Understanding the Nature of Dark Matter Halos and Galaxy Formation Through the

Lens of the Milky Way

by

## **Catherine Fielder**

Bachelor of Science, Texas Tech University, 2014

Master of Science, University of Pittsburgh, 2016

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

**Doctor of Philosophy** 

University of Pittsburgh

2022

#### UNIVERSITY OF PITTSBURGH

## KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Catherine Fielder

It was defended on

June 29th 2022

and approved by

Jeffrey Newman, Department of Physics and Astronomy, University of Pittsburgh Andrew Zentner, Department of Physics and Astronomy, University of Pittsburgh Rachel Bezanson, Department of Physics and Astronomy, University of Pittsburgh Michael Hatridge, Department of Physics and Astronomy, University of Pittsburgh

Matthew Walker, Department of Physics, Carnegie Mellon University

Copyright  $\bigodot$  by Catherine Fielder 2022

# Understanding the Nature of Dark Matter Halos and Galaxy Formation Through the Lens of the Milky Way

Catherine Fielder, PhD

University of Pittsburgh, 2022

The Universe is full of a zoo of galaxies that serve as beacons tracing the underlying structure of dark matter in the Universe. Galaxies exhibit a diverse array of structural properties and observed brightnesses, all of which provide clues to their evolution and the underlying structure of space. The galaxy in which we reside, the Milky Way, provides an exclusive opportunity for testing models of galaxy formation and cosmology courtesy of our proximity to the stars and dust that shape the Galaxy. Data-model comparisons of Milky Way observations to simulations, and observations of Milky Way-like galaxies are a compelling way to improve our understanding of galaxy formation and evolution.

However, we still currently lack insight into how the Milky Way fits in with the general galaxy population, an issue stemming from our location within the disk of the Milky Way. Dust blocks a large portion of our view of the Milky Way and we are unable to capture bulk integrated properties like we can for other galaxies. The focus of this thesis is on utilizing state-of-the-art statistical techniques and models to surpass these issue, addressing discrepancies in both simulations and observations, to improve our understanding of the Milky Way and how it connects to other galaxies as a means for answering questions on galaxy evolution.

# **Table of Contents**

Ack	now	ledgements	xviii
1.0	Inti	roduction	1
	1.1	Background	1
		1.1.1 The Standard Cosmological Model	1
		1.1.2 A Brief History of (Dark) Matter and Large Scale Structure	3
		1.1.3 Galaxy Formation and Evolution	5
		1.1.4 The Milky Way, A Laboratory for Galaxy Evolution	7
		1.1.5 Connecting the Galactic and Extra-Galactic	9
	1.2	Structure of Thesis	11
2.0	Illu	minating Dark Matter Halo Density Profiles Without Subhaloes	15
	2.1	Introduction	15
	2.2	Data	17
		2.2.1 Zoom-In Simulations of Two Mass Regimes	18
		2.2.2 Density Profiles and Subhalo Removal	19
	2.3	Analysis	21
		2.3.1 Density Profile Parameterization	23
		2.3.2 Profile Fitting Procedure	26
		2.3.3 Assessing Fits	29
	2.4	Results	32
		2.4.1 Identifying Best Fit Forms for Stacked Profiles	32
		2.4.1.1 Effective Power Law Index	35
		2.4.2 Impact of Subhaloes on Individual Halo Profiles	38
		2.4.3 The Concentration–Mass Relation With and Without Subhaloes $\ . \ .$	44
		2.4.4 Robustness Checks and Caveats	48
	2.5	Summary and Conclusion	51
	2.6	Other Halo Profiles	54

	2.7	Subhalo Removal with ROCKSTAR	55
	2.8	Best Fit Profile Values	59
3.0	Pre	dictably Missing Satellites: Subhalo Abundances in Milky Way-like haloes .	61
	3.1	Introduction	61
	3.2	Milky Way Halo Properties	66
		3.2.1 Zoom-in Simulations	66
		3.2.2 Concentration	67
		3.2.3 Spin	68
		3.2.4 Shape	70
		3.2.5 Halo Merger History	71
		3.2.6 The Milky Way Halo Compared to Other Dark Matter haloes	71
	3.3	Subhalo Abundances in Milky Way-Like haloes	74
		3.3.1 Halo Properties and Subhalo Abundances	74
		3.3.2 Predicting Milky Way Subhalo Abundances	79
		3.3.3 Halo Properties and Subhalo Scaling Relations	93
	3.4	Conclusion	94
	3.5	Concentration and Adiabatic Contraction	100
	3.6	Numerical and Mathematical Techniques	104
		3.6.1 Regression Modeling	104
		3.6.2 Nearest Neighbors	105
		3.6.3 Model Selection	106
		3.6.4 Maximum Likelihood Fitting of a Power-Law Model for Subhalo Abun-	
		dance	114
		3.6.5 Fitting Functions for the Subhalo Cumulative Velocity Function	121
4.0	Cor	nstraining the Milky Way's Ultraviolet to Infrared SED with Gaussian Pro-	
	cess	s Regression	124
	4.1	Introduction	124
	4.2	Observational Data	129
		4.2.1 Photometry	129
		4.2.1.1 SDSS Galaxies	129

		4.2.1.2	GALEX–SDSS–WISE Legacy Catalog	131
		4.2.1.3	DESI Legacy Imaging Surveys	131
	4.2.2	SDSS-h	pased Properties	132
		4.2.2.1	MPA-JHU Masses and Star Formation Rates	132
		4.2.2.2	Simard et al. Bulge and Disk Decompositions	133
		4.2.2.3	Galaxy Zoo 2 Bar Presence	134
	4.2.3	Milky V	Way Properties	136
4.3	Gaus	sian Pro	ocess Regression for Predicting Milky Way Photometry	139
	4.3.1	Limitat	tions of Using Analogue Galaxies	139
	4.3.2	Gaussia	an Process Regression: A Powerful Method for Interpolation and	
		Predict	ion	142
		4.3.2.1	Choice of Kernel for GPR	144
		4.3.2.2	Optimizing Training Samples	145
		4.3.2.3	Comparison to Results from Analogue Samples	145
		4.3.2.4	Characterizing Sources of Uncertainty	146
		4.3.2.5	Summary of the GPR Algorithm for Determining Milky Way	
			Photometric Properties	150
4.4	Resul	lts		153
	4.4.1	The M	ilky Way Compared to the Broader Galaxy Population	154
		4.4.1.1	Optical colors	155
		4.4.1.2	UV colors	160
		4.4.1.3	Infrared/WISE colors	162
	4.4.2	The M	ultiwavelength Spectral Energy Distribution of the Milky Way .	165
		4.4.2.1	Algorithm for Calculating the SED for the Milky Way	165
		4.4.2.2	Interpreting the SED of the Milky Way	169
		4.4.2.3	Impact of Physical Parameters on the Estimated SED of the	
			Milky Way	175
		4.4.2.4	Exploration of Other Sources of Physical Parameter Measure-	
			ments	179
4.5	Sumr	nary and	d Conclusions	182

		4.5.1 Summary	182	
		4.5.2 Discussion: The Milky Way as a Red Spiral	183	
		4.5.3 Outlook	185	
	4.6	Data Summary	187	
		4.6.1 Galaxy Properties	187	
		4.6.2 Tabulated Photometric Predictions for the Milky Way	188	
	4.7	Testing the Accuracy of Our Gaussian Process Regression Methods	190	
	4.8	<i>K</i> -corrections for WISE Bands	194	
		4.8.1 Calculating the $K$ -correction $\ldots \ldots \ldots$	194	
		4.8.2 Photometric Offsets Between SDSS and DESI Legacy Survey Imaging	198	
	4.9	Eddington Bias Corrections	201	
5.0	Em	pirically Driven K-corrections At Low Redshift	203	
	5.1	Introduction	203	
	5.2	Methods	208	
		5.2.1 Observational Data	208	
		5.2.2 <i>K</i> -corrections from Previous Work $\ldots \ldots \ldots \ldots \ldots \ldots \ldots$	209	
		5.2.3 Deriving Data Driven $K$ -corrections $\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$	211	
	5.3	Results	221	
		5.3.1 Distributions of Rest-frame Colors	221	
		5.3.2 Extrapolations to $z=0$	226	
		5.3.3 Spectral Energy Distributions	228	
	5.4	Conclusions	233	
	5.5	Tables of Polynomial Functions	236	
6.0	Cor	clusion	239	
	6.1	Summary of Previous Chapters	239	
	6.2	Potential Future Work	241	
App	oend	x. Modern Milky Way Analog Searches	245	
Bibliography 249				

# List of Tables

2.1	Table of convenient functions for the NFW, generalized NFW, Einasto,	
	and generalized Einasto profiles explored in this work	25
2.2	Statistics for evaluating goodness-of-fit and model suitability for the	
	NFW profile, generalized NFW profile, Einasto profile, and the gener-	
	alized Einasto profile.	34
2.3	Tables of the median concentrations and scatter for the simulated haloes.	47
2.4	Fit values for the density profiles explored in this work	60
3.1	Table of Spearman correlation coefficients, $\rho$ , and the corresponding $p$ -	
	values, $p$ , amongst all of the host halo properties considered in our analysis.	76
3.2	Table of $68\%$ and $95\%$ confidence intervals for the total abundance of	
	subhaloes above the resolution limit ( $V_{\rm max}^{\rm frac} > 0.065$ ) predicted for the	
	Milky Way	89
3.3	Table of Akaike Information Criterion (AIC) Bayesian Information Crite-	
	rion (BIC), $\chi^2$ , and $\sigma_{\rm RMS}$ values for the models used to estimate subhalo	
	abundance for the Milky Way within this work.	109
4.1	The passbands and corresponding power and uncertainties for the pre-	
	dicted SED of the Milky Way	168
4.2	X - r color estimates for each photometric band X used in this work	191
4.3	Additional color estimates for the Milky Way, updating values from Table	
	1 and Table 3 of Ref. [219] in addition to values for WISE colors 1 $$	91
4.4	Absolute magnitude estimates for the Milky Way, updating Table 1 and	
	Table 3 of Ref. [219] with additional absolute magnitude estimates be-	
	yond the optical	192

# List of Figures

2.1	Visualization of subhalo removal in a 2D projection for one of the Milky	
	Way-mass host haloes.	22
2.2	Analytic density profile derivatives for a Milky Way-mass (MMMZ) type	
	halo	27
2.3	The density profile of Halo $829$ in the Milky Way-mass (MMMZ) simu-	
	lations, the same halo shown in Fig. 2.1.	30
2.4	The stacked density profile from all 45 Milky Way-mass (MMMZ) host	
	haloes and 96 RHAPSODY haloes, including fit residuals for the NFW,	
	generalized NFW, Einasto and generalized Einasto profile	36
2.5	Effective power law index of the density profiles as a function of radius.	39
2.6	The effective power law indices of the density profiles as a function of	
	radius for the NFW profile, generalized NFW profile, Einasto profile,	
	and generalized Einasto profile.	40
2.7	Three individual halo profiles and their respective fits from the MW-mass	
	simulations and cluster mass simulations	42
2.8	The $\chi^2$ of the individual halo fits to the generalized Einasto profile with	
	subhaloes (the x-axis) plotted against the $\chi^2$ of the individual halo fits	
	without subhaloes (the y-axis)	43
2.9	The concentration of the MW-mass MMMZ haloes and the cluster-mass	
	RHAPSODY haloes calculated from individual fits to the generalized	
	Einasto profile and the NFW profile.	46
2.10	Visualization of subhalo removal in a 2D projection for an alternative	
	method for excluding subhaloes	57
2.11	A stacked density profile of all 45 of the MW-mass MMMZ host haloes	
	and 96 RHAPSODY haloes, including and alternate subhalos excluded	
	definition.	58

3.1	Plots of all possible combinations of the host halo parameters investigated	
	in this work.	73
3.2	Cumulative velocity functions $N_{\rm sat}(> V_{\rm max}^{\rm sat}/V_{\rm max}^{\rm host} = V_{\rm max}^{\rm frac})$ of subhaloes	
	for samples split according to various host halo properties. $\ldots$ .	80
3.3	Cumulative velocity function for the Milky Way from a one-parameter	
	scaling relation model which predicts subhalo abundance as a function	
	of halo concentration and accompanying residual	84
3.4	Cumulative velocity function for the Milky Way from a three-parameter	
	scaling relation model which predicts subhalo abundance as a function	
	of halo concentration, spin, and shape, and accompanying residual	85
3.5	Cumulative velocity function for the Milky Way from a three-parameter	
	scaling relation model which predicts subhalo abundance as a function	
	of halo scale factor at the last major merger, spin, and shape, and ac-	
	companying residual.	86
3.6	A CVF of one parameter models depicted with all possible theoretical	
	values of concentration for haloes at the same mass and morphology as	
	the Milky Way.	87
3.7	A CVF of one parameter models depicted with estimated values of Milky	
	Way host halo parameters	88
3.8	Histograms of predicted total cumulative subhalo abundance below the	
	resolution limit normalized by the mean measured subhalo abundance	91
3.9	Rotation curves based on the Gnedin contraction model for two initial	
	NFW concentration values of 10 and 20.	102
3.10	The contracted concentration as a result from our fit as a function of the	
	initial NFW concentration values input into CONTRA	103
3.11	Residual subhalo abundance as a function of the parent halo's NFW	
	concentration parameter	110
3.12	Residual subhalo abundance as a function of the parent halo's spin, con-	
	centration, and shape	111

3.13	The distribution of the coefficient of $c_{\rm NFW}$ from linear regression derived	
	from one thousand bootstraps with replacement of the simulated halo	
	samples.	113
3.14	Distributions of parameter coefficients from bootstrap re-samplings of	
	the subhalo abundance data for the three-parameter regression	115
3.15	Illustration of the three-parameter power-law model for subhalo abun-	
	dance as a function of the parent halo's spin, concentration, and shape.	117
3.16	Distributions of parameter exponents from bootstrap re-samplings of the	
	subhalo abundance data for the three-parameter power-law Poisson max-	
	imum likelihood fit	119
4.1	Total number of galaxies within an allowed tolerance $(N_{\rm gal} < N_{\sigma})$ as a	
	function of the maximum deviation away from the Galactic values allowed	
	for each parameter, in units of the uncertainty in the Milky Way value	
	of that parameter, $N_{\sigma}$	141
4.2	A comparison of galaxy samples and results from the Milky Way ana-	
	logues method versus Gaussian Process regression.	147
4.3	A breakdown of uncertainty due to scatter at fixed properties and scatter	
	due to Milky Way measurement uncertainty.	151
4.4	Contributions to the variance in rest-frame ${}^0(g-r)$ color for Gaussian	
	process regression employing varying sets of galaxy physical parameters.	152
4.5	Rest-frame optical color as a function of stellar mass	157
4.6	Rest-frame optical color as a function of log specific star formation rate	
	$(\log (SFR/M_*))$ in units of $yr^{-1}$ .	159
4.7	As Fig. 4.6, but for UV-optical colors	161
4.8	WISE color-color diagrams for both our parent sample and the predicted	
	results for the Milky Way from Gaussian process regression	164
4.9	The predicted SED for the Milky Way from a six-parameter Gaussian	
	process regression.	170

4.10	Postage stamp images corresponding to galaxies within the Ref. $[50]$ SED	
	atlas with $\chi^2$ values below the 95% upper limit when compared to the	
	estimated SED of the Milky Way.	171
4.11	Isolating the contributions of each physical parameter to the SED. Each	
	panels shows the effect of varying one parameter while fixing the other	
	five parameters to the fiducial Milky Way values	176
4.12	Distribution of the various galaxy physical properties used to predict	
	the Milky Way SED, viewed as both two-dimensional projections and	
	one-dimensional histograms.	189
4.13	Histograms depicting the difference between the actual color or abso-	
	lute magnitude of galaxies in an independent test sample from the value	
	predicted by GPR, normalized by the predicted error resulting from the	
	GPR fit $(\frac{\Delta}{\sigma})$	195
4.14	Offsets between SDSS and DESI Legacy Survey rest-frame $r$ -band abso-	
	lute magnitude, plotted as a function of SDSS rest-frame color, $^{0}(g-r)$ .	200
5.1	Observed $(r - i)$ color plotted as a function of redshift for our galaxy	
	sample. We have color coded the points by rest-frame ${}^0(g-r)$ color	
	where purple corresponds to more blue galaxies and red corresponds to	
	more red galaxies. It is evident that observed color is correlated with	
	rest-frame color (and redshift). By obtaining the fit coefficients in bins	
	of the rest-frame color (see Fig. 5.2) we can quantify how the observed	
	color changes relative to the rest-frame color across redshift space. $\ .$ .	213
5.2	An example of our first order polynomial fits. Here $i$ -band is the band we	
	desire to $K$ -correct and $r$ -band serves as the anchor. We plot the observed	
	(r-i) color as a function of redshift for our sample after it has been split	
	into 22 bins in ${}^0(g-r)$ with ~ 5435 objects per bin. Points from 3 of	
	these bins are plotted, which are labeled according to the average $^0(g-r)$	
	color per bin. The black dashed lines represent the fits to Equation $40$	
	for each bin plotted, performed using a Huber regression. The fits are a	
	sensible representation of the data and not swayed by outliers. $\ldots$ .	214

5.3Linear term coefficients  $(a_1)$  from the initial fit (i.e., the fit of observed color as a function of redshift in bins of rest-frame color,  $^{0}(g-r))$  as function of the central  ${}^{0}(g-r)$  of each bin are plotted as blue points. For reference, three of the initials fits from which the  $a_1$ 's are derived are depicted in Fig. 5.2. We perform a linear fit for  $a_1$  as a function of the center of our  ${}^{0}(g-r)$  bins with a Huber regression (Equation 42), which is plotted here as a black dashed line. A linear fit is a sufficient approximation due to the relatively small scatter. None of the  $a_1$  points in this band were excluded from the fit but the NMAD cutoff, but the Huber regression is robust to the reddest point. 2155.4Histograms of rest-frame colors comparing K-correction methods for GALEX FUV/NUV, SDSS uriz, and 2MASS J/H/Ks bands. We show our calculated rest-frame colors derived from KCORRECT (blue line) and GSWLC-M2 (orange line) and those calculated using the KCORRECT software (blue shaded) or in the GSWLC-M2 catalog (orange shaded). In general, our rest-frame colors derived from both KCORRECT and GSWLC-M2 are in excellent agreement with each other. In the 2MASS bands, our rest-frame colors derived from KCORRECT and GSWLC-M2 are self-consistent and in agreement with the KCORRECT colors; however, the GSWLC-M2 colors show more peaked distributions likely caused by uncertainties in modeling unusual older stellar populations (TP-AGB) stars, post-AGB stars, and extreme HB stars). The offsets in  ${}^{0}(g-r)$  are likely due to the fact that our rest-frame colors are derived from  $^{0}(g-r)$ for both catalogs. Overall, our rest-frame colors for these bands are wellmatched to each other and to those from KCORRECT and GSWLC-M2.

. 223

5.5Same as Fig. 5.4 but for the WISE bands. GSWLC-M2 did not derive WISE colors, but we still derive rest-frame WISE colors using  $^{0}(g - r)$ from both KCORRECT and GSWLC-M2. Our results derived from either source agree well with each other. Additionally, histograms match well with those of KCORRECT, except for the W4 band, which we suspect is a result of a template error in KCORRECT. This discrepancy in W4 highlights an advantage of deriving K-corrections independent of templates 2245.6Contour plots of galaxies in two narrow redshift bins 0.04 < z < 0.045(blue) and 0.075 < z < 0.08 (orange). We plot a variety of colors across our wavelength range as a function of  ${}^{0}(g-r)$ . The left column shows these colors in the observed-frame and the right shows these colors in the rest-frame, using our K-correction results. We plot narrow redshift bins as the observed SEDs of the galaxies should be approximately the same. It is evident that for all colors the contours of both redshift bins align closely in the rest-frame in contrast to the observed-frame. This is the expected result for successful K-corrections and similar results are observed in KCORRECT and GSWLC-M2 (save for (W4 - r)). . . . . 2275.7Plots of the mean difference between observed and rest-frame color (e.g.,  $\Delta(u-r) = (u-r) - (u-r)$  as a function of 12 redshift bins for optical colors. Blue points depict our K-corrected results with a linear initial fit, and orange points depict results for a quadratic fit. Grey points show results from the corresponding source denoted at the column title. The colored dashed lines depict a linear extrapolation to redshift 0. Our results all approach the origin closely, as they should by construction and under the expectation that are redshift 0 the K-correction should be 0. However, this is not necessarily the case for K-corrections that rely on 2295.8Same as Fig. 5.7 but for UV and near-IR bands. Our results perform just as well as if not better than those from SED fitting. . . . . . . . . 230 5.9Derived SEDs for two galaxies from our sample which share the same  ${}^{0}(g-r) = 0.667$  but reside at two different redshifts. The lower redshift galaxy is plotted with blue points and offset from the higher redshift galaxy in the y-axis for clarity. Stars are plotted at the observed fluxes. Triangles are plotted at the luminosity determined from the rest-frame magnitudes derived via KCORRECT (left panel) and GSWLC-M2 (right panel). Circles with error bars show our resulting luminosities from our K-correction method. As a means of comparison we plot the spectrum of NGC 4138 from the Ref. [50] galaxy SED atlas in gray, with dashed lines showing modeled portions of the spectra and solid lines showing the observed portions of the spectra. This template had the lowest  $\chi^2$ of the difference between the photometry of the two galaxies shown and the template (formalized in r-band). While we do not expect a direct match to a template, we provide it as a means of comparison to help guide the eye. We can conclude that our K-correction method provides reasonable estimates for even the weakly constrained WISE bands, where KCORRECT fails in W4 and GSWLC lacks. The rest-frame W4 prediction of the higher redshift galaxy lies below the y-axis range and the rest-frame prediction for the lower redshift galaxy overlaps with the observed photometry of the higher redshift galaxy. Given the redshifts of these objects and our data driven approach, our K-corrected magnitudes cannot significantly deviate from the observed photometry, which allows us to K-correct even low signal-to-noise bands like W3 and W4. . . . 234A1 Log star formation rate as a function of log mass for the Milky Way analogues selected in Ref. [219] and in this work. 248

#### Acknowledgements

Foremost I am grateful to my primary adviser, Jeffrey Newman, for his guidance and support over the years. Under Jeff's tutelage his enthusiasm and knowledge have been a source of inspiration and excitement that I myself feel in my work. I am especially appreciative of his endless patience and assistance in working through problems. Likewise, I am appreciate of the guidance of my co-adviser, Andrew Zentner. My writing skills and research ethic were forged by Andrew's attention to detail and exuberance for research.

I am just as indebted to Brett Andrews, who has also served as a mentor and close collaborator. He has demonstrated to me how to be a scientist in my own right with his research ethic and leadership skills. I would also like to offer my gratitude to Rachel Bezanson who has been an incredible role model within the department and offered advice in the often not spoken of management side of research with collaborators and growing a career. It is thanks to the assistance of Brett and Rachel that I felt prepared in taking the next steps of my own academic career as a post doc. I would also like to thank Yao-Yuan Mao, who mentored and collaborated with me at the fledgling stages of my time in graduate school. His expertise taught me many of the key computing and modeling tools necessary for astronomers, which has served as a critical foundation for my research going forwards. Also the efforts of Yao, in addition to Brett, Rachel and the other astronomy faculty, have forged the astronomy group into a welcome and inclusive place. As such, I too aspire to boost the students and working environment in any department I find myself in.

I am especially thankful for the scientists that have been willing to collaborate with me on papers as well as those offering insight on various topics and issue. In particular I would like to thank Nick Boardman and Gail Zasowski in addition to the entire Milky Way as a Galaxy group in SDSS IV. I also thank Samir Salim, Karen Masters, and Heidi Wu. They have all shown me the invaluable merits of working together to tackle questions, and I aspire to have the same quality of mentorship they have shown me towards the junior students I will work with. I thank my committee members, Michael Hatridge and Matthew Walker, for their guidance and advice in helping me grow into the researcher I am today. I also thank other faculty and staff members of the department, especially Carlos Badenes, Brian Batell, Ayres Freitas, Gracie Gollinger, John Hillier, Leyla Hirschfield, Arthur Kasowsky, and Michael Wood-Vasey, for providing advice, assistance, resources, and making the department a warm and friendly place.

I thank my dear friends in my student cohort, in particular Brian Flores, Nathan Herring, and Olivia Lanes. I will always look fondly upon all of the hard work we put in together solving problems on the whiteboard at all hours of the day, attempting to grasp advanced physics concepts. At last all of the endless hours of studying have paid off for us and I know we will all do incredible things in our careers.

I also thank other past and current students, especially Christine Mazzola Daher, Biprateep Dey, Daniel Doucette, Travis Hurst, Yasha Kaushal, Tim Licquia, Lorena Mezini, Alan Pearl, Daniel Perrefort, Kara Ponder, Troy Raen, David Setton, Antonia Villareal, Kevin Wilk, and Rongpu Zhou. You have all been a wealth of research knowledge, advice, and friendship. I will dearly miss our random conversations over coffee in the astronomy office. I also offer a special thanks to my friends outside of Physics, especially Natalie Bobeck, Steve Sailor, and Dave Oakley, for listening to my practice talks and tearing me away from the computer.

Finally, I would like to acknowledge my fiancée, Justin Stickel. His love and compassion have kept me going through some of the hardest times of my life. With his support throughout the course of my research I have grown more than I ever could have alone. He has always been by my side these past six and a half years, cheering me on, giving me advice, and helping me improve my coding skills.

#### 1.0 Introduction

#### 1.1 Background

The focus of this dissertation is on leveraging the Milky Way to fill in some of our knowledge gaps on the complexities of galaxy formation and the underlying nature of dark matter. The overarching goal is to answer how the Milky Way - both on the dark matter side and luminous matter side - fits in with other galaxies in the Universe. The chapters within are designed as a two-pronged approach to this goal, first focusing on the theoretical side and then on the observational side. These chapters are generally self contained meaning that readers can focus on sections more relevant to their interests. The remainder of Chapter 1 introduces readers with a limited knowledge on extragalactic astronomy to topics covered in the remainder of the text as a means for setting up the concepts behind my models and key results.

#### 1.1.1 The Standard Cosmological Model

Our modern Cosmology is based upon the cosmological principal, the idea that the Universe is isotropic and homogeneous, and Einstein's theory of relativity, where the mass distribution of the Universe governs the structure of space-time. These pillars have led to the success of our Big Bang cosmological model, in which we describe the formation of the elements and eventually stars and galaxies in the Universe.

The cosmological model was first defined by observations dependent on luminous matter. If we disperse the incoming light from one of the many objects we observed in the sky, we find that the spectrum produced by the atomic signatures are shifted to longer wavelengths (or redshifted) compared to what we expect from their rest-frame positions. Edwin Hubble's 1929 discovery of the recession of distant galaxies, with velocities proportional to their distance, was our first definitive piece of evidence of an expanding Universe [171] in which most galaxies are moving away from us. From this initial discovery a successful model of Cosmology has been built. Subsequent observations of the Cosmic Microwave Background (CMB) in 1964 confirmed the prediction of a Hot Big Bang in which a uniform background of radiation caused by the high temperatures and density of the early Universe was observed, marking when the Universe had cooled enough from expansion for photons to travel freely through space [290]. Primordial abundances of nuclei heavier than hydrogen have been directly predicted and measured, confirming nucleosynthesis driven by the Big Bang. The concept of a Universe originating from an initial small, hot, and dense state with no given center or edge yet a finite age and expanding space serves as the backbone for our cosmological model well supported by observations.

However, these imprints left by luminous matter are only a portion of our modern concordance cosmological model. Our model now also includes two more mysterious constituents of the Universe – dark matter and dark energy. The assembly of luminous gravitationaly bound stars, gas and dust within galaxies is enforced by more massive and extended material that we cannot see. Dark matter was first postulated by Fritz Zwicky in 1933 after finding that galaxies in the coma cluster moved too fast to be bound by visible matter [402]. Additional observations of galactic rotation curves, lensed galaxy systems, and the structure of the Universe over the latter half of the 20<sup>th</sup> century have generally convinced astronomers of the existence of dark matter and the pivotal connection it plays in galaxy formation and evolution. Even more surprising was the discovery of the *accelerated* expansion of space in the 1990s by cosmologists measuring distant supernovae [305, 292] - objects further away from us are moving faster than what is expected from typical Universe expansion. Now our current cosmological paradigm consists of cold dark matter, believed to be composed of relic massive stable particles with random motions at non-relativistic speeds (hence cold) when the Universe was young, and dark energy, most commonly thought to be a type of "vacuum energy" (i.e., an underlying background energy contributing to the cosmological constant). Thus we refer to our concordance cosmological model as " $\Lambda$ CDM" where  $\Lambda$  represents dark energy, alongside cold dark matter (CDM).

#### 1.1.2 A Brief History of (Dark) Matter and Large Scale Structure

As exemplified by Hubble's law, the study of cosmology is closely linked to the study of galaxies. Because of the fixed universal speed of light, the distance of a galaxy and hence its redshift (z) is linked directly to a specific age of the Universe. Thus more distant objects emitted their light at a time when the Universe was younger. But before we address the study of galaxies, we need to address the underlying structure of the Universe that built up these galaxies. Otherwise we are left with little explanation for the vast dichotomy of galaxies that we observe today. The cosmological principal asserts that the Universe is homogeneous. While at very large scales (>200 Mpch<sup>-1</sup>) this is true we know that the cosmological principal cannot hold at smaller scales, otherwise galaxies like the Milky Way we live in would never have formed.

Our standard "Big Bang" cosmological model proposes tiny (quantum) fluctuations in the density field in the fledgling stage of the Universe  $(z \sim \infty)$ . As the Universe expanded significantly and rapidly during a stage of inflation not long after, these quantum fluctuations were amplified as slight density fluctuations before they could self-annihilate. These fluctuations left their mark as small temperature anisotropies that we observe in the CMB. During expansion the Universe cooled and matter came to be the dominant constituent of the Universe instead of radiation (post  $z \sim 3600$ ). These over-densities became magnified by their gravitational field, increasing their density with respect to the background and attracting matter. Gravitational instability amplified these perturbations further, and started to form the dust clouds that cooled and fragmented into the first stars ( $z \sim 100$ ). Cumulatively more massive structure formed, yielding galaxies ( $z \sim 10$ ) and ultimately leading to the massive galaxy systems we observed at present day ( $z \sim 0$ ), such as galaxy clusters, superclusters, and walls (and consequently voids) (see e.g., Ref. [75]). The growth of structure in the Universe is especially governed by gravitational instability.

The structures that we see today could not have been built from luminous matter alone, but instead by dark matter, which accounts for 85% of the total mass of the Universe. Without dark matter the galaxy formation epoch would have begun considerably later than what we observe. Luminous matter is subject to radiation pressure, which works in opposition to the gravitational instability which allows for compact structures to form (like in stars, which are radiation pressure supported against gravitational collapse). Conversely, dark matter only interacts gravitationaly and is unimpeded by any other forces. Thus dark matter must have amplified the initial density perturbations in the young Universe, allowing for these structures to build up. Once the perturbations have grown sufficiently, a region may reach a critical density that is many times higher than the background density of the Universe. At this point the perturbations stop growing linearly and begin to undergo nonlinear collapse.

Ultimately, dark matter collapses into a network of self-bound objects we refer to as dark matter halos. While the precise definition of a halo can vary between research studies, astronomers agree that dark matter halos are approximately virialized objects <sup>1</sup> [80]. The virialized region is a few hundred times the background density of the Universe. The background density contrast is typically used to define the boundary of the halo, the most common being the virial radius, which is dependent on cosmology and time [52]. It is with the formation of dark matter halos that baryonic matter begins to follow suit via gravitational attraction and ultimately forming structures we observe. Consequently the study of dark matter halos is a critical aspect for studying the formation and evolution of galaxies.

N-body simulations serve as one of the key methods for investigations of dark matter halos since gravitational dynamics are far too complex to by analytically explored. These simulations take a co-moving cube to represent the Universe and place in particles that interact under gravitational dynamics (see e.g., Ref. [366, 1, 13, 166, 336]). N-body simulations have unveiled many aspects of non-linear structure formation. For example, numerical simulations in the late 90's revealed the existence of much lower mass halos within the virial radius of a more massive halo [260, 198]. While small halos merging with larger ones may experience mass stripping and disruption, some are dense enough to survive the process (or at minimum their cores). This results in smaller self-bound halos residing within the virial boundary of the more massive halo [188, 393]. We refer to these objects as subhalos (their observational counterparts are satellite galaxies), which we expect to be the direct result of hierarchical assembly  $^2$ . The study of halos provides key clues on how galaxies assemble and

<sup>&</sup>lt;sup>1</sup>A virialized objects is a stable system of gravitationaly interacting particles.

<sup>&</sup>lt;sup>2</sup>Hierarchical assembly refers to the creation of larger structure by the consecutive merging of smaller structures.

evolve, as the assembly of baryons necessitates the assembly of dark matter. The properties of a galaxy reflect those of the halo in which it lives.

#### **1.1.3 Galaxy Formation and Evolution**

Dark matter constitutes the underlying pieces for which galaxies are able to form by attracting baryonic matter into said dark matter gravitational potential wells. As the baryons cooled radiatively and contracted at the center of dark matter halos, their angular momentum was conserved resulting in the formation of a disk [113]. Baryonic matter can undergo such a process independent of dark matter, but would taken a significantly longer time without dark matter. While the precise details of disk formation and consequentially star formation are highly complex and still not understood to a high degree of certainty, we have a general theory on the formation of galaxies that is intimately tied with that of the dark matter halo. Galaxies serve as our beacons for probing the underlying geometry of space and as the houses of star (and therefore metal) formation, making them one of the key building blocks of the Universe.

Because of our inability to perform large scale experiments in space, our ability to query galaxy evolution models hinges on comparisons to observed galaxies. In the earlier 20<sup>th</sup> century addressing questions of galaxy evolution was limited to targeted observations with small sample sizes due to technological limits. The past few decades have birthed large, wide-scale surveys spanning wavelengths from the UV ( $\lambda \sim 100 - 400nm$ ) to the radio ( $\lambda \sim 1mm$  – several km). One notable survey is the Sloan Digital Sky Survey (SDSS; [390]), which has produced deep multi-band ( $\lambda = 3543 - 9134$ Å) CCD imaging and spectroscopic follow-up for > 10<sup>6</sup> galaxies out to a median redshift of  $z \sim 0.1$ . Combinations of photometric and spectroscopic data provide detailed information on both the light (e.g., colors, magnitudes) and physical properties (e.g., mass, star formation rate) of galaxies. Galaxy surveys of this scale (and forthcoming larger surveys) span the extensive breadth of properties that galaxies exhibit [28], their population distributions, and map the underlying 3D structure of the Universe traced by galaxies.

Rest-frame color and absolute magnitude are some of the easiest quantities to measure

in most galaxies, with known redshift (z), regardless of their distance from us. One of the most notable discoveries of underlying galactic distributions is the color-magnitude diagram (CMD), or integrated color as a function of magnitude. In this parameter space, the SDSS sample revealed that galaxies tend to fall into one of two populations: passively-evolving red galaxies with older stellar populations; or blue galaxies that are actively forming stars. In the optical CMD, these two galaxy populations are commonly referred to as the "red sequence" and "blue cloud" which also followed with general galaxy morphological type. This dichotomy was established by SDSS and and other surveys to high significance at both low redshift ( $z \sim 0.1$ ; [338, 11]) and high redshift ( $z \sim 1$ ; [20, 373]).

The majority of the population of galaxies that fall into the blue cloud are spiral galaxies. These galaxies are generally actively forming stars. Much of their light is dominated by the hot and consequently bright and *blue* O- and B-stars. As a result, the spectral energy distribution (SED; a characterization the galaxy's flux as a function of wavelength) is dominated in the UV and optical, and re-emission from excited dust in the infrared, by these stars. Spiral galaxies have their stars, gas, and dust organized into rotationally supported disks. These disks have an exponential decrease in brightness from the center [131], with results of various instabilities super-imposed within the disk, such as rings and the titular spiral arms - the major site of star formation [359]. The centers of spiral galaxies tend to contain older stellar populations where the central bulge resides and where bars often form [329], thus the central region of spiral galaxies tends to be redder than the remainder of the disk.

Conversely, the majority of the galaxy population that falls within the red sequence are elliptical galaxies. These galaxies are spheroidal or ellipsoidal in shape, exhibiting very little morphological features and smooth surface brightness profiles that diminish more gradually than those of spiral galaxies [323, 152]. Elliptical galaxies contain much smaller reservoirs of cold gas resulting in little if any star formation. As a result of the minimal star formation, the stellar populations within these galaxies are old and metal rich (lacking in young shortlived O- and B-stars) which gives them their red color and bright optical-red colors in their respective SED. Despite the lack of young, bright stars elliptical galaxies can far exceed the brightness of elliptical galaxies.

The region of the CMD between these two distinct galaxy populations is often referred

to as the "green valley". This locus is thought to contain a transitional population of galaxies that are passively evolving in the sense that star formation rates are diminished Ref. [20, 120], though they still can contain some younger stars. The increase in the fraction of red galaxies over time has led many astronomers to conclude that a galaxy first lives in the star forming blue cloud and then transitions into the green valley and ultimately into the red sequence in complete quiescence, with the galaxy growing more and more red over time due to the ageing stellar population. Thus comparisons of galaxy populations like those done with the CMD are a key tool for unlocking the answers to galaxy evolution.

### 1.1.4 The Milky Way, A Laboratory for Galaxy Evolution

Many remaining unknowns of galaxy formation and evolution, such as the origins of galaxy color bimodality, can be answered by comparing galaxy populations and by making direct comparisons to the Milky Way. Our residence within a massive, star forming galaxy provides a major advantage for seeking answers. The Milky Way is the optimal laboratory for testing galaxy evolution theories and the best benchmark for comparisons, given our proximity to fully resolved constituents that make up a galaxy of the luminosity of the Milky Way at the present epoch (z = 0). The Sun's location within nearly the mid-plane of the disk and approximately two-thirds radially from the center [183] places us in the prime location to study the creation of stars in nearby stellar nurseries, like the Taurus molecular cloud only 140pc away.

Moreover, the Milky Way is the only galaxy where we can measure the properties of large samples of individual, resolved stars. Detailed observations of stellar magnitudes and colors (e.g., 2MASS [330], the VVV Survey [309]), stellar spectra (e.g., APOGEE [115], GALAH [86]), and stellar distances and proper motions (Gaia [132]) - have covered over a billion stars in the Milky Way. While we are unable to fully map the extent of the Milky Way, for stars that are close enough these surveys have created an unprecedented 3-D map of the Milky Way, further revealing the composition, formation, and evolution of the Galaxy (e.g., Ref. [42, 300, 53]) through the assembly of stellar mass over cosmic time.

The Milky Way also offers a unique perspective for obtaining constraints on  $\Lambda CDM$ .

Studies on satellites of the Milky Way have enabled constraints on possible dark matter paradigms [267] and test the validity of predictions from N-body simulation such as the number density of satellites [201, 367, 178], which are challenging to measure outside of the Local Group. Stellar measurements can also be compared to simulations to test consistencies between  $\Lambda$ CDM and the Milky Way, tackling outstanding questions like the shape of the MW dark matter halo [82].

These in-depth observations of the Milky Way have brought forth a plethora of dedicated simulations made to produce Milky Way-like galaxies that match a number of these observed properties. Via computer, astronomers can track the dynamical evolution and interactions of dark matter, gas, and stars within synthetic version of the Milky Way, with initial conditions set by cosmological observations and prescriptions of physical laws guiding the process. Generally these simulations are limited by mass resolution, or the number of particles within the volume, which determine how many stars or how much gas a single particle represents. One example of such simulations are the Latte simulations [376]. They have produced possible iterations of the Milky Way, which have allowed us to extrapolate our local measurements to explore evolutionary histories of the Milky Way, such as the formation of disk components [226, 139, 315] and the bulge of the Milky Way [88]. The recipes used to model the evolution within simulations contain a number of free model parameters. These parameters can be tuned to construct evolutionary paths with results that directly match the compositions and structure of Milky Way stars and gas, or can be used to explore various contributions certain properties have to galactic evolution. Thus the connection between MW observation and simulation is a key tool in understanding the constituents of galaxy evolution.

In some respects, the Milky Way is a typical large galaxy of its morphology, reflective of its low density environment [85]. Given its stellar mass the Milky Way has a typical star formation rate and bulge-to-total mass ratio when compared to similar galaxies [36]. It has also been shown to reasonably obey the Tully-Fisher relation [218], which links maximum rotational velocity to luminosity in spiral galaxies [348]. The Milky Way ( $\sim 10^{12.1} M_{\odot}$ ) is one of two primary members of the Local Group, the other being Andromeda/M31 ( $\sim 10^{12.2} M_{\odot}$ ). As a whole, the Local Group is a relatively low mass system consisting of an assortment of spiral and dwarf ( $\leq 10^8 M_{\odot}$ ) galaxies, which resides in a low-density filament in the outskirts of the Virgo super-cluster [349]. This means that the Milky Way is considered to be in a relatively low-density environment, which is common for spiral galaxies.

Yet in other respects the Milky Way appears to be relatively unique. The massive satellites of the Milky Way - the Large and Small Magellanic clouds - are quite rare with only a few percent or less of analogous systems found in surveys and simulations [222, 104, 119]. Additionally, in contrast to M31, the Milky Way has not had any major mergers (objects with mass fractions above 0.3 relative to the MW merging with the MW) within the past 10 Gyr. One of the most unique features of the Milky Way is its anomalously small disk scale length discovered by Ref. [217]. This raises the question of the uniqueness of the Milky Way and complicates how we compare the Milky Way to other galaxies.

#### 1.1.5 Connecting the Galactic and Extra-Galactic

In order to connect our detailed observations within the Milky Way to the populations of observed galaxies we must first understand how the Milky Way compares to other galaxies. However, comparing colors and luminosities of the Milky Way to external galaxies is not trivial, regardless of whether we compare to observed galaxies or to mock images from high-resolution hydrodynamical simulations. Our location in the plane of the Milky Way is a double-edged sword; while we can explore star formation and galaxy evolution to an unrivaled level of detail, we are also inhibited in measuring the Milky Way the same way we measure external galaxies. Much of our view of the Galaxy is obscured by interstellar dust, especially at UV and optical wavelengths [62, 320]. This is due to the progressively efficient absorption and scattering of light by dust as a decreasing function of wavelength. Stars outside of the local solar region are reddened as a result of the dust obscuration. Additionally, determining the bulk integrated light of stellar populations in the Milky Way is challenging due to the spread of stars over large and varying distances, which is further exacerbated by correspondingly large and varying dust extinction along lines of sight to the Earth. This makes the study of any portions of the Galactic disk beyond the solar neighborhood exceptionally difficult, and results in a fragmented picture of the Milky Way. Integrated properties that are relatively painless to obtain in external galaxies (though dust obscuration can affect these observations as well, see e.g., [243, 245]), are impossible to obtain directly within our own Galaxy [265]. Hence much of our understanding of the Milky Way is still poorly constrained.

Recent work, such as that by Ref. [219], has began to tackle the question of where the Milky Way lies on the optical CMD in an effort to enable comparisons of the Milky Way to other spirals and the full galaxy population. This study was also limited, lacking critical constraints at short (UV) and long (IR) wavelength ranges which are established to be much more sensitive to star formation and could improve classification of the current evolutionary stage of the Milky Way. Constraints on the full UV-IR SED of the Milky Way will be the subject of Chapter 4 which focuses on capturing the full bulk photometric properties in the Milky Way. By exploiting the assumption that the Milky Way should not be incredibly unusual amongst its peers (i.e., galaxies of the same physical properties like mass and star formation rate), it is straightforward to build a model to determine Milky Way photometric properties trained on the photometric properties of other galaxies.

Another necessary task for making connections between the Milky Way and other galaxies hinges on comparisons between the Milky Way and simulations. The measurable structural and kinematic parameters of the Milky Way play a large role in providing comparisons for numerical simulations of synthetics galaxies. Modern high-resolution simulations, such as Eris [157], APOSTLE [316], Latte [376], and the DC Justice League suite [9] can construct mock galaxy images that are representative of the Milky Way. Much of our understanding of the details of how the Milky Way built up its stellar populations and observed structure is a result from these simulations. However, the link between observations and simulations is precarious, where assumptions that are erroneous on one side can propagate into the other. It is difficult to establish whether simulations match the Milky Way because models of the physics of star formation and feedback are correct, or instead they mimic Milky Way-like properties for a galaxy and dark matter halo with incorrect characteristics (mass, merger history, etc.) when using mis-tuned physical models. One way to address these issues is to consider what it means for a simulated galaxy to be Milky Way-like, which is one of the focal points of Chapter 3. This chapter explores constraints on Milky Way dark matter halos and the connections the halo properties have to real observables like satellite abundance, all which are shaped by the evolutionary history of the Galaxy. Likewise, it is not unusual for observers and simulators to make different assumptions about the systems they are exploring; a key issue explored in Chapter 2 that addresses potential double counting of subhalo contribution to the dark matter halo mass profile. Advances to the observation-simulation connection also depend on our improved understanding of the Milky Way in the context of the general galaxy population. Currently simulators must resort to comparing their simulated Milky Ways to very general galaxy populations that, while superficially resembling the Milky Way, have a wide range of other global properties [157]. This issue is also explored in Chapter 4, where a detailed outside-in picture of the Milky Way can dramatically alleviate the comparisons simulators make. Our ability to interpret our observations of the Milky Way into assertions about galaxies in general hinges upon the nuanced details within modeling.

#### **1.2** Structure of Thesis

In general this dissertation is centered on utilizing statistical tools and techniques to surpass many of the hurdles that inhibit our knowledge of Milky Way properties and out ability to connect the Milky Way to the galaxy population from both observation and simulation. This thesis has four major components, with additional supplementary material included at the end. Below I outline the motivation and findings of each chapter.

• Chapter 2 — Before placing the Milky Way in an extra-galactic context, I first focus on the disconnect between simulations and observations - a crucial first step before comparisons to the Milky Way can be made. Because dark matter has no electromagnetic interactions, our knowledge on the nests of galaxy formation is extracted by indirect observation, like gravitational lensing, and inferred from models of the mass distributions of halos. However, there is a disparity between the models of halos inferred from simulations versus those used to interpret observations. In simulations, the ensemble of the host halo and subhalos are considered as one unit, while some lensing inferences treat subhalos separately from the host. Additionally, subhalos are a key indicator of the evolutionary stage and history of a dark matter halo, so their treatment in models must be carefully considered. By employing least squared fitting to various forms of dark matter halo density distributions ( $\rho(r)$ ), I explore the impact subhalos have on these systems. I show that without including subhalos in the density distributions these distributions change significantly in shape, indicating that lens modeling must be more discerning with treatments of subhalos. This chapter also provides evidence for the very nuanced connection between environment, galaxy evolution, and the reflected properties exhibited by dark matter halos, especially for halos at the mass of the Milky Way and higher. For example, I find that the central concentration of material in a halo, which is correlated with the mass of the halo, is partially explained by subhalos. I also find that scatter amongst halos at fixed mass is also directly tied to the presence of subhalos. Both of these pieces of evidence indicate the intricate ties between halo evolution and the properties they exhibit.

Chapter 3 — In this chapter I place the Milky Way's dark matter halo in the context of other dark matter halos that are within the same mass range of the Milky Way. One of the unique predictions of  $\Lambda CDM$  is the abundance of dark matter subhalos. Observed subhalos around the Milky Way cause detectable perturbations which in turn can be used to constrain halo properties and test the validity of  $\Lambda$ CDM. But connecting the Milky Way to other galaxies, observed or simulated, is even further complicated by the discrepancies at small scales  $(<1 \text{ Mpc}h^{-1})$  between the predictions of N-body simulations and observations within the Local Group. One aspect of these discrepancies is the focus of this chapter, an issue coined the "missing satellites problem", or an over-prediction by simulations of the number of subhalos of a particular rotation speed relative to the number of observed satellites of the Milky Way. Simulators have attempted to fix this apparent discrepancy by tuning the baryonic physics in hydrodynamical simulations to reduce the number of predicted satellites. However, the halo comparisons that show this discrepancy do not use all of the information we have on the Milky Way's halo. Part of the missing satellites problem can be attributed to predicting the subhalo mass function from dark matter-only simulations that only match the Milky Way in mass and not in other important second order parameters such as concentration, spin, and time of the last major merger. To address this, in this chapter I use high-resolution zoom-in simulations of Milky Way-mass dark matter halos to investigate halo properties beyond mass and their correlation with subhalo abundance in order to determine subhalo populations for a halo whose properties match our own. I find that the Milky Way's halo has a higher concentration, lower spin, and quieter merger history than the average halo of the same mass. With a Poisson maximum likelihood model I make a more accurate prediction for the Milky Way's subhalo abundance, resulting in the conclusion that the Milky Way has as many as 30-50% fewer subhalos than typical halos of the same mass (i.e., neglecting matches in second order parameters). We must carefully consider what we call "Milky Way-like" before making imprudent comparisons.

- Chapter 4 This chapter focuses in on constructing a full outside-in portrait of the Milky Way. Here I construct the first SED of the Milky Way, spanning from the UV to the IR. This serves as a critical tool for allowing us to compare the Milky Way directly to objects measured in extragalactic surveys at a variety of wavelengths. A full Milky Way SED also offers constraints to hydrodynamical simulations that currently lack measurements of non-optical luminosities necessary for constraining star formation history and galactic extinction. I expand upon the work of Ref. [219] by increasing increasing both the number of physical galaxy parameters studied (disk scale length, bulge-to-total mass ratio, and bar probability in addition to stellar mass and star formation rate) - due to the unusually compact disk and quiescent evolutionary history of the Milky Way - as well as the range of photometric colors and magnitudes (in the band passes of GALEX FUV/NUV, 2MASS J/H/Ks, and WISE W1 - W4 in addition to SDSS ugriz). Due to the complexity of this model, I developed a Gaussian process regression (GPR) model that is trained on a broad galaxy sample and uses the fiducial MW properties and their uncertainties to predict individual photometric properties for the MW. I further confirm the optical "green valley" transitional classification of the Milky Way on the CMD in the process, but note that in the UV and IR the Milky Way is more consistent with the star forming "blue cloud". In fact, the Milky Way appears to be consistent with the population of red spiral galaxies, objects that are bright and red, likely undergoing effects from bar induced star formation suppression.
- Chapter 5 Luminosity measurements of various objects in the Universe are critical

for understanding how stars and galaxies have evolved over time. However, astronomers are challenged to correct for the redshifting of light to longer wavelengths between the time the light is emitted from the objects and then subsequently observed in a telescope. This difference, which must be accounted for in photometry, is referred to as the Kcorrection. Currently our K-correction methods rely upon some form of SED fitting utilizing empirical or model galaxies, which have a slew of limitations. For example, methods that utilize a somewhat small number of templates can result in inaccurate Kcorrections for galaxies that poorly match the templates. Similarly, one is left the dark in obtaining K-corrections for wavelengths that do not have available or poorly constrained templates, such as the IR bands, which became a notable issue in utilizing the WISE bands in Chapter 4. In this chapter I have developed an empirically driven approach that regresses off of a color for which SEDs *are* well constrained. By relying on the assumptions that at low redshift galaxy SEDs fall into a single parameter family and the K-correction for a given band is a simple (polynomial) function of a galaxy's rest-frame color (a degenerate combination of dust, metallicity, and star formation history) determined in some other pair of bands I have found that K-corrections are easily calculable at low redshift. With this approach, as long as one has access to a single rest-frame color that has been K-corrected already, one can determine rest-frame colors for any other pair of bands. This became a useful tool for determining the infrared colors of the Milky Way, and going forward may be instrumental in many aspects of astronomy.

• Chapter 6 – This chapter summarizes the results and implications this dissertation has accrued about how the Milky Way and its dark matter halo fits in with other galaxies and the underlying dark structure of the Universe. I illustrate the new comprehensive portrait of the Milky Way, a galaxy now no longer thought to be a standard blue spiral like any other. I conclude with future avenues of study to better pinpoint our understanding of the evolution of our Milky Way with a focus on statistical techniques exploited in this dissertation.

#### 2.0 Illuminating Dark Matter Halo Density Profiles Without Subhaloes

This chapter has been accepted for publication in the Monthly Notices of the Royal Astronomical Society ©: 2020 Catherine Fielder. Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved. Minor modifications have been made to the text.

### 2.1 Introduction

The gravitational collapse of over-dense patches of the universe in the Lambda–Cold Dark Matter (ACDM) cosmology culminates in the formation of virialized dark matter-dominated structures called dark matter haloes. Much of the modeling necessary to undertake modern data analyses for the purpoes of understanding cosmology, the nature of dark matter, or the evolution of galaxies relies on our understanding of various properties of dark matter haloes. Halo properties and their evolution have been investigated theoretically and in great detail [379, 188, 73, 332, 22, 398, 158, 225]. Similarly there are a number of observational probes of the structure of dark matter haloes, including gravitational lensing [253, 235, 76, 181, 282, 351], X-ray surface brightness/temperature [271], and line-of-sight velocity dispersion of either stars [15, 279, 389] or satellites [254, 261]. The delicate marriage of dark matter halo theory and observation is fundamental for our understanding of galaxy formation and evolution.

*N*-body simulations have been instrumental in determining the structures and mass assembly histories of haloes. Haloes grow hierarchically. Large haloes accumulate their masses through mergers with smaller haloes, which leave a lasting imprint. On dynamical timescales ( $\sim$ 3–4 Gyr at z = 0), in-falling haloes may be stripped of their mass and/or disrupted, while those that survive become dense, self-bound objects orbiting within the larger host halo (see e.g., [188, 393, 394, 56]). Therefore, host dark matter haloes are composed of (1) a relatively smooth component and (2) *subhaloes*, the gravitationally bound haloes that reside within the larger hosts. The smooth component of a halo consists primarily of disrupted subhaloes from earlier mergers [298, 391], but there may also be a sub-dominant contribution acquired via smooth accretion [368].

A basic characteristic of a dark matter halo is its density profile, which describes the way that mass is distributed throughout the bound object. Despite the fact that halo formation is a complex process, *N*-body simulations have shown that the mass distributions of dark matter haloes are well described by the same equilibrium density profile at all masses. The most widely used halo density profile is the spherically averaged Navarro–Frenk–White (NFW) profile [273, 274]. The NFW profile is a two-parameter (one-parameter at fixed mass) functional form that describes the halo mass distribution as a function of distance from the halo center:

$$\rho(r) = \frac{\rho_s}{\frac{r}{r_s}(1 + \frac{r}{r_s})^2},$$
(1)

where  $\rho_s$  is the local density. The length scale,  $r_s$ , is referred to as the characteristic radius or scale radius; it corresponds to the radius at which  $d \ln \rho / d \ln r = -2$  in the case of the NFW profile.

The degree to which the mass within a halo is distributed with respect to its radius is often quantified by the dimensionless halo concentration parameter,  $c = \frac{r_{\text{vir}}}{r_s}$ . The virial radius,  $r_{\text{vir}}$ , can be thought of as the "outer edge" of the halo that characterizes the halo size <sup>1</sup>. Halo concentration has a well-known dependence upon both mass (the concentrationmass relation) and redshift [274, 55, 199]. Profile parameters of individual haloes, such as concentration, are typically determined by building a spherically averaged density profile from simulation data and fitting this profile to the NFW form.

The contributions of substructure are handled differently in observational and theoretical studies, which provides challenges to compare results between them. In simulation analyses, spherical averages of haloes typically include all halo substructure. In observational investigations, such as the analyses of gravitational lens systems, the spherically averaged halo profile is typically modeled with a particular form, the parameters of which are then inferred from the data [257, 220]. However, in such observational studies, it is often the case that the observed substructures (e.g., satellite galaxies) within a lensed system are assigned

 $<sup>{}^{1}</sup>r_{\rm vir} = \left(\frac{3}{4\pi} \frac{M_{\rm vir}}{\rho(\langle r_{\rm vir})}\right)^{1/3}$ , which is derived from the virial theorem for a gravitating body.

their own, distinct density profiles in addition to the profile attributed to the host system [279, 280, 94, 144]. Thus there arises the risk that these subhaloes are being double counted. If the subhaloes are to be treated as discrete units with their own properties, one needs to use halo density profiles developed to exclude the mass within substructures. Likewise, if one were to exclude subhaloes from the host halo and then model the subhaloes separately, there is no guarantee that the same *host* halo mass profile (e.g., NFW) would still be an accurate representation of the host. These concerns are the focus of the present work.

In this work we examine the impact of subhaloes on host halo mass profiles, and the ability of analytic profiles like the NFW profile to accurately describe halo mass distributions that do not include the mass associated with subhaloes. To do this, we make use of the density profiles of haloes from which we have removed the mass associated with subhaloes, a technique first used by Ref. [224] and Ref. [385]. Even when we exclude subhaloes we are still spherically averaging the haloes. In order to explore a wide range of masses, we use two sets of high-resolution zoom-in simulations. One simulation focuses on Milky Way-mass haloes and the other on cluster-sized haloes.

In Section 2.2 we describe the simulations used in our work, the halo finder, and the halo density calculations. Section 2.3 details our analysis methods of the halo profiles through statistical methods and fitting to different mass profile prescriptions. Section 2.4 presents the results for both stacked and individual haloes in order to compare different functional forms for the density profile. In Section 2.5 we summarize our results and conclude that the mass distributions of the smooth component of the halo and the combination of the smooth and subhalo components are distinctly different, especially in the outer region of the halo mass distribution and describe the implications of such a result with a more universal mass profile, the generalized Einasto profile.

#### 2.2 Data

In the following section, we outline the simulations and halo finder used, the methods for calculating simulated halo density profiles, and the procedures used to excluding mass
associated with subhaloes.

## 2.2.1 Zoom-In Simulations of Two Mass Regimes

In this work we use two sets of zoom-in, gravity-only simulations. The first set consists of 45 zoom-in simulations focusing on haloes of approximately the mass of the Milky Way's halo and initially presented in Ref. [237], which we refer to as the Mao et al. Milky way Mass Zoom-in (MMMZ) simulations in the remainder of this chapter. The second simulation suite comprises 96 cluster-mass zoom-in simulations known as the RHAPSODY simulations, first presented in Ref. [385]. All analyses performed in this chapter are applied to both sets of simulations, allowing us to test for consistency of results across several orders of magnitude in halo mass and to test for mass dependence. In both cases we use the present-day (z = 0) snapshots.

The haloes in the MMMZ simulations cover a very narrow range in virial mass ( $M_{\rm vir} = 10^{12.1\pm0.03} \,\mathrm{M}_{\odot}$ ). The cosmological parameters for the simulations are as follows: matter density  $\Omega_{\rm M} = 0.286$ ; dark energy density  $\Omega_{\Lambda} = 1 - \Omega_{\rm M} = 0.714$ ; Hubble parameter  $h = H_0/100 = 0.7$ ; mass fluctuation amplitude  $\sigma_8 = 0.82$ ; and scalar spectral index  $n_{\rm s} = 0.96$ .

These MMMZ haloes were selected for high-resolution re-simulation from a parent c125– 1024 dark-matter-only simulation run with L-GADGET (see [17, 214]). The high-resolution zoom-in regions have a particle mass of  $m_{\rm p} = 3.0 \times 10^5 \ h^{-1} {\rm M}_{\odot}$  and a softening length of 170  $h^{-1}$  pc comoving. We use four times the softening length, or 0.68  $h^{-1}$  kpc, as the minimum length scale down to which the density profiles are fitted. The lower limit to the subhalo maximum circular velocity  $V_{\rm max}$  for convergence is approximately 10 km s<sup>-1</sup>. For more details on the MMMZ simulations, refer to Ref. [237].

The RHAPSODY simulations also span a very narrow mass range at the cluster-size scale,  $M_{\rm vir} = 10^{14.8\pm0.05} \,\mathrm{M_{\odot}}$ . The cosmological parameters used for the RHAPSODY simulations are very similar to those of MMMZ, specifically:  $\Omega_{\rm M} = 0.25$ ;  $\Omega_{\Lambda} = 0.75$ ; h = 0.7;  $\sigma_8 = 0.8$ ; and  $n_{\rm s} = 1$ . The differences between MMMZ and RHAPSODY cosmologies will have negligible effect on any comparisons performed in this chapter, as halo concentrations only depend weakly on cosmological parameters, especially at z = 0 [223]. These RHAPSODY zoom-in simulations were selected from one of the CARMEN simulations from the LArge Suite of DArk MAtter Sim-ulations (LasDamas; [248]) with a volume of 1  $h^{-1}$ Gpc and 1120<sup>3</sup> particles. We use the higher-resolution version of these simulations (RHAPSODY 8K), which have a particle mass of  $m_{\rm p} = 1.3 \times 10^8 h^{-1} M_{\odot}$ . And the minimum length scale that the density profiles are fitted down to is four times the softening length, equivalent to 13  $h^{-1}$ kpc.

We note that the choice of  $4\times$  the softening length as our minimum length scale is to be consistent with previous studies [385, 237], even though the softening lengths do not necessary set the convergence limit of halo mass profiles [294, 225]. Nevertheless, the minimum length scale has little effect on our results as the main change in profiles when subhaloes are excluded are in the outskirts.

## 2.2.2 Density Profiles and Subhalo Removal

In this work, we endeavor to study the density distribution of the "smooth" components of host haloes. We use ROCKSTAR (version 0.99.9-RC3+) to identify haloes according to a virial definition. The ROCKSTAR halo finder uses phase-space information in order to distinguish subhaloes from the host halo's background density, which we take advantage of for this work. Using the particle catalogs we define the *smooth* component of any host halo to be the mass not associated with any ROCKSTAR-identified subhalo. For more details on ROCKSTAR, see Ref. [18] or access the publicly available code at bitbucket.org/gfcstanford/rockstar.

There exists no unambiguous way to define the smooth component of a halo. Our operational definition is dependent upon the ROCKSTAR algorithm for halo identification. Further details regarding this choice and other possible ambiguities are given in Section 2.4.4 and 2.7.

For the purposes of this work, we define the smooth component of a host halo to be those particles not explicitly associated with a self-bound subhalo by ROCKSTAR. By default > 50% of particles must be bound for ROCKSTAR to read out the subhalo. We have used this definition to construct the following processed simulation data sets.

• subhalo-included: the set of all particles associated with the host and all of its subhaloes

(in ROCKSTAR this corresponds to all particles that reside within the virial radius of the host halo).

- *subhalo-excluded:* particles that are associated with the host but *not* explicitly associated with any subhalo listed in the ROCKSTAR halo catalog. This sample excludes subhaloes but includes diffuse substructures, such as streams, that do not meet the binding criteria in ROCKSTAR.
- *subhalo-only:* the set of particles explicitly associated with at least one subhalo. This corresponds to the set of particles in the subhalo-included set that are not within the subhalo-excluded set.

In all cases, only particles within the virial radius of the host halo are considered. As is evident, the *subhalo-included* data set consists of the union of particles within the *subhaloexcluded* and *subhalo-only* data sets, and the *subhalo-excluded* and *subhalo-only* data sets are mutually exclusive. Readers interested in further details of these definitions can see 2.7.

It is useful to get a visual impression of the substructure removal procedure. Fig. 2.1 shows a scatter plot of particles within a plane of width  $\Delta z = \pm 1$ kpc of the center of the host halo for one of the MMMZ haloes. The host halo is centered at (x = 0, y = 0). The left (blue) panel shows the combined distribution of host halo and subhalo particles; i.e., the *subhalo-included* sample as defined above. This is the set of particles that would generally be taken to correspond to a single dark matter halo in N-body simulation analyses and in the vast majority of studies of halo properties.

The middle (orange) panel of Fig. 2.1 depicts the subhalo-excluded sample. There is a noticeable feature aside from the smooth central component of the host halo; at  $(x = -0.01, y = 0.09h^{-1}\text{Mpc})$  there is an over-density that was not listed in the ROCKSTAR halo catalog due to its low bound fraction. Objects like these are in the tidal stream stage of merging. This definition of halo particles more closely maps to the dark matter associated with the central galaxy of the halo, and hence is closer to the *de facto* halo definition assumed in many observational analyses. For comparison, the right (green) panel depicts only the particles associated with those subhaloes that are listed in the ROCKSTAR halo table, or *subhalo-only*. By construction, the combination of the orange and green points is identical to the blue points. While it appears that there is a remaining central halo component, these particles exhibit a large, coherent velocity to the halo center when observed in velocity space. This over-density represents a substructure passing near the host halo center.

It is evident that the distribution of mass without subhaloes is much smoother with many fewer density peaks, demonstrating the effectiveness of our subhalo exclusion procedure.

The smooth, subhalo-excluded, components of halo profiles contain less total mass than the virial masses of the full halo profiles. For instance, for the MMMZ haloes, ~ 20 per cent of the mass is in subhaloes on average, while the mass fraction in subhaloes for the RHAP-SODY haloes is ~ 36 per cent. This difference is accounted for by the relation that more massive haloes contain more subhaloes [394, 145, 335] rather than a reflection of resolution, as the RHAPSODY haloes have a slightly lower resolution that the MMMZ haloes. As a consequence of this decrease in mass, the smooth components of the haloes will not separately satisfy the same virial over-density criterion that the subhalo-included profiles satisfy. This introduces an ambiguity in the physical boundaries of the haloes with subhaloes removed. We choose to use the original virial radius,  $r_{\rm vir}$ , as the halo boundary even in the case with subhaloes removed. We note that quantities listed with the subscript "nosub" are computed when subhaloes are excluded. We also note that when individual haloes are examined we include their snapshot reference ID.

## 2.3 Analysis

In this section, we describe our analysis methods. First, we summarize the functional forms of the analytic dark matter halo profiles used to fit the simulation data, including a new functional form that we refer to as the "generalized Einasto" profile. We then describe the algorithms used to fit these profiles to simulated halo density profiles. Lastly, we specify the statistics used to assess the quality of fits.



Figure 2.1: Visualization of subhalo removal in a 2D projection for one of the MMMZ host haloes. The particles shown here are all restricted to be within the virial radius of the halo, denoted by the black dashed line. In addition we plot a subset of the particles restricted to be near the z value of the halo center, requiring  $z_{\text{host}} - 0.001 < z < z_{\text{host}} + 0.001 h^{-1}$ Mpc. The left (blue) panel shows the standard case where all host and subhalo particles are considered together as making up the dark matter halo. High density areas away from the center corresponding to subhaloes can be easily seen. The middle (orange) panel corresponds to the case where we have excluded subhalo particles; his case may be more appropriate for comparison with observational techniques, as discussed in Section 2.2. The particle distribution is much smoother and more uniform, but some substructure (not associated with systems that ROCKSTAR defines as self-bound subhaloes) is still visible. In the right panel (green) we show only those particles associated with subhaloes that are included in the ROCKSTAR catalog. By construction the orange distribution and the green distribution added together correspond to the blue slice.

#### 2.3.1 Density Profile Parameterization

Halo density profiles have been modeled and parameterized in various different ways. The two following families of functions have been used most frequently to describe dark matter halo densities:

 Double Power-Law: These profiles asymptote to different power laws at small and large radii. Profiles of the double power-law type generally have functional forms similar to the NFW profile described by Equation 1. Previous work has explored a generalized NFW (gNFW) five-parameter profile in order to account for deviations of dark matter haloes from the standard NFW profile, particularly to allow for variations in the slopes at the inner (cusp/core) and outer regions of haloes [164, 397, 90]. The gNFW profile can be written as

$$\rho(r) = \frac{2^{(\beta-\gamma)/\alpha}\rho_s}{\left(\frac{r}{r_s}\right)^{\gamma} \left[1 + \left(\frac{r}{r_s}\right)^{\alpha}\right]^{(\beta-\gamma)/\alpha}},\tag{2}$$

where  $\alpha$ ,  $\beta$ , and  $\gamma$  are all constants. The generalized NFW double power law has a slope of  $-\gamma$  at small radii and  $-\beta$  at large radii; the  $\alpha$  parameter governs the rate of transition between these slopes. The standard NFW profile has  $(\alpha, \beta, \gamma) = (1, 3, 1)$ , corresponding to an inner power law index of -1 and an outer power law index of -3. Many studies use the NFW profile, as the simplicity of having fewer free parameters can be advantageous, despite the increased fidelity with which a gNFW profile can represent simulation data.

2. Continuously Varying Power-Law: Functional forms of this type allow for a gradual flattening of the slope of the density profile towards the inner part of the halo. A well-known member of the continuously varying power-law family is the Einasto profile [114], which has come to be used extensively as recent work has shown it to be a better description of the distribution of matter in haloes than other two or three-parameter profiles [137, 275, 108, 199]. The standard three-parameter Einasto profile is

$$\rho(r) = \rho_s \exp\left(-\frac{2}{\alpha} \left[\left(\frac{r}{r_s}\right)^{\alpha} - 1\right]\right),\tag{3}$$

where  $\rho_s$  and  $r_s$  are the scale density and scale radius, defined similarly as in the NFW case. The Einasto profile differs from the NFW profile most at small and large radii.

The inner slope of the Einasto profile goes to 0 as r approaches 0, and the outer slope does not asymptote to a constant value.

In the main body of this chapter, we focus on four specific profiles. These are (i) the NFW profile (Equation 1, two parameters), (ii) the generalized NFW profile (Equation 2, five parameters), and (iii) the Einasto profile (Equation 3, three parameters), which serve as standards to compare to prior work, as well as (iv) a four-parameter "generalized Einasto" (gEinasto) profile (a profile that was first presented in Ref. [297] and studied in Ref. [342]) that we find best describes halo density profiles:

$$\rho(r) = \rho_s \left(\frac{r}{r_s}\right)^{-\gamma} \exp\left(-\frac{2}{\alpha} \left[\left(\frac{r}{r_s}\right)^{\alpha} - 1\right]\right),\tag{4}$$

where  $\gamma$  modifies the inner density profile slope compared to the [114] form. This profile's asymptotic behavior tends to  $-\gamma$  at small radii<sup>2</sup>.

In 2.6 we describe the other functional forms that we tested, but do not present detailed results for each of these profiles for the sake of brevity. Our qualitative results carry over to these cases as well. As we will show, the gEinasto profile, (iv), is the best-performing model and will be used to demonstrate most of our results. We use the NFW model, (i), as a standard for comparison.

In order to enable comparisons across different types of halo profiles, we calculate halo concentration using the radius at which the profile fit has a derivative equal to -2,  $r = r_{-2}$ . We define concentration as  $c_{-2} = r_{\rm vir}/r_{-2}$ . For each of the aforementioned profiles, the local power-law indices and values of  $r_{-2}$  are listed in Table 2.1. Note that in the case of the gEinasto profile, if  $\gamma > 2$ , then the profile will never have  $d \ln \rho/d \ln r \ge -2$ . We also include for reference  $\rho_{-2} = \rho(r = r_{-2})$  which are useful for the case in which  $\rho_s$  is held constant when fitting profiles.

In Fig. 2.2 we show an example of the local power law indices as a function of halo-centric distance for each of the analytic profiles that we investigate for pedagogical purposes. Red lines depict the NFW (solid) and generalized NFW (dashed) profiles, and purple lines depict the Einasto (dash-dotted) and generalized Einasto (dotted) profiles. The parameters of

 $<sup>^{2}</sup>$ As we were finalizing this manuscript, Ref. [213] presented a cored-Einasto profile for the description of density profiles in galaxy formation simulations, akin to the cored-Sersic model of Ref. [153]

	(i.) NFW	(ii.) Gen. NFW	(iii.) Einasto	(iv.)Gen.Einasto
$\frac{\mathrm{d}\ln\rho(r)}{\mathrm{d}\ln r} =$	$-\frac{3r+r_s}{r+r_s}$	$-\gamma + \frac{(\gamma - \beta)(\frac{r}{r_s})^{\alpha}}{1 + (\frac{r}{r_s})^{\alpha}}$	$-2\left(\frac{r}{r_s}\right)^{\alpha}$	$-\gamma - 2\left(\frac{r}{r_s}\right)^{\alpha}$
$r_{-2} =$	$r_s$	$r_s \frac{(\gamma-2)}{(2-\beta)}^{1/\alpha}$	$r_s$	$r_s \left(rac{2-\gamma}{2} ight)^{1/lpha}$
$\rho_{-2} =$	$\frac{\rho_s}{4}$	$\frac{2^{(\beta-\gamma)/\alpha}\rho_s}{\left(\frac{(\gamma-2)^{1/\alpha}}{2-\beta}\right)^{\gamma} \left[1 + \left(\frac{(\gamma-2)}{(2-\beta)^{\alpha}}\right)\right]^{(\beta-\gamma)/\alpha}}$	$ ho_s$	$\rho_s \left(\frac{2-\gamma}{2}\right)^{-\gamma/\alpha} \exp\left(\frac{\gamma}{\alpha}\right)$

Table 2.1: Table of convenient functions for the density profiles discussed in this work. In order these are the local power-law index, the radius at which the profile has a derivative equal to -2,  $r_{-2}$ , and the density of the profile at which the radius is equal to -2,  $\rho_{-2}$ .

these profiles are representative of a Milky Way-mass halo stack. Notice that for illustrative purposes, the plot extends over an unusually large range of halo-centric distances. We do this in order to illustrate the asymptotic behaviors of the profiles, as each profile has unique behavior at its asymptotes. The parameterization of the generalized Einasto profile allows us to decrease the asymptotic slope in the inner region of the halo.

## 2.3.2 **Profile Fitting Procedure**

Much of the analysis in this chapter relies on fits of halo density profiles to the functional forms described in Section 2.3.1. We perform fits on both stacked and individual halo profiles. Stacked profiles have compelling uses in both theory (individual halo profiles are noisy, making stacks useful for the study of general trends) and observation (e.g. the stacking of weak lenses to improve the inference of halo properties from low signal-to-noise lenses). Individual halo profiles allow us to study halo-to-halo variations. Our fitting procedure is as follows:

- 1. From ROCKSTAR catalogs and particle tables, we compute halo density profiles as follows. We bin the distribution of particle distances relative to the center of the host halo, r, into 90 logarithmically spaced bins between  $r/r_{\rm vir} = 10^{-3}$  and 1. We calculate the density of each bin by dividing mass by bin volume,  $v_i = \frac{4}{3}\pi(r_{i+1}^3 - r_i^3)$ , where *i* is the bin index.
- 2. To compute stacked profiles, we first normalize each profile by the total mass of its *subhalo-included* profile. Normalization is not necessary for individual halo fits.
- 3. To construct a stack, we calculate the mean density of the number-weighted profiles, creating a number-weighted stack. Stacks are constructed in scaled units  $(r/r_{\rm vir})$ . We calculate standard error of the density in each bin,  $SE_i = \sigma_i/\sqrt{N}$ , where  $\sigma_i$  is the standard deviation of the density of the given radial bin across all haloes of the stack, and N is the number of haloes in the stack. We use the standard error as the uncertainty in the fits.



Figure 2.2: Analytic density profile derivatives for a Milky Way-mass (MMMZ) type halo. The parameters of each profile were determined from the subhalo-*included* fits to the MMMZ stacked haloes Section 2.3.1 (NFW, generalized NFW, Einasto, and generalized Einasto). This plot extends over a large range of radii in order to clearly depict clearly the asymptotic behaviors of each profile. For reference the region between the gray vertical bands designates the typical range that a halo profile is examined in with the left band at the resolution limit and the right band at the virial radius. The red lines depict the NFW (solid) and generalized NFW (dashed) profiles, and the purple lines depict the Einasto (dash-dot) and generalized Einasto (dotted) profiles. Our parameterisation of the generalized Einasto profile allows for a shallower inner slope in the halo.

- 4. For individual halo profiles we do not need to calculate the mean. In these cases, we use  $\sigma$ , the standard deviation of the density of all the haloes in the respective simulation, as the uncertainty.
- 5. We then mask all bins with bin centers below four times the softening length of the simulation from being used in the fits  $(4 \times l_{\text{soft}} = 0.68 \ h^{-1} \text{ kpc co-moving for MMMZ}$  and  $4 \times l_{\text{soft}} = 13 \ h^{-1} \text{ kpc}$  for RHAPSODY).
- 6. Finally, we fit each functional form to the density profiles (whether derived from a stack or an individual halo) by minimizing the usual χ<sup>2</sup> function [see Eq. (6) below] using scipy.optimize.curve\_fit from the scipy Python package [364]. To expedite the fits, we include by-eye initial guesses for each free parameter. The fitting errors are those described in points (iii) and (iv). We apply the following set of bounds on each parameter to avoid un-physical solutions.
  - a.  $\rho_s > 0$ ,
  - b.  $0 < r_s < R_{\rm vir}$ ,
  - c.  $0 < \alpha < 5$ ,
  - $\mathrm{d.} \quad 0<\beta<10,$
  - e.  $0 < \gamma < 5$  for gNFW or  $-5 < \gamma < 5$  for gEinasto.

The details of the fitting procedure are as follows. First, we choose as the independent variable for each binned density  $r_i = \sqrt{r_{\text{bin,outer}} \times r_{\text{bin,inner}}}$ , the geometric mean of the radial bin edges. We evaluate the analytic density profiles at each of these values of  $r_i$ . We allow  $\rho_s$  to be a free parameter in the fits, but one may choose instead to keep it fixed, which guarantees that the fitted profile will satisfy the integral definition of the halo virial mass. There are studies that have allowed  $\rho_s$  to be a free parameters [55, 371, 97, 223] and studies that have not [272, 89, 69]. Also note that when we plot stacked profiles we normalize the profile to the mean mass of the haloes (with subhaloes *included*) so that all profiles are plotted in absolute units. For example, we plot  $r^2\rho(r)$  in units of  $M_{\odot} \text{ kpc}^{-1}$ .

Fig. 2.3 is the first depiction of a single halo's density profile in this chapter and is shown as an example. The plot corresponds to MMMZ halo 829, the same object whose projected density is shown in Fig. 2.1. On the y-axis we plot the quantity  $r^2 \times \rho(r)$  instead of just  $\rho(r)$ to make it easier to see differences amongst density profiles, given the large dynamic range of density. The gray, vertical band at the left represents the resolution limit of four times the softening length, which is equivalent to 0.68  $h^{-1}$  kpc co-moving. The meanings of the colors are matched to Fig. 2.1. The solid blue curve represents the subhalo-included density of Halo 829; i.e., the quantity normally calculated from N-body simulations, incorporating both host halo and subhalo mass. The solid orange curve depicts the subhalo-excluded density profile; mass associated with detected subhaloes is not counted in this case. The solid green curve shows a subhalo only density profile for reference. While in the inner region of the halo the two profiles are quite similar they deviate from each other in the outer region.

As part of our analyses, we also scrutinize the local power-law index,  $d \ln \rho/d \ln r$ , from both the numerical halo density profile data and the fits to the analytic profiles. Examining the local power-law index of the profiles yields insight into the profile parameters (e.g.,  $\alpha$ ,  $\beta$ ,  $\gamma$ ) that will best represent the numerical data. The derivatives of the numerical data are calculated numerically using the three-point (quadratic) Lagrangian interpolation estimator [165, 27].

#### 2.3.3 Assessing Fits

We have employed a variety of statistics to assess the quality of fits for each functional form described in Section 2.3.1. The statistics that we will examine are the root-mean-square fractional residual,  $\chi^2$ , and Akaike and Bayesian information criteria.

The first statistic we use to characterize fit accuracy is the root-mean-square fractional residual between a halo profile and its best fit,

$$\text{fRMS} = \sqrt{\frac{1}{N_{\text{bins}}} \sum_{i} \left[\frac{\rho_{\text{pred}}(r_i)}{\rho_{\text{data}}(r_i)} - 1\right]^2},$$
(5)

where  $\rho_{\text{data}}(r_i)$  is the density estimated for the simulated halo in the *i*<sup>th</sup> radial bin,  $\rho_{\text{pred}}(r_i)$  is the value of the density predicted by the model to which the data is being fit. The fRMS



Figure 2.3: The density profile of Halo 829 in the Milky Way-mass (MMMZ) simulations, the same halo shown in Fig. 2.1. The y-axis shows  $r^2$  times the density, as a function of distance from the center of the halo. The gray region represents radii below the resolution limit of the simulations, as discussed in Section 2.2. The blue, orange, and green curves are based on particles shown in the correspondingly colored panels in the previous figure; in all figures, blue indicates subhalo-included distributions, orange indicates subhalo-excluded distributions, and green indicates subhalo only distributions for reference. In the inner region of the halo the subhaloes included and subhaloes excluded profiles are very similar to each other, but they deviate more in the outer parts of the halo; this is a common trend among haloes.

residual value provides an easily interpret-able measure of the accuracy of fits; if fRMS = 0.1, the best fit of a given type typically deviates from the measured halo density profile by  $\approx 1$ per cent (in an RMS sense, so that deviations are added in quadrature). The closer fRMS is to zero, the better the match is between the best fit to a given functional form and the measured density profiles.

We use the well-known  $\chi^2$  statistic to assess quality of fit and to perform model selection from among the analytic density profiles that we propose above. The  $\chi^2$  statistic is

$$\chi^2 = \sum_i \left[ \frac{\rho_{\text{pred}}(r_i) - \rho_{\text{data}}(r_i)}{\text{SE}_i} \right]^2,\tag{6}$$

where  $mathrmSE_i$  is the standard error per bin described in Section 2.3.2, and the sum over i is a sum over all of the bins not excluded by the resolution cut. To avoid confusion, we denote  $\chi^2$  values derived from fits to stacked density profiles as  $\chi^2_{\text{stack}}$ . For individual profiles, we calculate the  $\chi^2$  values resulting from each of the individual halo profile fits. In order to discuss the fit quality for this ensemble of fits, we examine the median  $\chi^2$  value resulting from the set of fits and designate it as  $\chi^2_{\text{median}}$ .

We aim to identify the profiles that best represent the halo profiles in our simulations; however, we cannot draw this conclusion directly from the fRMS and/or  $\chi^2$  values because the functional forms used have varying numbers of free parameters. Adding additional parameters to a fitting function will always decrease both the minimum value of  $\chi^2$  and fRMS. To assess whether or not the additional parameters have intrinsic explanatory power, it is necessary to determine whether or not the decrease in the minimum value of  $\chi^2$  is significant in comparison to the decrease in  $\chi^2$  one would expect from adding parameters that simply fit the noise in the data. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are statistics that are frequently used to quantify whether or not the improvement of the fit (i.e., decrease in  $\chi^2$ ) is sufficient to conclude that the additional parameters have explanatory power [147, 148]. The AIC and BIC are

$$AIC \equiv 2k + \chi^2, \text{ and} \tag{7}$$

$$BIC \equiv \ln(n)k + \chi^2, \tag{8}$$

where k is the number of free parameters in a given model and n is the number of data points used to evaluate the fit. Differences in AIC/BIC of more than ten are generally considered to provide very strong evidence of a superior model. For example, if AIC is reduced by more than ten upon introducing a model with more parametric freedom, one concludes that the new model is a superior fit to the data with intrinsic explanatory power.

# 2.4 Results

In this section, we compare the subhaloes-included and -excluded halo density profiles, and assess how well each of the various analytic forms of halo density profile can describe the simulated data sets. In so doing, we will show that the subhaloes-included profiles are very different from the subhaloes-excluded density profiles, particularly in the outer regions of the halo, and characterize those differences.

## 2.4.1 Identifying Best Fit Forms for Stacked Profiles

The focus of this subsection is to identify which analytic halo profile fitting functions best describe the stacked profiles of simulated dark matter haloes.

Table 2.2 presents a variety of summary statistics for assessing the fit quality for each of the proposed analytic density profiles introduced in Section 2.3.1. Blue or orange colors are used to indicate tables of statistics for the subhalo-included or subhalo-excluded cases, respectively. All goodness-of-fit statistics provided are the values for fits to stacked halo profiles, with the exception of the last column, which lists the median chi-squared value from the set of fits to each individual halo in the sample. The profiles are listed in Table 2.2 in order of BIC from lowest (best) to highest (worst). It is evident that fRMS and  $\chi^2$  generally follow the same rank ordering as BIC. As a reference, the best-fit values of the parameters from all of the profile fits are provided in Table 2.4 of 2.8.

The generalized Einasto profile is the most effective model for the simulated halo density profiles, considering both Milky-Way-mass haloes (MMMZ) and cluster-mass haloes (RHAP- SODY) and for both the subhalo-included and subhalo-excluded data. It is favored over the other profiles considered by all statistics used here, except for the median chi-squared value across haloes. In this case, with the exception of the the subhaloes-excluded MMMZ data, the generalized NFW profile yields marginally smaller values. The smaller  $\chi^2$  values for the individual haloes are likely a result of the increased model complexity of the generalized NFW profile.

Having established the utility of the generalized Einasto profile, we will focus on this profile for most of the remainder of this chapter. In most of the figures that follow, we will present detailed results for the gEinasto profile and the standard NFW profile. We include the standard NFW profile because it is widely studied and therefore enables comparison with previous literature. The corresponding rows in Table 2.2 are highlighted to indicate these profiles.

We now turn our attention to the optimal model, the generalized Einasto profile, and the comparison model, the NFW profile. Fig. 2.4a and Fig. 2.4b show the stacked fits in both mass regimes. As before blue denotes subhalo-included and orange denotes subhaloexcluded. The vertical gray band indicates radii below the simulation resolution limit (see Section 2.2.1). The fit to the generalized Einasto profile is shown by the thick dashed lines. The bands around the data indicate the halo-to-halo scatter (or intrinsic scatter) of the profiles used in the stack for each bin, or the standard deviation of the halo density per bin. It is evident that the intrinsic scatter across halos of the same mass is smaller than the change in the profiles without subhaloes in the outer halo region. Therefor the effects we are seeing are not merely due to the intrinsic scatter of dark matter halo profiles. These are among the same fits used to generate Table 2.2.

The lower panels of Fig. 2.4a and Fig. 2.4b depict the ratio of the density profile fits to the simulated profiles. The x-axes align with the top panels while the y-axes show  $\mathscr{R}(\rho(r)) = \rho(r)_{\text{fit}}/\rho(r)_{\text{simulation}}$ . This ratio would lie along the black dash-dot horizontal line at  $\mathscr{R}(\rho(r)) = 1$  for a perfect fit; the smaller the deviation from the  $\mathscr{R}(\rho(r)) = 1$  line, the closer a functional form is to the stacked profile. The fits to each profile are labeled and shown in the same order as Table 2.2 of increasing BIC.

In terms of deciding upon the functional form that most faithfully represents simulated

Models	fRMS	$\chi^2_{stack}$	BIC	$\chi^2_{ m median}$					
Gen. Einasto	0.017	44.481	61.588	7.21					
Gen. NFW	0.021	70.878	92.261	6.382					
Einasto	0.033	192.502	205.332	8.949					
NFW	0.137	1843.308	1851.861	28.389					
MW-Mass Subhaloes Excluded									
Models	fRMS	$\chi^2_{ m stack}$	BIC	$\chi^2_{ m median}$					
Gen. Einasto	0.016	35.602	52.709	5.974					
Gen. NFW	0.022	66.045	87.428	6.068					
Einasto	0.055	430.868	443.698	9.744					
NFW	0.104	2026.734	2035.287	35.095					
Cluster-Mass Subhaloes Included									
Models	fRMS	$\chi^2_{ m stack}$	BIC	$\chi^2_{ m median}$					
Gen. Einasto	0.027	135.334	151.842	8.924					
Gen. NFW	0.032	188.193	208.829	7.343					
Einasto	0.041	333.317	345.698	12.841					
NFW	0.075	818.921	827.175	19.402					
Cluster-Mass Subhaloes Excluded									
Models	fRMS	$\chi^2_{stack}$	BIC	$\chi^2_{ m median}$					
Gen. Einasto	0.018	33.268	49.776	4.122					
Gen. NFW	0.027	75.339	95.975	3.505					
Einasto	0.062	428.287	440.668	8.826					
NFW	0.096	1419.749	1428.003	23.233					

# MW-Mass Subhaloes Included

Table 2.2: Statistics for evaluating goodness-of-fit and model suitability for the NFW profile, generalized NFW profile, Einasto profile, and the generalized Einasto profile. The blue tables represent fits where mass from subhaloes is included in the density profile, as has generally been done in previous simulation studies; orange tables represent cases where subhalo mass has been excluded. The first two tables correspond to the Milky Way-mass (MMMZ) host haloes and the second two tables correspond to the cluster-mass (RHAPSODY) host haloes. The two models used in the following figures are highlighted. We first list statistics calculated from fits to stacks (averages) of the haloes in each respective simulation; specifically, the the root mean square fractional residual (fRMS), the  $\chi^2$  of the best fit to the stacked haloes ( $\chi^2_{stack}$ ), and the Bayesian Information Criteria. The final column, ( $\chi^2_{median}$ ), represents the median of the chi-squared values calculated from fits to the density profiles of all individual haloes from a simulation set. The statistics used are described in more detail in Section 2.3.3. In almost all cases, the generalized Einasto profile provides the most balance of fit quality and limited model complexity; the BIC values in particular provide strong evidence that this functional form performs optimally at describing halo density profiles.

dark matter haloes, Table 2.2, Fig. 2.4a, and Fig. 2.4b all show that at both mass ranges we have the same result: the gEinasto profile is a better representation of the simulation stacks than the NFW profile or Einasto profile, and is comparable to the generalized NFW profile. The gEinasto profile has the possible disadvantage of two additional free parameters compared to the NFW profile, but this result holds even using metrics, such as the BIC and AIC, that account for additional parameter freedom. These conclusions hold for both the subhalo-included and the subhalo-excluded data sets.

Despite being described faithfully by the same functional form (the gEinasto profile), it is apparent that the subhalo-included and subhalo-excluded profiles are notably different. In particular, subhalo-excluded density profiles drop more steeply at large radii than the subhalo-included density profiles. These differences manifest themselves in different best-fit profile parameters. The mass deficit at large halo-centric distance are a result of the fact that subhaloes are preferentially found in the outer regions of haloes [394]. It is likely the case that the shallow outer slope of the NFW profile is a result of subhalo contribution, which we explore further in the following subsection. Without subhaloes, the host halos likely have higher concentrations as a result of this deficit of mass in the outer region, which we explore further in Section 2.4.3.

#### 2.4.1.1 Effective Power Law Index

One of the clearest ways to see *why* the NFW (and other functional forms) do not perform as well as the gEinasto profile is by comparing the local power-law indices, defined as  $d \ln \rho / d \ln r$ , of the simulated profiles to various functional forms. Deviations from these asymptotic behaviors imply limitations to the quality of the fit that the corresponding profile can achieve. The technique for calculating the density derivatives is described in Section 2.3.1.

In Fig. 2.5a and Fig. 2.5b, we show the local power-law indices corresponding to the profiles shown in Fig. 2.4a and Fig. 2.4b. The results shown in Fig. 2.5a and Fig. 2.5b are *not* new fits to the local power-law indices of the profiles. Rather, they show the local power-law indices implied by the fits to local density discussed above. In both plots the thin solid lines



Figure 2.4: (a) Upper panel: The stacked density profile from all 45 Milky Way-mass (MMMZ) host haloes. The gray band at low r indicates the adopted resolution limit of four times the softening length. The blue curve depicts the density profile when subhalo mass is included, as conventionally done in N-body analyses, while the orange curve corresponds to the case where subhaloes have been excluded. Both curves have been fit to a generalized Einasto profile, as defined by Equation 4 and discussed in Section 2.3.1. The thick blue dashed line is a generalized Einasto fit to the blue curve, and the thick orange dashed line is a fit to the orange curve. The filled regions around the simulation data correspond to the halo-to-halo scatter of the density profiles used in the fit, or the standard deviation of the density of these haloes as described in Section 2.3.2. Lower panel: Ratio of the density profile of the fit to the simulation data. The horizontal dashed line denotes where a perfect fit would lie. Each residual panel is labeled according to the fit to the given profile, shown in the same order as Table 2.2. It is evident that a generalized Einasto profile is a better fit to the simulation than the standard NFW profile or Einasto profile in both cases as expected from Table 2.2. While the fit to the generalized NFW profile is comparable to the generalized Einasto profile, the generalized Einasto profile provides a better fit quality and smaller model complexity. (b) As in panel (a), but for the stacked density profile of the 96 cluster-mass (RHAPSODY) host haloes. In both mass ranges it is apparent that mass in subhaloes has the largest effect on the density profiles in their outer portions.

depict the local power-law indices of the simulation data and the thick dashed lines represent the parameters resulting from the fit to the generalized Einasto profile. The shaded regions around the fits to the gEinasto profile define the 68 and 95 percentile confidence regions of the fit. These percentile regions are calculated from bootstrap re-sampling (re-sampling with replacement) of the set of host halo density profiles 1000 times. For each bootstrap sample, the fit of the stack is performed again and the corresponding power law indices are computed by substitution of the fit parameters. Then we determine the confidence regions.

Fig. 2.5a and Fig. 2.5b further demonstrate that the generalized Einasto profile is able to capture the behavior of the simulation for both the subhalo-included and the subhaloexcluded models. It is also notable that the subhalo-excluded profiles are steeper than their counterparts due to the mass deficit in the outer region of the haloes.

We compare the subhalo-excluded power-law indices for all four profiles of interest (NFW, generalized NFW, Einasto, and generalized Einasto) in Fig. 2.6a and Fig. 2.6b. The steep outer slopes of the subhalo-excluded simulation data set cannot be well described by either the NFW profile or the Einasto profiles.

For the Einasto profile, the inner power-law index approaches 0 when  $r \ll r_{\rm s}$ . The parameter  $\alpha$  allows the profile to steepen as r increases, but no single value of  $\alpha$  can capture the rate of increase of the slope on all scales. The gEinasto profile improves upon the Einasto profile by introducing a distinct parameter to capture the inner profile power-law index, making it approach to  $-\gamma$  for  $r \ll r_{\rm s}$ . With this additional freedom,  $\alpha$  in the gEinasto profile can be tuned to match the halo density profiles at  $r \gtrsim r_{\rm s}$ . In this way, the generalized Einasto profile is able to capture the shallowness of the subhalo-included profiles and the steepness of the subhalo-excluded profiles at large radii.

The contrast between the outer power-law indices of the subhalo-included and subhaloexcluded profiles has a profound implication: the outer power-law index of the mass distribution of haloes is determined largely by subhaloes. A comparison between the results shown in Fig. 2.6a and Fig. 2.6b to the general profile shown in Fig. 2.2 shows that the asymptotic behavior of the given profile at  $r \sim r_{\rm vir}$  drives the fit more than the asymptotic behavior at the inner region. Profiles with relatively shallow slope at large radii, such as the standard NFW profile with  $d \ln \rho/d \ln r = -3$ , do not faithfully describe the smooth (subhalo-excluded) components of halos. This should be considered in any application for which one must model the smooth component of the host halo and the subhaloes associated with the host halo independently.

### 2.4.2 Impact of Subhaloes on Individual Halo Profiles

So far we have found, using stacked profiles and fit quality statistics, that the generalized Einasto profile describes well the halo mass distribution both with or without the presence of the mass in subhaloes. Haloes are dynamically evolving systems, which results in halo-tohalo variation for haloes of the same mass. A functional form that fits well to a stack may not necessarily be a good fit to the individual haloes that contributed to the stack due to this variation. In this subsection, we investigate fits to the profiles of individual simulated dark matter haloes, rather than to stacked profiles. We apply the same fitting procedures described in Section 2.3.2.

For the sake of brevity we show a selection of 3 haloes from each simulation in Fig. 2.7a and Fig. 2.7b. We select the haloes closest to the  $33^{rd}$ ,  $66^{th}$ , and  $99^{th}$  percentiles in  $\chi^2$  from the subhalo-*included* fits. This corresponds to Halo 088 (483), Halo 530 (266), and Halo 606 (517) for the MMMZ (RHAPSODY) haloes. We used the same fitting procedures as Fig. 2.4a and Fig. 2.4b in axes and colors, the only difference being that the profiles are not stacked. The error-bars are representative of the standard deviation of all haloes in the respective simulation (the halo-to-halo scatter) of which the calculation is describe in (iv) of Section 2.3.2.

Some of these profiles have notable peaks in the outer region or troughs in the inner region. We investigated these deviations from the fits for these haloes. These features are caused by massive subhaloes and/or active mergers. Consequently, these features are not well described by a single monotonic profile function when subhaloes are included, and in some cases even when subhaloes are excluded (as in Halo 088), because the particles are mixed in.

Despite the noisiness of the individual halo profiles, these results are very similar to those in Fig. 2.4a and Fig. 2.4b. According to the residuals the generalized Einasto profile provides



Figure 2.5: (a) Effective power law index of the density profiles (i.e., logarithmic derivatives of the profiles) as a function of radius for the Milky Way-mass (MMMZ) haloes. Plotted here are the derivatives derived from a stack, or average, of the profiles from all of the MMMZ host haloes. The solid lines depict the derivative of the stacked simulation data, which is calculated using a 3-point derivative algorithm numerically. The dashed line shows the local power-law index implied from the fit to the generalized Einasto profile. We re-sample the 45 hosts with replacement and re-calculate the fits of these new samples to the generalized Einasto profile. Then we calculate the 68 and 95 percent regions of the local power-law index. The shaded blue and orange regions around the fit represent these bootstrapped errors. As before, the gray region represents radii below the resolution limit of the simulations. The generalized Einasto profile provides a good fit to the simulations except the very extreme outer region where the profile falls off. (b) Same as (a) but for the cluster-mass (RHAPSODY) haloes; the generalized Einasto profile is also a good descriptor of the simulations in this mass range.



Figure 2.6: (a) Similar to Fig. 2.5a, this figure shows the effective power law indices of the density profiles as a function of radius for the MW-mass haloes for all four profiles (NFW profile, generalized NFW profile, Einasto profile, and generalized Einasto profile) explored herein. We show results for the subhalo-excluded models only, as all profiles provide good descriptions of the subhalo-included simulation data set. The solid lines depict the derivative of the stacked simulation data, and the dashed line shows the power-law index implied from the fit of the simulation data to the respective profile. (b) Same as (a) but for the cluster-mass haloes. For both mass regimes the NFW profile has a much shallower outer slope than the simulation data. The Einasto profile has a much less severe but shallow prediction as well. This shallow slope means the profiles do not have enough flexibility to match the simulation data when subhaloes are excluded. It is evident that the additional parameter of the generalized Einasto profile over the standard Einasto profile allows for a larger flexibility in describing the halo density profiles.

a tighter fit than the NFW profile. Additionally, as before, the subhalo-excluded profiles are much smoother. While the shapes of the profiles are not completely identical they are much more similar to each other.

To assess fit quality, Fig. 2.8 shows the  $\chi^2$  of the individual halo fits to the generalized Einasto profile, where the *x*-axis is the  $\chi^2$  with subhaloes and the *y*-axis is the  $\chi^2$  without subhaloes. Red points mark MMMZ haloes and purple points mark RHAPSODY haloes. The black, diagonal line corresponds to the case in which both values of  $\chi^2$  are equal. The square, diamond, and star points correspond to the halos whose profiles are shown in Fig. 2.7a and Fig. 2.7b.

It is evident that the  $\chi^2$  values for the fits to the gEinasto profiles are generally smaller when subhaloes are excluded. This is the case for both simulation mass ranges, but the effect is more pronounced in the RHAPSODY haloes. This decrease in  $\chi^2$  is likely a result of the halo density profiles yielding a much smoother mass distribution after subhaloes are removed. Because we are fitting to a smooth functional form, this form can describe the halo density profile with smaller residuals.

We further investigate if this change in  $\chi^2$  is correlated with the fraction of mass in subhaloes, defined as  $M_{\rm vir,nosubs}/M_{\rm vir}$ , where  $M_{\rm vir,nosubs}$  is the total mass within  $R_{\rm vir}$  but excluding particles associated with subhaloes. The points in Fig. 2.8 are shaded according to the subhalo abundance metric; halos marked in darker colors have a smaller mass fraction in subhaloes. The haloes with a greater portion of their mass in subhaloes tend to have worse  $\chi^2$  when subhaloes are included, an effect which is substantially alleviated without subhaloes. This follows the narrative that the halo density profiles are much smoother without subhaloes and thus better described by an analytic profile. However, there is a decent scatter about this trend - the haloes with the most mass in subhaloes don't necessarily have the worst fits to the generalized Einasto profile. Thus subhaloes abundance alone cannot describe this change in  $\chi^2$ .

Overall we have shown that in addition to stacked haloes individual halo profiles can be well described by the generalized Einasto profile. Both by eye inspection (examining the profile fits) and the results of the  $\chi^2$  of the fits provide strong evidence that the generalized Einasto fits to the individual profiles are acceptable. We also show that regardless of stacks



Figure 2.7: (a) Three individual halo profiles and their respective fits from the MW-mass simulations (the axes and colors are the same as those in Fig. 2.4a). These haloes were selected according to the  $\chi^2$  of their fits *including* subhaloes. We have marked three haloes that are at the 33<sup>rd</sup> (Halo 088), 66<sup>th</sup> (Halo 530), and 99<sup>th</sup> (Halo 606) percentiles in  $\chi^2$ , representative of a good, average, and not very good fit respectively. Error bars are representative of the halo-to-halo scatter, plotted every 10 bins starting at either the 0<sup>th</sup> bin (for subhalo-included; blue) or the 5<sup>th</sup> bin (for subhalo-excluded; orange). According to the residuals there is strong evidence that individual haloes tend to be better described by the generalized Einasto profile than the NFW profile. (b) As in (a) but for the cluster-mass host haloes. The haloes at the 33<sup>rd</sup>, 66<sup>th</sup>, and 99<sup>th</sup> percentiles in  $\chi^2$  before subhalo removal are Halo 483, Halo 266, and Halo 517. As in (a), the individual haloe profiles are better fit to the generalized Einasto profile than the NFW profile.



Figure 2.8: The  $\chi^2$  of the individual halo fits to the generalized Einasto profile with subhaloes (the x-axis) plotted against the  $\chi^2$  of the individual halo fits without subhaloes (the y-axis). Red points mark MW-mass (MMMZ) haloes and purple points mark cluster-mass (RHAPSODY) haloes. For reference, the black line across the diagonal shows  $\chi^2 = \chi^2_{nosub}$ . The three points marked here by the square, diamond, and star correspond to the halo profiles shown in Fig. 2.7a and Fig. 2.7b. All of the points are colored according to a proxy for subhalo number, the halo mass fraction in subhaloes  $M_{\rm vir,nosub}/M_{\rm vir}$ . The darker the shade, the less total halo mass in subhaloes (or the fewer subhaloes are excluded. Nearly all of the RHAPSODY haloes have improved fits after subhaloes are excluded. In general the haloes with a larger mass fraction in subhaloes have smaller  $\chi^2$  after subhaloes are excluded from the fits. Without subhaloes the halo profiles are smoother and have smaller residuals compared to analytic profiles. However, the scatter indicates that subhalo abundance alone does not account for the change in  $\chi^2$ .

or individual profiles, halo density profiles exhibit a mass deficit in the outer halo region when subhaloes are excluded. Additionally the subhalo excluded component of the haloes are smoother than their subhalo included counterparts, which allows them to be better described by a smooth functional form.

## 2.4.3 The Concentration–Mass Relation With and Without Subhaloes

It is now well known that mean halo concentration is a slowly declining function of halo mass (e.g., [55, 116, 371, 396, 107, 137, 227, 199, 223, 295, 101]). This mass dependence of halo concentration is widely thought to be caused by the fact that larger haloes assemble their masses later, on average, than their less-massive counterparts. In this section, we study the degree to which subhaloes themselves influence the mass dependence of halo concentrations. We emphasize that the concentrations calculated when subhaloes are excluded use the original virial radius of the halo with subhaloes. There is no unambiguous way to define a halo without its subhaloes so here we take the simplest approach of maintaining the same halo "edge" in both cases.

The best-fit values of halo concentration for subhalo-included and -excluded are shown in Fig. 2.9. Red points denote the MMMZ haloes and purple points denote the RHAPSODY haloes. Open circles mark values calculated from the generalized Einasto fits and open triangles mark values calculated from the NFW fits. We also show corresponding histograms of the concentration values, filled histograms represent the generalized Einasto profile results and outlined histograms represent NFW profile results. The concentration parameter is calculated as described in Section 2.3.1.

It is evident that haloes of different mass ranges exhibit different changes in concentration. The MMMZ haloes roughly maintain the same concentrations, with or without subhaloes, from the generalized Einasto fits. However these same haloes have increased concentrations resulting from the NFW fits. In contrast, the RHAPSODY haloes almost all have higher concentrations without subhaloes regardless of the fit profile. In part this effect is a result of the more massive RHAPSODY haloes having a larger fraction of their mass in subhaloes and in part a result of the cluster mass haloes having smaller concentrations to start with so  $r_{-2}$  encompasses more subhaloes. However, despite this change the concentrations of the RHAPSODY haloes and MMMZ haloes do not match without subhaloes. This implies that the concentration dependence on halo mass is not purely decided by the presence of subhaloes — the smooth central halo component is also impacted by halo formation history.

In Table 2.3 we quantify the changes exhibited by the concentration after subhaloes are excluded, with similar conclusions to those described above. The errors presented are calculated via bootstraps of the haloes. In the first and third columns we present the median concentrations of the individual halo fits (for MMMZ and RHAPSODY haloes respectively). The generalized Einasto profile median concentration has a smaller increase without subhaloes compared to the NFW profile (and even decreases minimally for the MMMZ haloes). Because subhaloes are typically in the outskirts of haloes, the exclusion of the mass in subhaloes results in an increase in concentration [394].

Next we calculate the scatter of the concentration at a fixed halo mass,  $\sigma_{\log c_{-2}}$ , shown in columns two and four of Table 2.3. This is done by calculating the inter-quartile range (IQR) of  $\log c_{-2}$ . This gives an estimate of  $\sigma$ , where we assume  $\log c_{-2}$  follows a normal distribution such that  $\sigma = IQR/1.349$ . When subhaloes are excluded, the scatter in both simulations is much smaller, indicating less scatter amongst halo fitting parameters. In general the scatter is smaller in the more massive RHAPSODY haloes compared to the MMMZ haloes, which is an expected result from, e.g., Ref. [278, 107]. The subhalo-included results fall within the range of values estimated for the NFW profiles (see e.g., [180, 55, 371, 76, 278, 107]).

The final column of Table 2.3 shows the ratio of the median concentration of the Milky Way-mass MMMZ haloes to the median concentration of the cluster-sized RHAPSODY haloes,  $\frac{\text{MMMZ med}(c_{-2})}{\text{RHAPSODY med}(c_{-2})}$ . This ratio is larger than one in all cases, reflecting the fact that concentration is a slowly decreasing function of halo mass. We find that the value of this ratio computed from the *subhaloes-excluded* concentrations is smaller than the value for the standard, *subhaloes-included* concentrations. The median concentrations of the haloes of the different masses exhibit a higher degree of similarity in the *subhaloes-excluded* case. Halo concentrations still exhibit a non-negligible mass dependence when subhalos are excluded. Thus subhaloes increase the mass dependence of concentrations but do not completely explain the mass trend.



Figure 2.9: The concentration of the MW-mass MMMZ haloes (red) and the cluster-mass RHAPSODY haloes (purple) calculated from individual fits to the generalized Einasto profile (open circles) and the NFW profile (open triangles) as described in Section 2.3.1. The horizontal axis is the concentration when including the mass in subhaloes, denoted by  $c_{-2}$  With Subhaloes, and the vertical axis is the concentration when excluding the mass in subhaloes as denoted by  $c_{-2}$  Without Subhaloes. We also show corresponding histograms of the concentrations for the gEinasto profile (filled) and the NFW profile (outlined). A solid black line shows y = x; points on this line result in the same concentration values regardless of the presence of subhaloes. The MMMZ haloes show little to no change in gEisnasto concentrations without subhaloes. For NFW concentrations, there is a larger increase in both mass ranges when halos are excluded. We find that a concentration–mass relation remains, even with the exclusion of subhalo mass.

MW-Mass			Cluster-Mass		c-M Relation
Models	Median $c_{-2}$	$\sigma_{log_{10}C_{-2}}$	Median c_2	$\sigma_{log_{10}c_{-2}}$	$C_{-2}^{MW}/C_{-2}^{cluster}$
NFW	$13.666 \pm 0.7$	$0.141 \pm 0.042$	$5.484 \pm 0.17$	$0.103 \pm 0.05$	$2.492 \pm 0.155$
Gen. Einasto	$11.633 \pm 0.65$	$0.216 \pm 0.033$	$4.632 \pm 0.15$	$0.137 \pm 0.044$	$2.511 \pm 0.163$
 MW-Mass		Cluster-Mass			
	MW-Mass		Cluster-Mass		c-M Relation
Models	MW-Mass Median c <sub>-2</sub>	$\sigma_{log_{10}c_{-2}}$	Cluster-Mass Median c <sub>-2</sub>	$\sigma_{log_{10}c_{-2}}$	c-M Relation $c_{-2}^{MW}/c_{-2}^{cluster}$
Models NFW	MW-Mass Median c <sub>-2</sub> 15.987 ± 0.48	$\sigma_{log_{10}c_{-2}}$ 0.118 ± 0.038	Cluster-Mass Median c <sub>-2</sub> 8.46 ± 0.2	$\sigma_{log_{10}c_{-2}}$ 0.077 ± 0.027	c-M Relation $c_{-2}^{MW}/c_{-2}^{cluster}$ 1.89 ± 0.087

Table 2.3: Tables of the median concentrations and scatter for the simulated haloes. The concentrations are calculated as described in Section 2.3.1. As before the blue table shows results of subhalo-included models and the orange table shows results for subhalo-excluded models. The median concentrations are simply the medians of the values computed for each individual halo for the given model, with errors from bootstrapping. We also list concentrations of the stacked halo profiles in Table 2.4. The scatter  $\sigma_{\log c_{-2}}$  describes the scatter in the concentration–mass relation. Concentration is simply a transformation of the halo parameters and makes it easy to study the differences, namely that the scatter is much smaller for subhalo-excluded models indicating that their parameters are more similar. The final column  $c_{-2}^{MW}/c_{-2}^{cluster}$  is the fraction of the median concentration is much smaller for subhalo-excluded models models indicate that subhalo-excludes have an effect on the concentration–mass relation but do not completely explain the trend.

#### 2.4.4 Robustness Checks and Caveats

We have performed several tests to ensure the robustness of our results. First, we examined the results when stacking subsets of haloes according to additional halo properties to see if this had an effect on the fit assessment (such as subhalo abundance). We have tabulated the same statistics as those in Table 2.2 when stacking subsets of the haloes split according to their subhalo mass fraction (see Section 2.4.2 for further discussion on this quantity) and other such proxies for subhalo abundance such as metrics for halo formation time (as halo formation time is anti-correlated with subhalo count). In general the trends were the same as those shown for the full stack, barring a couple of edge cases where the stack was influenced by a few haloes undergoing mergers.

In general, we present fRMS values calculated using the logarithmically spaced bins defined in (i) of Section 2.3.2. Logarithmic bins have the effect of preferentially weighting the inner portions of halo profiles relative to a binning scheme linear in radius, r. To do that we computed a linear fRMS by interpolating densities onto a linear grid and recomputing each of the fits. We have confirmed that using a linear binning scheme does not alter the qualitative results and preserves the rank ordering of fRMS among haloes in the sample.

A separate concern is that the halo profiles have not turned over yet at the virial radius, especially for the more massive RHAPSODY haloes. The edge of a halo is arbitrarily defined. Because of this there is the possibility that the virial radius may not be the physical edge of the halo. In terms of density profiles this means that the density is not yet at the point where the slope is diverging. We have examined and fit profiles extending to both  $1.5 \times r_{\rm vir}$ and  $2 \times r_{\rm vir}$  and find that this has no effect on our qualitative results with minor effect on our quantitative results.

In addition to these consistency checks, our results and conclusions are subject to several caveats. First, concentrations were determined via a specific algorithm for halo profile fitting. While we expect the qualitative aspects of our conclusions to be unaltered by the application of distinct algorithms, some of the quantitative details of the results may be sensitive to the algorithm used to determine halo concentrations.

Second, the detailed results presented here are dependent on what one defines as "sub-

structure." In Section 2.2.2 we define streams, caustics, and loosely self-bound objects that are actively being disrupted as part of the smooth halo component. Including these objects in the substructure component would reduce the mass of the smooth halo component and may also systematically alter the structure of the smooth halo component. For further detail and examples of an alternative way of defining the smooth halo component see 2.7 and relevant figures. Despite having a vastly different smooth halo component, the generalized Einasto profile still has the smallest BIC, fRMS,  $\chi^2$ , etc for subhalo excluded fits.

In this work we use the default binding criterion defined in ROCKSTAR (see Ref. [18] for further detail), or the limit that a subhalo must meet in order to be self bound. Using different halo finders and different binding criteria can result in slightly different substructure abundances. However, we do not expect the differences between results from various halo finders to be severe, as Ref. [285] found that substructure presences agreed quite well across halo finders. But these small difference in turn can slightly alter what is counted as "substructure" and what is attributed to the smooth halo component. While we use the criterion that 50 per cent of particles must be bound to a subhalo, Ref. [18] finds that effects don't manifest until a threshold of 15 per cent or lower. We expect our results to hold, at least qualitatively, for different subhalo self-binding criteria especially because the subhalo mass fraction is generally small compared to the overall halo mass.

Third, as with all simulation-based studies, the simulations that we analyzed had finite force and mass resolution. We again expect that our conclusions are not compromised by finite resolution, but several quantitative aspects of the results may be resolution dependent. Most notably, we provide best-fit parameters for each profile in 2.8. It may be tempting to take the inner profile power-law indices ( $\gamma$ ) to represent the asymptotic inner power-law indices of halo density profiles, as this is a quantity of interest across many sub-disciplines. However, we caution that the specific power-law indices quoted are likely resolution dependent and our analysis cannot to rule out profiles that become shallower than  $\rho \propto r^{-1}$  at small radii.

Resolution can alter our quantitative results in at least one additional way. As resolution increases, smaller subhaloes will be resolved within the simulation. Consequently, removing "subhaloes" according to the definition used herein is inherently resolution dependent. Higher resolution simulations will contain smaller subhaloes not present in relatively lower-resolution counterpart simulations. It is possible that this additional substructure may alter our results. Therefore it is useful to consider the amount of mass in subhaloes that may not be identified as mass associated with subhaloes due to our resolution. An estimate of this non-identified subhalo mass is calculated using an analytic power-law subhalo abundance function [335]. We find that the cumulative mass in subhaloes below our resolution limit to be approximately  $\sim 15\%$  of the mean MMMZ halo mass.

While 15% is a non-negligible mass fraction, we emphasize that it is the *spatial bias* of these very low-mass subhaloes relative to the overall halo mass distribution that would cause our results to change. If these subhaloes are not biased relative to the overall dark matter halo mass distribution, then their removal constitutes merely a change in profile normalization and not a change in profile shape. Several studies have examined the spatial distribution of subhaloes with respect to their hosts. These studies generally find that the most massive subhaloes do exhibit a spatial bias with respect to the overall halo matter distribution (massive subhaloes preferentially lie in halo outskirts) and that this bias decreases with decreasing halo mass [393, 394, 392, 340, 269, 303, 335, 135, 155]. This suggests that the subhaloes that are not identified within our simulation will not induce significant qualitative changes to our results. A definitive answer to this question awaits further, higher-resolution simulations. We close this discussion of subhaloes that may be missing due to resolution by noting that the definition of a halo or a subhalo that is most suitable depends upon the data with which the prediction is to be compared and the specific analyses that will be performed.

Fourth, we have only examined two narrow halo mass bins here. Therefore, we cannot make specific statements about the detailed mass dependence of any of the effects that we have explored, including the concentration–mass relation. An interesting extension of this work would be to consider the profiles of haloes over a wider range of masses in order to construct an improved halo concentration–mass relation based on smooth profiles.

Lastly, recent work such as that by Ref. [64] finds that satellites are more concentrated than subhaloes. Baryonic effects may alter our quantitative results along these lines; while in N-body simulations the subhalo mass loss occurs only in the outer regions, it may occur at smaller radii when baryonic physics is introduced. For example, Ref. [102] showed that substructure orbiting in the inner galaxy region can be destroyed by disk shocking. Substructure abundances can be reduced by as much as a half [140, 317]. While this effect predominantly occurs within the inner 30 kpc of the halo, this is within the range that we see a notable offset between subhaloes included and subhaloes excluded. This could mean that the effect subhaloes have on host halo density profiles is less severe, but we do expect some effect to remain. Another interesting extension of our work would be to explore the effects of removing subhaloes in hydro-dynamical models.

## 2.5 Summary and Conclusion

In this chapter, we have investigated the density profiles of the smooth components of host dark matter haloes and compared them with conventional halo density profiles. Typically, the density profiles of host dark matter haloes are analyzed including all of the mass associated with subhaloes within these hosts. Here, we isolate the smooth components of the host haloes by removing the mass associated with subhaloes (following earlier work by Ref. [385]), and study the resultant smooth host density profile.

We examine the difference between the smooth and conventional density profiles for a set of high-resolution simulations of 45 Milky Way-sized haloes (the MMMZ haloes, with  $M_{\rm vir} = 10^{12.1\pm0.03} M_{\odot}$ ), and a set of simulations of 96 cluster-sized haloes (the RHAPSODY haloes, with  $M_{\rm vir} = 10^{14.8\pm0.05} M_{\odot}$ ). Considering profiles at different masses is a priority because the amount of halo substructure is known to increase systematically with halo mass [394]. However, the prerequisite of high resolution precludes an exploration of a wide range of halo masses. Studying high-resolution simulations of Milky Way- and cluster-sized haloes is a compromise between these considerations.

We have drawn four primary conclusions from our work, which can be summarized as follows.

(i) The density profiles of the smooth components (i.e., excluding mass within subhaloes) of host dark matter haloes decline more steeply at large radii compared to conventional density profiles that include both smooth mass and mass within subhaloes (see Fig. 2.3,

Fig. 2.5a, Fig. 2.5b, Fig. 2.6a, Fig. 2.6b).

- (ii) A single functional form, the generalized Einasto (gEinasto) profile, describes all of the profiles that we have studied, including both smooth (subhaloes-excluded) and conventional (subhaloes-included) profiles, with a smaller residual error than either the often-used NFW or Einasto profiles (see Table 2.2, Fig. 2.4a, Fig. 2.4b, Fig. 2.7a, Fig. 2.7a).
- (iii) We find that concentrations  $(c_{-2})$  derived from the smooth halo density profiles exhibit a weaker dependence upon mass than the concentrations derived from conventional density profiles including subhaloes. This indicates that substructure plays an important role in establishing the concentration-mass relation (see Table 2.3, Fig. 2.8, Fig. 2.9).
- (iv) Concentrations derived from the density profiles of the smoothed components of haloes exhibit exhibit smaller scatter at fixed mass than conventional concentrations. This indicates that substructure plays a role in establishing the distribution of halo concentrations at fixed mass (see Table 2.3).

Each of these conclusions has a variety of important consequences. We elaborate on each of points (i)–(iv) in turn below.

The prevalence of substructure is a natural consequence of CDM. It is also known that subhaloes are distributed within their hosts differently than the mass distribution [270, 394]. One consequence of this difference is that the smooth component of a halo has a density profile that is different from its total mass density profile including subhaloes, declining more rapidly than the full halo density profile. This distinction may be relevant to studies that aim to understand the nature of the nearly universal density profiles of haloes, and may impact analyses of a variety of observations. The effect that we measure is also broadly consistent with recent work describing the influence of mergers on halo concentrations [369].

Consider the analysis of a hypothetical gravitational lens system as an illustration of the importance of the distinction between the mass in the smooth component of the host and its subhaloes. The mass distribution of the lens system can be constrained through the observation of the magnified/distorted images of the source galaxies behind the lens system. A common strategy is to treat the bulk of the lens mass as an NFW profile. However, visible substructures (satellite galaxies) or invisible substructure are treated as haloes with their own, distinct profiles. The problem with this scheme is that the NFW profile used to describe the main lens system is already calibrated to include the mass in substructure. This new, composite lens system, built from an NFW main halo and distinct subhaloes, no longer represents the mass distributions found in CDM halo simulations. One might informally say that the subhaloes are "double counted." The model would represent the predictions of simulations more faithfully if the main lens were modeled using a profile calibrated to the smooth component of the host halo alone. This work provides such profiles.

Similarly, halo density profiles determine the luminosities of extra-galactic sources of dark matter annihilation. The boost factors associated with the balance between the smooth halo component and subhaloes could be altered by the differences between the two as we have shown such profiles are markedly different.

We demonstrated that the "generalized Einasto" (gEinasto) profile introduced in Ref. [297] provides a better description of dark matter halo density profiles than either the NFW or Einasto profiles (or several other candidate profiles, see 2.6), in support of the conclusions of Ref. [342]. Indeed, in all cases that we have studied, including MMMZ (Milky Way-sized) haloes and RHAPSODY (cluster-sized) haloes, both with and without including the mass within subhaloes in the density profiles, the gEinasto profile provides a superior description of the dependence of halo density on halo-centric distance. As evinced in Fig. 2.2, Fig. 2.6a, and Fig. 2.6b, the Einasto profile has the shortcoming that it cannot describe the variation of the local power-law index d ln  $\rho/d \ln r$  as a function of r with a single value of  $\alpha$  for the entire halo. The gEinasto profile is an Einasto-type profile with an additional free inner power-law index. This additional freedom enables the Einasto profile to describe halo density over the entire range of resolved halo radii. A gEinasto fit to stacked halo density profiles describes the haloes with residuals smaller than a few per cent in all cases.

Concentration  $(c_{-2})$  is a convenient dimensionless characterization of the scale at which a halo profile steepens. Here we find that the concentrations of the smooth halo density profiles have a weaker dependence on mass than conventional concentrations, which is shown in Fig. 2.9 and Table 2.3. First, we find that the concentrations for the subhalo-excluded profiles
(for the gEinasto profile) are up to  $\sim 30$  per cent higher than concentrations derived from the conventional halo mass distribution for individual haloes (and up to  $\sim 35$  per cent higher for the stacks). This is in agreement with results from past work (e.g., Ref. [394, 237, 124]) that indicated that haloes with higher concentrations tend to have fewer subhaloes. We also find higher halo concentrations when subhaloes are removed. However, a mass dependence on concentration remains, even without subhaloes in the picture. Thus the concentration–mass relation is only partially explained by subhaloes.

At fixed mass we also find that the smooth profile concentrations have a significantly smaller scatter than concentrations with subhaloes included. Some scatter remains after subhalo removal. One possible and well explored source of this scatter is the halo formation history, where recently forming haloes typically have lower concentration (see e.g., Ref. [371, 396]). Another possible source of this remaining scatter is environment, as work by Ref. [228] found that more concentrated haloes live in denser environments at fixed mass. This is consistent with the now well-known concentration-dependent clustering of haloes [372, 138].

We have shown that subhaloes have a prominent effect on the profiles of dark matter haloes at larger radii  $(r \gtrsim r_{-2})$ , and have given fitting formulae that encapsulate this effect. As we collect ever more precise data at a variety of observational facilities, we hope that this new accounting for halo mass will enable more powerful and less biased data analyses. The effort to understand the role of subhaloes in determining halo properties is only beginning. Given this work, it is reasonable to suspect that subhaloes may influence a variety of halo properties. While a detailed exploration is beyond the scope of this work, future studies of such effects may yield tools which can further improve data analyses and deeper insights into the formation and evolution of dark matter haloes.

# 2.6 Other Halo Profiles

There have been numerous papers discussing various profiles that describe dark matter haloes. Most are variations of the double power-law generalized NFW profile (Equation 2) with power-law indices  $\alpha$ ,  $\beta$  and  $\gamma$  set to various specific values. There has been more recent work that also studies varieties of continuously varying power laws, namely modifications to the Einasto profile (Equation 3). In addition to the primary results presented in the main text, we investigated fitting dark matter halo density profiles to the following analytic forms.

- 1. The generalized NFW profile with various constraints on  $\alpha$ ,  $\beta$ , or  $\gamma$ . E.g.,  $(\alpha, \beta, \gamma) = (1, 3, \gamma); (1, \beta, \gamma); (\alpha, \beta, 1.58).$
- 2. The Generalized Moore profile [260]  $(\alpha, \beta, \gamma) = (3 \gamma, 3, \gamma)$ .
- 3. The Denhen & McLaughlin Profile [89]  $(\alpha, \beta, \gamma) = (\frac{4}{9}, \frac{31}{9}, \frac{7}{9}).$
- 4. The Generalized Denhen & McLaughlin Profile [89]  $(\alpha, \beta, \gamma) = (\frac{3-\gamma}{5}, \frac{18-\gamma}{5}, \gamma).$
- 5. The DiCintio [97] model  $(\alpha, \beta, \gamma) = (0.84, 2.85, 1.09)$  using the equations in their paper, and  $M_* = 5 \times 10^{10}$  and  $M_{\text{halo}} = 1.3 \times 10^{12}$  for the Milky Way from Ref. [28].
- 6. A log parabola (or curved power law), as it was a good descriptor of the effective power law index of our profiles (i.e., the derivative of the log of the density with respect to log r, which is discussed in Section 2.4.1.1). This is expressed by  $\rho(r) = (\frac{r}{r_s})^{-\alpha \beta \ln (r/r_s)}$ .

None of these profiles performed as well as the generalized NFW or generalized Einasto profiles. We mention them here for completeness for the reader.

#### 2.7 Subhalo Removal with Rockstar

Here we continue the discussion of how subhaloes are excluded from calculations. Cleanly identifying particles that belong to detectable subhaloes and distinguishing them from particles that belong only to the host and not to any subhalo is not straightforward. Objects identified as subhaloes in the halo table do not correspond one-to-one to the set of overdensities visible in particle distributions. For example, ROCKSTAR does not list subhaloes with low self-bounded particle fractions in the halo catalog (< 50 per cent self-bound particles). These are structures that are very diffuse and do not fall within the strict "subhalo" definition and are more commonly associated with tidal streams. However, some of these objects are very small and not real, so one must define a mass and self-bound fraction cut in order to get high purity results. Please refer to the ROCKSTAR paper, Ref. [18], for details on binding criteria.

In Section 2.2.2 our "subhalo-excluded" particles are those that are not associated with any subhalo listed in the ROCKSTAR halo catalog. This sample includes objects that do not meet this bounded-ness threshold in ROCKSTAR. However, it is also possible to select particles that are *only* tagged as host particles. This would mean excluding both particles that belong to substructure that meets ROCKSTAR's criteria *and* particles that are part of other more diffuse substructures.

In Fig. 2.10 we show a 2D halo slice ( $\Delta z = \pm 1 \text{kpc}$ ) comparing this alternative subhaloexcluded method to that of Section 2.2.2 and Fig. 2.1. The left panel depicts the particles of this alternative subhalo-excluded model in black. The right panel of Fig. 2.10 depicts the particles that are in our fiducial model but are not included in the alternative subhaloes excluded model. There is a significant portion of mass in diffuse material that is not yet part of the host halo according to ROCKSTAR, but likely would be in a few time steps. Thus we use this more realistic interpretation as our host halo.

The alternative subhalo-excluded method is an even smoother depiction of the host halo. If we then plotted the opposite of this to show subhaloes only, there would be a much more notable buildup of particles at the center of the halo caused by the diffuse substructure. This alternative subhalo-excluded method is less physical as haloes would indeed contain diffuse over-density peaks near the center as a result of hierarchical buildup. Hence we warn the reader about using this alternative subhalo-excluded definition.

Fig. 2.11a and Fig. 2.11b show what these differences look like when shown as a halo density profile. These plots are identical to the simulation data depicted in our other stacked figures, i.e., the upper panels of Fig. 2.4a and Fig. 2.4b. In addition we provide a stacked profile for the alternative method of subhalo-excluded depicted by the black line. Using this definition of subhalo exclusion more strongly impacts the inner portion of the halo in addition to the outer portion of the halo. This method over-suppresses the inner mass of a realistic halo, as we expect the diffuse/smooth halo component to partially consist of old disrupted subhaloes.



Figure 2.10: *Left*: alternative subhalo-excluded particles, depicted in black. This figure has the same z-axis cut at Fig. 2.1 and the virial radius of the host halo is overlaid. *Right*: The subset of particles that exist in the fiducial subhalo-excluded method presented in the chapter that do not exist in the set of the alternative subhalo-excluded particles. I.e. particles in the middle orange panel of Fig. 2.1 that do not overlap with the particles shown in the left panel of this plot. It is apparent that in this alternative subhalo-excluded method there is no evidence of other minor over-densities and the host is completely smooth. Additionally, there is a significant portion of mass that does exist in diffuse material that is not yet classified as part of the host halo in our fiducial model.



Figure 2.11: (a) The same simulation data depicted in the upper panel of Fig. 2.4a, i.e. a stacked density profile of all 45 of the MW-mass MMMZ host haloes. The blue curve depicts the density profile for subhalo-included, the orange curve depicts our simulations after subhaloes have been excluded, and the black curve depicts the simulations for the alternative method of full subhalo-excluded. (b) The same data depicted in the upper panel of Fig. 2.4b, or a stacked density profile of all 96 of the cluster-mass RHAPSODY host haloes depicting both methods of subhalo-excluded. It is evident that this method excludes a significant portion of mass in the inner halo region in addition to the outer halo, which is likely a less physical representation of an observed halo.

# 2.8 Best Fit Profile Values

For reference in Table 2.4 we provide the best fit parameters for the stacked haloes to each profile discussed in the text in addition to the concentration computed from these parameters, as per Section 2.3.1. We emphasize that the  $r_s$  provided is not the inverse of the concentration, as our concentrations are calculated from  $r_{-2}$  and not  $r_s$ .

# MW-Mass Subhaloes Included

Models	r <sub>s</sub>	α	β	γ	C_2
NFW	0.0758	N/A	N/A	N/A	12.85
G.NFW	1.0	0.58	5.1 1.2		10.52
Einasto	0.0859	0.17	N/A	N/A	11.88
G.Einasto	0.6821	0.34	N/A	0.98	10.53

# MW-Mass Subhaloes Excluded

Models	r <sub>s</sub>	α	β	γ	C_2
NFW	0.0606	N/A	N/A	N/A	15.86
G.NFW	0.6013	0.7	5.8	1.3	11.51
Einasto	0.0707	0.2	N/A	N/A	13.91
G.Einasto	0.5457	0.4	N/A	1.1	11.78

**Cluster-Mass Subhaloes Included** 

Models	r <sub>s</sub>	α	β	γ	C_2	
NFW	0.1837	N/A	N/A	N/A	5.44	
G.NFW	0.9962	0.78	4.64	1.19	4.49	
Einasto	0.2041	0.19	N/A N/A		4.9	
G.Einasto	0.9999	0.49	N/A	1.04	4.42	

Cluster-Mass Subhaloes Excluded

Models	r <sub>s</sub>	α	β	γ	C <sub>-2</sub>
NFW	0.118	N/A	N/A	N/A	8.48
G.NFW	0.5961	1.05	4.9	1.26	6.34
Einasto	0.1308	0.26	N/A	N/A N/A	
G.Einasto	0.5688	0.64	N/A	1.11	6.29

Table 2.4: Fit values for the profiles of interest in this work, described in Section 2.3.1. The  $r_s$  are scaled by the median  $r_{\rm vir}$  as in our plots. These are fit values that result from the stacked haloes after being fit to each listed profile. Values marked as "N/A" mean that this profile does not have that parameter in it.

## 3.0 Predictably Missing Satellites: Subhalo Abundances in Milky Way-like haloes

This chapter has been accepted for publication in the Monthly Notices of the Royal Astronomical Society ©: 2019 Catherine Fielder. Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved. Minor modifications have been made to the text.

# 3.1 Introduction

Simulations of structure formation based upon the  $\Lambda$ CDM (cold dark matter with a cosmological constant) cosmological model successfully describe a wide range of observations, particularly at large scales ( $r \gtrsim 10 \text{ Mpc} h^{-1}$ ). However, a number of observations on smaller scales ( $r \lesssim 1 \text{ Mpc} h^{-1}$ ) exhibit possible discrepancies with the standard model [128, 258, 198, 260, 43, 93, 57]. Many of these discrepancies have been observed in the Milky Way (MW), the galaxy we are able to study in the most detail.

In this chapter we focus most directly on an issue known as the 'missing satellites problem' (MSP). The MSP is the apparent over-prediction of the abundance of satellite haloes of a particular velocity dispersion (or rotation speed) within a  $\Lambda$ CDM model relative to the number of galaxies of similar velocities that have actually been observed in the Milky Way (e.g., [198, 260, 56]). We expect substructure to survive the hierarchical assembly of dark matter (DM) haloes, as the dense cores of merging haloes are not strongly affected by tidal interactions [188, 393, 56] but the question of how much substructure there should be remains.

We describe objects in terms of maximum circular velocity,  $V_{\text{max}}$ , which is a measure of the depth of the potential well of a dark matter halo. The critical range of maximum circular velocity we are interested in for studying the MSP is the range of  $10 - 30 \text{ km s}^{-1}$ , where other physical effects do not dominate (gas cooling on the low velocity end and photoionization feedback on the high velocity end). There has only been ~ 26 dwarf satellites observed of this size, compared to roughly ~ 350 predicted from CDM simulations [344, 56] with  $V_{\rm max} > 10 \,\rm km \, s^{-1}$  within 400 kpc of a Milky Way-size host. Of these ~ 100 should be massive enough to host observable galaxies [56]. The MSP is the difference between the number of predicted subhaloes of the size necessary to host galaxies with the observed stellar kinematics of the classical Milky Way dwarf satellites and the number of actual dwarf satellites. Within the  $\Lambda$ CDM framework, most such subhaloes cannot host dwarf satellites comparable in size to the classical MW dwarfs. Many previous works have tried to understand this discrepancy via some baryonic processes (e.g., supernova feedback) or through exotic physics (e.g., alternative models of dark matter).

The missing satellites problem is not the only example of potential small-scale discrepancies from ACDM expectations found in the Local Group. The simplest possible relationship between galaxies and dark matter would place the known satellites of the Milky Way into the largest subhaloes surrounding it [57]. The 'too-big-to-fail' (TBTF) problem was identified based upon ACDM N-body simulations [335, 99] which showed that the most massive subhaloes were too dense to host the the brightest Milky Way satellites [43, 44]. Naively, these dense subhaloes should form stars more efficiently than their more diffuse counterparts, yet they do not appear to host the Local Group dwarfs [343, 197].

A number of solutions have been proposed to alleviate or eliminate the MSP and TBTF problems. Some of the most well-studied proposals include:

1. Survey incompleteness: The missing objects could be too dim or too diffuse to have been detected yet. This question could partially be reconciled with future survey projects such as LSST, a hope that has gained traction after the numerous satellite galaxy discoveries from SDSS and DES [29, 106], although most of these dwarfs are predominately much smaller than the regime that we are interested in  $(V_{\text{max}} \leq 10 \text{ km s}^{-1})$ . Recent work by Ref. [195] argues that there is no longer any MSP based upon our current understanding of the suppression of star formation in the satellites and the recent discovery of faint dwarfs after being completeness corrected to the Milky Way's virial radius. They predict that the number of satellites that inhabit the Milky Way is consistent with CDM predictions. However, there is still some issue with objects smaller than Segue I ( $V_{\text{max}} = 10^{+7.0}_{-1.6} \text{ km s}^{-1}$  from Ref. [178]) in their model being over-predicted. In this case, the observed abundance of satellites can be explained, but the question of why subhaloes of similar potential well

depth may host galaxies of very different luminosities would still be open.

- 2. Host halo properties: Ref. [368] pointed out that the Milky Way's mass is uncertain by roughly a factor of two; if the MW mass is on the low side of this range, then both the missing satellites and TBTF problems are greatly alleviated. In addition, several studies have shown that there is significant scatter in subhalo abundances among host haloes at fixed host halo mass [299, 179, 358]. Both Ref. [299] and Ref. [179] argued that the TBTF problem is only marginally significant given this large scatter, even if the mass of the Milky Way halo is on the high end of the observed range.
- 3. Baryonic effects: Both MSP and TBTF are issues that arise when comparing observations with N-body simulations. However, baryonic physics can potentially change the observable satellite population substantially. Galaxies in small subhaloes may be dim because their host haloes did not form stars significantly prior to reionization, at which point their gas was stripped by the UV background [58, 23, 259, 40]. Milky Way subhaloes may have experienced a stronger suppression caused by the radiation from the MW itself. Photoionization is expected to affect haloes in the range of  $V_{\rm max} \sim 30 \,\rm km s^{-1}$ [111, 58, 23, 40, 316, 57].

Likewise, supernova feedback heats and blows out the interstellar medium, which can suppress star formation in low-mass dark matter haloes [91, 262, 318, 202]. This process generally effects larger subhaloes of  $V_{\rm max} \sim 100 \text{ km s}^{-1}$ . Winds and UV radiation from massive stars may also deposit energy into the gas within a subhalo, enhancing this effect [66].

Hydrodynamic simulations have backed the baryon solution up, showing that adding baryonic physics into dark matter-only models can substantially alleviate the missing satellites problem [401, 316, 400]. For example, dynamical friction resulting in angular momentum transfer from baryons to dark matter alters the dark matter density profile at the centres of low-mass galaxies [47], making them more cored than cuspy. These objects have less visible stars due to stronger effects of tidal stripping on cored density profiles [92].

Tidal stripping by the host halo or the parent galaxy during accretion could have torn apart some fraction of the lower mass satellites [100, 28, 140, 268].

4. Non-cold dark matter and exotic physics: Dark matter properties which deviate from CDM could reduce the number of subhaloes of appropriate potential well depth to host the observed satellite galaxies, thus alleviating or eliminating the MSP and the TBTF issue [333, 334, 393, 370]. Additionally Ref. [186] and Ref. [393] argue that non-standard inflationary scenarios could alter the initial conditions for structure formation, yielding fewer satellites at lower mass scales and alleviating small-scale issues such as the MSP.

Among these possible solutions, the study of the impact of host halo properties has mostly been limited to host halo mass. In N-body simulations, including the original simulation from which the haloes in this work were drawn, the abundance of resolved subhaloes is found to be directly proportional to the mass of the host halo over several orders of magnitude in mass (e.g., [136, 204, 45]). However, other halo properties, such as halo concentration and spin, correlate with the subhalo population as well; incorporating their effects will yield improved predictions of subhalo populations. This will enable more direct comparisons of the satellite abundances around the Milky Way to the subhalo populations of a halo whose properties match our Galaxy's.

In this chapter, our first goal is to use high resolution ACDM zoom-in simulations for Milky Way-mass haloes (described in Section 3.2.1) to explore the nature of the host-tohost scatter in subhalo populations. Our study seeks, in part, to determine which host halo properties determine the distribution of subhalo abundance at fixed host halo mass. We then make predictions for substructure abundances in the Milky Way which incorporate observational constraints on the properties of its host dark matter halo, and explore the impact on the missing satellites problem.

A variety of data has provided constraints on various properties of the dark matter halo which hosts the Milky Way other than its total mass. For example, kinematics of halo stars [87] and masers [277] have constrained the Milky Way halo concentration by constraining its mass distribution, while the tidal stream of the Sagittarius dwarf [212, 362] has constrained the halo's shape. We may grossly infer its angular momentum as well. For example, Ref. [218] show that the Milky Way's stellar disk has a scale length roughly a factor of two lower than would be typical given its mass (or luminosity) and rotation speed, lying further from the luminosity-velocity-radius relation than roughly 90% of spirals. This small scale length would be expected to be related to its halo spin parameter [256]. As a result, the atypical scale length of the Milky Way suggests that its dark matter halo may also be unusual; we wish to explore any impact this has on the abundance of subhaloes in our Galaxy.

The "tailor-made" prediction of Milky Way subhalo abundance which we obtain in this chapter is relevant to our interpretation of the small-scale challenges to  $\Lambda$ CDM in many ways. Suppose, for instance, that we assume the most conservative scenario in which any small-scale issues can be completely resolved by invoking baryonic processes rather than by introducing any new physics. We will show in this chapter that it is reasonable to expect that the Milky Way host halo includes roughly one-fourth fewer than average smaller subhaloes and as much as three-fifths fewer than average larger subhaloes than would be typical given its mass, based upon the correlations of satellite abundance with other Milky Way halo properties. In that case, the impact of baryonic physics on the MSP and TBTF problems must be significantly smaller than has been assumed in the past (as otherwise we have observed *more* satellites around the Milky Way than would be expected in  $\Lambda$ CDM).

This has important consequences for the tuning of the parametrized models of baryonic physics used in simulating galaxy evolution. For example, if we suppose that the MSP is resolved largely by baryonic feedback, then this feedback may need to be significantly less efficient than previously thought, since the Milky Way should be expected to have fewer subhaloes to begin with than a typical halo of its mass. Such alterations to baryonic physics models may have extensive impacts on simulations of dwarf galaxy formation and of galaxy evolution more generally. In addition, our study also provides a theoretical context for interpreting the observed satellite luminosity functions of Milky Way-like hosts that are outside the Local Volume [141], with which we can test how impactful host halo properties are on the satellite populations by correlating them with proxies for halo properties such as disk scale length. The general strategy that we describe and advocate in this manuscript will become increasingly useful as more and more becomes known about the Milky Way galaxy and the halo of the Milky Way.

The structure of this chapter is as follows. First, in Section 3.2 we describe the basic host dark matter halo properties we focus on in this chapter – spin, shape, concentration, and merger history – and describe what is known about each for the Milky Way. In Section 3.3, we

show that host halo properties are strongly correlated with the total abundance of subhaloes above a threshold in circular velocity. We then examine what model based upon the host halo properties provides the best predictions for satellite abundance for Milky Way-like hosts, and evaluate that model using the observed properties of the Milky Way host halo. In Section 3.4 we conclude that due to the somewhat unusual formation history of the Milky Way's host halo, we expect that it should have fewer subhaloes than typical for its mass, and discuss some implications and caveats. Additionally, in Section 3.5 we explain the details of obtaining a rough concentration estimate for the Milky Way when incorporating adiabatic contraction, and in Section 3.6 we discuss in detail the numerical and mathematical techniques utilized in this work and provide fitting functions for estimating subhalo abundance based upon host halo properties.

## 3.2 Milky Way Halo Properties

The following subsections detail the host halo properties we use in our analyses and include approximate estimates for the Milky Way. We stress that the values for the Milky Way are subject to both measurement and modelings errors. The most important take away is the approximate rank of the Milky Way in order to compare it to other dark matter haloes.

#### 3.2.1 Zoom-in Simulations

In the analyses presented here, we use a set of zoom-in cosmological simulations consisting of 45 Milky Way-mass haloes. These haloes were selected from a 125 Mpc  $h^{-1}$  parent simulation containing 1024<sup>3</sup> particles. The cosmological parameters for the simulations are  $\Omega_{\rm M} = 0.286$ ,  $\Omega_{\Lambda} = 1 - \Omega_{\rm M} = 0.714$ , h = 0.7, mass fluctuation amplitude  $\sigma_8 = 0.82$ , and scalar spectral index  $n_{\rm s} = 0.96$ . All of the Milky Way-analog haloes selected for re-simulation fall within the mass range of  $M_{\rm vir} = 10^{12.1\pm0.03} \,\mathrm{M}_{\odot}$ . The mass of the highest-resolution particles in the zoom-in simulations is  $m_{\rm p} = 3.0 \times 10^5 \,\mathrm{M}_{\odot} \,h^{-1}$ . The softening length within the highest-resolution region is 170 pc  $h^{-1}$  co-moving. The lower limit in  $V_{\rm max}$  for convergence is approximately  $10 \,\mathrm{km \, s^{-1}}$ . For more details on the simulation suite, refer to Ref. [237].

We use the ROCKSTAR halo finder to identify haloes and subhaloes within each simulation. Halo masses and radii are defined as virial values using the virial over-density threshold  $\Delta_{\rm vir}$ , which has a value of  $\approx 340$  given the cosmological parameters from the previous paragraph. Subhaloes are haloes whose centers lie within the virial radius  $(R_{\rm vir})$  of a halo that has a larger maximum circular velocity,  $V_{\rm max} = \max[GM(< R)/R]^{1/2}$ . We refer to any halo that is not a subhalo as a host halo. In each simulation, we select every subhalo that lies within 200 kpc of the host halo's center and has  $V_{\rm max} > 10 \,\mathrm{km \, s^{-1}}$ . Generally in this chapter, we use  $V_{\rm max}$  as a measure of the potential well depth in a subhalo. In the following, we will often quote subhalo circular velocities in units of the maximum circular velocity of the host halo, namely  $V_{\rm max}^{\rm sat}/V_{\rm max}^{\rm host}$ , because subhalo demographics are approximately self-similar when scaled in this way. Our resolution limit corresponds to a limit on this ratio of  $V_{\rm max}^{\rm sat}/V_{\rm max}^{\rm host} > 0.065$ .

Halo properties (e.g., concentration, spin, and so on; see below) are all computed as described in the ROCKSTAR documentation Ref. [18].

## **3.2.2** Concentration

On average, CDM haloes can be described by a universal density profile which is approximated well by the Navarro, Frenk, & White (NFW) profile of Ref. [273] (defined in Equation 1). The scale radius of a halo is most often expressed through the concentration parameter  $c_{\rm NFW}$ , which is the ratio of the halo virial radius to the halo scale radius,

$$c_{\rm NFW} = \frac{r_{\rm vir}}{r_{\rm s}}.$$
(9)

The concentration parameter characterizes the degree to which the mass of the halo is concentrated toward the halo center. Concentration is known to be a slowly-declining function of halo mass [274, 55]; it has previously been found to correlate with subhalo abundance [237].

We can compare the concentration of simulated haloes in the mass range of the Milky Way to the approximate constraints of the Milky Way halo concentration, along with the haloes of galaxies of the same Hubble type (SBb/c in the case of the Milky Way; [55, 274]). ACDM models of  $10^{12} M_{\odot}$  haloes have placed  $c_{\rm NFW}$  in the range of 11–21. However, the constraint on "normal" Sb galaxies according to Ref. [200] is expected to be ~ 10–17, based on statistical ensembles of haloes. It is challenging to pin down an exact estimate of concentration for Milky Way-sized haloes in dark matter-only simulation, as baryons are expected to cause haloes to adiabatically contract and drive up their concentrations [113, 34, 256, 97], while feedback and/or mergers can reduce halo concentrations. Concentration measurements of the range 18–24 are expected for a Milky Way dark matter halo, where the dark matter halo has been constrained with observations of dynamical tracers in the Milky Way halo [15, 65, 87, 277, 184, 395]. These observations are expected to reflect some degree of contraction of the halo, so we use a publicly available code, CONTRA [149], in order to estimate a non-contracted concentration for the Milky Way. This process is discussed in detail in Section 3.5. We determine the concentration of the Milky Way to be  $c_{\rm NFW}^{\rm MW} = 15.13^{+1.35}_{-2.58}$ , which we will use in our modeling.

### 3.2.3 Spin

It has long been thought that proto-dark matter haloes acquire angular momentum due to tidal torques from nearby over-densities [287, 103, 112, 378]. The resulting angular momentum is often parameterized using a dimensionless quantity called the spin parameter. The two most common definitions of the halo spin parameter are

$$\lambda_{\rm P} = \frac{J_{\rm vir} |E|^{1/2}}{G M_{\rm vir}^{5/2}} \tag{10}$$

and

$$\lambda_{\rm B} = \frac{J_{\rm vir}}{\sqrt{2}M_{\rm vir}r_{\rm vir}V_{\rm vir}},\tag{11}$$

where  $M_{\rm vir}$ ,  $J_{\rm vir}$ , E,  $V_{\rm vir}$ , and  $R_{\rm vir}$  are respectively halo virial mass, total angular momentum within the virial radius, total energy of the halo relative to a zero point at infinity, halo virial velocity, and the halo virial radius. The first of these,  $\lambda_{\rm P}$ , is generally referred to as the Peebles spin parameter and quantifies the angular momentum of the halo in units of the angular momentum necessary to support the halo assuming that all particles are on circular orbits [288]. The Bullock spin parameter,  $\lambda_{\rm B}$ , is a convenient definition for cases in which halo energies are not readily available (e.g., in most numerical simulations; [54]). In this work we use only the Bullock spin parameter.

Dark matter haloes are mostly supported by the random motions of their particles instead of rotation, so typical values of the spin parameter are quite small, with the median spin value being ~ 0.05 and ranging from 0.02 - 0.11 [14]. Since spin characterizes the angular momentum of the halo, there is expected to be a correlation between halo spin parameter and galaxy morphological type (with rotationally-supported galaxies found in haloes of greater spin). For example, Ref. [365] and Ref. [200] have estimated that Sb galaxies should reside in haloes with spins in the range  $\lambda_{\rm P}^{\rm Sb} \sim 0.02$  to 0.10 (90% confidence region). Simple galaxy formation and evolution models, such as the classic model of Ref. [256], suggest that for a fixed host halo circular velocity, more compact disks form within haloes of lower spin. Therefore, given that the Milky Way has a more concentrated stellar disk than is typical for a galaxy of its mass, we would also expect it to have a lower spin [255]. As an example, the Ref. [256] model for disk formation in hierarchical cosmologies in particular predicts that

$$\lambda_{\rm P} = \frac{2.0 R_{\rm d} V_{\rm c} (|E|)^{1/2}}{G M_{\rm vir}^{3/2}} \left(\frac{m_{\rm d}}{j_{\rm d}}\right),\tag{12}$$

with the relation from Ref. [55]

$$\lambda_{\rm B} \simeq \lambda_{\rm P} f(c_v)^{1/2},\tag{13}$$

where  $j_{\rm d} = \frac{J_{\rm d}}{J}$  is the angular momentum fraction or the disk angular momentum divided by the total angular momentum,  $m_{\rm d} = \frac{M_{\rm d}}{M}$  is the disk mass fraction or the disk mass divided by the total mass,  $R_{\rm d}$  is the disk scale-length,  $V_{\rm c}$  is the circular velocity, and  $f(c_v) \simeq [2/3 + (c_{\rm NFW}/21.5)^{0.7}]$  [256, 236]. For typical concentrations of  $c_{\rm NFW} \simeq 10$ ,  $f(c_v)$  is of order unity. This is what we will adopt for our work.

For simplicity, we use the Ref. [256] assumption that  $\frac{m_d}{j_d} = 1$  for an NFW halo. This is based upon the assumption that the specific angular momentum of what forms the disk and the halo are the same; however, recent simulations suggest that this idealization may be suspect [341, 177]. Recent estimates of the Milky Way parameters include disk scale length ( $R_d = 2.71^{+0.22}_{-0.20}$  kpc from Ref. [218], consistent with other measurements [250, 28]), circular velocity ( $V_c = 219 \pm 20$  km s<sup>-1</sup> from Ref. [304] and Ref. [109]), and virial mass  $(M_{\rm vir} = 1.3 \pm 0.3 \times 10^{12} {\rm M}_{\odot})$ . Putting these together, we estimate  $\lambda_{\rm B}^{\rm MW} = 0.0332 \pm 0.0105$  for the Milky Way. For comparison, Ref. [184] uses contrasting values for the Milky Way based on kinematics from giant stars,  $R_{\rm d} = 4.9 \pm 0.4$  kpc and  $M_{\rm vir} = 0.8^{+0.31}_{-0.16} \times 10^{12} {\rm M}_{\odot}$ , which yields  $\lambda_{\rm B} = 0.0975 \pm 0.0432$  for the Milky Way (or  $\lambda_{\rm B} = 0.0596 \pm 0.0193$  if we use the same  $M_{\rm vir}$  as in our calculation). In all three cases errors are calculated by propagation of errors.

Our estimate of  $\lambda_{\rm B}^{\rm MW} = 0.0332 \pm 0.0105$  is consistent with the results of Ref. [109], which use a slightly different approach and slightly different (older) estimates of Milky Way properties. This  $\lambda_{\rm B}^{\rm MW}$  will be the Milky Way mean and  $\sigma$  used in our analysis. We emphasize that we are not looking to calculate a precise  $\lambda_{\rm B}$  for the Milky Way - we are more interested in the rank of the Milky Way's spin, and whether it correlates with disk size, as it would in a Ref. [256] model. Although recent studies have challenged this model, we will take the classic approach. Additionally, as we will show in Section 3.3, the spin parameter has a sub-dominant effect on subhalo abundance.

## **3.2.4** Shape

Dark matter haloes in  $\Lambda$ CDM are not spherical but, rather, more nearly triaxial ellipsoids. The shapes of CDM haloes are commonly described by the ratios of their principal axis ratios, b/a (the intermediate-to-long axis ratio) and c/a (the short-to-long axis ratio). Generally, CDM haloes are close to prolate [7], with  $c \sim b < a$ . Therefore, for simplicity we quantify halo shape using c/a. Halo shape is highly dependent on the merger history of the halo. The more recent a merger, the less spherical a halo will be and the longest axis of the halo typically correlates with the impact direction of the most recent merger event. Generally, haloes at fixed mass that have formed earlier tend to be more spherical ( $c/a \sim 1.0$ ; [321, 377]), and more massive haloes tend to be less spherical (c/a < 1.0; [228]).

Constraints on the shape of the dark matter halo which hosts the Milky Way are relatively weak. Through various gas and stellar stream measurements, density profile estimates, and simulations, the Milky Way is approximated to be quasi-spherical, with  $c/a^{\text{MW}} \simeq 0.72-0.8$ according to Ref. [212] and Ref. [362] when incorporating measurements from the Sagittarius stream. Therefore the estimated value for the Milky Way we use in our analysis is c/a = 0.76,  $\sigma_{c/a} = .02.$ 

#### 3.2.5 Halo Merger History

Halo merger history is also correlated with subhalo abundance. Earlier forming haloes have been shown to have less substructure [394, 179, 238]. Host haloes that assembled earlier are expected to end up with less mass in subhaloes because there has been more time for accretion by the host.

There is good evidence from chemo-dynamical studies [306] and other work that the Milky Way has had a quieter accretion history than typical. It appears that the Galaxy has not had any substantial mergers since the formation of its galactic disk ( $\approx$  9–12 Gyr ago; [352]).

For simplicity, we characterize merger history in terms of the scale factor of the Universe at the time of a halo's last major merger,  $a_{\text{LMM}} = \frac{1}{1+z_{\text{LMM}}}$ . For the zoom-in simulations used in this chapter, we define a major merger as one with a mass ratio > 0.3 (which implies that a merger like the current Milky Way–Sagittarius merger would be classified as a minor merger, as expected by current mass estimates). Using the limits of 9–12 Gyr for the Milky Way's last major merger we convert to z = 1.33-3.55 and  $a_{\text{LMM}}^{\text{MW}} = 0.43-0.22$  respectively of which we select the median  $a_{\text{LMM}}^{\text{MW}} = 0.325$ ), using WMAP9 cosmological parameters (which match the simulation parameters closely; cf. Section 3.2.1). The choice of cosmology has a minuscule effect on this estimate in comparison to the uncertainty in the time since the last major merger.

### 3.2.6 The Milky Way Halo Compared to Other Dark Matter haloes

Figure 3.1 depicts the joint distributions of the spin ( $\lambda_{\rm B}$ ), concentration ( $c_{\rm NFW}$ ), shape (c/a), and last major merger scale ( $a_{\rm LMM}$ ) parameters of the 45 Milky Way-size halo simulations that we study. Each host halo from a zoom-in simulation is depicted as a circular or triangular point in purple or blue respectively. The blue triangular haloes correspond to the five nearest neighbors to the Milky Way in the multi-dimensional space consisting of all the parameters plotted, as we will discuss in Section 3.6.2. Our simulated Milky Way-like haloes

exhibit the same correlations between these properties found in prior work. For example concentration and spin are known to be anti-correlated [228], and shape and merger history are expected to be correlated [7].

The region of Fig. 3.1 in which Sb galaxies' haloes are thought to reside is shown by the orange dashed region. We highlight this regime to enable comparisons to the estimated parameters of the Milky Way halo, since the Milky Way is generally classified as an SBb/c galaxy. In the case of  $a_{\rm LMM}$ , a value is quoted only for our Galaxy, denoted by the black dashed lines, as we are not aware of prior work on major merger scale parameters for Sb galaxy dark matter haloes.

The black points with the error bars represent the estimated parameters of the Milky Way host halo from the literature; their provenance is described individually above. As is evident, the Milky Way lies closer to the outskirts of the multidimensional distribution in each projection. In particular, the Milky Way halo appears to have a somewhat low spin, high concentration, more spherical shape, and a longer lookback time to the last major merger than a typical halo of the same mass. The concentration, shape, and lookback time of the Milky Way is expected to be consistent with the Galaxy's more compact stellar disk than average at fixed luminosity.

We have demonstrated that the Milky Way may well be an outlier in the distributions of several halo properties. One of these properties, concentration, was previously found to correlate with subhalo abundance [394, 237, 179]. Indeed, it is not unreasonable to suspect that other properties correlate with subhalo abundance as well, particularly because halo spin and shape are so strongly associated with halo merger activity. In the following section, we will investigate the correlations between host halo properties and subhalo abundance and use these correlations to make predictions for subhalo demographics within haloes resembling that of the Milky Way.



Figure 3.1: Plots of all possible combinations of the host halo parameters investigated in this work. The orange dashed region indicates the range of estimated values for Sb galaxy host haloes; these have morphology similar to the Milky Way (which is estimated to be SBb/c). The black dashed regions shown in  $a_{\rm LMM}$  are values quoted only for the Milky Way due to no known work on major merger scales of Sb galaxy dark matter haloes. The black point with errorbars represent the estimated properties for the Milky Way host halo described in Section 3.2. We emphasize that these are approximate estimates and in many cases are difficult to constrain well. The blue triangle points are the 5 nearest neighbors to the Milky Way within this four-dimensional parameter space; their identification is discussed in Section 3.6.2. The Spearman correlation coefficient and *p*-values for all of the correlations shown here are presented in Table 3.1. The Milky Way lies at the outskirts of each of these projections. Known relationships between host halo properties have been reproduced by our simulations.

### 3.3 Subhalo Abundances in Milky Way-Like haloes

In this section we investigate the relationships between the host halo properties described above and the abundance of subhaloes in each simulated Milky Way-like halo. We will then incorporate the correlations observed into a prediction for subhalo abundances within haloes resembling that in which the Milky Way resides.

## 3.3.1 Halo Properties and Subhalo Abundances

We begin with a simple statistical search for correlations. Table 3.1 shows the Spearman correlation coefficient between the host halo properties. The Spearman (or ranked) correlation coefficient ( $\rho$ ) measures the strength and direction of a monotonic relationship between two ranked variables [266]. The coefficient can range from +1 to -1, where the extremes indicate that each of the variables is a strictly monotone function of the other, so that that ranks within lists of the two variables are perfectly associated. The p values is a way of investigating whether we can accept or reject the null hypothesis that there is no monotonic association between the two variables. We set our threshold at p < 0.05, or a less than 5% chance that the relationship found (or any stronger relationship) would happen if the null hypothesis were true. In Table 3.1 the numbers above the diagonal denote  $\rho$  and the numbers below the diagonal denote p. The table is color-coded according to the correlation coefficient: values of  $\rho$  near 1.0 are shown as red while  $\rho$  near -1.0 corresponds to blue. The Spearman correlation is sensitive to both linear and non-linear relationships, and (unlike the Pearson correlation coefficient) is robust to outliers. We use the Spearman correlation because we are interested in testing for general monotonic relationships between dark matter halo properties.

Focusing on the first column and row, we conclude that  $c_{\text{NFW}}$ ,  $a_{\text{LMM}}$ , and  $\lambda_{\text{B}}$  are all significantly correlated with subhalo abundance. The shape parameter c/a still has a relationship with subhalo abundance, but not as strong as for the other host properties.

Although no significant correlation between the abundance of subhaloes and host halo mass is found here, this is almost certainly due to the small mass range of the zoom-in haloes, such that the variations in other parameters dominate. In the parent simulations from which the re-simulated haloes were drawn, the number of subhaloes is on average directly proportional to host halo mass.

We next investigate how each host property individually influences subhalo abundance. In Fig. 3.2 we present the mean cumulative velocity function (CVF) of the subhaloes when host haloes are divided into subsets according to their properties. The CVFs (top panels) and ratio of each CVF to the average halo CVF (bottom panels) are separated into quartiles according to each host halo property considered here:  $c_{\rm NFW}$ ,  $\lambda_{\rm B}$ , c/a, or  $a_{\rm LMM}$ . The vertical axis of the CVF (upper) plots is the average cumulative number of subhaloes above a threshold in velocity, while the horizontal axis is the corresponding maximum subhalo velocity normalized by the maximum velocity of the respective host. The vertical axis in the lower panel corresponds to the cumulative number of subhaloes for a particular quartile divided by the average cumulative number of subhaloes all hosts.

The quartile curves shown in each panel are determined by percentiles in each respective host halo property. The > 75<sup>th</sup> percentile bin for each panel includes 12 host haloes, while the other percentiles have 11 host haloes each. For each halo we determine how many subhaloes are in each of 20 logarithmically spaced  $V_{\text{max}}^{\text{sat}}/V_{\text{max}}^{\text{host}} = V_{\text{max}}^{\text{frac}}$  bins; from this we can determine the cumulative number of subhaloes for each halo summing down to a given bin of  $V_{\text{max}}^{\text{frac}}$ . The mean CVF for the haloes in each quartile are shown as the red to orange lines. Portions of the plots which lie below the resolution limit described in Section 3.2.1 (i.e., with  $V_{\text{max}}^{\text{frac}} < 0.065$ ) are indicated by the hatched region.

The black points represent the 11 classical satellites of the Milky Way, using values from Ref. [388, 360, 185, 209, 44, 301, 249] and Ref. [178] compiled in Table 1 of Ref. [178]. The only kinematic information available for the Milky Way dwarf spheroidals is the line-ofsight velocities of stars, which can be used to constrain the dynamical mass of the dwarf. In the case of Sculptor, Draco, Leo II, Fornax, Sextans, Carina, Leo I, and Ursa Minor, Ref. [209] and Ref. [44] use the Via Lactea II simulation or the Aquarius suite of simulations, respectively, to assign weights to subhaloes in the simulations according to how well they match the dynamical mass of each respective Milky Way satellite, and then use the weighted average of  $V_{\text{max}}$  for those subhaloes as an estimate of the satellite's  $V_{\text{max}}$  value. For example,

	$N_{ m sub}$	$M_{ m vir}$	$c_{\rm NFW}$	$a_{\rm LMM}$	$\lambda_{\mathrm{B}}$	cla	_	<b>-</b> 1
N <sub>sub</sub>		$\rho = 9.47 \times 10^{-2}$	$\rho = -8.57 \times 10^{-1}$	$\rho = 5.64 \times 10^{-1}$	$\rho = 3.98 \times 10^{-1}$	$\rho = -2.29 \times 10^{-1}$		
$M_{ m vir}$	$p = 5.36 \times 10^{-1}$		$\rho=-9.78\times10^{-2}$	$\rho = -3.32 \times 10^{-2}$	$\rho = -8.66 \times 10^{-2}$	$\rho = -6.61 \times 10^{-2}$		
C <sub>NFW</sub>	$p = 5.80 \times 10^{-14}$	$p = 5.23 \times 10^{-1}$		$\rho = -6.68 \times 10^{-1}$	$\rho = -3.87 \times 10^{-1}$	$\rho = 4.75 \times 10^{-1}$		
$a_{\rm LMM}$	$p = 5.44 \times 10^{-5}$	$p = 8.29 \times 10^{-1}$	$p = 5.30 \times 10^{-7}$		$\rho = 5.38 \times 10^{-1}$	$\rho = -5.81 \times 10^{-1}$		U
$\lambda_{ m B}$	$p = 6.83 \times 10^{-3}$	$p = 5.71 \times 10^{-1}$	$p = 8.57 \times 10^{-3}$	$p = 1.39 \times 10^{-4}$		$\rho = -2.84 \times 10^{-1}$		
c/a	$p = 1.31 \times 10^{-1}$	$p = 6.66 \times 10^{-1}$	$p = 9.82 \times 10^{-4}$	$p = 2.85 \times 10^{-5}$	$p = 5.90 \times 10^{-2}$			1
								1

Table 3.1: Table of Spearman correlation coefficients,  $\rho$ , and the corresponding *p*-values, *p*, amongst all of the host halo properties considered in our analysis: the number of subhaloes  $(N_{\rm sub})$ , the mass within the virial radius  $(M_{\rm vir})$ , concentration  $(c_{\rm NFW})$ , the scale factor at the time of the last major merger  $(\alpha_{\rm LMM} = \frac{1}{1+z_{\rm LMM}})$ , the Bullock spin parameter  $(\lambda_{\rm B})$ , and the halo shape (parameterized by the ratio of tertiary to major axis length, c/a). Above the diagonal and denoted by  $\rho$  are the correlation coefficients. Below the diagonal and denoted by  $\rho$  are the correlation. A value p < 0.05 corresponds to a correlation that is statistically significant at  $> 2\sigma$ , or a < 5% chance this relationship would be found if the null hypothesis of of no monotonic association between variables is true. The values are colored by the strength of their correlation: 1.0 is red and -1.0 is blue, as described in the color bar. Subhalo number is most strongly correlated with  $c_{\rm NFW}$ ,  $\alpha_{\rm LMM}$ , and  $\lambda_{\rm B}$  in our simulations, followed by c/a and very weakly by  $M_{\rm vir}$  (which is expected given our small mass range).

Ref. [44] computes a distribution function of possible  $V_{\text{max}}$  by assigning a weight from the estimated likelihood that each subhalo from their six randomly-selected Milky Way-mass host haloes is consistent with the given satellite's mass. For the case of the Large Magellanic Cloud (LMC) Ref. [360] uses proper motions and line of sight velocity measurements of stars in the LMC in concordance with a model of a flat rotating disk to estimate the circular velocity. A similar treatment is done for the Small Magellanic Cloud (SMC). In the case of Sagittarius, Ref. [178] uses the relation  $V_{\text{max}} = 2.2\sigma_{\text{line of sight}}$  [301] with the line of sight velocity dispersion measurement ( $\sigma_{\text{line of sight}}$ ) from Ref. [249] in order to estimate its  $V_{\text{max}}$ . The dwarf spheroidal estimates from Ref. [209, 44] and Ref. [360] are all consistent (within errors) with this relation.

A key assumption made is that the simulated haloes of a given mass will match the kinematics of Milky Way satellites' haloes of the same estimated mass. In particular, the stellar content of these satellites is only in the very central region (~ 1 kpc) so extrapolation beyond this stellar distribution is necessary to constrain  $V_{\text{max}}$ . Although using  $V_{\text{max}}$  is less subject to extrapolating issues than the total mass,  $V_{\text{max}}$  will still depend substantially on the assumed distribution of dark matter for a galaxy of the observed size [393]. The shape of the dark matter density profile may vary substantially from an NFW profile. In the case for the work by Ref. [44], the density profiles are not assumed NFW or Einasto with all properties computed from the raw particle data in order to get around this issue (but a general profile is still indeed assumed).

To normalize these values to  $V_{\text{max}}^{\text{frac}}$  we use the average  $V_{\text{max}}$  of the Milky Way after doing 10,000 bootstraps perturbing the measured value from Ref. [388] and Ref. [178] of  $V_{\text{max}}^{\text{MW}} = 170 \pm 15 \text{ km s}^{-1}$  by a Gaussian of the error. This value was determined by using line of sight kinematic data and connecting it with simulation data by finding the best matched probability distributions. Our resulting  $V_{\text{max}}^{\text{host}} = 170.22 \text{ km s}^{-1}$  for the Milky Way. This is consistent with our Milky Way-mass host haloes, that have an average  $V_{\text{max}}^{\text{host}} = 174.06 \text{ km s}^{-1}$ .

The shaded gray regions around the Milky Way satellite points indicate the 68% and 95% confidence regions from the effect of measurement errors on each satellite  $V_{\text{max}}$ . For each satellite, we generate 10,000 Gaussian-distributed values randomly drawn from the

errors in each  $V_{\text{max}}^{\text{sat}}$  and then perturb the estimated value for that satellite by the generated value. We emphasize that the goal of including the Milky Way satellite points is strictly for reference and not direct comparison, as our dark matter-only simulations do not include the baryonic physics, feedback mechanisms, etc. (see Section 3.1) that would be necessary to make the  $V_{\text{max}}^{\text{frac}}$  values from the simulations directly comparable to the Milky Way satellite characteristics.

The bottom panel of each plot, which depicts the ratio of each quartile's mean CVF to the overall average CVF, shows the differences amongst the quartiles for a given property more clearly than the raw cumulative velocity functions plotted in the top panel. A dotted horizontal line at  $N_{\text{sat}}(>V_{\text{max}}^{\text{sat}})/\langle N \rangle = 1$  indicates where there would be no difference between a given quartile and the mean. The orange regions around  $N_{\text{sat}}(>V_{\text{max}}^{\text{sat}})/\langle N \rangle = 1$  corresponds to the 68% and 95% confidence region about this value from Poisson errors for a quartile of 11 haloes.

The differences between the CVFs of quartiles divided according to a given property allow us to investigate the relationship between that property and subhalo abundance. In Fig. 3.2 it is clear that at low velocities, the separation between the extreme quartiles for every property shown is larger than the  $2\sigma$  Poisson error. We see the most significant separation when we divide samples according to  $c_{\rm NFW}$ ; this property also has the strongest correlation to  $N^{\rm sub}$  (cf. Table 3.1), consistent with the results from Ref. [394] and the model developed by Ref. [237]. The next strongest effect is associated with  $a_{\rm LMM}$ , as expected from predictions from e.g., Ref. [394] and Ref. [179]. Interestingly c/a shows a more significant separation than  $\lambda_{\rm B}$  in Fig. 3.2, in contrast to Table 3.1. The higher concentration, lower spin, more spherical, or earlier forming haloes – that is, those which are most similar to the estimated properties of the Milky Way dark matter halo in each characteristic – are all associated with having fewer subhaloes.

We can conclude that at low velocities this set of four host halo properties can help to predict subhalo abundance, given their correlations with that quantity. We expect this to be the case at higher velocities as well, but there is too much noise due to low counts per bin to draw a statistically significant conclusion from Fig. 3.2 at high velocities. Physically we expect there to be far more subhaloes with low  $V_{\text{max}}$  than high, given the mass function of subhaloes [260, 58, 337, 205], making it the most important region to probe.

Having shown that we can identify host halo properties that correlate with subhalo abundance, we next investigate what *combinations* of these parameters provides the best predictions of subhalo abundances for Milky Way-like dark matter haloes.

# 3.3.2 Predicting Milky Way Subhalo Abundances

To address this question, we have built power-law scaling relation models which utilize various combinations of halo properties as predictors for the cumulative number of subhaloes above a given value of  $V_{\text{max}}$ , in order to produce more accurate predictions of the subhalo abundance for the Milky Way. We describe these models in detail in Section 3.6.3 but summarize them here.

The first model considered is a relatively simple one, incorporating only  $c_{\rm NFW}$  to predict subhalo abundance; we refer to it as our "one-parameter model" hereafter (the fit does incorporate a second parameter setting the scale of the overall subhalo numbers at a given velocity, however). This approach can be motivated by Ref. [237]'s conclusion that halo concentration provides sufficient information to predict subhalo abundance in haloes of a given mass. We compare predictions from this simple model to results from a power-law model built using an optimized combination of the examined host halo properties, which has greater statistical explanatory power; we will refer to it as our "three-parameter model", though again it also incorporates a normalization factor. We also compare to a model that does not have the best statistical explanatory power, but does not include concentration.

Specifically, the robust three-parameter model includes concentration, spin, and shape (as well as the assumption, motivated by tests with larger simulations, that subhalo abundance is proportional to mass). This specific set of parameters was chosen because it had lower Akaike and Bayesian Information criteria (AIC and BIC) than other models considered, which included all combinations of the halo parameters used in this chapter; quadratic terms combining those parameters; and first order polynomial cross terms e.g.,  $c_{\rm NFW} \times c/a$ , that had as many as 5 total parameters (apart from a constant term). These low information criterion values indicate that this model provides a better fit for subhalo abundances,



Figure 3.2: Cumulative velocity functions  $N_{\rm sat}(>V_{\rm max}^{\rm sat}/V_{\rm max}^{\rm host} = V_{\rm max}^{\rm frac})$  of subhaloes for samples split according to various host halo properties. The black points indicate the 11 classical Milky Way satellites  $(M_{\rm vir} \simeq 10^{10} {\rm M_{\odot}})$ . The gray region around each point represent 68% and 95% confidence regions as a result of Gaussian perturbations with amplitude given by the error in each satellite's  $V_{\rm max}^{\rm frac}$ . Separate cumulative velocity functions are shown for haloes divided into quartiles by, from left to right: (a) concentration  $(c_{\rm NFW})$ , (b) spin  $(\lambda_{\rm B})$ , (c) shape (c/a), and (d) scale factor at last major merger  $(a_{\rm LMM})$ ; color labeling is indicated in each panel's legend. In the bottom panel of each plot is shown the ratio of each curve to the average cumulative velocity function including all host haloes. These normalized plots include a black dashed line at  $N_{\rm frac}(>V_{\rm max})/\langle N \rangle = 1$  to indicate where there is no separation between the quartiles. The orange regions around  $N_{\rm sat}(>V_{\rm max}^{\rm frac})/\langle N \rangle = 1$  indicate the 68% and 95% confidence regions for a case with no deviation from the mean value due to Poisson errors (given the average number of subhaloes and the total number of host haloes that fall in a quartile where we use 11). Portions of the figures below the resolution limit of the simulations used,  $V_{\rm max}^{\rm sat}/V_{\rm max}^{\rm tot} = 0.065$  (described in Section 3.2.1), are depicted by the hatched regions. At low velocities the cumulative velocity functions exhibit statistically significant separations amongst various host halo properties.

given the number of free parameters in the model, than any others considered. Details of this evaluation are given in Section 3.6.3. This model is expected to provide additional information as opposed to just additional degrees of freedom, given its low AIC and BIC values. Although both models provide useful predictions of subhalo abundances, we would expect the three-parameter model to always provide a more accurate prediction than the one-parameter model, as the one-parameter model is a special case of the three-parameter model (so further optimization via the other parameters can only improve performance; see Table 3.3 and discussion). We also show in Table 3.3 that the  $\chi^2$  value of three-parameter model is 20% better than for the one-parameter model.

The comparison three-parameter model includes spin, shape, and scale factor at the time of the last major merger. This model is chosen for comparison due to the dubious nature of the concentration measurement of the Milky Way, as well as a means for comparing to a model without concentration. This model's information criteria are also shown in Table 3.3, which makes it evident that this model is not the best combination of parameters for predicting subhalo abundance. We use these three models to estimate subhalo abundances for the Milky Way; by comparing their results we can assess the robustness of our predictions.

We obtain the best predictions of subhalo abundances with a model where the average number of subhaloes in a halo of given properties has a power-law dependence on all relevant parameters, as described in Section 3.6.3 and Section 3.6.4. We define such a power-law model as:

$$N_{\rm sub}^{\rm pred}(>V_{\rm max}^{\rm frac}) = k \times \prod_i x_i^{\alpha_i},\tag{14}$$

where  $N_{\text{sub}}^{\text{pred}}$  is the predicted average subhalo abundance based on halo properties, k sets the scale of the abundances, and  $\alpha_i$  is the exponent for the  $i^{\text{th}}$  halo parameter  $x_i$  used in the model (e.g.,  $c_{\text{NFW}}$ ).

Because we are trying to predict the cumulative subhalo abundance for the Milky Way even at relatively high velocity thresholds where most haloes have few subhaloes, the Gaussian assumption which underlies the method of least-squares linear regression is not valid for this problem (following the usual rule of thumb that the Poisson distribution can be safely approximated by a Gaussian only for N > 25). Instead, we rely on a Poisson maximum likelihood method to fit models, as it should provide accurate results even in this regime. Specifically, we determine the parameter values which maximize the likelihood of the observed set of subhaloes in the simulations. Given the properties for the Milky Way discussed in Section 3.2 in combination with the results from the maximum likelihood fits, we make a prediction for  $N^{\text{sub}}(>V_{\text{max}}^{\text{frac}})$  for the Milky Way in 20 separate bins of  $V_{\text{max}}^{\text{frac}}$ , i.e. we fit a separate model for each threshold of  $V_{\text{max}}^{\text{frac}}$ . The equations and algorithms underlying our methods are discussed in detail in Section 3.6.4.

For each bin in  $V_{\text{max}}^{\text{frac}}$  we can predict a cumulative number of subhaloes for the Milky Way down to that velocity threshold by substituting in the estimates of the Galaxy's parameters discussed in Section 3.2 for the  $x_i$  in Equation 14, and using the k and  $\alpha_i$  values resulting from the model fit for that bin. For example, in the case of the one-parameter model we use the estimate of  $c_{\text{NFW}} = 15.13$  for the Milky Way, and the k and  $\alpha_{c_{\text{NFW}}}$  that result from the Poisson maximum likelihood fit for a particular velocity threshold to obtain a prediction for the corresponding element of the Milky Way CVF.

Fig. 3.3, Fig. 3.4, and Fig. 3.5 depict the results of our model fits, evaluated using the properties of the Milky Way host halo determined in Section 3.2.6. The purple lines show the predicted cumulative velocity functions (i.e., the subhalo abundance for each threshold in velocity fraction considered) for the Milky Way from the one- and three-parameter models, respectively. The dashed grey region indicates the resolution limit of the simulations (which begins to have effects below  $V_{\text{max}}^{\text{sat}}/V_{\text{max}}^{\text{host}} = V_{\text{max}}^{\text{frac}} = 0.065$ ). The solid black line shows the average CVF for the 45 host haloes. In the bottom panels the dashed-dot black line at  $N_{\text{sat}}(> V_{\text{max}}^{\text{frac}})/\langle N \rangle = 1$  is shown for comparison to indicate no separation between the predictions and the average CVF depicted by the solid black line (similar to Fig. 3.2).

Over-plotted in blue are the cumulative velocity functions for the same 5 nearest neighbors to the Milky Way that were indicated in Fig. 3.1, Fig. 3.11, and Fig. 3.12 for comparison. The CVF for the Milky Way classical satellites is again depicted by the black points, as discussed in Section 3.3.

Uncertainties on our model predictions have been obtained via bootstrap re-sampling. Specifically, we randomly select a set of size N = 45 of our simulated Milky Way-mass host haloes with replacement (i.e., allowing the same host halo to be selected more than once) and compute the maximum likelihood fit again 10,000 times. For each bootstrap sample, we also draw a new set of Milky Way halo properties from the probability distributions for each defined by their uncertainties (Gaussian for spin, shape, and scale factor at the last major merger, and two half-Gaussians for concentration due to asymmetric errors, see Section 3.6.4). We can identify 68% and 95% confidence intervals from these bootstrap samples as the regions containing the middle 68% and 95% of bootstrap results at a given velocity, respectively. The resulting confidence intervals, which reflect only uncertainties in Milky Way halo properties and the parameters of the subhalo abundance fits, are depicted by the darker and lighter purple regions.

Additionally, with this bootstrapped sample we can investigate the enlarged range of observed values expected from Poisson scatter about the mean subhalo abundance. For every bin in each bootstrap sample, we randomly generate a value from a Poisson distribution with mean given by the predicted number of subhaloes within that bin  $(N_{\text{sub},l}^{\text{pred}})$   $V_{\text{max},l}^{\text{frac}} - N_{\text{sub},l+1}^{\text{pred}}(> V_{\text{max},l}^{\text{frac}})$ , where *l* indicates velocity bin number and  $N_{\text{sub},l}^{\text{pred}}(> V_{\text{max},l}^{\text{frac}})$  is the predicted cumulative number of subhaloes down to the minimum velocity of that bin). To generate a CVF for that bootstrap sample we then add together the randomly-generated values cumulatively, starting with the highest velocity bin. The resulting values incorporate both the bootstrap uncertainties and the Poisson scatter in subhalo abundances, with the effects of covariance between velocity bins resulting from our use of cumulative counts properly accounted for. The 68% and 95% confidence intervals derived from the distribution of bootstrap values with Poisson scatter included are depicted by the darker and lighter orange regions, respectively.

The distributions of predicted subhalo abundances for the lowest-velocity bin above the resolution limit ( $V_{\text{max}}^{\text{frac}} > 0.065$ ) considering only uncertainties in Milky Way parameters and model fits (purple) or including Poisson variation as well (orange) are shown in Fig. 3.8. In this plot, shaded regions depict distributions just above the resolution limit and outlined histograms correspond to a bin at the high  $V_{\text{max}}^{\text{frac}}$  end before bin counts are dominated by zero. The extents of the 68% and 95% confidence intervals corresponding to these histograms are listed in Table 3.2.

In both models that include concentration the predicted CVF for the Milky Way lies



Figure 3.3: Top panel: Cumulative velocity function for the Milky Way from a one-parameter scaling relation model which predicts subhalo abundance as a function of halo concentration. The purple line corresponds to the prediction of the model at each threshold value of  $V_{\rm max}^{\rm frac}$  for the Milky Way, based on the MW halo's estimated properties. The solid black line corresponds to the average CVF for all of the 45 zoom-in host haloes, while the blue lines are the CVFs for the five nearest neighbors to the Milky Way in parameter space, as indicated in Fig. 3.1 and described in Section 3.6.2. The darker and lighter purple regions indicate the 68% and 95% confidence regions about the purple prediction line due to uncertainties in both fit parameters (evaluated via bootstrap re-sampling) and in Milky Way properties (evaluated by re-drawing values from Milky Way parameter uncertainties before evaluating the models). The darker and lighter orange regions indicate 68% and 95% confidence regions which incorporate Poisson scatter in subhalo abundances as well. Bottom panel: Ratios of the CVFs to the average CVF of all simulated Milky Way-like haloes. The one-parameter model predicts that the Milky Way's host halo should have (at 68% confidence) 12–30% fewer subhaloes than average at the low  $V_{\rm max}^{\rm frac}$  end and 19–45% fewer subhaloes than average dark matter halo of the same mass.



Figure 3.4: As Fig. 3.3 but for a three-parameter scaling relation model, which predicts subhalo abundance as a function of concentration, spin, and halo shape. The three-parameter model has a less significant prediction than the one-parameter model due to the buildup of error. This model predicts that the Milky Way's host halo should have 1–24% fewer subhaloes than average at the low  $V_{\rm max}^{\rm frac}$  end (at 68% confidence) and 22–52% fewer subhaloes at the high  $V_{\rm max}^{\rm frac}$  end. Again, additional error from Poisson scatter can dwarf this effect, especially at high  $V_{\rm max}^{\rm frac}$ .



Figure 3.5: As Fig. 3.4 but a three-parameter scaling relation which predicts subhalo abundance as a function of spin, shape, and scale factor at the last major merger (we have essentially exchanged  $c_{\rm NFW}$  for  $a_{\rm LMM}$ ). The predictions for this model are substantially different from the models that include concentration. This model predicts that the Milky Way's host halo should have up to 17% fewer to 2% more ubhaloes than average at the low  $V_{\rm max}^{\rm frac}$  end (at 68% confidence), and 4–35% fewer subhaloes at the high  $V_{\rm max}^{\rm frac}$  end. We emphasize that the blue lines denote the same haloes discussed in Section 3.6.2 and shown in previous plots (Fig. 3.1, Fig. 3.3, Fig. 3.4, etc). Namely these are the haloes most similar to the Milky Way across all four parameter spaces, which includes concentration.



Figure 3.6: A CVF of one parameter models depicted with all possible theoretical values of concentration for haloes at the same mass and morphology as the Milky Way (see Section 3.2.2). The model is computed in exactly the same way as for Fig. 3.3, and then plotted for various concentration values. haloes of concentration greater than  $\sim 12$  are consistent with having fewer subhaloes. The high- $V_{\rm max}^{\rm frac}$  tail is more sensitive to the concentration than the overall normalization.



Figure 3.7: A CVF of one parameter models depicted with estimated values of Milky Way host halo parameters (concentration in green, shape in orange, spin in blue, and scale factor at the last major merger in pink), as described in Section 3.2. It is clear that concentration is the strongest predictor for a far from average subhalo abundance, which is expected given Fig. 3.6, and the tight correlation between concentration and subhalo abundance seen in Table 3.1. The next strongest deviation comes from shape, followed by scale factor at the last major merger, and lastly spin (which is very close to the average).

Uncertainty	68%	95%
One-para. Bootstraps	0.70-0.88	0.64 - 1.05
Three-para. Bootstraps	0.76-0.99	0.69–1.19
Three-para. No $c_{\rm NFW}$ Bootstraps	0.83-1.02	0.71 - 1.13
One-para. Poisson	0.68-0.90	0.59-1.09
Three-para. Poisson	0.74 - 1.01	0.64-1.22
Three-para. No $c_{\rm NFW}$ Poisson	0.29-1.05	0.62–1.17

Table 3.2: Table of 68% and 95% confidence intervals for the total abundance of subhaloes above the resolution limit ( $V_{\text{max}}^{\text{frac}} > 0.065$ ) predicted for the Milky Way (or what we refer to in the text as  $low V_{\text{max}}^{\text{frac}}$ ). All calculations are done using the power-law models defined by Equation 14. We consider separate confidence intervals for the mean subhalo abundance of a Milky Way-like halo, which include only the uncertainties from subhalo abundance fitting and Milky Way halo parameter uncertainties (labeled as 'Bootstrap' confidence intervals here); and confidence intervals which also include the impact of Poisson scatter about that mean abundance (labeled as 'Poisson'). We provide results for the one-parameter model, the three-parameter model without concentration. All confidence intervals are calculated from the total cumulative predicted subhalo abundance above the resolution limit of  $V_{\text{max}}^{\text{frac}} = 0.065$  divided by the mean total measured subhalo abundance ( $N_{\text{sub}}^{\text{pred}}/\langle N_{\text{sub}}^{\text{meas}}\rangle$ ) above this limit. These confidence intervals correspond directly to the shaded regions depicted visually in Fig. 3.8.
below the average for the Milky Way-mass dark matter haloes up to  $1\sigma$ . However, this is not the case for the model that does not include concentration. This means that concentration is a critical component when examining subhalo abundances in relation to host halo properties, which is evident from Fig. 3.6 and Fig. 3.7 (which is also not surprising since, as discussed in Section 3.3.1 and seen in Table 3.1 concentration is by far the parameter most correlated with subhalo abundance). The former depicts one-parameter models based on the total theoretical range of concentration values of a halo of similar mass or morphology to the Milky Way (see Section 3.2.2. These models are identical to the one-parameter model, but a different concentration is used for plotting (i.e., the  $x_i$  in Equation 14). The mean concentration of haloes in our simulations is 11.63. A halo with a concentration larger than that mean is consistent what having a noticeably smaller predicted subhalo abundance. Additionally we find that the tail at the high  $V_{\max}^{\text{frac}}$  end is more sensitive to concentration than the overall normalization (concentrations smaller than 9 predict more massive subhaloes than average). The latter depicts one-parameter models for the estimated values of the Milky Way host halo properties as described in Section 3.2. This shows that concentration is the strongest predictor for a Milky Way-like halo to having fewer subhaloes, followed by shape, scale factor at the last major merger, and spin last.

Based upon the two superior models, we can conclude that the Milky Way host halo could have fewer subhaloes than would be typical of a halo of its mass (or should, if we take the mean to be representative). At the low  $V_{\text{max}}^{\text{frac}}$  end, at 68% confidence not accounting for Poisson scatter we should expect 12–30% fewer subhaloes than average based on the one-parameter model, or up to 22% fewer subhaloes based on the three parameter model, when incorporating uncertainties in the subhalo abundance fits and Milky Way parameter values. These percentage ranges are similar, but larger at high  $V_{\text{max}}$ , when Poisson scatter is accounted for. At the high  $V_{\text{max}}^{\text{frac}}$  end the mean prediction is even smaller compared to the average (19–52% fewer subhaloes), but uncertainties in fit parameters and fractional Poisson scatter are greater, making the presence of satellites as large as the Magellanic Clouds rare but not extraordinary. Additionally, despite the three-parameter model without concentration being much less robust (see Table 3.3), it still predicts as many as 17% fewer subhalo at low  $V_{\text{max}}^{\text{frac}}$  and up to 35% fewer subhaloes at high  $V_{\text{max}}^{\text{frac}}$  at 1 $\sigma$ .



Figure 3.8: Histograms of predicted total cumulative subhalo abundance below the resolution limit normalized by the mean measured subhalo abundance, where errors constituting the purple histograms include uncertainties in fitting power-law models and uncertainties in the Milky Way host halo parameters (based off just the bootstraps), whereas the orange histogram also incorporates uncertainties due to Poisson scatter about the bootstrapped relations. The filled histograms correspond to the low  $V_{\text{max}}^{\text{frac}}$  end just above the resolution limit, and the outlined histograms correspond to the high  $V_{\text{max}}^{\text{frac}}$  end before the subhalo counts are dominated by 0. The incorporation of Poisson scatter leads to a similar spread in predicted subhalo abundance at low  $V_{\text{max}}^{\text{frac}}$  and a much larger spread at high  $V_{\text{max}}^{\text{frac}}$ . The 68% and 95% confidence regions derived from these distributions are provided in Table 3.2. The one-parameter model predicts slightly fewer subhaloes for the Milky Way than the three-parameter model, which both predict fewer subhaloes for the Milky Way than the model that does not include concentration.

The differences between the predicted cumulative velocity functions of the two concentration based models are relatively small. This also demonstrates that a simple concentrationbased model is adequate for describing subhalo abundances in Milky Way-like haloes to first order, consistent with the results from Ref. [237] and Fig. 3.7. This is no surprise, given  $c_{\rm NFW}$ 's close relationship to the other parameters. For example, when we repeat the same analysis with a one-parameter model using the next-most correlated property in Table 3.1,  $a_{\rm LMM}$ , only 0-20% fewer subhaloes than average are predicted at low  $V_{\rm max}^{\rm frac}$ , while the results are unstable at high  $V_{\rm max}^{\rm frac}$ .  $c_{\rm NFW}$  is significantly more tightly correlated with  $N_{\rm sub}$  than the other host halo properties considered.

As discussed in Section 3.6.3, the three-parameter model predicts subhalo abundances for the simulated haloes with smaller scatter than the one-parameter model. As a result, at first glance one would expect this model to also yield more compact confidence intervals for the abundance of subhaloes around the Milky Way, but in Table 3.2 the opposite holds true (e.g., the 95% confidence interval including all sources of scatter spans 41% in the one-parameter model versus 50% in the three-parameter model with concentration). This can be explained at least in part by the uncertainties in Milky Way halo properties beyond concentration, the effects of which will alter the confidence intervals for the three-parameter model, but not for the one-parameter case. Specifically, the three-parameter will be a better model in the future when there are better constraints on the properties of the Milky Way. Otherwise the errors on the parameters propagate into the errors of the subhalo abundance estimate.

Fig. 3.3, Fig. 3.4, and Fig. 3.5 make clear that at  $V_{\text{max}}^{\text{frac}}$  above ~0.2 the typical halo whose properties match the Milky Way's will have even fewer subhaloes compared to the average a deficiency in excess of 60%. This makes it somewhat more surprising that the MW should have any relatively large satellites such as the Large and Small Magellanic Clouds (LMC and SMC) or Sagittarius (the SMC and Sagittarius correspond to the two highest- $V_{\text{max}}^{\text{sat}}$  points in each figure; the LMC is off the plot range with  $V_{\text{max}} = 91.7 \text{ km s}^{-1}$ ), as more massive satellites are rarer in the haloes most like the one which hosts our Galaxy.

With these results we can define fits for the parameter exponents,  $\alpha_i$ , and the scale k as functions of  $V_{\text{max}}^{\text{frac}}$  for both the one-parameter and three-parameter models which include concentration, as those are the superior models. These functions can be used in order to

determine a CVF for any set of host haloes. The functions and process is described in Section 3.6.5.

#### 3.3.3 Halo Properties and Subhalo Scaling Relations

As briefly mentioned previously, the issue of 'too-big-to-fail' (TBTF) refers to the overabundance of specifically massive and dense subhaloes predicted from CDM simulations in comparison to the number of luminous satellites that the Milky Way has been observed to host [43, 44, 343]. The TBTF has been formulated in many different ways; the most well known is through N-body simulations that show Milky Way-mass hosts should host 6–10 subhaloes with a potential well depth of  $V_{\rm max} > 25 \text{ km s}^{-1}$  [229, 43, 178]. This is not consistent with observations around the Milky Way; only the LMC and SMC fit into this criteria. Another formulation of TBTF is in terms of density. The central densities of Milky Way satellites that have been inferred from kinematics are too low compared to the central densities of their dark matter subhalo counterparts [43, 299, 178].

In quantitative terms, the density formulation of the TBTF is the statement that the classical dwarfs imply a larger radius at which the maximum velocity is reached  $(R_{\text{max}})$  for a given maximum velocity  $(V_{\text{max}})$  than is typical of CDM subhaloes, or that the subhaloes are denser then their kinematics suggest. We use  $R_{\text{max}}$  and  $V_{\text{max}}$  because NFW profiles consist of two parameters, and these are easily related to observed stellar kinematics. Subhaloes fall on a narrow line in  $R_{\text{max}}$ - $V_{\text{max}}$  space, seen in Ref. [393] and subsequent papers. If the structural relation of  $R_{\text{max}}$  and  $V_{\text{max}}$  varies systematically as a function of host halo property (as is the case in our CVFs), this could hold an interesting implication for TBTF — an implied relationship between subhalo density and host halo property.

We have used our model for satellite abundances in Milky Way-like haloes to investigate whether the TBTF issue can be related to host halo properties. Specifically, we have measured the relationship between the maximum circular velocity of subhaloes and their maximum radius, dividing haloes up into quartiles based on host halo properties as above. However, the results were inconclusive: i.e., any separation in the  $R_{\text{max}}$ - $V_{\text{max}}$  plane for samples divided into dark matter host halo property quartiles is relatively weak. We find no statistically significant separation when dividing host haloes into quartiles according to their halo concentration, spin, shape, or major merger scale. Anything unusual in the properties of the Milky Way halo cannot be used to explain structural differences between the observed satellite properties and the expected characteristics of subhaloes; rather, we find the bestfit scaling relations to be essentially independent of halo properties. Improvements to this analysis would require a much larger sample of Milky Way-mass haloes to be re-simulated at high resolution.

# 3.4 Conclusion

In this chapter, we have utilized Milky Way-mass zoom-in simulations to investigate the sources of the scatter in subhalo abundances at fixed mass (see Fig. 3.2). We particularly focus on predicting subhalo abundance conditioned on properties of a host halo. Recent studies of the Milky Way have revealed that the Milky Way has an unusually small disk [28, 218], which in standard galaxy formation theory would be related to unusual host halo properties; we have sought to determine if the particular properties of our Galaxy's halo would also cause its expected satellite galaxy abundance to be unusual. The aspects of the halo we have investigated are its concentration ( $c_{\rm NFW}$ ), spin ( $\lambda_{\rm B}$ ), shape (c/a), and scale factor at last major merger ( $a_{\rm LMM}$ ), as discussed in Section 3.2.

First, we conclude that based on current estimates of its properties the Milky Way's host dark matter halo indeed lies at an extremum compared to haloes of its mass from N-body simulations (see Fig. 3.1). In particular, the Milky Way lies away from the median in the projections across the full parameter space (higher-than average  $c_{\rm NFW}$ , lower than average  $\lambda_{\rm B}$ , more spherical than average c/a, and a very small  $a_{\rm LMM}$ ). Next, from our N-body simulations we have determined that the host halo properties considered are significantly correlated with subhalo abundances for Milky Way-mass dark matter haloes (Table 3.1); as a result, they can be used to predict the cumulative velocity function of a given halo (Fig. 3.2). haloes with lower-than average concentration host a greater number of subhaloes than haloes with higher-than average concentrations. Similarly, lower-than average spin haloes host fewer subhaloes than higher-than average spin haloes, lower-than average (less spherical) shaped haloes host fewer subhaloes than higher-than average shaped haloes, and earlier forming haloes host fewer subhaloes than later forming haloes. In concordance with estimates for the Milky Way, it should be expected that the Milky Way should host fewer subhaloes.

Using the results from the simulations, we have built two sets of scaling-relation models that predict subhalo abundance above a given threshold velocity based upon the properties of a dark matter halo. In the first model, we predict the subhalo abundance based on a single parameter (at fixed halo mass), namely concentration ( $c_{\rm NFW}$ ). Our second model was a three-parameter model that conditioned subhalo abundance on  $c_{\rm NFW}$ , spin ( $\lambda_{\rm B}$ ), and host halo shape (c/a). We also compare to a three-parameter model that does not include concentration, due to the current limited understanding of halo contraction making the Milky Way concentration a rough estimate. We then evaluate these models with the estimated properties of the Milky Way's host dark matter halo to predict subhalo abundances for our Galaxy.

The conclusion of this analysis is that we should expect a host halo similar to the Milky Way's to possess fewer subhaloes than the average halo of its mass. However, the error on actual measurements of the Milky Way dark matter halo also make the range of predicted subhaloes consistent with no effect. We will focus our discussion on the impact of the likely scenario that a halo like the Milky Way has fewer subhaloes when utilizing host halo parameters beyond mass. The central predicted values of our best models are well below the mean, which implies a significant probability of a subhalo count deficit. This result is summarized in Figures 3.3 and 3.4. Both classes of model yield the same basic result: the Milky Way is predicted to have 1–30% fewer subhaloes at low circular velocities ( $V_{\rm max}^{\rm frac}$ ) and 19-52% fewer at high circular velocities than a typical halo of its mass, at 68% confidence when considering only model fitting and Milky Way parameter uncertainties. The decrement with respect to the average cumulative velocity function of dark matter haloes is itself a function of  $V_{\rm max}$  (i.e., a function of subhalo mass). The effect is much larger than estimated uncertainties in the fitting and in propagated Milky Way parameters. Current observations have detected approximately 26 satellites around the Milky Way with  $V_{\rm max}^{\rm sat} > 10 \text{ km/s}$  (this does not include completeness correction or any other corrections). The mean prediction of the one- and three-parameter models estimate 127–135 total subhaloes around the Milky Way, of which some fraction should host observable satellites. At  $1\sigma$  below the mean our models predict as few as 118–129 subhaloes around the Milky Way. In comparison the mean subhalo number for the Milky Way mass host haloes in our zoom-ins is 169 total subhaloes with  $V_{\text{max}}^{\text{sat}} > 10 \text{ km/s}$ . The similar results from both models indicates that the dominant effect is the relationship between halo concentration,  $c_{\text{NFW}}$ , and satellite abundance,  $N^{\text{sat}}$ , as exemplified by Fig. 3.5,Fig. 3.6, and Fig. 3.7. It appears that a  $c_{\text{NFW}}$ -based model is generally adequate for predicting subhalo abundance in the mass range of the Milky Way's dark matter halo, though more complicated models can yield smaller errors.

Additionally, we have found that variations in host halo properties do not have a statistically significant impact on the structure of dark matter subhaloes themselves (at least as assessed using properties connected to TBTF within our sample; see Section 3.3.3). Only the subhalo numbers (and hence the MSP) have been impacted by taking host halo properties into account. A set of new halo re-simulations at a resolution  $8 \times$  higher than those used in this work are now under way and may enable improved investigation of the TBTF problem.

The results described above suggest that a non-negligible fraction of the 'missing satellites' problems is a result of the unusual formation history of the Milky Way. The halo of the Milky Way formed early with very few recent major mergers, which resulted in a more spherically-shaped halo. This also would be expected to lead to a more centrallyconcentrated dark matter halo — consistent with the estimates shown in Fig. 3.1 as well as results from, e.g., Ref. [371, 396] and Ref. [223]. This lack of major mergers should also lead to a relatively small angular momentum of the Milky Way halo. This is consistent with previous results showing that at all masses, haloes with lower spin tend to be in less-dense regions and less strongly clustered [138, 122, 363], such that many reside in environments resembling our Local Group. All of these halo characteristics correlate with having a smaller number of satellites. The results of our models are all consistent with scenarios where the Milky Way's low satellite abundance compared to simple  $\Lambda$ CDM predictions may in part be related to its quiet accretion history as was speculated in Ref. [219, 218].

However, the low subhalo abundance we predict for the Milky Way dark matter halo based on its properties is not on its own sufficient to explain the missing satellites issue. Other factors, such as baryonic physics, must still play a role. In Section 3.1 we listed several of the numerous solutions to the MSP. We will discuss how our results tie in with those solutions below.

1. Baryonic effects: Although our work has shown that we should expect there to be fewer subhaloes than previously anticipated for the Milky Way, the observations and predictions still do not match. At the low velocity end we predict 5× as many subhaloes as have been detected to date. Baryonic physics that causes any satellites in these subhaloes to be difficult to detect could address this problem; the strength of baryonic effects required would be smaller than previously estimated, however. Our results suggest that the Milky Way begins with a state of up to ~ 30% fewer small subhaloes and up to ~ 50% fewer larger subhaloes than average. Baryonic effects do not need to be as efficient as proposed and can use insight from host halo parameters to better tune models.

Ref. [401] and Ref. [47] suggest that a combination of supernova feedback and enhanced tidal disruption caused by the presence of a baryonic disk can resolve the missing satellites problem. However, simulations which recover a Milky Way-like population of satellites in a *typical* galaxy of Milky Way mass must be somewhat miss-tuned. That is, either feedback or tidal effects from the baryonic disk must be weaker than was assumed in the simulations of Ref. [47] in order to get a set of satellites like those that surround our Galaxy when starting out with fewer subhaloes than are typical.

Concentration is an indicator of formation time that may contain more information about the global formation of a halo than simply the time of the last major merger. In host haloes that accrete their substructure earlier, tidal stripping will have longer to operate, depleting subhalo abundances. This implies that the subhaloes within our Galaxy's halo were likely largely accreted relatively early compared to those surrounding other galaxies (barring those associated with the LMC and SMC which may be on their first infall [26, 25]) within the mass range of the Milky Way's halo. This may cause tidal stripping to have stronger effects in the Milky Way system than is typical for a galaxy of its mass.

2. Non-cold dark matter and exotic physics: Similar arguments apply to more exotic physics as to baryonic effects. If, for example, a non-CDM dark matter model and/or some modification to inflation were invoked to alleviate the MSP and TBTF issue, those

modifications would need to be weaker than previously assumed, given the smaller Milky Way subhalo abundance that we predict.

To summarize, when exploring potential factors affecting the missing satellites problem and tuning models to match the Milky Way's satellite population, it is important to ensure that those models predict that a galaxy will have a Milky Way-like satellite population, not for an average halo of Milky Way host mass, but rather for one which has Milky Way-like properties across the board.

We note that uncertainties on the mass of the Milky Way's dark matter halo are sufficient that it may be as small as one half of the value assumed when selecting haloes for resimulation in Ref. [237]. Since the number of subhaloes is proportional to halo mass in the parent simulations from which those haloes were drawn, we would correspondingly expect a 50% smaller subhalo population for a typical galaxy of Milky Way halo mass than what the simulation results give [368]. In combination with the differences between the expected population of Milky Way subhaloes and the population in a more typical galaxy implied by our model fits, this would mean in net that a  $3-4\times$  reduction in the number of subhaloes of the Milky Way halo (and hence the strength of the missing satellites problem) compared to what was previously assumed is entirely possible.

One outstanding anomaly is that, although the Milky Way overall has fewer satellites than might be expected for the typical halo of its mass, it actually has more of the most massive satellites than average. Previous numerical and observational studies have estimated that there is a 2.5–11% chance that a halo of the mass of the Milky Way hosts two subhaloes as large and luminous as the Magellanic Clouds (e.g., Ref. [60]). Taking our Galaxy's halo properties into account only increases this contrast. Fig. 3.3, Fig. 3.4 and Fig. 3.5 show that at  $V_{\text{max}}^{\text{frac}}$  above 0.2 the Milky Way subhalo abundance is expected to be further below the average for a halo of its mass than at lower velocities (with a decrement of up to 50% of the standard prediction, considering only model fitting and Milky Way parameter uncertainties). This makes the 'too-big-to-fail' problem even more of a puzzle. The Milky Way's host halo properties predict in our model that these large companions are very unlikely in concordance with TBTF, yet the Milky Way hosts two very massive companions. However, we note that the large Poisson scatter in this regime tends to dwarf the suppression of subhalo abundances. It would be interesting to follow up this work with an enlarged suite of re-simulated haloes that is more well-suited to addressing the abundances of these large, rare subhaloes.

There are several important caveats to our results which are important to keep in mind when interpreting this work. First, existing constraints on the Milky Way's halo properties are in many cases only rough estimates; this can propagate through into predictions of satellite populations. We have discussed the issue of concentration in detail in Section 3.2.2 and Section 3.5, where adiabatic contraction results in higher concentrated and less NFW-like haloes which is an issue that has not been explored in detail in the literature. Additionally one might worry particularly about estimates of our Galaxy's spin parameter, as there is significant doubt that the connection between disks and haloes is as tight as implied by the Ref. [256] model. Simple models of galaxy formation assume that angular momentum conservation during collapse leads spiral galaxies to have specific angular momenta comparable to the haloes they reside in [121, 256, 55]. However, simulation work by Ref. [177] shows that the halo spin parameter is not significantly correlated with galaxy size. Nonetheless, we remind the reader that it is concentration, and not spin, that drives the bulk of the reduction in subhalo abundance at fixed halo mass; changing the estimated spin parameter by a factor of two has a  $\sim 5\%$  effect on the predicted subhalo abundance at low  $V_{\rm max}^{\rm frac}$  and no noticeable effect at high  $V_{\rm max}^{\rm frac}$ .

In this chapter, we have focused entirely on the problem of differences between the subhalo abundance in Milky Way-like haloes versus the typical halo of the same mass. However, in exploring this problem we have investigated two related issues which we will focus on in upcoming work. First, we have investigated the influence subhaloes have on measurements of dark matter halo properties in simulations. This will provide us with a better understanding of the quantities used as input for semi-analytic models, as well as potential differences between estimates of halo properties for real galaxies versus measurements of those properties in simulations. Second, we have explored the relationship between subhalo abundances and the galaxy-halo connection. An interesting extension of this work which we have not pursued to date would be a comparison of subhalo abundances with the observed population of satellites around M 31, or with satellite populations around Milky Way analog galaxies. Disk scale length (compared to expectations from the mass-size relation of galaxies) could potentially be used to separate high-spin from low-spin haloes of comparable mass, enabling an empirical test of whether satellite abundance correlates with halo properties beyond mass. Through such studies, data from current and upcoming surveys (e.g., the SAGA Survey; [141]) have the potential to provide us with more insight into the Missing Satellites and Too Big to Fail problems.

#### 3.5 Concentration and Adiabatic Contraction

In order to compare the estimated concentration of the Milky Way to our dark matter only data, we must explore how adiabatic contraction of the halo effects the concentration measurement. We do this by employing CONTRA, a publicly available code that calculates the contraction of a dark matter halo as a result of a central population of baryons [149]. This code assumes a spherically symmetric distribution of matter for the dark matter halo (which we know is not entirely accurate), and that the velocity distribution is isotropic (which we know is also unlikely to be the case - the velocity distribution varies throughout the halo). However, it is the most simple model to begin with.

We explore results based upon the updated model for halo contraction from Ref. [149]. The Blumenthal [34] model is the original standard model that has since been updated. The Blumenthal model treats a halo as spherically symmetric, which undergoes homologous contraction - spherical shells that contact in radius but do not cross each other, with particle orbits that are circular, and angular momentum is conserved. The Gnedin model adds in a correction for gas dissipation, which better accounts for effect from mergers and feedback from star formation. It essentially allows for eccentricity in particle orbits that better reflects a complex formation scenario.

The parameters used for input into CONTRA are those determined for the Milky Way from Ref. [28], were the baryonic fraction  $f_b = 0.07 \pm 0.001$ , the baryon scale length  $R_b = 0.00957$  incorporating the scale length of the disk from Ref. [218], and we assume no velocity anisotropy. We calculate an NFW fit to circular velocity where contracted concentration is a free parameter (we also studied a model allowing mass to also be a free parameter which yielded similar results):

$$V(c_{\rm NFW}) = V_{\rm vir} \sqrt{\frac{1}{x} \frac{\ln\left(1 + c_{\rm NFW}x\right) - \frac{c_{\rm NFW}x}{(1 + c_{\rm NFW}x)}}{\ln\left(1 + c_{\rm NFW}\right) - \frac{c_{\rm NFW}}{1 + c_{\rm NFW}}}}$$
(15)

where  $x = \frac{r}{R_{\text{vir}}}$  with r as the contracted positions of the dark matter particles from Contra,  $V_{\text{vir}} = \sqrt{\frac{GM_{\text{vir}}}{R_{\text{vir}}}}, M_{\text{vir}} = 1.3 \times 10^{12} M_{\odot}$  for the Milky Way from Ref. [28], R vir = 282 kpc, and G is Newton's gravitational constant. Data output from CONTRA is log-spaced in 80 bins, so all of our analysis has the same structure, and the errors are linear. In order for us to avoid fitting to a more complex model (which would constitute its own paper), we fit in the outer region of the halo (r > 20 kpc) in rotation curve space in order to mitigate effects from the bulge and the disk.

This fit was calculated using scipy.optimize.curve\_fit from Ref. [364], which is a form of non-linear least squares fitting. The results are presented for two concentration values for the Gnedin model in Fig. 3.9. The blue curves correspond to an initial (pre-contraction) concentration of 10, and the orange curves correspond to an initial concentration of 20. This spans the approximate theoretical range of NFW concentrations for the Milky Way. In both cases the solid lines depict what the rotation curve looks like without any contraction (Equation 15). The over-plotted dashed lines represent the contracted results output from CONTRA, and the dashed-dot lines are the fit to Equation 15. It is evident that the fit is not a good depiction of the CONTRA data. This implies that for a fit to contracted data for a halo that most resembles the Milky Way, NFW is not the best model even in the outer regions of the halo where the fit matches the form of the non-contracted model better than the contracted data.

Fig. 3.10 shows the concentrations as a result of the contraction plotted as a function of their initial values. The differences are small but notable. haloes starting with a higher concentration experience slightly more contraction, which is anticipated. Although we reproduce the expected near 1:1 variance in concentration, the fits themselves are poor.

Despite the inaccuracies in the fits to the adiabatically-contracted profiles, we can take the resultant concentration from the fit in order to calculate NFW concentration as a function of contracted concentration. This is done using the Ref. [364] 1D interpolator. There are a range of estimated values for the Milky Way concentration using various dynamical tracers



Figure 3.9: Rotation curves based on the Gnedin contraction model for two initial NFW concentration values of 10 and 20, depicted by blue and orange respectively. These concentration values fully span the theoretical range for haloes of the same mass as the Milky Way. In both cases the solid lines depict what the rotation curve looks like without any contraction, which perfectly matches to the fit of Equation 15 by design. The over-plotted dashed lines represent the contracted results output from CONTRA, and the dashed-dot lines are the fit to Equation 15. It is evident that the fit is not a good depiction of the CONTRA data.



Figure 3.10: The contracted concentration as a result from our fit as a function of the initial NFW concentration values input into CONTRA, denoted by blue points. For comparison the black line shows the 1:1 ratio which would indicate no change in concentration. The differences are relatively small, where, for example, a halo of NFW concentration of 10 is expected to have a concentration of 13.95 after contraction.

(a majority being blue horizontal branch stars) and various models. We incorporate results from Ref. [15, 65, 87, 277, 184] and Ref. [395] in order to calculate the best estimate for the Milky Way's concentration post-contraction with a median and median interval of  $19.85^{+3.5}_{-5.085}$ . We re-map this estimate and errors into non-contracted space, yielding a best result for the Milky Way of  $c_{\text{NFW}}^{\text{MW}} = 15^{+1.35}_{-2.58}$ .

It is erroneous to say that a contracted halo similar to the Milky Way has a concentration in the same sense that an NFW halo does, for the case of a simple model. One reason for the bad fitting could be the level that the Milky Way itself deviates from an NFW profile an NFW profile is the average for haloes and only good down to  $\sim 10 - 20\%$ . Another is the fact that there is velocity anisotropies in the halo. These are issue that needs to be explored further in the field, but is beyond the context of this chapter.

# 3.6 Numerical and Mathematical Techniques

In the appendices which follow, we discuss details of the numerical and mathematical techniques used in this work. First, in Section 3.6.1 we describe our exploratory linear regression modeling using a variety of combinations of host halo properties as predictors for subhalo abundance. Next, in Section 3.6.2, we discuss the methods used to select the five simulated haloes with properties nearest to those of the Milky Way, which are indicated as blue points or lines in Figs. 3.1, 3.3, 3.4, 3.11, 3.12. In Section 3.6.3, we discuss the selection of the two models used in our final analyses based upon their information criteria. In Section 3.6.4, we discuss in detail the Poisson maximum likelihood methods used to fit power-law scaling models for the subhalo abundance. Lastly in Section 3.6.5 we provide a general fitting function for satellite abundance.

# 3.6.1 Regression Modeling

Our exploration of models for the dependence of subhalo abundance on halo properties starts with simple linear regression modeling. We begin by developing models for the total abundance of subhaloes per host halo, above the resolution limit described in Section 3.2.1 equal to  $V_{\text{max}}^{\text{frac}} = 0.065$ . In that regime, there are > 100 subhaloes per host halo, so Poisson errors in the abundance of subhaloes can be approximated well by a Gaussian distribution. As a result, if we restrict ourselves to purely linear models for simplicity, ordinary least squares (OLS) regression provides an appropriate analysis technique for our data.

Specifically, we regress the total number of subhaloes each host has (our dependent variable) against the matrix of host halo properties considered in this work (constituting the independent variables in this problem). The host halo property distributions are more uniformly distributed in linear space (in the cases of concentration and shape) or log space (for spin, last major merger scale, and mass), so before we take any further steps the log of  $\lambda_{\rm B}$  and  $a_{\rm LMM}$  are taken. Additionally, before the regression is performed (but after log transformations are applied), each set of properties is normalized to have mean zero and variance one, enabling coefficients of different properties to be directly compared to each other. We utilize the Ref. [286] sklearn.linear\_model.LinearRegression class for the OLS regression. We can compare the least-squares predictions to the actual number of subhaloes in each halo to test to what degree a given model explains the overall subhalo abundance.

### 3.6.2 Nearest Neighbors

Throughout this work we highlight the simulated host haloes that are most similar to the Milky Way in their properties. Specifically, the blue points or curves in our figures correspond to the five nearest neighbors to the Milky Way properties in the multidimensional parameter space of  $\lambda_{\rm B}$ ,  $c_{\rm NFW}$ , c/a, and  $a_{\rm LMM}$  (the simulated haloes are already selected to have approximately the same mass).

We identify these neighbors incorporating the results of the regression fits described in Section 3.6.1. Explicitly, we begin by transforming the estimates of Milky Way host halo properties described in Section 3.2 to the same scale as the halo properties used in regression by subtracting off the mean and dividing by the standard deviation of the values of that property amongst the simulated host haloes. We then multiply each host halo property (and the corresponding Milky Way values) by the regression coefficients for that parameter from an OLS linear regression. We can then define a 'distance' by the square root of the sum of the squares of the differences of each of these re-normalized and re-weighted properties. This distance will be smallest for those haloes which most closely resemble the Milky Way *in the properties which most strongly determine subhalo abundances*.

For convenience we determine the nearest neighbors using the scipy.spatial.KDTree() function from Ref. [286]. This function allows neighbors – that is, those haloes with smallest distance from a given point in parameter space, using the metric defined above – to be identified using a multidimensional binary tree, also known as a k-d tree.

### 3.6.3 Model Selection

In Section 3.3.2 we compare predictions for subhalo abundances from two models of differing complexity that include concentration. These were selected out of a set of more than twenty different models incorporating various linear combinations of host halo properties and their products, with up to five parameters per model in total. The parameters for each model were determined using simple linear OLS regression as discussed in Section 3.6.1. We also compare to an inferior model that does not include concentration because of the current difficult nature of measuring concentration for the Milky Way.

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are two statistics commonly used for informing model selection [5, 10]. Adding additional parameters will always tend to increase likelihood values (or decrease chi-squared) when fitting models to a given data set, even if those parameters do not truly have intrinsic explanatory power. The information criteria were introduced in order to penalize goodness-of-fit measures based upon the number of free parameters in a model in order to mitigate this issue. We have calculated both quantities for all models investigated to determine which ones provide the best fits to the data given their level of complexity; the two models we focus on in this work had the lowest AIC and BIC values of any models considered that lack degeneracies when given in power-law form. The information criteria can be defined as

$$AIC \equiv 2k - 2\ln(\hat{L}); \tag{16}$$

$$BIC \equiv \ln(n)k - 2\ln(\hat{L}), \tag{17}$$

where k is the number of free parameters in the model,  $\hat{L}$  is the maximum likelihood value, and n is the number of data points (in our case n = 45, corresponding to the 45 host haloes). For Gaussian errors  $2\ln(\hat{L}) = -\chi^2$  plus a constant, so we use the latter quantity for simplicity in this case (as the number of subhaloes is large enough that the Poisson distribution is very close to Gaussian, and only *differences* in the information criteria are meaningful, so constants do not matter). We calculate  $\chi^2$  as

$$\chi^{2} = \sum_{j} \frac{(N_{\rm sub}^{j,\rm meas} - N_{\rm sub}^{j,\rm pred})^{2}}{\sigma_{j}^{2}},$$
(18)

where  $N_{\rm sub}^{j,\rm meas}$  is the measured total subhalo abundance in the simulation above the resolution limit  $V_{\rm max}^{\rm frac} = 0.065$  for the  $j^{\rm th}$  halo,  $N_{\rm sub}^{j,\rm pred}$  is the predicted total subhalo abundance from a given model for the  $j^{\rm th}$  halo, and  $\sigma_j$  is the uncertainty in the  $j^{\rm th}$  subhalo abundance derived from the data,  $\sigma_j = \sqrt{N_{\rm sub}^{j,\rm meas}}$ ; in this case we calculate  $\sigma_j$  from the data values rather than a model both because N is large (so using data values should give us a good approximation to the predicted uncertainties if we knew the true mean abundance) and to enable apples-to-apples comparisons of  $\chi^2$  / AIC / BIC values between different models.

We also use the root-mean-square deviation ( $\sigma_{\text{RMS}}$ ) as a measure of the difference between predicted values from a given model (in our case  $N_{\text{sub}}^{\text{pred}}$ ) and the measured values ( $N_{\text{sub}}^{\text{meas}}$ ). We calculate this quantity as

$$\sigma_{\rm RMS} = \sqrt{\langle (N_{\rm sub}^{\rm meas} - N_{\rm sub}^{\rm pred})^2 \rangle}.$$
(19)

In other words,  $\sigma_{\text{RMS}}$  is a measure of how accurately a model is able to predict the subhalo abundances measured in our simulations [374]. It is worth keeping in mind that the rootmean-square deviation, like chi-squared, is sensitive to outliers.

Table 3.3 provides the AIC, BIC,  $\chi^2$ , and  $\sigma_{\rm RMS}$  values for linear versions of the models used in Section 3.3.2, in addition to power-law version of the models (which we describe in detail below) and a model which assumes that subhalo abundance is only determined by halo mass (in which case the least-squares prediction is simply the average subhalo abundance, as halo mass is the same for all the re-simulated galaxies). In this section we focus on the linear models, as those were used to select the best combinations of parameters to investigate. In our case a three-parameter model that includes linear dependence on halo spin, concentration, and shape fares better on both information criteria,  $\chi^2$ , and  $\sigma_{\rm RMS}$  than the one-parameter model that incorporates concentration alone. This three-parameter model performed better on both information criteria than almost all of the linear OLS regression models investigated, which included models linear in all the host halo parameters presented in Section 3.2, quadratic terms in those quantities, and cross terms multiplying pairs of halo properties, with up to five total parameters, not including a constant term. The only exceptions were models with cross terms that become degenerate with halo properties when converted into power law form, as in the models used in Section 3.3.2. We expect this model to be a more robust predictor for subhalo abundance. The three-parameter model without concentration is a much less robust model for predicting subhalo abundance. We include solely for comparison due to issue with current Milky Way concentration measurements as described in Section 3.5.

Visual depictions of the one-parameter and three-parameter OLS fits are shown in Fig. 3.11 and Fig. 3.12. The scatter plots in each panel shows the residual value of the number of subhaloes when the prediction from a linear model is subtracted, *excluding that model's dependence on the independent variable shown in that panel*. We refer to this as  $N_{\text{sub}}^{\text{meas}} - N_{\text{sub}}^{\neg x_i}$ where  $N_{\text{sub}}^{\text{meas}}$  is the number of total subhaloes per host halo across the simulations above the effective resolution limit of  $V_{\text{max}}^{\text{frac}} = 0.065$  and  $N_{\text{sub}}^{\neg x_i}$  is the predicted subhalo number, excluding the effect of host halo property  $x_i$ . In Fig. 3.11, this means that what is plotted on the y axis is the number of subhaloes minus the constant term of the fit; in Fig. 3.12 the constant term and the dependence on all parameters but the one plotted are removed.

The over-plotted line in each panel is  $m_i \times x_i$ , where  $m_i$  is the coefficient from the regression fit for the property  $x_i$ ; i.e., each line is the prediction of the regression model for the dependence on that parameter. Our methods are based off those presented in section 7 of Ref. [296]. In each plot the blue triangle points correspond to the nearest neighbors to the Milky Way, as discussed in Section 3.6.2. These scatter plots demonstrate that the results of the regression look generally sensible, but the correlated pattern of residuals indicates that a linear fit is an imperfect representation of the data; as a result, we utilize power law dependencies in our final models.

Model	AIC	BIC	$\chi^2$	$\sigma_{ m RMS}$
Linear				
One-para.	146.26	149.88	142.26	22.34
Three-para.	136.26	143.49	128.26	21.50
Three-para. No $c_{\rm NFW}$	462.11	469.34	454.11	39.92
Power law				
One-para.	142.85	146.46	138.84	22.17
Three-para.	125.62	132.85	117.62	20.59
Three-para. No $c_{\rm NFW}$	450.16	457.39	442.16	39.21
Mass-only model	687.24	698.05	685.24	49.69

Table 3.3: Table of Akaike Information Criterion (AIC) Bayesian Information Criterion (BIC),  $\chi^2$ , and  $\sigma_{\rm RMS}$  values for the models used to estimate subhalo abundance for the Milky Way within this work. Values are first listed for linear models which utilize one halo parameter (concentration) or three (concentration, spin parameter, and shape); values are then listed for power-law models incorporating the same quantities. We also provide statistics for a model which does not incorporate any host halo parameters, which implies that subhalo abundance is based on mass alone (as all of the re-simulated haloes have roughly the same mass). The three-parameter models have lower values for AIC, BIC,  $\chi^2$ , and  $\sigma_{\rm RMS}$  in all cases compared to a one-parameter model like that used in Ref. [237], indicating that such models provide a better fit to the data even when penalizing them for their extra free parameters. We present results based on both one-parameter and three-parameter power law model as superior to the others considered, and all halo-parameter-based models are vastly superior to the results when ignoring halo properties. Differences in AIC/BIC of > 10 are generally considered to provide very strong evidence of a superior fit.



Figure 3.11: Residual subhalo abundance as a function of the parent halo's NFW concentration parameter. In this and the figures below, the residual plotted on the y-axis is the observed ( $N^{\text{sat}}$  for a given halo minus the predicted  $N^{\text{sat}}$ ) for that halo from the regression model, where all terms except for the one dependent on the x-axis value are included in the prediction (in this case, this simply means that the constant term from regression has been subtracted from the total subhalo abundances above the effective resolution limit of  $V_{\text{max}}^{\text{frac}} = 0.065$ ). The over-plotted line is  $m \times c_{\text{NFW}}$  where m is the coefficient from the regression for the property  $c_{\text{NFW}}$ , i.e. the term that was omitted from the regression model in calculating the residual. If the model is a good fit to the data the line should be a good representation of the plotted points. The blue triangle points correspond to the nearest neighbors to the Milky Way in parameter space, identified as discussed in Section 3.6.2. The tightness of the points about the line indicates that a linear function of  $c_{\text{NFW}}$  provides a useful prediction of subhalo number; the systematic pattern of residuals about the line at low  $c_{\text{NFW}}$  indicates that a non-linear dependence should give an even better fit.



Figure 3.12: Residual subhalo abundance as a function of the parent halo's spin, concentration, and shape. The residual plotted on the y axis is the observed  $(N_{\text{sub}}^{\text{meas}} \text{ subhalo abundance for a given halo minus the predicted subhalo abundance for that halo from the regression model, where all terms except for the one dependent on the x axis value are included in the prediction <math>(N_{\text{sub}}^{\neg x_i})$ ). The over-plotted line in each panel corresponds to the best-fit regression coefficient for a given parameter times the value of that parameter (from left to right:  $\lambda_{\text{B}}$ ,  $c_{\text{NFW}}$ , or c/a). This corresponds to the prediction of the regression model for the dependence on each quantity, with dependence on all other parameters subtracted off (as well as the constant term of the fit). The blue triangle points indicate the five nearest neighbors to the Milky Way, identified as discussed in Section 3.6.2. The results show that a linear model based on  $\lambda_{\text{B}}$ ,  $c_{\text{NFW}}$ , and c/a provides improved predictions for subhalo abundance compared to one which depends on concentration alone.

Fig. 3.13 depicts the distribution of coefficients derived when OLS regression for the one-parameter is applied to 10,000 bootstrap re-samplings of the set of re-simulated haloes. Bootstraps provide a reliable way to obtain errors from regression fitting even in the presence of covariances or incorrect error models. The distribution of regression coefficients amongst the bootstrap samples approximates the true PDFs for those parameters; we can use statistics of the distribution of these values as estimates for parameter uncertainties. In Fig. 3.13 the  $c_{\rm NFW}$  coefficients are always non-zero, which indicates that the improvement to predictions for subhalo abundance from including this parameter is statistically significant.

Similarly, Fig. 3.14 shows the distribution of coefficients derived when OLS linear regression is applied to 10,000 bootstrap re-samplings with replacement of the three-parameter model. In the histograms for  $\lambda_{\rm B}$ , only 3.9% of the coefficients are  $\geq 0$  and in the histogram for c/a, 17.5% of the coefficients are  $\geq 0$ . This means that these two terms are not as dominant as  $c_{\rm NFW}$  in the subhalo prediction, and in particular c/a is noticeably outweighed by the other two parameters (typically for significance we look for < 5% of the coefficients crossing over the x = 0 line).

For multi-parameter models it is useful to also examine how the coefficients of each parameter correlate with each other, which helps illuminate degeneracies between parameters in the fitting. As can be seen in Fig. 3.14,  $c_{\rm NFW}$  and  $\lambda_{\rm B}$  are the most important parameters for this model since they show little correlation, indicating that they each provide almost totally distinct information. In contrast, the scatter plot between the  $\lambda_{\rm B}$  and c/acoefficients amongst the bootstrap samples shows significant covariance between them. This indicates that there is likely some redundant information between the two halo parameters (which is no surprise given their relationship to each other). We conclude that although the three-parameter model provides useful predictions for subhalo abundance, not all of these parameters are contributing equally. We also point out that the more parameters there are in a model, errors on these parameters will propagate into subhalo abundance predictions. Current estimates of the Milky Way host halo properties are not very exact - therefore at current times a three-parameter model will likely have larger errors than a one-parameter model (as seen in Fig. 3.3 and Fig. 3.4) but in the future when measurements are more accurate the three-parameter model is expected to provide a better prediction for subhalo



Figure 3.13: The distribution of the coefficient of  $c_{\rm NFW}$  from linear regression derived from one thousand bootstraps with replacement of the simulated halo samples. The statistics of the distribution of coefficients from bootstrapping can be used to characterize the uncertainties in the regression parameters. In this case, because the  $c_{\rm NFW}$  coefficients are always non-zero, we have firmly established that this parameter can be used to improve predictions of subhalo abundance.

abundance.

#### **3.6.4** Maximum Likelihood Fitting of a Power-Law Model for Subhalo Abundance

The linear one-parameter and three-parameter models presented above have been established to be better predictors for subhalo abundance than mass alone. In this section we investigate the improvements possible when models based upon power-law scaling relations, rather than a linear combination of parameters, are used. We also implement improved fitting methodologies that work well even in the small- $N_{\rm sub}$  regime.

Whereas we can write a linear model for subhalo abundance as

$$N_{\rm sub}^{\rm pred} = k + \sum_{i} \alpha_i \times x_i, \tag{20}$$

where k corresponds to the intercept of a regression fit and  $\alpha_i$  is the coefficient for the  $i^{\text{th}}$  halo parameter  $x_i$  used in the regression (e.g.,  $c_{\text{NFW}}$ ), we can similarly define a power-law model as

$$N_{\rm sub}^{\rm pred} = k \prod_{i} x_i^{\alpha_i}; \tag{21}$$

in this case the  $\alpha_i$  are the exponents of the parameters  $x_i$ , rather than their linear coefficients.

In the previous appendices, we investigated subhalo abundance models for a case where the net number of subhaloes per halo was large, such that the Poisson distribution may be closely approximated by a Gaussian and the assumptions of least-squares regression apply. However, for higher velocity thresholds where there may be only a few subhaloes per halo, this assumption breaks down. Instead, we adopt a Poisson distribution-based maximum likelihood approach which provides secure results even in this domain. For Poisson-distributed counts the natural logarithm of the likelihood for the number of subhaloes of the *j*th halo,  $\ln P(N_{sub}^{meas}|N_{sub}^{pred})$ , is given by the formula

$$\ln \mathcal{P}(\mathcal{N}_{\text{sub}}^{j,\text{meas}}|\mathcal{N}_{\text{sub}}^{j,\text{pred}}) = N_{\text{sub}}^{j,\text{meas}} \ln N_{\text{sub}}^{j,\text{pred}} - N_{\text{sub}}^{j,\text{pred}} - \ln \Gamma(N_{\text{sub}}^{j,\text{meas}} + 1),$$
(22)

where  $N_{\text{sub}}^{j,\text{meas}}$  is the observed number of subhaloes for a given velocity threshold for the  $j^{\text{th}}$  host;  $N_{\text{sub}}^{j,\text{pred}}$  is the predicted number of subhaloes for the  $j^{\text{th}}$  host from a given model and the same velocity threshold; and  $\Gamma$  indicates the standard Gamma function. We can



Figure 3.14: Distributions of parameter coefficients from bootstrap re-samplings of the subhalo abundance data for the three-parameter regression depicted in Fig. 3.12. The histograms correspond to how many times each value of a given coefficient occurs amongst the bootstrap samples. We can treat the distribution of coefficients from the bootstraps as a PDF for those coefficients to determine errors in regression fits. The scatter plots show pairs of the coefficients from each bootstrap fit plotted against each other. The scatter plot for  $\lambda_{\rm B}$  vs c/a shows that the coefficients have a significant correlation with each other. It is clear that not all halo parameters contribute equally to our ability to predict subhalo number; however, the AIC and BIC results provide strong evidence that a three-parameter model is superior to one that relies on concentration alone.

maximize the likelihood of a model given the set of halo simulations by maximizing the product of their individual likelihoods (as they are independent draws from the underlying distributions the net likelihood is the product of the individual likelihoods). However, this is equivalent to maximizing the sum of the values of the log likelihoods for each halo, i.e.,  $\sum_{j} \ln P(N_{sub}^{j,meas}|N_{sub}^{j,pred})$ , so we do the latter.

Our fit values of k and the power law exponents  $(\alpha_i)$  correspond to the values which maximize the total log likelihood,  $\sum \ln P(N_{sub}^{meas}|N_{sub}^{pred})$ . We determine this values by minimizing the negative of the total log likelihood using the scipy.optomize.minimize() function [364, 291] with the Nelder-Mead solver [276]. This function requires an initial guess which we construct by linear regression for the ln of  $N_{sub}^{meas}$  in terms of the ln's of the  $x_i$ . The intercept of this linear fit should correspond to the ln of the k parameter in Equation 21, while the power-law exponents  $\alpha_i$  should correspond to the linear coefficients of this regression.

Goodness-of-fit statistics for the power-law Poisson Maximum Likelihood fits of both the one-parameter model and three-parameter model for the total subhalo abundance above the effective resolution limit of  $V_{\text{max}}^{\text{frac}} = 0.065$  are given in Table 3.3. In both cases a power-law model provides a better fit than the equivalent linear model. The  $\Delta$ AIC and  $\Delta$ BIC for the linear versus the power-law models are > 3 in the case of one-parameter models and >~ 10 for three-parameter models, indicating that in each case the power-law model provides a superior representation of the data.

Similarly to Fig. 3.11, and Fig. 3.12, we can plot the equivalent of a residual plot for the three-parameter power law model, which we provide in Fig. 3.15. In this case the y-axis differs, as for a power-law model ratios, not differences, are more meaningful. Hence we plot  $N_{\text{sub}}^{\text{meas}}/N_{\text{sub}}^{\neg x_i}$ , where  $N_{\text{sub}}^{\neg x_i}$  is still the predicted subhalo abundance without including the host parameter  $x_i$ , but now using a power-law model instead of a linear one. Over-plotted in orange is the result of the power-law fit for the quantity plotted in a given panel,  $x_i^{\alpha_i}$ . For reference we also indicate the five nearest neighbors to the Milky Way as blue triangle points. Comparing Fig. 3.15 to Fig. 3.12, it is evident that the power-law Poisson maximum likelihood regression does a better job than the OLS linear regression as a predictor for subhalo abundance, and once more  $c_{\text{NFW}}$  is the parameter that has the greatest predictive power.



Figure 3.15: Illustration of the three-parameter power-law model for subhalo abundance as a function of the parent halo's spin, concentration, and shape (from left to right:  $\lambda_{\rm B}$ ,  $c_{\rm NFW}$ , and c/a). The y-axis shows the observed ( $N_{\rm sub}^{\rm meas}$  subhalo abundance for a given halo divided by the predicted subhalo abundance for that halo from a power-law model as described in Equation 21, where all terms except for the one dependent on the x axis value are included in the prediction ( $N_{\rm sub}^{-x_i}$ )). The over-plotted orange line in each panel corresponds to the parameter shown on the x-axis raised to the power of the best-fit exponent from our Poisson maximum likelihood fit. The blue triangle points indicate the five nearest neighbors to the Milky Way, identified as discussed in Section 3.6.2. This result shows that a power-law is a better fit to the data than a linear fit, and that  $c_{\rm NFW}$  has the greatest predictive power for subhalo abundance.

To determine the errors in the exponents from the power-law Poisson maximum likelihood fit we again use bootstrap re-sampling. Specifically, we produce 10,000 bootstrap re-samples with replacement of the ensemble of host haloes and perform the fit for each sample. Fig. 3.16 displays the results of this process for the power-law fits; it may be compared to Fig. 3.14. The dependence upon halo concentration is strongest, while the dependence upon spin is relatively weak and covariant with the shape dependence. Unlike in the case of the linear model (Fig. 3.14) there appears to be a significant degeneracy between  $c_{\text{NFW}}$  and c/a.

In order to predict the cumulative subhalo abundance for the Milky Way as a function of fractional velocity, we carry out this fitting procedure for a series of values for the minimum  $V_{\text{max}}^{\text{frac}}$ . Specifically, our procedure is as follows:

(1) We select twenty logarithmically-spaced values of  $V_{\text{max}}^{\text{frac}}$  to serve as the minimum fractional velocity for each of twenty bins, and determine the total number of subhaloes above every minimum value for each host halo.

(2) For each of the twenty minimum values of  $V_{\text{max}}^{\text{frac}}$  we fit a power-law model for predicting  $N_{\text{sub}}$  of the form given by Equation 21 by maximizing the total Poisson likelihood across all 45 host haloes (using Equation 22), as described above.

(3) We obtain a prediction for the abundance of subhaloes for the Milky Way above each fractional velocity threshold by evaluating the corresponding power-law fit with the Milky Way's estimated halo properties (described in Section 3.2) substituted in. That is, given the  $\alpha_i$ 's and k values from a Poisson maximum likelihood fit, we can predict the cumulative subhalo number for the Milky Way corresponding to each minimum value of  $V_{\text{max}}^{\text{sat}}/V_{\text{max}}^{\text{host}}$  from the equation

$$N^{\text{sat}}(>V_{\text{max}}^{\text{frac}}) = k \prod_{i} x_{i,MW}^{\alpha_{i}}, \tag{23}$$

where  $x_{i,MW}$  is the estimated value of the *i*'th halo property for the Milky Way's host halo. These predictions correspond to the solid purple curves in Fig. 3.3, Fig. 3.4, and Fig. 3.5.

(4) We then perform bootstrap re-sampling amongst the host haloes in order to calculate the uncertainty in our power-law fits, and in the parameter estimates for the Milky Way. For each bootstrap we perturb the measured Milky Way host halo properties by a random draw from a normal distribution with standard deviation set by that property's estimated error. The case of concentration is a bit unique because our concentration estimate does not have



Figure 3.16: Distributions of parameter exponents from bootstrap re-samplings of the subhalo abundance data for the three-parameter power-law Poisson maximum likelihood fit depicted in Fig. 3.15. The histograms correspond to how many times each value of a given coefficient occurs amongst the bootstrap samples. The scatter plots show pairs of the coefficients from each bootstrap fit plotted against each other. The scatter plot for c/a vs  $c_{\rm NFW}$  and c/a vs  $\lambda_{\rm B}$  show that the coefficients have a correlation with each other. This implies that c/a is not contributing as much as the other two parameters in predicting subhalo number. However, the statistics in Table 3.3 show that the three-parameter model is still the superior model for predicting subhalo abundance for Milky Way-mass dark matter haloes.

symmetric errors. We instead build two half-Gaussians with the same mean and construct the CDF of this distribution. With this we can use the PDF as a look-up table and generate random variables that correspond to that value of concentration. At the same time we select a new sample of the simulated host haloes with replacement. With this new sample we then refit for the power-law model parameters via Poisson maximum likelihood and evaluate each model with the corresponding estimated Milky Way halo properties for each bin. The resulting 68% and 95% confidence intervals of the cumulative velocity functions for each bootstrap sample correspond to the semitransparent purple regions in Fig. 3.3, Fig. 3.4, and Fig. 3.5.

(5) Last, with this bootstrapped set of samples we can determine the impact of Poisson scatter on the range of values possible for the Milky Way. For each bootstrap, in each bin of  $V_{\text{max}}^{\text{frac}}$  we compute the predicted number of subhaloes for the Milky Way per bin for a non-cumulative satellite abundance. Then we draw randomly from a Poisson distribution using this non-cumulative predicted abundance as the mean. Last, these Poisson values are cumulatively summed in order to determine the error corresponding to the scatter of the 45 haloes about the model prediction. The 68 and 95 percent confidence regions for this result are depicted in orange in Fig. 3.3, Fig. 3.4, and Fig. 3.5. Applying a log-normal error distribution yields very similar results to the Poisson scatter utilized here. We also compare the errors for the total cumulative number of subhaloes from the bootstraps alone and the additional impact of Poisson scatter in Fig. 3.8. The incorporation of Poisson noise leads to a much wider spread in predicted subhalo abundance at higher  $V_{\text{max}}$  and a comparable spread at low  $V_{\text{max}}$ . Percentile confidence regions for these two error models on total subhalo abundance normalized by the mean measured subhalo abundance are shown numerically in Table 3.2 and visually in Fig. 3.8.

The end result of this process is both a best-fit, best-estimate cumulative velocity function for the Milky Way and a set of additional CVFs whose distribution reflects the uncertainties in fitting power-law models, uncertainties in the parameters of the Milky Way host halo, and uncertainties due to Poisson scatter about the relation.

#### 3.6.5 Fitting Functions for the Subhalo Cumulative Velocity Function

Our best-fit models for the cumulative velocity function of subhaloes as a function of host halo parameters can be approximately described by a simple set of formulae. Our goal is to provide easily-computed values which can be substituted into Equation 23 to predict the cumulative velocity function for any dark matter halo. We take advantage of the fact that (due to the self-similarity of dark matter haloes of different masses and the virial scaling relations), the number of subhaloes above a given  $V_{\text{max}}^{\text{frac}}$  should have minimal dependence on halo mass. That is, we expect the number of satellites above a given *fraction* of a halo's  $V_{\text{max}}$  to be similar regardless of halo mass, but in a more massive halo, the satellites that are above that fractional threshold will also be more massive (this also leads to the conclusion that there will be more total satellites above a fixed minimum  $V_{\text{max}}$  in more massive haloes, roughly proportional to the halo mass).

To define an approximate model for the CVF, we need to specify values of k and  $\alpha_i$  as a function of the minimum subhalo fractional velocity considered,  $V_{\text{max}}^{\text{frac}}$ . Specifically, the results from Poisson maximum likelihood one-parameter power-law fits can be accurately represented using a function that is linear at low velocities and quadratic at high velocities. We have fit for this function using the Numpy curve\_fit() routine applied to the values of  $\alpha_{\text{cNFW}}$  as a function of  $V_{\text{max}}^{\text{frac}}$  obtained from the Poisson maximum likelihood results, using weights calculated from the standard deviation of the bootstrap results for this parameter at a given velocity. From this we obtain a fit:

$$\alpha_{c_{\rm NFW}} = \begin{cases} 0.12V^{\dagger} - 0.78, & V^{\dagger} < 0, \\ -66.26V^{\dagger 2} - 1.02V^{\dagger} - 0.78, & V^{\dagger} \ge 0, \end{cases}$$
(24)

where  $V^{\dagger} = V_{\text{max}}^{\text{frac}} - 0.12$ . This fit provides a good representation of the dependence of the power-law exponent on velocity over the range  $0.05 \le V_{\text{max}}^{\text{frac}} \le 0.25$ .

We find that a single quadratic function of  $V_{\text{max}}^{\text{frac}}$  is sufficient for characterizing the dependence of the power-law pre-factor k on velocity. In order to ensure a self-consistent set of fitting functions, we must obtain new values of k for each threshold velocity while forcing  $\alpha_{c_{\text{NFW}}}$  to have the value predicted by Equation 24 for  $\alpha_{c_{\text{NFW}}}$ . Specifically, we take  $k = \langle N_{\text{sub}}^{\text{meas}}/N_{k=1}^{\text{sat}}(>V_{\text{max}}^{\text{sat}})\rangle$ , where  $N_{\text{sub}}^{\text{meas}}/N_{k=1}^{\text{sat}}(>V_{\text{max}}^{\text{sat}})$  is the satellite abundance estimate

for each host halo taking k=1 and the value of  $\alpha_{c_{\text{NFW}}}$  from Equation 24; i.e.,  $N_{k=1}^{\text{sat}}(>V_{\text{max}}^{\text{sat}}) = c_{\text{NFW}}^{\alpha_{c_{\text{NFW}}}}$ . By a least-squares fit to the values of the logarithm of k as a function of threshold velocity, we obtain

$$\ln k = 205.58 (V_{\rm max}^{\rm frac})^2 - 73.66 V_{\rm max}^{\rm frac} + 10.93.$$
<sup>(25)</sup>

The results of Equation 24 and Equation 25 can then be substituted into Equation 21 to obtain a prediction for the subhalo abundance at a given threshold in velocity based on a dark matter halo's concentration. The results of these fitting formulae match the mean predictions for the Milky Way (corresponding to the purple line in Fig. 3.3) to better than 2.7% RMS over the range in velocities  $0.05 \leq V_{\text{max}}^{\text{frac}} \leq 0.25$ .

We similarly have derived fitting functions which can be used to approximate our threeparameter model fits. We begin by noting that the exponents of the spin and shape parameters from our bootstrap samples are consistent with being constant at all velocities. As a result, for our fitting formula we treat them as fixed at their median values across all bootstrap samples from the original maximum likelihood calculation, corresponding to

$$\alpha_{\lambda_{\rm B}} = 0.03; \tag{26}$$

$$\alpha_{c/a} = 0.26. \tag{27}$$

We then perform new fits for a new k and  $\alpha_{c_{\text{NFW}}}$  via Poisson maximum likelihood, including spin and shape in the model but forcing their exponents to have the fixed values described above.

Then we proceed the same as for the one-parameter model. We find that the concentration exponent for the three-parameter model can be approximated well by

$$\alpha_{c_{\rm NFW}} = \begin{cases} 0.10V^{\dagger} - 0.91, & V^{\dagger} < 0, \\ -71.86V^{\dagger 2} + 2.02V^{\dagger} - 0.91, & V^{\dagger} \ge 0, \end{cases}$$
(28)

where  $V^{\dagger} = V_{\text{max}}^{\text{frac}} - 0.11$  in this case. Once again, we use the fitting formula predictions for  $\alpha^{c_{\text{NFW}}}$  to estimate the k term at each velocity, and then fit for  $\ln k$  as a function of  $V_{\text{max}}^{\text{frac}}$  via least-squares. We then find

$$\ln k = 208.21 (V_{\text{max}}^{\text{frac}})^2 - 74.11 V_{\text{max}}^{\text{frac}} + 11.47.$$
<sup>(29)</sup>

The results from Equation 28 and Equation 29 can be substituted into Equation 21 to obtain a prediction based on a dark matter halo's concentration, spin, and shape for subhalo abundance above a given velocity threshold. The results of these sets of fitting formulae match the mean Milky Way prediction (the solid purple line in Fig. 3.4) to 6.9% difference RMS for  $0.05 \leq V_{\text{max}}^{\text{frac}} \leq 0.25$ .

We note that if realizations of a cumulative velocity function incorporating Poisson statistics are desired, care must be taken to ensure that covariances between values at different velocities resulting from the use of a cumulative quantity are properly accounted for. Step five in Section 3.6.4 describes the procedure which we have employed for this purpose.

# 4.0 Constraining the Milky Way's Ultraviolet to Infrared SED with Gaussian Process Regression

This chapter has been accepted for publication in the Monthly Notices of the Royal Astronomical Society ©: 2021 Catherine Fielder. Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved. Minor modifications have been made to the text.

# 4.1 Introduction

Within the Milky Way we have a unique opportunity to study the nuances of galactic properties, allowing us to test galaxy formation and evolution models at an unrivaled level of detail. For example, chemical abundances for hundreds of thousands of stars have been obtained from spectroscopic surveys [233, 240], and stellar surveys that catalog distance and dynamical measurements on millions of stars have been performed [132]. This exhaustive stellar information has helped constrain the Milky Way's evolutionary history. In turn, high-resolution dynamical simulations have been able to produce galaxies of increased similarity to the Milky Way [157, 316, 376], matching fundamental galaxy properties such as dwarf satellite populations and reproducing characteristics of our Galaxy's gas, dust, and stellar components. Comparisons between Milky Way stellar data and high-resolution hydrodynamical simulations of Milky Way-like galaxies are an increasingly useful way to improve our understanding of galaxy formation.

However, the marriage between observations and models is delicate: incorrect assumptions on one side can propagate into the other. Simulators must make choices about how to implement crucial parameters that affect the galaxy evolution process, such as the gas density threshold for star formation to occur and the efficiency with which it proceeds; this is sometimes done by attempting to match observed properties of the Milky Way. But without knowing how our Galaxy fits in amongst the broader galaxy population, it is difficult to determine whether simulations match the Milky Way because they have the correct physics or because they have incorrectly tuned parameters that match by coincidence or design.

This is complicated by the exceptional difficulty of obtaining a global picture of the Milky Way, given our location in the disk and the obscuration caused by interstellar dust. As a result, there are properties that we can easily measure in external galaxies that are impossible to measure directly within our own, making it difficult to determine where we fit within the broader galaxy population.

Creating an outside-in picture of the Milky Way that spans a multitude of broad-band wavelengths will enable simulators to more accurately tune their physics assumptions, as it will then be possible to test whether quantities that can only be determined from large-scale stellar surveys and those that can be measured directly only for extra-galactic objects are reproduced. The most basic, easiest-to-measure intrinsic quantities we can use to study galaxies are their luminosities and colors; once redshift is known, these can be inferred from broad-band photometry. Hence the focus of this chapter will be in determining these properties for the Milky Way. This will enable our Galaxy to be placed on standard color-magnitude and color-color diagrams and result in a multi-wavelength spectral energy distribution (SED) for the Milky Way.

The radiation emitted by a galaxy is characterized by its spectral energy distribution (SED), or flux as a function of wavelength. Galaxy SEDs contain the imprints of the physical processes occurring within - the stellar population's ages and abundances (i.e., the star formation history and metallicity of the galaxy), the dust and gas content and the chemistry and physical state of the interstellar medium (ISM). Because different sources dominate the emission at different wavelengths, long-wavelength-baseline SEDs allow one to disentangle the contributing effects. This makes SEDs one of the best direct probes for studying galaxy formation and evolution from both an observational and theoretical modeling perspective.

However, comparing colors and luminosities of the Milky Way to external galaxies is not trivial, regardless of whether we compare to observed galaxies or to mock images from highresolution hydrodynamical simulations (such as Eris [157], APOSTLE [316], and Latte [376]). Much of our view of the Galaxy is obscured by interstellar dust, especially at UV and optical wavelengths [62, 320]. Stars outside of the local solar region are reddened as a result of the
dust obscuration. Determining the integrated light of stellar populations in the Milky Way is challenging due to the spread of stars over large and varying distances, with correspondingly large and varying dust extinction along lines of sight to the Earth. This makes the study of any portions of the Galactic disk beyond the solar neighborhood exceptionally difficult, and results in a fragmented picture of the Milky Way. Integrated properties that are relatively painless to obtain in external galaxies (though dust obscuration can affect these observations as well, see e.g., Ref. [243, 245]), are impossible to obtain directly within our own Galaxy (e.g., [265]). As a result, simulators often must resort to comparing their simulated Milky Ways to very general galaxy populations (such as sets of Sbc or late type galaxies) that, while superficially resembling the Milky Way, have a wide range of other global properties (e.g., [157]).

In an effort to circumvent our limited view of the Milky Way, we can study galaxies that mimic the properties of our Galaxy but can be observed from outside, which we label Milky Way Analogues (MWAs). This method hinges on the Copernican assumption that the Milky Way should not be extraordinary among a galaxy population that shares some key properties with it. These comparisons are enabled by working within volume-limited subsets of large surveys, which ensure that the observed objects constitute a representative population. Previous work suggests that galaxies with similar stellar mass and star formation rates are also similar in other properties, as the observed galaxy population is well-matched by models that parameterize galactic star formation histories with a limited collection of curves [18, 146, 2, 190]. Even further, Ref. [19] showed mass and star formation rate are strongly correlated with the photometric properties of a galaxy. Therefore we can exploit the fact that two galaxies of identical mass and star formation rate should have similar luminosities and colors, with some scatter given the range of galaxy photometric properties at fixed physical parameters. Ref. [219], utilized this to constrain the Milky Way's optical colors and magnitudes based on the range of observed properties of MWAs that were matched in stellar mass and star formation rate. MWAs also allow direct comparison of properties of our Galaxy to its closest peers (e.g., [217, 218, 129, 35, 206]) and have been a successful tool for improving our understanding of the Milky Way in an extra-galactic context.

The Milky Way, however, has some characteristics that are atypical (at the  $< 2\sigma$  level)

amongst its peers - e.g., the Milky Way has an unusually compact disk (i.e., a small disk scalelength) [41, 28, 217], and an unusually quiescent merger history (from observation; [352, 306], and simulation; e.g., [124, 63]). The deviations of the Milky Way from the average suggest that we should consider parameters beyond just stellar mass and star formation rate in order to identify samples of objects that more closely resemble the Milky Way.

Galaxy morphological characteristics such as disk scale length  $(R_d)$  and bulge-to-total ratio (B/T) are tied to a galaxy's evolutionary history and therefore should connect to its photometric properties [61, 307] as well as to the ways in which the Milky Way is atypical. We would therefore wish to incorporate these properties in addition to stellar mass and star formation rate in defining an MWA. However, as the number of parameters required to match the Milky Way increases, the number of MWAs correspondingly reduces dramatically. For example Ref. [129] only found 179 analogues when selecting on stellar mass, bulge-to-total mass ratio, and morphology; Ref. [36] and Ref. [35] found no MWAs within  $1\sigma$  of the MW when selecting on stellar mass, star formation rate, bulge-to-total ratio, and disk scale length in either the SDSS-IV MaNGA survey [59] or a larger photometric sample drawn from the GALEX-SDSS-WISE Legacy catalog (GSWLC; Ref. [312]) respectively.

Ref. [219] found that the color of the Milky Way is consistent the green valley region of the CMD as it has been defined using purely optical passbands. Characterizing the UV and IR colors of the Milky Way can provide more sensitive probes of whether it would be classified as in the process of quenching if seen from outside, allowing us to better understand what type of population the Milky Way may belong to. Ref. [219] speculated that the Milky Way might belong to the population of massive "red spiral" galaxies, which are characterized by their red optical colors despite ongoing star formation [244, 81]. Galaxies within this population may be moving into the green valley due to slow quenching (cf. Ref. [319]). This conjecture can only be fully tested by examining wavelengths outside of the optical range; in g - r, the colors of massive spiral galaxies on the star-forming main sequence (a population that should include the Galaxy) overlap with both the red sequence and the blue cloud [81, 310]. However, samples of MWAs that have high-quality photometry over a broader wavelength range will have reduced numbers due to the limited coverage of sufficiently deep photometry in GALEX. To address the lack of analogues when multi-dimensional parameter spaces are used, and the smaller overall sample size resulting from the increase in wavelength coverage, we introduce a Gaussian process regression (GPR) approach in this work. GPR is an emergent tool in astrophysics. For example, Ref. [37] employed GPR to emulate results of cosmological simulations, while Ref. [150] used GPR to detect and classify exoplanets. We can use GPR to leverage information from a wider variety of galaxies, instead of just the closest Milky Way analogues, in order to extract information from large-scale trends between galaxy physical and photometric properties. Thanks to the probabilistic framework that underlies GPR, we obtain uncertainty estimates for all predicted quantities for free. The primary result from this chapter will be an ultraviolet to infrared SED of the Milky Way as viewed face-on, determined via GPR based on star formation history and structural parameters (i.e., galaxy physical parameters) that have been measured well for both the Milky Way and galaxies from the Sloan Digital Sky Survey [4].

The chapter is organized in the following manner. In Section 4.2 we describe the observational data used, including the external galaxy data in Section 4.2.1 and Section 4.2.2, and estimated properties of the Milky Way in Section 4.2.3. Section 4.3 details our new Gaussian process regression-based methodology. In Section 4.4 we compare the luminosity and colors of the Milky Way at multiple wavelengths and the Milky Way's predicted SED to properties of other galaxies. Finally we summarize our results, and discuss implications and future work in Section 4.5. 4.6 provides a summary of the galaxy parameters and tables that list predicted photometry for the Milky Way. 4.7 describes tests of the accuracy of the GPR procedures used here, and 4.8 describes how we address the systematic corrections needed for k-corrections and Eddington bias.

In this chapter all magnitudes are reported in the AB system, except for the Johnson-Cousin UBVRI magnitudes which are presented in the Vega system. Absolute magnitudes are derived using a Hubble constant  $H_0 = 100 \text{ km s}^{-1}\text{Mpc}^{-1}$ , so they are equivalent to  $M_y - 5 \log h$  (where  $M_y$  is the y-band absolute magnitude and  $h = H_0/100$ ) for other values of h. For other properties in which measurements for the Milky Way are compared to extragalactic galaxy measurements we assume  $H_0 = 70 \text{ km s}^{-1}\text{Mpc}^{-1}$  (h = 0.7) in accordance with Ref. [219], for a standard flat  $\Lambda$ CDM cosmology with  $\Omega_m = 0.3$ . Parameters such as log stellar mass and log star formation rate can be modified for different h values by subtracting  $2 \log h/0.7$ . We do this to avoid confusion and to allow for potential updates to future h measurements.

## 4.2 Observational Data

In this section we describe the many galaxy catalogs utilized in this work. We break this up by photometry (Section 4.2.1) and inferred galaxy properties (Section 4.2.2), with the Milky Way measurements included in the final subsection (Section 4.2.3).

### 4.2.1 Photometry

# 4.2.1.1 SDSS Galaxies

The sample of galaxies that we use as a starting point originates from the eighth data release (DR8; Ref. [4]) of the Sloan Digital Sky Survey III (SDSS-III; Ref. [390]). DR8 provides both images and photometry of thousands (almost  $10^6$ ) of local galaxies. The optical broadband passbands, u, g, r, i, and z were the subjects of previous Milky Way analogue work by Ref. [219] and are used in this study in addition to bands outside of the optical range.

We make use of both the "model" and "cmodel" magnitudes from SDSS. The former refers to magnitudes derived from the better of either a de Vaucouleurs or an exponential profile fit to the galaxy surface brightness distribution. These types of magnitudes are expected to produce the highest signal-to-noise estimate of galaxy colors; thus when we refer to galaxy colors derived from SDSS we will be using model magnitudes for the calculations. Alternatively, cmodel magnitudes are derived from the best fit to a linear combination of a de Vaucouleurs and exponential profile. These magnitudes provide the best estimate of the total flux of a galaxy in each passband. When we refer to galaxy absolute magnitudes for SDSS bands we will use cmodel magnitudes.

k-corrections on these passbands to rest-frame z = 0 were calculated via the KCORRECT

v4.2 software [31], as described in Ref. [219]. This provided AB absolute magnitudes for the SDSS *ugriz* photometry. Additionally, KCORRECT was used to convert the SDSS *ugriz* photometry to restframe Johnsons-Cousins *UVBRI* Vega magnitudes in order to make easy comparisons to literature values. Results are presented with the adoption of the Ref. [31] and Ref. [219] notation, where an absolute magnitude of passband y at redshift z is denoted as  ${}^{z}M_{y}$ .

Our main galaxy sample is derived from the volume-limited sample presented in Ref. [219]. A volume-limited sample is required for accurate results from Milky Way analogues in order to alleviate a radial selection effect known as Malmquist bias, i.e., the preferential inclusion of intrinsically bright galaxies. At higher redshifts within the main SDSS sample [339], only the most luminous galaxies will be brighter than the sample magnitude limit and followed-up spectroscopically. By using a volume-limited sample we ensure that galaxies within the range of the Milky Way's parameters are included equally at all distances considered. Ref. [219] determined the limits for their volume-limited sample from an initial draw of Milky Way analogues from the full SDSS DR8 parent catalog without any redshift cuts. Then in  ${}^0(g-r)$ vs.  ${}^{0}M_{r}$  (i.e., restframe g-r color derived using z = 0 passbands versus r-band absolute magnitude, again evaluated with the z = 0 passband) color-magnitude space a maximum redshift was chosen such that all objects as low in luminosity as the faintest Milky Way analogues would still be included at that z. A minimum redshift was also applied to limit the impact of the finite SDSS fiber aperture on measured galaxy properties. The resulting volume-limited sample contains a total of 124,232 target galaxies within the redshift range of 0.03 < z < 0.09. Some initial cuts on SDSS quality flags were also employed; for further details on the construction of this volume-limited sample refer to Section 3.1 of [219]. All cross-matches from SDSS to other catalogs presented here were constructed only using the volume-limited sample. Both the SDSS sample used here and the cross-matched catalogs presented in this chapter are available at our catalog GitHub repository<sup>1</sup>.

<sup>&</sup>lt;sup>1</sup>https://github.com/cfielder/Catalogs

# 4.2.1.2 GALEX–SDSS–WISE Legacy Catalog

Photometry in ultraviolet and infrared wavelengths used in this work comes from the GSWLC [312, 311]. We use GSWLC-M2, the medium-deep catalog of GSWLC-2, which covers 49% of the SDSS DR10 footprint. While this reduces the number of targets for study, the improved signal-to-noise in the UV-imaging over the shallow catalog enables tighter results. SDSS Photometry between DR7/DR8 and DR10 is the same, so no issue arises between cross matches of the SDSS volume-limited sample and the GSWLC-M2 sample. In order to account for any differences in astrometry we consider matches to be separations within 1.5 arcseconds.

The ultraviolet sample in the GSWLC catalog originates from the GALEX survey (FUV, NUV), and has been corrected for galactic reddening and calibration errors [241]. UV detections are available for 74% of SDSS targets in the GSWLC-M2 catalog. The infrared sample in the GSWLC catalog originated from the 2MASS and WISE surveys. The 2MASS photometry (JHKs) is available for 48% of SDSS targets and WISE photometry (W1, W2, W3, W4) is available for 41% of SDSS targets.

# 4.2.1.3 DESI Legacy Imaging Surveys

Data release 8 of the DESI (Dark Energy Spectroscopic Instrument) Legacy imaging surveys includes g, r, z, and WISE photometry of sources detected in DECam or BASS/MzLS imaging [96]. We use these catalogs for all WISE-band photometry presented here, as the matched-model measurements from the Legacy Surveys Tractor catalogs go substantially deeper and have lower errors than other public WISE data products. We also calculate k-corrections for the WISE bands using this photometry; we discuss our procedures for this, which allow accurate k corrections in the IR without making any assumptions about SED templates at those wavelengths, briefly in Section 4.8.1 and more extensively in Chapter 5.

DESI Legacy and SDSS contain photometry that comes from differing filter sets and detectors. DESI Legacy North uses BASS and MOSAIC filters while DESI Legacy South used DECam filters [96]. Our corrections for offsets between the DESI Legacy Survey and SDSS photometry is described in Section 4.8.2.

## 4.2.2 SDSS-based Properties

Only one of the galaxy properties which we use to create SEDs is recorded in the standard SDSS photometric catalogs and required no further analysis; specifically, the best fit axis ratio (b/a) of a galaxy. We use the b/a value determined from an exponential fit to the galaxy's surface brightness density in the r-band throughout this work.

Axis ratio is used here as a proxy for galaxy inclination. Disk galaxies with very small axis ratios are likely heavily tilted from our perspective, in contrast to face-on disks which would typically have axis ratios of  $b/a \sim 0.9$  [72, 234]. Highly-inclined star forming galaxies suffer greatly from dust obscuration which can effect their colors and magnitudes by making them appear redder and fainter than their intrinsic properties [79, 245, 263, 203]. This effect is important to consider when we are making predictions for a disk galaxy like the Milky Way.

# 4.2.2.1 MPA-JHU Masses and Star Formation Rates

The MPA-JHU galaxy property catalog provides total mass and star formation rate estimates for galaxies within SDSS DR7. Ref. [219] developed an updated version of this catalog using the SDSS DR8 photometry. To summarise, the masses are calculated by fitting stellar population synthesis model to the galaxy's photometry (instead of spectral features, as in Ref. [189, 133]), similar to Ref. [313]. The star formation rates are calculated by emission-line modelling based upon Ref. [46] with some updates. In order to account for aperture bias Ref. [219] follows the method of Ref. [313] in calculating photometry for the light that falls outside of the fibre and fitting stochastic stellar population synthesis models to it. Thus each galaxy SFR measurement consists of a combination of the SFR measured from inside and outside of the SDSS fibre. For more details on these calculations refer to Section 2.2.2 of Ref. [219] and Ref. [46]. In both cases a Kroupa initial mass function (broken power law) is assumed [207].

Each galaxy in our volume-limited sample is assigned a posterior probability distribution function (PDF) of log stellar mass and log star formation rate, as well as a corresponding cumulative distribution function (CDF) determined by the posterior. Nominal values used for the log  $M_*$  and log SFR of each object are taken to be the mean value of the posterior. The standard deviation ( $\sigma$ ) on these values are calculated from percentiles (P) in the CDF provided in the MPA-JHU measurements; we take  $\sigma = (P_{84} - P_{16})/2$ , i.e., half the difference between the 16<sup>th</sup> and 84<sup>th</sup> percentile values. This gives us estimates of the stellar mass and star formation rates as well as their associated errors for each galaxy in the sample.

The GSWLC-2 catalogs which we use for photometry also provide computations of stellar masses and star formation rates. We chose to use the MPA-JHU results instead in order to avoid any systematic effects that may result from using the same measurements for our galaxy properties that were used to determine these derived quantities. We have tested how our results change if we use GSWLC-2 masses and star formation rates and find minimal differences, as summarized in Section 4.4.2.4.

# 4.2.2.2 Simard et al. Bulge and Disk Decompositions

In order to characterize the light-weighted bulge-to-total ratios (B/T) and disk scale lengths ( $\mathbf{R}_d$ ) for SDSS galaxies, we use the catalog of Ref. [327]. This work performed galaxy image decompositions for objects within the Legacy area in SDSS via the GIM2D software package [328].

Specifically, we use the catalog of fits based on composites of a Sersic n = 4 bulge and a pure exponential disk (Sersic n = 1). For our analysis we use the B/T computed in r band, as it is expected to be more stable than g-based measurements. As a check we have performed our analyses using  $B/T_g$  instead of  $B/T_r$  and found minimal differences (cf. Section 4.4.2.4). It is worth noting that Ref. [208] finds that these bulge+disk decompositions can be somewhat inaccurate when applied to strongly-barred galaxies, which may lead to some biases within our galaxy sample. However, fits optimized for barred galaxies have been performed for only samples of a few thousand objects, inadequate for our purposes, so we rely on the [327] results here. The Ref. [328] bulge and disk decompositions are derived using a  $H_0 = 70 \text{ kms}^{-1}\text{Mpc}^{-1}$ .

## 4.2.2.3 Galaxy Zoo 2 Bar Presence

The presence of bars have been speculated to be correlated to the star formation history within a galaxy. However, the sense of the effect is unknown; a bar may be related to an increase in star formation [162, 161, 95, 174], no effect [293, 242, 67], or decreased star formation [361, 98, 130], along with other potential effects on color and metallicity [247]. Regardless of the source of any correlations, we wish to incorporate any possible differences having a bar may cause when we determine the Milky Way's SED, since our Galaxy does exhibit clear evidence of a bar [32]. The Galaxy Zoo 2 (GZ2) catalog [381, 160] contains identifications of detailed morphological features in disk galaxies, such as spiral arms and bar presence. The galaxies classified in Galaxy Zoo 2 are a sub-sample of the brightest and largest galaxies in the SDSS Main Galaxy Sample.

The Galaxy Zoo projects are open public projects in which members of the community identify whether they found a variety of galaxy features in the images provided. In the Galaxy Zoo 2 catalog the number of raw votes for each morphological feature is weighted (to account for user consistency) and adjusted to mitigate the impact of redshift-dependent biases, yielding a corrected fraction of volunteers who identified a given morphological feature for each galaxy. In our case we focus on the debiased fraction of volunteers who identified a galaxy as having a bar, which we denote by  $p_{bar}$  (following Galaxy Zoo labelling conventions; however, this is a fraction of votes, not a probability). Above some threshold in  $p_{bar}$  (after cuts in related parameters) we can be confident that a given galaxy indeed hosts a bar. Ref. [381] developed the initial version of Galaxy Zoo 2 bar thresholds; more conservative thresholds were later defined by Ref. [134].

Often, when one uses Galaxy Zoo results it is necessary to consider responses to previous questions that influence whether the question of interest is even presented to the volunteers. For example, as described in Ref. [381], a voter is only asked "Is there a sign of a bar feature through the center of the galaxy?" if they first selected that the object has "features or disk" when asked "Is the galaxy simply smooth and rounded, with no sign of a disk?", and then responded in the negative when asked "Could this be a disk viewed edge-on?". One would then identify galaxies that have received a large fraction of "yes" votes as containing bars. It is worth noting that Ref. [381] states galaxies that receive fewer than 10 net votes for a given question may not have a reliable classification. Therefore to construct a bar sample using the minimum allowances of Ref. [381], using the debiased vote fractions from Ref. [160] (which was an improvement to the debiasing of Ref. [381]), one would use the cuts in the second and third row of the third column of Table 3 in Ref. [381] with  $p_{\text{bar}} \ge 0.3$  as recommended by Ref. [134] and additional voting number thresholds  $N_{\text{not edgeon}} \ge 10$ , and  $N_{\text{bar}} \ge 10$ .

However, here we do not want to consider only information on the relationship between galaxy star formation history, structural properties, and photometric properties that comes from those galaxies that most definitively host bars. Rather, it is desirable for the training sample for our Gaussian process regression model to include objects spanning a broad range of parameter space: this yielded smaller net prediction errors in Milky Way photometric parameters than when small, restricted samples are used for training. For additionally discussion on this choice refer to Section 4.4.2.4.

To summarize, the primary set of data that we employ in this chapter consists of a cross-match between the original SDSS DR8 volume-limited sample reported in Ref. [219], an updated version of the MPA-JHU catalog [46] of masses and star formation rates in SDSS, the Ref. [327] morphological bulge-disk decomposition catalog, the GSWLC-2 medium-deep photometry catalog [312, 311], the Galaxy Zoo 2 catalog [381, 160], and the DESI Legacy survey DR8 catalog [96]. These cross-matches were performed using the astropy coordinates package. We required that objects be separated by less than 1.5 arcseconds from a counterpart in the SDSS volume-limited sample to be considered a match, and discard any galaxies that are not included in all catalogs considered.

Our sample is smaller than the original volume-limited sample as a result. The MPA-JHU catalog contains all objects from the volume-limited sample, so we begin with the same set of 124,232 galaxies utilised by Ref. [219]. After matching to Ref. [327], 123,167 galaxies remain; some objects are lost due to minor differences between the DR7 and DR8 SDSS catalogs. When we require GSWLC-M2 measurements, we are left with only 60,857 galaxies, as roughly half of the SDSS footprint is covered by medium-deep GALEX, 2MASS, and WISE photometry (see Section 4.2.1.2). After matching to GZ2 29,836 galaxies remain, a consequence of the brighter magnitude limit used to select GZ2 objects compared to the original SDSS Main Galaxy Sample (see Section 4.2.2.3). We note that due to the brighter magnitude limit of GZ2, our final sample is no longer volume-limited. This would yield biased results if we were measuring aggregate properties of Milky Way analogues. However, for our GPR methodology this only modulates the density of our training set within parameter space, causing larger prediction errors due to the sparser sampling, but not leading to a bias.

Finally, matching to DESI Legacy leads to only a minor reduction in galaxy number as it covers a super-set of the SDSS area with deeper photometry (but different reduction algorithms). After these cross matches we then remove objects with photometric values of "NaN", infinity, or -99, which all indicate missing photometry in one catalog or another. This is only done when necessary for the evaluation of the GPR. For our WISE *k*-correction calculations (see 4.8.1) we will exclude objects with large WISE photometry errors in a given band, which we do not propagate into our main galaxy sample.

The final data sample consists of 29,588 galaxies in total from redshift 0.03 < z < 0.09, which is publicly available at our catalog GitHub. The parameters which we will use to predict photometric properties are stellar mass (M<sub>\*</sub>), star formation rate (SFR), galaxy axis ratio as a proxy for inclination (b/a), bulge-to-total ratio (B/T), disk scale length ( $R_d$ ), and corrected bar vote fraction ( $p_{\text{bar}}$ ). Covariances amongst these parameters are minimal, as discussed further in 4.6.1; we show joint distributions for these parameters in Fig. 4.12.

## 4.2.3 Milky Way Properties

A number of the properties of the Milky Way used in this study have been derived using the Bayesian mixture model meta-analysis method first presented in Ref. [216]. This technique combines information from multiple measurements in order to obtain aggregate constraints on the Milky Way's properties, taking into account the possibility that individual measurements could be incorrect or have their errors miss-estimated. This Bayesian method is combined with Monte Carlo simulations in order to account for uncertainties on the Sun's measured Galactocentric radius,  $R_0$ ; the Galactic exponential disk scale length; and uncertainties in the local surface density of stellar mass. The inferred mass of the Milky Way's stellar disk depends upon all three of these parameters. Ref. [154] recently obtained a greatly improved geometric measurement of our distance from the center of the Milky Way,  $R_0 = 8.178 \pm 0.026$  kpc (Ref. [216] used the value  $8.3 \pm 0.35$ kpc from Ref. [143]). We have rerun the Bayesian mixture model inference from Ref. [219] and Ref. [217] with this updated measurement of the Galactocentric distance to obtain updated measurements of the Milky Way's mass, bulge-to-total mass ratio, and disk scale length; see those papers for all details of the data sets used and the calculations (the star formation rate estimate from Ref. [216] is not affected by the value of  $R_0$ , so we adopt it unchanged). The resulting Milky Way parameters used in this study are as follows:

- Bulge Mass  $M_*^B = 0.90 \pm 0.06 \times 10^{10} M_{\odot}$
- Disk Mass  $M^D_* = 4.58^{+1.18}_{-0.94} \times 10^{10} M_{\odot}$
- Total Stellar Mass  $M_* = 5.48^{+1.18}_{-0.94} \times 10^{10} M_{\odot}$
- Star Formation Rate SFR=  $1.65 \pm 0.19 \text{ M}_{\odot} \text{yr}^{-1}$  [216]
- Log Specific Star Formation Rate  $\log \frac{\text{SFR}}{M_*} = -10.52 \pm 0.10$
- Bulge-to-total Mass Ratio  $B/T = 0.16 \pm 0.03$
- Disk Scale Length  $R_d = 2.48^{+0.14}_{-0.15}$  kpc

The revised stellar mass estimate for the Milky Way is smaller than the previous estimate, but by an amount that is significantly smaller than the previously-estimated uncertainties.

By design these physical parameters are constructed such that they can be directly compared to the extragalactic catalogs used to predict photometric properties. For example, the stellar mass and star-formation rate estimates for the Milky Way are constructed assuming a Kroupa initial mass function [207] and an exponential disk model, which is the same way in which mass and star formation rates were calculated in the MPA-JHU catalog [46] that we use in this analysis. The only parameter where this is not fully the case is the bulge-to-total ratio, B/T. For the Milky Way we can securely estimate only a mass-weighted value for this quantity. In external galaxies, however, the mass-weighted B/T is much more difficult to obtain, and light-weighted measurements tend to be more reliable. However, our predicted Milky Way properties do not change significantly when we switch from r-band-based to gband-based B/T values, even though mass-to-light ratios differ significantly between these bands, suggesting that this is not a major issue. Our analysis also depends on two additional parameters that are determined independently of any Bayesian mixture model meta-analysis. These quantities have values assigned to them based on their meaning in their respective catalogs and our understanding of the Milky Way:

- Axis ratio (inclination proxy)  $b/a = 0.9 \pm 0.1$
- Bar vote fraction  $p_{\rm bar} = 0.45 \pm 0.15$

Galaxy inclination has a strong effect on color and luminosity measurements for disk galaxies. As mentioned in Section 4.2.2, dust alters the observed colors and magnitudes of star-forming galaxies that are highly inclined or edge-on. Our perspective within the Milky Way makes it somewhat equivalent to being edge-on to us (though our position within the Galaxy, rather than outside, does cause some differences). However, photometric properties are most cleanly determined for those objects which are observed face-on. Therefore, we predict the SED that would be observed for the Milky Way for axis ratio values drawn from a uniform distribution spanning from b/a = 0.8 to 1.0, consistent with the intrinsic axis ratios of spiral galaxy disks as described by Ref. [234] and in Section 4.2.2. Our results should therefore correspond to the properties of our Galaxy if it were observed face-on. While axis ratio is a good proxy for inclination, it is not a perfect substitute. For example, Ref. [355] found that disk galaxies that are rounder  $(b/a \sim 1)$  tend to be older and therefore intrinsically redder. This means that small biases could result from treating the Milky Way as having a face-on axis ratio. However, there is no straightforward way to avoid this, and the effect should be small compared to other sources of error.

The Milky Way exhibits clear evidence that it contains a bar (e.g., [32, 325]). However, very few Galaxy Zoo 2 galaxies have  $p_{\text{bar}} = 1.0$ , and those galaxies with the highest bar vote fractions are expected to have very strong bars, which may not match our Galaxy. Therefore we assume that in Galaxy Zoo 2 the Milky Way would have a vote fraction above the threshold for defining a bar, but not a value higher than the bulk of barred galaxies. Ref. [134] and Ref. [381] find that  $p_{\text{bar}} \geq 0.3$  serves as a reliable threshold between bar presence and lack thereof. Galaxies with  $0.3 \leq p_{\text{bar}} < 0.5$  likely have weaker bars while galaxies with  $p_{\text{bar}} > 0.5$  likely have stronger bars. Because the bar strength of the Milky Way's bar as it would be determined from outside our Galaxy is not well-constrained, we treat the bar vote fraction for the Milky Way as uniformly distributed between  $p_{\text{bar}} = 0.3$  and  $p_{\text{bar}} = 0.6$ . Choosing a larger mean vote fraction has small effect on our results, given the large range of fractions considered (compared to the distribution of  $p_{\text{bar}}$  in GZ2).

## 4.3 Gaussian Process Regression for Predicting Milky Way Photometry

In this subsection we describe how Gaussian process regression can be used to estimate photometric properties for the Milky Way. First we explain the need to transition from Milky Way analogue-based methods to GPR when we consider higher-dimensionality parameter spaces in Section 4.3.1. In Section 4.3.2 we explain the basic concepts behind GPR, and in Section 4.3.2.3 we highlight the fundamental differences between GPR and analogue galaxy methods. Section 4.3.2.1 describes the kernel used to set up our GPR, which guides how information is propagated from training objects to predictions. We briefly describe the computational limitations of the GPR implementation we are using in Section 4.3.2.2. Lastly, Section 4.3.2.4 investigates the contributions of various sources of uncertainty to our GPR predictions.

# 4.3.1 Limitations of Using Analogue Galaxies

Using Milky Way analogues to predict the photometric properties of the Milky Way, as was done in Ref. [219], has been a very useful methodology but also has limitations. Of particular concern is the dramatic reduction in MWA sample size that occurs as the number of parameters that must be matched increases; we would like to move from the two parameters considered by Ref. [219] (M<sub>\*</sub>and SFR), to a total of six, adding b/a, B/T,  $R_d$ , and  $p_{\rm bar}$ . Requiring that analogues be Milky Way-like in more ways should reduce the spread in photometric properties of the resulting sample, potentially enabling stronger constraints. There are only limited correlations between these six parameters (cf. Fig. 4.12 in 4.6.1), so degeneracies between them are minimal: they each add new information. For instance, galaxies with stellar mass and star formation rate matching the Milky Way exhibit bulge-tototal ratios ranging from zero to one: structural and star-formation history parameters carry distinct information. However, while matching on additional galaxy parameters produces a population of analogues that must each be closer in properties to the Milky Way, the resulting MWA sample becomes much smaller (with as few as ~ 5 analogues in the sample that are within  $3\sigma$  of the Milky Way in all of the properties considered, and none within  $2\sigma$ for every aspect).

The reduction in the size of analogue galaxy samples as more parameters are considered is illustrated in Fig. 4.1. We plot the total number of galaxies that are within a given number of  $\sigma$  for every Milky Way parameter considered (where  $\sigma$  represents the uncertainty in the Milky Way value for a given property) as a function of the number of  $\sigma$  used as a threshold. The lightest shade of blue denotes the number of analogues within a given threshold when only considering stellar mass (M<sub>\*</sub>) and axis ratio (b/a). The consecutive additions indicated in the legend represent the inclusion of the listed parameter in addition to all previous ones; i.e., we consecutively incorporate star formation rate, disk scale length, bulge-to-total ratio, and finally bar vote fraction. Hence, the darkest purple line shows the number of analogues when using all six parameters (which we have used in order of decreasing constraining power on Milky Way colors). The  $\sigma$  tolerances used for each parameter are the same Milky Way measurement errors defined in Section 4.2.3, *except* for  $p_{\text{bar}}$ . The significant uncertainty we fiducially ascribe to bar vote fraction would cause the 5 parameter and 6 parameter lines to be degenerate with one another; to avoid confusion we use  $\sigma_{p_{\text{bar}}} = 0.05$  instead of 0.15 when constructing this plot.

The previous work by Ref. [219] would most closely correspond to the +SFR (3 parameters), medium blue line. If one were to restrict to objects within  $\pm 2\sigma$  of the Milky Way value for all parameters employed, there would be a total of zero Milky Way analogue galaxies when using five or more parameters. In the six-dimensional space that we employ below, there are only 200 galaxies that are within even  $\pm 6\sigma$  of the Milky Way value in all six properties; at that extreme, analogue samples would be selecting objects that are not very close to the Milky Way at all. The lack of close analogues in high-dimensional parameter spaces makes constraints on Milky Way properties from the MWA method weak in that limit, with



Figure 4.1: Total number of galaxies within an allowed tolerance  $(N_{\text{gal}} < N_{\sigma})$  as a function of the maximum deviation away from the Galactic values allowed for each parameter, in units of the uncertainty in the Milky Way value of that parameter,  $N_{\sigma}$ . The lightest shade of blue denotes galaxy counts when analogues are selected using only stellar mass  $(M_*)$  and axis ratio (b/a). Consecutive parameter additions are cumulative. Thus the darkest purple line incorporates all six parameters of stellar mass, axis ratio, star formation rate (SFR), disk scale length  $(R_d)$ , bulge-to-total ratio (B/T), and bar vote fraction  $(p_{\text{bar}})$ . The  $\sigma$  values used for each parameters are defined in Section 4.2.3, except in the case of bar vote fraction, for which we use a smaller error ( $\sigma_{p_{\text{bar}}} = 0.05$  instead of 0.15) for illustrative purposes. Note that there are very few galaxies even within large tolerance windows when using several parameters, e.g., only ~ 10 MWAs within  $3\sigma$  for 5 parameters simultaneously, and no objects that are close to the Milky Way in all aspects, since the MWA method scales poorly with increased dimensionality.

correspondingly large uncertainties.

# 4.3.2 Gaussian Process Regression: A Powerful Method for Interpolation and Prediction

To address the lack of Milky Way analogue galaxies in our multi-dimensional parameter space, we have developed alternative methods for predicting Milky Way properties based upon Gaussian process regression. In this sub-section we summarise the basic properties of GPR relevant for this work. For in-depth discussion, we refer the reader to Ref. [302] and Ref. [151].

Gaussian process regression (sometimes called kriging) is effectively a method of interpolation where information from training data is accounted for by a smooth and continuous weighting function, called a "kernel" or covariance function. The joint probability distribution of the values of a Gaussian process at any finite set of points in parameter space will be a multi-variate Gaussian (with a number of dimensions set by the number of points in the set); the kernel specifies how the covariance between points depends upon their separation. The kernel should be a smooth and continuous function, with a length scale (which governs how far information propagates from a given point) that is optimized by training on the observed data. We can then *predict* what the value and the uncertainty of the desired quantity would be at any arbitrary point in space by applying this kernel to the training data. This is in contrast to other supervised learning algorithms which typically make single-valued, "point" predictions rather than predicting PDFs. It can be shown that GPR yields the minimum variance out of any unbiased interpolation method that depends only linearly on the training data; this makes GPR an **optimal interpolation algorithm**.

For our application of GPR, the galaxy sample described in Section 4.2 will serve as the training set. The six parameters we defined in Section 4.2.2 (stellar mass, star formation rate, etc.) serve as the features we will use for prediction. Our goal is to determine an optimised mapping from these physical parameters to a single output photometric parameter in our catalogue, e.g., the *r*-band absolute magnitude,  ${}^{0}M_{r}$ .

Once our training data is selected we then go into the model-selection phase of GPR, during which the mean function and covariance function (or kernel) used for GPR are selected and tuned. We detail our selection of the covariance function in Section 4.3.2.1. Effectively, the kernel determines how information from a given training point will be propagated to make predictions at other points in parameter space. Hyper-parameters describing the kernel are tuned at this step to maximise the log-marginal likelihood of the training data. After this step we consider the model to be "fit".

Finally we enter the inference phase of GPR. At any point in our six-dimensional parameter space, we can now determine the posterior probability distribution for the parameter of interest by applying the kernel to the training data. When evaluated at a single point in parameter space, a Gaussian process corresponds to a 1-D Gaussian; we thus obtain both a predicted mean for the property of interest (e.g.,  ${}^{0}M_{r}$ ) and the standard deviation of the Gaussian describing its uncertainty. In our example we would pass in a set of physical parameters measured for the Milky Way and obtain a predicted value for the  ${}^{0}M_{r}$  of the Milky Way, as well as the uncertainty in that value.

In reality we do not only query the GPR at the mean measured Milky Way properties presented in Section 4.2.3; rather, we perform random draws from the PDFs describing  $M_*$ , SFR, b/a, B/T,  $R_d$ , and  $p_{bar}$  in order to incorporate the uncertainties in the Milky Way's measured properties into our analysis. For log  $M_*$ , log SFR, B/T, and  $R_d$  we assume a normal distribution. For b/a and  $p_{bar}$  we draw from uniform distributions, as described in Section 4.2.3.

We perform these random draws 1000 times (so that we have 1000 full sets of Milky Way parameters). We then evaluate the GPR predictions for Milky Way photometric properties at each of these points in parameter space. This gives us a prediction and error estimate corresponding to each draw from the PDFs of Milky Way characteristics. Thus we end up with 1,000 total predictions for each Milky Way photometric property. Our mean prediction for the Milky Way in a given photometric band corresponds to the arithmetic mean of all these predictions.

In this work all Gaussian Process regression calculations have been done with the PYTHON scikit-learn Gaussian Process module, sklearn.gaussian\_process [286]. For details on the implementation of this module, refer to Ref. [286] and the scikit-learn documentation.

The following sub-sections provide more details on some aspects of our GPR methods

and their advantages.

# 4.3.2.1 Choice of Kernel for GPR

In this work we use a combination of two kernels for Gaussian Process Regression: a Radial Basis Function kernel and a white noise kernel. The Radial Basis Function (RBF) kernel decreases proportionally to  $\exp -\gamma D^2$ , where  $\gamma$  is a free parameter and D is the Euclidean distance between points; this kernel will cause the covariance between the predicted values from GPR at different points in parameter space to decrease as a Gaussian in distance as the separation between those points increases.

However, there is also scatter in galaxy photometric properties even for objects measured to have the same physical properties. In order to capture that, we also incorporate a white noise kernel, which models the spread in values for the predicted property at a fixed point in parameter space with normally distributed noise [302]. The net covariance used for the Gaussian process regression is then the sum of the distance-dependent covariance from the RBF kernel and the (diagonal) covariance matrix corresponding to the white noise kernel.

For a given set of training data, there is a nearly endless number of functions that can fit the given data points, each one a realization of the Gaussian process. The kernel creates a prior on the GP to constrain which functions from that set are most likely to describe the parameter space. The posterior is then determined using the training data values. Due to this probabilistic approach, the Gaussian process provides both predicted values and uncertainties at any points within the parameter space. Uncertainties due to the finite training sample size and its distribution in parameter space and those corresponding to intrinsic variation between training objects that have the same physical parameters are both captured. In regions of parameter space that are poorly constrained by the training data, the prediction uncertainties are correspondingly larger.

The kernels we use in this work are available in the sklearn.gaussian\_process.kernels base class. For the white noise kernel (WhiteKernel in scikit-learn) we initialize the noise level to be 1; similarly, we initialize the length scale for the RBF kernel to be 1 for each parameter. We opt to normalize the output photometric property to have mean zero and variance one across the training set, which helps to ensure that these initial guesses will have the right order of magnitude. The noise level and length scales are then optimized and the regression model is built via the sklearn.gaussian\_process.GaussianProcessRegressor class. In our model we allow the optimizer to restart 10 times in order to find the kernel parameters that maximize the likelihood without being trapped in a false maximum.

# 4.3.2.2 Optimizing Training Samples

The computation time and required memory for the scikit – learn implementation of GPR scales as the number of data points used to train the model squared and cubed, respectively. As a result, we find that the maximum training sample size we can use without running out of memory on the computers used for this work is ~ 6,000; it is infeasible to train from our entire catalog when using this GPR implementation.

We have therefore tested the effects of either restricting to objects with physical parameters within some tolerance of the MW fiducial values or randomly selecting a subset of objects in order to reduce the training set size. We have focused on the root mean squared error (RMSE) of predicted Milky Way photometry for the NUV, r, and J bands for this optimization. We use five-fold cross-validation for all the tests; i.e., we always train with 80% of the data and test with 20%, but rotate what objects are used for training and testing through the whole data set, and only retain the values for an object when it was in the test set. This provides unbiased estimates of the RMSE for a training set 80% as large as the one we actually have. We find that the combination that offered the lowest RMSE across all bands while keeping computational time manageable was to randomly select 2,000 galaxies out of the set of objects that are within  $12\sigma$  of the Milky Way for every parameters of interest. We therefore adopted this training strategy for all results below.

## 4.3.2.3 Comparison to Results from Analogue Samples

GPR can provide more accurate predictions than many other techniques thanks to its ability to leverage information from both nearby objects in the training set as well as from more distant objects that characterize larger-scale trends. In our application, this allows the GPR to map from the Milky Way's physical properties to its photometric properties much more accurately than if we had only used the few objects that are similar to the Milky Way in all respects (i.e., those which would be classified as MWAs) to inform the mapping.

Fig. 4.2 illustrates the fundamental difference between how properties are constrained by the Milky Way analogue selection method versus Gaussian process regression. For simplicity's sake we perform this comparison based on only three parameters (stellar mass, star formation rate, and axis ratio, the same ones utilized in Ref. [219]), as the analogue method starts to break down when more parameters are included. We also do not correct for Eddington bias in either measurement (q.v. 4.9) for simplicity. Objects included in a set of MWAs based on 5,000 samples from the distribution of possible Milky Way properties via methods equivalent to those from Ref. [219] are depicted by the orange points. The analogues fall within a narrow range of stellar mass, limiting the set of objects that contribute information. The orange star represents the mean prediction for the Milky Way's  $^{0}(g-r)$ color ( $^{0}(g-r) = 0.682$ ) resulting from this set (derived via the Hodges–Lehmann robust estimator, Ref. [168]). The orange ellipse depicts the  $1\sigma$  confidence region for  ${}^{0}(g-r)$  color and stellar mass. In contrast to the Milky Way analogues, the sample of galaxies used to train the GPR are shown by purple points. These cover a much broader range of parameter space than the MWAs. Similarly, the prediction for the Milky Way's  $^{0}(g-r)$  color using GPR is shown by a purple star (with  ${}^{0}(g-r) = 0.668$ ), along with the corresponding  $1\sigma$  confidence region which is shaded in purple. When we have many analogues, both techniques yield very similar results, but unlike MWAs the GPR technique still provides strong constraints when we consider many parameters at once.

## 4.3.2.4 Characterizing Sources of Uncertainty

We can quantify the contributions of different sources of uncertainty to our GPR estimates by changing how we perform the regression. We illustrate our methods by evaluating how the prediction from a six-parameter Gaussian process regression fit changes as a single parameter varies. The left panel of Fig. 4.3 shows one of the physical parameters being regressed from, star formation rate, on the x axis and the target value,  ${}^{0}(g-r)$  color, on the



Figure 4.2: A comparison of galaxy samples and results from the Milky Way analogues method versus Gaussian Process regression. For this analysis analogues were selected based upon stellar mass, star formation rate, and axis ratio, and we use the same parameters to predict color via GPR. Orange points represent Milky Way analogues selected by methods equivalent to Ref. [219], while purple points indicate the galaxies used to train the Gaussian process regression. The predicted mean Milky Way  ${}^{0}(g-r)$  color is denoted by a star in the corresponding color, and ellipses depict a  $1\sigma$  joint confidence region for color and stellar mass. Both methods yield a similar predicted  ${}^{0}(g-r)$  color for the Milky Way when we have large numbers of analogues. In contrast to the small window that the Milky Way analogues lie in, the GPR utilizes a wider variety of galaxies to capture larger-scale trends.

y axis. We chose this pair as galaxy star formation rate is expected to correlate with galaxy color well at fixed stellar mass.

First we isolate the scatter in color at fixed properties; this corresponds to the contribution of the white noise kernel to the covariance function of the GPR. To determine the magnitude of this scatter we query the GPR at the Milky Way's fiducial physical properties to obtain a predicted PDF of  $^{0}(g - r)$  at this point in parameter space from which we can draw samples. The standard deviation of the color of these samples corresponds to the scatter encoded in the white noise kernel.

In the panel at the right of Fig. 4.3 we plot a histogram of 10,000 possible  ${}^{0}(g-r)$  colors drawn from the GP's predicted PDF, evaluating it at the fiducial value of the Milky Way's SFR. In the left panel the lavender star denotes the mean predicted  ${}^{0}(g-r)$  for the Milky Way with this model, and the error bar corresponds to the standard deviation of the sample values. This error bar therefore corresponds to the  $1\sigma$  scatter in color at fixed properties. The gray shaded area corresponds to the  $\pm 1\sigma$  band for the GPR prediction of color for a range of log (SFR) values. By construction the half-width of the band must match the standard deviation of samples from the PDF at fixed properties.

To instead isolate the contribution to errors resulting from the uncertainty in Milky Way properties, we determine the distribution of mean predicted colors evaluated at varying values of SFR drawn from the PDF for the Milky Way. We perform 1000 draws from the fiducial MW log SFR PDF in total and evaluate the GPR mean predicted  ${}^{0}(g-r)$  for each. In Fig. 4.3 the dark purple points in the left panel shows the resulting predictions, which all fall on a continuous curve by construction. The panel to the right shows the histogram of this set of predictions in purple. We quantify the scatter in color attributable to uncertainties in the Milky Way's measurements via the standard deviation of the  ${}^{0}(g-r)$  values at these 1,000 points.

To illustrate the full range of values obtained via GPR we plot ten samples from the distribution of predictions for each of the thousand Milky Way SFR draws as faint blue points in Fig. 4.3. These samples vary in color due to *both* the scatter in color at fixed properties and due to the uncertainty in Milky Way properties.

We can extend these same ideas in 1-D to evaluate the relative contributions of uncer-

tainties in Milky Way properties and of the scatter in properties at fixed color to the error in Milky Way  ${}^{0}(g-r)$  color, for any GPR model of interest. The key difference from Fig. 4.3 is that, in order to quantify the full scatter due to uncertainties in Milky Way properties, we allow all the parameters to vary, not only SFR. We present the results of this analysis for GPR models based on two to six physical parameters in Fig. 4.4. We use a stacked bar plot to display the contribution to the variance from each error source, where the x-axis is labeled according to the physical parameters used to train the GPR (where additions are all cumulative, so entries further to the right incorporate more parameters). Since independent errors will add in quadrature, the contribution to the net variance from each factor is proportional to the height of its bar. The variance due to the scatter at fixed properties is shown in a lighter violet shade, while that resulting from the uncertainty in Milky Way properties is shown in dark purple. The scatter at fixed properties to the majority of the error in the Gaussian process regression for  ${}^{0}(g-r)$  color.

We have also evaluated the contribution to uncertainties resulting from the finite size of the training sets used. This scatter is isolated by varying the randomly-selected training sample 100 times. For each training sample we evaluate the mean predicted value from GPR at the Milky Way's fiducial physical parameters. The contribution to uncertainties from the finite size of training samples is obtained by calculating the variance of the GPR predictions across all of the training sets. These values are minute: the variances are an order of magnitude smaller than error attributed to the scatter due to Milky Way measurements. Thus any contribution to errors resulting from the finite training sample size is negligible.

We have performed the same error budget test for colors in the UV, near-IR, and mid-IR. The results mirror those presented in Fig. 4.4: the errors are dominated by scatter at fixed properties, followed by scatter from the Milky Way measurements. In all cases errors attributable to finite training set size are negligible compared to other sources. While the cumulative variance decreases for every parameter added when we predict colors, this is not the case for absolute magnitudes. In that case, the cumulative variance decreases as we add parameters until the sixth parameter,  $p_{\rm bar}$ , is incorporated. At that point, the variance increases and the scatter due to finite training becomes more important. For this reason all absolute magnitude predictions within this chapter are performed using only 5 parameters,

excluding  $p_{\text{bar}}$ .

While uncertainties in Milky Way characteristics will contribute to the random errors in the derived photometric properties of our Galaxy, uncertainties in the physical parameters of the *training* galaxies can cause *systematic* errors. If the density of objects in parameter space varies quickly (with non-negligible second or higher derivatives), objects will more often scatter from well-populated regions of parameter space into sparser regions than vice versa. The resulting systematic shift in the measured distribution of parameters compared to the underlying distribution with no scatter is known as Eddington bias.

In the context of this work, Eddington bias will lead to shifts in the color and luminosity predicted for the Milky Way. We derive corrections for Eddington bias using methods similar to those of Ref. [219]; we detail our procedures in 4.9. The estimated Eddington bias is subtracted off from the GPR-predicted colors and luminosities for the Milky Way to produce our final estimates for the Galaxy's photometric properties and likewise the uncertainty on the Eddington bias calculations is propagated into our final error estimates. In general Eddington bias has small but nonzero effects on our results ( $< 1\sigma$  for almost all parameters, as listed in 4.6).

In the following sections, the errors on GPR results presented include the contributions from scatter at fixed properties, uncertainties in Milky Way properties, and uncertainty from Eddington bias.

# 4.3.2.5 Summary of the GPR Algorithm for Determining Milky Way Photometric Properties

Here we summarize the steps taken to predict Milky Way photometric properties via GPR. Our method proceeds as follows:

- 1. Construct the training sample by restricting to objects within  $12\sigma$  of the Milky Way in all physical parameters considered and then randomly down-sampling to 2,000 objects.
- 2. Adopt the combination of a Radial Basis Function (RBF) kernel and a white noise kernel as the covariance function to be used for GPR.
- 3. Train the GPR using a single photometric property (normalized to have mean zero and



Figure 4.3: A breakdown of uncertainty due to scatter at fixed properties and scatter due to Milky Way measurement uncertainty. We plot results from a six parameter Gaussian process regression trained to predict  ${}^{0}(q-r)$  color. For this example we vary only one of the input parameters, the star formation rate (SFR). The light violet color corresponds to results when we use a fixed training set and evaluate the GPR at fixed MW properties; this isolates the variation attributable to the white noise kernel, corresponding to the scatter in color at fixed galaxy properties. The dark purple color corresponds to results when the training set remains fixed but the SFR value is drawn randomly from the fiducial MW PDF, so that the contribution to the color error attributable to the uncertainty in the star formation rate of the Milky Way can be evaluated (the other five parameters used to train the GPR are held constant for simplicity for this example). In the left panel the shaded region depicts the  $\pm 1\sigma$  range predicted by the GPR fit, plotted out to  $\pm 3\sigma$  of the Milky Way's SFR. The predicted  ${}^{0}(g-r)$  color and  $1\sigma$  error for the Milky Way's fiducial SFR (and hence only including error at fixed properties) is shown by the star-shaped point and error bar in light purple. The predicted  ${}^{0}(g-r)$  color from each of 5,000 draws from the SFR PDF is shown by the purple points, which accounts for scatter due to the SFR measurement. These predictions perfectly trace the GPR fitand tend to fall within  $\sim 3\sigma$  of the Milky Way's SFR, by construction. For reference we also show 10 samples from the GPR-predicted PDF for each of the 5,000 random SFR values as faint blue points. The distribution of these points reflects both the scatter of galaxy colors at fixed properties and the uncertainty in the MW measurements. Histograms of the colors for each sample, whose distributions correspond to the scatter at fixed properties and the scatter due to Milky Way SFR measurement uncertainties, are shown in the right panel. It is evident that the spread in GPR predictions at fixed properties is much larger than the scatter that results from uncertainties in the Milky Way's SFR.



Figure 4.4: Contributions to the variance in rest-frame  ${}^{0}(g-r)$  color for Gaussian process regression employing varying sets of galaxy physical parameters. The x-axis shows the set of parameters used to predict color; they increase cumulatively as we go from left to right, in order from the most constraining to the least constraining parameter. The variance decreases monotonically as the number of parameters increases. In every case the scatter in color at fixed properties dominates errors; uncertainties in the physical properties of the Milky Way are sub-dominant. Contributions to uncertainties due to having a finite training set are small enough to be considered negligible, enough so that they would not be visible on this plot if we included them.

variance one) as the output or "y" value and the physical galaxy parameters as the "x" values. This training will tune the hyperparameters of the kernel.

- Perform 1,000 random draws from the PDFs that describe the fiducial Milky Way's properties. This will allow us to incorporate uncertainties in the Milky Way measurements into our results.
- 5. Use GPR to apply the optimized kernel to the training set and predict the photometric property of interest. For each randomly-drawn set of physical properties for the Milky Way, we obtain the mean prediction, predicted variance, and a set of 1,000 values drawn from the GPR-predicted PDF corresponding to that position in physical parameter space (which we refer to as a set of samples).
- 6. The mean photometric prediction for the Milky Way is then calculated as the mean of the set of GPR output means at the position of each MW draw. The error on the prediction is calculated as the standard deviation of the values from the complete set of samples generated, allowing us to incorporate both uncertainties associated with the scatter at fixed properties and errors resulting from the uncertainties in MW properties.

The code used to construct the GPR is provided on our GP GitHub page for public use here<sup>2</sup>. At this site we provide sample code for determining photometry estimates, addressing systematics, and constructing an SED.

# 4.4 Results

Via GPR predictions for Milky Way photometric properties across the spectrum, we can produce a comprehensive outside-in portrait of the Milky Way SED, allowing comparisons to the colors and luminosities of other galaxies. In this section we apply a variety of diagnostics from the literature, such as color-luminosity, color-mass, and color-color diagrams, in order to assess how the Milky Way compares to the broader population. We also construct a multi-wavelength SED for the Milky Way and compare our results to templates from the literature.

<sup>&</sup>lt;sup>2</sup>https://github.com/cfielder/GPR-for-Photometry

## 4.4.1 The Milky Way Compared to the Broader Galaxy Population

As discussed in Section 4.3, we have predicted the Milky Way colors and luminosities based upon the six parameters of stellar mass (M<sub>\*</sub>), star formation rate (SFR), axis ratio (b/a), bulge-to-total ratio (B/T), disk scale length  $(R_d)$ , and bar vote fraction  $(p_{\text{bar}})$ . In the following color diagrams, all magnitudes and colors are presented as rest-frame AB magnitudes (evaluating all passbands at redshift zero).

Our quantitative results are summarized in Table 4.2-Table 4.4. The values provided correspond to the mean rest-frame predictions based upon the Gaussian process regression derived via the methods presented in Section 4.3, and have been corrected for Eddington bias as described in 4.9. colors and magnitudes are all calculated independently of one another. For example, we use GPR to predict  ${}^{0}(g - r)$  galaxy color directly, as opposed to deriving this value by subtracting the predicted  ${}^{0}M_{r}$  from the predicted  ${}^{0}M_{g}$ . For SDSS photometry, our derived colors are based upon model magnitudes, as these yield the most accurate color estimates for SDSS galaxies; however, the absolute magnitudes provided are based upon **cmodel** magnitudes, as those most accurately represent the total brightness of an object.

Log-spaced density contours corresponding to the cross-matched galaxy sample of 29,836 galaxies described in Section 4.2.1 and Section 4.2.2 are plotted in gray-scale on all of the following color-based diagrams. We also overlay red and blue ellipses which denote the rough locus of the red sequence and blue cloud, respectively, in each plot. These shadings are intended to guide the eye and should not be interpreted in a quantitative manner. In a corner of each plot we provide error bars that correspond to the mean uncertainties in each galaxy property being plotted for the training set.

In each diagram we also show the locations of the 36 red spiral galaxies selected in Ref. [244] that overlap with our cross-matched sample (out of 294 in the original catalog). This sample of objects was selected based upon their color, presence of spiral features, and shape/structural parameters from SDSS. They are required to have color  ${}^{0}(g - r) >$  $0.63 - 0.02({}^{0}M_{r} + 20)$ , overlapping the blue edge of the red sequence. They are also selected to have a spiral likelihood  $p_{\rm spiral} \geq 0.8$  in the prescription of Ref. [12], and are required to have visible arms in Galaxy Zoo 1  $p_{\rm CW} > 0.8$  or  $p_{\rm ACW} > 0.8$  [221] in order to ensure they have spiral morphology. These objects are also selected to be approximately face-on (equivalent to an axis ratio requirement b/a > 0.63), as dust reddening is expected to have a substantial impact on the apparent colors of spirals [245]. However, in that paper the axis ratio values were calculated via r-band isophotal measurements, while ours are determined from an exponential profile fit. Therefore we apply a profile-fit-based cut of b/a > 0.6 to this sample to enable a more direct comparison to our face-on results for the Milky Way. Finally, Ref. [244] requires that the red spiral sample contain galaxies with an SDSS  $f_{deV} \leq 0.5$ , where  $f_{deV}$  is defined as the weight of the de Vaucouleurs profile in the best-fit linear combination with the exponential profile matched to the object's image. This ensures that S0 galaxies do not contaminate the sample, although they are already only a small percentage of the GZ1 sample.

The resulting red spiral sample is represented by red points in our plot. We overlay the positions of these objects in each parameter space to help assess the consistency of the inferred properties of the Milky Way with this population. Two objects whose  ${}^{0}(g-r)$  colors in the cross-matched catalog differed by > 0.1 mag from the photometry used in Ref. [244] due to changes in SDSS pipelines were excluded.

#### 4.4.1.1 Optical colors

We first present results at optical wavelengths, as they allow us to compare directly to previous work done with Milky Way analogues in Ref. [219]. We focus on the SDSS *ugriz* bands (cf. Section 4.2.1.1). Fig. 4.5 presents predictions for Milky Way optical optical colors as a function of stellar mass  $(M_*)$  in solar mass units.

The upper panel shows  ${}^{0}(u-r)$  color and the lower panel shows  ${}^{0}(g-r)$  color versus mass. Both panels have overlaid dashed reference lines which can be used to distinguish general regions of the diagrams. The top portion contains galaxies that are on the red sequence, while the middle portion contains green valley galaxies, and the lower portion corresponds to the blue cloud. The dashed lines bracketing the green valley in the upper panel correspond to  ${}^{0}(u-r) = -0.24 + 0.25 M_{*}$  and  ${}^{0}(u-r) = -0.75 + 0.25 M_{*}$  [319]. In the lower panel the plotted lines correspond to  ${}^{0}(g-r) = 0.6 + 0.06(M_{*} - 10)$  and  ${}^{0}(g-r) = (0.6 + 0.06(M_{*} - 10)) + 0.1$  [251].

Previous results from Ref. [219] are plotted in orange. The star represents the mean prediction and the ellipse encompasses the  $1\sigma$  confidence region. We remind the reader that these constraints were determined based only on stellar mass and star formation rate, along with a cut on axis ratio. In comparison, the results of the 6-parameter Gaussian process regression are plotted in purple. In red we plot the GPR result evaluated with an axis ratio of b/a = 0.3 rather than 0.9, to illustrate the impact that the assumed inclination has on the inferred SED. The GPR confidence regions are calculated from the covariance between the samples drawn from the regression predictions; the distribution of these samples incorporates both uncertainties in Milky Way properties and scatter in colors at fixed properties (cf. Section 4.3.2.4). In the lower right corner of each panel we show the mean error in optical color and log stellar mass amongst the galaxies in our final sample. Per-object uncertainties in the optical colors account for roughly half of the total scatter in our GPR color prediction for a face-on Milky Way. In contrast, the average error in stellar mass in the training sample does not affect the uncertainty in the stellar mass of the Milky Way (it will, however, contribute to Eddington biases, as discussed in Section 4.8. Note that Ref. [219] did not use the same stellar mass as we do, reflecting our updated estimate for the mass of the Milky Way (cf. Section 4.2.3).

In both  ${}^{0}(u-r)$  and  ${}^{0}(g-r)$  our results are consistent with, though marginally redder than, those reported in Ref. [219]. This is no surprise as we do not expect the Milky Way to move far in optical color space when constraints tighten. Even when we make predictions for a much more inclined Milky Way our results do not change dramatically in the optical, with shifts well within the uncertainties in both our face-on results and those from Ref. [219], although the color does become marginally redder as expected. In the optical the Milky Way appears to lie in the "saddle" of the the galaxy-color bimodality, implying that the Milky Way is redder than the average spiral galaxy in the local Universe in the optical bands. That said, if one were to only consider spiral galaxies of similar mass to our Galaxy, the Milky Way is not as unusually red as it would be if compared to lower-mass spiral galaxies.

The green valley (and by extension the galaxy color-bimodality) has been used as a basic tool to distinguish transitional galaxies from the general galaxy population. Transitional



Figure 4.5: Rest-frame optical color as a function of stellar mass. The gray-scale, log-spaced contours depict the density of galaxies in our cross-matched sample. The dashed gray lines correspond to divisions of the galaxy population used in the literature. We expect "red sequence" galaxies to be in the upper portion of each plot, with the "blue cloud" corresponding to the bluest colors. The region between the two lines corresponds to the "green valley" population. For  ${}^{0}(u - r)$  we use the divisions of Ref. [319] and in  ${}^{0}(g - r)$  we follow Ref. [251] (cf. Section 4.4.1.1). The results and  $1\sigma$  confidence region from Ref. [219] are marked in orange. Our results from applying GPR to all six galaxy physical parameters considered are marked in purple; the  $1\sigma$ region is determined by the covariance between Gaussian process samples. For comparison, in lighter red we show the six-parameter GPR result obtained when setting the axis ratio of the Milky Way to  $b/a = 0.3 \pm 0.1$ rather than 0.9. The stellar masses differ between our prediction and Ref. [219] due to the updates to the mass estimate for the Milky Way described in Section 4.2.3. Red points correspond to members of the red spiral galaxy sample of Ref. [244]. In the lower right corner of each panel we depict error bars representative of the mean uncertainties for galaxies in the comparison sample. Our results are consistent with Ref. [219] and indicate that at optical wavelengths the Milky Way is redder than the typical star-forming spiral galaxy, in addition to being more massive.

galaxies have lower specific star formations rates (sSFR = SFR/M<sub>\*</sub>) than a star-forming galaxy of the same mass; specific star formation rate can be used as a proxy for the evolutionary state of a galaxy and its star forming history. Ref. [310] defines the transitional region in sSFR space to be below the sSFR of massive Sbc galaxies, as these Sbc's are the earliest galaxy type expected to proceed with regular star formation free of quenching, but above the sSFR at which galaxies appear to no longer be star forming in the UV. As described in that work, this range corresponds to  $-11.8 < \log (SFR/M_*) < -10.8$ . Note that here and throughout this chapter log refers to the base 10 logarithm.

In Fig. 4.6 we plot the same rest-frame colors as in Fig. 4.5 but as a function of log specific star formation rate. The vertical dashed lines denote the transitional region in log SFR/M<sub>\*</sub>, as defined by Ref. [310]. The region with log sSFR above -10.8 corresponds to galaxies that are actively forming stars while objects with log specific star formation rate below -11.8 are quiescent; transitional objects are between them. In the lower panel the green horizontal lines correspond to the green valley definition of Ref. [210] evaluated with the predicted r band absolute magnitude for the Milky Way ( ${}^{0}M_{r} = -20.65$ ). Galaxies residing within this range in  ${}^{0}(g - r)$  are expected to reside within the green valley. Galaxies above this designation are expected to lie in or near the red sequence, and galaxies below are expected to lie in or near the blue cloud.

Based on its specific star formation rate, the Milky Way must lie within the star-forming population, rather than in the transitional range. While one might expect an object that meets optical definitions of the green valley to have a transitional sSFR this is not necessarily true, as galaxies of different evolutionary states can share the same optical color (see e.g., [81, 310]).

In  ${}^{0}(g - r)$  if we take the green valley to be 0.1 in width, as defined by Ref. [252], the Galaxy would either fall within or be redder than the green valley, consistent with the results shown in Fig. 4.5. Despite ongoing star formation, more massive spiral galaxies tend to be redder in the optical than their lower-mass counterparts [244]. However, our estimated properties for the Milky Way lie in the middle of the distributions of colors, masses, and sSFRs of the red spiral sample from Ref. [244]; these objects are redder in the optical than is typical for even the most massive spirals.



Figure 4.6: Rest-frame optical color as a function of log specific star formation rate (log (SFR/M<sub>\*</sub>)) in units of yr<sup>-1</sup>. The vertical dashed gray lines designate divisions of galaxy populations according to their specific star formation rates, following the definitions of Ref. [310]: quiescent galaxies are at left, transitional objects are in the middle, and star-forming objects are at right. The Milky Way lies on the star-forming side of these divisions. The green horizontal lines in the bottom panel correspond to the green valley definition of Ref. [210], evaluated at the *r*-band absolute magnitude of the Milky Way. According to the prescription by Ref. [252] the green valley has a width of 0.1 in  $^{0}(g - r)$ , leading to the limits shown here here. The Milky Way mean value falls above this "green valley" region. While the optical color of the Milky Way is redder than most star-forming galaxies in the local Universe, based on its specific star formation rate the Milky Way would not be considered a transitional galaxy.

As these plots exemplify, differentiating between galaxy populations based only on optical photometry is challenging. For instance, in the lower panel in Fig. 4.6 we can see that red sequence, transitional, and star-forming objects can all have colors of  ${}^{0}(g - r) \sim 0.7$ . In Fig. 4.5 the blue cloud becomes difficult to distinguish from the red sequence at high masses as the most massive spirals have lower sSFRs and, therefore, redder colors. In the following subsections we investigate constraints on the color of the Milky Way at UV and IR wavelengths where galaxy populations may separate more clearly.

# 4.4.1.2 UV colors

We utilize far-ultraviolet (FUV) and near-ultraviolet (NUV) photometry from GALEX provided in the GSWLC-M2 catalogue [241, 312, 311], as discussed in Section 4.2.1.2. Thermal emission from massive stars with lifetimes < 100 Myr peaks at ultraviolet wavelengths, while lower-mass, longer-lived stars play a larger role at optical wavelengths [310, 350]. Because UV radiation is produced by short-lived but high-luminosity stars it provides a sensitive indicator of *recent* star formation. As a result, UV photometry can more clearly differentiate star-forming from quiescent galaxies than optical measurements can.

Much as in Section 4.4.1.1, we can use our GPR results to place the Milky Way on UVbased diagnostic diagrams from the literature. In Fig. 4.7 we plot  ${}^{0}(FUV-r)$  and  ${}^{0}(NUV-r)$ UV-optical colors versus specific star formation rate. The contours and vertical reference lines shown are defined in the same way as in Fig. 4.6. For  ${}^{0}(NUV - r)$  we show horizontal lines corresponding to the "green valley" definition of Ref. [310], bounded at  $4 < {}^{0}(NUV - r) < 5$ . As before, our face-on MW prediction and the corresponding  $1\sigma$  confidence region are plotted in purple and the inclined MW prediction is plotted in red; unlike in the optical, there are no previous estimates of Milky Way properties in this space that we could plot. Much as in the optical, uncertainties in the UV photometry for individual objects account for roughly half of the total scatter ascribed to our Milky Way UV predictions.

Compared to typical star-forming galaxies in the local Universe, the Milky Way has redder than average UV colors and lower than average sSFR. The Milky Way appears to lie on the blue side of the  ${}^{0}(NUV - r)$  green valley border, in contrast to its location in the



Figure 4.7: As Fig. 4.6, but for UV-optical colors. The vertical reference lines come from Ref. [310] where objects with a log sSFR > -10.8 are actively forming stars, objects with log sSFR < -11.8 are quiescent, and objects in between are considered transitional. In  $^{0}(NUV - r)$  we also plot the bounds of the UV-optical "green valley" as defined in Ref. [310]. Our predictions for the Milky Way show that it lies on the star forming side of the transitional regions, and is most likely on the blue side of the UV-optical green valley border. Our Galaxy lies in a similar region of these color spaces to the red spiral sample of Ref. [244].
optical (cf. Fig. 4.5 and Fig. 4.6). This reflects the limited discriminating power of optical color; the green valley is only 0.1 mag wide in  ${}^{0}(g-r)$ , allowing objects to easily scatter over its borders due to even small photometric errors or inclination effects, but it spans an entire magnitude in  ${}^{0}(NUV - r)$ . A more inclined Milky Way is predicted to be notably more red in the UV than in the optical, so much so that it could be consistent with the UV green valley in color.

As in the optical, our estimates for the Milky Way in the UV-sSFR plane lie in the middle of the Ref. [244] red spiral population. Red spiral galaxies tend to lie outside of the UV green valley as they have star formation rates comparable to typical blue spirals of the same mass [81], and UV color is more sensitive to recent star formation rate than the optical is.

## 4.4.1.3 Infrared/WISE colors

The infrared data for our galaxy sample originates from the 2MASS and WISE surveys, as included in the GSWLC-M2 [312, 311] and DESI Legacy catalogs [96], respectively (cf. Section 4.2.1). Similar to in the ultraviolet, the infrared brightness of a galaxy is sensitive to recent star formation due to re-emission of UV photons absorbed by dust. The infrared colors of galaxies also exhibit a color bi-modality, but star-forming galaxies exhibit *redder* IR colors than the passively-evolving population, rather than bluer. Instead of the "green valley," the region between the star-forming and quiescent populations in the IR is commonly referred to as the infrared transition zone (IRTZ), following Ref. [6].

In Fig. 4.8a and Fig. 4.8b we plot WISE color-color diagrams for the cross-matched galaxy sample in addition to the GPR prediction for the Milky Way. The Milky Way color is poorly constrained in some WISE bands due to the lower signal-to-noise of these detections. If we compare our covariance ellipse in the  ${}^{0}(W1-W2)$  direction to the average errors in the photometry (lower right error bar) the photometric errors account for a modest fraction of the total uncertainties in our GPR prediction. However, in  ${}^{0}(W2-W3)$  and  ${}^{0}(W3-W4)$  errors in the photometry for individual objects dominate the estimated uncertainties in the Milky Way GPR predictions. The vertical lines in these plots designate the IRTZ from Ref. [6];

objects with  ${}^{0}(W2-W3) > 0.565$  in AB magnitudes correspond to late-type galaxies, those with  ${}^{0}(W2-W3) < -1.035$  are early-type galaxies, and in between lies the IRTZ. Magnitudes were converted from Vega to AB magnitudes via the prescription of Ref. [176].

As before, we show predictions for the Milky Way if it were approximately face-on or more steeply inclined by evaluating the GPR at different axis ratios. An inclined Milky Way appears the be most notably different in the W3 band, which traces prominent dust emission features, particularly those associated with polycyclic aromatic hydrocarbons [383]. It appears that an inclined spiral galaxy would be measured to be more IR bright compared to a face-on counterpart matching it in all other ways; this may represent a systematic effect related to data processing, since the dust emission would be expected to be optically thin. We find similar results in Section 4.4.2.3.

In both diagrams the prediction for the Milky Way lies on the star-forming side of the infrared transition zone. If we compare Fig. 4.8a to the classification scheme in Figure 12 of Ref. [383] (note that this requires converting our AB magnitudes to Vega magnitudes), the Milky Way lies within the region of color space they label as typical for spiral galaxies, as would be expected. Similarly, the classification scheme of Figure 11.b of Ref. [175] would place the Milky Way in the intermediate disk region, consistent with the expectation for a massive spiral galaxy. Intermediate disk objects are thought to be in transition towards being quenched due to star formation rates that are decreasing with time; our estimate for the Milky Way's star formation rate is slightly below average for a spiral galaxy of the same mass, consistent with this picture.

The infrared results mirror what we find from the UV: the Milky Way is still forming enough stars to appear bright in the IR due to re-emission from dust. Galaxies are expected to transition in the optical before they do in the infrared [6, 350]. If we follow the narrative of Ref. [6] and Ref. [331], the Milky Way may be in the early transition phase from star-forming to quiescent. It is brighter and redder in the optical compared to the typical star-forming galaxy. In the UV the Milky Way is on the blue side of the green valley but near it. In the IR the Milky Way's inferred color is more typical for a star-forming galaxy, though uncertainties are substantial. This would track with the expectation that a galaxy transitions in the optical before it does in the IR.



Figure 4.8: WISE color-color diagrams for both our parent sample (log density contours) and the predicted results for the Milky Way from Gaussian process regression. Reference lines from Ref. [6] designate the infrared transition zone  $(1.035 <^0 (W2 - W3) < 0.565;$  IRTZ). The Milky Way appears to lie on the star forming side of the IRTZ, much closer to the median colors of typical spiral galaxies, in contrast to the UV and IR. Again our Galaxy lies in a similar region of these color spaces to the red spiral sample of Ref. [244].

As before, the inferred colors of the Milky Way in the mid-IR are consistent with the range of values for the Ref. [244] red spiral sample, though some WISE bands are not well constraining due to the low signal-to-noise of the underlying measurements used for prediction. Based on this, it remains plausible that the Milky Way is a part of the red spiral population. We discuss how the Milky Way's colors compare to other galaxy populations further in Section 4.5.2.

#### 4.4.2 The Multiwavelength Spectral Energy Distribution of the Milky Way

Thus far, we have focused on predictions for a single Milky Way color at a time. However, we can assemble color information across all passbands to construct a spectral energy distribution (SED) for the Galaxy. SEDs, which quantify the the total energy of emitted photons as a function of wavelength or frequency, are valuable tools in the study of galaxies. Many physical characteristics of galaxies can alter their SEDs - the age of their stellar population, stellar abundances, gas and dust content, inter-stellar medium (ISM) chemistry, details of star formation history, and the presence of an AGN can all leave distinct signposts that give observers insight into the formation and evolution of a given galaxy (see e.g., [326]). Because these effects each tend to alter the SED at specific portions of the spectral range, with broad enough wavelength coverage and detailed enough spectral information one can disentangle the dominant processes in a given galaxy.

Detailed modeling of the Milky Way's SED will be the focus of follow-up work. In this work, we will present a proof-of-concept for a GPR-constructed SED for the Milky Way and provide an initial analysis of its properties. In the following sub-section we outline our GPR-based methods for determining the SED of the Galaxy before presenting quantitative results and assessing the effects the galaxy physical parameters used for prediction each has on the SED.

## 4.4.2.1 Algorithm for Calculating the SED for the Milky Way

We work in frequency ( $\nu$ ) space instead of wavelength ( $\lambda$ ) space when calculating the SED of the Milky Way, as spectral energy distributions are most typically presented in units

of energy per unit frequency. Our algorithm for calculating the SED proceeds as follows:

# 1. Estimate ${}^{0}M_{r}$ and colors for the Milky Way –

Our SED is calculated in reference to the *r*-band. Therefore using the GPR (described in Section 4.3) we predict the *r*-band AB absolute magnitude  $({}^{0}M_{r})$  and all colors with restframe *r* as the reference band; i.e.,  ${}^{0}(x - r)$  where *x* spans from FUV to W4 (e.g.,  ${}^{0}(FUV - r), ..., {}^{0}(W4 - r))$ .

Eddington bias is subtracted off separately from our predicted colors and  ${}^{0}M_{r}$  before we combine them. Similarly, the uncertainty in the Eddington bias is added in quadrature to the uncertainty in the GP calculations (which incorporates both scatter at fixed properties and errors due to the Milky Way property uncertainties; cf. Section 4.3.2.4). For details on the Eddington bias calculations refer to 4.9.

## 2. Calculate flux ratios –

We calculate the flux in each band  $f_{\nu,x}$  relative to the flux in the *r*-band  $f_{\nu,r}$  via the relation

$$\log\left(\frac{f_{\nu,x}}{f_{\nu,r}}\right) = \frac{{}^{0}(x-r)}{-2.5},\tag{30}$$

where  ${}^{0}(x-r)$  is the Eddington bias-corrected color in the x band compared to the r band, which we have predicted via GPR.

## 3. Calculate luminosity –

From the r-band absolute magnitude combined with the flux ratios it is straightforward to convert to luminosity. We first calculate the r band luminosity as

$$\log(L_{\nu,r}) = \frac{({}^{0}M_{r} - 34.04)}{-2.5} \left[ \log\left(\frac{W}{Hz}\right) \right],$$
(31)

where  ${}^{0}M_{r}$  is the Eddington bias-corrected *r*-band absolute magnitude for the Milky Way obtained via GPR. This formula is derived via the relation for converting flux to AB magnitudes in combination with the area of a 10 pc radius sphere to convert flux to luminosity. The luminosity in any other band can then be calculated via the relation

$$\log\left(L_{\nu,x}\right) = \log\left(\frac{f_{\nu,x}}{f_{\nu,r}}\right) + \log\left(L_{\nu,r}\right).$$
(32)

We can then add log frequency  $(\log \nu)$  to obtain  $\log \nu L_{\nu,x}$ .

## 4. Calculate errors –

To convert uncertainties in magnitudes and colors to uncertainties in  $\log \nu L_{\nu,x}$  we make use of propagation of errors. First to determine  $\sigma_{\log \nu L_{\nu,r}}$  we calculate the partial derivative of  $\log \nu L_{\nu,r}$  (which is Equation 31 +  $\log \nu$ ) with respect to  ${}^{0}M_{r}$ . This yields

$$\sigma_{\log\nu L_{\nu,r}} = 0.4\sigma_{^0M_r}.\tag{33}$$

Errors in the other bands are calculated in a similar manner, but they depend on both the error in color and the error in  ${}^{0}M_{r}$ :

$$\sigma_{\log\nu L_{\nu,x}} = 0.4 \times \sqrt{\sigma_{0(x-r)}^2 + \sigma_{0M_r}^2}.$$
(34)

In the plots that follow we do not plot the contribution to errors from  $\sigma_{{}^{0}M_{r}}$  as it is fully covariant across all bands; as a result, when templates are normalized to match the observed SED, any error in  ${}^{0}M_{r}$  would simply change the normalization. We do provide its value for reference. Thus the error bars presented in the Milky Way SED plots are equivalent to  $0.4\sigma_{0(x-r)}$ .

The color predictions used to derive the luminosities used for our SED are provided for reference in Table 4.2 and the value of  ${}^{0}M_{r}$  is provided in the absolute magnitude table Table 4.4. We present our estimated luminosities and associated uncertainties, incorporating Eddington bias corrections in each case, in Table 4.1.

Passband	$\lambda_{eff} \; [\mu m]$	$\log \nu L_{\nu} \ [\log W]$	$\sigma_{\log \nu L_{\nu}} \ [\log W]$
FUV	0.155	35.53	0.20
NUV	0.2275	35.63	0.20
u	0.354	36.01	0.06
g	0.4750	36.44	0.02
r	0.622	36.62	0.09
i	0.763	36.66	0.01
Z	0.905	36.70	0.02
J	1.25	36.67	0.04
Н	1.65	36.54	0.05
Ks	2.15	36.39	0.06
W1	3.368	35.59	0.08
W2	4.618	35.58	0.10
W3	12.082	35.61	0.20
W4	22.194	35.49	0.30

Table 4.1: The passbands and corresponding power and uncertainties for the predicted SED of the Milky Way, as plotted in Fig. 4.9. These values have already had Eddington bias subtracted out.

#### 4.4.2.2 Interpreting the SED of the Milky Way

In Fig. 4.9 we present our full predicted spectral energy distribution for the Milky Way, along with a variety of empirical template galaxy spectra from the literature. We plot  $\log \nu L_{\nu}$ (the power emitted per log interval in frequency) on the vertical axis, in units of log Watts. The horizontal axis corresponds to restframe wavelength in units of  $\mu m$ ; each photometric band used is labeled at its effective wavelength along the top of the plot. The black open circles represent the estimates for the Milky Way's luminosity along with their associated errors, calculated as described in Section 4.4.2.1. We also show the error bar corresponding to the uncertainty in  ${}^{0}M_{r}$  near the bottom of the Fig. 4.9a for reference. The numerical values for the Milky Way SED corresponding to the plotted points are provided in Table 4.1.

While detailed fitting of the Milky Way's SED using physical models lies beyond the scope of this chapter, we will compare to observed SEDs of individual galaxies and composite galaxy templates from the literature as a sanity check on the realism of our results. Using photometry for extragalactic samples to constrain the SED of the Milky Way, and then comparing the results to observed galaxy photometry (albeit for different objects) is somewhat circular. However, given that our analysis has treated every band completely independently, there were no guarantees that we should get a sensible SED when combining GPR results across the spectrum.

In Fig. 4.9a we compare our predicted Milky Way SED to templates from Ref. [21], which are refinements to the templates from Ref. [74] and (for starburst galaxies) Ref. [196]. The Ref. [74] templates were based upon averaging the observed SEDs of relatively blue galaxies of a given morphological type. Given the broad range of observed SEDs for objects with similar morphological classification, these should not be considered universally applicable for all galaxies of a given type; however, we use the same labeling as Ref. [21] for consistency with the literature.

In this and successive plots, we have normalized all templates to match the estimated Milky Way SED in the r band. We normalize in an optical band as those bands have the smallest errors in the predicted SED; which particular optical band we choose has minimal effect on our comparisons. It is apparent that the Milky Way SED is generally consistent



Figure 4.9: The predicted SED for the Milky Way from a six-parameter Gaussian process regression, depicted by black open circles in all panels; the method of calculation is described in Section 4.4.2.1. (a) We compare to empirical galaxy templates from Ref. [21] (re-calibrated from Ref. [74]) normalized to match the Milky Way in the r-band. The Milky Way SED is consistent with the 'Sbc' template from this set, which is labeled as such because it was originally based on the average of the SEDs of two blue galaxies of morphological type Sbc. (b) We compare the SED of the Milky Way to the most closely-matched templates from Ref. [50], which are based on spectra and model fits to bright nearby galaxies. We show all templates whose  $\chi^2$  values in comparison to the Milky Way SED are below the 95% upper limit value of  $\chi^2$  for 13 degrees of freedom; again we normalize in the r-band. The two galaxies with the smallest  $\chi^2$  (which fall below the 68% upper limit) have their photometry plotted as round points, which are been offset in the wavelength direction for clarity. Their accompanying spectra from the SED atlas are also plotted, without an offset. We also show as fainter curves the SEDs for the remaining two galaxies with potentially matching SEDs, which have  $\chi^2$ values below the 68% limit. Images of the four best-fitting galaxies are provided in Fig. 4.10. Portions of the spectra from the SED atlas that are based on models are depicted using dotted lines, while those that are directly based on observations are shown using solid lines. (c) The optical and near-IR portion of the Milky Way SED, with Ref. [50] templates, as in (b). (d) The mid-IR portion of (b). The GPR method produces a Milky Way SED that is consistent in shape with both composite SED templates and individual observed SEDs.



Figure 4.10: Postage stamp images corresponding to galaxies within the Ref. [50] SED atlas with  $\chi^2$  values below the 95% upper limit when compared to the estimated SED of the Milky Way. At the top of each image we indicate the galaxy NGC number as well as the  $\chi^2$  value for the comparison of its SED to the Galactic one; we present them in order from smallest to largest  $\chi^2$ . At the bottom left of each tile is a letter marked (A) or (B) which denote the source of the given image. (A) images are from Ref. [117] and (B) images are from Ref. [322]. The border around each image matches the color used for its SED in Fig. 4.9b. It is evident that galaxies that may have similar SEDs to the Milky Way exhibit a broad range of visual morphologies, further emphasizing the lack of a unique mapping between morphology and galaxy SED.

with the [21] "Sbc" galaxy template. The galaxies used to construct this template, M51 (NGC 5194) and NGC 2903 are undergoing moderate amounts of star formation, with log specific star formation rates of -10.1 and -10.4 (versus -10.5 for the Milky Way) [264]. The Milky Way is generally expected to have an SBbc morphological type (see, e.g., Ref. [28] for a recent review on Milky Way structure, as well as Ref. [167, 191, 110]), though given how it was constructed, we should not read too much into the agreement of our SED with the Ref. [21] "Sbc" template in particular. It is clear, however, that the GPR method yields results that resemble composite SED templates from the literature.

Ref. [21] provides only a sparse set of composite templates that may not match the SEDs of an individual galaxy. Additionally, those templates do not span the full wavelength range of the Milky Way SED we have produced. Thus we also compare to the set of 129 observed galaxy spectral energy distributions from Ref. [50]. This SED atlas encompasses a variety of bright galaxies in the very local Universe. Unlike the templates shown in Fig. 4.9a, these correspond to SEDs of individual galaxies (not averages) with minimal modeling used to interpolate between photometric and spectroscopic coverage.

The data tables from Ref. [50] provide extinction-corrected photometry as well as a variety of summary values such as luminosity distance. Because the magnitudes are presented in the AB system we can use the relation:

$$\log f_{\nu} = \frac{m_{AB} - 8.9}{-2.5},\tag{35}$$

to calculated flux, where  $f_{\nu}$  is flux in units of Jansky and  $m_{AB}$  is the observed magnitude in each band. We neglect k-corrections as these galaxies are very nearby  $(3.1 < D_L < 249.2 \text{Mpc})$ at the most extreme), so the corrections are generally negligible. We then can use the fluxluminosity relation  $L_{\nu} = 4\pi D_L^2 f_{\nu}$  to calculate the luminosity in each band. Via propagation of errors the uncertainty in  $\log \nu L_{\nu}$  for these templates is equivalent to  $\sigma_{\log \nu L_{\nu}} = 0.4\sigma_{m_{AB}}$ .

We take a few further steps before comparing the Ref. [50] SEDs to our Galaxy's. First, we have obtained the axis ratios (as a proxy for inclination) for 89 out of the 129 galaxies in the Ref. [50] sample from the Siena Galaxy Atlas, which have been distributed as part of DESI Legacy imaging surveys Data Release 9. All galaxies with b/a below 0.5 or unknown axis ratios were excluded from comparisons. Lower axis ratios should correspond to highlyinclined galaxies for which reddening will strongly affect the SED [353, 234], making them inappropriate comparisons to our face-on SED for the Milky Way. We then normalise the observed SEDs to match the Galactic SED in the r-band, as was done for the Ref. [21] templates before.

Finally, we calculate the  $\chi^2$  difference between our predicted SED for the Milky Way and each of the galaxy SEDs presented in the Ref. [50] atlas. We emphasize that no fitting is performed in this comparison other than matching in the r band. We then calculate  $\chi^2$ using log quantities, as that is the space in which we perform our predictions and for which errors are (by construction) symmetric:

$$\chi^2 = \sum \left( \frac{\log \nu L_{\nu}^{\text{atlas}} - \log \nu L_{\nu}^{\text{MW}}}{\sigma_{\log \nu L_{\nu}}} \right)^2, \tag{36}$$

where  $\sigma_{\log \nu L_{\nu}}$  combines in quadrature the total error in the MW SED for a given band, the uncertainties in the Ref. [50] photometry, and  $\log_{10}$  (1.1), which corresponds to a 10% error in  $\nu L_{\nu}$ . This extra error is added to account for systematic uncertainties in the photometry for a given band relative to others; if this were not included, optical bands would dominate the  $\chi^2$  value due to their small nominal uncertainties. We calculate  $\chi^2$  using the 14 bandpasses in which we have measured the Milky Way's predicted SED, which yields 13 total degrees of freedom (one is lost due to the *r*-band normalization performed).

Fig. 4.9b over-plots the Ref. [50] SEDs for galaxies whose  $\chi^2$  values fall within the 68% upper limits for a  $\chi^2$  distribution with 13 degrees of freedom (corresponding to  $\chi^2 = 14.8$ ). We also examine galaxies that fall below the 95% limit (with  $\chi^2 < 22.4$ ), but exclude them from plotting for brevity. Objects below the 68% upper limit (four in total) are clearly consistent with the SED of the Milky Way, while those between the 68 and 95 percent limits (comprising three objects) are in some tension with the SED of the Milky Way, but could still be a match. While we calculate  $\chi^2$  using the broadband photometry for each galaxy, we also plot the full SED from Ref. [50] for each of these galaxies for reference.

In Fig. 4.9b the two galaxies with the smallest  $\chi^2$  are labeled and plotted with the highest opacity in teal and gold. For these two galaxies we also plot the observed photometry as points offset in the wavelength direction so they are easier to compare to the Milky Way

values. The higher  $\chi^2$  objects are plotted with low opacity in pale blue. We also provide more detailed plots of two separate ranges of the SEDs. The optical through near-IR regime is depicted in Fig. 4.9c, while the near- to mid-IR SED is depicted in Fig. 4.9d. Portions of the spectra that are based on observations are plotted with solid lines, while modeled portions are plotted with dotted lines.

We also provide images for the four Ref. [50] atlas galaxies that fall below the 68% limit in Fig. 4.10. At the top of each postage stamp image we list the galaxy NGC number, as well as the  $\chi^2$  value for comparing photometry for that galaxy to our Milky Way SED. The tiles are presented in order from smallest to largest  $\chi^2$  value. At the bottom left of each tile is a letter marking (A) or (B) which refers to the source for each given image. (A) images are from Ref. [117] and (B) images are from Ref. [322]. The borders that surround each image match the color coding in Fig. 4.9b.

It is clear from this comparison that our GPR method produces results whose spectral shape is comparable to observed galaxy SEDs. The small total number of SEDs within the Ref. [50] atlas that are consistent with the Milky Way is no surprise; after cutting on inclination we are reduced to only 70 objects, the majority of which are early-type galaxies, leaving a limited number of examples to cover the full range of star-forming galaxies. We emphasize that there remains a need for more careful investigation and modeling of the Milky Way SED we have obtained; this will be the topic of follow-up work.

However, we will briefly comment on the galaxies from the Ref. [50] atlas whose SEDs are most consistent with the Milky Way. NGC 4138 is an Hubble type SA(r)0 galaxy which contains an AGN and a star-forming ring. Using estimates for mass from Ref. [182]  $(2.92 \times 10^{10} \text{ M}_{\odot})$  and Ref. [187]  $(6.23 \times 10^{10} \text{ M}_{\odot})$  and the star formation rate estimates from Ref. [380]  $(0.14 \text{ M}_{\odot}\text{yr}^{-1})$  and Ref. [49]  $(0.2 \text{ M}_{\odot}\text{yr}^{-1})$  yield a log sSFR of ~ -11.3 to -11.5, significantly smaller than the Milky Way value (-10.52). NGC 4138 has a  $^{0}(g - r)$  color of  $0.73 \pm 0.05$  Ref. [50], matching the Milky Way value.

In contrast, NGC 3351/M95 is a galaxy of Hubble type SB(r)b, versus SBb or SBc for the Galaxy; it contains a pseudo-bulge [314, 127, 50], much as the Milky Way is conjectured to possess [28]. According to measurements provided by Ref. [215, 142] NGC 3351 has a log specific star formation rate of -10.43 per year (SFR = 0.940 M<sub> $\odot$ </sub>yr<sup>-1</sup>), which is comparable to the log sSFR of the Milky Way of -10.52. According to the measurements complied by Ref. [50], NGC 3351 has a  ${}^{0}(g - r)$  color of  $0.74 \pm 0.05$ , very similar to that of the Milky Way's ( $0.73 \pm 0.05$ , see Table 4.2). It matches the Milky Way SED equally well as NGC 4138 at most wavelengths, with the exception of the longest-wavelength W3 and W4 bands. However, we note that the photometry for NGC 3351 from Ref. [83] does not show the strong red slope in the mid-IR seen in the SED atlas spectrum. If the Ref. [83] measurements were used in the mid-IR, the agreement with the Milky Way SED would be significantly better.

Two other galaxies in the atlas have SEDs for which the  $\chi^2$  value when compared to the Milky Way SED are below the 68% significance level. NGC 5055 is a galaxy of Hubble type SAbc [50] with log sSFR of -10.53 as tabulated in Ref. [192]. NGC 3265 is of Hubble type SA(rs)0 pec [8] with log sSFR of -9.12 [192]. The set of objects whose SEDs are consistent with the Milky Way's span a diverse range of visual morphologies.

#### 4.4.2.3 Impact of Physical Parameters on the Estimated SED of the Milky Way

Thus far we present results based upon a set of fiducial values (with uncertainties) for the Milky Way's stellar mass, star formation rate, axis ratio, disk scale length, bulge-to-total mass ratio, and bar presence (presented in Section 4.2.3). Because of the generality of the Gaussian process regression fit, we can vary each of these parameters one at a time and test its impact on the inferred SED.

In our previous analyses we randomly drew values from the distributions of Milky Way properties and made predictions based on each of those draws, which were then combined, as described in Section 4.3. This allowed us to incorporate the uncertainties in the Milky Way's physical properties into our results. In this subsection, however, we will neglect these uncertainties in order to isolate the effect of changing the central values for each parameter. We have performed a similar analysis to the one presented below by sampling from Milky Way uncertainties as a cross-check; the results are very similar so we do not present them here for brevity.

In this analysis we keep the values for the five Milky Way parameters *not* being studied at their fiducial mean. We then choose a discrete set of values for the sixth parameter at which



Figure 4.11: Isolating the contributions of each physical parameter to the SED. Each panels shows the effect of varying one parameter while fixing the other five parameters to the fiducial Milky Way values (see Section 4.2.3). We vary most parameters over a range of  $\pm 2\sigma$  from the Milky Way's fiducial values. In the case of axis ratio, we explore a wider range of values to convey better the impact of inclination on the observed SED. The upper-left panel depicts the different SEDs that would be inferred for different galaxy axis ratios, with the expected results that an inclined Milky Way would look much fainter than face-on. In the remainder of the panels we depict the log of the predicted SED divided by the SED evaluated for fiducial Milky Way parameters in order to make differences more clear; the plotted quantity is thus  $\Delta(\log(\nu L_{\nu})) = \log(\frac{\nu L_{\nu}}{\nu L^{MW}})$ . The predicted SED for the Milky Way value is plotted with a dashed black line and open circles at the passbands. On the color bars the Milky Way's value is marked by a horizontal white line. In b/a the GP captures the effects of dust reddening at higher inclinations. If the Milky Way were to have a higher SFR, the SED would be brighter in both the UV and IR. If the Milky Way's disk were more extended we would observe an increased UV brightness and decreased mid-IR brightness. We caution the reader that predictions for disk scale lengths more than  $1\sigma$  below the Milky Way value may not be reliable. Relatively few galaxies of the mass of Milky Way have sizes smaller than the Milky Way, causing the GPR to be poorly trained for small  $R_d$  values [218]. Changes in the Milky Way's B/T on the SED would be minimal. As we decrease the intensity of the Milky Way's bar it seems to have a minimal effect, with a slightly increased UV brightness. In general these follow our expectations of galaxy evolution which we discuss further in Section 4.4.2.3.

to evaluate the SED via GPR. In the case of star formation rate, bulge-to-total mass ratio, and bar vote fraction we select 8 values that lie evenly spaced between  $\pm 2\sigma$  of the fiducial mean value for the Milky Way (inclusive), in addition to evaluating at the nominal value. For axis ratio we step through a wide range of possible galaxy axis ratios instead of focusing around 0.9 in order to capture the full effects of inclination on the Galaxy SED. We have excluded mass from this exercise, as changing mass would radically alter the normalization of the predicted SED.

For each of the values that we step through for the given test parameter, the GPR is evaluated as before. We apply a  $12\sigma$  cutoff on the training sample for all parameters *except* the one being varied. By focusing on one parameter at a time we can explore the impact each has on the predicted SED for the Milky Way and can assess whether the GP is able to capture the expected correlations between galaxy properties.

The results of this analysis are presented in Fig. 4.11. In the color bars for each panel the lighter shades correspond to smaller values for the parameter being varied and darker shades correspond to larger values. In all panels the fiducial value for the Milky Way is marked by a horizontal white line on the color bar.

The upper left panel shows the GPR-predicted SED for each axis ratio value considered, with axes similar to Fig. 4.9. The SED evaluated with b/a = 0.9, the fiducial axis ratio value used for the Milky Way, is marked by open points. In this panel one can see the effects of inclination reddening first hand, an effect that has been seen repeatedly in analyses in spiral galaxies [324, 353, 234, 244]. The cross-section for dust extinction and scattering generally increases with decreasing wavelength, causing reddening effects to be the strongest at the shortest wavelengths. Thus we expect the SEDs of galaxies to appear redder the more inclined they are [387]. The increased attenuation at higher inclinations for galaxies in the training sample causes the GPR to predict a redder SED as the inclination increases (lower b/a). We find a decreased brightness in the UV/optical and an increased brightness in the IR as disks are viewed more edge-on, an effect also observed by SED modelers (see e.g., [281]).

In the remainder of the panels we plot SEDs *divided by* the SED evaluated at the fiducial Milky Way values, as effects are more subtle and would be difficult to discern in an unnormalized plot. We include a plot of this type based on varying axis ratio in the lower left panel, though we use a larger y-range than the other normalized plots due to the large dynamic range spanned.

An SED which matches the prediction for the fiducial MW values exactly would fall along the horizontal line at  $\log(\nu L_{\nu}) - \log(\nu L_{\nu}^{MW}) = 0$ . The photometric predictions for the Milky Way nominal parameters hence correspond to the open circles along this line.

Star formation rate has the clearest effect on the SED, as seen in the top middle panel. The amount of UV flux is a sensitive indicator of star formation as it is dominated by hot, massive, short-lived stars. These stars contribute to the flux at optical and near-IR wavelengths, but are subdominant there; but in the mid-IR the SED responds strongly to star formation due to light from hot stars that is reprocessed by dust. The GPR predictions reflect all of these phenomena.

The upper right panel depicts the effect of disk scale length on the predicted SED. We note that the predictions become unstable at shorter scale lengths due to the small number of Milky Way-mass galaxies with radius smaller than our Galaxy in the training samples (cf. Ref. [218]), so results for disk scale lengths more than  $1\sigma$  below the Milky Way value may not be robust. Outside of that regime, it is clear that Milky Way-like galaxies with shorter scale lengths exhibit significantly less flux in the UV than those with longer scale lengths when M<sub>\*</sub>, SFR, etc. are all held fixed, along with smaller effects at optical-IR wavelengths.

A smaller disk, with other properties held fixed, could imply that the gas within the disk is denser and dust columns are correspondingly greater. This would in turn cause the SED to look fainter in the UV relative to the IR compared to if the disk were more extended. We have tested the effect of varying b/a and  $R_d$  simultaneously and find that we can compensate for a change in one with a change in the other almost perfectly. Given the long-standing scenario that inclination effects on the SEDs of disc galaxies are driven by dust [234], it is reasonable to hypothesize that the effects of varying  $R_d$  must relate to varying dust impact as well.

The bottom middle panel shows the effect of varying only the bulge-to-total ratio. Overall, the impact on the SED is small. We see that if the Milky Way were to have a more massive bulge, it would appear slightly fainter in the UV. This could reflect the fact that bulges have older stellar populations than spiral disks; the effect may be subtle here as to first order the effect is captured by variation in specific star formation rate. Bulges are also susceptible to dust reddening [347, 105], so a larger bulge may also suffer from greater reddening effects at short wavelengths.

The lower right panel shows the variation in SEDs as we change the bar vote fraction. As this quantity increases we see a decrease in UV brightness and an increase in mid-IR brightness. In general we speculate that most effects caused by a bar may have been captured in the variation with other measured properties (particularly star formation rate). The observed trend with  $p_{\text{bar}}$  could cohere with a narrative of bar-induced star formation suppression (see e.g., [346, 163]), which would affect the UV bands the most and lead to a galaxy looking more red once the bar instability has caused the consumption of cold gas in the disk [244]. However, this would not explain why galaxies with stronger bars are brighter in the mid-IR while simultaneously being fainter in the UV. Alternatively, one could speculate that processes associated with bars could modulate the properties of interstellar dust (e.g., by affecting the ability of gas to cool and form molecular clouds); having more dust (at fixed  $M_*$ , SFR, and inclination) should cause a reduction of flux in the UV and an increase in flux in the mid-IR, as observed here.

## 4.4.2.4 Exploration of Other Sources of Physical Parameter Measurements

As described in Section 4.2.2.1 and Section 4.2.2.2, there exist multiple options for the values used for the stellar mass, star formation rate, and bulge-to-total ratio for SDSS galaxies. In this subsection we discuss the impact on our results of using different measurements of galaxy parameters or different methods of defining our training samples.

The GSWLC-M2 catalog [312, 311] includes estimates of stellar masses and star formation rates computed based on the photometry within the catalog. While the M<sub>\*</sub> values in the GSWLC-M2 catalog closely match those from the MPA-JHU catalog [46] used for our main results, the SFRs presented in the GSWLC-M2 catalog are far less bimodal than those presented in the MPA-JHU catalog; a significant number of galaxies would be classified as star-forming in GSWLC that would be considered quiescent based on the MPA-JHU catalog.

We have produced a Milky Way SED using the GSWLC-M2 stellar masses and star

formation rates as features in place of the MPA-JHU values, but otherwise following the same methods used to produce the results in Section 4.4.1 and Section 4.4.2. The impact on predictions in the optical and IR is small. The effect is more notable in the UV, where the Milky Way is predicted to be closer to the mean color of star-forming galaxies with the same sSFR (corresponding to smaller values of  ${}^{0}(FUV - r)$  and  ${}^{0}(NUV - r)$  in Fig. 4.7). Even so, the predicted UV colors are still within  $0.5\sigma$  of those resulting from using MPA-JHU M<sub>\*</sub> and SFRs. In addition to being brighter in the NUV and FUV, the predicted SED is also marginally fainter in W3 and W4 compared to the results presented in Fig. 4.9, with shifts that are again well below  $1\sigma$  in each band.

Overall we find that using the stellar masses and star formation rates derived from the GSWLC-M2 catalog instead of the MPA-JHU catalog has little effect on our predictions for the Milky Way, and is subdominant to other sources of uncertainty.

There are also multiple options for which band to measure bulge-to-total ratios; the Ref. [327] catalog contains bulge and disk decompositions performed both in the g and r bands. For our main results we use the r-band value,  $B/T_r$ , but have also tested the impact of instead using  $B/T_g$  on our results presented in Section 4.4.1 and Section 4.4.2.

Overall we find only very small effects on predicted colors from changing to  $B/T_g$ . All results are well within a few hundredths of a magnitude of the previous values, except in the case of  ${}^{0}(u-r)$ . For that value, the predicted value for the Milky Way from the GPR trained on  $B/T_g$  is larger than when we use  $B/T_r$  (corresponding to redder color; cf. Fig. 4.6), though still within  $0.5\sigma$  of our primary result. As a consequence, the predicted SED for the Milky Way using  $B/T_g$  instead of  $B/T_r$  is almost identical to before.

As discussed in Section 4.2.2.3, using citizen science votes from Galaxy Zoo 2 requires consideration of responses to previous questions that influence whether the question of interest is even asked. If we wished to select a pure sample of barred galaxies, it would be necessary to ensure not only that a high fraction of people who were asked whether a bar is present voted in the affirmative, but also that a substantial total number of people voted on each of the preceding questions in order to minimize errors in vote fractions (see, e.g., Ref. [381]). However, with GPR we may gain more information about trends that influence photometric properties by including a broader set of objects, so we do not necessarily wish to exclude non-barred or non-spiral galaxies from the training sample. We explore here how changing the treatment of GZ2 votes influences our GPR predictions.

We have tested what impact restricting the training set to only barred, face-on spirals would have by testing how predicted colors for the Milky Way vary when using training samples with a variety of different constraints: (1) a control sample without any restrictions based on Galaxy Zoo 2 vote results; (2) a sample where if the number of votes on whether or not a galaxy has a bar,  $N_{\text{bar}}$ , is less than ten we set  $p_{\text{bar}} = 0$ ; (3) a sample using the bar selection cuts from Ref. [381] Table 3, column 3, rows 2 and 3 in addition to the vote count thresholds mentioned in Section 4.2.2.3; or (4) a sample using the same cuts as Ref. [381] except setting  $p_{\text{bar}} = 0$  when  $N_{\text{bar}} < 10$ , rather than rejecting objects with low  $N_{\text{bar}}$  from the set entirely.

Applying all of the Ref. [381] cuts reduces the size of the training sample by roughly an order of magnitude, degrading the ability of GPR to predict colors and increasing net errors. Furthermore, requiring  $N_{\rm bar} \geq 10$  not only shrinks the size of the sample but also greatly biases the luminosity distribution of the training sample compared to a volumelimited sample, which may result in biases in inferred photometry. We have explored how the GPR-predicted Milky Way colors change for each of these four training sample definitions (but otherwise using the methodologies described in Section 4.3). The results from (2) compared to our fiducial case, (1), are nearly identical, so the particular values of  $p_{bar}$ assigned to objects with poorly-constrained vote counts cannot have had a large systematic impact on our predicted Milky Way photometry. We also find that restricting to training sample (3) or (4) yields much ( $\gtrsim 2\times$ ) larger errors on all predictions. Results from (3) and (4) are still within  $1\sigma$  of those from (1) and (2), however. Therefore we conclude that with GPR we get better predictions when we include more objects (including some with noisier vote fractions) than when we instead restrict training to just the best-constrained objects. Therefore we perform no GZ2-based cuts on the galaxy sample used for training, and instead include objects spanning the full range of bar, face-on, and features vote fractions in the training set.

#### 4.5 Summary and Conclusions

#### 4.5.1 Summary

In this work we have set out to estimate a full SED for the Milky Way, spanning wavelengths from the UV to the IR. Our central motivation is twofold: (1) to improve our understanding of how the Milky Way compares to the general galaxy population and by doing so (2) guide the tuning of parameters in simulations in order to create more realistic galaxies.

The previous work by Ref. [219] constrained the optical colors and luminosity of the Milky Way using Milky Way analogue galaxies selected based on their stellar mass and star formation rate, obtaining the best constraints on the Milky Way's photometric properties available previous to this work. Here, we have been able to reduce the uncertainties on these constraints further by incorporating information from additional parameters such as disk scale length and bulge-to-total ratio, that also connect to a galaxy's evolutionary history [61, 307], and have for the first time developed predictions for Milky Way photometry at wavelengths beyond the optical.

We have shown that the Milky Way analogue method breaks down when we attempt to match the Galaxy in many physical parameters; the number of Milky Way analogues rapidly approaches zero in higher-dimensional spaces (cf. Fig. 4.1). Expanding to a wider wavelength range requires information from data sets that do not cover the full SDSS footprint, making the problem worse. We instead have predicted the photometric properties of the Milky Way using Gaussian process regression, which provides an optimal means of interpolating information from a limited training set. We have performed a series of tests throughout this chapter that have demonstrated that GPR is able to produce realistic and reliable photometric predictions.

We have compared predictions for the Milky Way to the broader local galaxy population in color-mass, color-specific star formation rate, and color-color diagrams. As exemplified by Fig. 4.5, we obtain similar results in the optical to those reported by Ref. [265] and Ref. [219], though with reduced errors, further confirming the Milky Way has optical colors consistent with the green valley population. For the first time we have also predicted UV (Fig. 4.7) and IR colors Fig. 4.8 for the Milky Way, which provide more sensitive diagnostics of the evolutionary status of a galaxy. We find that in both these regimes the Milky Way appears to lie on the star-forming side of the green valley.

In this work we have determined the luminosity and colors of the Milky Way for GALEX FUV and NUV, SDSS ugriz, 2MASS JHKs, and WISE W1-W4 bands in an entirely selfconsistent way, giving us unprecedented constraints on its spectral energy distribution. We have constructed the first multi-wavelength SED for the Milky Way. This SED has a shape consistent with both composite galaxy templates (Fig. 4.9a) and observed SEDs of individual galaxies (Fig. 4.9b). The GPR method produces a realistic SED with errors and captures previously known galaxy property correlations, such as those between reddening in spiral galaxies and viewing angle or between star-formation rate and UV and IR flux (Fig. 4.11). High-resolution hydrodynamical simulators [157, 316, 376] no longer have to compare their mocks of the Milky Way blindly to photometric constraints from broad galaxy populations that span a wide range of properties. Rather, it should now be possible to tune the treatment of star formation efficiency, threshold gas density for star formation, and dust properties to produce galaxies which match the photometric properties of the Milky Way directly, while simultaneously exploiting those properties that we can measure well from inside the Galaxy.

## 4.5.2 Discussion: The Milky Way as a Red Spiral

As previously suggested in a variety of works (e.g., Ref. [310], Ref. [319], and Ref. [219]), definitions of the green valley that rely only on optical bands may lead to misleading conclusions. The Milky Way has a specific star formation rate that is higher than the canonical values for green valley galaxies,  $\log \text{sSFR} = -10.52$  as compared to  $\sim -11.8 < \log \text{sSFR} < -10.8$ for transitioning galaxies from Ref. [310], even though it has red optical colors for a starforming object. However, at UV and IR wavelengths the colors of the Milky Way more clearly place it amongst the star-forming population. This combination of red optical colors when viewed face-on with significant star formation evident at UV and IR wavelengths is characteristic of the previously-identified population of red spiral galaxies. A population of red spiral galaxies in clusters was first identified by Ref. [356]. Since then, these "passive spirals" have been identified at a range of redshifts and in multiple datasets. As noted by to Ref. [81], for galaxies with a stellar mass above  $10^{10} M_{\odot}$  like the Milky Way  $(M_* = 5.48^{+1.18}_{-0.94} \times 10^{10} M_{\odot})$ , the blue cloud and red sequence overlap in their optical colors (this is also consistent with findings by [310]). This makes optical photometry a poor choice for constraining the star formation activity for galaxies like our own. However, these massive objects still exhibit a distinct color bi-modality in the UV, as shown by Ref. [386, 310] and is evident from comparing Fig. 4.6 and Fig. 4.7. In comparison to their lower-mass counterparts, massive galaxies produced the great majority of their stars at earlier epochs [39]. This causes the optical colors of massive galaxies to be dominated by relatively old stellar populations as opposed to probing recent star formation activity [386, 71, 81]. Ref. [159] and Ref. [399] provide evidence that this is the case for red spirals. Direct or re-radiated light from young stars still dominates the red spiral SEDs at UV and IR wavelengths, however.

Reflecting that, Ref. [81] and Ref. [331] both find that red spiral galaxies tend to be UV bright; this can be driven by a relatively small amount of total star formation. In order to facilitate comparison of the Milky Way to the red spiral galaxy population, we have overplotted the red spirals from the Ref. [244] catalog (based on a Galaxy Zoo 2 and optical color selection) that are also part of our cross-matched galaxy catalog on all color-mass, color-sSFR, and color-color diagrams presented here. In each diagram the Milky Way falls near the middle of the red spiral population.

Ref. [81] notes that 85-90% of objects in their red sample maintain SFRs of ~  $1M_{\odot}$  yr<sup>-1</sup>; Ref. [244] found that red spiral galaxies selected from Galaxy Zoo 2 typically had lower rates of ongoing star formation than blue spirals of the same mass, but still non-negligible. For comparison the Milky Way has a SFR of  $1.65 \pm 0.19 \text{ M}_{\odot}\text{yr}^{-1}$  [216], while the average star formation rate of galaxies of approximately the same mass as the Milky Way ( $\pm 0.3$  in log stellar mass) with B/T < 0.75 (to exclude ellipticals) within our cross-matched galaxy sample is  $1.69 \text{ M}_{\odot}\text{yr}^{-1}$ ; the Galaxy is very close to average in this respect.

Ref. [244] also finds that red spiral galaxies have a significantly higher bar fraction compared to blue spirals of the same mass; 70% versus 27%. This matches with the clear evidence that the Milky Way possesses a bar [32, 28, 325]. Ref. [244] notes that one possible evolutionary scenario for red spirals is bar-driven gas inflows. This removes gas from the outer disk and funnels it into central star formation, which in turn causes the disk to appear more and more red over time [246, 308, 68, 130].

Ref. [247] and Ref. [130] find that barred spirals tend to have redder colors than their unbarred counterparts. It thus may be the case that the bar has played a role in the colors that we observe for the Milky Way in this work. For example, bar quenching may play a role in the development of red spirals, as Ref. [12] finds many high-stellar-mass red spirals in the field and Ref. [331] notes that field red spirals most likely evolve primarily via secular evolution due to the lack of nearby galaxies. In the case of the Milky Way, work by Ref. [163] and Ref. [194] find that the bar may have played a substantial role in the star formation history of the Milky Way (leading to a significant decrease in star formation 9-10 Gyr ago, and thereby causing the observed pattern of chemical abundances in the disk). Although the effect of bar vote fraction in Fig. 4.11 is small, it may be that the effects of a bar are primarily captured by other parameters (e.g., SFR).

We note that Ref. [118] studied a population somewhat similar to red spirals, which they labelled "red misfits". Ref. [118] define this population as corresponding to objects with log(sSFR) > -10.8 and restframe g - r > 0.67 (i.e., specific star formation rate measured to be above the value for the saddle point in the bimodal distribution and color redder than the saddle point in the color bimodality). Based on these divisions, the Milky Way almost certainly meets this definition (which is less stringent than most red spiral classifications).

#### 4.5.3 Outlook

As seen in Fig. 4.10, there is a significant diversity in the set of galaxies that have SEDs consistent with the Milky Way, given the measurement uncertainties in both our results and the Brown SED atlas. The goal of this chapter has been to construct the Milky Way's UV-to-IR SED to enable comparisons to samples of external galaxies and to improve the tuning of simulations. However, in follow-up work we will fit the estimated SED of the Milky Way using population synthesis models to obtain more detailed constraints on how the star formation history, dust reddening properties, and metallicity of the Galaxy would be interpreted from outside (see e.g., [78]). This will require proper treatment of covariances between different photometric bands; we will address this by employing multioutput Gaussian process regression in this future work.

The longest-wavelength WISE W3 and W4 bands could have substantial discriminating power on what SEDs are consistent with the Milky Way's, if they only had smaller errors, as is evident in Fig. 4.9. However, currently these bands are poorly measured compared to the optical or near-IR; for most objects used in training the Milky Way SED, the signal-to-noise ratio in these bands is below one. Given the low effects of dust extinction in these bands, investigation of the flux ratio (or color) in these bands across the all-sky WISE imaging, potentially combining modeling of smooth components of the Milky Way with mapping of the contributions from dust, may provide an alternative method to constrain the color of the Milky Way at the longest wavelengths. If luminosities in the W3 and W4 bands can be measured relative to the luminosity in W2, long-wavelength measurements could be effectively anchored well to the SED presented here; measuring such relative quantities should be affected less by modeling uncertainties than absolute measurements would be.

The SED presented in the chapter (or future improved versions) can be used to identify multiwavelength Milky Way analogue galaxies by matching in unresolved photometric properties. If we do not need to require detailed morphological measurements or citizen science inspection of images it would greatly increase the size of the parent catalogs that could be used to identify MWAs, which could be useful for a variety of follow-up studies such as determining gas masses for the Milky Way or studying environments of Milky Way-like galaxies.

The Milky Way appears to be atypical in its satellite population and mass assembly history. For example, Ref. [119] finds that the assembly history of the Milky Way is only reproduced in 0.65% of Milky Way-mass EAGLE galaxies. This assembly history should be closely related to the local environment surrounding our Galaxy, and environment has been found to play a key role in the formation and evolution of galaxies (e.g., [357, 289, 33]).

In future work we plan to explore how incorporating measures of galaxy environment (e.g., measures of the local overdensity of galaxies) within a GPR model affect the predicted SED of the Milky Way. The noisiness of environment measures [170] and the impact of SDSS fiber collisions on Local Group-like systems (as typically only one galaxy out of two close neighbors would be observed, causing analogues of a Milky Way-M31 pair to be missed) may limit the information that may be gained from this, however. While we anticipate the environment to have a small impact on the Galactic SED compared to the dominant effects of stellar mass and star formation rate on galaxy colors [156], assessing the local environments of the most Milky Way-like galaxies may allow us to explore and to what extent our Galaxy's environment has shaped its exhibited characteristics.

Gaussian process regression can be useful for a variety of studies beyond the photometric estimates for the Milky Way considered here. For this reason the authors have provided their analysis code on our project GitHub for full public access for adaption to any other project, under a CC BY-SA 4.0 license.

## 4.6 Data Summary

#### 4.6.1 Galaxy Properties

In this subsection we summarize the distribution of properties of the galaxy sample used for training the GPR (which is described in Section 4.2). This sample consists of crossmatches between the SDSS DR8 volume-limited sample reported in Ref. [219], the MPA-JHU catalog of galaxy stellar masses and star formation rates [46], the Ref. [327] morphological catalog, the Ref. [312, 311] GSWLC-M2 photometric catalog, the DESI Legacy imaging survey DR8 [96], and the Galaxy Zoo 2 catalog [381, 160].

In Fig. 4.12 we present a corner plot for the distributions of the various galaxy physical properties used to train the GPR: mass (M<sub>\*</sub>), star formation rate (SFR), axis ratio (b/a), bulge-to-total ratio (B/T), disk scale length  $(R_d)$ , and bar vote fraction  $(p_{\text{bar}})$ . In each contour plot the fiducial value for the Milky Way (from Section 4.2.3) is marked by a black star, while in the histograms the fiducial value is marked by a vertical dashed black line. Overall we can conclude that covariances are fairly weak for most parameter pairs, with the exceptions of the well-known M<sub>\*</sub>-  $R_d$  (or mass-size) relation, the bimodal distribution of objects in the  $M_*$ -SFR plane (corresponding to the red sequence/blue cloud division), and weaker correlations between galaxy bulge-to-total ratio or bar vote fraction and star formation rate.

## 4.6.2 Tabulated Photometric Predictions for the Milky Way

In this section we present the estimates of the colors and luminosities for the Milky Way which we have obtained via Gaussian process regression. Table 4.2, Table 4.3, and Table 4.4 update Table 1 and 3 from in Ref. [219], and also incorporate results for the non-SDSS bands used in this work. In addition, Table 4.2 tabulates the colors used to derive the luminosities presented in Table 4.1, a necessary ingredient for our SED calculation.

In Table 4.4 all absolute magnitudes are calculated assuming a Hubble constant of  $H_0 = 100h$ km s<sup>-1</sup> Mpc<sup>-1</sup>. Results for *FUV*, *NUV*, *ugriz*, *JHKs*, and *W*1–*W*4 are all presented in the AB magnitude system, as in the body of the chapter. The Johnsons-Cousins *UBVRI* values are provided in the Vega magnitude system. These were converted from the *ugriz* measurements via the KCORRECT V4.2 software on an object-by-object basis, as in Ref. [219]. We remind the reader that absolute magnitudes are computer with 5 physical parameters instead of 6, due to the increased cumulative variance when  $p_{\text{bar}}$  is incorporated in the GPR for them. Therefore we do *not* include derivatives with respect to  $p_{\text{bar}}$  for them, as bar vote fraction was not used in the predictions of these quantities.

All tabulated results are presented for rest-frame z = 0 passbands. We note that each table row is calculated independently from all others; for example,  ${}^{0}(FUV-r)$  is not calculated from subtracting the predicted  ${}^{0}M_{r}$  from the predicted  ${}^{0}M_{FUV}$ , but rather from a Gaussian process regression that predicts  ${}^{0}(FUV-r)$  directly. This will in general yield smaller errors for color calculations, as some of the errors in absolute magnitudes from different bands will be covariant. We derive color predictions from model magnitudes while absolute magnitudes are determined from cmodel magnitudes, again to minimize errors.

In these tables, the "Corrected Value" column corresponds to the predicted color or absolute magnitude from the GPR, derived as the mean predicted value from 1,000 draws from the fiducial distributions of Milky Way physical properties (as described in Section 4.3),



Figure 4.12: Distribution of the various galaxy physical properties used to predict the Milky Way SED, viewed as both two-dimensional projections and one-dimensional histograms. In the contour plots the fiducial values for the Milky Way are designated by a black star, while in the histograms the fiducial value is indicated by a vertical black dashed line. The covariances are weak in almost all parameter combinations, except for the well-known  $M_*$ -  $R_d$  (mass-size) relation, the bimodal distribution of galaxies in the  $M_*$ -SFR plane, and weaker correlations of bulge-to-total ratio and bar vote fraction with star formation rate.

after Eddington bias has been subtracted. The amount subtracted is tabulated for reference in the "Bias Removed" column. The errors in the "Corrected Value" columns have had the uncertainty in the Eddington bias added in quadrature and represent  $1\sigma$  errors.

We also tabulate the derivatives of each photometric property with respect to every galaxy physical property used for the prediction, allowing the photometry presented here to be updated for values of Milky Way parameters that differ from the fiducial values used in this work (or, conversely, to correct for a re-calibration of extragalactic values). The methods applied are directly adapted from those presented in Section 5 of Ref. [219], but we provide a brief summary here.

To calculate the derivatives we offset each of the Galaxy's measured properties by  $\pm 0.1$  multiplied by the fiducial error in that property ( $\sigma$ ). We then evaluate the GPR using these "offset" properties, rather than performing random draws from the Milky Way distributions. For example, in order to calculate  $\partial({}^{0}M_{r})/\partial(\log SFR)$  we would query the GPR for predictions of the Milky Way values at  $\log SFR - \sigma_{SFR}/10$ , the nominal  $\log SFR$ , and  $\log SFR + \sigma_{SFR}/10$ , while all other parameters (stellar mass, disk scale length, etc.) are held at the fiducial Milky Way values. This yields three total predictions, two with offset SFR values and one at the nominal SFR. We then use these three values to calculate the derivative using the three-point Lagrangian interpolation method. Because these derivatives are small, they are sensitive to the limited training sample size required for the scikit – learn implementation of GPR (refer to Section 4.3.2.2 for details). In order to mitigate this effect we repeat this process for ten different training sets, re-evaluating the GPR prediction at the offset physical parameter and nominal value each time and calculating the derivative again. The derivatives presented here for each parameter are then computed as the average of the ten derivative values.

#### 4.7 Testing the Accuracy of Our Gaussian Process Regression Methods

In order to have confidence in the results we obtain from the Gaussian process regression, it is necessary to test the accuracy of the predicted photometry and error estimates. As in

Property	Corrected Value	Bias Removed	$\partial/\partial(\log M_*)$	$\partial/\partial(\log {\rm SFR})$	$\partial/\partial(b/a)$	$\partial/\partial R_d$	$\partial/\partial(B/T)$	$\partial/\partial p_{\rm bar}$
$^{0}(FUV-r)$	$4.24\pm0.59$	$0.16\pm0.05$	1.70	-1.03	-1.50	-0.15	0.51	0.14
$^{0}(NUV-r)$	$3.60\pm0.38$	$0.16\pm0.05$	1.34	-0.85	-1.13	-0.12	0.76	0.09
$^{0}(u-r)$	$2.14\pm0.14$	$0.06\pm0.01$	2.02	-0.96	-0.58	-0.29	-0.19	-0.05
$^{0}(g-r)$	$0.73\pm0.04$	$0.05\pm0.01$	0.13	-0.13	-0.09	0.00	0.07	0.01
$^{0}(i-r)$	$-0.33\pm0.03$	$0.03\pm0.00$	-0.08	0.05	0.06	0.01	-0.03	0.00
$^{0}(z-r)$	$-0.61\pm0.06$	$0.04\pm0.01$	-0.16	0.09	0.14	0.01	-0.03	-0.02
$^{0}(J-r)$	$-0.89\pm0.11$	$0.02\pm0.01$	-0.24	0.05	0.16	0.03	0.03	-0.01
$^{0}(H-r)$	$-1.11\pm0.14$	$0.03\pm0.01$	-0.28	0.05	0.09	0.03	0.02	-0.03
$^{0}(Ks-r)$	$-0.78\pm0.15$	$0.02\pm0.02$	-0.29	0.05	0.20	0.03	0.03	-0.02
$^{0}(W1 - r)$	$-0.17\pm0.19$	$0.02\pm0.01$	-0.30	-0.12	0.44	0.09	-0.09	-0.03
$^{0}(W2 - r)$	$0.41\pm0.26$	$0.00\pm0.01$	-0.25	-0.19	0.67	0.11	-0.11	-0.03
$^{0}(W3 - r)$	$-0.30\pm0.48$	$0.42\pm0.04$	0.40	-1.18	0.54	0.12	0.06	0.07
$^{0}(W4 - r)$	$-1.01\pm0.70$	$0.06\pm0.03$	1.08	-1.54	0.84	0.23	-0.19	-0.38

Table 4.2: X - r color estimates for each photometric band X used in this work; these values are used to calculate the luminosities presented in Table 4.1, and have been determined by the methods described in Section 4.3. The corrected value column refers to our final estimate after accounting for Eddington bias (cf. Section 4.9). colors presented here are in the AB magnitude system.

Property	Corrected Value	Bias Removed	$\partial/\partial(\log M_*)$	$\partial/\partial(\log {\rm SFR})$	$\partial/\partial(b/a)$	$\partial/\partial R_d$	$\partial/\partial(B/T)$	$\partial/\partial p_{\rm bar}$
$^{0}(u-g)$	$1.49\pm0.11$	$0.12\pm0.01$	0.65	-0.47	-0.40	-0.15	-0.35	-0.02
$^{0}(r-i)$	$0.30\pm0.03$	$0.01\pm0.00$	0.07	-0.04	-0.05	-0.01	0.02	0.00
$^{0}(i-z)$	$0.27\pm0.03$	$0.01\pm0.00$	0.06	-0.03	-0.11	0.00	0.01	0.00
$^{0}(U - V)$	$1.11\pm0.12$	$0.13\pm0.01$	0.42	-0.41	-0.24	0.08	-0.25	0.02
$^{0}(U - B)$	$0.30\pm0.08$	$0.12\pm0.02$	0.69	-0.28	-0.39	0.04	-0.68	0.09
$^{0}(B - V)$	$0.80\pm0.05$	$0.04\pm0.01$	0.17	-0.16	-0.08	-0.01	0.08	0.01
$^{0}(V - R)$	$0.54\pm0.03$	$0.00\pm0.00$	0.06	-0.04	-0.06	0.00	0.03	0.01
$^{0}(R - I)$	$0.60\pm0.04$	$0.01\pm0.00$	0.08	-0.06	-0.09	-0.01	0.02	0.01
$^{0}(W1 - W2)$	$-0.57\pm0.08$	$0.00\pm0.01$	0.03	0.10	-0.07	-0.03	0.01	0.02
$^{0}(W2 - W3)$	$0.91\pm0.33$	$0.00\pm0.02$	-0.33	0.92	-0.08	-0.02	-0.38	-0.02
$^{0}(W3 - W4)$	$0.40\pm0.49$	$0.02\pm0.01$	-0.65	0.09	0.25	-0.02	0.21	0.34

Table 4.3: Additional color estimates for the Milky Way, updating values from Table 1 and Table 3 of Ref. [219] in addition to values for WISE colors (as plotted in Fig. 4.8). The SDSS ugriz and WISE W1-W4 values are presented in AB magnitudes. The Johnsons-Cousins UBVRI bands have been converted from the ugriz measurements via KCORRECT on an object-by-object basis, as in Ref. [219], and are presented in the Vega magnitude system.

Property	Corrected Value	Bias Removed	$\partial/\partial(\log M_*)$	$\partial/\partial(\log \mathrm{SFR})$	$\partial/\partial(b/a)$	$\partial/\partial R_d$	$\partial/\partial(B/T)$
$^{0}M_{FUV} - 5logh$	$-16.34 \pm 0.66$	$0.37\pm0.06$	-0.04	-1.31	-1.84	-0.27	0.71
$^{0}M_{NUV} - 5logh$	$-17.12\pm0.49$	$0.25\pm0.06$	-0.23	-1.12	-1.52	-0.26	0.48
$^{0}M_{u} - 5logh$	$-18.53\pm0.33$	$0.24\pm0.03$	-1.69	-1.14	-1.08	-0.35	0.36
$^{0}M_{g}-5 logh$	$-19.91\pm0.24$	$0.17\pm0.02$	-1.24	-0.52	-0.7	-0.15	-0.04
$^{0}M_{r}-5logh$	$-20.68\pm0.53$	$0.23\pm0.03$	-1.29	-0.52	-0.52	-0.14	-0.17
$^{0}M_{i} - 5 logh$	$-20.95\pm0.21$	$0.10\pm0.02$	-1.34	-0.41	-0.35	-0.16	0.01
$^{0}M_{z}-5logh$	$-21.34\pm0.24$	$0.14\pm0.03$	-1.48	-0.38	-0.33	-0.14	0.13
$^{0}M_{U} - 5logh$	$-19.47\pm0.39$	$0.24\pm0.03$	-0.48	-1.02	-1.11	-0.36	-0.59
$^{0}M_{B} - 5 logh$	$-19.59\pm0.25$	$0.16\pm0.02$	-1.15	-0.65	-0.58	-0.16	0.00
$^{0}M_{V}-5logh$	$-20.48\pm0.24$	$0.09\pm0.02$	-1.24	-0.50	-0.46	-0.18	-0.06
$^{0}M_{R}-5logh$	$-20.91\pm0.21$	$0.04\pm0.02$	-1.34	-0.46	-0.53	-0.18	-0.09
$^{0}M_{I} - 5 logh$	$-21.65\pm0.2$	$0.02\pm0.02$	-1.28	-0.42	-0.29	-0.13	0.06
$^{0}M_{J}-5logh$	$-21.65\pm0.20$	$0.01\pm0.02$	-1.72	-0.40	-0.35	-0.12	-0.11
$^{0}M_{H} - 5logh$	$-21.75\pm0.21$	$0.06\pm0.02$	-1.85	-0.39	-0.18	-0.10	-0.10
$^{0}M_{Ks} - 5 logh$	$-21.56\pm0.20$	$0.07\pm0.03$	-1.87	-0.36	-0.15	-0.10	0.02
$^{0}M_{W1} - 5logh$	$-21.80\pm0.30$	$0.03\pm0.01$	-2.31	-0.3	-0.04	0.01	0.04
$^{0}M_{W2} - 5logh$	$-21.19\pm0.30$	$0.00\pm0.01$	-2.34	-0.32	0.01	0.03	0.02
$^{0}M_{W3} - 5logh$	$-22.21\pm0.48$	$0.03\pm0.05$	-1.73	-1.28	0.64	0.05	0.28
$^{0}M_{W4} - 5logh$	$-22.39\pm0.67$	$0.01\pm0.04$	-0.75	-1.34	0.49	0.10	-0.01

Table 4.4: Absolute magnitude estimates for the Milky Way, updating Table 1 and Table 3 of Ref. [219] with additional absolute magnitude estimates beyond the optical. GALEX UV, SDSS ugriz, 2MASS J - Ks and Wise W1 - W4 magnitudes are all presented in the AB system. The Johnsons-Cousins UBVRI magnitudes are presented in the Vega system.

any supervised machine learning problem, we want to ensure that the given inputs map correctly to the given outputs. In order to enable an unbiased evaluation of our model we must split our data into separate training and test samples, as we do not want to evaluate a model with information that was used to train it. To do this we use the **scikit – learn** function **sklearn.model\_selection.test\_train\_split**. We use a split of 75% of our data for training and 25% of the data for testing. We do not construct a separate validation set as we do not have many hyper parameters to tune (the kernel defaults work well for our case, as discussed in Section 4.3.2.1).

For each photometric property that we study, we train the GPR using only the physical and photometric properties of the galaxies within the training sample. With that trained model we then make predictions for a given photometric property using the testing sample's physical properties. This gives us predicted absolute magnitudes and colors for every galaxy in our testing sample, which we can then compare to their actual observed values.

Fig. 4.13 depicts the results of this training and testing analysis for all of the bands relevant to constructing the SED. We present histograms based on the differences between the intrinsic value of a photometric property and the predicted value from the GPR,  $\Delta$ , for each galaxy in the testing sample. We then divide these  $\Delta$  values by the standard deviation predicted by the GPR for that galaxy,  $\sigma$  to obtain the normalized residual,  $\frac{\Delta}{\sigma}$ . By focusing on this quantity we can assess whether a bias in a given galaxy is present and how large it is compared to the prediction uncertainties, and can additionally evaluate whether the error estimates from the GPR are too small (too optimistic) or too large (too conservative) compared to actual deviations.

In each histogram, we overlay a Gaussian with a mean of zero and a variance of one in gray. One would expect the distribution of  $\frac{\Delta}{\sigma}$  to follow this distribution if there is no bias in the predictions and the error values are all accurate. We also list the means and medians of  $\frac{\Delta}{\sigma}$  for each band on the corresponding histogram. Some quantities exhibit a small skew, but invariably the mean offsets are  $< 0.05\sigma$  and the median offsets are  $< 0.2\sigma$ , so any bias is far subdominant to other sources of errors. In general the distribution of  $\frac{\Delta}{\sigma}$  is tighter than the expected Gaussian, indicating that the  $\sigma$  estimates from GPR are slightly too large (i.e., overly conservative).

In order to assess what may cause the small offsets, we have evaluated the histograms of  $\frac{\Delta}{\sigma}$  separately for face-on galaxies (which we take to be those with axis ratio b/a > 0.5) and more inclined objects ( $b/a \leq 0.5$ ). We find that most of the slight skewing that is observable in Fig. 4.13 is driven by the galaxies with higher inclinations. In this regime galaxy color can change quickly and non-linearly with b/a, which can explain why the GPR struggles to accurately predict colors for these galaxies.

Overall, we find that the GPR performs well at mapping from galaxy physical parameter inputs to photometric properties, with biases that are much smaller than the predicted uncertainties, and error estimates that err on the conservative side. We therefore can use the method with confidence to make predictions for the photometric properties of the Milky Way.

## 4.8 K-corrections for WISE Bands

#### 4.8.1 Calculating the *K*-correction

When we observe extragalactic objects their spectrum is redshifted into the "observed" reference frame. In order to convert from observed flux to rest-frame values, astronomers utilize K-corrections, which account for the shifting of rest-frame light into different passbands as redshift increases. While most of the photometry that we have used in this work (as described in Section 4.2.1) has readily available K-corrections into rest-frame bands, measurements in the WISE bands [383] do not.

Typically, *K*-corrections for a given band are calculated by calculating the effect on observed colors of redshifting a theoretical or empirical spectral energy distribution (SED) template; existing software packages such as KCORRECT Ref. [31] have been designed to perform this task. However, this approach is not feasible at wavelengths where SED templates are not available or poorly constrained.

For this reason, in the WISE bands we adopt a new, almost purely empirical approach which builds on the template-based methods presented in Ref. [16]. Our K-correction tech-



Figure 4.13: Histograms depicting the difference between the actual color or absolute magnitude of galaxies in an independent test sample from the value predicted by GPR, normalized by the predicted error resulting from the GPR fit  $(\frac{\Delta}{\sigma})$ . For reference we have overlaid a Gaussian with a mean of zero and variance of one in each panel; in the ideal case, the normalized residuals would follow this distribution. For reference we also provide the means and medians of  $\frac{\Delta}{\sigma}$  for each panel. The  $\frac{\Delta}{\sigma}$  distribution for each property matches the expectation well, with biases of  $< 0.2\sigma$  and RMS residuals slightly smaller than would be expected from the predicted errors. Most of the minor skewing is attributed to galaxies with small axis ratios, for which color changes quickly with inclination. Results for all quantities necessary for deriving the Milky Way SED are shown here. We can conclude that the GPR is able to accurately map from galaxy physical properties to colors and absolute magnitudes.

niques are detailed in Chapter 5. While we do not go into detail here, we summarize the strategy below.

We rely on the assumption that the K-correction needed for a given band is a simple (polynomial) function of a galaxy's rest-frame color determined in some pair of bands; i.e., that galaxy SEDs constitute a one-parameter family. This color can be determined in the optical where more conventional K-correction methods are effective; we just need some way to sort SEDs into a one-parameter family. The assumption that one parameter is sufficient is not perfect, as there is some diversity in SEDs at fixed rest-frame color, but it is good to first order as variations in dust, specific star formation rate, and metallicity all have coarsely similar effects on the spectrum of a galaxy. Because we are focused on K corrections over a limited redshift range near z = 0, getting things right to this level is sufficient to make K-correction errors far subdominant to other sources of uncertainty in the WISE bands.

Our goal is to produce a function that takes as inputs the apparent magnitude in some band, the redshift, and the rest-frame color in some pair of bands (which can be determined by conventional K-correction methods, we will use  ${}^{0}(g-r)$  for this example), and returns the rest-frame absolute magnitude in the desired band. We determine this function in two steps. First, we determine polynomial relationships between the *observed* color in a pair of bands (one of which serves as the anchor, and one of which is the desired target band) and redshift, in bins of the rest-frame color in another (e.g., optical) pair of bands which is provided as one of the inputs. This relationship is determined via a second order polynomial fit; i.e., we parameterize the observed color  $(X - r)_{observed} = f(z) = a_0 + a_1 z + a_2 z^2$ . For each bin in rest-frame color, we determine the fit coefficients separately. This is similar to the technique applied in Ref. [50]; see, e.g., Fig. 2 of that work.

We then fit for the linear dependence of each of the fit coefficients on the mean rest-frame color of a bin; i.e., we now treat the coefficients as functions  $a_1({}^0(g-r))$  and  $a_2({}^0(g-r))$ . We ignore the value of  $a_0$  as it corresponds to the predicted mean rest-frame color at z = 0; rather than assuming this mean color is appropriate for all objects, we will instead only make small adjustments to the observed photometry. To do this, we first fit for  $a_1({}^0(g-r))$  and  $a_2({}^0(g-r))$  via a linear regression algorithm which is robust to outliers, specifically the Huber regression method implemented by scikit – learn (sklearn.linear\_model.HuberRegressor). We

assume a linear dependence between each coefficient and the rest-frame reference color ( $^{0}(g - r)$  here).

The result of this process is a function that predicts the observed color in a new pair of bands, as a function of the rest-frame color in some other pair of bands and the redshift. At z = 0, the observed color and rest-frame color must agree, by definition. Hence, the *difference* between the prediction for the observed color for an object at a given redshift, and the prediction for the observed color at z = 0 (which corresponds to the value of  $a_0(^0(g - r))$ , must be identical to the K-correction for the target color. If one subtracts off the K-corrected reference band (which again can be an optical band where K-corrections are already well-determined), one is left with the K-corrected unknown band on its own.

We apply this (small) correction to the *observed* photometry to map to the rest-frame equivalent (i.e., we do not simply adopt the color in the new bands predicted from the restframe color in known bands – which could entail large shifts from the observed photometry for some objects – but rather we only make use of the small *offset* between observed and rest-frame colors that was fit across the sample). To describe the process symbolically, we use the rest-frame  ${}^{0}(g-r)$  color for an object to obtain the  $a_1$  and  $a_2$  values for the mapping of color in some band X relative to r, X - r. Then we can obtain the rest-frame color in the new band  ${}^{0}(X - r) = (X - r)_{\text{observed}} - a_1 z - a_2 z^2$ . Finally, we can determine the rest-frame absolute magnitude in band X via the expression  ${}^{0}M_X = {}^{0}M_r + {}^{0}(X - r)$ .

The result of this process is a simple but effective empirical method for obtaining Kcorrections for bands where SED templates are not well known, which can be applied so long
as K-corrections have already been determined in another set of better-characterized bands.
In this work, we use  ${}^{0}(g - r)$  color to determine the K-correction mappings as a function of
rest-frame color and redshift, and WISE-r colors as the target in each case. We have selected
these bands as g and r both have small photometric uncertainties and well-characterized Kcorrections. We will present more details of our procedures and tests of their effectiveness in
Ref. [123].

When evaluating the fits for the the K-corrections we exclude objects that have large WISE photometric errors in the W1 and W2 channels. We do not place requirements on W3 or W4 errors as they are invariably large. Specifically, in W1 we perform the K-correction fits
restricting to galaxies with errors  $\sigma_{m_{W1}} < 0.125$ , while in W2 we require errors  $\sigma_{m_{W2}} < 0.25$ . Objects with larger errors *are* still included when performing our GPR analyses.

# 4.8.2 Photometric Offsets Between SDSS and DESI Legacy Survey Imaging

As described in Section 4.2.1.3, we rely on Legacy Survey catalogs for WISE photometry, as the *Tractor*-based measurements have lower uncertainties than other public catalogs. Because of the use of a matched object model across all passbands, colors in WISE bands relative to optical bands will be determined with greater accuracy than total magnitudes in a single band. However, because the filters and instruments used for the Legacy Surveys have different transmission and response curves and photometry was performed with differing analysis pipelines (BASS g and r and MOSAIC-3 z filters were used in the northern portion of the Legacy Survey footprint, while in the south DECam was used for g, r, and z), the rest-frame r band absolute magnitudes for a given object should differ between the SDSS and Legacy data.

To calculate the WISE minus SDSS r band colors used to construct the Milky Way SED, we therefore must take the WISE minus Legacy Survey r band color (which should be measured self-consistently due to the use of a common model) and apply a small correction to compensate for the differences between Legacy r and SDSS r. To do this we proceed as follows:

- Calculate absolute magnitudes for DESI Legacy bands We calculate rest-frame absolute magnitudes for all objects in the cross-matched catalog in the BASS and DECam r and SDSS g and r bands, using the KCORRECT [31] software. We base these calculations only on the SDSS ugriz photometry for our galaxy sample.
- 2. Calculate the filter offsets in r We now wish to determine the offsets between restframe r absolute magnitude in Legacy Survey and SDSS filters, as a function of rest-frame color:  $\Delta_r = M_{r,BASS} - M_{r,SDSS}$  for the North, or  $\Delta_r = M_{r,DECam} - M_{r,SDSS}$  for the South. In Fig. 4.14 we plot  $\Delta_r$  as a function of SDSS  ${}^{0}(g-r)$  for our sample. Almost all objects fall along a linear relationship between the r-band offset and color.
- 3. Fit for offsets as a function of  ${}^{0}(g-r)$  We perform a robust least-squares fit using the

scikit – learn [286] Huber regression function with  ${}^{0}(g-r)$  as the independent variable and  $\Delta_r$  as the dependent variable. In Fig. 4.14 we plot these fits with black dashed lines.

Using the coefficients resulting from this fit, we can convert a rest-frame color referenced to Legacy Survey r (e.g.,  ${}^{0}(W1 - r_{\text{DECam}})$  to one referenced to SDSS r ( ${}^{0}(W1 - r_{\text{SDSS}})$ , for this example) by evaluating the fit line at the rest-frame  ${}^{0}(g - r)$  color for a given object. Our GPR predictions for WISE absolute magnitudes and WISE-r colors require use of this correction; however, for WISE colors such as (W1-W2) the dependence on the offset cancels out.

We apply this correction in tandem with calculating the K-corrected WISE colors, before training the Gaussian process regression. In the previous sub-section we expressed our calculation of absolute magnitudes as  ${}^{0}M_{X} = {}^{0}M_{r} + {}^{0}(X - r)$ . Re-written to explicitly use color relative to the SDSS r band, this expression becomes

$${}^{0}M_{X} = {}^{0}M_{r_{\rm SDSS}} + {}^{0}(X - r_{\rm SDSS}) = {}^{0}M_{r_{\rm SDSS}} + ({}^{0}M_{X} - {}^{0}M_{r_{\rm SDSS}}).$$
(37)

However, we have measured the  $X - r_{\text{Legacy}}$  color, not  $X - r_{\text{SDSS}}$ . In order to account for the offset between Legacy and SDSS photometry we therefore require the correction derived above:

$${}^{0}M_{X} = {}^{0}M_{r_{\rm SDSS}} + ({}^{0}M_{X} - {}^{0}M_{r_{\rm Legacy}}) + ({}^{0}M_{r_{\rm Legacy}} - {}^{0}M_{r_{\rm SDSS}}),$$
(38)

where  $({}^{0}M_{r_{\text{Legacy}}} - {}^{0}M_{r_{\text{SDSS}}})$  is the offset between rest-frame SDSS r and DESI Legacy Survey r, or  $\Delta_r$  above. Applying this correction before performing further analysis allows us to train the GPR based upon the combined North and South DESI Legacy catalogs rather than training with each separately.

The magnitude of the offsets applied to the WISE photometric band estimates are small, spanning from -0.14 to 0.05 in magnitude. Compared to the photometric errors of these bands in Table 4.2 and Table 4.4, the errors on the photometry far outshine any error attributed to the DESI Legacy - SDSS offset (especially for W3 and W4).



Figure 4.14: Offsets between SDSS and DESI Legacy Survey rest-frame r-band absolute magnitude, plotted as a function of SDSS rest-frame color,  ${}^{0}(g-r)$ . The upper panel shows objects in the northern portion of the Legacy Survey, which utilized the BASS filters and instrument in r. The lower panel shows objects in the southern Legacy region, where DECam was used. The dashed black line depicts our linear fit for  $\Delta_r$  as a function of  ${}^{0}(g-r)$ . This fit follows the locus of objects well. With this fit we can obtain the offsets between colors measured relative to Legacy Survey and SDSS photometry and subtract those offsets to construct predicted WISE-based colors.

#### 4.9 Eddington Bias Corrections

Uncertainties in stellar mass, star formation rate, disc scale length, etc. can lead to biases in the inferred photometric properties of the Milky Way. When the distribution of objects in parameter space varies non-linearly, scatter from errors will preferentially move objects from more sparsely populated regions to rarer locations compared to the opposite situation. This can lead to an "Eddington" bias in inferred properties. For example, because of the rarity of massive galaxies, a galaxy with a large stellar mass measurement is more likely to have an intrinsic stellar mass that is smaller than the measured value than one that is larger, as there is a much larger number of objects that can up-scatter compared to the number that down-scatter. As a result, the stellar mass measurements used when training the Gaussian process regression would then be biased high. Similar effects can occur with star formation rate, disk scale length, etc.

We quantify this bias through an empirical approach based upon the methods presented in Ref. [219]. We perturb the measured values of the galaxy sample used to train the GPR with Gaussian noise sampled from the measured errors. Specifically this means we repeatedly add to each galaxy's measured  $M_*$ , SFR, B/T, etc. a value randomly drawn from a Gaussian distribution centered at zero with a standard deviation of that object's error in the given quantity (e.g.,  $\sigma_{\log M_*}$ ), and determine the effects that adding errors to each quantity has. We make the simplifying assumption that the bar vote fraction should not significantly contribute to the Eddington bias (as uncertainties in that quantity are difficult to characterize, but also should have limited effects on photometry).

To obtain predicted photometry for a given noise level, we run the GPR in a similar manner to the methods used in Section 4.3, except we now train the regression on the perturbed sample before predicting photometric quantities. We then predict the photometric quantity of interest by evaluating each model assuming the fiducial value for each Milky Way physical property (i.e., we use a fixed set of values for all evaluations rather than sampling from the MW distributions). In our application, we perform this procedure on samples with Gaussian noise applied to each physical property from one to four times successively.

In the original cross-matched training set, the values for each physical property have been

perturbed from their true, intrinsic values due to measurement errors; we define this case as corresponding to having noise applied n = 1 times. However, we wish to evaluate what the measurement would be if there were no noise; i.e., for the case where n = 0. We therefore wish to characterize the difference in the property of interest between the case where noise has been applied n or n-1 times, and evaluate for n = 1 to determine the correction needed to remove Eddington bias.

To determine this value we perform use a least squares quadratic fit to the set of differences between GPR predictions as we add more noise: i.e., we fit a relation of the form  $P_n - P_{n-1} = An^2 + Bn + C$ , where  $P_n$  is the prediction from GPR when noise has been applied a total of n times and A, B, and C are coefficients of the fit. We use the resulting fit to extrapolate to  $P_1 - P_0$ , which should correspond to the effect of Eddington bias. The mean Eddington bias at n = 1 is then quantified as the sum of the coefficients A + B + C from the quadratic fit of the set of differences  $P_n - P_{n-1}$ . In practice, we perform the entire analysis (adding random noise repeatedly, training the GPR, fitting for  $P_n - P_{n-1}$ , and adding the coefficients) 25 times. The effect of Eddington bias for a given band is then calculated as the mean of the set of 25 values. We then subtract this value from the GPR-predicted mean for a given photometric property of the Milky Way to obtain a corrected value. The bias removed in each of our predictions is documented in the third column of Table 4.2-Table 4.4.

The uncertainty in each of the 25 Eddington bias estimates (i.e., in the quadratic function evaluated at the point n = 1) can be calculated via the square root of the sum of the covariance matrix for the coefficients of the least-squares quadratic fit (as when n = 1 the quadratic result is simply A + B + C). The uncertainty in the Eddington bias in each band is therefore equivalent to  $\sigma_{\text{bias}} = \frac{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}{n}$ , where  $\sigma_i$  is the estimated uncertainty from the i'th sample and here n = 25. As a check we have also computed errors via the standard error on the mean Eddington bias and find comparable (though somewhat more optimistic) results. The Eddington bias uncertainty is combined in quadrature with the uncertainty estimated from sampling the GPR to produce the errors on the corrected values in the tables in Section 4.6.

## 5.0 Empirically Driven K-corrections At Low Redshift

This chapter is a draft with the intention to submit for peer-review and publication in the Monthly Notices of the Royal Astronomical Society.

# 5.1 Introduction

Broad band luminosity measurements are critical for understanding galaxy evolution, but require correcting for the redshifting of light across photometric band passes. The redshifting of a galaxy spectra is equivalent to shifting the filter transmission curve through which the galaxy is observed. Hence two galaxies with identical spectral energy distributions (SEDs) but at different redshifts will usually have different fluxes in the same observed band pass. This difference is referred to as the K-correction [173, 283]. K-corrections are especially important for comparing populations of galaxies at varying redshift, as we want equivalent measurements of the SEDs of these galaxies. K-corrections serve as a critical equalizer in order to make practical comparisons among various galaxy surveys across many redshift ranges. For objects where the entire spectrum has been observed, K-corrections are straightforward to calculate, as the spectrum can be shifted to accommodate redshift corrections. However, such results cannot always be trusted since spectra are typically of central galactic regions, so the appropriate K-correction for integrated photometry will likely differ. Considering we have yet to enter the era of easily available spatially resolved spectroscopy in wide-field surveys, we must depend on photometric K-corrections.

To determine the intrinsic brightness of an object, or the brightness in the rest-frame, we use a transformation that acts upon the observed-frame photometry. The K-correction between a bandpass R used to observe a galaxy at redshift z and rest bandpass Q is defined by [283, 169, 31]

$$M_Q = m_R - DM(z) - K_{QR}(z) + 5\log h.$$
(39)

Wherein  $M_Q$  is the absolute magnitude,  $m_R$  is the observed apparent magnitude, DM(z) is

the distance modulus,  $K_{QR}(z)$  is the K-correction between band R and band Q, and h is the Hubble constant. The distance modulus is defined as  $DM(z) = 5 \log(\frac{d_L}{10pc})$  where  $d_L$  is the luminosity distance. For brevity we do not include the full and rigorous formalism for the generalized definition of the K-correction, which is presented in Ref. [169] (see Eq. 8 and 10).

The key ingredients for determining K-corrections for a given galaxy are (1) some knowledge of the spectrum of the object, which characterizes the flux emitted by the galaxy across a large wavelength range and (2) the filter curves of the instrument used to make the observations, because the transmission function of each filter affects the collected photons slightly differently. Oftentimes we do not have much information on the full spectrum of the objects we observe, instead we merely have a handful of photometric points that sample various parts of the spectrum. This requires us to either use analytical approximations to quantify how the observed galaxy's color changed with redshift, or to utilize templates to reconstruct the observed galaxy's SED.

Historically, template matching has been the oft used approach. There are two families of SED templates that are used in this matching to determine K-corrections empirical templates derived from a representative set of real galaxy spectra (e.g., Ref. [74, 196, 50]), and synthetic templates derived from theoretical stellar population synthesis (SPS) models [51, 239]. SPS models characterize star formation and evolution within a galaxy to construct a realistic SED. Therefore theoretical templates, when used in combination with dust attenuation models, have the benefit of providing other physical quantities like estimates of stellar mass and star formation rate in addition to K-corrected photometry.

There are a number of ways through which galaxy K-corrections have been calculated in the literature. We briefly review K-correction methods here, but this list is not exhaustive. We then discuss the caveats to using these methods.

Early work modeled K-corrections as a simple function of redshift and galaxy morphological type, such as in the work by Ref. [74]. The authors used empirical SED templates for four morphological galaxy types (E, Sbc, Scd and Irr galaxies) to obtain K-corrections using the best data available at the time. However, this method is somewhat oversimplified based on our current understanding of galaxy evolution, as galaxies that may share the same morphology can have a substantially diverse sets of spectral properties [354].

It is more common to obtain K-corrections by modeling galaxy SEDs as a function of wavelength. Using the observed colors of the galaxy in question, it is matched in some way (typically a fit) to a template SED that has been shifted to the observed galaxy's redshift. Then the template is used to characterize the SED of the source galaxy. Maximum likelihood fitting of photometry to theoretical SPS based SED models has been often utilized to calculate K-corrections such as in work by e.g., Ref. [20] and Ref. [48]. As mentioned above, SPS models make reasonable approximations of real galaxy SEDs. However, while this method works well for red galaxies, blue galaxies are not as tightly constrained due to effects of star formation, dust attenuation, and nebular emission lines which have degenerate effects on galaxy SEDs. Even worse, SPS models alone may struggle to capture the full diversity of galaxy SEDs. Ref. [30] introduced K-correction calculations derived from matching photometric observations to templates constructed from *combinations* of Ref. [51] SPS models, dust attenuation models, and nebular emission lines, which resulted in more realistic SEDs. This was the first instance of using multiple components to construct galaxy SEDs, which has served as the backbone of the commonly used KCORRECT V4.3 software of Ref. [31]. KCORRECT has become the industry standard for determining K-corrections and will serve as a comparison for our own results.

More recently, K-corrections have been estimated with analytical functions of redshift, which are parameterized with a property that characterizes the galaxy. Ref. [382] fit second order polynomials to the K-corrections as a function of rest-frame color with the Ref. [196] empirical templates. Because their work was limited to three bands (*BRI* in DEEP2, the Deep Evolutionary Exploratory Probe; [84]), the authors were unable to use the fitting methods previously discussed such as that by Ref. [30] – three bands are not enough to constrain the possible SED templates used for deriving K-corrections owing to the lack of DEEP2 filter availability in KCORRECT at the time. Ref. [70] empirically showed that K-corrections can be approximated as solely a function of redshift and a single color, using 190,275 observed galaxies spanning the redshift range 0.03 < z < 0.6. The authors derived K-corrections defined by polynomials that are fifth order in observed color (they used (g - r)) and third order in redshift for nine filters (ugrizYJHK), using both KCORRECT and PEGASE.2 [126] models (a type of stellar evolution model used in constructing synthetic SEDs). Work by Ref. [284] takes a similar approach, deriving a linear relationship between the K-correction, (g-r) color, and redshift for a number of galaxies of SDSS DR7 that were processed by KCORRECT v4.3. Ref. [16] did a more in-depth study on which observed colors best characterise each optical band (ugriz), deriving second order polynomials in color from the 129 empirical SEDs from the Ref. [50] Atlas.

While K-corrections are derived reasonably well with templates there are also many limitations to using these templates. First, there are a limited number of template SEDs. As a result, small template numbers or limited template components may not span the full range of possible colors exhibited by a diverse galaxy sample. This is especially concerning for galaxies that have less commonplace SEDs which could result in inaccurate K-corrections due to template mismatches. Empirical templates may provide a more representative SED sample than theoretical templates. However, many of these templates only cover ultraviolet and optical wavelengths, lacking infrared coverage. Additionally, many of the older commonly used empirical templates such as those from Ref. [74] and Ref. [196] are active galactic nuclei (AGN) spectra, which can lead to systematic errors due to the AGN outshining much of the galaxy with strong emission lines and blue continua. A number of these older empirical templates were also based on the centers of bright galaxies. Ref. [382] found that the use of these templates (a portion of the [196] templates) resulted in calculations too red by 0.08 magnitudes. Empirical templates are also still limited by the galaxies for which robust observations have been obtained, meaning some galaxy families do not have well observed SEDs. Theoretical SEDs may also cover a limited wavelength range for the same reason. Methods that employ model based templates are also limited by the quality of the models used.

Regardless of which family of template is used, depending on templates also limits our ability to determine K-corrections at wavelengths where the templates are not available, poorly constrained, or models are lacking. For example, the WISE bands probe the interstellar dust content of galaxies due to their sensitivity to polycyclic aromatic hydrocarbon emission features, the small-grain dust continuum, and the thermal emission tail of larger dust grains [384], which makes them an important component of a galaxy's SED and characteristics. There have been few options for determining WISE K-corrections other than with KCORRECT. Additionally, while KCORRECT can calculate K-corrections for the WISE photometric bands, these fits do *not* employ modelling for dust emission features, which are increasingly important for bands W2 through W4. The lack of this feature can result in incorrect K-corrections.

In this work we have developed a data driven approach to determining K-corrections that significantly limits the extent to which templates are relied upon and circumvents the need for templates in less constrained filters. Conceptually our method is somewhat similar to that of Ref. [382], Ref. [284] and Ref. [50]. We still utilize the approximation that the K-correction needed for a given band is a simple (polynomial) function of a galaxy's redshift and a quantity that characterizes the galaxy. However, in this work we use rest-frame color determined in some other pair of bands. This rest-frame color is determined from templates where SEDs are exceptionally well constrained, such as the optical rest-frame (g - r) band. Operating under the assumption that at low redshift galaxy SEDs fall into a single parameter family, we show that with a single color we can anchor the SED and interpolate all other K-corrected colors. With this approach, as long as one has access to a rest-frame color that has been K-corrected already, one can determine rest-frame colors for any other bands. We caution that this assumption generally only applies at low redshift (z < 1) and becomes much more complex at higher redshift, where the assumption that SEDs fall into a single parameter family [77, 231, 230] breaks down. We presented a cursory version of this approach in Ref. [125] in order to determine our rest-frame WISE photometry and expand upon the method in this work.

This chapter is organized as follows. In Section 5.2.1, we briefly described the data used in our calculations: spanning from GALEX, SDSS, 2MASS, and WISE detections. Section 5.2.3 details our method for determining K-corrections. In Section 5.3 we compare our computed rest-frame colors from our K-corrections to literature values. Last, we summarize and discuss our findings in Section 5.4.

All magnitudes and colors are presented in the AB system in this work. Absolute magnitudes are determined using a Hubble constant  $H_0 = 100 \text{ km s}^{-1} \text{Mpc}^{-1}$ , so they are equivalent to  $M_y - 5 \log h$  (where  $M_y$  is the y-band absolute magnitude and  $h = H_0/100$ ) for other values of h. Photometry is presented from adopting the notation used in Ref. [31] and Ref. [219], where an absolute magnitude of passband y at redshift z is denoted as  ${}^{z}M_{y}$ .

# 5.2 Methods

#### 5.2.1 Observational Data

The data used in this study originates from the GALEX-SDSS-WISE-Legacy Catalogue 2 (GSWLC-2) of Ref. [312, 311], within the redshift range of 0.01 < z < 0.09. This data spans across the filter range of the GALEX ultraviolet survey FUV/NUV bands [241], the SDSS optical ugriz bands (DR10; [3]), the 2MASS near-infrared JHKs bands [330], and the WISE mid-infrared W1/W2/W3/W4 bands [384]. Specifically, the WISE data is from the "unWISE" reduction of Ref. [211] which is more appropriate for galaxies than the official pipeline. We use photometric data from the medium-deep GSWLC catalog, where the depth pertains to the UV photometry from GALEX, which unlike other bands is generally of much less uniform depth - going from shallow to very deep. While using the medium-deep instead of the shallow catalog decreases the number of galaxies at our disposal by roughly one half, the increased signal-to-noise in the UV imagining is a worthwhile trade-off.

For objects within the GSWLC-M2 catalog we have both the observed photometry (and associated errors) in addition to rest-frame values from the UV to the near IR (excluding WISE). These rest-frame values were obtained via fits to SED models, as described in Ref. [312], which we summarize in Section 5.2.2. For convenience we convert the observed flux  $(f_{\nu})$  in Jy to AB magnitudes using the relation  $m_{AB} = -2.5 \log_{10} (f_{\nu}) + 8.90$ . As a result all magnitudes presented in this work are in the AB system.

In order to keep our galaxy sample at a maximum we split our data into four different catalogs: (1) a catalogs consisting of the SDSS ugriz optical bands, (2) a catalog consisting of the GALEX FUV/NUV bands and the ugriz optical bands which we call UV+optical, (3) a catalog consisting of the 2MASS J/H/Ks near-IR bands and the ugriz optical bands which we call near-IR+optical, and (4) a catalog consisting of the WISE W1/W2/W3/W4 mid-IR

bands, the J/H/Ks near-IR bands, and the *ugriz* optical bands which we call IR+optical. This way we can ensure that results from runs of KCORRECT are not driven by a specific set of bands. For example, an IR prediction can be entirely driven by a fit in the UV+optical depending on the signal-to-noise of the measurement, particularly because extrapolation is common for IR SED models.

Before we take further steps in our analysis we exclude all objects that have photometric values of "NaN", infinity, or 0, which all indicate missing photometry. When evaluating the fits for the K-corrections we use a limited redshift range 0.04 < z < 0.09 in order to avoid selection effects at low redshift (0.01 < z < 0.04), but compute K-corrections for the full redshift range of our sample. For the fits in the FUV/NUV bands we also exclude objects that have large GALEX photometric errors. Specifically, in FUV we perform the K-correction fits restricting to galaxies with errors  $\sigma_{m_{FUV}} < 0.12$ , while in NUV we require errors  $\sigma_{m_{NUV}} < 0.1$ . We apply similar cuts for objects that have large WISE photometric errors. These correspond to  $\sigma_{m_{W1}} < 0.013$ ,  $\sigma_{m_{W2}} < 0.025$ ,  $\sigma_{m_{W3}} < 0.15$ , and  $\sigma_{m_{W4}} < 0.25$ . These cuts were determined by examining distributions of the errors and how they biased the K-correction results. Before analyzing our results we do a final clean to remove any "NaN" or infinity from each catalog for the rare cases that an object did not have a proper fit in KCORRECT or our own analysis. Within our final wavelength separated catalogs the optical catalog contains 148,704 galaxies, the UV+optical catalog contains 28,318 galaxies. the near-IR+optical catalog contains 74,038 galaxies, and the IR+optical catalog contains 40,361 galaxies. Each spans from redshift 0.01 < z < 0.09. The catalogs containing our K-corrected results and comparison values are publicly available at the catalog section of our catalog GitHub page.<sup>1</sup>.

### 5.2.2 K-corrections from Previous Work

We will be comparing our results from our fit calculations (described in the following subsection Section 5.2.3) to K-corrections from two other sources as a means to validate our results.

<sup>&</sup>lt;sup>1</sup>https://github.com/cfielder/K-corrections/tree/main/Catalogs

1. We generated K-corrections with the KCORRECT v4.3 software package Ref. [31]. KCOR-RECT is a robust method for determining K-corrections using a combination of SED templates with SPS models, dust models, and nebular emission models. KCORRECT uses a non-negative matrix factorization algorithm that creates model based template sets. A set of five template SEDs are used, which are derived from combinations of 450 [51] SPS models across a wide range of age and metallicity, and 35 ionized gas emission models from Ref. [193]. After the templates are generated and reduced into 5 sets, linear combinations of these templates are fit to the measured photometry and respective photometric errors for each galaxy for which K-corrections are being calculated with a  $\chi^2$  minimization technique. This constructs an estimated SED for that galaxy. From the estimated SED K-corrections and other physical parameters can be determined.

We ran KCORRECT on the observed photometry and associated errors within the GSWLC-M2 catalog for bands from the UV to IR composed of combinations of subsets of the wavelength range as described in Section 5.2.1 in addition to the full wavelength range. KCORRECT can compute K-corrections for all filters considered in this chapter (FUVthrough W4). We added an additional error term in quadrature to our inverse variance maggies equivalent to 1% of the respective observed maggy for the given band to allow for more flexibility in the models.

2. GSWLC [312, 311] produces K-corrections by performing SED fitting using model SEDs. For this catalog a Bayesian approach to SED fitting was employed ([313], see also Ref. [189, 46, 345] for the basis of this method) utilizing the CIGALE SED fitting code [38], which used a combination of [51] SPS models, dust models, and nebular emission models. In this approach millions of distinct models are considered individually, rather than utilizing a linear combination of a smaller number of models (templates) like that performed by KCORRECT. Additionally, the GSWLC fits employ more recently updated models. In contrast to the best-fit ( $\chi^2$  minimization) approach of SED fitting, such as that employed by KCORRECT, the Bayesian approach determines the full probability distribution of any parameter. This has the advantage of more robust parameter characterization and uncertainty. The absolute magnitudes derived in the GSWLC catalogs are determined from the best fit model by comparison between the redshift model fluxes and the observations.

We simply adopt these rest-frame magnitudes directly from the catalog, as the GSWLC-M2 catalog serves as the basis for the data used in the analysis presented here so no additional conversions or calculations are necessary.

Results from KCORRECT and GSWLC-M2 provide two standards of comparison for our own derived *K*-corrections. One of the innate difficulties in determining *K*-corrections is the lack of a "true" answer for any observed galaxy. This makes it challenging to probe any differences between our own data driven approach and *K*-corrections from other template based approaches. Our goal is a method for which results are comparable to those from other approaches in regimes where we trust the SED templates and a technique that is easier to implement than those used by KCORRECT and GSWLC.

# 5.2.3 Deriving Data Driven K-corrections

Our approach to determining K-corrections is based on the simplifying assumption that galaxy SEDs at low redshift are a single parameter family. This assumption stems from applications of dimensionality reduction methods, like principal component analysis, that find galaxy spectra can be described with only a single parameter [231, 230, 77]. Likewise, previous work has shown that K-corrections can be approximated by an analytical function of redshift parameterized by a single quantity that characterizes the galaxy, like observed color [70, 284, 16] or  $D_n4000$  [375].

In this work, we leverage rest-frame color as our single parameter, i.e., it serves as our means of mapping to the SED shape. We choose a rest-frame color that can be welldetermined by another K-correction method and bootstrap off of that reference color to determine K-corrections for the rest of the SED. While the assumption of this direct mapping of a single rest-frame color to SED shape is not perfect due to the ambiguity of contributions to the SED at fixed rest-frame color [78, 232], it is sufficient to first order out to z = 1 [230], whereas our galaxy sample only spans from 0.01 < z < 0.09.

Our approach for calculating K-corrections is the construction of a function that produces rest-frame absolute magnitude given (1) the apparent magnitude of the band that needs to be corrected, (2) the apparent magnitude of another band to serve as an anchor, (3) restframe color in a pair of bands, and (4) redshift. The band that we refer to as the "anchor" is necessary for connecting the observed-frame to the rest-frame. Our method requires a K-corrected galaxy color calculated from another method (such as the ones described in Section 5.2.2). We recommend the selection of high signal-to-noise bands for the reference rest-frame color and the anchor band, hence our use of (g - r) and  $m_r$  in this work. Some caution is warranted in using  ${}^0(g-r)$ , as galaxies that span a large range in (NUV - r) (4–6; i.e., green valley and truly quiescent galaxies) will end up with the same (g - r) of ~ 0.75. For the UV bands we tested other rest-frame colors such as  ${}^0(u - r)$  and yielded comparable results to the  ${}^0(g - r)$  derivation.

To determine the function from which we derive rest-frame colors, we perform a series of polynomial fits that are outlined in the following algorithm. These fits are tempered by various error cuts that determine which quantities are used for which fits. In the following, Ywill refer to our generalized target band, which can be any band of interest except r, which is the anchor band.

1. The initial fits: Determine fits of the observed (anchor - target band) color as a function of redshift, in bins of rest-frame color.

We approximate our K-corrections as an analytical function of redshift parameterized by rest-frame color. Thus we define a K-correction as  $K_{corr} = f(z) \times (r - Y)$ , where z is redshift, f(z) is a polynomial function to be determined, r is the anchor band, and Y is the band for which a K-correction is needed. First we determine the polynomial relationships between the observed color in a pair of bands (one of which serves as the anchor, and one of which is the desired target band) and redshift, in bins of the restframe color in another (e.g., optical) pair of bands. The relationship between these three quantities is exemplified in Fig. 5.1 where we plot observed (r - i) color as a function of redshift, with the points colored by rest-frame (g - r). It is clear that the change in (r - i) can be quantified by the rest-frame color and redshift. We use small bins of rest-frame color (~ 20 in total) with an approximately equal number of objects per bin. In the work presented here we use bins of rest-frame (g - r) color, but any other reliable color measurement can be used.



Figure 5.1: Observed (r-i) color plotted as a function of redshift for our galaxy sample. We have color coded the points by rest-frame  ${}^{0}(g-r)$  color where purple corresponds to more blue galaxies and red corresponds to more red galaxies. It is evident that observed color is correlated with rest-frame color (and redshift). By obtaining the fit coefficients in bins of the rest-frame color (see Fig. 5.2) we can quantify how the observed color changes relative to the rest-frame color across redshift space.



Figure 5.2: An example of our first order polynomial fits. Here *i*-band is the band we desire to *K*-correct and *r*-band serves as the anchor. We plot the observed (r - i) color as a function of redshift for our sample after it has been split into 22 bins in  ${}^{0}(g - r)$  with ~ 5435 objects per bin. Points from 3 of these bins are plotted, which are labeled according to the average  ${}^{0}(g - r)$  color per bin. The black dashed lines represent the fits to Equation 40 for each bin plotted, performed using a Huber regression. The fits are a sensible representation of the data and not swayed by outliers.



Figure 5.3: Linear term coefficients  $(a_1)$  from the initial fit (i.e., the fit of observed color as a function of redshift in bins of rest-frame color,  ${}^0(g-r)$ ) as function of the central  ${}^0(g-r)$  of each bin are plotted as blue points. For reference, three of the initials fits from which the  $a_1$ 's are derived are depicted in Fig. 5.2. We perform a linear fit for  $a_1$  as a function of the center of our  ${}^0(g-r)$  bins with a Huber regression (Equation 42), which is plotted here as a black dashed line. A linear fit is a sufficient approximation due to the relatively small scatter. None of the  $a_1$  points in this band were excluded from the fit but the NMAD cutoff, but the Huber regression is robust to the reddest point.

In each bin we perform a polynomial fit to determine the relationship between observed color and redshift. We compute both linear and quadratic fits for all bands for comparison purposes but note our work favors the linear approach, which is further discussed in Section 5.3.2. Our plots and tables generally show results from linear fits for the initial polynomial fits. Our function is defined as

$$(r - Y)_{\text{obs}} = f(z) = a_0 + a_1 z + a_2 z^2, \tag{40}$$

where  $a_0$  is the constant term,  $a_1$  is the linear coefficient, and  $a_2$  is the quadratic coefficient. We parameterize Equation 40 using a Huber regression technique [172], a method that is robust to outliers. These fits are implemented via scikit – learn (sklearn.linear\_model.HuberRegressor; [286]), where we set the parameter that controls the number of samples to be counted as outliers  $\epsilon = 1.01$ . The smaller  $\epsilon$  is the more robust to outliers the model is (where 1 is the minimum). For the linear case  $a_2$  is set to 0. Thus, for each bin in rest-frame color we determine the fit coefficients  $a_0$ ,  $a_1$ , and  $a_2$ separately (or just  $a_0$  and  $a_1$  for the linear case).

We take the additional step of weighting in these fits. The weights are based upon redshift such that higher redshift objects do not dominate the fits for observed color vs. z in the redder  ${}^{0}(g - r)$  bins since these bins are more populated. For a single bin in  ${}^{0}(g - r)$  we split the objects into 40 bins in redshift space. Using these redshift bins and the counts in each bin  $N_{bin}$  we build a function  $N_{bin} = f(z_{cen})$  where  $z_{cen}$  is the center of the redshift bin. This fit is performed using the Ref. [364] scipy.interpolate.interp1d method. With this function we pass in the redshift of each individual galaxy in the given  ${}^{0}(g - r)$  bin to obtain the  $N_{bin}$ . So this means we figure out which redshift bin the galaxy belongs in and the associated number of galaxies that belonged in that bin  $(N_{bin})$ . This value is used for the weighting for that specific galaxy, which means that galaxies that fall into the same redshift bin will have the same weighting. The weights are defined as:

$$w = \frac{1}{(NMAD(r-Y)_{obs})^2} \frac{\langle N_{bin} \rangle}{N_{bin}},$$
(41)

where  $(NMAD(r - Y)_{obs})$  is the normalized median absolute deviation of the observed color and  $\langle N_{bin} \rangle$  is the mean number of galaxies in the 40 bins. Generally this down-weights bins with large numbers of objects which is especially important in the redder  ${}^{0}(g-r)$  bins that predominantly have higher redshift objects skewing our fits. These weights are then utilized in the aforementioned Huber regression of Equation 40 by assigning them as sample\_weight = w.

As an example we show fits where Y is the SDSS *i*-band. A linear fit for observed (r-i) as a function of redshift is illustrated in Fig. 5.2. We have elected to present 3 bins in  ${}^{0}(g-r)$  instead of the full set of 22 for clarity. Each bin contains approximately 5435 galaxies. The points show all of the galaxies within the respective bin, while the black dashed lines show the result of the linear fit using Equation 40 ( $a_2 = 0$ ). While the bin with blue points could also be fit but a quadratic, overall results are more stable by employing linear fits.

2. The secondary fits: Determine fits of the previous fit coefficients as a function of mean rest-frame color.

Second, we fit for the linear dependence of each of the fit coefficients  $(a_1 \text{ and } a_2)$  on the *mean* rest-frame color within a bin. Namely we determine  $a_1$  (mean rest-frame color) and  $a_2$  (mean rest-frame color). We express these fits as

$$a_1 = b_1 \langle {}^0(g - r) \rangle + b_0 \tag{42}$$

and

$$a_2 = c_1 \langle {}^0(g - r) \rangle + c_0, \tag{43}$$

where  $b_0$ ,  $b_1$ ,  $c_0$ , and  $c_1$  are fit coefficients and intercepts. These fits are calculated using the Huber regression as before. Therefore we assume a linear dependence between each coefficient and the rest-frame reference color. For these secondary fits we also tested up to fifth order polynomials as well as logarithmic and exponential functions, but found that a linear fit performed just as well if not better than these more complex functions. However, not all calculated  $a_1$ 's and  $a_2$ 's are necessarily used in this fit, as in many bands the  $a_1$ 's or  $a_2$ 's have a large amount of scatter. We use an NMAD cutoff to determine which  $a_1$ 's and  $a_2$ 's are fit in addition to the Huber's robustness to outlier, as we found that in some bands the Huber still struggled with some of the fits. The NMAD cutoff is derived from the residual of the initial fit. In each  ${}^{0}(g-r)$  bin we calculate the difference between the predicted  $(r - Y_{\text{observed}})$  and the actual  $(r - Y_{\text{observed}})$  for each galaxy. Then we calculate the normalized median absolute deviation for the residual in the respective  ${}^{0}(g-r)$  bin. Bins where the NMAD < 2.5min(NMAD) are excluded from the fits of  $a_1$ and  $a_2$  as a function of mean rest-frame color.

Our *i*-band example continues in Fig. 5.3, which illustrates the secondary fit. In blue points we plot the  $a_1$ 's determined from the fitting procedure in (i) as a function of the mean  ${}^0(g - r)$  for each of the bins that underwent the initial fit. The black dashed line marks the resulting linear fit from the Huber regression for this "secondary" fit (Equation 42). For this band, none of the  $a_1$ 's were excluded by the NMAD cutoff.

For full data sets where a large number of galaxies require k-corrections, binning in a restframe color is not necessary once  $a_1$  (mean rest-frame color) have been determined. The fit coefficients can be evaluated for any given object given its rest-frame color ( ${}^0(g-r)$ ) in our example). We do not determine a fit for  $a_0^0(g-r)$  because at a redshift of 0  $a_0$ corresponds to the mean rest-frame color, based on the simple argument that at redshift 0  $(r-Y)_{\text{observed}}$  must be equal to  ${}^0(r-Y)$  as there are no redshift effects. In terms of Equation 40  $f(0) = a_0 = {}^0(r-Y)$ . We cannot simply solve for  $a_0$  as then we would assume the mean rest-frame color for all objects. Instead we make small adjustments to the observed photometry with  $a_1$  to obtain  ${}^0(r-Y)$ .

# 3. Determine whether to use a "constant" $a_1$ .

As mentioned above, our final derived K-corrections (see Fig. 5.9) will all be determined from linear fits ( $a_2 = 0$  in Equation 40 and we do not use Equation 43). However, for some bands it is better to use a constant linear term  $a_1$  instead of a rest-frame color dependent  $a_1$ . In these cases the resultant  $a_1$ 's from the initial fit are all approximately the same value, causing the secondary fit to be approximately a flat line. For these bands instead of calculating  $a_1$  as a function of mean rest-frame color, as described in (ii), we simply use the median of the  $a_1$ 's. The results become far more stable if we switch to using a constant value for  $a_1$ , as opposed to letting it remain as a function.

To determine whether to use a constant  $a_1$  we take a statistical approach. First we

determine the NMAD of the  $a_1$ 's predicted for each  ${}^0(g-r)$  bin (which we refer to as  $a_{1,\text{pred}}$ ) from the fit of  $a_1$  as a function of mean  ${}^0(g-r)$  (Equation 42). We refer to this quantity as  $\sigma_c$  ( $\sigma_c = \text{NMAD}(a_{1,\text{pred}})$ ). Then we determine the NMAD of  $a_1 - a_{1,\text{pred}}$  which we call  $\sigma$ . The  $\Delta\chi^2$  between the constant and the linear fit is defined as

$$\Delta \chi^2 = N(1 - \frac{\sigma_c^2}{\sigma^2}) \tag{44}$$

where N is the number of  ${}^{0}(g-r)$  bins used in the initial fit in (i) that pass the NMAD cutoff in (ii). Then we define  $\Delta AIC$  (the Akaike Information Criterion, a robustness test; Ref. [5]) as

$$\Delta AIC = 2 + \Delta \chi^2, \tag{45}$$

as we have 2 free parameters from the linear fit. For photometric bands where  $\Delta AIC \geq -10$  we use a linear fit for  $a_1$  and for photometric bands where  $\Delta AIC \geq -10$  we use a "constant"  $a_1$ , meaning that we calculate the median of the  $a_1$ 's determined in (i). Bands for which we use a "constant"  $a_1$  derived from KCORRECT  ${}^0(g-r)$  are: JHKs, and W1/W2/W3/W4. Bands for which we use a constant  $a_1$  derived from GSWLC-M2  ${}^0(g-r)$  are: z, Ks, W1, W2, and W4. All other bands use what we refer to as a "linear"  $a_1$ , referring to solving Equation 42.

# 4. Using the coefficient fits and algebra, determine colors and absolute magnitudes.

At this stage we now have a function that predicts the observed color (r - Y) as a function of the rest-frame color in another pair of bands and the redshift. Using our previous discussion of  $a_0$  in (ii) we can re-write Equation 40 as:

$$(r - Y)_{\text{obs}} = {}^{0}(r - Y) + a_1 z + a_2 z^2$$
(46)

Thus by subtracting  $a_1 z$  and  $a_2 z^2$  from observed color, rest-frame color can be obtained (i.e., a k-corrected color). We can simplify our rest-frame color as

$${}^{0}(r-Y) = (r-Y)_{\text{obs}} - a_{1}z - a_{2}z^{2} = m_{r} - m_{Y} - a_{1}z - a_{2}z^{2}$$
(47)

If one is interested in absolute magnitude instead of color, then one *must* have access to a k-corrected version of the anchor band. To solve for  ${}^{0}M_{Y}$  instead of  ${}^{0}(r-Y)$ , color can be split into magnitude components in the following way:

$${}^{0}(r-Y) = {}^{0}M_{r} - {}^{0}M_{Y}{}^{0}M_{Y} = {}^{0}M_{r} + {}^{0}(r-Y).$$
(48)

 $^{0}M_{Y}$  is the final k-corrected absolute magnitude.

### 5. Determine errors on K-corrections.

Errors attributed to this K-correction method are determined statistically. Because our final results are derived from either a linear fit for the  $a_1$ 's or constant  $a_1$ 's we present this calculation for linear fits.

The formula to describe our errors can be derived from Equation 47 using propagation of errors. The resulting formula is:

$$\sigma(^{0}(r-Y)) = \sqrt{\sigma_{m_{r}}^{2} + \sigma_{m_{Y}}^{2} + (z\sigma_{a_{1}})^{2}}.$$
(49)

We have dropped the  $a_1^2 \sigma_z^2$  term as  $\sigma_z$  is negligible for this local sample of galaxies. Thus we need to determine  $\sigma_{a_1}$  for results with either linear  $a_1$  or constant  $a_1$ . For linear  $a_1$ ,  $\sigma_{a_1}$  is calculated by a bootstrap method. For each band we bootstrap the  $a_1$ 's and corresponding mean  ${}^0(g - r)$  of the  $\sim 20$  bins 100 times. In each bootstrap iteration we perform a Huber regression as in (ii), Equation 42, which yields 100  $b_0$ 's and  $b_1$ 's. Then for each galaxy we can determine 100  $a_1$ 's with Equation 42 by plugging in the galaxy's  ${}^0(g - r)$ . Lastly we calculate the standard deviation of the bootstrapped  $a_1$ 's and multiply by redshift, which gives us our  $z\sigma_{a_1}$  term.

For constant  $a_1$  bootstrapping is not necessary. Instead we define  $\sigma_{a_1} = \frac{\text{NMAD}(a_1)}{\sqrt{0.64N}}$ . We define  $\sigma_{a_1}$  this way because we use the median  $a_1$ , and the standard deviation of the median scales as  $\frac{1}{\sqrt{0.64N}}$ .

For convenience purposes we keep systematic (photometry measurement error, e.g.,  $\sigma_{m_y}$ ) documented separately from errors attributed to the *K*-correction method  $(z\sigma_{a_1})$ . When applicable errors are combined in quadrature. Tables of  $b_0$  and  $b_1$  for determining  $a_1$ , and the median  $a_1$ 's for our data are available in Table 5.1 (KCORRECT derived) and Table 5.2 (GSWLC-M2 derived) of Section 5.5. The functions used to determine our  $b_0$ 's,  $b_1$ 's, median  $a_1$ 's and other related quantities in addition to sample code are provided at our *K*-correction GitHub<sup>2</sup> page for public use.

# 5.3 Results

We test our results by comparing to two other K-correction methods performed on the same data: the standard software KCORRECT, and SED fitting results from the GSWLC-M2 catalog [312, 311] which are described in Section 5.2.1. In our work the rest-frame values from which we determine our fits are from either the GSWLC-M2 catalog calculations or the KCORRECT calculations. Depending on which K-correction method we wish to compare our calculations to we use the respective rest-frame color.

While there are discrepancies between K-corrections calculated between any two methods, with our results in general agreement with other methods for well-behaved bands we can be confident in our simplified approach.

# 5.3.1 Distributions of Rest-frame Colors

We start with a generalized comparison to KCORRECT and GSWLC-M2. In Fig. 5.4 normalized histograms depict results for colors across the wavelength range. We derive these respective to the catalog to which they are being compared. For example we derive  ${}^{0}(u-r)$  utilizing  ${}^{0}(g-r)$  from KCORRECT when comparing to KCORRECT results. The step histograms denote our derived results, and the lightly shaded histograms denote results from the two comparison methods. Blue histograms come from KCORRECT and orange histograms come from GSWLC.

Foremost, we do not expect all results to be 1:1 with each other due to the nature of fitting and how the *K*-corrections are approximated in each method. The relatively large

<sup>&</sup>lt;sup>2</sup>https://github.com/cfielder/K-corrections

scatter of individual K-corrections has been exemplified in the literature (see e.g., [284], Fig. 3, [70], Fig. 2). We do find that, in general, the distributions of our K-corrected colors are in good agreement with those derived from other methods. In many cases our results are in better agreement with the respective comparison work than KCORRECT and GSWLC-M2 are with each other. However, there are some notable differences.

Our results in the 2MASS bands from either derived source are in excellent agreement with each other and the results from KCORRECT. However, the GSWLC-M2 results are more peaked. Such discrepancies in the near-IR 2MASS bands in the GSWLC-M2 catalog may result from a couple of factors. Templates in the IR are typically of lower quality than in the optical, particularly because there are few spectra fully observed from the UV-IR. Therefore some assumptions are required in the less constrained bands. For example, there is uncertainty of how asymptotic giant branch (AGB) stars contribute to stellar population synthesis models. In particular, thermally pulsating AGB (TP-AGB) stars may contribute a significant portion of light to the IR portion of an SED but how much is unknown [78]. The IR portion of the SED is also further complicated by contributions of dust emission. It is likely the case here that the discrepancy of the GSWLC-M2 results is driven by the SED fits preferentially constrained in the UV and optical where models have more certainty.

While small, we note that discrepancies in the FUV, NUV, and u bands are likely a result of (1) the high sensitivity of UV co lours to minimal amounts of star formation, (2) the uncertainty of contributions of post-AGB and extreme horizontal branch stars, and (3) complexity of dust absorption and scattering [78]. Simple stellar populations, which are the backbone of SPS model SED based approaches, may not adequately capture the UV sensitivity to star formation, and dust and unusual stellar populations further complicate these effects.

In Fig. 5.4 it appears that there are also discrepancies of note in the optical bands (particularly g and i). However, across wavelength space the differences between results are no more than ~ 0.1 magnitudes and agree on average at the 0.05 magnitude level after exploration of the NMAD and root mean squared error between our results and the respective literature result. Differences are more pronounced in the optical given the narrow range of color space that they span. While the  ${}^{0}(g-r)$  results presented here are calculated in bins of



Figure 5.4: Histograms of rest-frame colors comparing K-correction methods for GALEX FUV/NUV, SDSS uriz, and 2MASS J/H/Ks bands. We show our calculated rest-frame colors derived from KCORRECT (blue line) and GSWLC-M2 (orange line) and those calculated using the KCORRECT software (blue shaded) or in the GSWLC-M2 catalog (orange shaded). In general, our rest-frame colors derived from both KCORRECT and GSWLC-M2 are in excellent agreement with each other. In the 2MASS bands, our rest-frame colors derived from KCORRECT and GSWLC-M2 are self-consistent and in agreement with the KCORRECT colors; however, the GSWLC-M2 colors show more peaked distributions likely caused by uncertainties in modeling unusual older stellar populations (TP-AGB stars, post-AGB stars, and extreme HB stars). The offsets in  ${}^0(g - r)$  are likely due to the fact that our rest-frame colors are derived from  ${}^0(g - r)$  for both catalogs. Overall, our rest-frame colors for these bands are well-matched to each other and to those from KCORRECT and GSWLC-M2.



Figure 5.5: Same as Fig. 5.4 but for the WISE bands. GSWLC-M2 did not derive WISE colors, but we still derive rest-frame WISE colors using  ${}^{0}(g-r)$  from both KCORRECT and GSWLC-M2. Our results derived from either source agree well with each other. Additionally, histograms match well with those of KCORRECT, except for the W4 band, which we suspect is a result of a template error in KCORRECT. This discrepancy in W4 highlights an advantage of deriving K-corrections independent of templates in poorly constrained bands.

 ${}^{0}(g-r)$ , which is redundant, we also derived  ${}^{0}(g-r)$  from computing  ${}^{0}(g-i)$  and  ${}^{0}(r-i)$  in bins of  ${}^{0}(g-r)$  and then combining results. The results were similar, so we present the more straightforward result here. We want to emphasize that the bluest galaxies in KCORRECT and GSWLC-M2 have derived  ${}^{0}(g-r)$  colors redder than ours starting at  $z \sim 0.03$ . Additionally, the rest-frame color derived from the SED fitting techniques is *redder* than the observed color for these very blue objects, contrary to the expected blue-ward shift in the optical post K-correction. The likely culprit is either a lack of SED templates for the most extremely blue galaxies, or some lacking or excess contribution from the dust absorption and nebular emission models. Investigations in small redshift bins of rest-frame (g-r) color show that the red sequence is more consistent at different redshift with our K-correction method than  ${}^{0}(g-r)$  color derived using KCORRECT.

Fig. 5.5 depicts the same comparisons as Fig. 5.4 for the WISE bands. The GSWLC-M2 did not perform SED fitting directly with the WISE bands [311], so those values are not included. Overall, we observe very consistent results with those of KCORRECT except in the W4 band. This miss-match in distributions between results of KCORRECT and the distributions of the observed photometry and our own derived rest-frame photometry can largely be attributed to the lack of dust emission modeling in KCORRECT. Therefore any rest-frame W2 through W4 results obtained via KCORRECT could be incorrect due to the lack of these types of models.

To check the performance of our K-corrections we examine colors in small redshift bins. We choose narrow bins as there should be minimal color variance across the galaxy sample, meaning they should all have approximately the same observed frame SED. Fig. 5.6 shows results of this analysis for a selection of colors across our wavelength range. We plot contours for 0.04 < z < 0.045 (blue) and 0.075 < z < 0.08 (orange). On our x-axis we plot  ${}^{0}(g - r)$ . The left column depicts contours using observed colors and the right column depicts contours for rest-frame colors. These specific rest-frame colors are derived from KCORRECT  ${}^{0}(g - r)$ . For objects that have been properly K-corrected, the contours should line up better in the rest-frame column than the observed-frame column. This is indeed the case for our bands, which indicates that our K-corrections are working as expected. Plots made for KCORRECT and GSWLC-M2 rest-frame colors can be made, with similar results, save for (W4-r) which is dramatically offset in the KCORRECT results. While (FUV - r) exhibits minimal change, this is also the case for KCORRECT and GSWLC-M2.

#### 5.3.2 Extrapolations to z=0

At redshift 0, by definition a K-correction must be 0. This means that any difference between observed colors (or magnitudes) and rest-frame colors (or magnitudes) at redshift 0 must also be 0. As we showed in Section 5.2.3 our K-corrections are defined to follow this relationship. However, this is not enforceable for methods that incorporate SED/template fitting. We compare our K-correction colors to that of KCORRECT and GSWLC-M2 in Fig. 5.7 and Fig. 5.8 for optical and UV/near-IR bands, respectively. We plot the color difference as a function of redshift where, for example,  $\Delta(u - r) = (u - r) - (u - r)$  or the difference between the observed-frame color and rest-frame color. We specifically plot the means of these color differences across 12 bins in redshift space. Blue points correspond to K-corrected results using our method with a linear fit and orange points correspond to K-corrected results using our method with a quadratic fit. Grey points correspond to results from KCORRECT (left column) and GSWLC (right column).

The blue points also have representative error bars plotted that include both systematic errors and K-correction errors, for which the K-correction errors are described in (v) of Section 5.2.3. We must take a few additional steps for these plots, as we are showing the means in bins. The error attributed to the K-correction does *not* diminish when determining means. So for the K-correction error we simply compute  $\langle \sigma_{a_1} z \rangle$ . We define the random error on the mean as  $\sigma_{sys} = \frac{\sigma(\Delta)}{\sqrt{N_{bin}}}$ . The total error on the means is thus these two contributions added in quadrature.

Because our galaxy sample only extends down to redshift 0.01, we must extrapolate to z = 0. This is done with a linear interpolation for color difference as a function of the redshift of the bin center  $(z_{cen})$ , utilizing the scipy.interpolate.interp1d class from Ref. [364]. These extrapolations are plotted as dashed lines, which ideally should pass through the origin. Our linear and quadratic fits do so in most cases (with some scatter), in contrast to fits from KCORRECT and GSWLC-M2 for many bands. This directly demonstrates the im-



Figure 5.6: Contour plots of galaxies in two narrow redshift bins 0.04 < z < 0.045 (blue) and 0.075 < z < 0.08 (orange). We plot a variety of colors across our wavelength range as a function of  ${}^{0}(g - r)$ . The left column shows these colors in the observed-frame and the right shows these colors in the rest-frame, using our *K*-correction results. We plot narrow redshift bins as the observed SEDs of the galaxies should be approximately the same. It is evident that for all colors the contours of both redshift bins align closely in the rest-frame in contrast to the observed-frame. This is the expected result for successful *K*-corrections and similar results are observed in KCORRECT and GSWLC-M2 (save for (W4 - r)).

proved accuracy of our *K*-correction method compared to SED-fitting *K*-correction methods. Additionally these figures exemplify that in most cases a linear function of redshift is a better choice than a quadratic one for deriving our relationship between observed color, redshift, and rest-frame color. First, the linear fit derives very comparable results to the quadratic fit and is much simpler to implement. Second, the linear fit results more consistently pass through the origin than the quadratic fit results. It is also worth noting that despite the difference in the techniques, the relative shapes of the results in both KCORRECT and GSWLC-M2 agree well. Likewise the relative shapes of our results, regardless for which source the rest-frame band is used, are also in agreement in both linear and quadratic initial fits.

#### 5.3.3 Spectral Energy Distributions

Spectral energy distributions (SEDs) are an excellent test of our K-correction results. In this subsection we compare our K-corrected rest-frame photometry to that of our other sources of K-corrected photometry and the observed photometry for a couple of galaxies. Because our galaxy sample is at such a low redshift, the K-correction in each band should be relatively small and mirror the shape of the observed SED.

To determine luminosity we use similar derivations to those presented in Section 4.3 of Ref. [125], which we also summarize here. Because the *r*-band serves as the anchor in our *K*-correction derivations we can derive fluxes relative to the *r*-band. We use  ${}^{0}M_{r}$  from KCORRECT and GSWLC-M2 respective to which *K*-correction source we compare to. The luminosity of *r*-band is derived as

$$\log\left(\nu_r L_{\nu,r}\right) = \log\nu_r + \frac{({}^{0}M_r - 34.04)}{-2.5},\tag{50}$$

which is derived from the relation for converting flux to AB magnitudes in combination with the area of a 10 pc radius sphere to convert flux to luminosity. The error on this quantity via propagation of errors is equivalent to

$$\sigma_{\log(\nu_r L_{\nu,r})} = 0.4 \sigma_{M_r}.$$
(51)



Figure 5.7: Plots of the mean difference between observed and rest-frame color (e.g.,  $\Delta(u-r) = (u-r) - (u-r)$ ) as a function of 12 redshift bins for optical colors. Blue points depict our K-corrected results with a linear initial fit, and orange points depict results for a quadratic fit. Grey points show results from the corresponding source denoted at the column title. The colored dashed lines depict a linear extrapolation to redshift 0. Our results all approach the origin closely, as they should by construction and under the expectation that are redshift 0 the K-correction should be 0. However, this is not necessarily the case for K-corrections that rely on SED fitting.



Figure 5.8: Same as Fig. 5.7 but for UV and near-IR bands. Our results perform just as well as if not better than those from SED fitting.

For the other bands we use our derived colors  ${}^{0}(r - Y)$  to first determine the flux ratio relative to the *r*-band

$$\log\left(\frac{f_{\nu,r}}{f_{\nu,Y}}\right) = \frac{{}^{0}(r-Y)}{-2.5}.$$
(52)

This formula is based on the definition of observed AB magnitude. We can then derive luminosity for other bands utilizing the equivalence of flux ratios to luminosity ratios. Our final SED luminosities are

$$\log\left(\nu_Y L_{\nu,Y}\right) = \log\left(L_{\nu,r}\right) - \log\left(\frac{f_{\nu,r}}{f_{\nu,Y}}\right) + \log\left(\nu_Y\right),\tag{53}$$

where  $\log(\nu_Y L_{\nu,Y})$  is in units of Watts. The corresponding errors are defined as

$$\sigma_{\log\nu L_{\nu,x}} = 0.4\sqrt{\sigma_{m_r}^2 + \sigma_{m_Y}^2 + (z\sigma_{a_1})^2 \sigma_{0_{M_r}}^2},\tag{54}$$

which can be derived from propagation of errors and plugging Equation 47 into Equation 52.

In Fig. 5.9, we plot the SEDs of two galaxies from our sample. These galaxies share the same  ${}^{0}(g-r)$  color of 0.667 but reside at two different redshifts with the lower redshift galaxy plotted in blue and the higher redshift galaxy plotted in orange. The normalization of the lower redshift object is offset such that the two galaxies can be plotted without overlap. Open circles with error-bars show results from our work, which are derived using either a linear or constant  $a_1$  in Section 5.2.3. For all bands other than r, we do not plot the  $\sigma_{{}^{0}M_r}$ term in Equation 54 as this error is covariant across all bands. This quantity is small and challenging to infer from the plots;  $\sigma_{{}^{0}M_r} = 0.0012$  for the low redshift plot and  $\sigma_{{}^{0}M_r} = 0.0032$ for the higher redshift plot. Triangular points show the K-corrections for these galaxies from KCORRECT (left panel) and GSWLC-M2 (right panel). Lastly, star points show the observed fluxes of the same galaxies, normalized to match in the r-band.

As a means to compare these two galaxies we also plot an SED template galaxy from the Ref. [50] nearby galaxy SED atlas - NGC 4138. The magnitudes of the templates were converted to fluxes (as described in 4.2.2 of Ref. [125] utilizing the AB magnitude definition), and then normalized in r-band. For each of the 129 galaxy SED templates in the atlas we calculate the chi-squared of the difference between the brightnesses of each of our selected galaxies and that of the templates:

$$\chi^2 = \sum \left( \frac{\log \nu L_{\nu}^{\text{atlas}} - \log \nu L_{\nu}^{\text{galaxy}}}{\sigma_{\log \nu L_{\nu}}} \right)^2.$$
(55)

 $\sigma_{\log \nu L_{\nu}}$  combines in quadrature the total error in the galaxy SED for a given band, the uncertainties in the Ref. [50] photometry, and  $\log_{10}$  (1.1), which corresponds to a 10% error in  $\nu L_{\nu}$ . This additional 10% error is added to account for systematic uncertainties in the photometry such that optical bands do no dominate the  $\chi^2$  values given their intrinsically small uncertainties.

The template plotted in gray in Fig. 5.9 corresponds to the spectrum of the NGC 4138 template, which had the smallest  $\chi^2$  compared to *both* galaxies selected from our sample. Dashed portions of the plotted template indicate regions of the spectrum that were modeled, while solid portions of the template indicate regions of the spectrum that were observed. We do not expect this template to directly match the photometry of either selected galaxy but to provide a means of comparison. This spectrum also exemplifies the use of modeling required in the UV and IR for empirical SED templates.

One can draw several conclusions from Fig. 5.9. First, our K-corrected photometry is reasonably close to that of other methods of producing K-corrected photometry in addition to the observed photometry. At our low redshift range the rest-frame photometry should not be substantially different from the observed photometry, which is the case for our results. When combined, we have results that resemble an SED reasonably well, while making minimal assumptions about how an SED *should* look.

Second, it is evident that K-CORRECT fails in giving a reasonable result for W4 band K-corrections which is further supported by the results shown in Fig. 5.5. In Fig. 5.9 the brightness of the lower redshift galaxy's W4 band from KCORRECT is so low that it overlaps with the other galaxy's observed brightness. For the higher redshift galaxy the W4 point is below the y-axis range. Our K-correction approach supplies reasonable results for W4. By construction our method cannot stray too far from the observed photometry, so even bands like W4 that have low signal-to-noise ratios can be reasonably constrained. In that same

vein our K-correction method has more flexibility than standard SED fitting where multiple bands are forced to match the templates.

# 5.4 Conclusions

In this work, we present a new, empirically-driven method for obtaining K-corrections. Our method is inspired by the polynomial based K-correction methods of Ref. [16] and Ref. [70], where we exploit the parameterisation of K-corrections by an analytical function of redshift and color [382, 284]. At low redshift (z < 1), SEDs fall into a single parameter family where a single rest-frame color allows us to infer the full SED shape - an inference made from results describing spectra by a single parameter in reduced dimensional space [77, 231, 230]. Our method limits the dependence on SED templates by interpolating using a rest-frame color for which SEDs are well-constrained. We perform a series of linear fits to construct a function that maps observed color and redshift to rest-frame color.

There are a number of pieces of evidence that prove our data driven approach to determining K-corrections is yielding sensible results. Throughout this work we compare to the oft-used KCORRECT v4.3 software of Ref. [31] and the derived results in the GSWLC-M2 catalog [312, 311], both of which utilize SED fitting in different ways (see Section 5.2.2). Our K-corrected photometry is in agreement with K-corrected photometry of the same objects for both comparison methods, particularly at UV and optical wavelengths (see e.g., Fig. 5.4). Likewise our K-corrected photometry produces sensible looking galaxy SEDs, similar to both comparison methods (see Fig. 5.9). We also have better agreement across redshift in color-color space with our rest-frame colors than our observed-frame colors, as exemplified in Fig. 5.6. This is the expected behavior of photometry that has been correctly K-corrected. Lastly, our K-corrections go to zero at redshift zero, as shown in Fig. 5.7 and Fig. 5.8, as expected by the definition of the K-correction . However, SED fitting methods cannot directly enforce this, as exemplified by their differences in the figures. Our method is driven by and stays close to the observations in the band of interest rather than follow a model which can yield larger deviations. This allows us to K-correct bands that are poorly


Figure 5.9: Derived SEDs for two galaxies from our sample which share the same  ${}^{0}(g-r) = 0.667$  but reside at two different redshifts. The lower redshift galaxy is plotted with blue points and offset from the higher redshift galaxy in the y-axis for clarity. Stars are plotted at the observed fluxes. Triangles are plotted at the luminosity determined from the rest-frame magnitudes derived via KCORRECT (left panel) and GSWLC-M2 (right panel). Circles with error bars show our resulting luminosities from our K-correction method. As a means of comparison we plot the spectrum of NGC 4138 from the Ref. [50] galaxy SED atlas in gray, with dashed lines showing modeled portions of the spectra and solid lines showing the observed portions of the spectra. This template had the lowest  $\chi^2$  of the difference between the photometry of the two galaxies shown and the template (formalized in r-band). While we do not expect a direct match to a template, we provide it as a means of comparison to help guide the eye. We can conclude that our K-correction method provides reasonable estimates for even the weakly constrained WISE bands, where KCORRECT fails in W4and GSWLC lacks. The rest-frame W4 prediction of the higher redshift galaxy lies below the y-axis range and the rest-frame prediction for the lower redshift galaxy overlaps with the observed photometry of the higher redshift galaxy. Given the redshifts of these objects and our data driven approach, our K-corrected magnitudes cannot significantly deviate from the observed photometry, which allows us to K-correct even low signal-to-noise bands like W3 and W4.

constrained by templates and have low signal-to-noise ratios.

There are also notable discrepancies between our K-corrected photometry and that of KCORRECT and GSWLC-M2, which mostly occur in the infrared bands. First, as seen in Fig. 5.4, the rest-frame photometry in the 2MASS JHKs bands is more peaked compared to the other histograms. However, our derivations from both KCORRECT and GSWLC-M2 are in excellent agreement with each other and that of KCORRECT. This highlights one of advantage of our approach. SED model uncertainties in the IR due to contributions from AGB stars may result in incorrect K-corrections. Higher signal-to-noise optical portions may also drive SED fits resulting in incorrect K-corrections. Empirical SED models are also more poorly constrained at non-optical wavelengths due to a lack of spectral measurements that in turn produce uncertain SED fits. The other notable discrepancy is that of the WISE W4 band, noted in Fig. 5.5 and Fig. 5.9. The predicted KCORRECT W4 band restframe photometry is severely offset from our predictions, while our results utilizing either KCORRECT or GSWLC-M2 optical colors are in good agreement with each other and with KCORRECT results in the other WISE bands. We suspect that the KCORRECT templates are making un-physical assumptions, yielding an unexpected result. Most notable is the lack of dust emission modeling in KCORRECT, a feature of increasing importance for the WISE bands. In contrast, our data-driven approach entirely avoids issues of template fit failures and model issues by only relying on well-constrained rest-frame optical colors.

Our data driven K-correction approach is particularly suitable for calculations in bands where templates are not well constrained or readily available. In our previous work of Ref. [125] in constructing a UV-IR SED for the Milky Way we required WISE-band K-corrections - bands sensitive to dust content and dust luminosity (which are consequently a constraint on star formation rate, see e.g., Ref. [311]). Because WISE-band K-corrections were not readily accessible we developed the approach presented here for determining K-corrections

#### 5.5 Tables of Polynomial Functions

In this section we present the derived intercept  $(b_0)$  and coefficient  $(b_1)$  for the secondary fits described in (ii) of Section 5.2.3. For bands where a "constant"  $a_1$  is a statistically better choice (see (iii) of Section 5.2.3) in which we use the median of the  $a_1$ 's determined in the first fit ((i) of Section 5.2.3) we present the median  $a_1$  instead (med.  $a_1$ ). Table 5.1 presents results where the KCORRECT  ${}^0(g-r)$  serves as the reference rest-frame band and Table 5.2 presents results where GSWLC-M2  ${}^0(g-r)$  serves as the reference rest-frame band.

The  $b_0$  and  $b_1$  quantities specifically come from Equation 42. The reader can use these quantities and their chosen rest-frame color band to determine  $a_1$ . Then using Equation 46 one can obtain rest-frame color for the bands of interest. For example, with optical SDSS r as the anchor band, if one wished to determine rest-frame quantities for optical z-band with Table 5.1 or Table 5.2,  $a_1$  can be solved as:  $a_1 = b_0 = b_1^0(g - r)$  where  ${}^0(g - r)$  is the rest-frame color of the galaxy or galaxies of interest. Then one can solve for rest-frame color by calculating  ${}^0(r - z) = (r - z)_{obs} - a_1 z$ . For bands that use median  $a_1$ , the first equation can be skipped. To determine, for example, rest-frame W4 measurements one can simply compute  ${}^0(r - W4) = (r - W4)_{obs} - (med. <math>a_1)z$ .

In addition to these tables we provide code for determining med.  $a_1$ ,  $b_0$  and  $b_1$  or other related quantities at our GitHub repository for full public access for adaption to any other project, under a CC BY-SA 4.0 license.

KCORRECT Derived Coefficients				
Passband	med. $a_1$	$b_0$	$b_1$	
FUV		-18.557	30.780	
NUV		-7.311	13.429	
u		2.503	-7.553	
g		0.172	-2.849	
i		2.609	-2.887	
$\mathbf{Z}$		1.816	-1.440	
J	1.009			
Н	1.189			
Ks	3.946			
W1	3.945			
W2	5.520			
W3	6.252			
W4	9.394			

Table 5.1: Table of KCORRECT  ${}^{0}(g-r)$  derived quantities for the median  $a_1$ 's, intercept  $(b_0)$ , and slope  $(b_1)$  for all bands considered in this work. Depending on the band it is statistically more sensible to either simply used the median derived  $a_1$ 's or to fully solve  $a_1 = b_0 + b_1^0(g-r)$ . In either case, rest-fame colors can then be determined with the relation  ${}^{0}(r-Y) = (r-Y)_{obs} - a_1 z$  where Y is the band for which rest-frame quantities are to be derived.

GSWLC-M2 Derived Coefficients				
Passband	med. $a_1$	$b_0$	$b_1$	
FUV		-11.243	19.597	
NUV		-4.251	8.790	
u		3.986	-9.244	
g		1.089	-4.069	
i		2.088	-2.089	
Z	0.777			
J		4.345	-4.506	
Н		6.955	-7.721	
Ks	3.687			
W1	3.548			
W2	5.391			
W3		21.736	-25.185	
W4	9.277			

Table 5.2: Same as Table 5.1 but for quantities derived utilizing GSWLC-M2  $^{0}(g-r)$ . Note that the same passbands may work preferentially with median  $a_1$ 's for one catalog that do not in the other.

# 6.0 Conclusion

In this dissertation I have focused on the use of statistical techniques and models to address inconsistencies between simulations and observations and determine various global properties of the Milky Way. This has allowed me to paint a more comprehensive portrait of the Milky Way than ever before, putting together the puzzle pieces on how to bridge the gap between the galactic and extragalactic, both theoretically and observationally. The Milky Way is the ultimate laboratory for deciphering the many facets of galaxy formation and evolution, which makes the task of placing the Milky Way amongst it peers and correctly comparing it to simulations of crucial importance.

# 6.1 Summary of Previous Chapters

In Chapter 2 I explored one of the disconnects between N-body simulations and models of halos used to interpret lensing observations. This study serves as a backbone in interpreting how we connect observations and simulations and demonstrates the important role subhalos have in our studies of halo evolution. In this chapters I utilize least squared fitting in order to test a variety of dark matter halo density distributions and how they are impacted by the presence of subhalos. Overall in this chapter I find that density profiles of the smooth halo component (without subhalos) differ substantially from the conventional halo density profile, declining more rapidly at large radii. I also find that concentrations derived from smooth density profiles exhibit less scatter at fixed mass and a weaker mass dependence than standard concentrations. Both smooth and standard halo profiles can be described by a generalized Einasto profile, an Einasto profile. These results hold for both Milky Waymass and cluster-mass halos. This new characterization of smooth halo profiles can be useful for many analyses, such as lensing and dark matter annihilation, in which the smooth and clumpy components of a halo should be accounted for separately. In Chapter 3 I directly compare subhalo abundances between different dark matter halos of the same mass of the Milky Way. This chapter addresses one of the small scale discrepancies in ACDM - the missing satellites problem - and places the Milky Way dark matter halo in a theoretical extragalactic context. Motivated by recent studies that identified ways in which the Milky Way is atypical, I investigate how the properties of dark matter halos with mass comparable to our Galaxy's — including concentration, spin, shape, and scale factor of the last major merger — correlate with the subhalo abundance. Using zoom-in simulations of Milky Way-like halos, I build two maximum likelihood models of subhalo abundance as functions of host halo properties. From these models I conclude that the Milky Way most likely has fewer subhalos than the average halo of the same mass. A result that implies that models tuned to explain the missing satellites problem assuming typical subhalo abundances for our Galaxy may be over-correcting. This also raises the issue of simulators calling an object Milky Way-like by considering mass alone, which this chapter demonstrates is ill-advised.

The studies explored in Chapter 4 are aimed at directly quantifying bulk photometric properties of the Milky Way. I demonstrated the use of a machine learning technique - Gaussian process regression - to more astronomical applications. This model is trained on SDSS galaxies to map galaxy properties (stellar mass, apparent axis ratio, star formation rate, bulge-to-total ratio, disk scale length, and bar vote fraction) to UV (GALEX FUV/NUV), optical (SDSS ugriz) and IR (2MASS JHKs and WISE W1/W2/W3/W4) fluxes and uncertainties. With these models I estimate the photometric properties of the MW, resulting in a full UV-to-IR spectral energy distribution (SED) as it would be measured externally, viewed face-on. I confirm that the Milky Way lies in the green valley in optical diagnostic diagrams, but show for the first time that the MW is in the star-forming region in standard UV and IR diagnostics—characteristic of the population of red spiral galaxies. Although the GPR method predicts one band at a time, the resulting MW UV-IR SED is consistent with SEDs of local spirals with characteristics broadly similar to the MW, suggesting that these independent predictions can be combined reliably. This UV-IR SED will be invaluable for reconstructing the MW's star formation history using the same tools employed for external galaxies, allowing comparisons of results from *in situ* measurements to those from the methods used for extra-galactic objects.

Chapter 5 explores an issue that stemmed from Chapter 4 - namely that we were unable to obtain rest-frame brightnesses for a couple of our mid-infrared bands (W3 and W4) due to the lack of tools available. Rest-frame photometric properties depend on K-corrections, which convert between flux in observed band(s) to flux in rest-frame band(s). K-corrections serve as a critical means of comparing galaxies at various redshifts. Said corrections often rely on fits to empirical or theoretical spectral energy distribution (SED) templates of galaxies. For bands in which such constraints are lacking this chapter address this shortcoming by developing an empirically-driven approach to K-corrections. I perform a polynomial fit for the K-correction as a function of a galaxy's rest-frame color determined in well-constrained bands (e.g.,  $^{0}(g-r)$ ) and redshift, exploiting the fact that galaxy SED's can be described as a one parameter family at low redshift (0.01 < z < 0.09). For bands well-constrained by SED templates, the empirically-driven K-corrections are comparable to the SED fitting method of KCORRECT and SED template fitting employed in the GSWLC-M2 catalog (the updated medium-deep GALEX-SDSS-WISE Legacy Cataloger). This method dramatically outperforms the available SED fitting K-corrections for WISE W4 and mitigates incorrect template assumptions and enforces the K-correction to be 0 at z = 0. Such a tool can be useful in many aspects of astronomy, beyond my own.

## 6.2 Potential Future Work

The conclusions of each chapter contains several ideas on potential direct extensions of this work. Here I focus on a number of ways in which we can continue narrowing our picture of the Milky Way, how it compares to galaxies, and ways in which the Milky Way can tell us about dark matter and galaxy formation.

The UV-to-IR SED for the Milky Way which I have constructed in Chapter 4, can serve as a critical tool for assessing models used in constructing hydrodynamical simulations of the Milky Way. Namely do the SEDs derived from simulated Milky Ways match the SED derived from observations. By fitting this SED with one or multiple of the many available SED fitting codes, we can determine detailed star formation histories, dust reddening properties, and metallicites of the the Milky Way and/or Milky Way analogs. This enables a variety of checks that can be performed to test consistency between properties that are inferred from SEDs and those measured from stars. For example, we have measured the SFR of the MW (see Ref. [216]), so we expect analysis of the SED to recover this result. These checks provide tests of models (such as population synthesis or dust models) used to generate galaxy properties and assign photometry, extinction, etc. to hydrodynamical simulations of Milky Way-like galaxies. This modeling can also allow us to test the relatively unconstrained contributions of certain objects to SEDs such as the effect of thermally-pulsing asymptotic giant branch stars to the near-IR luminosity in stellar population synthesis models.

The results of Chapter 3 show that the Milky Way appears to be atypical in its satellite population and mass assembly history. Ref. [119] finds that the assembly history of the Milky Way is only reproduced in 0.65% of Milky Way-mass EAGLE galaxies. This assembly history should be closely related to the local environment surrounding our Galaxy, and environment has been found to play a key role in the formation and evolution of galaxies [357, 289, 33]. Due to its proximity, the Milky Way's environment has been studied extensively. A number of future projects on the study of galaxy environments can prove useful in our understanding of galaxy evolution. For example, Ref. [217] found that the Milky Way has an anomalously small disk scale length. Standard galaxy evolution theory assumes that merger history and the angular momenta of the halo is correlated with the disk size. In Chapter 3 I found that halos similar to the Milky Way in properties beyond mass typically also have fewer subhalos and more quiet merger histories. By selecting disk galaxies in SDSS and utilizing the upcoming DESI bright galaxy sample for environmental measurements we can directly determine if galaxies with shorter scale lengths (like the MW) also possess fewer satellites. This project will also allow us to assess whether dark matter halo properties correlate with scale length in the manner predicted by standard disk formation theory, establishing to what degree additional physics (e.g., feedback effects) might be needed to explain the observed trends. One could also build a model using Gaussian process regression in order to predict galaxy environment from a limited set of observed properties, particularly deviations from the luminosity-size relation. Galaxy environments tend to be noisy (e.g., if one were to rely on the morphology-density relation). A model that relies on galaxy properties instead can prove to be much more robust and useful for surveys in which it is difficult to measure local environments.

The gas content of a galaxy is an important counterpart to stellar populations as gas flows through the stellar life cycle. Neutral atomic hydrogen (HI) is a tracer of the overall gas content of a galaxy. It has been well established that galaxies undergoing star formation have a much larger neutral gas content than galaxies that are not. Cold gas, namely molecular hydrogen (H2) and the CO molecule, probes gas cloud counts and the fuel supply for star formation. Determining the HI mass of the Milky Way is important for a variety of studies such as comparisons of star formation efficiency between our Galaxy and extragalactic objects, studies of HI scaling relations, and exploring the origins of the Kennicutt-Schmidt law. However, such gas measurements are poorly constrained within the Milky Way due to distance ambiguities and the sensitivity to assumptions about MW structure. By utilizing a GPR model, like that introduced in Chapter 4, trained on physical galaxy properties to infer HI, H2 and CO gas masses from HI-MaNGA/ALMA/ALMaQUEST data in order to predict bulk gas masses for the Milky Way we can bypass issues resulting from our location in the disk. These predictions can be compared to current gas mass inferences for the Milky Way and allow us to more directly constrain the current fuel for star formation within the Milky Way. This in turn will allow us to better pinpoint the gas accretion history of the Milky Way and tie it in to the known merger history which will provide insight into the mechanisms that formed the bulge (or pseudo-bulge) of the Milky Way.

Galaxy spectra are a useful tool for determining the abundances of stellar populations within a galaxy, which reveal details such as elemental abundances and gas content. A full predicted spectrum for the Milky Way will allow us to link the MW to external galaxies more directly, which can greatly constrain our determination of how the MW has evolved and enable much more precise MW simulation efforts. Such a project would expound upon the GPR technique demonstrated in Chapter 4, in which one could us a GPR in order to predict every wavelength of the MW's spectrum. Instead of training the regression to predict broadband photometric fluxes, the Gaussian process can fit for the relationships between galaxy properties and the flux at each wavelength using spectra obtained from SDSS and/or the DESI bright galaxy survey. This model will be completely empirical (beyond the modeling used to determine galactic properties), and would not rely on any templates or other assumptions other than the relationship between galaxy properties and spectra; this means this model can also be adapted for other galaxies where full spectra cannot be obtained (e.g., large surveys like Vera Rubin LSST where exhaustive spectral follow-up is not feasible).

### Appendix Modern Milky Way Analog Searches

This Appendix provides the updated code for finding Milky Way analog galaxies, original released in Ref. [219]. While selecting analogues is no longer the technique we use to constrain Milky Way photometric predictions, Milky Way analogues (and analogue finding in general) still have a variety of relevant applications to address other science questions. For example, we have been working with students who are searching for stellar analogues using our code.

The statistical method of using Milky Way analogues was presented first in Ref. [219], as a way to determine the Galaxy's photometric properties to high precision. The physical properties used in this work included stellar mass and star formation rate, properties that have been shown to be closely correlated with galaxy luminosity and color [19], in addition to an inclination cut for high fidelity results. This work relies on the key assumption that galaxies that match the Milky Way in stellar mass and star formation rate should also have similar photometric properties to the Milky Way.

In summary, this was done by constructing a sub-sample of galaxies whose distributions of stellar mass and star formation rate match the probability distributions of these quantities for the Milky Way. To select an analogue, first a pair of values is drawn from the fiducial Milky Way stellar mass and star formation rate probability distribution function (independently). Then from the sample of galaxies an object within 1% of the drawn pair is selected. If one cannot be found tolerance of up to 3% is allowed. This draw and search is performed 5,000 times. Afterwards objects that are edge on and disk dominated are excluded. This yielded 3,402 objects of which 935 were unique galaxies. Duplicates are necessary so that the property values for the Milky Way analogues match the posterior distribution of the Milky Way properties (stellar mass and star formation rate).

There are several drawbacks to this analogue finding method, which we address in this paper with both an updated algorithm in addition to a new method for determining Milky Way photometric properties without the use of analogue galaxies (see Section 4.3.2). One such drawback, which we focus on in this subsection, is the slow algorithm. Another highdimensional drawback, which we expand upon in Section 4.3.1, is the diminished number of Milky Way analogue galaxies as we expand to higher-dimensional parameter space. We would like to add that despite these drawbacks the application of Milky Way analogues (and analogue finding in general) is still relevant to many other studies outside of photometric predictions.

The original analogue search algorithm is a simple geometric search for objects in the catalog within that lie within up to 3% of the Milky Way draw. While a first pass of selecting analogues only takes a few minutes, calculating systematics such as Eddington bias can require many iterations of selecting analogue samples (see Section 4.9 for details on Eddington bias). This process can be very time intensive ( $\sim 1$  day in time space). The time insensitivity is one of the major motivators to switching to a different type of search algorithm, namely a binary tree.

We have translated the code from IDL to Python and made use of a k-d tree, which is a binary tree with points in k-dimensional space [24]. This tree is a very useful structure for multidimensional searching. This search method is significantly more efficient than the geometric search - analogues can be selected in an order of seconds, and systematics can be calculated in minutes. First all values within the volume-limited catalog are standardized by the Milky Way properties. For example the mean Milky Way stellar mass is subtracted from the stellar mass of each galaxy in the catalog and then the resulting value is divided by the standard deviation of the Milky Way's stellar mass. This ensures that all of the data is on the same scale when put into the tree. Finally, using the Ref. [364] spatial.cKdDTree we construct a k-d tree from this scaled data.

Then, as in the original algorithm, a point can be drawn randomly from the fiducial probability distribution function of the Milky Way parameters of interest. For the purposes of this example we will consider a pair of values in stellar mass and star formation rate  $(M_*, \text{SFR})$ . This pair of values must also be standardized, as the tree contains standardized observations. Using this pair we then query the k-d tree for the 7 nearest neighbors to this pair within the tree. We chose 7 as our optimum values via k-s statistical tests for  $M_*$  and SFR as a function of neighbor number - 7 neighbors had the optimum convergence. We also employ an upper bound to the neighbor distance. All neighbors selected must be within  $6 - \sigma$  of the Milky Way mean properties.

After the 7 nearest neighbors to this draw in Milky Way parameters space are selected, a single analogue must be chosen from these neighbors. We elect to assign weights to each neighbor based on their distance in parameter space from the query point. Our weights (w)are equal to

$$w = \frac{e^{-d^2/2}}{\sum (e^{-d^2/2})},\tag{56}$$

where d is the Euclidean distance in parameter space of the neighbor (e.g.,  $d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$  for two n-dimensional vectors). Then a Milky Way analogue is selected from a random draw amongst these neighbors, using these weights. The analysis then proceeds as outlined in Ref. [219].

We present in Fig. A1 a comparison in log star formation rate (log (SFR)  $[M_{\odot}yr^{-1}]$ ) versus log mass (log ( $M_*$ )  $[M_{\odot}]$ ) space of Milky Way analogous selected from the Ref. [219] method in orange and this updated method in purple. The inner and outer ellipses mark the  $1\sigma$  and  $2\sigma$  covariance of the analogues. These ellipses overlap almost identically from both samples. In the parameter histograms we also include the fiducial Milky Way probability distribution function for comparison, denoted by a black line. The analogues from both algorithms overlap closely and match the Milky Way PDF well. In addition our optical photometric results via this new algorithm are within  $1\sigma$  of the values reported in Ref. [219] and often are almost identical. This streamlined algorithm for selecting Milky Way analogues drastically reduces how time intensive analogue selection and analysis can be. Additionally it is easy to scale the k-d tree method to more than two parameters, which will allow us to better constrain the Milky Way photometric properties. This updated code is publicly available at the Milky Way Analogue GitHub page<sup>1</sup>.

<sup>&</sup>lt;sup>1</sup>https://github.com/cfielder/Milky-Way-Analogs



Figure A1: Log star formation rate (log(SFR)  $[M_{\odot}yr^{-1}]$ ) as a function of log mass (log(M<sub>\*</sub>)  $[M_{\odot}]$ ) for the Milky Way analogues selected in Ref. [219] (orange) and in this work (purple). Points that appear darker have been selected more times as an analogue than the fainter points. The corresponding ellipses note the  $1\sigma$  and  $2\sigma$  covariance for the analogue sample which overlap almost perfectly from both sources. We also include histograms of the selected analogue's mass and star formation rate. For reference the black lines indicate the fiducial Milky Way PDF for of stellar mass and star formation rate. It is evident that the new analogue selection algorithm yields very similar results to the previous algorithm, and that the new method matches with the fiducial Milky Way PDF slightly better.

# Bibliography

- S. J. Aarseth. Dynamical evolution of clusters of galaxies, I. Monthly Notices of the Royal Astronomical Society, 126:223, 1 1963.
- [2] Louis E. Abramson, Rik J. Williams, Andrew J. Benson, Juna A. Kollmeier, and John S. Mulchaey. The Circular Velocity Function of Group Galaxies. *The Astrophysical Journal*, 793(1):49, 9 2014.
- [3]Christopher P. Ahn, Rachael Alexandroff, Carlos Allende Prieto, Friedrich Anders, Scott F. Anderson, Timothy Anderton, Brett H. Andrews, Eric Aubourg, Stephen Bailey, Fabienne A. Bastien, Julian E. Bautista, Timothy C. Beers, Alessandra Beifiori, Chad F. Bender, Andreas A. Berlind, Florian Beutler, Vaishali Bhardwaj, Jonathan C. Bird, Dmitry Bizyaev, Cullen H. Blake, Michael R. Blanton, Michael Blomqvist, John J. Bochanski, Adam S. Bolton, Arnaud Borde, Jo Bovy, Alaina Shelden Bradley, W. N. Brandt, Dorothée Brauer, J. Brinkmann, Joel R. Brownstein, Nicolás G. Busca, William Carithers, Joleen K. Carlberg, Aurelio R. Carnero, Michael A. Carr, Cristina Chiappini, S. Drew Chojnowski, Chia-Hsun Chuang, Johan Comparat, Justin R. Crepp, Stefano Cristiani, Rupert A. C. Croft, Antonio J. Cuesta, Katia Cunha, Luiz N. da Costa, Kyle S. Dawson, Nathan De Lee, Janice D. R. Dean, Timothée Delubac, Rohit Deshpande, Saurav Dhital, Anne Ealet, Garrett L. Ebelke, Edward M. Edmondson, Daniel J. Eisenstein, Courtney R. Epstein, Stephanie Escoffier, Massimiliano Esposito, Michael L. Evans, D. Fabbian, Xiaohui Fan, Ginevra Favole, Bruno Femenía Castellá, Emma Fernández Alvar, Diane Feuillet, Nurten Filiz Ak, Hayley Finley, Scott W. Fleming, Andreu Font-Ribera, Peter M. Frinchaboy, J. G. Galbraith-Frew, D. A. García-Hernández, Ana E. García Pérez, Jian Ge, R. Génova-Santos, Bruce A. Gillespie, Léo Girardi, Jonay I. González Hernández, III Gott, J. Richard, James E. Gunn, Hong Guo, Samuel Halverson, Paul Harding, David W. Harris, Sten Hasselquist, Suzanne L. Hawley, Michael Hayden, Frederick R. Hearty, Artemio Herrero Davó, Shirley Ho, David W. Hogg, Jon A. Holtzman, Klaus Honscheid, Joseph Huehnerhoff, Inese I. Ivans, Kelly M. Jackson, Peng Jiang, Jennifer A. Johnson, K. Kinemuchi, David Kirkby, Mark A. Klaene, Jean-Paul Kneib, Lars Koesterke, Ting-Wen Lan, Dustin Lang, Jean-Marc Le Goff, Alexie Leauthaud, Khee-Gan Lee, Young Sun Lee, Daniel C. Long, Craig P. Loomis, Sara Lucatello, Robert H. Lupton, Bo Ma, III Mack, Claude E., Suvrath Mahadevan, Marcio A. G. Maia, Steven R. Majewski, Elena Malanushenko, Viktor Malanushenko, A. Manchado, Marc Manera, Claudia Maraston, Daniel Margala, Sarah L. Martell, Karen L. Masters, Cameron K. McBride, Ian D. McGreer, Richard G. McMahon, Brice Ménard, Sz. Mészáros, Jordi Miralda-Escudé, Hironao Miyatake, Antonio D. Montero-Dorta. Francesco Montesano, Surhud More, Heather L. Morrison, Demitri Muna, Jeffrey A. Munn, Adam D. Myers, Duy Cuong Nguyen, Robert C. Nichol, David L. Nidever, Pasquier Noterdaeme, Sebastián E. Nuza, Julia E. O'Connell, Robert W. O'Connell,

Ross O'Connell, Matthew D. Olmstead, Daniel J. Oravetz, Russell Owen, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Kaike Pan, John K. Parejko, Prachi Parihar, Isabelle Pâris, Joshua Pepper, Will J. Percival, Ignasi Pérez-Ràfols, Hélio Dotto Perottoni, Patrick Petitjean, Matthew M. Pieri, M. H. Pinsonneault, Francisco Prada, Adrian M. Price-Whelan, M. Jordan Raddick, Mubdi Rahman, Rafael Rebolo, Beth A. Reid, Jonathan C. Richards, Rogério Riffel, Annie C. Robin, H. J. Rocha-Pinto, Constance M. Rockosi, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Graziano Rossi, Arpita Roy, J. A. Rubiño-Martin, Cristiano G. Sabiu, Ariel G. Sánchez, Basílio Santiago, Conor Sayres, Ricardo P. Schiavon, David J. Schlegel, Katharine J. Schlesinger, Sarah J. Schmidt, Donald P. Schneider, Mathias Schultheis, Kris Sellgren, Hee-Jong Seo, Yue Shen, Matthew Shetrone, Yiping Shu, Audrey E. Simmons, M. F. Skrutskie, Anže Slosar, Verne V. Smith, Stephanie A. Snedden, Jennifer S. Sobeck, Flavia Sobreira, Keivan G. Stassun, Matthias Steinmetz, Michael A. Strauss, Alina Streblyanska, Nao Suzuki, Molly E. C. Swanson, Ryan C. Terrien. Aniruddha R. Thakar, Daniel Thomas, Benjamin A. Thompson, Jeremy L. Tinker, Rita Tojeiro, Nicholas W. Troup, Jan Vandenberg, Mariana Vargas Magaña, Matteo Viel, Nicole P. Vogt, David A. Wake, Benjamin A. Weaver, David H. Weinberg, Benjamin J. Weiner, Martin White, Simon D. M. White, John C. Wilson, John P. Wisniewski, W. M. Wood-Vasey, Christophe Yèche, Donald G. York, O. Zamora, Gail Zasowski, Idit Zehavi, Gong-Bo Zhao, Zheng Zheng, and Guangtun Zhu. The Tenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Apache Point Observatory Galactic Evolution Experiment. Astronomical JournalAstrophysical Journal, Supplement, 211(2):17, April 2014.

Hiroaki Aihara, Carlos Allende Prieto, Deokkeun An, Scott F. Anderson, Éric [4]Aubourg, Eduardo Balbinot, Timothy C. Beers, Andreas A. Berlind, Steven J. Bickerton, Dmitry Bizyaev, Michael R. Blanton, John J. Bochanski, Adam S. Bolton, Jo Bovy, W. N. Brand t, J. Brinkmann, Peter J. Brown, Joel R. Brownstein, Nicolas G. Busca, Heather Campbell, Michael A. Carr, Yanmei Chen, Cristina Chiappini, Johan Comparat, Natalia Connolly, Marina Cortes, Rupert A. C. Croft, Antonio J. Cuesta, Luiz N. da Costa, James R. A. Davenport, Kyle Dawson, Saurav Dhital, Anne Ealet, Garrett L. Ebelke, Edward M. Edmondson, Daniel J. Eisenstein, Stephanie Escoffier, Massimiliano Esposito, Michael L. Evans, Xiaohui Fan, Bruno Femenía Castellá, Andreu Font-Ribera, Peter M. Frinchaboy, Jian Ge, Bruce A. Gillespie, G. Gilmore, Jonay I. González Hernández, J. Richard Gott, Andrew Gould, Eva K. Grebel, James E. Gunn, Jean-Christophe Hamilton, Paul Harding, David W. Harris, Suzanne L. Hawley, Frederick R. Hearty, Shirley Ho, David W. Hogg, Jon A. Holtzman, Klaus Honscheid, Naohisa Inada, Inese I. Ivans, Linhua Jiang, Jennifer A. Johnson, Cathy Jordan, Wendell P. Jordan, Eyal A. Kazin, David Kirkby, Mark A. Klaene, G. R. Knapp, Jean-Paul Kneib, C. S. Kochanek, Lars Koesterke, Juna A. Kollmeier, Richard G. Kron, Hubert Lampeitl, Dustin Lang, Jean-Marc Le Goff, Young Sun Lee, Yen-Ting Lin, Daniel C. Long, Craig P. Loomis, Sara Lucatello, Britt Lundgren, Robert H. Lupton, Zhibo Ma, Nicholas MacDonald, Suvrath Mahadevan, Marcio A. G. Maia, Martin Makler, Elena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Claudia Maraston, Daniel Margala, Karen L. Masters, Cameron K. McBride, Peregrine M. McGehee, Ian D. McGreer, Brice Ménard, Jordi Miralda-Escudé, Heather L. Morrison, F. Mullally, Demitri Muna, Jeffrey A. Munn, Hitoshi Murayama, Adam D. Myers, Tracy Naugle, Angelo Fausti Neto, Duy Cuong Nguyen, Robert C. Nichol, Robert W. O'Connell, Ricardo L. C. Ogando, Matthew D. Olmstead, Daniel J. Oravetz, Nikhil Padmanabhan, Nathalie Palanque-Delabrouille, Kaike Pan, Parul Pandey, Isabelle Pâris, Will J. Percival, Patrick Petitjean, Robert Pfaffenberger, Janine Pforr, Stefanie Phleps, Christophe Pichon, Matthew M. Pieri, Francisco Prada, Adrian M. Price-Whelan, M. Jordan Raddick, Beatriz H. F. Ramos, Céline Reylé, James Rich, Gordon T. Richards, Hans-Walter Rix, Annie C. Robin, Helio J. Rocha-Pinto, Constance M. Rockosi, Natalie A. Roe, Emmanuel Rollinde, Ashley J. Ross, Nicholas P. Ross, Bruno M. Rossetto, Ariel G. Sánchez, Conor Savres, David J. Schlegel, Katharine J. Schlesinger, Sarah J. Schmidt, Donald P. Schneider, Erin Sheldon, Yiping Shu, Jennifer Simmerer, Audrey E. Simmons, Thirupathi Sivarani, Stephanie A. Snedden, Jennifer S. Sobeck, Matthias Steinmetz, Michael A. Strauss, Alexander S. Szalay, Masayuki Tanaka, Aniruddha R. Thakar, Daniel Thomas, Jeremy L. Tinker, Benjamin M. Tofflemire, Rita Tojeiro, Christy A. Tremonti, Jan Vandenberg, M. Vargas Magaña, Licia Verde, Nicole P. Vogt, David A. Wake, Ji Wang, Benjamin A. Weaver, David H. Weinberg, Martin White, Simon D. M. White, Brian Yanny, Naoki Yasuda, Christophe Yeche, and Idit Zehavi. The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III. The Astrophysical Journal Supplementary Series, 193(2):29, 4 2011.

- [5] Hirotugu Akaike. A new look at the statistical model identification. pages AC-19(6):716–723. IEEE Transactions on Automatic Control, 1974.
- [6] Katherine Alatalo, Sabrina L. Cales, Philip N. Appleton, Lisa J. Kewley, Mark Lacy, Ute Lisenfeld, Kristina Nyland, and Jeffrey A. Rich. Catching Quenching Galaxies: The Nature of the WISE Infrared Transition Zone. *The Astrophysical Journal, Letters*, 794(1):L13, 10 2014.
- [7] Brandon Allgood, Ricardo A. Flores, Joel R. Primack, Andrey V. Kravtsov, Risa H. Wechsler, Andreas Faltenbacher, and James S. Bullock. The shape of dark matter haloes: dependence on mass, redshift, radius and formation. *Monthly Notices of the Royal Astronomical Society*, 367:1781–1796, 4 2006.
- [8] H. B. Ann, Mira Seo, and D. K. Ha. A Catalog of Visually Classified Galaxies in the Local (z ~ 0.01) Universe. The Astrophysical Journal Supplementary Series, 217(2):27, 4 2015.
- [9] Elaad Applebaum, Alyson M. Brooks, Charlotte R. Christensen, Ferah Munshi, Thomas R. Quinn, Sijing Shen, and Michael Tremmel. Ultrafaint Dwarfs in a Milky

Way Context: Introducing the Mint Condition DC Justice League Simulations. Astrophysical Journal, 906(2):96, 1 2021.

- [10] A. C. Atkinson. Likelihood ratios, posterior odds and information criteria. *Journal of Econometrics*, 16:15–20, 1981.
- [11] I. K. Baldry, M. L. Balogh, R. Bower, K. Glazebrook, and R. C. Nichol. Color bimodality: Implications for galaxy evolution. In Roland E. Allen, Dimitri V. Nanopoulos, and Christopher N. Pope, editors, *The New Cosmology: Conference on Strings and Cosmology*, volume 743 of *American Institute of Physics Conference Series*, pages 106–119, 12 2004.
- [12] Steven P. Bamford, Robert C. Nichol, Ivan K. Baldry, Kate Land, Chris J. Lintott, Kevin Schawinski, Anže Slosar, Alexander S. Szalay, Daniel Thomas, Mehri Torki, Dan Andreescu, Edward M. Edmondson, Christopher J. Miller, Phil Murray, M. Jordan Raddick, and Jan Vandenberg. Galaxy Zoo: the dependence of morphology and colour on environment<sup>\*</sup>. *Monthly Notices of the Royal Astronomical Society*, 393(4):1324–1352, 3 2009.
- [13] Josh Barnes and Piet Hut. A hierarchical O(N log N) force-calculation algorithm. Nature, 324(6096):446–449, 12 1986.
- [14] Joshua Barnes and George Efstathiou. Angular Momentum from Tidal Torques. Astrophysical Journal, 319:575, 8 1987.
- [15] Giuseppina Battaglia, Amina Helmi, Heather Morrison, Paul Harding, Edward W. Olszewski, Mario Mateo, Kenneth C. Freeman, John Norris, and Stephen A. Shectman. The radial velocity dispersion profile of the Galactic halo: constraining the density profile of the dark halo of the Milky Way. *Monthly Notices of the Royal Astronomical Society*, 364(2):433–442, 12 2005.
- [16] Richard Beare, Michael J. I. Brown, and Kevin Pimbblet. An Accurate New Method of Calculating Absolute Magnitudes and K-corrections Applied to the Sloan Filter Set. *The Astrophysical Journal*, 797(2):104, 12 2014.
- [17] Matthew R. Becker. Connecting Galaxies with Halos Across Cosmic Time: Stellar mass assembly distribution modeling of galaxy statistics. *arXiv e-prints*, page arXiv:1507.03605, 7 2015.

- [18] Peter S. Behroozi, Risa H. Wechsler, and Hao-Yi Wu. The ROCKSTAR Phasespace Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *Astrophysical Journal*, 762(2):109, 1 2013.
- [19] Eric F. Bell and Roelof S. de Jong. Stellar Mass-to-Light Ratios and the Tully-Fisher Relation. *The Astrophysical Journal*, 550(1):212–229, 3 2001.
- [20] Eric F. Bell, Christian Wolf, Klaus Meisenheimer, Hans-Walter Rix, Andrea Borch, Simon Dye, Martina Kleinheinrich, Lutz Wisotzki, and Daniel H. McIntosh. Nearly 5000 Distant Early-Type Galaxies in COMBO-17: A Red Sequence and Its Evolution since z~1. The Astrophysical Journal, 608(2):752–767, 6 2004.
- [21] N. Benítez, H. Ford, R. Bouwens, F. Menanteau, J. Blakeslee, C. Gronwall, G. Illingworth, G. Meurer, T. J. Broadhurst, M. Clampin, M. Franx, G. F. Hartig, D. Magee, M. Sirianni, D. R. Ardila, F. Bartko, R. A. Brown, C. J. Burrows, E. S. Cheng, N. J. G. Cross, P. D. Feldman, D. A. Golimowski, L. Infante, R. A. Kimble, J. E. Krist, M. P. Lesser, Z. Levay, A. R. Martel, G. K. Miley, M. Postman, P. Rosati, W. B. Sparks, H. D. Tran, Z. I. Tsvetanov, R. L. White, and W. Zheng. Faint Galaxies in Deep Advanced Camera for Surveys Observations. *The Astrophysical Journal Supplementary Series*, 150(1):1–18, 1 2004.
- [22] A. J. Benson, S. Cole, C. S. Frenk, C. M. Baugh, and C. G. Lacey. The nature of galaxy bias and clustering. *Monthly Notices of the Royal Astronomical Society*, 311(4):793–808, 2 2000.
- [23] A. J. Benson, C. S. Frenk, C. G. Lacey, C. M. Baugh, and S. Cole. The effects of photoionization on galaxy formation - II. Satellite galaxies in the Local Group. *Monthly Notices of the Royal Astronomical Society*, 333:177–190, 6 2002.
- [24] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. Commun. ACM, 18(9):509–517, 9 1975.
- [25] G. Besla, N. Kallivayalil, L. Hernquist, R. P. van der Marel, T. J. Cox, and D. Kereš. Simulations of the Magellanic Stream in a First Infall Scenario. Astrophysical Journal, 721:L97–L101, 10 2010.
- [26] Gurtina Besla, Nitya Kallivayalil, Lars Hernquist, Brant Robertson, T. J. Cox, Roeland P. van der Marel, and Charles Alcock. Are the Magellanic Clouds on Their First Passage about the Milky Way? Astrophysical Journal, 668:949–967, 10 2007.

- [27] P.R. Bevington and D.K. Robinson. *Data Analysis and Reduction for the Physical Sciences*. McGraw Hill, 3 edition, 2002.
- [28] Joss Bland-Hawthorn and Ortwin Gerhard. The Galaxy in Context: Structural, Kinematic, and Integrated Properties. Annual Review of Astronomy and Astrophysics, 54:529–596, 9 2016.
- [29] M. R. Blanton, M. A. Bershady, B. Abolfathi, F. D. Albareti, C. Allende Prieto, A. Almeida, J. Alonso-García, F. Anders, S. F. Anderson, B. Andrews, and et al. Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe. Astronomical Journal, 154:28, 7 2017.
- [30] Michael R. Blanton, David W. Hogg, Neta A. Bahcall, Ivan K. Baldry, J. Brinkmann, István Csabai, Daniel Eisenstein, Masataka Fukugita, James E. Gunn, Željko Ivezić, D. Q. Lamb, Robert H. Lupton, Jon Loveday, Jeffrey A. Munn, R. C. Nichol, Sadanori Okamura, David J. Schlegel, Kazuhiro Shimasaku, Michael A. Strauss, Michael S. Vogeley, and David H. Weinberg. The Broadband Optical Properties of Galaxies with Redshifts 0.02<z<0.22. Astrophysical Journal, 594(1):186–207, September 2003.</p>
- [31] Michael R. Blanton and Sam Roweis. K-Corrections and Filter Transformations in the Ultraviolet, Optical, and Near-Infrared. Astronomical Journal, 133(2):734–754, 2 2007.
- [32] Leo Blitz and David N. Spergel. Direct Evidence for a Bar at the Galactic Center. *The Astrophysical Journal*, 379:631, 10 1991.
- [33] Asa F. L. Bluck, J. Trevor Mendel, Sara L. Ellison, David R. Patton, Luc Simard, Bruno M. B. Henriques, Paul Torrey, Hossen Teimoorinia, Jorge Moreno, and Else Starkenburg. The impact of galactic properties and environment on the quenching of central and satellite galaxies: a comparison between SDSS, Illustris and L-Galaxies. *Monthly Notices of the Royal Astronomical Society*, 462(3):2559–2586, 11 2016.
- [34] George R. Blumenthal, S. M. Faber, Ricardo Flores, and Joel R. Primack. Contraction of Dark Matter Galactic Halos Due to Baryonic Infall. Astrophysical Journal, 301:27, 2 1986.
- [35] N. Boardman, G. Zasowski, A. Seth, J. Newman, B. Andrews, M. Bershady, J. Bird, C. Chiappini, C. Fielder, A. Fraser-McKelvie, A. Jones, T. Licquia, K. L. Masters, I. Minchev, R. P. Schiavon, J. R. Brownstein, N. Drory, and R. R. Lane. Milky Way analogues in MaNGA: multiparameter homogeneity and comparison to the Milky Way. *Monthly Notices of the Royal Astronomical Society*, 491(3):3672–3701, 1 2020.

- [36] Nicholas Fraser Boardman, Gail Zasowski, Jeffrey Newman, Brett Andrews, Catherine Fielder, Matthew Bershady, Jonathan Brinkmann, Niv Drory, Dhanesh Krishnarao, Richard Lane, Ted Mackereth, Karen Masters, and Guy Stringfellow. Are the Milky Way and Andromeda unusual? A comparison with Milky Way and Andromeda analogues. *Monthly Notices of the Royal Astronomical Society*, 498(4):4943–4954, 11 2020.
- [37] Sebastian Bocquet, Katrin Heitmann, Salman Habib, Earl Lawrence, Thomas Uram, Nicholas Frontiere, Adrian Pope, and Hal Finkel. The Mira-Titan Universe. III. Emulation of the Halo Mass Function. *The Astrophysical Journal*, 901(1):5, 9 2020.
- [38] M. Boquien, D. Burgarella, Y. Roehlly, V. Buat, L. Ciesla, D. Corre, A. K. Inoue, and H. Salas. CIGALE: a python Code Investigating GALaxy Emission. Astronomy and Astrophysics, 622:A103, February 2019.
- [39] A. Boselli, G. Gavazzi, J. Donas, and M. Scodeggio. 1.65 Micron (H Band) Surface Photometry of Galaxies. VI. The History of Star Formation in Normal Late-Type Galaxies. Astronomical Journal, 121(2):753–767, 2 2001.
- [40] Mia S. Bovill and Massimo Ricotti. Pre-Reionization Fossils, Ultra-Faint Dwarfs, and the Missing Galactic Satellite Problem. *Astrophysical Journal*, 693:1859–1870, 3 2009.
- [41] Jo Bovy and Hans-Walter Rix. A Direct Dynamical Measurement of the Milky Way's Disk Surface Density Profile, Disk Scale Length, and Dark Matter Profile at 4 kpc <~R <~9 kpc. *The Astrophysical Journal*, 779(2):115, 12 2013.
- [42] Jo Bovy, Hans-Walter Rix, and David W. Hogg. The Milky Way Has No Distinct Thick Disk. *Astrophysical Journal*, 751(2):131, 6 2012.
- [43] Michael Boylan-Kolchin, James S. Bullock, and Manoj Kaplinghat. Too big to fail? The puzzling darkness of massive Milky Way subhaloes. *Monthly Notices of the Royal Astronomical Society*, 415:L40–L44, 7 2011.
- [44] Michael Boylan-Kolchin, James S. Bullock, and Manoj Kaplinghat. The Milky Way's bright satellites as an apparent failure of ΛCDM. Monthly Notices of the Royal Astronomical Society, 422:1203–1218, 5 2012.
- [45] Michael Boylan-Kolchin, Volker Springel, Simon D. M. White, and Adrian Jenkins. There's no place like home? Statistics of Milky Way-mass dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 406:896–912, 8 2010.

- [46] J. Brinchmann, S. Charlot, S. D. M. White, C. Tremonti, G. Kauffmann, T. Heckman, and J. Brinkmann. The physical properties of star-forming galaxies in the low-redshift Universe. *Monthly Notices of the Royal Astronomical Society*, 351(4):1151–1179, 7 2004.
- [47] Alyson M. Brooks, Michael Kuhlen, Adi Zolotov, and Dan Hooper. A Baryonic Solution to the Missing Satellites Problem. *Astrophysical Journal*, 765:22, 3 2013.
- [48] Michael J. I. Brown, Arjun Dey, Buell T. Jannuzi, Kate Brand, Andrew J. Benson, Mark Brodwin, Darren J. Croton, and Peter R. Eisenhardt. The Evolving Luminosity Function of Red Galaxies. Astrophysical Journal, 654(2):858–877, January 2007.
- [49] Michael J. I. Brown, John Moustakas, Robert C. Kennicutt, Nicolas J. Bonne, Huib T. Intema, Francesco de Gasperin, Mederic Boquien, T. H. Jarrett, Michelle E. Cluver, J. D. T. Smith, Elisabete da Cunha, Masatoshi Imanishi, Lee Armus, Bernhard R. Brandl, and J. E. G. Peek. Calibration of Ultraviolet, Mid-infrared, and Radio Star Formation Rate Indicators. *The Astrophysical Journal*, 847(2):136, 10 2017.
- [50] Michael J. I. Brown, John Moustakas, J. D. T. Smith, Elisabete da Cunha, T. H. Jarrett, Masatoshi Imanishi, Lee Armus, Bernhard R. Brandl, and J. E. G. Peek. An Atlas of Galaxy Spectral Energy Distributions from the Ultraviolet to the Mid-infrared. *The Astrophysical Journal Supplementary Series*, 212(2):18, 6 2014.
- [51] G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. Monthly Notices of the Royal Astronomical Society, 344(4):1000–1028, October 2003.
- [52] Greg L. Bryan and Michael L. Norman. Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons. *The Astrophysical Journal*, 495(1):80–99, 3 1998.
- [53] Sven Buder, Karin Lind, Melissa K. Ness, Diane K. Feuillet, Danny Horta, Stephanie Monty, Tobias Buck, Thomas Nordlander, Joss Bland-Hawthorn, Andrew R. Casey, Gayandhi M. de Silva, Valentina D'Orazi, Ken C. Freeman, Michael R. Hayden, Janez Kos, Sarah L. Martell, Geraint F. Lewis, Jane Lin, Katharine J. Schlesinger, Sanjib Sharma, Jeffrey D. Simpson, Dennis Stello, Daniel B. Zucker, Tomaž Zwitter, Ioana Ciucă, Jonathan Horner, Chiaki Kobayashi, Yuan-Sen Ting, Rosemary F. G. Wyse, and Galah Collaboration Wyse. The GALAH Survey: chemical tagging and chronochemodynamics of accreted halo stars with GALAH+ DR3 and Gaia eDR3. *Monthly Notices of the Royal Astronomical Society*, 510(2):2407–2436, 2 2022.

- [54] J. S. Bullock, A. Dekel, T. S. Kolatt, A. V. Kravtsov, A. A. Klypin, C. Porciani, and J. R. Primack. A Universal Angular Momentum Profile for Galactic Halos. *Astro*physical Journal, 555:240–257, 7 2001.
- [55] J. S. Bullock, T. S. Kolatt, Y. Sigad, R. S. Somerville, A. V. Kravtsov, A. A. Klypin, J. R. Primack, and A. Dekel. Profiles of dark haloes: evolution, scatter and environment. *Monthly Notices of the Royal Astronomical Society*, 321(3):559–575, 3 2001.
- [56] James S. Bullock. Notes on the Missing Satellites Problem. *arXiv e-prints*, page arXiv:1009.4505, 9 2010.
- [57] James S. Bullock and Michael Boylan-Kolchin. Small-Scale Challenges to the ΛCDM Paradigm. Annual Review of Astronomy and Astrophysics, 55:343–387, 8 2017.
- [58] James S. Bullock, Andrey V. Kravtsov, and David H. Weinberg. Reionization and the Abundance of Galactic Satellites. *Astrophysical Journal*, 539:517–521, 8 2000.
- [59] K. Bundy, M. A. Bershady, D. R. Law, R. Yan, N. Drory, N. MacDonald, D. A. Wake, B. Cherinka, J. R. Sánchez-Gallego, A.-M. Weijmans, D. Thomas, C. Tremonti, K. Masters, L. Coccato, A. M. Diamond-Stanic, A. Aragón-Salamanca, V. Avila-Reese, C. Badenes, J. Falcón-Barroso, F. Belfiore, D. Bizyaev, G. A. Blanc, J. Bland-Hawthorn, M. R. Blanton, J. R. Brownstein, N. Byler, M. Cappellari, C. Conroy, A. A. Dutton, E. Emsellem, J. Etherington, P. M. Frinchaboy, H. Fu, J. E. Gunn, P. Harding, E. J. Johnston, G. Kauffmann, K. Kinemuchi, M. A. Klaene, J. H. Knapen, A. Leauthaud, C. Li, L. Lin, R. Maiolino, V. Malanushenko, E. Malanushenko, S. Mao, C. Maraston, R. M. McDermid, M. R. Merrifield, R. C. Nichol, D. Oravetz, K. Pan, J. K. Parejko, S. F. Sanchez, D. Schlegel, A. Simmons, O. Steele, M. Steinmetz, K. Thanjavur, B. A. Thompson, J. L. Tinker, R. C. E. van den Bosch, K. B. Westfall, D. Wilkinson, S. Wright, T. Xiao, and K. Zhang. Overview of the SDSS-IV MaNGA Survey: Mapping nearby Galaxies at Apache Point Observatory. Astrophysical Journal, 798:7, 1 2015.
- [60] Michael T. Busha, Risa H. Wechsler, Peter S. Behroozi, Brian F. Gerke, Anatoly A. Klypin, and Joel R. Primack. Statistics of Satellite Galaxies around Milky-Way-like Hosts. Astrophysical Journal, 743:117, 12 2011.
- [61] Michele Cappellari. Structure and Kinematics of Early-Type Galaxies from Integral Field Spectroscopy. Annual Review of Astronomy and Astrophysics, 54:597–665, 9 2016.

- [62] Jason A. Cardelli, Geoffrey C. Clayton, and John S. Mathis. The Relationship between Infrared, Optical, and Ultraviolet Extinction. *The Astrophysical Journal*, 345:245, 10 1989.
- [63] Edoardo Carlesi, Yehuda Hoffman, Stefan Gottlöber, Noam I. Libeskind, Alexander Knebe, Gustavo Yepes, and Sergey V. Pilipenko. On the mass assembly history of the Local Group. *Monthly Notices of the Royal Astronomical Society*, 491(2):1531–1539, 1 2020.
- [64] Scott G. Carlsten, Jenny E. Greene, Annika H. G. Peter, Johnny P. Greco, and Rachael L. Beaton. Radial Distributions of Dwarf Satellite Systems in the Local Volume. *arXiv e-prints*, page arXiv:2006.02444, 6 2020.
- [65] Riccardo Catena and Piero Ullio. A novel determination of the local dark matter density. *Journal of Cosmology and Astro-Particle Physics*, 2010:004, 8 2010.
- [66] Daniel Ceverino and Anatoly Klypin. The Role of Stellar Feedback in the Formation of Galaxies. *Astrophysical Journal*, 695:292–309, 4 2009.
- [67] S. Chapelon, T. Contini, and E. Davoust. Starbursts in barred spiral galaxies. V. Morphological analysis of bars. *Astronomy and Astrophysics*, 345:81–92, 5 1999.
- [68] Edmond Cheung, E. Athanassoula, Karen L. Masters, Robert C. Nichol, A. Bosma, Eric F. Bell, S. M. Faber, David C. Koo, Chris Lintott, Thomas Melvin, Kevin Schawinski, Ramin A. Skibba, and Kyle W. Willett. Galaxy Zoo: Observing Secular Evolution through Bars. *The Astrophysical Journal*, 779(2):162, 12 2013.
- [69] Hillary L. Child, Salman Habib, Katrin Heitmann, Nicholas Frontiere, Hal Finkel, Adrian Pope, and Vitali Morozov. Halo Profiles and the Concentration-Mass Relation for a ΛCDM Universe. Astrophysical Journal, 859(1):55, 5 2018.
- [70] Igor V. Chilingarian, Anne-Laure Melchior, and Ivan Yu. Zolotukhin. Analytical approximations of K-corrections in optical and near-infrared bands. *Monthly Notices of the Royal Astronomical Society*, 405(3):1409–1420, July 2010.
- [71] Igor V. Chilingarian and Ivan Yu. Zolotukhin. A universal ultraviolet-optical colourcolour-magnitude relation of galaxies. *Monthly Notices of the Royal Astronomical Society*, 419(2):1727–1739, 1 2012.

- [72] Jungyeon Cho and Changbom Park. Internal Extinction in the Sloan Digital Sky Survey Late-Type Galaxies. *The Astrophysical Journal*, 693(2):1045–1055, 3 2009.
- [73] S. Cole, A. Aragon-Salamanca, C. S. Frenk, J. F. Navarro, and S. E. Zepf. A recipe for galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 271:781–806, 12 1994.
- [74] G. D. Coleman, C. C. Wu, and D. W. Weedman. Colors and magnitudes predicted for high redshift galaxies. Astronomical JournalAstrophysical Journal, Supplement, 43:393–416, July 1980.
- [75] Matthew Colless, Gavin Dalton, Steve Maddox, Will Sutherland, Peder Norberg, Shaun Cole, Joss Bland-Hawthorn, Terry Bridges, Russell Cannon, Chris Collins, Warrick Couch, Nicholas Cross, Kathryn Deeley, Roberto De Propris, Simon P. Driver, George Efstathiou, Richard S. Ellis, Carlos S. Frenk, Karl Glazebrook, Carole Jackson, Ofer Lahav, Ian Lewis, Stuart Lumsden, Darren Madgwick, John A. Peacock, Bruce A. Peterson, Ian Price, Mark Seaborne, and Keith Taylor. The 2dF Galaxy Redshift Survey: spectra and redshifts. *Monthly Notices of the Royal Astronomical Society*, 328(4):1039–1063, 12 2001.
- [76] Julia M. Comerford and Priyamvada Natarajan. The observed concentration-mass relation for galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 379(1):190–200, 7 2007.
- [77] A. J. Connolly, A. S. Szalay, M. A. Bershady, A. L. Kinney, and D. Calzetti. Spectral Classification of Galaxies: an Orthogonal Approach. Astronomical Journal, 110:1071, September 1995.
- [78] Charlie Conroy. Modeling the Panchromatic Spectral Energy Distributions of Galaxies. Annual Review of Astronomy and Astrophysics, 51(1):393–455, 8 2013.
- [79] Charlie Conroy, David Schiminovich, and Michael R. Blanton. Dust Attenuation in Disk-dominated Galaxies: Evidence for the 2175 Å Dust Feature. The Astrophysical Journal, 718(1):184–198, 7 2010.
- [80] A. Cooray and R. Sheth. Halo models of large scale structure. *Physics Reports*, 372:1–129, 12 2002.
- [81] L. Cortese. Are passive red spirals truly passive?. The current star formation activity of optically red disc galaxies. *Astronomy and Astrophysics*, 543:A132, 7 2012.

- [82] Biwei Dai, Brant E. Robertson, and Piero Madau. Around the Way: Testing ΛCDM with Milky Way Stellar Stream Constraints. *Astrophysical Journal*, 858(2):73, 5 2018.
- [83] D. A. Dale, D. O. Cook, H. Roussel, J. A. Turner, L. Armus, A. D. Bolatto, M. Boquien, M. J. I. Brown, D. Calzetti, I. De Looze, M. Galametz, K. D. Gordon, B. A. Groves, T. H. Jarrett, G. Helou, R. Herrera-Camus, J. L. Hinz, L. K. Hunt, R. C. Kennicutt, E. J. Murphy, A. Rest, K. M. Sandstrom, J. D. T. Smith, F. S. Tabatabaei, and C. D. Wilson. Updated 34-band Photometry for the Sings/KINGFISH Samples of Nearby Galaxies. *The Astrophysical Journal*, 837(1):90, 3 2017.
- [84] Marc Davis, Sandra M. Faber, Jeffrey Newman, Andrew C. Phillips, Richard S. Ellis, Charles C. Steidel, C. Conselice, Alison L. Coil, D. P. Finkbeiner, David C. Koo, Puragra Guhathakurta, B. Weiner, Ricardo Schiavon, C. Willmer, Nicholas Kaiser, Gerard A. Luppino, Gregory Wirth, Andrew Connolly, Peter Eisenhardt, M. Cooper, and B. Gerke. Science Objectives and Early Results of the DEEP2 Redshift Survey. In Puragra Guhathakurta, editor, *Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II*, volume 4834 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pages 161–172, February 2003.
- [85] Maria E. de Rossi, Patricia B. Tissera, Gabriella De Lucia, and Guinevere Kauffmann. Milky Way type galaxies in a ACDM cosmology. *Monthly Notices of the Royal Astronomical Society*, 395(1):210–217, 5 2009.
- [86] G. M. De Silva, K. C. Freeman, J. Bland-Hawthorn, S. Martell, E. Wylie de Boer, M. Asplund, S. Keller, S. Sharma, D. B. Zucker, T. Zwitter, B. Anguiano, C. Bacigalupo, D. Bayliss, M. A. Beavis, M. Bergemann, S. Campbell, R. Cannon, D. Carollo, L. Casagrande, A. R. Casey, G. Da Costa, V. D'Orazi, A. Dotter, L. Duong, A. Heger, M. J. Ireland, P. R. Kafle, J. Kos, J. Lattanzio, G. F. Lewis, J. Lin, K. Lind, U. Munari, D. M. Nataf, S. O'Toole, Q. Parker, W. Reid, K. J. Schlesinger, A. Sheinis, J. D. Simpson, D. Stello, Y. S. Ting, G. Traven, F. Watson, R. Wittenmyer, D. Yong, and M. Žerjal. The GALAH survey: scientific motivation. *Montly Notices of the Royal Astronomical Society*, 449(3):2604–2617, 5 2015.
- [87] A. J. Deason, V. Belokurov, N. W. Evans, and J. An. Broken degeneracies: the rotation curve and velocity anisotropy of the Milky Way halo. *Monthly Notices of the Royal Astronomical Society*, 424:L44–L48, 7 2012.
- [88] Victor P. Debattista, Oscar A. Gonzalez, Robyn E. Sanderson, Kareem El-Badry, Shea Garrison-Kimmel, Andrew Wetzel, Claude-André Faucher-Giguère, and Philip F. Hopkins. Formation, vertex deviation, and age of the Milky Way's bulge: input from a cosmological simulation with a late-forming bar. *Monthly Notices of the Royal* Astronomical Society, 485(4):5073–5085, 6 2019.

- [89] Walter Dehnen and Dean E. McLaughlin. Dynamical insight into dark matter haloes. Monthly Notices of the Royal Astronomical Society, 363(4):1057–1068, 11 2005.
- [90] Avishai Dekel, Guy Ishai, Aaron A. Dutton, and Andrea V. Maccio. Dark-matter halo profiles of a general cusp/core with analytic velocity and potential. *Monthly Notices of the Royal Astronomical Society*, 468(1):1005–1022, 6 2017.
- [91] Avishai Dekel and Joseph Silk. The Origin of Dwarf Galaxies, Cold Dark Matter, and Biased Galaxy Formation. *Astrophysical Journal*, 303:39, 4 1986.
- [92] A. Del Popolo. The Cusp/Core Problem and the Secondary Infall Model. Astrophysical Journal, 698:2093–2113, 6 2009.
- [93] Antonino Del Popolo and Morgan Le Delliou. Small Scale Problems of the ΛCDM Model: A Short Review. Galaxies, 5:17, 2 2017.
- [94] Giulia Despali, Simona Vegetti, Simon D. M. White, Carlo Giocoli, and Frank C. van den Bosch. Modelling the line-of-sight contribution in substructure lensing. *Monthly Notices of the Royal Astronomical Society*, 475(4):5424–5442, 4 2018.
- [95] Nicholas Devereux. The Spatial Distribution of 10 Micron Luminosity in Spiral Galaxies. *The Astrophysical Journal*, 323:91, 12 1987.
- [96]Arjun Dey, David J. Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, Joseph R. Findlay, Doug Finkbeiner, David Herrera, Stéphanie Juneau, Martin Landriau, Michael Levi, Ian McGreer, Aaron Meisner, Adam D. Myers, John Moustakas, Peter Nugent, Anna Patej, Edward F. Schlaffy, Alistair R. Walker, Francisco Valdes, Benjamin A. Weaver, Christophe Yèche, Hu Zou, Xu Zhou, Behzad Abareshi, T. M. C. Abbott, Bela Abolfathi, C. Aguilera, Shadab Alam, Lori Allen, A. Alvarez, James Annis, Behzad Ansarinejad, Marie Aubert, Jacqueline Beechert, Eric F. Bell, Segev Y. BenZvi, Florian Beutler, Richard M. Bielby, Adam S. Bolton, César Briceño, Elizabeth J. Buckley-Geer, Karen Butler, Annalisa Calamida, Raymond G. Carlberg, Paul Carter, Ricard Casas, Francisco J. Castander, Yumi Choi, Johan Comparat, Elena Cukanovaite, Timothée Delubac, Kaitlin DeVries, Sharmila Dey, Govinda Dhungana, Mark Dickinson, Zhejie Ding, John B. Donaldson, Yutong Duan, Christopher J. Duckworth, Sarah Eftekharzadeh, Daniel J. Eisenstein, Thomas Etourneau, Parker A. Fagrelius, Jay Farihi, Mike Fitzpatrick, Andreu Font-Ribera, Leah Fulmer, Boris T. Gänsicke, Enrique Gaztanaga, Koshy George, David W. Gerdes, Satya Gontcho A. Gontcho, Claudio Gorgoni, Gregory Green, Julien Guy, Diane Harmer, M. Hernand ez, Klaus Honscheid, Lijuan Wendy Huang, David J. James, Buell T. Jannuzi, Linhua Jiang, Richard Joyce, Armin Karcher, Sonia Karkar, Robert

Kehoe, Jean-Paul Kneib, Andrea Kueter-Young, Ting-Wen Lan, Tod R. Lauer, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Michael Lesser, Laurence Perreault Levasseur, Ting S. Li, Justin L. Mann, Robert Marshall, C. E. Martínez-Vázquez, Paul Martini, Hélion du Mas des Bourboux, Sean McManus, Tobias Gabriel Meier, Brice Ménard, Nigel Metcalfe, Andrea Muñoz-Gutiérrez, Joan Najita, Kevin Napier, Gautham Narayan, Jeffrey A. Newman, Jundan Nie, Brian Nord, Dara J. Norman, Knut A. G. Olsen, Anthony Paat, Nathalie Palanque-Delabrouille, Xiyan Peng, Claire L. Poppett, Megan R. Poremba, Abhishek Prakash, David Rabinowitz, Anand Raichoor, Mehdi Rezaie, A. N. Robertson, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Gregory Rudnick, Sasha Safonova, Abhijit Saha, F. Javier Sánchez, Elodie Savary, Heidi Schweiker, Adam Scott, Hee-Jong Seo, Huanyuan Shan, David R. Silva, Zachary Slepian, Christian Soto, David Sprayberry, Ryan Staten, Coley M. Stillman, Robert J. Stupak, David L. Summers, Suk Sien Tie, H. Tirado, Mariana Vargas-Magaña, A. Katherina Vivas, Risa H. Wechsler, Doug Williams, Jinyi Yang, Qian Yang, Tolga Yapici, Dennis Zaritsky, A. Zenteno, Kai Zhang, Tianmeng Zhang, Rongpu Zhou, and Zhimin Zhou. Overview of the DESI Legacy Imaging Surveys. Astronomical Journal, 157(5):168, 5 2019.

- [97] Arianna Di Cintio, Chris B. Brook, Aaron A. Dutton, Andrea V. Macciò, Greg S. Stinson, and Alexander Knebe. A mass-dependent density profile for dark matter haloes including the influence of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 441(4):2986–2995, 7 2014.
- [98] S. Díaz-García, F. D. Moyano, S. Comerón, J. H. Knapen, H. Salo, and A. Y. K. Bouquin. Distribution of star formation in galactic bars as seen with H $\alpha$  and stacked GALEX UV imaging. *Astronomy and Astrophysics*, 644:A38, 12 2020.
- [99] J. Diemand, M. Kuhlen, P. Madau, M. Zemp, B. Moore, D. Potter, and J. Stadel. Clumps and streams in the local dark matter distribution. *Nature*, 454:735–738, 8 2008.
- [100] Jürg Diemand, Michael Kuhlen, and Piero Madau. Formation and Evolution of Galaxy Dark Matter Halos and Their Substructure. Astrophysical Journal, 667:859–877, 10 2007.
- [101] Benedikt Diemer and Andrey V. Kravtsov. A Universal Model for Halo Concentrations. Astrophysical Journal, 799(1):108, 1 2015.
- [102] Elena D'Onghia, Volker Springel, Lars Hernquist, and Dusan Keres. Substructure Depletion in the Milky Way Halo by the Disk. Astrophysical Journal, 709(2):1138– 1147, 2 2010.

- [103] A. G. Doroshkevich. Spatial structure of perturbations and origin of galactic rotation in fluctuation theory. *Astrophysics*, 6:320–330, 10 1970.
- S. P. Driver, D. T. Hill, L. S. Kelvin, A. S. G. Robotham, J. Liske, P. Norberg, I. K. Baldry, S. P. Bamford, A. M. Hopkins, J. Loveday, J. A. Peacock, E. Andrae, J. Bland-Hawthorn, S. Brough, M. J. I. Brown, E. Cameron, J. H. Y. Ching, M. Colless, C. J. Conselice, S. M. Croom, N. J. G. Cross, R. de Propris, S. Dye, M. J. Drinkwater, S. Ellis, Alister W. Graham, M. W. Grootes, M. Gunawardhana, D. H. Jones, E. van Kampen, C. Maraston, R. C. Nichol, H. R. Parkinson, S. Phillipps, K. Pimbblet, C. C. Popescu, M. Prescott, I. G. Roseboom, E. M. Sadler, A. E. Sansom, R. G. Sharp, D. J. B. Smith, E. Taylor, D. Thomas, R. J. Tuffs, D. Wijesinghe, L. Dunne, C. S. Frenk, M. J. Jarvis, B. F. Madore, M. J. Meyer, M. Seibert, L. Staveley-Smith, W. J. Sutherland, and S. J. Warren. Galaxy and Mass Assembly (GAMA): survey diagnostics and core data release. *Monthly Notices of the Royal Astronomical Society*, 413(2):971–995, 5 2011.
- [105] Simon P. Driver, Cristina C. Popescu, Richard J. Tuffs, Jochen Liske, Alister W. Graham, Paul D. Allen, and Roberto de Propris. The Millennium Galaxy Catalogue: the B-band attenuation of bulge and disc light and the implied cosmic dust and stellar mass densities. *Monthly Notices of the Royal Astronomical Society*, 379(3):1022–1036, 8 2007.
- [106] A. Drlica-Wagner, K. Bechtol, E. S. Rykoff, E. Luque, A. Queiroz, Y.-Y. Mao, R. H. Wechsler, J. D. Simon, B. Santiago, B. Yanny, E. Balbinot, S. Dodelson, A. Fausti Neto, D. J. James, T. S. Li, M. A. G. Maia, J. L. Marshall, A. Pieres, K. Stringer, A. R. Walker, T. M. C. Abbott, F. B. Abdalla, S. Allam, A. Benoit-Lévy, G. M. Bernstein, E. Bertin, D. Brooks, E. Buckley-Geer, D. L. Burke, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, M. Crocce, L. N. da Costa, S. Desai, H. T. Diehl, J. P. Dietrich, P. Doel, T. F. Eifler, A. E. Evrard, D. A. Finley, B. Flaugher, P. Fosalba, J. Frieman, E. Gaztanaga, D. W. Gerdes, D. Gruen, R. A. Gruendl, G. Gutierrez, K. Honscheid, K. Kuehn, N. Kuropatkin, O. Lahav, P. Martini, R. Miquel, B. Nord, R. Ogando, A. A. Plazas, K. Reil, A. Roodman, M. Sako, E. Sanchez, V. Scarpine, M. Schubnell, I. Sevilla-Noarbe, R. C. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, D. Tucker, V. Vikram, W. Wester, Y. Zhang, J. Zuntz, and DES Collaboration. Eight Ultra-faint Galaxy Candidates Discovered in Year Two of the Dark Energy Survey. Astrophysical Journal, 813:109, 11 2015.
- [107] Alan R. Duffy, Joop Schaye, Scott T. Kay, and Claudio Dalla Vecchia. Dark matter halo concentrations in the Wilkinson Microwave Anisotropy Probe year 5 cosmology. *Monthly Notices of the Royal Astronomical Society*, 390(1):L64–L68, 10 2008.

- [108] Aaron A. Dutton and Andrea V. Macciò. Cold dark matter haloes in the Planck era: evolution of structural parameters for Einasto and NFW profiles. *Monthly Notices of* the Royal Astronomical Society, 441(4):3359–3374, 7 2014.
- [109] Aaron A. Dutton and Frank C. van den Bosch. The angular momentum of disc galaxies: implications for gas accretion, outflows, and dynamical friction. *Monthly Notices of the Royal Astronomical Society*, 421:608–620, 3 2012.
- [110] Yu. N. Efremov. On the spiral structure of the Milky Way Galaxy. Astronomy Reports, 55(2):108–122, 2 2011.
- [111] G. Efstathiou. Suppressing the formation of dwarf galaxies via photoionization. Monthly Notices of the Royal Astronomical Society, 256:43P-47P, 5 1992.
- [112] G. Efstathiou and B. J. T. Jones. The rotation of galaxies: numerical investigations of the tidal torque theory. *Monthly Notices of the Royal Astronomical Society*, 186:133– 144, 1 1979.
- [113] O. J. Eggen, D. Lynden-Bell, and A. R. Sandage. Evidence from the motions of old stars that the Galaxy collapsed. *Astrophysical Journal*, 136:748, 11 1962.
- [114] J. Einasto. On the Construction of a Composite Model for the Galaxy and on the Determination of the System of Galactic Parameters. *Trudy Astrofizicheskogo Instituta Alma-Ata*, 5:87–100, January 1965.
- Daniel J. Eisenstein, David H. Weinberg, Eric Agol, Hiroaki Aihara, Carlos Allende 115 Prieto, Scott F. Anderson, James A. Arns, Éric Aubourg, Stephen Bailey, Eduardo Balbinot, Robert Barkhouser, Timothy C. Beers, Andreas A. Berlind, Steven J. Bickerton, Dmitry Bizyaev, Michael R. Blanton, John J. Bochanski, Adam S. Bolton, Casey T. Bosman, Jo Bovy, W. N. Brandt, Ben Breslauer, Howard J. Brewington, J. Brinkmann, Peter J. Brown, Joel R. Brownstein, Dan Burger, Nicolas G. Busca, Heather Campbell, Phillip A. Cargile, William C. Carithers, Joleen K. Carlberg, Michael A. Carr, Liang Chang, Yanmei Chen, Cristina Chiappini, Johan Comparat, Natalia Connolly, Marina Cortes, Rupert A. C. Croft, Katia Cunha, Luiz N. da Costa, James R. A. Davenport, Kyle Dawson, Nathan De Lee, Gustavo F. Porto de Mello, Fernando de Simoni, Janice Dean, Saurav Dhital, Anne Ealet, Garrett L. Ebelke, Edward M. Edmondson, Jacob M. Eiting, Stephanie Escoffier, Massimiliano Esposito, Michael L. Evans, Xiaohui Fan, Bruno Femenía Castellá, Leticia Dutra Ferreira, Greg Fitzgerald, Scott W. Fleming, Andreu Font-Ribera, Eric B. Ford, Peter M. Frinchaboy, Ana Elia García Pérez, B. Scott Gaudi, Jian Ge, Luan Ghezzi, Bruce A. Gillespie, G. Gilmore, Léo Girardi, J. Richard Gott, Andrew Gould, Eva K.

Grebel, James E. Gunn, Jean-Christophe Hamilton, Paul Harding, David W. Harris, Suzanne L. Hawley, Frederick R. Hearty, Joseph F. Hennawi, Jonay I. González Hernández, Shirley Ho, David W. Hogg, Jon A. Holtzman, Klaus Honscheid, Naohisa Inada, Inese I. Ivans, Linhua Jiang, Peng Jiang, Jennifer A. Johnson, Cathy Jordan, Wendell P. Jordan, Guinevere Kauffmann, Eyal Kazin, David Kirkby, Mark A. Klaene, G. R. Knapp, Jean-Paul Kneib, C. S. Kochanek, Lars Koesterke, Juna A. Kollmeier, Richard G. Kron, Hubert Lampeitl, Dustin Lang, James E. Lawler, Jean-Marc Le Goff, Brian L. Lee, Young Sun Lee, Jarron M. Leisenring, Yen-Ting Lin, Jian Liu, Daniel C. Long, Craig P. Loomis, Sara Lucatello, Britt Lundgren, Robert H. Lupton, Bo Ma, Zhibo Ma, Nicholas MacDonald, Claude Mack, Suvrath Mahadevan, Marcio A. G. Maia, Steven R. Majewski, Martin Makler, Elena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Claudia Maraston, Daniel Margala, Paul Maseman, Karen L. Masters, Cameron K. McBride, Patrick McDonald, Ian D. McGreer, Richard G. McMahon, Olga Mena Requejo, Brice Ménard, Jordi Miralda-Escudé, Heather L. Morrison, Fergal Mullally, Demitri Muna, Hitoshi Murayama, Adam D. Myers, Tracy Naugle, Angelo Fausti Neto, Duy Cuong Nguyen, Robert C. Nichol, David L. Nidever, Robert W. O'Connell, Ricardo L. C. Ogando, Matthew D. Olmstead, Daniel J. Oravetz, Nikhil Padmanabhan, Martin Paegert, Nathalie Palanque-Delabrouille, Kaike Pan, Parul Pandey, John K. Parejko, Isabelle Pâris, Paulo Pellegrini, Joshua Pepper, Will J. Percival, Patrick Petitjean, Robert Pfaffenberger, Janine Pforr, Stefanie Phleps, Christophe Pichon, Matthew M. Pieri, Francisco Prada, Adrian M. Price-Whelan, M. Jordan Raddick, Beatriz H. F. Ramos, I. Neill Reid, Celine Reyle, James Rich, Gordon T. Richards, George H. Rieke, Marcia J. Rieke. Hans-Walter Rix, Annie C. Robin, Helio J. Rocha-Pinto, Constance M. Rockosi, Natalie A. Roe, Emmanuel Rollinde, Ashley J. Ross, Nicholas P. Ross, Bruno Rossetto, Ariel G. Sánchez, Basilio Santiago, Conor Sayres, Ricardo Schiavon, David J. Schlegel, Katharine J. Schlesinger, Sarah J. Schmidt, Donald P. Schneider, Kris Sellgren, Alaina Shelden, Erin Sheldon, Matthew Shetrone, Yiping Shu, John D. Silverman, Jennifer Simmerer, Audrey E. Simmons, Thirupathi Sivarani, M. F. Skrutskie, Anže Slosar, Stephen Smee, Verne V. Smith, Stephanie A. Snedden, Keivan G. Stassun, Oliver Steele, Matthias Steinmetz, Mark H. Stockett, Todd Stollberg, Michael A. Strauss. Alexander S. Szalay, Masayuki Tanaka, Aniruddha R. Thakar, Daniel Thomas, Jeremy L. Tinker, Benjamin M. Tofflemire, Rita Tojeiro, Christy A. Tremonti, Mariana Vargas Magaña, Licia Verde, Nicole P. Vogt, David A. Wake, Xiaoke Wan, Ji Wang, Benjamin A. Weaver, Martin White, Simon D. M. White, John C. Wilson, John P. Wisniewski, W. Michael Wood-Vasey, Brian Yanny, Naoki Yasuda, Christophe Yèche, Donald G. York, Erick Young, Gail Zasowski, Idit Zehavi, and Bo Zhao. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. Astronomical Journal, 142(3):72, 9 2011.

[116] Vincent R. Eke, Julio F. Navarro, and Matthias Steinmetz. The Power Spectrum Dependence of Dark Matter Halo Concentrations. Astrophysical Journal, 554(1):114– 125, 6 2001.

- [117] ESA/Hubble. ESA/Hubble image site, 2021.
- [118] Fraser A. Evans, Laura C. Parker, and Ian D. Roberts. Red Misfits in the Sloan Digital Sky Survey: properties of star-forming red galaxies. *Monthly Notices of the Royal Astronomical Society*, 476(4):5284–5302, 6 2018.
- [119] Tilly A. Evans, Azadeh Fattahi, Alis J. Deason, and Carlos S. Frenk. How unusual is the Milky Way's assembly history? *Monthly Notices of the Royal Astronomical Society*, 497(4):4311–4321, 10 2020.
- [120] S. M. Faber, C. N. A. Willmer, C. Wolf, D. C. Koo, B. J. Weiner, J. A. Newman, M. Im, A. L. Coil, C. Conroy, M. C. Cooper, M. Davis, D. P. Finkbeiner, B. F. Gerke, K. Gebhardt, E. J. Groth, P. Guhathakurta, J. Harker, N. Kaiser, S. Kassin, M. Kleinheinrich, N. P. Konidaris, R. G. Kron, L. Lin, G. Luppino, D. S. Madgwick, K. Meisenheimer, K. G. Noeske, A. C. Phillips, V. L. Sarajedini, R. P. Schiavon, L. Simard, A. S. Szalay, N. P. Vogt, and R. Yan. Galaxy Luminosity Functions to z~1 from DEEP2 and COMBO-17: Implications for Red Galaxy Formation. *The Astrophysical Journal*, 665(1):265–294, 8 2007.
- [121] S. M. Fall and G. Efstathiou. Formation and rotation of disc galaxies with haloes. Monthly Notices of the Royal Astronomical Society, 193:189–206, 10 1980.
- [122] Andreas Faltenbacher and Simon D. M. White. Assembly Bias and the Dynamical Structure of Dark Matter Halos. Astrophysical Journal, 708:469–473, 1 2010.
- [123] Catherine E. Fielder, Brett Andrews, and Jeffrey A. Newman. Model Independent K-corrections. In preparation, 2021.
- [124] Catherine E. Fielder, Yao-Yuan Mao, Jeffrey A. Newman, Andrew R. Zentner, and Timothy C. Licquia. Predictably missing satellites: subhalo abundances in Milky Waylike haloes. *Monthly Notices of the Royal Astronomical Society*, 486(4):4545–4568, 7 2019.
- [125] Catherine E. Fielder, Jeffrey A. Newman, Brett H. Andrews, Gail Zasowski, Nicholas F. Boardman, Tim Licquia, Karen L. Masters, and Samir Salim. Constraining the Milky Way's Ultraviolet to Infrared SED with Gaussian Process Regression. arXiv e-prints, page arXiv:2106.14900, June 2021.

- [126] M. Fioc and B. Rocca-Volmerange. PEGASE: a UV to NIR spectral evolution model of galaxies. Application to the calibration of bright galaxy counts. Astronomy and Astrophysics, 500:507–519, October 1997.
- [127] David B. Fisher, Niv Drory, and Maximilian H. Fabricius. Bulges of Nearby Galaxies with Spitzer: The Growth of Pseudobulges in Disk Galaxies and its Connection to Outer Disks. *The Astrophysical Journal*, 697(1):630–650, 5 2009.
- [128] Ricardo A. Flores and Joel R. Primack. Observational and Theoretical Constraints on Singular Dark Matter Halos. Astrophysical Journal, 427:L1, 5 1994.
- [129] Amelia Fraser-McKelvie, Michael Merrifield, and Alfonso Aragón-Salamanca. From the outside looking in: what can Milky Way analogues tell us about the star formation rate of our own galaxy? Monthly Notices of the Royal Astronomical Society, 489(4):5030–5036, 11 2019.
- [130] Amelia Fraser-McKelvie, Michael Merrifield, Alfonso Aragón-Salamanca, Thomas Peterken, Katarina Kraljic, Karen Masters, David Stark, Francesca Fragkoudi, Rebecca Smethurst, Nicholas Fraser Boardman, Niv Drory, and Richard R. Lane. SDSS-IV MaNGA: The link between bars and the early cessation of star formation in spiral galaxies. *Monthly Notices of the Royal Astronomical Society*, 499(1):1116–1125, 11 2020.
- [131] K. C. Freeman. On the Disks of Spiral and S0 Galaxies. Astrophysical Journal, 160:811, June 1970.
- [132]Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, C. A. L. Bailer-Jones, M. Biermann, D. W. Evans, L. Ever, F. Jansen, C. Jordi, S. A. Klioner, U. Lammers, L. Lindegren, X. Luri, F. Mignard, C. Panem, D. Pourbaix, S. Randich, P. Sartoretti, H. I. Siddiqui, C. Soubiran, F. van Leeuwen, N. A. Walton, F. Arenou, U. Bastian, M. Cropper, R. Drimmel, D. Katz, M. G. Lattanzi, J. Bakker, C. Cacciari, J. Castañeda, L. Chaoul, N. Cheek, F. De Angeli, C. Fabricius, R. Guerra, B. Holl, E. Masana, R. Messineo, N. Mowlavi, K. Nienartowicz, P. Panuzzo, J. Portell, M. Riello, G. M. Seabroke, P. Tanga, F. Thévenin, G. Gracia-Abril, G. Comoretto, M. Garcia-Reinaldos, D. Teyssier, M. Altmann, R. Andrae, M. Audard, I. Bellas-Velidis, K. Benson, J. Berthier, R. Blomme, P. Burgess, G. Busso, B. Carry, A. Cellino, G. Clementini, M. Clotet, O. Creevey, M. Davidson, J. De Ridder, L. Delchambre, A. Dell'Oro, C. Ducourant, J. Fernández-Hernández, M. Fouesneau, Y. Frémat, L. Galluccio, M. García-Torres, J. González-Núñez, J. J. González-Vidal, E. Gosset, L. P. Guy, J. L. Halbwachs, N. C. Hambly, D. L. Harrison, J. Hernández, D. Hestroffer, S. T. Hodgkin, A. Hutton, G. Jasniewicz, A. Jean-Antoine-Piccolo, S. Jordan, A. J. Korn, A. Krone-Martins, A. C. Lan-

zafame, T. Lebzelter, W. Löffler, M. Manteiga, P. M. Marrese, J. M. Martín-Fleitas. A. Moitinho, A. Mora, K. Muinonen, J. Osinde, E. Pancino, T. Pauwels, J. M. Petit. A. Recio-Blanco, P. J. Richards, L. Rimoldini, A. C. Robin, L. M. Sarro, C. Siopis, M. Smith, A. Sozzetti, M. Süveges, J. Torra, W. van Reeven, U. Abbas, A. Abreu Aramburu, S. Accart, C. Aerts, G. Altavilla, M. A. Alvarez, R. Alvarez, J. Alves, R. I. Anderson, A. H. Andrei, E. Anglada Varela, E. Antiche, T. Antoja, B. Arcay, T. L. Astraatmadja, N. Bach, S. G. Baker, L. Balaguer-Núñez, P. Balm, C. Barache, C. Barata, D. Barbato, F. Barblan, P. S. Barklem, D. Barrado, M. Barros, M. A. Barstow, S. Bartholomé Muñoz, J. L. Bassilana, U. Becciani, M. Bellazzini, A. Berihuete, S. Bertone, L. Bianchi, O. Bienaymé, S. Blanco-Cuaresma, T. Boch, C. Boeche, A. Bombrun, R. Borrachero, D. Bossini, S. Bouquillon, G. Bourda, A. Bragaglia, L. Bramante, M. A. Breddels, A. Bressan, N. Brouillet, T. Brüsemeister, E. Brugaletta, B. Bucciarelli, A. Burlacu, D. Busonero, A. G. Butkevich, R. Buzzi, E. Caffau, R. Cancelliere, G. Cannizzaro, T. Cantat-Gaudin, R. Carballo, T. Carlucci, J. M. Carrasco, L. Casamiquela, M. Castellani, A. Castro-Ginard, P. Charlot, L. Chemin, A. Chiavassa, G. Cocozza, G. Costigan, S. Cowell, F. Crifo, M. Crosta, C. Crowley, J. Cuypers, C. Dafonte, Y. Damerdji, A. Dapergolas, P. David, M. David, P. de Laverny, F. De Luise, R. De March, D. de Martino, R. de Souza, A. de Torres, J. Debosscher, E. del Pozo, M. Delbo, A. Delgado, H. E. Delgado, P. Di Matteo, S. Diakite, C. Diener, E. Distefano, C. Dolding, P. Drazinos, J. Durán, B. Edvardsson, H. Enke, K. Eriksson, P. Esquej, G. Eynard Bontemps, C. Fabre, M. Fabrizio, S. Faigler, A. J. Falcão, M. Farràs Casas, L. Federici, G. Fedorets, P. Fernique, F. Figueras, F. Filippi, K. Findeisen, A. Fonti, E. Fraile, M. Fraser, B. Frézouls, M. Gai, S. Galleti, D. Garabato, F. García-Sedano, A. Garofalo, N. Garralda, A. Gavel, P. Gavras, J. Gerssen, R. Geyer, P. Giacobbe, G. Gilmore, S. Girona, G. Giuffrida, F. Glass, M. Gomes, M. Granvik, A. Gueguen, A. Guerrier, J. Guiraud, R. Gutiérrez-Sánchez, R. Haigron, D. Hatzidimitriou, M. Hauser, M. Haywood, U. Heiter, A. Helmi, J. Heu, T. Hilger, D. Hobbs, W. Hofmann, G. Holland, H. E. Huckle, A. Hypki, V. Icardi, K. Janßen. G. Jevardat de Fombelle, P. G. Jonker, A. L. Juhász, F. Julbe, A. Karampelas, A. Kewley, J. Klar, A. Kochoska, R. Kohley, K. Kolenberg, M. Kontizas, E. Kontizas, S. E. Koposov, G. Kordopatis, Z. Kostrzewa-Rutkowska, P. Koubsky, S. Lambert, A. F. Lanza, Y. Lasne, J. B. Lavigne, Y. Le Fustec, C. Le Poncin-Lafitte, Y. Lebreton, S. Leccia, N. Leclerc, I. Lecoeur-Taibi, H. Lenhardt, F. Leroux, S. Liao, E. Licata, H. E. P. Lindstrøm, T. A. Lister, E. Livanou, A. Lobel, M. López, S. Managau, R. G. Mann, G. Mantelet, O. Marchal, J. M. Marchant, M. Marconi, S. Marinoni, G. Marschalkó, D. J. Marshall, M. Martino, G. Marton, N. Mary, D. Massari, G. Matijevič, T. Mazeh, P. J. McMillan, S. Messina, D. Michalik, N. R. Millar, D. Molina, R. Molinaro, L. Molnár, P. Montegriffo, R. Mor, R. Morbidelli, T. Morel, D. Morris, A. F. Mulone, T. Muraveva, I. Musella, G. Nelemans, L. Nicastro, L. Noval, W. O'Mullane, C. Ordénovic, D. Ordóñez-Blanco, P. Osborne, C. Pagani, I. Pagano, F. Pailler, H. Palacin, L. Palaversa, A. Panahi, M. Pawlak, A. M. Piersimoni, F. X. Pineau, E. Plachy, G. Plum, E. Poggio, E. Poujoulet, A. Prša, L. Pulone, E. Racero. S. Ragaini, N. Rambaux, M. Ramos-Lerate, S. Regibo, C. Reylé, F. Riclet, V. Ripepi, A. Riva, A. Rivard, G. Rixon, T. Roegiers, M. Roelens, M. Romero-Gómez, N. Rowell, F. Royer, L. Ruiz-Dern, G. Sadowski, T. Sagristà Sellés, J. Sahlmann, J. Salgado, E. Salguero, N. Sanna, T. Santana-Ros, M. Sarasso, H. Savietto, M. Schultheis, E. Sciacca, M. Segol, J. C. Segovia, D. Ségransan, I. C. Shih, L. Siltala, A. F. Silva, R. L. Smart, K. W. Smith, E. Solano, F. Solitro, R. Sordo, S. Soria Nieto, J. Souchay, A. Spagna, F. Spoto, U. Stampa, I. A. Steele, H. Steidelmüller, C. A. Stephenson, H. Stoev, F. F. Suess, J. Surdej, L. Szabados, E. Szegedi-Elek, D. Tapiador, F. Taris, G. Tauran, M. B. Taylor, R. Teixeira, D. Terrett, P. Teyssand ier, W. Thuillot, A. Titarenko, F. Torra Clotet, C. Turon, A. Ulla, E. Utrilla, S. Uzzi, M. Vaillant, G. Valentini, V. Valette, A. van Elteren, E. Van Hemelryck, M. van Leeuwen, M. Vaschetto, A. Vecchiato, J. Veljanoski, Y. Viala, D. Vicente, S. Vogt, C. von Essen, H. Voss, V. Votruba, S. Voutsinas, G. Walmsley, M. Weiler, O. Wertz, T. Wevers, L. Wyrzykowski, A. Yoldas, M. Žerjal, H. Ziaeepour, J. Zorec, S. Zschocke, S. Zucker, C. Zurbach, and T. Zwitter. Gaia Data Release 2. Summary of the contents and survey properties. Astronomy and Astrophysics, 616:A1, 8 2018.

- [133] Anna Gallazzi, Stéphane Charlot, Jarle Brinchmann, Simon D. M. White, and Christy A. Tremonti. The ages and metallicities of galaxies in the local universe. *Monthly Notices of the Royal Astronomical Society*, 362(1):41–58, 9 2005.
- [134] Melanie A. Galloway, Kyle W. Willett, Lucy F. Fortson, Carolin N. Cardamone, Kevin Schawinski, Edmond Cheung, Chris J. Lintott, Karen L. Masters, Thomas Melvin, and Brooke D. Simmons. Galaxy Zoo: the effect of bar-driven fuelling on the presence of an active galactic nucleus in disc galaxies. *Monthly Notices of the Royal Astronomical Society*, 448(4):3442–3454, 4 2015.
- [135] L. Gao, J. F. Navarro, C. S. Frenk, A. Jenkins, V. Springel, and S. D. M. White. The Phoenix Project: the dark side of rich Galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 425(3):2169–2186, 9 2012.
- [136] L. Gao, S. D. M. White, A. Jenkins, F. Stoehr, and V. Springel. The subhalo populations of ACDM dark haloes. *Monthly Notices of the Royal Astronomical Society*, 355:819–834, 12 2004.
- [137] Liang Gao, Julio F. Navarro, Shaun Cole, Carlos S. Frenk, Simon D. M. White, Volker Springel, Adrian Jenkins, and Angelo F. Neto. The redshift dependence of the structure of massive Λ cold dark matter haloes. *Monthly Notices of the Royal* Astronomical Society, 387(2):536–544, 6 2008.
- [138] Liang Gao and Simon D. M. White. Assembly bias in the clustering of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 377(1):L5–L9, 4 2007.
- [139] Shea Garrison-Kimmel, Philip F. Hopkins, Andrew Wetzel, Kareem El-Badry, Robyn E. Sanderson, James S. Bullock, Xiangcheng Ma, Freeke van de Voort, Zachary Hafen, Claude-André Faucher-Giguère, Christopher C. Hayward, Eliot Quataert, Dušan Kereš, and Michael Boylan-Kolchin. The origin of the diverse morphologies and kinematics of Milky Way-mass galaxies in the FIRE-2 simulations. *Monthly Notices of the Royal Astronomical Society*, 481(3):4133–4157, December 2018.
- [140] Shea Garrison-Kimmel, Andrew Wetzel, James S. Bullock, Philip F. Hopkins, Michael Boylan-Kolchin, Claude-André Faucher-Giguère, Dušan Kereš, Eliot Quataert, Robyn E. Sanderson, Andrew S. Graus, and Tyler Kelley. Not so lumpy after all: modelling the depletion of dark matter subhaloes by Milky Way-like galaxies. *Monthly Notices of the Royal Astronomical Society*, 471(2):1709–1727, 10 2017.
- [141] Marla Geha, Risa H. Wechsler, Yao-Yuan Mao, Erik J. Tollerud, Benjamin Weiner, Rebecca Bernstein, Ben Hoyle, Sebastian 3i, Phil J. Marshall, Ricardo Muñoz, and Yu Lu. The SAGA Survey. I. Satellite Galaxy Populations around Eight Milky Way Analogs. Astrophysical Journal, 847:4, 9 2017.
- [142] K. George, P. Joseph, C. Mondal, S. Subramanian, A. Subramaniam, and K. T. Paul. Insights on bar quenching from a multiwavelength analysis: The case of Messier 95. Astronomy and Astrophysics, 621:L4, 1 2019.
- [143] S. Gillessen, F. Eisenhauer, S. Trippe, T. Alexand er, R. Genzel, F. Martins, and T. Ott. Monitoring Stellar Orbits Around the Massive Black Hole in the Galactic Center. *The Astrophysical Journal*, 692(2):1075–1109, 2 2009.
- [144] Daniel Gilman, Simon Birrer, Anna Nierenberg, Tommaso Treu, Xiaolong Du, and Andrew Benson. Warm dark matter chills out: constraints on the halo mass function and the free-streaming length of dark matter with eight quadruple-image strong gravitational lenses. *Monthly Notices of the Royal Astronomical Society*, 491(4):6077–6101, 2 2020.
- [145] Carlo Giocoli, Giuseppe Tormen, and Frank C. van den Bosch. The population of dark matter subhaloes: mass functions and average mass-loss rates. *Monthly Notices of the Royal Astronomical Society*, 386(4):2135–2144, 6 2008.
- [146] Michael D. Gladders, Augustus Oemler, Alan Dressler, Bianca Poggianti, Benedetta Vulcani, and Louis Abramson. The IMACS Cluster Building Survey. IV. The Lognormal Star Formation History of Galaxies. *The Astrophysical Journal*, 770(1):64, 6 2013.

- [147] Stephanie Glen. Akaike's information criterion: Definition, and formulas, 2015.
- [148] Stephanie Glen. Bayesian information criterion (bic) / schwarz criterion, 2018.
- [149] O. Y. Gnedin, A. V. Kravtsov, A. A. Klypin, and D. Nagai. Response of Dark Matter Halos to Condensation of Baryons: Cosmological Simulations and Improved Adiabatic Contraction Model. *Astrophysical Journal*, 616:16–26, 11 2004.
- [150] Tyler A. Gordon, Eric Agol, and Daniel Foreman-Mackey. A Fast, Two-dimensional Gaussian Process Method Based on Celerite: Applications to Transiting Exoplanet Discovery and Characterization. Astronomical Journal, 160(5):240, 11 2020.
- [151] Jochen Görtler, Thilo Spinner, Dirk Streeb, Daniel Weiskopf, and Oliver Deussen. Uncertainty-Aware Principal Component Analysis. arXiv e-prints, page arXiv:1905.01127, 5 2019.
- [152] Alister W. Graham. Elliptical and Disk Galaxy Structure and Modern Scaling Laws. In Terry D. Oswalt and William C. Keel, editors, *Planets, Stars and Stellar Systems. Volume 6: Extragalactic Astronomy and Cosmology*, volume 6, pages 91–140. Springer, 2013.
- [153] Alister W. Graham, Peter Erwin, I. Trujillo, and A. Asensio Ramos. A New Empirical Model for the Structural Analysis of Early-Type Galaxies, and A Critical Review of the Nuker Model. *Astronomical Journal*, 125(6):2951–2963, 6 2003.
- [154] Gravity Collaboration, R. Abuter, A. Amorim, M. Bauböck, J. P. Berger, H. Bonnet, W. Brand ner, Y. Clénet, V. Coudé Du Foresto, P. T. de Zeeuw, J. Dexter, G. Duvert, A. Eckart, F. Eisenhauer, N. M. Förster Schreiber, P. Garcia, F. Gao, E. Gendron, R. Genzel, O. Gerhard, S. Gillessen, M. Habibi, X. Haubois, T. Henning, S. Hippler, M. Horrobin, A. Jiménez-Rosales, L. Jocou, P. Kervella, S. Lacour, V. Lapeyrère, J. B. Le Bouquin, P. Léna, T. Ott, T. Paumard, K. Perraut, G. Perrin, O. Pfuhl, S. Rabien, G. Rodriguez Coira, G. Rousset, S. Scheithauer, A. Sternberg, O. Straub, C. Straubmeier, E. Sturm, L. J. Tacconi, F. Vincent, S. von Fellenberg, I. Waisberg, F. Widmann, E. Wieprecht, E. Wiezorrek, J. Woillez, and S. Yazici. A geometric distance measurement to the Galactic center black hole with 0.3% uncertainty. Astronomy and Astrophysics, 625:L10, 5 2019.
- [155] Brendan F. Griffen, Alexander P. Ji, Gregory A. Dooley, Facundo A. Gómez, Mark Vogelsberger, Brian W. O'Shea, and Anna Frebel. The Caterpillar Project: A Large Suite of Milky Way Sized Halos. Astrophysical Journal, 818(1):10, 2 2016.

- [156] Ruth Grützbauch, Robert W. Chuter, Christopher J. Conselice, Amanda E. Bauer, Asa F. L. Bluck, Fernando Buitrago, and Alice Mortlock. Galaxy properties in different environments up to z<sup>~</sup> 3 in the GOODS NICMOS Survey. *Monthly Notices of the Royal* Astronomical Society, 412(4):2361–2375, 4 2011.
- [157] Javiera Guedes, Simone Callegari, Piero Madau, and Lucio Mayer. Forming Realistic Late-type Spirals in a ACDM Universe: The Eris Simulation. *The Astrophysical Journal*, 742(2):76, 12 2011.
- [158] Quan Guo, Shaun Cole, Vincent Eke, Carlos Frenk, and John Helly. Spatial and luminosity distributions of galactic satellites. *Monthly Notices of the Royal Astronomical Society*, 434(3):1838–1848, 9 2013.
- [159] Cai-Na Hao, Yong Shi, Yanmei Chen, Xiaoyang Xia, Qiusheng Gu, Rui Guo, Xiaoling Yu, and Songlin Li. Spatially Resolved Studies of Local Massive Red Spiral Galaxies. *The Astrophysical Journal, Letters*, 883(2):L36, 10 2019.
- [160] Ross E. Hart, Steven P. Bamford, Kyle W. Willett, Karen L. Masters, Carolin Cardamone, Chris J. Lintott, Robert J. Mackay, Robert C. Nichol, Christopher K. Rosslowe, Brooke D. Simmons, and Rebecca J. Smethurst. Galaxy Zoo: comparing the demographics of spiral arm number and a new method for correcting redshift bias. *Monthly Notices of the Royal Astronomical Society*, 461(4):3663–3682, 07 2016.
- [161] T. G. Hawarden, J. H. Huang, and Q. S. Gu. Infrared Emission from Barred Spiral Galaxies. In R. Buta, D. A. Crocker, and B. G. Elmegreen, editors, *IAU Colloq. 157: Barred Galaxies*, volume 91 of *Astronomical Society of the Pacific Conference Series*, page 54, 1 1996.
- [162] T. G. Hawarden, C. M. Mountain, S. K. Leggett, and P. J. Puxley. Enhanced star formation - the importance of bars in spiral galaxies. *Monthly Notices of the Royal Astronomical Society*, 221:41P-45, 8 1986.
- [163] M. Haywood, M. D. Lehnert, P. Di Matteo, O. Snaith, M. Schultheis, D. Katz, and A. Gómez. When the Milky Way turned off the lights: APOGEE provides evidence of star formation quenching in our Galaxy. *Astronomy and Astrophysics*, 589:A66, 5 2016.
- [164] Lars Hernquist. An Analytical Model for Spherical Galaxies and Bulges. Astrophysical Journal, 356:359, 6 1990.

- [165] F. B. Hildebrand. Introduction to Numerical Analysis. Dover Books, 2 edition, 1987.
- [166] R. W. Hockney and J. W. Eastwood. Computer simulation using particles. CRC Press, 1988.
- [167] P. W. Hodge. The Hubble type of the Milky Way Galaxy. Publications of the ASP, 95:721–723, 10 1983.
- [168] J. L. Hodges and E. L. Lehmann. Estimates of location based on rank tests. Ann. Math. Statist., 34(2):598–611, 06 1963.
- [169] David W. Hogg, Ivan K. Baldry, Michael R. Blanton, and Daniel J. Eisenstein. The K correction. arXiv e-prints, pages astro-ph/0210394, October 2002.
- [170] David W. Hogg, Michael R. Blanton, Jarle Brinchmann, Daniel J. Eisenstein, David J. Schlegel, James E. Gunn, Timothy A. McKay, Hans-Walter Rix, Neta A. Bahcall, J. Brinkmann, and Avery Meiksin. The Dependence on Environment of the Color-Magnitude Relation of Galaxies. *The Astrophysical Journal, Letters*, 601(1):L29–L32, 1 2004.
- [171] Edwin Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. Proceedings of the National Academy of Science, 15(3):168–173, 3 1929.
- [172] P. J. Huber. Robust Estimation of a Location Parameter. Annals of Mathematical Statistics, 35:73–101, 1964.
- [173] M. L. Humason, N. U. Mayall, and A. R. Sandage. Redshifts and magnitudes of extragalactic nebulae. Astronomical Journal, 61:97–162, January 1956.
- [174] E. Hummel, J. M. van der Hulst, R. C. Kennicutt, and W. C. Keel. Environment impact on the nuclear radio activity in spiral galaxies. Astronomy and Astrophysics, 236:333, 9 1990.
- [175] T. H. Jarrett, M. E. Cluver, C. Magoulas, M. Bilicki, M. Alpaslan, J. Bland-Hawthorn, S. Brough, M. J. I. Brown, S. Croom, S. Driver, B. W. Holwerda, A. M. Hopkins, J. Loveday, P. Norberg, J. A. Peacock, C. C. Popescu, E. M. Sadler, E. N. Taylor, R. J. Tuffs, and L. Wang. Galaxy and Mass Assembly (GAMA): Exploring the WISE Web in G12. *The Astrophysical Journal*, 836(2):182, 2 2017.

- T. H. Jarrett, M. Cohen, F. Masci, E. Wright, D. Stern, D. Benford, A. Blain, S. Carey, R. M. Cutri, P. Eisenhardt, C. Lonsdale, A. Mainzer, K. Marsh, D. Padgett, S. Petty, M. Ressler, M. Skrutskie, S. Stanford, J. Surace, C. W. Tsai, S. Wheelock, and D. L. Yan. The Spitzer-WISE Survey of the Ecliptic Poles. *The Astrophysical Journal*, 735(2):112, 7 2011.
- [177] Fangzhou Jiang, Avishai Dekel, Omer Kneller, Sharon Lapiner, Daniel Ceverino, Joel R. Primack, Sandra M. Faber, Andrea V. Macciò, Aaron Dutton, Shy Genel, and Rachel S. Somerville. Is the dark-matter halo spin a predictor of galaxy spin and size? ArXiv e-prints, page arXiv:1804.07306, 4 2018.
- [178] Fangzhou Jiang and Frank C. van den Bosch. Comprehensive assessment of the too big to fail problem. Monthly Notices of the Royal Astronomical Society, 453:3575–3592, 11 2015.
- [179] Fangzhou Jiang and Frank C. van den Bosch. Statistics of dark matter substructure
  I. Model and universal fitting functions. Monthly Notices of the Royal Astronomical Society, 458(3):2848–2869, 5 2016.
- [180] Y. P. Jing. The Density Profile of Equilibrium and Nonequilibrium Dark Matter Halos. Astrophysical Journal, 535(1):30–36, 5 2000.
- [181] David E. Johnston, Erin S. Sheldon, Risa H. Wechsler, Eduardo Rozo, Benjamin P. Koester, Joshua A. Frieman, Timothy A. McKay, 8 E. Evrard, Matthew R. Becker, and James Annis. Cross-correlation Weak Lensing of SDSS galaxy Clusters II: Cluster Density Profiles and the Mass-Richness Relation. arXiv e-prints, page arXiv:0709.1159, 9 2007.
- [182] Katherine P. Jore, Adrick H. Broeils, and Martha P. Haynes. A Counter-Rotating Disk in the Normal SA Galaxy NGC 4138. Astronomical Journal, 112:438, 8 1996.
- [183] Mario Jurić, Željko Ivezić, Alyson Brooks, Robert H. Lupton, David Schlegel, Douglas Finkbeiner, Nikhil Padmanabhan, Nicholas Bond, Branimir Sesar, Constance M. Rockosi, Gillian R. Knapp, James E. Gunn, Takahiro Sumi, Donald P. Schneider, J. C. Barentine, Howard J. Brewington, J. Brinkmann, Masataka Fukugita, Michael Harvanek, S. J. Kleinman, Jurek Krzesinski, Dan Long, Jr. Neilsen, Eric H., Atsuko Nitta, Stephanie A. Snedden, and Donald G. York. The Milky Way Tomography with SDSS. I. Stellar Number Density Distribution. Astrophysical Journal, 673(2):864–914, 2 2008.

- [184] Prajwal Raj Kafle, Sanjib Sharma, Geraint F. Lewis, and Joss Bland-Hawthorn. On the Shoulders of Giants: Properties of the Stellar Halo and the Milky Way Mass Distribution. Astrophysical Journal, 794:59, 10 2014.
- [185] Nitya Kallivayalil, Roeland P. van der Marel, Gurtina Besla, Jay Anderson, and Charles Alcock. Third-epoch Magellanic Cloud Proper Motions. I. Hubble Space Telescope/WFC3 Data and Orbit Implications. Astrophysical Journal, 764:161, 2 2013.
- [186] Marc Kamionkowski and Andrew R. Liddle. The Dearth of Halo Dwarf Galaxies: Is There Power on Short Scales? *Physical Review Letters*, 84:4525–4528, 5 2000.
- [187] Susan Alice Joan Kassin. Dark and luminous matter in bright spiral galaxies. PhD thesis, The Ohio State University, Ohio, USA, 1 2004.
- [188] G. Kauffmann, S. D. M. White, and B. Guiderdoni. The formation and evolution of galaxies within merging dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 264:201–218, 9 1993.
- [189] Guinevere Kauffmann, Timothy M. Heckman, Simon D. M. White, Stéphane Charlot, Christy Tremonti, Jarle Brinchmann, Gustavo Bruzual, Eric W. Peng, Mark Seibert, Mariangela Bernardi, Michael Blanton, Jon Brinkmann, Francisco Castander, Istvan Csábai, Masataka Fukugita, Zeljko Ivezic, Jeffrey A. Munn, Robert C. Nichol, Nikhil Padmanabhan, Aniruddha R. Thakar, David H. Weinberg, and Donald York. Stellar masses and star formation histories for 10<sup>5</sup> galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 341(1):33–53, 5 2003.
- [190] Daniel D. Kelson, Rik J. Williams, Alan Dressler, Patrick J. McCarthy, Stephen A. Shectman, John S. Mulchaey, Edward V. Villanueva, Jeffrey D. Crane, and Ryan F. Quadri. The Carnegie-Spitzer-IMACS Redshift Survey of Galaxy Evolution since z = 1.5. I. Description and Methodology. *The Astrophysical Journal*, 783(2):110, 3 2014.
- [191] Jr. Kennicutt, R. C. Stellar Structure of the Galaxy: Overview. In Charles E. Woodward, Michael D. Bicay, and J. Michael Shull, editors, *Tetons 4: Galactic Structure*, *Stars and the Interstellar Medium*, volume 231 of *Astronomical Society of the Pacific Conference Series*, page 2, 1 2001.
- [192] R. C. Kennicutt, D. Calzetti, G. Aniano, P. Appleton, L. Armus, P. Beirão, A. D. Bolatto, B. Brandl, A. Crocker, K. Croxall, D. A. Dale, J. Donovan Meyer, B. T. Draine, C. W. Engelbracht, M. Galametz, K. D. Gordon, B. Groves, C. N. Hao, G. Helou, J. Hinz, L. K. Hunt, B. Johnson, J. Koda, O. Krause, A. K. Leroy, Y. Li,

S. Meidt, E. Montiel, E. J. Murphy, N. Rahman, H. W. Rix, H. Roussel, K. Sandstrom, M. Sauvage, E. Schinnerer, R. Skibba, J. D. T. Smith, S. Srinivasan, L. Vigroux, F. Walter, C. D. Wilson, M. Wolfire, and S. Zibetti. KINGFISH—Key Insights on Nearby Galaxies: A Far-Infrared Survey with Herschel: Survey Description and Image Atlas. *Publications of the ASP*, 123(910):1347, 12 2011.

- [193] L. J. Kewley, M. A. Dopita, R. S. Sutherland, C. A. Heisler, and J. Trevena. Theoretical Modeling of Starburst Galaxies. Astrophysical Journal, 556(1):121–140, July 2001.
- [194] S. Khoperskov, M. Haywood, P. Di Matteo, M. D. Lehnert, and F. Combes. Bar quenching in gas-rich galaxies. *Astronomy and Astrophysics*, 609:A60, 1 2018.
- [195] S. Y. Kim, A. H. G. Peter, and J. R. Hargis. Missing Satellites Problem: Completeness Corrections to the Number of Satellite Galaxies in the Milky Way are Consistent with Cold Dark Matter Predictions. *Physical Review Letters*, 121(21):211302, 11 2018.
- [196] Anne L. Kinney, Daniela Calzetti, Ralph C. Bohlin, Kerry McQuade, Thaisa Storchi-Bergmann, and Henrique R. Schmitt. Template Ultraviolet to Near-Infrared Spectra of Star-forming Galaxies and Their Application to K-Corrections. *The Astrophysical Journal*, 467:38, 8 1996.
- [197] Evan N. Kirby, James S. Bullock, Michael Boylan-Kolchin, Manoj Kaplinghat, and Judith G. Cohen. The dynamics of isolated Local Group galaxies. *Monthly Notices* of the Royal Astronomical Society, 439:1015–1027, 3 2014.
- [198] Anatoly Klypin, Andrey V. Kravtsov, Octavio Valenzuela, and Francisco Prada. Where Are the Missing Galactic Satellites? Astrophysical Journal, 522:82–92, 9 1999.
- [199] Anatoly Klypin, Gustavo Yepes, Stefan Gottlöber, Francisco Prada, and Steffen Heß. MultiDark simulations: the story of dark matter halo concentrations and density profiles. Monthly Notices of the Royal Astronomical Society, 457(4):4340–4359, 4 2016.
- [200] Anatoly Klypin, HongSheng Zhao, and Rachel S. Somerville. ACDM-based Models for the Milky Way and M31. I. Dynamical Models. Astrophysical Journal, 573:597–613, 7 2002.

- [201] S. Koposov, V. Belokurov, N. W. Evans, P. C. Hewett, M. J. Irwin, G. Gilmore, D. B. Zucker, H. W. Rix, M. Fellhauer, E. F. Bell, and E. V. Glushkova. The Luminosity Function of the Milky Way Satellites. *Astrophysical Journal*, 686:279–291, 10 2008.
- [202] Sergey E. Koposov, Jaiyul Yoo, Hans-Walter Rix, David H. Weinberg, Andrea V. Macciò, and Jordi Miralda Escudé. A Quantitative Explanation of the Observed Population of Milky Way Satellite Galaxies. Astrophysical Journal, 696:2179–2194, 5 2009.
- [203] Ehsan Kourkchi, R. Brent Tully, J. Don Neill, Mark Seibert, Hélène M. Courtois, and Alexandra Dupuy. Global Attenuation in Spiral Galaxies in Optical and Infrared Bands. *The Astrophysical Journal*, 884(1):82, 10 2019.
- [204] Andrey V. Kravtsov, Andreas A. Berlind, Risa H. Wechsler, Anatoly A. Klypin, Stefan Gottlöber, Brandon Allgood, and Joel R. Primack. The Dark Side of the Halo Occupation Distribution. Astrophysical Journal, 609:35–49, 7 2004.
- [205] Andrey V. Kravtsov, Alexey Vikhlinin, and Daisuke Nagai. A New Robust Low-Scatter X-Ray Mass Indicator for Clusters of Galaxies. Astrophysical Journal, 650:128–136, 10 2006.
- [206] D. Krishnarao, C. Tremonti, F. Amelia, K. Kraljic, N. Boardman, K. Masters, R. Benjamin, L. Haffner, A. Jones, Z. Pace, M. Bershady, J. Brinkmann, J. Brownstein, N. Drory, K. Zhang, D. Bizyaev, and K. Pan. Optical Emission, Bars, LI(N)ERS and Beyond: Bridging the Milky Way with Extragalactic Surveys. In American Astronomical Society Meeting Abstracts #235, volume 235 of American Astronomical Society Meeting Abstracts, page 182.02, 1 2020.
- [207] Pavel Kroupa and Carsten Weidner. Galactic-Field Initial Mass Functions of Massive Stars. *The Astrophysical Journal*, 598(2):1076–1078, 12 2003.
- [208] Sandor J. Kruk, Chris J. Lintott, Steven P. Bamford, Karen L. Masters, Brooke D. Simmons, Boris Häußler, Carolin N. Cardamone, Ross E. Hart, Lee Kelvin, Kevin Schawinski, Rebecca J. Smethurst, and Marina Vika. Galaxy Zoo: secular evolution of barred galaxies from structural decomposition of multiband images. *Monthly Notices of the Royal Astronomical Society*, 473(4):4731–4753, 2 2018.
- [209] Michael Kuhlen. The Dark Matter Annihilation Signal from Dwarf Galaxies and Subhalos. *Advances in Astronomy*, 2010:162083, 1 2010.

- [210] C. N. Lackner and J. E. Gunn. Astrophysically motivated bulge-disc decompositions of Sloan Digital Sky Survey galaxies. *Monthly Notices of the Royal Astronomical Society*, 421(3):2277–2302, 4 2012.
- [211] Dustin Lang, David W. Hogg, and David J. Schlegel. WISE Photometry for 400 Million SDSS Sources. Astronomical Journal, 151(2):36, February 2016.
- [212] David R. Law and Steven R. Majewski. The Sagittarius Dwarf Galaxy: A Model for Evolution in a Triaxial Milky Way Halo. *Astrophysical Journal*, 714:229–254, 5 2010.
- [213] Alexandres Lazar, James S. Bullock, Michael Boylan-Kolchin, T. K. Chan, Philip F. Hopkins, Andrew S. Graus, Andrew Wetzel, Kareem El-Badry, Coral Wheeler, Maria C. Straight, Dušan Kereš, Claude-André Faucher-Giguère, Alex Fitts, and Shea Garrison-Kimmel. A dark matter profile to model diverse feedback-induced core sizes of ACDM haloes. arXiv e-prints, page arXiv:2004.10817, 4 2020.
- [214] Benjamin V. Lehmann, Yao-Yuan Mao, Matthew R. Becker, Samuel W. Skillman, and Risa H. Wechsler. The Concentration Dependence of the Galaxy-Halo Connection: Modeling Assembly Bias with Abundance Matching. Astrophysical Journal, 834(1):37, 1 2017.
- [215] Adam K. Leroy, Fabian Walter, Elias Brinks, Frank Bigiel, W. J. G. de Blok, Barry Madore, and M. D. Thornley. The Star Formation Efficiency in Nearby Galaxies: Measuring Where Gas Forms Stars Effectively. Astronomical Journal, 136(6):2782– 2845, 12 2008.
- [216] Timothy C. Licquia and Jeffrey A. Newman. Improved Estimates of the Milky Way's Stellar Mass and Star Formation Rate from Hierarchical Bayesian Meta-Analysis. *The Astrophysical Journal*, 806(1):96, 6 2015.
- [217] Timothy C. Licquia and Jeffrey A. Newman. Sizing Up the Milky Way: A Bayesian Mixture Model Meta-analysis of Photometric Scale Length Measurements. *The As*trophysical Journal, 831(1):71, 11 2016.
- [218] Timothy C. Licquia, Jeffrey A. Newman, and Matthew A. Bershady. Does the Milky Way Obey Spiral Galaxy Scaling Relations? *The Astrophysical Journal*, 833:220, 12 2016.

- [219] Timothy C. Licquia, Jeffrey A. Newman, and Jarle Brinchmann. Unveiling the Milky Way: A New Technique for Determining the Optical Color and Luminosity of Our Galaxy. *The Astrophysical Journal*, 809:96, 8 2015.
- [220] M. Limousin, J-P. Kneib, and P. Natarajan. Galaxy Galaxy Lensing as a Probe of Galaxy Dark Matter Halos. *arXiv e-prints*, pages astro-ph/0606447, 6 2006.
- [221] Chris Lintott, Kevin Schawinski, Steven Bamford, Anže Slosar, Kate Land, Daniel Thomas, Edd Edmondson, Karen Masters, Robert C. Nichol, M. Jordan Raddick, Alex Szalay, Dan Andreescu, Phil Murray, and Jan Vandenberg. Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies. *Monthly Notices* of the Royal Astronomical Society, 410(1):166–178, 1 2011.
- [222] Lulu Liu, Brian F. Gerke, Risa H. Wechsler, Peter S. Behroozi, and Michael T. Busha. How Common are the Magellanic Clouds? *Astrophysical Journal*, 733(1):62, 5 2011.
- [223] Aaron D. Ludlow, Sownak Bose, Raúl E. Angulo, Lan Wang, Wojciech A. Hellwing, Julio F. Navarro, Shaun Cole, and Carlos S. Frenk. The mass-concentration-redshift relation of cold and warm dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 460(2):1214–1232, 8 2016.
- [224] Aaron D. Ludlow, Julio F. Navarro, Volker Springel, Mark Vogelsberger, Jie Wang, Simon D. M. White, Adrian Jenkins, and Carlos S. Frenk. Secondary infall and the pseudo-phase-space density profiles of cold dark matter haloes. *Monthly Notices of* the Royal Astronomical Society, 406(1):137–146, 7 2010.
- [225] Aaron D. Ludlow, Joop Schaye, and Richard Bower. Numerical convergence of simulations of galaxy formation: the abundance and internal structure of cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 488(3):3663–3684, 9 2019.
- [226] Xiangcheng Ma, Philip F. Hopkins, Andrew R. Wetzel, Evan N. Kirby, Daniel Anglés-Alcázar, Claude-André Faucher-Giguère, Dušan Kereš, and Eliot Quataert. The structure and dynamical evolution of the stellar disc of a simulated Milky Way-mass galaxy. *Monthly Notices of the Royal Astronomical Society*, 467(2):2430–2444, 5 2017.
- [227] Andrea V. Macciò, Aaron A. Dutton, and Frank C. van den Bosch. Concentration, spin and shape of dark matter haloes as a function of the cosmological model: WMAP1, WMAP3 and WMAP5 results. *Monthly Notices of the Royal Astronomical Society*, 391(4):1940–1954, 12 2008.

- [228] Andrea V. Macciò, Aaron A. Dutton, Frank C. van den Bosch, Ben Moore, Doug Potter, and Joachim Stadel. Concentration, spin and shape of dark matter haloes: scatter and the dependence on mass and environment. *Monthly Notices of the Royal* Astronomical Society, 378(1):55–71, 6 2007.
- [229] Piero Madau, Jürg Diemand, and Michael Kuhlen. Dark Matter Subhalos and the Dwarf Satellites of the Milky Way. *Astrophysical Journal*, 679:1260–1271, 6 2008.
- [230] Darren S. Madgwick, Alison L. Coil, Christopher J. Conselice, Michael C. Cooper, Marc Davis, Richard S. Ellis, Sandy M. Faber, Douglas P. Finkbeiner, Brian Gerke, Puragra Guhathakurta, Nick Kaiser, David C. Koo, Jeffrey A. Newman, Andrew C. Phillips, Charles C. Steidel, Benjamin J. Weiner, Christopher N. A. Willmer, R. Yan, and Deep2 Survey Team. The DEEP2 Galaxy Redshift Survey: Spectral Classification of Galaxies at z~1. Astrophysical Journal, 599(2):997–1005, December 2003.
- [231] Darren S. Madgwick, Rachel Somerville, Ofer Lahav, and Richard Ellis. Galaxy spectral parametrization in the 2dF Galaxy Redshift Survey as a diagnostic of star formation history. *Monthly Notices of the Royal Astronomical Society*, 343(3):871–879, August 2003.
- [232] Gladis Magris C., Juan Mateu P., Cecilia Mateu, Gustavo Bruzual A., Ivan Cabrera-Ziri, and Alfredo Mejía-Narváez. On the Recovery of Galaxy Properties from SED Fitting Solutions. *Publications of the ASP*, 127(947):16, January 2015.
- Steven R. Majewski, Ricardo P. Schiavon, Peter M. Frinchaboy, Carlos Allende Pri-[233]eto, Robert Barkhouser, Dmitry Bizyaev, Basil Blank, Sophia Brunner, Adam Burton, Ricardo Carrera, S. Drew Chojnowski, Kátia Cunha, Courtney Epstein, Greg Fitzgerald, Ana E. García Pérez, Fred R. Hearty, Chuck Henderson, Jon A. Holtzman, Jennifer A. Johnson, Charles R. Lam, James E. Lawler, Paul Maseman, Szabolcs Mészáros, Matthew Nelson, Duy Coung Nguyen, David L. Nidever, Marc Pinsonneault, Matthew Shetrone, Stephen Smee, Verne V. Smith, Todd Stolberg, Michael F. Skrutskie, Eric Walker, John C. Wilson, Gail Zasowski, Friedrich Anders, Sarbani Basu, Stephane Beland, Michael R. Blanton, Jo Bovy, Joel R. Brownstein, Joleen Carlberg, William Chaplin, Cristina Chiappini, Daniel J. Eisenstein, Yvonne Elsworth, Diane Feuillet, Scott W. Fleming, Jessica Galbraith-Frew, Rafael A. García, D. Aníbal García-Hernández, Bruce A. Gillespie, Léo Girardi, James E. Gunn, Sten Hasselquist, Michael R. Hayden, Saskia Hekker, Inese Ivans, Karen Kinemuchi, Mark Klaene, Suvrath Mahadevan, Savita Mathur, Benoît Mosser, Demitri Muna, Jeffrey A. Munn, Robert C. Nichol, Robert W. O'Connell, John K. Parejko, A. C. Robin, Helio Rocha-Pinto, Matthias Schultheis, Aldo M. Serenelli, Neville Shane, Victor Silva Aguirre, Jennifer S. Sobeck, Benjamin Thompson, Nicholas W. Troup,

David H. Weinberg, and Olga Zamora. The Apache Point Observatory Galactic Evolution Experiment (APOGEE). Astronomical Journal, 154(3):94, 9 2017.

- [234] Ariyeh H. Maller, Andreas A. Berlind, Michael R. Blanton, and David W. Hogg. The Intrinsic Properties of SDSS Galaxies. *The Astrophysical Journal*, 691(1):394–406, 1 2009.
- [235] Rachel Mandelbaum, Christopher M. Hirata, Tamara Broderick, Uroš Seljak, and Jonathan Brinkmann. Ellipticity of dark matter haloes with galaxy-galaxy weak lensing. Monthly Notices of the Royal Astronomical Society, 370(2):1008–1024, 8 2006.
- [236] S. Mao, H. J. Mo, and S. D. M. White. The evolution of galactic discs. Monthly Notices of the Royal Astronomical Society, 297:L71–L75, 7 1998.
- [237] Yao-Yuan Mao, Marc Williamson, and Risa H. Wechsler. The Dependence of Subhalo Abundance on Halo Concentration. *Astrophysical Journal*, 810(1):21, 9 2015.
- [238] Yao-Yuan Mao, Andrew R. Zentner, and Risa H. Wechsler. Beyond assembly bias: exploring secondary halo biases for cluster-size haloes. *Monthly Notices of the Royal Astronomical Society*, 474(4):5143–5157, 3 2018.
- [239] Claudia Maraston. Evolutionary population synthesis: models, analysis of the ingredients and application to high-z galaxies. Monthly Notices of the Royal Astronomical Society, 362(3):799–825, September 2005.
- [240] S. L. Martell, S. Sharma, S. Buder, L. Duong, K. J. Schlesinger, J. Simpson, K. Lind, M. Ness, J. P. Marshall, M. Asplund, J. Bland-Hawthorn, A. R. Casey, G. De Silva, K. C. Freeman, J. Kos, J. Lin, D. B. Zucker, T. Zwitter, B. Anguiano, C. Bacigalupo, D. Carollo, L. Casagrande, G. S. Da Costa, J. Horner, D. Huber, E. A. Hyde, P. R. Kafle, G. F. Lewis, D. Nataf, C. A. Navin, D. Stello, C. G. Tinney, F. G. Watson, and R. Wittenmyer. The GALAH survey: observational overview and Gaia DR1 companion. *Monthly Notices of the Royal Astronomical Society*, 465(3):3203–3219, 3 2017.
- [241] C. Martin and GALEX Team. The Galaxy Evolution Explorer Early Data. In Matthew Colless, Lister Staveley-Smith, and Raylee A. Stathakis, editors, *Maps of the Cosmos*, volume 216, page 221, 1 2005.
- [242] L. Martinet and D. Friedli. Bar strength and star formation activity in late-type barred galaxies. *Astronomy and Astrophysics*, 323:363–373, 7 1997.

- [243] Karen L. Masters, Riccardo Giovanelli, and Martha P. Haynes. Internal Extinction in Spiral Galaxies in the Near-Infrared. *Astronomical Journal*, 126(1):158–174, 7 2003.
- [244] Karen L. Masters, Moein Mosleh, A. Kathy Romer, Robert C. Nichol, Steven P. Bamford, Kevin Schawinski, Chris J. Lintott, Dan Andreescu, Heather C. Campbell, Ben Crowcroft, Isabelle Doyle, Edward M. Edmondson, Phil Murray, M. Jordan Raddick, Anže Slosar, Alexander S. Szalay, and Jan Vandenberg. Galaxy Zoo: passive red spirals. *Monthly Notices of the Royal Astronomical Society*, 405(2):783–799, 6 2010.
- [245] Karen L. Masters, Robert Nichol, Steven Bamford, Moein Mosleh, Chris J. Lintott, Dan Andreescu, Edward M. Edmondson, William C. Keel, Phil Murray, M. Jordan Raddick, Kevin Schawinski, Anže Slosar, Alexander S. Szalay, Daniel Thomas, and Jan Vandenberg. Galaxy Zoo: dust in spiral galaxies. *Monthly Notices of the Royal Astronomical Society*, 404(2):792–810, 5 2010.
- [246] Karen L. Masters, Robert C. Nichol, Martha P. Haynes, William C. Keel, Chris Lintott, Brooke Simmons, Ramin Skibba, Steven Bamford, Riccardo Giovanelli, and Kevin Schawinski. Galaxy Zoo and ALFALFA: atomic gas and the regulation of star formation in barred disc galaxies. *Monthly Notices of the Royal Astronomical Society*, 424(3):2180–2192, 8 2012.
- [247] Karen L. Masters, Robert C. Nichol, Ben Hoyle, Chris Lintott, Steven P. Bamford, Edward M. Edmondson, Lucy Fortson, William C. Keel, Kevin Schawinski, Arfon M. Smith, and Daniel Thomas. Galaxy Zoo: bars in disc galaxies. *Monthly Notices of* the Royal Astronomical Society, 411(3):2026–2034, 3 2011.
- [248] Cameron McBride, A. A. Berlind, R. Scoccimarro, M. Manera, J. Tinker, M. Busha, R. Wechsler, H. Wu, and F. van den Bosch. LasDamas: Accurate Determination of the Abundance of Galaxy Clusters. In American Astronomical Society Meeting Abstracts #217, volume 217 of American Astronomical Society Meeting Abstracts, page 249.07, 1 2011.
- [249] A. McConnachie, J. Richardson, and D. Mackey. Andromeda and the Milky Way: Twin sisters, distant relations, or strangers in the night? In *European Physical Journal Web of Conferences*, volume 19, page 01003, 2 2012.
- [250] Paul J. McMillan. Mass models of the Milky Way. Monthly Notices of the Royal Astronomical Society, 414:2446–2457, 7 2011.
- [251] J. Trevor Mendel, Luc Simard, Sara L. Ellison, and David R. Patton. Towards a physical picture of star formation quenching: the photometric properties of recently

quenched galaxies in the Sloan Digital Sky Survey. Monthly Notices of the Royal Astronomical Society, 429(3):2212–2227, 3 2013.

- [252] Alexander J. Mendez, Alison L. Coil, Jennifer Lotz, Samir Salim, John Moustakas, and Luc Simard. AEGIS: The Morphologies of Green Galaxies at 0.4 < z < 1.2. The Astrophysical Journal, 736(2):110, 8 2011.
- [253] M. Meneghetti, M. Bartelmann, K. Dolag, L. Moscardini, F. Perrotta, C. Baccigalupi, and G. Tormen. Strong lensing efficiency of galaxy clusters in dark energy cosmologies. *Astronomy and Astrophysics*, 442(2):413–422, 11 2005.
- [254] David Merritt. The Distribution of Dark Matter in the Coma Cluster. Astrophysical Journal, 313:121, 2 1987.
- [255] H. J. Mo and S. Mao. Cosmological Formation of Disk Galaxies. In Dynamics of Galaxies: from the Early Universe to the Present, volume 197, page 145. IOP Publishing, 1 2000.
- [256] H. J. Mo, Shude Mao, and Simon D. M. White. The formation of galactic discs. Monthly Notices of the Royal Astronomical Society, 295:319–336, 4 1998.
- [257] Ole Möller, Priyamvada Natarajan, Jean-Paul Kneib, and A. W. Blain. Probing the Mass Distribution in Groups of Galaxies using Gravitational Lensing. Astrophysical Journal, 573(2):562–575, 7 2002.
- [258] Ben Moore. Evidence against dissipation-less dark matter from observations of galaxy haloes. *Nature*, 370:629–631, 8 1994.
- [259] Ben Moore, Juerg Diemand, Piero Madau, Marcel Zemp, and Joachim Stadel. Globular clusters, satellite galaxies and stellar haloes from early dark matter peaks. *Monthly Notices of the Royal Astronomical Society*, 368:563–570, 5 2006.
- [260] Ben Moore, Sebastiano Ghigna, Fabio Governato, George Lake, Thomas Quinn, Joachim Stadel, and Paolo Tozzi. Dark Matter Substructure within Galactic Halos. Astrophysical Journal, Letters, 524(1):L19–L22, 10 1999.
- [261] Surhud More, Frank C. van den Bosch, Marcello Cacciato, Ramin Skibba, H. J. Mo, and Xiaohu Yang. Satellite kinematics - III. Halo masses of central galaxies in SDSS. *Monthly Notices of the Royal Astronomical Society*, 410(1):210–226, 1 2011.

- [262] Masao Mori, Yuzuru Yoshii, and Ken'ichi Nomoto. Dissipative Process as a Mechanism of Differentiating Internal Structures between Dwarf and Normal Elliptical Galaxies in a Cold Dark Matter Universe. Astrophysical Journal, 511:585–594, 2 1999.
- [263] L. Morselli, A. Renzini, P. Popesso, and G. Erfanianfar. The effect of disc inclination on the main sequence of star-forming galaxies. *Monthly Notices of the Royal Astronomical Society*, 462(3):2355–2365, 11 2016.
- [264] J. C. Muñoz-Mateos, A. Gil de Paz, S. Boissier, J. Zamorano, T. Jarrett, J. Gallego, and B. F. Madore. Specific Star Formation Rate Profiles in Nearby Spiral Galaxies: Quantifying the Inside-Out Formation of Disks. *The Astrophysical Journal*, 658(2):1006–1026, 4 2007.
- [265] Simon J. Mutch, Darren J. Croton, and Gregory B. Poole. The Mid-life Crisis of the Milky Way and M31. The Astrophysical Journal, 736(2):84, 8 2011.
- [266] Jerome L Myers, A. (Arnold) Well, and ProQuest (Firm). Research design and statistical analysis. Mahwah, N.J.: Lawrence Erlbaum Associates, 2nd ed edition, 2003. Description based upon print version of record.
- [267] E. O. Nadler, A. Drlica-Wagner, K. Bechtol, S. Mau, R. H. Wechsler, V. Gluscevic, K. Boddy, A. B. Pace, T. S. Li, M. McNanna, A. H. Riley, J. García-Bellido, Y. Y. Mao, G. Green, D. L. Burke, A. Peter, B. Jain, T. M. C. Abbott, M. Aguena, S. Allam, J. Annis, S. Avila, D. Brooks, M. Carrasco Kind, J. Carretero, M. Costanzi, L. N. da Costa, J. De Vicente, S. Desai, H. T. Diehl, P. Doel, S. Everett, A. E. Evrard, B. Flaugher, J. Frieman, D. W. Gerdes, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, S. R. Hinton, K. Honscheid, D. Huterer, D. J. James, E. Krause, K. Kuehn, N. Kuropatkin, O. Lahav, M. A. G. Maia, J. L. Marshall, F. Menanteau, R. Miquel, A. Palmese, F. Paz-Chinchón, A. A. Plazas, A. K. Romer, E. Sanchez, V. Scarpine, S. Serrano, I. Sevilla-Noarbe, M. Smith, M. Soares-Santos, E. Suchyta, M. E. C. Swanson, G. Tarle, D. L. Tucker, A. R. Walker, W. Wester, and DES Collaboration. Constraints on Dark Matter Properties from Observations of Milky Way Satellite Galaxies. *Physical Review Letters*, 126(9):091101, 3 2021.
- [268] Ethan O. Nadler, Yao-Yuan Mao, Gregory M. Green, and Risa H. Wechsler. Modeling the connection between subhalos and satellites in milky way–like systems. *The Astrophysical Journal*, 873(1):34, 3 2019.
- [269] Daisuke Nagai and Andrey V. Kravtsov. The Radial Distribution of Galaxies in Λ Cold Dark Matter Clusters. Astrophysical Journal, 618(2):557–568, 1 2005.

- [270] Daisuke Nagai and Andrey V. Kravtsov. The Radial Distribution of Galaxies in Λ Cold Dark Matter Clusters. Astrophysical Journal, 618(2):557–568, 1 2005.
- [271] Daisuke Nagai, Alexey Vikhlinin, and Andrey V. Kravtsov. Testing X-Ray Measurements of Galaxy Clusters with Cosmological Simulations. Astrophysical Journal, 655(1):98–108, 1 2007.
- [272] J. F. Navarro, E. Hayashi, C. Power, A. R. Jenkins, C. S. Frenk, S. D. M. White, V. Springel, J. Stadel, and T. R. Quinn. The inner structure of ΛCDM haloes -III. Universality and asymptotic slopes. *Monthly Notices of the Royal Astronomical Society*, 349(3):1039–1051, 4 2004.
- [273] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. The Structure of Cold Dark Matter Halos. Astrophysical Journal, 462:563, 5 1996.
- [274] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. A Universal Density Profile from Hierarchical Clustering. *Astrophysical Journal*, 490(2):493–508, 12 1997.
- [275] Julio F. Navarro, Aaron Ludlow, Volker Springel, Jie Wang, Mark Vogelsberger, Simon D. M. White, Adrian Jenkins, Carlos S. Frenk, and Amina Helmi. The diversity and similarity of simulated cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 402(1):21–34, 2 2010.
- [276] J.A. Nelder and R. Mead. A simplex method for function minimization. *Comput. J.*, 7:308–313, 1965.
- [277] Fabrizio Nesti and Paolo Salucci. The Dark Matter halo of the Milky Way, AD 2013. Journal of Cosmology and Astro-Particle Physics, 2013:016, 7 2013.
- [278] Angelo F. Neto, Liang Gao, Philip Bett, Shaun Cole, Julio F. Navarro, Carlos S. Frenk, Simon D. M. White, Volker Springel, and Adrian Jenkins. The statistics of Λ CDM halo concentrations. Monthly Notices of the Royal Astronomical Society, 381(4):1450–1462, 11 2007.
- [279] Andrew B. Newman, Tommaso Treu, Richard S. Ellis, David J. Sand, Carlo Nipoti, Johan Richard, and Eric Jullo. The Density Profiles of Massive, Relaxed Galaxy Clusters. I. The Total Density Over Three Decades in Radius. Astrophysical Journal, 765(1):24, 3 2013.

- [280] A. M. Nierenberg, T. Treu, G. Brammer, A. H. G. Peter, C. D. Fassnacht, C. R. Keeton, C. S. Kochanek, K. B. Schmidt, D. Sluse, and S. A. Wright. Probing dark matter substructure in the gravitational lens HE 0435-1223 with the WFC3 grism. Monthly Notices of the Royal Astronomical Society, 471(2):2224–2236, 10 2017.
- [281] S. Noll, D. Burgarella, E. Giovannoli, V. Buat, D. Marcillac, and J. C. Muñoz-Mateos. Analysis of galaxy spectral energy distributions from far-UV to far-IR with CIGALE: studying a SINGS test sample. Astronomy and Astrophysics, 507(3):1793–1813, 12 2009.
- [282] Nobuhiro Okabe, Graham P. Smith, Keiichi Umetsu, Masahiro Takada, and Toshifumi Futamase. LoCuSS: The Mass Density Profile of Massive Galaxy Clusters at z = 0.2. Astrophysical Journal, 769(2):L35, 6 2013.
- [283] J. B. Oke and Allan Sandage. Energy Distributions, K Corrections, and the Stebbins-Whitford Effect for Giant Elliptical Galaxies. Astrophysical Journal, 154:21, October 1968.
- [284] Ana Laura O'Mill, Fernanda Duplancic, Diego García Lambas, and Jr. Sodré, Laerte.
  Photometric redshifts and k-corrections for the Sloan Digital Sky Survey Data Release
  7. Monthly Notices of the Royal Astronomical Society, 413(2):1395–1408, May 2011.
- [285] Julian Onions, Alexander Knebe, Frazer R. Pearce, Stuart I. Muldrew, Hanni Lux, Steffen R. Knollmann, Yago Ascasibar, Peter Behroozi, Pascal Elahi, Jiaxin Han, Michal Maciejewski, Manuel E. Merchán, Mark Neyrinck, Andrés. N. Ruiz, Mario A. Sgró, Volker Springel, and Dylan Tweed. Subhaloes going Notts: the subhalo-finder comparison project. *Monthly Notices of the Royal Astronomical Society*, 423(2):1200– 1214, 6 2012.
- [286] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.
- [287] P. J. E. Peebles. Origin of the Angular Momentum of Galaxies. Astrophysical Journal, 155:393, 2 1969.
- [288] Phillip James Edwin. Peebles. *The large-scale structure of the universe*. Princeton Series in Physics, 1980.

- [289] Ying-jie Peng, Simon J. Lilly, Alvio Renzini, and Marcella Carollo. Mass and Environment as Drivers of Galaxy Evolution. II. The Quenching of Satellite Galaxies as the Origin of Environmental Effects. *The Astrophysical Journal*, 757(1):4, 9 2012.
- [290] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. The Astrophysical Journal, 142:419–421, 7 1965.
- [291] F. Perez and B. E. Granger. Ipython: A system for interactive scientific computing. Computing in Science Engineering, 9(3):21–29, 2007.
- [292] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and The Supernova Cosmology Project. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *The Astrophysical Journal*, 517(2):565–586, 6 1999.
- [293] Stephen M. Pompea and G. H. Rieke. A Test of the Association of Infrared Activity with Bars. *The Astrophysical Journal*, 356:416, 6 1990.
- [294] C. Power, J. F. Navarro, A. Jenkins, C. S. Frenk, S. D. M. White, V. Springel, J. Stadel, and T. Quinn. The inner structure of ACDM haloes - I. A numerical convergence study. *Monthly Notices of the Royal Astronomical Society*, 338(1):14–34, 1 2003.
- [295] Francisco Prada, Anatoly A. Klypin, Antonio J. Cuesta, Juan E. Betancort-Rijo, and Joel Primack. Halo concentrations in the standard Λ cold dark matter cosmology. Monthly Notices of the Royal Astronomical Society, 423(4):3018–3030, 7 2012.
- [296] Abhishek Prakash, Timothy C. Licquia, Jeffrey A. Newman, Ashley J. Ross, Adam D. Myers, Kyle S. Dawson, Jean-Paul Kneib, Will J. Percival, Julian E. Bautista, Johan Comparat, Jeremy L. Tinker, David J. Schlegel, Rita Tojeiro, Shirley Ho, Dustin Lang, Sandhya M. Rao, Cameron K. McBride, Guangtun Ben Zhu, Joel R. Brownstein, Stephen Bailey, Adam S. Bolton, Timothée Delubac, Vivek Mariappan, Michael R. Blanton, Beth Reid, Donald P. Schneider, Hee-Jong Seo, Aurelio Carnero Rosell, and Francisco Prada. The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Luminous Red Galaxy Target Selection. *The Astrophysical Journal Supplement Series*, 224:34, 6 2016.

- [297] P. Prugniel and F. Simien. The fundamental plane of early-type galaxies: nonhomology of the spatial structure. Astronomy and Astrophysics, 321:111–122, 5 1997.
- [298] Chris W. Purcell, James S. Bullock, and Andrew R. Zentner. Shredded Galaxies as the Source of Diffuse Intrahalo Light on Varying Scales. Astrophysical Journal, 666(1):20–33, 9 2007.
- [299] Chris W. Purcell and Andrew R. Zentner. Bailing out the Milky Way: variation in the properties of massive dwarfs among galaxy-sized systems. *Journal of Cosmology* and Astro-Particle Physics, 2012:007, 12 2012.
- [300] A. B. A. Queiroz, C. Chiappini, A. Perez-Villegas, A. Khalatyan, F. Anders, B. Barbuy, B. X. Santiago, M. Steinmetz, K. Cunha, M. Schultheis, S. R. Majewski, I. Minchev, D. Minniti, R. L. Beaton, R. E. Cohen, L. N. da Costa, J. G. Fernández-Trincado, D. A. Garcia-Hernández, D. Geisler, S. Hasselquist, R. R. Lane, C. Nitschelm, A. Rojas-Arriagada, A. Roman-Lopes, V. Smith, and G. Zasowski. The Milky Way bar and bulge revealed by APOGEE and Gaia EDR3. Astronomy and Astrophysics, 656:A156, 12 2021.
- [301] Valery Rashkov, Piero Madau, Michael Kuhlen, and Jürg Diemand. On the Assembly of the Milky Way Dwarf Satellites and Their Common Mass Scale. *Astrophysical Journal*, 745:142, 2 2012.
- [302] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. the MIT Press, 2006.
- [303] Darren Reed, Fabio Governato, Thomas Quinn, Jeffrey Gardner, Joachim Stadel, and George Lake. Dark matter subhaloes in numerical simulations. *Monthly Notices of* the Royal Astronomical Society, 359(4):1537–1548, 6 2005.
- [304] M. J. Reid, A. C. S. Readhead, R. C. Vermeulen, and R. N. Treuhaft. The Proper Motion of Sagittarius A\*. I. First VLBA Results. Astrophysical Journal, 524:816–823, 10 1999.
- [305] Adam G. Riess, Alexei V. Filippenko, Peter Challis, Alejandro Clocchiatti, Alan Diercks, Peter M. Garnavich, Ron L. Gilliland, Craig J. Hogan, Saurabh Jha, Robert P. Kirshner, B. Leibundgut, M. M. Phillips, David Reiss, Brian P. Schmidt, Robert A. Schommer, R. Chris Smith, J. Spyromilio, Christopher Stubbs, Nicholas B. Suntzeff, and John Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *The Astronomical Journal*, 116(3):1009–1038, September 1998.

- [306] G. R. Ruchti, J. I. Read, S. Feltzing, A. M. Serenelli, P. McMillan, K. Lind, T. Bensby, M. Bergemann, M. Asplund, A. Vallenari, E. Flaccomio, E. Pancino, A. J. Korn, A. Recio-Blanco, A. Bayo, G. Carraro, M. T. Costado, F. Damiani, U. Heiter, A. Hourihane, P. Jofré, G. Kordopatis, C. Lardo, P. de Laverny, L. Monaco, L. Morbidelli, L. Sbordone, C. C. Worley, and S. Zaggia. The Gaia-ESO Survey: a quiescent Milky Way with no significant dark/stellar accreted disc. *Monthly Notices of the Royal* Astronomical Society, 450:2874–2887, 7 2015.
- [307] Kanak Saha and Arianna Cortesi. Forming Lenticular Galaxies via Violent Disk Instability. *The Astrophysical Journal, Letters*, 862(1):L12, 7 2018.
- [308] Amélie Saintonge, Linda J. Tacconi, Silvia Fabello, Jing Wang, Barbara Catinella, Reinhard Genzel, Javier Graciá-Carpio, Carsten Kramer, Sean Moran, Timothy M. Heckman, David Schiminovich, Karl Schuster, and Stijn Wuyts. The Impact of Interactions, Bars, Bulges, and Active Galactic Nuclei on Star Formation Efficiency in Local Massive Galaxies. *The Astrophysical Journal*, 758(2):73, 10 2012.
- [309] R. K. Saito, D. Minniti, B. Dias, M. Hempel, M. Rejkuba, J. Alonso-García, B. Barbuy, M. Catelan, J. P. Emerson, O. A. Gonzalez, P. W. Lucas, and M. Zoccali. Milky Way demographics with the VVV survey. I. The 84-million star colour-magnitude diagram of the Galactic bulge. Astronomy and Astrophysics, 544:A147, 8 2012.
- [310] S. Salim. Green Valley Galaxies. Serbian Astronomical Journal, 189:1–14, 12 2014.
- [311] Samir Salim, Médéric Boquien, and Janice C. Lee. Dust Attenuation Curves in the Local Universe: Demographics and New Laws for Star-forming Galaxies and High-redshift Analogs. *The Astrophysical Journal*, 859(1):11, 5 2018.
- [312] Samir Salim, Janice C. Lee, Steven Janowiecki, Elisabete da Cunha, Mark Dickinson, Médéric Boquien, Denis Burgarella, John J. Salzer, and Stéphane Charlot. GALEX-SDSS-WISE Legacy Catalog (GSWLC): Star Formation Rates, Stellar Masses, and Dust Attenuations of 700,000 Low-redshift Galaxies. *The Astrophysical Journal Supplementary Series*, 227(1):2, 11 2016.
- [313] Samir Salim, R. Michael Rich, Stéphane Charlot, Jarle Brinchmann, Benjamin D. Johnson, David Schiminovich, Mark Seibert, Ryan Mallery, Timothy M. Heckman, Karl Forster, Peter G. Friedman, D. Christopher Martin, Patrick Morrissey, Susan G. Neff, Todd Small, Ted K. Wyder, Luciana Bianchi, José Donas, Young-Wook Lee, Barry F. Madore, Bruno Milliard, Alex S. Szalay, Barry Y. Welsh, and Sukyoung K. Yi. UV Star Formation Rates in the Local Universe. The Astrophysical Journal Supplementary Series, 173(2):267–292, 12 2007.

- [314] Allan Sandage and John Bedke. *The Carnegie atlas of galaxies*, volume 638 of . Carnegie Inst of Washington, 1994.
- [315] Robyn E. Sanderson, Andrew Wetzel, Sarah Loebman, Sanjib Sharma, Philip F. Hopkins, Shea Garrison-Kimmel, Claude-André Faucher-Giguère, Dušan Kereš, and Eliot Quataert. Synthetic Gaia Surveys from the FIRE Cosmological Simulations of Milky Way-mass Galaxies. Astrophysical Journal, Supplement, 246(1):6, 1 2020.
- [316] Till Sawala, Carlos S. Frenk, Azadeh Fattahi, Julio F. Navarro, Richard G. Bower, Robert A. Crain, Claudio Dalla Vecchia, Michelle Furlong, John. C. Helly, Adrian Jenkins, Kyle A. Oman, Matthieu Schaller, Joop Schaye, Tom Theuns, James Trayford, and Simon D. M. White. The APOSTLE simulations: solutions to the Local Group's cosmic puzzles. *Monthly Notices of the Royal Astronomical Society*, 457(2):1931–1943, 4 2016.
- [317] Till Sawala, Pauli Pihajoki, Peter H. Johansson, Carlos S. Frenk, Julio F. Navarro, Kyle A. Oman, and Simon D. M. White. Shaken and stirred: the Milky Way's dark substructures. *Monthly Notices of the Royal Astronomical Society*, 467(4):4383–4400, 6 2017.
- [318] Cecilia Scannapieco, Patricia B. Tissera, Simon D. M. White, and Volker Springel. Effects of supernova feedback on the formation of galaxy discs. *Monthly Notices of the Royal Astronomical Society*, 389:1137–1149, 9 2008.
- [319] Kevin Schawinski, C. Megan Urry, Brooke D. Simmons, Lucy Fortson, Sugata Kaviraj, William C. Keel, Chris J. Lintott, Karen L. Masters, Robert C. Nichol, Marc Sarzi, Ramin Skibba, Ezequiel Treister, Kyle W. Willett, O. Ivy Wong, and Sukyoung K. Yi. The green valley is a red herring: Galaxy Zoo reveals two evolutionary pathways towards quenching of star formation in early- and late-type galaxies. *Monthly Notices* of the Royal Astronomical Society, 440(1):889–907, 5 2014.
- [320] David J. Schlegel, Douglas P. Finkbeiner, and Marc Davis. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *The Astrophysical Journal*, 500(2):525–553, 6 1998.
- [321] Michael D. Schneider, Carlos S. Frenk, and Shaun Cole. The shapes and alignments of dark matter halos. *Journal of Cosmology and Astro-Particle Physics*, 2012:030, 5 2012.
- [322] SDSS. Sdss image gallery, 2021.

- [323] J. L. Sersic. Atlas de Galaxias Australes. Observatorio Astronòmico, 1968.
- [324] Zhengyi Shao, Quanbao Xiao, Shiyin Shen, H. J. Mo, Xiaoyang Xia, and Zugan Deng. Inclination-dependent Luminosity Function of Spiral Galaxies in the Sloan Digital Sky Survey: Implications for Dust Extinction. *The Astrophysical Journal*, 659(2):1159– 1171, 4 2007.
- [325] Juntai Shen and Xing-Wu Zheng. The bar and spiral arms in the Milky Way: structure and kinematics. *Research in Astronomy and Astrophysics*, 20(10):159, 10 2020.
- [326] L. Silva, A. Schurer, G. L. Granato, C. Almeida, C. M. Baugh, C. S. Frenk, C. G. Lacey, L. Paoletti, A. Petrella, and D. Selvestrel. Modelling the spectral energy distribution of galaxies: introducing the artificial neural network. *Monthly Notices of the Royal Astronomical Society*, 410(3):2043–2056, 1 2011.
- [327] Luc Simard, J. Trevor Mendel, David R. Patton, Sara L. Ellison, and Alan W. Mc-Connachie. A Catalog of Bulge+disk Decompositions and Updated Photometry for 1.12 Million Galaxies in the Sloan Digital Sky Survey. The Astrophysical Journal Supplementary Series, 196(1):11, 9 2011.
- [328] Luc Simard, Christopher N. A. Willmer, Nicole P. Vogt, Vicki L. Sarajedini, Andrew C. Phillips, Benjamin J. Weiner, David C. Koo, Myungshin Im, Garth D. Illingworth, and S. M. Faber. The DEEP Groth Strip Survey. II. Hubble Space Telescope Structural Parameters of Galaxies in the Groth Strip. The Astrophysical Journal Supplementary Series, 142(1):1–33, 9 2002.
- [329] B. D. Simmons, Thomas Melvin, Chris Lintott, Karen L. Masters, Kyle W. Willett, William C. Keel, R. J. Smethurst, Edmond Cheung, Robert C. Nichol, Kevin Schawinski, Michael Rutkowski, Jeyhan S. Kartaltepe, Eric F. Bell, Kevin R. V. Casteels, Christopher J. Conselice, Omar Almaini, Henry C. Ferguson, Lucy Fortson, William Hartley, Dale Kocevski, Anton M. Koekemoer, Daniel H. McIntosh, Alice Mortlock, Jeffrey A. Newman, Jamie Ownsworth, Steven Bamford, Tomas Dahlen, Sandra M. Faber, Steven L. Finkelstein, Adriano Fontana, Audrey Galametz, N. A. Grogin, Ruth Grützbauch, Yicheng Guo, Boris Häußler, Kian J. Jek, Sugata Kaviraj, Ray A. Lucas, Michael Peth, Mara Salvato, Tommy Wiklind, and Stijn Wuyts. Galaxy Zoo: CAN-DELS barred discs and bar fractions. *Monthly Notices of the Royal Astronomical Society*, 445(4):3466–3474, 12 2014.
- [330] M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, R. Capps, T. Chester, J. Elias, J. Huchra, J. Liebert, C. Lonsdale, D. G. Monet, S. Price, P. Seitzer, T. Jarrett, J. D. Kirkpatrick, J. E. Gizis, E. Howard,

T. Evans, J. Fowler, L. Fullmer, R. Hurt, R. Light, E. L. Kopan, K. A. Marsh, H. L. McCallon, R. Tam, S. Van Dyk, and S. Wheelock. The Two Micron All Sky Survey (2MASS). *The Astronomical Journal*, 131(2):1163–1183, 2 2006.

- [331] R. J. Smethurst, C. J. Lintott, B. D. Simmons, K. Schawinski, P. J. Marshall, S. Bamford, L. Fortson, S. Kaviraj, K. L. Masters, T. Melvin, R. C. Nichol, R. A. Skibba, and K. W. Willett. Galaxy Zoo: evidence for diverse star formation histories through the green valley. *Monthly Notices of the Royal Astronomical Society*, 450(1):435–453, 6 2015.
- [332] Rachel S. Somerville and Joel R. Primack. Semi-analytic modelling of galaxy formation: the local Universe. Monthly Notices of the Royal Astronomical Society, 310(4):1087–1110, 12 1999.
- [333] Jesper Sommer-Larsen and Alexandre Dolgov. Formation of Disk Galaxies: Warm Dark Matter and the Angular Momentum Problem. Astrophysical Journal, 551:608– 623, 4 2001.
- [334] David N. Spergel and Paul J. Steinhardt. Observational Evidence for Self-Interacting Cold Dark Matter. *Physical Review Letters*, 84:3760–3763, 4 2000.
- [335] V. Springel, S. D. M. White, C. S. Frenk, J. F. Navarro, A. Jenkins, M. Vogelsberger, J. Wang, A. Ludlow, and A. Helmi. Prospects for detecting supersymmetric dark matter in the Galactic halo. *Nature*, 456:73–76, 11 2008.
- [336] Volker Springel. The cosmological simulation code GADGET-2. Monthly Notices of the Royal Astronomical Society, 364(4):1105–1134, 12 2005.
- [337] Felix Stoehr, Simon D. M. White, Giuseppe Tormen, and Volker Springel. The satellite population of the Milky Way in a ACDM universe. *Monthly Notices of the Royal Astronomical Society*, 335:L84–L88, 10 2002.
- [338] Iskra Strateva, Željko Ivezić, Gillian R. Knapp, Vijay K. Narayanan, Michael A. Strauss, James E. Gunn, Robert H. Lupton, David Schlegel, Neta A. Bahcall, Jon Brinkmann, Robert J. Brunner, Tamás Budavári, István Csabai, Francisco Javier Castander, Mamoru Doi, Masataka Fukugita, Zsuzsanna Győry, Masaru Hamabe, Greg Hennessy, Takashi Ichikawa, Peter Z. Kunszt, Don Q. Lamb, Timothy A. McKay, Sadanori Okamura, Judith Racusin, Maki Sekiguchi, Donald P. Schneider, Kazuhiro Shimasaku, and Donald York. Color Separation of Galaxy Types in the Sloan Digital Sky Survey Imaging Data. Astronomical Journal, 122(4):1861–1874, 10 2001.

- [339] Michael A. Strauss, David H. Weinberg, Robert H. Lupton, Vijay K. Narayanan, James Annis, Mariangela Bernardi, Michael Blanton, Scott Burles, A. J. Connolly, Julianne Dalcanton, Mamoru Doi, Daniel Eisenstein, Joshua A. Frieman, Masataka Fukugita, James E. Gunn, Željko Ivezić, Stephen Kent, Rita S. J. Kim, G. R. Knapp, Richard G. Kron, Jeffrey A. Munn, Heidi Jo Newberg, R. C. Nichol, Sadanori Okamura, Thomas R. Quinn, Michael W. Richmond, David J. Schlegel, Kazuhiro Shimasaku, Mark SubbaRao, Alexander S. Szalay, Dan Vanden Berk, Michael S. Vogeley, Brian Yanny, Naoki Yasuda, Donald G. York, and Idit Zehavi. Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample. Astronomical Journal, 124(3):1810–1824, 9 2002.
- [340] James E. Taylor and Arif Babul. The evolution of substructure in galaxy, group and cluster haloes - II. Global properties. Monthly Notices of the Royal Astronomical Society, 364(2):515-534, 12 2005.
- [341] A. F. Teklu, R.-S. Remus, K. Dolag, A. M. Beck, A. Burkert, A. S. Schmidt, F. Schulze, and L. K. Steinborn. Connecting Angular Momentum and Galactic Dynamics: The Complex Interplay between Spin, Mass, and Morphology. *Astrophysical Journal*, 812:29, 10 2015.
- [342] Balša Terzić and Alister W. Graham. Density-potential pairs for spherical stellar systems with Sérsic light profiles and (optional) power-law cores. *Monthly Notices of the Royal Astronomical Society*, 362(1):197–212, 9 2005.
- [343] Erik J. Tollerud, Michael Boylan-Kolchin, and James S. Bullock. M31 satellite masses compared to ACDM subhaloes. *Monthly Notices of the Royal Astronomical Society*, 440:3511–3519, 6 2014.
- [344] Erik J. Tollerud, James S. Bullock, Louis E. Strigari, and Beth Willman. Hundreds of Milky Way Satellites? Luminosity Bias in the Satellite Luminosity Function. Astrophysical Journal, 688:277–289, 11 2008.
- [345] Christy A. Tremonti, Timothy M. Heckman, Guinevere Kauffmann, Jarle Brinchmann, Stéphane Charlot, Simon D. M. White, Mark Seibert, Eric W. Peng, David J. Schlegel, Alan Uomoto, Masataka Fukugita, and Jon Brinkmann. The Origin of the Mass-Metallicity Relation: Insights from 53,000 Star-forming Galaxies in the Sloan Digital Sky Survey. Astrophysical Journal, 613(2):898–913, October 2004.
- [346] A. D. Tubbs. The inhibition of star formation in barred spiral galaxies. *The Astro*physical Journal, 255:458–466, 4 1982.

- [347] R. J. Tuffs, C. C. Popescu, H. J. Völk, N. D. Kylafis, and M. A. Dopita. Modelling the spectral energy distribution of galaxies. III. Attenuation of stellar light in spiral galaxies. *Astronomy and Astrophysics*, 419:821–835, 6 2004.
- [348] R. B. Tully and J. R. Fisher. A new method of determining distances to galaxies. Astronomy and Astrophysics, 54:661–673, 2 1977.
- [349] R. Brent Tully, Hélène Courtois, Yehuda Hoffman, and Daniel Pomarède. The Laniakea supercluster of galaxies. *Nature*, 513(7516):71–73, 9 2014.
- [350] Sarah E. Tuttle and Stephanie Tonnesen. BreakBRD Galaxies. I. Global Properties of Spiral Galaxies with Central Star Formation in Red Disks. *The Astrophysical Journal*, 889(2):188, 2 2020.
- [351] Keiichi Umetsu, Adi Zitrin, Daniel Gruen, Julian Merten, Megan Donahue, and Marc Postman. CLASH: Joint Analysis of Strong-lensing, Weak-lensing Shear, and Magnification Data for 20 Galaxy Clusters. Astrophysical Journal, 821(2):116, 4 2016.
- [352] M. Unavane, Rosemary F. G. Wyse, and Gerard Gilmore. The merging history of the Milky Way. *Monthly Notices of the Royal Astronomical Society*, 278:727–736, 2 1996.
- [353] Cayman T. Unterborn and Barbara S. Ryden. Inclination-Dependent Extinction Effects in Disk Galaxies in the Sloan Digital Sky Survey. *The Astrophysical Journal*, 687(2):976–985, 11 2008.
- [354] Emir Uzeirbegovic, Garreth Martin, and Sugata Kaviraj. How the spectral energy distribution and galaxy morphology constrain each other, with application to morphological selection using galaxy colours. *Monthly Notices of the Royal Astronomical Society*, 510(3):3849–3857, March 2022.
- [355] Jesse van de Sande, Nicholas Scott, Joss Bland-Hawthorn, Sarah Brough, Julia J. Bryant, Matthew Colless, Luca Cortese, Scott M. Croom, Francesco d'Eugenio, Caroline Foster, Michael Goodwin, Iraklis S. Konstantopoulos, Jon S. Lawrence, Richard M. McDermid, Anne M. Medling, Matt S. Owers, Samuel N. Richards, and Rob Sharp. A relation between the characteristic stellar ages of galaxies and their intrinsic shapes. *Nature Astronomy*, 2:483–488, 4 2018.
- [356] S. van den Bergh. A new classification system for galaxies. *The Astrophysical Journal*, 206:883–887, 6 1976.

- [357] Frank C. van den Bosch, Daniel Aquino, Xiaohu Yang, H. J. Mo, Anna Pasquali, Daniel H. McIntosh, Simone M. Weinmann, and Xi Kang. The importance of satellite quenching for the build-up of the red sequence of present-day galaxies. *Monthly Notices* of the Royal Astronomical Society, 387(1):79–91, 6 2008.
- [358] Frank C. van den Bosch and Fangzhou Jiang. Statistics of Dark Matter Substructure: II. Comparison of Model with Simulation Results. *ArXiv e-prints*, page arXiv:1403.6835, 3 2014.
- [359] P. C. van der Kruit and K. C. Freeman. Galaxy Disks. Annual Review of Astronomy and Astrophysics, 49(1):301–371, 9 2011.
- [360] Roeland P. van der Marel and Nitya Kallivayalil. Third-epoch Magellanic Cloud Proper Motions. II. The Large Magellanic Cloud Rotation Field in Three Dimensions. *Astrophysical Journal*, 781:121, 2 2014.
- [361] Matias Vera, Sol Alonso, and Georgina Coldwell. Effect of bars on the galaxy properties. *Astronomy and Astrophysics*, 595:A63, 10 2016.
- [362] C. Vera-Ciro and A. Helmi. Constraints on the Shape of the Milky Way Dark Matter Halo from the Sagittarius Stream. *Astrophysical Journal, Letters*, 773:L4, 8 2013.
- [363] Antonio S. Villarreal, Andrew R. Zentner, Yao-Yuan Mao, Chris W. Purcell, Frank C. van den Bosch, Benedikt Diemer, Johannes U. Lange, Kuan Wang, and Duncan Campbell. The immitigable nature of assembly bias: the impact of halo definition on assembly bias. *Monthly Notices of the Royal Astronomical Society*, 472:1088–1105, 11 2017.
- [364] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake Vand erPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1. 0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17:261–272, 2020.
- [365] Maya Vitvitska, Anatoly A. Klypin, Andrey V. Kravtsov, Risa H. Wechsler, Joel R. Primack, and James S. Bullock. The Origin of Angular Momentum in Dark Matter Halos. *The Astrophysical Journal*, 581(2):799–809, 12 2002.

- [366] S. von Hoerner. Die numerische Integration des n-Körper-Problemes für Sternhaufen.
  I. Zeitschrift fuer Astrophysik, 50:184–214, 1 1960.
- [367] M. G. Walker, M. Mateo, E. W. Olszewski, J. Peñarrubia, N. Wyn Evans, and G. Gilmore. A universal mass profile for dwarf spheroidal galaxies? *Astrophysical Journal*, 704:1274–1287, 10 2009.
- [368] J. Wang, J. F. Navarro, C. S. Frenk, S. D. M. White, V. Springel, A. Jenkins, A. Helmi, A. Ludlow, and M. Vogelsberger. Assembly history and structure of galactic cold dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 413(2):1373–1382, 5 2011.
- [369] Kuan Wang, Yao-Yuan Mao, Andrew R. Zentner, Johannes U. Lange, Frank C. van den Bosch, and Risa H. Wechsler. Concentrations of Dark Haloes Emerge from Their Merger Histories. arXiv e-prints, page arXiv:2004.13732, 4 2020.
- [370] Mei-Yu Wang, Annika H. G. Peter, Louis E. Strigari, Andrew R. Zentner, Bryan Arant, Shea Garrison-Kimmel, and Miguel Rocha. Cosmological simulations of decaying dark matter: implications for small-scale structure of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 445:614–629, 11 2014.
- [371] Risa H. Wechsler, James S. Bullock, Joel R. Primack, Andrey V. Kravtsov, and Avishai Dekel. Concentrations of Dark Halos from Their Assembly Histories. Astrophysical Journal, 568(1):52–70, 3 2002.
- [372] Risa H. Wechsler, Andrew R. Zentner, James S. Bullock, Andrey V. Kravtsov, and Brandon Allgood. The Dependence of Halo Clustering on Halo Formation History, Concentration, and Occupation. *Astrophysical Journal*, 652(1):71–84, 11 2006.
- [373] Benjamin J. Weiner, Andrew C. Phillips, S. M. Faber, Christopher N. A. Willmer, Nicole P. Vogt, Luc Simard, Karl Gebhardt, Myungshin Im, D. C. Koo, Vicki L. Sarajedini, Katherine L. Wu, Duncan A. Forbes, Caryl Gronwall, Edward J. Groth, G. D. Illingworth, R. G. Kron, Jason Rhodes, A. S. Szalay, and M. Takamiya. The DEEP Groth Strip Galaxy Redshift Survey. III. Redshift Catalog and Properties of Galaxies. *The Astrophysical Journal*, 620(2):595–617, 2 2005.
- [374] Eric W. Weisstein. Root-mean-square, 2018.

- [375] Eduard Westra, Margaret J. Geller, Michael J. Kurtz, Daniel G. Fabricant, and Ian Dell'Antonio. Empirical Optical k-Corrections for Redshifts. *Publications of the ASP*, 122(897):1258, November 2010.
- [376] Andrew R. Wetzel, Philip F. Hopkins, Ji-hoon Kim, Claude-André Faucher-Giguère, Dušan Kereš, and Eliot Quataert. Reconciling Dwarf Galaxies with ΛCDM Cosmology: Simulating a Realistic Population of Satellites around a Milky Way-mass Galaxy. *The Astrophysical Journal, Letters*, 827(2):L23, 8 2016.
- [377] S. D. M. White. Formation and Evolution of Galaxies. In Cosmology and Large Scale Structure, page 349, 1 1996.
- [378] S. D. M. White and M. J. Rees. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. *Monthly Notices of the Royal Astronomical Society*, 183:341–358, 5 1978.
- [379] Simon D. M. White and Carlos S. Frenk. Galaxy Formation through Hierarchical Clustering. Astrophysical Journal, 379:52, 9 1991.
- [380] Theresa Wiegert and Jayanne English. Kinematic classification of non-interacting spiral galaxies. *New Astronomy*, 26:40–61, 1 2014.
- [381] Kyle W. Willett, Chris J. Lintott, Steven P. Bamford, Karen L. Masters, Brooke D. Simmons, Kevin R. V. Casteels, Edward M. Edmondson, Lucy F. Fortson, Sugata Kaviraj, William C. Keel, Thomas Melvin, Robert C. Nichol, M. Jordan Raddick, Kevin Schawinski, Robert J. Simpson, Ramin A. Skibba, Arfon M. Smith, and Daniel Thomas. Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 435(4):2835–2860, 11 2013.
- [382] C. N. A. Willmer, S. M. Faber, D. C. Koo, B. J. Weiner, J. A. Newman, A. L. Coil, A. J. Connolly, C. Conroy, M. C. Cooper, M. Davis, D. P. Finkbeiner, B. F. Gerke, P. Guhathakurta, J. Harker, N. Kaiser, S. Kassin, N. P. Konidaris, L. Lin, G. Luppino, D. S. Madgwick, K. G. Noeske, A. C. Phillips, and R. Yan. The Deep Evolutionary Exploratory Probe 2 Galaxy Redshift Survey: The Galaxy Luminosity Function to z~1. Astrophysical Journal, 647(2):853–873, August 2006.
- [383] Edward L. Wright, Peter R. M. Eisenhardt, Amy K. Mainzer, Michael E. Ressler, Roc M. Cutri, Thomas Jarrett, J. Davy Kirkpatrick, Deborah Padgett, Robert S. McMillan, Michael Skrutskie, S. A. Stanford, Martin Cohen, Russell G. Walker, John C. Mather, David Leisawitz, III Gautier, Thomas N., Ian McLean, Dominic

Benford, Carol J. Lonsdale, Andrew Blain, Bryan Mendez, William R. Irace, Valerie Duval, Fengchuan Liu, Don Royer, Ingolf Heinrichsen, Joan Howard, Mark Shannon, Martha Kendall, Amy L. Walsh, Mark Larsen, Joel G. Cardon, Scott Schick, Mark Schwalm, Mohamed Abid, Beth Fabinsky, Larry Naes, and Chao-Wei Tsai. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *Astronomical Journal*, 140(6):1868–1881, 12 2010.

- [384] Edward L. Wright, Peter R. M. Eisenhardt, Amy K. Mainzer, Michael E. Ressler, Roc M. Cutri, Thomas Jarrett, J. Davy Kirkpatrick, Deborah Padgett, Robert S. McMillan, Michael Skrutskie, S. A. Stanford, Martin Cohen, Russell G. Walker, John C. Mather, David Leisawitz, III Gautier, Thomas N., Ian McLean, Dominic Benford, Carol J. Lonsdale, Andrew Blain, Bryan Mendez, William R. Irace, Valerie Duval, Fengchuan Liu, Don Royer, Ingolf Heinrichsen, Joan Howard, Mark Shannon, Martha Kendall, Amy L. Walsh, Mark Larsen, Joel G. Cardon, Scott Schick, Mark Schwalm, Mohamed Abid, Beth Fabinsky, Larry Naes, and Chao-Wei Tsai. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. Astronomical Journal, 140(6):1868–1881, December 2010.
- [385] Hao-Yi Wu, Oliver Hahn, Risa H. Wechsler, Yao-Yuan Mao, and Peter S. Behroozi. Rhapsody. I. Structural Properties and Formation History from a Statistical Sample of Re-simulated Cluster-size Halos. Astrophysical Journal, 763(2):70, 2 2013.
- [386] Ted K. Wyder, D. Christopher Martin, David Schiminovich, Mark Seibert, Tamás Budavári, Marie A. Treyer, Tom A. Barlow, Karl Forster, Peter G. Friedman, Patrick Morrissey, Susan G. Neff, Todd Small, Luciana Bianchi, José Donas, Timothy M. Heckman, Young-Wook Lee, Barry F. Madore, Bruno Milliard, R. Michael Rich, Alex S. Szalay, Barry Y. Welsh, and Sukyoung K. Yi. The UV-Optical Galaxy Color-Magnitude Diagram. I. Basic Properties. *The Astrophysical Journal Supplementary Series*, 173(2):293–314, 12 2007.
- [387] Ting Xiao, Tinggui Wang, Huiyuan Wang, Hongyan Zhou, Honglin Lu, and Xiaobo Dong. Dust reddening in star-forming galaxies. Monthly Notices of the Royal Astronomical Society, 421(1):486–501, 03 2012.
- [388] X. X. Xue, H. W. Rix, G. Zhao, P. Re Fiorentin, T. Naab, M. Steinmetz, F. C. van den Bosch, T. C. Beers, Y. S. Lee, E. F. Bell, C. Rockosi, B. Yanny, H. Newberg, R. Wilhelm, X. Kang, M. C. Smith, and D. P. Schneider. The Milky Way's Circular Velocity Curve to 60 kpc and an Estimate of the Dark Matter Halo Mass from the Kinematics of 2400 SDSS Blue Horizontal-Branch Stars. Astrophysical Journal, 684:1143–1158, 9 2008.

- [389] Akın Yıldırım, Remco C. E. van den Bosch, Glenn van de Ven, Aaron Dutton, Ronald Läsker, Bernd Husemann, Jonelle L. Walsh, Karl Gebhardt, Kayhan Gültekin, and Ignacio Martín-Navarro. The massive dark halo of the compact early-type galaxy NGC 1281. Monthly Notices of the Royal Astronomical Society, 456(1):538–553, 2 2016.
- Donald G. York, J. Adelman, Jr. Anderson, John E., Scott F. Anderson, James Annis, 390 Neta A. Bahcall, J. A. Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, William N. Boroski, Steve Bracker, Charlie Briegel, John W. Briggs, J. Brinkmann, Robert Brunner, Scott Burles, Larry Carey, Michael A. Carr, Francisco J. Castander, Bing Chen, Patrick L. Colestock, A. J. Connolly, J. H. Crocker, István Csabai, Paul C. Czarapata, John Eric Davis, Mamoru Doi, Tom Dombeck, Daniel Eisenstein, Nancy Ellman, Brian R. Elms, Michael L. Evans, Xiaohui Fan, Glenn R. Federwitz, Larry Fiscelli, Scott Friedman, Joshua A. Frieman, Masataka Fukugita, Bruce Gillespie, James E. Gunn, Vijay K. Gurbani, Ernst de Haas, Merle Haldeman, Frederick H. Harris, J. Hayes, Timothy M. Heckman, G. S. Hennessy, Robert B. Hindsley, Scott Holm, Donald J. Holmgren, Chi-hao Huang, Charles Hull, Don Husby, Shin-Ichi Ichikawa, Takashi Ichikawa, Zeljko Ivezić, Stephen Kent, Rita S. J. Kim, E. Kinnev, Mark Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, John Korienek, Richard G. Kron, Peter Z. Kunszt, D. Q. Lamb, B. Lee, R. French Leger, Siriluk Limmongkol, Carl Lindenmeyer, Daniel C. Long, Craig Loomis, Jon Loveday, Rich Lucinio, Robert H. Lupton, Bryan MacKinnon, Edward J. Mannery, P. M. Mantsch, Bruce Margon, Peregrine McGehee, Timothy A. McKay, Avery Meiksin, Aronne Merelli, David G. Monet, Jeffrey A. Munn, Vijay K. Narayanan, Thomas Nash, Eric Neilsen, Rich Neswold, Heidi Jo Newberg, R. C. Nichol, Tom Nicinski, Mario Nonino, Norio Okada, Sadanori Okamura, Jeremiah P. Ostriker, Russell Owen, A. George Pauls, John Peoples, R. L. Peterson, Donald Petravick, Jeffrey R. Pier, Adrian Pope, Ruth Pordes, Angela Prosapio, Ron Rechenmacher, Thomas R. Quinn, Gordon T. Richards, Michael W. Richmond, Claudio H. Rivetta, Constance M. Rockosi, Kurt Ruthmansdorfer, Dale Sand ford, David J. Schlegel, Donald P. Schneider, Maki Sekiguchi, Gary Sergey, Kazuhiro Shimasaku, Walter A. Siegmund, Stephen Smee, J. Allyn Smith, S. Snedden, R. Stone, Chris Stoughton, Michael A. Strauss, Christopher Stubbs, Mark SubbaRao, Alexander S. Szalay, Istvan Szapudi, Gyula P. Szokoly, Anirudda R. Thakar, Christy Tremonti, Douglas L. Tucker, Alan Uomoto, Dan Vanden Berk, Michael S. Vogeley, Patrick Waddell, Shu-i. Wang, Masaru Watanabe, David H. Weinberg, Brian Yanny, Naoki Yasuda, and SDSS Collaboration. The Sloan Digital Sky Survey: Technical Summary. Astronomical Journal, 120(3):1579– 1587, 9 2000.
- [391] Jesús Zavala and Carlos S. Frenk. Dark Matter Haloes and Subhaloes. Galaxies, 7(4):81, 9 2019.
- [392] Andrew R. Zentner, Andreas A. Berlind, James S. Bullock, Andrey V. Kravtsov,

and Risa H. Wechsler. The Physics of Galaxy Clustering. I. A Model for Subhalo Populations. *Astrophysical Journal*, 624(2):505–525, 5 2005.

- [393] Andrew R. Zentner and James S. Bullock. Halo Substructure and the Power Spectrum. Astrophysical Journal, 598(1):49–72, 11 2003.
- [394] Andrew R. Zentner, Andrey V. Kravtsov, Oleg Y. Gnedin, and Anatoly A. Klypin. The Anisotropic Distribution of Galactic Satellites. Astrophysical Journal, 629(1):219– 232, 8 2005.
- [395] M. Zhai, X.-X. Xue, L. Zhang, C.-D. Li, G. Zhao, and C.-Q. Yang. The mass of the Galactic dark matter halo from ~9000 LAMOST DR5 K giants. *Research in Astronomy and Astrophysics*, 18:113, 8 2018.
- [396] D. H. Zhao, Y. P. Jing, H. J. Mo, and G. Börner. Mass and Redshift Dependence of Dark Halo Structure. Astrophysical Journal, Letters, 597(1):L9–L12, 11 2003.
- [397] Hongsheng Zhao. Analytical models for galactic nuclei. Monthly Notices of the Royal Astronomical Society, 278(2):488–496, 1 1996.
- [398] Zheng Zheng, Andreas A. Berlind, David H. Weinberg, Andrew J. Benson, Carlton M. Baugh, Shaun Cole, Romeel Davé, Carlos S. Frenk, Neal Katz, and Cedric G. Lacey. Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies. Astrophysical Journal, 633(2):791–809, 11 2005.
- [399] Shuang Zhou, Cheng Li, Cai-Na Hao, Rui Guo, Houjun Mo, and Xiaoyang Xia. Star formation histories of massive red spiral galaxies in the local universe. *arXiv e-prints*, page arXiv:2011.13749, 11 2020.
- [400] Qirong Zhu, Federico Marinacci, Moupiya Maji, Yuexing Li, Volker Springel, and Lars Hernquist. Baryonic impact on the dark matter distribution in Milky Waysized galaxies and their satellites. *Monthly Notices of the Royal Astronomical Society*, 458:1559–1580, 5 2016.
- [401] Adi Zolotov, Alyson M. Brooks, Beth Willman, Fabio Governato, Andrew Pontzen, Charlotte Christensen, Avishai Dekel, Tom Quinn, Sijing Shen, and James Wadsley. Baryons Matter: Why Luminous Satellite Galaxies have Reduced Central Masses. Astrophysical Journal, 761:71, 12 2012.

[402] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica* Acta, 6:110–127, 1 1933.