25 \\ er(a, b, x, 0.5), & \frac{a-b}{a} \leq 0.25 \end{cases} \quad (\text{B.5})$$

### B.3.2 Simulation Method

For each sorting method and RSM we evaluated the comparison function for 9 different error levels as well as the Weber error. We looked at multiple population sizes,  $n$ , to verify performance as a function of  $n$ . For each combination of method, comparison function, and population size we conducted multiple trials. Each trial consisted of generating a random list of  $n$  real numbers from a range of 0 to 1. We used the Java<sup>TM</sup> random number generator to select from both a uniform and a normal distribution, in separate trials. The results of the multiple trials were combined to determine the mean and standard deviation for the

number of comparisons and the accuracy metrics.

Within a trial, the sorting method was applied up to 10 times on the same list of data. The intent was to see if multiple sequential sorts would improve the overall sorting accuracy. For example, if a list is sorted with one pass of the quick sort algorithm with a comparison function with accuracy 0.75, the result will be a list that is not in correct rank order. Another pass of the quick sort algorithm with the same accuracy will likely move the list to be closer to the correct rank order. We expect that multiple passes of a sorting algorithm will continue to improve the accuracy of the rank order as long as the comparison function used has an accuracy  $> 0.5$ .

### B.3.3 Accuracy Metric

For each randomly generated list, we were able to use a traditional sorting method with a perfectly accurate comparison function to determine the true rank order. This made it possible for us to calculate the accuracy of each pass of the sorting methods. When comparing the actual ranking of a list of items,  $\tau$ , with the predicted ranking,  $\tau'$ , Hüllermeier et al. (2008) suggest using the Spearman rank correlation,  $\rho$ , and Kendall's Tau.

Spearman's Rho is defined as :

$$(\tau, \tau') \mapsto 1 - \frac{6 \sum_{i=1}^n (\tau(i) - \tau'(i))^2}{n(n^2 - 1)}. \quad (\text{B.6})$$

Kendall's Tau provides a measure of rank accuracy by counting the number of discordant, or rank reversed, pairs. It is defined as:

$$(\tau, \tau') \mapsto 1 - \frac{2|\{(i, j) | \tau(i) < \tau(j) \wedge \tau'(i) > \tau'(j)\}|}{n(n-1)/2}. \quad (\text{B.7})$$

The calculation of tau is an  $O(n^2)$  operation. Both accuracy measures cover the interval  $[-1, +1]$ . A value of 1 indicates the list is in correct rank order, -1 means inverse rank order, and a value close to 0 indicates an ordering that appears random.

## B.4 Results

Figure B.1 is a scatter plot showing the performance (Spearman’s Rho) of each method when ranking 1000 items chosen at random from a uniform distribution with comparison functions of various accuracies after one iteration of the sort for each trial. The y-error bars represent the one  $\sigma$  uncertainty in the value of  $\rho$  after at least 10 trials.

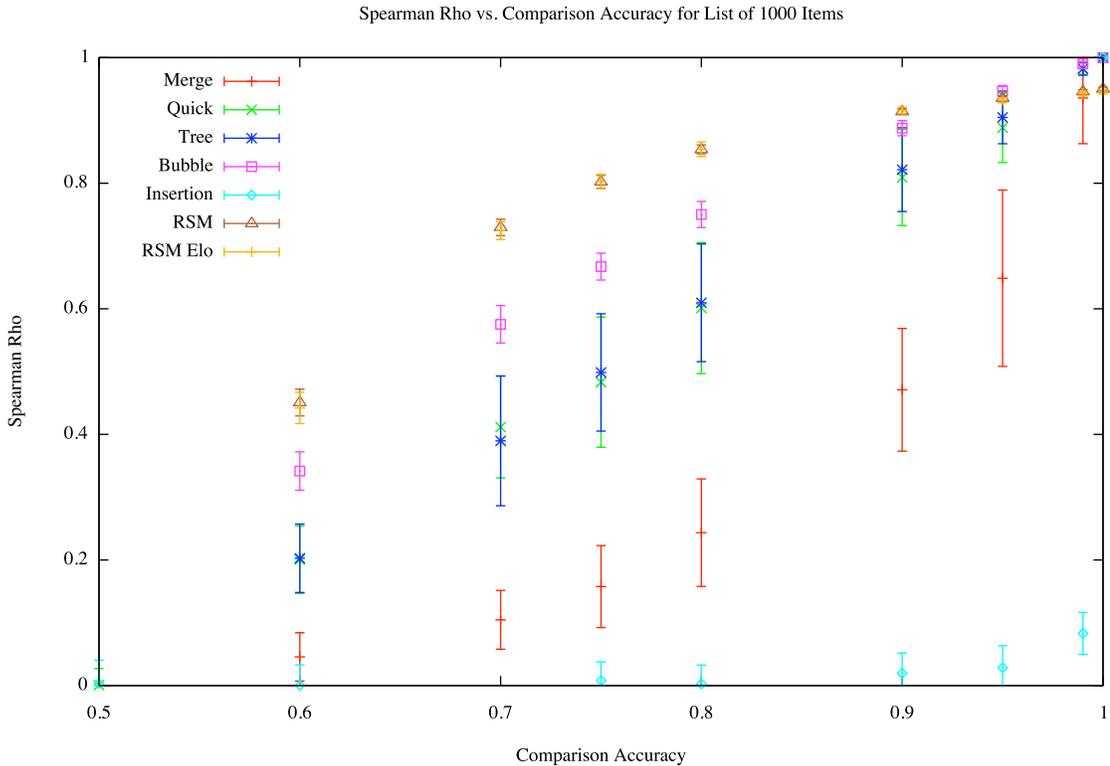


Figure B.1: Spearman Rho for  $n = 1000$

Among the  $O(n \log n)$  sorts, quick sort and tree sort perform very similarly. The merge sort appears to be less accurate after a single sort iteration. Overall, the best performing and worst performing sorts were  $O(n^2)$ , bubble sort and insertion sort respectively. For accuracies below 0.95, the best performing methods were the two RSM implementations. Their values improve with each comparison. The RSM methods do not have an internal convergence criteria, so the state of the list after  $n \log n$  comparisons was used to compute  $\rho$ .

For a perfectly accurate comparison function, neither RSM achieves perfect sort accuracy.

Increasing the list size to 100 000, does not change the performance order of the methods, see Figure B.2. The merge sort and insertion sort perform noticeably worse. The RSM methods appear to perform slightly better for the larger list size.

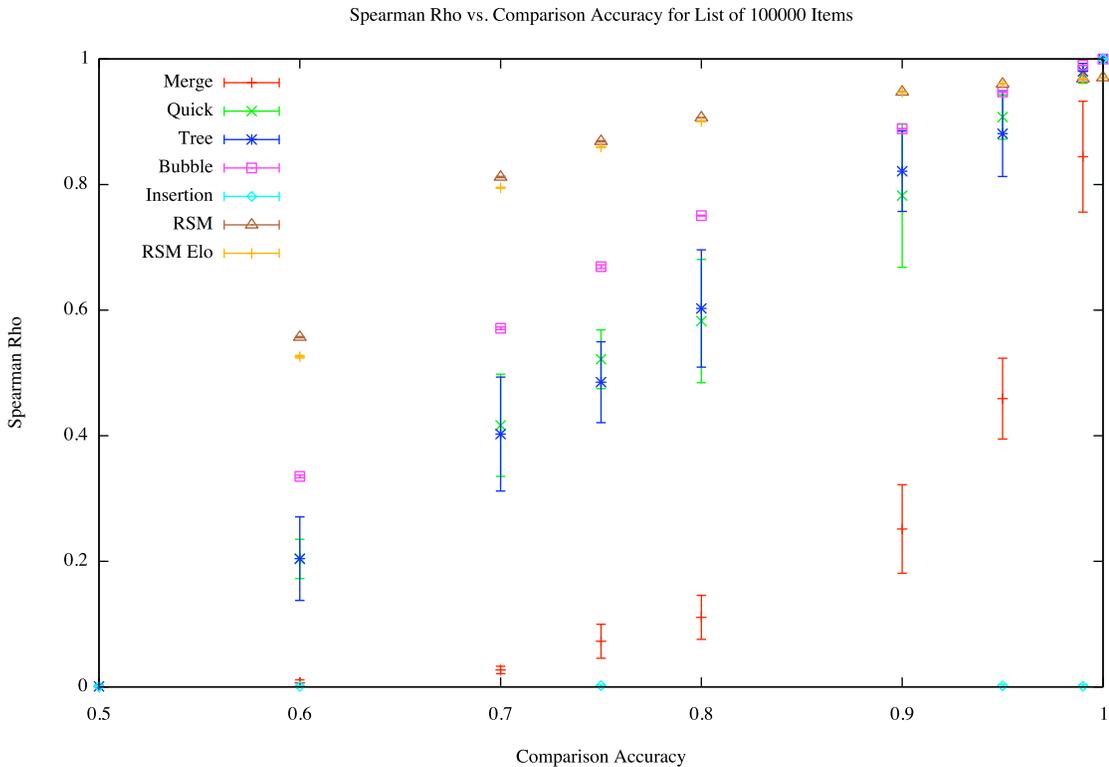


Figure B.2: Spearman Rho for  $n = 100\,000$

The sorting performance is similar when measured with Kendall Tau and for the normal, rather than uniform, distribution. However, there is a difference in performance between ranking the two types of distributions when using the Weber's law to model the error in the comparison function. Table B.1 shows the performance for each method under the Weber comparison function. All methods perform similarly to a comparison function with an accuracy in the range of  $[0.9, 1.0]$ . All of the sorting methods perform worse under the normal distribution than the uniform distribution. Conversely, the two RSM algorithms perform better with the normal distribution. The accuracy of insertion sort improves under

the Weber model than the random error model. Its performance is still much worse than the other methods.

Table B.1: Performance for a List of 1000 Items with Weber Comparison Function for Uniform and Normal Distributions

Method	$\rho$ - Uniform	$\rho$ - Normal	$\rho$ Difference	$\tau$ - Uniform	$\tau$ - Normal	Difference $\tau$
Merge	$0.959 \pm 0.004$	$0.859 \pm 0.013$	-0.1	$0.834 \pm 0.008$	$0.675 \pm 0.014$	-0.159
Quick	$0.959 \pm 0.005$	$0.859 \pm 0.019$	-0.1	$0.827 \pm 0.011$	$0.663 \pm 0.0215$	-0.164
Tree	$0.959 \pm 0.005$	$0.856 \pm 0.019$	-0.103	$0.827 \pm 0.01$	$0.661 \pm 0.02$	-0.166
Bubble	$0.977 \pm 0.003$	$0.868 \pm 0.012$	-0.109	$0.879 \pm 0.007$	$0.699 \pm 0.013$	-0.18
Insertion	$0.443 \pm 0.035$	$0.098 \pm 0.032$	-0.345	$0.307 \pm 0.026$	$0.064 \pm 0.022$	-0.243
RSM	$0.918 \pm 0.004$	$0.936 \pm 0.004$	0.018	$0.746 \pm 0.006$	$0.78 \pm 0.006$	0.034
RSM Elo	$0.907 \pm 0.005$	$0.934 \pm 0.003$	0.027	$0.726 \pm 0.007$	$0.772 \pm 0.005$	0.046

One method to improve the final rank order accuracy is to perform consecutive sorts of the same list for each trial. The performance of some sorting methods is dependent on the initial order of the list. For example, quick sort will perform poorly on a nearly sorted list. However, for even moderate accuracies such as 0.75, after only one sort iteration the list has a  $\rho$  of 0.5 when compared to the correct rank order. Figure B.3 shows the performance of each of the methods (excluding bubble sort) as the number of sort iterations is increased. Quick sort and tree sort do their best work on the first iteration. The other methods show at least incremental improvement with each additional iteration. Interestingly, the  $O(n^2)$  method, insertion sort, has several iterations in the  $O(n \log n)$  comparisons region. This indicates that an inaccurate comparison function causes the algorithm to converge very prematurely. The two RSM algorithms outperform the other methods.

Seeing that several of these methods do not achieve their best accuracy until after several iterations of  $O(n \log n)$  comparisons, we wondered if there was an easy improvement to each method. For every pairwise comparison, perform the comparison twice. If the two answers agree, accept it. If not perform a third, tie-breaker comparison. This is also known as a “Best of Three” competition. In general, the accuracy of a comparison function with errors can be improved by using a “Best of N” competition. The expected accuracy can be given as a probability of winning the match, where the probability,  $p$ , of winning any single comparison is the same as the accuracy.

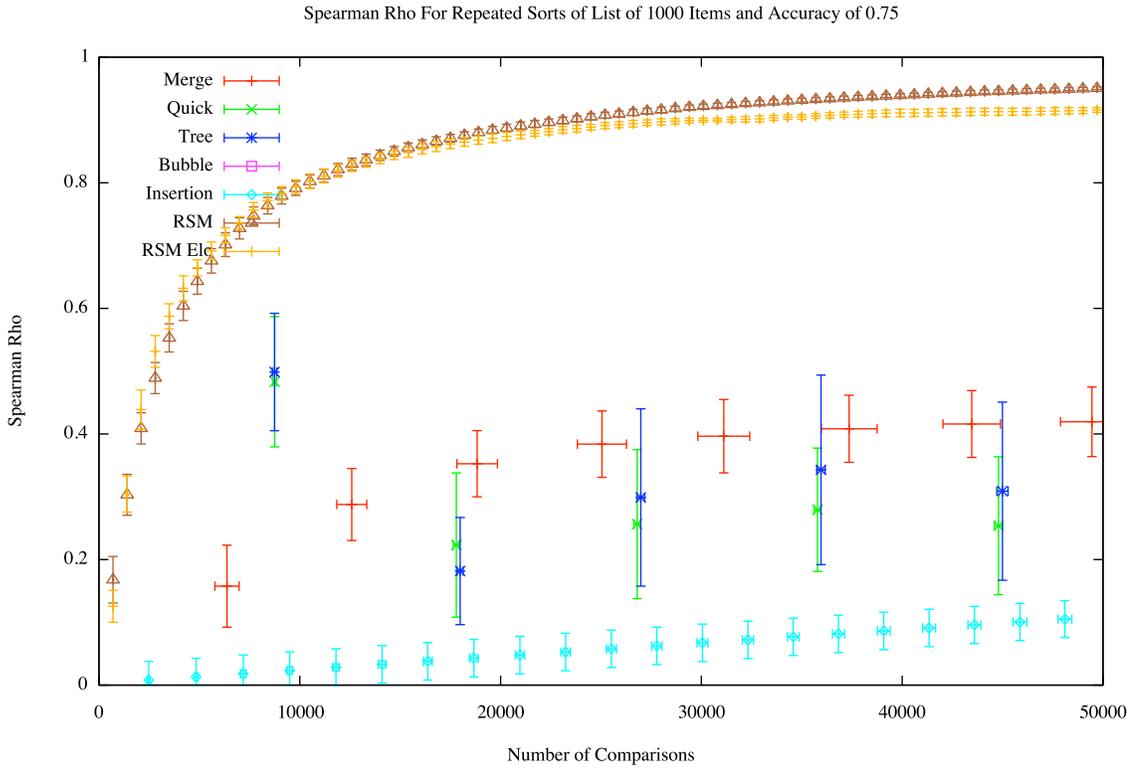


Figure B.3: Spearman Rho for Repeated Sorts of  $N = 1000$ , Accuracy = 0.75

$$P(k \text{ of } n) = \sum_{j=k}^n P_j(p)$$

$$P_j = \binom{j-1}{k-1} p^k (1-p)^{j-k} \quad (\text{B.8})$$

The expected number of comparisons, per pair, as a function of  $p$  then becomes

$$E(k \text{ of } n) = \sum_{j=k}^n j * (P_j(p) + P_j(1-p))$$

To verify, we performed a merge sort on several lists of 1000 numbers selected from a

uniform distribution. For an original accuracy of 0.75 we calculated the expected accuracy and number of comparisons for best of 3, best of 5, and best of 7. We then computed the rho and tau using the experimental procedure for the “Best of” competitions as well as single competitions using the expected accuracy of each.

Table B.2 contains the expected accuracy and the multiplier for the number of comparisons as the N in “Best of N” is increased with initial accuracy of 0.75. The third column shows the number of comparisons when performing sorting with only single comparisons when using the expected accuracy. The fourth column shows the multiplier, E, multiplied by the number of comparisons. The final column shows the actual number of comparisons for each “Best of N” trial. The agreement between the last two columns is well within one  $\sigma$  of each other.

Table B.2: “Best of N” Number of Comparisons for a List of 1000 Items with Initial Accuracy of 0.75

Best of N	Expected Accuracy	Expected Comparisons	Comparisons	E * Comparisons	“Best of” Comparisons
1	0.75	1	6400 $\pm$ 600	-	-
3	0.844	2.38	7180 $\pm$ 540	17100 $\pm$ 1280	17100 $\pm$ 1230
5	0.896	3.77	7780 $\pm$ 400	29400 $\pm$ 1510	29300 $\pm$ 1650
7	0.929	5.16	8110 $\pm$ 360	41800 $\pm$ 1860	41800 $\pm$ 1790

Table B.3 contains the  $\rho$  and  $\tau$  values. The value when using only single comparisons at the expected accuracy is given. Next, the value is given for using increasing N in “Best of N” with an initial accuracy of 0.75. Again, the values match the predicted values to within one  $\sigma$ .

Table B.3: “Best of N” Performance for a List of 1000 Items with Initial Accuracy of 0.75

Best of N	Expected Accuracy	$\rho$	“Best of” $\rho$	$\tau$	“Best of” $\tau$
1	0.75	0.168 $\pm$ 0.070	-	0.117 $\pm$ 0.048	-
3	0.844	0.316 $\pm$ 0.103	0.309 $\pm$ 0.103	0.231 $\pm$ 0.076	0.226 $\pm$ 0.075
5	0.896	0.453 $\pm$ 0.112	0.456 $\pm$ 0.127	0.356 $\pm$ 0.090	0.358 $\pm$ 0.101
7	0.929	0.572 $\pm$ 0.134	0.572 $\pm$ 0.124	0.480 $\pm$ 0.114	0.480 $\pm$ 0.105

The final area of investigation tracked the number of comparisons performed for the first sort iteration as a function of the accuracy of the comparison function. For merge sort and quick sort (as well as tree sort), the number of comparisons decreased with decreasing

accuracy. Similar to the poor behavior of the insertion sort, as the comparisons become less accurate, the  $O(n \log n)$  sort methods converge with fewer comparisons. The resulting list is likely not in correct rank order and the decreased number of comparisons is not presented here as a benefit, in and of itself, of inaccurate comparison functions.

The result in Figure B.4 demonstrates the relationship between comparison accuracy and number of comparisons. This relationship may allow one to determine the accuracy of the comparison function posteriori. For situations where one does not know the accuracy of the comparison function, or the relative accuracy of two different comparison functions, counting the number of comparisons may allow an estimate of the accuracy.

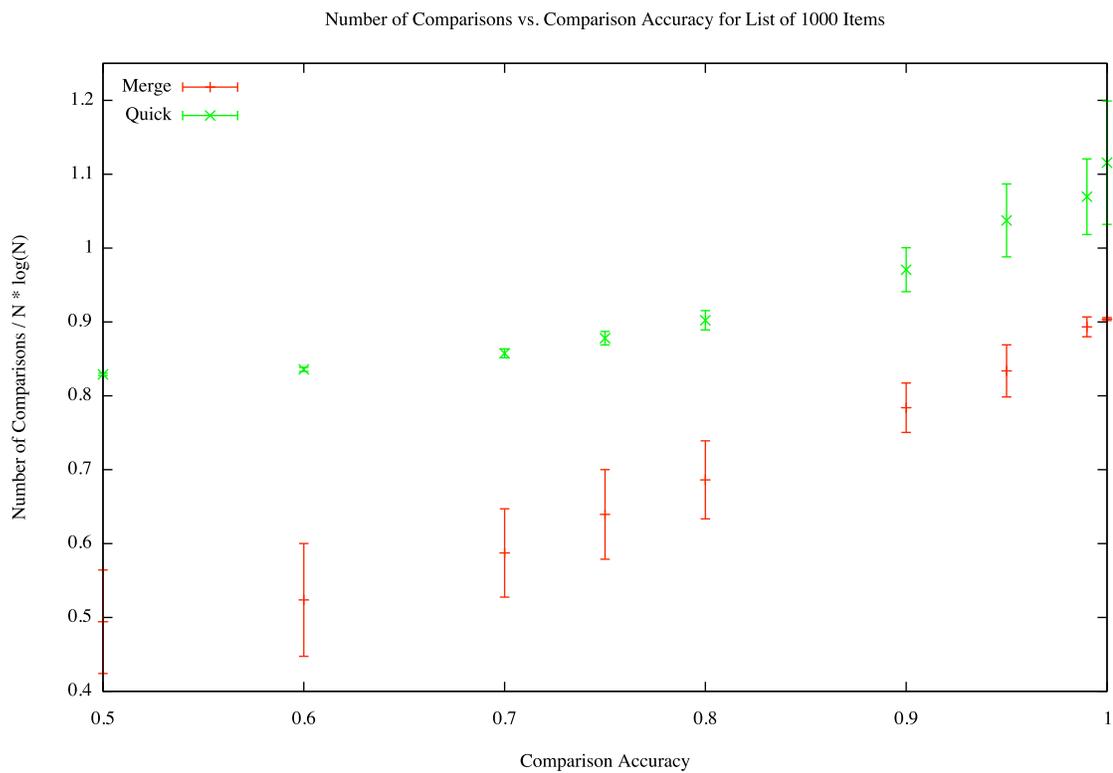


Figure B.4: Number of Comparisons vs. Comparison Accuracy

## B.5 Conclusion

In the presence of an inaccurate comparison function, the insertion sort method performs poorly and the Ranking by Scoring Methods perform best. Other sort methods perform well under the assumption of Weber-like errors in comparisons. For merge sort, the accuracy can be improved by repeating the sort of the entire list for several iterations. The accuracy of comparisons for all methods can be improved through the use of “Best of N” matches. The most consistent behavior was exhibited by the RSM algorithms that performed well for different error models and sample distributions. The RSM method out performs other methods when considering both accuracy and number of comparisons for Weber-like error comparison functions and comparison functions with random accuracies of  $\leq 0.9$ .

Additional work can be done to see if there are more ideal parameters for the Elo scoring algorithm to improve the accuracy of those results. The prospects of using the number of comparisons to estimate the accuracy of the comparison

# Appendix C: Target Preparation and Simulation Search Software

## C.1 Introduction

This guide describes how to perform the Target Preparation and Simulation Search activities of the Merger App Software. The purpose of the software is to determine likely orbit parameters that can be used to reconstruct, via simulation, the disturbed morphologies of interacting galaxies. This is accomplished by performing simple measurements on an image of the pair of galaxies. These measurements, when combined with astronomical catalog data, can be used to constrain the simulation parameters to a range of values. A search of the phase space represented by these values is conducted first with random samples used to rapidly perform simulations. The user is able to select a number of candidate matches. Through a process of review, and an optional enhancement step, the user can refine the list of candidate matches to determine the best set of simulation parameters.

### C.1.1 Running the Software

The software is written in Java<sup>TM</sup>. It requires a Java Runtime Environment, version 1.6 or later. It has mainly been tested with the Sun<sup>®</sup>(now Oracle<sup>®</sup>) virtual machine, but should work with other Java implementations as well. Once you have downloaded the software package and extracted it from the archive you should see several scripts in the directory.

To launch the target preparation script on Mac OS X, Unix, or Linux systems run the `tgt.sh` script. On windows run the `tgt.bat` script. The search software is launched via the `search.sh` or `search.bat` script.

Some common problems include not having java installed (the JRE, you don't need a browser plugin) or not having version 1.6 or later available as the default java on your system. Additionally, double-clicking the scripts from a file browser may not work. You

may have to open a terminal window or command prompt and execute the command that way to ensure the relative classpath definition works.

When the software runs, it will create a `.jspam` directory in your user home directory. In there is a preferences file called `merger.props`. This will be the default directory for downloading target information later on.

The function of the two activities are described below.

## **C.2 Target Preparation**

Preparing a target will result in the creation of several image files and a text file for simulation parameter ranges. The process includes identifying the target and obtaining an image on the “Target” tab. Next the centers of each galaxy are identified and have a distance and mass set on the “Disk Info” tab. On the “Disk Orientation” tab the disk size and orientation angles are set. On the final tab, “Parameter Ranges”, the software will estimate the likely ranges of the simulation parameters.

### **C.2.1 Target Tab**

The first step is to assign a name to this target and then pick a directory for where these files can be saved. This information is entered in the “Target Name” box and by clicking the “Pick Dir” button at the top of the software. See Figure C.1 below. Once the name is entered and the directory is picked, click Save.

The next step is to query an image server and a catalog database for information about and an image of the merger of interest. The Sloan Digital Sky Survey (SDSS) provides both of these online resources for each of its data releases. The software is able to query SDSS DR7, DR8, and DR9 (when it becomes available). The software is also capable of querying imagery from the Digitized Sky Survey (DSS) of the Space Telescope Science Institute. The catalog database used to support those images is the NASA/IPAC Extragalactic Database hosted at JPL/Caltech.

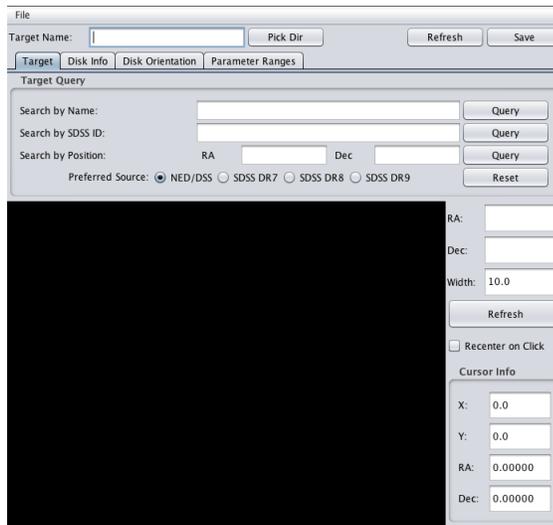


Figure C.1: The target identification screen.

To search for a target by name, type the name into the “Search by Name” box and click “Query”. You can also search the SDSS object ID or by a position on the sky. To search by position enter both Right Ascension and Declination in degrees.

Clicking “Query” will retrieve an image from the image server. DSS has image data that covers the whole sky. However, SDSS does not, so when using that as your source, the query may not return imagery.

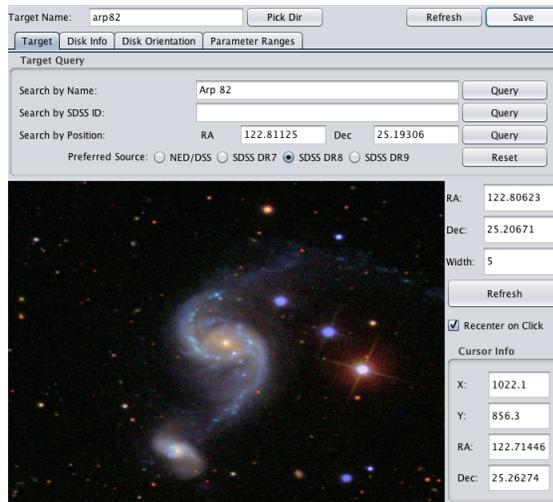


Figure C.2: Target tab with downloaded image.

Once an image has been downloaded, you can move the mouse cursor over the image to see the RA/Dec coordinates for each pixel. If you want to zoom on the image, the software will actually download a new copy of the image for you. Just update the value for the width, measured in arc minutes, and the software will download a new image with equal height and width. To recenter an image, make sure the check box next to “Recenter on click” is checked, and then the next time you click the left mouse button in the image, the software will download a new image centered at the RA/Dec that you clicked on. After you have the image centered on the primary or larger galaxy and scaled to show all of the features of both galaxies, click “Save” again. Figure 2 shows how it looks for Arp 82.

### C.2.2 Disk Info

On this tab you need to identify the center of each galaxy. Start with the primary tab selected on the left. Click the “Select” button. Now, click on the image at the center of the primary galaxy. This should be close to the center of the image. Once you’ve selected a center, the pixel coordinates and RA/Dec should appear in the coordinates boxes.

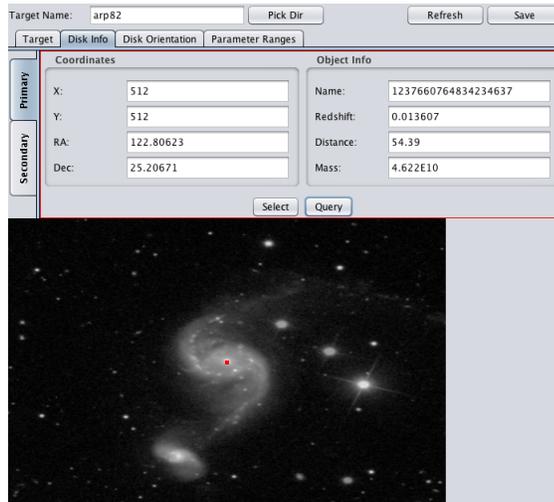


Figure C.3: Disk Info with primary center selected.

Next, click on the “Query” button to query the catalog database for information about the galaxy at that point in the sky. The software will retrieve the catalog name and redshift.

The redshift will be converted to a distance in Mpc. This distance will be combined with apparent magnitude information from the catalog to estimate absolute magnitude. From the absolute magnitude, the mass of the galaxy will be estimated. If the query does not turn up information, you can enter it by hand. Once the primary galaxy has been identified, click on the Secondary tab and repeat the “Select” and “Query” steps. The redshifts should be similar. Click “Save” to save your work.

### C.2.3 Disk Orientation

On this tab you will be able to remove the background of the image and set the size and shape of the initial disks for the simulation. Begin by clicking and dragging the “Background” slider up slowly. This will raise the threshold for the image floor. As the slider goes up, the less bright pixels are set to 0 brightness. Once you feel that most of the background directly around the galaxies has been removed, you can stop dragging the slider. Stars and other objects in the image that are not touching the galaxies will be removed later.

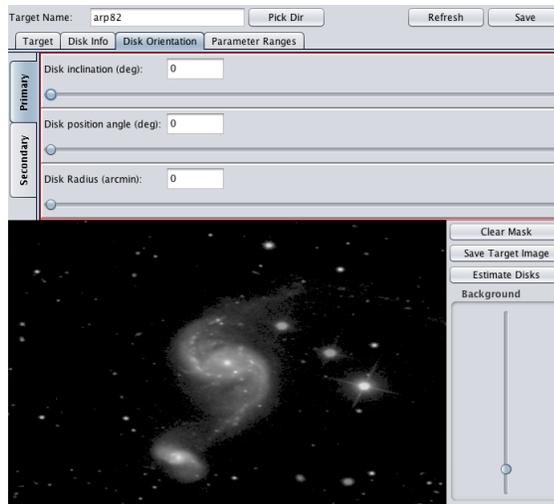


Figure C.4: Disk Orientation with background level raised.

If there is a star or other object in the image that is touching your galaxies, you can click and hold down the mouse while over the image to activate the eraser mode. The cursor will turn into an eraser, and you can use it to erase pixels from the image by moving your

mouse while holding down the button. The figure below shows where the nearby star has been erased.



Figure C.5: Image after the star to the right of the primary has been erased.

If you are satisfied that the background is correctly removed and other objects have been separated from the galaxies, you can save this target image by clicking “Save Target Image”. This saves only the pixels connected to the primary and secondary galaxies you defined.

Next click “Estimate Disks” to get the software to estimate disk shape and size for you.

You can adjust the orientation angles and disk size by using the sliders at the top. Generally the disk should be set to be a little bit smaller than the outer tidal features, like the tails in this image.

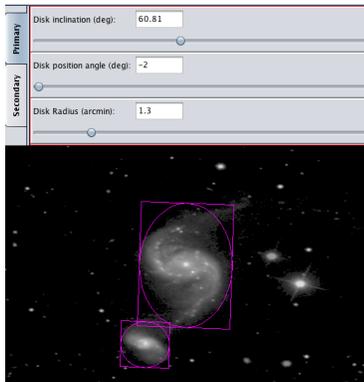


Figure C.6: Automated disk parameters.

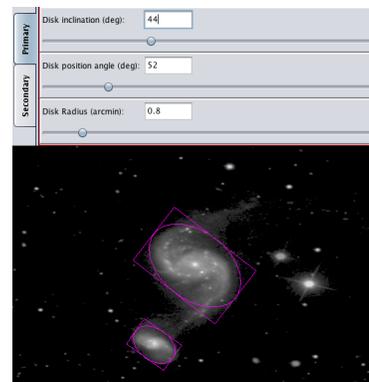


Figure C.7: Disks after user adjustments.

Once you have the disk shape and size as you like it, click “Save” at the top right.

## C.2.4 Parameter Ranges

The parameter range values are initially 0. The software will estimate ranges for you if you click “Calculate Simulation Ranges” at the bottom. Advanced users may wish to edit the ranges here. Once done, click “Save” again.

Parameter	Value	Parameter	Value
Min Rx:	-0.585323	Max Rx:	-0.585323
Min Ry:	-1.743519	Max Ry:	-1.743519
Min Rz:	-6.064874	Max Rz:	6.064874
Min Vx:	-1.91024	Max Vx:	1.91024
Min Vy:	-1.91024	Max Vy:	1.91024
Min Vz:	-1.91024	Max Vz:	1.91024
Min M1:	0.04622	Max M1:	4.622
Min M2:	0.009871	Max M2:	0.9871
Min Radius 1:	0.421904	Max Radius 1:	1.265713
Min Radius 2:	0.184583	Max Radius 2:	0.553749
Inc 1 Range:	60	Angle 1 Range:	60
Inc 2 Range:	60	Angle 2 Range:	60

Figure C.8: Parameter ranges.

With the parameter ranges set, you are now ready to begin running simulations of this pair of galaxies. If you would like to redo a previous step, go back to the desired tab and make changes. If you are unsatisfied with the results after a few tries, consider starting over by exiting and relaunching the application.

## C.3 Simulation Search

Using the initial conditions and parameter ranges you set up during target preparation, you can generate simulations in an attempt to match the shape of the galaxies in your target image. The process begins on the “Simulate” tab. Here you can view 8 simulations per batch and select candidate matches. There is a “Review” tab to allow you to see which simulations you have selected. The next tab is “Enhance” where you can try to improve upon the simulations that were randomly generated. On the “Evaluate” tab you will rank

the simulations to determine which is the best. There is also a “Statistics” tab to let you view information about your progress.

### C.3.1 Simulate

To begin running simulations, open the “File” menu and click on “Select Target”. From this dialog browse to the directory where you saved the tgt file in the previous activity and open it. You should see your target image loaded in the center square.

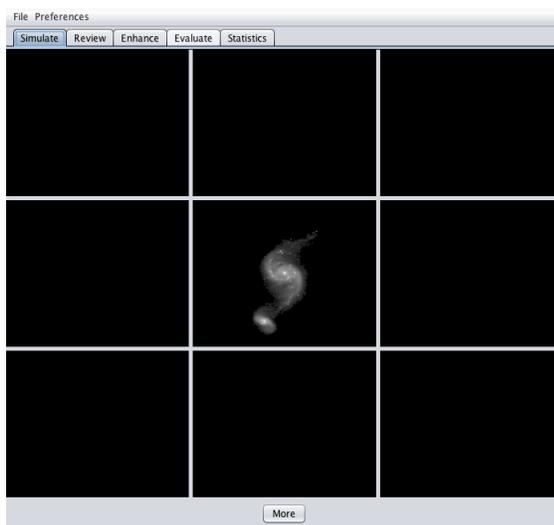


Figure C.9: Simulate tab with target loaded.

To have the software randomly select sets of simulation parameters and run them for you, click “More”. As interesting simulations appear, you can click on them. They will be highlighted with a red box and will be added to your list of selected simulations. If you change your mind about a simulation before you click “More” again, you can click the image a second time to turn off the red highlighting and remove it from your list. The goal at this stage is not perfection. Look for simulations that have at least a hint of having the proper shape, such as a tail that is too short or in the wrong position. Maybe you can find a simulation where the primary is correct, but the secondary is not. Add it to your selected list. You will need to review about 1000 simulations to develop several dozen candidate matches.



Figure C.10: Simulate screen with selected simulations.

### C.3.2 Preferences

The Preferences menu will let you edit simulation preferences. You can set the number of particles in each simulation and the size of the particles. There is an option to turn the animation on and off. You can also view the simulations in red/blue stereo if you have red/blue 3D glasses available. The “Color” tab will let you set the color for the simulation particles.

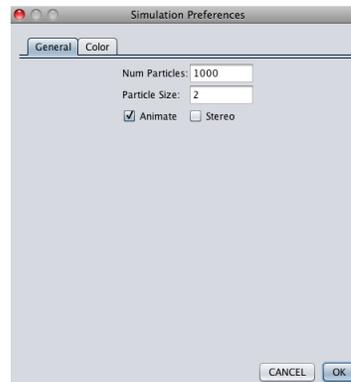


Figure C.11: Preferences.

### C.3.3 Review

The “Review” tab allows you to see the list of selected simulations. As you progress through the “Evaluate” activity, the order of simulations in the list will change so that the best simulations are at the top. To select a simulation for use on the “Enhance” tab, click on it and a yellow highlight box will appear around it.



Figure C.12: Review tab.

### C.3.4 Enhance

The “Enhance” tab allows you to alter the simulation parameters in an attempt to improve the match with the target image. On the interactive simulation screens like this one and on the “Simulate” tab, if you click and hold down the mouse you can rotate the simulation in 3D to see if from different angles. The “Reset View” button under the simulation will return it to its default position. Clicking “More ‘Stars’ ” will rerun the simulation using more particles. Clicking “Fewer ‘Stars’ ” will do the opposite. The increase in particle number is only temporary to provide an enhanced view of the simulation.

Simulation parameters are available to alter two at a time. Clicking on the various toggle buttons: Mass, Speed, Depth, Size, Angles 1, and Angles 2, will allow you to edit their values. There are tooltips available to explain which parameters are active. To adjust

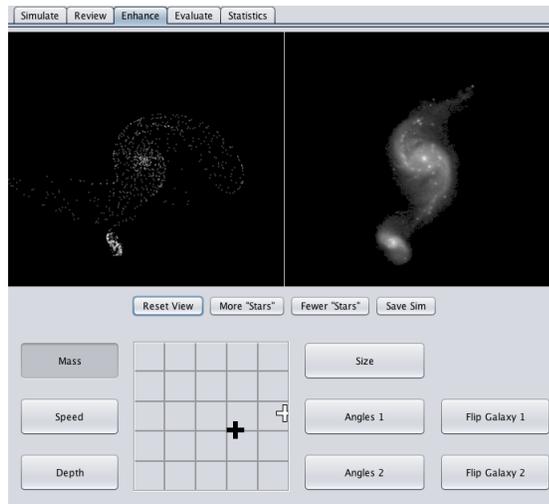


Figure C.13: Enhance tab.

the values, click on the black cross symbol, hold the mouse down, and then drag it around the gridded area. The tool tips will indicate which values are edited by moving up and down or left and right. Depending on the speed of your computer, the simulations will update immediately while you move the cursor around. Sometimes you may have to wait a few seconds after you stop moving the cursor for the simulation to catch up. The flip buttons allow you to reverse the rotation direction of galaxy 1 and 2 independently. If you think you've improved the simulation, click "Save Sim" to add it to the list of selected simulations.

### C.3.5 Evaluate

After you've selected at least 8 simulations, you can use the "Evaluate" tab to sort them by quality or fitness. This is accomplished by judging a series of head-to-head comparisons. Two simulations are displayed at a time. If you think the simulation on the left is better, click on it. If you think the one on the right is better, click on it. If you think neither simulation is particularly good, you can click on "Neither is a Good Match". Through a series of these competitions a fitness score is accumulated for each simulation. The ones that are more often judged to be the best will rise to the top.

The scoring algorithm does not have a termination condition. As you add more simulations, you will need to come back to this activity to properly evaluate them. For a given set of simulations, the more competitions you judge, the more accurate your score for them will be.



Figure C.14: Evaluate screen with Merger Wars.

### C.3.6 Statistics

The “Statistics” tab provides information on your progress. The number of random sims rejected refers to simulations that were evaluated by the system to have too small of an impulse to generate an interesting simulation and were never shown to you. The number of simulations run is how many you have seen on the “Simulate” tab. The number selected refers to how many you have clicked on. The number enhanced refers to how many you have added from the “Enhance” tab. The number evaluated refers to how many have had a score assigned. The number of merger wars competitions refers to the number of clicks you done on the “Evaluate” tab.

### C.3.7 Saving

Once you are satisfied with your results, you can save the set of simulation parameters in a text file. Each row will have a comma separated list of simulation parameters that you

Simulate		Review		Enhance		Evaluate		Statistics		
# Random Sims Rejected:	766	# Simulations Enhanced:	0							
# Simulations Run:	184	# Simulations Evaluated:	16							
# Simulations Selected:	19	# Merger Wars Competitions:	18							
States   Orbit Distribution										
Fitness	M1	M2	Rz	Vx	Vy	Vz	Q1	Q1	Q2	Q2
1.0	2.453...	0.959...	1.835...	0.553...	-1.35...	0.619...	29.59...	43.19...	56.56...	50.14...
1.0	3.644...	0.474...	-0.04...	-1.00...	0.438...	-0.64...	48.85...	72.70...	132.2...	72.37...
0.666	3.485...	0.955...	1.574...	0.963...	0.164...	0.458...	314.9...	66.97...	35.43...	56.07...
0.75	4.086...	0.824...	1.924...	0.231...	0.097...	0.103...	292.3...	36.41...	231.0...	31.04...
0.333	3.503...	0.303...	1.039...	0.620...	-1.35...	0.190...	38.12...	54.69...	212.5...	75.66...
0.5	0.777...	0.931...	2.141...	-0.56...	-0.70...	1.622...	325.2...	52.01...	321.1...	67.68...
0.333	3.186...	0.280...	-0.81...	1.389...	-0.92...	-0.54...	216.1...	64.96...	24.31...	63.30...
0.0	0.911...	0.795...	3.713...	0.014...	-0.59...	0.741...	35.96...	79.27...	344.2...	54.94...
0.0	1.247...	0.767...	0.575...	0.607...	-1.06...	1.354...	17.87...	41.77...	231.5...	45.20...
1.0	2.745...	0.767...	-2.12...	0.153...	0.015...	-0.09...	69.23...	57.38...	205.5...	41.91...
0.0	4.544...	0.921...	-0.43...	0.537...	-1.06...	0.493...	300.4...	38.64...	148.5...	67.91...
0.0	2.413...	0.466...	0.711...	0.681...	-1.67...	0.168...	344.5...	77.03...	54.74...	29.47...
1.0	4.195...	0.783...	-1.58...	0.555...	-1.35...	-1.41...	338.0...	69.06...	13.36...	44.66...
0.0	2.707...	0.326...	1.975...	0.418...	-1.51...	1.770...	306.3...	52.25...	67.03...	79.30...
0.0	3.352...	0.982...	-0.82...	0.094...	-0.82...	1.060...	322.6...	29.23...	220.7...	30.91...
0.0	4.020...	0.505...	-2.78...	-0.38...	0.147...	0.205...	307.6...	67.50...	135.4...	72.34...
0.0	3.060...	0.276...	-4.23...	-0.16...	-0.61...	-1.54...	115.9...	43.68...	23.40...	77.89...
0.0	2.059...	0.676...	2.268...	0.112...	-1.72...	1.375...	231.8...	29.87...	346.0...	37.25...
0.0	1.459...	0.222...	-1.14...	0.595...	-1.07...	-0.53...	321.9...	58.07...	340.6...	29.97...

Figure C.15: Statistics tab with table view of states

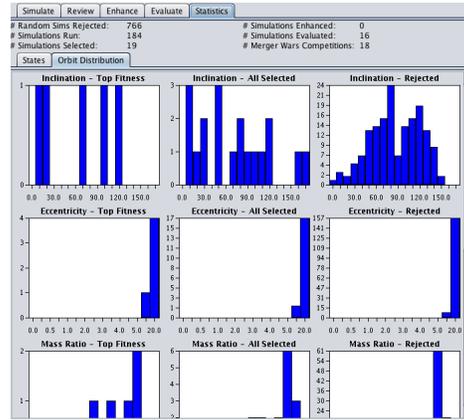


Figure C.16: Statistics tab with histogram view.

can use with the SPAM simulation software to recreate your target galaxy. The order of simulation parameters in the file is by fitness, with the best ones on top.

### C.3.8 About the Software

The software is based on the tools and interfaces he created for the Galaxy Zoo Mergers<sup>1</sup> project of the Zooniverse<sup>2</sup>. The software for computing the simulations is known as SPAM. It is described in the first two references below. The goals of the Galaxy Zoo Mergers project are discussed in the other reference. The SPAM code is available on the Astrophysics Source Code Library<sup>3</sup>. The author plans to soon make the Target Preparation and Simulation Search software available there as well.

<sup>1</sup><http://mergers.galaxyzoo.org>

<sup>2</sup><http://www.zooniverse.org>

<sup>3</sup><http://asterisk.apod.com/viewforum.php?f=35>

## Bibliography

- Antunes, A. and Wallin, J. (2001). Convergence on n-body plus SPH. In *Bulletin of the American Astronomical Society*, volume 199, page 1433.
- Antunes, A. and Wallin, J. (2007). Dynamical parameters for AM 0644741. *The Astrophysical Journal*, 670(1):261–268.
- Arp, H. (1966). Atlas of peculiar galaxies. *The Astrophysical Journal Supplement Series*, 14:1.
- Barnes, J. and Hut, P. (1986). A hierarchical  $O(n \log n)$  force-calculation algorithm. *Nature*, 324(6096):446–449.
- Barnes, J. E. (2011). Identikit 2: an algorithm for reconstructing galactic collisions. *Monthly Notices of the Royal Astronomical Society*, 413:2860–2872.
- Barnes, J. E. and Hernquist, L. (1992). Dynamics of interacting galaxies. *Annual Review of Astronomy and Astrophysics*, 30:705–742.
- Barnes, J. E. and Hibbard, J. E. (2009). IDENTIKIT 1: A modeling tool for interacting disk galaxies. *The Astronomical Journal*, 137(2):3071–3090.
- Baumgardt, H. and Makino, J. (2003). Dynamical evolution of star clusters in tidal fields. *Monthly Notices of the Royal Astronomical Society*, 340:227–246.
- Bell, E. F., McIntosh, D. H., Katz, N., and Weinberg, M. D. (2003). The optical and near-infrared properties of galaxies: I. luminosity and stellar mass functions. *The Astrophysical Journal Supplement Series*, 149:289–312.

- Bertin, E. and Arnouts, S. (1996). SExtractor: software for source extraction. *Astronomy and Astrophysics Supplement Series*, 117:393–404.
- Binney, J. and Tremaine, S. (2008). Galactic dynamics: Second edition. *Galactic Dynamics: Second Edition*, by James Binney and Scott Tremaine. ISBN 978-0-691-13026-2 (HB). Published by Princeton University Press, Princeton, NJ USA, 2008., -1.
- Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., and Price-Whelan, A. (2011). Improved background subtraction for the sloan digital sky survey images. *The Astronomical Journal*, 142:31.
- Boland, M. V. (1999). *Quantitative description and automated classification of cellular protein localization patterns in fluorescence microscope images of mammalian cells*. PhD thesis, Carnegie Mellon University.
- Borne, K. D. (1984). Interacting binary galaxies. I - a numerical model and preliminary results. *The Astrophysical Journal*, 287:503–522.
- Borne, K. D. (1988). Interacting binary galaxies. II - matching models to observations. *The Astrophysical Journal*, 330:38–50.
- Borne, K. D. and Richstone, D. O. (1991). A merger scenario for NGC 7252 - a tale of two tails. *The Astrophysical Journal*, 369:111–120.
- Bruzual A., G. and Charlot, S. (1993). Spectral evolution of stellar populations using isochrone synthesis. *The Astrophysical Journal*, 405:538–553.
- Chandrasekhar, S. and von Neumann, J. (1943). The statistics of the gravitational field arising from a random distribution of stars II. *The Astrophysical Journal*, 97:255–262.
- Ciotti, L. and Bertin, G. (1999). Analytical properties of the  $r1/m$  law. *Astronomy and Astrophysics*, 352:447–451.
- Cutri, R. M., Wright, E. L., Conrow, T., Bauer, J., Benford, D., Brandenburg, H., Dailey, J., Eisenhardt, P. R. M., Evans, T., Fajardo-Acosta, S., Fowler, J., Gelino, C., Grillmair,

- C., Harbut, M., Hoffman, D., Jarrett, T., Kirkpatrick, J. D., Leisawitz, D., Liu, W., Mainzer, A., Marsh, K., Masci, F., McCallon, H., Padgett, D., Ressler, M. E., Royer, D., Skrutskie, M. F., Stanford, S. A., Wyatt, P. L., Tholen, D., Tsai, C. W., Wachter, S., Wheelock, S. L., Yan, L., Alles, R., Beck, R., Grav, T., Masiero, J., McCollum, B., McGehee, P., Papin, M., and Wittman, M. (2012). Explanatory supplement to the WISE all-sky data release products. *Explanatory Supplement to the WISE All-Sky Data Release Products*.
- Darg, D. W., Kaviraj, S., Lintott, C. J., Schawinski, K., Sarzi, M., Bamford, S., Silk, J., Proctor, R., Andreescu, D., Murray, P., Nichol, R. C., Raddick, M. J., Slosar, A., Szalay, A. S., Thomas, D., and Vandenberg, J. (2009). Galaxy zoo: the fraction of merging galaxies in the SDSS and their morphologies. *arXiv:0903.4937*.
- De Jong, K. A. (2006). *Evolutionary Computation - A Unified Approach*. MIT Press, 1st edition.
- Elo, A. E. (1978). *The Rating Of Chess Players, Past & Present*. Arco Pub., 1st edition.
- Hammer, F., Flores, H., Puech, M., Yang, Y. B., Athanassoula, E., Rodrigues, M., and Delgado, R. (2009). The hubble sequence: just a vestige of merger events? *Astronomy and Astrophysics*, 507:1313–1326.
- Hernquist, L. (1993). N-body realizations of compound galaxies. *The Astrophysical Journal Supplement Series*, 86:389–400.
- Holincheck, A., Wallin, J., Borne, K., Lintott, C., Smith, A., Bamford, S., and Fortson, L. (2010a). Tasking citizen scientists from galaxy zoo to model galaxy collisions: Preliminary results, interface, analysis. In *Galaxy Wars: Stellar Populations and Star Formation in Interacting Galaxies*, volume 423, page 223.
- Holincheck, A., Wallin, J., Borne, K., Lintott, C. J., and Smith, A. (2010b). Building a catalog of dynamical properties of interacting galaxies in SDSS with the aid of citizen scientists. In *Bulletin of the American Astronomical Society*, volume 215, page 383.

- Hopkins, P. F., Cox, T. J., Hernquist, L., Narayanan, D., Hayward, C. C., and Murray, N. (2013). Star formation in galaxy mergers with realistic models of stellar feedback and the interstellar medium. *Monthly Notices of the Royal Astronomical Society*, 430:1901–1927.
- Hüllermeier, E., Fürnkranz, J., Cheng, W., and Brinker, K. (2008). Label ranking by learning pairwise preferences. *Artificial Intelligence*, 172(16-17):1897–1916.
- Kauffmann, G., Heckman, T. M., White, S. D. M., Charlot, S., Tremonti, C., Brinchmann, J., Bruzual, G., Peng, E. W., Seibert, M., Bernardi, M., Blanton, M., Brinkmann, J., Castander, F., Csabai, I., Fukugita, M., Ivezić, Z., Munn, J. A., Nichol, R. C., Padmanabhan, N., Thakar, A. R., Weinberg, D. H., and York, D. (2003). Stellar masses and star formation histories for 105 galaxies from the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society*, 341:33–53.
- Kennicutt, Robert C., J. (1998). Induced star formation. In Friedli, D., Martinet, L., and Pfenniger, D., editors, *Galaxies: Interactions and Induced Star Formation*, volume 26 of *Saas-Fee Advanced Courses*, pages 1–104. Springer Berlin Heidelberg.
- Knuth, D. E. (1973). *The Art of Computer Programming, Volume III: Sorting and Searching*. Addison-Wesley, Reading, MA.
- Lampton, M. (1997). Damping-undamping strategies for the levenberg-marquardt nonlinear least-squares method. *Comput. Phys.*, 11(1):110115.
- Lintott, C., Schawinski, K., Bamford, S., Slosar, A., Land, K., Thomas, D., Edmondson, E., Masters, K., Nichol, R., Raddick, J., Szalay, A., Andreescu, D., Murray, P., and Vandenberg, J. (2011). Galaxy Zoo 1 : Data release of morphological classifications for nearly 900,000 galaxies. *Monthly Notices of the Royal Astronomical Society*, 410:166–178.
- Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., Murray, P., and Vandenberg, J. (2008). Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society*, 389:1179–1189.

- Lupton, R., Gunn, J. E., Ivezi, Z., Knapp, G. R., Kent, S., and Yasuda, N. (2001). The SDSS imaging pipelines. In *Astronomical Data Analysis Software and Systems X*, volume 238 of *ASP Conference Series*, page 269.
- Mihos, C. (2002). Galaxy crash JavaLab. <http://burro.cwru.edu/JavaLab/GalCrashWeb/>.
- Navarro, J. F., Frenk, C. S., and White, S. D. M. (1996). The structure of cold dark matter halos. *The Astrophysical Journal*, 462:563.
- Petsch, H. P. and Theis, C. (2008). Modeling merging galaxies using MINGA: improving restricted n-body by dynamical friction. *Astronomische Nachrichten*, 329:1046–1049.
- Portegies Zwart, S. F., Belleman, R. G., and Geldof, P. M. (2007). High-performance direct gravitational n-body simulations on graphics processing units. *New Astronomy*, 12(8):641–650.
- Pyle, D. (1999). *Data Preparation for Data Mining*. Morgan Kaufmann, San Francisco, CA.
- Scarpa, R. (2006). Modified newtonian dynamics, an introductory review. In *1st Crisis in Cosmology Conference*, volume 822 of *AIP Conference Proceedings*, pages 253–265.
- Scudder, J. M., Ellison, S. L., Torrey, P., Patton, D. R., and Mendel, J. T. (2012). Galaxy pairs in the sloan digital sky survey - v. tracing changes in star formation rate and metallicity out to separations of 80 kpc. *Monthly Notices of the Royal Astronomical Society*, 426:549–565.
- Shamir, L., Orlov, N., Eckley, D. M., Macura, T., Johnston, J., and Goldberg, I. G. (2008). Wndchrm an open source utility for biological image analysis. *Source Code for Biology and Medicine*, 3(1):13. PMID: 18611266.
- Shamir, L., Wallin, J., and Holincheck, A. (2013). Automatic quantitative morphological analysis of interacting galaxies. *Submitted for publication*.

- Shi, F., Kong, X., Wicker, J., Chen, Y., Zi-Qiang, Gong, Dong-Xin, and Fan (2012). Star formation rate indicators in wide-field infrared survey preliminary release. *Journal of Astrophysics and Astronomy*, 33:213–220.
- Silverstein, D. A. and Farrell, J. E. (2001). Efficient method for paired comparison. *Journal of Electronic Imaging*, 10(2):394.
- Simard, L., Mendel, J. T., Patton, D. R., Ellison, S. L., and McConnachie, A. W. (2011). A catalog of Bulge+Disk decompositions and updated photometry for 1.12 million galaxies in the sloan digital sky survey. *The Astrophysical Journal Supplement Series*, 196:11.
- Smith, B. J., Carver, D. C., Pfeiffer, P., Perkins, S., Barkanic, J., Fritts, S., Southerland, D., Manchikalapudi, D., Baker, M., Luckey, J., Franklin, C., Moffett, A., and Struck, C. (2010). The automatic galaxy collision software. In *Galaxy Wars: Stellar Populations and Star Formation in Interacting Galaxies*, volume 423, page 227.
- Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., and Pearce, F. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636.
- Struck, C. (1999). Galaxy collisions. *Physics Reports*, 321:1–137.
- Struck, C. (2006a). Galaxy collisions - dawn of a new era. In *Astrophysics Update 2*, page 115. Praxis Publishing Ltd.
- Struck, C. (2006b). Triggering and tempering star formation in colliding galaxies. In Charmandaris, V., Rigopoulou, D., and Kylafis, N., editors, *Studying Galaxy Evolution with Spitzer and Herschel*.
- Theis, C. and Harfst, S. (2000). Modeling interacting galaxies using a parallel genetic algorithm. In *Dynamics of Galaxies: from the Early Universe to the Present*, volume 197, page 357.

- Theis, C. and Kohle, S. (2001). Multi-method-modeling of interacting galaxies. I. a unique scenario for NGC 4449? *Astronomy and Astrophysics*, 370:365–383.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4):273–286.
- Toomre, A. and Toomre, J. (1972). Galactic bridges and tails. *The Astrophysical Journal*, 178:623–666.
- Wahde, M. (1998). Determination of orbital parameters of interacting galaxies using a genetic algorithm. description of the method and application to artificial data. *Astronomy and Astrophysics Supplement Series*, 132:417–429.
- Wahde, M. and Donner, K. J. (2001). Determination of the orbital parameters of the m 51 system using a genetic algorithm. *Astronomy and Astrophysics*, 379:115–124.
- Wallin, J., Holincheck, A., Borne, K., Lintott, C., Smith, A., Bamford, S., and Fortson, L. (2010). Tasking citizen scientists from galaxy zoo to model galaxy collisions. In *Galaxy Wars: Stellar Populations and Star Formation in Interacting Galaxies*, volume 423 of *ASP Conference Series*, page 217.
- Wallin, J. F. (1990). Dynamical and photometric models of star formation in tidal tails. *The Astronomical Journal*, 100:1477–1488.
- Wallin, J. F., Schombert, J. M., and Struck-Marcell, C. (1990). Observations and models of star formation in the tidal features of interacting galaxies. In *Paired and Interacting Galaxies: International Astronomical Union Colloquium No. 124*, volume 3098, pages 727–730.
- West, A. A., Garcia-Appadoo, D. A., J. Dalcanton, J., Disney, M. J., Rockosi, C. M., Ivezić, Z., Bentz, M. C., and Brinkmann, J. (2010). HI selected galaxies in the sloan digital sky survey I: Optical data. *The Astronomical Journal*, 139(2):315–328.

- Willett, K. (2011). Update on the "Violin clef" merger: redshifts and merger zoo. <http://blog.galaxyzoo.org/2011/10/24/update-on-the-violin-clef-merger-redshifts-and-merger-zoo/>.
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., Ressler, M. E., Cutri, R. M., Jarrett, T., Kirkpatrick, J. D., Padgett, D., McMillan, R. S., Skrutskie, M., Stanford, S. A., Cohen, M., Walker, R. G., Mather, J. C., Leisawitz, D., Gautier, T. N., McLean, I., Benford, D., Lonsdale, C. J., Blain, A., Mendez, B., Irace, W. R., Duval, V., Liu, F., Royer, D., Heinrichsen, I., Howard, J., Shannon, M., Kendall, M., Walsh, A. L., Larsen, M., Cardon, J. G., Schick, S., Schwalm, M., Abid, M., Fabinsky, B., Naes, L., and Tsai, C.-W. (2010). The wide-field infrared survey explorer (WISE): mission description and initial on-orbit performance. *The Astronomical Journal*, 140:1868–1881.
- Yan, L., Donoso, E., Tsai, C.-W., Stern, D., Assef, R. J., Eisenhardt, P., Blain, A. W., Cutri, R., Jarrett, T., Stanford, S. A., Wright, E., Bridge, C., and Riechers, D. A. (2013). Characterizing the mid-infrared extragalactic sky with WISE and SDSS. *The Astronomical Journal*, 145:55.

## Curriculum Vitae

Anthony Holincheck lives with his wife and children in Northern Virginia. He works as a software developer providing tools to help customers manage, and make sense out of, their large data sets.